Intro to Probability

Michael L. Littman

CS 22 2020

April 3, 2020
Overview

Set Theory and Probability (17.4)
- Probability Spaces (17.4.1)
- Probability Rules from Set Theory (17.4.2)
- Uniform Probability Spaces (17.4.3)
- Infinite Probability Spaces (17.4.4)
Definitions of probability spaces

**Definition:** A countable *sample space* $S$ is a nonempty countable set.

**Definition:** An element $\omega \in S$ is called an *outcome*.

**Definition:** A subset of $S$ is called an *event*.

**Definition:** A *probability function* on a sample space $S$ is a function $\Pr : S \rightarrow \mathbb{R}$ such that

- $\Pr[\omega] \geq 0$ for all $\omega \in S$, and
- $\sum_{\omega \in S} \Pr[\omega] = 1$.

**Definition:** A sample space together with a probability function is called a *probability space*. For any event $E \subseteq S$, the probability of $E$ is defined to be the sum of the probabilities of outcomes in $E$:

$$\Pr[E] := \sum_{\omega \in E} \Pr[\omega].$$
Extensions

Probability theory and set theory and counting fit together really well.

We just defined probability for finite probability spaces. Can be defined for uncountable sets like the set of real numbers.

We’ll extend to countably infinite, like the integers.
Sum rule

**Rule:** If \( \{E_0, E_1, \ldots, \} \) is collection of disjoint events, then

\[
\Pr \left[ \bigcup_{n \in \mathbb{N}} E_n \right] = \sum_{n \in \mathbb{N}} \Pr[E_n].
\]

Like the sum rule in counting.

Example: If a donut shop produces 60% chocolate glazed, 10% creme filled, and 30% vanilla filled, then the probability of filled donuts is 40%, the sum of the probabilities of the two filled styles.
Complement rule

<table>
<thead>
<tr>
<th>set</th>
<th>name</th>
<th>Pr</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>chocolate</td>
<td>0.60</td>
</tr>
<tr>
<td>$B$</td>
<td>jelly</td>
<td>0.25</td>
</tr>
<tr>
<td>$A \cap B$</td>
<td>both</td>
<td>0.15</td>
</tr>
<tr>
<td>$\overline{A} \cap \overline{B}$</td>
<td>neither</td>
<td>0.30</td>
</tr>
</tbody>
</table>

Since we know that either $A$ and $\overline{A}$ are disjoint and cover all possibilities, the sum rule tells us that $\Pr[A] + \Pr[\overline{A}] = 1$.

**Rule:** $\Pr[\overline{A}] = 1 - \Pr[A]$.

Example: The chance that someone doesn’t like chocolate is (in symbols) $\Pr[\overline{A}] = 1 - \Pr[A] = 0.40$. 
Difference Rule

<table>
<thead>
<tr>
<th>set</th>
<th>name</th>
<th>Pr</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>chocolate</td>
<td>0.60</td>
</tr>
<tr>
<td>$B$</td>
<td>jelly</td>
<td>0.25</td>
</tr>
<tr>
<td>$A \cap B$</td>
<td>both</td>
<td>0.15</td>
</tr>
<tr>
<td>$\overline{A} \cap \overline{B}$</td>
<td>neither</td>
<td>0.30</td>
</tr>
</tbody>
</table>

**Rule:** $\Pr[B - A] = \Pr[B] - \Pr[A \cap B]$

Example: The chance that someone likes chocolate but not jelly is (in symbols) $\Pr[A - B] = \Pr[A] - \Pr[A \cap B] = 0.45$.

Proof: Follows from the Sum Rule because $B$ is the union of the disjoint sets $B - A$ and $A \cap B$. 
Inclusion-Exclusion

<table>
<thead>
<tr>
<th>set</th>
<th>name</th>
<th>Pr</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>chocolate</td>
<td>0.60</td>
</tr>
<tr>
<td>$B$</td>
<td>jelly</td>
<td>0.25</td>
</tr>
<tr>
<td>$A \cap B$</td>
<td>both</td>
<td>0.15</td>
</tr>
<tr>
<td>$\overline{A} \cap \overline{B}$</td>
<td>neither</td>
<td>0.30</td>
</tr>
</tbody>
</table>

**Rule:** $\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B]$

Example: The chance that someone likes either chocolate or jelly is (in symbols) $\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B] = 0.70$.

Proof: Follows from the Sum and Difference Rules, because $A \cup B$ is the union of the disjoint sets $A$ and $B - A$. 
Boole’s Inequality

<table>
<thead>
<tr>
<th>set</th>
<th>name</th>
<th>Pr</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>chocolate</td>
<td>0.60</td>
</tr>
<tr>
<td>$B$</td>
<td>jelly</td>
<td>0.25</td>
</tr>
<tr>
<td>$A \cap B$</td>
<td>both</td>
<td>0.15</td>
</tr>
<tr>
<td>$\bar{A} \cap \bar{B}$</td>
<td>neither</td>
<td>0.30</td>
</tr>
</tbody>
</table>

**Rule:** $\Pr[A \cup B] \leq \Pr[A] + \Pr[B]$

Example: The chance that someone likes either chocolate or jelly is (in symbols) $\Pr[A \cup B] \leq \Pr[A] + \Pr[B] = 0.85$.

Proof by inclusion-exclusion and the fact that probabilities are non-negative.
Monotonicity Rule

<table>
<thead>
<tr>
<th>set</th>
<th>name</th>
<th>Pr</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>chocolate</td>
<td>0.60</td>
</tr>
<tr>
<td>$B$</td>
<td>jelly</td>
<td>0.25</td>
</tr>
<tr>
<td>$A \cap B$</td>
<td>both</td>
<td>0.15</td>
</tr>
<tr>
<td>$\overline{A} \cap \overline{B}$</td>
<td>neither</td>
<td>0.30</td>
</tr>
</tbody>
</table>

**Rule:** If $A \subseteq B$, then $\Pr[A] \leq \Pr[B]$

Example: The chance that someone likes chocolate must at least as big as the chance they like both chocolate and jelly.

Proof: $\Pr[B] = \Pr[A \cup (B - A)] = \Pr[A] + \Pr[B - A] \geq \Pr[A]$. First equality because $A \subseteq B$. Then, they add together by the sum rule because the two sets are disjoint. Then, the last inequality holds because probabilities are non-negative.
Conjunction fallacy

From Kahneman and Tversky (1982):

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

Which is more probable?

1. Linda is a bank teller.
2. Linda is a bank teller and is active in the feminist movement.
3. Same.
4. Can’t tell.

Vote?
Union bound

Rule:

\[
\Pr[E_1 \cup E_2 \cup \cdots \cup E_n] \leq \Pr[E_1] + \Pr[E_2] + \cdots + \Pr[E_n].
\]

Example: The probability that a student has conflict with an exam is 0.001. What’s the probability that any of 300 students have a conflict? Can’t assume independence because groups of students take classes together, do sports together. Can’t get an exact answer with the information provided.

Union bounds says probability of student 1 or 2 or 3 ... 300 less than or equal to 300 \times 0.001 = 0.3.

Used in machine learning all the time.
Uniform

**Definition**: A finite probability space $S$ is said to be *uniform* if $Pr[\omega]$ is the same for every outcome $\omega \in S$.

In finite spaces, for any $E \subseteq S$,

$$Pr[E] = \frac{|E|}{|S|}.$$ 

Examples: Sides of a die, cards in a deck.

Contrast with: Vowels vs. consonants, primes vs. composites.
Counting example

What’s the probability that 5 coin flips leads to a palindromic sequence?

What’s the space of possibilities $S$? The results of 5 coin flips: HTTHH. $|S| = 2^5 = 32$.

What’s the event of interest $E$? Palindromic results: TTHTT. $|E| = 2^3 = 8$. That’s because the first 3 flips are “free”, then the 4th flip must match the 2nd and the fifth flip must match the first.

The probability, therefore, is $|E|/|S| = 2^3 / 2^5 = 1 / 2^2 = 1/4$. 
Three-sided coin

If we want a uniform distribution over two options, we can flip a coin (H/T). If we want a uniform distribution over four options, we can flip two coins (HH/HT/TH/TT). What if we want a uniform distribution over three options?

We could flip two coins and say it’s option 1 if both heads, option 2 if both tails, and option 3 if mismatch. Problem? Yes. The probability of mismatch is $\Pr[HT] + \Pr[TH] = \frac{1}{2}$. Not $\frac{1}{3}-\frac{1}{3}-\frac{1}{3}$.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HH</td>
<td>1/4</td>
</tr>
<tr>
<td>HT</td>
<td>1/4</td>
</tr>
<tr>
<td>TH</td>
<td>1/4</td>
</tr>
<tr>
<td>TT</td>
<td>1/4</td>
</tr>
</tbody>
</table>
Repeat the trial

We could flip two coins and say it’s option 1 if HH, option 2 if TT, option 3 if HT, and do over if TH. Problem? Maybe. There is an infinite number of outcomes...

Procedure selects option 1 if:

- HH on the first trial
- or TH on the first trial and HH on the second trial
- or TH on the first two trials and HH on the third trial
- ...

\[
\Pr(\text{option 1}) = \frac{1}{4} + \frac{1}{4} \times \frac{1}{4} + \frac{1}{4} \times \frac{1}{4} \times \frac{1}{4} + \ldots
\]

\[
= \sum_{i=1}^{\infty} \frac{1}{4^i}
\]

\[
= \frac{1}{4} \times \sum_{i=0}^{\infty} \frac{1}{4^i}
\]

\[
= \frac{1}{4} \left( \frac{1}{1-\frac{1}{4}} \right)
\]

\[
= \frac{1}{4} \times \frac{4}{3} = \frac{1}{3}.
\]
Aside: Geometric sum

\[ x = \sum_{i=0}^{\infty} p^i \]
\[ x = p^0 + p^1 + p^2 + p^3 + \ldots \] (the sum we want)
\[ px = p^1 + p^2 + p^3 + p^4 + \ldots \] (expand)
\[ p^0 + px = p^0 + p^1 + p^2 + p^3 + p^4 + \ldots \] (multiply by \( p \))
\[ p^0 + px = x \] (add \( p^0 \))
\[ p^0 = x - px \] (defn of \( x \))
\[ 1 = x(1 - p) \] (subtract \( px \))
\[ \frac{1}{1-p} = x \] (factor/simplify)
\[ \frac{1}{1-p} = x \] (divide by \( 1 - p \))
Infinite sample space

\[ S = \{ HH, HT, TT, TH : HH, TH : HT, TH : TT, TH : TH : HH, TH : TH : HT, TH : TH : TT, \ldots \} \]
\[ = \{ (TH)^n : HH, (TH)^n : HT, (TH)^n : TT | n \in \mathbb{N} \} \]

The probability space is:

\[ \Pr((TH)^n : HH) = \Pr((TH)^n : HT) = \Pr((TH)^n : TT) = 1/4^{n+1}. \]

Note: \[ \sum_{n=0}^{\infty} 3 \times 1/4^{n+1} = 3/4 \times \frac{1}{1-1/4} = 3/4 \times 4/3 = 1. \]

Non-negative and sums to one, valid probability space!