CS 440/ECE 448 Lecture 2: Probability

Mark Hasegawa-Johnson, 1/2022

All content CC-BY-4.0 unless otherwise noted

CC-BY-3.0 Image by Diacritica, 2010
https://commons.wikimedia.org/wiki/File:6sided_dice.jpg
Outline

• Motivation: Why use probability?
• Random variables
• The axioms of probability
• Conditional probability
• Mutually exclusive vs. Independent vs. Conditionally Independent
Why use probability?

• Stochastic environment: outcome of an action might be truly random.
• Multi-agent environment:
  • If other players are rational and their goals are known, then you don’t need probability; you just work out what their rational actions will be.
  • If other players have unknown goals, then model them as random.
• Unknown environment: outcome of an action is not truly random, but you don’t know what the outcome will be.
  • In this case, “probability” measures your belief: $P(Q|A)=$the degree to which you believe that action $A$ will produce outcome $Q$.
• Computational complexity:
  • Instead of searching $10^{100}$ possible outcomes, you could randomly choose 1000 paths to try, and then choose the best of those.
Why NOT use probability?

• Multi-agent environment:
  • Maybe it’s better to find out what the other players really want?

• Unknown environment:
  • Maybe it’s better to learn the rules of the game?

• Computational complexity:
  • Maybe it’s better to do a complete search, instead of just a partial search?

Notice: these are quantitative questions. “Better” requires some metric: how much better, and with what probability?
What is probability?

• Latin *probabilis* = probable, commendable, believable, from *probare* = to test something
• If tested, it will (probably) turn out to be true
Outline

• Motivation: Why use probability?
• Random variables
• The axioms of probability
• Conditional probability
• Mutually exclusive vs. Independent vs. Conditionally Independent
Random variables

• A random variable is a function that maps from the outcomes of an experiment to a set of values
• Example: throw four dice, all different colors. X = number of pips showing on the green die.
• Then run the experiment...

• In this particular outcome, X=3.
• In some other outcome, X would have taken a different value.
• We totally ignore aspects of the outcome that are irrelevant to X, e.g., the pips on the red, purple, and blue dice.
Notation: $P(X = x)$ is a number

- Capital letters are random variables. Small letters are values that the random variable might take.
- "$X = 3$" is a possible outcome of the experiment, which we call an "event." We denote the probability of that event as $P(X = 3)$:

$$P(X = 3) = \frac{1}{6}$$
Notation: $P(X)$ is a table

Let’s use $P(X)$ to mean the complete probability table, specifying $P(X = x)$ for all possible values of $x$:

$$P(X) = \begin{array}{ccccccc}
 x & 1 & 2 & 3 & 4 & 5 & 6 \\
 P(X = x) \hline
 1 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\
\end{array}$$

We call this table of numbers, $P(X)$, a **probability distribution**.
Outline

• Motivation: Why use probability?
• Random variables
• The axioms of probability
• Conditional probability
• Mutually exclusive vs. Independent vs. Conditionally Independent
Outcomes, events, and binary random variables

• Let’s define an ”outcome” to be a particular result of the experiment (e.g., particular settings for the red, blue, purple and green dice).

• Let’s define an “event” to be a set of possible outcomes. For example, the event $X = 3$ is defined to be the set of all outcomes in which $X = 3$.

• A binary random variable is equal to true if the event occurred, and false otherwise. For example, we could define a binary random variable $A$ in this way:

$$ A = \begin{cases} 
T & \text{if } X = 3 \\
F & \text{if } X \neq 3 
\end{cases} $$
Outcomes, events, and binary random variables

• We will do a LOT in this class with binary random variables, because they are so useful. For example, consider a weather-prediction task:

$$A = \begin{cases} T & \text{if it is raining} \\ F & \text{otherwise} \end{cases}$$

• A binary random variable is an indicator of an event. The random variable is “T” if the event occurs, and “F” otherwise.
The axioms of probability

Axiom 1: every event has a non-negative probability.
\[ P(A = T) \geq 0 \]

Axiom 2: If an event always occurs, we say it has probability 1.
\[ \Omega = \begin{cases} T & \text{always} \\ F & \text{never} \end{cases} \]
\[ P(\Omega = T) = 1 \]

Axiom 3: probability measures behave like set measures.
\[ P(A \lor B = T) = P(A = T) + P(B = T) - P(A \land B = T) \]
Axiom 3: probability measures behave like set measures.

Area of the whole rectangle is $P(\Omega = T) = 1$.

Area of this circle is $P(A = T)$.

Area of this circle is $P(B = T)$.

Area of their intersection is $P(A \cap B = T)$.

Area of their union is

$$P(A \cup B = T) = P(A = T) + P(B = T) - P(A \cap B = T)$$
Example

• A = “it will rain tomorrow.” Suppose $P(A = T) = 0.4$.
• B = “it will snow tomorrow.” Suppose $P(B = T) = 0.2$.
• $A \land B = “it will both rain and snow tomorrow.”$ Suppose $P(A \land B = T) = 0.1$

Then the probability that it will either rain or snow tomorrow is

$$P(A \lor B = T) =$$

$$P(A = T) + P(B = T) - P(A \land B = T) =$$

$$0.4 + 0.2 - 0.1 =$$

$$0.5$$
Outline

• Motivation: Why use probability?
• The axioms of probability
• Random variables
• Conditional probability
• Mutually exclusive vs. Independent vs. Conditionally Independent
Joint and Conditional probabilities: definitions

- \( P(A \land B = T) \) is the probability that both event A and event B happen. This is called their **joint probability**.
- \( P(B = T | A = T) \) is the probability that event B happens, given that event A happens. This is called the **conditional probability** of B given A.

**Example:**
- A = “it will rain tomorrow”
- B = “it will snow tomorrow”
- \( P(A \land B = T) \) = probability that it will both snow and rain
- \( P(B = T | A = T) \) = probability that it will snow, given that it rains
Joint probabilities are usually given in the problem statement.

Area of the whole rectangle is $P(\text{True}) = 1$.

Suppose $P(A = T) = 0.4$

Suppose $P(B = T) = 0.2$

Suppose $P(A \land B = T) = 0.1$
Conditioning events change our knowledge!
For example, given that A is true...

Most of the events in this rectangle are no longer possible!

Only the events inside this circle are now possible.
Conditioning events change our knowledge! For example, given that \( A \) is true...

If \( A \) always occurs, then by the axioms of probability, the probability of \( A = T \) is 1. We can say that

\[
P(A = T | A = T) = 1.
\]

The probability of \( B \), given \( A \), is the size of the event \( A \land B \), expressed as a fraction of the size of the event \( A \):

\[
P(B = T | A = T) = \frac{P(A \land B = T)}{P(A = T)}
\]
Joint and Conditional distributions of random variables

• $P(X, Y)$ is the **joint probability distribution** over all possible outcomes $P(X = x, Y = y)$.

• $P(X|Y)$ is the **conditional probability distribution** of outcomes $P(X = x|Y = y)$.

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$
Joint and Conditional distributions of random variables

Example:

$X =$ number of pips on the bone die.

$Y = X \text{ modulo } 2.$

The joint probability $P(X = 5, Y = 1) = \frac{1}{6}$.

Their joint distribution is:

$$P(X, Y) = \begin{array}{c|c|c|c|c|c|c|c}
 y & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
\hline
 0 & 0 & \frac{1}{6} & 0 & \frac{1}{6} & 0 & \frac{1}{6} & 0 \\
 1 & \frac{1}{6} & 0 & \frac{1}{6} & 0 & \frac{1}{6} & 0 & 0 \\
\end{array}$$
Joint and Conditional distributions of random variables

- Suppose we’re given the complete table $P(X = x, Y = y)$, and we want to find $P(X = 5|Y = 1)$. How do we do that?

<table>
<thead>
<tr>
<th>$P(X = x, Y = y)$</th>
<th>$x$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>$\frac{1}{6}$</td>
<td>$\frac{1}{6}$</td>
<td>$\frac{1}{6}$</td>
<td>$\frac{1}{6}$</td>
<td>$\frac{1}{6}$</td>
<td>$\frac{1}{6}$</td>
<td>$\frac{1}{6}$</td>
</tr>
</tbody>
</table>

Bone die found at Cantonment Clinch. CC-BY-3.0, Colby Kirk, 2007.
Joint and Conditional distributions of random variables

- Suppose we’re given the complete table $P(X = x, Y = y)$, and we want to find $P(X = 5|Y = 1)$. How do we do that?
- Well, we know that the event $Y = 1$ occurred, so we eliminate all outcomes in which $Y \neq 1$

<table>
<thead>
<tr>
<th>$P(X = x, Y = y)$</th>
<th>$x$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>$y$</td>
<td>1/6</td>
</tr>
</tbody>
</table>

Bone die found at Cantonment Clinch. CC-BY-3.0, Colby Kirk, 2007.
Joint and Conditional distributions of random variables

- Suppose we’re given the complete table $P(X = x, Y = y)$, and we want to find $P(X = 5|Y = 1)$. How do we do that?
- Well, we know that the event $Y = 1$ occurred, so we eliminate all outcomes in which $Y \neq 1$.
- But we know that the sum of all entries should be $P(\text{True})=1$, so we renormalize the table so that it adds up to 1.

$P(X|Y = 1) = \frac{P(X = x, Y = y)}{P(Y = 1)}$
Joint and Conditional distributions of random variables

- Thus, the conditional probability is 
  \[ P(X = 5|Y = 1) = \frac{1}{3}. \]
Outline

• Motivation: Why use probability?
• Random variables
• The axioms of probability
• Conditional probability
• Mutually exclusive vs. Independent vs. Conditionally Independent
Mutually exclusive events

Mutually exclusive events never occur simultaneously:
\[ P(A \lor B = T) = P(A = T) + P(B = T) - P(A \land B = T) \]
\[ = P(A = T) + P(B = T) \]
Examples of mutually exclusive events

• If A is the event “tomorrow it rains,” and B is the event “tomorrow it does not rain,” then A and B are mutually exclusive.
• If A is the event “the number on the die is 5 or 6,” and B is the event “the number on the die is 1 or 2”, then A and B are mutually exclusive.
Independent events

Independent events occur with equal probability, regardless of whether or not the other event has occurred:

\[ P(A = T | B = T) = P(A = T) \]
Examples of independent events

• If A is the event “it rains,” and B is the event “the stock market goes down,” those are probably independent events.

• If A is the event “the number on the green die is 5”, and B is the event “the number on the red die is 2”, those are probably independent events.
Independent events

We can re-write the definition of independent events in an interesting and useful way, by plugging in the definition of conditional probability:

\[ P(A = T|B = T) = \frac{P(A \cap B = T)}{P(B = T)} = P(A = T) \]
Independent events: A useful alternate definition

Re-arranging terms in the previous slide gives us this alternative definition of independent events:

\[ P(A \land B = T) = P(A = T)P(B = T) \]
Independent vs. Mutually Exclusive

- **Independent events:**
  \[ P(A \land B = T) = P(A = T)P(B = T) \]

- **Mutually exclusive events:**
  \[ P(A \lor B = T) = P(A = T) + P(B = T) \]

Don’t confuse them! Mutually exclusive events are not independent. Quite the contrary. Think about the set pictures.

\[ P(A \land B = T) = \]

\[ 0 \]

\[ P(A \land B = T) = P(A = T)P(B = T) \]
Conditionally independent events

Events A and B are conditionally independent, given C, if

\[ P(A = T | B = T, C = T) = P(A = T | C = T) \]
Events $A$ and $B$ are conditionally independent, given $C$, if

$$P(A = T|B = T, C = T) = \frac{P(A \land B = T|C = T)}{P(B = T|C = T)} = P(A = T|C = T)$$
Conditionally independent events

Events A and B are conditionally independent, given C, if

\[ P(A \land B = T|C = T) = P(A = T|C = T)P(B = T|C = T) \]
Toothache = patient has a toothache

Cavity = the patient has a cavity

Catch = dentist’s probe catches on something in the mouth

Independence ≠ Conditional Independence
These Events are not Independent

• If the patient has a toothache, then it’s likely he has a cavity. Having a cavity makes it more likely that the probe will catch on something.
  
  \[ P(\text{Catch} = T|\text{Toothache} = T) > P(\text{Catch} = T) \]

• If the probe catches on something, then it’s likely that the patient has a cavity. If he has a cavity, then he might also have a toothache.
  
  \[ P(\text{Toothache} = T|\text{Catch} = T) > P(\text{Toothache} = T) \]

• So Catch and Toothache are not independent
...but they are Conditionally Independent

- Here are some reasons the probe might not catch, despite having a cavity:
  - The dentist might be really careless
  - The cavity might be really small
- Those reasons have nothing to do with the toothache!
  \[ P(\text{Catch} = T|\text{Cavity} = T, \text{Toothache} = T) = P(\text{Catch} = T|\text{Cavity} = T) \]
- **Catch** and **Toothache** are **conditionally independent** given knowledge of **Cavity**
...but they are Conditionally Independent

These statements are all equivalent:

\[ P(\text{Catch} = T|\text{Cavity} = T, \text{Toothache} = T) = P(\text{Catch} = T|\text{Cavity} = T) \]
\[ P(\text{Toothache} = T|\text{Cavity} = T, \text{Catch} = T) = P(\text{Toothache} = T|\text{Cavity} = T) \]
\[ P(\text{Toothache} \land \text{Catch}|\text{Cavity}) = P(\text{Toothache}|\text{Cavity}) \cdot P(\text{Catch}|\text{Cavity}) \]

...and they all mean that \text{Catch} and \text{Toothache} are \textbf{conditionally independent} given knowledge of \text{Cavity}
Summary

• A random variable is a function that maps from the outcome of an experiment to a particular value.
• The axioms of probability are (1) every probability is non-negative, (2) an event that always occurs has probability 1.0, (3) probability measures behave like set measures.

\[
P(B = T | A = T) = \frac{P(A \land B = T)}{P(A = T)}
\]

• A and B are **mutually exclusive** iff \( P(A \land B = T) = 0 \)
• A and B are **independent** iff \( P(A = T | B = T) = P(A = T) \)
• A and B are **conditionally independent given C** iff:

\[
P(A = T | B = T, C = T) = P(A = T | C = T)
\]