Continuing Probability.

Wrap up: Probability Formalism.

Events, Conditional Probability, Independence, Bayes’ Rule
Probability Space: Formalism

Simplest physical model of a uniform probability space:

A bag of identical balls, except for their color (or a label). If the bag is well shaken, every ball is equally likely to be picked.

\[ \Omega = \{ \text{white, red, yellow, grey, purple, blue, maroon, green} \} \]

\[ Pr[\text{blue}] = \frac{1}{8}. \]
Probability Space: Formalism

Simplest physical model of a non-uniform probability space:

\[ \Omega = \{ \text{Red, Green, Yellow, Blue} \} \]

\[ Pr[\text{Red}] = \frac{3}{10}, \quad Pr[\text{Green}] = \frac{4}{10}, \quad \text{etc.} \]

Note: Probabilities are restricted to rational numbers: \( \frac{N_k}{N} \).
Probability Space: Formalism

Physical model of a general non-uniform probability space:

The roulette wheel stops in sector $\omega$ with probability $p_\omega$.

$$\Omega = \{1, 2, 3, \ldots, N\}, Pr[\omega] = p_\omega.$$
An important remark

- The random experiment selects **one and only one** outcome in $\Omega$.
- For instance, when we flip a fair coin **twice**
  - $\Omega = \{HH, TH, HT, TT\}$
  - The experiment selects **one** of the elements of $\Omega$.
- In this case, it's wrong to think that $\Omega = \{H, T\}$ and that the experiment selects two outcomes.
- Why? Because this would not describe how the two coin flips are related to each other.
- For instance, say we glue the coins side-by-side so that they face up the same way. Then one gets $HH$ or $TT$ with probability 50% each. This is not captured by ‘picking two outcomes.’
Lecture 15: Summary

Modeling Uncertainty: Probability Space

1. Random Experiment
2. Probability Space: $\Omega; Pr[\omega] \in [0, 1]; \sum_\omega Pr[\omega] = 1.$
3. Uniform Probability Space: $Pr[\omega] = 1/|\Omega|$ for all $\omega \in \Omega.$
CS70: On to Calculation.

Events, Conditional Probability, Independence, Bayes’ Rule

1. Probability Basics Review
2. Events
3. Conditional Probability
4. Independence of Events
5. Bayes’ Rule
Setup:

- Random Experiment.
  Flip a fair coin twice.
- Probability Space.
  - **Sample Space:** Set of outcomes, Ω.
    \[ Ω = \{HH, HT, TH, TT\} \]
    (Note: **Not** \[ Ω = \{H, T\} \] with two picks!)
  - **Probability:** \( Pr[ω] \) for all \( ω \in Ω \).
    \[ Pr[HH] = \cdots = Pr[TT] = 1/4 \]
    1. \( 0 \leq Pr[ω] \leq 1 \).
    2. \( \sum_{ω ∈ Ω} Pr[ω] = 1 \).
Set notation review

Figure: Two events

Figure: Union (or)

Figure: Difference (A, not B)

Figure: Complement (not)

Figure: Intersection (and)

Figure: Symmetric difference (only one)
Probability of exactly one ‘heads’ in two coin flips?

Idea: Sum the probabilities of all the different outcomes that have exactly one ‘heads’: $HT, TH$.

This leads to a definition!

**Definition:**

- An **event**, $E$, is a subset of outcomes: $E \subset \Omega$.
- The **probability of** $E$ is defined as $Pr[E] = \sum_{\omega \in E} Pr[\omega]$. 

![Sample Space Diagram](image)

**Uniform Probability Space**

$Pr[\omega] = \frac{1}{|\Omega|}$

$Pr[E] = \frac{|E|}{|\Omega|}$
Event: Example

Physical experiment

Probability model

\[ \Omega = \{\text{Red, Green, Yellow, Blue}\} \]

\[ Pr[\text{Red}] = \frac{3}{10}, \quad Pr[\text{Green}] = \frac{4}{10}, \quad \text{etc.} \]

\[ E = \{\text{Red, Green}\} \Rightarrow Pr[E] = \frac{3 + 4}{10} = \frac{3}{10} + \frac{4}{10} = Pr[\text{Red}] + Pr[\text{Green}] \]
Probability of exactly one heads in two coin flips?

Sample Space, $\Omega = \{HH, HT, TH, TT\}$.

Uniform probability space:
$Pr[HH] = Pr[HT] = Pr[TH] = Pr[TT] = \frac{1}{4}$.

Event, $E$, “exactly one heads”: $\{TH, HT\}$.

$$Pr[E] = \sum_{\omega \in E} Pr[\omega] = \frac{|E|}{|\Omega|} = \frac{2}{4} = \frac{1}{2}.$$
Example: 20 coin tosses.

20 coin tosses

Sample space: $\Omega = \text{set of 20 fair coin tosses.}$
$\Omega = \{T, H\}^{20} \equiv \{0, 1\}^{20}; \quad |\Omega| = 2^{20}$.

What is more likely?

$\omega_1 := (1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1), \text{ or}$
$\omega_2 := (1, 0, 1, 1, 0, 0, 0, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 0, 0)$?

Answer: Both are equally likely: $Pr[\omega_1] = Pr[\omega_2] = \frac{1}{|\Omega|}$.

What is more likely?

$(E_1)$ Twenty Hs out of twenty, or
$(E_2)$ Ten Hs out of twenty?

Answer: Ten Hs out of twenty.

Why? There are many sequences of 20 tosses with ten Hs; only one with twenty Hs. $\Rightarrow Pr[E_1] = \frac{1}{|\Omega|} \ll Pr[E_2] = \frac{|E_2|}{|\Omega|}$.

$|E_2| = \binom{20}{10} = 184,756$. 
Probability of $n$ heads in 100 coin tosses.

$\Omega = \{H, T\}^{100}$; $|\Omega| = 2^{100}$.

Event $E_n = \text{`n heads'}$; $|E_n| = \binom{100}{n}$

$p_n := Pr[E_n] = \frac{|E_n|}{|\Omega|} = \frac{\binom{100}{n}}{2^{100}}$

Observe:

- Concentration around mean: Law of Large Numbers;
- Bell-shape: Central Limit Theorem.
Roll a red and a blue die.

\[
Pr[\text{Sum to 7}] = \frac{6}{36} \quad Pr[\text{Sum to 10}] = \frac{3}{36}
\]
Exactly 50 heads in 100 coin tosses.

Sample space: $\Omega = \text{set of 100 coin tosses} = \{H, T\}^{100}$.  
$|\Omega| = 2 \times 2 \times \cdots \times 2 = 2^{100}$.

Uniform probability space: $Pr[\omega] = \frac{1}{2^{100}}$.

Event $E =$ “100 coin tosses with exactly 50 heads”  

$|E| = \binom{100}{50}$.

Choose 50 positions out of 100 to be heads.

$|E| = \binom{100}{50}$.

$$Pr[E] = \frac{\binom{100}{50}}{2^{100}}.$$
Calculation.
Stirling formula (for large $n$):

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

$$\binom{2n}{n} \approx \frac{\sqrt{4\pi n(2n/e)^{2n}}}{\left[\sqrt{2\pi n(n/e)^n}\right]^2} \approx \frac{4^n}{\sqrt{\pi n}}.$$

$$Pr[E] = \frac{|E|}{|\Omega|} = \frac{|E|}{2^{2n}} = \frac{1}{\sqrt{\pi n}} = \frac{1}{\sqrt{50\pi}} \approx .08.$$
Exactly 50 heads in 100 coin tosses.

\[ Pr[n \text{Hs out of } 2n] = \frac{\binom{2n}{n}}{2^{2n}} \]

0.08
Probability is Additive

Theorem

(a) If events $A$ and $B$ are disjoint, i.e., $A \cap B = \emptyset$, then

$$Pr[A \cup B] = Pr[A] + Pr[B].$$

(b) If events $A_1, \ldots, A_n$ are pairwise disjoint, i.e., $A_k \cap A_m = \emptyset$, $\forall k \neq m$, then

$$Pr[A_1 \cup \cdots \cup A_n] = Pr[A_1] + \cdots + Pr[A_n].$$

Proof:

Obvious.
Consequences of Additivity

Theorem

(a) \( \Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B]; \)  
    (inclusion-exclusion property)

(b) \( \Pr[A_1 \cup \cdots \cup A_n] \leq \Pr[A_1] + \cdots + \Pr[A_n]; \)  
    (union bound)

(c) If \( A_1, \ldots A_N \) are a partition of \( \Omega \), i.e., 
    pairwise disjoint and \( \bigcup_{m=1}^{N} A_m = \Omega \), then 
    \( \Pr[B] = \Pr[B \cap A_1] + \cdots + \Pr[B \cap A_N]. \)  
    (law of total probability)

Proof:

(b) is obvious.

Proofs for (a) and (c)? Next...
Inclusion/Exclusion

\[ Pr[A \cup B] = Pr[A] + Pr[B] - Pr[A \cap B] \]

Another view. Any \( \omega \in A \cup B \) is in \( A \cap \bar{B}, A \cup B, \text{ or } \bar{A} \cap B \). So, add it up.
Total probability

Assume that $\Omega$ is the union of the disjoint sets $A_1, \ldots, A_N$.

Then,

$$Pr[B] = Pr[A_1 \cap B] + \cdots + Pr[A_N \cap B].$$

Indeed, $B$ is the union of the disjoint sets $A_n \cap B$ for $n = 1, \ldots, N$.

In “math”: $\omega \in B$ is in exactly one of $A_i \cap B$.

Adding up probability of them, get $Pr[\omega]$ in sum.

..Did I say...

Add it up.
Roll a Red and a Blue Die.

$E_1 = \text{`Red die shows 6'}$; $E_2 = \text{`Blue die shows 6'}$

$E_1 \cup E_2 = \text{`At least one die shows 6'}$

$Pr[E_1] = \frac{6}{36}, \hspace{0.5cm} Pr[E_2] = \frac{6}{36}, \hspace{0.5cm} Pr[E_1 \cup E_2] = \frac{11}{36}.$
Conditional probability: example.

Two coin flips. First flip is heads. Probability of two heads? \( \Omega = \{ HH, HT, TH, TT \} \); Uniform probability space. Event \( A = \) first flip is heads: \( A = \{ HH, HT \} \).

New sample space: \( A \); uniform still.

Event \( B = \) two heads.

The probability of two heads if the first flip is heads. \textbf{The probability of } \( B \) \textbf{ given } \( A \) is \( 1/2 \).
A similar example.

Two coin flips. At least one of the flips is heads. 
→ Probability of two heads?

Ω = \{HH, HT, TH, TT\}; uniform.
Event \(A = \) at least one flip is heads. \(A = \{HH, HT, TH\}\).

New sample space: \(A\); uniform still.

Event \(B = \) two heads.

The probability of two heads if at least one flip is heads. The probability of \(B\) given \(A\) is \(1/3\).
Conditional Probability: A non-uniform example

\[ \Omega = \{\text{Red, Green, Yellow, Blue}\} \]

\[
Pr[\text{Red}|\text{Red or Green}] = \frac{3}{7} = \frac{Pr[\text{Red} \cap (\text{Red or Green})]}{Pr[\text{Red or Green}]} 
\]
Another non-uniform example

Consider $\Omega = \{1, 2, \ldots, N\}$ with $Pr[n] = p_n$. Let $A = \{3, 4\}, B = \{1, 2, 3\}$.

$$Pr[A|B] = \frac{p_3}{p_1 + p_2 + p_3} = \frac{Pr[A \cap B]}{Pr[B]}.$$
Yet another non-uniform example

Consider $\Omega = \{1, 2, \ldots, N\}$ with $Pr[n] = p_n$.
Let $A = \{2, 3, 4\}, B = \{1, 2, 3\}$.

$$Pr[A|B] = \frac{p_2 + p_3}{p_1 + p_2 + p_3} = \frac{Pr[A \cap B]}{Pr[B]}.$$
**Definition:** The conditional probability of $B$ given $A$ is

$$Pr[B|A] = \frac{Pr[A \cap B]}{Pr[A]}.$$
More fun with conditional probability.

Toss a red and a blue die, sum is 4, What is probability that red is 1?

\[
Pr[B|A] = \frac{|B \cap A|}{|A|} = \frac{1}{3}; \text{ versus } Pr[B] = 1/6.
\]

\(B\) is more likely given \(A\).
Yet more fun with conditional probability.

Toss a red and a blue die, sum is 7, what is probability that red is 1?

\[
Pr[B|A] = \frac{|B \cap A|}{|A|} = \frac{1}{6}; \text{ versus } Pr[B] = \frac{1}{6}.
\]

Observing \( A \) does not change your mind about the likelihood of \( B \).
Emptiness..

Suppose I toss 3 balls into 3 bins. $A =$“1st bin empty”; $B =$“2nd bin empty.” What is $Pr[A | B]$?

$\Omega = \{1, 2, 3\}^3$

$Pr[ B ] = Pr[\{(a, b, c) \mid a, b, c \in \{1, 3\}\}] = Pr[\{1, 3\}^3] = \frac{8}{27}$

$Pr[A \cap B] = Pr[(3, 3, 3)] = \frac{1}{27}$

$Pr[A | B] = \frac{Pr[A \cap B]}{Pr[B]} = \frac{(1/27)}{(8/27)} = 1/8; \text{ vs. } Pr[A] = \frac{8}{27}$.

$A$ is less likely given $B$: If second bin is empty the first is more likely to have balls in it.
Three Card Problem

Three cards: Red/Red, Red/Black, Black/Black.
Pick one at random and place on the table. The upturned side is a Red. What is the probability that the other side is Black?
Can’t be the BB card, so...prob should be 0.5, right?

$R$: upturned card is Red; $RB$: the Red/Black card was selected.

Want $P(RB|R)$.

What’s wrong with the reasoning that leads to $\frac{1}{2}$?

\[
P(RB|R) = \frac{P(RB \cap R)}{P(R)}
\]

\[
= \frac{\frac{1}{3} \cdot 1}{\frac{1}{3} \cdot 1 + \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot 0}
\]

\[
= \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}
\]

Once you are given $R$: it is twice as likely that the RR card was picked.
Gambler’s fallacy.

Flip a fair coin 51 times.

A = “first 50 flips are heads”

B = “the 51st is heads”

\[ Pr[B \mid A] \]?

\[ A = \{HH\cdots HT, HH\cdots HH\} \]

\[ B \cap A = \{HH\cdots HH\} \]

Uniform probability space.

\[ Pr[B \mid A] = \frac{|B \cap A|}{|A|} = \frac{1}{2}. \]

Same as \( Pr[B] \).

The likelihood of 51st heads does not depend on the previous flips.
Recall the definition:

\[ Pr[B|A] = \frac{Pr[A \cap B]}{Pr[A]} . \]

Hence,

\[ Pr[A \cap B] = Pr[A] Pr[B|A] . \]

Consequently,

\[ Pr[A \cap B \cap C] = Pr[(A \cap B) \cap C] \]
\[ = Pr[A \cap B] Pr[C|A \cap B] \]
\[ = Pr[A] Pr[B|A] Pr[C|A \cap B] . \]
Theorem Product Rule
Let $A_1, A_2, \ldots, A_n$ be events. Then

$$Pr[A_1 \cap \cdots \cap A_n] = Pr[A_1] Pr[A_2 | A_1] \cdots Pr[A_n | A_1 \cap \cdots \cap A_{n-1}].$$

Proof: By induction.
Assume the result is true for $n$. (It holds for $n = 2$.) Then,

$$Pr[A_1 \cap \cdots \cap A_n \cap A_{n+1}] = Pr[A_1 \cap \cdots \cap A_n] Pr[A_{n+1} | A_1 \cap \cdots \cap A_n]$$

$$= Pr[A_1] Pr[A_2 | A_1] \cdots Pr[A_n | A_1 \cap \cdots \cap A_{n-1}] Pr[A_{n+1} | A_1 \cap \cdots \cap A_n],$$

so that the result holds for $n + 1$. \qed
Correlation

An example.
Random experiment: Pick a person at random.
Event $A$: the person has lung cancer.
Event $B$: the person is a heavy smoker.

Fact:

$$Pr[A|B] = 1.17 \times Pr[A].$$

Conclusion:

- Smoking increases the probability of lung cancer by 17%.
- Smoking causes lung cancer.
Correlation

Event $A$: the person has lung cancer. Event $B$: the person is a heavy smoker. $\Pr[A|B] = 1.17 \times \Pr[A]$.

A second look.

Note that

$$\Pr[A|B] = 1.17 \times \Pr[A] \iff \frac{\Pr[A \cap B]}{\Pr[B]} = 1.17 \times \Pr[A]$$
$$\iff \Pr[A \cap B] = 1.17 \times \Pr[A]\Pr[B]$$
$$\iff \Pr[B|A] = 1.17 \times \Pr[B].$$

Conclusion:

- Lung cancer increases the probability of smoking by 17%.
- Lung cancer causes smoking. Really?
Causality vs. Correlation

Events $A$ and $B$ are **positively correlated** if

$$Pr[A \cap B] > Pr[A]Pr[B].$$

(E.g., smoking and lung cancer.)

$A$ and $B$ being positively correlated does not mean that $A$ causes $B$ or that $B$ causes $A$.

Other examples:

- Tesla owners are more likely to be rich. That does not mean that poor people should buy a Tesla to get rich.
- People who go to the opera are more likely to have a good career. That does not mean that going to the opera will improve your career.
- Rabbits eat more carrots and do not wear glasses. Are carrots good for eyesight?
Total probability

Assume that $\Omega$ is the union of the disjoint sets $A_1, \ldots, A_N$.

Then,

$$Pr[B] = Pr[A_1 \cap B] + \cdots + Pr[A_N \cap B].$$

Indeed, $B$ is the union of the disjoint sets $A_n \cap B$ for $n = 1, \ldots, N$. Thus,

Total probability

Assume that $\Omega$ is the union of the disjoint sets $A_1, \ldots, A_N$.

Is your coin loaded?
Your coin is fair w.p. 1/2 or such that $\Pr[H] = 0.6$, otherwise.
You flip your coin and it yields heads.
What is the probability that it is fair?

**Analysis:**

$A = \text{‘coin is fair’}, B = \text{‘outcome is heads’}$

We want to calculate $Pr[A|B]$.
We know $Pr[B|A] = 1/2, Pr[B|\bar{A}] = 0.6, Pr[A] = 1/2 = Pr[\bar{A}]$

Now,

$$Pr[B] = Pr[A \cap B] + Pr[\bar{A} \cap B] = Pr[A]Pr[B|A] + Pr[\bar{A}]Pr[B|\bar{A}]$$
$$= (1/2)(1/2) + (1/2)0.6 = 0.55.$$

Thus,

$$Pr[A|B] = \frac{Pr[A]Pr[B|A]}{Pr[B]} = \frac{(1/2)(1/2)}{(1/2)(1/2) + (1/2)0.6} \approx 0.45.$$
Is your coin loaded?

A picture:

Imagine 100 situations, among which
\[ m := 100 \left( \frac{1}{2} \right) \left( \frac{1}{2} \right) \] are such that \( A \) and \( B \) occur and
\[ n := 100 \left( \frac{1}{2} \right) \left( 0.6 \right) \] are such that \( \bar{A} \) and \( B \) occur.

Thus, among the \( m + n \) situations where \( B \) occurred, there are \( m \) where \( A \) occurred.

Hence,

\[
Pr[ A \mid B ] = \frac{m}{m + n} = \frac{ \left( \frac{1}{2} \right) \left( \frac{1}{2} \right) }{ \left( \frac{1}{2} \right) \left( \frac{1}{2} \right) + \left( \frac{1}{2} \right) 0.6 .}
\]
Independence

**Definition:** Two events $A$ and $B$ are independent if

$$Pr[A \cap B] = Pr[A]Pr[B].$$

Examples:

- When rolling two dice, $A =$ sum is 7 and $B =$ red die is 1 are independent;
- When rolling two dice, $A =$ sum is 3 and $B =$ red die is 1 are not independent;
- When flipping coins, $A =$ coin 1 yields heads and $B =$ coin 2 yields tails are independent;
- When throwing 3 balls into 3 bins, $A =$ bin 1 is empty and $B =$ bin 2 is empty are not independent;
Independence and conditional probability

**Fact:** Two events $A$ and $B$ are **independent** if and only if

$$Pr[A|B] = Pr[A].$$

Indeed: $Pr[A|B] = \frac{Pr[A \cap B]}{Pr[B]}$, so that

$$Pr[A|B] = Pr[A] \iff \frac{Pr[A \cap B]}{Pr[B]} = Pr[A] \iff Pr[A \cap B] = Pr[A] Pr[B].$$
Bayes Rule

Another picture: We imagine that there are $N$ possible causes $A_1, \ldots, A_N$.

Imagine 100 situations, among which $100p_nq_n$ are such that $A_n$ and $B$ occur, for $n = 1, \ldots, N$.

Thus, among the $100 \sum_m p_m q_m$ situations where $B$ occurred, there are $100p_nq_n$ where $A_n$ occurred.

Hence,

$$Pr[A_n|B] = \frac{p_nq_n}{\sum_m p_m q_m}.$$
Why do you have a fever?

Using Bayes’ rule, we find

\[ Pr[\text{Flu}|\text{High Fever}] = \frac{0.15 \times 0.80}{0.15 \times 0.80 + 10^{-8} \times 1 + 0.85 \times 0.1} \approx 0.58 \]

\[ Pr[\text{Ebola}|\text{High Fever}] = \frac{10^{-8} \times 1}{0.15 \times 0.80 + 10^{-8} \times 1 + 0.85 \times 0.1} \approx 5 \times 10^{-8} \]

\[ Pr[\text{Other}|\text{High Fever}] = \frac{0.85 \times 0.1}{0.15 \times 0.80 + 10^{-8} \times 1 + 0.85 \times 0.1} \approx 0.42 \]

These are the posterior probabilities. One says that ‘Flu’ is the Most Likely a Posteriori (MAP) cause of the high fever.
Bayes’ Rule Operations

Bayes’ Rule is the canonical example of how information changes our opinions.
Thomas Bayes

Portrait used of Bayes in a 1936 book,[1] but it is doubtful whether the portrait is actually of him.[2] No earlier portrait or claimed portrait survives.

Born  
c. 1701  
London, England

Died  
7 April 1761 (aged 59)  
Tunbridge Wells, Kent, England

Residence  
Tunbridge Wells, Kent, England

Nationality  
English

Known for  
Bayes' theorem

Thomas Bayes

A Bayesian picture of Thomas Bayes.

Fig. 3. Joshua Bayes (1671–1746).
Testing for disease.

Let's watch TV!!
Random Experiment: Pick a random male.
Outcomes: \((test, disease)\)

\(A\) - prostate cancer.
\(B\) - positive PSA test.

- \(Pr[A] = 0.0016, (0.16 \% \text{ of the male population is affected.})\)
- \(Pr[B|A] = 0.80 \text{ (80\% chance of positive test with disease.)}\)
- \(Pr[B|\overline{A}] = 0.10 \text{ (10\% chance of positive test without disease.)}\)


Positive PSA test \((B)\). Do I have disease?

\(Pr[A|B]???\)
Bayes Rule.

Using Bayes’ rule, we find

\[
P[A|B] = \frac{0.0016 \times 0.80}{0.0016 \times 0.80 + 0.9984 \times 0.10} = 0.013.
\]

A 1.3% chance of prostate cancer with a positive PSA test.

Surgery anyone?

Impotence...

Incontinence..

Death.
Summary

Events, Conditional Probability, Independence, Bayes’ Rule

Key Ideas:

- **Conditional Probability:**
  \[ \Pr[A|B] = \frac{\Pr[A \cap B]}{\Pr[B]} \]

- **Independence:** \( \Pr[A \cap B] = \Pr[A] \Pr[B] \).

- **Bayes’ Rule:**
  \[ \Pr[A_n|B] = \frac{\Pr[A_n] \Pr[B|A_n]}{\sum_m \Pr[A_m] \Pr[B|A_m]} . \]

  \( \Pr[A_n|B] = \) posterior probability; \( \Pr[A_n] = \) prior probability.

- **All these are possible:**
  \( \Pr[A|B] < \Pr[A] \); \( \Pr[A|B] > \Pr[A] \); \( \Pr[A|B] = \Pr[A] \).