14: Conditional Expectation

Lisa Yan
May 6, 2020
<table>
<thead>
<tr>
<th></th>
<th>Title</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Conditional distributions</td>
<td>14a_conditional_distributions</td>
</tr>
<tr>
<td>11</td>
<td>Web server requests, redux</td>
<td>14b_web_servers</td>
</tr>
<tr>
<td>14</td>
<td>Conditional expectation</td>
<td>14c_cond_expectation</td>
</tr>
<tr>
<td>20</td>
<td>Law of Total Expectation</td>
<td>14d_law_of_total_expectation</td>
</tr>
<tr>
<td>24</td>
<td>Exercises</td>
<td>LIVE</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>$T$</th>
<th>$Y = 1$</th>
<th>$Y = 2$</th>
<th>$Y = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>−1</td>
<td>.06</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>0</td>
<td>.29</td>
<td>.14</td>
<td>.09</td>
</tr>
<tr>
<td>1</td>
<td>.30</td>
<td>.08</td>
<td>.02</td>
</tr>
</tbody>
</table>

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

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

<table>
<thead>
<tr>
<th></th>
<th>$Y = 1$</th>
<th>$Y = 2$</th>
<th>$Y = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = -1$</td>
<td>.06</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>$T = 0$</td>
<td>.29</td>
<td>.14</td>
<td>.09</td>
</tr>
<tr>
<td>$T = 1$</td>
<td>.30</td>
<td>.08</td>
<td>.02</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$Y = 1$</th>
<th>$Y = 2$</th>
<th>$Y = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = -1$</td>
<td>.75</td>
<td>.125</td>
<td>?</td>
</tr>
<tr>
<td>$T = 0$</td>
<td>.56</td>
<td>.27</td>
<td>.17</td>
</tr>
<tr>
<td>$T = 1$</td>
<td>.75</td>
<td>.2</td>
<td>.05</td>
</tr>
</tbody>
</table>
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?

<table>
<thead>
<tr>
<th>$T$</th>
<th>$Y = 1$</th>
<th>$Y = 2$</th>
<th>$Y = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-1$</td>
<td>0.09</td>
<td>0.04</td>
<td>0.08</td>
</tr>
<tr>
<td>0</td>
<td>0.45</td>
<td>0.61</td>
<td>0.75</td>
</tr>
<tr>
<td>1</td>
<td>0.46</td>
<td>0.35</td>
<td>0.17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$T$</th>
<th>$Y = 1$</th>
<th>$Y = 2$</th>
<th>$Y = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-1$</td>
<td>0.75</td>
<td>0.125</td>
<td><strong>0.125</strong></td>
</tr>
<tr>
<td>0</td>
<td>0.56</td>
<td>0.27</td>
<td>0.17</td>
</tr>
<tr>
<td>1</td>
<td>0.75</td>
<td>0.2</td>
<td>0.05</td>
</tr>
</tbody>
</table>

**Joint PMF**

<table>
<thead>
<tr>
<th></th>
<th>$Y = 1$</th>
<th>$Y = 2$</th>
<th>$Y = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T = -1$</td>
<td>0.06</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>$T = 0$</td>
<td>0.29</td>
<td>0.14</td>
<td>0.09</td>
</tr>
<tr>
<td>$T = 1$</td>
<td>0.30</td>
<td>0.08</td>
<td>0.02</td>
</tr>
</tbody>
</table>

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

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

$0.30 / (0.06 + 0.29 + 0.30)$

Conditional PMFs also sum to 1 conditioned on different events!
Extended to Amazon

P(bought item $X$ | bought item $Y$)
Quick check

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

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)

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:
  \[ X \sim \text{Poi}(\lambda p), \quad Y \sim \text{Poi}(\lambda (1 - p)) \]
- Two big parts of our derivation:
  \[ P(X = n, Y = m) = P(X = n | N = n + m)P(N = n) \]
  \[ X | N = n + m \sim \text{Bin}(n + m, p) \]
Web server requests, redux

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

• Let $X = \# \text{ requests from humans/day.}$ \hspace{1cm} X \sim \text{Poi}(\lambda_1)$
• Let $Y = \# \text{ requests from bots/day.}$ \hspace{1cm} Y \sim \text{Poi}(\lambda_2)$
• $X$ and $Y$ are independent. \hspace{1cm} \rightarrow X + Y \sim \text{Poi}(\lambda_1 + \lambda_2)$

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

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

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

\[
= \binom{n}{k} \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^k \left( \frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{n-k} \quad \quad X|X + Y \sim \text{Bin} \left( X + Y, \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)
\]
Conditional Expectation
Conditional expectation

Recall 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

\[ E[X|Y = y] = \sum_x x P(X = x|Y = y) = \sum_x x p_{X|Y}(x|y) \]
It’s been so long, our dice friends

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

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:

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

3. Law of total expectation (next time)
It's been so long, our dice friends

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

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

It’s been so long, our dice friends

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

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


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

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

\[
= \sum_{d_1} d_1P(D_1 = d_1) + d_2 \sum_{d_1} P(D_1 = d_1)
\]

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

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

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

3. Law of total expectation:

\[ E[X] = E[E[X|Y]] \quad \text{what}?! \]
Proof of Law of Total Expectation

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

\[ = \sum_y P(Y = y) \sum_x xP(X = x|Y = y) \]

\[ = \sum_y \left( \sum_x xP(X = x|Y = y)P(Y = y) \right) = \sum_y \left( \sum_x xP(X = x, Y = y) \right) \]

\[ = \sum_x \sum_y xP(X = x, Y = y) = \sum_x x \sum_y P(X = x, Y = y) \]

\[ = \sum_x xP(X = x) \]

\[ = E[X] \quad \text{...what?} \]
Another way to compute $E[X]$:

$$E[E[X|Y]] = \sum_{y} P(Y = y)E[X|Y = y] = E[X]$$

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 $P(Y = y)$)

```python
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
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

Lisa Yan, CS109, 2020
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 $X_1, X_2, \ldots, X_n$ are called conditionally independent given $Y$ if:

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

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

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

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

Flu  Under-grad

Fever  Tired

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:
- \( P(F_{ev} = 1|T = 0, F_{lu} = 1) = P(F_{ev} = 1|F_{lu} = 1) \)
- \( P(F_{lu} = 1, U = 0) = P(F_{lu} = 1)P(U = 0) \)
Constructing a Bayesian Network

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?

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

- $P(T = 1 | F_{lu} = 0, U = 0) = P(F_{ev} = 1 | F_{lu} = 1) = 0.9$
- $P(T = 1 | F_{lu} = 0, U = 1) = P(F_{ev} = 1 | F_{lu} = 0) = 0.05$
- $P(T = 1 | F_{lu} = 1, U = 0) = P(F_{ev} = 1 | F_{lu} = 1) = 0.05$
- $P(T = 1 | F_{lu} = 1, U = 1) = P(F_{ev} = 1 | F_{lu} = 1) = 0.05$
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.

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

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

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

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

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

\[ = 0.095 \]
3. \( P(F_{lu} = 1 | U = 1, T = 1) \)?

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

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

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

1. Compute joint probabilities

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

2. Definition of conditional probability

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

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

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

Can we use sampling to do approximate inference?

Yes.
Conditional Expectation
+ General Inference

Lisa Yan
July 22, 2020
Conditional Expectation

Conditional Distributions

Expectation
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

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

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
Quick check

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

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
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

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

- 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
Analyzing recursive code

```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 $Y = \text{return value of } \text{recurse}().$

What is $E[Y]$?
Analyzing recursive code

```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 $E[Y]$?


When $X = 1$, return 3.
Analyzing recursive code

```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 $Y = \text{return value of } \text{recurse()}. \text{What is } E[Y]?$


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

What is $E[Y | X = 2]$?
B. $E[Y + 5] = 5 + E[Y]$
C. $5 + E[Y | X = 2]$
Analyzing recursive code

```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 $Y = \text{return value of recurse()}$. What is $E[Y]$?


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

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

If $Y$ discrete

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

When $X = 2$, return $5 + \text{a future return value of recurse()}$. 

Lisa Yan, CS109, 2020
Analyzing recursive code

```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 $Y = \text{return value of } \text{recurse}()$. What is $E[Y]$?


When $X = 3$, return

$7 + \text{a future return value of } \text{recurse}()$.

$E[Y|X = 3] = E[7 + Y]$
Analyzing recursive code

```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 $Y = \text{return value of recurse()}$. What is $E[Y]$?


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

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

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

$$E[Y] = 15$$

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

$$E[X|Y = y] = \sum_x xp_{X|Y}(x|y) = \sum_x xp_X(x) = E[X]$$
Interlude for jokes/announcements
CRACK

BOOM

WHOA! WE SHOULD GET INSIDE!

IT'S OKAY! LIGHTNING ONLY KILLS ABOUT 45 AMERICANS A YEAR, SO THE CHANCES OF DYING ARE ONLY ONE IN 7,000,000. LET'S GO ON!

THE ANNUAL DEATH RATE AMONG PEOPLE WHO KNOW THAT STATISTIC IS ONE IN SIX.

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!

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:
- \( P(\text{Fever} = 1|T = 0, \text{Flu} = 1) = P(\text{Fever} = 1|\text{Flu} = 1) \)
- \( P(\text{Flu} = 1, U = 0) = P(\text{Flu} = 1)P(U = 0) \)
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

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

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

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

\[ = 0.122 \]
Solving inference questions precisely is possible, but sometimes tedious.

Can we use sampling to do approximate inference?

Yes.
**Inference via math**

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

= 0.122

(from pre-lecture video)

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

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

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

Step 0:
Have a fully specified Bayesian Network

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

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

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

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

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

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)

Stanford University
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

# 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]
# 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)
    #
    # 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)

    # 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]
# 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)
```

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

```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 $(U = 1, T = 1)$. 

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

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

```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 $P(F_u = 1|U = 1, T = 1)$?

```
def reject_inconsistent(samples, outcome):
    consistent_samples = []
    for sample in samples:
        if check_consistent(sample, outcome):
            consistent_samples.append(sample)
    return consistent_samples
```

# Inference question

```
# 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)
```

---

Conditional event = samples with $F_{56} = 1$, $U = 1$, $T = 1$.

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

```python
def reject_inconsistent(samples, outcome):
    ... return consistent_samples
```

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

Lisa Yan, CS109, 2020
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)}

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

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

Chemical present?

Chemical detected?

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

What if random variables are continuous?

What if you run out of time/computational power?

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

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

\[ P(T = 1|F_{lu} = 0, U = 0) = 0.1 \]
\[ P(T = 1|F_{lu} = 0, U = 1) = 0.8 \]
\[ P(T = 1|F_{lu} = 1, U = 0) = 0.9 \]
\[ P(T = 1|F_{lu} = 1, U = 1) = 1.0 \]
Congratulations on finishing the midterm 😊