Stat 200A, Fall 2018
A. Adhikari and Z. Bartha

WEEK 1 SOLUTIONS

You are expected to do all these problems, but for **Homework 1** please turn in **only Problems 2, 3, and 7** on **Thursday August 30 at the start of lecture**.

1. Applying the Rules

(a) I choose one of the 12 months at random. Independently, my friend chooses one of the 12 months at random. Which of the following is the chance that we choose the same month?

$$(i)\ \frac{1}{12}\cdot\frac{1}{11} \qquad (ii)\ \frac{1}{12}\cdot\frac{1}{12} \qquad (iii)\ \frac{1}{12} \qquad (iv)\ \frac{1}{11}$$

(iii) since

$$\begin{aligned}&\mathbb{P}(\text{we choose the same month})\\ &\quad = \mathbb{P}(\text{we both choose January}) + \ldots + \mathbb{P}(\text{we both choose December})\\ &\quad = \frac{1}{12}\cdot\frac{1}{12} + \ldots + \frac{1}{12}\cdot\frac{1}{12} = \frac{1}{12}.\end{aligned}$$

(b) A coin is tossed 5 times. What is the chance of getting at least one head?

$\mathbb{P}(\text{at least one head}) = 1 - \mathbb{P}(\text{all tails}) = 1 - \left(\frac{1}{2}\right)^5 = \frac{31}{32}$.

(c) A coin is tossed 5 times. What is the chance that both faces appear at least once?

$\mathbb{P}(\text{both faces at least once}) = 1 - \mathbb{P}(\text{all heads}) - \mathbb{P}(\text{all tails}) = 1 - \left(\frac{1}{2}\right)^5 - \left(\frac{1}{2}\right)^5 = \frac{15}{16}$.

(d) A coin that lands heads with probability p is tossed repeatedly. What is the chance that k tosses are needed to get the first head?

$\mathbb{P}(\text{first head is on toss } k) = \mathbb{P}(\text{all tails for the first } k-1 \text{ tosses}) \cdot \mathbb{P}(\text{head for the } k\text{th toss}) = (1-p)^{k-1}p$.

(e) A coin that lands heads with probability p is tossed repeatedly. What is the chance that more than k tosses are needed to get the first head?

$\mathbb{P}(\text{the first } k \text{ tosses are all tails}) = (1-p)^k$.

2. Exact Calculations and Approximations

(a) A coin that lands heads with chance $1/N$ is tossed n times. Find the chance that there are no heads. Find a simple exponential approximation to this chance assuming that $N \to \infty$ and $n \to \infty$ in such a way that $n/N \to \mu$ for some positive number μ.

$\mathbb{P}(\text{all the } n \text{ tosses are tails}) = \left(1 - \frac{1}{N}\right)^n \to e^{-\mu}$, since

$$\log\left(1 - \frac{1}{N}\right)^n = n\log\left(1 - \frac{1}{N}\right) \sim n\left(-\frac{1}{N}\right) \to -\mu.$$

(b) There is a set of n members, from whom I take a sample of n members at random with replacement. In Stat 201B you will see that this is one of the two main aspects of randomization in the bootstrap method.

What is the chance that my sample contains all n members of the set? Guess the limit of this chance as $n \to \infty$.

$$\frac{\text{\# of orders in which we can pick all } n \text{ members}}{\text{\# of all possible sequences}} = \frac{n!}{n^n} \to 0$$

as $n \to \infty$, for example because $\frac{n!}{n^n} \le \frac{1}{n} \cdot \frac{2}{n} \cdot \ldots \cdot \frac{\lfloor n/2 \rfloor}{n} \le \left(\frac{1}{2}\right)^{\lfloor n/2 \rfloor} \to 0$.

(c) Stirling's formula is an approximation for factorials. It says that for large integers n,

$$n! \quad \sim \quad \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

The symbol $\sim$ stands for "asymptotically equivalent", which means that the ratio of the two sides goes to 1 as n tends to ∞. Use Stirling's formula to approximate the chance in Part (b) when n is large.

$$\frac{n!}{n^n} \sim \frac{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n}{n^n} = \frac{\sqrt{2\pi n}}{e^n}.$$

(d) A coin is tossed $2n$ times. Find the chance of getting exactly n heads using the method in the hint below. Approximate the chance using Stirling's formula and find its limit as n gets large. Later in the course we will see why the answer does not contradict the law of averages.

[Hint: Count the number of all possible sequences of $2n$ heads and tails, and then count the number that contain exactly n heads.]

$$\begin{aligned}
\mathbb{P}(\text{exactly } n \text{ heads}) &= \frac{\text{\# of sequences with } n \text{ heads and } n \text{ tails}}{\text{\# of sequences of length } 2n} \\
&= \frac{\binom{2n}{n}}{2^{2n}} = \frac{(2n)!}{n! \cdot n! \cdot 2^{2n}} \\
&\sim \frac{\sqrt{2\pi \cdot 2n}\left(\frac{2n}{e}\right)^{2n}}{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n \sqrt{2\pi n}\left(\frac{n}{e}\right)^n 2^{2n}} = \frac{1}{\sqrt{\pi n}}.
\end{aligned}$$

3. Bounds

Even if you don't know a probability exactly, it is sometimes enough to know a range in which it must lie. Let's start with a warm-up.

(a) In a class, 60% of the students have read *The Merchant of Venice* and 10% have read *King Lear*. Fill in the four blanks below.

- The percent of students who have read both of those works is at least 0% and at most 10%.
- The percent of students who have read at least one of those works is at least 60% and at most 70%.

(b) Let $A_1, A_2, \ldots, A_n$ be events. Use induction to prove *Boole's Inequality*:

$$P(\bigcup_{i=1}^{n} A_i) \ \leq \ \sum_{i=1}^{n} P(A_i)$$

The statement is trivial for $n = 1$. For $n = 2$ we have

$$\begin{aligned}\mathbb{P}(A_1 \cup A_2) &= \mathbb{P}(A_1 \setminus A_2) + \mathbb{P}(A_1 \cap A_2) + \mathbb{P}(A_2 \setminus A_1)\\ &\leq \mathbb{P}(A_1 \setminus A_2) + 2\mathbb{P}(A_1 \cap A_2) + \mathbb{P}(A_2 \setminus A_1)\\ &= \mathbb{P}(A_1) + \mathbb{P}(A_2).\end{aligned}$$

Now let's assume that it is true for $n = k$ and prove it for $n = k + 1$:

$$\mathbb{P}\left(\bigcup_{i=1}^{k+1} A_i\right) = \mathbb{P}\left(\left(\bigcup_{i=1}^{k} A_i\right) \bigcup A_{k+1}\right) \leq \mathbb{P}\left(\bigcup_{i=1}^{k} A_i\right) + \mathbb{P}(A_{k+1}) \leq \sum_{i=1}^{k} \mathbb{P}(A_i) + \mathbb{P}(A_{k+1}) = \sum_{i=1}^{k+1} \mathbb{P}(A_i)$$

by using the $n = 2$ and the $n = k$ cases.

(c) A data scientist can estimate four parameters based on the same random sample. For each estimate, the chance that it is good is 95%. Find a lower bound on the chance that all four estimates are good. Why is it not OK to say that the chance has to be $(0.95)^4$?

$$\begin{aligned}\mathbb{P}(\text{all four are good}) &= 1 - \mathbb{P}(\text{at least one is bad})\\ &= 1 - \mathbb{P}\left(\bigcup_{i=1}^{4} \{\text{the } i\text{th is bad}\}\right)\\ &\geq 1 - \sum_{i=1}^{4} \mathbb{P}(\text{the } i\text{th is bad})\\ &= 1 - 4 \cdot 0.05 = 0.8.\end{aligned}$$

We can't simply multiply the probabilities since we don't know whether the different estimates are independently good or not.

4. Symmetry and Simple Random Samples

Randomly shuffle a standard deck of 52 cards. That means all shuffles are equally likely. Formally, you're working with the uniform distribution on all permutations of the 52 cards.

A standard deck contains four *suits*, called hearts, diamonds, spades, and clubs. The first two suits are red and the others are black. Each suit consists of 13 cards, labeled Ace, 2, 3, 4, ..., 10, Jack, Queen, King. These labels are called *ranks*.

(a) How many shuffles are there? In how many shuffles is the 17th card red? What is the chance that the 17th card is red? Does the number 17 appear in your answer?

There are 52! shuffles, the 17th card is red in half of them, 52!/2, by symmetry, so the chance of this happening is 1/2. By symmetry, the answer is the same for all positions.

(b) I deal a poker hand. That's 5 cards dealt at random without replacement from the deck. What is the chance that the last two cards that I deal are aces? How would your answer have been affected had I dealt a bridge hand (13 cards) instead of a poker hand?

$\binom{4}{2}/\binom{52}{2}$ in point cases: by symmetry it doesn't matter how many cards I am dealing, the probability is the same as if I just looked at the top two cards of the deck.

(c) I shuffle a deck and deal all the cards one by one. What is the chance that the ace of spades appears before all the red cards?

1/27, since we can restrict our attention just to the order in which the 16 red cards and the ace of spades appear, and all the orders are equally likely. Hence, the probability of any specific card of these 27 coming first is the same, 1/27.

(d) A *simple random sample* is a sample drawn uniformly at random without replacement from a finite population. Suppose a population consists of N individuals of whom G are "good" and $B = N - G$ are "bad". I draw a simple random sample of n individuals (you can assume $n \leq N$). Let A_i be the event that the ith individual drawn is good. For each i in the range 1 through n, find $P(A_i)$. This result will be used repeatedly in the course.

$\mathbb{P}(A_i) = \mathbb{P}(A_1)$ for any i by symmetry: we can just reorder our sample so that the originally ith pick comes first. And $\mathbb{P}(A_1) = G/N$.

5. Inclusion-Exclusion Let A, B, and C be events on an outcome space.

(a) Use the additivity axiom to show that

$$P(A \cup B) \;=\; P(A) + P(B) - P(AB)$$

This is the first *inclusion-exclusion* formula for the chance of a union. You include both the individual events and exclude the intersection.

$$\begin{aligned}\mathbb{P}(A) + \mathbb{P}(B) &= [\mathbb{P}(A \setminus B) + \mathbb{P}(AB)] + [\mathbb{P}(B \setminus A) + \mathbb{P}(AB)] \\ &= [\mathbb{P}(A \setminus B) + \mathbb{P}(B \setminus A) + \mathbb{P}(AB)] + \mathbb{P}(AB) \\ &= \mathbb{P}(A \cup B) + \mathbb{P}(AB)\end{aligned}$$

(b) Prove that

$$P(A \cup B \cup C) \;=\; P(A) + P(B) + P(C) - P(AB) - P(BC) - P(CA) + P(ABC)$$

Please don't just argue "by picture". Start by setting $D = A \cup B$, and use Part (a) twice.

$$
\begin{aligned}
\mathbb{P}(A \cup B \cup C) &= \mathbb{P}(D \cup C) \\
&= \mathbb{P}(D) + \mathbb{P}(C) - \mathbb{P}(DC) \\
&= \mathbb{P}(A \cup B) + \mathbb{P}(C) - \mathbb{P}((A \cup B) \cap C) \\
&= \mathbb{P}(A \cup B) + \mathbb{P}(C) - \mathbb{P}(AC \cup BC) \\
&= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB) + \mathbb{P}(C) - [\mathbb{P}(AC) + \mathbb{P}(BC) - \mathbb{P}(ABC)] \\
&= \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(AB) - \mathbb{P}(AC) - \mathbb{P}(BC) + \mathbb{P}(ABC)
\end{aligned}
$$

(c) The result of Part (b) can be extended to n events $A_1, A_2, \ldots, A_n$ as follows:

$$
\begin{aligned}
&P(\bigcup_{i=1}^{n} A_i) \\
&= \sum_{i=1}^{n} P(A_i) - \sum\sum_{1 \le i < j \le n} P(A_i A_j) + \sum\sum\sum_{1 \le i < j < k \le n} P(A_i A_j A_k) - \cdots + (-1)^{n+1} P(A_1 A_2 \ldots A_n)
\end{aligned}
$$

You could prove it by induction using the method of Part (b), but you don't have to do that here. We'll prove it later using a less cumbersome process. For now, just count the number of terms in each sum on the right hand side. How many terms are there in the first sum? How many in the second? How many in the third?

The number of terms in the mth sum is $\binom{n}{m}$.

(d) A die is rolled n times. What is the chance that all six faces appear?

Let A_i be the event that face i does not appear. Then

$$
\begin{aligned}
\mathbb{P}(\text{all six faces appear}) &= 1 - \mathbb{P}(\text{at least one face does not appear}) \\
&= 1 - \mathbb{P}(A_1 \cup A_2 \cup \ldots \cup A_6) \\
&= 1 - \sum_{i=1}^{6} \mathbb{P}(A_i) + \sum_{1 \le i < j \le 6} \mathbb{P}(A_i A_j) - \sum_{1 \le i < j < k \le 6} \mathbb{P}(A_i A_j A_k) + \ldots + \mathbb{P}(A_1 A_2 \ldots A_6) \\
&= 1 - 6 \cdot \left(\frac{5}{6}\right)^n + \binom{6}{2} \cdot \left(\frac{4}{6}\right)^n - \binom{6}{3} \cdot \left(\frac{3}{6}\right)^n + \binom{6}{4} \left(\frac{2}{6}\right)^n - \binom{6}{5} \cdot \left(\frac{1}{6}\right)^n + 0.
\end{aligned}
$$

6. Fixed Points of a Random Permutation

A *fixed point* of a permutation is an element that remains in its original place after the permutation. In this exercise you will examine the distribution of the number of fixed points in a random permutation.

Consider the outcome space consisting of all $n!$ permutations of the integers 1 through n. A *random permutation* is the result of one draw made uniformly at random from this outcome space.

This abstract setting is usually more colorfully described, for example in terms of letters and envelopes. There are n envelopes labeled 1 through n, and n letters labeled 1 through n. The letters are permuted randomly into the envelopes, one letter per envelope. That is, the letters are placed one by one into envelopes picked uniformly at random without replacement.

Suppose you have a random permutation of the integers 1 through n. In what follows, i and j are integers between 1 and n.

(a) Find the chance that i is a fixed point of the permutation. That is, find the chance that letter i falls in envelope i.

(b) Find the chance that both i and j are fixed points.

(c) A *derangement* is a permutation that has no fixed points: none of the letters falls in the matching envelope. Find the chance of a derangement.

[Hint: Use the complement and inclusion-exclusion.]

(d) For $1 \le k \le n$, let $p_{k,n}$ be the chance that there are exactly k fixed points in a random permutation of n letters. So your answer to Part (c) is $p_{0,n}$. Show that

$$p_{k,n} \;=\; \frac{1}{k!} p_{0,n-k}$$

(e) What should the value of $\sum_{k=0}^{n} p_{k,n}$ be? Why? No calculations needed.

(f) What you have found in Part (d) is called the *distribution of the number of fixed points*. This is a probability distribution on a finite set, namely, the values $0, 1, \ldots, n$, but it is better understood in the limit as n gets large. For each fixed $k \ge 0$, let $p_k = \lim_{n\to\infty} p_{k,n}$. Find p_k and hence find $\sum_{k=0}^{\infty} p_k$.

See the PROB 140 textbook: `http://prob140.org/textbook/chapters/Chapter_05/03_The_Matching_Problem`.

7. Using Recursion

A coin lands heads with probability p. My friend and I take turns tossing the coin, with my friend tossing first. Whoever gets the first head wins. Find the chance that my friend wins, in two ways:

(a) For each i, find the chance that my friend wins on toss i, and sum.

My friend wins on toss i (i odd) if and only if in all the previous rounds both of us toss tails and on the ith toss my friends tosses head. The chance of this is $(1-p)^{i-1}p$. Summing this up and substituting $i = 2j+1$:

$$\mathbb{P}(\text{my friend wins}) = \sum_{i>0 \text{ odd}} (1-p)^{i-1}p = \sum_{j=0}^{\infty}(1-p)^{2j}p = p\sum_{j=0}^{\infty}\left[(1-p)^2\right]^j = \frac{p}{1-(1-p)^2} = \frac{1}{2-p}.$$

(b) Let x be the chance that you are trying to find. Set up an equation for x, using the observation that either my friend wins on the first toss, or we both get tails and then the game starts over. Solve for x and confirm that it's the same as what you got in Part (a). This is an example of how arranging probability calculations in the right way can reduce the amount of algebra you have to do.

$$\begin{aligned} x &= \mathbb{P}(\text{my friend wins}) \\ &= \mathbb{P}(\text{my friend wins on the first toss}) + \mathbb{P}(\text{we both toss tails and my friend wins later}) \\ &= p + (1-p)^2 x, \end{aligned}$$

hence, $x = p + (1-p)^2 x$ which has the solution $x = \frac{p}{1-(1-p)^2} = \frac{1}{2-p}$.

(c) Should the answer to Parts (a) and (b) be bigger than 1/2, equal to 1/2, or less than 1/2? Give an intuitive justification of your choice and then prove that your choice is correct. Yes, you can assume that $0 < p < 1$.

Intuitively, $x > 1/2$ since my friend starts the game, otherwise the game is symmetric between us. More precisely, in each round if only one of us tosses a head, that person wins – this gives us equal chance – but if both of us toss heads, then my friend wins. Looking at the formula that we got also proves this: if $0 < p < 1$, then $1/2 < 1/(2-p) < 1$.

8. Why Study Balls in Boxes?

The setting of "balls in boxes" is standard in probability, and students tend to think that it's artificial. In fact it provides a general way of looking at trials where the outcomes fall into natural categories. By varying the probabilistic assumptions, you get varying distributions of the number of outcomes in a category.

For now we'll just look at trials that result in two categories, which you can think of heads and tails, or successes and failures, or zeros and ones. We'll call them Category 0 and Category 1.

This can be modeled by "b balls in $B = 2$ boxes," as follows:

Each ball is a trial. Each trial is labeled "0" or "1" depending on the category in which it results. That's it. There are no probabilistic assumptions yet.

As an example, suppose you have $b = 5$ balls and $B = 2$ boxes. Then, if you observe the category (box) in which each trial (ball) results, you might get a sequence like 11010. You could also step back from this level of detail and simply note the count in each category. We'll use $(2, 3)$ to mean "2 trials resulted in Category 0, and 3 trials in Category 1."

(a) Suppose you observe just the count in each category, that is, you observe (b_0, b_1) where integers b_0 and b_1 are non-negative with $b_0 + b_1 = b$. How many possible pairs (b_0, b_1) are there?

There are $b + 1$ possible pairs: $(0, b), (1, b-1), \ldots, (b, 0)$.

(b) In statistical mechanics, the *Bose-Einstein model* assumes all of these pairs to be equally likely. Under this model, what is the chance that the count in Category 1 is equal to b_1, for b_1 in the range 0 through b?

$1/(b+1)$

(c) Suppose you observe the box into which each ball falls; you'll get a sequence of length b, each of whose elements is either 0 or 1. How many such sequences are there? If you fix a pair of counts in each category, say (b_0, b_1) with $b_0 + b_1 = b$, how many sequences of length b result in this pair of counts?

Since each of the b elements of the sequence can be either 0 or 1, the number of possible sequences is 2^b. If we fix the number of 1's to be b_1, there are $\binom{b}{b_1}$ ways to pick their positions, the other elements will be 0. Hence, there are $\binom{b}{b_1}$ sequences with b_1 1's.

(d) The *Maxwell-Boltzmann model* assumes all sequences of length b are equally likely. Under this model, what is the chance that the count in Category 1 is equal to b_1, for b_1 in the range 0 through b?

Based on part (c), $\frac{\#\text{ of sequences with } b_1 \text{ 1's}}{\text{total } \# \text{ of sequences}} = \frac{\binom{b}{b_1}}{2^b}$.