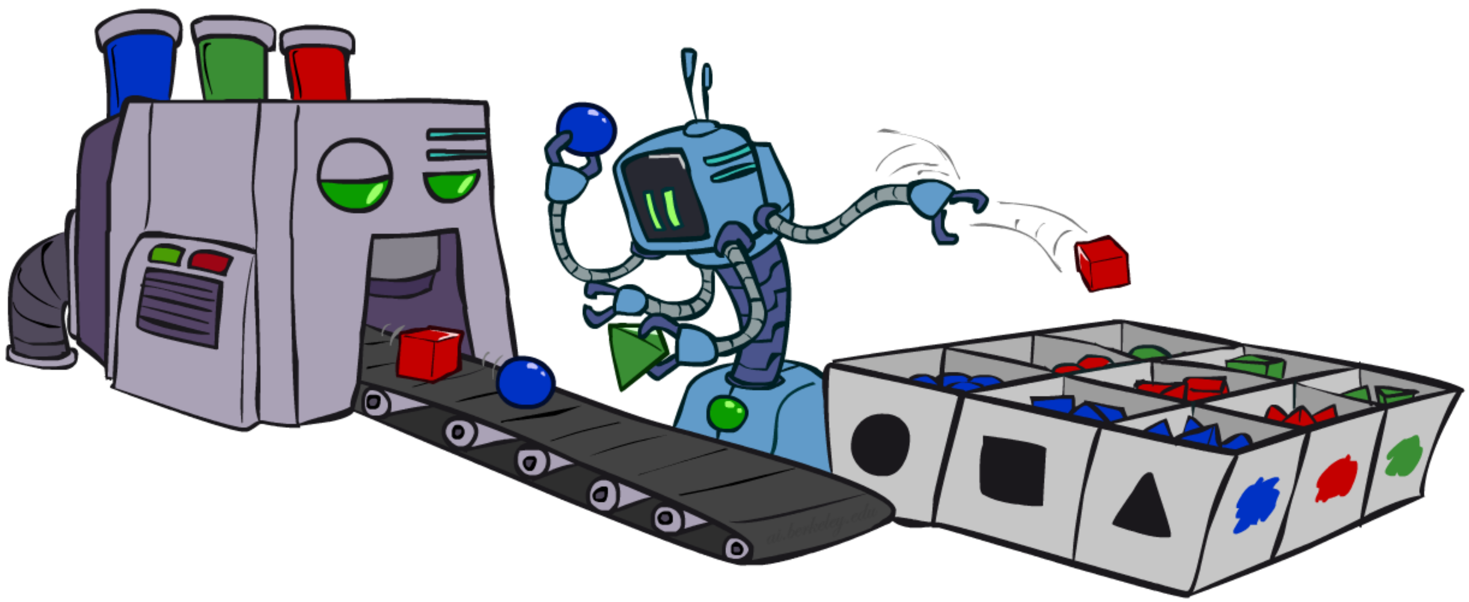
AI: Representation and Problem Solving

Bayes Nets Sampling



Instructors: Nihar B. Shah and Tuomas Sandholm

Slide credits: CMU AI and ai.berkeley.edu

Today: Sampling

- For random variables $X_1,\ldots,X_n$
  - How to get a sample from $P(X_1,\ldots,X_n)$?
  - How to get a sample from $P(X_5, X_6 \mid X_3=x_3, X_7=x_7)$?

- **Why do we need this?**

Reason 1: Inference

- Estimating posterior probabilities (P(Query | evidence)) can be computationally expensive

- Instead, sampling from the posterior distribution can be easier
  - Recall Monte Carlo approach from earlier
  - Given enough samples, counts converge to true probability

- **Use that to approximate the posterior probability**

Warm up

Prior Sampling: Given N samples from P(A,B,C), what does the value $\frac{count(+a,-b,+c)}{N}$ approximate?

A. $P(+a,-b,+c)$
B. $P(+c \mid +a,-b)$
C. $P(+c \mid -b)$
D. $P(+c)$

In fact, $\lim_{N\to\infty} \frac{\boldsymbol{count(+a,-b,+c)}}{\boldsymbol{N}} = \boldsymbol{P(+a,-b,+c)}$

Warm-up

Given these N=10 samples from P(A,B,C):

What is the approximate value for $P(-a, +b, -c)$?

A. 1/10
B. 5/10
C. 1/4
D. 1/5

Counts

+a	+b	+c	0
+a	+b	-c	0
+a	-b	+c	3
+a	-b	-c	0
-a	+b	+c	4
-a	+b	-c	1
-a	-b	+c	2
-a	-b	-c	0

Warm-up

Given these N=10 samples from P(A,B,C):

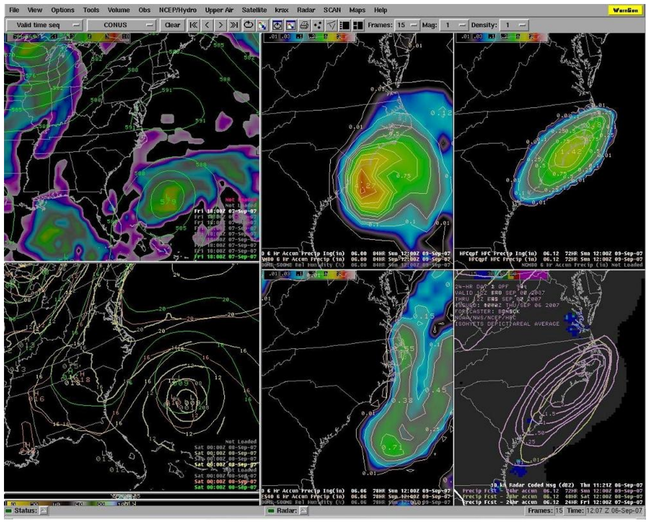
What is the approximate value for $P(-c| -a, +b)$?

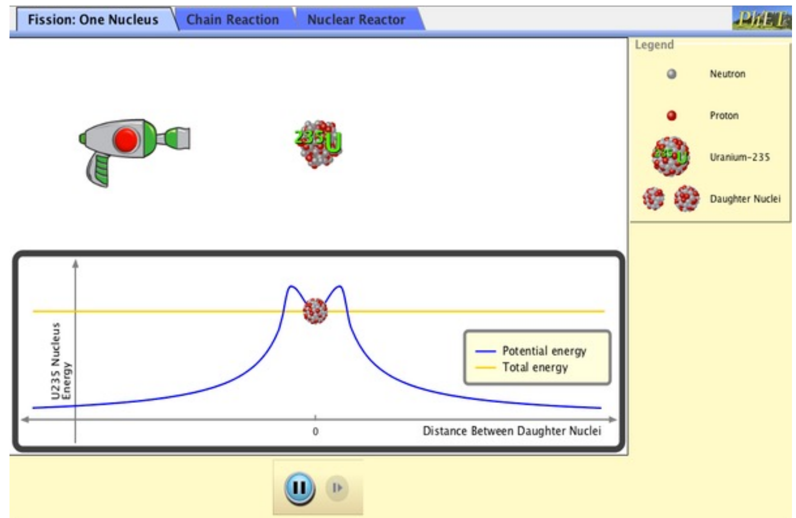
A. 1/10
B. 5/10
C. 1/4
D. 1/5

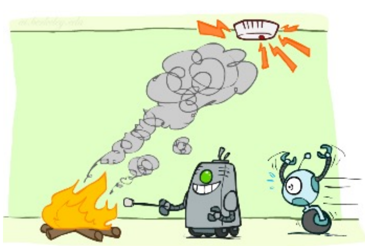
Counts

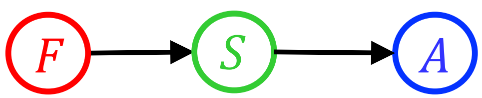
+a	+b	+c	0
+a	+b	-c	0
+a	-b	+c	3
+a	-b	-c	0
-a	+b	+c	4
-a	+b	-c	1
-a	-b	+c	2
-a	-b	-c	0

Reason 2: Simulations





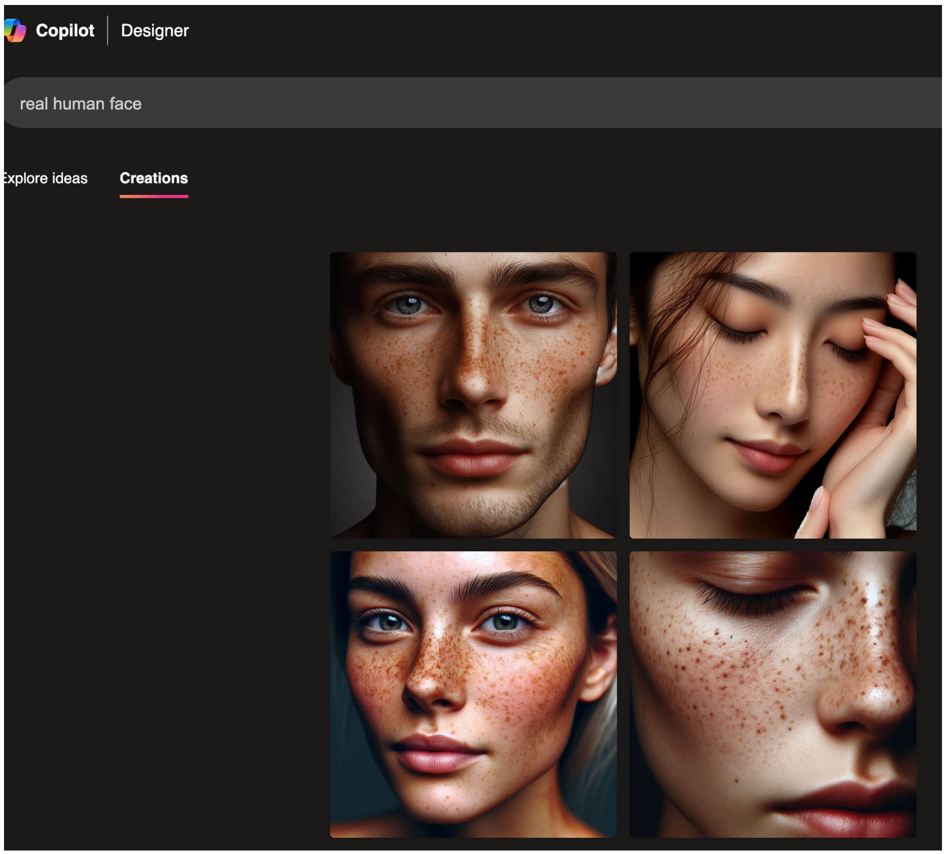




Fire department wants to conduct a drill

- Simulate daily conditions by drawing from P(F, S, A)
- Simulate situation of an alarm by drawing from P(F, S | A=+a)

Cool connection: GenAI Image generation

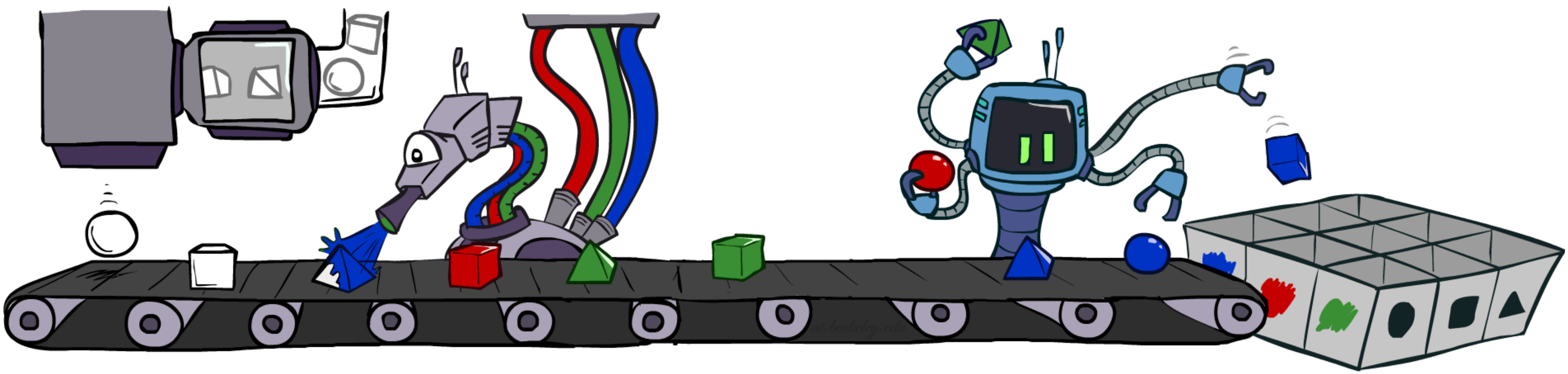


- This is at its core a sampling problem
- It is generating random samples from the distribution, in this example, of human images
- The distribution is unknown and hard to specify
- Techniques much more advanced than what we'll study here

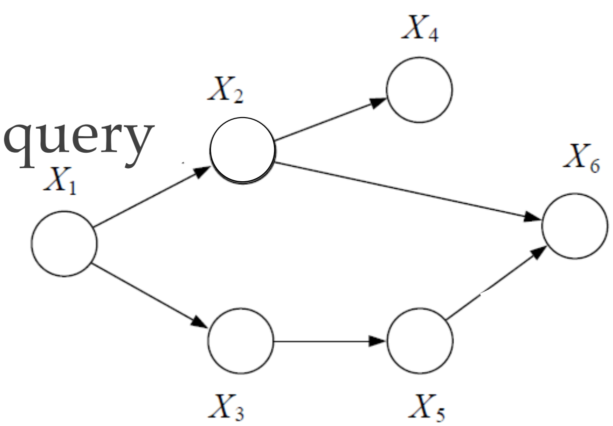
Sampling in Bayes' Nets

- Prior Sampling
- Rejection Sampling
- Likelihood Weighting
- Gibbs Sampling

Prior Sampling

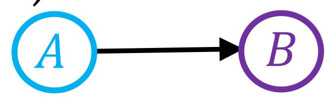


- **Given a Bayes net, how to sample from $P(X_1,\ldots,X_n)$?**
  - Certain applications (e.g., simulations) need sample from entire joint distribution
  - To answer a conditional or marginal probability query
    - Approximate joint distribution based on samples
    - Answer desired query from it



Example

- How would you sample from P(A,B)?
  - You have access to P(A) and P(B | A)



$P(A)$

+a	1/2
-a	1/2

$P(B|A)$

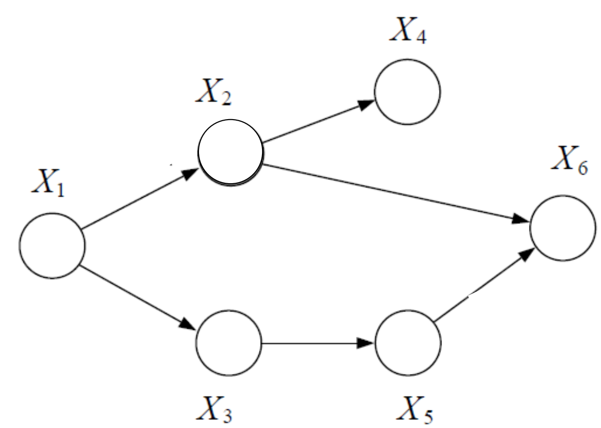
+a	+b	1/10
	-b	9/10
-a	+b	1/2
	-b	1/2

- P(A, B) = P(A) P(B | A)
- First draw a sample a~P(A)
- Then draw b~P(B | A=a)
- Thus (a, b) is a sample from P(A,B)

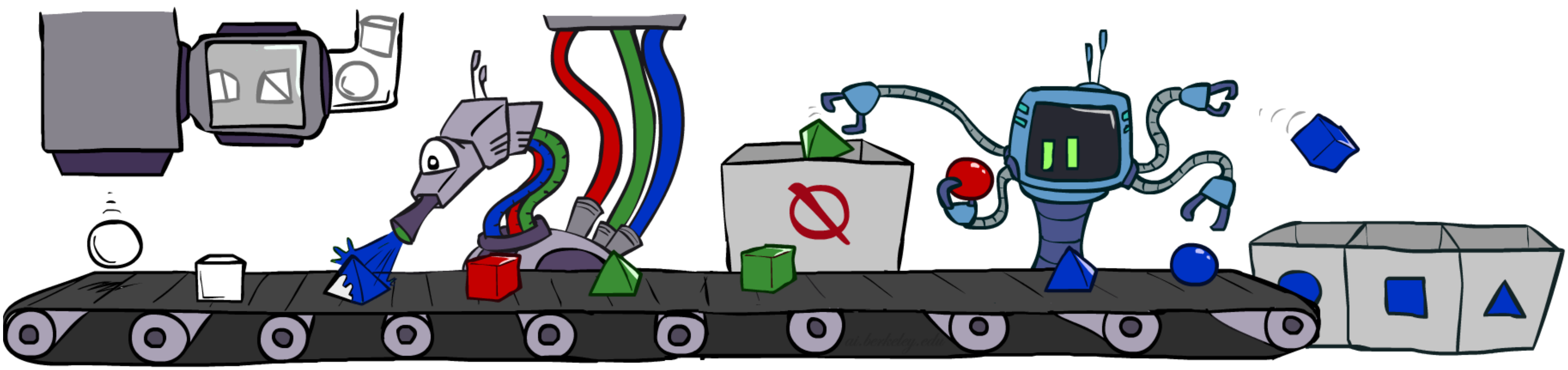
Prior Sampling

- **Given a Bayes net, how to sample from $P(X_1,\ldots,X_n)$?**
  - You have access to the CPTs used to construct the Bayes net

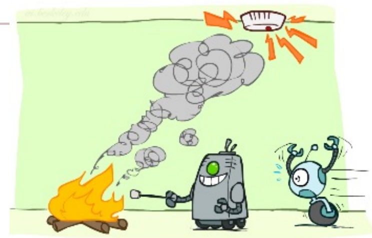
- Consider a topological ordering of the nodes of the Bayes net (say, it is $X_1,\ldots,X_n$)
- For i=1, 2, …, n
  - Sample $x_i \sim P(X_i \mid Parents(X_i)$=their sampled values)
- Return $(x_1, x_2, \ldots, x_n)$

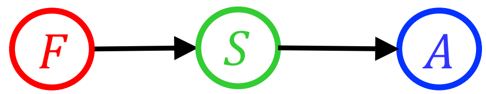


Rejection Sampling



How to sample conditionals?





- How to get a sample from P(F | A=+a)
  - E.g., for a fire department drill

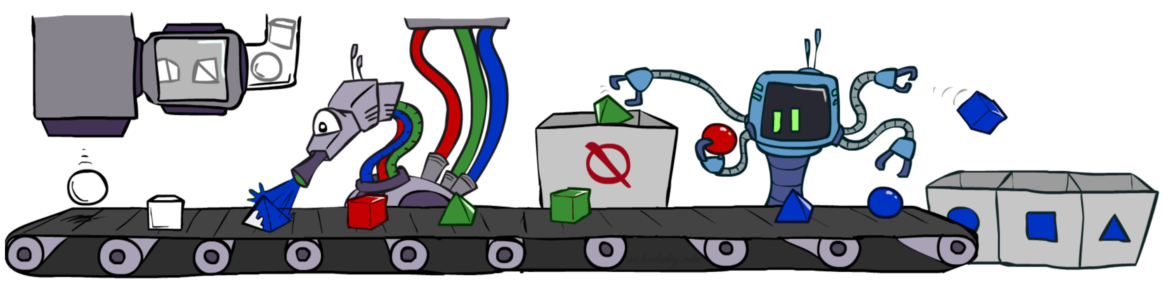
Rejection sampling:

- Use prior sampling to get a sample (f, s, a) ~ P(F, S, A)
- If a = -a, then discard this sample and go back to the step above
- If a = +a, then return the sampled value of f

Rejection Sampling

For given values of variable(s) $X_e=x_e$, want to draw a sample from P(other X's | $X_e=x_e$)

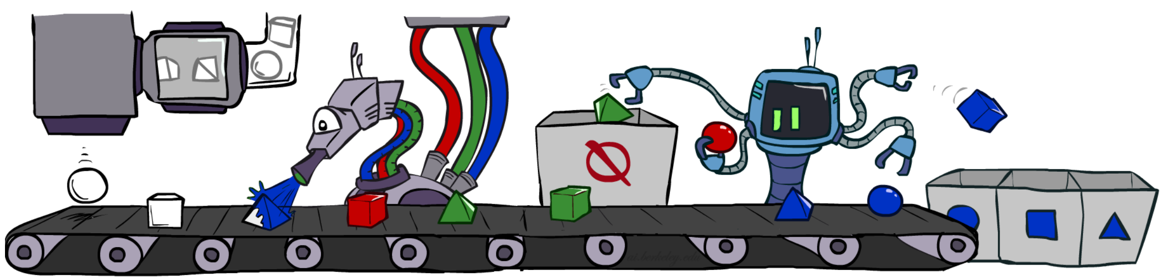
- Sample $(x_1, \ldots, x_n) \sim P(X_1,\ldots,X_n)$
  - If sample for X_e is different from given evidence
    - Discard sample and go back to first step
- Return sampled value



Rejection Sampling in Bayes nets

For given values of variable(s) $X_e=x_e$, want to draw a sample from P(other X's | $X_e=x_e$)

- For i=1, 2, …, n (assumed topological ordering of graph)
  - Sample x_i from $P(X_i \mid Parents(X_i) = \text{sampled values})$
  - If i is in evidence set e, and sampled x_e is different from given evidence
    - Restart from first step, starting again from i=1
- Return sampled value



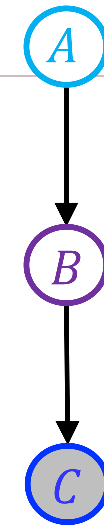
Question

Consider rejection sampling under evidence C=+c. Suppose you draw 10 samples, and observe the following counts.

Why don't you observe any samples with –c?

Counts $N(A, B, C)$

+a	+b	+c	4
+a	+b	-c	
+a	-b	+c	3
+a	-b	-c	
-a	+b	+c	2
-a	+b	-c	
-a	-b	+c	1
-a	-b	-c	



Question

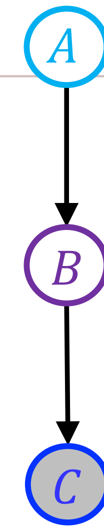
Consider rejection sampling under evidence C=+c. Suppose you draw 10 samples, and observe the following counts.

Approximately, what is P(+a,+b | +c)?

1) 1/10
2) 1/20
3) ¼
4) 1/2

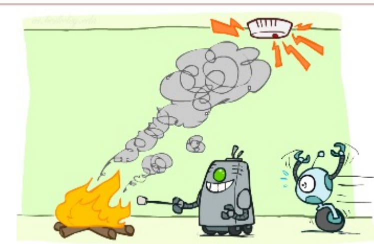
Counts $N(A,B,C)$

+a	+b	+c	4
+a	+b	-c	
+a	-b	+c	3
+a	-b	-c	
-a	+b	+c	2
-a	+b	-c	
-a	-b	+c	1
-a	-b	-c	



Problem with rejection sampling

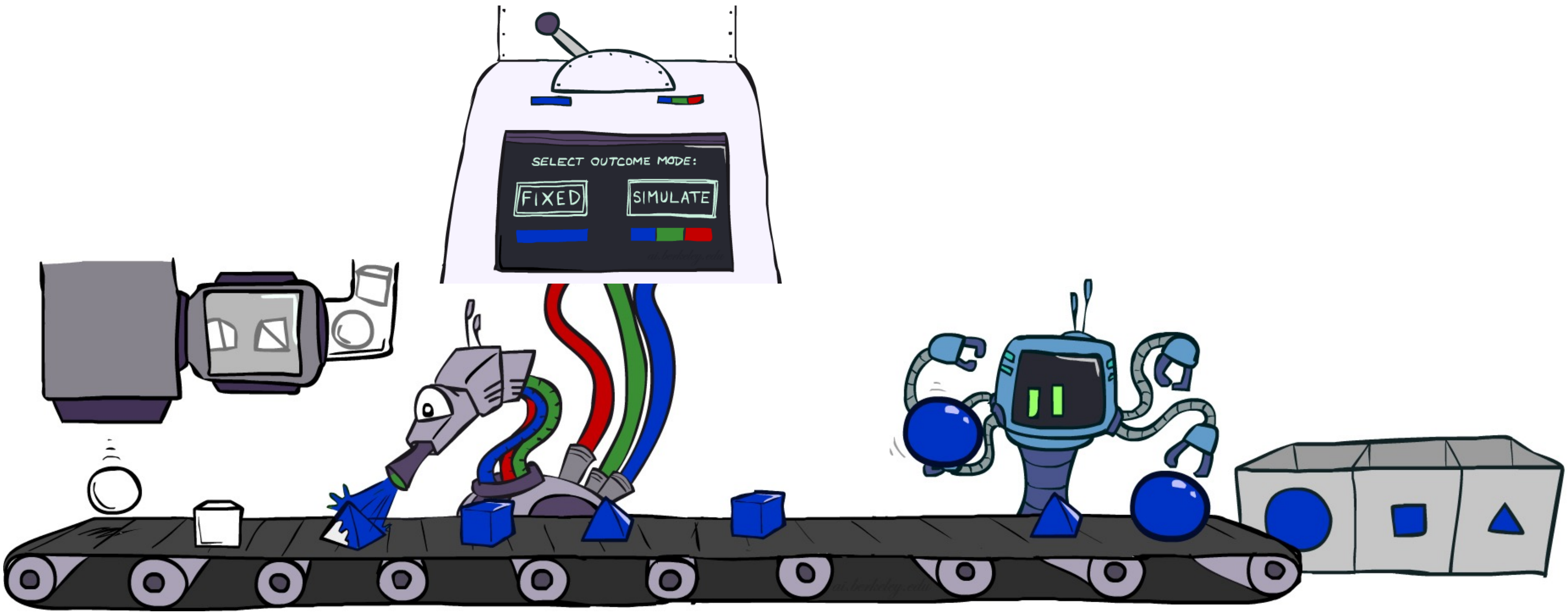
“If a = -a, then discard this sample…”



Can be very wasteful!

E.g., if P(evidence) is low, then will have to discard a large fraction of samples!

Likelihood Weighting



Likelihood reweighing: Main idea

- In rejection sampling, we were drawing samples from $P(X_1,\ldots,X_n)$ without regard to given evidence

- Instead:
  - Let's fix evidence variables $X_e=x_e$
  - Sample the remaining variables
  - Due to the "fixing", the distribution of the sampling may have issues
  - Do some reweighing to address these issues

Likelihood weighted Sampling

For given values of **variable** $X_e=x_e$, want to obtain P(other X's | $X_e=x_e$)

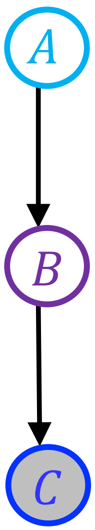
- For i=1, 2, …, n (assumed toplogical ordering of graph)
  - If i = e
    - Set x_e as the given value
    - Let $w = P(X_e=x_e \mid Parents(X_e) = \text{sampled values})$
  - Else
    - Sample x_i from $P(X_i \mid Parents(X_i)=\text{sampled or given values})$
- Return sample along with weight w

Instead of sampling X_e, we set it to given evidence x_e and give a weight to this sample as its probability conditioned on sampled values of its parents

$$P(X_1=x_1, \dots, X_{e-1}=x_{e-1}, X_{e+1}=x_{e+1}, \dots, X_n=x_n \mid X_e=x_e) = \frac{\sum_{\text{samples}} \mathbb{I}\{\text{sample}=(x_1, \dots, x_{e-1}, x_{e+1}, \dots, x_n)\} * \text{weight of sample}}{\sum_{\text{samples}} \text{weight of sample}}$$

Question

Suppose e = B. We want to estimate P(A, C | B = +b) via likelihood reweighing.



$P(A)$

+a	1/2
-a	1/2

$P(B|A)$

+a	+b	1/10
	-b	9/10
-a	+b	1/2
	-b	1/2

$P(C|B)$

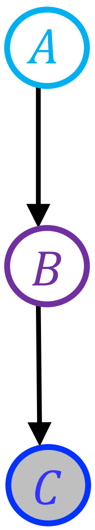
+b	+c	4/5
	-c	1/5
-b	+c	1
	-c	0

- What variable should we sample first? What distribution?
- Sample A via P(A)
- Suppose you draw +a
- What should you do next?
- Next variable in topological ordering is the evidence variable B
- Set weight w = P(B=+b | A = +a)
- w=1/10
- What next?
- Sample C via P(C|B=+b)
- Suppose you draw –c

- Output sample (+a, -c) with weight 1/10

Question

Suppose e = B. We want to estimate P(A, C | B = +b) via likelihood reweighing.



$P(A)$

+a	1/2
-a	1/2

$P(B|A)$

+a	+b	1/10
	-b	9/10
-a	+b	1/2
	-b	1/2

$P(C|B)$

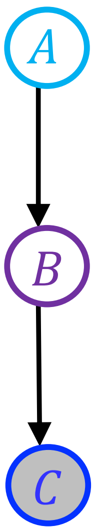
+b	+c	4/5
	-c	1/5
-b	+c	1
	-c	0

- What are possible values that weight w can take?

- Recall w = P(B = +b | A)
- Thus w can take values 1/10 or 1/2

Question

Suppose e = B. We want to estimate P(A, C | B = +b) via likelihood reweighing.



$P(A)$

+a	1/2
-a	1/2

$P(B|A)$

+a	+b	1/10
	-b	9/10
-a	+b	1/2
	-b	1/2

$P(C|B)$

+b	+c	4/5
	-c	1/5
-b	+c	1
	-c	0

- Suppose we have the following samples:
  - 4 samples (+a, -c) each with weight 1/10
  - 2 samples (-a, -c) each with weight 1/2
  - 2 samples (-a, +c) each with weight 1/2
  - 1 sample (+a, +c) with weight 1/10

- What is our estimate of P(A=+a, C=+c | B = +b)?

- $\frac{1/10}{\frac{1}{10}+2*\frac{1}{2}+2*\frac{1}{2}+4*\frac{1}{10}} = \frac{1}{25}$

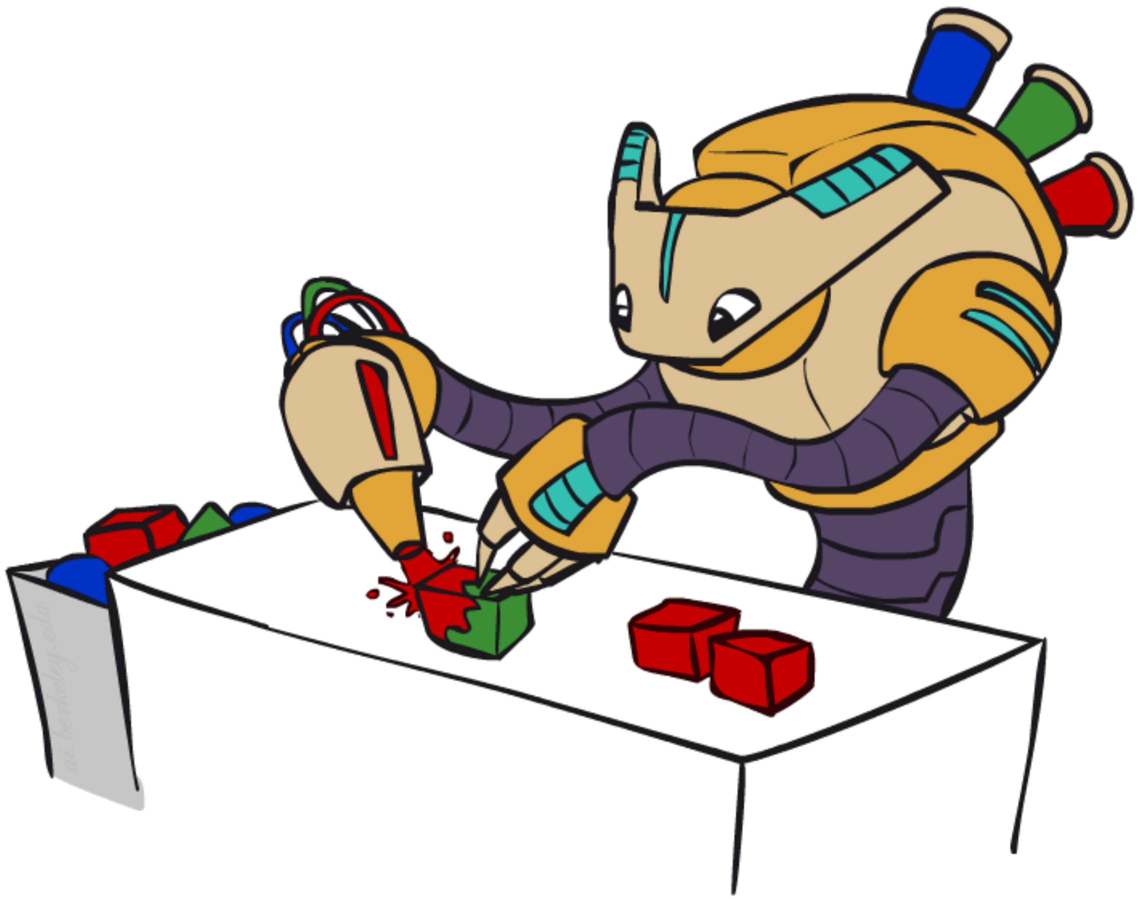
Likelihood weighted Sampling

What if there are multiple evidence (given) variables?

- Initialize weight w=1
- For i=1, 2, …, n (assumed toplogical ordering of graph)
  - If i is an evidence (given) variable
    - Set x_e as the given value
    - $w = w * P(X_e = x_e \mid Parents(X_e) = \text{sampled values})$
  - Else
    - Sample x_i from $P(X_i \mid Parents(X_i) = \text{sampled or given values})$
- Return sample along with weight w

$$P(\text{other variables} = \text{value} \mid X_e = x_e) = \frac{\sum_{samples} \mathbb{I}\{\text{sample}=\text{value}\} * \text{weight of sample}}{\sum_{samples} \text{weight of sample}}$$

Gibbs Sampling



GenAI Image Generation

- Suppose you want to generate images
- These images don't actually exist and you want to generate new ones
- Popular technique: Diffusion models
- Also: Generative Adversarial Networks (GANs)
- At its core, this involves sampling from some unknown crazy distribution!
- Today: Let's understand Gibbs sampling via a toy version of this

Image generation: Toy example

- Image comprises a foreground and a background
  - Image foreground ϵ {Cow, human, airplane, car}
  - Image background ϵ {Buildings, sky, grass}
- Want to generate an image randomly~P(Foreground, Background)
- What is P(Foreground = cow)? What is P(Background = Buildings)?
  - Hard to tell
- What is P(Foreground = cow | Background = grass)?
- What is P(Background = grass | Foreground = airplane)?
- **Marginals are hard to specify or estimate but conditionals are easier!**

Gibbs sampling

- You have access to $P(X_i \mid X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n)$ for all i
- Want to sample from $P(X_1,\ldots,X_n)$

- Initialize some values $(x_1,\ldots,x_n)$
- Repeat many times:
  - For $i = 1,\ldots,n$:
    - Let $x_i \sim P(X_i \mid X_1 = x_1,\ldots,X_{i-1}=x_{i-1},X_{i+1}=x_{i+1},\ldots,X_n=x_n)$
    - Note: this will overwrite the previous value of x_i
- Output $x_1,\ldots,x_n$

Gibbs sampling: Toy example

- Initialize Foreground = cow, Background = sky
  - Draw from P(Foreground | Background=sky) to get airplane
  - Draw from P(Background | Foreground= airplane) to get sky
  - Draw from P(Foreground | Background=sky) to get human
  - Draw from P(Background | Foreground=human) to get buildings
  - Draw from P(Foreground | Background=buildings) to get car
  - Draw from P(Background | Foreground=car) to get grass

- Output (Foreground=car, Background=grass)

Gibbs sampling of conditional

- You are given $X_e = x_e$
- You have access to $P(X_i | X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n)$ for every remaining variable i
- Want to sample from P(other variables | $X_e = x_e$)

- Initialize some values for all other variables
- Repeat many times:
  - For every variable i not in e
    - Let $x_i \sim P(X_i | X_1 = x_1,\ldots,X_{i-1}=x_{i-1},X_{i+1}=x_{i+1},\ldots,X_n=x_n)$
    - Note: this will overwrite the previous value of x_i
- Output $x_1,\ldots,x_n$

Poll

- Consider two variables A and B, taking values {-a,+a} and {-b,+b} respectively.
- To avoid pathological cases, suppose P(A=a,B=b)>0 for every (a,b).
- You are given access to P(A|B) and P(B|A).
- Gibbs sampling produces P(A,B) from these two conditionals. But one may wonder whether the two conditionals even specify P(A,B) uniquely or whether they leave some ambiguity. To this end, work out the following.

- State true or false: From this data, one can always recover P(A, B).