A Brief Compilation

of Guides, Walkthroughs, and Problems

Discrete Mathematics and Probability Theory

at the University of California, Berkeley

Alvin Wan

Contents

0.1 Purpose

This compilation is (unofficially) written for the Spring 2016 CS70: Discrete Mathematics and Probability Theory class taught by Professor Satish Rao and Professor Jean Walrand at UC Berkeley. It's primary purpose is to offer additional practice problems and walkthroughs to build intuition, as a supplement to official course notes and lecture slides. Including more difficult problems in walkthroughs, there are over 35 exam-level problems.

0.1.1 Contributors

A Special Thanks to Sinho Chewi for spending many hours suggesting improvements, catching bugs, and discussing ideas and solutions for problems with me. Additionally, thanks to Dibya Ghosh and Blake Tickell, who helped review problems for clarity and correctness.

0.1.2 Structure

Each chapter is structured so that this book can be read on its own. A minimal guide at the beginning of each section covers essential materials and misconceptions but does not provide a comprehensive overview. Each guide is then followed by walkthroughs covering classes of difficult problems and 3-5 exam-level (or harder) problems that I've written specifically for this book. *Note: As of Spring 2016, not all chapters have problems. However, all chapters have at least a walkthrough. This will be amended in Fall 2016.*

0.1.3 Breakdown

For the most part, guides are "cheat sheet"s for select chapters from official course notes, with additional comments to help build intuition.

For more difficult parts of the course, guides may be accompanied by breakdowns and analyses of problem types that might not have been explicitly introduced in the course. These additional walkthroughs will attempt to provide a more regimented approach to solving complex problems.

Problems are divvied up into two parts: (1) walkthroughs - a string of problems that "evolve" from the most basic to the most complex - and (2) exam-level questions, erring on the side of difficulty where needed. The hope is that with walkthroughs, students can reduce a relatively difficult problem into smaller, simpler subproblems.

0.1.4 Resources

Additional resources, including 20+ quizzes with 80 practice questions, and other random worksheets and problems are posted online at `alvinwan.com/cs70`.

Chapter 1

Modular Arithmetic and Polynomials

1.1 Guide

1.1.1 Modular Arithmetic

In modulo p, only the numbers $\{0, 1, ..., p-1\}$ exist. Additionally, division is not well-defined. Instead, we define a *multiplicative inverse*. We know that outside a modulo field, for any number n, an inverse n^{-1} multiplied by itself is 1. $(n \cdot n^{-1} = 1)$ Thus, we extend the definition of an inverse to the modulo field in this manner, where for any number n,

$$n \cdot n^{-1} = 1 \pmod{p}$$

Do not forget that division, and thus, *fractions* do not exist in a modulo.

1.1.2 Polynomial Properties

For polynomials, we have two critical properties.

1. A polynomial of degree d has at most d roots.
2. A polynomial of degree d is uniquely identified by $d+1$ distinct points.

Note that taking a polynomial over a Galois Field of modulo p (denoted $GF(p)$) simply means that all operations and elements in that field are (mod p).

1.1.3 Fermat's Little Theorem

Fermat's Little Theorem states that if p is prime, for any a, $a^p = a \pmod{p}$. If p does not divide a, we additionally know the following.

$$a^{p-1} = 1 \pmod p$$

Applying Fermat's Little Theorem repeatedly until the exponent of a is less than $p-1$ gives us an interesting corollary: $a^y = a^{y \bmod p-1} \pmod p$

1.1.4 Lagrange Interpolation

For a given set of points, compute first the Δ for each coordinate, where

$$\Delta_i = \Pi_{i \neq j} \frac{(x - x_j)}{(x_i - x_j)}$$

Then, to recover the original polynomial, multiply all Δ_i by the respective y_is.

$$P(x) = \sum_i \Delta_i y_i$$

1.1.5 Error-Correcting Codes

Across a lossy channel, where at most k packets are lost, send $n + k$ packets. Across a corruption channel, where at most k packets are corrupted, send $n+2k$ packets. To recover a P across a corruption channel, apply Berkelamp-Welsh.

1.1.6 RSA

In RSA, we have a public key (N, e), where N is a product of two primes, p and q, and e is co-prime to $(p-1)(q-1)$. Here are ENCRYPT and DECRYPT.

$$E(x) = x^e \pmod N$$
$$D(y) = y^d = x \pmod N$$

Why are they defined this way? We have that $y = E(x)$, so we plug in:

$$D(E(x)) = E(x)^d = (x^e)^d = x^{ed} = x^1 \pmod N$$

If the above equation $x^{ed} = x$ is satisfied, then $D(y)$ returns the original message. How do we generate d? By Fermat's Little Theorem's corollary, we know $ed = 1 \pmod{(p-1)(q-1)}$. Given we have e, we see that we can compute d if and only if we know p and q. Thus, breaking RSA equates factorizing N into p, q.

1.1.7 Secret Sharing

In a secret-sharing problem, our goal is to create a secret such that only a group meeting specific requirements can uncover it. We will explore secret sharing problems in *3.2 Secret-Sharing Walkthrough.*

1.2 Secret-Sharing Walkthrough

We begin with the most elementary form of secret-sharing, which requires consensus from some subset of k people before the secret can be revealed.

Question: Basic

Construct a scheme that requires that at least k of n people to come together, in order to unlock the safe.

Answer Polynomial of degree $k-1$.

We need at least k people, meaning this polynomial should require k points. Thus, we create a polynomial of degree $k-1$, and distribute n distinct points along this polynomial to n people. The secret is this polynomial evaluated at 0.

Question: Combining Polynomials

Develop a scheme that requires x_1 people from group A and x_2 people from B.

Answer polynomials of $x_1 - 1$, $x_2 - 1$ degrees, and a 1-degree polynomial using $P_1(0)$, $P_2(0)$

Create a polynomial p_1 of degree $x_1 - 1$ for A and a second polynomial p_2 of degree $x_2 - 1$ for B. Use the secrets of p_1 and p_2 ($p_1(0)$ and $p_2(0)$) to create a third polynomial p_3 of degree 1. The secret is $p_3(0)$.

Question: Combining Polynomials Generalized

Construct a scheme that requires x_i from each of the n groups of people.

Answer n P_i's of $x_i - 1$ degree, and 1 $n-1$-degree polynomial using $P_i(0)$ for all i.

Create n polynomials with degree $x_i - 1$ for the ith group. Use the secrets $(\forall i, p_i(0))$ of all n polynomials to create an $n+1$th polynomial of degree $n-1$. The root of this $n+1$th polynomial is the secret.

Question: Re-weighting

Each group elects o_i officials. Construct a scheme that requires $a_i \leq o_i$ officials from each group, where 10 citizens can replace an official.

Answer $10a_i - 1$ polynomials, where each official gets 10 points and each citizen gets 1

Create a polynomial of degree $10a_i - 1$, and give each of the a_i officials 10 points each. Then, give each citizen 1 point each. Use the secrets $(\forall i, p_i(0))$ of all n polynomials to create an $n+1$th polynomial of degree $n-1$. Since each official has 10 times the number of packets for the same polynomial, any 10 citizens can "merge" to become a single official.

Question: Erasure Channel

Construct a scheme that requires k workers to unlock the safe. Make sure to specify, if at most m workers do not respond to requests, how many workers we need to ensure we can unlock the safe?

Answer $k+m$, one $k-1$-degree polynomial

The intuitive response to simply request the number of people that may not respond in addition to the number of people we need to unlock the safe. Thus, we need to request $k+m$ workers to reconstruct a degree $k-1$-degree polynomial.

Question: Corruption Channel

Construct a scheme that requires k workers to unlock the safe. Make sure to specify, if at most m workers *mis-remember information*, how many people we need to ensure we can unlock the safe.

Answer $k + 2m$, one $k - 1$-degree polynomial

Per our knowledge of corruption channels, we need to $n + 2k$ packets, or in this case, $k + 2m$ workers, where again, we construct a degree $k - 1$ polynomial.

Chapter 2

Counting

2.1 Guide

Counting bridges discrete mathematics and probability theory, to some degree providing a transition from one to the next. Albeit a seemingly trivial topic, this section provides foundations for probability.

2.1.1 Fundamental Properties

We have two counting properties, as follows:

1. If, for each of k items, we have $\{n_1, n_2, ..., n_k\}$ options, the total number of possible combinations is $n_1 \cdot n_2 \cdots n_k = \Pi_i n_i$.
2. To find the total number of un-ordered combinations, divide the number of ordered combinations by the number of orderings. $\binom{n}{k} = \frac{n!}{(n-k)!k!}$

2.1.2 Stars and Bars

Given n balls and k bins, count all the ways to distribute n balls among k bins. Given a list of n balls, we need to slice this list in $k-1$ places to get k partitions.

In other words, given all of our *stars* (balls, n) and bars ("slices," $k-1$), we have $n+k-1$ total items. We can either choose to place our $k-1$ bars or our n stars. Thus, the total number of ways to distribute n balls among k bins:

$$\binom{n+k-1}{n} = \binom{n+k-1}{k-1}$$

2.1.3 Order, Replacement, and Distinguishability

This is a list of translations, from "counting" to "English". One of the most common mistakes over-counting or under-counting the number of combinations.

- Indistinguishable balls $\implies$ Order doesn't matter, combinations
- Distinguishable balls $\implies$ Order does matter, permutations
- Only one ball in each bin $\implies$ Without replacement
- Multiple balls allowed in each bin $\implies$ With replacement

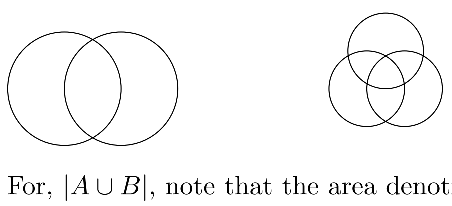
2.1.4 Combinatorial Proofs

Several of the following are broadly applicable, for all sections in probability. However, we will introduce them here, as part of a set of approaches you can use to tackle combinatorial proofs.

- Addition is OR, and multiplication is AND.
  - "Expand" coefficients. $2\binom{n}{2} = \binom{n}{2} + \binom{n}{2}$, so consider all pairs from n OR consider all pairs from (another) n.
  - Distribute quantities. $\binom{n}{2}^{a+b} = \binom{n}{2}^a \binom{n}{2}^b$, so we consider all pairs from n, a times AND consider all pairs from (another) n, b times.
- Switch between equivalent forms for combinations, to see which makes more sense.
  - Rewrite quantities as "choose 1". $n = \binom{n}{1}$, so we pick one from n items.
  - Toggle between the two quantities. $\binom{a+b}{b} = \binom{a+b}{a}$, as choosing b is the same as choosing a from $a + b$.
- Try applying the first rule of counting as well.
  - 2^n is the equivalent of picking all (possibly empty) subsets. In other words, we consider 2 possibilities for all n items $\{\text{INCLUDE}, \text{DON'T INCLUDE}\}$.
  - $\binom{n}{k}k! = \frac{n!}{(n-k)!}$, which is k samples from n items *without* replacement.
- Make sure to not prove equality mathematically, or attempt to just write in words what happens mathematically.

2.1.5 Inclusion Exclusion Principle

This section prevents over-counting; we can visualize this as Venn Diagrams.



For, $|A \cup B|$, note that the area denoting $A \cap B$ is duplicated. So, we subtract the intersection to get the space of all A and B. Thus,

$$|A \cup B| = |A| + |B| - |A \cap B|$$
$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |B \cap C| - |C \cap A| + |A \cap B \cap C|$$

2.2 Stars and Bars Walkthrough

Stars and bars can come in many different types and flavors. The following build on one case, building in complexity. Here are several other stars and bars variations:

- All possible phone numbers, given that the digits sum to a particular value.
- Distributing chairs among rooms, given we have only a particular number of chairs.
- Bringing at least k of each sock, given you can only fit n socks in your suitcase.

Notice that in all of the mentioned scenarios, there is some "bound" on the number of items we can distribute. That should immediately trigger "stars and bars".

<u>QUESTION: BASIC</u>

How many ways are there to sprinkle 10 oreo pieces on our 3 scoops of ice cream? Assume that each scoop is a different flavor of ice cream. (i.e., Each scoop is distinguishable.)

ANSWER $\binom{12}{2}$

This reduces to a stars and bars problem, where we have 3 possible "buckets" (2 bars) and 10 stars. Note that to specify 3 buckets, we need only 2 bars to separate the 10 stars. Thus, we can choose the 2 bars or the 10 balls from 12 slots.

<u>QUESTION: AT LEAST</u>

How many ways are there to sprinkle 10 sprinkles on 3 scoops, such that the first scoop gets at least 5 pieces?

ANSWER $\binom{7}{2}$

This reduces to a second stars and bars problem. Simply, we first give the first scoop 5 pieces. Why do this? This means that regardless of however many *additional* pieces we distribute it, the first scoop will have at least 5 pieces.

We are then left with a sub-stars-and-bars problem, with $10 - 5 = 5$ sprinkles and 3 scoops. We proceed as usual, noting this is 5 stars and 2 bars.

<u>QUESTION: AT MOST</u>

Assume that each scoop can only hold a maximum of 8 pieces. How many ways are there to sprinkle 10 sprinkles on 3 scoops?

ANSWER $\binom{12}{2} - \binom{3}{1}\binom{2}{1} - \binom{3}{1}$

First, we count all the possible ways to distribute 10 oreo sprinkles among 3 scoops. This is the answer to the last quiz, problem 3: $\binom{12}{2}$

Then, we count the number of invalid combinations. The only invalid combinations are when a scoop has 9 or more sprinkles. We consider each case:

1. One scoop has 9 sprinkles. There are $\binom{3}{1}$ ways to pick this one scoop with 9 sprinkles. Then, there are two other scoops to pick from, to give the final sprinkle, making $\binom{2}{1}$ ways to "distribute" the last sprinkle.
2. One scoop has 10 sprinkles. There are $\binom{3}{1}$ ways to pick this one scoop. There are no more sprinkles for the other scoops.

Thus, we take all combinations and then substract both invalid combinations. We note that the invalid combinations are mutually exclusive.

Question: At Least and At Most

Assume that each scoop can only hold a maximum of 8 pieces and a minimum of 2. How many ways are there to sprinkle 14 sprinkles on our 3 scoops of ice cream?

Answer $\binom{10}{2} - \binom{3}{1}\binom{2}{1} - \binom{3}{1}$

Given the first problem in this walkthrough, we know that we can reduce the problem to a sub-stars-and-bars problem. We first distribute 2 sprinkles to each scoop, guaranteeing that each scoop will have at least 2 sprinkles distributed to it.

Then, we count all the possible ways to distribute 8 oreo sprinkles among 3 scoops. This is - by stars and bars - $\binom{10}{2}$.

Then, we count the number of invalid combinations. Since each scoop already has 2 sprinkles, it can take at most 6 more, to satisfy the 8-sprinkles maximum. Thus, the invalid combinations are when we distribute 7 or more sprinkles to a single scoop.

We consider each case:

1. One scoop has 7 sprinkles. There are $\binom{3}{1}$ ways to pick this one scoop with 7 sprinkles. Then, there are two other scoops to pick from, to give the final sprinkle, making $\binom{2}{1}$ ways to "distribute" the last sprinkle.
2. One scoop has 8 sprinkles. There are $\binom{3}{1}$ ways to pick this one scoop. There are no more sprinkles for the other scoops.

Thus, we take all combinations and then substract both invalid combinations. We note that the invalid combinations are mutually exclusive, making $\binom{10}{2} - \binom{3}{1}\binom{2}{1} - \binom{3}{1}$.

2.3 Problems

1. If we roll a standard 6-sided die 3 times, how many ways are there to roll a sum total of 14 pips where all rolls have an even number of pips?
2. Given a standard 52-card deck and a 5-card hand, how many unique hands are there with at least 1 club AND no aces?
3. Given a standard 52-card deck and a 5-card hand, how many unique hands are there with at least 1 club OR no aces?
4. Given a standard 52-card deck and a 3-card hand, how many unique hands are there with cards that sum to 15? (*Hint: Each card is uniquely identified by both a number and a suit. This problem is more complex than phone numbers.*)

Chapter 3

Probability

3.1 Guide

3.1.1 Random Variables

Let Ω be the sample space. A random variable is by definition a function mapping events to real numbers. $X : \Omega \to \mathbb{R}, X(\omega) \in \mathbb{R}$. An *indicator variable* is a random variable that only assumes values $\{0, 1\}$ to denote success or failure for a single trial. Note that for an indicator, expectation is equal to the probability of success:

$$E[X_i] = 1 \cdot P[X_i = 1] + 0 \cdot P[X_i = 0] = P[X_i = 1]$$

3.1.2 Law of Total Probability

The law of total probability states that $Pr[A] = Pr[A|B]Pr[B] + Pr[A|\bar{B}]Pr[\bar{B}]$, if the only values of B are B and $\bar{B}$. More generally speaking, for a set of B_i that partition Ω,

$$Pr[A] = \sum_i Pr[A|B_i]Pr[B_i]$$

Do not forget this law. On the exam, students often forget to multiply by $Pr[B_i]$ when computing $Pr[A]$.

3.1.3 Conditional Probability

Conditional probability gives us the probability of an event given *priors*. By definition, the probability of A given B is defined to be

$$Pr[A|B] = \frac{Pr[A \cap B]}{Pr[B]}$$

3.1.4 Bayes' Rule

Bayes' expands on this idea, using the Law of Total Probability.

$$Pr[A|B] = \frac{Pr[B|A]Pr[A]}{Pr[B]}$$

3.1.5 Independence

Note that if the implication only goes in one direction, the converse is *not* necessarily true. This is a favorite for exams, where the crux of a True-False question may be rooted in a converse of one of the following implications.

$$X, Y \text{INDEPENDENT} \Leftrightarrow (Pr[XY] = Pr[X]Pr[Y])$$
$$X, Y \text{INDEPENDENT} \Rightarrow (E[XY] = E[X]E[Y])$$
$$X, Y \text{INDEPENDENT} \Rightarrow (\text{VAR}(X + Y) = \text{VAR}(X) + \text{VAR}(Y))$$
$$X, Y \text{INDEPENDENT} \Rightarrow (\text{COV}(X, Y) = 0)$$

Using the above definition for independence with probabilities, we have the following corollary.

$$X, Y \text{INDEPENDENT} \Leftrightarrow (Pr[X|Y] = \frac{Pr[X, Y]}{Pr[Y]} = \frac{Pr[X]Pr[Y]}{Pr[Y]} = Pr[X])$$

3.1.6 Symmetry

Given a set of trials, the principle of symmetry states that the probability of each trial is independent of other trials, *without additional information.* See *3.2 Symmetry Walkthrough* for more details and concrete examples.

3.2 Symmetry Walkthrough

Let us take 10 marbles from a bag of N marbles, of which $r > 10$ are red, *without* replacement. Let X_i be an indicator equal to 1 if and only if the ith marble is red.

QUESTION: SINGLE UNCONDITIONED, NONE CONDITIONED

With what probability is the first marble red? The second? Third? The tenth?

ANSWER All: $\frac{r}{N}$

The probability of the first marble being red is $\frac{r}{N}$. However, by symmetry, the probability of each marble being red is *also* $\frac{r}{N}$! We know this is true, because we are not given any information about any of the marbles. With or without replacement, symmetry applies. However, as we will see, symmetry breaks down or applies to a limited degree when we are given information and condition our probabilities.

QUESTION: SINGLE UNCONDITIONED, SINGLE CONDITIONED

Given that the first marble is red, with what probability is the second marble red? The third? The tenth?

ANSWER $\forall 1 < i \leq 10$, $\frac{r-1}{N-1}$

We are given that the first marble is red. As a result, we are actually computing $Pr[X_i = 1|X_1 = 1]$. Since one red marble has already been removed, we have that for the second marble, $Pr[X_2 = 1|X_1 = 1] = \frac{r-1}{N-1}$. Again, by symmetry, we in fact know that this is true for all $i > 1$, as we do not have additional information that would tell us otherwise.

QUESTION: SINGLE UNCONDITIONED, MULTIPLE CONDITIONED

Given that the first and fifth marbles are red, with what probability is the second marble red? The tenth?

ANSWER $\frac{r-2}{N-2}$

The "position" of the marble that we have information about, does not matter. Thus, again applying symmetry, we have that all remaining marbles have probability $\frac{r-2}{N-2}$ of being red.

QUESTION: MULTIPLE UNCONDITIONED, "NONE" CONDITIONED

What is the probability that the first two marbles are red? The second and third? The ninth and tenth?

ANSWER $\frac{r}{N}\frac{r-1}{N-1}$

Again, by symmetry, we can argue that regardless of *which* two marbles, the probability that any pair is red, is the probability that one pair is red. Note that symmetry doesn't apply *within* the pair, however. When considering the "second" marble, we know that the "first" marble is necessarily red. When computing the probability that the first and second marbles are red, we are then computing $Pr[X_1]Pr[X_2|X_1]$, which is $\frac{r}{N}\frac{r-1}{N-1}$.

Question: Multiple Unconditioned, Multiple Conditioned

Given that the first marble is red and the second is not red, what is the probability that the seventh marble is red and the ninth marble is not red?

Answer $\frac{r-1}{N-2}\frac{N-r-1}{N-3}$

At this point, it should be apparent that we could have asked for the probability that *any* two marbles are red. For the seventh marble, which is red, we consider the number of red marbles we have no information about, which is $r-1$ and the total number of marbles we have no information about, $N-2$. This makes the probability that the seventh marble is red, $\frac{r-1}{N-2}$.

For the tenth marble, there are now $N-3$ marbles we do not have information about. There are $r-2$ red marbles we do not have information about. Thus, the numebr of remaining non-red marbles is $(N-3)-(r-2)=N-r-1$, making the probability that the tenth marble is not red, $\frac{N-r-1}{N-3}$.

3.3 Problems

1. We sample k times at random, *without* replacement, a coin from our wallet. There are p pennies, n nickels, d dimes, and q quarters, making N total coins. Given that the first three coins are pennies, what is the probability that we will sample 2 nickels, 2 pennies, and 1 dime next, in any order?
2. We are playing a game with our apartment-mate Wayne. There are three coins, one biased with probability p of heads and the other two fair coins. First, each player is assigned one of three coins uniformly at random. Players then flip simultaneously, where each player earns h points per head. The winning player is the one with the most points. If Wayne earns k points after n coin flips, what is the probability that Wayne has the biased coin?
3. Let X and Y be the results from two numbers, chosen uniformly randomly in the range $\{1, 2, 3, ..., k\}$. Define $Z = |X - Y|$.

   (a) Find the probability that $Z < k - 2$.
   (b) Find the probability that $Z \geq 2$.

4. Consider a 5 in. $\times$ 3 in. board, where each square inch is occupied by a single tile. A monkey hammers away at the board, choosing a position uniformly at random; assume a single strike completely removes a tile from that position. (Note that the monkey can choose to strike a position with no tiles.) By knocking off tiles, the monkey can create digits. For example, the following would form the digit "3", where 0 denotes a missing tile and 1 denotes a tile.

$$\begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

   (a) Given the monkey strikes n times, where $n > 15$, what is the probability that the monkey knocks out tiles, such that the board forms "8"?
   (b) Given the monkey strikes n times, where $n > 15$, what is the probability that the monkey knocks out tiles, such that the board forms "2"?
   (c) Given the monkey strikes n times, where $n > 15$, what is the probability that the monkey knocks out tiles, such that the board forms an even number?
   (d) Which digit is most likely to occur?

Chapter 4

Expectation

4.1 Guide

With expectation, we begin to see that some quantities no longer make sense. Expressions that we compute the expectation for may in fact be far detached from any intuitive meaning. We will specifically target how to deal with these, in the below regurgitations of expectation laws and definitions.

4.1.1 Expectation Definition

Expectation is, intuitively, the mean. We multiply all *values* by the *probability* that that value occurs. The formula is as follows:

$$E[X] = \sum_{x \in \mathbb{R}} x Pr[X = x]$$

However, we also know the following.

$$E[g(X)] = \sum_{x \in \mathbb{R}} g(x) Pr[X = x]$$

Said another way, the expression in $E[...]$, which we will call $g(X)$ can be needlessly complex. To solve such an expectation, simply plug in the expression into the summation. For example, say we need to solve for $E[X^{2/5}]$. This makes little intuitive sense. However, we know the expression in terms of X affects only the *value*.

$$E[X^{2/5}] = \sum x^{2/5} Pr[X = x]$$

This also extends to multiple random variables, so

$$E[g(X,Y)] = \sum_{x,y \in \mathbb{R}} g(x,y)Pr[X = x, Y = y]$$

For example, we have that

$$E[(X+Y)^2] = \sum_{x,y \in \mathbb{R}} (x+y)^2 Pr[X = x, Y = y]$$

4.1.2 Linearity of Expectation

Regardless of independence, the linearity of expectation always holds. Said succinctly, it is true that $E[\sum_i a_i X_i] = \sum_i a_i E[X_i]$. Said once more in a less dense format, using constants a_i and random variables x_i:

$$E[a_1 X_1 + a_2 X_2 + \cdots + a_n X_n] = a_1 E[X_1] + a_2 E[X_2] + \cdots + a_n E[X_n]$$

Given a more complex combination of random variables, apply linearity of expectation to solve.

4.1.3 Conditional Expectation

Here, we expand our definition of expectation.

$$E[Y|X = x] = \sum_y y Pr[Y = y|X = x]$$

We know how to solve for $Pr[Y = y|X = x]$, using definitions from the last chapter.

4.1.4 Law of Total Expectation

The Law of Total Expectation states simply that

$$E[E[Y|X]] = E[Y]$$

However, the projection property provides a more general form, showing that the law of total expectation is actually a special case where $f(x) = 1$.

$$E[E[Y|X]f(X)] = E[Yf(X)]$$

4.2 Linearity of Expectation Walkthrough

Let us begin with a simple linearity of expectation problem, where of the random variables are independent. However, as we will see, linearity of expectation always holds, regardless of dependence between the involved random variables. From this walkthrough, you should learn:

- When a question asks "how many", *immediately* think "indicator variable". The indicator variable is then used to indicate whether or not the ith trial is a success. Sum over all indicators to compute the number of successes.
- **Option 1 for computing expectation**: Directly compute a complex expression for random variables, using the fact that $E[g(X,Y)] = \sum g(x,y)Pr[X = x, Y = y]$
- **Option 2 for computing expectation**: We can expand an expectation that makes little intuitive sense, such as $E[(X+Y)^2] = E[X^2+2XY+Y^2]$. Using linearity of expectation, we can then compute each term separately.

QUESTION: VARIABLES

Knowing that $E[X] = 3, E[Y] = 2, E[Z] = 1$, compute $E[1X + 2Y + 3Z]$.

ANSWER 10

Use linearity of expectation to expand $E[X+2Y+3Z] = E[X]+2E[Y]+3E[Z] = 3 + 2(2) + 3(1) = 10$.

QUESTION: INDEPENDENCE

Let $Z = X_1 + X_2 + X_3 + X_4$, where each X_i is the number of pips for a dice roll. What is $E[Z]$?

ANSWER 14

We begin by computing $E[Z] = E[X_1 + X_2 + X_3 + X_4]$, which, by linearity of expectation, is the same as $E[X_1] + E[X_2] + E[X_3] + E[X_4]$. All of the X_is are the same, making $\forall i, E[X_i] = \frac{7}{2}$. This makes $E[Z] = 4E[X_1] = 4\frac{7}{2} = 14$.

We will now see a more surprising application of linearity of expectation, where the random variables are dependent.

QUESTION: DEPENDENCE

Consider a bag of k red marbles and k blue marbles. *Without replacement*, we pull 4 marbles from the bag in sequential order; what is the expected number of red marbles?

ANSWER 2

The question asks for "how many," so we know we need indicators to count successes. Let us define $Z = \sum_{i=1}^{4} X_i$, where each X_i is 1 if the ith marble is red. We know, $E[Z] = E[\sum_{i=1}^{4} X_i]$, and by linearity of expectation

$$E[Z] = \sum_{i=1}^{4} E[X_i]$$

For all indicators, $E[X] = P[X = 1]$.

$$= \sum_{i=1}^{4} P[X_i = 1]$$
$$= \sum_{i=1}^{4} \frac{1}{2}$$
$$= 4\frac{1}{2} = 2$$

As linearity of expectation shows, the probability of a red marble on the first marble is the same as the probability of a red marble on the fourth. This *symmetry* applies to any set of samples where we are not given additional information about the samples drawn.

Question: Algebra with Independence

Consider three random variables X, Y, and Z, where X is the outcome of a dice roll, Y is the outcome of another dice roll, and $Z = Y + X$ (Z be the sum of the two dice roll outcomes). Compute $E[Z]$ and $E[Z^2]$. Is $E[Z^2] = E[Z]^2$?

Answer 7, 38.5, No

First, we'll compute $E[Z]$.

$$E[Z] = E[X + Y]$$
$$= E[X] + E[Y]$$
$$= \frac{7}{2} + \frac{7}{2}$$
$$= 7$$

$E[Z^2]$ is not guaranteed to be 49. $E[Z^2] \neq E[Z]^2$ *necessarily*. Whereas it is possible, it is not guaranteed.

$$E[Z^2] = E[(X + Y)^2]$$
$$= E[X^2 + 2XY + Y^2]$$
$$= E[X^2] + 2E[XY] + E[Y^2]$$

We will compute each term separately, to provide clarity.

First, we compute $2E[XY]$. Since, X and Y are independent, we can rewrite this as

$$2E[X]E[Y] = 2(\frac{7}{2})(\frac{7}{2}) = \frac{49}{2}$$

Second, we compute $E[X^2]$. This is given in the previous part.

Third, note that $E[X^2] = E[Y^2]$, since these are both dice rolls. Thus,

$$E[X^2] + 2E[XY] + E[Y^2]$$
$$= \frac{91}{6} + \frac{49}{2} + \frac{91}{6}$$
$$= \frac{329}{6}$$

This makes $E[Z^2] = 54.8$, which is not $E[Z]^2 = 49$.

Question: More Dependence

Angie goes to the music building daily to practice singing. Each day, she chooses one of n pieces at random. Given some $m < n$ of the pieces are arias, let us impose an absolute ordering on these arias. If Angie practices for k days, how many times will Angie practice all arias in sequential order, without repeating an aria? (Note that this means Angie will necessarily spend m days, practicing one aria a day, to finish one sequence.)

Answer: $(k - m + 1)\frac{1}{n^m}$

The question asks for "how many," so we know we need indicators to count successes. We define X to be the total number of sequences and $X = \sum_i X_i$ where X_i is be 1 iff Angie begins a sequence on day i. This means that the last day Angie can begin a sequence is $k - m + 1$. Thus, we actually consider $k - m + 1$ trials.

Now, we compute the probability that, of m trials, Angie picks exactly the right m arias in sequential order. Thus, the probability for a particular X_i is $\frac{1}{n^m}$.

$$E[X] = E[\sum_i X_i]$$
$$= \sum_i E[X_i]$$
$$= (k - m + 1)E[X_i]$$
$$= (k - m + 1)\frac{1}{n^m}$$

4.3 Dilution Walkthrough

For the following problems, some answers are so complex syntactically that placing an upside-down miniature version would render the answer completely illegible. As a consequence, most answers in this section are placed right-side up in plain sight.

It is now year 3000 where watermelon is a sacred fruit. Everyone receives N watermelons at birth. However, citizens of this future must participate in the watermelon ceremonies, annually. At this ritual, citizens can choose to pick 1 melon at random, to replace with a cantaloupe.

QUESTION: TWO CASES

Given that a citizen has m watermelons at the nth year, what are all the possible number of watermelons that this citizen can have in the $n+1$th year, and what is the probability that each of these situations occur?

$$X_{n+1} = \begin{cases} m & w.p. 1 - \frac{m}{N} \\ m-1 & w.p. \frac{m}{N} \end{cases}$$

In the first case, our number of watermelons does not change. This only occurs if our pick is a cantaloupe. Since there are m watermelons, there are $N-m$ cantaloupes. Thus, the probability of picking a single cantaloupe is $\frac{N-m}{N} = 1 - \frac{m}{N}$.

The second case falls out, as there are m watermelons, making $\frac{m}{N}$.

QUESTION: THREE CASES

Let us suppose that a citizen now picks **two** watermelons at random, at this ritual. Given that a citizen has m watermelons at the nth year, what are all the possible number of watermelons that this citizen can have in the $n+1$th year, and what is the probability that each of these situations occur?

$$X_{n+1} = \begin{cases} m & w.p. \frac{(N-m)(N-m-1)}{N(N-1)} \\ m-1 & w.p. \frac{2m(N-m)}{N(N-1)} \\ m-2 & w.p. \frac{m(m-1)}{N(N-1)} \end{cases}$$

In the first case, our number of watermelons does not change. This only occurs if both of our picks are cantaloupes. Since there are m watermelons, there are $N-m$ cantaloupes. Thus, the probability of picking a single cantaloupe is $\frac{N-m}{N}$. The probability of picking two cantaloupes is $\frac{(N-m)(N-m-1)}{N(N-1)}$.

Likewise, if the number of watermelons decreases by 1, we have chosen one watermelon and one cantaloupe. This means we either chose the watermelon second and the cantaloupe first $\frac{N-m}{N}\frac{m}{N-1}$, or we chose the watermelon first and the cantaloupe second $\frac{m}{N}\frac{N-m}{N-1}$. Summed together, we have that the probability of one watermelon and one cantaloupe is $\frac{2m(N-m)}{N(N-1)}$.

Finally, the probability of picking two watermelons is $\frac{m(m-1)}{N(N-1)}$.

QUESTION: CONDITIONAL EXPECTATION, TWO CASES

Again, let us consider the original scenario, where each citizen picks only 1 melon at random at the ritual. Given that a citizen has m watermelons at the nth year, how many watermelons will a citizen then have in year $n+1$, on average?

ANSWER $X_n(1-\frac{1}{N})$

We are effectively computing $E[X_{n+1}|X_n = m]$. We already considered all possible values of X_{n+1} with their respective probabilities. So,

$$\begin{aligned}
E[X_{n+1}|X_n = m] &= \sum_x E[X_{n+1} = x|X_n = m] \\
&= \sum_x xPr[X_{n+1} = x|X_n = m] \\
&= m(1-\frac{m}{N}) + (m-1)\frac{m}{N} \\
&= m - \frac{m^2}{N} + \frac{m^2}{N} - \frac{m}{N} \\
&= m - \frac{m}{N} \\
&= m(1-\frac{1}{N})
\end{aligned}$$

Since $X_n = m$, we substitute it in.

$$= X_n(1-\frac{1}{N})$$

QUESTION: CONDITIONAL EXPECTATION, THREE CASES

Let each citizen pick **two** melons per ritual. Given that a citizen has m watermelons at the nth year, how many watermelons will a citizen then have in year $n+1$, on average?

ANSWER $m(1-\frac{2}{N})$

We are effectively computing $E[X_{n+1}|X_n = m]$. We already considered all possible values of X_{n+1} with their respective probabilities. So,

$$E[X_{n+1}|X_n = m] = \sum_x xPr[X_{n+1} = x|X_n = m]$$

$$\begin{aligned}
&= \frac{m(N-m)(N-m-1) + (m-1)2m(N-m) + (m-2)m(m-1)}{N(N-1)} \\
&= \frac{m(N-1)(N-2)}{N(N-1)} \\
&= m\frac{N-2}{N}
\end{aligned}$$

Since $X_n = m$, we substitute it in.

$$E[X_{n+1}|X_n] = X_n(1 - \frac{2}{N})N$$

QUESTION: EXPECTATION

Let each citizen pick **two** melons per ritual. After n years, compute the average number of watermelons a particular citizen will have left.

ANSWER $(1 - \frac{2}{N})^{n-1}N$

We are now computing $E[X_n]$. First, we note that the law of total expectation allows us to conclude the following.

$$\begin{aligned} E[X_{n+1}] &= E[E[X_{n+1}|X_n]] \\ &= E[X_n(1 - \frac{2}{N})] \\ &= (1 - \frac{2}{N})E[X_n] \end{aligned}$$

We have the following relationship.

$$E[X_n] = (1 - \frac{2}{N})E[X_{n-1}]$$

Since $E[X_n]$ is recursively defined, we see that the constant in front of $E[X_{n-1}]$ will simply be multiplied repeatedly. Thus, we can express this in terms of $E[X_1]$.

$$E[X_n] = (1 - \frac{2}{N})^{n-1}E[X_1]$$

Finally, we note that we began with N watermelons, so $E[X_1] = N$.

$$E[X_n] = (1 - \frac{2}{N})^{n-1}N$$

QUESTION: ALGEBRA

Let each citizen pick **two** melons per ritual. If all citizens begin with 100 watermelons, what is the minimum number of years such that the expected number of cantaloupes a citizen has is at least 99?

ANSWER 229

Just plug into our expression for $E[X_n]$. 99 cantaloupes means 1 watermelon. Thus, we are solving for $E[X_n] = 1$.

$$E[X_n] = (1 - \frac{2}{100})^{n-1}100 = 1$$

$$(\frac{49}{50})^{n-1} = \frac{1}{100}$$
$$\log{(\frac{49}{50})^{n-1}} = \log\frac{1}{100}$$

Using the log rule, $\log a^n = n \log a$, we get:

$$(n-1)\log\frac{99}{100} = \log\frac{1}{100}$$
$$n-1 = \frac{\log\frac{1}{100}}{\log\frac{49}{50}}$$
$$n = 1 + \frac{\log\frac{1}{100}}{\log\frac{49}{50}}$$
$$n \approx 229$$

It will take approximately 229 years.

4.4 Problems

1. Consider a set of books B. The spine of each book has a single character inscribed, and this character is the only distinguishing characteristic. Placed side-by-side in the right order, the spines of all books in B spell some string s, "RaoWalrandRule". What is the expected number of times s shows up, if a monkey picks n books from B uniformly at random and places them, uniformly at random, on a shelf? The books may be upside-down, but the spines are always facing out. *Hint: Do repeating characters affect the probability?*
2. Four square inch tiles make up a 2 in. $\times$ 2 in. insignia, and each tile is distinct. We have 257^2 such tiles. When the tiles are randomly arranged in a 257×257 grid, what is the expected number of properly-formed insignias? Assume that tiles can be rotated in any of 4 orientations but not flipped.
3. Let X_i be the event that the ith dice roll has an even number of pips. Let Z be the product of all X_i, from 1 to 10. Formally, $Z = \Pi_{i=1}^{10} X_i$. Let Y be the sum of all X_i, from 1 to 10. Formally, $Y = \sum_{i=1}^{10} X_i$.

   (a) Compute $E[Y|Z]$. Remember that your result is a function of Z.

   (b) Using your derivation from part (a), compute $E[Y|Z > 0]$. Explain your answer intuitively.

   (c) Using your derivation from part (a), compute $E[Y|Z = 0]$.

Chapter 5

Distributions and Estimation

5.1 Guide

Distributions help us model common patterns in real-life situations. In the end, being able to recognize distributions quickly and effectively is critical to completing difficult probability problems.

5.1.1 Important Distributions

Here are several of the most important distributions and when to use them.

- **Binomial Distribution**: Number of successes in n trials, where each trial has probability p of success
- **Geometric Distribution**: Number of trials until the first success, where each trial is independent and each trial has probability p of success
- **Poisson Distribution**: The probability of k successes per n trials or a unit of time, given some average number of successes

5.1.2 Combining Distributions

- The minimum across k geometric distributions, each with the same parameter p, is $X' \sim \textsc{Geo}((1-p)^k)$.
- The sum of any k Poisson distributions is another Poisson distribution $X' \sim \textsc{Pois}(\sum_i^k \lambda_i)$.

5.1.3 Variance

Variance is by definition equal to $E[(X - E[X])^2]$. After a brief derivation, we get that $\textsc{Var}(X)$ is then the following.

$$\text{Var}(X) = E[X^2] - E[X]^2$$

Now, we discuss how to algebraically manipulate variance. First, note that shifting all values of X by a constant c will not change the variance. Second, pulling out constants from inside the variance squares the value.

$$\text{Var}(X + c) = \text{Var}(X)$$
$$\text{Var}(aX) = a^2\text{Var}(X)$$

5.1.4 Covariance

Covariance is by definition equal to $E[(X - E[X])(Y - E[Y])]$. After a brief derivation, we get that $\text{Cov}(X, Y)$ is then equal to the following.

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$$

Now, we discuss how to algebraically manipulate Cov. First, we can split sums. Second, we can move constants out and apply the constant to either variable.

$$\text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y)$$
$$a\text{Cov}(X, Y) = \text{Cov}(aX, Y) = \text{Cov}(X, aY)$$

5.1.5 Linearity of Variance

The variance of two random variables summed together is.

$$\text{Var}(X + Y) = \text{Var}(X) + 2\text{Cov}(X, Y) + \text{Var}(Y)$$

However, if X and Y are independent, we have that $\text{Cov}(X, Y) = 0$ and $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$. More generally, if all X_is are independent, the *linearity of variance* holds $\text{Var}(X_1 + ...X_n) = \text{Var}(X_1) + ... + \text{Var}(X_n)$, or:

$$\text{Var}(\sum_i X_i) = \sum_i \text{Var}(X_i)$$

If all X_i share a common distribution with identical parameters, then we also know $\sum_i \text{Var}(X_i) = n \cdot \text{Var}(X_i)$.

5.1.6 Linear Regression

$$L[Y|X] = E[Y] + \frac{\text{Cov}(X, Y)}{\text{Var}(X)}(X - E[X])$$

5.2 Variance Walkthrough

We will begin with the most basic form of variance, with independent random variables. There are several important takeaways from this walkthrough.

1. If events X_i are mutually independent, then $\text{VAR}(\sum_i X_i) = \sum_i \text{VAR}(X_i)$.
2. The converse of statement is *not* necessarily true.
3. Otherwise, we apply the more general definition of variance, that $\text{VAR}(X) = E[X^2] - E[X]^2$.

QUESTION: VARIANCE ALGEBRA

Let X be the number of pips for a single dice roll. Compute $\text{VAR}(6X^2 + 3)$.

ANSWER 5369

Shifting a distribution (or similarly, a random variable's values) by a constant does not affect the variance.

$$\text{VAR}(6X^2 + 3) = \text{VAR}(6X^2)$$

We pull out the constant, and substitute the equation in.

$$\begin{aligned} &= 36\text{VAR}(X^2) \\ &= 36(E[X^4] - E[X^2]^2) \end{aligned}$$

We know that $E[X^2] = \frac{1^2+2^2+3^2+4^2+5^2+6^2}{6} = \frac{91}{6}$. Likewise, $E[X^4] = \frac{1^4+2^4+3^4+4^4+5^4+6^4}{6} = \frac{2275}{6}$. Finally, plug in and simplify.

$$\begin{aligned} &= 36(\frac{2275}{6} - (\frac{91}{6})^2) \\ &= 36(\frac{2275}{6} - \frac{8281}{36}) \\ &= 6 \cdot 2275 - 8281 \\ &= 5369 \end{aligned}$$

QUESTION: INDEPENDENT EVENTS

Let X and Y be the number of pips for two separate dice rolls. Compute $\text{VAR}(\sqrt{6}X - \sqrt{6}Y + 3)$.

ANSWER 35

We first know that the shift does not affect variance. We apply linearity of variance, as X and Y are independent. Thus,

$$\begin{aligned}\textsc{Var}(\sqrt{6}X - \sqrt{6}Y + 3) &= \textsc{Var}(\sqrt{6}X - \sqrt{6}Y)\\ &= \textsc{Var}(\sqrt{6}X) + \textsc{Var}(-\sqrt{6}Y)\\ &= 6(\textsc{Var}(X) + \textsc{Var}(Y))\end{aligned}$$

We can compute $\textsc{Var}(X) = E[X^2] - E[X]^2$ separately. We know that $E[X] = \frac{1+2+3+4+5+6}{6} = \frac{7}{2}$. Likewise, $E[X^2] = \frac{1^2+2^2+3^2+4^2+5^2+6^2}{6} = \frac{91}{6}$. Thus,

$$\begin{aligned}\textsc{Var}(X) &= E[X^2] - E[X]^2\\ &= \frac{91}{6} - (\frac{7}{2})^2\\ &= \frac{91}{6} - \frac{49}{4}\\ &= \frac{35}{12}\end{aligned}$$

Now, we substitute in.

$$\begin{aligned}\textsc{Var}(\sqrt{6}X - \sqrt{6}Y + 3) &= 6(\textsc{Var}(X) + \textsc{Var}(Y))\\ &= 6(\frac{35}{12} + \frac{35}{12})\\ &= 6\frac{35}{6}\\ &= 35\end{aligned}$$

Chapter 6

Bounds

6.1 Guide

6.1.1 Markov's Inequality

Markov's inequality offers a bound in one direction. Intuitively, it gives us an upper bound on the probability we are greater than some value. Keep in mind that $a \geq 0$, and X cannot take negative values.

$$Pr[X > a] \leq \frac{E[X]}{a}$$

More generally, for a strictly non-negative, monotonically increasing function f,

$$Pr[X > a] \leq \frac{E[f(X)]}{f(a)}$$

6.1.2 Chebyshev's Inequality

Keep in mind that $a \geq 0$, and intuitively, it gives an *upper bound* on the probability that we are *more* than a distance a from the mean.

$$Pr[|X - E[X]| \geq \alpha] \leq \frac{\text{Var}(X)}{\alpha^2}$$

However, we may be interested in a *lower bound* on the probability that we are *less* than a distance a from the mean. Thus, if Chebyshev's offers a bound of p, we are actually interested in $1 - p$.

6.1.3 Law of Large Numbers

Let $E[\bar{X}]$ be the average across all samples of data. Let $E[X]$ be the actual average. The Law of Large Numbers states that with many i.i.d. random variables, $E[\bar{X}]$ approaches $E[X]$.

6.2 Confidence Intervals Walkthrough

This question was taken from the Spring 2016 CS70 Discussion 12.

On the planet Vegas, everyone carries a coin. Many people are honest and carry a fair coin (heads on one side and tails on the other), but a fraction p of them cheat and carry a trick coin with heads on both sides. You want to estimate p with the following experiment: you pick a random sample of n people and ask each one to flip his or her coin. Assume that each person is independently likely to carry a fair or a trick coin.

<u>QUESTION: ESTIMATION</u>

Given the results of your experiment, how should you estimate p?

We are looking for $\tilde{p}$, the fraction of people with trick coins, so let us begin by assuming that the fraction of people with trick coins is $\tilde{p}$.

Let $\tilde{q}$ be the fraction of people we *observe* with heads, in terms of $\tilde{p}$.

$$\tilde{q} = (1)\tilde{p} + \frac{1}{2}(1 - \tilde{p})$$

This implies that

$$2\tilde{q} = 2\tilde{p} + (1 - \tilde{p})$$
$$\tilde{p} = 2\tilde{q} - 1$$

Note that $\tilde{p}$ is the fraction of people we *think* have trick coins. This is different from p, which is the *actual* fraction of people with trick coins. To express the *actual* p in terms of our *actual* q, we rewrite the tilde expressions. Note that this is in theory dependent on the people we sample, so this is only an approximate equality that should be true as we approach an infinite number of samples.

$$p \approx 2q - 1$$

<u>QUESTION: CHEBYSHEV'S</u>

How many people do you need to ask to be 95% sure that your answer is off by at most 0.05?

We are looking for the difference between $\tilde{p}$ and p to be less than 0.05 with probability 95%. We first note that Chebyshev's inequality naturally follows, as Chebyshev's helps us find distance from the mean with a certain probability. Formally, this is Chebyshev's:

$$Pr[|X - \mu| \geq a] \leq \frac{var(X)}{a^2}$$

However, we are interested in finding an n so that we are off by *at most* 0.05 with probability *95%*. This is equivalent to being off by *at least* 0.05 with probability *5%*. The latter is answerable by Chebyshev's.

Then, we follow three steps.

Step 1 : Fit to $|X - \mu| \geq a$

We first only deal with "your answer is off by at most 0.05". We can re-express this mathematically, with the following:

$$|\tilde{p} - p| < 0.05$$

We don't have $\tilde{p}$, however, so we plug in q for our $\tilde{p}$.

$$|(2\tilde{q} - 1) - (2q - 1)| \leq 0.05$$
$$|2\tilde{q} - 2q| \leq 0.05$$
$$|\tilde{q} - q| \leq 0.025$$

First, note that with infinitely many samples, the fraction $\tilde{q}$ should naturally converge to become the fraction q.

$$\mu_{\tilde{q}} = q$$

We can thus transform this to something closer to our form!

$$|\tilde{q} - \mu_{\tilde{q}}| \leq 0.025$$

.

However, we need to incorporate the number of people we are sampling. So, we multiply all by n.

$$|\tilde{q}n - \mu_{\tilde{q}}n| \leq 0.025n$$

Let us consider again: what is $\tilde{q}$? We know that $\tilde{q}$ was previously defined to be the fraction of people that we *observe* to have heads. We are inherently asking for the number of heads in n trials. In other words, we want k successes among n trials, so this sounds calls for a Bernoulli random variable! We will define X_i to be 1 if the ith person tells us heads. This makes

$$\tilde{q} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

To make our life easier, let us define another random variable $Y = \tilde{q}n$.

$$Y = \tilde{q}n = \sum_{i=1}^{n} X_i$$

.

Seeing that this now matches the format we need, our α is 0.025. Our final form is

$$Pr[|Y - qn| \geq 0.025n] \leq \frac{var(Y)}{(n0.025)^2}$$

Equivalently,

$$Pr[|Y - qn| < 0.025n] \geq 1 - \frac{var(Y)}{(n0.025)^2}$$

Step 2 : Compute $\frac{var(Y)}{a^2}$

We first compute $var(X_i)$.

$$\begin{aligned} var(X_i) &= E[X_i^2] - E[X_i]^2 \\ &= q - q^2 \\ &= q(1-q) \end{aligned}$$

We then compute $var(Y)$.

$$\begin{aligned} var(Y) &= var(\sum_{i=1}^{n} X_i) \\ &= \sum_{i=1}^{n} var(X_i) \\ &= nvar(X_i) \\ &= nq(q-1) \end{aligned}$$

Thus, we have the value of our right-hand-side.

$$\begin{aligned} \frac{var(Y)}{a^2} &= \frac{nq(1-q)}{(n0.025)^2} \\ &= \frac{q(1-q)}{n(0.025)^2} \end{aligned}$$

Step 3 : Compute Bound

We now consider the remainder of our question: "How many people do you need to ask to be 95% sure...". Per the first paragraph right before step 1, we are actually interested in the probability of 5%. Thus, we want the following.

$$\frac{var(Y)}{a^2} = \frac{q(1-q)}{n(0.025)^2} = 0.05$$

.

We have an issue however: there are two variables, and we don't know q. However, we can upper bound the quantity $q(1-q)$. Since Chebyshev's computes an upper bound for the probability, we can substitute $q(1-q)$ for its maximum value.

$$q(1-q) = q^2 - q$$

To find it's maximum, we take the derivative and set equal to 0.

$$q' = 2q - 1 = 0$$
$$q = \frac{1}{2}$$

This means that $q(1-q)$ is maximized at $q = \frac{1}{2}$, making the maximum value for $q(1-q)$, $\frac{1}{2}(1-\frac{1}{2}) = \frac{1}{2}(\frac{1}{2}) = \frac{1}{4}$. We now plug in $\frac{1}{4}$.

$$1 - \frac{q(1-q)}{n(0.025)^2} = 0.95$$
$$\frac{q(1-q)}{n(0.025)^2} = 0.05$$
$$\frac{1/4}{n(0.025)^2} = 0.05$$
$$\frac{1}{4n(0.025)^2} = \frac{1}{20}$$
$$\frac{5}{(0.025)^2} = n$$
$$n = 8000$$

Chapter 7

Markov Chains

7.1 Guide

Markov Chains are closely tied to both linear algebra and differential equations. We will explore connections with both to build a better sense of how markov chains work.

7.1.1 Definition

Formally, a Markov Chain is a countable set of random variables that satisfy the *memoryless (Markov) property*, where transitions to the next state depend only on the current state.

7.1.2 Characterization

We are interested in three properties of Markov Chains: (1) reducibility, (2) periodicity, and (3) transience.

- A Markov chain is irreducible if it can go from every state i to every other state j, possibly in multiple steps.
- $d(i) :=$ g.c.d$\{n > 0 \mid P^n(i,i) = Pr[X_n = i|X_0 = i] > 0\}, i \in \mathcal{X}$ where $d(i) = 1$ if and only if the state i is aperiodic. The Markov chain is aperiodic if and only if all states are aperiodic.
- A distribution π is invariant for the transition probability matrix P if it satisfies the following balance equation: $\pi = \pi P$. If a time-dependent distribution converges $\lim_{n\to\infty} \pi^n = \pi$, the resulting distribution is then called the *stationary distribution* or *steady state distribution.*

7.1.3 Transition Probability Matrices

The Transition Probability Matrix (TPM) is written so that the rows sum to 1. Each i, jth entry corresponds to the probability that we transition *from* state i (row) to state j (column).

7.1.4 Balance Equations

Balance equations consider incoming edges in the Markov chain, where each $\pi(i)$ is the sum of all previous states $\pi(h)$ multiplied by the probability that h transitions to i. More succinctly,

$$\forall i, \pi(i) = \sum_h P_{hi} \cdot \pi(h)$$

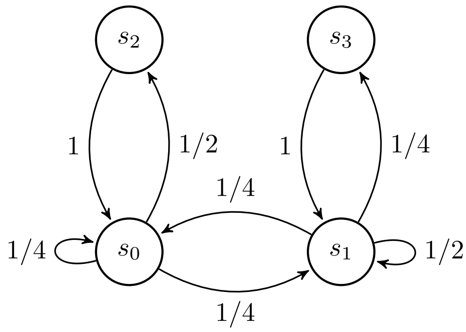
Note that we can obtain the balance equations by left-multiplying the TPM by $\vec{\pi} = [\pi(0)\pi(1)\ldots\pi(n)]^T$

7.1.5 Important Theorems

- Any finite, irreducible Markov chain has a unique invariant distribution.
- Any irreducible, aperiodic Markov chain has a unique invariant distribution that it will converge to, independent of the chain's initial state.

7.2 Hitting Time Walkthrough

"Hitting Time" questions ask for the expected amount of time to "hit" a state. Note that for in CS70, we consider only discrete-time Markov chains. The first 3 variants introduced below reduce to solving a system of linear equations.



Question: Linear Algebra with Transition Probability

Let X_i be the number of steps needed to reach X_3. Compute $E[X_0]$.

Answer 16

Begin by writing the First Step equations.

$$\beta(s_0) = 1 + \frac{1}{4}\beta(s_0) + \frac{1}{4}\beta(s_1) + \frac{1}{2}\beta(s_2)$$
$$\beta(s_1) = 1 + \frac{1}{4}\beta(s_0) + \frac{1}{2}\beta(s_1) + \frac{1}{4}\beta(s_3)$$
$$\beta(s_2) = 1 + \beta(s_0)$$
$$\beta(s_3) = 0$$

Bring all constants to one side and all $\beta(s_i)$ to the right.

$$-1 = -\frac{3}{4}\beta(s_0) + \frac{1}{4}\beta(s_1) + \frac{1}{2}\beta(s_2)$$
$$-1 = \frac{1}{4}\beta(s_0) - \frac{1}{2}\beta(s_1) + \frac{1}{4}\beta(s_3)$$
$$-1 = \beta(s_0) - \beta(s_2)$$
$$0 = \beta(s_3)$$

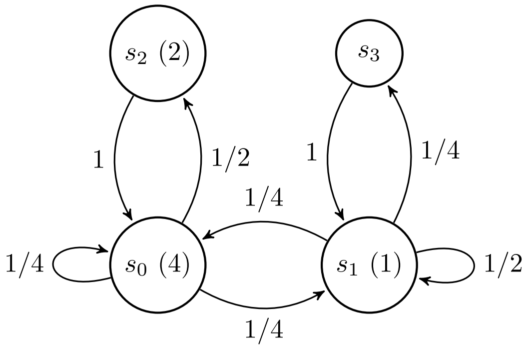
Write the system of equations as a matrix.

$$\begin{bmatrix} -3/4 & 1/4 & 1/2 & 0 & -1 \\ 1/4 & -1/2 & 0 & 1/4 & -1 \\ 1 & 0 & -1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Finally, solve the system of equations.

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 16 \\ 0 & 0 & 0 & 0 & 10 \\ 0 & 0 & 0 & 0 & 17 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

This means $\beta(s_0) = 16$.



Question: Linear Algebra with Transition Probability, Wait Time

Let X_i be the number of steps needed to reach X_3. In the Markov Chain above, the number in parentheses for a state represents the number of steps needed to "pass through" that state. Compute $E[X_0]$.

Begin by writing the First Step equations.

$$\beta(s_0) = 4 + \frac{1}{4}\beta(s_0) + \frac{1}{4}\beta(s_1) + \frac{1}{2}\beta(s_2)$$
$$\beta(s_1) = 1 + \frac{1}{4}\beta(s_0) + \frac{1}{2}\beta(s_1) + \frac{1}{4}\beta(s_3)$$
$$\beta(s_2) = 2 + \beta(s_0)$$
$$\beta(s_3) = 0$$

Bring all constants to one side and all $\beta(s_i)$ to the right.

$$-4 = -\frac{3}{4}\beta(s_0) + \frac{1}{4}\beta(s_1) + \frac{1}{2}\beta(s_2)$$
$$-1 = \frac{1}{4}\beta(s_0) - \frac{1}{2}\beta(s_1) + \frac{1}{4}\beta(s_3)$$
$$-2 = \beta(s_0) - \beta(s_2)$$
$$0 = \beta(s_3)$$

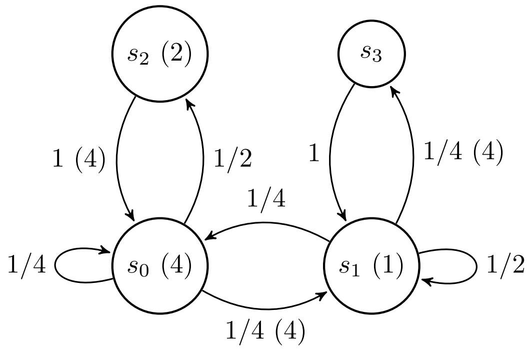
Write the system of equations as a matrix.

$$\begin{bmatrix} -3/4 & 1/4 & 1/2 & 0 & -4 \\ 1/4 & -1/2 & 0 & 1/4 & -1 \\ 1 & 0 & -1 & 0 & -2 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Finally, solve the system of equations.

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 44 \\ 0 & 0 & 0 & 0 & 24 \\ 0 & 0 & 0 & 0 & 46 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

This means $\beta(s_0) = 44$.



QUESTION: LIN. ALG. WITH TRANS. PROB., WAIT TIME, TRANS. TIME

Let X_i be the number of steps needed to reach X_3. In the Markov Chain above, the number in parentheses for a state represents the number of steps needed to "pass through" that state. The numbers in parentheses for an edge represent the number of steps needed to "pass through" that edge. If no number is specified, the edge takes only 1 step. Compute $E[X_0]$.

Begin by writing the First Step equations.

$$\beta(s_0) = 4 + \frac{1}{4}\beta(s_0) + \frac{1}{4}(\beta(s_1) + 4) + \frac{1}{2}\beta(s_2)$$

$$\beta(s_1) = 1 + \frac{1}{4}\beta(s_0) + \frac{1}{2}\beta(s_1) + \frac{1}{4}(\beta(s_3) + 4)$$

$$\beta(s_2) = 2 + (\beta(s_0) + 4)$$

$$\beta(s_3) = 0$$

Simplify all constants.

$$\beta(s_0) = 4 + \frac{1}{4}\beta(s_0) + \frac{1}{4}\beta(s_1) + 1 + \frac{1}{2}\beta(s_2)$$

$$\beta(s_1) = 1 + \frac{1}{4}\beta(s_0) + \frac{1}{2}\beta(s_1) + \frac{1}{4}\beta(s_3) + 1$$
$$\beta(s_2) = 2 + \beta(s_0) + 4$$
$$\beta(s_3) = 0$$

Bring all constants to one side and all $\beta(s_i)$ to the right.

$$-5 = -\frac{3}{4}\beta(s_0) + \frac{1}{4}\beta(s_1) + \frac{1}{2}\beta(s_2)$$
$$-2 = \frac{1}{4}\beta(s_0) - \frac{1}{2}\beta(s_1) + \frac{1}{4}\beta(s_3)$$
$$-6 = \beta(s_0) - \beta(s_2)$$
$$0 = \beta(s_3)$$

Write the system of equations as a matrix.

$$\begin{bmatrix} -3/4 & 1/4 & 1/2 & 0 & -5 \\ 1/4 & -1/2 & 0 & 1/4 & -2 \\ 1 & 0 & -1 & 0 & -6 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Finally, solve the system of equations.

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 72 \\ 0 & 0 & 0 & 0 & 48 \\ 0 & 0 & 0 & 0 & 78 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

This means $\beta(s_0) = 72$.

Chapter 8

Solutions

This section contains completely-explained solutions for each of the problems provided. Each one of these problems is designed to be at exam-level or harder, err'ing on the side of difficulty. The goal is touch on all major topics presenting in that chapter. In each of the following solutions, we identify "Takeaways" for every question at the bottom. You should understand just how the solution appeals to those takeaways, and on the exam, be prepared to apply tips and tricks presented here.

8.1 Counting

1. If we roll a standard 6-sided die 3 times, how many ways are there to roll a sum total of 14 pips where all rolls have an even number of pips?

   Solution: We can reduce all of our dice to "3-sided die" that contain only even numbers. Additionally, we can consider a reduced subproblem. In the original problem, we can only combine any number of 2, 4, 6 for a total of 14. This is the same as counting the number of ways to combine 1, 2, 3 for a total of 7.

   Distributing x to a dice roll is the same as assigning it $2x$ pips. Since a dice roll can have at most 6 pips, x is at most 3 for a single roll. Since dice do not have a 0 side, x is at least 1. Thus, we are distributing 7 balls among 3 bins with at most 3 balls and at least 1 ball for a single bin.

   By *2.2 Stars and Bars Walkthrough: At Least*, we first distribute 1 ball to each bin, reducing the problem to 4 balls and 3 bins, for $\binom{6}{2}$

   By *2.2 Stars and Bars Walkthrough: At Most*, we can identify two classes of invalid combinations:

   - Distribute all 4 balls to one bin: $\binom{3}{1}$.
   - Distribute 3 balls to one bin. There are then 2 other bins to pick from, for the last ball: $\binom{3}{1}\binom{2}{1}$

   In sum, we then have (1) all ways to distribute 7 balls, with at least 1 in each bin MINUS (2) all the ways to get more than 3 balls in a single bin.

   $$\binom{6}{2} - \binom{3}{1} - \binom{3}{1}\binom{2}{1}$$

   Takeaway: Reduce complex stars and bars to the most basic form.

   Alternate Solution:

   First, we distribute 1 ball to each bin, reducing the problem to distributing 4 balls among 3 bins such that each bin contains no more than 2 balls. The possibilities can be enumerated: $2+2+0, 2+0+2, 0+2+2, 1+1+2, 1+2+1, 2+1+1$. Hence, there are 6 total ways.

   Takeaway: When the options are few enough, enumerate.

2. Given a standard 52-card deck and a 5-card hand, how many unique hands are there with at least 1 club AND no aces?

   Solution: Let A be the event, "at least 1 club", and B be the event "no aces". We are looking for $|A \cap B|$.

   Note that computing $|A \cap B|$ is potentially tedious. Instead of considering A, "at least 1 club", it is simpler to consider $\bar{A}$, "no clubs". Thus, we rewrite $|A \cap B| = |B| - |\bar{A} \cap B|$. To compute $|\bar{A} \cap B|$, we examine all combinations with no aces and no clubs. We are drawing from $52-4-12 =$

36 cards, making $|\bar{A} \cap B| = \binom{36}{5}$. Consider all hands with no aces. This is $|B| = \binom{48}{5}$. Thus, we have

$$\begin{aligned} |A \cap B| &= |B| - |\bar{A} \cap B| \\ &= \binom{48}{5} - \binom{36}{5} \end{aligned}$$

3. Given a standard 52-card deck and a 5-card hand, how many unique hands are there with at least 1 club OR no aces?

   Solution: Again, let A be the event, "at least 1 club", and B be the event "no aces". We are looking for $|A \cup B| = |A| + |B| - |A \cap B|$.

   From the previous part, we have $|A \cap B| = |B| - |\bar{A} \cap B|$. Thus, we first simplify the original $|A \cup B|$ algebraically.

$$\begin{aligned} |A \cup B| &= |A| + |B| - |A \cap B| \\ &= |A| + |B| - (|B| - |\bar{A} \cap B|) \\ &= |A| + |\bar{A} \cap B| \end{aligned}$$

   We now compute $|A|$. Again, $|\bar{A}|$ is easier to compute, so we consider $|A| = |\Omega| - |\bar{A}|$. $|\bar{A}|$, or combinations without any clubs, is $\binom{39}{5}$. Thus, the number of combinations with at least one club is $|A| = \binom{52}{5} - \binom{39}{5}$. We now compute $|A \cup B| = |A| + |\bar{A} \cap B|$

$$\binom{52}{5} - \binom{39}{5} + \binom{36}{5}$$

   Takeaway: Draw a Venn Diagram, and compute simpler portion.

4. Given a standard 52-card deck and a 3-card hand, how many unique hands are there with cards that sum to 15? (*Hint: Each card is uniquely identified by both a number and a suit. This problem is more complex than phone numbers.*)

   Solution:

   Numbers

   First, a card has a maximum value of 13. Implicitly, we are thus looking for all the ways to distribute 15 among 3 bins, where any bin has at most 13 and at least 1. First, we distribute a value of 1 to each bin, so the problem reduces down to distributing 12 among 3 bins, where each bin is at most 12. By stars and bars, the answer here is just $\binom{14}{2}$.

   Suits

   There are a total of $\binom{4}{1}^3$ ways to pick suits. The only problem is when we assign the same number *and* suit to two different cards. We will thus consider all invalid combinations.

- It is possible that each card is a 5. There are 4^3 total assignments of suits to the cards, and only $4!/1! = 24$ of them are valid. Hence, there are $4^3 - 24$ invalid combinations here.
- Let us count the number of invalid combinations in which exactly 2 of the cards are assigned the same number. The number that the cards are assigned can range from 1 to 7, excluding 5 (since if two cards are 5, this forces the third card to also be 5, which corresponds to the previous case of all 3 cards having the same value). There are $\binom{3}{2}$ ways to choose the locations of the numbers, $\binom{6}{1}$ ways to choose the number which is repeated, 4 ways to choose the suit of the repeated numbers (in an invalid combination, the repeated numbers have the same suit), and 4 ways to choose the suit of the last number, for a total of $3 \cdot 6 \cdot 4 \cdot 4 = 288$ invalid combinations.

We thus have (1) all ways to distribute 15 validly MULTIPLIED BY (2) all possible suit choices and finally, MINUS invalid suit and number assignments.

$$\binom{14}{2}\binom{4}{1}^3 - (4^3 - 24) - 288 = 5496$$

Takeaway: Beware of over-counting.

8.2 Probability

1. We sample k times at random, *without* replacement, a coin from our wallet. There are p pennies, n nickels, d dimes, and q quarters, making N total coins. Given that the first three coins are pennies, what is the probability that we will sample 2 nickels, 2 pennies, and 1 dime next, in any order?

   Solution: *In a previous version, this problem was more complex but so was the answer. The question is now simplified.*

   We have a $\frac{n}{N-3}$ of picking a nickel and likewise $\frac{n}{N-3}$ of picking another nickel. We repeat for the pennies to obtain $\frac{p-4}{N-5}$ (as the first three coins are given to be pennies) and $\frac{p-5}{N-6}$. Finally, the probability of picking a dime next is $\frac{d}{N-7}$.

   Now, we need to discount order. There are a total of $\binom{5}{2}$ ways to pick slots for pennies, and from the remaining 3 slots, $\binom{3}{1}$ ways to pick a slot for the dime. Multiplying the two together, we obtain our final answer.

$$\binom{5}{2}\binom{3}{1}\frac{n}{N-3}\frac{n-1}{N-4}\frac{p-4}{N-5}\frac{p-5}{N-6}\frac{d}{N-7}$$

2. We are playing a game with our apartment-mate Wayne. There are three coins, one biased with probability p of heads and the other two fair coins. First, each player is assigned one of three coins uniformly at random. Players then flip simultaneously, where each player earns h points per head. The winning player is the one with the most points. If Wayne earns k points after n coin flips, what is the probability that Wayne has the biased coin?

   Solution: Let X be the number of heads that Wayne obtains. Let Y be the event that Wayne has the biased coin. By Bayes' Rule, we know the following.

$$Pr[Y|X] = \frac{Pr[X|Y]Pr[Y]}{Pr[X|Y]Pr[Y] + Pr[X|\bar{Y}]Pr[\bar{Y}]}$$

   We will first compute the easiest terms. We know the following probabilities. Given that we have three coins with one biased, the probability of a biased coin is $\frac{1}{3}$ and for an unbiased coin, $\frac{2}{3}$.

$$Pr[Y] = \frac{1}{3}$$
$$Pr[\bar{Y}] = \frac{2}{3}$$

   We will now compute both conditionals, given Y and $\bar{Y}$. Restated in English, we are computing the probability of $\frac{k}{h}$ success given n trials. The number of successes follows the $Bin(n, p)$ distribution, and the general formula is

$$Pr[Bin(n,p) = k] = \binom{n}{k}(1-p)^{n-k}p^k$$

For the biased coin, the probability of heads is p (Y). For the fair coin, the probability of heads is $\frac{1}{2}$ ($\bar{Y}$). To simplify, we will define $k' = \frac{k}{h}$, which gives us the number of heads that Wayne obtained.

$$Pr[X = k'|Y] = \binom{n}{k'}(1-p)^{n-k'}p^{k'}$$

$$Pr[X = k'|\bar{Y}] = \binom{n}{k'}\frac{1}{2^n}$$

We have computed all values, so we plug into Bayes' Rule and simplify.

$$\begin{aligned}
Pr[Y|X] &= \frac{Pr[X|Y]Pr[Y]}{Pr[X|Y]Pr[Y] + Pr[X|\bar{Y}]Pr[\bar{Y}]} \\
&= \frac{Pr[X|Y]\frac{1}{3}}{Pr[X|Y]\frac{1}{3} + Pr[X|\bar{Y}]\frac{2}{3}} \\
&= \frac{Pr[X|Y]}{Pr[X|Y] + Pr[X|\bar{Y}]2} \\
&= \frac{\binom{n}{k'}(1-p)^{n-k'}p^{k'}}{\binom{n}{k'}(1-p)^{n-k'}p^{k'} + \binom{n}{k'}\frac{1}{2^n}2} \\
&= \frac{(1-p)^{n-k'}p^{k'}}{(1-p)^{n-k'}p^{k'} + 2^{1-n}}
\end{aligned}$$

Takeaway: Remember Bayes' Rule.

3. Let X and Y be the results from two numbers, chosen uniformly randomly in the range $\{1, 2, 3, ..., k\}$. Define $Z = |X - Y|$.

   (a) Find the probability that $Z < k - 2$.

   Solution: Instead of computing $Pr[Z < k - 2]$, it is easier to compute $1 - Pr[Z < k - 2] = Pr[Z \geq k - 2]$, as this includes only two possible values for Z.

   $$Pr[Z \geq k - 2] = Pr[Z = k - 2] + Pr[Z = k - 1]$$

   We can now compute each probability independently. There are a total of k^2 possible combinations:

   - For $Pr[Z = k - 1]$, we have 1 possible combination of numbers $(k, 1)$, with 2 ways of rolling those combinations $(k, 1)$ or $(1, k)$, making 2 total combinations. $\frac{2}{k^2}$

- For $Pr[Z = k-2]$, we have 2 possible combinations of numbers $((k, 2), (k-1, 1))$ with 2 ways to roll each, making 4 total combinations. $\frac{4}{k^2}$

$$Pr[Z \geq k-2] = \frac{4}{k^2} + \frac{2}{k^2} = \frac{6}{k^2}$$

We thus solve for $Pr[Z < k-2]$.

$$1 - Pr[Z < k-2] = Pr[Z \geq k-2]$$
$$Pr[Z < k-2] = 1 - Pr[Z \geq k-2]$$

Now, we can plug in to get our final expression.

$$Pr[Z < k-2] = 1 - Pr[Z \geq k-2]$$
$$= 1 - \frac{6}{k^2}$$

Takeaway: Use counting where applicable.

(b) Find the probability that $Z \geq 2$.

Solution: Again, we apply the same trick in the previous part. It is easier to compute the probability for $Pr[Z < 2] = 1 - Pr[Z \geq 2]$. Thus, we have only two possible values of Z to account for.

$$Pr[Z < 2] = Pr[Z = 0] + Pr[Z = 1]$$

There are a total of k^2 combinations.

- We know that $Z = 0$ implies that both rolls yielded the same number. Thus, the first roll has k options and for each value the first roll assumes, the second has 1, making k total combinations. $\frac{k}{k^2}$
- We know that $Z = 1$ implies that the rolls are within 1 of each other. Let us consider two cases, (1) $X \in \{1, k\}$ or (2) $1 < X < k$. If the first roll is 1 or k, then the second roll has only one option each, 2 or $k-1$, respectively. If the first roll $1 < X < k$ (k-2 possibilities), then Y has two options each, making $2(k-2)$ possible combinations. $\frac{2+2(k-2)}{k^2} = \frac{2(k-1)}{k^2}$

$$Pr[Z < 2] = \frac{k}{k^2} + \frac{2(k-1)}{k^2} = \frac{3k-2}{k^2}$$

We thus solve for $Pr[Z \geq 2]$.

$$Pr[Z < 2] = 1 - Pr[Z \geq 2]$$
$$Pr[Z \geq 2] = 1 - Pr[Z < 2]$$

Now, plug in our solution for $Pr[Z < 2]$.

$$\begin{aligned} Pr[Z \geq 2] &= 1 - Pr[Z < 2] \\ &= 1 - \frac{3k-2}{k^2} \\ &= \frac{k^2 - 3k + 2}{k^2} \\ &= \frac{(k-2)(k-1)}{k^2} \end{aligned}$$

4. Consider a 5 in. × 3 in. board, where each square inch is occupied by a single tile. A monkey hammers away at the board, choosing a position uniformly at random; assume a single strike completely removes a tile from that position. (Note that the monkey can choose to strike a position with no tiles.) By knocking off tiles, the monkey can create digits. For example, the following would form the digit "3", where 0 denotes a missing tile and 1 denotes a tile.

$$\begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

(a) Given the monkey strikes n times, where $n > 15$, what is the probability that the monkey knocks out tiles, such that the board forms "8"?

Solution: To knock out an 8, the monkey needs to achieve the following board.

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

We can consider this as balls and bins, where we have n balls being thrown into 15 bins. Except, all balls are distinguishable. The total number of ways to strike 15 tiles is 15^n.

We wish to avoid the 2 tiles and to strike all other 13 tiles. Thus, we are throwing n balls into 13 bins, where each bin has at least 1. Since order matters, we cannot apply stars and bars directly.

i. Consider all the ways to distribute n balls among 13 bins, 13^n
ii. Subtract all cases where 1 bin is left empty. First, we choose the bin that is left empty $\binom{13}{1}$, then we distribute the n balls among the remaining 12 bins, 12^n. Together, this is $-\binom{13}{1}12^n$.

iii. By inclusion-exclusion, we have double-counted all cases where 2 bins are left empty. Choose the 2 bins that are left empty, $\binom{13}{2}$. Distribute to the remaining bins, 11^n. This is $+\binom{13}{2}11^n$.

We notice a pattern, which is that for $0 \leq i \leq 13$, we select i to be the number of bins that are empty. Then, we distribute n balls among the remaining i bins. By inclusion-exclusion, we include and exclude alternately. The total number of ways to distribute n balls among 13 bins, where each bin receives at least one ball is thus

$$\sum_{i=0}^{13}(-1)^i\binom{13}{i}(13-i)^n$$

The probability that this digit occurs is thus

$$\frac{1}{15^n}\sum_{i=0}^{13}(-1)^i\binom{13}{i}(13-i)^n$$

Generally, let a be the number of 0s in the matrix. The probability that the monkey forms this particular number is

$$\frac{1}{15^n}\sum_{i=0}^{a}(-1)^i\binom{a}{i}(a-i)^n$$

Takeaway: Don't overthink, and consider counting.

(b) Given the monkey strikes n times, where $n > 15$, what is the probability that the monkey knocks out tiles, such that the board forms "2"?

Solution: To knock out a 2, the monkey needs to achieve one of two boards.

$$\begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

By the same logic as the previous part, we can see that there are 4 positions we avoid and 11 positions we must hit. Using the general form above, we see that the number of 0s (a) is 11. Thus, the probability of a 2 is:

$$\frac{1}{15^n}\sum_{i=0}^{11}(-1)^i\binom{11}{i}(11-i)^n$$

(c) Given the monkey strikes n times, where $n > 15$, what is the probability that the monkey knocks out tiles, such that the board forms an even number?

Solution: First, let us count the number of 0s required to form each digit.

- digit 0 : 12 zeros
- digit 2 : 11 zeros
- digit 4 : 9 zeros
- digit 6 : 12 zeros
- digit 8 : 13 zeros

Let the probability that a digit with a zeros occurs be

$$p_a = \frac{1}{15^n} \sum_{i=0}^{a} (-1)^i \binom{a}{i} (a-i)^n$$

Applying the p_a, we see that the probability that any of these even numbers occurs is:

$$2p_{12} + p_{11} + p_9 + p_{13}$$

(d) Which digit is most likely to occur?

Solution: The probability that a digit occurs is dependent solely upon the number of zeros present in that digit. Thus, we look for the digit with the most zeros. We enumerate all numbers and corresponding 1s and 0s.

Digit	ones	zeros
0	3	12
1	6	9
2	4	11
3	4	11
4	6	9
5	4	11
6	3	12
7	6	9
8	2	13
9	3	12

We note that 8 has the most zeros. Thus, the digit 8 is most likely to occur.

8.3 Expectation

1. Consider a set of books B. The spine of each book has a single character inscribed, and this character is the only distinguishing characteristic. Placed side-by-side in the right order, the spines of all books in B spell some string s, "RaoWalrandRule". What is the expected number of times s shows up, if a monkey picks n books from B uniformly at random and places them, uniformly at random, on a shelf? The books may be upside-down, but the spines are always facing out. *Hint: Do repeating characters affect the probability?*

   Solution: This is similar to the last problem in the *4.2 Linearity of Expectation Walkthrough*. However, we are drawing letters from s instead of the entire alphabet. We realize the question the is asking for "expected number of (successes)" once more, so we immediately define our indicator variable, X_i to be 1 if one instance of s ends at the ith position along the shelf.

   The string is 14 letters long, meaning that the first 13 books on the shelf cannot be the position where an instance of s ends. This makes $X = \sum_{i=14}^{n} X_i$ the total number of times s appears on the shelf. By linearity of expectation:

$$E[X] = E[\sum_{i=14}^{n} X_i]$$
$$= \sum_{i=14}^{n} E[X_i]$$
$$= \sum_{i=14}^{n} Pr[X_i = 1]$$

   We now compute the probability of success, which is the probability that s is spelled, given some random sampling of letters from s. For the first letter, sampling randomly from s give us a $\frac{3}{14}$ probability of retrieving an R, as there are 14 total letters, 3 of which are R. We proceed to compute the product of these probabilities, one for each letter.

$$= \sum_{i=14}^{n} \frac{3 \cdot 3 \cdot 1 \cdot 1 \cdot 3 \cdot 2 \cdot 3 \cdot 3 \cdot 1 \cdot 1 \cdot 3 \cdot 1 \cdot 2 \cdot 1}{14^{14}}$$
$$= \sum_{i=14}^{n} \frac{3^6 \cdot 2^2}{14^{14}}$$
$$= (n-13)\frac{3^6 \cdot 2^2}{14^{14}}$$

   Takeaway: Be clever with how you define "success" for your indicator variables. Remember linearity of expectation always holds.

2. Four square inch tiles make up a 2 in. $\times$ 2 in. insignia, and each tile is distinct. We have 257^2 such tiles. When the tiles are randomly arranged in a 257×257 grid, what is the expected number of properly-formed insignias? Assume that tiles can be rotated in any of 4 orientations but not flipped.

   Solution: There are $256 \cdot 256 = 4^8$ interior vertices. Let X_i be 1 if the ith interior vertex is at the center of a valid insignia. Thus, we know that $X = \sum_i^{4^8} X_i$ is the total number of valid insignias. We can apply linearity of expectation, and then invoke properties of an indicator variable.

$$\begin{aligned} E[X] &= E[\sum_i^{4^8} X_i] \\ &= \sum_i^{4^8} E[X_i] \\ &= 4^8 E[X_i] \\ &= 4^8 Pr[X_i = 1] \end{aligned}$$

   We can use counting to compute the probability that vertex i is the center of a valid insignia. First, there are 4 valid orientations for the insignia. Second, we now compute all possible combinations of 4 tiles. Each of the four tiles is chosen at random and then rotated at random, making 4 possible tiles with 4 possible orientations each (16), for each tile. Thus there are a total of $16^4 = 4^8$ combinations.

$$Pr[X_i = 1] = \frac{4}{4^8} = \frac{1}{4^7} = \frac{1}{4^7}$$

   We now plug it back in to solve for $E[X]$.

$$\begin{aligned} E[X] &= 4^8 Pr[X_i = 1] \\ &= 4^8 \frac{1}{4^7} \\ &= 4 \end{aligned}$$

   The takeaway below was mentioned prior, but it is worth mentioning again.

   Takeaway: Define your indicator variables cleverly.

3. Let X_i be the event that the ith dice roll has an even number of pips. Let Z be the product of all X_i, from 1 to 10. Formally, $Z = \Pi_{i=1}^{10} X_i$. Let Y be the sum of all X_i, from 1 to 10. Formally, $Y = \sum_{i=1}^{10} X_i$.

(a) Compute $E[Y|Z]$. Remember that your result is a function of Z.

Solution: We will begin by making algebraic manipulations. First, we plug in Y. Then, by linearity of expectation, we move the summation out.

$$E[Y|Z] = E[\sum_{i=1}^{10} X_i|Z]$$
$$= \sum_{i=1}^{10} E[X_i|Z]$$
$$= 10E[X_i|Z]$$

Separately, we will now compute $E[X_i|Z]$. We note that for an indicator variable i, $E[X_i] = Pr[X_i = 1]$. The analogous, conditional form states the following.

$$E[X_i|Z] = Pr[X_i = 1|Z]$$

Applying Bayes' Rule, we then get the following.

$$= \frac{Pr[Z|X_i = 1]Pr[X_i = 1]}{Pr[Z|X_i = 1]Pr[X_i = 1] + Pr[Z|X_i = 0]Pr[X_i = 0]}$$

First, note that $Pr[X_i = 1]$ - the probability of rolling an even number of pips - is $\frac{1}{2}$. The probability of rolling an odd, $Pr[X_i = 0]$ is also $\frac{1}{2}$. Thus, this simplifies

$$= \frac{Pr[Z|X_i = 1]\frac{1}{2}}{Pr[Z|X_i = 1]\frac{1}{2} + Pr[Z|X_i = 0]\frac{1}{2}}$$
$$= \frac{Pr[Z|X_i = 1]}{Pr[Z|X_i = 1] + Pr[Z|X_i = 0]}$$

We will additionally assign Z to a value k. After solving $E[X_i|Z = k]$, we can then substitute all instances of k with Z for $E[X_i|Z]$.

$$E[X_i|Z = k] = \frac{Pr[Z = k|X_i = 1]}{Pr[Z = k|X_i = 1] + Pr[Z = k|X_i = 0]}$$

We now substitute Z in. For $Pr[Z = k|X_i = 0]$, since Z is the product of all X_j, including $X_i = 0$, then $Z = 0$. Thus, $Pr[Z = k|X_i = 0] = Pr[k = 0]$.

$$E[X_i|Z = k] = \frac{Pr[\Pi_{j=1}^{10} X_j = k|X_i = 1]}{Pr[\Pi_{j=1}^{10} X_j = k|X_i = 1] + Pr[k = 0]}$$

Consider the probability specified in the numerator. $Pr[X_i = 1|\Pi_{j=1}^{10} X_j]$. We know that since X_i is 1, it does not affect the product. Thus, this probability reduces to $Pr[\Pi_{j=1,i\neq j}^{10} X_j]$.

$$= \frac{Pr[\Pi_{j=1,i\neq j}^{10} X_j = k]}{Pr[\Pi_{j=1,i\neq j}^{10} X_j = k] + Pr[k = 0]}$$

Finally, to convert $E[X_i|Z = k]$ back to $E[X_i|Z]$, we substitute Z for all k. After that, we use our first result to obtain an expression for $E[Y|Z]$. We can continue to reason about the numerator, but we will leave that for part (c); a more intuition-based approach is shown in the alternate solution that follows.

$$E[X_i|Z] = \frac{Pr[\Pi_{j=1,i\neq j}^{10} X_j = Z]}{Pr[\Pi_{j=1,i\neq j}^{10} X_j = Z] + Pr[Z = 0]}$$

$$E[Y|Z] = 10 \frac{Pr[\Pi_{j=1,i\neq j}^{10} X_j = Z]}{Pr[\Pi_{j=1,i\neq j}^{10} X_j = Z] + Pr[Z = 0]}$$

Takeaway: Be able to switch between (1) algebraic manipulations and (2) intuition.

Alternate Solution:

Since Z can only be 0 or 1, let us consider the two cases. First, when $Z = 1$, then every X_i must be 1, which means that the sum of X_i's must be 10. Therefore, $E[Y|Z = 1] = 10$.

Otherwise, $Z = 0$. Computing this case requires a bit more work. From the definition:

$$E[Y] = \sum_{y=0}^{10} yPr[Y = y]$$

Conditioned on $Z = 0$, we know that at least one of the X_i's is 0. This rules out the possibility that every $X_i = 1$, which occurs with probability 2^{-10}. The probability that at least one X_i is 0 is $1-2^{-10}$. The conditional distribution $Pr(Y = y|Z = 0)$ therefore has to be "rescaled":

$$Pr[Y = y|Z = 0] = \frac{Pr[Y = y, Z = 0]}{Pr[Z = 0]} = \frac{Pr[Y = y]}{1 - 2^{-10}}, \quad 0 \leq y \leq 9$$

Therefore, the conditional expectation is

$$\begin{aligned} E[Y|Z=0] &= \sum_{y=0}^{9} y Pr[Y=y|Z=0] \\ &= \frac{1}{1-2^{-10}} \sum_{y=0}^{9} y \Pr[Y=y] \\ &= \frac{1}{1-2^{-10}} \left(\sum_{y=0}^{10} y \Pr[Y=y] - 10 \cdot \frac{1}{2^{10}} \right) \\ &= \frac{1}{1-2^{-10}} \left(E[Y] - 10 \cdot \frac{1}{2^{10}} \right) \\ &= \frac{1}{1-2^{-10}} \left(5 - 10 \cdot \frac{1}{2^{10}} \right) \end{aligned}$$

We have obtained the two different cases for $E[Y|Z=z]$, which is enough to specify $E[Y|Z]$ as a function of Z. Just for fun, we can "stitch" the two expressions into one:

$$E[Y|Z] = 10Z + \frac{1}{1-2^{-10}} \left(5 - 10 \cdot \frac{1}{2^{10}} \right) (1-Z)$$

(b) Using your derivation from part (a), compute $E[Y|Z>0]$. Explain your answer intuitively.

Solution: To gain some insight, we then evaluate this expression for $Z>0$. Notice that if $Z>0$, then $Pr[Z=0]=0$. Thus,

$$E[Y|Z>0] = 10 \frac{Pr[\Pi_{j=1,i\neq j}^{10} X_j = Z]}{Pr[\Pi_{j=1,i\neq j}^{10} X_j = Z] + 0} = 10$$

This makes sense, as if $Z>0$, then $\forall i, X_i \neq 0$. Thus, all $\forall i, X_i = 1$, making $Y = \sum_{i=1}^{10} X_i = 10$.

(c) Using your derivation from part (a), compute $E[Y|Z=0]$.

Solution: Next, we evaluate this expression at $Z=0$. Note that $Pr[Z=0]=1$.

$$E[Y|Z=0] = 10 \frac{Pr[\Pi_{j=1,i\neq j}^{10} X_j = 0]}{Pr[\Pi_{j=1,i\neq j}^{10} X_j = 0] + 1}$$

We will now compute $Pr[\Pi_{j=1,i\neq j}^{10} X_j = 0]$ separately. Note that the probability the product is 0, is the probability that at least one of the $X_j, i \neq j$ is 0. As a result, we consider the probability that *none* of the X_j are 0, and subtract that from 1. Formally,

$$\begin{aligned} &Pr[\Pi_{j=1,i\neq j}^{10} X_j = 0] \\ &= Pr[\cup_{j=1,i\neq j}^{10} X_j = 0] \end{aligned}$$

$$= 1 - Pr[\cap_{j=1,i\neq j}^{10} X_j \neq 0]$$

Since the probability of all $Pr[X_j \neq 0]$ are independent, we can multiply all of the probabilities together.

$$= 1 - \Pi_{j=1,i\neq j}^{10} Pr[X_j \neq 0]$$

For all X_j, $Pr[X_j \neq 0]$ is the probability that a dice roll is not even, which is $\frac{1}{2}$.

$$= 1 - \Pi_{j=1,i\neq j}^{10} \frac{1}{2}$$

$$= 1 - \frac{1}{2^9}$$

Plugging it back in, we get that

$$E[Y|Z=0] = 10\frac{1-\frac{1}{2^9}}{1-\frac{1}{2^9}+1}$$

$$= 10\frac{1-\frac{1}{2^9}}{2-\frac{1}{2^9}}$$

$$= 10\frac{2^9-1}{2^{10}-1}$$

Takeaway: Apply DeMorgan's to handle grotesque probabilities.