Independence

Learning Goals

1. Be able to recognize independent events
2. Use independence rules to calculate probabilities
3. Recognize and use *conditional* independencies

Summary

Two events A and B are called **independent** if:

$$P(AB) = P(A)P(B) \qquad P(A|B) = P(A)$$

Otherwise, they are called **dependent** events

Two events A and B are

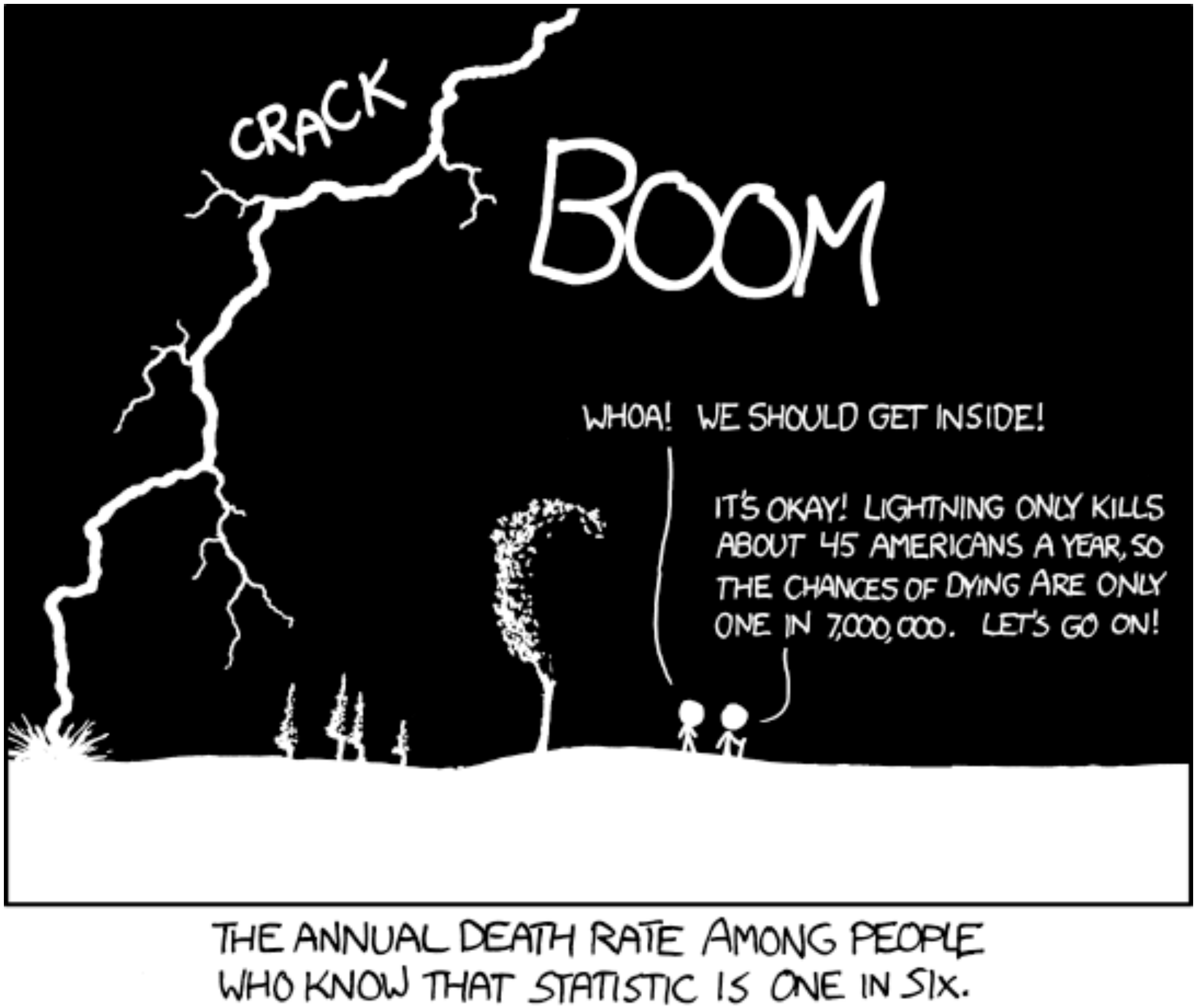
conditionally independent on C if:

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

$$P(A|BC) = P(A|C)$$

Review

The Tragedy of Conditional Prob



Thanks xkcd! `http://xkcd.com/795/`

And vs Condition

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

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

A Few Useful Formulas

- For any events A and B:

$P(A\ B) = P(B\ A)$ (Commutativity)

$P(A\ B) = P(A \mid B)\, P(B)$ (Chain rule)

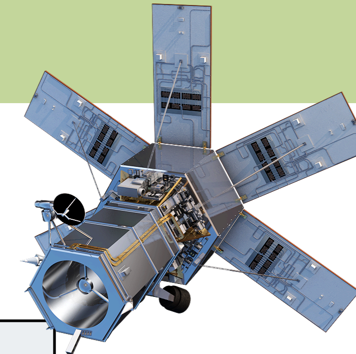
$= P(B \mid A)\, P(A)$

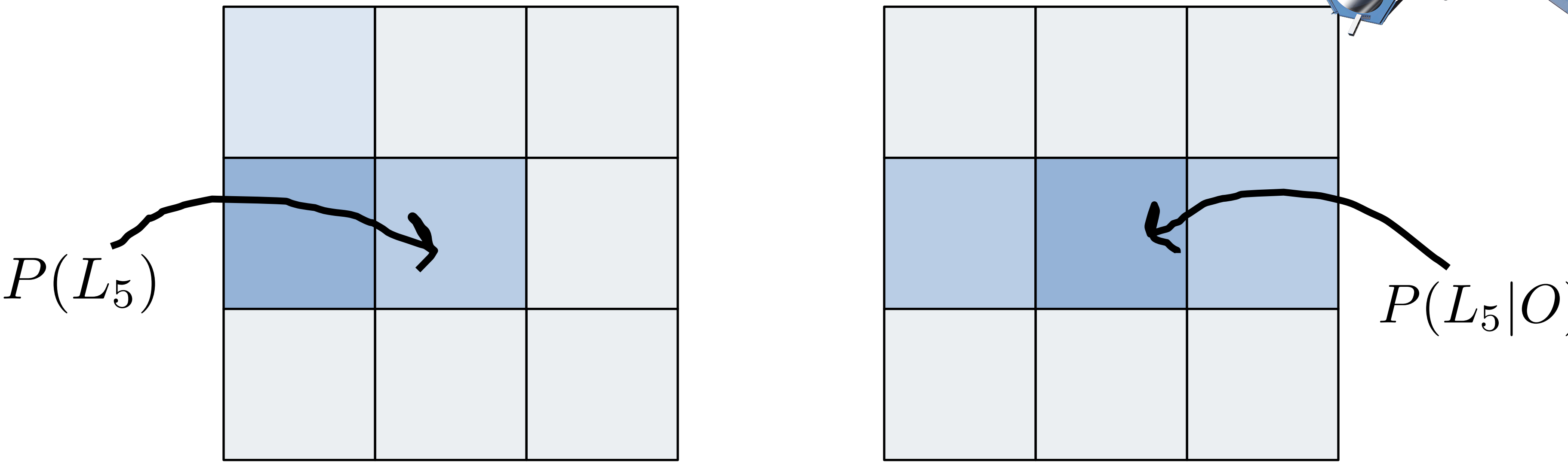
$P(A\ B^c) = P(A) - P(AB)$ (Intersection)

$P(A) + P(A^c) = 1$ (Total Probability)

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}$$ (Bayes Theorem)

Bayes: Update Belief





Before Observation

After Observation

$$P(L_5|O) = \frac{P(O|L_5)P(L_5)}{P(O)}$$

Generality of Conditional Probability

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

$$P(A\,B \mid E) = P(B\,A \mid E)$$

$$P(A\,B \mid E) = P(A \mid B\,E)\,P(B \mid E)$$

$$P(A \mid B\,E) = \frac{P(B \mid A\,E)\,P(A \mid E)}{P(B \mid E)} \qquad \text{(Bayes' Thm.)}$$

- Can think of E as "everything you already know"
- Formally, P(• | E) satisfies 3 axioms of probability

BAE's Theorem?

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

End Review

Today, start with a cool program


G_1
G_2
G_3
G_4
G_5
T

G_1 G_2 G_3 G_4 G_5 T

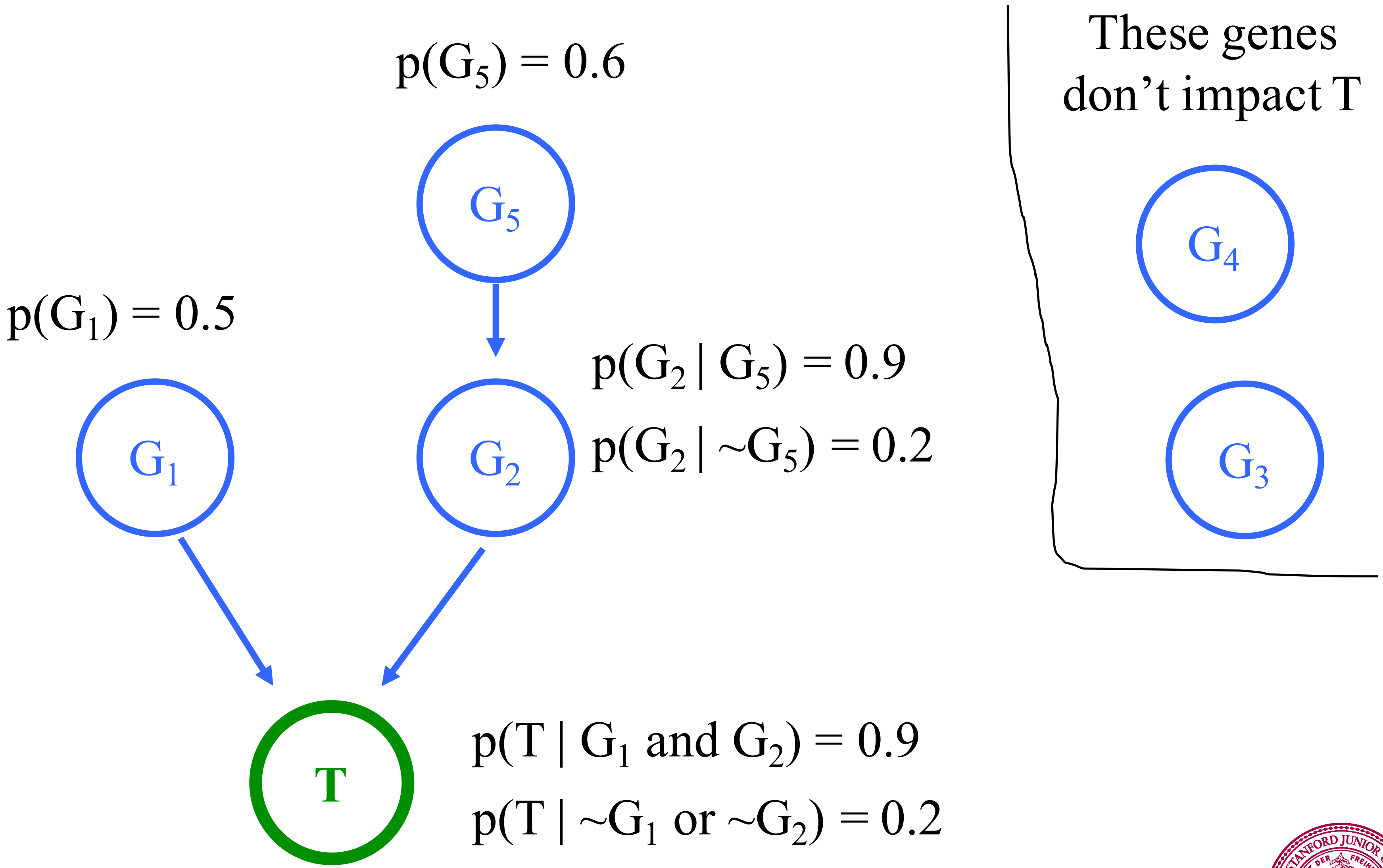
dna.txt — dna

```
False,True,False,False,True,False
True,True,False,True,True,False
True,True,False,True,True,True
False,True,False,True,True,False
False,True,False,False,True,False
True,True,False,True,True,True
False,False,True,False,False,False
False,False,True,False,True,False
True,False,False,True,False,False
False,True,False,True,True,False
True,False,False,True,False,False
True,False,True,True,False,False
False,True,False,False,True,False
False,False,True,True,False,False
True,True,False,False,True,True
True,False,True,True,False,False
True,True,True,True,True,True
True,False,True,False,False,True
False,True,False,True,True,True
False,False,True,False,False,False
False,False,False,True,True,False
False,True,False,False,True,False
True,True,False,True,True,True
False,True,False,True,True,False
True,False,False,False,False,True
False,False,True,True,False,True
False,False,False,True,False,False
False,True,True,False,False,True
False,True,False,False,True,True
False,False,False,False,False,True
False,True,False,True,True,False
True,False,False,True,False,False
True,True,False,True,True,True
True,True,False,False,True,True
True,True,False,True,True,True
False,False,False,True,False,False
```

100,000 samples

6 observations per sample

Discovered Pattern



We’ve gotten ahead of ourselves

Source: The Ho

Start at the beggining

Source: The Ho

Independence

Two events A and B are called **independent** if:

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

Or, equivalently:

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

Otherwise, they are called **dependent** events

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 *<u>independent</u>*

- 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) ≠ P(E) P(G) → E and G *<u>dependent</u>*

Intuition through proofs:

Independence with 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.

Independence

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

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

$$
\begin{aligned}
P(AB^C) &= P(A) - P(AB) && \text{By Intersection Rule} \\
&= P(A) - P(A)P(B) && \text{By independence} \\
&= P(A)[1 - P(B)] && \text{Factoring} \\
&= P(A)P(B^C) && \text{Since } P(B) + P(B^C) = 1
\end{aligned}
$$

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

Independence

Let A and B be independent

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

From our first proof

A and B^C are independent

From our second proof

And thus:

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

Since A and BC are independent

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

Put it all together

Intuitively, if A and B are independent, knowing whether B holds gives us no information about A

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* ≤ *n*) it holds that:

$$P(E_{s_1}, E_{s_2}, E_{s_3}, \ldots E_{s_r})$$
$$= P(E_{s_1})P(E_{s_2})P(E_{s_3})\ldots P(E_{s_r})$$

- <u>Example</u>: 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

Generating Random Bits

- A computer produces a series of random bits, with probability *p* of producing a 1.
  - Each bit generated is an independent trial
  - E = first *n* bits are 1's, followed by a single 0
  - What is P(E)?
- Solution
  - P(first *n* 1's) = P(1st bit=1) P(2nd bit=1) ... P(nth bit=1)
    $= p^n$
  - P(*n+1* bit=0) = (1 – p)
  - P(E) = P(first *n* 1's) P(*n+1* bit=0) = p^n (1 – p)

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) = $p^k(1-p)^{n-k}$

- P(exactly *k* heads on *n* coin flips) =?

Important Result

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…

The coin flips are independent!

$$P(F_i) = p^k (1-p)^{n-k}$$

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!

Moment of Crystallization

Add vs Multiply?


Batman vs Superman
COMING SOON
#BATMAN v SUPERMAN
SEE IT IN 3D

Add vs Multiply

+ vs ×

Add vs Multiply

Probabilities

multiply

P(AB)	Generally:	P(A)P(B|A)
	Independent:	P(A)P(B)
P(A∪B)	Generally:	P(A) + P(B) – P(AB)
	Mutually Exclusive:	P(A) + P(B)

add

Add vs Multiply

Counting

multiply

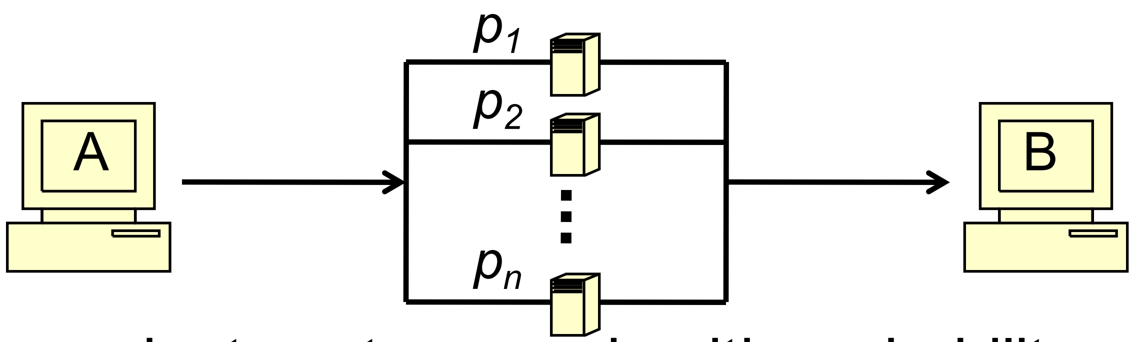
A then B	Generally:	*Not handled*
	Independent Counts:	$\lvert A \rvert \lvert B \rvert$
$A \cup B$	Generally:	$\lvert A \rvert + \lvert B \rvert - \lvert AB \rvert$
	Mutually Exclusive:	$\lvert A \rvert + \lvert B \rvert$

add

Combining with Previous Skills

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)?

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

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

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

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)
  - $P(F_1^c F_2^c \ldots F_k^c) = P(\text{no strings hashed to buckets 1 to } k)$
    $= (1 - p_1 - p_2 - \ldots - p_k)^m$
  - $P(E) = 1 - (1 - p_1 - p_2 - \ldots - p_k)^m$

The Hardest Example

- *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) $= P(F_1 F_2 \ldots F_k) = 1 - P((F_1 F_2 \ldots F_k)^c)$

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

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

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

Phew…

Conditional Independence

Recall, two events A and B are **independent** if:

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

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

Two events E and F are **conditionally independent** on C if:

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

$$P(A|BC) = P(A|C)$$

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?

$P(E|F)$

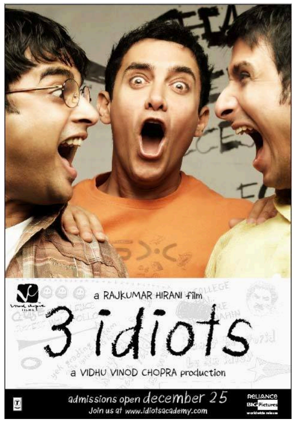
$$P(E|F) = \frac{P(EF)}{P(F)} = \frac{\#\text{people who watched both}}{\#\text{people who watched } F}$$

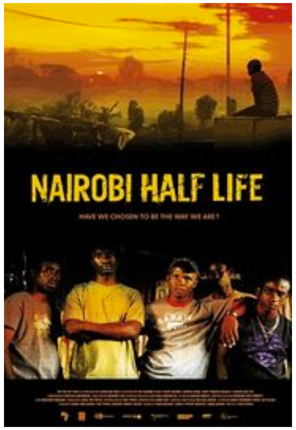
$P(E|F) = 0.42$

Conditioned on watching a set of movies?

Netflix and Learn

Each event corresponds to watching a particular movie





E_1 E_2 E_3 E_4

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

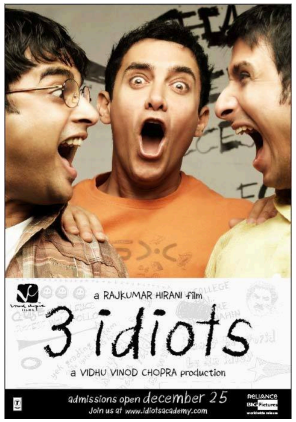
Is E_4 independent of E_1,E_2,E_3?

Netflix and Learn

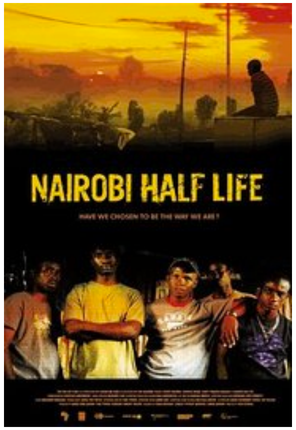
Is E_4 independent of E_1, E_2, E_3?

E_1

E_2



E_3



E_4

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

Netflix and Learn

- 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 24 movies not in the set

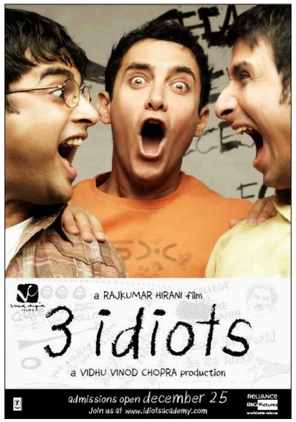
$$P(E) = \frac{\binom{4}{4}\binom{12996}{24}}{\binom{13000}{30}} = 10^{-11}$$

Choose 30 movies from netflix

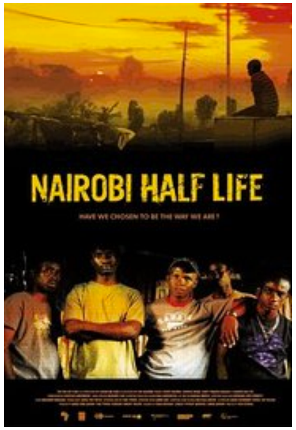
Netflix and Learn

E_1

E_2



E_3



E_4

Netflix and Learn

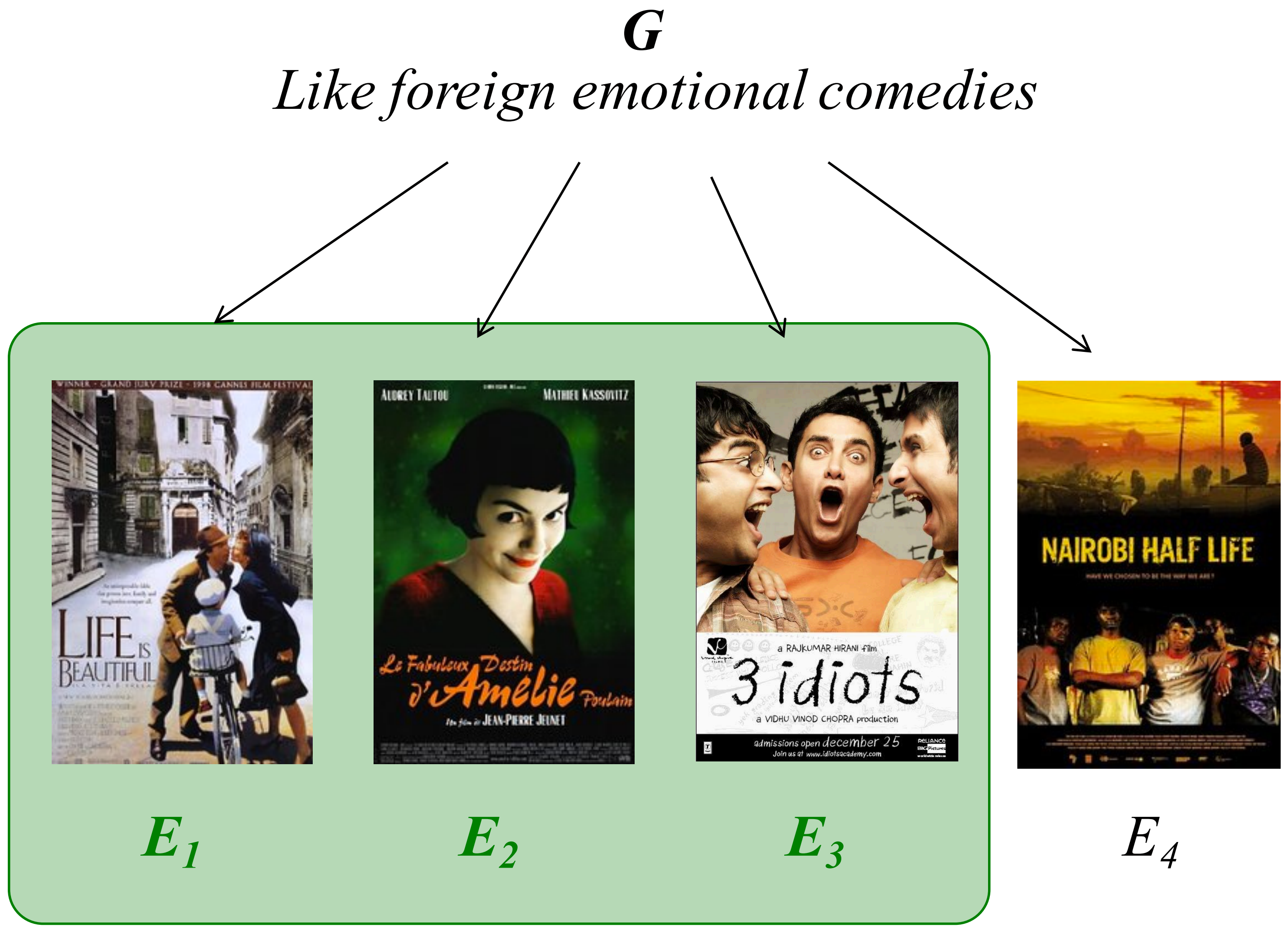
G

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 G

Netflix and Learn

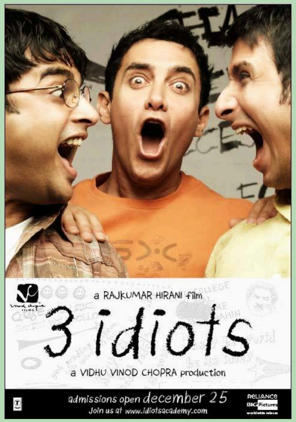


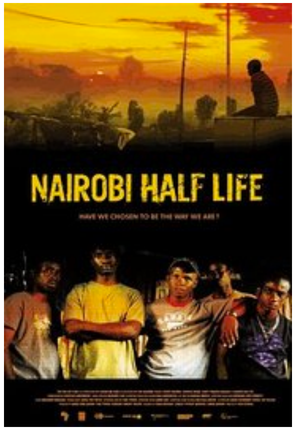
Assume E_1, E_2, E_3 and E_4 are conditionally independent given G

Netflix and Learn

G

Like foreign emotional comedies

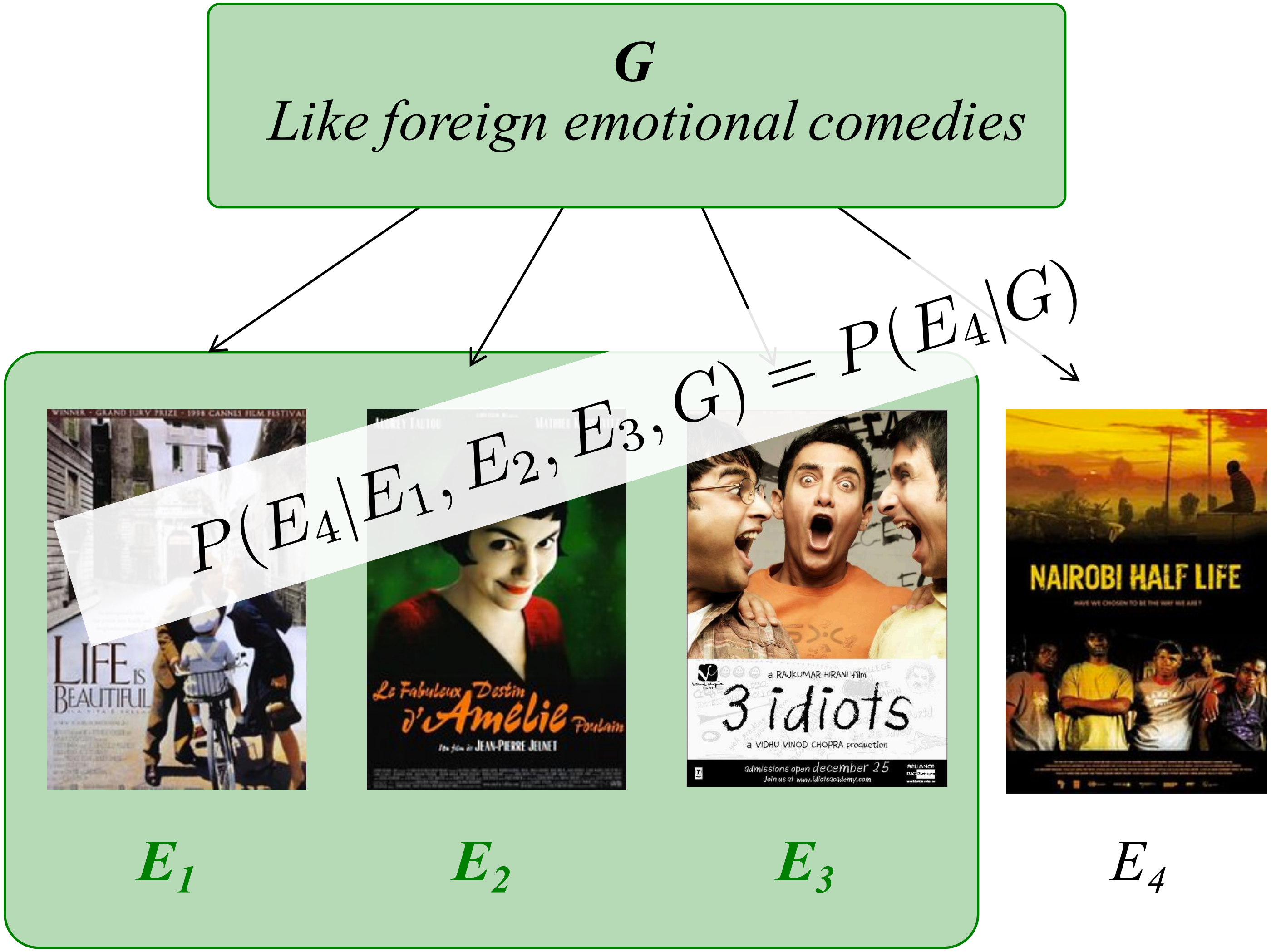




E_1 ***E_2*** ***E_3*** *E_4*

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

Netflix and Learn



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

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

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

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"*

When we introduced conditions

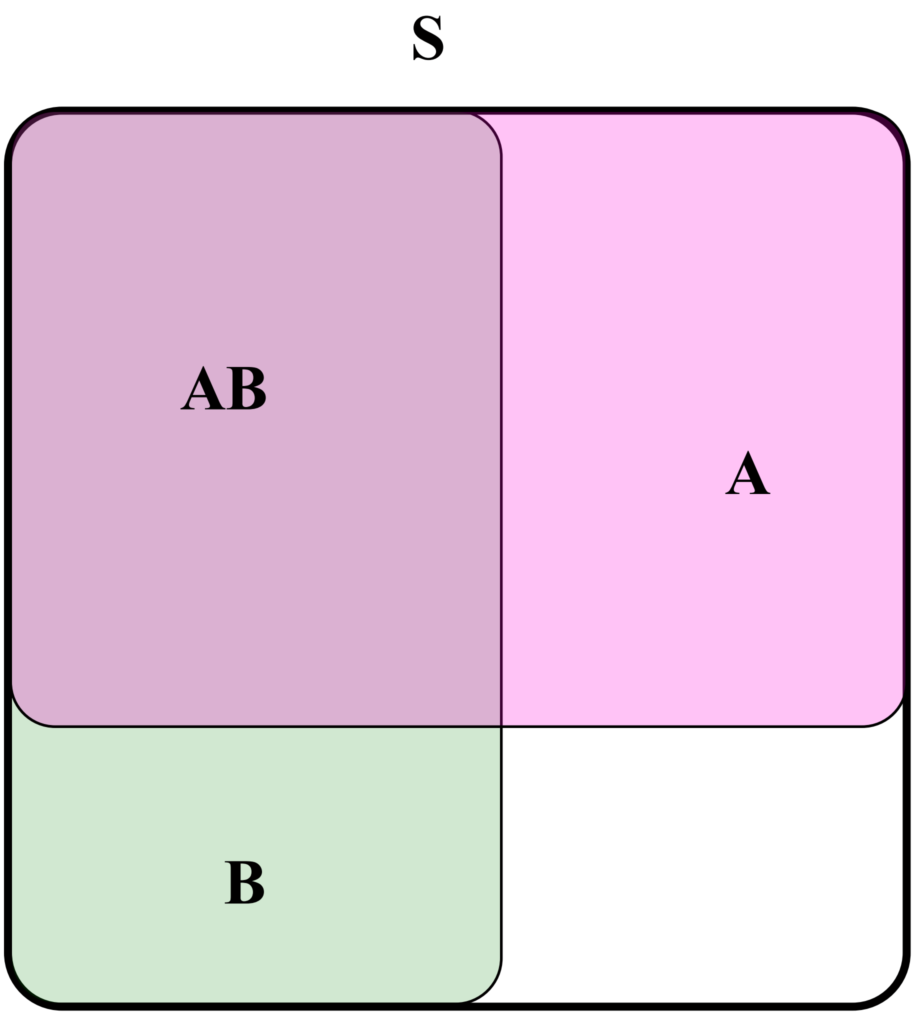
Identities of probability remain the same

But sometimes independence / dependence relationships change

What the fish?

What does independence look like?

Independence



Independence Definition 1:

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

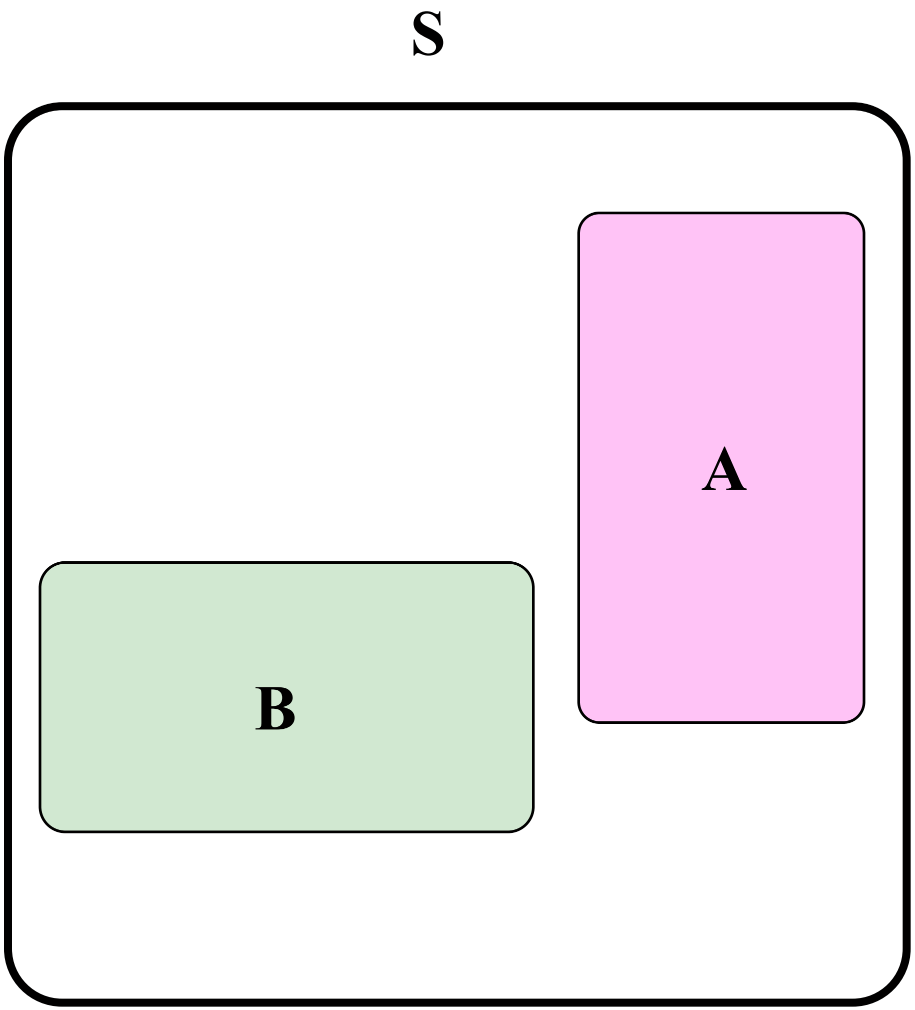
$$\frac{|AB|}{|S|} = \frac{|A|}{|S|} \times \frac{|B|}{|S|}$$

Independence Definition 2:

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

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

Independence?

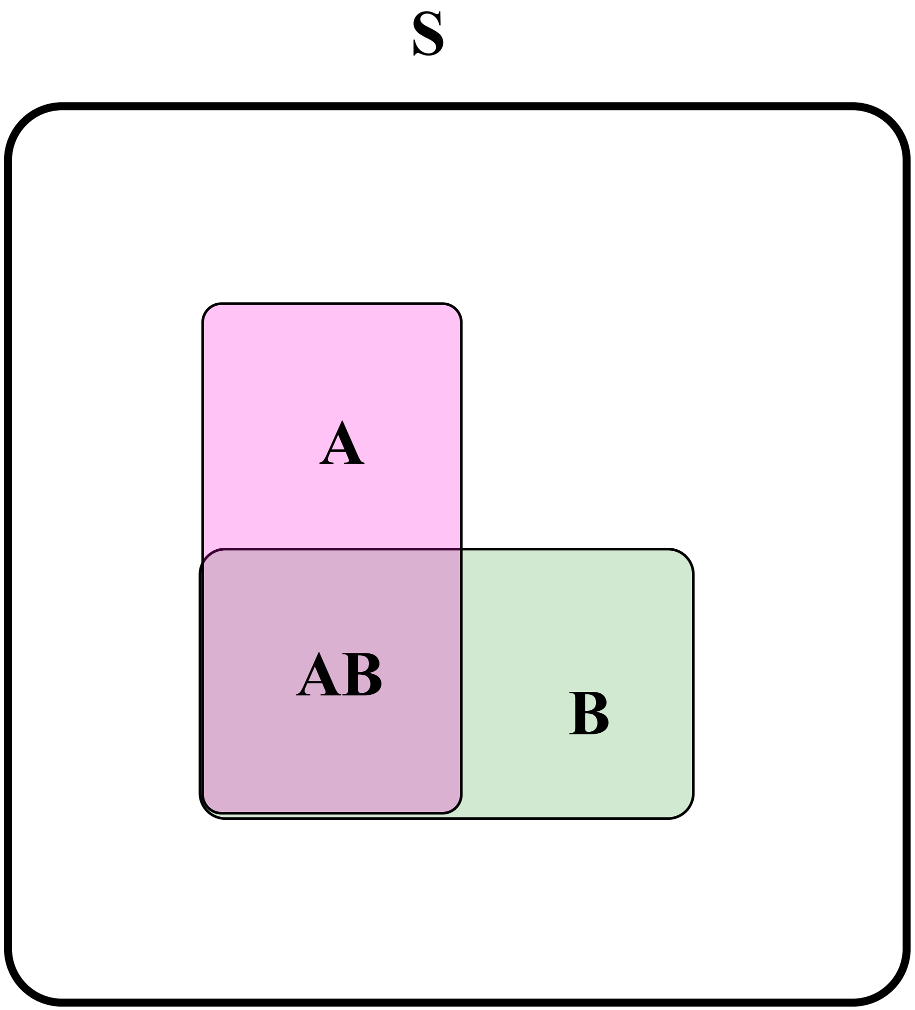


Independence Definition 1:

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

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

Independence?



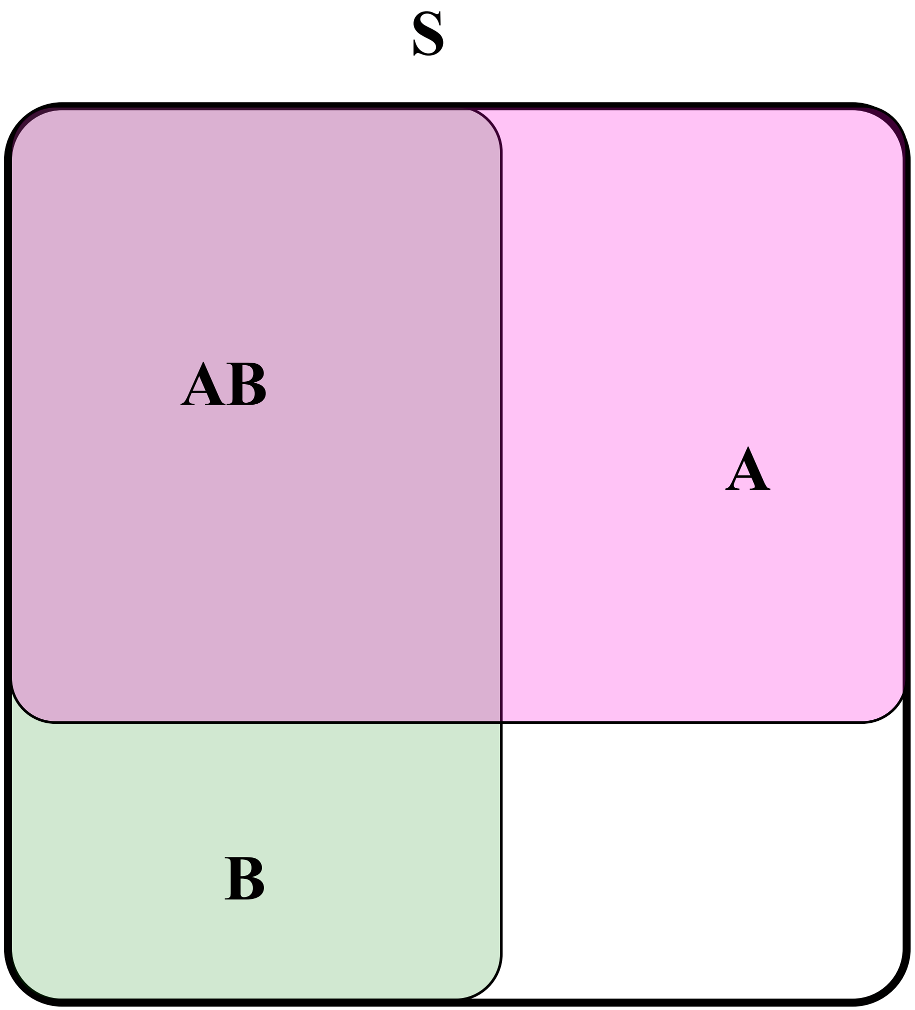
Independence Definition 2:

$$P(A|B) \stackrel{?}{=} P(A)$$

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

$$\frac{1}{2} \neq \frac{2}{16}$$

Independence



Independence Definition 1:

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

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

Independence Definition 2:

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

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

Friday Night Fever

- Population of 10,000 people.
  - Of those, 300 have Malaria (event M) and 200 have Bacterial Infection (event B). 6 people have both.
  - Have Fever if and only if you have Malaria or Bacteria.
  - Are M and B independent?
- Solution:
  - $P(M) = 300 / 10{,}000 = 0.03$
  - $P(B) = 200 / 10{,}000 = 0.02$
  - $P(MB) = 6 / 10{,}000 = 0.0006$
  - $P(M)P(B) = 0.0006$
  - $P(M)P(B) = P(MB)$
  - Independent

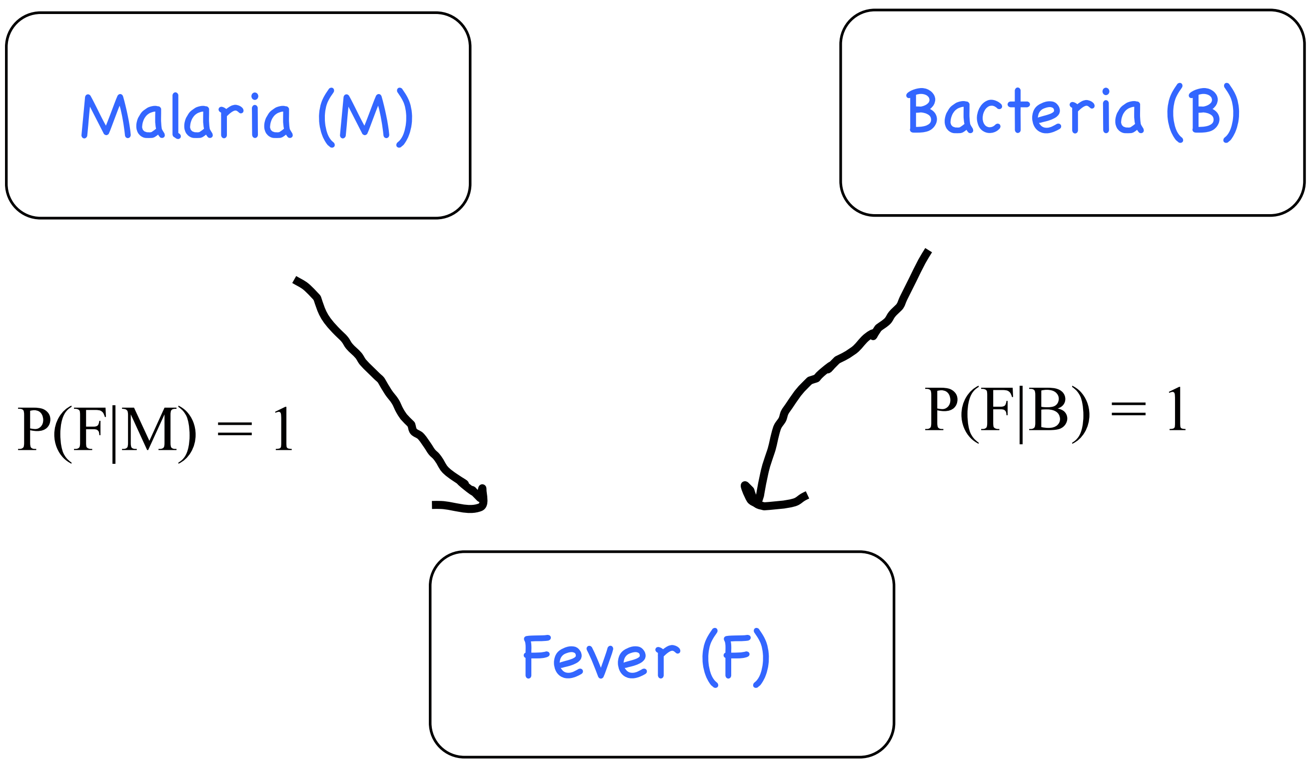
Causality

Malaria does not cause Bacteria and
Bacteria does not cause Malaria

This is 9/10 important

*This is a “causal” diagram. It helps explain why things are independent

Causality

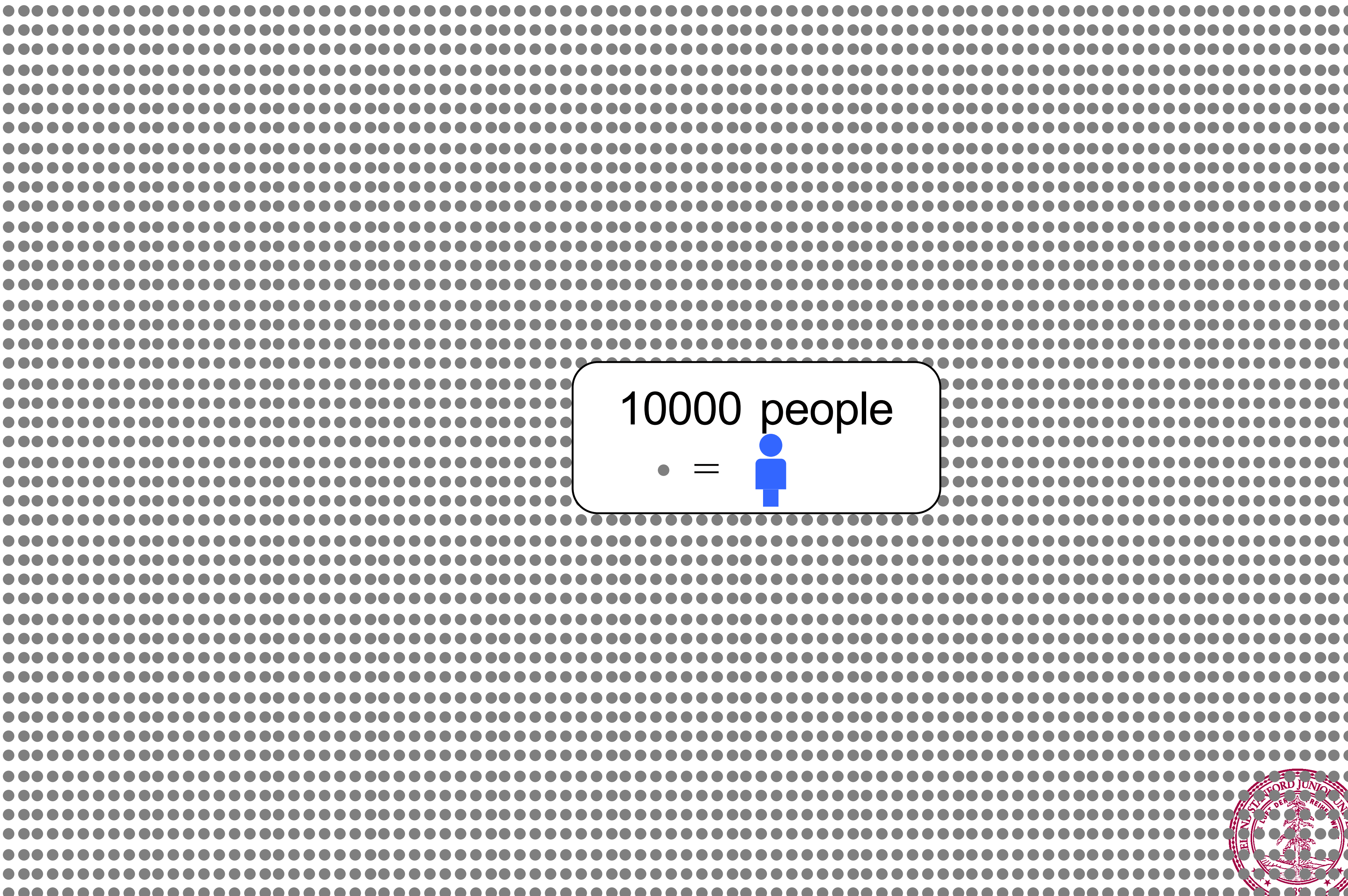


*This is a "causal" diagram. It helps explain why things are independent

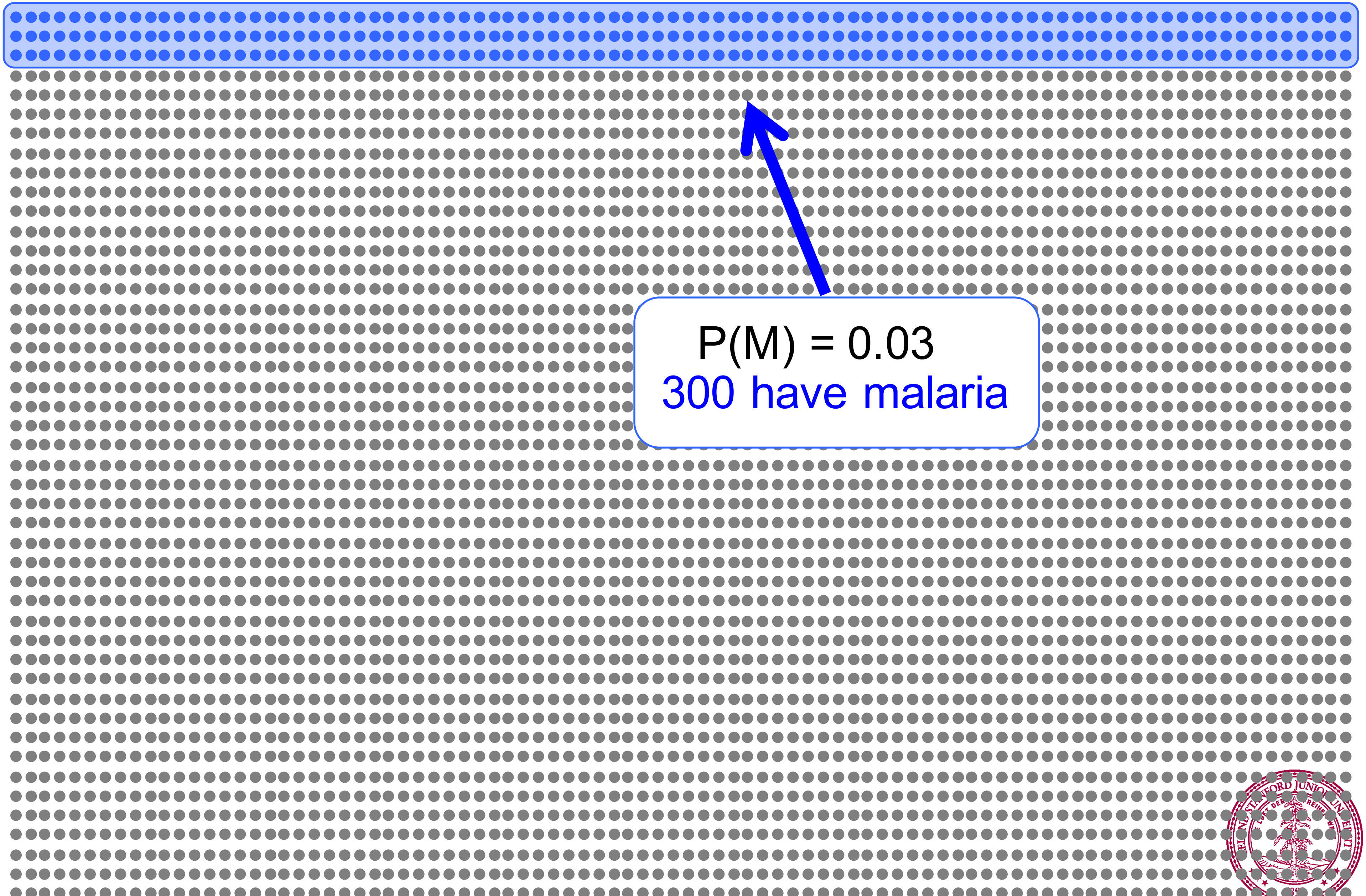
Friday Night Fever

- Population of 10,000 people.
  - Of those, 300 have Malaria (event M) and 200 have Bacterial Infection (event B). 6 people have both.
  - Have Fever if and only if you have Malaria or Bacteria.
  - Are M and B independent given F?
- Solution:
  - Total people with Fever = 200+300 – 6 = 494
  - $P(M|F) = 300 / 494 = 0.61$
  - $P(B|F) = 200 / 494 = 0.40$
  - $P(MB|F) = 6 / 494 = 0.012$
  - $P(M|F)P(B|F) = 0.224$
  - $P(M|F)P(B|F) \neq P(MB|F)$
  - Conditionally dependent

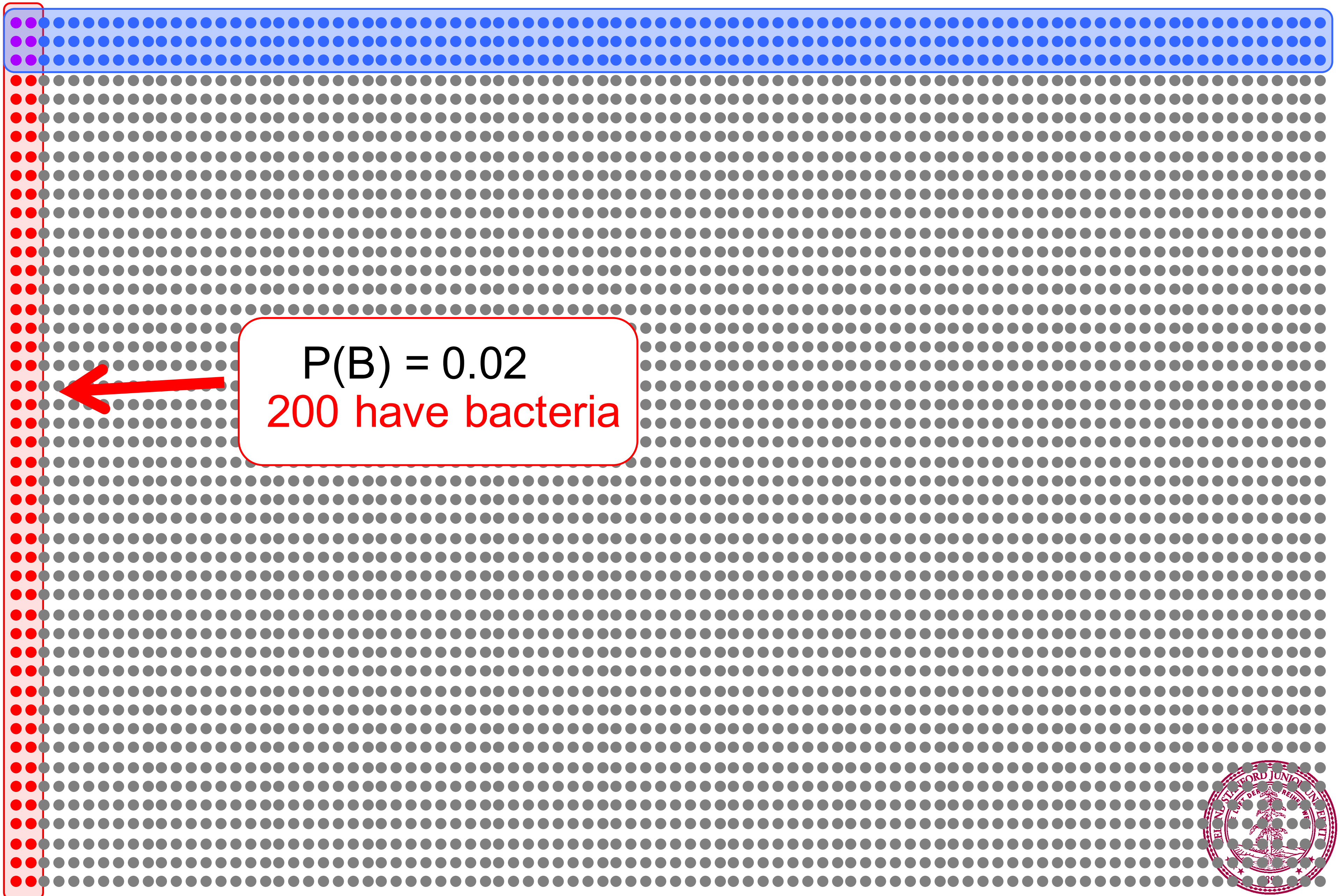
Conditional Dependence



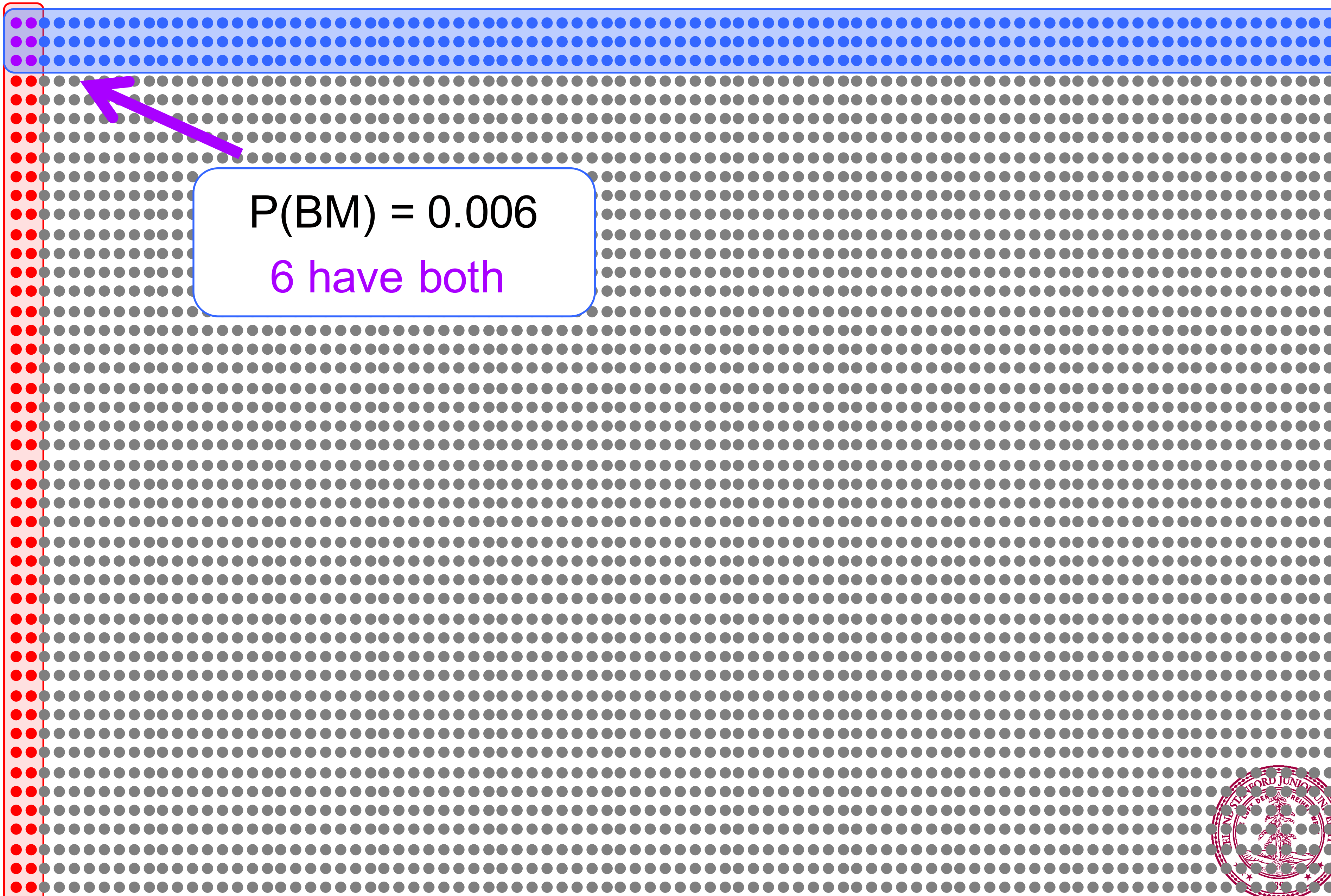
Conditional Dependence



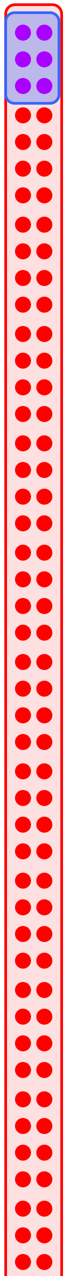
Conditional Dependence



Conditional Dependence



Conditional Dependence

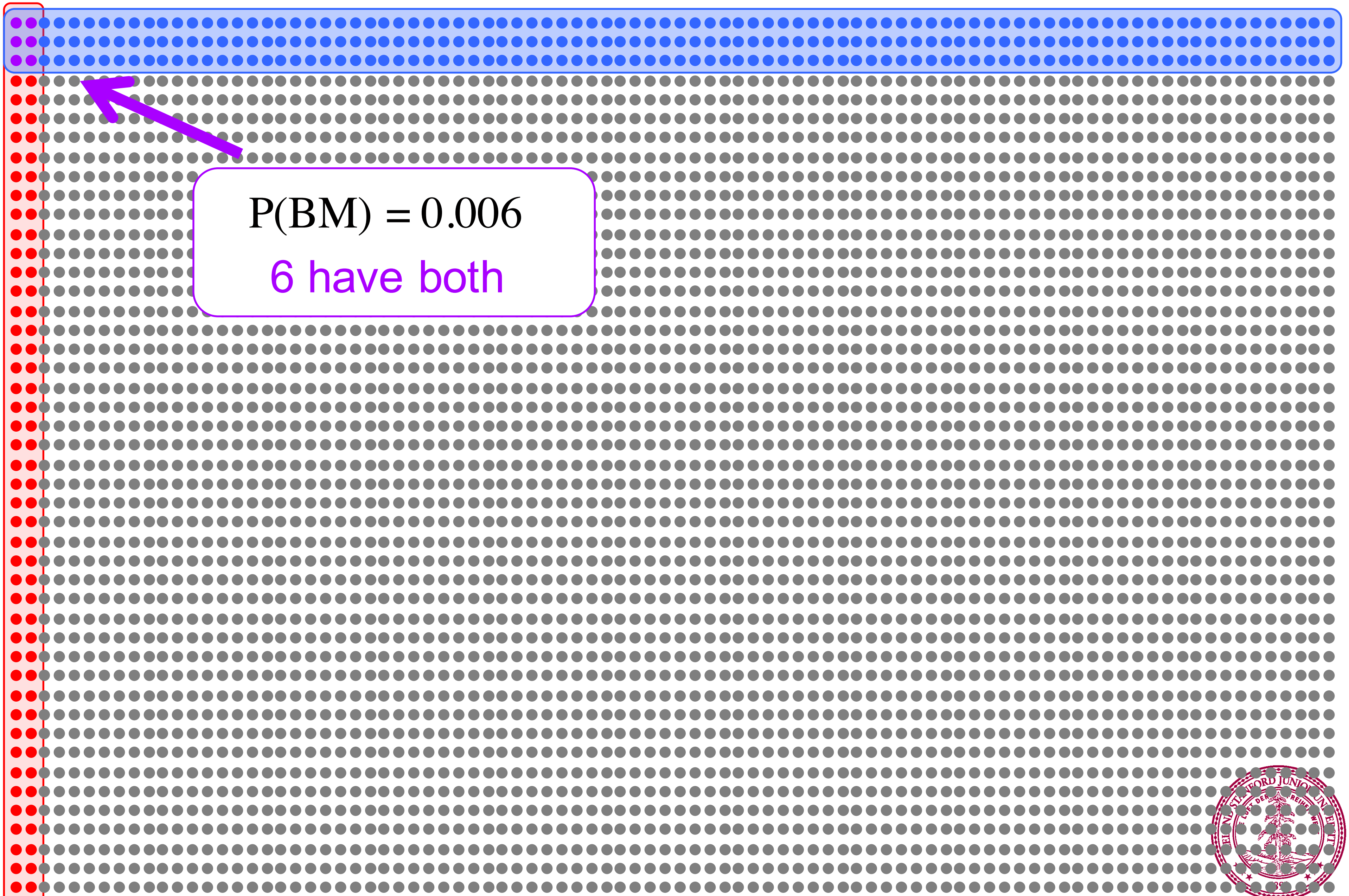


If we condition on B, the same ratio of people have malaria

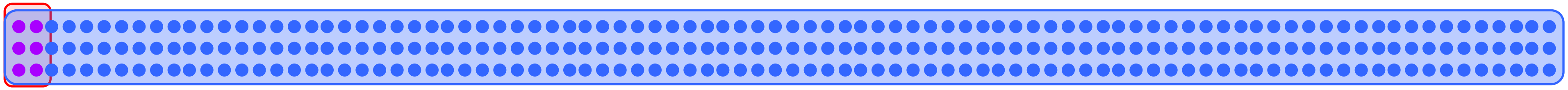
$$P(M|B) = 6/200 = 0.03$$
$$P(M) = 300/10000 = 0.03$$
$$P(M) = P(M|B)$$

That's the math definition of independence

Conditional Dependence



Conditional Dependence

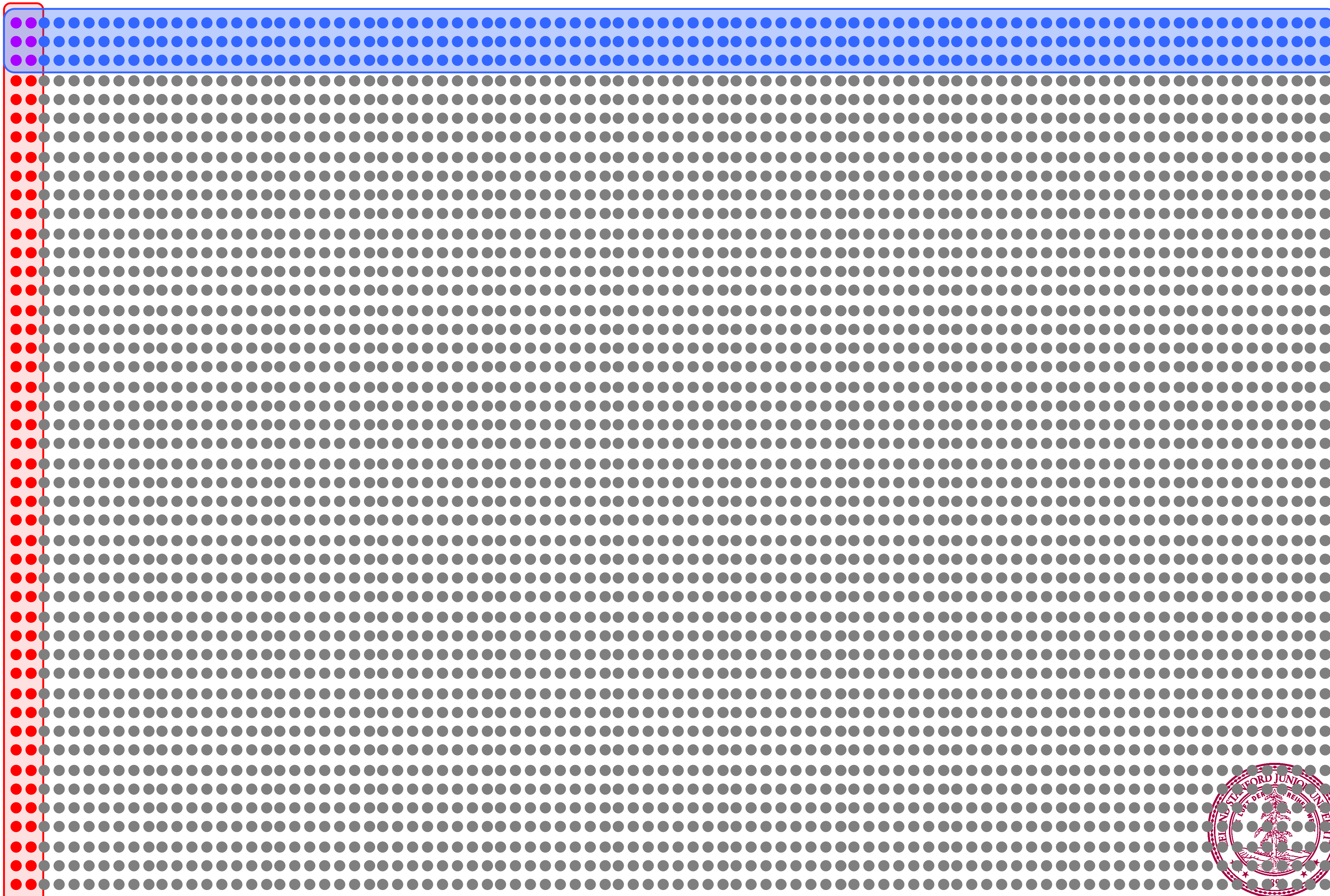


If we condition on M, the same ratio of people have bacteria

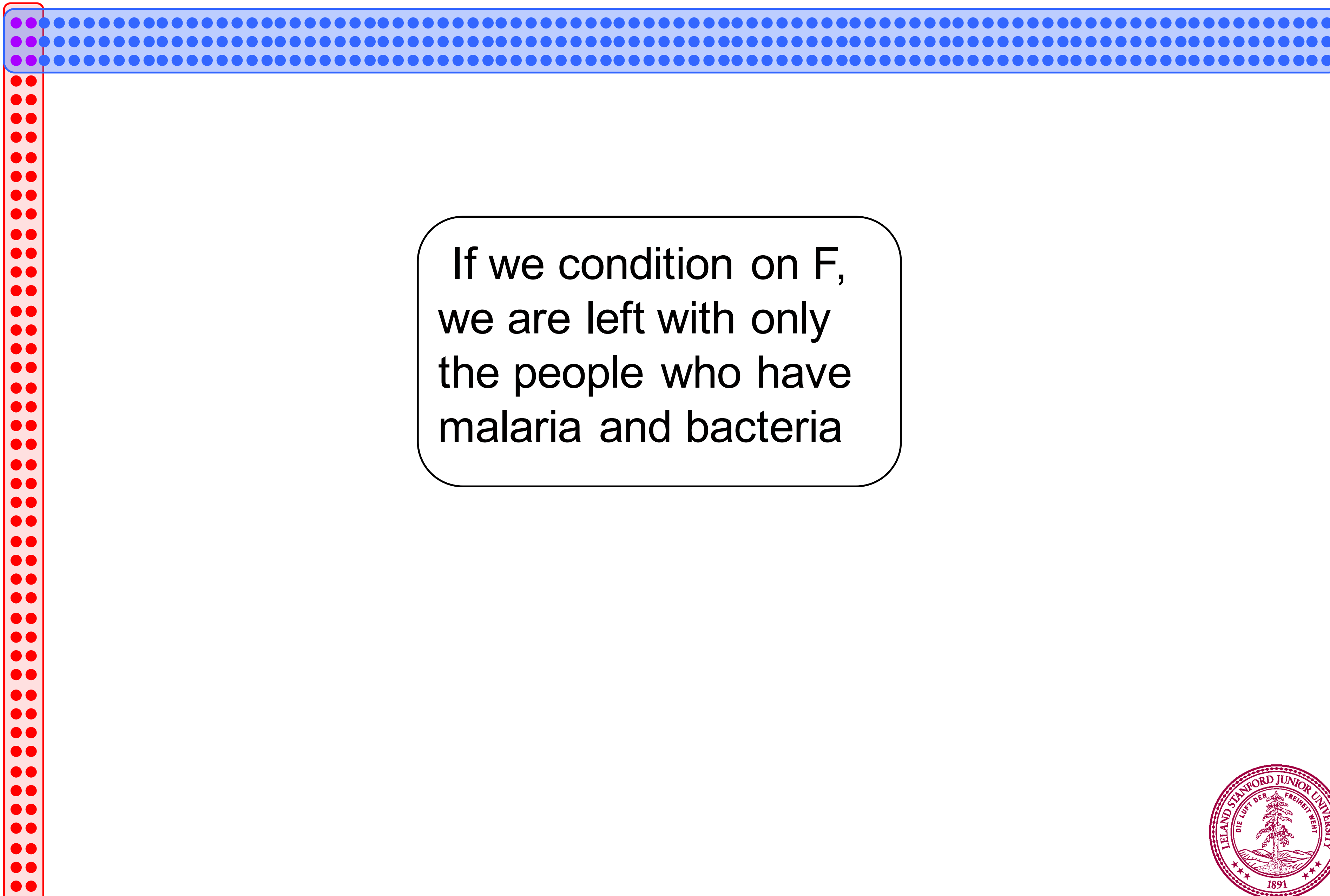
There it is again!

$$P(B|M) = 6/300 = 0.02$$
$$P(B) = 200/10000 = 0.02$$
$$P(B|M) = P(B)$$

Conditional Dependence



Conditioned on Fever

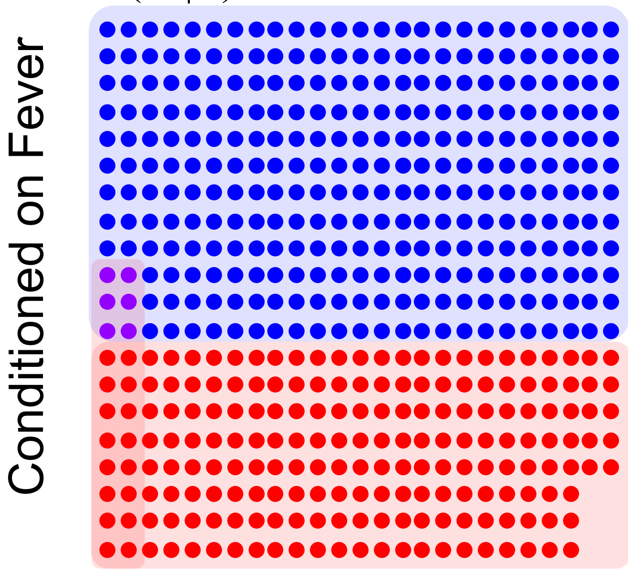


If we condition on F, we are left with only the people who have malaria and bacteria

Conditioned on Fever

P(B|F) = 200/494 = 0.40

P(M|F) = 300/494 = 0.61

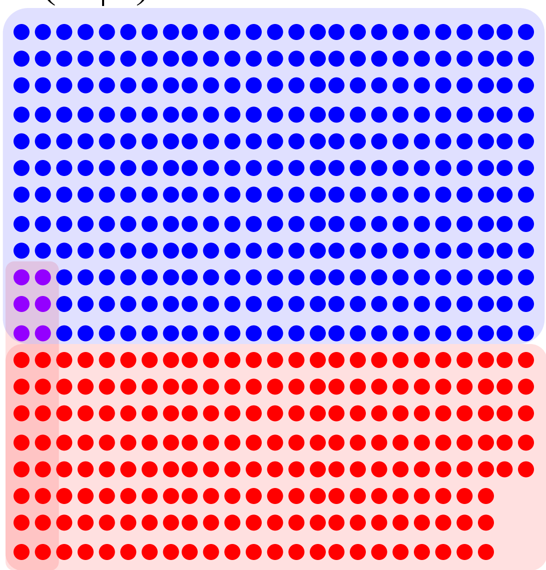


Conditioned on Fever

$P(B|F) = 200/494 = 0.40$

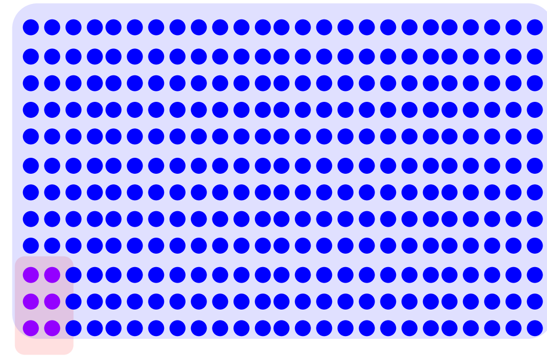
$P(M|F) = 300/494 = 0.61$

Conditioned on Fever



Test shows Malaria →

Conditioned on Fever + Malaria



$P(B|MF) = 6/300 = 0.02$

$P(B|F) \neq P(B|MF)$

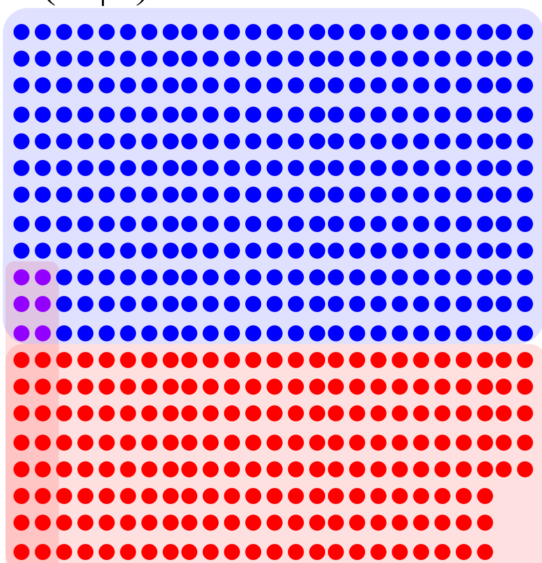
That's the math definition of conditional dependence

Conditioned on Fever

$P(B|F) = 200/494 = 0.40$

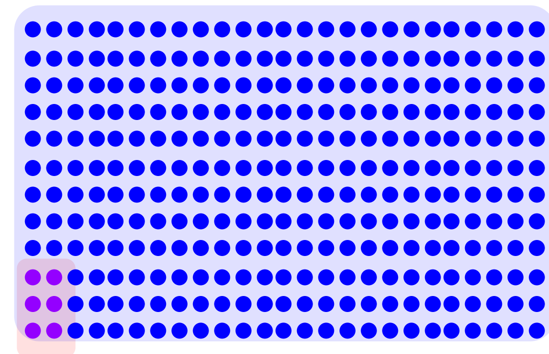
$P(M|F) = 300/494 = 0.61$

Conditioned on Fever



Test shows Malaria

Conditioned on Fever + Malaria



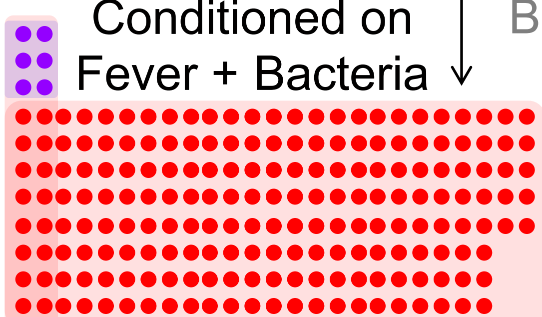
$P(B|MF) = 6/300 = 0.02$

$P(B|F) \ != P(B|MF)$

That's the math definition of conditional dependence

Test shows Bacteria

Conditioned on Fever + Bacteria

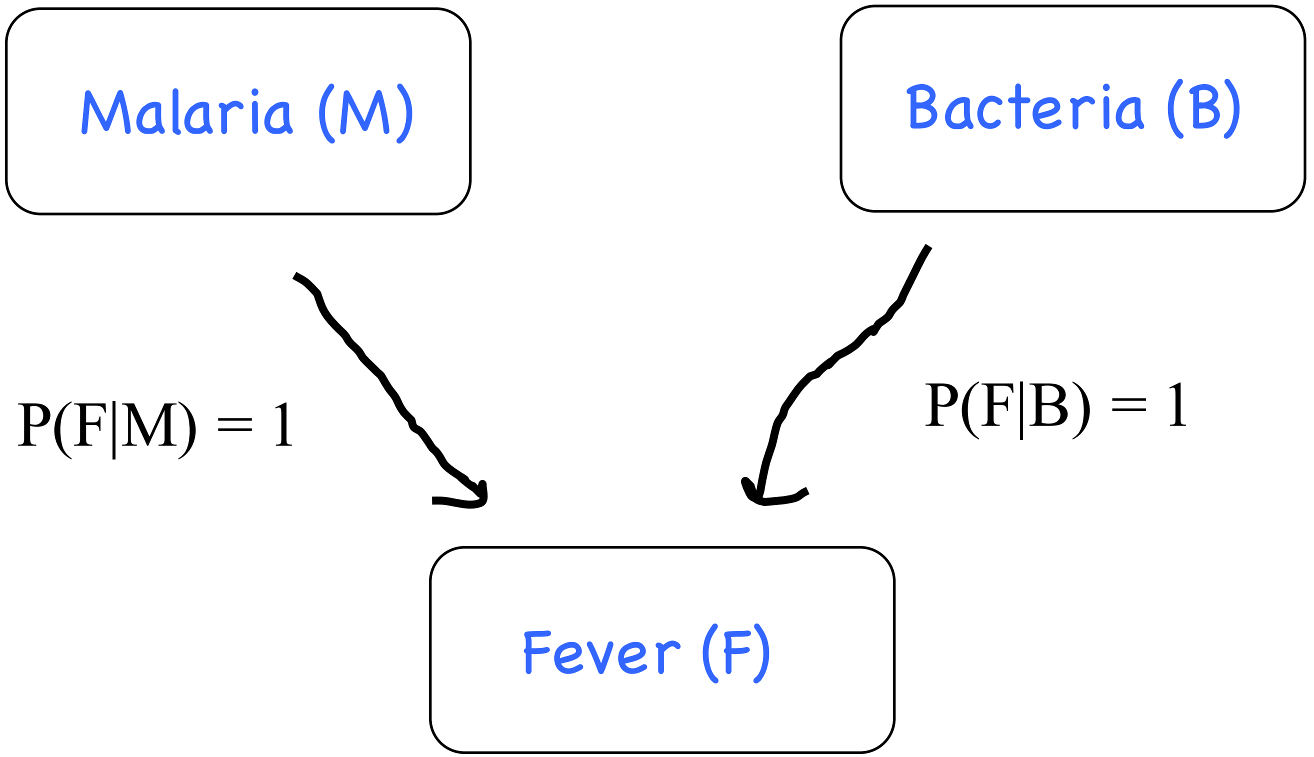


$P(M|BF) = 6/200 = 0.03$

$P(M|F) \neq P(M|BF)$

If we condition on F, the events bacteria and malaria become dependent

Conditional Dependence



*This is a “causal” diagram. It helps explain why things are independent

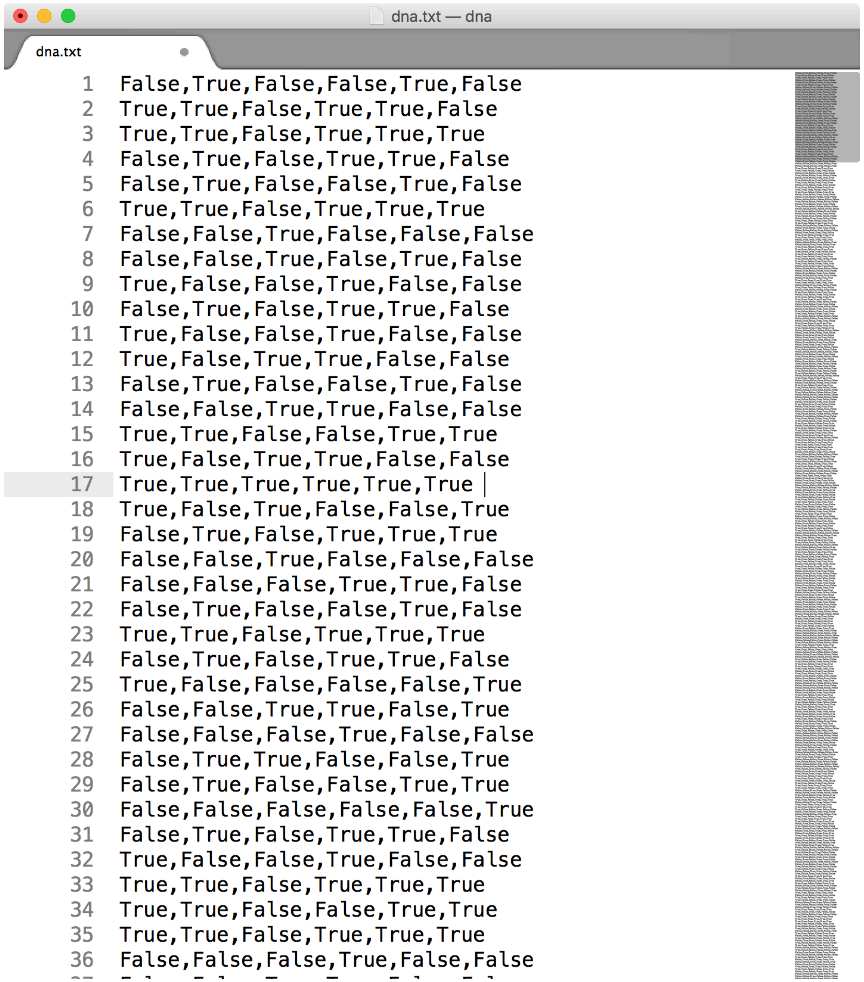
Parents With a Common Child

Say two independent parents have a common child:
When conditioned on the child they are no longer independent

And Here We Are


G_1
G_2
G_3
G_4
G_5
T

G_1 G_2 G_3 G_4 G_5 T




```
False,True,False,False,True,False
True,True,False,True,True,False
True,True,False,True,True,True
False,True,False,True,True,False
False,True,False,False,True,False
True,True,False,True,True,True
False,False,True,False,False,False
False,False,True,False,True,False
True,False,False,True,False,False
False,True,False,True,True,False
True,False,False,True,False,False
True,False,True,True,False,False
False,True,False,False,True,False
False,False,True,True,False,False
True,True,False,False,True,True
True,False,True,True,False,False
True,True,True,True,True,True
True,False,True,False,False,True
False,True,False,True,True,True
False,False,True,False,False,False
False,False,False,True,True,False
False,True,False,False,True,False
True,True,False,True,True,True
False,True,False,True,True,False
True,False,False,False,False,True
False,False,True,True,False,True
False,False,False,True,False,False
False,True,True,False,False,True
False,True,False,False,True,True
False,False,False,False,False,True
False,True,False,True,True,False
True,False,False,True,False,False
True,True,False,True,True,True
True,True,False,False,True,True
True,True,False,True,True,True
False,False,False,True,False,False
```

100,000 samples

6 observations per sample

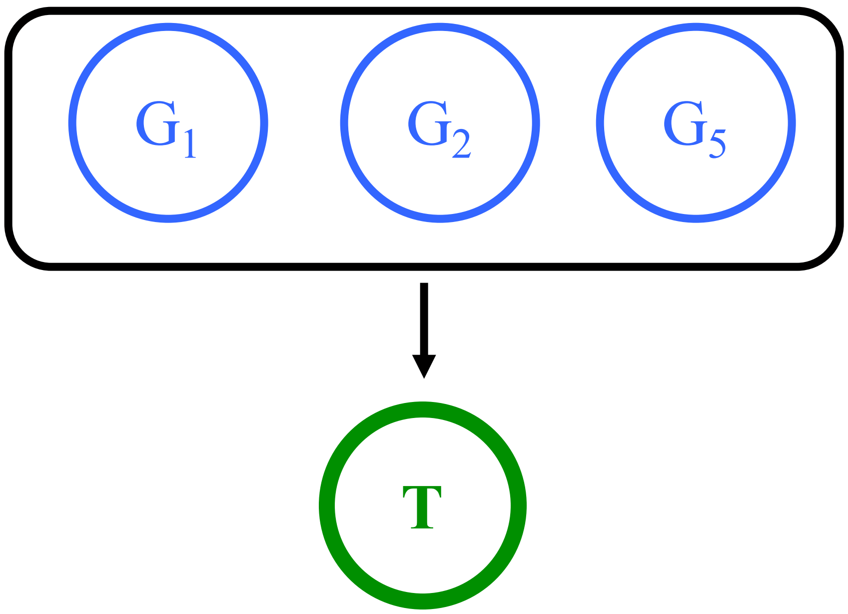
Correlation does not imply causation

Independence implies lack of causation

Model Discovery

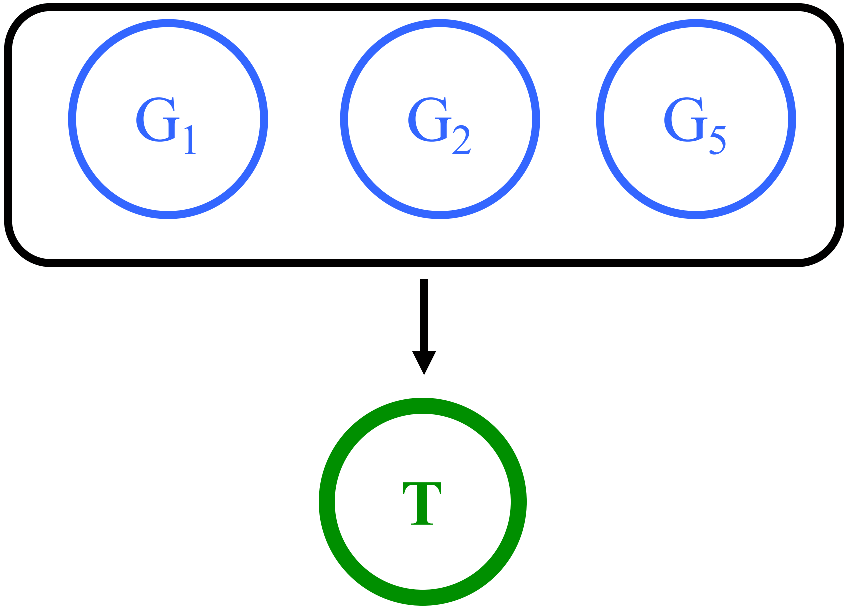
```
p(G1) = 0.500
p(G2) = 0.545
p(G3) = 0.299
p(G4) = 0.701
p(G5) = 0.600
p(T) = 0.390
```

```
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
```



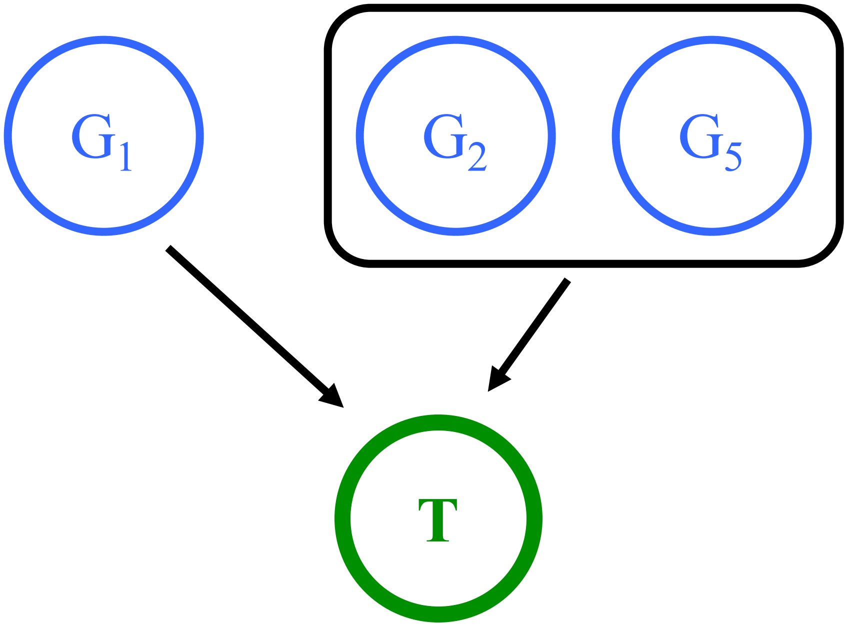
Model Discovery

```
T is independent of G3
T is independent of G4
G1 is independent of G2
G1 is independent of G5
T is independent of G5 | G2
```



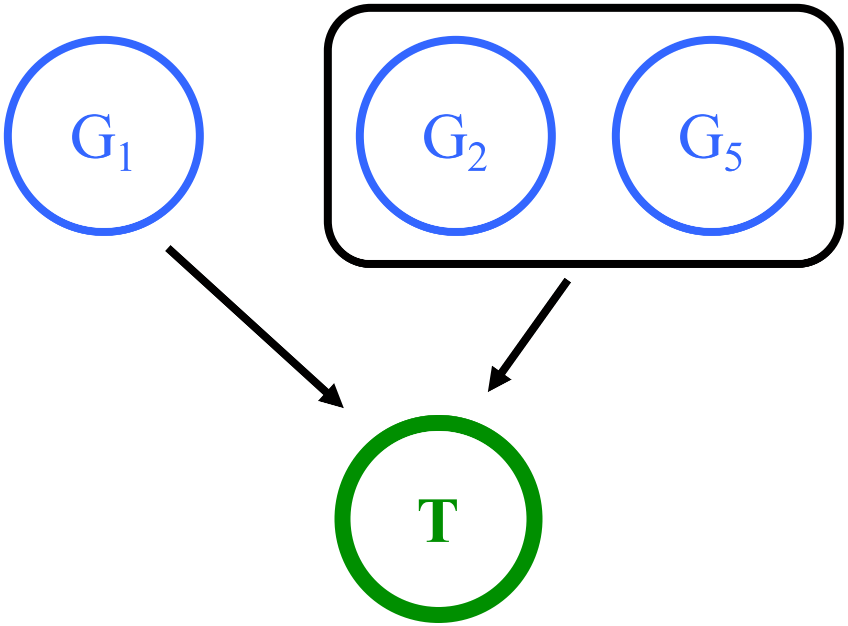
Model Discovery

```
T is independent of G3
T is independent of G4
G1 is independent of G2
G1 is independent of G5
T is independent of G5 | G2
```



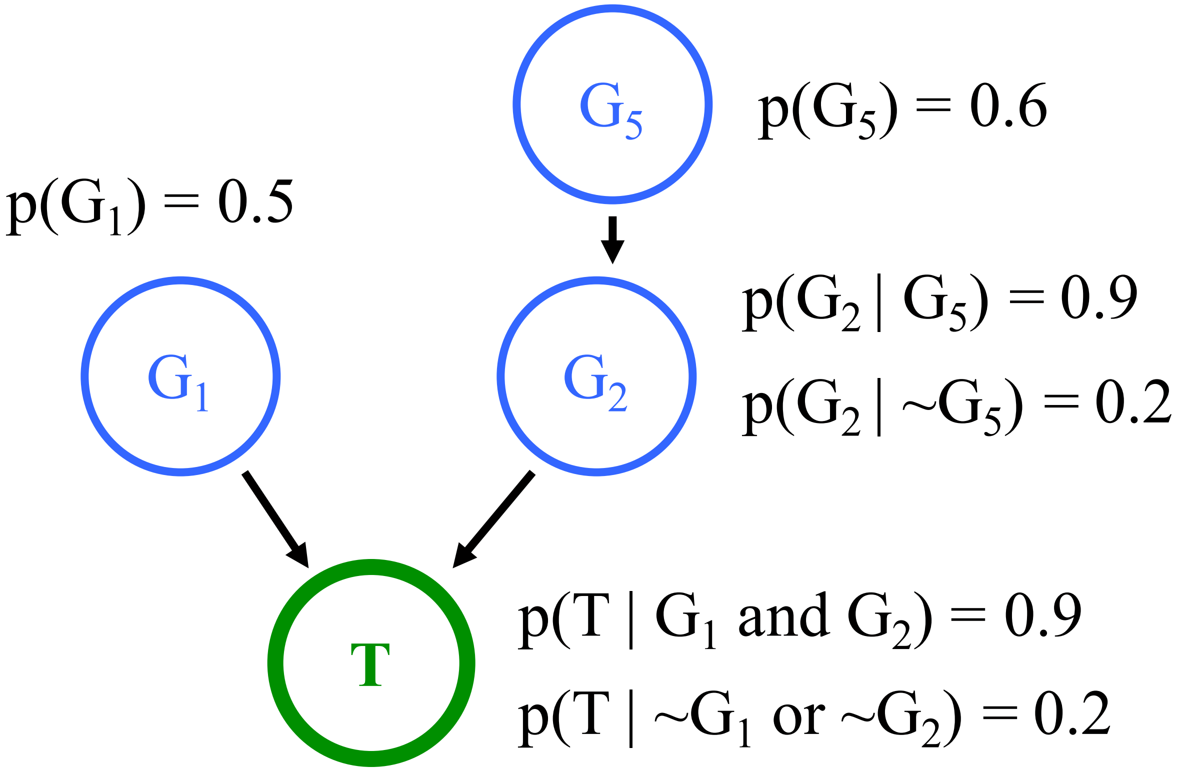
Model Discovery

```
T is independent of G3
T is independent of G4
G1 is independent of G2
G1 is independent of G5
T is independent of G5 | G2
```



Model Discovery

```
T is independent of G3
T is independent of G4
G1 is independent of G2
G1 is independent of G5
T is independent of G5 | G2
```



Summary

Two events A and B are called **<u>independent</u>** if:

$$P(AB) = P(A)P(B) \qquad P(A|B) = P(A)$$

Otherwise, they are called **<u>dependent</u>** events

Two events A and B are **<u>conditionally independent</u>** on C if:

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

$$P(A|BC) = P(A|C)$$

Advanced Reading

Chow–Liu tree - Wikipedia

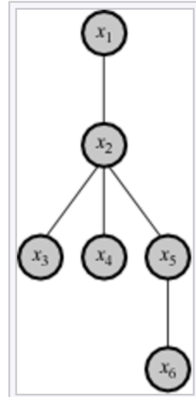
Chris Piech

Secure https://en.wikipedia.org/wiki/Chow–Liu_tree

Chow–Liu tree

From Wikipedia, the free encyclopedia

In probability theory and statistics **Chow–Liu tree** is an efficient method for constructing a second-order product approximation of a joint probability distribution, first described in a paper by Chow & Liu (1968). The goals of such a decomposition, as with such Bayesian networks in general, may be either data compression or inference.



A first-order dependency tree representing the product on the left.

Contents [hide]

The Chow–Liu representation [edit]

The Chow–Liu method describes a joint probability distribution $P(X_1, X_2, \ldots, X_n)$ as a product of second-order conditional and marginal distributions. For example, the six-dimensional distribution $P(X_1, X_2, X_3, X_4, X_5, X_6)$ might be approximated as

$$P'(X_1, X_2, X_3, X_4, X_5, X_6) = P(X_6|X_5)P(X_5|X_4)P(X_4|X_3)P(X_3|X_2)P(X_2|X_1)P(X_1)$$

where each new term in the product introduces just one new variable, and the product can be represented as a first-order