# Formalizing Probability

What do we assign probability to?
Intuitively, we assign them to possible *events* (things that might happen, *outcomes* of an experiment)

Formally, we take a *sample space* to be a *set*.

- Intuitively, the sample space is the set of possible outcomes, or possible ways the world could be.

An *event* is a subset of a sample space.

We assign probability to events: that is, to subsets of a sample space.

Sometimes the hardest thing to do in a problem is to decide what the sample space should be.

- There's often more than one choice
- A good thing to do is to try to choose the sample space so that all outcomes (i.e., elements) are equally likely
    - This is not always possible or reasonable

# Choosing the Sample Space

**Example 1:** We toss a coin. What's the sample space?

- Most obvious choice: {heads, tails}
- Should we bother to model the possibility that the coin lands on edge?
- What about the possibility that somebody snatches the coin before it lands?
- What if the coin is biased?

**Example 2:** We toss a die. What's the sample space?

**Example 3:** Two distinguishable dice are tossed together. What's the sample space?

- (1,1), (1,2), (1,3), . . . , (6,1), (6,2), . . . , (6,6)

What if the dice are indistinguishable?

**Example 4:** You're a doctor examining a seriously ill patient, trying to determine the probability that he has cancer. What's the sample space?

**Example 5:** You're an insurance company trying to insure a nuclear power plant. What's the sample space?

# Probability Measures

A *probability measure* assigns a real number between 0 and 1 to every subset of (event in) a sample space.

- Intuitively, the number measures how likely that event is.
- Probability 1 says it's certain to happen; probability 0 says it's certain not to happen
- Probability acts like a *weight* or *measure*. The probability of separate things (i.e., disjoint sets) adds up.

Formally, a probability measure Pr on S is a function mapping subsets of $S$ to real numbers such that:

1. For all $A \subseteq S$, we have $0 \leq \Pr(A) \leq 1$
2. $\Pr(\emptyset) = 0$; $\Pr(S) = 1$
3. If $A$ and $B$ are disjoint subsets of $S$ (i.e., $A \cap B = \emptyset$), then $\Pr(A \cup B) = \Pr(A) + \Pr(B)$.

It follows by induction that if $A_1, \ldots, A_k$ are pairwise disjoint, then

$$\Pr(\cup_{i=1}^k A_i) = \Sigma_i^k \Pr(A_i).$$

- This is called *finite additivity*; it's actually more standard to assume a countable version of this, called *countable additivity*

In particular, this means that if $A = \{e_1, \ldots, e_k\}$, then

$$\Pr(A) = \sum_{i=1}^{k} \Pr(e_i).$$

In finite spaces, the probability of a set is determined by the probability of its elements.

# Equiprobable Measures

Suppose $S$ has $n$ elements, and we want Pr to make each element equally likely.

- Then each element gets probability $1/n$
- $\Pr(A) = |A|/n$

In this case, Pr is called an *equiprobable measure*.

**Example 1:** In the coin example, if you think the coin is fair, and the only outcomes are heads and tails, then we can take $S = \{\text{heads,tails}\}$, and
$\Pr(\text{heads}) = \Pr(\text{tails}) = 1/2$.

**Example 2:** In the two-dice example where the dice are distinguishable, if you think both dice are fair, then we can take $\Pr((i,j)) = 1/36$.

- Should it make a difference if the dice are indistinguishable?

# Equiprobable measures on infinite sets

Defining an equiprobable measure on an infinite set can be tricky.

**Theorem:** There is no equiprobable measure on the positive integers.

**Proof:** By contradiction. Suppose Pr is an equiprobable measure on the positive integers, and $\Pr(1) = \epsilon > 0$.

There must be some $N$ such that $\epsilon > 1/N$.
Since $\Pr(1) = \cdots = \Pr(N) = \epsilon$, we have

$$\Pr(\{1, \ldots, N\}) = N\epsilon > 1 \text{ — a contradiction}$$

But if $\Pr(1) = 0$, then $\Pr(S) = \Pr(1) + \Pr(2) + \cdots = 0$.

# Some basic results

How are the probability of $E$ and $\overline{E}$ related?

- How does the probability that the dice lands either 2 or 4 (i.e., $E = \{2, 4\}$) compare to the probability that the dice lands 1, 3, 5, or 6 ($\overline{E} = \{1, 3, 5, 6\}$)

**Theorem 1:** $\Pr(\overline{E}) = 1 - \Pr(E)$.

**Proof:** $E$ and $\overline{E}$ are disjoint, so that

$$\Pr(E \cup \overline{E}) = \Pr(E) + \Pr(\overline{E}).$$

But $E \cup \overline{E} = S$, so $\Pr(E \cup \overline{E}) = 1$.
Thus $\Pr(E) + \Pr(\overline{E}) = 1$, so

$$\Pr(\overline{E}) = 1 - \Pr(E).$$

**Theorem 2:** $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$.

$$\begin{aligned}
A &= (A - B) \cup (A \cap B) \\
B &= (B - A) \cup (A \cap B) \\
A \cup B &= (A - B) \cup (B - A) \cup (A \cap B)
\end{aligned}$$

So

$$\begin{aligned}
\Pr(A) &= \Pr(A - B) + \Pr(A \cap B) \\
\Pr(B) &= \Pr(B - A) + \Pr(A \cap B) \\
\Pr(A \cup B) &= \Pr(A - B) + \Pr(B - A) + \Pr(A \cap B)
\end{aligned}$$

The result now follows.

Remember the Inclusion-Exclusion Rule?

$$|A \cup B| = |A| + |B| - |A \cap B|$$

This follows easily from Theorem 2, if we take Pr to be an equiprobable measure. We can also generalize to arbitrary unions.

# Disclaimer

- Probability is a well defined mathematical theory.
- Applications of probability theory to “real world” problems is not.
- Choosing the sample space, the events and the probability function requires a “leap of faith”.
- We cannot prove that we chose the right model but we can argue for that.
- Some examples are easy some are not:
  - Flipping a coin or rolling a die.
  - Playing a lottery game.
  - Guessing in a multiple choice test.
  - Determining whether or not the patient has AIDS based on a test.
  - Does the patient have cancer?

# Conditional Probability

One of the most important features of probability is that there is a natural way to *update* it.

**Example:** Bob draws a card from a 52-card deck. Initially, Alice considers all cards equally likely, so her probability that the ace of spades was drawn is 1/52. Her probability that the card drawn was a spade is 1/4.

Then she sees that the card is black. What should her probability now be that

- the card is the ace of spades?
- the card is a spade?

A reasonable approach:

- Start with the original sample space
- Eliminate all outcomes (elements) that you now consider impossible, based on the observation (i.e., assign them probability 0).
- Keep the relative probability of everything else the same.
  - Renormalize to get the probabilities to sum to 1.

What should the probability of $B$ be, given that you've observed $A$? According to this recipe, it's

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

$\Pr(\text{A}\spadesuit \mid \text{black}) = (1/52)/(1/2) = 1/26$
$\Pr(\text{spade} \mid \text{black}) = (1/4)/(1/2) = 1/2.$

A subtlety:

- What if Alice doesn't completely trust Bob? How do you take this into account? Two approaches:

  (1) Enlarge sample space to allow more observations.

  (2) Jeffrey's rule:

$$\Pr(\text{A}\spadesuit \mid \text{black}) \cdot \Pr(\text{Bob telling the truth}) +$$
$$\Pr(\text{A}\spadesuit \mid \text{red}) \cdot \Pr(\text{Bob lying}).$$

# The second-ace puzzle

Alice gets two cards from a deck with four cards:
A♠, 2♠, A♡, 2♡.

| | | |
|---|---|---|
| A♠ A♡ | A♠ 2♠ | A♠ 2♡ |
| A♡ 2♠ | A♡ 2♡ | 2♠ 2♡ |

Alice then tells Bob "I have an ace".

She then says "I have the ace of spades".

The situation is similar if if Alice says "I have the ace of hearts".

*Puzzle*: Why should finding out which particular ace it is raise the conditional probability of Alice having two aces?

# The Monty Hall Puzzle

- You're on a game show and given a choice of three doors.
  - Behind one is a car; behind the others are goats.
- You pick door 1.
- Monty Hall opens door 2, which has a goat.
- He then asks you if you still want to take what's behind door 1, or to take what's behind door 3 instead.

Should you switch?

# Justifying Conditioning

Suppose that after learning $B$, we update Pr to $\Pr_B$. Some reasonable assumptions:

C1. $\Pr_B$ is a probability distribution

C2. $\Pr_B(B) = 1$

C3. If $C_1, C_2 \subseteq B$, then $\Pr_B(C_1)/\Pr_B(C_2) = \Pr(C_1)/\Pr(C_2)$.

- The relative probability of subsets of $B$ doesn't change after learning $B$.

**Theorem:** C1–C3 force conditioning:

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

**Proof:** By C1, $\Pr_B(B) + \Pr_B(\overline{B}) = 1$.
By C2, $\Pr_B(B) = 1$, so $\Pr_B(\overline{B}) = 0$.
General property of probability:
If $C \subseteq C'$, then $\Pr_B(C) \leq \Pr_B(C')$ (Why?)
So if $C \subseteq \overline{B}$, then $\Pr_B(C) = 0$

By C1,

$$\Pr_B(A) = \Pr_B(A \cap B) + \Pr_B(A \cap \overline{B}) = \Pr_B(A \cap B).$$

Since $A \cap B \subseteq B$, by C3,

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

By C3, $\Pr_B(B) = 1$, so $\Pr_B(A \cap B) = \Pr(A \cap B)/\Pr(B)$.

# The Second-Child Problem

Suppose that any child is equally likely to be male or female, and that the sex of any one child is independent of the sex of the other. You have an acquaintance and you know he has two children, but you don't know their sexes. Consider the following four cases:

1. You visit the acquaintance, and a boy walks into the room. The acquaintance says "That's my older child."
2. You visit the acquaintance, and a boy walks into the room. The acquaintance says "That's one of my children."
3. The acquaintance lives in a culture, where male children are always introduced first, in descending order of age, and then females are introduced. You visit the acquaintance, who says "Let me introduce you to my children." Then he calls "John [a boy], come here!"
4. You go to a parent-teacher meeting. The principal asks everyone who has at least one son to raise their hands. Your acquaintance does so.

In each case, what is the probability that the acquaintance's second child is a boy?

- The problem is to get the right sample space

# Independence

Intuitively, events $A$ and $B$ are independent if they have no effect on each other.

This means that observing $A$ should have no effect on the likelihood we ascribe to $B$, and similarly, observing $B$ should have no effect on the likelihood we ascribe to $A$.

Thus, if $\Pr(A) \neq 0$ and $\Pr(B) \neq 0$ and $A$ is independent of $B$, we would expect

$$\Pr(B|A) = \Pr(B) \text{ and } \Pr(A|B) = \Pr(A).$$

Interestingly, one implies the other.

$\Pr(B|A) = \Pr(B)$ iff $\Pr(A \cap B)/\Pr(A) = \Pr(B)$ iff

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

Formally, we say $A$ and $B$ are *(probabilistically) independent* if

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

This definition makes sense even if $\Pr(A) = 0$ or $\Pr(B) = 0$.

# Example

Alice has two coins, a fair one $f$ and a loaded one $l$.

- $l$'s probability of landing $H$ is $p > 1/2$.

Alice picks $f$ and flips it twice.

- What is the sample space?

$$\Omega = \{(H,H),(H,T),(T,H),(T,T)\}.$$

- What is Pr?
- By symmetry this should be an equiprobable space.

Let $H_1 = \{(H,H),(H,T)\}$ and let $H_2 = \{(H,H),(T,H)\}$.

$H_1$ and $H_2$ are independent:

- $H_1 = \{(H,H),(H,T)\} \Rightarrow \Pr(H_1) = 2/4 = 1/2$.
- Similarly, $P(H_2) = 1/2$.
- $H_1 \cap H_2 = \{(H,H)\} \Rightarrow \Pr(H_1 \cap H_2) = 1/4$.
- So, $\Pr(H_1 \cap H_2) = \Pr(H_1) \cdot \Pr(H_2)$.

Alice next picks $l$ and flips it twice.

- The sample space is the same as before:

$$\Omega = \{(H,H),(H,T),(T,H),(T,T)\}.$$

- We now define Pr by *assuming* the flips are independent:
  - $\Pr\{(H,H)\} = \Pr(H_1 \cap H_2) := p^2$
  - $\Pr\{(H,T)\} = \Pr(H_1 \cap \bar{H}_2) := p(1-p)$
  - $\Pr\{(T,H)\} = \Pr(\bar{H}_1 \cap H_2) := (1-p)p$
  - $\Pr\{(T,T)\} = \Pr(\bar{H}_1 \cap \bar{H}_2) := (1-p)^2$.
- $H_1$ and $H_2$ are now independent by construction:

$$\begin{aligned} \Pr(H_1) &= \Pr\{(H,H),(H,T)\} = \\ &= p^2 + p(1-p) = p[p+(1-p)] = p. \end{aligned}$$

  Similarly, $\Pr(H_2) = \Pr\{(H,H),(T,H)\} = p$
  $\Pr(H_1 \cap H_2) = \Pr(H,H) = p^2$.

- For either coin, the two flips are independent.

What if Alice randomly picks a coin and flips it twice?

- What is the sample space?

  $\Omega = \{(f,(H,H)),(f,(H,T)),(f,(T,H)),(f,(T,T)),$
  $(l,(H,H)),(l,(H,T)),(l,(T,H)),(l,(T,T))\}.$

  The sample space has to specify which coin is picked!

- How do we construct Pr?
- E.g.: $\Pr(f, H, H)$ should be probability of getting the fair times the probability of getting heads with the fair coin: $1/2 \times 1/4$
- Follows from the following general result:

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

- So with $F$, $L$ denoting the events $f$ (respectively, $l$) was picked,

$$\begin{aligned}\Pr\{(f,(H,H))\} &= \Pr(F \cap (H_1 \cap H_2)) \\ &= \Pr(H_1 \cap H_2|F)\Pr(F) \\ &= 1/2 \cdot 1/2 \cdot 1/2.\end{aligned}$$

  Analogously, we have for example

$$\Pr\{(l,(H,T))\} = p(1-p) \cdot 1/2.$$

Are $H_1$ and $H_2$ independent now?

**Claim.** $\Pr(A) = \Pr(A|E)\Pr(E) + \Pr(A|\bar{E})\Pr(\bar{E})$

**Proof.** $A = (A \cap E) \cup (A \cap \bar{E})$, so

$$\Pr(A) = \Pr(A \cap E) + \Pr(A \cap \bar{E}).$$

$\Pr(H_1) = \Pr(H_1|F)\Pr(F) + \Pr(H_1|L)\Pr(L) = p/2 + 1/4.$

Similarly, $\Pr(H_2) = p/2 + 1/4$.

However,

$$\begin{aligned} & \Pr(H_1 \cap H_2) \\ = \ & \Pr(H_1 \cap H_2|F)\Pr(F) + \Pr(H_1 \cap H_2|L)\Pr(L) \\ = \ & p^2/2 + 1/4 \cdot 1/2 \\ \neq \ & (p/2 + 1/4)^2 \\ = \ & \Pr(H_1) \cdot \Pr(H_2). \end{aligned}$$

So $H_1$ and $H_2$ are dependent events.

# Probability Trees

Suppose that the probability of rain tomorrow is .7. If it rains, then the probability that the game will be cancelled is .8; if it doesn't rain, then the probability that it will be cancelled is .1. What is the probability that the game will be played?

The situation can be described by a tree:

Similar trees can be used to describe

- Sequential decisions
- Randomized algorithms

# Bayes' Theorem

**Bayes Theorem:** Let $A_1, \ldots, A_n$ be mutually exclusive and exhaustive events in a sample space $S$.

- That means $A_1 \cup \ldots \cup A_n = S$, and the $A_i$'s are pairwise disjoint: $A_i \cap A_j = \emptyset$ if $i \neq j$.

Let $B$ be any event in $S$. Then

$$\Pr(A_i|B) = \frac{\Pr(A_i)\Pr(B|A_i)}{\Sigma_{j=1}^{n} \Pr(A_j)\Pr(B|A_j)}.$$

**Proof:**

$$B = B \cap S = B \cap (\cup_{j=1}^{n} A_j) = \cup_{i=1}^{n}(B \cap A_j).$$

Therefore, $\Pr(B) = \Sigma_{j=1}^{n} \Pr(B \cap A_j)$.

Next, observe that $\Pr(B|A_i) = \Pr(A_i \cap B)/\Pr(A_i)$. Thus,

$$\Pr(A_i \cap B) = \Pr(B|A_i)\Pr(A_i).$$

Therefore,

$$\begin{aligned}
\Pr(A_i|B) &= \tfrac{\Pr(A_i \cap B)}{\Pr(B)} \\
&= \tfrac{\Pr(B|A_i)\Pr(A_i)}{\Sigma_{j=1}^{n} \Pr(B \cap A_j)} \\
&= \tfrac{\Pr(B|A_i)\Pr(A_i)}{\Sigma_{j=1}^{n} \Pr(B|A_j)\Pr(A_j)}
\end{aligned}$$

# Example

In a certain county, 60% of registered voters are Republicans, 30% are Democrats, and 10% are Independents. 40% of Republicans oppose increased military spending, while 65% of the Democrats and 55% of the Independents oppose it. A registered voter writes a letter to the county paper, arguing against increased military spending. What is the probability that this voter is a Democrat?

$S = \{\text{registered voters}\}$
$A_1 = \{\text{registered Republicans}\}$
$A_2 = \{\text{registered Democrats}\}$
$A_3 = \{\text{registered independents}\}$
$B = \{\text{voters who oppose increased military spending}\}$

We want to know $\Pr(A_2|B)$.

We have

$$
\begin{array}{lll}
\Pr(A_1) = .6 & \Pr(A_2) = .3 & \Pr(A_3) = .1 \\
\Pr(B|A_1) = .4 & \Pr(B|A_2) = .65 & \Pr(B|A_3) = .55
\end{array}
$$

Using Bayes' Theorem, we have:

$$\begin{aligned}
\Pr(A_2|B) &= \frac{\Pr(B|A_2)\times\Pr(A_2)}{\Pr(B|A_1)\times\Pr(A_1)+\Pr(B|A_2)\times\Pr(A_2)+\Pr(B|A_3)\times\Pr(A_3)} \\
&= \frac{.65\times.3}{(.4\times.6)+(.65\times.3)+(.55\times.1)} \\
&= \frac{.195}{.49} \\
&\approx .398
\end{aligned}$$

# AIDS

Suppose we have a test that is 99% effective against AIDS. Suppose we also know that .3% of the population has AIDS. What is the probability that you have AIDS if you test positive?

$S = \{\text{all people}\}$ (in North America??)
$A_1 = \{\text{people with AIDS}\}$
$A_2 = \{\text{people who don't have AIDS}\}$ $(A_2 = \overline{A_1})$
$B = \{\text{people who test positive}\}$

$$\Pr(A_1) = .003 \qquad \Pr(A_2) = .997$$

Since the test is 99% effective:

$$\Pr(B|A_1) = .99 \qquad \Pr(B|A_2) = .01$$

Using Bayes' Theorem again:

$$\begin{aligned} \Pr(A_1|B) &= \tfrac{.99 \times .003}{(.99 \times .003) + (.01 \times .997)} \\ &\approx \tfrac{.003}{.003 + .01} \\ &\approx .23 \end{aligned}$$