15: General Inference

Jerry Cain
April 30, 2021
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

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

Lisa Yan, Chris Piech, Mehran Sahami, and Jerry Cain, CS109, Spring 2021
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 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):

$$\log P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n|Y = y) = \sum_{i=1}^{n} \log P(X_i = x_i|Y = y)$$
Lec. 12: Independence of multiple random variables

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

for $r = 1, \ldots, n$:

for every subset $E_1, E_2, \ldots, E_r$:

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

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

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

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

(Hint: independence of RVs implies independence of all events)
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. 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:
- \( P(F_v = 1|T = 0, F_u = 1) = P(F_v = 1|F_u = 1) \)
- \( P(F_u = 1, U = 0) = P(F_u = 1)P(U = 0) \)
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?

- $P(F_{lu} = 1) = 0.1$
- $P(U = 1) = 0.8$
- $P(F_{ev} = 1|F_{lu} = 1) = 0.9$
- $P(F_{ev} = 1|F_{lu} = 0) = 0.05$
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?

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

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

- \( P(F_{ev} = 1|F_{lu} = 1) = 0.9 \)
- \( P(F_{ev} = 1|F_{lu} = 0) = 0.05 \)
- \( P(T = 1|F_{lu} = 0, U = 0) = 0.1 \)
- \( P(T = 1|F_{lu} = 0, U = 1) = 0.8 \)
- \( P(T = 1|F_{lu} = 1, U = 0) = 0.9 \)
- \( P(T = 1|F_{lu} = 1, U = 1) = 1.0 \)
Inference (I):
Math
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

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

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

Compute joint probabilities using chain rule.

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

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

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

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

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

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

   1. Compute joint probabilities
      \[
      P(F_{lu} = 1, U = 1, F_{ev} = 1, T = 1)
      \]
      \[
      \ldots
      \]
      \[
      P(F_{lu} = 0, U = 1, F_{ev} = 0, T = 1)^
      \]

   2. Definition of conditional probability
      \[
      \frac{\sum_y P(F_{lu} = 1, U = 1, F_{ev} = y, T = 1)}{\sum_x \sum_y P(F_{lu} = x, U = 1, F_{ev} = y, T = 1)} = 0.122
      \]

\[
\begin{align*}
P(F_{ev} = 1|F_{lu} = 1) &= 0.9 \\
P(F_{ev} = 1|F_{lu} = 0) &= 0.05 \\
P(T = 1|F_{lu} = 0, U = 0) &= 0.1 \\
P(T = 1|F_{lu} = 0, U = 1) &= 0.8 \\
P(T = 1|F_{lu} = 1, U = 0) &= 0.9 \\
P(T = 1|F_{lu} = 1, U = 1) &= 1.0
\end{align*}
\]
Solving inference questions precisely is possible, but sometimes tedious.

Can we use sampling to do approximate inference?

**Yes.**
15: General Inference

Jerry Cain
April 30, 2021
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:
- \( P(F_{ev} = 1|T = 0, F_{lu} = 1) = P(F_{ev} = 1|F_{lu} = 1) \)
- \( P(F_{lu} = 1, U = 0) = P(F_{lu} = 1)P(U = 0) \)
Inference via math

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

= 0.122

(from pre-lecture video)

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

Flu  Under-
grad

Fever  Tired

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

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

Flu → Undergrad

Fever ↓ Tired

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

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

Solving inference questions precisely is possible, but sometimes tedious.

Can we use sampling to do approximate inference?
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 $P(F_{lu} = 1|U = 1, T = 1)$?

```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...

[Samples...]

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)
```

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

Why would this definition of approximate probability make sense?
Why would this approximate probability make sense?

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

Why would this definition of approximate probability make sense?

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:

\[ P(E) = \lim_{n \to \infty} \frac{n(E)}{n} \quad n = \# \text{ of total trials} \]

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

[flu, und, fev, tir]
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!!
# Method: Make Sample
# create a single sample from the joint distribution
# based on the medical "WebMD" Bayesian Network

```python
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]
```

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

```
\begin{align*}
P(F_{ev} = 1 | F_{lu} = 1) &= 0.9 \\
P(F_{ev} = 1 | F_{lu} = 0) &= 0.05
\end{align*}
```

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

Flu

Undergrad

Fever

Tired

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

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

Flu
Under-grad
Fever
Tired

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

# 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]
Interlude for announcements
Announcements: CS109 contest

Do something cool and creative with probability

Grand Prize:  
Two lowest quizzes replaced with 100%

Finalists:  
Lowest quiz replaced with 100%

Optional Proposal:  Mon. 05/17, 11:59pm
Due:  Mon. 05/31, 11:59pm

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, 1, 1, 1]
[0, 1, 0, 0]
[0, 1, 0, 1]
[1, 1, 0, 1]
[0, 0, 1, 1]
[0, 1, 1, 1]
[0, 1, 0, 0]
[0, 0, 1, 0]
[1, 1, 1, 0]
[0, 1, 0, 1]
[0, 0, 0, 1]
[0, 1, 0, 1]
[0, 0, 1, 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)
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)
```
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 $P(F_{lu} = 1|U = 1, T = 1)$?

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

```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 $(F_{lu} = 1, U = 1, T = 1)$.
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):
    consistent_samples
    return consistent_samples
```

Conditional event: $F_{lu} = 1, U = 1, T = 1$.

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)
```

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!

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