05: Independence

Lisa Yan
April 15, 2020

Quick slide reference

3	Generalized Chain Rule	05a_chain
9	Independence	05b_independence_i
16	Independent Trials	05c_independence_ii
21	Exercises and deMorgan's Laws	LIVE

Generalized Chain Rule

Chain Rule

Definition of conditional probability:

$$P(E|F) = \frac{P(EF)}{P(F)}$$

The Chain Rule:

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

Generalized Chain Rule

$$P(E_1 E_2 E_3 \ldots E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1 E_2) \ldots P(E_n|E_1 E_2 \ldots E_{n-1})$$

Quick check

$P(E_1 E_2 E_3 \ldots E_n) =$
$\quad P(E_1)P(E_2|E_1) \ldots P(E_n|E_1 E_2 \ldots E_{n-1})$ Chain Rule

You are going to a friend's Halloween party.

Let C = there is candy
M = there is music
W = you wear a costume
E = no one wears your costume

An awesome party means that all of these events must occur.

What is $P(\text{awesome party}) = P(CMWE)$?

A. $P(C)P(M|C)P(W|CM)P(E|CMW)$
B. $P(M)P(C|M)P(W|MC)P(E|MCW)$
C. $P(W)P(E|W)P(CM|EW)$
D. A, B, and C
E. None/other

Quick check

$P(E_1 E_2 E_3 \dots E_n) = P(E_1)P(E_2|E_1) \dots P(E_n|E_1 E_2 \dots E_{n-1})$ Chain Rule

You are going to a friend's Halloween party.

Let C = there is candy
M = there is music
E = no one wears your costume
W = you wear a costume

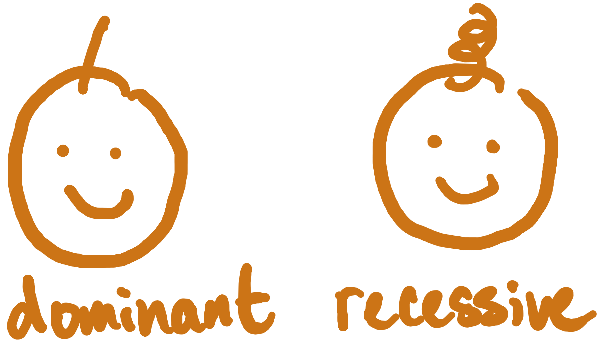
An awesome party means that all of these events must occur.

What is $P(\text{awesome party}) = P(CMEW)$?

A. $P(C)P(M|C)P(E|CM)P(W|CME)$
B. $P(M)P(C|M)P(E|MC)P(W|MCE)$
C. $P(W)P(E|W)P(CM|EW)$
D. A, B, and C
E. None/other

Chain Rule is a way of introducing "order" and "procedure" into probability.

Think of the children



Two parents both have an (A, a) gene pair.

- Each parent will pass on one of their genes (each gene equally likely) to their child.
- The probability of a single child having curly hair (recessive trait) is 0.25.
- There are three children.

What is the probability that all three children have curly hair?

Let E_1, E_2, E_3 be the events that child 1, 2, and 3 have curly hair, respectively.

$$P(E_1 E_2 E_3) = P(E_1) P(E_2 | E_1) P(E_3 | E_1 E_2)$$

Independence I

Independence

Two events E and F are defined as independent if:

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

Otherwise E and F are called dependent events.

If E and F are independent, then:

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

Intuition through proof

Independent events E and F ⟺ $P(EF) = P(E)P(F)$

Statement:

If E and F are independent, then $P(E|F) = P(E)$.

Proof:

$$P(E|F) = \frac{P(EF)}{P(F)}$$ Definition of conditional probability

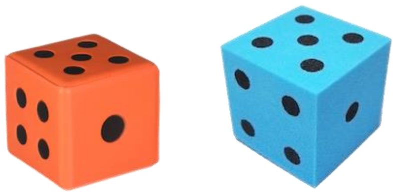
$$= \frac{P(E)P(F)}{P(F)}$$ Independence of E and F

$$= P(E)$$ Taking the bus to cancellation city

Knowing that F happened does not change our belief that E happened.

Dice, our misunderstood friends

Independent events E and F ⟺ $P(EF) = P(E)P(F)$, $P(E|F) = P(E)$



- Roll two 6-sided dice, yielding values D_1 and D_2.
- Let event E: $D_1 = 1$
  event F: $D_2 = 6$
  event G: $D_1 + D_2 = 5$

$G = \{(1,4), (2,3), (3,2), (4,1)\}$

1. Are E and F independent?

$$P(E) = 1/6$$
$$P(F) = 1/6$$
$$P(EF) = 1/36$$

☑ <u>independent</u>

2. Are E and G independent?

$$P(E) = 1/6$$
$$P(G) = 4/36 = 1/9$$
$$P(EG) = 1/36 \neq P(E)P(G)$$

✗ <u>dependent</u>

Generalizing independence

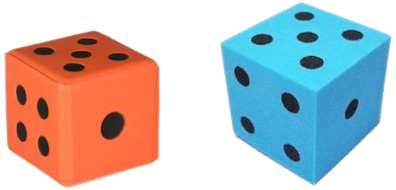
Three events E, F, and G are independent if:

$P(EFG) = P(E)P(F)P(G)$, and
$P(EF) = P(E)P(F)$, and
$P(EG) = P(E)P(G)$, and
$P(FG) = P(F)P(G)$

n events $E_1, E_2, \ldots, E_n$ are independent if:

for $r = 1, \ldots, n$:
for every subset $E_1, E_2, \ldots, E_r$:
$P(E_1, E_2, \ldots, E_r) = P(E_1)P(E_2) \cdots P(E_r)$

Dice, increasingly misunderstood (still our friends)



- Each roll of a 6-sided die is an independent trial.
- Two rolls: D_1 and D_2.
- Let event E: $D_1 = 1$
  event F: $D_2 = 6$
  event G: $D_1 + D_2 = 7$

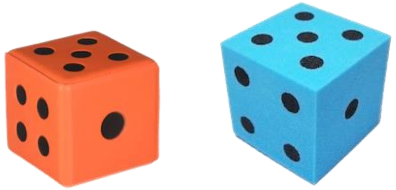
$G = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$

1. Are E and F independent? ✔
2. Are E and G independent?
3. Are F and G independent?
4. Are E, F, G independent?

$$P(EF) = 1/36$$

Dice, increasingly misunderstood (still our friends)

- Each roll of a 6-sided die is an independent trial.
- Two rolls: D_1 and D_2.
- Let event E: $D_1 = 1$
  event F: $D_2 = 6$
  event G: $D_1 + D_2 = 7$



$G = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$

1. Are E and F independent? ✓
2. Are E and G independent? ✓
3. Are F and G independent? ✓
4. Are E, F, G independent? ✗

$P(EF) = 1/36$

Pairwise independence is not sufficient to prove independence of >2 events!

Independence II

Independent trials

We often are interested in experiments consisting of n **independent trials**.

- n trials, each with the same set of possible outcomes
- n-way independence: an event in one subset of trials is independent of events in other subsets of trials

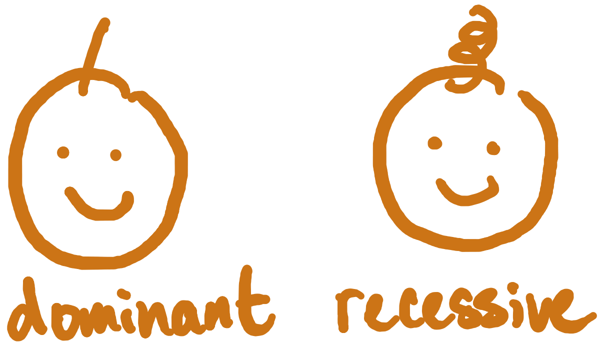
Examples:

- Flip a coin n times
- Roll a die n times
- Send a multiple choice survey to n people
- Send n web requests to k different servers

Think of the children as independent trials

Two parents both have an (A, a) gene pair.

- Each parent will pass on one of their genes (each gene equally likely) to their child.
- The probability of a single child having curly hair (recessive trait) is 0.25.
- There are three children. **Each child is an independent trial**.



What is the probability that all three children have curly hair?

Let E_1, E_2, E_3 be the events that child 1, 2, and 3 have curly hair, respectively.

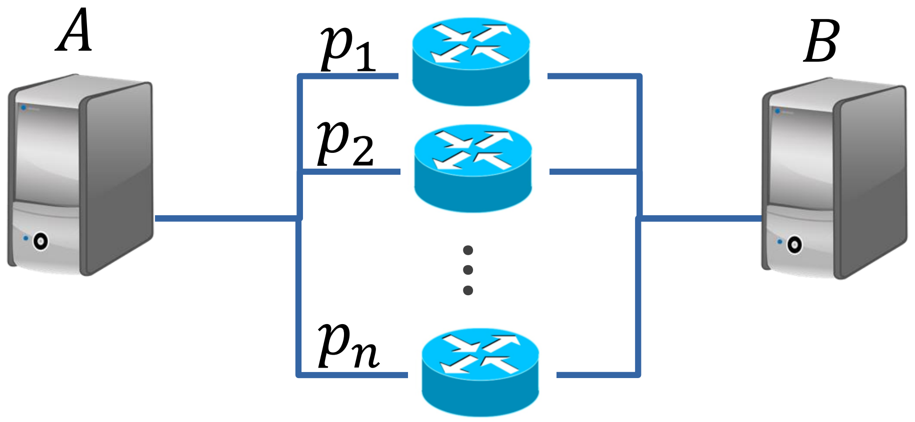
$$P(E_1 E_2 E_3) = P(E_1)P(E_2|E_1)P(E_3|E_1 E_2)$$

Network reliability

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 $P(E)$?

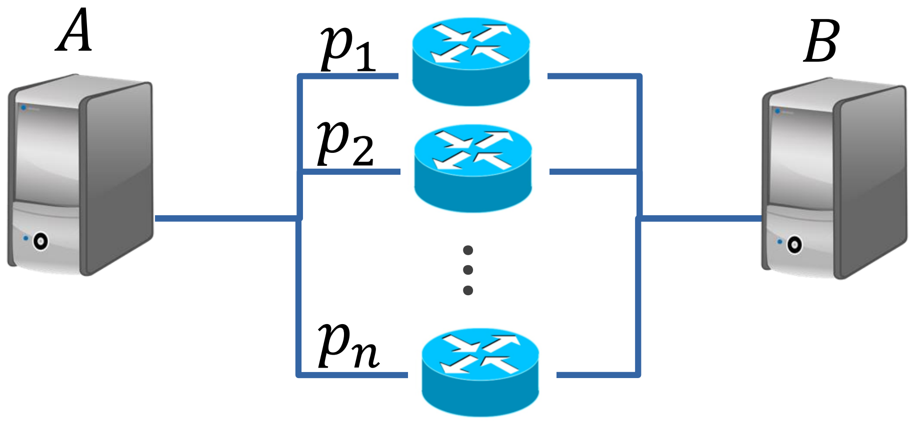


Network reliability

Consider the following parallel network:

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

What is $P(E)$?



$$
\begin{aligned}
P(E) &= P(\geq 1 \text{ one router works}) \\
&= 1 - P(\text{all routers fail}) \\
&= 1 - (1 - p_1)(1 - p_2)\cdots(1 - p_n) \\
&= 1 - \prod_{i=1}^{n}(1 - p_i)
\end{aligned}
$$

≥ 1 with independent trials: take complement

Conditional Independence

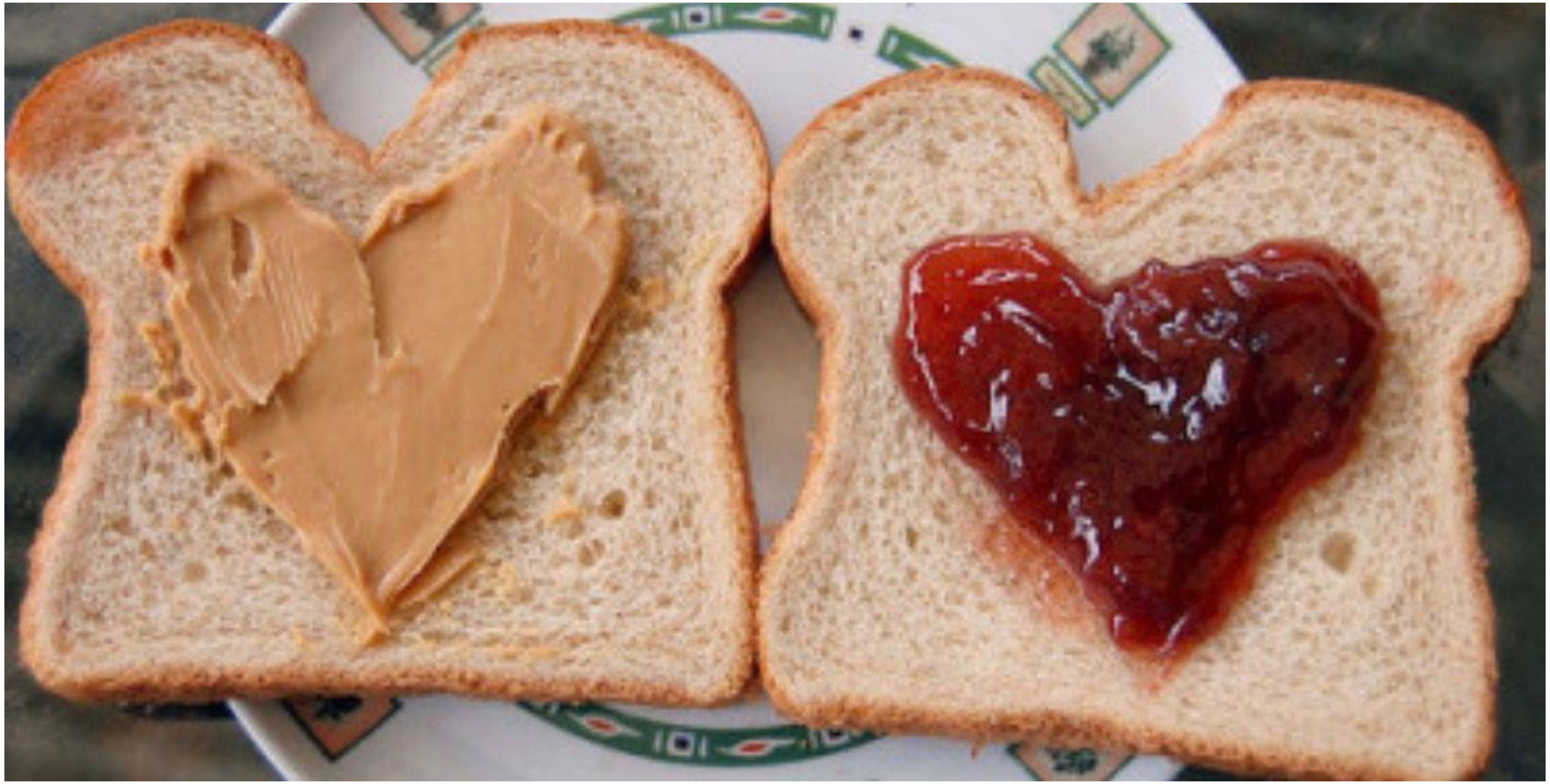
Conditional Paradigm

For any events A, B, and E, you can condition consistently on E, and all formulas still hold:

Axiom 1	$0 \leq P(A\|E) \leq 1$
Corollary 1 (complement)	$P(A\|E) = 1 - P(A^C\|E)$
Transitivity	$P(AB\|E) = P(BA\|E)$
Chain Rule	$P(AB\|E) = P(B\|E)P(A\|BE)$
Bayes' Theorem	$P(A\|BE) = \dfrac{P(B\|AE)P(A\|E)}{P(B\|E)}$

BAE 's theorem?

Conditional Independence



Conditional Probability Independence

Conditional Independence

Independent events E and F ⟺ $P(EF) = P(E)P(F)$, $P(E|F) = P(E)$

Two events A and B are defined as conditionally independent given E if:

$$P(AB|E) = P(A|E)P(B|E)$$

An equivalent definition:

A. $P(A|B) = P(A)$
B. $P(A|BE) = P(A)$
C. $P(A|BE) = P(A|E)$

Conditional Independence

Independent events E and F ⟺ $P(EF) = P(E)P(F)$
$P(E|F) = P(E)$

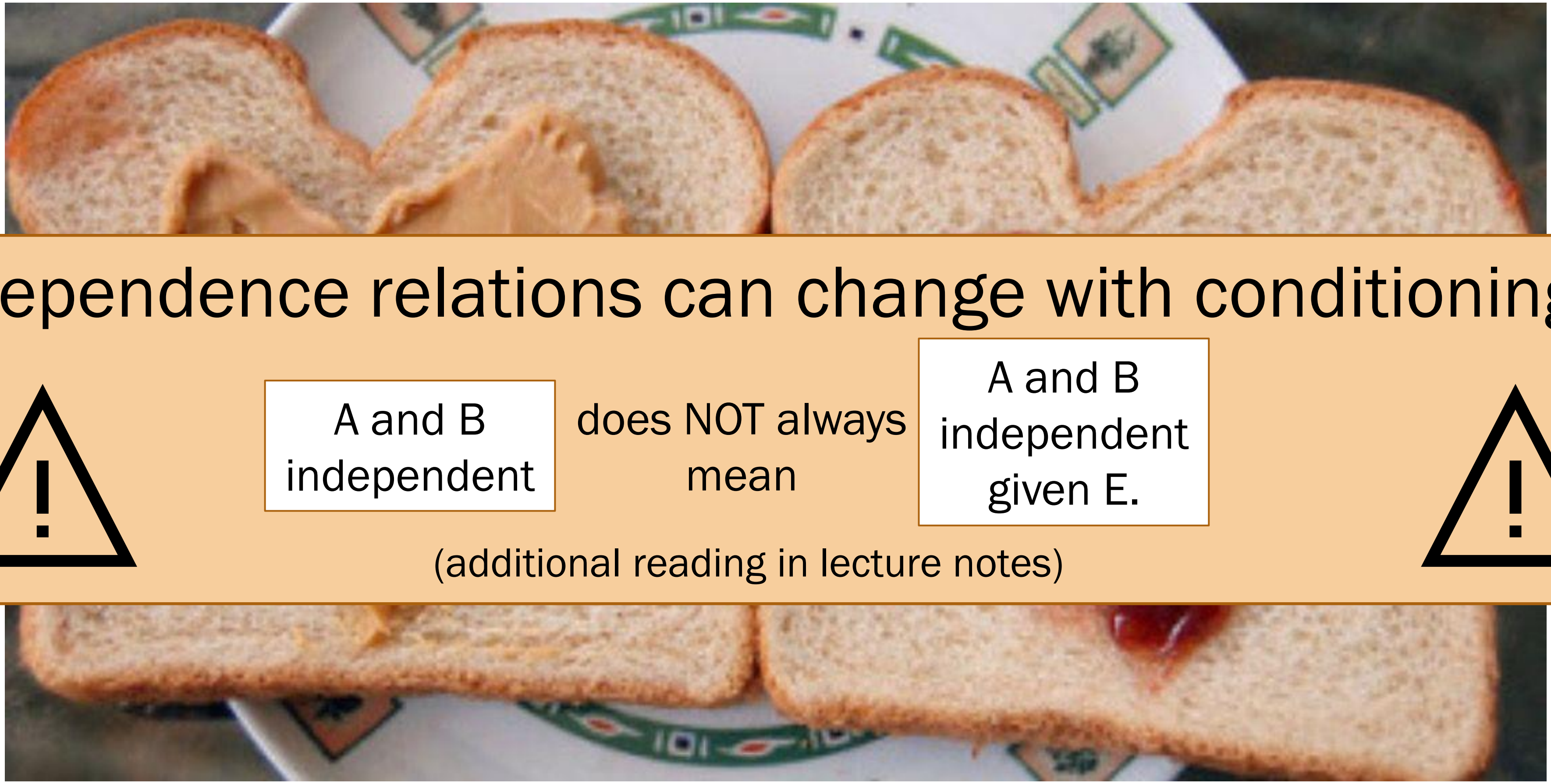
Two events A and B are defined as conditionally independent given E if:

$$P(AB|E) = P(A|E)P(B|E)$$

An equivalent definition:

A. $P(A|B) = P(A)$
B. $P(A|BE) = P(A)$
C. $P(A|BE) = P(A|E)$

Conditional Independence



Independence relations can change with conditioning.

A and B independent does NOT always mean A and B independent given E.

(additional reading in lecture notes)

Conditional Probability

Independence

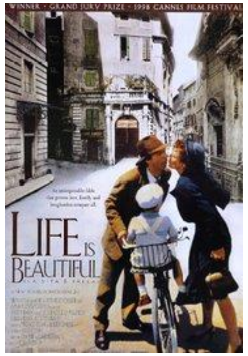
Netflix and Condition

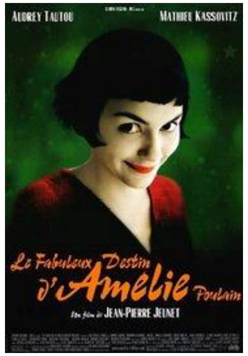
Review

Let E = a user watches Life is Beautiful.

Let F = a user watches Amelie.

What is $P(E)$?





$$P(E) \approx \frac{\text{\# people who have watched movie}}{\text{\# people on Netflix}} = \frac{10{,}234{,}231}{50{,}923{,}123} \approx 0.20$$

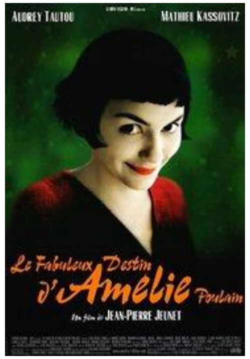
What is the probability that a user watches Life is Beautiful, given they watched Amelie?

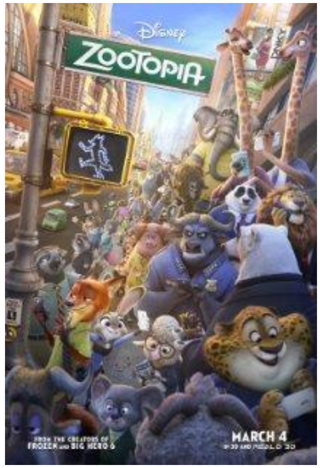
$$P(E|F) = \frac{P(EF)}{P(F)} = \frac{\text{\# people who have watched both}}{\text{\# people who have watched Amelie}} \approx 0.42$$

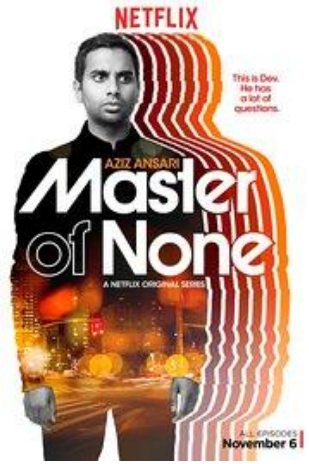
Netflix and Condition

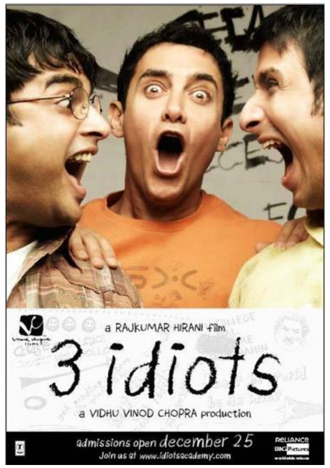
Review

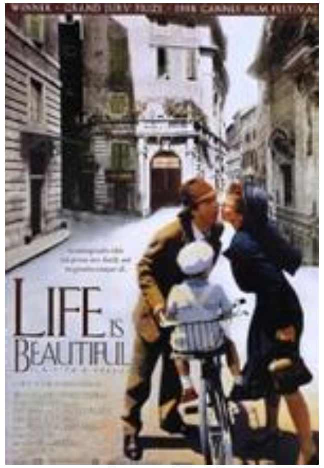
Let E be the event that a user watches the given movie.
Let F be the event that the same user watches Amelie.









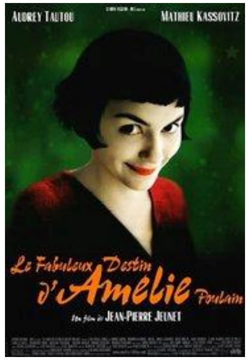


$P(E) = 0.19$	$P(E) = 0.32$	$P(E) = 0.20$	$P(E) = 0.09$	$P(E) = 0.20$
$P(E\|F) = 0.14$	$P(E\|F) = 0.35$	$P(E\|F) = 0.20$	$P(E\|F) = 0.72$	$P(E\|F) = 0.42$

Independent!

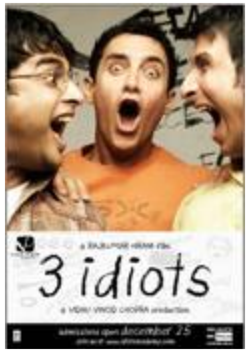
Netflix and Condition (on many movies)

Watched:



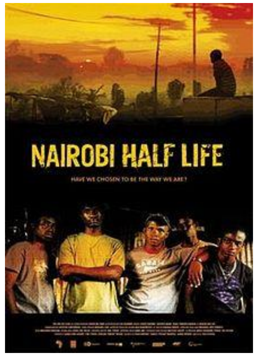
E_1

E_2



E_3

Will they watch?



E_4

What if $E_1 E_2 E_3 E_4$ are not independent? (e.g., all international emotional comedies)

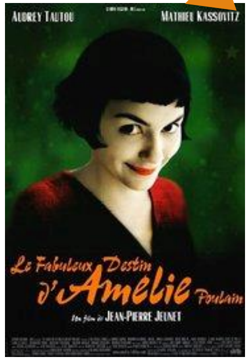
$$P(E_4 | E_1 E_2 E_3) = \frac{P(E_1 E_2 E_3 E_4)}{P(E_1 E_2 E_3)} = \frac{\text{\# people who have watched all 4}}{\text{\# people who have watched those 3}}$$

We need to keep track of an exponential number of movie-watching statistics

Netflix and Condition (on many movies)

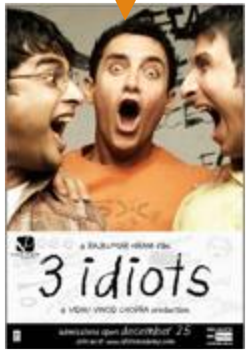
K: likes international emotional comedies

Watched:



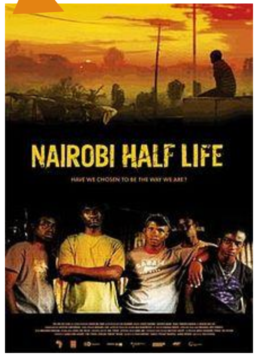
E_1

E_2



E_3

Will they watch?



E_4

Assume: $E_1 E_2 E_3 E_4$ are *conditionally independent given* K

$$P(E_4|E_1E_2E_3) = \frac{P(E_1E_2E_3E_4)}{P(E_1E_2E_3)}$$

$$P(E_4|E_1E_2E_3K) = P(E_4|K)$$

An easier probability to store and compute!

Conditional independence is a Big Deal

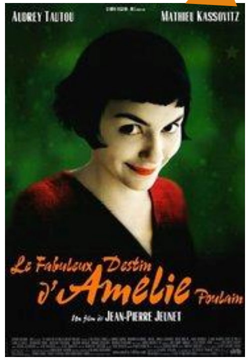
Conditional independence is a practical, real-world way of decomposing hard probability questions.

> "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"

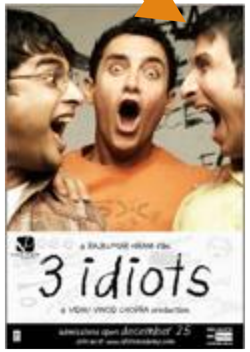
Netflix and Condition

K: likes international emotional comedies

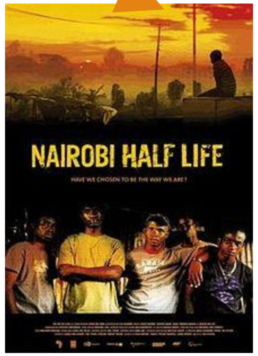


E_1

E_2



E_3



E_4

Challenge: How do we determine K? Stay tuned in 6 weeks' time!

$E_1E_2E_3E_4$ are dependent

$E_1E_2E_3E_4$ are conditionally independent given K

Dependent events can become conditionally independent.
And vice versa: Independent events can become conditionally dependent.

05: Independence (live)

Oishi Banerjee and Cooper Raterink
Adapted from Lisa Yan
July 1, 2020

Independence

Two events E and F are defined as independent if:

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

Equivalently, for independent events E and F,

- $P(E|F) = P(E)$

Think

Slide 24 has two questions to think over by yourself. We'll go over it together afterwards.

Post any clarifications here!

https://us.edstem.org/courses/667/discussion/84208

Think by yourself: 2 min

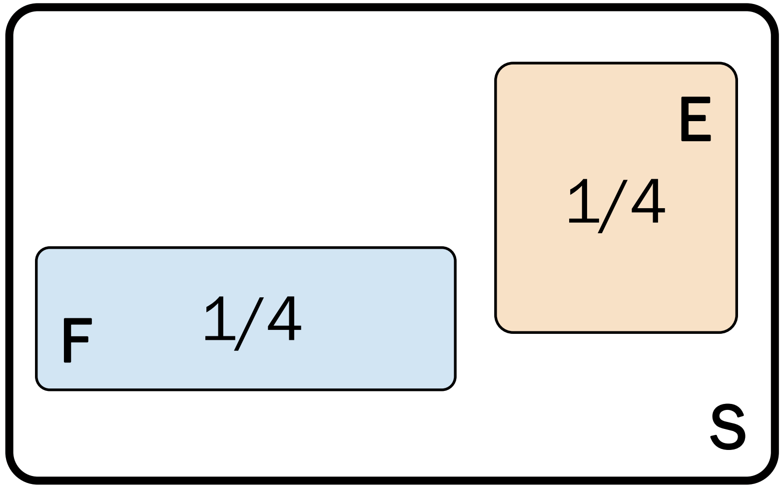
(by yourself)

Independence?

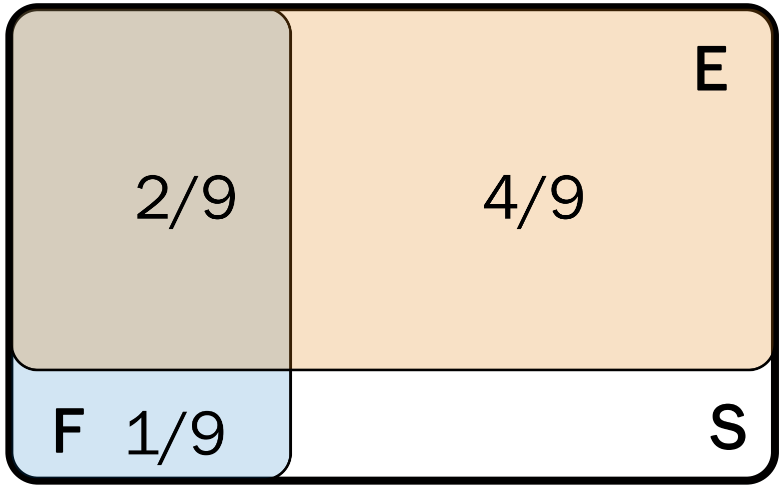
Independent events E and F $\Leftrightarrow$ $P(EF) = P(E)P(F)$, $P(E|F) = P(E)$

1. Two events E and F are independent if:
   - A. Knowing that F happens means that E can't happen.
   - ✓ B. Knowing that F happens doesn't change probability that E happened.
2. Are E and F independent in the following pictures (not to scale)?

A. ✗



B. ✓



Be careful:
- Independence is NOT mutual exclusion.
- Independence is difficult to visualize graphically.

(by yourself)

Independence

Two events E and F are defined as <u>independent</u> if:

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

For independent events E and F,

- $P(E|F) = P(E)$
- E and F^C are independent.

new

Independence of complements

Statement:

If E and F are independent, then E and F^C are independent.

Proof:

$P(EF^C) = P(E) - P(EF)$	Intersection
$= P(E) - P(E)P(F)$	Independence of E and F
$= P(E)[1 - P(F)]$	Factoring
$= P(E)P(F^C)$	Complement
E and F^C are independent	Definition of independence

Knowing that F **did or didn't happen** does not change our belief that E happened.

Independence

Review

Two events E and F are defined as <u>independent</u> if:

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

For independent events E and F,

- $P(E|F) = P(E)$
- E and F^C are independent

Independent trials are when we observe independent sub-experiments, each of which has the same set of possible outcomes.

Breakout Rooms

Check out the questions on the next slide (Slide 30). Post any clarifications here!

https://us.edstem.org/courses/667/discussion/84208

Breakout rooms: 5 min. Introduce yourself!

(biased) Coin Flips

Suppose we flip a coin n times. Each coin flip is an **independent trial** with probability p of coming up heads. Write an expression for the following:

1. $P(n$ heads on n coin flips$)$
2. $P(n$ tails on n coin flips$)$
3. $P($first k heads, then $n - k$ tails$)$
4. $P($exactly k heads on n coin flips$)$

(biased) Coin Flips

Suppose we flip a coin n times. Each coin flip is an **independent trial** with probability p of coming up heads. Write an expression for the following:

1. $P(n$ heads on n coin flips$)$
2. $P(n$ tails on n coin flips$)$
3. $P($first k heads, then $n-k$ tails$)$
4. $P($exactly k heads on n coin flips$)$

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

# of mutually exclusive outcomes

P(a particular outcome's k heads on n coin flips)

Make sure you understand #4! It will come up again.

Interlude for announcements

Announcements

Sections start tomorrow!

Live on Zoom: 12:30-1:30 PT Thursday (recorded)

Problem Set 1

due: 1pm PT Wednesday
1st late due date: 1pm PT **Monday** (because Friday is a holiday)

Problem Set 2

due: 1pm Friday July 10

Partial answer checkers for psets 1 and 2 available on Gradescope!

Probability of events

E or F
$P(E \cup F)$

Mutually exclusive?

Just add!

$P(E) + P(F)$

Inclusion-Exclusion Principle

$P(E) + P(F) - P(E \cap F)$

E and F
$P(EF)$

Independent?

Just multiply!

Chain Rule

Probability of events

E or F
$P(E \cup F)$

Mutually exclusive?

Just add!
$P(E) + P(F)$

Inclusion-Exclusion Principle
$P(E) + P(F) - P(E \cap F)$

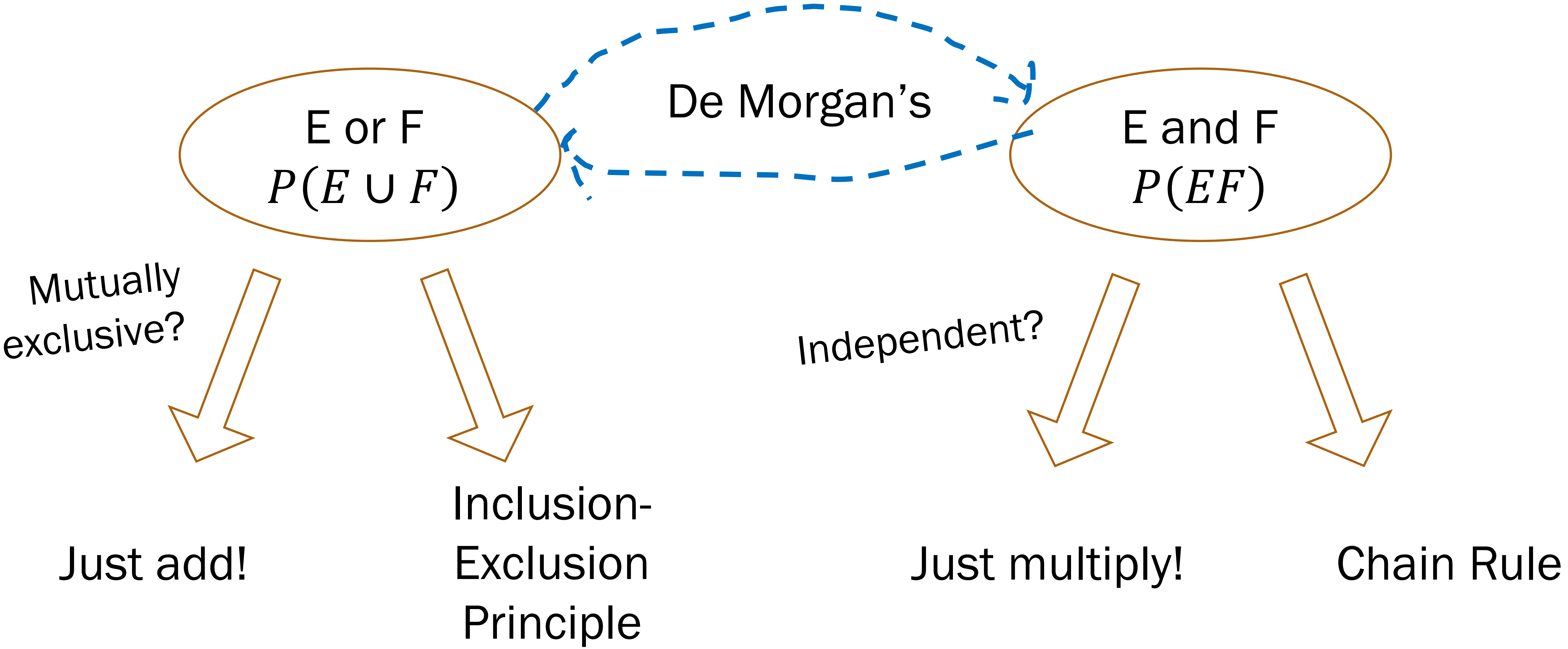
E and F
$P(EF)$

Independent?

Just multiply!
$P(E)P(F)$

Chain Rule
$P(E)P(F|E)$
or
$P(F)P(E|F)$

Probability of events



Augustus De Morgan

Augustus De Morgan (1806–1871):

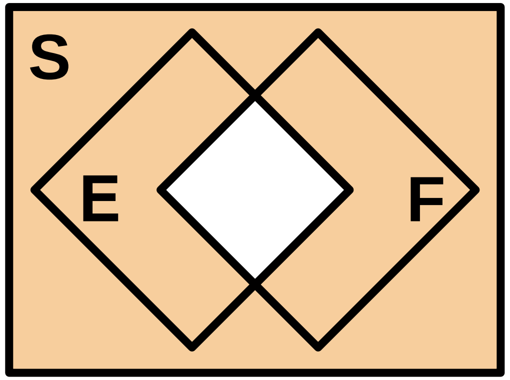
British mathematician who wrote the book *Formal Logic* (1847).

He looked remarkably similar to Jason Alexander (George from Seinfeld)

(but that's not important right now)

De Morgan's Laws

DeMorgan's lets you switch from AND to OR.



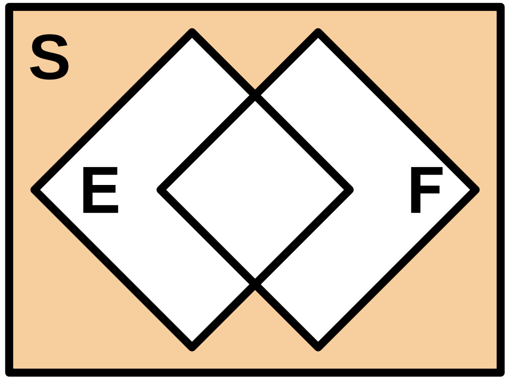
$$(E \cap F)^C = E^C \cup F^C$$

$$\left(\bigcap_{i=1}^{n} E_i\right)^C = \bigcup_{i=1}^{n} E_i^C$$

In probability:

$$P(E_1 E_2 \cdots E_n)$$
$$= 1 - P\left((E_1 E_2 \cdots E_n)^C\right)$$
$$= 1 - P\left(E_1^C \cup E_2^c \cup \cdots \cup E_n^c\right)$$

Great if E_i^C mutually exclusive!



$$(E \cup F)^C = E^C \cap F^C$$

$$\left(\bigcup_{i=1}^{n} E_i\right)^C = \bigcap_{i=1}^{n} E_i^C$$

In probability:

$$P(E_1 \cup E_2 \cup \cdots \cup E_n)$$
$$= 1 - P\left((E_1 \cup E_2 \cup \cdots \cup E_n)^C\right)$$
$$= 1 - P(E_1^c E_2^c \cdots E_n^c)$$

Great if E_i independent!

Think, then Breakout Rooms

Check out the questions on the next slide (Slide 40). **These are challenging problems.**

Post any clarifications here!

https://us.edstem.org/courses/667/discussion/84208

Think by yourself: 2 min

Breakout rooms: 5 min

Hash table fun

- m strings are hashed (unequally) into a hash table with n buckets.
- Each string hashed is an independent trial w.p. p_j of getting hashed into bucket j.

What is $P(E)$ if

1. E = bucket 1 has $\geq$ 1 string hashed into it?

2. E = at least 1 of buckets 1 to k has $\geq$ 1 string hashed into it?

Hash table fun

- m strings are hashed (unequally) into a hash table with n buckets.
- Each string hashed is an independent trial w.p. p_j of getting hashed into bucket j.

What is $P(E)$ if

1. E = bucket 1 has $\geq$ 1 string hashed into it?

Define S_i = string i is hashed into bucket 1
S_i^C = string i is not hashed into bucket 1

$P(S_i) = p_1$
$P(S_i^C) = 1 - p_1$

Hash table fun

- m strings are hashed (unequally) into a hash table with n buckets.
- Each string hashed is an independent trial w.p. p_j of getting hashed into bucket j.

What is $P(E)$ if

1. E = bucket 1 has $\geq$ 1 string hashed into it?

Define S_i = string i is hashed into bucket 1
S_i^C = string i is not hashed into bucket 1

$P(S_i) = p_1$
$P(S_i^C) = 1 - p_1$

WTF (not-real acronym for Want To Find):

$$P(E) = P(S_1 \cup S_2 \cup \cdots \cup S_m)$$

$$= 1 - P\left((S_1 \cup S_2 \cup \cdots \cup S_m)^C\right)$$ Complement

$$= 1 - P\left(S_1^C S_2^C \cdots S_m^C\right)$$ De Morgan's Law

$$= 1 - P\left(S_1^C\right)P\left(S_2^C\right) \cdots P(S_m^C) = 1 - \left(P\left(S_1^C\right)\right)^m$$ S_i independent trials

$$= 1 - (1 - p_1)^m$$

More hash table fun: Possible approach?

- m strings are hashed (unequally) into a hash table with n buckets.
- Each string hashed is an independent trial w.p. p_j of getting hashed into bucket j.

What is $P(E)$ if

1. E = bucket 1 has $\geq$ 1 string hashed into it?
2. E = at least 1 of buckets 1 to k has $\geq$ 1 string hashed into it?

Define F_j = bucket j has at least one string in it

$$P(E) = P(F_1 \cup F_2 \cup \cdots \cup F_k)$$
$$= 1 - P\left((F_1 \cup F_2 \cup \cdots \cup F_k)^C\right)$$
$$= 1 - P\left(F_1^C F_2^C \cdots F_k^C\right)$$
$$? = 1 - P\left(F_1^C\right) P\left(F_2^C\right) \cdots P\left(F_k^C\right)$$

⚠ F_i bucket events are *dependent*!
So we cannot approach with complement.

More hash table fun

- m strings are hashed (unequally) into a hash table with n buckets.
- Each string hashed is an independent trial w.p. p_j of getting hashed into bucket j.

What is $P(E)$ if

1. E = bucket 1 has ≥ 1 string hashed into it?
2. E = at least 1 of buckets 1 to k has ≥ 1 string hashed into it?

Define F_j = bucket j has at least one string in it

$$P(E) = P(F_1 \cup F_2 \cup \cdots \cup F_k)$$

$$= 1 - P\left((F_1 \cup F_2 \cup \cdots \cup F_k)^C\right)$$

$$= 1 - P\left(F_1^C F_2^C \cdots F_k^C\right)$$

$$= P(\text{no strings hashed to buckets 1 to } k)$$

$$= \left(P(\text{string hashed outside bkts 1 to } k)\right)^m$$

$$= (1 - p_1 - p_2 \ldots - p_k)^m$$

$$= 1 - (1 - p_1 - p_2 \ldots - p_k)^m$$

The fun never stops with hash tables

- m strings are hashed (unequally) into a hash table with n buckets.
- Each string hashed is an independent trial w.p. p_i of getting hashed into bucket i.

1. E = bucket 1 has $\geq$ 1 string hashed into it.
2. E = at least 1 of buckets 1 to k has $\geq$ 1 string hashed into it.
3. E = each of the buckets 1 to k has $\geq$ 1 string hashed into it.

What is $P(E)$? See the lecture notes for the solution!

Dice, increasingly misunderstood (still our friends)

- Each roll of a 6-sided die is an independent trial.
- Two rolls: D_1 and D_2.
- Let event E: $D_1 = 1$
  event F: $D_2 = 6$
  event G: $D_1 + D_2 = 7$
  event H: $D_1 + D_2 = 2$

$G = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$

$H = \{(1,1)\}$

☑ Are E and F independent?

$P(E) = 1/6$
$P(F) = 1/6$
$P(EF) = 1/36 = 1/6 \cdot 1/6$

☑ Are F and G independent?

$P(F) = 1/6$
$P(G) = 1/6$
$P(FG) = 1/36 = 1/6 \cdot 1/6$

✗ Are F and H independent?

$P(F) = 1/6$
$P(H) = 1/36$
$P(FH) = 0 \neq 1/6 \cdot 1/36$

Dice, increasingly misunderstood (still our friends)

- Each roll of a 6-sided die is an independent trial.
- Two rolls: D_1 and D_2.
- Let event E: $D_1 = 1$

  event F: $D_2 = 6$

  event G: $D_1 + D_2 = 7$

$G = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$

Are E and F independent?

$P(E) = 1/6$

$P(F) = 1/6$

$P(EF) = 1/36$

✓

Are E and F independent, <u>when conditioning on G?</u>

✕

$P(E \mid G) = 1/6$

$P(F \mid G) = 1/6$

$P(EF \mid G) = 1/6 \neq 1/6 \cdot 1/6$

Independence doesn't imply conditional independence!

Dice, increasingly misunderstood (still our friends)

- Each roll of a 6-sided die is an independent trial.
- Two rolls: D_1 and D_2.
- Let event G: $D_1 + D_2 = 7$
  event H: $D_1 + D_2 = 2$
  event J: $D_1 + D_2 > 2$

✗ Are H and J independent?

$$P(H) = 1/36$$
$$P(J) = 1 - 1/36 = 35/36$$
$$P(HJ) = 0 \neq 1/36 \cdot 35/36$$

✓ Are H and J independent, when conditioning on G?

$$P(H \mid G) = 0$$
$$P(J \mid G) = 1$$
$$P(HJ \mid G) = 0 = 0 \cdot 1$$

Conditional independence doesn't imply independence either!

Happy 4th of July!

(P.S. Next week's recordings are a bit heavier than usual, so get started early!)