14: Conditional Expectation

Lisa Yan

May 6, 2020

Quick slide reference

3	Conditional distributions	14a_conditional_distributions
11	Web server requests, redux	14b_web_servers
14	Conditional expectation	14c_cond_expectation
20	Law of Total Expectation	14d_law_of_total_expectation
24	Exercises	LIVE

Discrete conditional distributions

Discrete conditional distributions

Recall the definition of the conditional probability of event E given event F:

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

For discrete random variables X and Y, the **conditional PMF** of X given Y is

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

Different notation, same idea:

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

Discrete probabilities of CS109

Each student responds with:

Year Y

- 1: Frosh/Soph
- 2: Jr/Sr
- 3: Co-term/grad/NDO

Timezone T (12pm California time in my timezone is):

- -1: AM
- 0: noon
- 1: PM

Joint PMF

	$Y=1$	$Y=2$	$Y=3$
$T=-1$	.06	.01	.01
$T=0$	.29	.14	.09
$T=1$	.30	.08	$\boxed{.02}$

$P(Y=3, T=1)$

Joint PMFs sum to 1.

Discrete probabilities of CS109

The below are **conditional probability tables** for conditional PMFs
(A) $P(Y = y|T = t)$ and (B) $P(T = t|Y = y)$.

1. Which is which?
2. What's the missing probability?

Joint PMF

	$Y = 1$	$Y = 2$	$Y = 3$
$T = -1$	.06	.01	.01
$T = 0$	.29	.14	.09
$T = 1$	.30	.08	.02

	$Y = 1$	$Y = 2$	$Y = 3$
$T = -1$	.09	.04	.08
$T = 0$	.45	.61	.75
$T = 1$	.46	.35	.17

	$Y = 1$	$Y = 2$	$Y = 3$
$T = -1$	.75	.125	?
$T = 0$	.56	.27	.17
$T = 1$	.75	.2	.05

Discrete probabilities of CS109

The below are **conditional probability tables** for conditional PMFs
(A) $P(Y = y|T = t)$ and (B) $P(T = t|Y = y)$.

1. Which is which?
2. What's the missing probability?

Joint PMF

	$Y = 1$	$Y = 2$	$Y = 3$
$T = -1$	.06	.01	.01
$T = 0$	.29	.14	.09
$T = 1$	.30	.08	.02

(B) $P(T = t|Y = y)$

	$Y = 1$	$Y = 2$	$Y = 3$
$T = -1$	.09	.04	.08
$T = 0$	.45	.61	.75
$T = 1$	.46	.35	.17

.30/(.06+.29+.30)

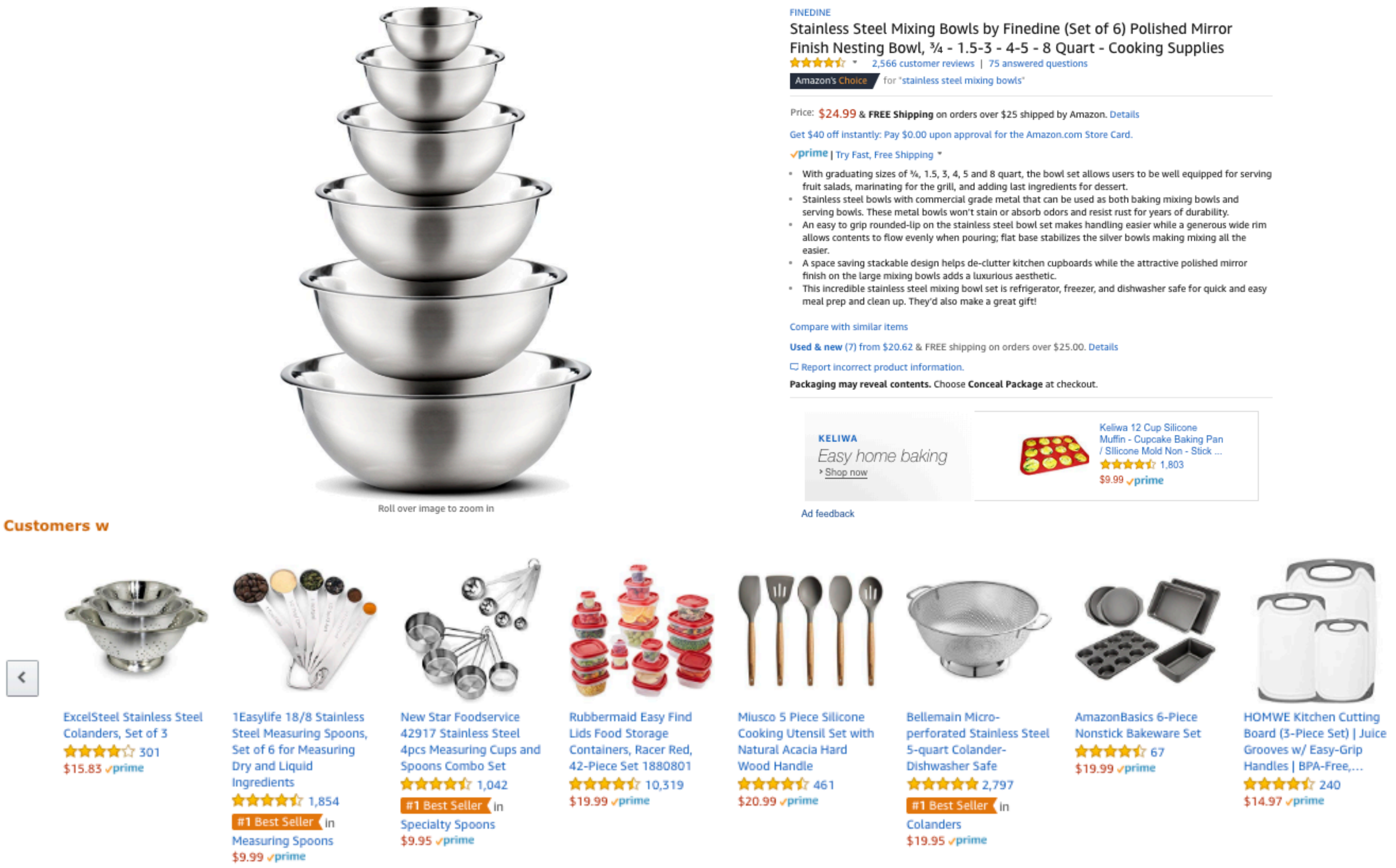
(A) $P(Y = y|T = t)$

	$Y = 1$	$Y = 2$	$Y = 3$
$T = -1$	.75	.125	.125
$T = 0$	.56	.27	.17
$T = 1$	.75	.2	.05

1-.75-.125

Conditional PMFs also sum to 1 conditioned on different events!

Extended to Amazon



P(bought item X | bought item Y)

Quick check

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

Number or function?

1. $P(X = 2|Y = 5)$

2. $P(X = x|Y = 5)$

3. $P(X = 2|Y = y)$

4. $P(X = x|Y = y)$

True or false?

5. $\sum_x P(X = x|Y = 5) = 1$

6. $\sum_y P(X = 2|Y = y) = 1$

7. $\sum_x \sum_y P(X = x|Y = y) = 1$

8. $\sum_x \left(\sum_y P(X = x|Y = y) P(Y = y) \right) = 1$

Quick check

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

Number or function?

1. $P(X = 2|Y = 5)$

   number

2. $P(X = x|Y = 5)$

   1-D function

3. $P(X = 2|Y = y)$

   1-D function

4. $P(X = x|Y = y)$

   2-D function

True or false?

5. $\sum_x P(X = x|Y = 5) = 1$ true

6. $\sum_y P(X = 2|Y = y) = 1$ false

7. $\sum_x \sum_y P(X = x|Y = y) = 1$ false

8. $\sum_x \left(\sum_y P(X = x|Y = y)P(Y = y) \right) = 1$ true

Web server requests, redux

Web server requests (Lecture: Independent RVs)

Review

Let $N = \#$ of requests to a web server per day. Suppose $N \sim \text{Poi}(\lambda)$.

- Each request independently comes from a human (prob. p), or bot $(1-p)$.
- Let X be # of human requests/day, and Y be # of bot requests/day.

Are X and Y independent? What are their marginal PMFs?

Our approach:

- Yes, independent Poisson random variables:

$$X \sim \text{Poi}(\lambda p), Y \sim \text{Poi}\big(\lambda(1-p)\big)$$

- Two big parts of our derivation:
  - $P(X = n, Y = m) = P(X = n | N = n + m) P(N = n)$
  - $X | N = n + m \sim \text{Bin}(n + m, p)$

A conditional distribution, $X|N$!

Web server requests, redux

(Note: this is a different problem setup from the previous slide)

Consider the number of requests to a web server per day.

- Let X = # requests from humans/day. $X \sim \text{Poi}(\lambda_1)$
- Let Y = # requests from bots/day. $Y \sim \text{Poi}(\lambda_2)$
- X and Y are independent. $\rightarrow X + Y \sim \text{Poi}(\lambda_1 + \lambda_2)$

What is $P(X = k | X + Y = n)$?

$$P(X = k | X + Y = n) = \frac{P(X = k, Y = n - k)}{P(X + Y = n)} = \frac{P(X = k)P(Y = n - k)}{P(X + Y = n)} \quad (X,Y \text{ indep.})$$

$$= \frac{e^{-\lambda_1}\lambda_1^k}{k!} \cdot \frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!} \cdot \frac{n!}{e^{-(\lambda_1+\lambda_2)}(\lambda_1 + \lambda_2)^n} = \frac{n!}{k!\,(n-k)!} \cdot \frac{\lambda_1^k \lambda_2^{n-k}}{(\lambda_1 + \lambda_2)^n}$$

$$= \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n-k}$$

$$X | X + Y \sim \text{Bin}\left(X + Y, \frac{\lambda_1}{\lambda_1 + \lambda_2}\right)$$

Conditional Expectation

Conditional expectation

Recall the the conditional PMF of X given $Y = y$:

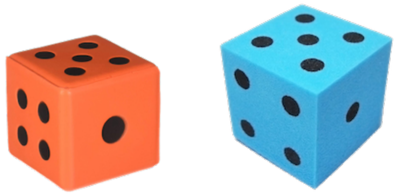
$$p_{X|Y}(x|y) = P(X = x|Y = y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

The **conditional expectation** of X given $Y = y$ is

$$E[X|Y = y] = \sum_x xP(X = x|Y = y) = \sum_x x p_{X|Y}(x|y)$$

It's been so long, our dice friends

$$E[X|Y=y] = \sum_x x p_{X|Y}(x|y)$$



- Roll two 6-sided dice.
- Let roll 1 be D_1, roll 2 be D_2.
- Let S = value of $D_1 + D_2$.

1. What is $E[S|D_2 = 6]$?

$$E[S|D_2 = 6] = \sum_{x=7}^{12} xP(S = x|D_2 = 6)$$

$$= \left(\frac{1}{6}\right)(7 + 8 + 9 + 10 + 11 + 12)$$

$$= \frac{57}{6} = 9.5$$

Intuitively: $6 + E[D_1] = 6 + 3.5 = 9.5$

Let's prove this!

Properties of conditional expectation

1. LOTUS:

$$E[g(X)|Y = y] = \sum_x g(x)p_{X|Y}(x|y)$$

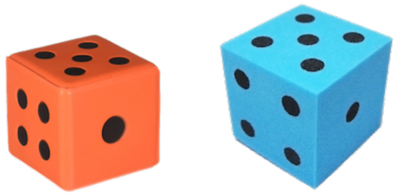
2. Linearity of conditional expectation:

$$E\left[\sum_{i=1}^{n} X_i \mid Y = y\right] = \sum_{i=1}^{n} E[X_i|Y = y]$$

3. Law of total expectation (next time)

It's been so long, our dice friends

$$E[X|Y=y] = \sum_x x p_{X|Y}(x|y)$$

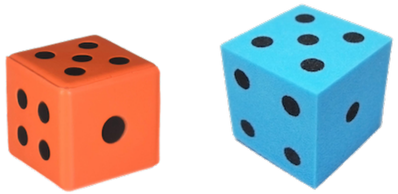


- Roll two 6-sided dice.
- Let roll 1 be D_1, roll 2 be D_2.
- Let S = value of $D_1 + D_2$.

1. What is $E[S|D_2 = 6]$? $\frac{57}{6} = 9.5$
2. What is $E[S|D_2]$?
   A. A function of S
   B. A function of D_2
   C. A number
3. Give an expression for $E[S|D_2]$.

It's been so long, our dice friends

$$E[X|Y=y] = \sum_x x p_{X|Y}(x|y)$$



- Roll two 6-sided dice.
- Let roll 1 be D_1, roll 2 be D_2.
- Let S = value of $D_1 + D_2$.

1. What is $E[S|D_2 = 6]$? $\frac{57}{6} = 9.5$
2. What is $E[S|D_2]$?
   A. A function of S
   B. A function of D_2
   C. A number
3. Give an expression for $E[S|D_2]$.

$$E[S|D_2 = d_2] = E[D_1 + d_2 | D_2 = d_2]$$

$$= \sum_{d_1} (d_1 + d_2) P(D_1 = d_1 | D_2 = d_2)$$

$$= \sum_{d_1} d_1 P(D_1 = d_1) + d_2 \sum_{d_1} P(D_1 = d_1)$$

($D_1 = d_1, D_2 = d_2$ independent events)

$$= E[D_1] + d_2 = 3.5 + d_2$$

$$E[S|D_2] = 3.5 + D_2$$

Law of Total Expectation

Properties of conditional expectation

1. LOTUS:

$$E[g(X)|Y = y] = \sum_x g(x)p_{X|Y}(x|y)$$

2. Linearity of conditional expectation:

$$E\left[\sum_{i=1}^{n} X_i \mid Y = y\right] = \sum_{i=1}^{n} E[X_i|Y = y]$$

3. Law of total expectation:

$$\boxed{E[X] = E\big[E[X|Y]\big]}$$ what?!

Proof of Law of Total Expectation

$E[X] = E\big[E[X|Y]\big]$

$$E\big[E[X|Y]\big] = E[g(Y)] = \sum_y P(Y=y)E[X|Y=y]$$ (LOTUS, $g(Y) = E[X|Y]$)

$$= \sum_y P(Y=y) \sum_x xP(X=x|Y=y)$$ (def of conditional expectation)

$$= \sum_y \left(\sum_x xP(X=x|Y=y)P(Y=y) \right) = \sum_y \left(\sum_x xP(X=x,Y=y) \right)$$ (chain rule)

$$= \sum_x \sum_y xP(X=x,Y=y) \quad = \sum_x x \sum_y P(X=x,Y=y)$$ (switch order of summations)

$$= \sum_x xP(X=x)$$ (marginalization)

$$= E[X]$$ ...what?

Another way to compute $E[X]$

$E[X] = E[E[X|Y]]$

$$E[E[X|Y]] = \sum_y P(Y = y)E[X|Y = y] = E[X]$$

If we only have a conditional PMF of X on some discrete variable Y, we can compute $E[X]$ as follows:

1. Compute expectation of X given some value of $Y = y$
2. Repeat step 1 for all values of Y
3. Compute a weighted sum (where weights are $P(Y = y)$)

```
def recurse():
  if (random.random() < 0.5):
      return 3
  else: return (2 + recurse())
```

Useful for analyzing recursive code!!

15: General Inference

Lisa Yan
May 8, 2020

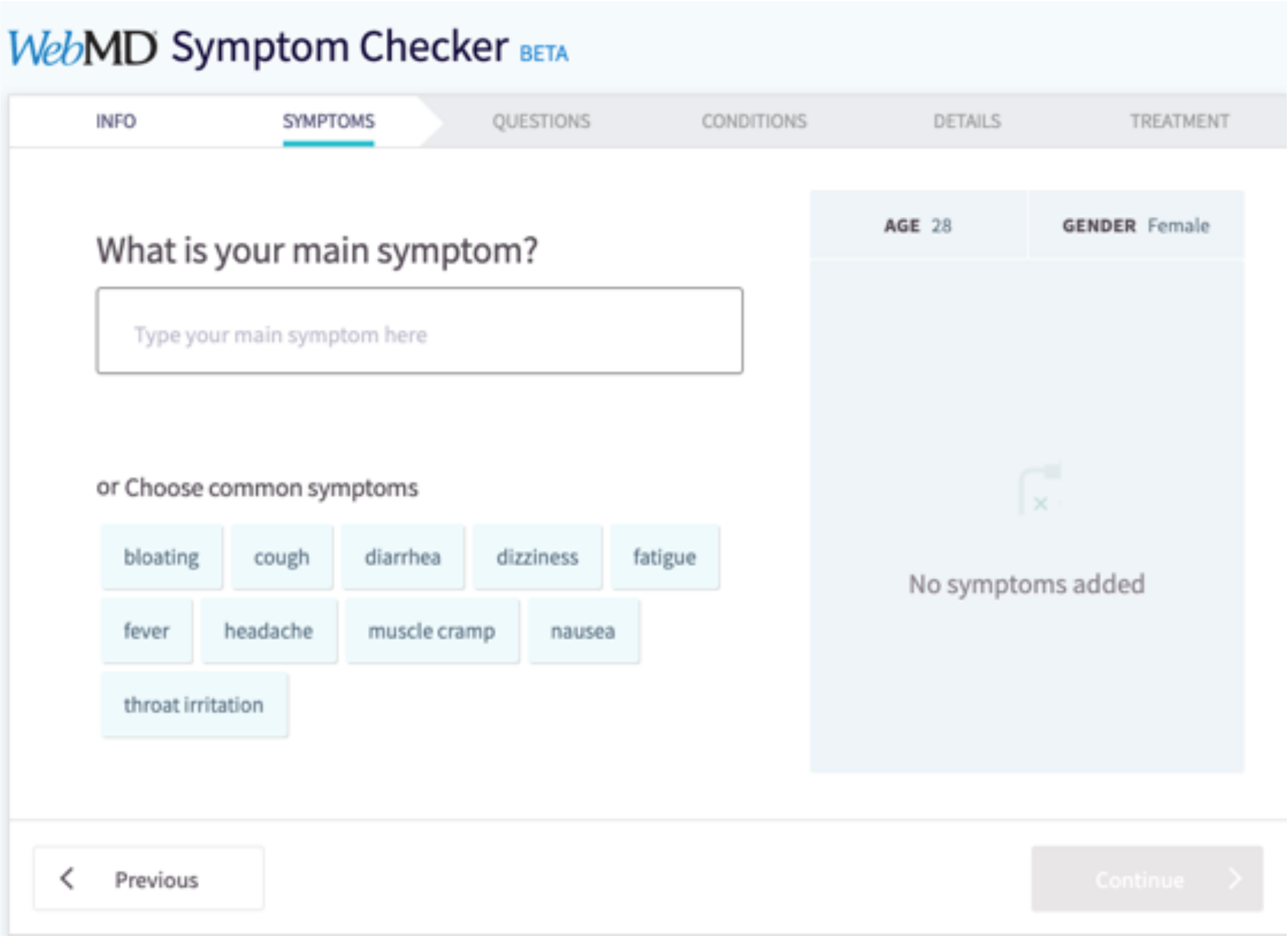
Quick slide reference

3	General Inference: intro	15a_inference
15	Bayesian Networks	15b_bayes_nets
22	Inference (I): Math	15c_inference_math
29	Inference (II): Rejection sampling	LIVE
69	Inference (III): Gibbs sampling (extra)	(no video)

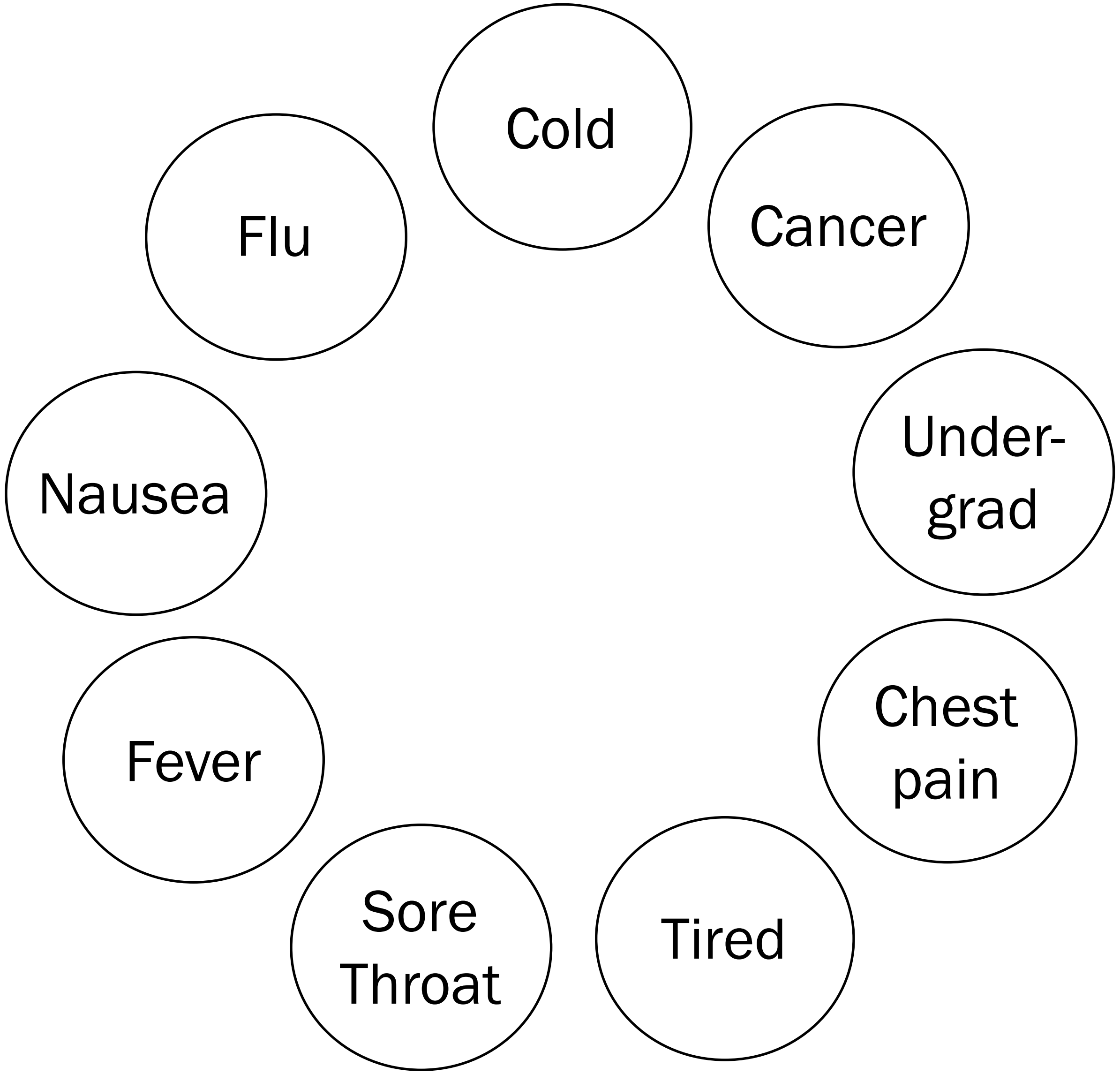
General Inference: Introduction

Inference

Inference


WebMD Symptom Checker BETA
INFO
SYMPTOMS
QUESTIONS
CONDITIONS
DETAILS
TREATMENT
What is your main symptom?
Type your main symptom here
or Choose common symptoms
bloating
cough
diarrhea
dizziness
fatigue
fever
headache
muscle cramp
nausea
throat irritation
AGE 28
GENDER Female
No symptoms added
Previous
Continue

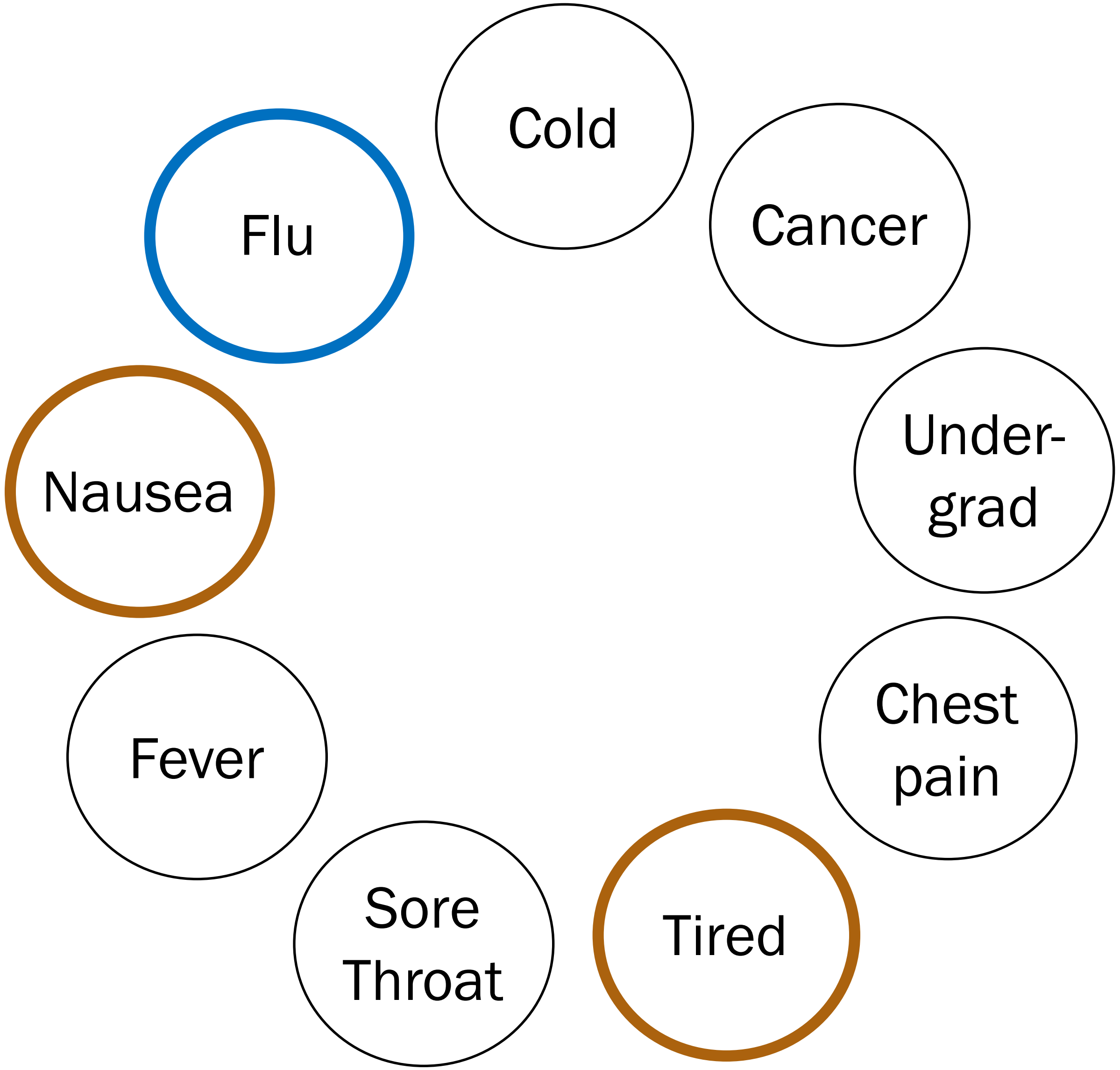
Inference



General inference question:

Given the values of some random variables, what is the conditional distribution of some other random variables?

Inference

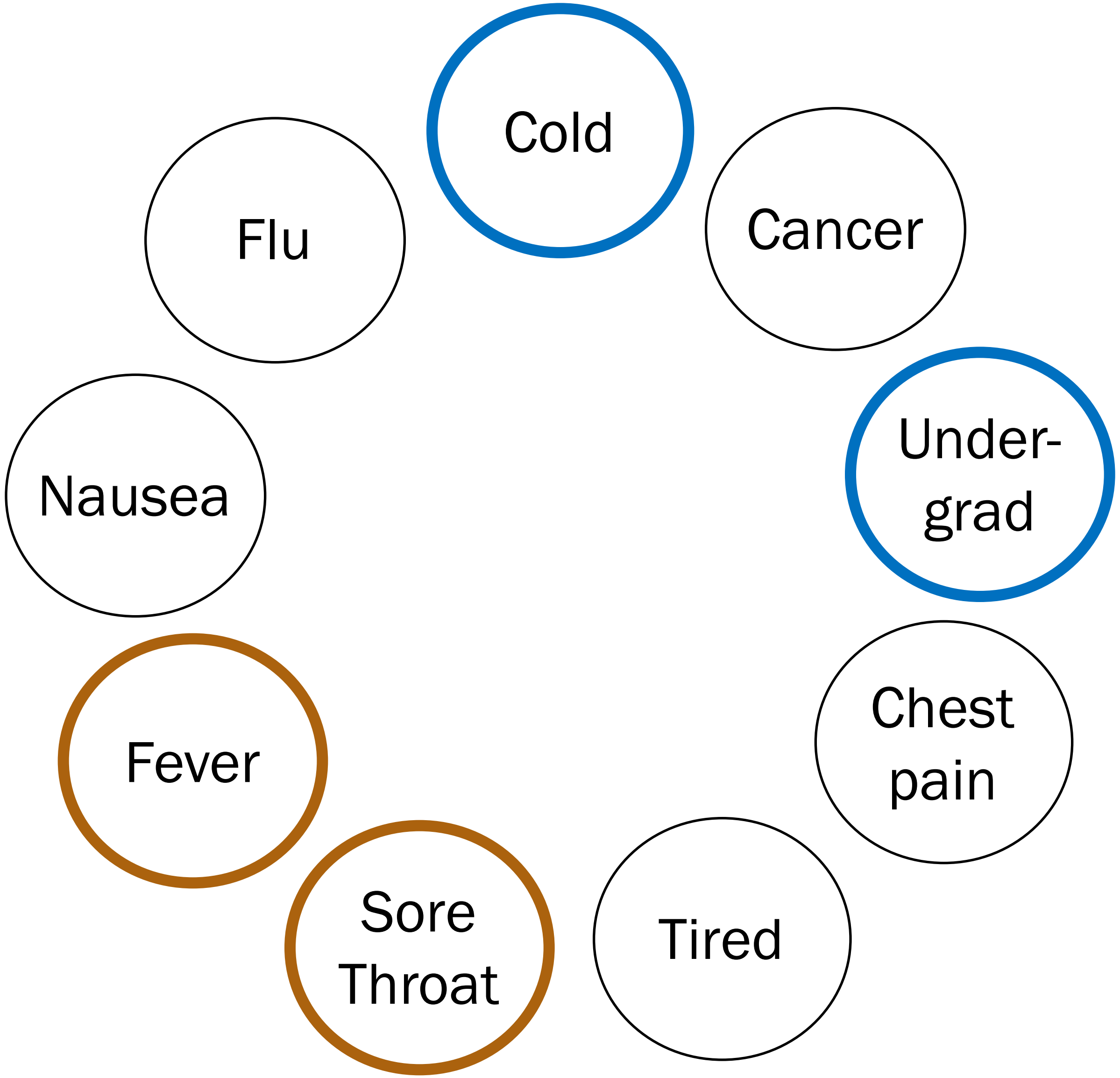


One inference question:

$$P(F = 1 | N = 1, T = 1)$$

$$= \frac{P(F = 1, N = 1, T = 1)}{P(N = 1, T = 1)}$$

Inference

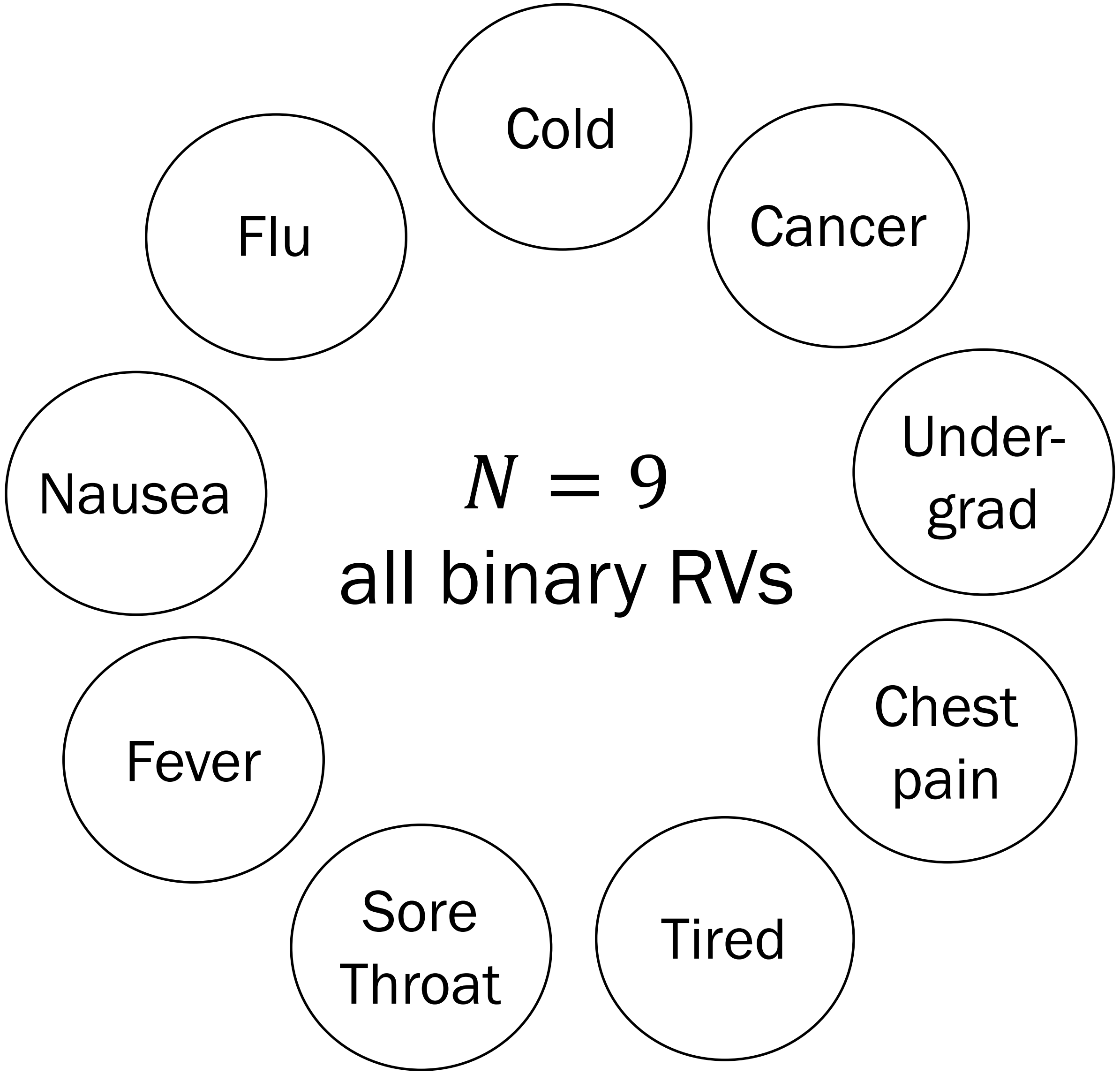


Another inference question:

$$P(C_o = 1, U = 1 | S = 0, F_e = 0)$$

$$= \frac{P(C_o = 1, U = 1, S = 0, F_e = 0)}{P(S = 0, F_e = 0)}$$

Inference

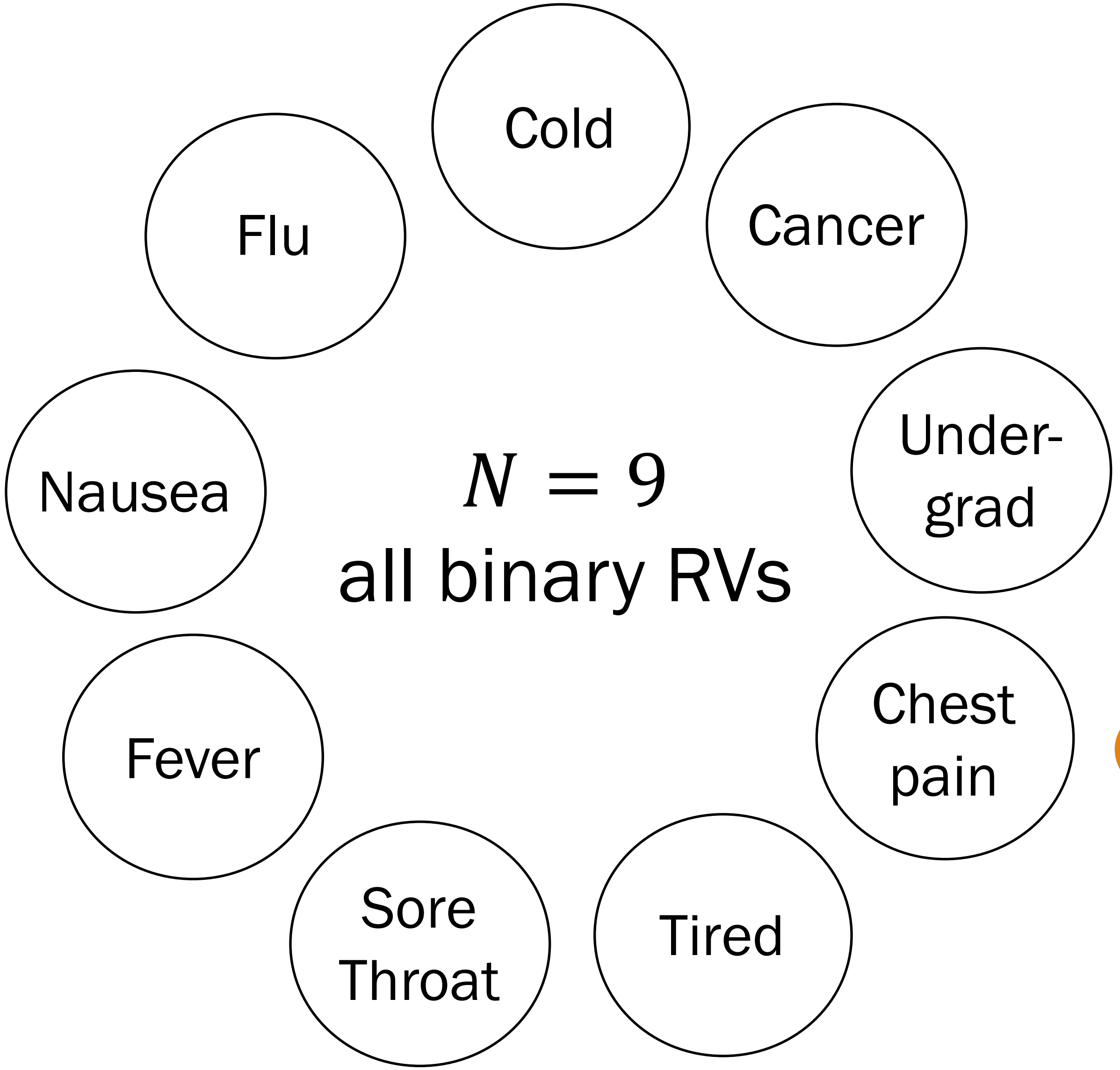


If we knew the **joint distribution**, we can answer all probabilistic inference questions.

What is the size of the joint probability table?

A. 2^{N-1} entries
B. N^2 entries
C. 2^N entries
D. None/other/don't know

Inference



$N = 9$
all binary RVs

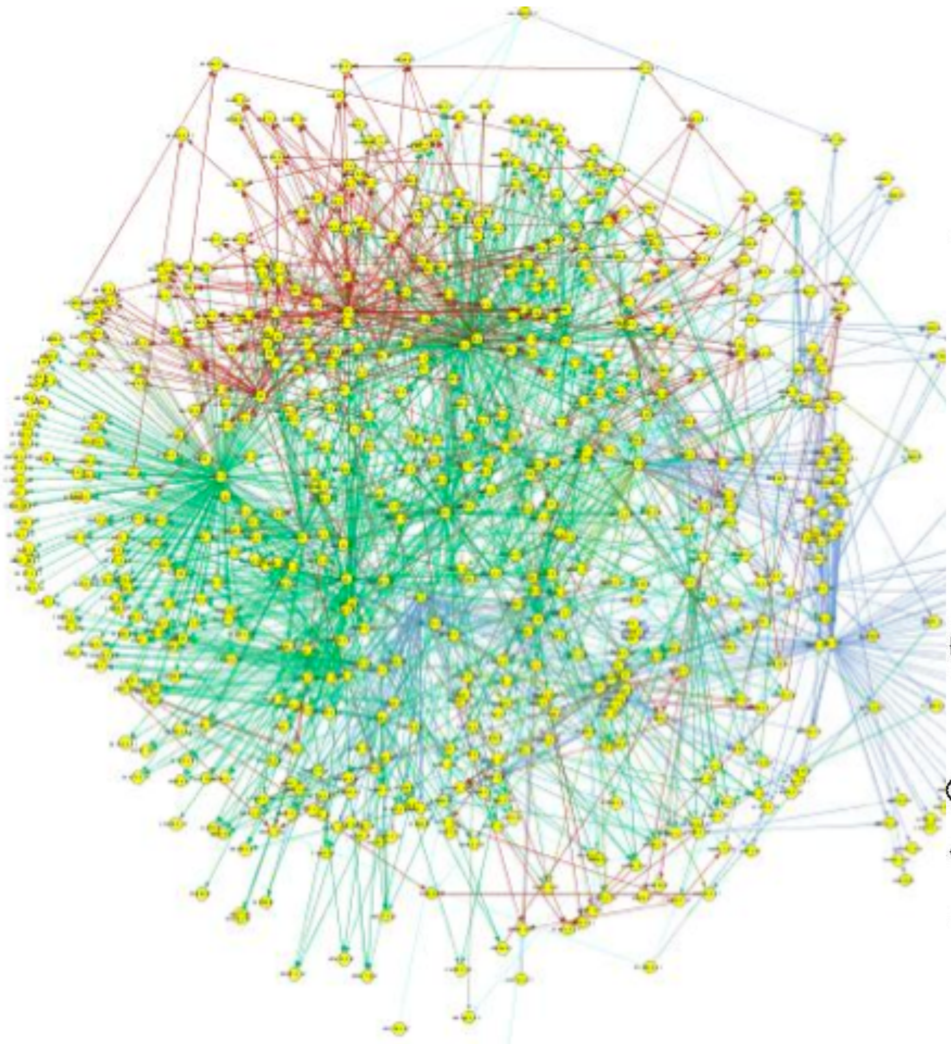
If we knew the **joint distribution**, we can answer all probabilistic inference questions.

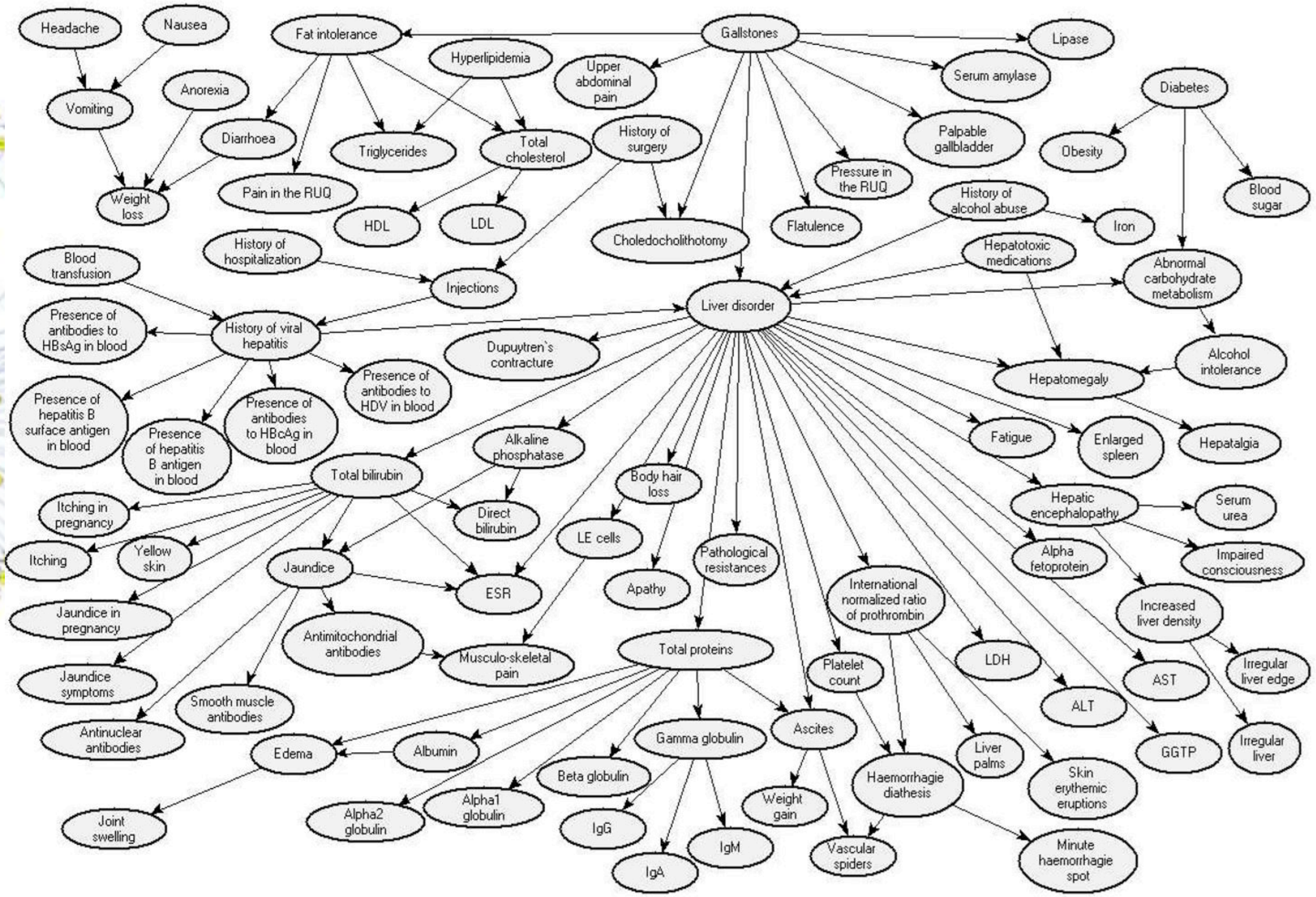
What is the size of the joint probability table?

A. 2^{N-1} entries
B. N^2 entries
C. 2^N entries
D. None/other/don't know

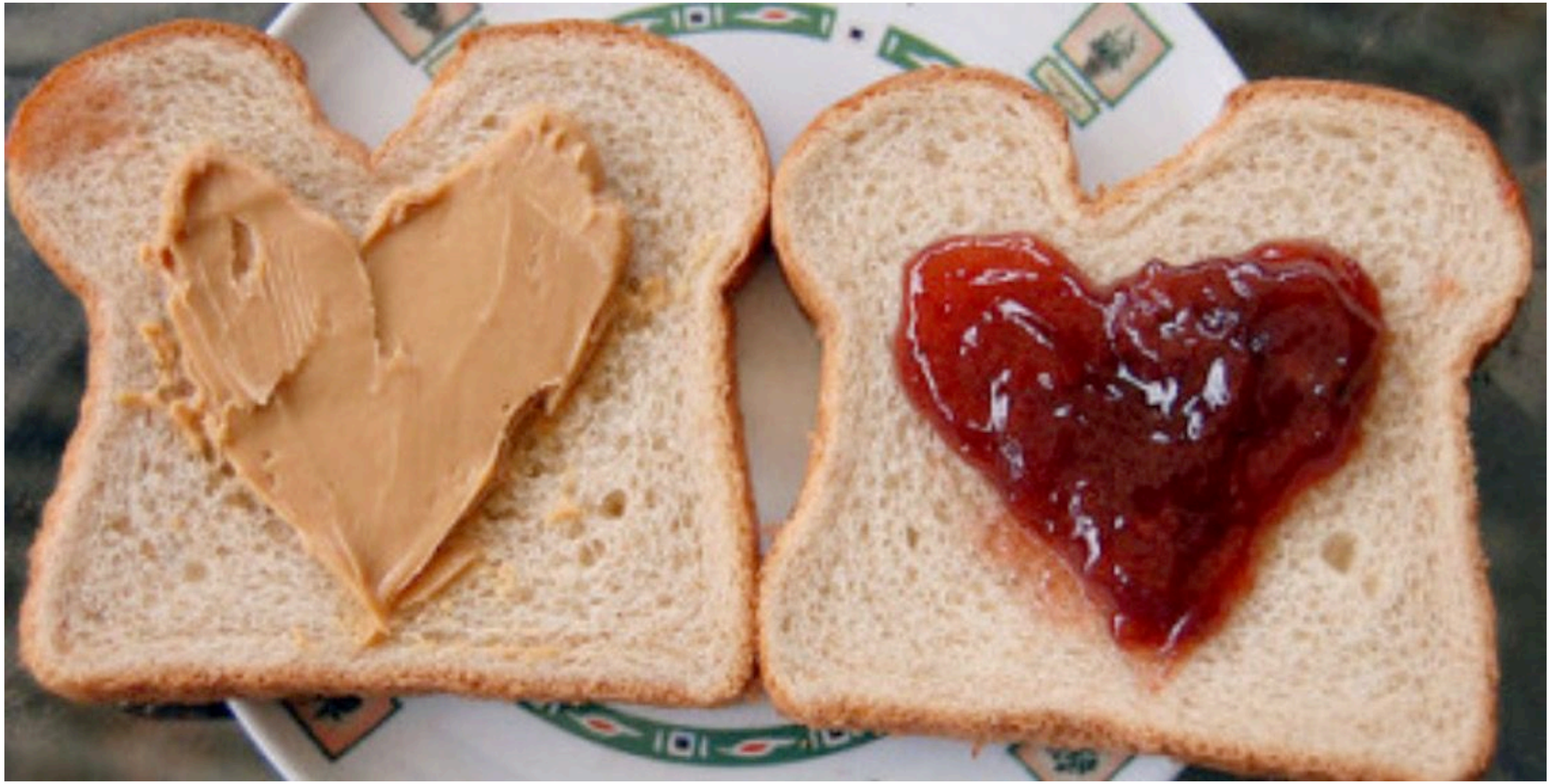
Naively specifying a joint distribution is often intractable.

N can be large…





Conditionally Independent RVs



~~Conditional Probability~~
Conditional Distributions

~~Independence~~
Independent RVs

Conditionally Independent RVs

Recall that two events A and B are conditionally independent given E if:

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

n discrete random variables $X_1, X_2, \dots, X_n$ are called **conditionally independent given Y** if:

for all $x_1, x_2, \dots, x_n, y$:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n | Y = y) = \prod_{i=1}^{n} P(X_i = x_i | Y = y)$$

This implies the following (cool to remember for later):

$$\log P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n | Y = y) = \sum_{i=1}^{n} \log P(X_i = x_i | Y = y)$$

Lec. 12: Independence of multiple random variables

Errata

Recall independence of n events $E_1, E_2, \dots, E_n$:

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

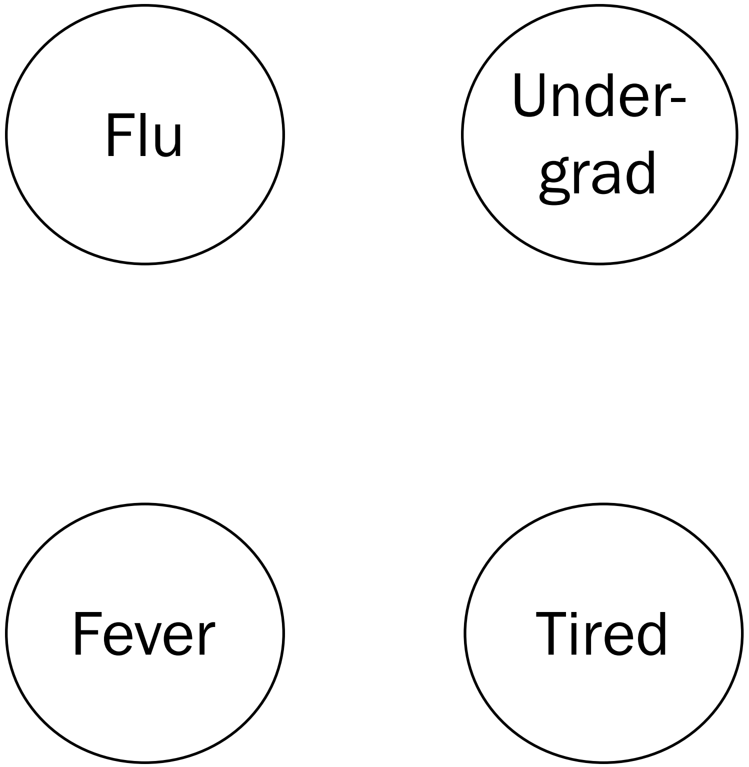
We have independence of n discrete random variables $X_1, X_2, \dots, X_n$ if for all $x_1, x_2, \dots, x_n$:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i)$$

Errata (edited May 3): **Removed the independent RV requirement for all subsets of size** $r = 1, \dots, n$. Do you see why this requirement is unnecessary?
(Hint: independence of RVs implies independence of all events)

Bayesian Networks

A simpler WebMD



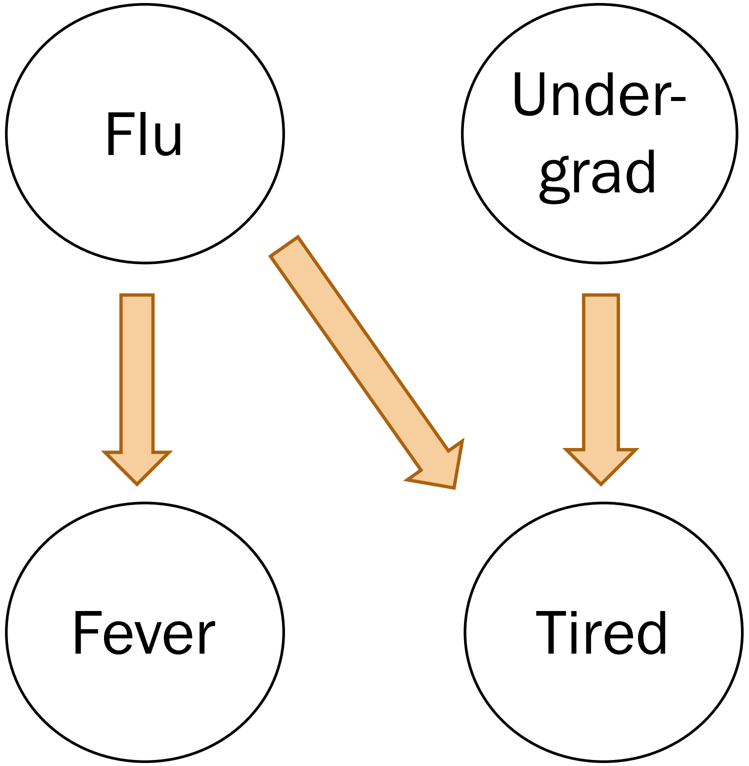
Great! Just specify $2^4 = 16$ joint probabilities…?

$$P(F_{lu} = a, F_{ev} = b, U = c, T = d)$$

What would a Stanford flu expert do?

Describe the joint distribution using causality!!!

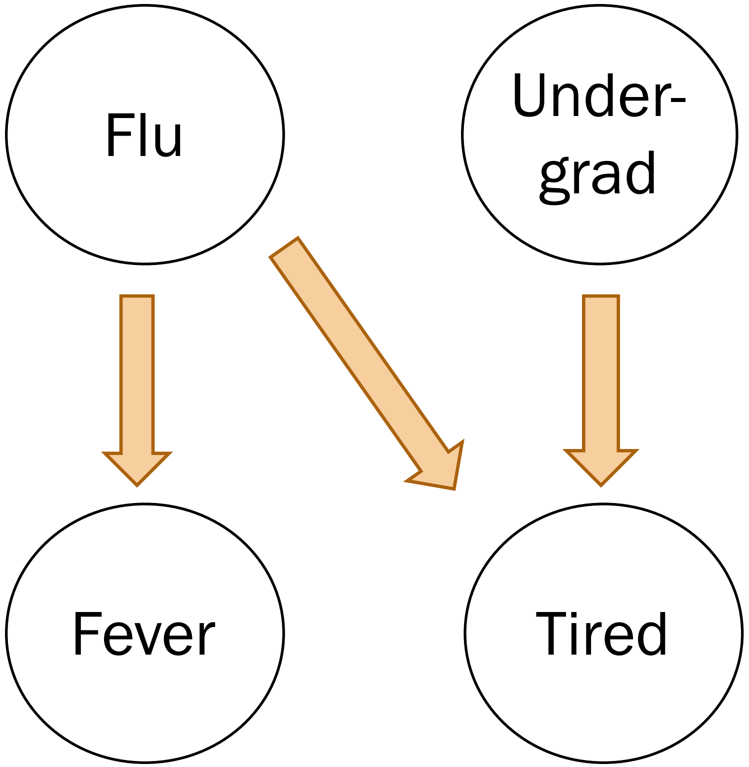
Constructing a Bayesian Network



What would a Stanford flu expert do?

1. Describe the joint distribution using causality.
2. **<u>Assume conditional independence.</u>**

Constructing a Bayesian Network



In a Bayesian Network,

Each random variable is conditionally independent of its non-descendants, given its parents.

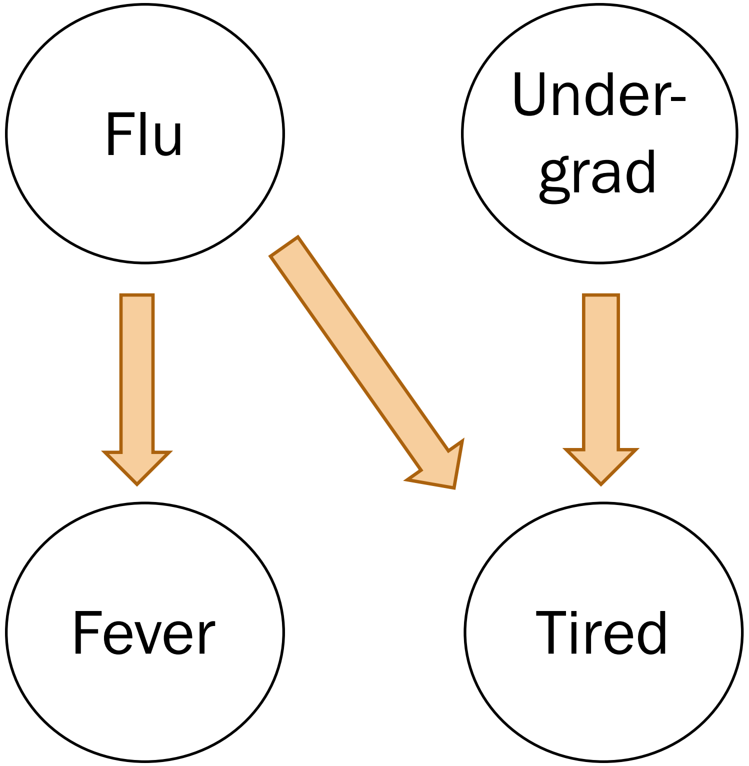
- Node: random variable
- Directed edge: conditional dependency

Examples:

- $P(F_{ev} = 1 | T = 0, F_{lu} = 1) = P(F_{ev} = 1 | F_{lu} = 1)$
- $P(F_{lu} = 1, U = 0) = P(F_{lu} = 1)P(U = 0)$

Constructing a Bayesian Network

$P(F_{lu} = 1) = 0.1$ $\quad$ $P(U = 1) = 0.8$



$P(F_{ev} = 1 | F_{lu} = 1) = 0.9$
$P(F_{ev} = 1 | F_{lu} = 0) = 0.05$

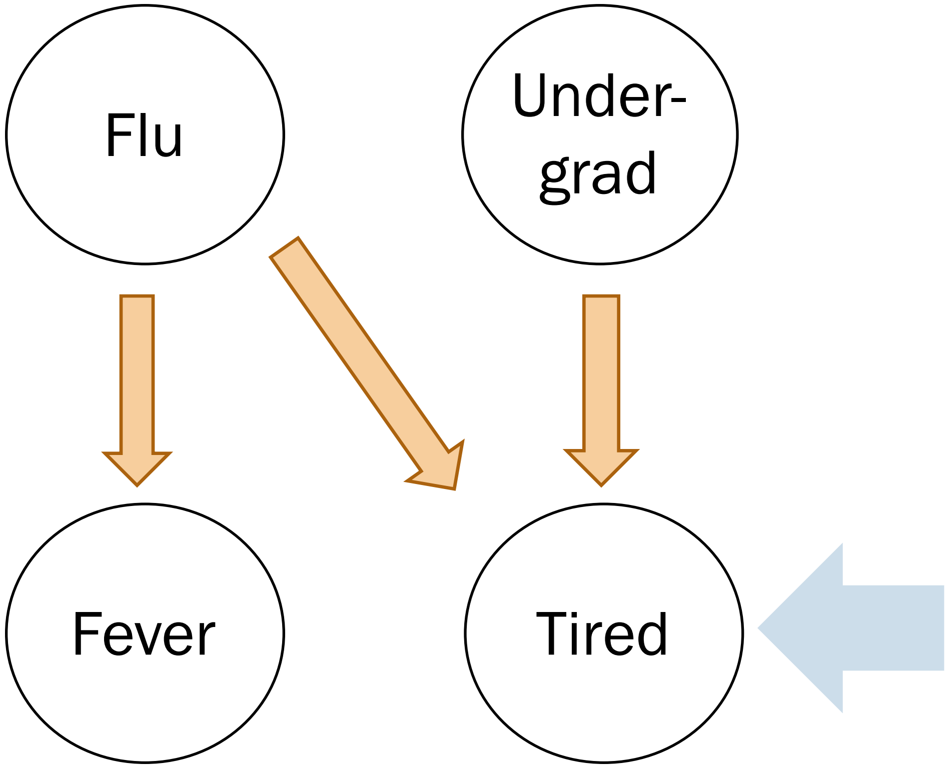
What would a Stanford flu expert do?

1. Describe the joint distribution using causality.
2. ✅ Assume conditional independence.
3. Provide $P(\text{values}|\text{parents})$ for each random variable

What conditional probabilities should our expert specify?

Constructing a Bayesian Network

$P(F_{lu} = 1) = 0.1$ $\quad$ $P(U = 1) = 0.8$



$P(F_{ev} = 1 | F_{lu} = 1) = 0.9$
$P(F_{ev} = 1 | F_{lu} = 0) = 0.05$

What would a Stanford flu expert do?

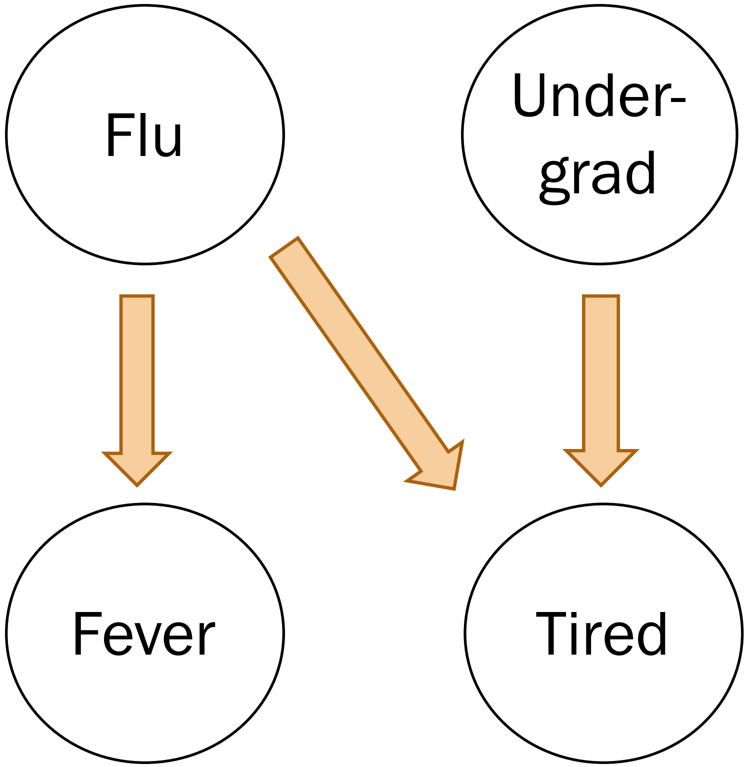
1. Describe the joint distribution using causality.
2. Assume conditional independence.
3. Provide $P(\text{values}|\text{parents})$ for each random variable

What conditional probabilities should our expert specify?

$P(T = 1 | F_{lu} = 0, U = 0)$
$P(T = 1 | F_{lu} = 0, U = 1)$
$P(T = 1 | F_{lu} = 1, U = 0)$
$P(T = 1 | F_{lu} = 1, U = 1)$

Using a Bayes Net

$P(F_{lu} = 1) = 0.1$ $\quad$ $P(U = 1) = 0.8$



$P(F_{ev} = 1 | F_{lu} = 1) = 0.9$
$P(F_{ev} = 1 | F_{lu} = 0) = 0.05$

$P(T = 1 | F_{lu} = 0, U = 0) = 0.1$
$P(T = 1 | F_{lu} = 0, U = 1) = 0.8$
$P(T = 1 | F_{lu} = 1, U = 0) = 0.9$
$P(T = 1 | F_{lu} = 1, U = 1) = 1.0$

What would a CS109 student do?

1. Populate a Bayesian network by asking a Stanford flu expert
   or
   by using reasonable assumptions

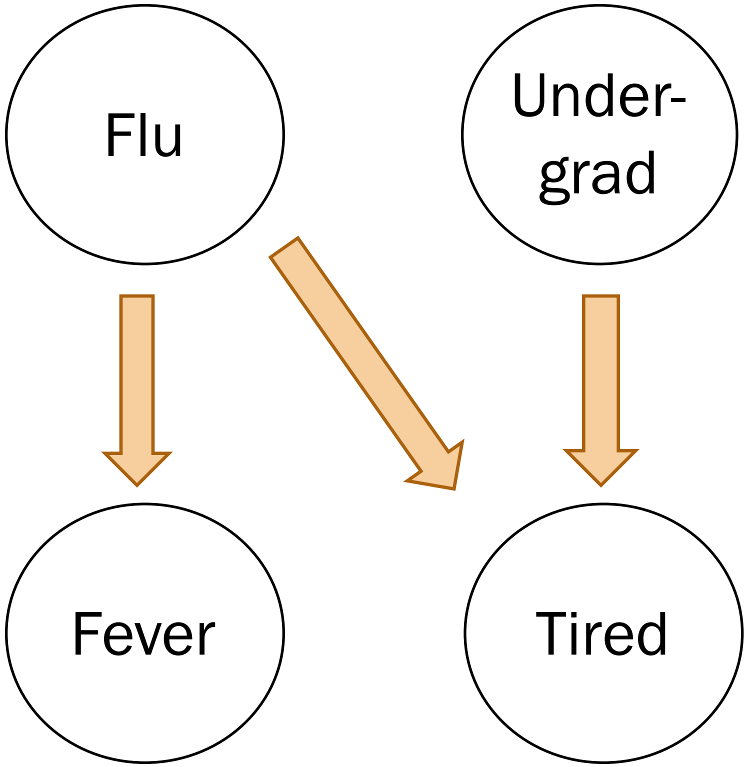
2. **Answer inference questions**

Our focus today

Inference (I): Math

Bayes Nets: Conditional independence

Review



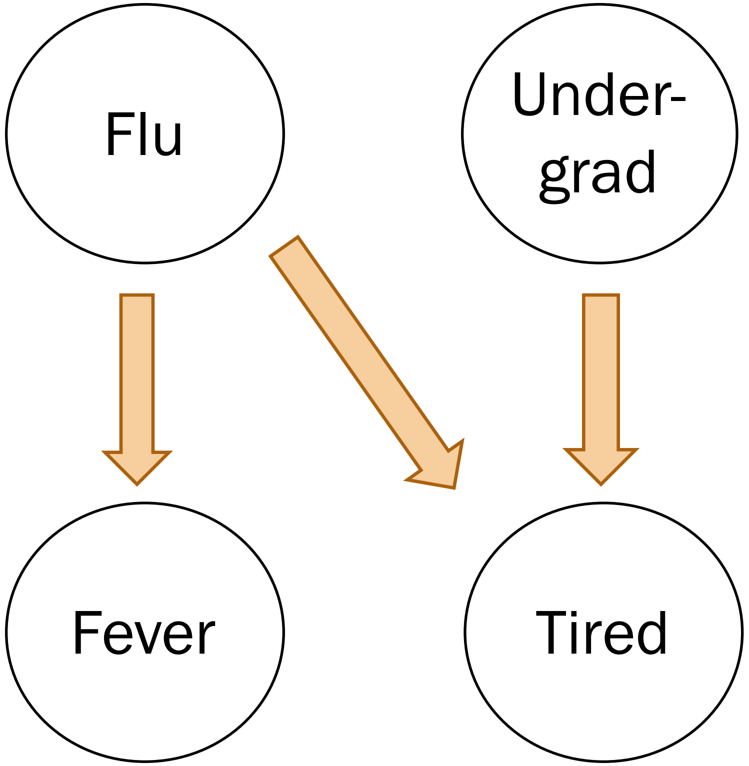
In a Bayesian Network,

Each random variable is conditionally independent of its non-descendants, given its parents.

- Node: random variable
- Directed edge: conditional dependency

Inference via math

$P(F_{lu} = 1) = 0.1$ $P(U = 1) = 0.8$



$P(F_{ev} = 1|F_{lu} = 1) = 0.9$
$P(F_{ev} = 1|F_{lu} = 0) = 0.05$

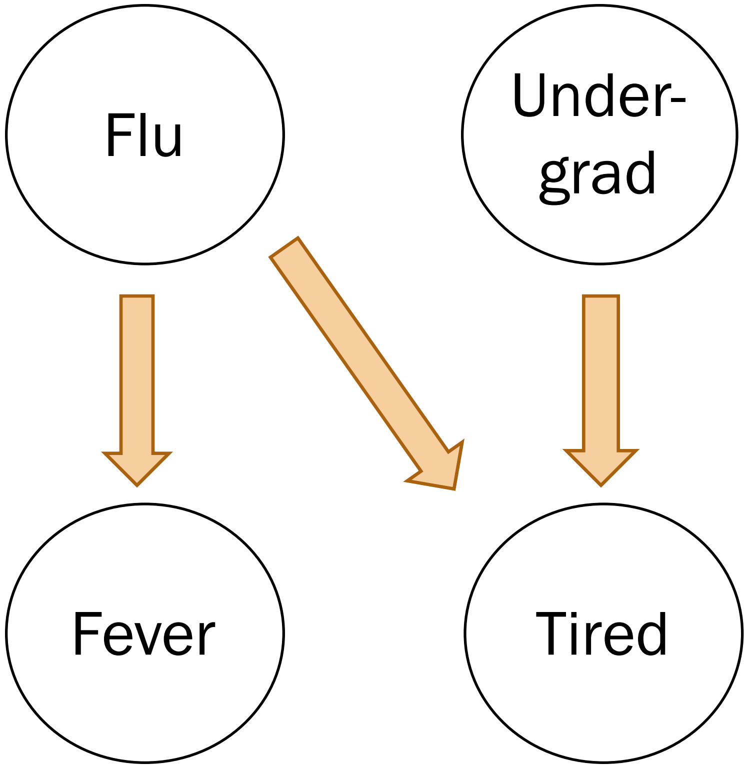
$P(T = 1|F_{lu} = 0, U = 0) = 0.1$
$P(T = 1|F_{lu} = 0, U = 1) = 0.8$
$P(T = 1|F_{lu} = 1, U = 0) = 0.9$
$P(T = 1|F_{lu} = 1, U = 1) = 1.0$

1. $P(F_{lu} = 0, U = 1, F_{ev} = 0, T = 1)$?

   Compute joint probabilities using chain rule.

Inference via math

$P(F_{lu} = 1) = 0.1$ $\quad$ $P(U = 1) = 0.8$



$P(F_{ev} = 1 | F_{lu} = 1) = 0.9$
$P(F_{ev} = 1 | F_{lu} = 0) = 0.05$

$P(T = 1 | F_{lu} = 0, U = 0) = 0.1$
$P(T = 1 | F_{lu} = 0, U = 1) = 0.8$
$P(T = 1 | F_{lu} = 1, U = 0) = 0.9$
$P(T = 1 | F_{lu} = 1, U = 1) = 1.0$

2. $P(F_{lu} = 1 | F_{ev} = 0, U = 0, T = 1)$?

1. Compute joint probabilities

$$P(F_{lu} = 1, F_{ev} = 0, U = 0, T = 1)$$
$$P(F_{lu} = 0, F_{ev} = 0, U = 0, T = 1)$$

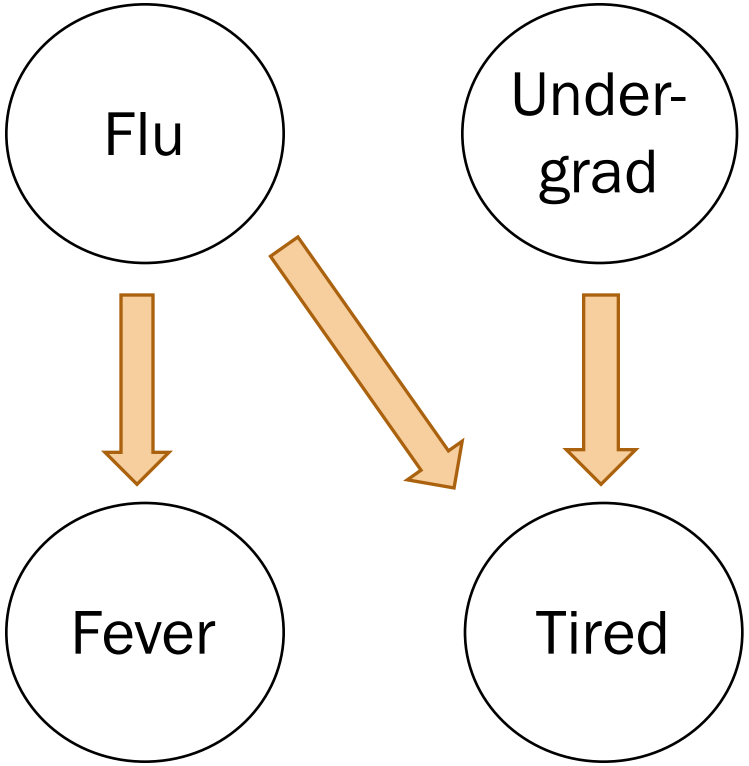
2. Definition of conditional probability

$$\frac{P(F_{lu} = 1, F_{ev} = 0, U = 0, T = 1)}{\sum_x P(F_{lu} = x, F_{ev} = 0, U = 0, T = 1)}$$

$$= 0.095$$

Inference via math

$P(F_{lu} = 1) = 0.1$ $\quad$ $P(U = 1) = 0.8$



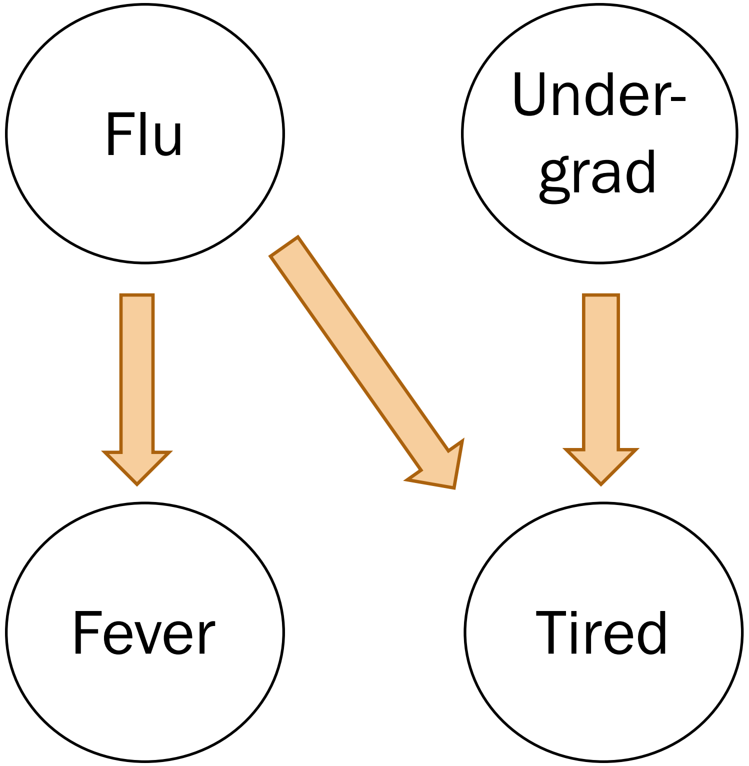
$P(F_{ev} = 1 | F_{lu} = 1) = 0.9$
$P(F_{ev} = 1 | F_{lu} = 0) = 0.05$

$P(T = 1 | F_{lu} = 0, U = 0) = 0.1$
$P(T = 1 | F_{lu} = 0, U = 1) = 0.8$
$P(T = 1 | F_{lu} = 1, U = 0) = 0.9$
$P(T = 1 | F_{lu} = 1, U = 1) = 1.0$

3. $P(F_{lu} = 1 | U = 1, T = 1)$?

Inference via math

$P(F_{lu} = 1) = 0.1$ $\quad$ $P(U = 1) = 0.8$



$P(F_{ev} = 1|F_{lu} = 1) = 0.9$
$P(F_{ev} = 1|F_{lu} = 0) = 0.05$

$P(T = 1|F_{lu} = 0, U = 0) = 0.1$
$P(T = 1|F_{lu} = 0, U = 1) = 0.8$
$P(T = 1|F_{lu} = 1, U = 0) = 0.9$
$P(T = 1|F_{lu} = 1, U = 1) = 1.0$

3. $P(F_{lu} = 1|U = 1, T = 1)$?
   1. Compute joint probabilities

      $$P(F_{lu} = 1, U = 1, F_{ev} = 1, T = 1)$$

      ...

      $$P(F_{lu} = 0, U = 1, F_{ev} = 0, T = 1)?$$

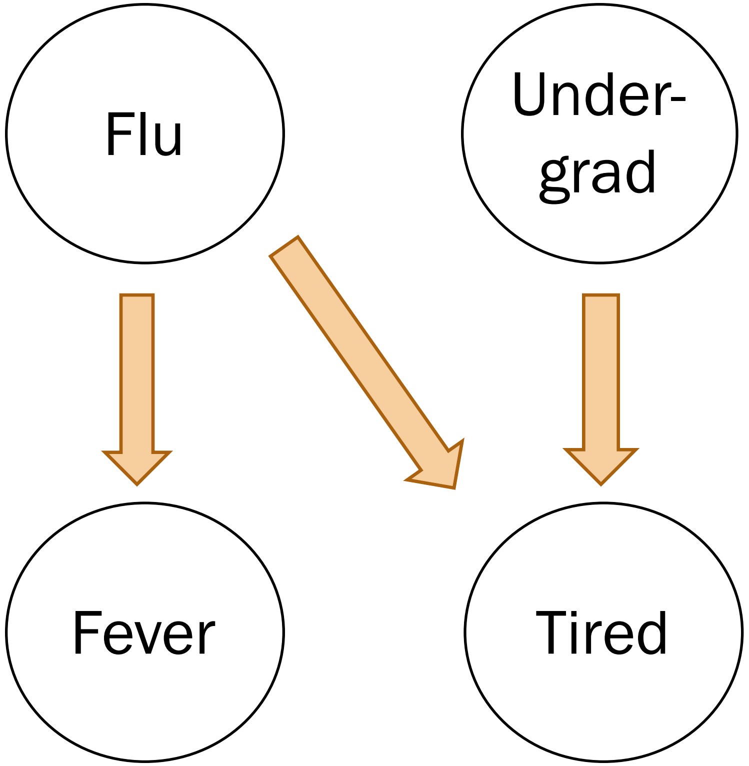
   2. Definition of conditional probability

      $$\frac{\sum_y P(F_{lu} = 1, U = 1, F_{ev} = y, T = 1)}{\sum_x \sum_y P(F_{lu} = x, U = 1, F_{ev} = y, T = 1)}$$

$$= 0.122$$

Inference via math

$P(F_{lu} = 1) = 0.1$ $P(U = 1) = 0.8$



$P(F_{ev} = 1|F_{lu} = 1) = 0.9$
$P(F_{ev} = 1|F_{lu} = 0) = 0.05$

$P(T = 1|F_{lu} = 0, U = 0) = 0.1$
$P(T = 1|F_{lu} = 0, U = 1) = 0.8$
$P(T = 1|F_{lu} = 1, U = 0) = 0.9$
$P(T = 1|F_{lu} = 1, U = 1) = 1.0$

Solving inference questions precisely is possible, but sometimes tedious.

Can we use sampling to do approximate inference?

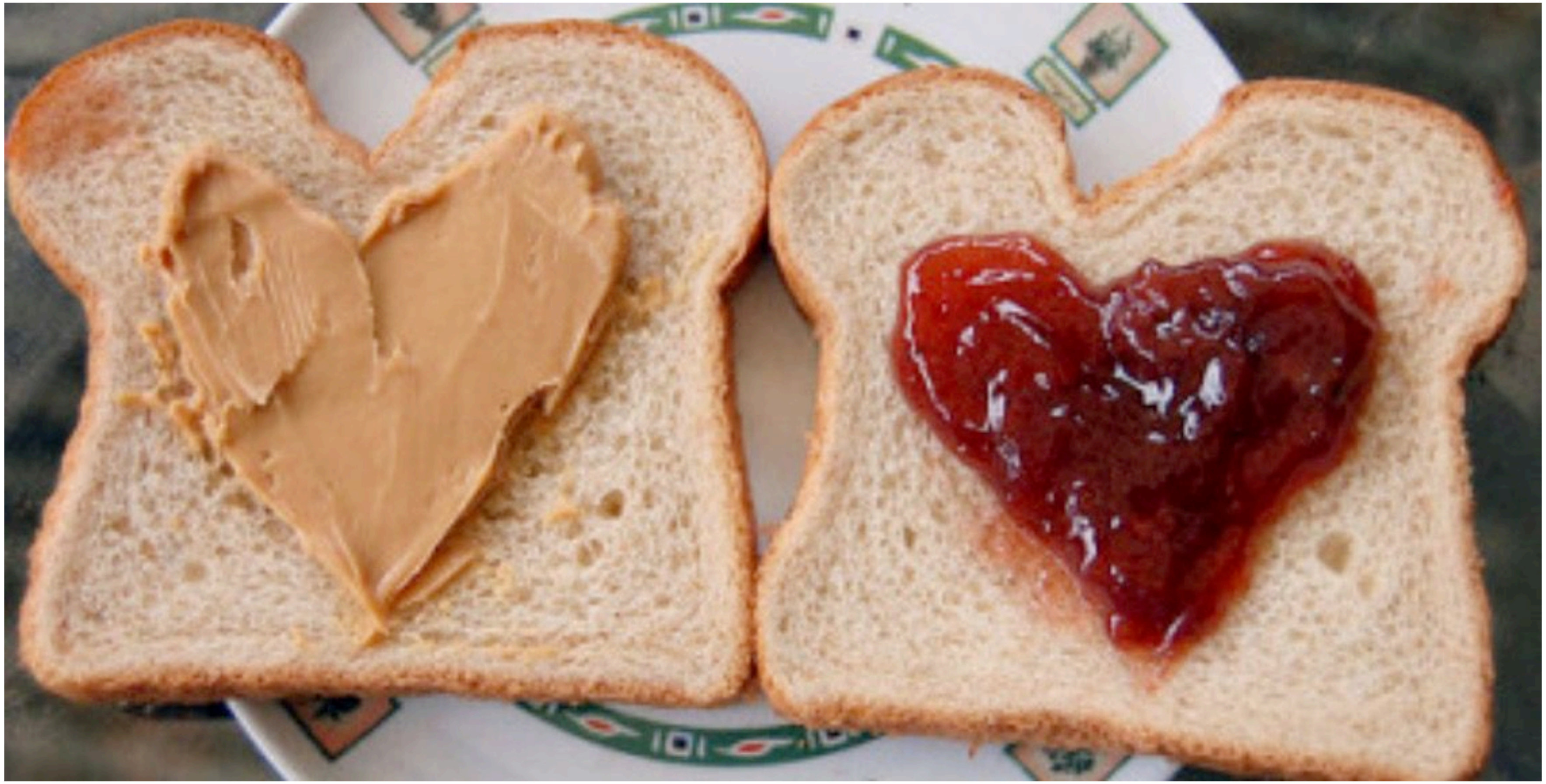
Yes.

(live) Conditional Expectation + General Inference

Lisa Yan

July 22, 2020

Conditional Expectation



Conditional Distributions   Expectation

Breakout Rooms

Check out the question on the next slide (Slide 28). Post any clarifications here!

https://us.edstem.org/courses/667/discussion/93799

Breakout rooms: 4 min. Introduce yourself!

Quick check

A. value
B. random variable, function of Y
C. random variable, function of X
D. function of X and Y
E. doesn't make sense

1. $E[X]$
2. $E[X, Y]$
3. $E[X + Y]$
4. $E[X|Y]$
5. $E[X|Y = 6]$
6. $E[X = 1]$

Quick check

A. value
B. random variable, function of Y
C. random variable, function of X
D. function of X and Y
E. doesn't make sense

1. $E[X]$
2. $E[X, Y]$
3. $E[X + Y]$
4. $E[X|Y]$
5. $E[X|Y = 6]$
6. $E[X = 1]$

Conditional Expectation

The conditional expectation of X given $Y = y$ is

$$E[X|Y = y] = \sum_x xP(X = x|Y = y) = \sum_x xp_{X|Y}(x|y)$$

- Interpret: $E[X|Y]$ is a random variable that takes on the value $E[X|Y = y]$ with probability $P(Y = y)$

The **Law of Total Expectation** states that

$$E\big[E[X|Y]\big] = \sum_y E[X|Y = y]P(Y = y) = E[X]$$

- Apply: $E[X]$ can be calculated as the expectation of $E[X|Y]$

Think

Slide 34 has a question to go over by yourself.

Post any clarifications here!

https://us.edstem.org/courses/667/discussion/93799

Think by yourself: 2 min

(by yourself)

Analyzing recursive code

$$E[X] = E\big[E[X|Y]\big] = \sum_y E[X|Y=y]P(Y=y)$$

```
def recurse():
  # equally likely values 1,2,3
  x = np.random.choice([1,2,3])
  if (x == 1): return 3
  elif (x == 2): return (5 + recurse())
  else: return (7 + recurse())
```

Let Y = return value of `recurse()`.
What is $E[Y]$?

Analyzing recursive code

$$E[X] = E\big[E[X|Y]\big] = \sum_y E[X|Y=y]P(Y=y)$$

```
def recurse():
  # equally likely values 1,2,3
  x = np.random.choice([1,2,3])
  if (x == 1): return 3
  elif (x == 2): return (5 + recurse())
  else: return (7 + recurse())
```

Let Y = return value of `recurse()`.

What is $E[Y]$?

$$E[Y] = \boxed{E[Y|X=1]}P(X=1) + E[Y|X=2]P(X=2) + E[Y|X=3]P(X=3)$$

$E[Y|X=1] = 3$

When $X = 1$, return 3.

Analyzing recursive code

If Y discrete
$$E[X] = E\big[E[X|Y]\big] = \sum_y E[X|Y=y]P(Y=y)$$

```python
def recurse():
  # equally likely values 1,2,3
  x = np.random.choice([1,2,3])
  if (x == 1): return 3
  elif (x == 2): return (5 + recurse())
  else: return (7 + recurse())
```

Let Y = return value of `recurse()`.
What is $E[Y]$?

$$E[Y] = E[Y|X=1]P(X=1) + E[Y|X=2]P(X=2) + E[Y|X=3]P(X=3)$$

$$E[Y|X=1] = 3$$

What is $E[Y|X=2]$?

A. $E[5] + Y$
B. $E[Y+5] = 5 + E[Y]$
C. $5 + E[Y|X=2]$

(by yourself)

Analyzing recursive code

If Y discrete
$$E[X] = E\big[E[X|Y]\big] = \sum_y E[X|Y=y]P(Y=y)$$

```
def recurse():
  # equally likely values 1,2,3
  x = np.random.choice([1,2,3])
  if (x == 1): return 3
  elif (x == 2): return (5 + recurse())
  else: return (7 + recurse())
```

Let Y = return value of `recurse()`.
What is $E[Y]$?

$$E[Y] = E[Y|X=1]P(X=1) + E[Y|X=2]P(X=2) + E[Y|X=3]P(X=3)$$

$E[Y|X=1] = 3$

When $X = 2$, return 5 +
a future return value of `recurse()`.

What is $E[Y|X=2]$?

A. $E[5] + Y$
B. $E[Y+5] = 5 + E[Y]$
C. $5 + E[Y|X=2]$

Analyzing recursive code

If Y discrete

$$E[X] = E\big[E[X|Y]\big] = \sum_y E[X|Y = y]P(Y = y)$$

```python
def recurse():
  # equally likely values 1,2,3
  x = np.random.choice([1,2,3])
  if (x == 1): return 3
  elif (x == 2): return (5 + recurse())
  else: return (7 + recurse())
```

Let Y = return value of `recurse()`.
What is $E[Y]$?

$$E[Y] = E[Y|X = 1]P(X = 1) + E[Y|X = 2]P(X = 2) + E[Y|X = 3]P(X = 3)$$

$E[Y|X = 1] = 3$

$E[Y|X = 2] = E[5 + Y]$

When $X = 3$, return $7 +$ a future return value of `recurse()`.

$$E[Y|X = 3] = E[7 + Y]$$

Analyzing recursive code

If Y discrete

$$E[X] = E\big[E[X|Y]\big] = \sum_y E[X|Y=y]P(Y=y)$$

```python
def recurse():
  # equally likely values 1,2,3
  x = np.random.choice([1,2,3])
  if (x == 1): return 3
  elif (x == 2): return (5 + recurse())
  else: return (7 + recurse())
```

Let Y = return value of `recurse()`.
What is $E[Y]$?

$$E[Y] = E[Y|X=1]P(X=1) + E[Y|X=2]P(X=2) + E[Y|X=3]P(X=3)$$

$$E[Y|X=1] = 3 \qquad E[Y|X=2] = E[5+Y] \qquad E[Y|X=3] = E[7+Y]$$

$$E[Y] = 3(1/3) + (5+E[Y])(1/3) + (7+E[Y])(1/3)$$

$$E[Y] = (1/3)(15 + 2E[Y]) = 5 + (2/3)E[Y]$$

$$E[Y] = 15$$

On your own: What is $\mathrm{Var}(Y)$?

Independent RVs, defined another way

If X and Y are independent discrete random variables, then $\forall x, y$:

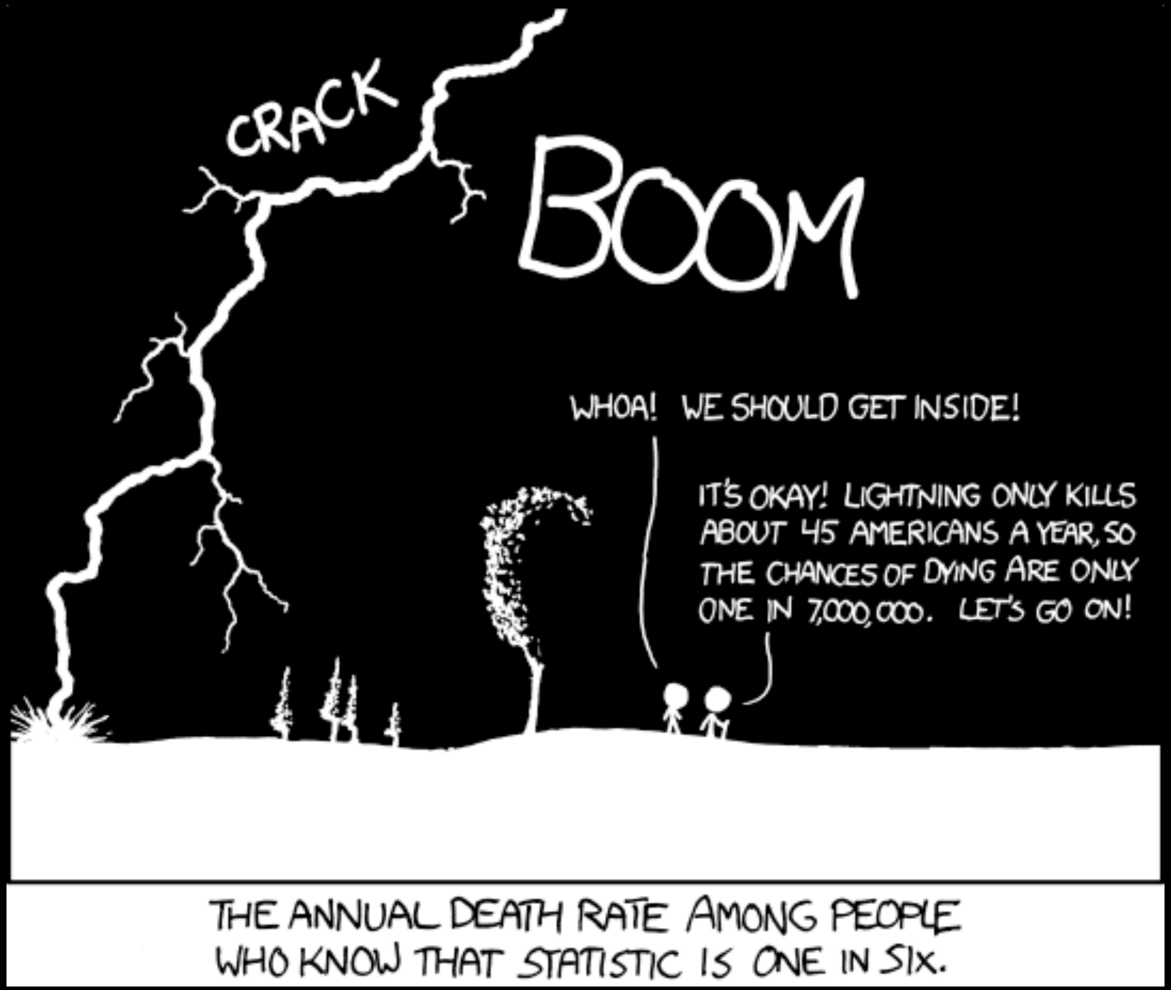
$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{P(X = x)P(Y = y)}{P(Y = y)} = P(X = x)$$

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x, y)}{p_Y(y)} = \frac{p_X(x)p_Y(y)}{p_Y(y)} = p_X(x)$$

Note for conditional expectation, independent X and Y implies

$$E[X|Y = y] = \sum_x x p_{X|Y}(x|y) = \sum_x x p_X(x) = E[X]$$

Interlude for jokes/announcements



https://xkcd.com/795/

Interesting probability news

U.S. Recession Model at 100% Confirms Downturn Is Already Here

"Bloomberg Economics created a model last year to determine America's recession odds."

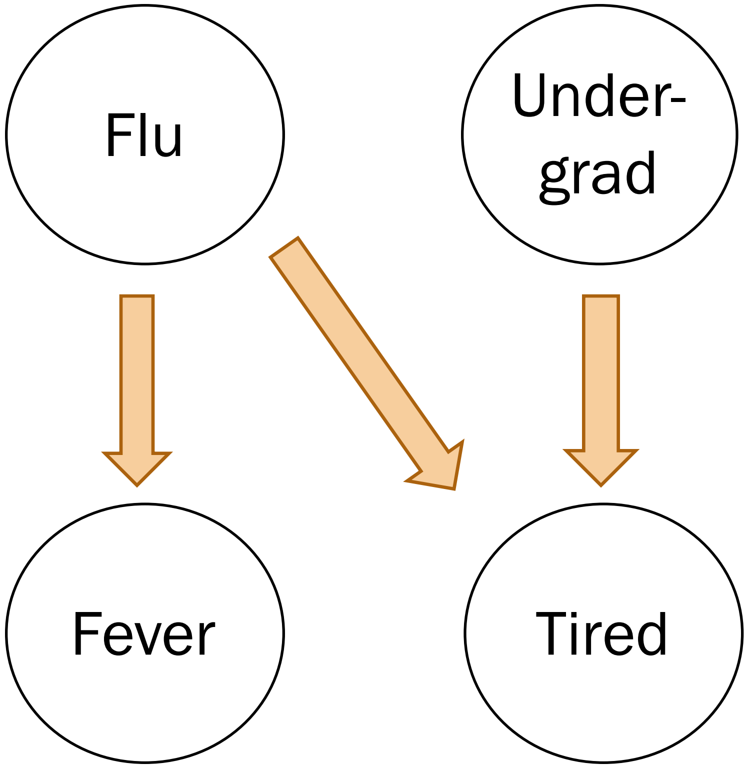
- I encourage you to read through and understand the parameters used to define this model!

100%

Chance of Recession Within 12 Months

https://www.bloomberg.com/graphics/us-economic-recession-tracker/

Constructing a Bayesian Network



In a Bayesian Network,

Each random variable is conditionally independent of its non-descendants, given its parents.

- Node: random variable
- Directed edge: conditional dependency

Examples:

- $P(F_{ev} = 1 | T = 0, F_{lu} = 1) = P(F_{ev} = 1 | F_{lu} = 1)$
- $P(F_{lu} = 1, U = 0) = P(F_{lu} = 1)P(U = 0)$

Breakout Rooms

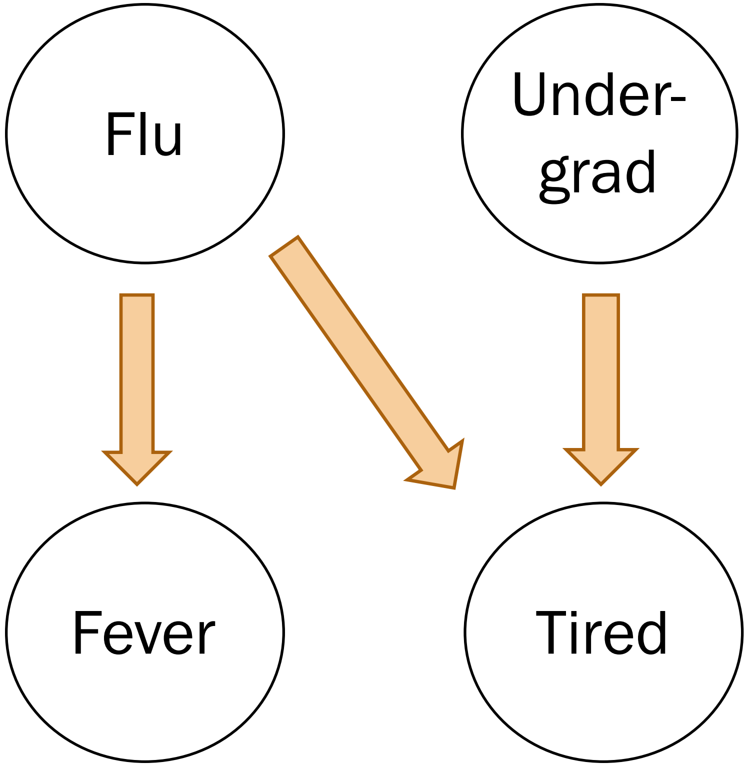
Check out the question on the next slide.
Post any clarifications here!

https://us.edstem.org/courses/667/discussion/93799

Breakout rooms: 4 min. Introduce yourself!

Inference via math

$P(F_{lu} = 1) = 0.1$ $\quad$ $P(U = 1) = 0.8$



What is $P(F_{lu} = 1 | U = 1, T = 1)$?

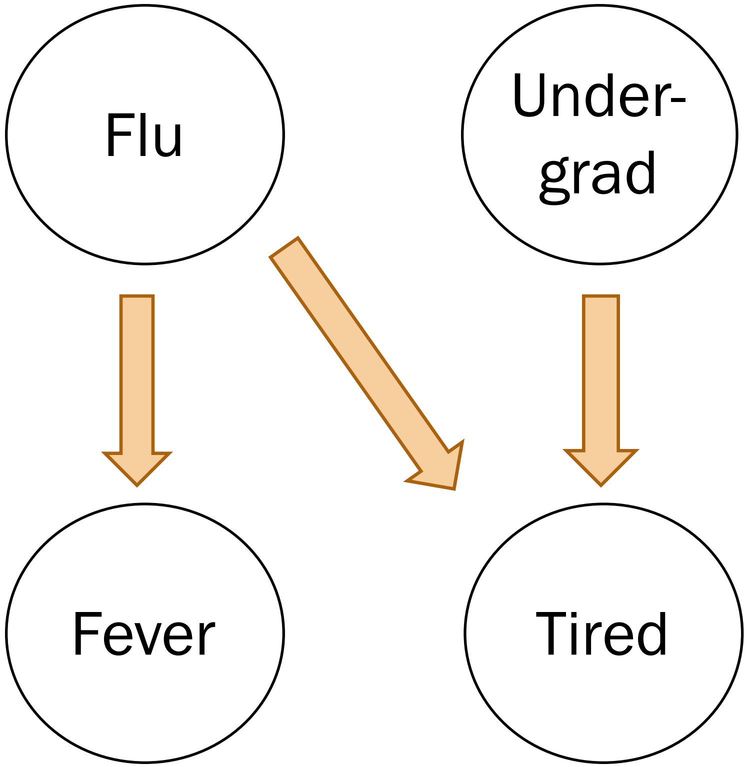
$$= 0.122$$

$P(F_{ev} = 1 | F_{lu} = 1) = 0.9$
$P(F_{ev} = 1 | F_{lu} = 0) = 0.05$

$P(T = 1 | F_{lu} = 0, U = 0) = 0.1$
$P(T = 1 | F_{lu} = 0, U = 1) = 0.8$
$P(T = 1 | F_{lu} = 1, U = 0) = 0.9$
$P(T = 1 | F_{lu} = 1, U = 1) = 1.0$

Inference via math

$P(F_{lu} = 1) = 0.1$ $\quad$ $P(U = 1) = 0.8$



$$P(F_{ev} = 1 | F_{lu} = 1) = 0.9$$
$$P(F_{ev} = 1 | F_{lu} = 0) = 0.05$$

$$P(T = 1 | F_{lu} = 0, U = 0) = 0.1$$
$$P(T = 1 | F_{lu} = 0, U = 1) = 0.8$$
$$P(T = 1 | F_{lu} = 1, U = 0) = 0.9$$
$$P(T = 1 | F_{lu} = 1, U = 1) = 1.0$$

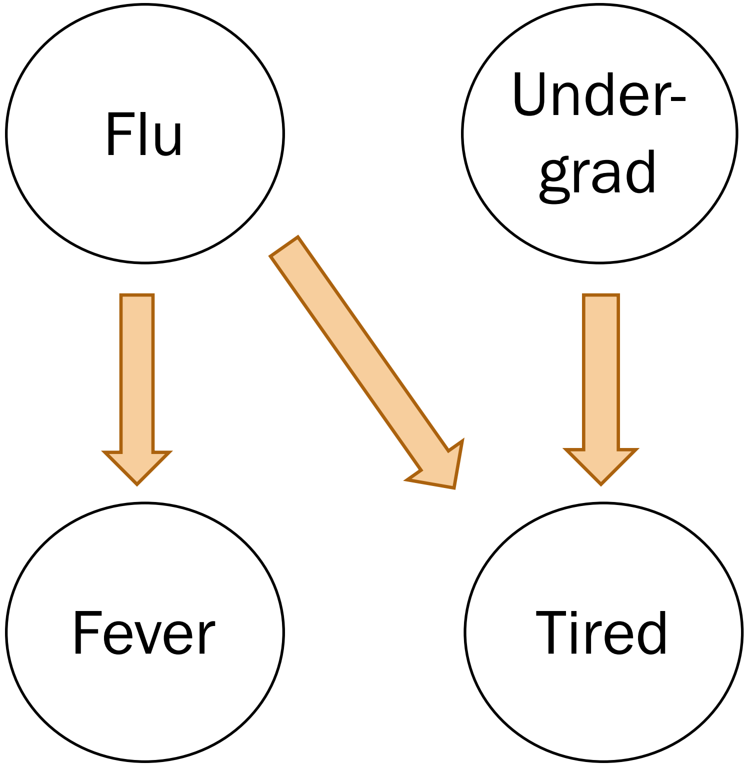
Solving inference questions precisely is possible, but sometimes tedious.

Can we use sampling to do approximate inference?

Yes.

Inference via math

$P(F_{lu} = 1) = 0.1$ $\quad$ $P(U = 1) = 0.8$



What is $P(F_{lu} = 1 | U = 1, T = 1)$?

$= 0.122$

(from pre-lecture video)

$P(F_{ev} = 1 | F_{lu} = 1) = 0.9$
$P(F_{ev} = 1 | F_{lu} = 0) = 0.05$

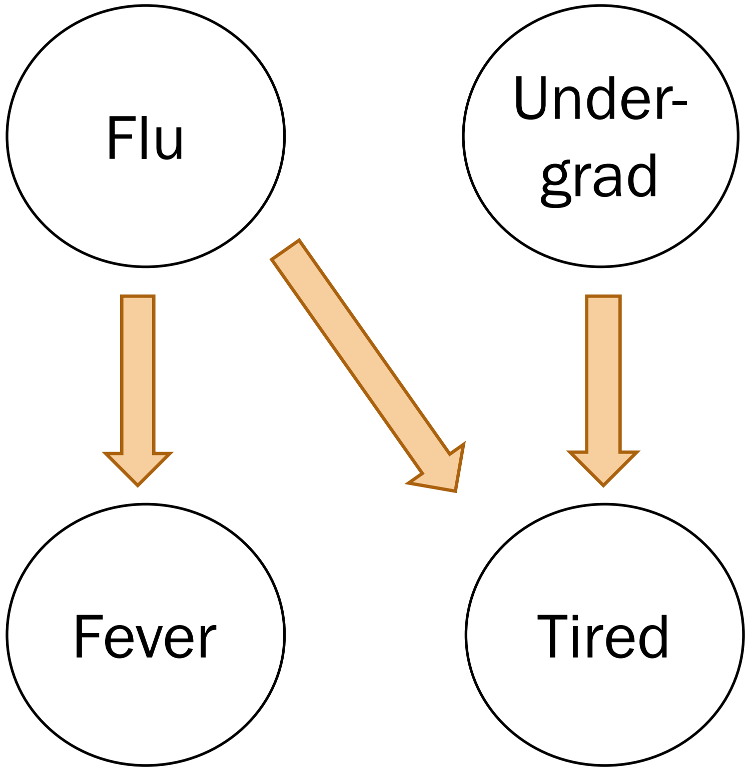
$P(T = 1 | F_{lu} = 0, U = 0) = 0.1$
$P(T = 1 | F_{lu} = 0, U = 1) = 0.8$
$P(T = 1 | F_{lu} = 1, U = 0) = 0.9$
$P(T = 1 | F_{lu} = 1, U = 1) = 1.0$

Rejection sampling algorithm

Step 0:

Have a fully specified Bayesian Network

$P(F_{lu} = 1) = 0.1$ $\quad$ $P(U = 1) = 0.8$



$P(F_{ev} = 1 | F_{lu} = 1) = 0.9$
$P(F_{ev} = 1 | F_{lu} = 0) = 0.05$

$P(T = 1 | F_{lu} = 0, U = 0) = 0.1$
$P(T = 1 | F_{lu} = 0, U = 1) = 0.8$
$P(T = 1 | F_{lu} = 1, U = 0) = 0.9$
$P(T = 1 | F_{lu} = 1, U = 1) = 1.0$

Rejection sampling algorithm

Inference question: What is $P(F_{lu} = 1 | U = 1, T = 1)$?

```
def rejection_sampling(event, observation):
  samples = sample_a_ton()
  samples_observation = …
        # number of samples with (U = 1, T = 1)
  samples_event =
        # number of samples with (F_lu = 1, U = 1, T
  return len(samples_event)/len(samples_observ
```

[flu, und, fev, tir]

```
Sampling...
[0, 1, 0, 1]
[0, 1, 0, 1]
[0, 1, 0, 1]
[0, 0, 0, 0]
[0, 1, 0, 1]
[0, 1, 1, 1]
[0, 1, 0, 0]
[1, 1, 1, 1]
[0, 0, 1, 1]
...
[0, 1, 0, 1]
Finished sampling
```

Rejection sampling algorithm

Inference question: What is $P(F_{lu} = 1 | U = 1, T = 1)$?

```
def rejection_sampling(event, observation):
  samples = sample_a_ton()
  samples_observation = …
        # number of samples with (U = 1, T = 1)
  samples_event =
        # number of samples with (F_lu = 1, U = 1, T = 1)
  return len(samples_event)/len(samples_observation)
```

$$\text{Approximate Probability} = \frac{\#\text{ samples with } (F_{lu} = 1, U = 1, T = 1)}{\#\text{ samples with } (U = 1, T = 1)}$$

Rejection sampling algorithm

Inference question: What is $P(F_{lu} = 1 | U = 1, T = 1)$?

Approximate Probability = $\dfrac{\text{\# samples with } (F_{lu} = 1, U = 1, T = 1)}{\text{\# samples with } (U = 1, T = 1)}$

Why would this definition of approximate probability make sense?

Think

Slide 40 has a question to go over by yourself.

Post any clarifications here!

https://us.edstem.org/courses/667/discussion/93799

Think by yourself: 2 min

(by yourself)

Why would this approximate probability make sense?

Inference question: What is $P(F_{lu} = 1 | U = 1, T = 1)$?

Approximate Probability = $\dfrac{\text{\# samples with } (F_{lu} = 1, U = 1, T = 1)}{\text{\# samples with } (U = 1, T = 1)}$

Recall our definition of probability as a frequency:

$$P(E) = \lim_{n \to \infty} \frac{n(E)}{n}$$

n = # of total trials

$n(E)$ = # trials where E occurs

Why would this approximate probability make sense?

Inference question: What is $P(F_{lu} = 1 | U = 1, T = 1)$?

Approximate Probability = $\dfrac{\text{\# samples with } (F_{lu} = 1, U = 1, T = 1)}{\text{\# samples with } (U = 1, T = 1)}$

Recall our definition of probability as a frequency:

$$P(E) = \lim_{n \to \infty} \frac{n(E)}{n}$$

n = # of total trials

$n(E)$ = # trials where E occurs

Rejection sampling algorithm

Inference question: What is $P(F_{lu} = 1 | U = 1, T = 1)$?

```
def rejection_sampling(event, observation):
  samples = sample_a_ton()
  samples_observation = …
        # number of samples with (U = 1, T = 1)
  samples_event =
        # number of samples with (F_lu = 1, U = 1, T
  return len(samples_event)/len(samples_observ
```

[flu, und, fev, tir]

```
Sampling...
[0, 1, 0, 1]
[0, 1, 0, 1]
[0, 1, 0, 1]
[0, 0, 0, 0]
[0, 1, 0, 1]
[0, 1, 1, 1]
[0, 1, 0, 0]
[1, 1, 1, 1]
[0, 0, 1, 1]
...
[0, 1, 0, 1]
Finished sampling
```

Rejection sampling algorithm

```
N_SAMPLES = 100000
# Method: Sample a ton
# -------------------
# create N_SAMPLES with likelihood proportional
# to the joint distribution
def sample_a_ton():
    samples = []
    for i in range(N_SAMPLES):
        sample = make_sample() # a particle
        samples.append(sample)
    return samples
```

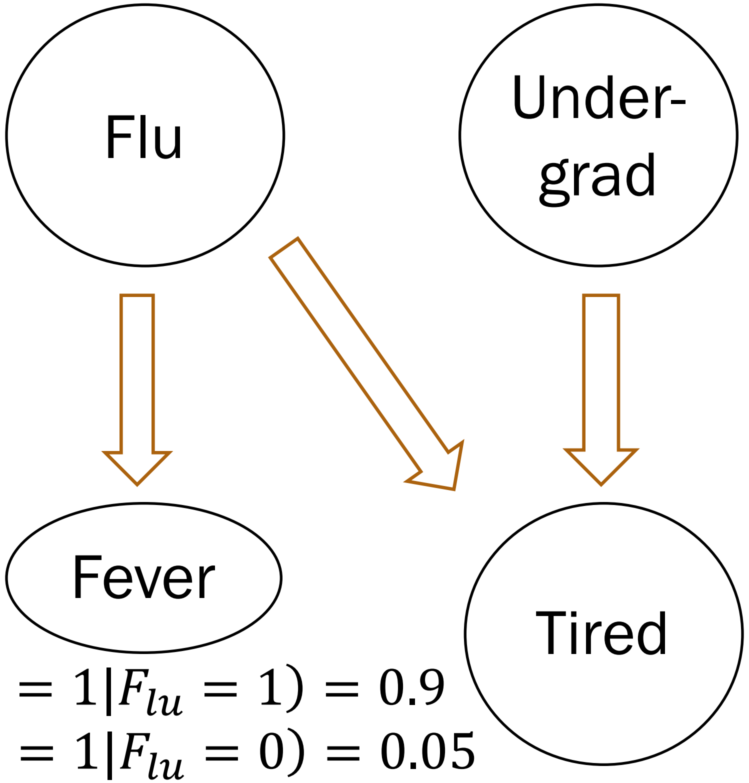
How do we make a sample $(F_{lu} = a, U = b, F_{ev} = c, T = d)$ according to the joint probability?

Create a sample using the Bayesian Network!!

Rejection sampling algorithm

```python
# Method: Make Sample
# -------------------
# create a single sample from the joint distribution
# based on the medical "WebMD" Bayesian Network
def make_sample():
    # prior on causal factors
    flu = bernoulli(0.1)
    und = bernoulli(0.8)

    # choose fever based on flu
    if flu == 1: fev = bernoulli(0.9)
    else:        fev = bernoulli(0.05)

    # choose tired based on (undergrad and flu)
    #
    # TODO: fill in
    #
    #

    # a sample from the joint has an
    # assignment to *all* random variables
    return [flu, und, fev, tir]
```

$P(F_{lu} = 1) = 0.1$ $\quad$ $P(U = 1) = 0.8$



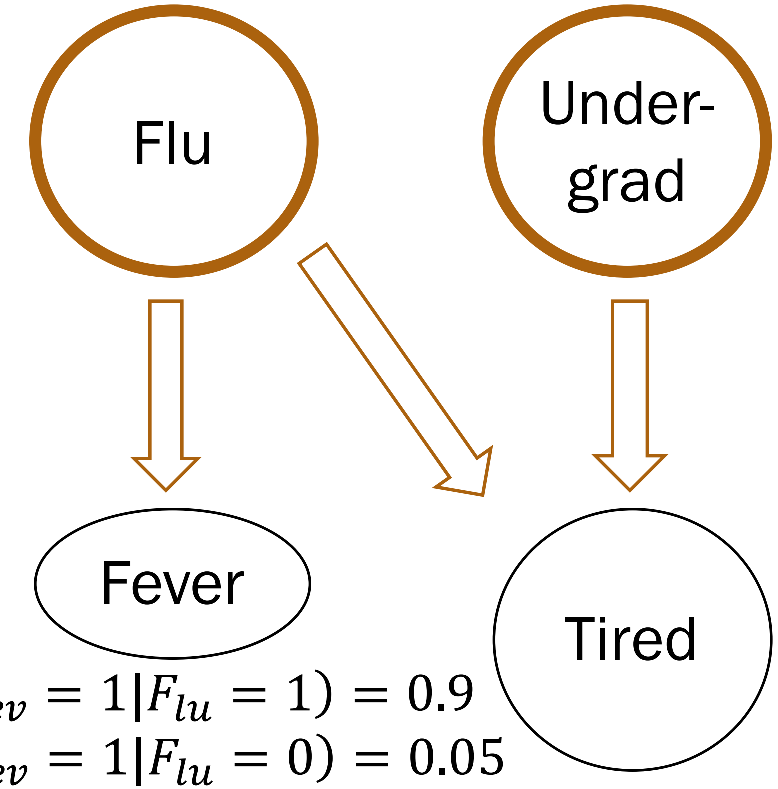
$$P(F_{ev} = 1 | F_{lu} = 1) = 0.9$$
$$P(F_{ev} = 1 | F_{lu} = 0) = 0.05$$

$$P(T = 1 | F_{lu} = 0, U = 0) = 0.1$$
$$P(T = 1 | F_{lu} = 0, U = 1) = 0.8$$
$$P(T = 1 | F_{lu} = 1, U = 0) = 0.9$$
$$P(T = 1 | F_{lu} = 1, U = 1) = 1.0$$

Rejection sampling algorithm

```
# Method: Make Sample
# -------------------
# create a single sample from the joint distribution
# based on the medical "WebMD" Bayesian Network
def make_sample():
    # prior on causal factors
    flu = bernoulli(0.1)
    und = bernoulli(0.8)

    # choose fever based on flu
    if flu == 1: fev = bernoulli(0.9)
    else:        fev = bernoulli(0.05)

    # choose tired based on (undergrad and flu)
    #
    # TODO: fill in
    #
    #

    # a sample from the joint has an
    # assignment to *all* random variables
    return [flu, und, fev, tir]
```

$P(F_{lu} = 1) = 0.1$ $\quad P(U = 1) = 0.8$



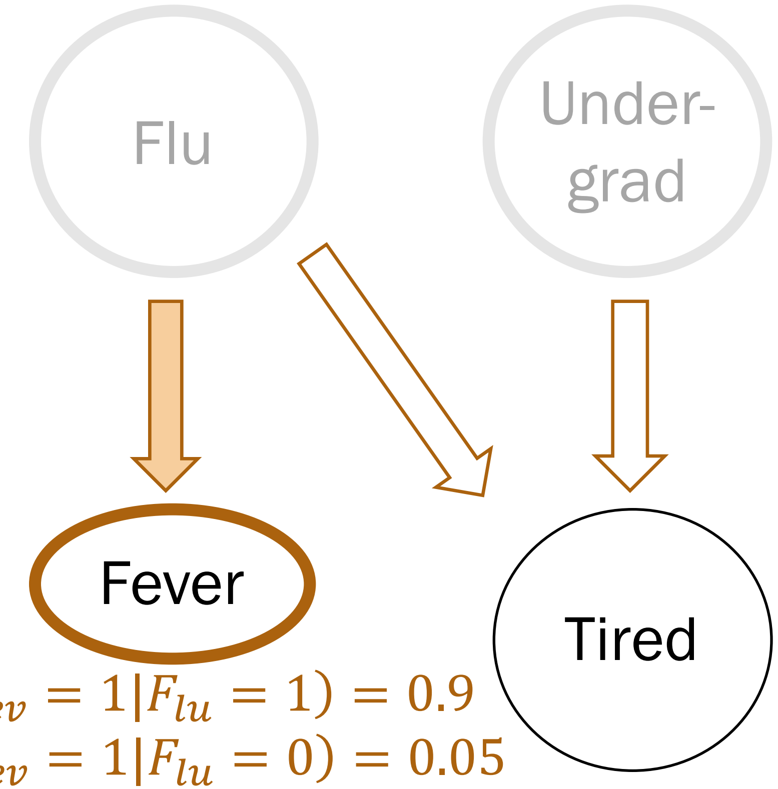
$P(F_{ev} = 1 | F_{lu} = 1) = 0.9$
$P(F_{ev} = 1 | F_{lu} = 0) = 0.05$

$P(T = 1 | F_{lu} = 0, U = 0) = 0.1$
$P(T = 1 | F_{lu} = 0, U = 1) = 0.8$
$P(T = 1 | F_{lu} = 1, U = 0) = 0.9$
$P(T = 1 | F_{lu} = 1, U = 1) = 1.0$

Rejection sampling algorithm

```python
# Method: Make Sample
# -------------------
# create a single sample from the joint distribution
# based on the medical "WebMD" Bayesian Network
def make_sample():
    # prior on causal factors
    flu = bernoulli(0.1)
    und = bernoulli(0.8)

    # choose fever based on flu
    if flu == 1: fev = bernoulli(0.9)
    else:        fev = bernoulli(0.05)

    # choose tired based on (undergrad and flu)
    #
    # TODO: fill in
    #
    #

    # a sample from the joint has an
    # assignment to *all* random variables
    return [flu, und, fev, tir]
```

$P(F_{lu} = 1) = 0.1$ $P(U = 1) = 0.8$



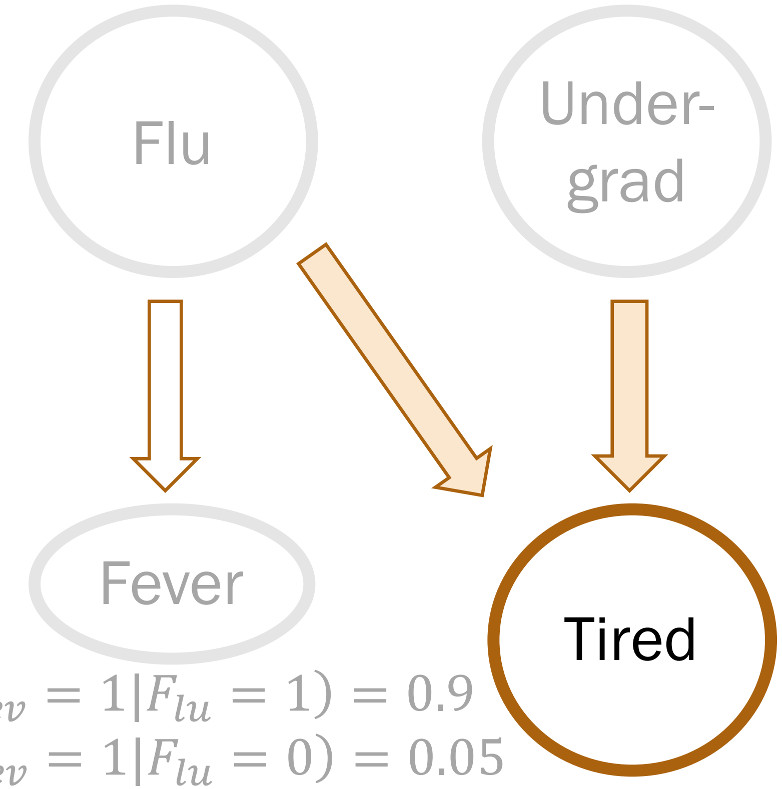
$P(F_{ev} = 1 | F_{lu} = 1) = 0.9$
$P(F_{ev} = 1 | F_{lu} = 0) = 0.05$

$P(T = 1 | F_{lu} = 0, U = 0) = 0.1$
$P(T = 1 | F_{lu} = 0, U = 1) = 0.8$
$P(T = 1 | F_{lu} = 1, U = 0) = 0.9$
$P(T = 1 | F_{lu} = 1, U = 1) = 1.0$

Rejection sampling algorithm

```python
# Method: Make Sample
# -------------------
# create a single sample from the joint distribution
# based on the medical "WebMD" Bayesian Network
def make_sample():
    # prior on causal factors
    flu = bernoulli(0.1)
    und = bernoulli(0.8)

    # choose fever based on flu
    if flu == 1: fev = bernoulli(0.9)
    else:        fev = bernoulli(0.05)

    # choose tired based on (undergrad and flu)
    #
    # TODO: fill in
    #
    #

    # a sample from the joint has an
    # assignment to *all* random variables
    return [flu, und, fev, tir]
```

$P(F_{lu} = 1) = 0.1$ $\quad P(U = 1) = 0.8$



$P(F_{ev} = 1 | F_{lu} = 1) = 0.9$

$P(F_{ev} = 1 | F_{lu} = 0) = 0.05$

$P(T = 1 | F_{lu} = 0, U = 0) = 0.1$

$P(T = 1 | F_{lu} = 0, U = 1) = 0.8$

$P(T = 1 | F_{lu} = 1, U = 0) = 0.9$

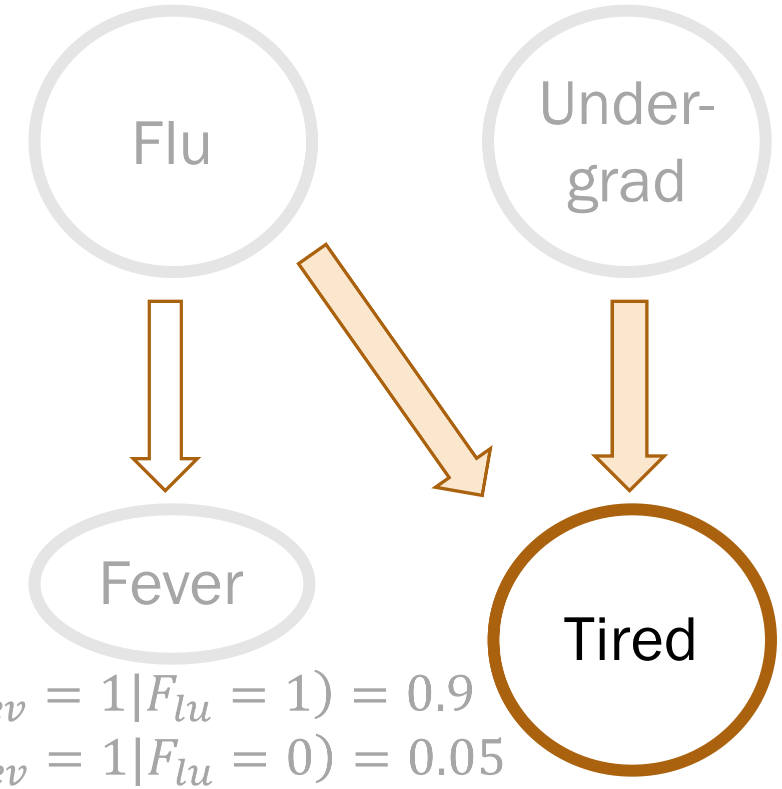
$P(T = 1 | F_{lu} = 1, U = 1) = 1.0$

Rejection sampling algorithm

```python
# Method: Make Sample
# --------------------
# create a single sample from the joint distribution
# based on the medical "WebMD" Bayesian Network
def make_sample():
    # prior on causal factors
    flu = bernoulli(0.1)
    und = bernoulli(0.8)

    # choose fever based on flu
    if flu == 1: fev = bernoulli(0.9)
    else:        fev = bernoulli(0.05)

    # choose tired based on (undergrad and flu)
    if   flu == 0 and und == 0: tir = bernoulli(0.1)
    elif flu == 0 and und == 1: tir = bernoulli(0.8)
    elif flu == 1 and und == 0: tir = bernoulli(0.9)
    else:                       tir = bernoulli(1.0)

    # a sample from the joint has an
    # assignment to *all* random variables
    return [flu, und, fev, tir]
```

$P(F_{lu} = 1) = 0.1$

$P(U = 1) = 0.8$



$P(F_{ev} = 1|F_{lu} = 1) = 0.9$

$P(F_{ev} = 1|F_{lu} = 0) = 0.05$

$P(T = 1|F_{lu} = 0, U = 0) = 0.1$

$P(T = 1|F_{lu} = 0, U = 1) = 0.8$

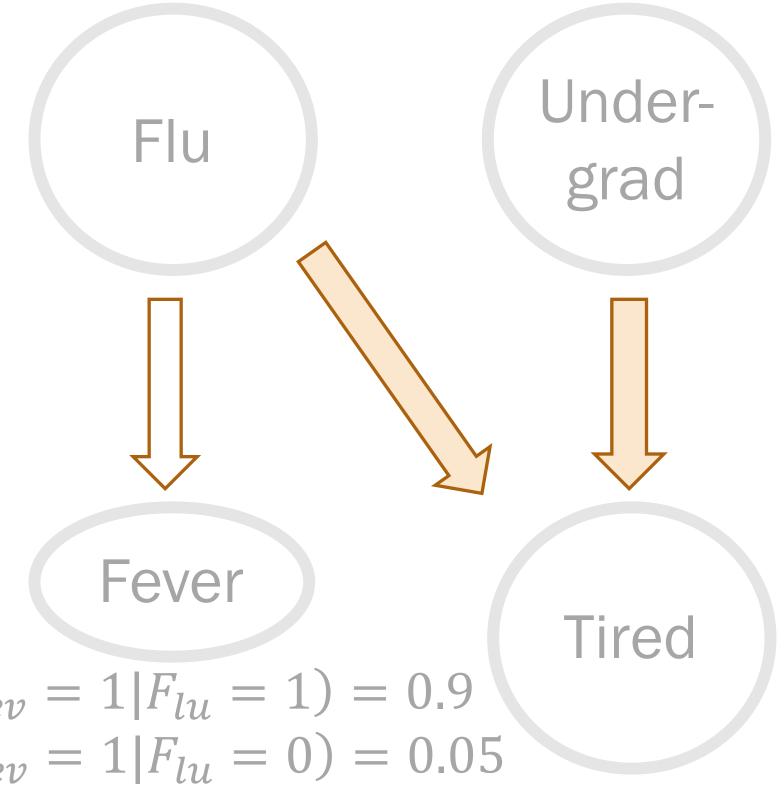
$P(T = 1|F_{lu} = 1, U = 0) = 0.9$

$P(T = 1|F_{lu} = 1, U = 1) = 1.0$

Rejection sampling algorithm

```python
# Method: Make Sample
# -------------------
# create a single sample from the joint distribution
# based on the medical "WebMD" Bayesian Network
def make_sample():
    # prior on causal factors
    flu = bernoulli(0.1)
    und = bernoulli(0.8)

    # choose fever based on flu
    if flu == 1: fev = bernoulli(0.9)
    else:        fev = bernoulli(0.05)

    # choose tired based on (undergrad and flu)
    if   flu == 0 and und == 0: tir = bernoulli(0.1)
    elif flu == 0 and und == 1: tir = bernoulli(0.8)
    elif flu == 1 and und == 0: tir = bernoulli(0.9)
    else:                       tir = bernoulli(1.0)

    # a sample from the joint has an
    # assignment to *all* random variables
    return [flu, und, fev, tir]
```

$P(F_{lu} = 1) = 0.1$ $\quad P(U = 1) = 0.8$



$P(F_{ev} = 1|F_{lu} = 1) = 0.9$

$P(F_{ev} = 1|F_{lu} = 0) = 0.05$

$P(T = 1|F_{lu} = 0, U = 0) = 0.1$

$P(T = 1|F_{lu} = 0, U = 1) = 0.8$

$P(T = 1|F_{lu} = 1, U = 0) = 0.9$

$P(T = 1|F_{lu} = 1, U = 1) = 1.0$

Rejection sampling algorithm

Inference question: What is $P(F_{lu} = 1 | U = 1, T = 1)$?

```
def rejection_sampling(event, observation):
  samples = sample_a_ton()
  samples_observation = …
          # number of samples with (U = 1, T = 1)
  samples_event =
          # number of samples with (F_lu = 1, U = 1, T = 1)
  return len(samples_event)/len(samples_observation)
```

Rejection sampling algorithm

Inference question: What is $P(F_{lu} = 1 | U = 1, T = 1)$?

```
def rejection_sampling(event, observation):
  samples = sample_a_ton()
  samples_observation =
        reject_inconsistent(samples, observation)
  samples_event =
        # number of samples with (F_lu = 1, U = 1, T = 1)
  return len(samples_event)/len(samples_observation)
```

Rejection sampling algorithm

Inference question: What is $P(F_{lu} = 1 | U = 1, T = 1)$?

```
def rejection_sampling(event, observation):
  samples = sample_a_ton()
  samples_observation =
        reject_inconsistent(samples, observation)
  samples_event =
        # number of samples with (F_lu = 1, U = 1, T = 1)
  return len(samples_event)/len(samples_observation)
```

Keep only samples that are consistent with the observation $(U = 1, T = 1)$.

Rejection sampling algorithm

Inference question: What is $P(F_{lu} = 1 | U = 1, T = 1)$?

```
def rejection_sampling(event, observation):
    samples = sample_a_ton()
    samples_observation =
            reject_inconsistent(samples, observation)
    samples_
    return
```

```
# Method: Reject Inconsistent
# -------------------
# Rejects all samples that do not align with the outcome.
# Returns a list of consistent samples.
def reject_inconsistent(samples, outcome):
    consistent_samples = []
    for sample in samples:
        if check_consistent(sample, outcome):
            consistent_samples.append(sample)
    return consistent_samples
```

$(U = 1, T = 1)$

Rejection sampling algorithm

Inference question: What is $P(F_{lu} = 1 | U = 1, T = 1)$?

```
def rejection_sampling(event, observation):
    samples = sample_a_ton()
    samples_observation =
              reject_inconsistent(samples, observation)
    samples_event =
              reject_inconsistent(samples_observation, event)
    return len(samples_event)/len(samples_observation)
```

Conditional event = samples with $(F_{lu} = 1, U = 1, T = 1)$.

Rejection sampling algorithm

Inference question: What is $P(F_{lu} = 1 | U = 1, T = 1)$?

```python
def rejection_sampling(event, observation):
    samples = sample_a_ton()
    samples_observation =
            reject_inconsistent(samples, observation)
    samples_event =
            reject_inconsistent(samples_observation, event)
    return l                                          ation)
```

Conditi $= 1)$.

```python
def reject_inconsistent(samples, outcome):
```

$(F_{lu} = x, U = 1, F_{ev} = y, T = 1)$ $\quad$ $(F_{lu} = 1)$

```python
    return consistent_samples
```

Rejection sampling algorithm

Inference question: What is $P(F_{lu} = 1 | U = 1, T = 1)$?

```
def rejection_sampling(event, observation):
    samples = sample_a_ton()
    samples_observation =
            reject_inconsistent(samples, observation)
    samples_event =
            reject_inconsistent(samples_observation, event)
    return len(samples_event)/len(samples_observation)
```

Approximate Probability = $\dfrac{\text{\# samples with } (F_{lu} = 1, U = 1, T = 1)}{\text{\# samples with } (U = 1, T = 1)}$

Rejection sampling

If you can sample enough from the joint distribution, you can answer most probability inference questions.

With enough samples, you can correctly compute:

- Probability estimates
- Conditional probability estimates
- Expectation estimates

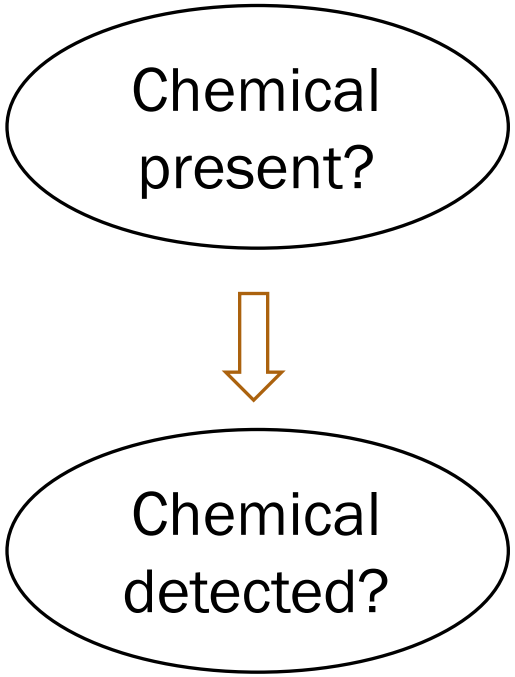
Because your samples are a representation of the joint distribution!

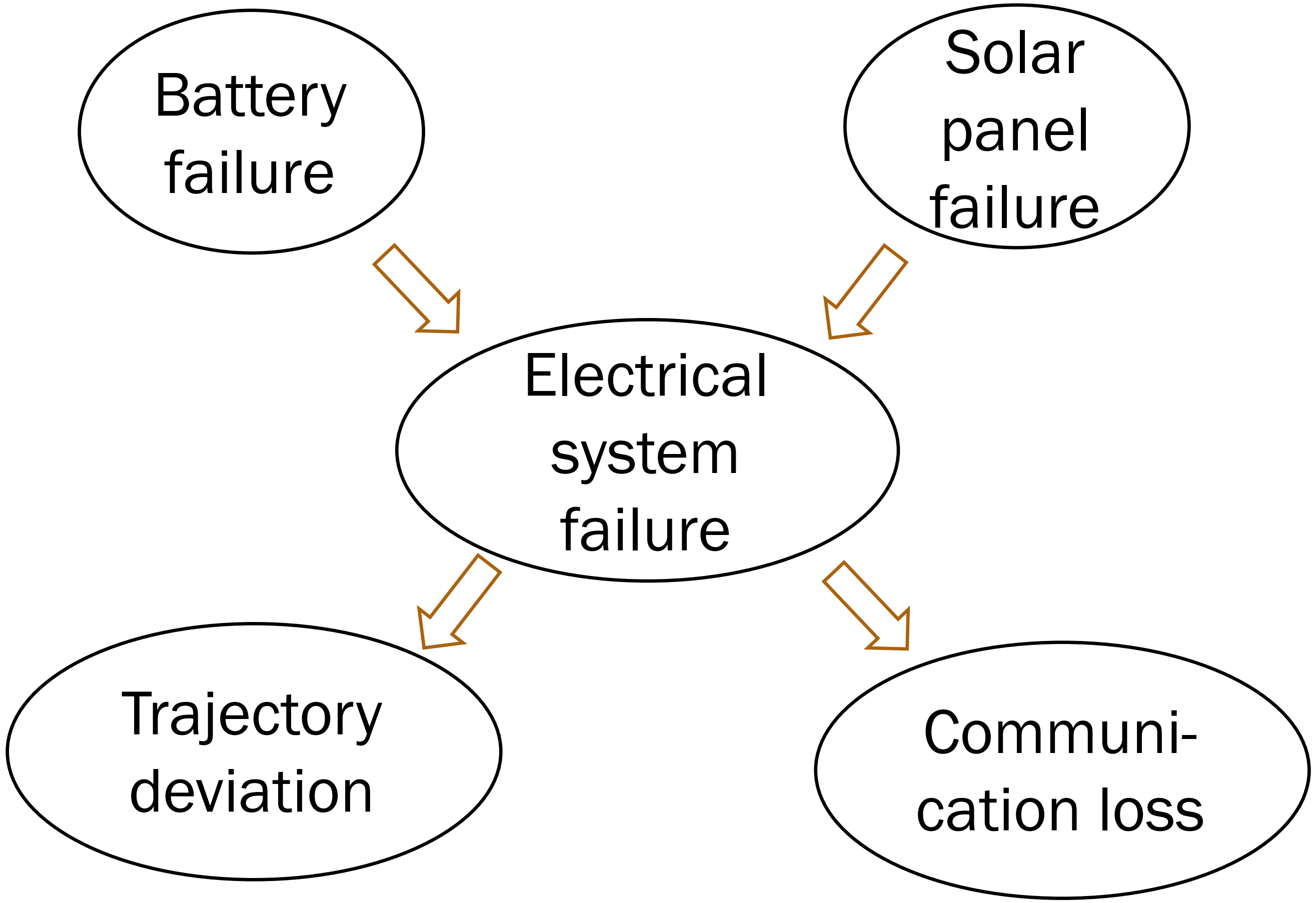
[flu, und, fev, tir]

```
Sampling...
[0, 1, 0, 1]
[0, 1, 0, 1]
[0, 1, 0, 1]
[0, 0, 0, 0]
[0, 1, 0, 1]
[0, 1, 1, 1]
[0, 1, 0, 0]
[1, 1, 1, 1]
[0, 0, 1, 1]
...
[0, 1, 0, 1]
Finished sampling
```

```
P(has flu | undergrad and is tired) = 0.122
```

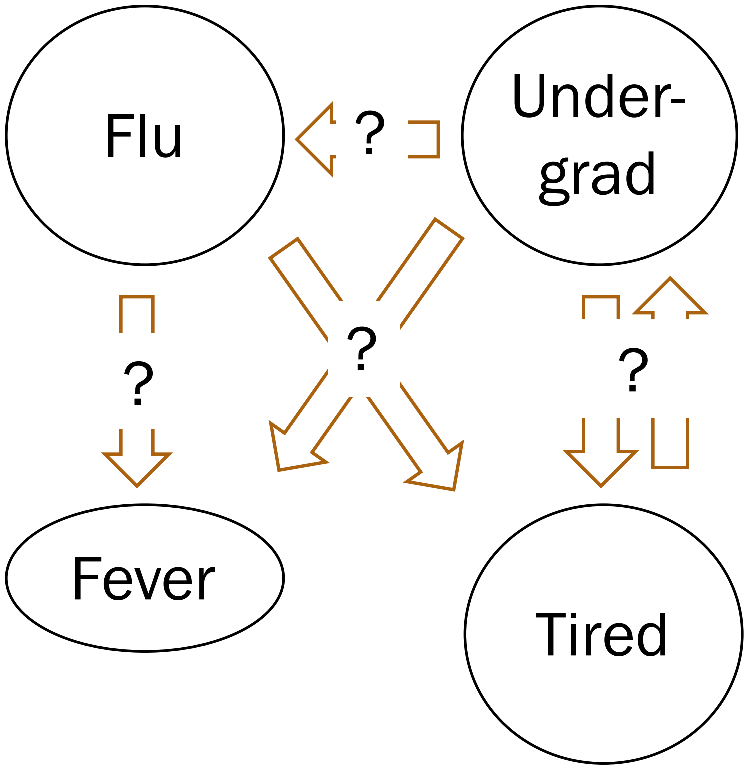
Other applications





Take CS238/AA228: Decision Making under Uncertainty!

Challenge with Bayesian Networks



What if we don't know the structure?

Take CS228: Probabilistic Graphical Models!

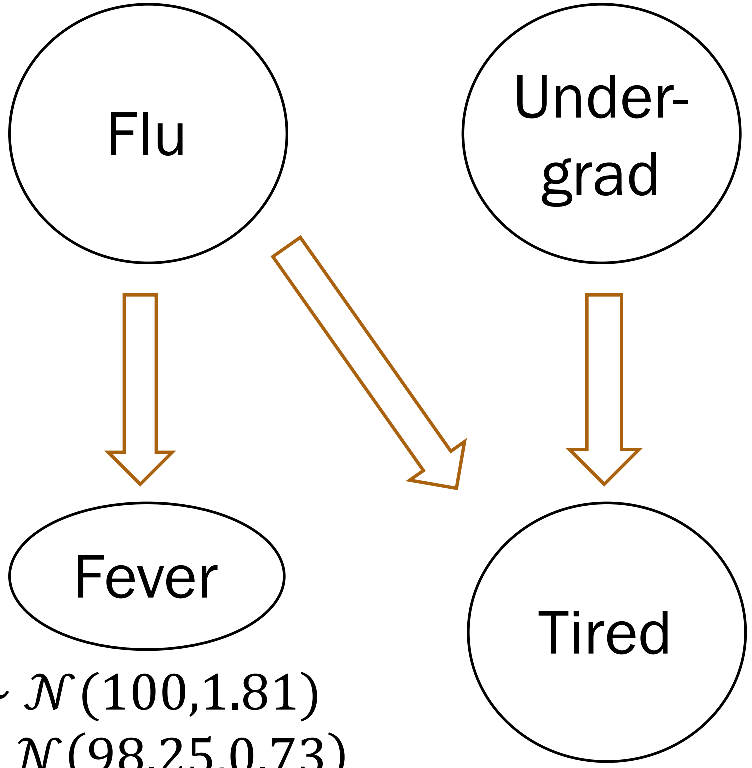
Disadvantages of rejection sampling

$$P(F_{lu} = 1 | F_{ev} = 99.4)?$$

What if random variables are continuous?

What if you run out of time/computational power?

$P(F_{lu} = 1) = 0.1$ $P(U = 1) = 0.8$



$F_{ev} | F_{lu} = 1 \sim \mathcal{N}(100, 1.81)$
$F_{ev} | F_{lu} = 0 \sim \mathcal{N}(98.25, 0.73)$

$P(T = 1 | F_{lu} = 0, U = 0) = 0.1$
$P(T = 1 | F_{lu} = 0, U = 1) = 0.8$
$P(T = 1 | F_{lu} = 1, U = 0) = 0.9$
$P(T = 1 | F_{lu} = 1, U = 1) = 1.0$

Congratulations on finishing the midterm ☺