Independence

Review

Chain Rule: $P(EF) = P(E|F)P(F)$

- Goes from an "and" to a conditional and vice versa
- Think about which event you want to condition on

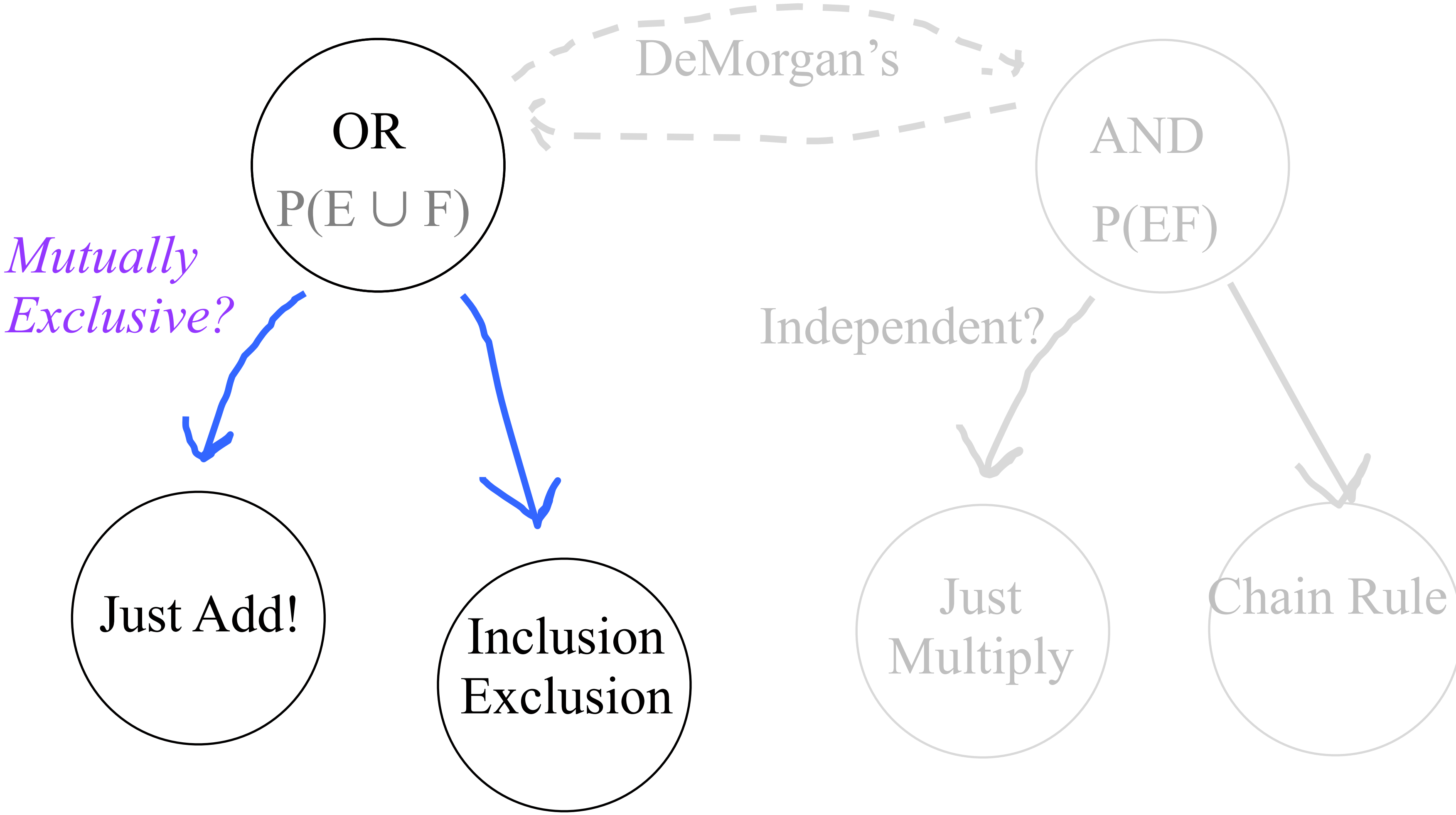
LotP: $P(E) = P(E|F)P(F) + \mathrm{P}(\mathrm{E}|F^C)\mathrm{P}(F^C)$

- We don't know about E but we do know about E|F
- Don't forget about the "and" version and "summation" version

Bayes: $P(E|F) = \dfrac{P(F|E)P(E)}{P(F)}$

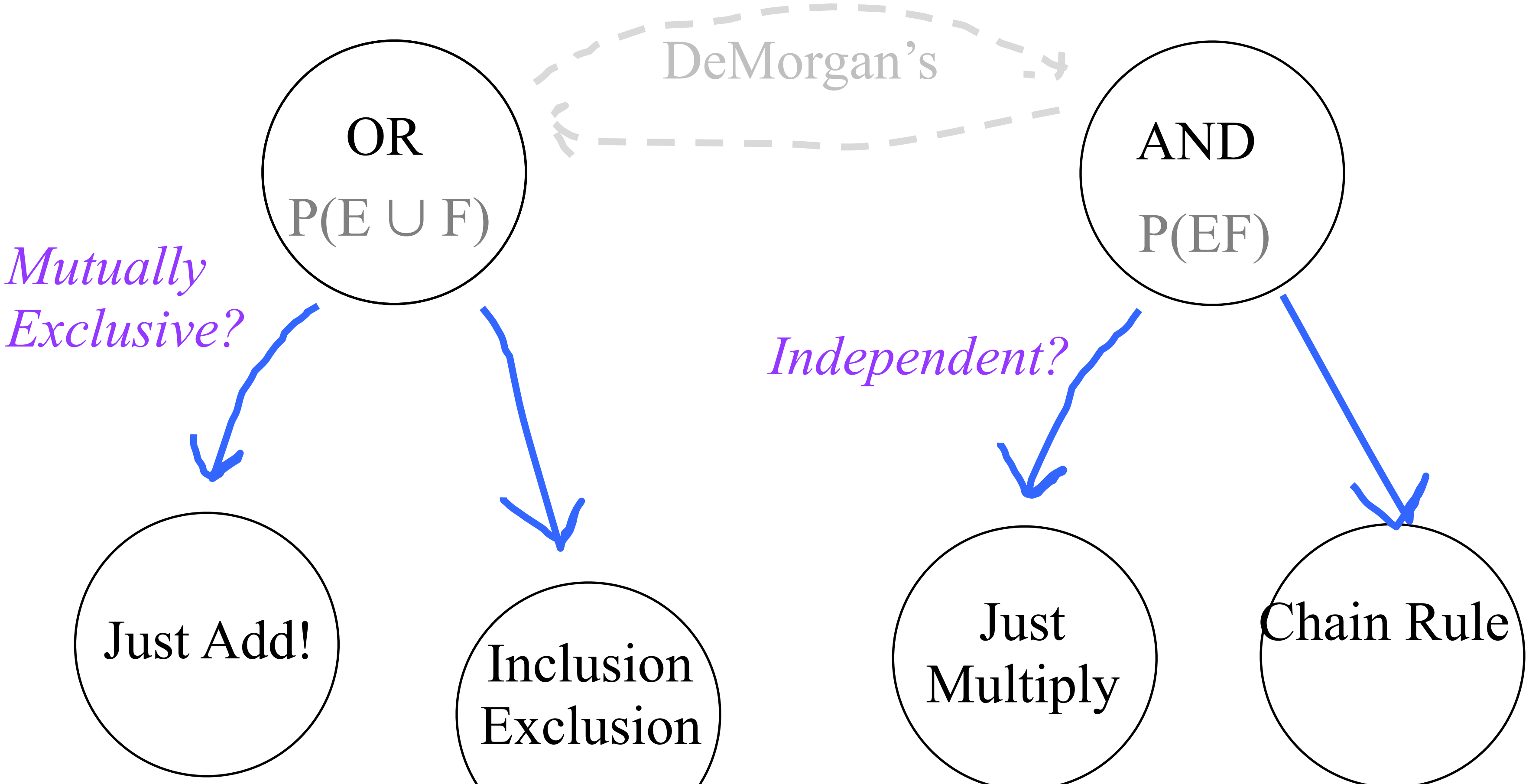
- Good for when E|F is hard but F|E is not so hard
- Common mistake: not trying chain rule first

Today



OR
P(E ∪ F)
DeMorgan's
AND
P(EF)
Mutually Exclusive?
Independent?
Just Add!
Inclusion Exclusion
Just Multiply
Chain Rule

Today



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

$$P(E \cup F) = P(E) + P(F) - P(EF)$$

Probability of “AND”


We the People
insure domestic Tranquility, provide for the common
and our Posterity, do ordain and establish this

Independence

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

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

Definition of conditional probability

$$= \frac{P(A)P(B)}{P(B)}$$

Since A and B are independent

$$= P(A)$$

Taking the bus to cancel city

Knowing that event B happened, doesn't change our belief that A will happen.

Dice, Our Misunderstood Friends

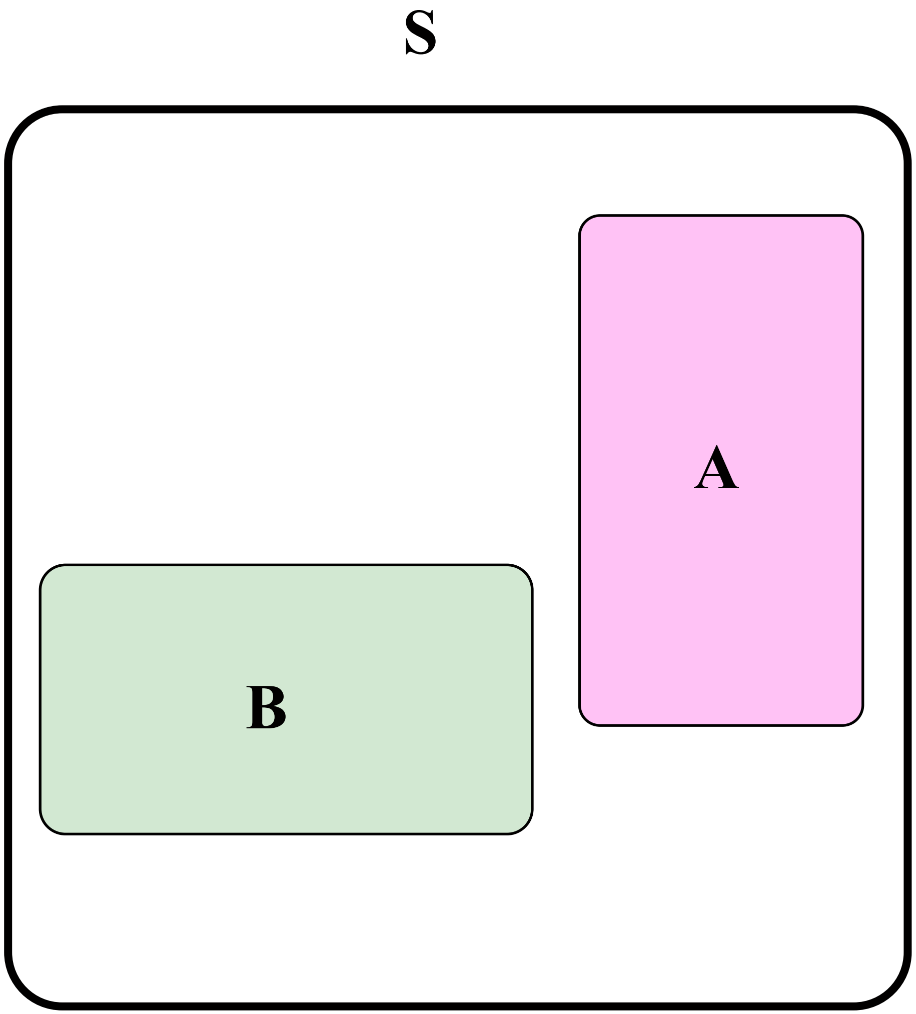
- Roll two 6-sided dice, yielding values D_1 and D_2
  - Let E be event: $D_1 = 1$
  - Let F be event: $D_2 = 1$

- What is P(E), P(F), and P(EF)?
  - $P(E) = 1/6$, $P(F) = 1/6$, $P(EF) = 1/36$
  - $P(EF) = P(E)\,P(F)$ → 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)?
  - $P(E) = 1/6$, $P(G) = 4/36 = 1/9$, $P(EG) = 1/36$
  - $P(EG) \neq P(E)\,P(G)$ → E and G *dependent*

What does independence look like?

Independence?

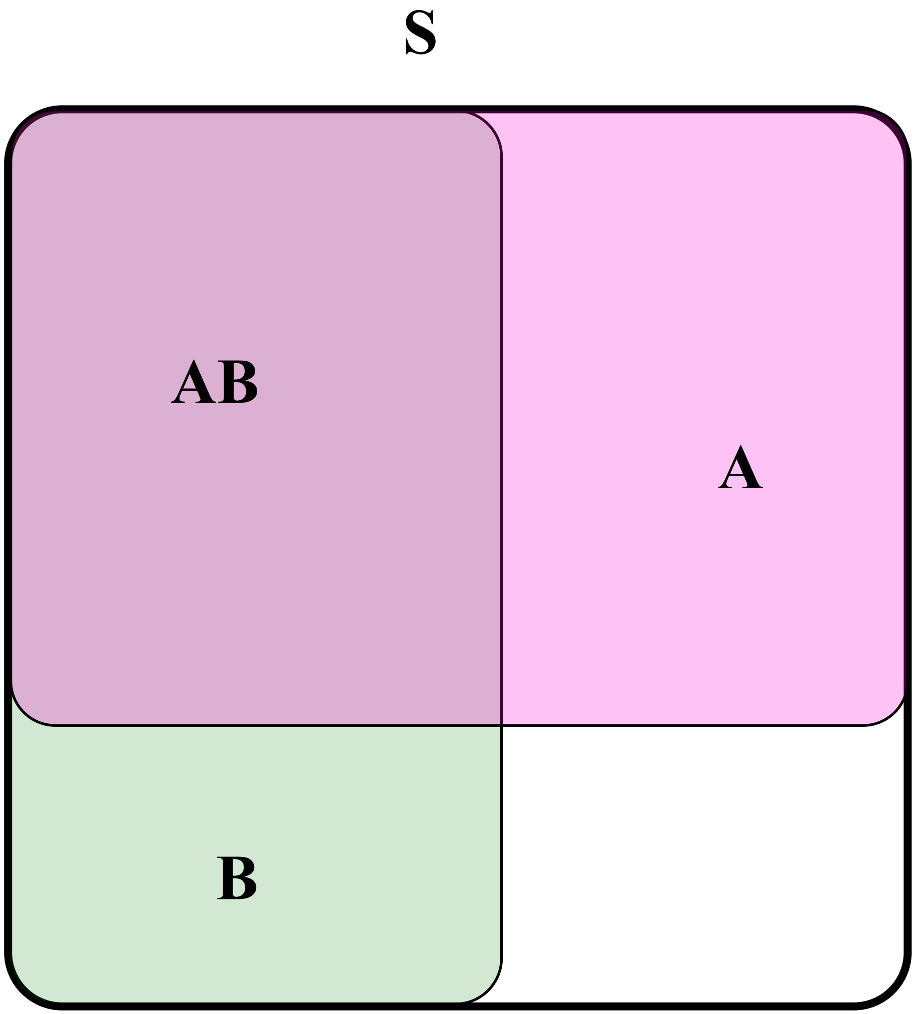


Independence Definition 1:

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

$$\frac{|AB|}{|S|} = \frac{|A|}{|S|} \times \frac{|B|}{|S|}$$

Independence



Independence Definition 1:

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

$$\frac{|AB|}{|S|} = \frac{|A|}{|S|} \times \frac{|B|}{|S|}$$

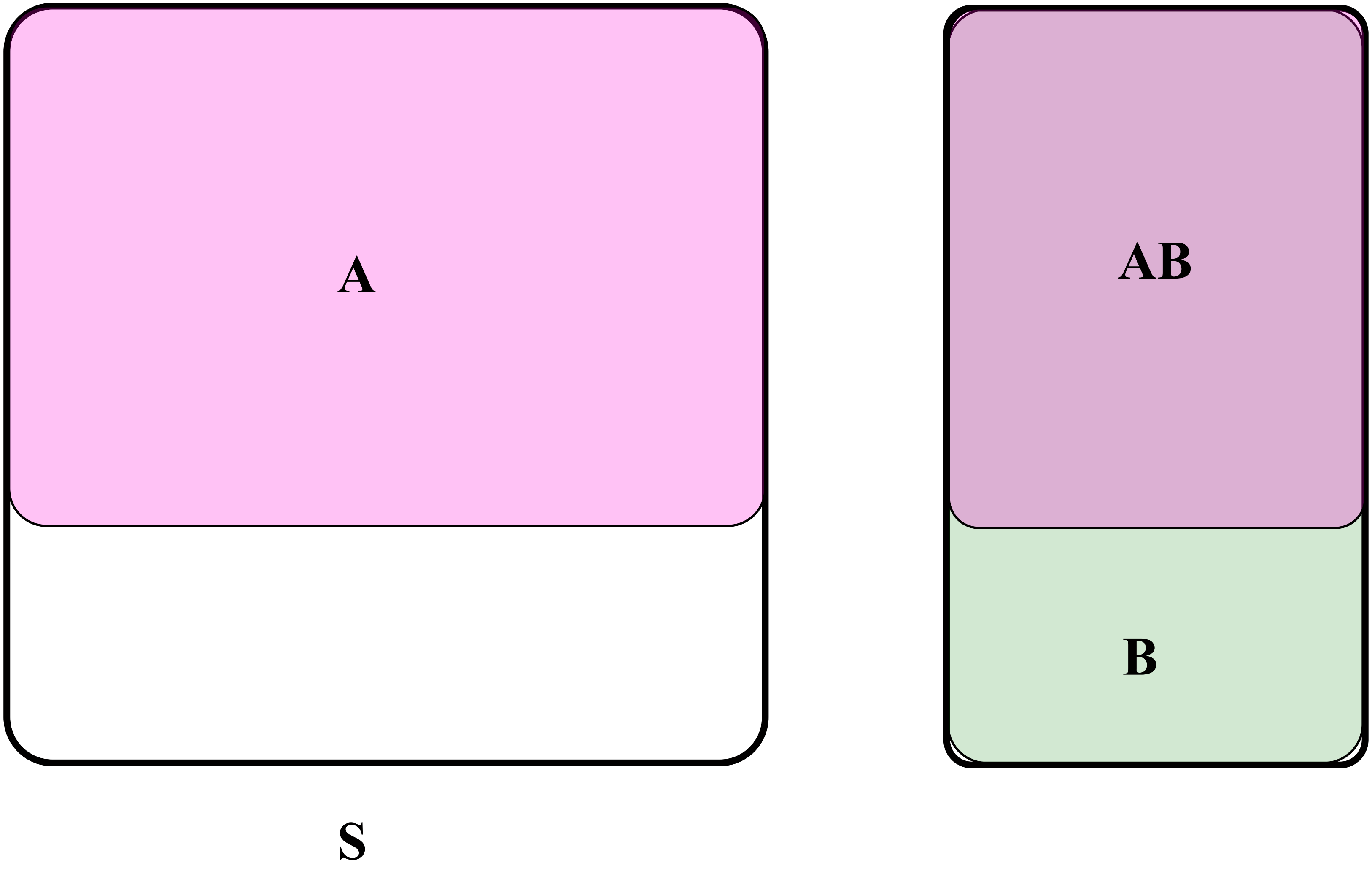
Independence Definition 2:

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

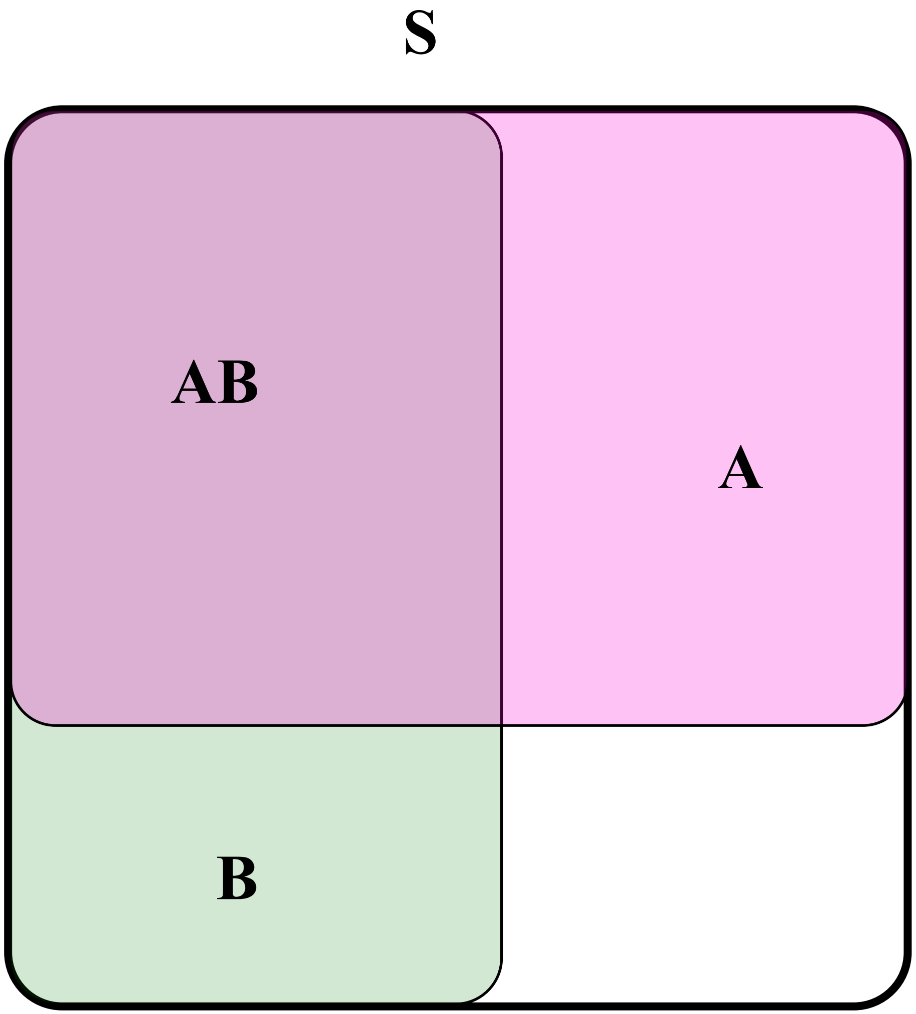
$$\frac{|AB|}{|B|} = \frac{|A|}{|S|}$$

Independence

This ratio, P(A)… … is the same as this one, P(A|B)



Independence



Independence Definition 1:

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

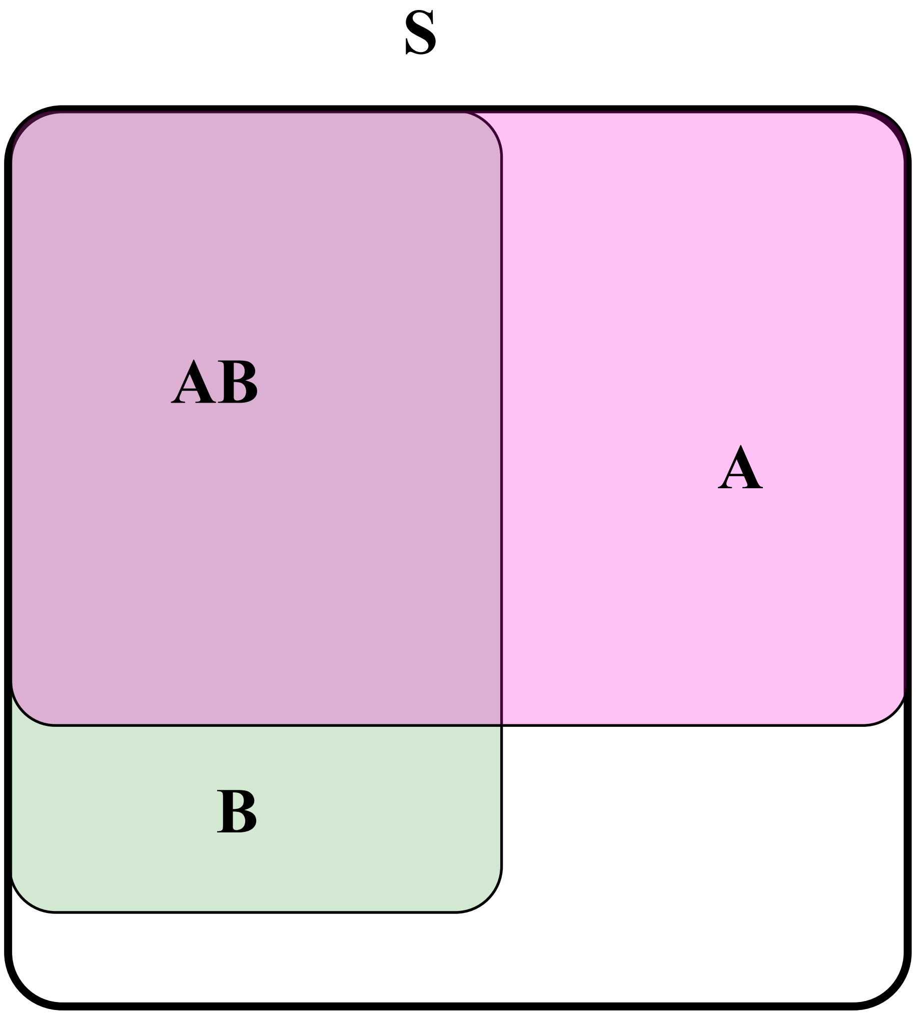
$$\frac{|AB|}{|S|} = \frac{|A|}{|S|} \times \frac{|B|}{|S|}$$

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:

Independence

Given independent events A and B, prove that A and B^C are independent

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

$$
\begin{aligned}
P(AB^C) &= P(A) - P(AB) && \text{By Total Law of Prob.} \\
&= P(A) - P(A)P(B) && \text{By independence} \\
&= P(A)[1 - P(B)] && \text{Factoring} \\
&= P(A)P(B^C) && \text{Since } P(B) + P(B^C) = 1
\end{aligned}
$$

So if A and B are independent A and B^C are also independent

Generalization

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:

$$P(E_{1'}E_{2'}E_{3'}\ldots E_{r'}) = P(E_{1'})P(E_{2'})P(E_{3'})\ldots P(E_{r'})$$

- 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

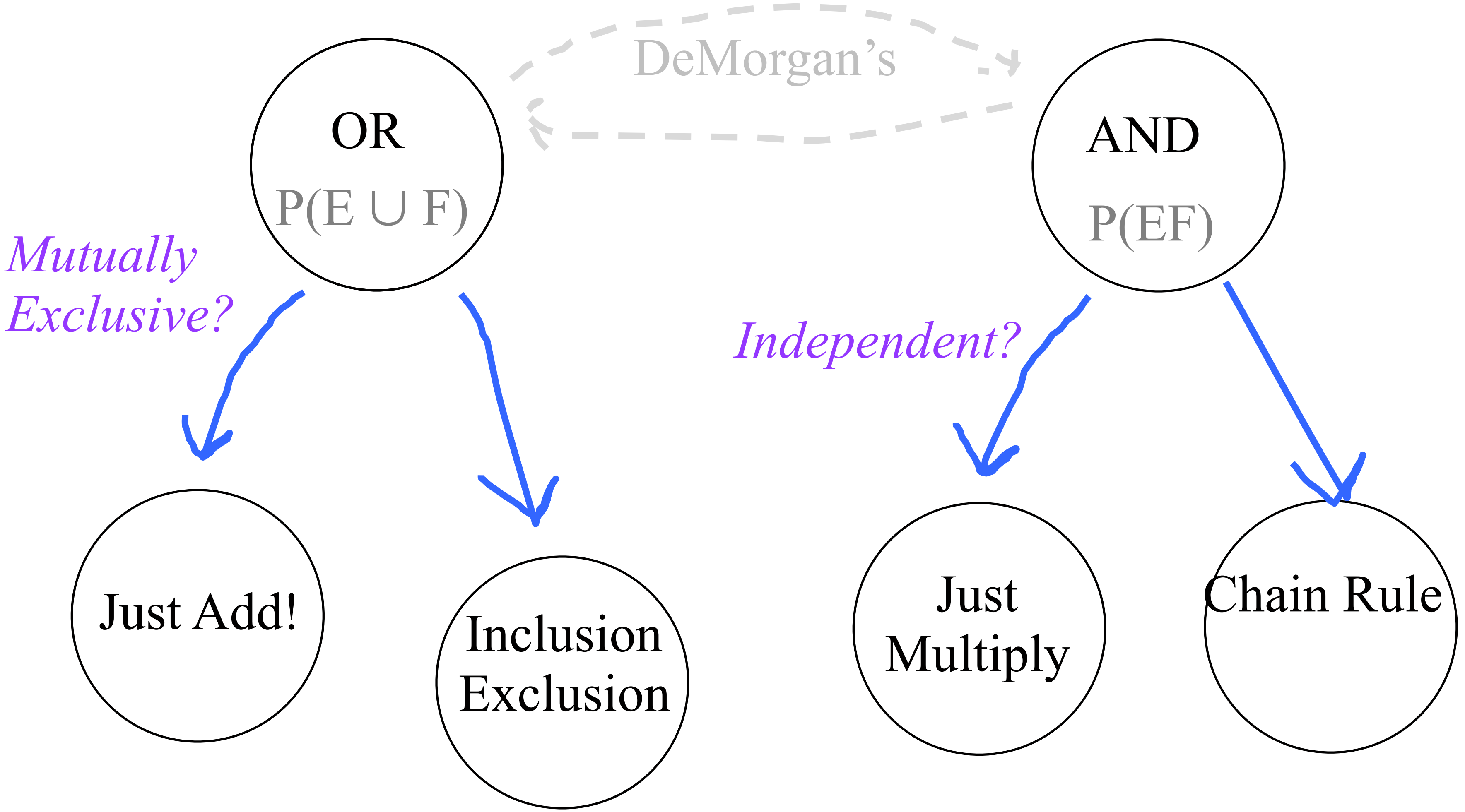
Math > Intuition

Two Dice

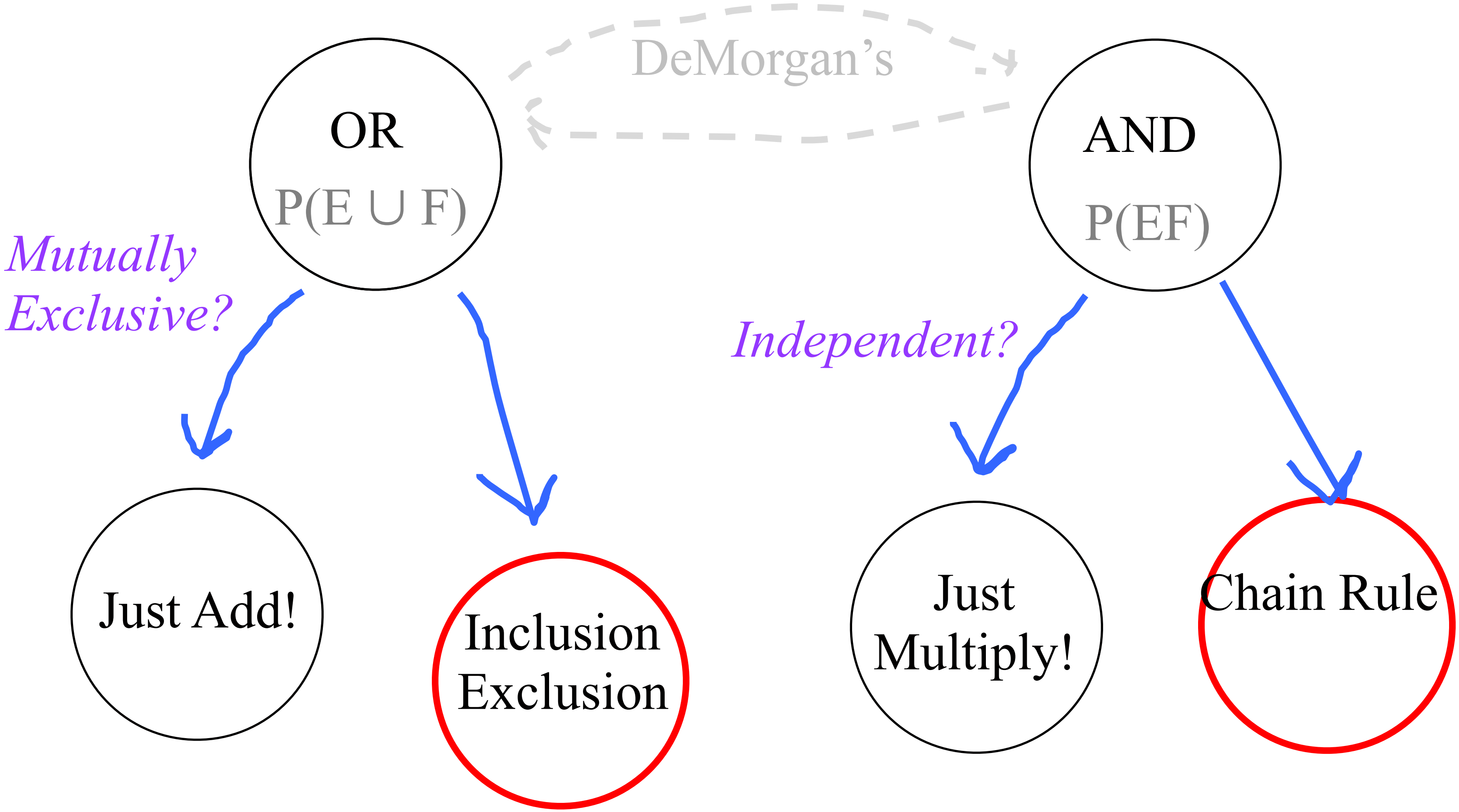
- Roll two 6-sided dice, yielding values D_1 and D_2
  - Let E be event: $D_1 = 1$
  - Let F be event: $D_2 = 6$
  - Are E and F independent? Yes!
- Let G be event: $D_1 + D_2 = 7$
  - Are E and G independent? Yes!
  - $P(E) = 1/6$, $P(G) = 1/6$, $P(E\ G) = 1/36$ [roll (1, 6)]
  - Are F and G independent? Yes!
  - $P(F) = 1/6$, $P(G) = 1/6$, $P(F\ G) = 1/36$ [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


DeMorgan's
OR
P(E ∪ F)
AND
P(EF)
Mutually Exclusive?
Independent?
Just Add!
Inclusion Exclusion
Just Multiply
Chain Rule

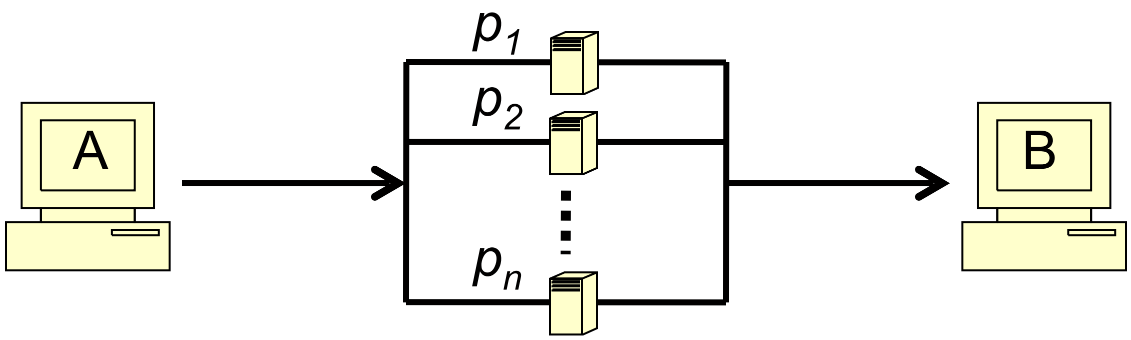
Today


DeMorgan's
OR
P(E ∪ F)
AND
P(EF)
Mutually Exclusive?
Independent?
Just Add!
Inclusion Exclusion
Just Multiply!
Chain Rule

Use the two properties (mutual exclusion and independence)

Sending a Message Through Network

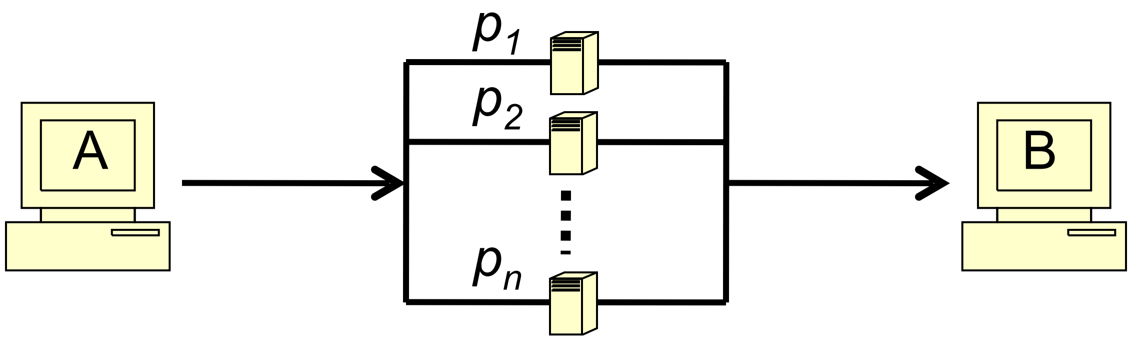
- Consider the following parallel network:



  - 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)?

Sending a Message Through Network

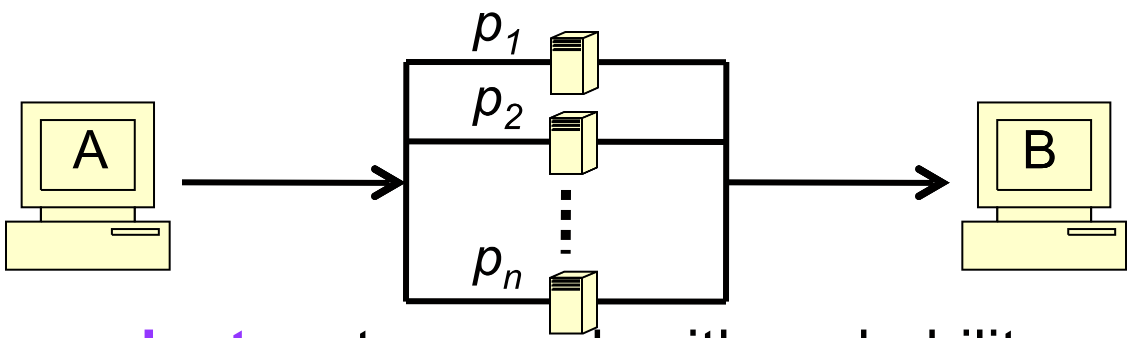
- Consider the following parallel network:



  - 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)?

Sending a Message Through Network

- 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)?
- Solution:
  - P(E) = 1 – P(all routers fail)

    $= 1 - (1 - p_1)(1 - p_2)\ldots(1 - p_n)$

$$1 - \prod_{i=1}^{n} (1 - p_i)$$

Coin Flips

- Say a coin comes up heads with probability p
  - Each coin flip is an **independent** trial

- P(*n* heads on *n* coin flips) = p^n
- P(*n* tails on *n* coin flips) = $(1 - p)^n$

- P(first *k* heads, then *n* – *k* tails) = $p^k (1 - p)^{n-k}$

- Consider a particular ordering (HTHTH). What is the probability of that *exact* ordering?

$$= p^3 \cdot (1 - p)^2$$

Explain…

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…. The coin flips are independent!

Ordering 2: H, T, H, T, T, T….

And so on… $P(F_i) = p^k (1-p)^{n-k}$

Let's make each ordering with *k* heads an event… F_i

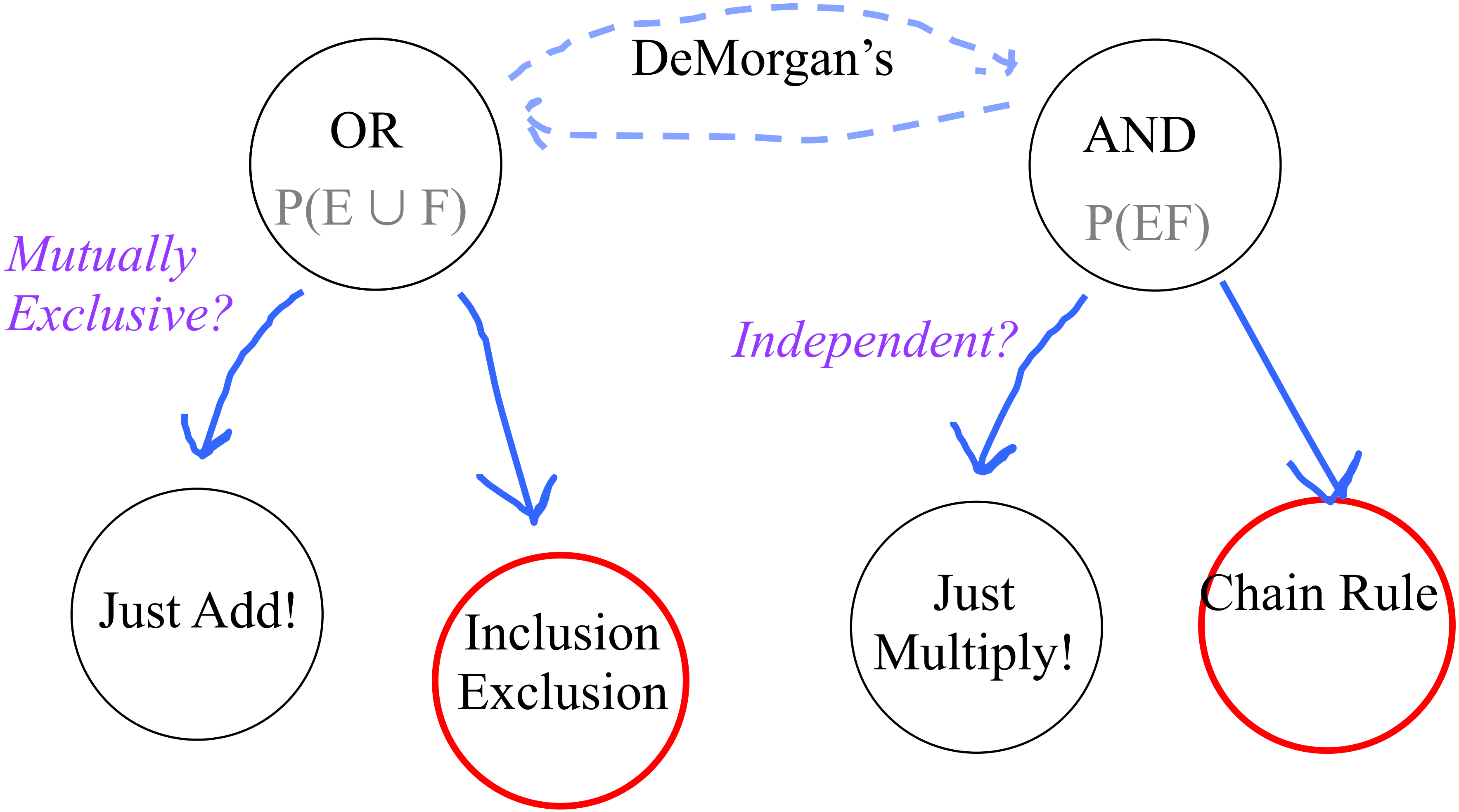
P(exactly *k* heads on *n* coin flips) = P(any one of the events)

P(exactly *k* heads on *n* coin flips) = P(F_1 or F_2 or F_3…)

Those events are mutually exclusive!

Stretch Break!

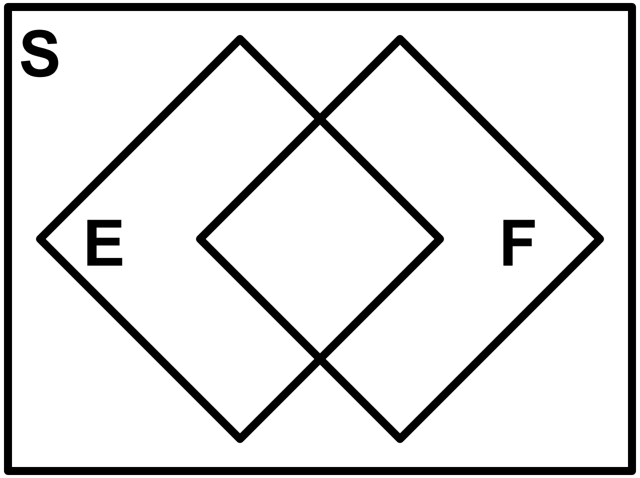
Today


DeMorgan's
OR
P(E ∪ F)
AND
P(EF)
Mutually
Exclusive?
Independent?
Just Add!
Inclusion
Exclusion
Just
Multiply!
Chain Rule

Sets Review

Set Operations Review

- Say E and F are subsets of S

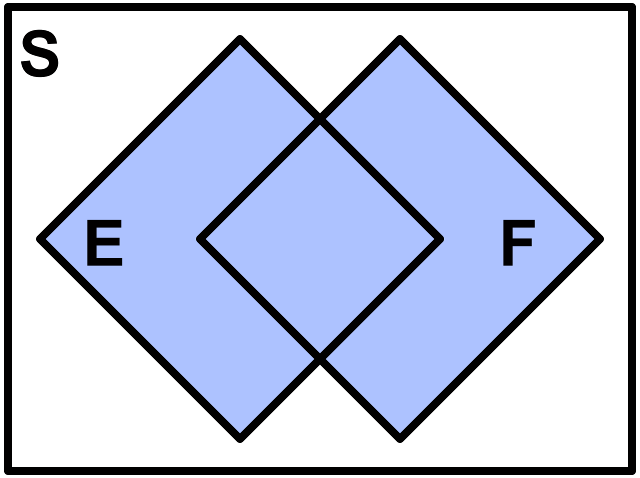


Set Operations Review

- Say E and F are events in S

Event that is in E <u>or</u> F

E ∪ F



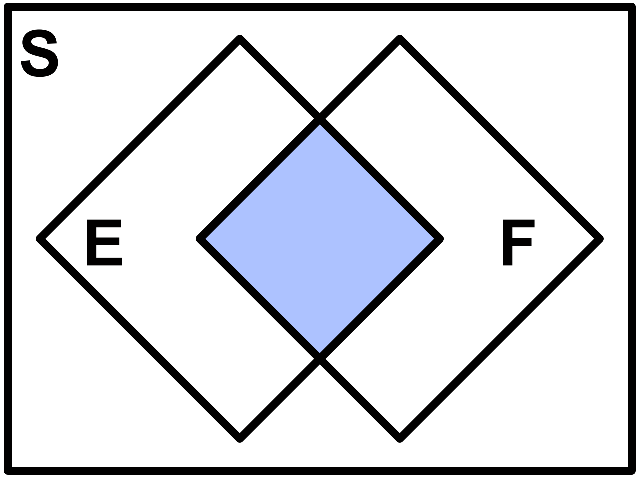
- S = {1, 2, 3, 4, 5, 6} die roll outcome
- E = {1, 2} F = {2, 3} E ∪ F = {1, 2, 3}

Set Operations Review

- Say E and F are events in S

Event that is in E <u>and</u> F

E ∩ F *or* **EF**

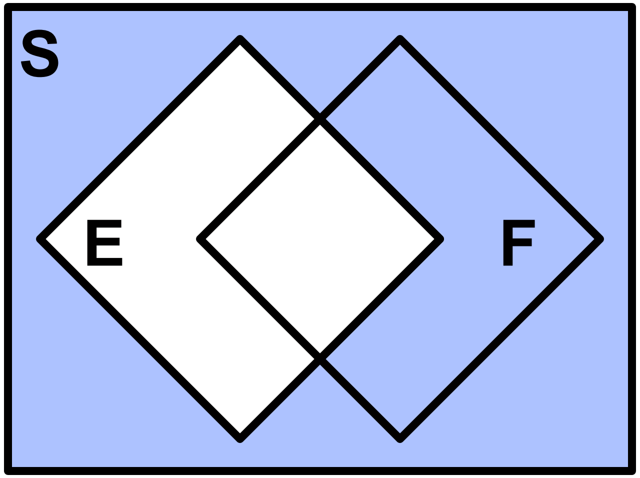


Set Operations Review

- Say E and F are events in S

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

E^c *or* **~E**

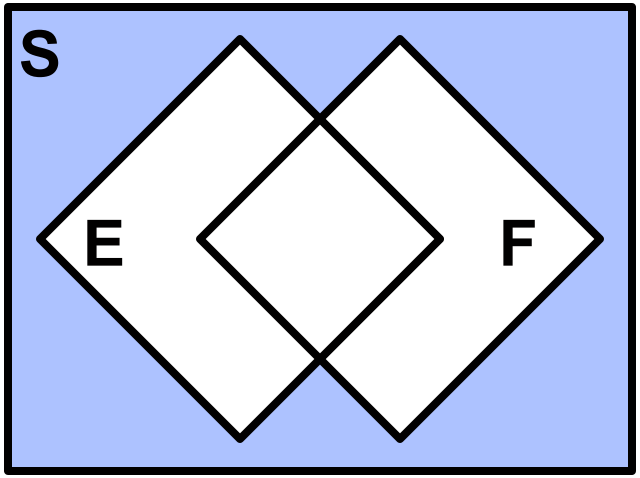


Set Operations Review

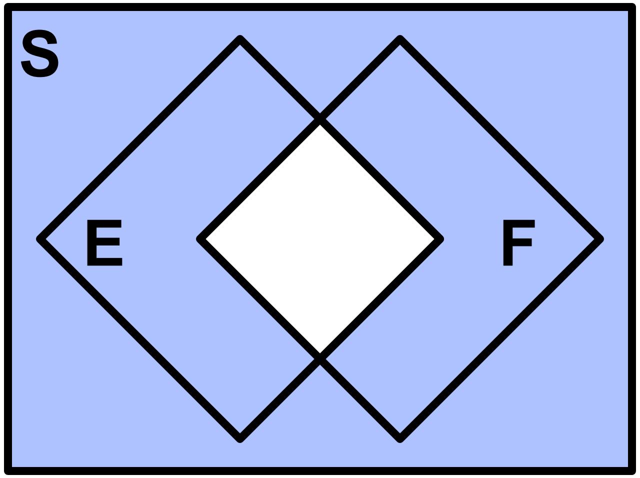
- Say E and F are events in S

DeMorgan's Laws

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



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



Augustus Demorgan

Jason Alexander

- British Mathematician who wrote the book “Formal Logic”in 1847
- Celebrity lookalike is Jason Alexander from Seinfeld.

Hash Tables

- *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 = at least one string hashed to first bucket
  - What is P(E)?
- Solution

To the white board

Hash Tables

- *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 = **at least one** string hashed to first bucket
  - What is P(E)?
- Solution
  - P that string 1 **or** 2 **or** .. n goes to bucket 1?
  - P that string 1 **and** 2 **and** … n goes to bucket not 1?
  - $P(E) = 1 - (1 - p_1)^m$

Yet More Hash Tables

- *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 = At least 1 of buckets 1 to *k* has ≥ 1 string hashed to it
- Solution
  - F_i = at least one string hashed into *i*-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(no strings hashed to buckets 1 to *k*)

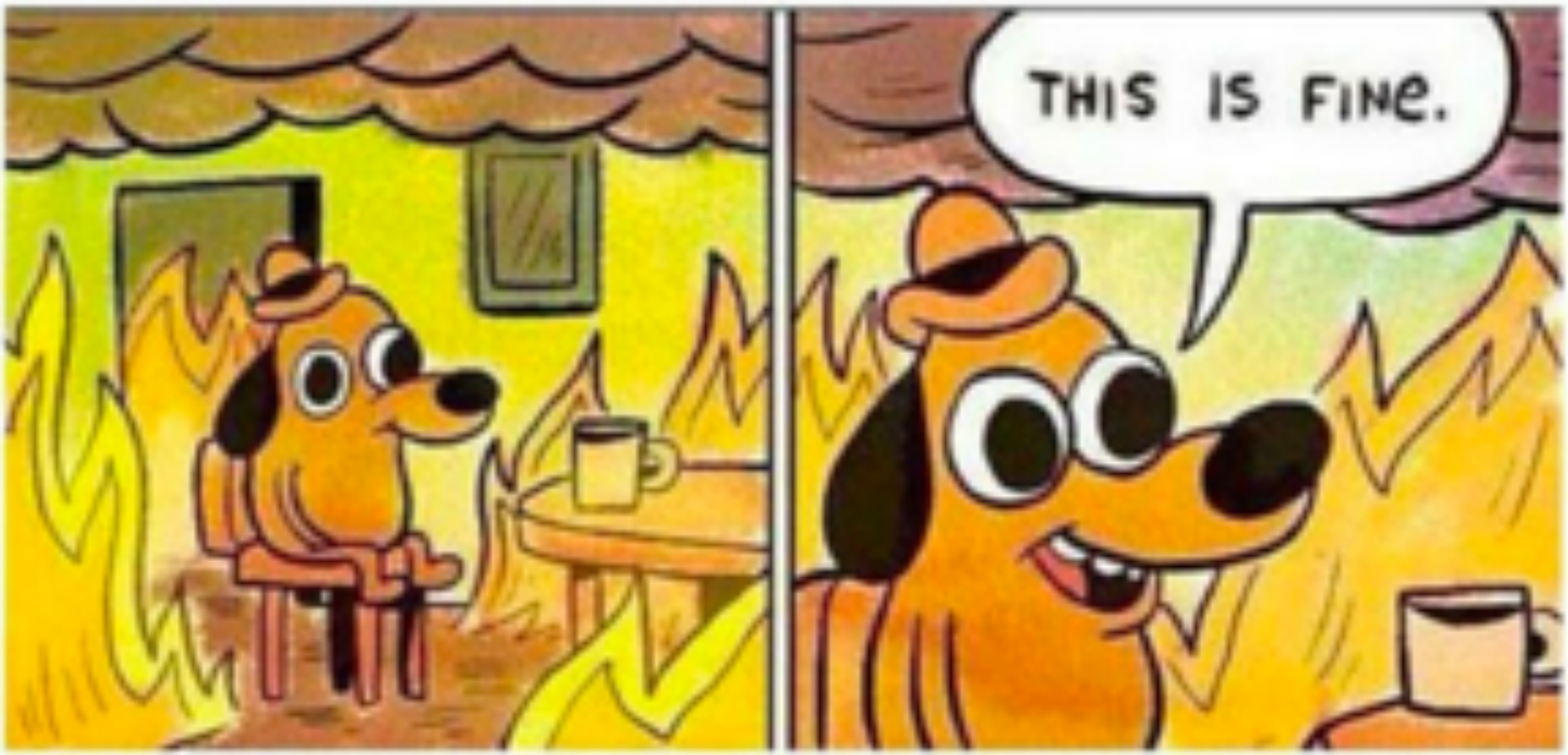
    $= (1 - p_1 - p_2 - \ldots - p_k)^m$
  - $P(E) = 1 - (1 - p_1 - p_2 - \ldots - p_k)^m$

No, Really, More Hash Tables

$=$

- 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 = Each of buckets 1 to k has ≥ 1 string hashed to it

No, Really, More Hash Tables



It is expected that this last example will take some review!

No, Really, More Hash Tables

- 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 = Each of buckets 1 to k has ≥ 1 string hashed to it
- Solution
  - F_i = at least one string hashed into i-th bucket
  - $P(E) = P(F_1 F_2 \ldots F_k) = 1 - P((F_1 F_2 \ldots F_k)^c)$

$$= 1 - P(F_1^c \cup F_2^c \cup \ldots \cup F_k^c) \quad \text{(DeMorgan's Law)}$$

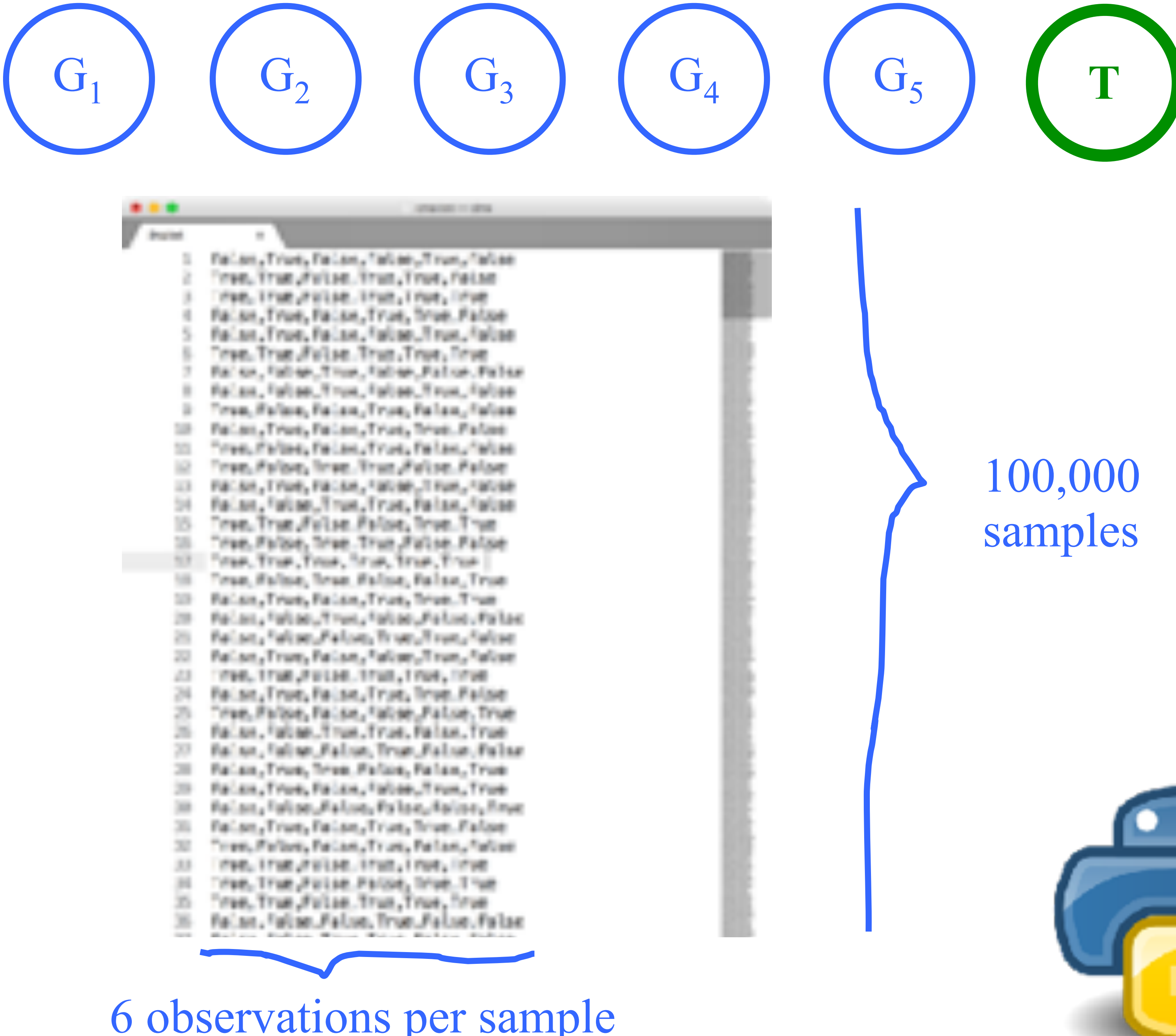
$$= 1 - \sum_{r=1}^{k} (-1)^{(r+1)} \sum_{i_1 < \ldots < i_r} P(F_{i_1}^{\ c} F_{i_2}^{\ c} \ldots F_{i_r}^{\ c})$$

where $P(F_{i_1}^{\ c} F_{i_2}^{\ c} \ldots F_{i_r}^{\ c}) = (1 - p_{i_1} - p_{i_2} - \ldots - p_{i_r})^m$

Here we are

Source: The Ho


G_1
G_2
G_3
G_4
G_5
T


G1
G2
G3
G4
G5
T
100,000
samples
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: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: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: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
```

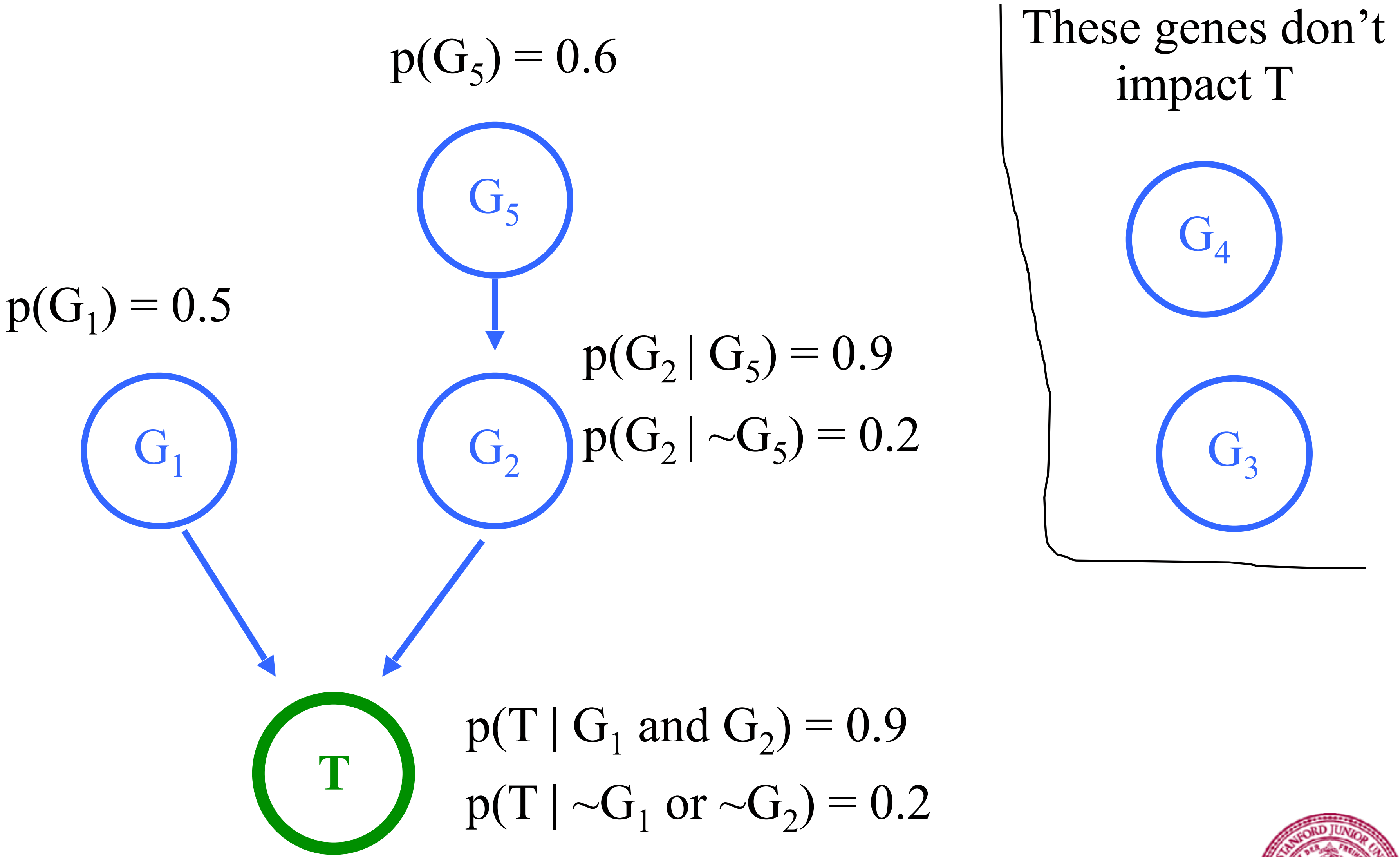
Discovered Pattern

```
Piech-2: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
```

Only Causal Structure that Fits



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

Conditional 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:

$$P(E|FG) = P(E|G)$$

NETFLIX

And Learn

Netflix and Learn

What is the probability that a user will watch Life is Beautiful?

$P(E)$

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

$P(E) = 10{,}234{,}231 \,/\, 50{,}923{,}123 = 0.20$

Netflix and Learn

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?

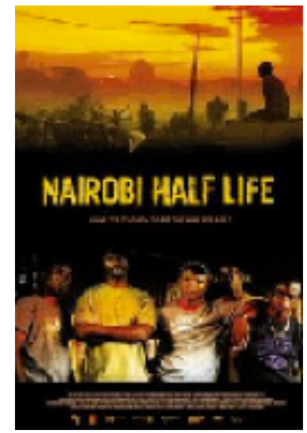
Netflix and Learn

Each event corresponds to liking a particular movie

E_1

E_2

E_3



E_4

$$P(E_4 | E_1, E_2, E_3)?$$

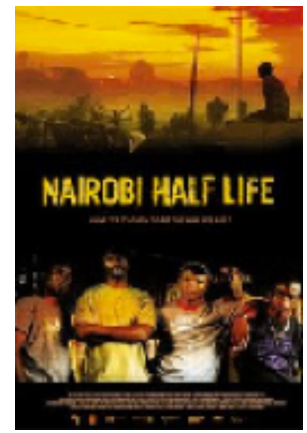
Netflix and Learn

Is E_4 independent of E_1, E_2, E_3?

E_1

E_2

E_3



E_4

$$P(E_4 | E_1, E_2, E_3) \stackrel{?}{=} P(E_4)$$

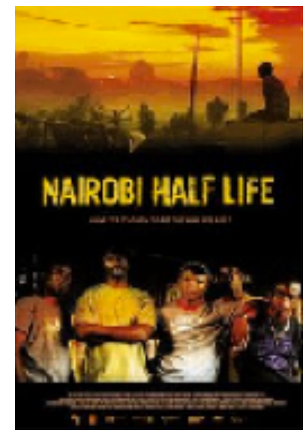
Netflix and Learn

Is E_4 independent of E_1, E_2, E_3?

E_1

E_2

E_3



E_4

$$P(E_4 | E_1, E_2, E_3) = \frac{P(E_1 E_2 E_3 E_4)}{P(E_1 E_2 E_3)}$$

Netflix and Learn

- 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:

Watch those four

Choose 26 movies not in the set

$$P(E) = \frac{\binom{4}{4}\binom{12996}{24}}{\binom{13000}{30}} = 10^{-11}$$

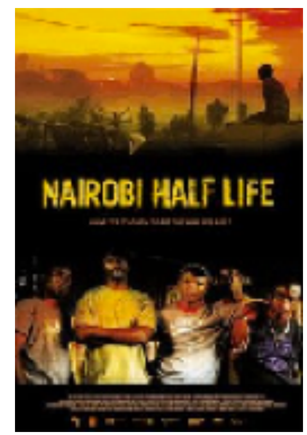
Choose 30 movies from netflix

Netflix and Learn

E_1

E_2

E_3

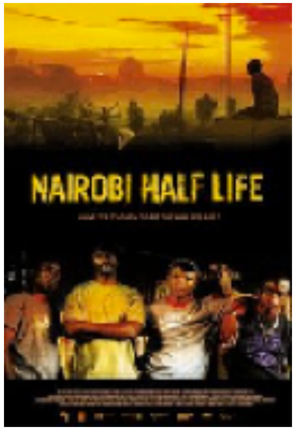


E_4

Netflix and Learn

K_1

Like foreign emotional comedies



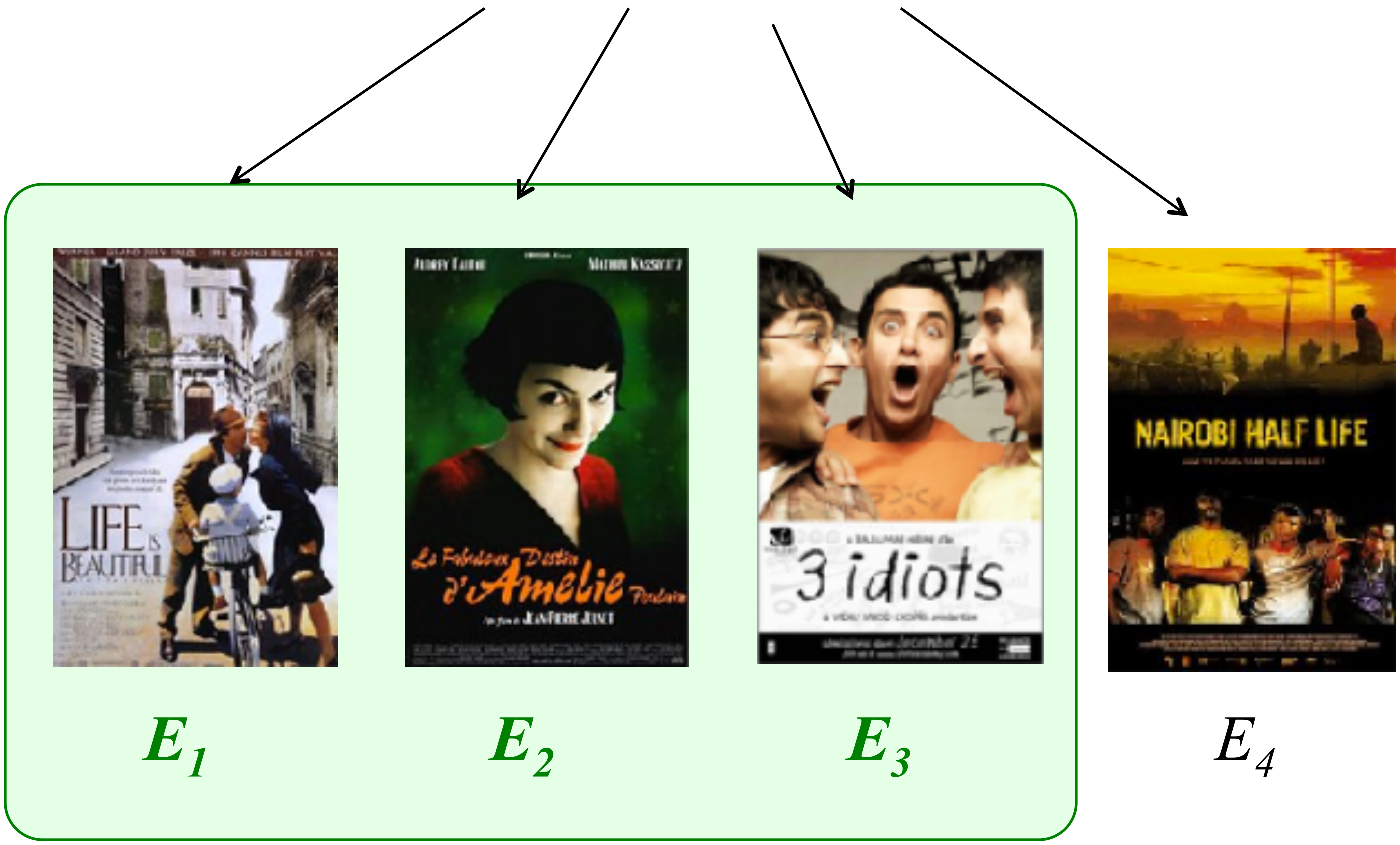
E_1 E_2 E_3 E_4

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

Netflix and Learn

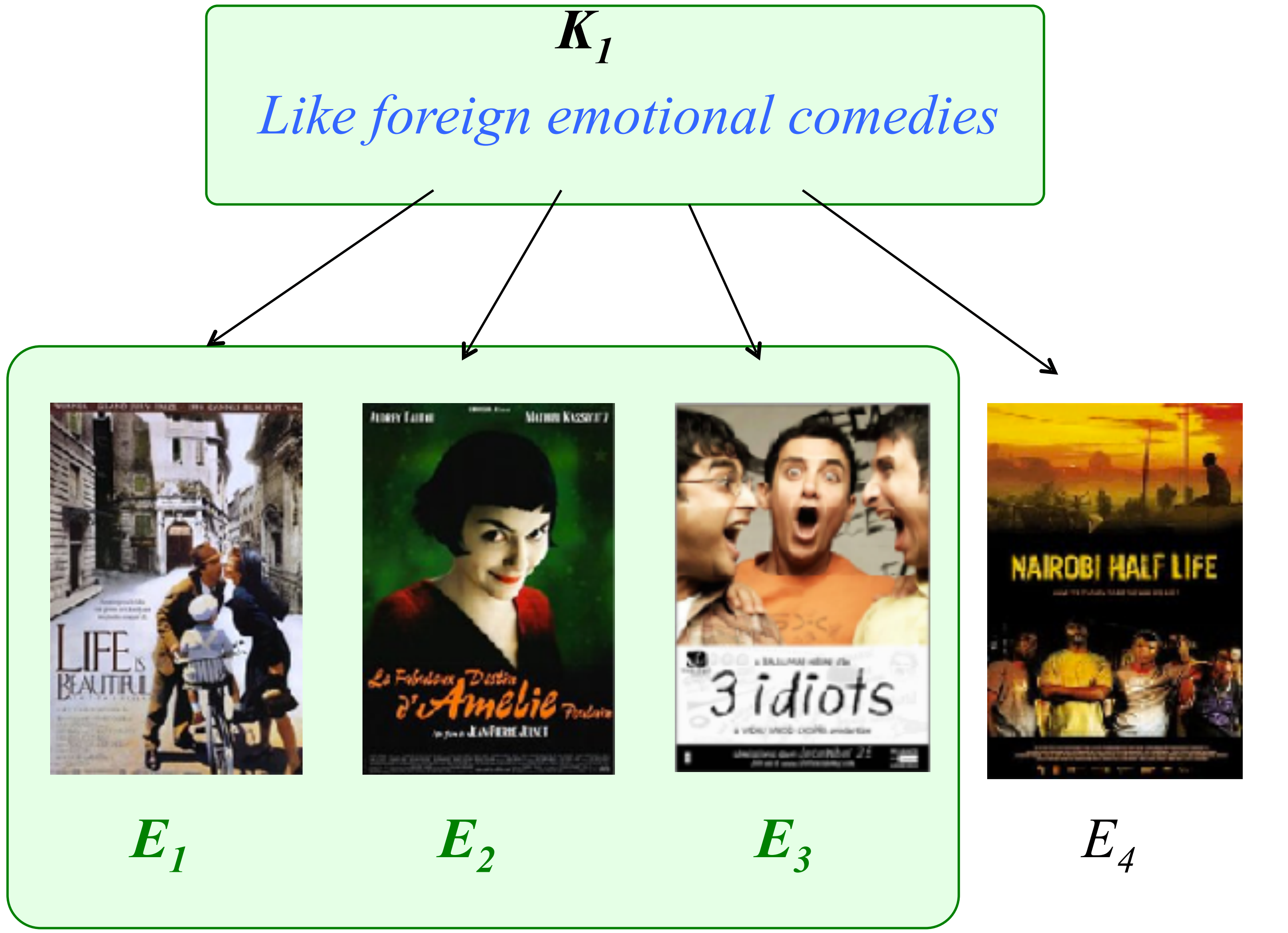
K_1

Like foreign emotional comedies



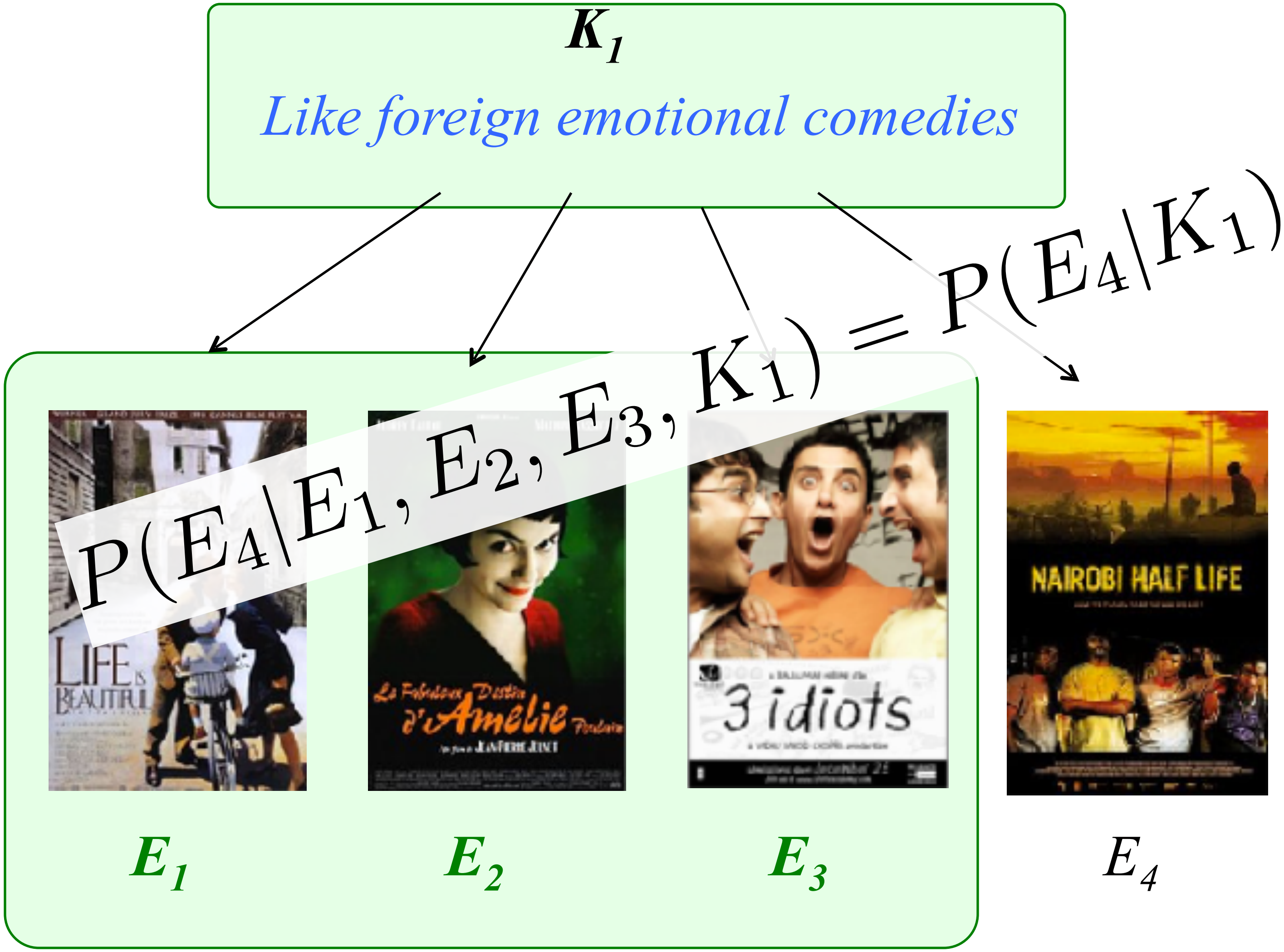
Assume E_1, E_2, E_3 and E_4 are conditionally independent given K_1

Netflix and Learn



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

Netflix and Learn



$$P(E_4 | E_1, E_2, E_3, K_1) = P(E_4 | K_1)$$

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.

Conditional Independence

If E and F are dependent,

that does not mean E and F will be dependent when another event happens.

Conditional Dependence

If E and F are independent,

that does not mean E and F will be independent when another event happens.

Big Deal

"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"*