A Brief Compilation

of Guides, Walkthroughs, and Problems

Probability Theory and Random Processes

at the University of California, Berkeley

Alvin Wan

Contents

0.1 Purpose 3
- 0.1.1 Contributors 3
- 0.1.2 Structure 3
- 0.1.3 Breakdown 3
- 0.1.4 Resources 3

1 Random Variables **4**

1.1 Guide 4
- 1.1.1 Random Variables 4
- 1.1.2 Law of Total Probability 4
- 1.1.3 Conditional Probability 4
- 1.1.4 Bayes' Rule 5
- 1.1.5 Independence 5
- 1.1.6 Symmetry 5

1.2 Problems 6

2 Expectation, Variance, Covariance **8**

2.1 Guide 8
- 2.1.1 Expectation 8
- 2.1.2 Linearity of Expectation 8
- 2.1.3 Linearity of Expectation 9
- 2.1.4 Law of Total Expectation 9
- 2.1.5 Conditional Expectation 9
- 2.1.6 Law of Iterated Expectation 9
- 2.1.7 States 10
- 2.1.8 Covariance 10
- 2.1.9 Properties of Covariance 10
- 2.1.10 Variance 10
- 2.1.11 Properties of Variance 10
- 2.1.12 Variance of Sum 11
- 2.1.13 Law of Total Variance 11

2.2 Problems 12

3 Bernoulli Processes **13**
3.1 Guide 13
3.1.1 Bernoulli Distribution 13
3.1.2 Binomial Distribution 13
3.1.3 Geometric Distribution 14
3.1.4 Negative Binomial Distribution 14
3.1.5 Merging 14
3.1.6 Splitting 14
3.1.7 Combining Distributions 14
3.2 Problems 15

4 Poisson Processes **17**
4.1 Guide 17
4.1.1 Poisson Distribution 17
4.1.2 Exponential Distribution 17
4.1.3 Gamma Distribution 18
4.1.4 Merging 18
4.1.5 Splitting 18
4.1.6 Combining Distributions 18
4.2 Problems 19

5 Confidence Intervals **21**
5.1 Guide 21
5.1.1 Markov's Inequality 21
5.1.2 Chebyshev's Inequality 21
5.1.3 Confidence Interval 21
5.1.4 Weak Law of Large Numbers 22
5.2 Problems 23

6 Markov Chains **24**
6.1 Guide 24
6.1.1 Definition 24
6.1.2 Irreducibility 24
6.1.3 Periodicity 24
6.1.4 Invariant Distributions 25
6.1.5 Balance Equations 25
6.1.6 Hitting Time 25
6.1.7 Probability of A before B 26
6.2 Problems 27

7 Transformations **28**
7.1 Guide 28
7.1.1 Moment Generating Function 28
7.1.2 Expectation 28
7.1.3 Affine Transformation of Random Variable 28
7.1.4 Sum of Independent Random Variables 28

7.1.5 Combining Distributions 29
7.1.6 Inversion Property 29
7.1.7 Distributions 29
7.2 Problems 30

8 Solutions **31**
8.1 Probability 32
8.2 Expectation, Variance, Covariance 39
8.3 Bernoulli Processes 43
8.4 Poisson Processes 47
8.5 Confidence Intervals 51
8.6 Markov Chains 55
8.7 Transformations 59

0.1 Purpose

This compilation is (unofficially) written for the Fall 2016 EE126: Probability Theory and Random Processes at UC Berkeley. Its 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.

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.

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`. See also a related book, abcDMPT for Discrete Mathematics and Probability Theory.

Chapter 1

Random Variables

1.1 Guide

1.1.1 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:

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

1.1.2 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 $Pr[A]$.

1.1.3 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

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

1.1.4 Bayes' Rule

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

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

1.1.5 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.

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

1.1.6 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.*

1.2 Problems

1. Consider a housing district, represented by a 8×8 grid. Tom is currently at the bottom-left, $(0, 0)$ and would like to get home to $(7, 7)$. If he goes up with probability p and right with probability $1 - p$, what is the probability he makes it home without visiting the graveyard at $(5, 6)$ or school at $(4, 7)$? (Note that by design, it is impossible to reach both the graveyard and school. This is to make the problem slightly simpler.)
2. Consider a game of cards with 3 other friends. You and your friends draw cards from a standard 52-card deck. Each player receives a random 4 cards randomly, and each player is assigned an ace. In total, each player has 5 cards.

   (a) One of your friends, Bob, claims he holds a full house. What is the probability he is bluffing? Recall that a full house is a hand containing two cards of one rank and three cards of another rank.
   (b) Each player randomly draws an additional card at random from the deck. Bob claims he *now* holds a full house. What is the probability he is bluffing?
   (c) The dealer verifies Bob's claim. Bob now claims that the sum of the values for his cards is at least 13. What is the probability that Bob is telling the truth?

3. Consider a loaded 6-sided die, where just one number is favored with probability p. The remaining numbers have equal probability $\frac{1-p}{5}$. You and five other friends must decide who will take out the trash today, for your shared apartment. Develop a scheme that allows you to pick a victim, uniformly at random.
4. Consider a standard 52-card deck. You are assigned a standard 5-card hand, where each card is drawn randomly from the deck.

   (a) What is the probability that we have exactly 1 ace?
   (b) What is the probability that we have exactly 3 clubs? (Hint: Use counting)
   (c) Given we have three clubs, what is the probability of an ace of clubs?
   (d) What is the probability that we have 3 clubs or 1 ace? (Not XOR, Hint: Think about inclusion-exclusion.)

5. Siqi purchases s packets of strawberry Pocky (S), c packets of chocolate Pocky (C), and m packets of mint Pocky (M), making n total packets of Pocky, for her friend Tyler. He has since given in to temptation and has elected to begin eating Pocky.

   (a) Tyler randomly pulls a Pocky packet out, then places the packet back inside his bag before drawing the next one. He repeats this m times. What is the probability that he sees exactly k chocolate packets?

(b) Forest randomly steals k packets from Tyler's bag of Pocky. What is the probability that Forest has at least one chocolate packet?

(c) Given that Tyler has picked six packets of Pocky in sequential order, what is the probability that Tyler picks S, C, S, C in that order? (Note that interspersing other packets with this packet is also valid, so $S[M]CS[S]C$ would also satisfy this condition.)

6. You are rolling a 100-sided die. Across n trials, let X be the number of rolls with an odd number of dots. Let Y be the number of rolls with an even number of dots.

(a) Assume n is divisible by 4. What is $\Pr(Y - X \geq \frac{n}{2})$?

(b) Assume n is odd. What is $\Pr(X > Y)$?

Chapter 2

Expectation, Variance, Covariance

2.1 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.

2.1.1 Expectation

Expectation of a random variable X is intuitively, the mean

$$\mathbb{E}[X] = \sum x \Pr(X = x)$$

Don't try to rationalize the following; taken as a whole, there is no intuitive meaning for $g(X)$.

$$\mathbb{E}[g(X)] = \sum g(x) \Pr(X = x)$$

The continuous analog is simply an integral over the values multiplied by the **probability density function (PDF)**, denoted $f_X(x)$ for the density of X given the value x.

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

2.1.2 Linearity of Expectation

The linearity of expectation **always holds**, regardless of the independence (or lack thereof). For example,

$$\mathbb{E}[2X + 3Y] = 2\mathbb{E}[X] + 3\mathbb{E}[Y]$$

More generally, take random variables X_i and constants α_i.

$$\mathbb{E}[\sum_i \alpha_i X_i] = \sum_i \alpha_i \mathbb{E}[X_i]$$

2.1.3 Linearity of Expectation

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

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

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

2.1.4 Law of Total Expectation

Here, we expand our definition of expectation.

$$\mathbb{E}[X] = \sum_y \mathbb{E}[X|Y = y] \Pr(Y = y)$$

2.1.5 Conditional Expectation

Note that the following is a *function of X*. In fact, $\mathbb{E}[Y|X]$ is a random variable, unlike $\mathbb{E}[X]$.

$$\mathbb{E}[Y|X] = \sum_y y Pr[Y = y|X]$$

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

2.1.6 Law of Iterated Expectation

The Law of Total Expectation states simply that

$$\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$$

2.1.7 States

We can thus model the evolution of a system over time using $\mathbb{E}[X_t|X_{t-1}]$. In other words, we can express the state at some time step t, X_t with a function g in terms of the state at the previous time step, $g(X_{t-1})$.

Consider $X(t+1) = \alpha X(t)$ for some constant α. In terms of $X(0)$, we have that $X(t)$ is

$$X(t) = \alpha^t X(0)$$

Consider $X(t+1) = \alpha X(t) + \beta$ for constants α, β. In terms of $X(0)$, $X(t)$ is

$$X(t) = \alpha^t X(0) + \beta(\sum_{i=0}^{t-1} \alpha^i)$$

(Note that the summation begins from $a^0 = 1$)

2.1.8 Covariance

Take two random variables X, Y. They do not need to be independent. We have the following expression. See 3.2 for a proof.

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

2.1.9 Properties of Covariance

Covariances sum, like vectors, and are symmetric.

- $\text{cov}(X, Y) = \text{cov}(Y, X)$
- $\text{cov}(A + B, Y) = \text{cov}(A, Y) + \text{cov}(B, Y)$

2.1.10 Variance

Variance of a random variable X is intuitively, the squared distance from the mean, or the *spread*.

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

2.1.11 Properties of Variance

- Variance is unaffected by a constant shift, α. Remember variance is the spread of our X.

$$\text{var}(X + \alpha) = \text{var}(X)$$

- A scalar constant is squared when taken out of variance.

$$\textsc{var}(\alpha X) = \alpha^2 \textsc{var}(X)$$

2.1.12 Variance of Sum

Take two random variables X, Y. They do not need to be independent. We have the following expression.

$$\textsc{var}(X + Y) = \textsc{var}(X) + \textsc{var}(Y) + 2\textsc{cov}(X, Y)$$

Here's a nifty trick: Say you know variance and the expected value of a random variable X. Then, we can compute $\mathbb{E}[X^2]$ fairly efficiently! As a matter of fact, it is $\mathbb{E}[X^2] = \textsc{var}(X) + \mathbb{E}[X]^2$

Note that if two random variables are independent, the **linearity of variance** applies, as $\textsc{cov}(X, Y) = 0$, we have that

$$\textsc{var}(X + Y) = \textsc{var}(X) + \textsc{var}(Y)$$

2.1.13 Law of Total Variance

The following is most useful when you do not know the variance of random variable directly, but given additional information, it is simpler to compute. Take random variables, X, N.

$$\textsc{var}(X) = \textsc{var}(\mathbb{E}[X|N]) + \mathbb{E}[\textsc{var}(X|N)]$$

2.2 Problems

1. Consider a new, innovative textile, a 6-sided cube. (Think of it as a 6-sided stamp) We place 37^3 textiles in an $37 \times 37 \times 37$ cube, orienting these cubes randomly. Four textiles together, oriented in the correct direction will print an insignia. The insignia is symmetric across both axes, all cubes are identical, and each textile has 4 printable surfaces of 6 sides. e.g., If either of the 2 remaining faces is facing up, that textile will not contribute to a valid insignia. How many fully-formed insignias can we expect to see across all 6 faces of this agglomerate n^3 shape?
2. There are n total students getting a photo taken. Let n be a perfect square so that $n = m^2$, for some integer m. The n students are randomly organized into a grid of $m \times m$. The cameraman stands at the head on a slightly elevated pedestal. Consider student X_i; he can see the cameraman if everyone before him is at most as tall as he is.

   (a) There are only two heights a, b among all n students, where $a < b$. Each person has height a with probability p. If everyone remains standing, compute the number of people with an obstructed view of the cameraman. Consider an infinite number of students, so the grid is $m \times \infty$.

   (b) Consider the scenario in the previous part. This time, return to the original finite number of students, with an $m \times m$ grid.

   (c) Now, consider three heights a, b, c among all students, where $a < b < c$. Each person has height a with probability p, height b with probability q, and height c with probability r, where $p + q + r = 1$. Consider a single row of infinitely many students. How many students of height b do we expect to count, until we see a student of height c?

   (d) Now again consider the original scenario. If there are three heights, a, b, c among all n students, compute the number of people with an obstructed view of the cameraman.

3. Sinho is eating from a bag of pistachios, and every time step, he flips a coin. If the coin lands heads, he randomly picks a pistachio. If the pistachio has not been cracked, he will crack it and eat the nut. Regardless of the pistachio's state, he returns the shell to his bag. If the coin lands tails, he digs around until he finds an uneaten pistachio and eats it, again returning the shell to his bag. He begins with 500 nuts.

   (a) In terms of the number of eaten pistachios at the current time step, X_i, how many pistachios will Sinho have eaten at time $i + 1$?

   (b) After n steps, how many pistachios has Sinho eaten? Assume $n < 500$, so he could not have eaten all the nuts.

Chapter 3

Bernoulli Processes

3.1 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 solving probability questions quickly.

3.1.1 Bernoulli Distribution

A single trial, whose probability of success is p.

- Parameter: p - the probability of success
- Probability: p
- Mean: p
- Variance: $p(1-p)$

We can then consider multiple trials, which yields the binomial distribution below.

3.1.2 Binomial Distribution

Number of successes in n trials, where each independent trial has probability p of success. Alternatively, the probability of k successes.

- Parameters:
  - n - number of trials
  - p - probability of success for a given trial
- Probability (of k successes): $\binom{n}{k}p^k(1-p)^{n-k}$
- Mean: np
- Variance: $np(1-p)$

3.1.3 Geometric Distribution

Number of trials until the first success, where each trial is independent and each trial has probability p of success. Alternatively, the probability of k failures before one success.

- Parameters:
  - p - probability of success for a given, independent trial
- Probability (of k trials): $(1-p)^{k-1}p$
- Mean: $\frac{1}{p}$
- Variance: $\frac{1-p}{p^2}$

We can consider a sum of geometric random variables (e.g., successive attempts at rolling a 6)

3.1.4 Negative Binomial Distribution

(a.k.a., Pascal's) Properties of a sum of geometrically-distributed random variables. Number of times until ith success.

- Parameters:
  - p - probability of success for a given, independent trial
  - t - number of random variables to add
- Probability (of k successes in t time steps): $\binom{t-1}{k-1}p^k(1-p)^{t-k}$
- Mean: $\frac{t}{p}$
- Variance: $\frac{t(1-p)}{p^2}$

For the following, consider $Y = Y_1 + Y_2 \dots Y_k$ and $Z = Z_1 + Z_2 \dots + Z_l$, where each $Y_i \sim \text{Geom}(p)$ and $Z_i \sim \text{Geom}(q)$.

3.1.5 Merging

Examine two Bernoulli processes. We count success if we see *at least* one arrival. This means that $X = \text{MERGE}(Y, Z)$ and that $X_i \sim \text{Geom}(p + q - pq)$.

3.1.6 Splitting

Now, examine a single Bernoulli process. For every arrival we flip a coin with bias q. Consider it a part of process A with heads and B otherwise. This means that $A = \text{SPLIT}(Y)$ and that $A_i \sim \text{Geom}(pq)$ and $B \sim \text{Geom}(p(1-q))$.

3.1.7 Combining Distributions

The minimum across k geometric distributions, each with the same parameter p, is $X \sim \text{GEO}((1-p)^k)$.

3.2 Problems

1. Every 10 minutes, Bob and Alice deliberates whether or not to do their CS70 homework. With probability p, Bob is ready to work on homework. With probability q, Alice is ready to work on homework. Otherwise, at least one of them is on Facebook. If they are distracted for more than an hour, all hope is lost, and they do not continue working on homework.

   (a) If *both* of them are ready to do homework, they spend 10 minutes finishing one problem. Consider N, the number of homework problems that Bob and Alice complete. Find the PMF of N.

   (b) Under this strategy, where both must agree to work on homework, how long does it take for Bob and Alice to finish a homework with 10 problems?

   (c) Now, if one between the two are ready to do homework, he/she will convince the other to work on homework, and they spend 10 minutes finishing one problem. Find the new PMF of N.

   (d) Under this strategy, where at least one must be ready to work on homework, how long does it take Bob and Alice to finish a homework with 10 problems?

2. Alice has given up on Bob and is now working on the problem set alone. Every 10 minutes, Alice deliberates whether or not to work on her CS70 homework. She chooses to work on her CS70 homework with probability p and is otherwise distracted by Facebook.

   (a) We pick a problem i from the problem set uniformly at random. How long do we expect Alice to take finishing i and *after finishing* $i-1$? Assume Alice takes the full 10 minutes to complete a problem.

   (b) We pick a problem i from the problem set uniformly at random. How long do we expect Alice to have spent on Facebook *before starting* i and *after finishing* $i-1$? Again, assume Alice takes the full 10 minutes to complete a problem.

   (c) Say we pick a random point in time. At this point in time, Alice has finished $i-1$ problems. How long can we expect the length of that interval, starting from the completion of problem $i-1$ to the completion of problem i?

   (d) Assume the time it takes for Alice to complete the ith problem is $T_i \sim U[0, 10]$. Re-compute the expected amount of time Alice spends on Facebook in between *starting* problem i and *finishing* problem $i-1$.

3. Let X_i be the number of rolls you need to see the ith 6. Let Y_i be the number of rolls you need to see the ith 6 after rolling the $i-1$th 6, so $Y_i = X_i - X_{i-1}$. Compute the following quantities, keeping in mind that each $Y_i \sim \text{Geom}(\frac{1}{6})$.

   (a) Compute $\mathbb{E}[Y_1 + Y_2 + Y_3 | Y_1 + Y_2]$.

(b) Compute $\mathbb{E}[Y_1 + Y_2 | Y_1 + Y_2 + Y_3]$.
(c) Compute $\mathbb{E}[X_1 | X_2]$.
(d) Let $Z = \min(X_1, X_2)$. Compute $\mathbb{E}[\max(X_1, X_2) | Z]$
(e) Compute $\mathbb{E}[\min(Y_1, Y_2) | X_2]$.

Chapter 4

Poisson Processes

4.1 Guide

4.1.1 Poisson Distribution

Number of successes per unit time, space etc., if there are many trials and our average of successes per unit time, space etc. is λ.

Continuous analog for Binomial Distribution Note that this is the binomial distribution as $n \to \infty$; this approximation in general holds when $n \geq 100$ and $p \leq 0.01$.

- Parameters:
  - λ - average per unit time, space etc., the "rate"
- Probability (of k successes): $\frac{\lambda^k}{k!}e^{\lambda}$
- Mean: λ
- Variance: λ

4.1.2 Exponential Distribution

Number of unit time intervals until next success.

Continuous analog of Geometric Distribution

- Parameters:
  - λ - average per unit time, space, the "rate"
- Probability (of k successes): $\lambda e^{-\lambda x}$
- Mean: $\frac{1}{\lambda}$
- Variance: $\frac{1}{\lambda^2}$

4.1.3 Gamma Distribution

(a.k.a., Erlang) Properties of a sum of exponentially-distributed random variables. Number of times until ith success.

Continuous analog of Negative Binomial

- Parameters:
  - λ - average per unit time, space for each exponential
  - k - number of exponential distributions
- Mean: $\frac{k}{\lambda}$
- Variance: $\frac{k}{\lambda^2}$

For the following, consider $Y = Y_1 + Y_2 \dots Y_k$ and $Z = Z_1 + Z_2 \cdots + Z_l$, where each $Y_i \sim \text{Expo}(\lambda_1)$ and $Z_i \sim \text{Expo}(\lambda_2)$.

4.1.4 Merging

Examine two Poisson processes. We count success if we see *at least* one arrival. This means that $X = \textsc{Merge}(Y, Z)$ and that $X_i \sim \text{Expo}(\lambda_1 + \lambda_2)$.

4.1.5 Splitting

Examine a single Poisson process. For every arrival we flip a coin with bias p and . This means that $X = \textsc{Split}(Y)$ and that $X_i \sim \text{Expo}(\lambda_1 p)$.

4.1.6 Combining Distributions

The minimum across k exponential distributions, each with the same parameter λ, is $X \sim \text{Expo}(k\lambda)$.

4.2 Problems

1. Every week Bob receives λ surveys. Approximate to 50 weeks in a year. Each survey is emailed to *at least* 2,000 CS students, where each promises gift cards for K_i lucky winners, where $K_i \sim N(5, \sqrt{5})$ The actual value of K_i may be different for each survey, but note they are i.i.d.

   (a) Bob completes a survey with probability p. How many surveys do we expect Bob to complete in a year?

   (b) Assume that survey winners are picked uniformly at random. Give an upper bound for the number of surveys we expect him to win, in a year, given that he fills out surveys with probability p.

   (c) Assume each survey is sent to exactly 2,000 CS students. Compute the variance in the number of surveys that we expect Bob to win. Is this an upper bound or a lower bound, given that surveys in reality go to *at least* 2,000 students? Assume $p = 1, \lambda = 0.01$.

2. Consider CalCentric, a new piece of software that sees λ bug reports every minute. With probability p_1, production support marks a request as as `SEV-1`, with probability p_2 marks it as `SEV-2` and otherwise as `SEV-3`. Three teams, T_1, T_2, T_3 have been created to handle each level of urgency. Let Y_i be the number of minutes until the next request. Let Z_i be the number of requests per hour. Let X_{ij} represent the time until the jth request for team i. Assume all teams have zero bug reports at midnight.

   (a) Find the PMF of Z_i, Y_i, and X_{ij}.

   (b) We pick a bug uniformly at random. How long do we expect team i to process bug j?

   (c) We pick a time uniformly at random across all 24 hours; let us call this time t, which is measured in minutes since midnight. How many total bug reports do we expect to have on file? How many bug reports for team i? For team i, how much time do we expect between the last bug report and the next?

   (d) For team i, what is the PDF of the time it takes between bugs at time t? Let this time be T.

   (e) Given the first bug of the day for CalCentric as a whole comes at time t_1, at what time of day do we expect team 2 to receive its third bug t_3?

   (f) Given that CalCentric has seen b bugs and that team 1 has b_1 bug reports to handle, how many bug reports to we expect team 2 to have? Assuming b_2 is more than the current number of bug reports assigned to team 2, how much more time do we expect until team 2 has b_2 bug reports?

3. With probability $\frac{1}{3}$, Alice takes her motorcycle to school, which takes 5 minutes. With probability $\frac{1}{2}$, Alice takes her bike to school. Otherwise,

she takes 20 minutes to walk to school. Each motorcycle tire has been worn out and has an expected lifetime of 2 hours.

(a) How much time does Alice spend walking before the first time she rides her motorcycle?
(b) In expectation, after how many days will Alice pop a tire?
(c) Alice has lost her bike and walks with probability $\frac{2}{3}$. If there are 180 days of school, how many days will Alice walk to school? Assume that Alice does not have a tire replacement; if she pops a tire, she cannot use her motorcycle for the remaining days.

Chapter 5

Confidence Intervals

5.1 Guide

5.1.1 Markov's Inequality

Markov's inequality states the following. Remember that $\alpha > 0$.

$$\Pr(X \geq \alpha) \leq \frac{E[X]}{\alpha}$$

This is a one-line derivation, using expectation.

$$E[X] = \sum a \Pr(a) \geq \sum_{a>\alpha} a \Pr(a) \geq \alpha \sum_{a>\alpha} \Pr(a) \geq \alpha \Pr(a \geq \alpha)$$

We can then re-arrange to obtain our final result $\Pr(a \geq \alpha) \leq \frac{E[X]}{\alpha}$.

5.1.2 Chebyshev's Inequality

Chebyshev's inequality states the following. Intuitively, it is the probability that we are *more* than some distance α from the mean. Remember that $\alpha > 0$.

$$\Pr(|X - \mu| \geq \alpha) \leq \frac{\text{VAR}(X)}{\alpha^2}$$

5.1.3 Confidence Interval

Consider some random variable X, its mean μ, some positive α, and probability p. A **confidence interval** is an interval of α distance from μ that we know X has probability p of falling in.

Remember the distinction between observable values and non-observable values. We usually have some p that we'd like to estimate but cannot observe directly.

As a result, we observe some q, express it in terms of p and then solve for p. This is then our estimate for p.

5.1.4 Weak Law of Large Numbers

The **Law of Large Numbers** states that as $n \to \infty$, our estimate of the mean approaches the true mean. More formally, $\Pr(|A_n - \mu| \geq \alpha) \to 0$, where $A_n = \frac{\sum_{i=1}^{n} X_i}{n}$

5.2 Problems

1. Bob is hanging a string of Christmas lights up. With probability p, a lightbulb is dead. There are 500 total lightbulbs, and Bob will tolerate at most 10 dead lightbulbs. If we wish to be 95% certain that we have 10 or fewer dead lightbulbs, find a lower bound on p.
2. Derek wishes to protect user privacy whilst performing large-scale analytics. He has a sample size of 2,000 users, where each user clicks on an advertisement with probability p. Assume all users act independently. He decides to "jiggle" the data. Specifically, he replaces $\frac{2}{5}$ of the data with randomly-generated samples - "click" with probability $\frac{1}{2}$ and "no click" otherwise.

   (a) Let q be the probability of clicks that we observe. Find the estimate for p, $\hat{p}$, in terms of q.

   (b) Apply Chebyshev's to find a 95% confidence interval for p.

3. Alice also wishes to protect user privacy. She creates n different surveys that are kept with probability p. For each of the following, consider whether or not the Law of Large Numbers still holds, when we consider S_n to be the number of user responses considered, given n survey responses. Prove your conjecture. Assume that $k << n$.

   (a) Fixing a constant k and assigning k randomly-selected users to a randomly-selected survey.

   (b) Fixing a constant k and for each group of $\frac{n}{k}$ assign to a random survey.

Chapter 6

Markov Chains

6.1 Guide

6.1.1 Definition

A Markov chain is a set of states X_i, where each X_{i+1} only depends on X_i. In this class, we only consider Markov Chains with finitely many states.

- P is the **transition matrix**. $P(i, j)$ gives us the probability of a transition from X_i to X_j.
- The current state of a Markov Chain or **distribution** (i.e., values at all the nodes) at time t is represented using π_t. So, $\pi_{t+1} = \pi_t P$.
- All values must sum to 1.

$$\sum_i \pi(i) = 1$$

- All transition probabilities for a single destination must sum to 1.

$$\sum_j \Pr(i, j) = 1$$

6.1.2 Irreducibility

A Markov Chain is **irreducible** iff there exists some path between every pair of states. (i.e., for each state, all other states are "reachable"). Note this means the Markov Chain must be a single connected component.

6.1.3 Periodicity

A state X_i is **aperiodic** if the length of all paths starting at X_i and ending at X_i has GCD 1. More formally,

$$d(i) := \text{G.C.D.}\{n > 0 | P^n(i,i) = \Pr(X_n = i | X_0 = i) > 0\}, i \in X$$

X_i is aperiodic if $d(i) = 1$.

A Markov Chain is aperiodic if all of its states are aperiodic.

6.1.4 Invariant Distributions

- An **invariant distribution** π with transition matrix P is a distribution such that $\pi = \pi P$.
- An irreducible Markov Chain always has a unique invariant distribution.

To solve for the **invariant distribution**, do the following:

1. Check that the Markov Chain is irreducible. Only then can we guarantee that there exists a unique invariant distribution. Note that invariant distributions may exist for other Markov Chains but *are not guaranteed.*
2. Write out all balance equations. As it turns out, this system of equations is dependent.
3. So, replace one equation with the requirement that all $\pi(i)$ sum to 1. Explicitly, remove an arbitrary balance equation, and add the equation $\pi(1) + \cdots + \pi(n) = \sum_{i=1}^{n} \pi(i) = 1$.
4. Write the balance equations as a matrix P. Solve the augmented matrix $[P|\pi]$.

6.1.5 Balance Equations

Balance equations specify transitions for a Markov Chain. Let $\pi(j)$ denote the value of state j. We express $\pi(j)$ in terms of all possible paths from i to j. So $\pi(j) = \sum_i P(i,j)\pi(i)$.

6.1.6 Hitting Time

To solve a **hitting time** problem, write out your balance equations and solve the linear system of equations. Remember that we consider the set of possible *destinations.* In other words,

$$\beta(i) = p\beta(i-1) + (1-p)\beta(i-2) + 1$$

For the above equation, this means that from state i we have probability p of reaching state $i-1$ next and probability $1-p$ of reaching state $i-2$. We assume that this transition takes "time" 1. Note that this transition could take a variable amount of time. In which case, add the amount of time the transition takes, instead of 1.

6.1.7 Probability of A before B

To solve **probability of A before B**, consider the probability of reaching any state in A from states neither in A nor in B. Then, consider the probability of reaching B from states in A. We again consider the set of possible *destinations*. Note there is no extra term.

$$\alpha(i) = p\alpha(i-1) + (1-p)\alpha(i-2)$$

The above means that from state i, we have probability p of reaching state $i-1$ and probability $1-p$ of reaching $i-2$.

6.2 Problems

1. Bob is walking along a small bridge, 3 feet wide and 10 feet long. Every time step, he walks 1 foot along the bridge. With probability p, he walks straight forward, and otherwise, he takes one step 1 foot forward and 1 foot to the right with probability $\frac{1}{2}$, forward 1 foot and left 1 foot with probability $\frac{1}{2}$. If he walks off the bridge, he falls into the water and foregoes the bridge to get to the other side. How many time steps do we expect Bob to walk before falling in the water, assuming he starts at the center of the bridge?
2. Let us consider the following scenarios. Use Markov Chains only when necessary, but consider your intuition first. Justify all of your answers.

   (a) Consider a coin with bias p. What is the probability that we see TH before HH?

   (b) Consider a fair dice. What is the probability that we see 5-3 before seeing 3-5?

3. Consider a die.

   (a) Let this die be fair. How much time do we expect until we roll 5 dots, followed by a roll of 3 dots?

   (b) How many 5s do we expect to see before we roll 5 dots, then 3 dots?

   (c) Let this die be loaded, favoring 5 with probability $\frac{6}{11}$. All others are have equal probability, or $\frac{1}{11}$. How much time do we expect until we roll 5 dots, followed by a roll of 3 dots?

   (d) Again with the loaded die, how many 5s do we expect to see before our first sequence of 5-3? (Note that we cannot simply take the answer from the previous part and multiply by $\frac{1}{6}$. This is because the duration of survival is positively correlated with the number of 5s we expect to see.)

Chapter 7

Transformations

7.1 Guide

7.1.1 Moment Generating Function

The MGF of a random variable X is

$$M_X(t) = \mathbb{E}[e^{tX}]$$

It is always true that $M_X(0) = 1$.

7.1.2 Expectation

We can relate this ith derivative of the MGF to the ith moment of X.

$$\frac{\partial^k t}{\partial t^k} M_X(0) = \mathbb{E}[X^k]$$

7.1.3 Affine Transformation of Random Variable

Consider random variables X, Y and constants a, b such that $Y = aX + b$.

$$M_Y(s) = \mathbb{E}[e^{s(aX+b)}] = \mathbb{E}[e^{saX+sb}] = e^{sb}\mathbb{E}[e^{saX}] = e^{sb}M_X(sa)$$

7.1.4 Sum of Independent Random Variables

Consider three random variables Z, X, Y such that $Z = X + Y$ and X, Y are independent.

$$M_Z(s) = M_{X+Y}(s) = \mathbb{E}[e^{s(X+Y)}] = \mathbb{E}[e^{sX}e^{sY}] = \mathbb{E}[e^{sX}]\mathbb{E}[e^{sY}] = M_X(s)M_Y(s)$$

7.1.5 Combining Distributions

For a random variable Y and various random variables $X_1, \dots X_n$, then

$$f_Y(y) = p_1 f_{X_1}(y) + \cdots + p_n f_{X_n}(y)$$

gives us the following.

$$M_Y(s) = p_1 M_{X_1}(s) + \cdots + p_n M_{X_n}(s)$$

7.1.6 Inversion Property

"The transform $M_X(s)$ associated with a random variable X uniquely determines the CDF of X, assuming that $M_X(s)$ is finite for all s in some interval $[-a, a]$, where a is a positive number."

7.1.7 Distributions

- Bernoulli(p): $M_X(s) = 1 - p + pe^s$
- Binomial(n, p): $M_X(s) = (1 - p + pe^s)^n$
- Geometric(p): $M_X(s) = \frac{pe^s}{1-(1-p)e^s}$
- Poisson(λ): $M_X(s) = e^{\lambda(e^s-1)}$
- Uniform(a, b) (discrete): $M_X(s) = \frac{e^{sa}(e^{s(b-a+1)}-1)}{(b-a+1)(e^s-1)}$
- Uniform(a, b) (continuous): $M_X(s) = \frac{e^{sb}-e^{sa}}{s(b-a)}$
- Exponential(λ): $M_X(s) = \frac{\lambda}{\lambda-s}$, where $s < \lambda$
- Normal(μ, σ^2): $M_X(s) = e^{\sigma^2 s^2/2)+\mu s}$

7.2 Problems

1. Find the PDF of Z given the following moment-generating function of z.

$$M_Z(s) = \frac{2}{3}\frac{5}{5-s} + \frac{1}{3}\frac{e^{s2} - e^s}{s}$$

2. Let X be the time until lightbulb 1 dies, where its average lifetime is λ_1. Let Y be the time until lightbulb 2 dies, where its average lifetime is λ_2. Without applying linearity of expectation, find the expectation of $Z = X + Y$.

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 Probability

1. Consider a housing district, represented by a 8×8 grid. Tom is currently at the bottom-left, $(0, 0)$ and would like to get home to $(7, 7)$. If he goes up with probability p and right with probability $1 - p$, what is the probability he makes it home without visiting the graveyard at $(5, 6)$ or school at $(4, 7)$? (Note that by design, it is impossible to reach both the graveyard and school. This is to make the problem slightly simpler.)

   Solution: Consider all possible ways to get to $(7, 7)$. We can view this as 16 actions, where we pick 8 to go up, $\binom{16}{8}$. We note that all paths will necessarily take 8 blocks up and 8 blocks to the right, so it is not necessary to consider the probability of taking a particular path.

   Now, consider all the ways to get to $(5, 6)$, which by the same logic is $\binom{11}{5}$. From $(5, 6)$, the number of ways to get to $(7, 7)$ is $\binom{3}{2}$. So, the number of ways to walk through the graveyard is $a = \binom{11}{5}\binom{3}{2}$. By the same logic, the number of paths passing through $(4, 7)$ is $b = \binom{11}{4}\binom{3}{3}$.

   In sum, we have the following

$$1 - \frac{a+b}{\binom{16}{8}} = 1 - \frac{\binom{11}{5}\binom{3}{2} + \binom{11}{4}}{\binom{16}{8}}$$

2. Consider a game of cards with 3 other friends. You and your friends draw cards from a standard 52-card deck. Each player receives a random 4 cards randomly, and each player is assigned an ace. In total, each player has 5 cards.

   (a) One of your friends, Bob, claims he holds a full house. What is the probability he is bluffing? Recall that a full house is a hand containing two cards of one rank and three cards of another rank.

   Solution: 0. To form a full house, either your friend has three aces or two aces. Neither is possible, because the four aces are distributed evenly, one to each friend.

   (b) Each player randomly draws an additional card at random from the deck. Bob claims he *now* holds a full house. What is the probability he is bluffing?

   Solution: We consider the probability that your friend *has* a full house and take the complement of that event.

$$1 - \Pr(FH)$$

   To compute the probability of a full house, we consider the possible full houses. Recall that none of these full houses can include aces, so we discount rank 1. We first pick our ranks: from the 12 remaining ranks, we pick 2. $\binom{12}{2}$. Then, we pick our suits. For each card, there are a total of 4 different suits. 4^5. In total, we have $\binom{48}{4}$ different possibilities, as we are considering the cards after distributing aces.

$$1 - \frac{\binom{12}{2}4^5}{\binom{48}{4}}$$

(c) The dealer verifies Bob's claim. Bob now claims that the sum of the values for his cards is at least 13. What is the probability that Bob is telling the truth?

Solution: We compute the probability that the sum is at most 12. This is the complement of achieving a sum at least 13.

$$\Pr(X \leq 13) = 1 - \Pr(X < 13)$$

Note there is only one possible way to achieve < 13. Since he has an ace, the values for cards involved in his full house must sum to at most 12. Considering that full house cannot involve his ace, he must have 3 2s and 2 3s. Note he cannot achieve any lower value with a full house and extra ace.

$$1 - \frac{4^5}{\binom{48}{4}}$$

3. Consider a loaded 6-sided die, where just one number is favored with probability p. The remaining numbers have equal probability $\frac{1-p}{5}$. You and five other friends must decide who will take out the trash today, for your shared apartment. Develop a scheme that allows you to pick a victim, uniformly at random.

   Solution: Roll until you see six rolls in a row with unique values. If any values repeat, restart the sequence. This ensures that we have outcomes of equal probability. Specifically, any sequence of 6 unique values has probability $p(\frac{1-p}{5})^5$ of occurring.

   The issue is that we now have 6! orderings of 6 distinct values. We need to translate this into one of 6 unique outcomes. To do this, we arbitrarily pick and fix an index to consider. Say, we consider the first number in the sequence. The value at the index will determine the value of that sequence. So, if we roll, $1-2-3-4-5-6$, then the value of that permutation is 1.

   Alternate Solution

   We can consider only the following permutations of values. Again, we discard any permutation that does not match one of the following. Make the first roll the value of our outcome:

   - 1-2-3-4-5-6

- 2-3-4-5-6-1
- 3-4-5-6-1-2

⋮

-

We see that each case has the same probability of occurring, and since we only consider these 6 outcomes, we have a $\frac{1}{6}$ chance of achieving any value in $\{1, 2, 3, 4, 5, 6\}$.

4. Consider a standard 52-card deck. You are assigned a standard 5-card hand, where each card is drawn randomly from the deck.

   (a) What is the probability that we have exactly 1 ace?

   Solution: It is easier to reason about this using counting. We pick an ace and then we pick from the non-Ace cards.

$$\frac{\binom{4}{1}\binom{48}{4}}{\binom{52}{5}}$$

   Alternatively, we can consider the possibility of picking an Ace $\frac{4}{52}$ and then picking four non-Ace cards, $\frac{48}{51}\frac{47}{50}\frac{46}{49}\frac{45}{48}$. Multiply by the number of ways to place the Ace among 5 cards, which is $\binom{5}{1} = 5$.

$$\frac{5 * 4 * 48 * 47 * 46 * 45}{52 * 51 * 50 * 49 * 48}$$

   (b) What is the probability that we have exactly 3 clubs? (Hint: Use counting)

   Solution: It is easier to reason about this using counting. We pick our three clubs. Since there are 13, we choose 3 from 13, and since there are 52 - 13 = 39 non-club cards, we pick 2 from 39.

$$\frac{\binom{13}{3}\binom{39}{2}}{\binom{52}{5}}$$

   (c) Given we have three clubs, what is the probability of an ace of clubs?

   Solution: In the denominator, we count all ways to pick three clubs, which is the numerator from part 2.

$$\binom{13}{3}\binom{39}{2}$$

   In the numerator, we count all ways to pick a 5-card hand with an ace of clubs. Given that our ace is a club, we only have two more clubs to choose and two non-clubs to choose.

$$\binom{12}{2}\binom{39}{2}$$

Thus, our final answer is the following.

$$\frac{\binom{12}{2}\binom{39}{2}}{\binom{13}{3}\binom{39}{2}} = \frac{3}{13}$$

(d) What is the probability that we have 3 clubs or 1 ace? (Not XOR, Hint: Think about inclusion-exclusion.)

Solution: We need to apply inclusion-exclusion. Recall that this states for two events A and B,

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$$

We have that A is from part 1 and B is from part 2. All that remains is to compute $\Pr(A \cap B)$ which is the probability of 3 clubs *and* exactly one ace. We have two cases; either the ace is a club or the ace is not a club. Let C be the event that the ace is a club. By some variant of the law of total probability:

$$\Pr(A \cap B) = \Pr(A \cap B \cap C) + \Pr(A \cap B \cap \bar{C})$$

We now compute both probabilities. Pick the ace of clubs, $\binom{1}{1}$. We pick our two non-ace clubs from 12 non-ace clubs $\binom{12}{2}$. Finally, we pick our non-ace, non-club from $52 - 4 - 13 + 1 = 36$.

$$\Pr(A \cap B \cap C) = \frac{\binom{12}{2}\binom{36}{2}}{\binom{52}{5}}$$

Given that our ace is *not* a club, we have all three non-Ace clubs to choose ($\binom{12}{3}$), not to mention a non-club, non-ace and a non-club ace card. We know there are $52 - 13 - 3 = 36$ non-club, non-ace cards and there are 3 non-club, ace cards. Thus, we have the following.

$$\Pr(A \cap B \cap \bar{C}) = \frac{\binom{12}{3}\binom{36}{1}\binom{3}{1}}{\binom{52}{5}}$$

Thus, we can combine these to get the following for $\Pr(A \cap B)$.

$$\frac{\binom{12}{2}\binom{36}{2}}{\binom{52}{5}} + \frac{\binom{12}{3}\binom{36}{1}\binom{3}{1}}{\binom{52}{5}}$$

We then have our final expression for $\Pr(A \cup B)$.

$$\frac{\binom{4}{1}\binom{48}{4}}{\binom{52}{5}} + \frac{\binom{13}{3}\binom{39}{2}}{\binom{52}{5}} - \left(\frac{\binom{12}{2}\binom{36}{2}}{\binom{52}{5}} + \frac{\binom{12}{3}\binom{36}{1}\binom{3}{1}}{\binom{52}{5}}\right)$$

5. Siqi purchases s packets of strawberry Pocky (S), c packets of chocolate Pocky (C), and m packets of mint Pocky (M), making n total packets of Pocky, for her friend Tyler. He has since given in to temptation and has elected to begin eating Pocky.

   (a) Tyler randomly pulls a Pocky packet out, then places the packet back inside his bag before drawing the next one. He repeats this m times. What is the probability that he sees exactly k chocolate packets?

   Solution: Let C be the number of chocolate packets that Tyler sees. Note that C is binomially distributed with m trials and the probability $\frac{c}{n}$ of successfully picking a chocolate packet on a given trial, $C \sim \textsc{Bin}(m, \frac{c}{n})$. We can intuit about the probability as well. First compute the probability. We have k successes, or $(\frac{c}{n})^k$ and $m-k$ failures, so $(1-\frac{c}{n})^{m-k}$. We then compute the number of ways to distribute k successes among m trials, which is $\binom{m}{k}$.

$$\binom{m}{k}(\frac{c}{n})^k(1-\frac{c}{n})^{m-k}$$

   (b) Forest randomly steals k packets from Tyler's bag of Pocky. What is the probability that Forest has at least one chocolate packet?

   Solution: Instead, let us consider the complement of this event. Let F be the number of chocolate Pocky packets that Forest has. Specifically, we consider the case where Forest has *zero* chocolate packets.

$$\Pr(F \geq 1) = 1 - \Pr(F < 1) = 1 - \Pr(F = 0)$$

   Note that the F is *not* binomially distributed, because the trials are not independent. Instead, we have the following probability that $F = 0$:

$$\frac{s+m}{n}\frac{s+m-1}{n-1}\cdots\frac{s+m-(k-1)}{n-(k-1)} = \frac{(s+m)!(n-k)!}{(s+m-k)!n!}$$

   Thus, we have our final answer:

$$1 - \frac{(s+m)!(n-k)!}{(s+m-k)!n!}$$

   Takeaway: Consider the complement.

   (c) Given that Tyler has picked six packets of Pocky in sequential order, what is the probability that Tyler picks S, C, S, C in that order? (Note that interspersing other packets with this packet is also valid, so $S[M]CS[S]C$ would also satisfy this condition.)

   Solution: Note that the two extra packets can be any flavor. Thus, we are implicitly multiplying by 1^2. It also does not matter where our sequence begins or ends, by symmetry. However, we need to pick the slots that our pattern occupies. $\binom{6}{4}$. We then consider the probability that our specific permutation occurs.

Let X_a be the event that we find a strawberry packet, X_b be the event we find a chocolate packet next, etc. We are thus interested in computing the following.

$$Pr[X_a \cap X_b \cap X_c \cap X_d]$$

The probability of a strawberry packet at some position is $\frac{s}{n}$. There are now $n-1$ packets remaining, of which $s-1$ are strawberry.

$$Pr[X_a] = \frac{s}{n}$$

The probability of a chocolate packet at some later position, regardless of which position it is, is $\frac{c}{n-1}$. There are now $n-2$ coins remaining, of which $c-1$ are pennies.

$$\Pr[X_b|X_a] = \frac{c}{n-1}$$

The probability of another strawberry at some later position is $\frac{s-1}{n-2}$. There are now $n-3$ coins remaining.

$$Pr[X_c|X_a, X_b] = \frac{s-1}{n-2}$$

The probability of another chocolate is $\frac{c-1}{n-3}$.

$$Pr[X_d|X_a, X_b, X_c] = \frac{c-1}{n-3}$$

Since all samples are made at random, we know the following holds.

$$\begin{aligned}
Pr[X_a \cap X_b \cap X_c \cap X_d] &= Pr[X_a]Pr[X_b|X_a]Pr[X_c|X_a, X_b]Pr[X_d|X_a, X_b, X_c] \\
&= \binom{6}{4}\frac{s}{n}\frac{c}{n-1}\frac{s-1}{n-2}\frac{c-1}{n-3} \\
&= \binom{6}{4}\frac{s(s-1)c(c-1)(n-4)!}{n!}
\end{aligned}$$

Takeaway: Apply symmetry when we have no additional information.

6. You are rolling a 100-sided die. Across n trials, let X be the number of rolls with an odd number of dots. Let Y be the number of rolls with an even number of dots.

(a) Assume n is divisible by 4. What is $\Pr(Y - X \geq \frac{n}{2})$?

Solution: Note that $X + Y = n$, so $Y - X = n - 2X$.

$$\begin{aligned}
\Pr(Y - X \geq \frac{n}{2}) &= \Pr(n - 2X \geq \frac{n}{2}) \\
&= \Pr(-2X \geq -\frac{n}{2}) \\
&= \Pr(X \leq \frac{n}{4})
\end{aligned}$$

Consider that X is binomially distributed, with n trials, where the success of each trial is $\frac{1}{2}$, $X \sim \textsc{Bin}(n, \frac{1}{2})$.

$$\Pr(X \leq \frac{n}{4}) = \sum_{i=1}^{n/4} \Pr(X = i) = \sum_{i=1}^{n/4} \binom{n}{i} \frac{1}{2^n} = \frac{1}{2^n} \sum_{i=1}^{n/4} \binom{n}{i}$$

(b) Assume n is odd. What is $\Pr(X > Y)$?

Solution: Intuitively, this is $\frac{1}{2}$. Note that for every combination where $X > Y$, we can swap all odd with even and even with odd, for another combination where $Y > X$. Thus, the number of permutations where $X > Y$ must be the same as the number of permutations where $Y > X$, so

$$\Pr(X > Y) = \Pr(Y > X)$$

We additionally know that since n is odd, these are the only two possibilities. Thus,

$$\Pr(X > Y) + \Pr(Y > X) = 1$$

Finally, this means that

$$\Pr(X > Y) = \Pr(Y > X) = \frac{1}{2}$$

Takeaway: Consider intuitive approaches before diving into the math.

8.2 Expectation, Variance, Covariance

1. Consider a new, innovative textile, a 6-sided cube. (Think of it as a 6-sided stamp) We place 37^3 textiles in an $37 \times 37 \times 37$ cube, orienting these cubes randomly. Four textiles together, oriented in the correct direction will print an insignia. The insignia is symmetric across both axes, all cubes are identical, and each textile has 4 printable surfaces of 6 sides. e.g., If either of the 2 remaining faces is facing up, that textile will not contribute to a valid insignia. How many fully-formed insignias can we expect to see across all 6 faces of this agglomerate n^3 shape?

   Solution: Let X_i be the number of insignias we observe on face i. For each face, consider the interior vertices. For example, a 3×3 (think tic-tac-toe grid) would have 4 interior vertices. More generally an $n \times n$ grid has a $(n-1)^2$ interior vertices. Let X_{ij} be the jth interior vertex for the ith face, which takes on value 1 of the jth interior vertex is the center of a valid insignia. For any indicator random variable, we know that

   $$E[X_{ij}] = \Pr(X_{ij} = 1)$$

   We consider the probability that a valid insignia is formed at vertex j. All cubes must have a valid face showing, making $(\frac{4}{6})^4 = (\frac{2}{3})^4$. Each face must also be rotated in the correct direction, which has probability $(\frac{1}{4})^4$ of occurring. Thus, we have that

   $$\Pr(X_{ij} = 1) = \frac{1}{6^4}$$

   For each face, we have $(n-1)^2 = 36^2 = 6^4$ interior vertices, and we have a total of 6 faces. Thus, we have

   $$E[X] = 6E[X_i] = 6(6^4)E[X_{ij}] = 6(6^4)\frac{1}{6^4} = 6$$

2. There are n total students getting a photo taken. Let n be a perfect square so that $n = m^2$, for some integer m. The n students are randomly organized into a grid of $m \times m$. The cameraman stands at the head on a slightly elevated pedestal. Consider student X_i; he can see the cameraman if everyone before him is at most as tall as he is.

   (a) There are only two heights a, b among all n students, where $a < b$. Each person has height a with probability p. If everyone remains standing, compute the number of people with an obstructed view of the cameraman. Consider an infinite number of students, so the grid is $m \times \infty$.

   Solution: Let X be the total number of students that can see the cameraman, where X_i is the number students *in column* i that can see the cameraman. We see that X_i is geometrically distributed with parameter $1-p$, with a maximum value of m, as X_i is the number of students of height a that stand before a student with height b, $X_i \sim \text{Geom}(1-p)$.

$$E[X] = mE[X_i] = \frac{m}{1-p}$$

(b) Consider the scenario in the previous part. This time, return to the original finite number of students, with an $m \times m$ grid.

Solution: Let X be the total number of students that can see the cameraman, where X_i is the number students *in column* i that can see the cameraman. We must split once more where X_{ij} denotes if the jth student in the ith row can see the cameraman. If the student has height a, then the student can see if everyone in front has height a. If the student has height b, the heights of students before do not matter.

$$\mathbb{E}[X_{ij}] = \Pr(X_{ij} = 1) = (p)p^{j-1} + (1-p)$$

Apply linearity of expectation twice:

$$\mathbb{E}[X] = m\mathbb{E}[X_i] = m\sum_j \mathbb{E}[X_j] = m\sum_j (p^j + (1-p))$$

(c) Now, consider three heights a, b, c among all students, where $a < b < c$. Each person has height a with probability p, height b with probability q, and height c with probability r, where $p + q + r = 1$. Consider a single row of infinitely many students. How many students of height b do we expect to count, until we see a student of height c?

Solution: Let N be the number of students we see before the first student of height c, including the student of height c. Since height c occurs with probability r, we have that $N \sim \text{Geom}(r)$. Let B be the number of students of height b that we see. Note $B \sim \text{Bin}(N-1, \frac{q}{p+q})$. Thus, we apply law of iterated expectations and then solve using conditional expectation.

$$E[B] = E[E[B|N]] = E[(N-1)\frac{q}{p+q}] = \frac{q}{p+q}(\frac{1}{r} - 1) = \frac{q(1-r)}{p+q}$$

(d) Now again consider the original scenario. If there are three heights, a, b, c among all n students, compute the number of people with an obstructed view of the cameraman.

Solution: As in part a, let X_{ij} denote whether or not the jth person can see. Note that if the jth student has height a, then all students before must have height a as well. If the jth student, all students before must have height

$$X_{ij} = p^j + q(p+q)^{j-1} + r$$

Again, apply linearity of expectation twice.

$$\mathbb{E}[X] = m\mathbb{E}[X_i] = m\sum_j \mathbb{E}[X_{ij}] = m\sum_j (p^j + q(p+q)^{j-1} + r)$$

3. Sinho is eating from a bag of pistachios, and every time step, he flips a coin. If the coin lands heads, he randomly picks a pistachio. If the pistachio has not been cracked, he will crack it and eat the nut. Regardless of the pistachio's state, he returns the shell to his bag. If the coin lands tails, he digs around until he finds an uneaten pistachio and eats it, again returning the shell to his bag. He begins with 500 nuts.

   (a) In terms of the number of eaten pistachios at the current time step, X_i, how many pistachios will Sinho have eaten at time $i+1$?

   Solution: With probability $\frac{1}{2}$, Sinho pulls a random pistachios from the bag. With probability $\frac{X_i}{500}$, Sinho picks an already-eaten pistachio. Otherwise, he picks an uneaten pistachio, eats it, and returns it to the bag, incrementing the number of eaten pistachios by 1.

   With probability $\frac{1}{2}$, Sinho digs around to find an uneaten pistachio and definitely increases the number of eaten pistachios.

$$\begin{aligned} E[X_{i+1}|X_i] &= \frac{1}{2}(\frac{X_i}{500}X_i + (1 - \frac{X_i}{500})(X_i+1)) + \frac{1}{2}(X_i+1) \\ &= \frac{1}{2}(\frac{X_i}{500}X_i + X_i + 1 - \frac{X_i}{500}X_i - \frac{X_i}{500}) + \frac{1}{2}(X_i+1) \\ &= \frac{1}{2}(\frac{499X_i}{500} + 1) + \frac{1}{2}(X_i+1) \\ &= \frac{999}{1000}X_i + 1 \end{aligned}$$

   (b) After n steps, how many pistachios has Sinho eaten? Assume $n < 500$, so he could not have eaten all the nuts.

   Solution:

   By the law of iterated expectation, we have that

$$E[X_n] = E[E[X_n|X_{n-1}]] = E[\frac{999}{1000}X_{n-1} + 1]$$

   We would like to re-express this in terms of $E[X_0] = 500$. We know that a system of the form $X(t+1) = \alpha X(t) + \beta$ has the following solution:

$$X(t) = \alpha^t X(0) + \beta(\frac{1-\alpha^{t-1}}{1-\alpha})$$

Thus, we can plug in

$$E[X_n] = (\frac{999}{1000})^n X_0 + \frac{1-(\frac{999}{1000})^{t-1}}{\frac{1}{1000}} = (\frac{999}{1000})^n 500 + 1000(1-(\frac{999}{1000})^{t-1})$$

8.3 Bernoulli Processes

1. Every 10 minutes, Bob and Alice deliberates whether or not to do their CS70 homework. With probability p, Bob is ready to work on homework. With probability q, Alice is ready to work on homework. Otherwise, at least one of them is on Facebook. If they are distracted for more than an hour, all hope is lost, and they do not continue working on homework.

   (a) If *both* of them are ready to do homework, they spend 10 minutes finishing one problem. Consider N, the number of homework problems that Bob and Alice complete. Find the PMF of N.

   Solution: This is a merging of Bernoulli processes.

   Consider the time it takes to complete a single homework problem. Since both Bob and Alice must agree to work on the problem, we note that X_i is geometrically distributed with probability pq, $X_i \sim \text{Geom}(pq)$. Additionally, an hour is comprised of 6 10-minute intervals, thus we are interested in $\Pr(X_i \leq 6) = 1 - \Pr(X_i > 6) = 1 - (1-pq)^6$, which gives us the probability that Bob and Alice finish another problem within an hour. Note that to terminate the process, we cannot have Bob and Alice focused for any 10-minute interval, $(1-pq)^6$.

   $$\Pr(N = k) = (1 - (1-pq)^6)^k (1-pq)^6$$

   (b) Under this strategy, where both must agree to work on homework, how long does it take for Bob and Alice to finish a homework with 10 problems?

   Solution: It will take infinite time, as we do not exclude the possibility that Alice and Bob lose all hope.

   (c) Now, if one between the two are ready to do homework, he/she will convince the other to work on homework, and they spend 10 minutes finishing one problem. Find the new PMF of N.

   Solution: This is another merging of of Bernoulli processes.

   Consider the time it takes to finish a single homework problem. At least one of Bob or Alice must agree to work, so we have $1-(1-p)(1-q) = p+q-pq$, or $X_i \sim \text{Geom}(1-(1-p)(1-q))$. We are again interested in $\Pr(X_i \leq 6) = 1 - \Pr(X_i > 6) = 1 - ((1-p)(1-q))^6$. Again, to terminate the process, both Bob and Alice must be distracted for the full hour, $((1-p)(1-q))^6$. We get an answer very similar to one from part a.

   $$\Pr(N = k) = (1 - ((1-p)(1-q))^6)^k ((1-p)(1-q))^6$$

(d) Under this strategy, where at least one must be ready to work on homework, how long does it take Bob and Alice to finish a homework with 10 problems?

Solution: Again, it will take infinite time, as we do not exclude the possibility that Alice and Bob lose all hope.

2. Alice has given up on Bob and is now working on the problem set alone. Every 10 minutes, Alice deliberates whether or not to work on her CS70 homework. She chooses to work on her CS70 homework with probability p and is otherwise distracted by Facebook.

(a) We pick a problem i from the problem set uniformly at random. How long do we expect Alice to take finishing i and *after finishing* $i-1$? Assume Alice takes the full 10 minutes to complete a problem.

Solution: First, consider the amount of time it takes Alice to finish her ith problem after finishing the $i-1$th problem. We can appeal to the memoryless property of the Geometric distribution to ignore the time it took her to finish the $i-1$ problems beforehand. Instead, we focus on the time from $i-1$ to i. This time, T_i, *including the 10 minutes it takes to finish the ith problem*, can be modeled as a geometric random variable with parameter p.

$$T_i \sim \text{Geom}(p)$$

Thus, take the expectation to get the number of 10-minute intervals Alice takes. Multiply by 10 to get the total amount of time.

$$\mathbb{E}[T_i] = \frac{10}{p} \text{ minutes}$$

(b) We pick a problem i from the problem set uniformly at random. How long do we expect Alice to have spent on Facebook *before starting* i and *after finishing* $i-1$? Again, assume Alice takes the full 10 minutes to complete a problem.

Solution: From the previous part, we have that $T_i \sim \text{Geom}(p)$. However, we need to subtract the time it takes to finish the last problem, to get the time that Alice is distracted by Facebook, between the i and $i-1$ problems.

Let Alice's duration of *distraction* be modeled by a shifted geometric random variable. $T \sim \text{Geom}(p) - 1$, as we ignore the 10 minutes it takes to finish the ith problem. After the $i-1$th problem, we expect

$$\mathbb{E}[T] = \mathbb{E}[\text{Geom}(p)] - 1 = \frac{1}{p} - 1$$

This translates into

$$\mathbb{E}[T] = 10(\frac{1}{p} - 1) \text{ minutes}$$

(c) Say we pick a random point in time. At this point in time, Alice has finished $i-1$ problems. How long can we expect the length of that interval, starting from the completion of problem $i-1$ to the completion of problem i?

Solution: This is a tidbit of renewal theory. We invoke the Inspector's Paradox, as we note that if we pick points in time, we are more likely to land in a longer interval than we are to land in a shorter interval. Thus, this is *not* just the expected length of an interval. Instead, we see that there is $\frac{1}{p}$ until the next completed problem, and $\frac{1}{p} - 1$ from the last completed problem. In sum, the length of the interval is thus

$$\frac{2}{p} - 1$$

(d) Assume the time it takes for Alice to complete the ith problem is $T_i \sim U[0, 10]$. Re-compute the expected amount of time Alice spends on Facebook in between *starting* problem i and *finishing* problem $i-1$.

Solution: We simply take the expectation from the previous part and add the amount of time after Alice finishes a problem and the start of the next 10-minute interval.

$$\mathbb{E}[T] = 10(\frac{1}{p} - 1) + \mathbb{E}[T_i] = \frac{10}{p} - 5$$

3. Let X_i be the number of rolls you need to see the ith 6. Let Y_i be the number of rolls you need to see the ith 6 after rolling the $i-1$th 6, so $Y_i = X_i - X_{i-1}$. Compute the following quantities, keeping in mind that each $Y_i \sim \text{Geom}(\frac{1}{6})$.

(a) Compute $\mathbb{E}[Y_1 + Y_2 + Y_3 | Y_1 + Y_2]$.

Solution: Intuitively, we know that $Y_1 + Y_2$ is a known constant, and we only need to compute $\mathbb{E}[Y_3] = 6$, since all $Y_i \sim \text{Geom}(\frac{1}{6})$.

$$\mathbb{E}[Y_1 + Y_2 + Y_3 | Y_1 + Y_2] = Y_1 + Y_2 + 6$$

We can see this using linearity of expectation explicitly.

$$\mathbb{E}[Y_1 + Y_2 | Y_1 + Y_2] + \mathbb{E}[Y_3 | Y_1 + Y_2] = Y_1 + Y_2 + \mathbb{E}[Y_3] = Y_1 + Y_2 + 6$$

(b) Compute $\mathbb{E}[Y_1 + Y_2 | Y_1 + Y_2 + Y_3]$.

Solution: We know that Y_i are all i.i.d., so we expect the following, by linearity of expectation:

$$\mathbb{E}[Y_1 + Y_2 | Y_1 + Y_2 + Y_3] = \frac{2}{3}(Y_1 + Y_2 + Y_3)$$

(c) Compute $\mathbb{E}[X_1 | X_2]$.

Solution: We can view the sum total and consider the the geometric in *reverse*. From the end of the second 6, we expect that the time of the last arrival for a Geom(p) is $\frac{1}{p} - 1$. In our case, we have Geom($\frac{1}{6}$), so the time since the last problem is 5. This makes

$$\mathbb{E}[X_1 | X_2] = X_2 - 5$$

(d) Let $Z = \min(X_1, X_2)$. Compute $\mathbb{E}[\max(X_1, X_2) | Z]$

Solution: This is simply the exponential between the second and first geometric processes, regardless of the order.

$$\mathbb{E}[\max(X_1, X_2) | Z] = Z + \mathbb{E}[Y_2] = Z + 6$$

(e) Compute $\mathbb{E}[\min(Y_1, Y_2) | X_2]$.

Solution: Consider the following.

$$\begin{aligned}\Pr(\min(Y_1, Y_2) \geq k) &= \Pr(\min(Y_1, X_2 - Y_1) \geq k) \\ &= \Pr(Y_1 \geq k, X_2 - Y_1 \geq k)\end{aligned}$$

Let us reason about this. We note that $Y_1 \geq k$ and that $Y_1 \leq X_2 - k$, so we consider the probability of falling within $X_2 - k - k + 1 = X_2 - 2k + 1$. We then sum over all possible values of k.

$$\sum_{k=1}^{X_2 - 1} \frac{X_2 - 2k + 1}{X_2 - 1}$$

8.4 Poisson Processes

1. Every week Bob receives λ surveys. Approximate to 50 weeks in a year. Each survey is emailed to *at least* 2,000 CS students, where each promises gift cards for K_i lucky winners, where $K_i \sim N(5, \sqrt{5})$ The actual value of K_i may be different for each survey, but note they are i.i.d.

   (a) Bob completes a survey with probability p. How many surveys do we expect Bob to complete in a year?

   Solution: This is Poisson thinning. Note that the number of surveys that are sent, N, is Poisson distributed with parameter λ, so $N \sim \text{Pois}(\lambda)$. Once receiving a survey, Bob fills it out with probability p, so we see that this is a new Poisson process distributed with parameter λp. We expect Bob to fill out λp surveys per week. Per year, he then fills about $50\lambda p$.

   (b) Assume that survey winners are picked uniformly at random. Give an upper bound for the number of surveys we expect him to win, in a year, given that he fills out surveys with probability p.

   Solution: Intuitively, we see that the number of surveys is on average $\lambda p 50$ and that the number of gift cards per survey is 5, so we expect $\lambda p 250$ gift cards to be distributed among 2000 students. We expect any student to the have $\frac{\lambda p}{8}$ chance to receive a gift card. Here is the formal computation:

   First, we compute the number of surveys that Bob wins a single survey, given K. Let W be the number of surveys that Bob wins, and W_i be the indicator that Bob wins a single survey.

$$\mathbb{E}[W_i|K] = \frac{K}{2000}$$

   Thus, we can apply the law of iterated expectation to get

$$\mathbb{E}[\mathbb{E}[W_i|K]] = \mathbb{E}[\frac{K}{2000}] = \frac{1}{400}$$

   Now, we can compute $W = \sum_i W_i$.

$$\mathbb{E}[W|N] = \mathbb{E}[NW_i|N] = N\mathbb{E}[W_i] = \frac{N}{400}$$

   By the law of iterated expectation, we have

$$\mathbb{E}[W] = \mathbb{E}[\mathbb{E}[W|N]] = \mathbb{E}[\frac{N}{400}] = \frac{50\lambda p}{400} = \frac{\lambda p}{8}$$

   (c) Assume each survey is sent to exactly 2,000 CS students. Compute the variance in the number of surveys that we expect Bob to win. Is this an upper bound or a lower bound, given that surveys in reality go to *at least* 2,000 students? Assume $p = 1, \lambda = 0.01$.

Solution: Again compute variance for W_i and then apply linearity of variance, by independence of each survey.

$$\text{VAR}(W_i|K) = \frac{K}{2000}(1 - \frac{K}{2000})$$

By the law of total variance, we have (using $\mathbb{E}[W_i|K]$ from the previous part)

$$\begin{aligned}
\text{VAR}(W_i) &= \mathbb{E}[\text{VAR}(W_i|K)] + \text{VAR}(\mathbb{E}[W_i|K]) \\
&= \frac{1}{400}(1 - \frac{1}{400}) + \frac{5}{2000^2} \\
&= \frac{399}{400^2} + \frac{1}{400^2(5^2)} \\
&= \frac{499}{200,000}
\end{aligned}$$

Apply linearity of variance, and we have

$$\text{VAR}(W|N) = N\text{VAR}(W_i) = \frac{499}{2 * 10^5}N$$

Finally, we apply law of total variance once more, using $\mathbb{E}[W|N]$ from the previous part.

$$\begin{aligned}
\text{VAR}(W) &= \mathbb{E}[\text{VAR}(W|N)] + \text{VAR}(\mathbb{E}[W|N]) \\
&= \frac{499}{2 * 10^5}\lambda 50p + \frac{5}{400^2} \\
&= \frac{499}{400 * 100^2} + \frac{5}{400^2} \\
&= \frac{1023}{400^2 5}
\end{aligned}$$

This is a lower bound for variance, since with more students, we have higher uncertainty.

2. Consider CalCentric, a new piece of software that sees λ bug reports every minute. With probability p_1, production support marks a request as as `SEV-1`, with probability p_2 marks it as `SEV-2` and otherwise as `SEV-3`. Three teams, T_1, T_2, T_3 have been created to handle each level of urgency. Let Y_i be the number of minutes until the next request. Let Z_i be the number of requests per hour. Let X_{ij} represent the time until the jth request for team i. Assume all teams have zero bug reports at midnight.

   (a) Find the PMF of Z_i, Y_i, and X_{ij}.

   Solution: $Z_i \sim \text{Pois}(60\lambda p_i)$, so we see that the PDF is $\Pr(Z_i = k) = \frac{(\lambda p_i)^k}{k!}e^{-\lambda p_i}$.

$Y_i \sim \text{Expo}(\lambda p_i)$, so the PDF is $f_{Y_i}(y) = \lambda e^{-\lambda y}$.

X_{ij} is Erlang with intensity j, where each exponential is Y_i. Thus, $f_{X_{ij}}(x) = \frac{\lambda^j x^{j-1}}{(j-1)!} e^{-\lambda x}$.

(b) We pick a bug uniformly at random. How long do we expect team i to process bug j?

Solution: This is simply the expected value of an exponential for team i, which is $E[T_i] = \frac{1}{\lambda p_i}$.

(c) We pick a time uniformly at random across all 24 hours; let us call this time t, which is measured in minutes since midnight. How many total bug reports do we expect to have on file? How many bug reports for team i? For team i, how much time do we expect between the last bug report and the next?

Solution: We expect there to be λt total bug reports. For team i, there should be $\lambda p_i t$ bug reports. We expect $\frac{2}{\lambda p_i}$ minutes, as we expect $\frac{1}{\lambda p_i}$ until the next arrival and likewise, $\frac{1}{\lambda p_i}$ from the last arrival.

(d) For team i, what is the PDF of the time it takes between bugs at time t? Let this time be T.

Solution: We take the PDF of Y_i and scale by the length of the segment, as there is a higher probability of landing in a longer segment.

$$f_T(t) = \frac{f_{Y_i}(y)y}{E[Y_i]} = \frac{y\lambda e^{-\lambda y}}{1/\lambda} = y\lambda^2 e^{-\lambda y}$$

(e) Given the first bug of the day for CalCentric as a whole comes at time t_1, at what time of day do we expect team 2 to receive its third bug t_3?

Solution: The first bug may or may not go to team 2. With probability p_2, it is assigned to team 2 and we consider an Erlang of mode 2. Otherwise, we consider an Erlang of mode 3.

$$\begin{aligned}\mathbb{E}[t_3|t_1] &= p_2(t_1 + \mathbb{E}[X_{2,3} - X_{2,1}]) + (1-p_2)\mathbb{E}[X_{2,3}] \\ &= p_2(t_1 + \frac{2}{\lambda p_2}) + (1-p_2)\frac{3}{\lambda p_2}\end{aligned}$$

(f) Given that CalCentric has seen b bugs and that team 1 has b_1 bug reports to handle, how many bug reports to we expect team 2 to have? Assuming b_2 is more than the current number of bug reports assigned to team 2, how much more time do we expect until team 2 has b_2 bug reports?

Solution: There are only $b - b_1$ remaining bug reports and of those, we expect p_2 to be assigned to team 2. As a result, we expect team 2 to have

$$(b - b_1)\frac{p_2}{p_2 + p_3}$$

bug reports. We now examine an Erlang with mode $b_2 - (b - b_1)p_2$. The mean of an Erlang distribution is $\frac{k}{\lambda p_2}$, where k is the mode. Thus, we have

$$\frac{b_2 - (b - b_1)p_2/(p_2 + p_3)}{\lambda p_2}$$

3. With probability $\frac{1}{3}$, Alice takes her motorcycle to school, which takes 5 minutes. With probability $\frac{1}{2}$, Alice takes her bike to school. Otherwise, she takes 20 minutes to walk to school. Each motorcycle tire has been worn out and has an expected lifetime of 2 hours.

   (a) How much time does Alice spend walking before the first time she rides her motorcycle? **Solution**: Consider N, the number of days until Alice first rides her motorcycle. Take X to be the number of days Alice chooses to walk.

   Note that $N \sim \text{Geom}(\frac{1}{3})$ and that $X \sim \text{Bin}(N-1, \frac{1/6}{2/3}) = \text{Bin}(N - 1, \frac{1}{4})$.

   $$20\mathbb{E}[X] = 20\mathbb{E}[\mathbb{E}[X|N]] = 20\mathbb{E}[(N-1)\frac{1}{4}] = 5\mathbb{E}[N-1] = 10$$

   (b) In expectation, after how many days will Alice pop a tire?

   Solution: In expectation, Alice uses her motorcycle every 3 days. The tire in expectation can last up to 2 hours or 24 rides, meaning Alice will last up to $3 * 24 = 72$ days.

   (c) Alice has lost her bike and walks with probability $\frac{2}{3}$. If there are 180 days of school, how many days will Alice walk to school? Assume that Alice does not have a tire replacement; if she pops a tire, she cannot use her motorcycle for the remaining days.

   Solution: The motorcycle, in expectation, can only be used 24 times. Thus, Alice is expected to walk $180 - 24 = 156$ days.

8.5 Confidence Intervals

1. Bob is hanging a string of Christmas lights up. With probability p, a lightbulb is dead. There are 500 total lightbulbs, and Bob will tolerate at most 10 dead lightbulbs. If we wish to be 95% certain that we have 10 or fewer dead lightbulbs, find a lower bound on p.

   Solution: Let X be the number of dead light bulbs. The first step is to fit the form $\Pr(|X - \mu| \geq \alpha)$.

   $$\begin{aligned}&\Pr(X \geq 10)\\ &\qquad = \Pr(X - 500p \geq 10 - 500p)\\ &\qquad \leq \Pr(\|X - 500p\| \geq 10 - 500p)\end{aligned}$$

   Second, we note that p is already included. We find $\text{VAR}(X)$, knowing that X is binomial.

   $$\text{VAR}(X) = 500p(1-p)$$

   Third, we plug into the bound and compute.

   $$\begin{aligned}\Pr(\|X - 500\| \leq 10 - 500p) &= 1 - \Pr(\|X - 500p\| \geq 10 - 500p)\\ &\leq 1 - \frac{\text{VAR}(X)}{(10-500p)^2}\\ &= 1 - \frac{500p(1-p)}{10^2(1-50p)} = 0.95\end{aligned}$$

   Plugging in, we get $p = 0.00700$ or 0.7% chance of lightbulb death.

2. Derek wishes to protect user privacy whilst performing large-scale analytics. He has a sample size of 2,000 users, where each user clicks on an advertisement with probability p. Assume all users act independently. He decides to "jiggle" the data. Specifically, he replaces $\frac{2}{5}$ of the data with randomly-generated samples - "click" with probability $\frac{1}{2}$ and "no click" otherwise.

   (a) Let q be the probability of clicks that we observe. Find the estimate for p, $\hat{p}$, in terms of q.

   Solution: First, express the probability of clicks that we observe, in terms of the true probability.

   $$\begin{aligned}q &= \frac{2}{5}\frac{1}{2} + \frac{3}{5}p\\ &= \frac{1}{5} + \frac{3}{5}p\end{aligned}$$

Let us now express p in terms of q, to obtain our estimate $\hat{p}$.

$$\hat{p} = \frac{5q-1}{3}$$

(b) Apply Chebyshev's to find a 95% confidence interval for p.

Solution: The first step is to fit the form $\Pr(|X-\mu| \geq \alpha)$.

$$\Pr(|\hat{p}-p| \geq \alpha) = \Pr(|\hat{q}-q| \geq \frac{3}{5}\alpha)$$

The second step is to define our random variable and incorporate n.

Define $Y = \frac{1}{n}\sum_i Y_i$, where Y_i is i is 1 if we observe user i clicks, after "jiggling" the data. Note that by the Weak Law of Large Numbers Y should converge to the true mean, or in other words, the true q. Applying the fact that $\textsc{var}(cX) = c^2\textsc{var}(X)$ for some constant c and then linearity of variance, we get the following:

$$\begin{aligned}
\textsc{var}(Y) &= \frac{1}{n^2}\textsc{var}(\sum_i Y_i) \\
&= \frac{1}{n^2}(\frac{2n}{5}p(1-p) + \frac{3n}{5}\frac{1}{4}) \\
&= \frac{1}{n}(\frac{2}{5}p(1-p) + \frac{3}{5}\frac{1}{4})
\end{aligned}$$

Note that this quantity is maximized when $p(1-p) = \frac{1}{4}$, so

$$\textsc{var}(Y) = \frac{1}{n}(\frac{2}{5}\frac{1}{4} + \frac{3}{5}\frac{1}{4}) = \frac{1}{4n}$$

Finally, we plug into the bound and compute.

$$\Pr(|Y-q| \geq \frac{3}{5}\alpha) \leq \frac{\textsc{var}(Y)}{(3/5\alpha)^2}$$

Note that we need an upper bound on falling *within* the interval. Thus, we need the complement of the provided bound.

Also note that each Y_i has probability q of success. Thus, the Bernoulli distribution gives us $\textsc{var}(Y_i) = q(1-q)$. We wish to maximize this for our upper bound. Note that for any probability q, this is maximized when $q = \frac{1}{2}$, making $\frac{1}{4}$.

$$\begin{aligned}
1 - \frac{\text{VAR}(Y_n)}{(3/5\alpha)^2} &= 0.95 \\
\frac{1}{4n(3/5\alpha)^2} &= 0.05 \\
\frac{25}{36n\alpha^2} &= \frac{5}{100} \\
\frac{1}{2,000\alpha^2} &= \frac{36}{500} \\
\frac{1}{\alpha^2} &= \frac{36}{4} \\
\alpha &= \frac{2}{6} = \frac{1}{3}
\end{aligned}$$

Thus, our bound is

$$[p - \frac{1}{3}, p + \frac{1}{3}]$$

3. Alice also wishes to protect user privacy. She creates n different surveys that are kept with probability p. For each of the following, consider whether or not the Law of Large Numbers still holds, when we consider S_n to be the number of user responses considered, given n survey responses. Prove your conjecture. Assume that $k << n$.

   (a) Fixing a constant k and assigning k randomly-selected users to a randomly-selected survey.

   Solution: Yes. Let us define an "indicator" to denote the number of users that the ith group of k users contributes. With probability p, we ignore the survey that the user group is assigned to and thus ignore the ith group.

$$X_i = \begin{cases} k & \text{w.p. p} \\ 0 & \text{o.w.} \end{cases}$$

   Note that $\mathbb{E}[X_i] = pk$ and $\text{VAR}(X_i) = k^2 p(1-p)$. Compute $A_n = \frac{S_n}{n}$.

$$\begin{aligned}
\mathbb{E}[A_n] &= \frac{1}{n} npk = pk \\
\text{VAR}(A_n) &= \frac{1}{n^2}\text{VAR}(S_n) = \frac{k^2}{n} p(1-p)
\end{aligned}$$

   Note that

$$\Pr(\|A_n - \mathbb{E}[A_n]\| > \epsilon) = \Pr(\|A_n - pk\| > \epsilon) \leq \frac{k^2 p(1-p)}{n\epsilon^2}$$

   which tends to 0 as $n \to \infty$. This thus obeys the Law of Large Numbers.

(b) Fixing a constant k and for each group of $\frac{n}{k}$ assign to a random survey.

Solution: No. Again, define X_i.

$$X_i = \begin{cases} \frac{n}{k} & \text{w.p. p} \\ 0 & \text{o.w.} \end{cases}$$

Note that $\mathbb{E}[X_i] = \frac{np}{k}$ and $\text{VAR}(X_i) = \frac{n}{k}p(1-p)$. Compute $A_n = \frac{S_n}{n}$.

$$\mathbb{E}[A_n] = \frac{p}{k}$$

$$\text{VAR}(A_n) = \frac{1}{n^2}\text{VAR}(S_n) = \frac{1}{n}\text{VAR}(X_i) = \frac{p(1-p)}{k}$$

Note that

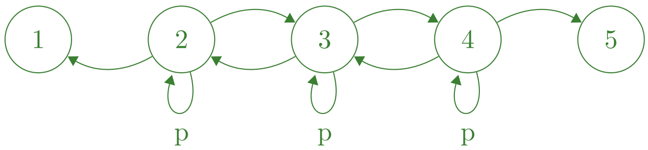
$$\Pr(\|A_n - \mathbb{E}[A_n]\| > \epsilon) = \Pr(\|A_n - \frac{p}{k}\| > \epsilon) \leq \frac{p(1-p)}{k\epsilon^2}$$

which is constant, regardless of n. Thus, this does not obey the Law of Large Numbers.

8.6 Markov Chains

1. Bob is walking along a small bridge, 3 feet wide and 10 feet long. Every time step, he walks 1 foot along the bridge. With probability p, he walks straight forward, and otherwise, he takes one step 1 foot forward and 1 foot to the right with probability $\frac{1}{2}$, forward 1 foot and left 1 foot with probability $\frac{1}{2}$. If he walks off the bridge, he falls into the water and foregoes the bridge to get to the other side. How many time steps do we expect Bob to walk before falling in the water, assuming he starts at the center of the bridge?

   Solution: Consider a five-state Markov Chain. With X_1, X_2, X_3, X_4, X_5, where X_1, X_5 represent falling into water.

   This is a hitting time problem, where our targets under consideration are both X_1, X_5.

   We will compute the time until falling into the water, for each state, $\beta(i)$.

$$\begin{aligned}
\beta(1) &= 0 \\
\beta(2) &= 1 + \frac{1-p}{2}\beta(1) + \frac{1-p}{2}\beta(3) + p\beta(2) \\
\beta(3) &= 1 + \frac{1-p}{2}\beta(2) + \frac{1-p}{2}\beta(4) + p\beta(3) \\
\beta(4) &= 1 + \frac{1-p}{2}\beta(3) + \frac{1-p}{2}\beta(5) + p\beta(4) \\
\beta(5) &= 0
\end{aligned}$$

   This is simply a linear system of equations, so we rearrange all constants onto one side and plug into a matrix form to row reduce.

$$\begin{aligned}
0 &= \beta(1) \\
-1 &= \frac{1-p}{2}\beta(1) + (p-1)\beta(2) + \frac{1-p}{2}\beta(3) + 0 + 0 \\
-1 &= 0 + \frac{1-p}{2}\beta(2) + (1+p)\beta(3) + \frac{1-p}{2}\beta(4) + 0 \\
-1 &= 0 + 0 + \frac{1-p}{2}\beta(3) + (p-1)\beta(4) + \frac{1-p}{2}\beta(5) \\
0 &= \beta(5)
\end{aligned}$$

   This is equivalent to the following augmented matrix:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ (1-p)/2 & p-1 & (1-p)/2 & 0 & 0 & -1 \\ 0 & (1-p)/2 & p-1 & (1-p)/2 & 0 & -1 \\ 0 & 0 & (1-p)/2 & p-1 & (1-p)/2 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Row reducing, we get the following matrix.

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 3/(1-p) \\ 0 & 0 & 1 & 0 & 0 & 4/(1-p) \\ 0 & 0 & 0 & 1 & 0 & 3/(1-p) \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Starting in the center of the bridge, we expect him to fall into the water after $\frac{4}{1-p}$ time steps.

2. Let us consider the following scenarios. Use Markov Chains only when necessary, but consider your intuition first. Justify all of your answers.

   (a) Consider a coin with bias p. What is the probability that we see TH before HH?

   Solution: We can use intuition. If we roll T at any point, we must see TH before HH; the probability would be 1. We can thus only fail if we see HH in the first two rolls. This is the complement of rolling two heads in the first two rolls.

   $$1 - p^2$$

   (b) Consider a fair dice. What is the probability that we see 5-3 before seeing 3-5?

   Solution: Consider all possible pairs of numbers, as our states. Note that the probability of reaching 5-3 from 3-5 is the same as the probability of reaching 3-5 from 5-3. By symmetry, we then expect the probability of hitting 5-3 before 3-5 to be

   $$\frac{1}{2}$$

3. Consider a die.

   (a) Let this die be fair. How much time do we expect until we roll 5 dots, followed by a roll of 3 dots?

   Solution: We note that for fair die, we expect 6 rolls until we see a 5. For each of those rolls, we expect 6 more rolls until we see a 3. Thus, we expect 36 rolls until we see a sequence 5-3.

(b) How many 5s do we expect to see before we roll 5 dots, then 3 dots?

Solution: We computed earlier that we expect the sequence of 5-3 to appear after 36 rolls. We know that the last two rolls are 5 and 3. Thus, we only have randomness in the first 34 rolls. We expect $\frac{1}{6}$ of these to be 5s. Thus, we expect $\frac{17}{3} + 1$ 5s.

(c) Let this die be loaded, favoring 5 with probability $\frac{6}{11}$. All others are have equal probability, or $\frac{1}{11}$. How much time do we expect until we roll 5 dots, followed by a roll of 3 dots?

Solution: First, let us consider the number of rolls it takes to see 5-3. Consider X_1, X_2, X_3, where X_1 is any set of two rolls that does not start with a 5, X_2 is any set of two rolls that starts with a 5, and X_3 is the state with 5-3. We consider time until X_3, and list our balance equations.

$$\begin{aligned}
\beta(1) &= 1 + \frac{5}{11}\beta(1) + \frac{6}{11}\beta(2) \\
\beta(2) &= 1 + \frac{4}{11}\beta(1) + \frac{6}{11}\beta(2) + \frac{1}{11}\beta(3) \\
\beta(3) &= 0
\end{aligned}$$

We effectively have two unknowns and two equations, we can solve by hand. Rewrite both equations

$$\begin{aligned}
x_1 &= \frac{11}{6} + x_2 \\
x_2 &= \frac{11}{5} + \frac{4}{5}x_1
\end{aligned}$$

Plug the second into the first equation and solve for x_1.

$$\begin{aligned}
x_1 &= \frac{11}{6} + \frac{11}{5} + \frac{4}{5}x_1 \\
\frac{1}{5}x_1 &= \frac{121}{30} \\
x_1 &= \frac{121}{25} = (\frac{11}{5})^2
\end{aligned}$$

Our answer is thus the following.

$$(\frac{11}{5})^2$$

(d) Again with the loaded die, how many 5s do we expect to see before our first sequence of 5-3? (Note that we cannot simply take the answer from the previous part and multiply by $\frac{1}{6}$. This is because the duration of survival is positively correlated with the number of 5s we expect to see.)

Solution: Using the states defined in the previous part, we modify our balance equations. Instead of counting time steps, we are now counting the number of 5s that show up. Thus, we add 1 anytime we exit node X_2. Otherwise, the setup looks identical to that of our equations in the previous part.

$$\beta(1) = \frac{5}{11}\beta(1) + \frac{6}{11}\beta(2)$$
$$\beta(2) = 1 + \frac{4}{11}\beta(1) + \frac{6}{11}\beta(2) + \frac{1}{11}\beta(3)$$
$$\beta(3) = 0$$

Again, we can solve by hand.

$$x_1 = x_2$$
$$x_2 = \frac{11}{5} + \frac{4}{5}x_1$$

Plug the second equation into the first to find x_1, and we have

$$x_1 = \frac{11}{5} + \frac{4}{5}x_1 = 11$$

Thus, the number of 5s we expect to see is 11.

8.7 Transformations

1. Find the PDF of Z given the following moment-generating function of z.

$$M_Z(s) = \frac{2}{3}\frac{5}{5-s} + \frac{1}{3}\frac{e^{s2} - e^s}{s}$$

Solution: We have $\frac{2}{3}$ probability of assuming an $E \sim \text{Expo}(5)$. Otherwise, we assume a shifted uniform $U \sim [1, 2]$.

Our PDF is thus the following.

$$f_Z(z) = \frac{2}{3}5e^{-5z} + \frac{1}{3}\frac{1}{3} = \frac{3}{5}5e^{-5z} + \frac{1}{9}$$

2. Let X be the time until lightbulb 1 dies, where its average lifetime is λ_1. Let Y be the time until lightbulb 2 dies, where its average lifetime is λ_2. Without applying linearity of expectation, find the expectation of $Z = X + Y$.

Solution: X, Y are independent, so $M_Z(s) = M_X(s)M_Y(s)$.

$$M_Z(s) = \frac{\lambda_1}{\lambda_1 - s}\frac{\lambda_2}{\lambda_2 - s}$$

To compute expectation, we take the derivative once, and evaluate at $s = 0$

$$\begin{aligned}&\frac{\lambda_1}{\lambda_1 - s}\frac{\lambda_2}{(\lambda_2 - s)^2} + \frac{\lambda_1}{(\lambda_1 - s)^2}\frac{\lambda_2}{\lambda_2 - s}|_{s=0}\\ &= \frac{1}{\lambda_2} + \frac{1}{\lambda_1}\end{aligned}$$