Motivation and Definitions

A **finite discrete probability space** is a pair \( (S, Pr) \) for some finite set \( S \) and some function \( Pr : S \to \mathbb{R} \), where for all \( x \in S \), \( Pr(x) > 0 \), and \( \sum_{x \in S} Pr(x) = 1 \). \( S \) is called the *sample space*, and the elements of \( S \) are called *outcomes*. \( Pr \) is called the *probability distribution function*.

Where does all this come from? Well, the *sample space* \( S \) is supposed to represent a set of mutually exclusive, mutually exhaustive outcomes - i.e. the set of all things that can happen.

To understand the *probability distribution function*, let’s think about our sample space as a box of area 1. Then, each outcome occupies a certain amount of *area* within the sample space. That is, the area is *distributed* among the outcomes according to the probability distribution function! The larger the area, the more likely the outcome is. For example,

<table>
<thead>
<tr>
<th>It rains</th>
<th>It doesn’t rain</th>
</tr>
</thead>
</table>

Here, we have that each outcome occupies the same amount of area in the box - i.e. each is equally likely. Because we have 2 outcomes, we have that the probability of each outcome is \( \frac{1}{2} \).

Let’s take a look at another way we could divide up the area of the box among the outcomes.

<table>
<thead>
<tr>
<th>It rains</th>
<th>It doesn’t rain</th>
</tr>
</thead>
</table>

Here, the probability of it raining is \( \frac{3}{4} \) so this outcome takes up \( \frac{3}{4} \) of the area within
the box. The remaining amount of area is \( \frac{1}{4} \), so this is the probability that it doesn’t rain.

**TASK:** Explain how the fact that the outcomes must be **mutually exclusive** and **mutually exhaustive** play a role in how we distribute the area of the box among the outcomes.

**TASK:** If we have \( n \) outcomes in \( S \), and \( Pr \) assigns each outcome an equal amount of area within our box, what is the probability of a particular outcome?

Note that a \( Pr \) that assigns each each outcome an equal amount of area within our box - i.e. each outcome has the same probability - is called a **uniform distribution**.

**TASK:** Suppose \( S \) has \( n \) outcomes, and each has the same probability. Explain why the probability the outcome of an *event* \( E \), which is a subset of \( S \), is \( |E|/n \).

We can generalize the result you just proved like this. Consider an event \( E = \{o_1...o_n\} \). The *probability of this event*, denoted \( Pr(E) \), is equal to \( Pr(o_1) + Pr(o_2) + ... + Pr(o_n) \). This is because the amount of area an event takes up is just the sum of the outcomes in the event.

**TASK:** Suppose \( p \) is the probability that event \( A \) will happen. What is the probability of \( A^c \) (where \( A^c \) denotes all outcomes in \( S \) not in \( A \))? 

**Checkpoint 1:** Call over a TA.
Conditional Probability and Independence

We sometimes want to know what the probability of something is, given that we know a certain subset of outcomes has happened (i.e. no outcome outside that subset can happen).

**TASK:** Prof. Littman has 2 coins: one fair, and one double heads. He picks one uniformly at random (i.e. each coin has probability of 1/2 of being picked), but he doesn’t tell us which one he picks. Prof. Littman then flips the coin he picks, and he then shouts out the result.

Suppose Prof. Littman shouts out Heads. What is the probability he flipped the fair coin, given that we know the coin flip resulted in Heads? *Hint:* How many ways can the coin flip result in heads, and how many of them start with the fair coin?

We can generalize our work here. Given two events $A, B \subseteq S$ where $Pr(B) > 0$, we define the *conditional probability of $A$ given $B$* as

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

We know that the only possible outcomes are those in $B$ (since $B$ is supposed to have happened). We therefore make $Pr(B)$ our denominator to renormalize our universe.

Looking at this formula, we also get a general definition for $Pr(A \cap B)$ (the probability of $A$ and $B$). We have that

$$
Pr(A \cap B) = Pr(A)Pr(B \mid A) = Pr(B)Pr(A \mid B).
$$

There are certain special situations where $Pr(B \mid A) = Pr(B)$; that is, where the fact that $A$ happened does not affect the probability of $B$ happening. We have a name for this situation.

**Definition:** Two events $A, B \subseteq S$ are *(pairwise)* independent if

$$
Pr(A \mid B) = Pr(A)
$$

If $Pr(B) = 0$, $B$ and any other event are independent.

Equivalently, $A$ and $B$ are independent if

$$
Pr(A \cap B) = Pr(A)Pr(B).
$$

**TASK:** Explain why these definitions are equivalent.
**TASK:** If two events $A$ and $B$ have an empty intersection, what is the probability of $A$ and $B$?

For each of the following pairs of events $A$ and $B$, identify whether they are independent, and justify why or why not.

i. Suppose we roll a fair die, and suppose $A =$ rolling an even number, and $B =$ rolling a number greater than three.

ii. Suppose we flip a fair coin three times, and suppose $A =$ the first coin is a tails, and $B =$ there is a run of exactly two heads (i.e. two, but not three, heads are flipped in a row).

**Checkpoint 2:** Call over a TA.
The Relationship Between Counting and Probability

Questions about probability are often closely tied to questions about counting. Often, we need to count two things—the set of all possible outcomes, and the set of outcomes we are considering.

**TASK:** Suppose $S$ is the set of all binary strings of length $n$, and the outcomes in $S$ are uniformly distributed. What is the probability of getting a string with $k$ 1’s (where $0 \leq k \leq n$)?

**TASK:** For what value(s) of $k$ is the probability of getting a string with $k$ 1’s the largest? Try it out for some values and see if you find a pattern.

**TASK:** Explain how we can use the work we just did to determine probabilities associated with flipping a fair coin $n$ times.

**Checkpoint 3:** Call over a TA.
Random Variables

From a probability space \((S, p)\), we can create a new probability space \((S', p')\), where \(S'\) is a partition of \(S\) - i.e. \(S' = \{P_1, P_2, ..., P_n\}\) - and where \(p' = Pr(P_i)\) for all \(i\).

Visually, we can take a box partitioned into \(m\) outcomes, and then partition these \(m\) outcomes in \(n\) groups. The amount of area each group takes up is just the sum of the outcomes within the group, so this box with \(m\) groups is just like having a probability space with \(m\) outcomes.

This leads us to one of the biggest, most famous misnomers in all of mathematics: random variables. A random variable is a function from a set of outcomes \(S\), often to \(\mathbb{R}\) or \(\mathbb{Z}\). We can think of this random variable as partitioning \(S\), where each event within the partition gets assigned a value. The diagram below shows what we mean.

![Diagram of random variables](image)

Here \(o_1, ..., o_5\) are outcomes in \(S\), and we can see that they are partitioned in 3 groups. Our random variable assigns each group a particular number. Let’s walk through some other more concrete examples of random variables.

We could have the random variable \(X\) on the sample space \{Heads, Tails\}, where \(X(\text{Heads}) = 1\), and \(X(\text{Tails}) = 0\).

**TASK:** If \(Pr(\text{Heads}) = 1/2\), and \(Pr(\text{Tails}) = 1/2\), then what is \(Pr(X = 1)\)? How about \(Pr(X = 0)\)?

We could also have the random variable \(Y\), where the domain of \(Y\) is \(S\), all sequences of coin flips of length \(n\), and where \(Y(s) = \text{the number of Heads in } s\).

**TASK:** If \(S\) is uniformly distributed, then what is \(Pr(Y = k)\), where \(0 \leq k \leq n\)?

*Hint:* You’ve already computed this.
**TASK:** Let $C_i$ be a random variable for the $i$th coin flip, $s_i$, in our sequence, $s$, of coin flips of length $n$, where $C_i$(Heads) = 1 and $C_i$(Tails) = 0. Let $C(s) = C_1(s_1) + C_2(s_2) + \ldots + C_n(s_n)$. Explain why $C(s) = Y(s)$.

**TASK:** Explain why $Pr(C = k) = \binom{n}{k} Pr(C_i = 1)^k Pr(C_i = 0)^{n-k}$.

**TASK:** Compute $Pr(C \geq k)$.

**TASK:** Compute $Pr(C = k | C_1 = 1)$.

**Checkpoint 4:** Call over a TA.
Famous Problems

Problem 1

Suppose the probability of a child being born with brown hair is $\frac{1}{2}$ and a child being born with non-brown hair is $\frac{1}{2}$. Consider a family with 2 children.

a. Given that at least one of their children is brown-haired, what is the probability the family has 2 brown-haired children?

b. Given that at least one of their children is brown-haired and born on a Tuesday, what is the probability the family has 2 brown-haired children? You may assume that the probability a child is born on a given day of the week is $\frac{1}{7}$.

Problem 2

A study claims that members of Group $B$ perform significantly better on tests than members of Group $A$, with 64% of Group $B$ passing as opposed to 38% of Group $A$. The full results of the study includes the following table:

<table>
<thead>
<tr>
<th></th>
<th>Group A</th>
<th>Group B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Prep</td>
<td>(\frac{16000}{20000} = 80%)</td>
<td>(\frac{15000}{20000} = 60%)</td>
</tr>
<tr>
<td>w/o Test Prep</td>
<td>(\frac{8000}{25000} = 30%)</td>
<td>(\frac{10000}{5000} = 20%)</td>
</tr>
<tr>
<td>Total</td>
<td>(\frac{11500}{30000} = 38%)</td>
<td>(\frac{16000}{30000} = 64%)</td>
</tr>
</tbody>
</table>

**TASK:** Give an interpretation that explains why members of Group $A$ actually perform better than members of Group $B$.

**TASK:** Explain how this situation can be generalized to situations where data could be used to mislead people.
Problem 3

You are on a game show, and on this game show, there are three doors. Behind two are goats, and behind the third is a shiny new car. You are asked to pick a door. Once you have made your selection, the game show host will open one of the remaining doors, revealing that it does not contain the shiny new car. The game show host then asks you if you would like to switch your selection to the other unopened door, or stay with your original choice. Should you switch? Use probabilistic concepts to make your decision.

Checkpoint 5: Call over a TA.