Math 116: Cryptography

Michael Pham

Spring 2025

Contents

Week 1
Introduction

1.1 Lecture – 1/22/2025

1.1.1 Administrivia

Basic Information

- Tues 5-6pm on Zoom.
- Fri 4-5pm at Evans 891.

The main focuses of this course includes:

- Cryptogrpahy
- Cryptonalysis
- Coding

Within this, we will learn about:

- Probability Theory
- Information Theory
- Abstract Algebra

The first two topics will comprise the first half of the course, while Abstract Algebra will be the focus of the second half.

We'll also touch on some historical background for the topics covered.

Grading Breakdown

The grading breakdown is as follows:

- Homework - 25%
- Midterm 1 - 25%
  - March 7

- Midterm 2 - 25%
  - April 23
- Final Product - 25%

We note that the midterms are not cumulative; that is, each midterm will cover content exclusively from its half of the course.

1.1.2 The Basics

The basic setup is as follows: suppose we have Alice and Bob, with Alice wanting to send a message to Bob.

However, the message Alice sends is secretive. Then, she must encrypt the message somehow; that is, performing some function on it which (hopefully) hides its meaning.

However, Bob must have some reasonable decryption function as well. Then, he can have some secret key to decrypt the ciphertext.

Definition 1.1 (Basic Terminology)**.** The **plaintext** is the original, unencrypted message.

The **ciphertext** is the encrypted message.

Now, we can imagine a third party Eve. Suppose Eve can see the ciphertext. Then, the goal is so that Eve can't figure out the plaintext within a reasonable time from the ciphertext.

Then, this encapsulates the two forces in these scenarios:

- Alice wanting to send messages to Bob without Eve knowing.
- Eve wanting to figure out the message.

Variations

Some variants we can have includes:

- Does Alice know the secret key K that Bob holds?
- Can Eve pretend to be Alice when attacking?
- What are Eve's resources?
- Does Eve know the encryption algorithm?

1.1.3 Early Encryption Methods

Scytale Cipher

This first cipher was supposedly used by the Ancient Greeks. We will walk through the following example:

Example 1.2 (Scytale Cipher Encryption)**.** Suppose we have the following plaintext: `HES GOT PADFOOT IN THE PLACE WHERE ITS HIDDEN`.

Then, we want to scramble this message somehow. Suppose we have a secret key then, which will be some small integer. In this case, let it be $K = 6$. Then, we will line the plaintext up into six columns:

Then, we will go down each column to create our ciphertext. Going through the encryption process, we

will get the following ciphertext: `HPTPEAILSDNAGFTCOOHETOE`.

In order to decrypt, we simply count the number of letters, divide into columns, and then read along the rows.

Example 1.3 (Scytale Cipher Decryption). Suppose $K = 5$, with the ciphertext being

`OBDFANYTBELWYIAOSFNIE`

Now, there are 21 letters. So, when putting them in columns, the length of each column will be $5, 4, 4, 4, 4$. So, we have:

And from there, we see that it reads: `ONE IF BY LAND TWO IF ...`

It's evident then that it's relatively simply to brute-force the key until some sort of sentence can be read. Note that we wouldn't want too large of a K as that'd lead to the letters not really being scrambled.

Note here that the Scytale Cipher is an example of a transposition cipher: that is, it permutes the plaintext.

Caesar Cipher

The basic premise of this is to simply write out our alphabet, and assign each letter its numerical position in the alphabet.

Now, let $1 \leq K \leq 25$. Then, get each letter, get its position, add our K to it, and replace it with the letter corresponding to the new value.

Note that we need to mod our result so that it wraps around.

To decrypt given a key, we can just subtract the value K and we're chilling.

Supposedly, this was used by Julius Caesar, and even the Russian Army in WWI.

The Caesar Cipher is an example of a substitution cipher: letters are mapped to some new letter by a function.

Now, these two ciphers fail under the same issue: due to how small K is, it's easy to just brute-force until a message is decrypted.

1.1.4 Substitution Ciphers

For a general substitution cipher, we note that we have $26!$ ways to substitute every letter; the first letter has 26 possibilities, then the next has 25, and so on.

This can be approximated using Sterling's Approximation to see that there are more than 10^{26} possibilities – that's a lot!

However, even then, substitution ciphers aren't that powerful: the English language isn't random. There are 26 letters in the alphabet, but certain letters are more frequent (namely, `E` appears 12% of the time, with T at 9% and so on).

Furthermore, certain "pairs" (bigrams) can appear together more frequently as well (for example, `TH` appears more than 1.5% of the time).

And we can further look at trigrams, and so on.

Then, by creating a frequency table of our ciphertext, we can make reasonable guesses on what letter is what.

1.2 Lecture – 1/24/2025

In today's lecture, we will be reviewing discrete probability.

1.2.1 Definitions

Definition 1.4 (Sample Space)**.** The Sample Space Ω is the set of completely determined outcomes.

Example 1.5. For example, suppose we flip a coin. In this scenario, $\Omega = \{H, T\}$ as we can have either heads or tails.

Or, if we roll two six-sided die, then $\Omega = \{(1,1), (1,2), \ldots, (6,6)\}$, with $|\Omega| = 36$.

Definition 1.6 (Event)**.** An event is a subset $A \subset \Omega$.

Definition 1.7 (Discrete Probability Function)**.** A discrete probability function $p : \Omega \to [0,1]$ is one such that $\sum_{\omega \in \Omega} p(\omega) = 1$.

We say then that $p(A) = \sum_{a \in A} p(a)$.

Example 1.8. Suppose we have a fair coin. That is, $p(H) = p(T) = \frac{1}{2}$.

Similarly, if we are rolling two fair, six-sided dice, let A denote the event in which the sum is 6. Then, we have that $A = \{(1,5), (2,4), \ldots\}$. In total, we will have five possibilities; in other words, $p(A) = \frac{5}{36}$.

1.2.2 Solving Probability Problems

We say that there are maneuvers to calculate $P(A)$:

- Symmetry, Inclusion-Exclusion, Complements, etc.

Example 1.9 (Symmetry)**.** Suppose we have an urn with 101 balls in it. Say we have 100 black balls and 1 red ball.

We repeatedly draw a ball from the urn without replacement. Then, we want to find the draw which is most likely to be red.

One way to figure this out is to observe that:

- $p(1) = \frac{1}{101}$.
- $p(2) = \frac{100}{101} \cdot \frac{1}{100} = \frac{1}{101}$.
- $p(3) = \frac{100}{101} \cdot \frac{99}{100} \cdot \frac{1}{99} = \frac{1}{101}$

Now, we observe that in all of these cases, the answer ends up being $\frac{1}{101}$...

So, instead, we can think of lining up the balls in a row, and our n^{th} draw is where the red ball is. In that case, the ball has the same chance of landing in each spot.

Definition 1.10. ...

Example 1.11 (Recursion). We observe that the sample space if $\Omega = \{M, V\}$.

Now, $p(M)$ denotes M firing the gun. And $p(V)$ denotes V firing the gun.

If we tried going by the first approach, it becomes pretty ugly, yielding us something along the lines of:

$$p(M) = \frac{1}{6} + \left(\frac{5}{6}\right)^2 (\frac{1}{6}) + \left(\frac{5}{6}\right)^4 (\frac{1}{6}) + \cdots$$

On the first turn, Mitya can fire the gun with a $\frac{1}{6}$ chance. Then, there is a $\frac{5}{6}$ chance it isn't: in this case, this can be viewed as a new game.

Then, we have:

$$\begin{aligned} p(M) &= \frac{1}{6} + \frac{5}{6}(1 - p(M)) \\ p(M) + \frac{5}{6}p(M) &= 1 \\ p(M) &= \frac{6}{11} \end{aligned}$$

Example 1.12 (Birthday Problem). Let us consider the following question: what is the probability that, among n people, at least one person has a birthday today.

What's the smallest n such that this is more likely than not (that is, the probability is $\geq \frac{1}{2}$).

First, if $n = 1$, then there's a $\frac{1}{365}$ chance their birthday is today.

In the case of $n = 2$, we might be tempted to say that the probability is $\frac{1}{365} + \frac{1}{365}$. However, there is overcounting – we must subtract the probability that they both have a birthday today. That is, $\frac{1}{365} + \frac{1}{365} - \frac{1}{365^2}$.

And so on.

However, instead, we can utilize complements. So, for n people, we note that for each person, there are 364 days that aren't today that they can be on:

$$p(n) = 1 - \frac{364^n}{365^n}$$

Going back to our original question, to find a big enough n, we can proceed as follows:

$$\begin{aligned} \ln(1 - p(n)) &< \ln\left(\frac{1}{2}\right) \\ n &> 256 \end{aligned}$$

Note that we can find 256 using Taylor approximations of $ln(1 - x) \approx -x$ and some algebraic manipulations.

Example 1.13 (Birthday Paradox). Suppose we want to find the probability that among n people, two of them have a common birthday.

Let $p(n)$ be the probability that two people have a common birthday. Again, let us check for each $n \in \mathbb{N}$.

First, for $n = 1$, $p(1) = 0$. For $n = 2$, we have that $p(2) = \frac{1}{365}$. For $n > 2$, it becomes harder to calculate. So, instead, we can use the complement:

For $1 - p(n)$, we have:

$$\begin{aligned} 1 - p(n) &= \left(\frac{365}{365}\right)\left(\frac{364}{365}\right)\cdots \\ &= 1 \cdot \left(1 - \frac{1}{365}\right)\cdots\left(1 - \frac{n}{365}\right) \end{aligned}$$

The Birthday Paradox will be relevant to crytography later.

Example 1.14 (Coin Flips). Suppose we are flipping a fair coin repeatedly. Now, we want to flip the coin until we see consecutive HH (then Alice wins) or HT comes up (then Bob wins).

Our sample space will be $\Omega = \{A, B\}$. Denote $p(A)$ to be the probability that Alice wins.

We can model this as the following: let S be our starting state. If we flip T, we go back to S. Otherwise, if H, then we worry about the outcome: in this case, there is a $\frac{1}{2}$ probability of either winning.

Now, let $p(x)$ denote the probability that Alice wins starting on x.

1.2.3 Distributions

If Ω is finite, the uniform distribution is a probability function $p : \Omega \to [0, 1]$ such that $p(\omega) = \frac{1}{|\Omega|}$ for every $\omega \in \Omega$.

However, we don't always have the uniform distribution. For example, we have some natural non-uniform distributions:

- Distribution of letters in the English alphabet.
- Benford's Law: numbers with small leading digits appear more frequently in natural datasets.

WEEK 2

SECOND WEEK WOES

2.1 Lecture – 1/27/2025

2.1.1 Combinatorial Analysis

Definition 2.1 (Factorial)**.** We define the factorial to be $n! = n(n-1)\cdots(1)$.

We typically use factorial to count things where order matters. An example is to count the number of substitution ciphers there are:

- For the first letter, we have 26 different letters it can be mapped to.
- Then, there are 25 letters.
- And so on...

Theorem 2.2 (Stirling's Approximation)**.** Stirling's Approximation states that:

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

Thus, we see that $n!$ is "super exponential."

Next up, we look at combinations:

Definition 2.3 (Combination)**.** We define the number of ways to choose k items out of n (where order doesn't matter) to be:

$$\begin{aligned}\binom{n}{k} &= \frac{n!}{k!(n-k)!} \\ &= \frac{n(n-1)\cdots(n-k+1)}{k!}\end{aligned}$$

From here, we note then that there is a symmetry property here. That is, we have $\binom{n}{k} = \binom{n}{n-k}$.

Example 2.4 (Powerball). In the Powerball, we have six balls. The first five we pick numbers from 1 to 69. The last is the "powerball", where we pick from 1 to 23.

In the first five slots, the order doesn't matter, while the powerball is its own thing.

So, we are now interested in the probability of winning the lottery. Then, suppose that each ticket has an equal chance of winning.

First, we find the number of tickets there are: this is simply $\binom{69}{5}$.

Principle of Inclusion-Exclusion

One form of this is to consider events $A_1, A_2, A_3 \in \Omega$. Then, the probability of $P(A_1 \cup A_2 \cup A_3) = \sum_1^3 P(A_i) - \sum_{j\neq i}^3 P(A_i \cap A_j) + \sum_1^3 P(A_i \cap A_j \cap A_k)$.

Example 2.5 (PIE Example). Let Ω be the integers in $[1, 100]$. Then, we have $A_d \subset \Omega$ be the set $A_d = \{n \in \Omega : d \mid n\}$.

Then, we want to find $P(A_2 \cup A_3 \cup A_5)$; that is, the probability the number we pick is divisible by 2, 3, or 5.

To solve this, we can apply PIE.

First, the probability that a number chosen is divisible by 2 is $P(A_2) = \frac{50}{100}$. For $P(A_3) = \frac{33}{100}$, and $P(A_5) = \frac{20}{100}$.

Next, we check $P(A_2 \cap A_3) = P(A_6) = \frac{16}{100}$ (numbers divisible by 6). The general pattern is just finding the numbers divisible by d, then round down.

So, we have $P(A_2 \cap A_5) = P(A_{10}) = \frac{10}{100}$, and $P(A_3 \cap A_5) = P(A_{15})\frac{6}{100}$.

Finally, we have $P(A_2 \cap A_3 \cap A_5) = P(A_{30}) = \frac{3}{100}$.

Using PIE then, we have:

$$\frac{50}{100} + \frac{33}{100} + \frac{20}{100} - \frac{16}{100} - \frac{10}{100} - \frac{6}{100} + \frac{3}{100} = \frac{74}{100}$$

Example 2.6 (Enigma). Suppose we want to find the probability that a uniformly random substitution cipher sends no letter to itself. This has relevance to the Enigma later on.

To do this, we can first consider its complement. First, consider the number of ciphers that sends at least one letter to itself.

Suppose we have n letters, and we denote A_i to be the event in which the i^{th} letter gets sent to itself.

To do this, we note that $P(A_i) = \frac{1}{n}$. For each pair i, j, we have that $P(A_i \cap A_j) = \left(\frac{1}{n}\right)\left(\frac{1}{n-1}\right)$.

In general then, we have $P(A_1 \cap A_2 \cap \cdots) = \frac{1}{n(n-1)(n-2)\cdots(n-k+1)} = \frac{(n-k)!}{n!}$.

Then, putting this together, we have:

$$\begin{aligned} P(A_1 \cup A_2 \cup \cdots \cup A_n) = n \cdot (\frac{1}{n}) \\ - \binom{n}{2} \frac{1}{n(n-1)} \\ + \binom{n}{3} \frac{1}{n(n-1)(n-2)} \\ \vdots \\ + (-1)^{k-1} \binom{n}{k} \frac{(n-k)!}{n!} \end{aligned}$$

Then, we have the probability to be:

$$\begin{aligned} 1 - P(A_1 \cup A_2 \cup \cdots \cup A_n) &= 1 - \sum_{k=1}^{n} (-1)^{k-1} \frac{1}{k!} \\ &= \sum_{k=0}^{n} (-1)^{k} \frac{1}{k!} \\ &\approx e^{-1} \end{aligned}$$

Example 2.7 (Poker). We pick five cards out of a standard deck of cards. That is, we have $\binom{52}{5}$ different number of hands.

Now, we want to look for specific patterns that counts. For example:

- Pair – two cards with the same number.
- Two Pair
- Three of a Kind
- Four of a Kind
- Full House – a Pair and a Three of a Kind

Then, the lower the probability of a hand is, the "strongest" it will be.

First, we look at pairs. We have 13 numbers we can pick from. Then, from the four suites, we have to pick two suites. That is, $13 \cdot \binom{4}{2}$. Then, for the remaining three cards, we have twelve numbers remaining to choose from, so $\binom{12}{3}$, with 4 different suites for each. Putting it all together, we have $13 \cdot \binom{4}{2} \cdot \binom{12}{3} \cdot 4^3$ choices. We divide by $\binom{52}{5}$ to get our final probability.

Next, for two pairs, we want to pick two numbers from the thirteen: that is, $\binom{13}{2}$. From there, we for each of the cards we have to pick two suites from for each pair. That is, $\binom{4}{2}^2$. For the final card, we have 11 numbers left, with 4 suites to choose from. Divide by the total number of hands for our final probability.

For Three of a Kind. So, we have:

$$13 \cdot \binom{4}{3} \cdot 122 \cdot 4^2$$

And divide it by $\binom{52}{5}$.

For Four of a Kind, we have:

$$13 \cdot 12 \cdot 4$$

And for a Full House, we have $13 \cdot \binom{4}{3} \cdot 12 \cdot \binom{4}{2}$.

2.2 Lecture – 1/29/2025

For today's lecture, we will look at Expected Value.

Definition 2.8 (Random Variable)**.** A random variable is a function $X : \Omega \to \mathbb{R}$.

Definition 2.9 (Expected Value)**.** The expected value of a random variable X, denoted by $\mathbb{E}[X]$, is equal to:

$$\sum_{\omega \in \Omega} X(\omega) P(\omega)$$

We note here that, in principal, this may not converge. That is, it could be infinity. For example, suppose we are taking a random walk on the number line. For each second, we flip a coin to go left or right, finding the expected value of flips to get back to the origin turns out to be infinity.

Example 2.10 (Powerball and Expected Value)**.** Suppose the jackpot is $\$100,000,000$. However, to play, we have to pay a $\$5$ ticket. Then, denote the probability of winning to be p; in that case, our expected value is $(100000000 - 5)p + -5(1 - p)$.

Thus, we see that we are losing roughly $\$4.66$ each time; it's financially inadvisable to play the powerball.

Example 2.11 (Flipping Coin)**.** Suppose we are flipping a coin with a probability p for heads, with $p \in [0, 1]$. Then, we want to find the expected number of flips until we see the first head.

We let X be a random variable denoted the number of flips until the first head. We want to thus find $\mathbb{E}[X]$.

Looking at the outcome space, we have $\Omega = \{1, 2, \ldots\}$, and $X(n) = n$. We want to find the probability for n flips.

Then, we observe that we have:

$$\begin{aligned}\mathbb{E}[X] &= (1)(p) + (2)(1-p)p + \cdots + (n)(1-p)^{n-1}p \\ &= \sum_{n=1}^{\infty} (1-p)^{n-1} pn\end{aligned}$$

To calculate this sum, we can first try it recursively:

Theorem 2.12 (Linearity of Expected Value)**.** For random variables $X_1, \ldots, X_n$, we have:

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

Example 2.13. Suppose we wanted to find the expected values of getting heads in N flips. Then, let X be the number of heads in N flips.

Then, we can let $X = X_1 + X_2 + \cdots + X_n$, where X_i is having i heads in N flips.

So, this is simply:

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[X_1 + \cdots + X_n] \\ &= \mathbb{E}\{X_1\} + \cdots + \mathbb{E}[X_n]\end{aligned}$$

Then, we have that $\mathbb{E}[X] = p(1) + (1-p)(0)$. And for each i, we have a probability of getting heads of p. Thus, the final expected value is just Np.

However, if each X_i depend on each other, it's a different matter.

Example 2.14 (Coupon Collector). Suppose we have n coupons. Each day, you receive a coupon uniformly at random.

Then, we want to find the expected number of days to collect all of the coupons.

To solve this, we can let $\Omega = \{\text{all possible sequences of coupons}\}$. Then, let X_i bethe number of days it takes to get the i^{th} new coupon after already having $i-1$ distinct coupons.

For example, we have $X_1 = 1$. Then, we observe that X_i is essentially seeing the number of flips until we get the first heads, where $p = \frac{n-(i-1)}{n}$.

Thus, we see that $\mathbb{E}[X_i] = \frac{n}{n-(i-1)}$. So, the final answer would be:

$$\begin{aligned}\mathbb{E}[X] &= \sum_{i=1}^{n} \frac{n}{n-(i-1)} \\ &= n\left(\frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{n}\right) \\ &\approx n\ln(n)\end{aligned}$$

Example 2.15. Suppose you're rolling a fair, six-sided die.

First, let X be the number of rolls to get a 6. In this case, we see that $\mathbb{E}[X] = 6$.

Next, let Y to be the number of rolls between $6's$. Then, $\mathbb{E}[Y] = 6$.

Now, suppose a bus arrives between rolls. Let X_i be the number of rolls from bus arrival to the next six. Then, $\mathbb{E}[X_1] = 6$.

Let X_2 be the number of rolls from bus arrival to the previous six. In this case, $\mathbb{E}[X_2] = 6$ since it should be symmetric either way.

Now, when looking at $\mathbb{E}[X_1 + X_2]$, this is looking at the number of rolls between $6's$ before and after bus arrived. However, this is in fact equal to 12.

Remark 2.16. We need to be careful about interpreting probability and expectation.

2.3 Lecture – 1/31/2025

Let (Ω, P) be a probability ensemble. Let us suppose then that we have events $A, B \subset \Omega$.

Then, we give the following definition:

Definition 2.17 (Conditional Probability)**.** We define the conditional probability of A given B, denoted by $P(A \mid B)$, to be:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

We note that we sometimes denote $P(A \cap B)$ by $P(A, B)$. Furthermore, if $P(B) = 0$, we would say that this isn't defined.

Example 2.18 (Bigrams)**.** Let us consider bigrams x_1x_2 in English text. Recall that the probability $P(x_1 = q)$ is very low. However, if we look at $P(x_1 = q \mid x_2 = u)$, we see then that the probability is higher. And if we had $P(x_2 = u \mid x_1 = q)$, the probability gets much higher.

Note that this example illustrates how $P(A \mid B) \neq P(B \mid A)$ necessarily.

Definition 2.19 (Independence)**.** We say that events $A, B \subset \Omega$ are independent if $P(A \mid B) = P(A)$. That is, knowing B doesn't affect the probability of A.

This can be rewritten further to say that:

$$P(A) = \frac{P(A \cap B)}{P(B)}$$
$$P(A)P(B) = P(A \cap B)$$

Example 2.20 (Independent Events)**.** Suppose we flip a coin twice. Let A denote the first being heads, and B being the second being heads. Then, clearly, these events are independent.

2.3.1 Properties of Conditional Probability

Theorem 2.21 (Product Rule)**.** We say that $P(A, B \mid C) = P(A \mid B, C)P(B \mid C)$.

Intuitively, this is just us finding the probability of A, B given C. Then, we first check for B given C, and then find A given the other two having had happened.

Theorem 2.22 (Bayes' Theorem)**.**

$$P(A \mid B) = P(B \mid A)\frac{P(A)}{P(B)}$$

2.3.2 Random Variables

Say we have X, Y random variables $\Omega \to \mathbb{R}$. We say then that X, Y are independent if $P(X = x, Y = y) = P(X = x)P(Y = y)$, for all $x, y \in \mathbb{R}$.

This is also equivalent to saying that $P(X = x \mid Y = y) = P(X = x)$.

Theorem 2.23 (Sum Rule)**.** We say that:

$$P(X = x) = \sum_{y \in \mathbb{R}} P(X = x \mid Y = y)$$

Example 2.24. Alice is testing for disease. She can either be positive or negative for the disease.

We denote this true state to be a random variable X, where $X = 1$ if Alice is positive, else 0.

Now, let's say that she tests for a disease. The result can also be positive or negative; we will use a random variable Y as an indicator variable.

Say then that the test is 95% accurate. That is, if Alice is positive, there's a 95% chance it shows up positive. And if Alice is positive, there's a 95% chance it shows up negative.

In other words, we have $P(Y = 1 \mid X = 1) = 0.95, P(Y = 0 \mid X = 0) = 0.95$.

Now, suppose Alice tested positive and we want to find the probability that Alice has the disease. We observe then that:

$$\begin{aligned} P(X = 1 \mid Y = 1) &= \frac{P(Y = 1 \mid X = 1)P(X = 1)}{P(Y = 1)} \\ &= \frac{P(Y = 1 \mid X = 1)P(X = 1)}{P(Y = 1 \mid X = 1)P(X = 1) + P(Y = 1 \mid X = 0)P(X = 0)} \end{aligned}$$

At this point, we can't proceed further without knowing about $P(X = 0)$ and $P(X = 1)$, and so on. So, let us suppose then that $P(X = 1) = 0.01$. Then, ew can proceed from here.

Remark 2.25. Often in practice, we need to make prior assumptions.

Example 2.26. Suppose a family has two cats. At least one is male. Then, what is the probability that both are male?

Well, to begin with, we need to make prior assumptions. For example, we may assume independence and uniform. In other words, the prior distribution is:

- $MM = 1/4$.
- $MF = 1/4$.
- $FM = 1/4$.
- $FF = 1/4$.

In this case, we have a probability of $1/3$ since we want at least one cat to be male; thus, we ignore FF.

A remark to be made here is that if the question was changed slightly to say what is the probability given the first cat was male, then the answer would change to $1/2$.

Typically in our problems, we have three things:

- $\mathcal{H}$: Hypothesis/Model.
- Θ: Parameters.
- D: Data/Outcome.

Example 2.27. Suppose we flip a coin 10 times and it comes up heads 9 times. This is our data.

Them, we can make hypotheses:

- H_0: The coin is fair.
- H_1: The coin is biased.

And we note that Θ can be the probability of H.

We can usually consider a function $D \mapsto P(D \mid \Theta, \mathcal{H})$. This is our conditional probability.

On the other hand, we can also look at $\Theta \mapsto P(D \mid \Theta, \mathcal{H}) = \mathcal{L}(\Theta \mid D)$, which is the likelihood.

We also have $P(D \mid \mathcal{H})$, which is the evidence for $\mathcal{H}$. It can also be referred to as the marginal likelihood.

And finally, we have the prior probability $P(\Theta \mid \mathcal{H})$. Meanwhile, $P(\Theta \mid D, H)$ is the posterior probability. We use "prior" and "posterior" because this is before/after we see the data; it's in relation to the data.

Example 2.28. In the Philippines, there's a game similar to the powerball.

Now, we have $\binom{55}{6}$ lottery tickets. The winning numbers was $9, 18, 27, 36, 45, 54$. And in this case, there were 433 winners; people began to suspect rigging.

Then, we can model it as follows:

- $\mathcal{H}_0$: Not rigged.
- $\mathcal{H}_1$: Rigged.

Week 3
Third Week of Hell

3.1 Lecture – 2/3/2025

Special Lecture

There is a lecture on Turing's work from $3-4$ at Banatao (310 Sutardja) by Wigderson.

Let us look at $\mathbb{Z}/2 = \{0, 1\}$. Then, we define the addition operation on this group to be:

- $0 \oplus 0 = 0$.
- $0 \oplus 1 = 1$.
- $1 \oplus 0 = 1$.
- $1 \oplus 1 = 0$.

This operation is sometimes called XOR, where $0 \leftrightarrow$ `False` and $1 \leftrightarrow$ `True`. This is an important interpretation in Computer Science.

Today, we will be looking at plaintexts that are bit strings. In other words, each character $c \in \{0, 1\}$.

3.1.1 Vernam Cipher

This is a cipher which fell into literature around 1919.

Example 3.1 (Vernam Cipher Example)**.** The plaintext s is some bit-string. For example, let $s = 1011$.

We also need some secret key k which is also a bit-string. In our example, let $k = 001101$.

Then, we can perform the XOR operation on our plaintext and key to encrypt our plaintext.

In our example, since our key is longer than our plaintext, we would truncate the last digits which is longer than our plaintext. Thus, we would have:

$$s \oplus k = 1000$$

To decrypt, we note that we can simply XOR our ciphertext with the key to get the original plaintext

again. This is because:

$$\begin{aligned} c \oplus k &= (s \oplus k) \oplus k \\ &= s \oplus (k \oplus k) \\ &= s \oplus 0 \\ &= s \end{aligned}$$

We note that we can think of this as a "one-time pad."

Remark 3.2. In the case where the plaintext is longer than the key, there are a number of ways to deal with this:

- We may repeat the secret key.
  - However, if this is the case, there can be some flaws that can be exploited. Namely, we can send long messages to find a repetition.

Typically, we should want our key to be longer than the plaintext.

Now, a common weakness we've seen in previous ciphers is how they don't hide the frequency of the letters. However, for the Vernam Cipher, if the key k is random, then the ciphertext will look random.

To see this, we suppose that we have:

$$\begin{aligned} s &= x_1 x_2 \cdots x_n \\ k &= k_1 k_2 \cdots k_n \end{aligned}$$

Now, we note that each k_i is uniform; that is, there is a $\frac{1}{2}$ probability it's 1, and $\frac{1}{2}$ it's 0. We make no such assumption for our plaintext s.

Now, to show that the ciphertext will in fact look random, we look at the probability that $P(x_i \oplus k_i = 0)$, we can use conditioning to see:

$$\begin{aligned} P(x_i \oplus k_i = 0) &= P(k_i = 0 \mid x_i = 0)P(x_i = 0) + P(k_i = 1 \mid x_i = 1)P(x_i = 1) \\ &= \frac{1}{2}P(x_i = 0) + \frac{1}{2}P(x_i = 1) \\ &= \frac{1}{2} \end{aligned}$$

Then, we note that $\mathbb{E}\left[x_i \oplus k_i\right] = \frac{1}{2}$. Then, by linearity of expectations, if we sum over all the bits, we have:

$$\sum_{i=1}^{n} \mathbb{E}\left[x_i \oplus k_i\right] = \frac{n}{2} = \mathbb{E}\left[\text{number of } 0\text{'s in the ciphertext}\right]$$

Weaknesses

Definition 3.3 (Shannon's Notion of Perfect Security)**.** Suppose that Alice sends a plaintext, which we will model as a random variable X, to Bob. Then, we have $X \to f_k(X)$.

We will also have a secret key K which is another random variable. And finally, we have a random variable Y for our ciphertext.

Then, we say that this algorithm has perfect secrecy if $P(X = x \mid Y = y) = P(X = x)$. In other words, X and Y are independent.

Remark 3.4. When discussing perfect security, we are implicitly constraining the length of the texts.

Example 3.5 (Security of Ciphers)**.** We can look at the Scytale Cipher to figure out if it has perfect security or not.

One thing is that we know the ciphertext has the exact same number of letters (and letters) as our plaintext. This means then that we can remove certain possible plaintexts from our possible plaintexts.

In other words, we have $P(X = x \mid Y = y) \neq P(X = x)$.

Similarly, for substitution ciphers, the frequency distribution of the ciphertext is the same as the plaintext. Thus, they aren't independent.

Now, we make the following claim:

Proposition 3.6. The Vernam Cipher is perfectly secure.

Proof. To begin with, we want to find $P(X = x_1x_2 \cdots x_n \mid Y = y_1y_2 \cdots y_n)$.

Then, we want to instead show that:

$$P(X = x_1 \cdots x_n, Y = y_1, \cdots, y_n) = P(X = x_1 \cdots x_n)P(Y = y_1 \cdots y_n)$$

Then, note that $Y = X \oplus K$. So, this is equal to:

$$\begin{aligned} P(X = x_1 \cdots x_n, K = (y_1 \oplus x_1) \cdots (y_n \oplus y_n)) &= P(X = x_1 \cdots x_n)P(k_1 = y_1 \oplus x_1 \cdots y_n \oplus x_n) \\ &= P(X = x_1 \cdots x_n) \prod_{i=1}^{n} P(k_i = y_i \oplus k_i) \\ &= \frac{1}{2^n} P(X = x_1 \cdots x_n) \\ P(Y = y_1 \cdots y_n)P(X = x_1 \cdots x_n) &= \frac{1}{2^n} P(X = x_1 \cdots x_n) \end{aligned}$$

■

Now, note that although the cipher is perfectly secure, it isn't perfect; there are still flaws to it.

For example, if we re-use a key, we see then that Eve has c_1, c_2. Then, we have:

$$\begin{aligned} (c_1 \oplus c_2) &= (p_1 \oplus k) \oplus (p_2 \oplus k) \\ &= p_1 \oplus p_2 \end{aligned}$$

Then, we note that c_1, c_2 are in-depth if encrypted with the same k.