15: General Inference

Lisa Yan
May 8, 2020
Quick slide reference

3  General Inference: intro
15  Bayesian Networks
22  Inference (I): Math
29  Inference (II): Rejection sampling
54  Inference (III): Gibbs sampling (extra)

15a_inference
15b_bayes_nets
15c_inference_math
LIVE
(no video)
General Inference: Introduction
Inference
Inference

WebMD Symptom Checker BETA

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
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

$N = 9$
all binary RVs
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

Naively specifying a joint distribution is often intractable.

\(N = 9\) all binary RVs
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 $X_1, X_2, ..., X_n$ are called conditionally independent given $Y$ if:

for all $x_1, x_2, ..., x_n, y$:

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

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

$$\log P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n | Y = y) = \sum_{i=1}^{n} \log P(X_i = x_i | Y = y)$$
Recall independence of $n$ events $E_1, E_2, ..., E_n$:

for $r = 1, ..., n$: for every subset $E_1, E_2, ..., E_r$:

$$P(E_1, E_2, ..., E_r) = P(E_1)P(E_2) \cdots P(E_r)$$

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

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

Errata (edited May 3): Removed the independent RV requirement for all subsets of size $r = 1, ..., n$. 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!!!
What would a Stanford flu expert do?

1. Describe the joint distribution using causality.

2. **Assume conditional independence.**
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:
- \( P(F_{ev} = 1 | T \neq 0, F_{lu} = 1) = P(F_{ev} = 1 | F_{lu} = 1) \)
- \( P(F_{lu} = 1, U = 0) = P(F_{lu} = 1)P(U = 0) \)
Constructing a Bayesian Network

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?

$P(F_u = 1) = 0.1 \quad P(U = 1) = 0.8$

$P(F_e = 1 | F_u = 1) = 0.9$
$P(F_e = 1 | F_u = 0) = 0.05$
Constructing a Bayesian Network

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?

- $P(T = 1 | F_{lu} = 0, U = 0) = 0.8$
- $P(T = 1 | F_{lu} = 0, U = 1) = 0.05$
- $P(T = 1 | F_{lu} = 1, U = 0) = 0.05$
- $P(T = 1 | F_{lu} = 1, U = 1) = 0.05$

$p(F_{lu} = 1) = 0.1$  \hspace{1cm}  $p(U = 1) = 0.8$

Flu $\rightarrow$ Under-grad $\rightarrow$ Fever $\rightarrow$ Tired

$p(F_{ev} = 1 | F_{lu} = 1) = 0.9$
$p(F_{ev} = 1 | F_{lu} = 0) = 0.05$
Using a Bayes Net

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

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

- Flu
- Undergrad
- Fever
- Tired

\[
P(F_{eu} = 1|F_{lu} = 1) = 0.9 \quad P(F_{eu} = 1|F_{lu} = 0) = 0.05
\]

\[
P(T = 1|F_{lu} = 0, U = 0) = 0.1 \quad P(T = 1|F_{lu} = 0, U = 1) = 0.8
\]

\[
P(T = 1|F_{lu} = 1, U = 0) = 0.9 \quad P(T = 1|F_{lu} = 1, U = 1) = 1.0
\]
Inference (I): Math
Bayes Nets: Conditional independence

In a Bayesian Network, each random variable is conditionally independent of its non-descendants, given its parents.

- Node: random variable
- Directed edge: conditional dependency

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

\[
P(F_{eu} = 1|F_{lu} = 1) = 0.9 \quad P(T = 1|F_{lu} = 0, U = 0) = 0.1
\]

\[
P(F_{eu} = 1|F_{lu} = 0) = 0.05 \quad P(T = 1|F_{lu} = 0, U = 1) = 0.8
\]

\[
P(F_{eu} = 1) = 0.8 \quad P(T = 1|F_{lu} = 1, U = 0) = 0.9
\]

\[
P(T = 1|F_{lu} = 1, U = 1) = 1.0
\]

Review
Inference via math

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

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

Compute joint probabilities using chain rule.

\[
P(F_{lu} = 0, U = 1, F_{ev} = 0, T = 1) = P(F_{lu} = 0) P(U = 1 | F_{lu} = 0) P(F_{ev} = 0 | F_{lu} = 0, U = 1) \times P(T = 1 | F_{ev} = 0, U = 1, F_{lu} = 0)
\]

\[
= P(F_{ev} = 0) P(U = 1) P(F_{ev} = 0 | F_{lu} = 0) \times P(T = 1 | F_{ev} = 0, U = 1)
\]

\[
= 0.9 \times 0.8 \times 0.95 \times 0.8
\]

\[
= 0.5472
\]
Inference via math

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

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

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

\[ P(F_{ev} = 1|F_{lu} = 1) = 0.9 \]
\[ P(F_{ev} = 1|F_{lu} = 0) = 0.05 \]
\[ P(T = 1|F_{lu} = 0, U = 0) = 0.1 \]
\[ P(T = 1|F_{lu} = 0, U = 1) = 0.8 \]
\[ P(T = 1|F_{lu} = 1, U = 0) = 0.9 \]
\[ P(T = 1|F_{lu} = 1, U = 1) = 1.0 \]
Inference via math

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

1. Compute joint probabilities

\[ P(F_{flu} = 1, U = 1, F_{ev} = 1, T = 1) \]

\[ \ldots \]

\[ P(F_{flu} = 0, U = 1, F_{ev} = 0, T = 1) \]

2. Definition of conditional probability

\[
\frac{\sum_y P(F_{lu} = 1, U = 1, F_{ev} = y, T = 1)}{\sum_x \sum_y P(F_{lu} = x, U = 1, F_{ev} = y, T = 1)} = 0.122
\]

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

\[ P(F_{ev} = 1|F_{lu} = 1) = 0.9 \]
\[ P(F_{ev} = 1|F_{lu} = 0) = 0.05 \]
\[ P(T = 1|F_{lu} = 0, U = 0) = 0.1 \]
\[ P(T = 1|F_{lu} = 0, U = 1) = 0.8 \]
\[ P(T = 1|F_{lu} = 1, U = 0) = 0.9 \]
\[ P(T = 1|F_{lu} = 1, U = 1) = 1.0 \]
Inference via math

Solving inference questions precisely is possible, but sometimes tedious.

Can we use sampling to do approximate inference?

Yes.

\[
\begin{align*}
P(F_{lu} = 1) &= 0.1 & P(U = 1) &= 0.8 \\

Flu & \quad & Under-grad \\
\downarrow & & \downarrow \\
Fever & \quad & Tired \\
\downarrow & & \downarrow \\
P(F_{ev} = 1|F_{lu} = 1) &= 0.9 & P(T = 1|F_{lu} = 0, U = 0) &= 0.1 \\
P(F_{ev} = 1|F_{lu} = 0) &= 0.05 & 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
\end{align*}
\]
15: General Inference

Lisa Yan
May 8, 2020
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:
- $P(F_{ev} = 1|T = 0, F_{lu} = 1) = P(F_{ev} = 1|F_{lu} = 1)$
- $P(F_{lu} = 1, U = 0) = P(F_{lu} = 1)P(U = 0)$
Check out the question on the next slide (Slide 31). Post any clarifications here!

https://us.edstem.org/courses/109/discussion/59206

Breakout rooms: 4 min. Introduce yourself!
Inference via math

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

\[ P(F_{lu} = 1|F_{ev} = 1, U = 1, T = 1)? \]

\[ P(F_{ev} = 1|F_{lu} = 1) = 0.9 \]
\[ P(F_{ev} = 1|F_{lu} = 0) = 0.05 \]
\[ P(T = 1|F_{lu} = 0, U = 0) = 0.1 \]
\[ P(T = 1|F_{lu} = 0, U = 1) = 0.8 \]
\[ P(T = 1|F_{lu} = 1, U = 0) = 0.9 \]
\[ P(T = 1|F_{lu} = 1, U = 1) = 1.0 \]

\[ 0.9 + 0.05 = 1 ? \]
\[ 1 = P(F_{ev}=1 | F_{lu}=1) + P(F_{ev}=0 | F_{lu}=1) \]
\[ 0.9 \]
\[ 1 = P(F_{ev}=1 | F_{lu}=0) + P(F_{ev}=0 | F_{lu}=0) \]
\[ 0.95 \]
Inference via math

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

\[
\begin{align*}
P(F_{lu} = 1|F_{ev} = 1, U = 1, T = 1) &= \frac{P(F_{lu} = 1, F_{ev} = 1, U = 1, T = 1)}{P(F_{ev} = 1, U = 1, T = 1)} \\
&= \frac{P(F_{lu} = 1) \cdot P(F_{ev} = 1|F_{lu} = 1) \cdot P(U = 1|F_{ev} = 1, F_{lu} = 1) \cdot P(T = 1|F_{ev} = 1, F_{lu} = 1, U = 1)}{P(F_{ev} = 1) \cdot P(U = 1) \cdot P(T = 1|F_{ev} = 1, U = 1)} \\
&= \frac{0.1 \cdot 0.9 \cdot 0.9 \cdot 1.0}{0.9 \cdot 0.8 \cdot 1.0} \\
&\approx 0.714
\end{align*}
\]

\[
\begin{align*}
P(F_{ev} = 1|F_{lu} = 1) &= 0.9 \\
P(F_{ev} = 1|F_{lu} = 0) &= 0.05 \\
P(T = 1|F_{lu} = 0, U = 0) &= 0.1 \\
P(T = 1|F_{lu} = 0, U = 1) &= 0.8 \\
P(T = 1|F_{lu} = 1, U = 0) &= 0.9 \\
P(T = 1|F_{lu} = 1, U = 1) &= 1.0
\end{align*}
\]
Inference via math

\[ P(F_{lu} = 1) = 0.1 \quad \Leftrightarrow \quad P(U = 1) = 0.8 \]

\[ P(F_{lu} = 1|F_{ev} = 1, U = 1, T = 1) \]

\[
\text{Numerator} \quad P(F_{lu} = 1, F_{ev} = 1, U = 1, T = 1) \\
= P(F_{lu} = 1) \cdot P(U = 1|F_{lu} = 1) \cdot P(F_{ev} = 1|F_{lu} = 1, U = 1) \cdot P(T = 1|F_{lu} = 1, U = 1, F_{ev} = 1)
\]

\[ = 0.1 \cdot 0.9 \cdot 0.8 \cdot 1 = 0.72 \]

\[ P(T = 1|F_{lu} = 0, U = 0) = 0.1 \]
\[ P(T = 1|F_{lu} = 0, U = 1) = 0.8 \]
\[ P(T = 1|F_{lu} = 1, U = 0) = 0.9 \]
\[ P(T = 1|F_{lu} = 1, U = 1) = 1.0 \]
Solving inference questions precisely is possible, but sometimes tedious.

Can we use sampling to do approximate inference?

Yes.
Inference via math

\[ P(F_{lu} = 1) = 0.1 \quad \text{and} \quad P(U = 1) = 0.8 \]

\[ P(F_{ev} = 1|F_{lu} = 1) = 0.9 \quad \text{and} \quad P(F_{ev} = 1|F_{lu} = 0) = 0.05 \]

(3) \quad \text{What is } P(F_{lu} = 1|U = 1, T = 1)?

\[ = 0.122 \]

(from pre-lecture video)
Rejection sampling algorithm

Step 0:
Have a fully specified Bayesian Network

\begin{align*}
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
\end{align*}
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 = ...  # 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)
```

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

Inference question: What is $P(\text{Fl}_u = 1|U = 1, T = 1)$?

```python
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 $(\text{Fl}_u = 1, U = 1, T = 1)$

return len(samples_event)/len(samples_observation)
```

Approximate Probability = \[
\frac{\# \text{ samples with } (\text{Fl}_u = 1, U = 1, T = 1)}{\# \text{ samples with } (U = 1, T = 1)}
\]
Rejection sampling algorithm

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

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

Why would this definition of approximate probability make sense?
Think

Slide 41 has a question to go over by yourself.

Post any clarifications here!

https://us.edstem.org/courses/109/discussion/59206

Think by yourself: 2 min
Why would this approximate probability make sense?

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

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

Recall our definition of probability as a frequency:

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

\( n = \text{# of total trials} \)

\( n(E) = \text{# trials where } E \text{ occurs} \)
Why would this approximate probability make sense?

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

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

Recall our definition of probability as a frequency:

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

$n = \# \text{ of total trials}$

$n(E) = \# \text{ trials where } E \text{ occurs}$
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 = ...
    # 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)]
```

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

Flu

Undergrad

Fever

Tired

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

\[
P(F_{ev} = 1|F_{lu} = 1) = 0.9
P(F_{ev} = 1|F_{lu} = 0) = 0.05
\]

\[
P(T = 1|F_{lu} = 0, U = 0) = 0.1
P(T = 1|F_{lu} = 0, U = 1) = 0.8
P(T = 1|F_{lu} = 1, U = 0) = 0.9
P(T = 1|F_{lu} = 1, U = 1) = 1.0
\]
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]
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]
Interlude for jokes/announcements
Announcements

Quiz #1
Solutions: after class
Grades: after class
Regrades: by next Friday

Problem Set 4
Out: later today
Due: Monday 5/18 10am
Covers: Up to and including today

Mid-quarter feedback form
Open until: next Friday

link

Giant sheet of all equations covered so far?
Announcements: CS109 contest

Do something cool and creative with probability

Replaces one “passing” work requirement

Optional Proposal: Sat. 5/23 11:59pm
Due: Week 10 Monday 6/8, 11:59pm

Winner 1st place $
Interesting probability news

http://www.intuitior.com/statistics/TwentyQs.html

CS109 Current Events Spreadsheet
# 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]
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)
    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]
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)
    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]
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 = ...
    # 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)
```

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

```python
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 \( P(F_{lu} = 1|U = 1, T = 1) \)?

```python
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 $P(F_{lu} = 1|U = 1, T = 1)$?

```python
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)
```

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

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

Inference question:

```python
# 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
```
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 len(samples_event)/len(samples_observation)
```

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

What is $P(F_{lu} = 1 | U = 1, T = 1)$?
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 len(samples_event) / len(samples_observation)
```

```
def reject_inconsistent(samples, outcome):
    Condition (F_{lu} = x, U = 1, F_{ev} = y, T = 1) \hspace{1cm} (F_{lu} = 1) = 1).
    return consistent_samples
```

What is \( P(F_{lu} = 1 | U = 1, T = 1) \)?
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 len(samples_event)/len(samples_observation)
```

Approximate Probability = \[
\frac{\text{# samples with } (F_{lu} = 1, U = 1, T = 1)}{\text{# samples with } (U = 1, T = 1)}
\]
To the code!
Rejection sampling

If you can sample enough from the joint distribution, you can answer any probability inference question.

With enough samples, you can correctly compute:

• Probability estimates
• Conditional probability estimates
• Expectation estimates

Because your samples are a representation of the joint distribution!

\[
P(\text{has flu} \mid \text{undergrad and is tired}) = 0.122
\]
Other applications

Chemical present?

Chemical detected?

Battery failure

Electrical system failure

Trajectory deviation

Solar panel failure

Communication loss

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

\[ P(F_{lu} = 1|F_{ev} = 1)? \]

What if we never encounter some samples?

[flu=0, und, fev=1, tir]

\[
\begin{align*}
P(F_{lu} = 1) &= 0.1 \\
P(U = 1) &= 0.8 \\
P(F_{ev} = 1|F_{lu} = 1) &= 0.9 \\
P(F_{ev} = 1|F_{lu} = 0) &= 0.05 \\
P(T = 1|F_{lu} = 0, U = 0) &= 0.1 \\
P(T = 1|F_{lu} = 0, U = 1) &= 0.8 \\
P(T = 1|F_{lu} = 1, U = 0) &= 0.9 \\
P(T = 1|F_{lu} = 1, U = 1) &= 1.0
\end{align*}
\]
Disadvantages of rejection sampling

What if we never encounter some samples?

What if random variables are continuous?

\[ P(F_{lu} = 1 | F_{ev} = 99.4) ? \]

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

\[ P(T = 1 | F_{lu} = 0, U = 0) = 0.1 \]
\[ P(T = 1 | F_{lu} = 0, U = 1) = 0.8 \]
\[ P(T = 1 | F_{lu} = 1, U = 0) = 0.9 \]
\[ P(T = 1 | F_{lu} = 1, U = 1) = 1.0 \]
Gibbs sampling (extra)
Gibbs Sampling (not covered)

Basic idea:
• Fix all observed events
• Incrementally sample a new value for each random variable
• Difficulty: More coding for computing different posterior probabilities

Learn in extra slides/extra notebook!
(or by taking CS228/CS238)