Today

Probability:
  Keep building it formally..
  And our intuition.
Poll: blows my mind.

Flip 300 million coins.
Which is more likely?
(A) 300 million heads.
(B) 300 million tails.
(C) Alternating heads and tails.
(D) A tail every third spot.

Given the history of the universe up to right now.

What is the likelihood of our universe?
(A) The likelihood is 1. Cuz here it is.
(B) As likely as any other. Cuz of probability.
(C) Well. Quantum. IDK- TBH.

Perhaps a philosophical (“wastebasket”) question.
Also, “cuz” == “because”
Probability Basics.

Probability Space.

1. **Sample Space**: Set of outcomes, $\Omega$.

2. **Probability**: $Pr[\omega]$ for all $\omega \in \Omega$.
   
   2.1 $0 \leq Pr[\omega] \leq 1$.
   
   2.2 $\sum_{\omega \in \Omega} Pr[\omega] = 1$.

Example: Two coins.

1. $\Omega = \{HH, HT, TH, TT\}$
   
   (Note: Not $\Omega = \{H, T\}$ with two picks!)

2. $Pr[HH] = \cdots = Pr[TT] = 1/4$
Consequences of Additivity

Theorem

(a) Inclusion/Exclusion: \( Pr[A \cup B] = Pr[A] + Pr[B] - Pr[A \cap B] \);
(b) Union Bound: \( Pr[A_1 \cup \cdots \cup A_n] \leq Pr[A_1] + \cdots + Pr[A_n] \);
(c) Law of Total Probability:

If \( A_1, \ldots, A_N \) are a partition of \( \Omega \), i.e., pairwise disjoint and \( \bigcup_{m=1}^{N} A_m = \Omega \), then

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

Proof Idea: Total probability.

Add it up!
What does Rao mean by “Add it up.”

(A) Organize intuitions/proofs around $Pr[\omega]$.
(B) Organize intuition/proofs around $Pr[\mathcal{A}]$.
(C) Some weird song whose refrain he heard in his youth.

(A), (B), and (C)
**Definition:** The conditional probability of $B$ given $A$ is

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

Note also:

\[ Pr[A \cap B] = Pr[B|A]Pr[A] \]
Product Rule

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


**Theorem** Product Rule
Let $A_1, A_2, \ldots, A_n$ be events. Then

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

\[ Pr[A \mid B] = \frac{Pr[A \cap B]}{Pr[B]}, \quad Pr[B \mid A] = \frac{Pr[A \cap B]}{Pr[A]} . \]

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

Bayes Rule: \( Pr[A \mid B] = \frac{Pr[B \mid A] Pr[A]}{Pr[B]} \).
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 \( P[A|B] \).

We know \( P[B|A] = 1/2, \; P[B|\bar{A}] = 0.6, \; Pr[A] = 1/2 = Pr[\bar{A}] \)

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,

\[
Pr[A|B] = \frac{Pr[A]Pr[B|A]}{Pr[B]} = \frac{(1/2)(1/2)}{(1/2)(1/2) + (1/2)0.6} \approx 0.45.
\]
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; $Pr[A \cap B] = \frac{1}{36}$, $Pr[A] Pr[B] = (\frac{1}{6})(\frac{1}{6})$.

- When rolling two dice, $A =$ sum is 3 and $B =$ red die is 1 are not independent; $Pr[A \cap B] = \frac{1}{36}$, $Pr[A] Pr[B] = (\frac{2}{36})(\frac{1}{6})$.

- When flipping coins, $A =$ coin 1 yields heads and $B =$ coin 2 yields tails are independent; $Pr[A \cap B] = \frac{1}{4}$, $Pr[A] Pr[B] = (\frac{1}{2})(\frac{1}{2})$.

- When throwing 3 balls into 3 bins, $A =$ bin 1 is empty and $B =$ bin 2 is empty are not independent; $Pr[A \cap B] = \frac{1}{27}$, $Pr[A] Pr[B] = (\frac{8}{27})(\frac{8}{27})$. 
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

$$Pr[A|B] = Pr[A] \iff \frac{Pr[A \cap B]}{Pr[B]} = Pr[A] \iff Pr[A \cap B] = Pr[A]Pr[B].$$
Conditional Probability: Review

Recall:

- \( \Pr[A \mid B] = \frac{\Pr[A \cap B]}{\Pr[B]} \).
- Hence, \( \Pr[A \cap B] = \Pr[B] \Pr[A \mid B] = \Pr[A] \Pr[B \mid A] \).
- \( A \) and \( B \) are positively correlated if \( \Pr[A \mid B] > \Pr[A] \),
  i.e., if \( \Pr[A \cap B] > \Pr[A] \Pr[B] \).
- \( A \) and \( B \) are negatively correlated if \( \Pr[A \mid B] < \Pr[A] \),
  i.e., if \( \Pr[A \cap B] < \Pr[A] \Pr[B] \).
- \( A \) and \( B \) are independent if \( \Pr[A \mid B] = \Pr[A] \),
  i.e., if \( \Pr[A \cap B] = \Pr[A] \Pr[B] \).
- Note: \( B \subset A \Rightarrow A \) and \( B \) are positively correlated.
  \( (\Pr[A \mid B] = 1 > \Pr[A]) \)
- Note: \( A \cap B = \emptyset \Rightarrow A \) and \( B \) are negatively correlated.
  \( (\Pr[A \mid B] = 0 < \Pr[A]) \)
Illustrations: Pick a point uniformly in the unit square

Which $A$ and $B$ are independent?

(A) Left.
(B) Middle.
(B) Right.

See next slide.
Illustrations: Pick a point uniformly in the unit square

- **Left:** $A$ and $B$ are independent. $Pr[B] = b; Pr[B|A] = b$.

- **Middle:** $A$ and $B$ are positively correlated. $Pr[B|A] = b_1 > Pr[B|\bar{A}] = b_2$. Note: $Pr[B] \in (b_2, b_1)$.

- **Right:** $A$ and $B$ are negatively correlated. $Pr[B|A] = b_1 < Pr[B|\bar{A}] = b_2$. Note: $Pr[B] \in (b_1, b_2)$. 
Bayes and Biased Coin

Pick a point uniformly at random in the unit square. Then

\[ Pr[A] = 0.5; \quad Pr[\bar{A}] = 0.5 \]
\[ Pr[B|A] = 0.5; \quad Pr[B|\bar{A}] = 0.6; \quad Pr[A \cap B] = 0.5 \times 0.5 \]
\[ Pr[B] = 0.5 \times 0.5 + 0.5 \times 0.6 = Pr[A]Pr[B|A] + Pr[\bar{A}]Pr[B|\bar{A}] \]
\[ Pr[A|B] = \frac{0.5 \times 0.5}{0.5 \times 0.5 + 0.5 \times 0.6} = \frac{Pr[A]Pr[B|A]}{Pr[A]Pr[B|A] + Pr[\bar{A}]Pr[B|\bar{A}]} \]
\[ \approx 0.46 = \text{fraction of } B \text{ that is inside } A \]
Bayes: General Case

Pick a point uniformly at random in the unit square. Then

\[ Pr[A_{n}] = p_{n}, \quad n = 1, \ldots, N \]
\[ Pr[B | A_{n}] = q_{n}, \quad n = 1, \ldots, N; \quad Pr[A_{n} \cap B] = p_{n} q_{n} \]
\[ Pr[B] = p_{1} q_{1} + \cdots + p_{N} q_{N} \]

\[ Pr[A_{n} | B] = \frac{p_{n} q_{n}}{p_{1} q_{1} + \cdots + p_{N} q_{N}} = \text{fraction of } B \text{ inside } A_{n}. \]
Bayes Rule

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

100 situations: $100p_nq_n$ where $A_n$ and $B$ occur, for $n = 1, \ldots, N$. In $100\sum_m p_mq_m$ occurrences of $B$, $100p_nq_n$ occurrences of $A_n$.

Hence,

$$Pr[A_n|B] = \frac{p_nq_n}{\sum_m p_mq_m}.$$ 

But, $p_n = Pr[A_n], q_n = Pr[B|A_n], \sum_m p_mq - m = Pr[B]$, hence,

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

Using Bayes’ rule, we find

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

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

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

The values 0.58, $5 \times 10^{-8}$, 0.42 are the posterior probabilities.
Why do you have a fever?

Our “Bayes’ Square” picture:

Note that even though $Pr[\text{Fever}|\text{Ebola}] = 1$, one has

$$Pr[\text{Ebola}|\text{Fever}] \approx 0.$$  

This example shows the importance of the prior probabilities.
Why do you have a fever?

We found

\[
Pr[\text{Flu}|\text{High Fever}] \approx 0.58, \\
Pr[\text{Ebola}|\text{High Fever}] \approx 5 \times 10^{-8}, \\
Pr[\text{Other}|\text{High Fever}] \approx 0.42
\]

‘Flu’ is Most Likely a Posteriori (MAP) cause of high fever. ‘Ebola’ is Maximum Likelihood Estimate (MLE) of cause: causes fever with largest probability.

Recall that

\[
p_m = Pr[A_m], q_m = Pr[B|A_m], Pr[A_m|B] = \frac{p_m q_m}{p_1 q_1 + \cdots + p_M q_M}.
\]

Thus,

\[\begin{align*}
\text{MAP} &= \text{value of } m \text{ that maximizes } p_m q_m. \\
\text{MLE} &= \text{value of } m \text{ that maximizes } q_m.
\end{align*}\]
Bayes’ Rule Operations

Bayes’ Rule: 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

A Bayesian picture of Thomas Bayes.
Testing for disease.

Random Experiment: Pick a random male.
Outcomes: \((test, disease)\)
\(A\) - prostate cancer.
\(B\) - positive PSA test.

- \(Pr[A] = 0.0016\), (.16 % of the male population is affected.)
- \(Pr[B|A] = 0.80\) (80% chance of positive test with disease.)
- \(Pr[B|\overline{A}] = 0.10\) (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.
Quick Review

Events, Conditional Probability, Independence, Bayes’ Rule

Key Ideas:

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

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

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

  \[ Pr[A_n \mid B] = \text{posterior probability}; \ Pr[A_n] = \text{prior probability} \]

- All these are possible:
  \[ Pr[A \mid B] < Pr[A]; \ Pr[A \mid B] > Pr[A]; \ Pr[A \mid B] = Pr[A] \]
Independence

Recall:

A and B are independent
⇔ \( Pr[A \cap B] = Pr[A]Pr[B] \)
⇔ \( Pr[A|B] = Pr[A] \).

Consider the example below:

\[
\begin{array}{c|c|c}
A_1 & 0.1 & 0.15 \\
A_2 & 0.25 & 0.25 \\
A_3 & 0.15 & 0.1 \\
\hline
B & & \end{array}
\]

Which are independent? (A) \((A_2, B)\)  (B) \((A_2, \bar{B})\)  (C) \((A_1, B)\).

\((A_2, B)\) are independent: \( Pr[A_2|B] = 0.5 = Pr[A_2] \).

\((A_2, \bar{B})\) are independent: \( Pr[A_2|\bar{B}] = 0.5 = Pr[A_2] \).

\((A_1, B)\) are not independent: \( Pr[A_1|B] = \frac{0.1}{0.5} = 0.2 \neq Pr[A_1] = 0.25. \)
Pairwise Independence

Flip two fair coins. Let

- $A = \text{`first coin is H'} = \{HT, HH\}$;
- $B = \text{`second coin is H'} = \{TH, HH\}$;
- $C = \text{`the two coins are different'} = \{TH, HT\}$.

$A, C$ are independent; $B, C$ are independent;
$A \cap B, C$ are not independent. ($Pr[A \cap B \cap C] = 0 \neq Pr[A \cap B]Pr[C]$.)

False: If $A$ did not say anything about $C$ and $B$ did not say anything about $C$, then $A \cap B$ would not say anything about $C$. 
Example 2

Flip a fair coin 5 times. Let $A_n = \text{‘coin n is H’}$, for $n = 1, \ldots, 5$.

Then, $A_m, A_n$ are independent for all $m \neq n$.

Also, $A_1$ and $A_3 \cap A_5$ are independent.

Indeed,

$Pr[A_1 \cap (A_3 \cap A_5)] = \frac{1}{8} = Pr[A_1]Pr[A_3 \cap A_5]$.

Similarly, $A_1 \cap A_2$ and $A_3 \cap A_4 \cap A_5$ are independent.

This leads to a definition ....
Definition Mutual Independence

(a) The events $A_1, \ldots, A_5$ are **mutually independent** if

$$Pr[\cap_{k \in K} A_k] = \prod_{k \in K} Pr[A_k], \text{ for all } K \subseteq \{1, \ldots, 5\}. $$

(b) More generally, the events $\{A_j, j \in J\}$ are **mutually independent** if

$$Pr[\cap_{k \in K} A_k] = \prod_{k \in K} Pr[A_k], \text{ for all finite } K \subseteq J. $$

Example: Flip a fair coin forever. Let $A_n = \text{‘coin } n \text{ is H.’}$ Then the events $A_n$ are mutually independent.
Theorem

(a) If the events \( \{A_j, j \in J\} \) are mutually independent and if \( K_1 \) and \( K_2 \) are disjoint finite subsets of \( J \), then

\[
\cap_{k \in K_1} A_k \text{ and } \cap_{k \in K_2} A_k \text{ are independent.}
\]

(b) More generally, if the \( K_n \) are pairwise disjoint finite subsets of \( J \), then the events

\[
\cap_{k \in K_n} A_k \text{ are mutually independent.}
\]

(c) Also, the same is true if we replace some of the \( A_k \) by \( \bar{A}_k \).
Conditional Probability: Review

Recall:

- \(Pr[A|B] = \frac{Pr[A \cap B]}{Pr[B]}\).
- Hence, \(Pr[A \cap B] = Pr[B]Pr[A|B] = Pr[A]Pr[B|A]\).
- \(A\) and \(B\) are positively correlated if \(Pr[A|B] > Pr[A]\),
  i.e., if \(Pr[A \cap B] > Pr[A]Pr[B]\).
- \(A\) and \(B\) are negatively correlated if \(Pr[A|B] < Pr[A]\),
  i.e., if \(Pr[A \cap B] < Pr[A]Pr[B]\).
- \(A\) and \(B\) are independent if \(Pr[A|B] = Pr[A]\),
  i.e., if \(Pr[A \cap B] = Pr[A]Pr[B]\).
- Note: \(B \subset A \Rightarrow A\) and \(B\) are positively correlated.
  \((Pr[A|B] = 1 > Pr[A])\)
- Note: \(A \cap B = \emptyset \Rightarrow A\) and \(B\) are negatively correlated.
  \((Pr[A|B] = 0 < Pr[A])\)
Quick Review.

Main results:

- **Bayes’ Rule:** $Pr[A_m|B] = \frac{p_m q_m}{p_1 q_1 + \cdots + p_M q_M}$.
- **Product Rule:**
  \[
  Pr[A_1 \cap \cdots \cap A_n] = Pr[A_1] Pr[A_2|A_1] \cdots Pr[A_n|A_1 \cap \cdots \cap A_{n-1}].
  \]