Independence
Today, start with a cool program
100,000 samples

6 observations per sample
Discovered Pattern

- $p(G_1) = 0.5$
- $p(G_5) = 0.6$
- $p(T \mid G_1 \text{ and } G_2) = 0.9$
- $p(T \mid \sim G_1 \text{ or } \sim G_2) = 0.2$

- $p(G_2 \mid G_5) = 0.9$
- $p(G_2 \mid \sim G_5) = 0.2$

These genes don’t impact $T$
We’ve gotten ahead of ourselves
Start at the beginning
And vs Condition

\[ P(AB) \text{ vs } P(A|B) \]

\[ P(AB) = P(A|B)P(B) \]
Today

- OR: $P(E \cup F)$
- AND: $P(EF)$

- DeMorgan’s
- Mutually Exclusive?
- Independent?
- Just Add!
- Inclusion Exclusion
- Just Multiply
- Chain Rule
Today

- OR
  - $P(E \cup F)$
- Inclusion Exclusion
- Mutually Exclusive?
- Just Add!

- AND
  - $P(EF)$
- Independent?
- Just Multiply
- Chain Rule

DeMorgan’s
Probability of “OR”
If events are mutually exclusive, probability of OR is simple:

\[ P(E \cup F) = P(E) + P(F) \]
If events are mutually exclusive, probability of OR is simple:

\[ P(E \cup F) = \frac{7}{50} + \frac{4}{5} = \frac{11}{50} \]
OR with Many Mutually Exclusive Events

\[ P(X_1 \cup X_2 \cup \cdots \cup X_n) = \sum_{i=1}^{n} P(X_i) \]
If events are *mutually exclusive* probability of OR is easy!
What about when they are not Mutually exclusive?
OR without Mutually Exclusivity

\[ P(E \cup F) = P(E) + P(F) - P(EF) \]

AKA Inclusion Exclusion
OR without Mutually Exclusivity

AKA Inclusion Exclusion

\[ P(E \cup F) = \frac{8}{50} + \frac{14}{50} - \frac{3}{50} = \frac{19}{50} \]
More than two sets?
Inclusion Exclusion with Three Sets

\[ P(E \cup F \cup G) = \]
Inclusion Exclusion with Three Sets

\[ P(E \cup F \cup G) = P(E) \]
Inclusion Exclusion with Three Sets

\[ P(E \cup F \cup G) = P(E) + P(F) \]
Inclusion Exclusion with Three Sets

\[ P(E \cup F \cup G) = P(E) + P(F) + P(G) \]
Inclusion Exclusion with Three Sets

\[ P(E \cup F \cup G) = P(E) + P(F) + P(G) - P(EF) \]
Inclusion Exclusion with Three Sets

\[ P(E \cup F \cup G) = P(E) + P(F) + P(G) - P(EF) - P(EG) \]
Inclusion Exclusion with Three Sets

\[ P(E \cup F \cup G) = P(E) + P(F) + P(G) \]

\[ -P(EF) - P(EG) - P(FG) \]
Inclusion Exclusion with Three Sets

\[ P(E \cup F \cup G) = P(E) + P(F) + P(G) \]
\[ - P(EF) - P(EG) - P(FG) \]
\[ + P(EFG) \]
General Inclusion Exclusion

\[ P(E_1 \cup E_2 \cup \cdots \cup E_n) = \sum_{r=1}^{n} (-1)^{r+1} Y_r \]

\[ Y_1 = \text{Sum of all events on their own} \quad \sum_{i} P(E_i) \]

\[ Y_2 = \text{Sum of all pairs of events} \quad \sum_{i,j \text{ s.t. } i \neq j} P(E_i \cap E_j) \]

\[ Y_3 = \text{Sum of all triples of events} \quad \sum_{i,j,k \text{ s.t. } i \neq j, j \neq k, i \neq k} P(E_i \cap E_j \cap E_k) \]

* Where \( Y_r \) is the sum, for all combinations of \( r \) events, of the probability of the union those events.
Today

**OR**

\[ P(E \cup F) \]

**Just Add!**

**Inclusion Exclusion**

Mutually Exclusive?

DeMorgan’s

**AND**

\[ P(EF) \]

Independent?

Just Multiply

Chain Rule
Today

\[ P(E \cup F) = P(E) + P(F) - P(EF) \]

\[ P(EF) = P(E|F)P(F) \]

DeMorgan’s

Mutually Exclusive?

Just Add!

Inclusion Exclusion

Independent?

Just Multiply

Chain Rule
Probability of “AND”
We the People of the United States, in Order to form a more perfect Union, establish Justice, insure domestic Tranquility, provide for the common Defense, promote the general Welfare, and secure the Blessings of Liberty to ourselves and our Posterity, do ordain and establish this Constitution for the United States of America.
Two events A and B are called independent if:

\[ P(AB) = P(A)P(B) \]

Otherwise, they are called dependent events.
If events are *independent* probability of AND is easy!

*You will need to use this “trick” with high probability*
Intuition through proofs

Let $A$ and $B$ be independent

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

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

\[ = P(A) \]

Definition of conditional probability

Since $A$ and $B$ are independent

Taking the bus to cancel city

Knowing that event $B$ happened, doesn’t change our belief that $A$ will happen.
• Roll two 6-sided dice, yielding values $D_1$ and $D_2$
  ▪ Let $E$ be event: $D_1 = 1$
  ▪ Let $F$ be event: $D_2 = 1$

• What is $P(E)$, $P(F)$, and $P(EF)$?
  ▪ $P(E) = 1/6$, $P(F) = 1/6$, $P(EF) = 1/36$
  ▪ $P(EF) = P(E) \times P(F) \Rightarrow E$ and $F$ independent

• Let $G$ be event: $D_1 + D_2 = 5$ \{$(1, 4), (2, 3), (3, 2), (4, 1)$\}
• What is $P(E)$, $P(G)$, and $P(EG)$?
  ▪ $P(E) = 1/6$, $P(G) = 4/36 = 1/9$, $P(EG) = 1/36$
  ▪ $P(EG) \neq P(E) \times P(G) \Rightarrow E$ and $G$ dependent
What does independence look like?
Independence Definition 1:

\[
P(AB) = P(A)P(B) \]

\[
\frac{|AB|}{|S|} = \frac{|A|}{|S|} \times \frac{|B|}{|S|}
\]
Independence

Independence Definition 1:

\[ P(AB) = P(A)P(B) \]

\[ \frac{|AB|}{|S|} = \frac{|A|}{|S|} \times \frac{|B|}{|S|} \]

Independence Definition 2:

\[ P(A|B) = P(A) \]

\[ \frac{|AB|}{|B|} = \frac{|A|}{|S|} \]
Independence

This ratio, $P(A)$…

… is the same as this one, $P(A|B)$
Independence Definition 1:

\[ P(AB) = P(A)P(B) \]

\[ \frac{|AB|}{|S|} = \frac{|A|}{|S|} \times \frac{|B|}{|S|} \]

Independence Definition 2:

\[ P(A|B) = P(A) \]

\[ \frac{|AB|}{|B|} = \frac{|A|}{|S|} \]
Dependence

Independence Definition 1:
\[ P(AB) = P(A)P(B) \]
\[ \frac{|AB|}{|S|} = \frac{|A|}{|S|} \times \frac{|B|}{|S|} \]

Independence Definition 2:
\[ P(A|B) = P(A) \]
\[ \frac{|AB|}{|B|} = \frac{|A|}{|S|} \]
More Intuition through proofs:
Given independent events $A$ and $B$, prove that $A$ and $B^C$ are independent

We want to show that $P(AB^C) = P(A)P(B^C)$

$$P(AB^C) = P(A) - P(AB)$$

By Total Law of Prob.

$$= P(A) - P(A)P(B)$$

By independence

$$= P(A)[1 - P(B)]$$

Factoring

$$= P(A)P(B^C)$$

Since $P(B) + P(B^C) = 1$

So if $A$ and $B$ are independent $A$ and $B^C$ are also independent
Generalized Independence

- General definition of Independence:
  Events $E_1$, $E_2$, ..., $E_n$ are independent if for every subset with $r$ elements (where $r \leq n$) it holds that:

$$P(E_1'E_2'E_3'...E_r') = P(E_1')P(E_2')P(E_3')...P(E_r')$$

- Example: outcomes of $n$ separate flips of a coin are all independent of one another
  - Each flip in this case is called a “trial” of the experiment
Math > Intuition
• Roll two 6-sided dice, yielding values $D_1$ and $D_2$
  ▪ Let $E$ be event: $D_1 = 1$
  ▪ Let $F$ be event: $D_2 = 6$
  ▪ Are $E$ and $F$ independent?    Yes!

• Let $G$ be event: $D_1 + D_2 = 7$
  ▪ Are $E$ and $G$ independent?    Yes!
  ▪ $P(E) = 1/6$,    $P(G) = 1/6$,    $P(E \cap G) = 1/36$    [roll (1, 6)]
  ▪ Are $F$ and $G$ independent?    Yes!
  ▪ $P(F) = 1/6$,    $P(G) = 1/6$,    $P(F \cap G) = 1/36$    [roll (1, 6)]
  ▪ Are $E$, $F$ and $G$ independent?    No!
  ▪ $P(EFG) = 1/36 \neq 1/216 = (1/6)(1/6)(1/6)$
New Ability
Today

OR

\[ P(E \cup F) \]

AND

\[ P(EF) \]

Mutually Exclusive?

Inclusion Exclusion

Independent?

DeMorgan’s

Just Add!

Just Multiply

Chain Rule

Independent?
Today

OR
P(E ∪ F)

AND
P(EF)

Mutually Exclusive?

Independent?

Just Add!

Inclusion Exclusion

Just Multiply!

Chain Rule

DeMorgan’s
Use the two properties (mutual exclusion and independence)
• Consider the following parallel network:

  \[ n \text{ independent routers, each with probability } p_i \text{ of functioning (where } 1 \leq i \leq n) \]

  \[ E = \text{functional path from A to B exists. What is } P(E)? \]
Consider the following parallel network:

- $n$ independent routers, each with probability $p_i$ of functioning (where $1 \leq i \leq n$)
- $E = \text{functional path from A to B exists}$. What is $P(E)$?
Consider the following parallel network:

- $n$ independent routers, each with probability $p_i$ of functioning (where $1 \leq i \leq n$)
- $E = $ functional path from A to B exists. What is $P(E)$?

Solution:

- $P(E) = 1 - P(\text{all routers fail})$
  
  \[ = 1 - (1 - p_1)(1 - p_2)\ldots(1 - p_n) \]
  
  \[ = 1 - \prod_{i=1}^{n} (1 - p_i) \]
• Say a coin comes up heads with probability $p$
  ▪ Each coin flip is an independent trial

• $P(n \text{ heads on } n \text{ coin flips}) = p^n$
• $P(n \text{ tails on } n \text{ coin flips}) = (1 - p)^n$

• $P(\text{first } k \text{ heads, then } n - k \text{ tails}) = p^k (1 - p)^{n-k}$

• Consider a particular ordering (THTHTHT). What is the probability of that exact ordering?

  $$= p^3 \cdot (1 - p)^2$$
P(exactly $k$ heads on $n$ coin flips)?

\[ \binom{n}{k} p^k (1-p)^{n-k} \]

Think of the flips as ordered:
- Ordering 1: T, H, H, T, T, T, ...
- Ordering 2: H, T, H, T, T, T, ...
And so on...

Let's make each ordering with $k$ heads an event... $F_i$

\[ P(F_i) = p^k (1-p)^{n-k} \]

The coin flips are independent!

P(exactly $k$ heads on $n$ coin flips) = P(any one of the events)

P(exactly $k$ heads on $n$ coin flips) = P($F_1$ or $F_2$ or $F_3$... )

Those events are mutually exclusive!
Today

DeMorgan’s

OR

P(E ∪ F)

Just Add!

Inclusion Exclusion

Mutually Exclusive?

AND

P(EF)

Just Multiply!

Independent?

Chain Rule

Independent?
Set Operations Review

- Say E and F are subsets of S
Set Operations Review

- Say $E$ and $F$ are events in $S$

  Event that is in $E$ or $F$

  $E \cup F$

- $S = \{1, 2, 3, 4, 5, 6\}$ die roll outcome
- $E = \{1, 2\}$ $F = \{2, 3\}$ $E \cup F = \{1, 2, 3\}$
• Say E and F are events in S

\[ S = \{1, 2, 3, 4, 5, 6\} \text{ die roll outcome} \]

\[ E = \{1, 2\} \quad F = \{2, 3\} \quad E \cap F = \{2\} \]

• **Note:** *mutually exclusive* events means \( E \cap F = \emptyset \)
• Say $E$ and $F$ are events in $S$

  Event that is not in $E$ (called complement of $E$)

  $E^c$ or $\sim E$

  - $S = \{1, 2, 3, 4, 5, 6\}$  die roll outcome
  - $E = \{1, 2\}$  $E^c = \{3, 4, 5, 6\}$
• Say E and F are events in S

DeMorgan’s Laws

\[(E \cup F)^c = E^c \cap F^c\]

\[(E \cap F)^c = E^c \cup F^c\]
Augustus De Morgan

- British Mathematician who wrote the book “Formal Logic” in 1847
- Celebrity lookalike is Jason Alexander from Seinfeld.
• $m$ strings are hashed (unequally) into a hash table with $n$ buckets
  ▪ Each string hashed is an independent trial, with probability $p_i$ of getting hashed to bucket $i$
  ▪ $E = \text{at least one string hashed to first bucket}$
  ▪ What is $P(E)$?

• Solution

*To the white board*
• $m$ strings are hashed (unequally) into a hash table with $n$ buckets
  ▪ Each string hashed is an independent trial, with probability $p_i$ of getting hashed to bucket $i$
  ▪ $E = \text{at least one}$ string hashed to first bucket
  ▪ What is $P(E)$?

• Solution

*To the white board*
• $m$ strings are hashed (unequally) into a hash table with $n$ buckets
  ▪ Each string hashed is an independent trial, with probability $p_i$ of getting hashed to bucket $i$
  ▪ $E = \text{At least 1 of}$ buckets 1 to $k$ has $\geq 1$ string hashed to it

• Solution
  ▪ $F_i = \text{at least one string hashed into } i\text{-th bucket}$
  ▪ $P(E) = P(F_1 \cup F_2 \cup \ldots \cup F_k) = 1 - P((F_1 \cup F_2 \cup \ldots \cup F_k)^c)$
    \[= 1 - P(F_1^c F_2^c \ldots F_k^c)\]  (DeMorgan’s Law)
  ▪ $P(F_1^c F_2^c \ldots F_k^c) = P(\text{no strings hashed to buckets 1 to } k)$
    \[= (1 - p_1 - p_2 - \ldots - p_k)^m\]
  ▪ $P(E) = 1 - (1 - p_1 - p_2 - \ldots - p_k)^m$
• $m$ strings are hashed (unequally) into a hash table with $n$ buckets
  - Each string hashed is an independent trial, with probability $p_i$ of getting hashed to bucket $i$
  - $E = \text{Each of$ buckets 1 to } k \text{ has } \geq 1 \text{ string hashed to it}$
No, Really, More Hash Tables

This is fine.
• $m$ strings are hashed (unequally) into a hash table with $n$ buckets
  - Each string hashed is an independent trial, with probability $p_i$ of getting hashed to bucket $i$
  - $E = \text{Each of buckets 1 to } k \text{ has } \geq 1 \text{ string hashed to it}$

• Solution
  - $F_i = \text{at least one string hashed into } i\text{-th bucket}$
  - $P(E) = P(F_1 F_2 \ldots F_k) = 1 - P((F_1 F_2 \ldots F_k)^c)$
    $= 1 - P(F_1^c \cup F_2^c \cup \ldots \cup F_k^c)$ \hspace{1cm} (DeMorgan’s Law)
    $= 1 - 
    \sum_{i=1}^{k} (-1)^{r+1} \sum_{i_1 < \ldots < i_r} P(F_{i_1}^c F_{i_2}^c \ldots F_{i_r}^c)$

where $P(F_{i_1}^c F_{i_2}^c \ldots F_{i_r}^c) = (1 - p_{i_1} - p_{i_2} - \ldots - p_{i_r})^m$
It is expected that this last example will take some review!
Here we are

Source: The Hobbit
100,000 samples

G₁ G₂ G₃ G₄ G₅ T

6 observations per sample
Discovered Pattern

```
|Piech-2:DNA piech$ python findStructure.py
size data = 100000
p(G1) = 0.500
p(G2) = 0.545
p(G3) = 0.299
p(G4) = 0.701
p(G5) = 0.600
p(T) = 0.390
p(T and G1) = 0.291, P(T)p(G1) = 0.195
p(T and G2) = 0.300, P(T)p(G2) = 0.213
p(T and G3) = 0.116, P(T)p(G3) = 0.117
p(T and G4) = 0.273, P(T)p(G4) = 0.273
p(T and G5) = 0.309, P(T)p(G5) = 0.234
```

```
... p(T and G5 | G2) = 0.450
p(T | G2)p(G5 | G2) = 0.450
```
Discovered Pattern

```
Piech-2:~$ python findStruct.py
size data = 100000
p(G1) = 0.500
p(G2) = 0.545
p(G3) = 0.299
p(G4) = 0.701
p(G5) = 0.600
p(T) = 0.390
p(T and G1) = 0.291, P(T)p(G1) = 0.195
p(T and G2) = 0.300, P(T)p(G2) = 0.213
p(T and G3) = 0.116, P(T)p(G3) = 0.117
p(T and G4) = 0.273, P(T)p(G4) = 0.273
p(T and G5) = 0.309, P(T)p(G5) = 0.234
```

\[ p(T \text{ and } G5 \mid G2) = 0.450 \]
\[ p(T \mid G2)p(G5 \mid G2) = 0.450 \]
Discovered Pattern

```
$ python findStructure.py
size data =  100000
p(G1) =  0.500
p(G2) =  0.545
p(G3) =  0.299
p(G4) =  0.701
p(G5) =  0.600
p(T) =  0.390
p(T and G1) =  0.291 , P(T)p(G1) =  0.195
p(T and G2) =  0.300 , P(T)p(G2) =  0.213
p(T and G3) =  0.116 , P(T)p(G3) =  0.117
p(T and G4) =  0.273 , P(T)p(G4) =  0.273
p(T and G5) =  0.309 , P(T)p(G5) =  0.234

... 

p(T and G5 | G2) =  0.450
p(T | G2)p(G5 | G2) =  0.450
```
Discovered Pattern

$\texttt{Piech-2:DNA piech}\$ python findStructure.py

size data = 100000

\begin{align*}
    p(G1) &= 0.500 \\
    p(G2) &= 0.545 \\
    p(G3) &= 0.299 \\
    p(G4) &= 0.701 \\
    p(G5) &= 0.600 \\
    p(T) &= 0.390 \\
    p(T \text{ and } G1) &= 0.291, \quad P(T)p(G1) = 0.195 \\
    p(T \text{ and } G2) &= 0.380, \quad P(T)p(G2) = 0.213 \\
    p(T \text{ and } G3) &= 0.116, \quad P(T)p(G3) = 0.117 \\
    p(T \text{ and } G4) &= 0.273, \quad P(T)p(G4) = 0.273 \\
    p(T \text{ and } G5) &= 0.309, \quad P(T)p(G5) = 0.234
\end{align*}

\[\vdots\]

\[p(T \text{ and } G5 \mid G2) = 0.450\]
\[p(T \mid G2)p(G5 \mid G2) = 0.450\]
Discovered Pattern

$p(T \text{ and } G5 \mid G2) = 0.450$

$p(T \mid G2)p(G5 \mid G2) = 0.450$
Only Causal Structure that Fits

- $p(G_1) = 0.5$
- $p(G_5) = 0.6$
- $p(G_2 | G_5) = 0.9$
- $p(\sim G_5) = 0.2$
- $p(T | G_1 \text{ and } G_2) = 0.9$
- $p(T | \sim G_1 \text{ or } \sim G_2) = 0.2$

These genes don’t impact T
[if I have time]
Phew!
Mutual exclusion
And
Independence

Are two properties of events that make it easy to calculate probabilities.
In the conditional paradigm, the formulas of probability are preserved.
Independence relationships can change with conditioning.

If E and F are independent, that does not mean they will still be independent given another event G.

There is additional reading about this in the course reader. You will explore this more in depth in CS228.
Two Great Tastes

Conditional Probability

Independence
Two events $E$ and $F$ are called **conditionally independent given** $G$, if

$$P(EF|G) = P(E|G)P(F|G)$$

Or, equivalently if:

$$P(E|FG) = P(E|G)$$
And Learn
What is the probability that a user will watch *Life is Beautiful*?

\[
P(E) = \lim_{n \to \infty} \frac{n(E)}{n} \approx \frac{\# \text{people who watched movie}}{\# \text{people on Netflix}}
\]

\[
P(E) = \frac{10,234,231}{50,923,123} = 0.20
\]
What is the probability that a user will watch Life is Beautiful, given they watched Amelie?

\[ P(E|F) \]

\[
P(E|F) = \frac{P(EF)}{P(F)} = \frac{\text{#people who watched both}}{\text{#people who watched } F}
\]

\[ P(E|F) = 0.42 \]
Conditioned on liking a set of movies?
Each event corresponds to liking a particular movie

\[ P(E_4|E_1, E_2, E_3)? \]
Is $E_4$ independent of $E_1, E_2, E_3$?
Is $E_4$ independent of $E_1, E_2, E_3$?

$$P(E_4|E_1, E_2, E_3) \overset{?}{=} P(E_4)$$
Is $E_4$ independent of $E_1, E_2, E_3$?

\[ P(E_4|E_1, E_2, E_3) = \frac{P(E_1E_2E_3E_4)}{P(E_1E_2E_3)} \]
What is the probability that a user watched four particular movies?

- There are 13,000 titles on Netflix
- The user watches 30 random titles
- E = movies watched include the given four.

Solution:

\[
P(E) = \frac{\binom{4}{4} \binom{12996}{24}}{\binom{13000}{30}} = 10^{-11}
\]
Netflix and Learn

$E_1$ $E_2$ $E_3$ $E_4$
Like foreign emotional comedies

Assume $E_1$, $E_2$, $E_3$ and $E_4$ are conditionally independent given $K_1$. 

Netflix and Learn
Like foreign emotional comedies

Assume $E_1$, $E_2$, $E_3$ and $E_4$ are conditionally independent given $K_1$. 

Netflix and Learn
Like foreign emotional comedies

Assume $E_1$, $E_2$, $E_3$ and $E_4$ are conditionally independent given $K_1$. 

Netflix and Learn
Like foreign emotional comedies

\[ P(E_4|E_1, E_2, E_3, K_1) = P(E_4|K_1) \]

Assume \( E_1, E_2, E_3 \) and \( E_4 \) are conditionally independent given \( K_1 \).
Conditional independence is a practical, real world way of decomposing hard probability questions.
If E and F are dependent, that does not mean E and F will be dependent when another event happens.
If $E$ and $F$ are independent, that does not mean $E$ and $F$ will be independent when another event happens.
“Exploiting conditional independence to generate fast probabilistic computations is one of the main contributions CS has made to probability theory”

-Judea Pearl wins 2011 Turing Award, “For fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning”