EECS 70 Discrete Mathematics and Probability Theory
Spring 2014 Anant Sahai

Probability review problems

1 Counting and basic probability

1. **Gotta seat them all.** There are 12 couples to be seated in the front row for an event. The row has exactly 24 seats, so it is guaranteed that each guest will have a seat, and that once all the guests have been seated, there will be no empty seat left.

   a) In how many ways can you seat the guests assuming that any guest cen be seated at any chair?

   b) In how many ways can you seat the guests assuming that all couples should be seated together?

   c) If you picked a random seating arrangement, what is the probability that all couples will be seated together?

2. **Arranging letters.** How many ways are there to arrange the letters of the word "SUPERMAN"

   a) On a straight line?

   b) On a straight line, such that "SUPER" occurs as a substring?

   c) On a straight line, such that "SUPER" occurs as a subsequence?

   d) On a circle?

   e) On a circle, such that "SUPER" occurs as a substring?

   f) On a circle, such that "SUPER" occurs as a subsequence?

3. **Starbucks.** At Starbucks, you can choose either a Tall, a Grande, or a Venti drink. Further, you can choose whether you want an extra shot of espresso or not. Furthermore, you can choose whether you want a Latte, a Cappuccino, an Americano, or a Frappucino.

   a) How many different kinds of drinks can you get (assuming drinks of different sizes are different) at Starbucks?

   b) Assume that 50% of drinks sold by Starbucks are Tall, 40% are Grande, and 10% are Venti. Assume that 10% of orders specify an extra shot of espresso. Further assume that all other choices (for example, Latte vs Cappuccino vs Americano vs Frappucino) are equally likely. Suppose you ordered a Tall Cappuccino with an extra shot, and someone else picked up your drink by mistake and left the store, so that you are forced to pick up his drink. What is the likelihood that this other person's drink is the same as yours, except for a possible difference in size?

   c) Using the probabilities from the previous part, what is the probability that two drinks sampled at random from the Starbucks pick up counter are the same, except for a possible size difference?

2 Conditional probability and Bayes' rule

1. **Which is the biased coin?** In this problem, we will use Bayesian inference to try and distinguish a biased coin from a fair coin.

   Suppose I have 2 coins. Let's call them A and B. Coin A is unbiased, *i.e.*, when you flip it, it comes up with a heads with probability exactly 0.5, or about half the time. Coin B is biased. When you flip it, it comes up with a heads with probablility 0.75, or about three fourths of the time.

   The problem is that outwardly, the two coins are identical and indistinguishable. I can't tell which is which just by looking at them, or weighing them, or subjecting them to any physical test. All I can do is flip the coins many times, and hope to tell them apart based on the outcomes of these flips.

   a) So I pick up one of these coins at random and flip it a 100 times, and I get 50 heads. Given the outcome above, what is the probability that the coin I picked is unbiased?

   b) Now I pick up the other coin and flip *it* a 100 times, and I get 75 heads. Given the two outcomes above (*i.e.*, the heads counts from both sets of 100 flips), what is the probability that the first coin that I picked is unbiased?

2. **Accounting fraud.** There are 3 ways a CEO can commit accounting fraud: (A) he can overstate revenue by booking fake orders, (B) he can overstate profits by recording near-term expenses as long-term capital expenses, and (C) he can understate liabilities by making overly optimistic assumptions about the growth prospects of the company's pension funds. In the United States, 50% of CEOs commit fraud (A), 20% of CEOs commit fraud (B), 15% of CEOs commit both fraud (A) and fraud (B), and all CEOs commit fraud (C).

   The probability of getting caught if a CEO does (A) is 0.01, independent of whether or not he commits other kinds of fraud. Similarly, the probability of getting caught if he does (B) is 0.05, and of getting caught if he does (C) is 0.001.

   a) What are the chances (rounded to 2 decimal places) that a CEO who perpetrates all 3 kinds of fraud will get caught?

   b) You read on the front page of the Wall Street Journal that the CEO of TrustMe, Inc. (for all practical purposes, a randomly chosen CEO) has been arrested for accounting fraud. What are the chances (rounded to 2 decimal places) that this CEO perpetrated at least 2 kinds of fraud?

   c) What are the chances (rounded to 2 decimal places) that the above CEO who was caught perpetrated exactly 2 kinds of fraud?

3. **Randomly chosen program.** I have 3 computer programs, P_1, P_2, and P_3. The first program (P_1) is just an infinite loop that prints the digit 0 over and over again. P_2 is somewhat more interesting. It is also an infinite loop, but it prints the digits 00001 over and over again. The program P_3 is the most interesting: it first chooses a random integer $N \geq 1$ drawn from a geometric distribution with $p = 0.01$, and then it prints N 1s, and then prints the digit 0 over and over again.

   I choose one of the above 3 programs at random (each with equal likelihood) and run it. I look at the 10^{th} and the 15^{th} digits produced as output by the chosen program. What are the chances that both these digits are 1?

3 Hashing, Load balancing, and Coupon collecting

1. **Car collection.** There are 4 different kinds of Hot wheels cars, and your five year old cousin wants to collect them all. He already has 3 of these cars. On his sixth birthday, he gets a gift from each guest in his birthday party. Two of the guests know how much he likes cars, and also know exactly which cars he has, so they will get him a random Hot wheels car that he doesn't have. Three of the guests know that he likes cars, but don't know exactly which cars he has, so they will get him a random Hot Wheels car, but not necessarily one that he doesn't have. One guest has no clue as to what he likes, so this guest just picks a toy at random from the toy store, which has 15 different toys in addition to the 5 Hot wheels cars. What is the probability that at the end of the party, the kid has all the 5 cars?

2. **Processes, servers and overloading.** I have M processes (jobs) and N servers that I can assign the jobs to. Any job may be assigned to any server. Suppose I assign each job to a randomly chosen server, with all servers being equally likely. We say that a server is overloaded if it is assigned greater than or equal to K jobs, where $K \leq M$. What is the probability that the first server is overloaded?

3. **Jobs and servers, but without immediate repetition.** I have $M \geq 2$ jobs and $N \geq M$ servers to assign the jobs to.

   I use the following system to assign the jobs: the first job is randomly assigned to one of the N servers (with all servers being equally likely). The second job is again assigned randomly to a server, except that the server that got the first job is excluded from the selection (any of the other $N-1$ servers are equally likely to get the job). Similarly, the third job is assigned randomly, except this time, the server that got the second job is excluded (any of the other $N-1$ servers have equal likelihood of being chosen for this job). And so on until all M jobs have been assigned.

   What is the probability that each of the M jobs goes to a different server?

4. **Bounding overload probabilities.** I have a load balancing set up with N servers, where I can guarantee that the probability of the i^{th} server being overloaded (where $1 \leq i \leq N$) is at most p, which is independent of i. Which of the following is true? Check all that apply.

   - Pr (no server is overloaded) $\leq 1 - Np$
   - Pr (no server is overloaded) $\geq 1 - Np$
   - Pr (at least one server is overloaded) $\geq p$
   - Pr (no server is overloaded) $\leq 1 - N(1-p)$
   - Pr (no server is overloaded) $\geq 1 - (1-p)^{2N}$
   - Pr (at least one server is overloaded) $\leq Np$

4 Probability distributions, expectation, and variance

1. **Simple binomial distribution.** Suppose I toss a biased coin (that comes up heads with a probability of 0.8) 5 times. Calculate the probability distribution of the number of heads. Calculate the mean and variance of this distribution.

2. **Expectation and variance with scaling.** Suppose we have a random variable X, whose expectation is μ and whose variance is σ^2. What is the expectation and variance of the random variable $Y = kX$, where k is a constant?

3. **Expectation and variance with adding.** Suppose we have a random variable X, whose expectation is μ and whose variance is σ^2. Let us define k independent and identically distributed random variables X_i, $1 \le i \le k$, where each X_i has the same distribution as that of X. What is the expectation and variance of the random variable $Z = \sum_{i=1}^{k} X_i$. How does this compare to the expectation and variance of $Y = kX$ that you calculated in the previous problem?

4. **Drunk man.** We have a drunk man walking along a straight road. At time $t = 0$, the man is in position $x = 0$. At each time point, he moves to the right (positive x direction) with probability 0.6, to the left with probability 0.3, and stays exactly where he is with probability 0.1. Calculate the following:

   a) The probability distribution of the man's position at times $t = 0$, $t = 1$, $t = 2$, and $t = 3$.
   b) The expected value of the man's position at times $t = 0$, $t = 1$, $t = 2$, and $t = 3$.
   c) The variances of the man's position at times $t = 0$, $t = 1$, $t = 2$, and $t = 3$.

5. **An algorithm for generating prime numbers.** A number of applications, such as RSA, require the generation of large primes. Let's say I propose the following algorithm to generate an N-bit prime number, where $N > 10$:

```
def genPrime (N):
    n = 2
    while n is not prime:
        n = N-bit number picked uniformly at random
    return n
```

   Let $\pi(x)$ be the number of primes that are less than or equal to x. A famous result from mathematics states that:

$$\pi(x) > \frac{x}{\ln x}, \text{ for } x \ge 17$$

   a) Assuming that N is sufficiently large, what can you say about the probability distribution of the number of times the `while` loop of the above algorithm runs? Among the distributions that you have studied in this course, which distribution most closely resembles the number of runs of the `while` loop?
   b) Find an upper bound for the expected number of `while` loop runs before a prime is generated by the above algorithm.

6. **Horse race.** There are 4 horses in a race. Their probabilities of winning the race are 0.65, 0.25, and 0.1. The odds that the house gives you are $0.1/1$, $3/1$, and $12/1$ respectively (meaning, if you bet on the first horse and it wins, you will get \$0.10 in profit for every dollar you invest, and if the horse loses, you will lose your entire money). Assuming you have \$100 with you, and you want to maximize your expected profit, how should you divide your money amongst the horses (you also have the option of keeping all or part of the money without betting on any horse)? And if you allocate the money this way, what will be the variance of your net profit?

7. **Coin flips, parity, and independence.** Let X_i, $i \ge 1$ be a sequence of independent fair coin flips, where the random variable $X_i = 1$ if the i^{th} coin flip produces a heads, and $X_i = 0$ otherwise. Let

$$Y_i = \sum_{j=1}^{i} X_i \pmod 2, \text{ and}$$

$$Z_i = \sum_{j=1}^{i} X_i \pmod 3$$

a) Are the Y_is pairwise independent?

b) Are the Z_is pairwise independent?

c) Find $E[Y_3]$ and $\sigma^2_{Y_3}$.

d) Find $E[Z_3]$ and $\sigma^2_{Z_3}$.

e) Find $E[Y_1Y_3]$ and $\sigma^2_{Y_1Y_3}$.

f) Find $E[Z_1Z_3]$ and $\sigma^2_{Z_1Z_3}$.

8. **Extending the geometric distribution for multiple successes.** In class, you saw that the geometric distribution with parameter p can be interpreted as the "number of tries upto and including your first success", given that the probability of success is p at each try. But what if we change the termination condition? As in the regular geometric distribution, let us say that the probability of success in any one trial is $p = 0.5$, and it is independent of other trials. However, unlike the traditional geometric distribution (which terminates at the first success) let us say our new termination criterion is that we need to succeed at least two consecutive times. Under this new criterion, let Y be the "number of tries upto and including the first consecutive pair of successes". Calculate the following:

   a) The number of arrangements of the letters {`H, T`}, of length n, such that the last letter is a `H`, and with no consecutive pair of `H`s. ***Hint:*** *do this part and the next together, and try to express the number of arrangements as a recurrence relation. Solve the recurrence relation. Does the solution remind you of a famous mathematical sequence?*

   b) The number of arrangements of the letters {`H, T`}, of length n, such that the last letter is a `T`, and with no consecutive pair of `H`s.

   c) The distribution of Y.

   d) The expected value of Y.

   e) **Bonus. This is a hard problem.** The variance of Y.