

Warm-up as you walk in

Given these $N=10$ observations of the world:

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

- A. $1/10$
- B. $5/10$
- C. $1/4$
- D. $1/5$
- E. I'm not sure

$-a, -b, +c$
 $+a, -b, +c$
 $-a, -b, +c$
 $-a, +b, +c$
 $+a, -b, +c$
 $-a, +b, -c$
 $-a, +b, +c$
 $-a, +b, +c$
 $+a, -b, +c$
 $-a, +b, +c$

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

Announcements

Midterm 2

- Regrade requests due by tomorrow

Assignments

- P4
 - Due Tonight! (or up until Saturday) 10pm
- HW9
 - Due Tues 4/11 10pm
- P5
 - Out 4/11, Due 4/27
- HW10
 - Out 4/18, Due Tues 4/25 10pm

Announcements

Final Exam

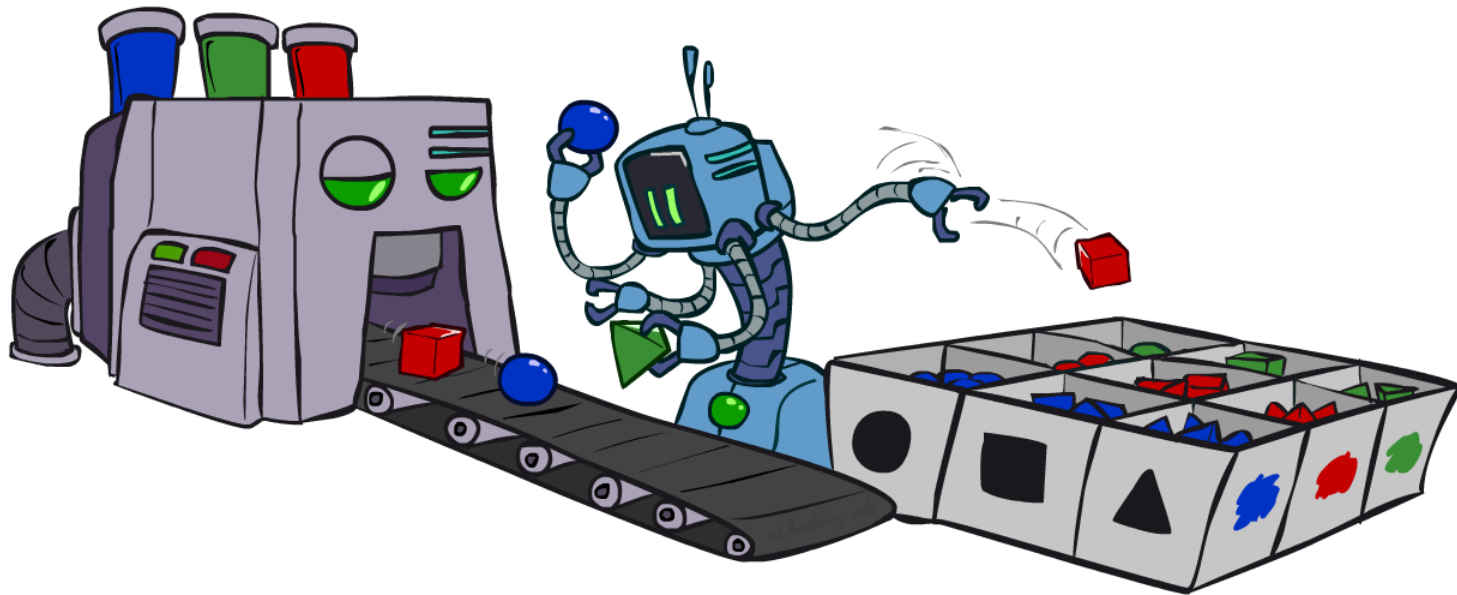
- Cumulative
 - Roughly half of final exam on final 1/3 of the course
 - The other half of the exam covering other topics
- 3 cheat sheets and calculator allowed

TA next semester!

- CSD application <https://www.ugrad.cs.cmu.edu/ta/F23>

AI: Representation and Problem Solving

Bayes Nets Sampling



Instructor: Stephanie Rosenthal

Slide credits: CMU AI and <http://ai.berkeley.edu>

Bayes Nets

✓ Part I: Representation

✓ Part II: Exact inference

✓ ▪ Enumeration (always exponential complexity)

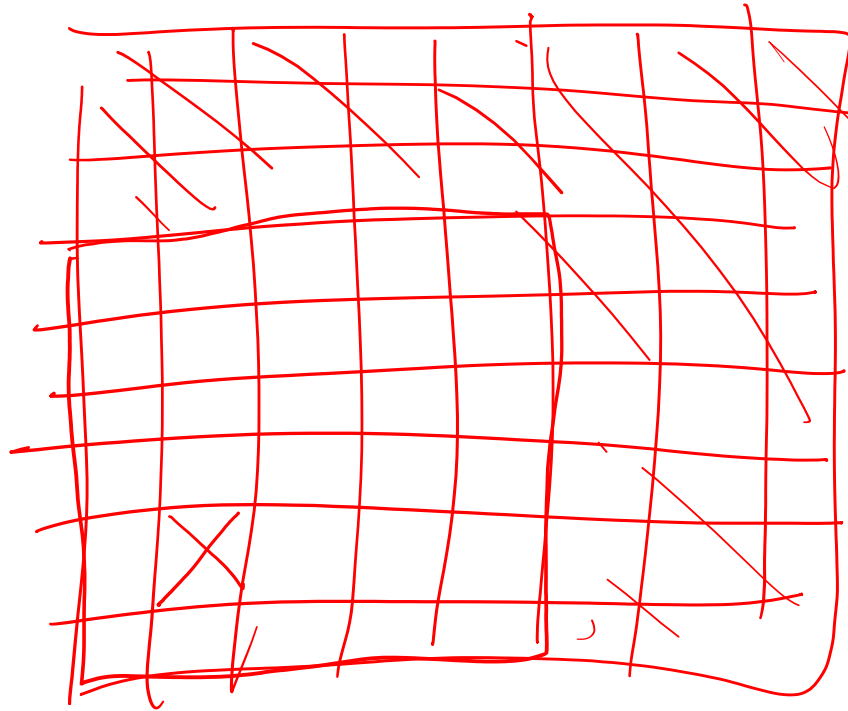
✓ ▪ Variable elimination (worst-case exponential complexity, often better)

✓ ▪ Inference is NP-hard in general

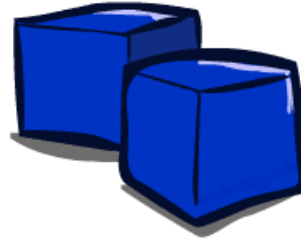
Part III: Approximate Inference

Inference vs Sampling

Motivation for Approximate Inference



Approximate Inference: Sampling



Warm-up as you walk in

Given these N=10 observations of the world:

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

- A. 1/10
- B. 5/10
- C. 1/4
- D. 1/5
- E. I'm not sure

$$\frac{P(-a, +b, -c)}{P(-a, +b)} = \frac{1}{5}$$

- a, -b, +c
- +a, -b, +c
- a, -b, +c
- a, +b, +c
- +a, -b, +c
- a, +b, -c
- a, +b, +c
- a, +b, +c
- +a, -b, +c
- a, +b, +c

list

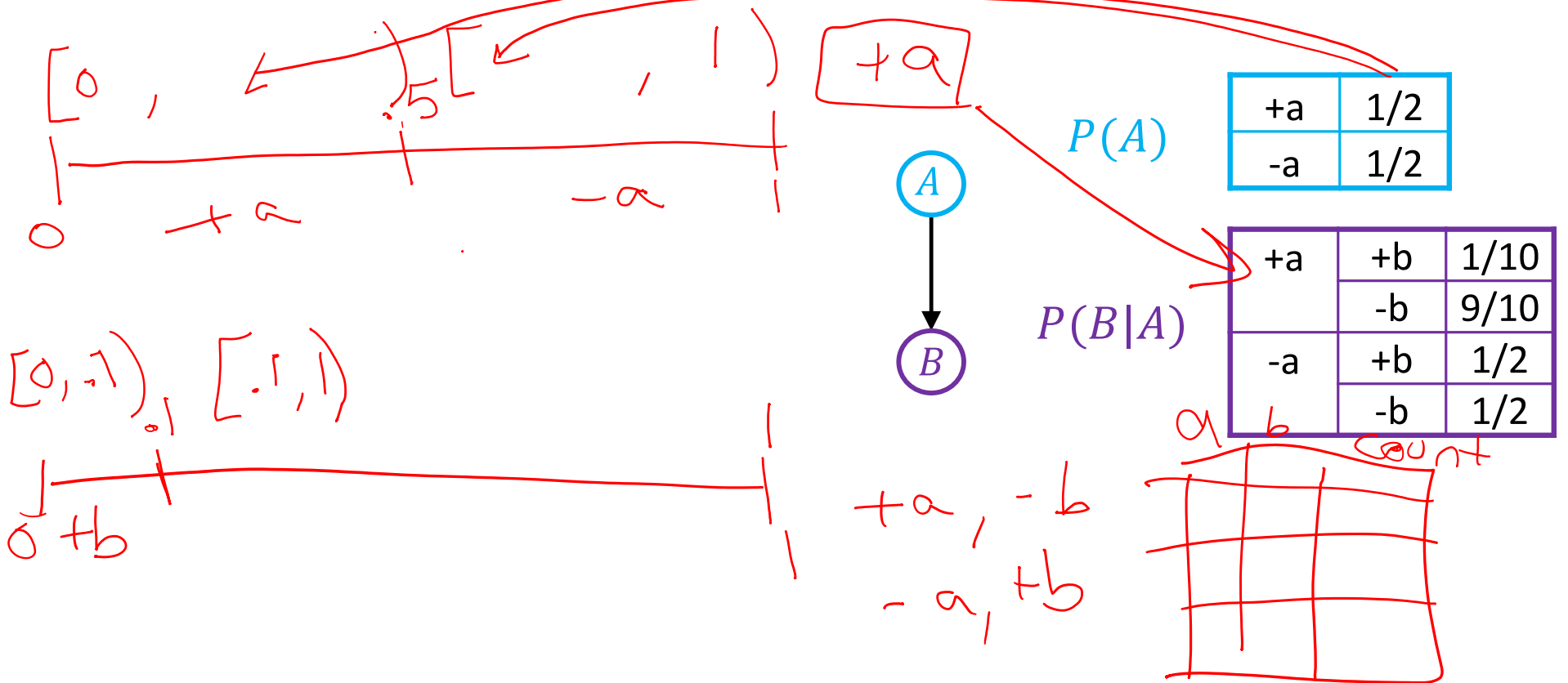
table

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

Sampling

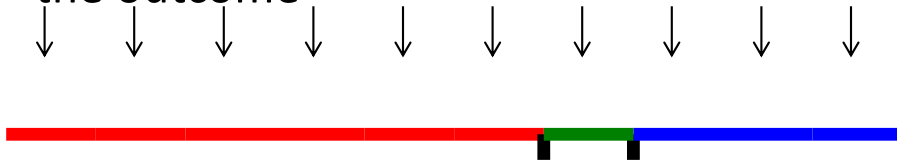
How would you sample from a conditional distribution?



Sampling

Sampling from given distribution

- Step 1: Get sample u from uniform distribution over $[0, 1)$
 - e.g. `random()` in python
- Step 2: Convert this sample u into an outcome for the given distribution by having each outcome associated with a sub-interval of $[0,1)$ with sub-interval size equal to probability of the outcome

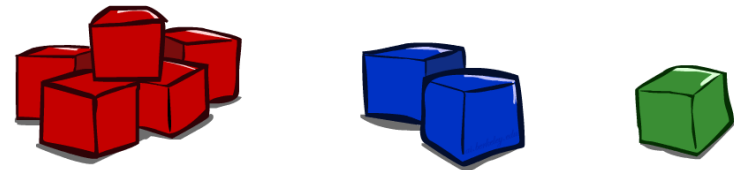


Example

C	P(C)
red	0.6
green	0.1
blue	0.3

- $0 \leq u < 0.6, \rightarrow C = red$
- $0.6 \leq u < 0.7, \rightarrow C = green$
- $0.7 \leq u < 1, \rightarrow C = blue$

- If `random()` returns $u = 0.83$, then our sample is $C = blue$
- E.g, after sampling 8 times:



Sampling in Bayes' Nets

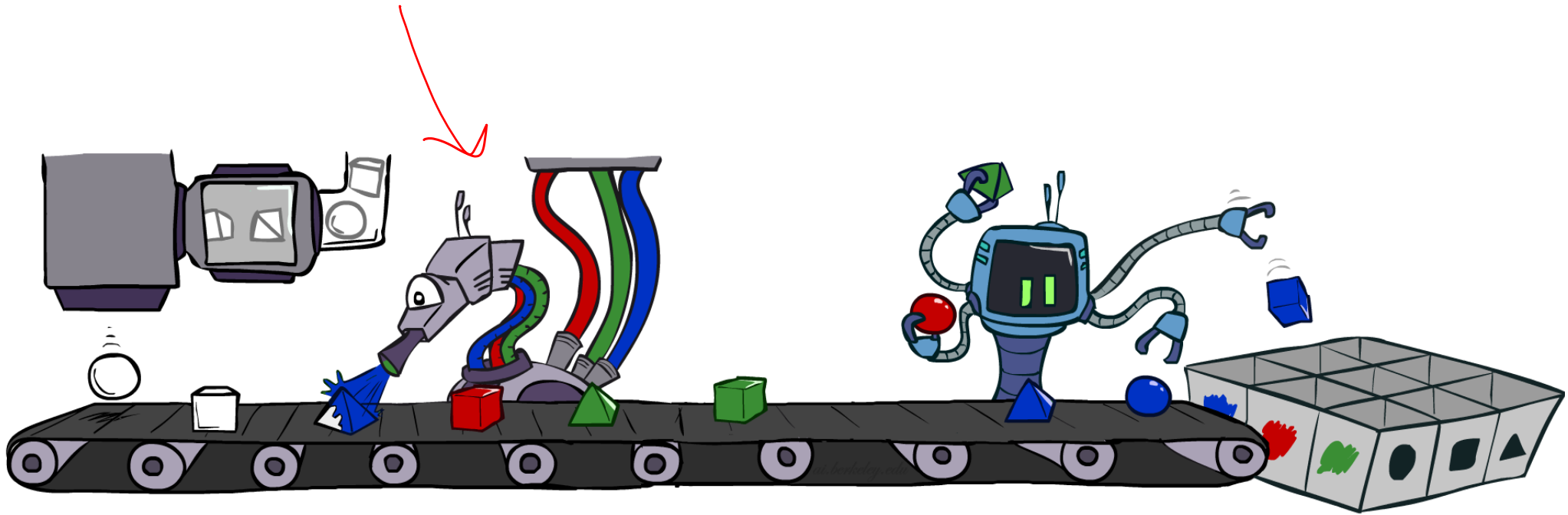
Prior Sampling

Rejection Sampling

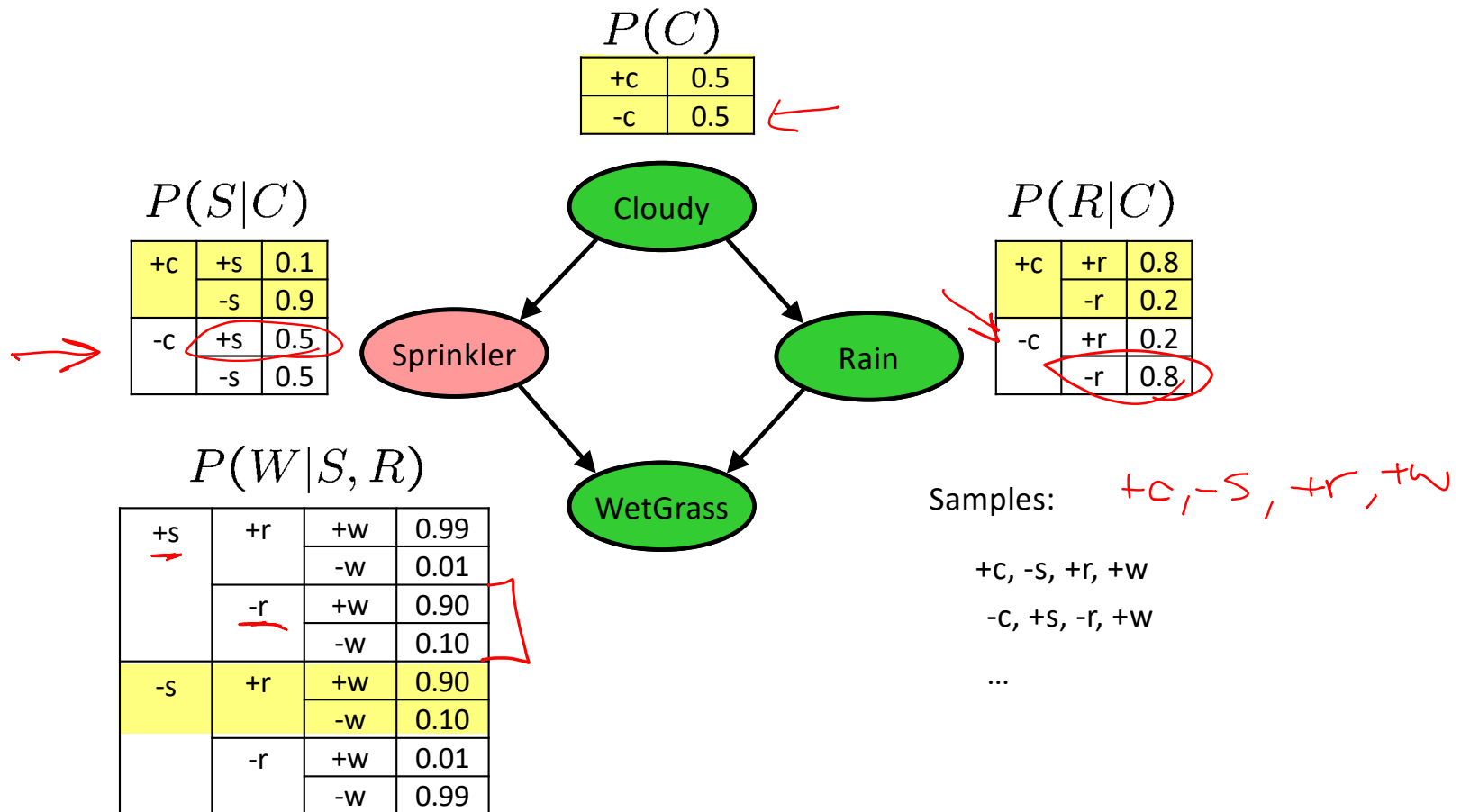
Likelihood Weighting

Gibbs Sampling

Prior Sampling



Prior Sampling



Prior Sampling

n variables

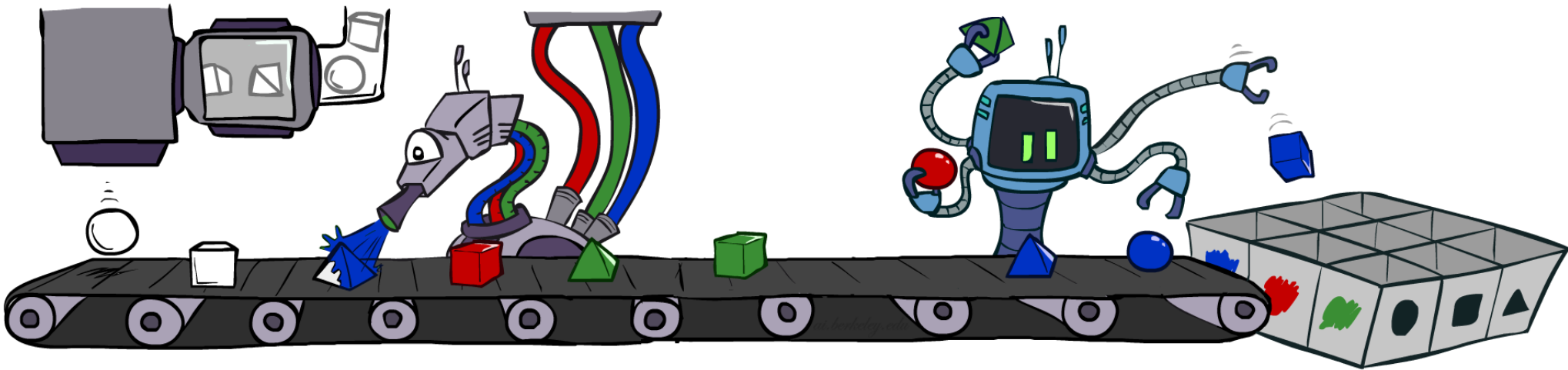
repeat

For $i=1, 2, \dots, n$

- Sample x_i from $P(X_i \mid \text{Parents}(X_i))$

Return (x_1, x_2, \dots, x_n)

1 sample



Poll 1

Prior Sampling: What does the value $\frac{N(+a, -b, +c)}{N}$ approximate?

- A. $P(+a, -b, +c)$
- B. $P(+c \mid +a, -b)$
- C. $P(+c \mid -b)$
- D. $P(+c)$
- E. I don't know

$$P(A, B, C) = P(A)P(B|A)P(C|B)$$

table count

$$N(+a, -b, +c)$$

N ← #samples



$P(A)$



$P(B|A)$



$P(C|B)$



#samples

Poll 1

Prior Sampling: What does the value $\frac{N(+a, -b, +c)}{N}$ approximate?

- A. $P(+a, -b, +c)$
- B. $P(+c \mid +a, -b)$
- C. $P(+c \mid -b)$
- D. $P(+c)$
- E. I don't know



Poll 2

How many $\{-a, +b, -c\}$ samples out of $N=1000$ should we expect?

- A. 1
- B. 50
- C. 125
- D. 200
- E. I have no idea

1000

500 $-a$

250 $-a, +b$

50

$-a, +b, -c$



$P(A)$

+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

Poll 2

How many $\{-a, +b, -c\}$ samples out of $N=1000$ should we expect?

- A. 1
- B. 50
- C. 125
- D. 200
- E. I have no idea



$P(A)$

+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

Probability of a sample

Given this Bayes Net & CPT,
what is $P(+a, +b, +c)$?

Algorithm: Multiply probability of
each node given parents:

- $w = 1.0$
- for $i=1, 2, \dots, n$
 - Set $w = w * P(x_i | \text{Parents}(X_i))$
- return w



$P(A)$

+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

Prior Sampling

This process generates samples with probability:

$$S_{PS}(x_1 \dots x_n) = \prod_{i=1}^n P(x_i | \text{Parents}(X_i)) = P(x_1 \dots x_n)$$

...i.e. the BN's joint probability

Let the number of samples of an event be $N_{PS}(x_1 \dots x_n)$

$$\begin{aligned} \text{Then } \lim_{N \rightarrow \infty} \hat{P}(x_1, \dots, x_n) &= \lim_{N \rightarrow \infty} N_{PS}(x_1, \dots, x_n) / N \\ &= S_{PS}(x_1, \dots, x_n) \\ &= P(x_1 \dots x_n) \end{aligned}$$

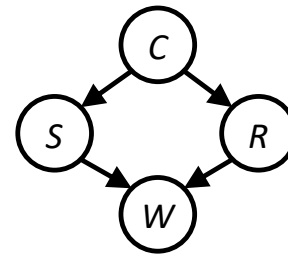
$S_{PS}(x_1 \dots x_n)$ = the probability that a sample $(x_1 \dots x_n)$ is drawn using Prior Sampling

Example

We'll get a bunch of samples from the BN:

+c, -s, +r, +w
+c, +s, +r, +w
-c, +s, +r, -w
+c, -s, +r, +w
-c, -s, -r, +w

$$P(w) = \sum_{c,s,r} P(c,s,r,w)$$



If we want to know $P(W)$

- We have counts $\langle +w:4, -w:1 \rangle$
- Normalize to get $P(W) = \langle +w:0.8, -w:0.2 \rangle$
- This will get closer to the true distribution with more samples
- Can estimate anything else, too
- What about $P(C | +w)$? $P(C | +r, +w)$? $P(C | -r, -w)$?
- Fast: can use fewer samples if less time (what's the drawback?)

Practice Prior Sampling

~~Use a random number generator to pick a number between 0 and 100,
divide random number by 100 to get a probability~~

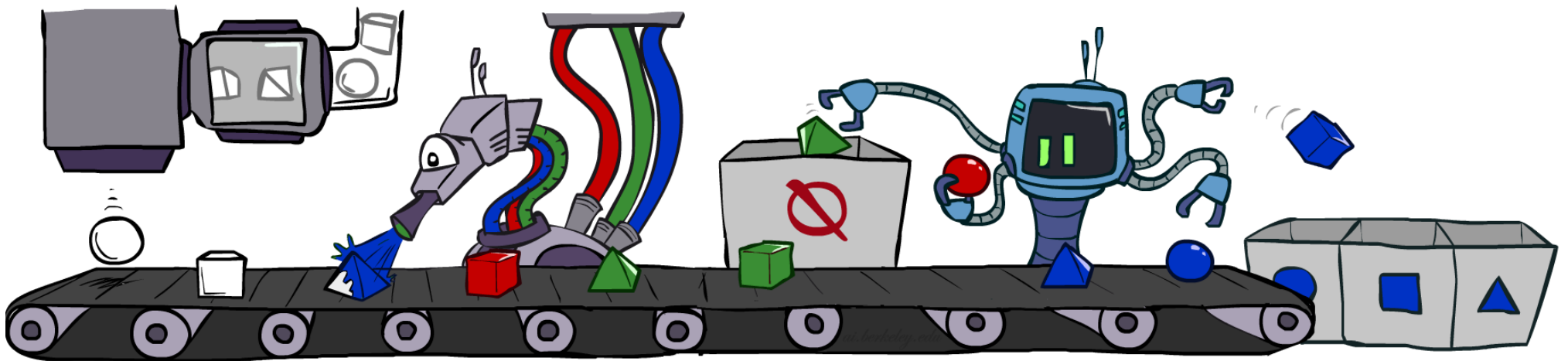
Use the random #s provided!

Check the relevant probability distribution to determine what value is selected.

Continue selecting values for each variable until you have a single sample

Record the sample, repeat the process for 9 more samples

Rejection Sampling



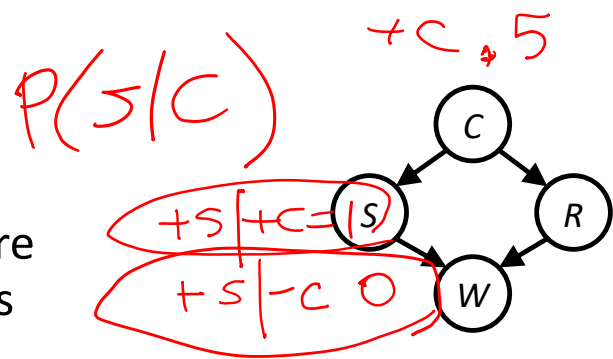
Rejection Sampling

Let's say we want $P(C)$

- No point keeping all samples around
- Just tally counts of C as we go

Let's say we want $P(C | +s)$

- Same thing: tally C outcomes, but ignore (reject) samples which don't have $S=+s$
- This is called rejection sampling
- It is also consistent for conditional probabilities (i.e., correct in the limit)



- +C, -S, +r, +W
- +C, +S, +r, +W
- C, +S, +r, -W
- +C, -S, +r, +W
- C, -S, -r, +W

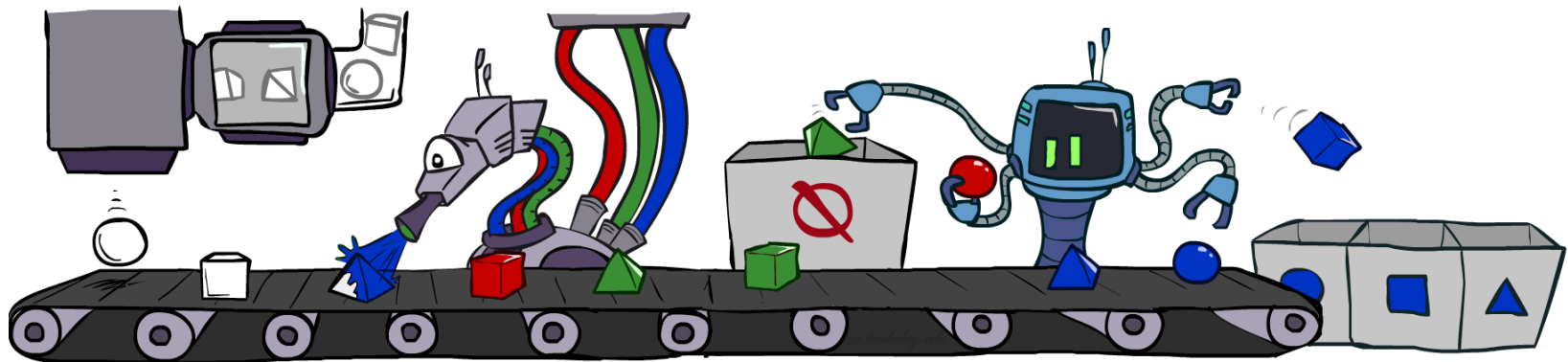
Rejection Sampling

IN: evidence instantiation

For $i=1, 2, \dots, n$

- Sample x_i from $P(X_i \mid \text{Parents}(X_i))$
- If x_i not consistent with evidence
 - Reject: Return, and no sample is generated in this cycle

Return (x_1, x_2, \dots, x_n)



Poll 3

What queries can we (approximately) answer with rejection sampling samples (evidence: $+c$)?

- A. $P(+a, -b, +c)$
- B. $P(+a, -b \mid +c)$** ←
- C. Both
- D. Neither
- E. I have no idea

Counts $N(A, B, C)$

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

Pr



Poll 3

What queries can we (approximately) answer with rejection sampling samples (evidence: $+c$)?

- A. $P(+a, -b, +c)$
- B. $P(+a, -b \mid +c)$
- C. Both ← If we also have total number of attempts
- D. Neither
- E. I have no idea

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	



Practice Rejection Sampling

~~Use a random number generator to pick a number between 0 and 100,
divide random number by 100 to get a probability~~

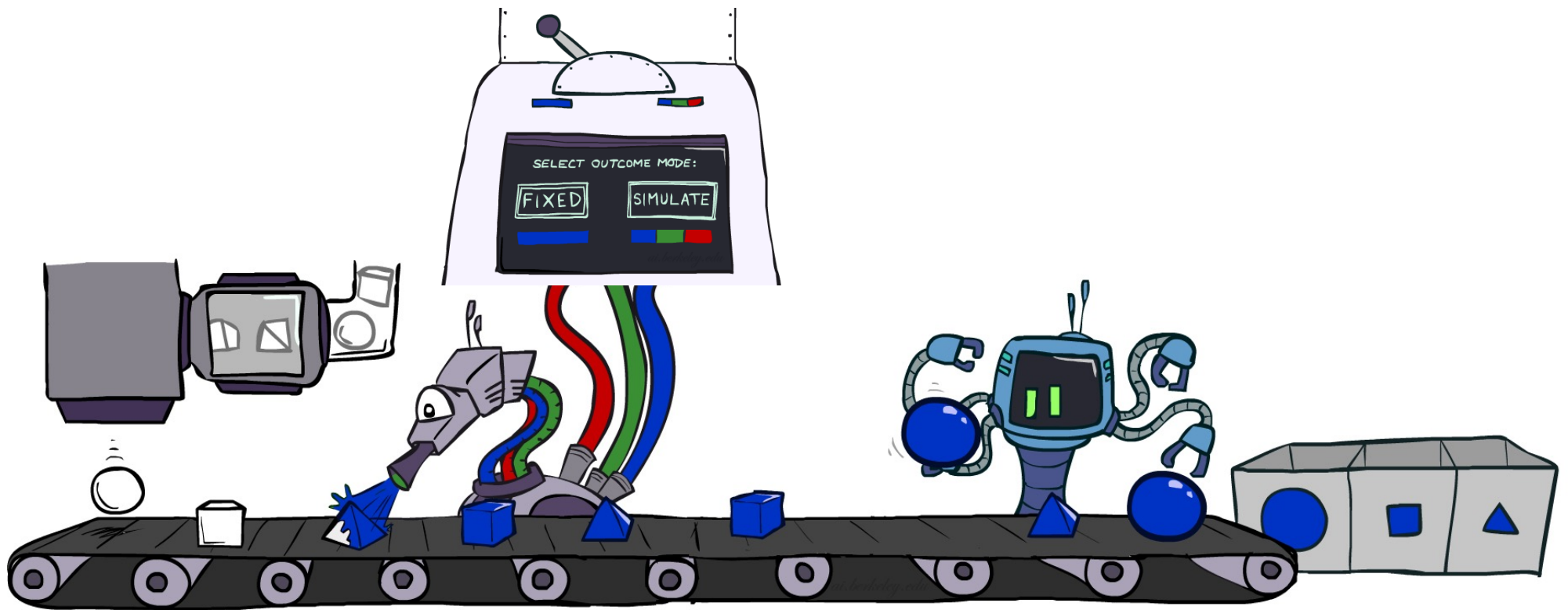
Check the relevant probability distribution to determine what value is selected.

REJECT ENTIRE SAMPLE IF VALUE IS INCONSISTENT WITH EVIDENCE

Continue selecting values for each variable until you have a single sample

Record the sample, repeat the process for 9 more samples

Likelihood Weighting

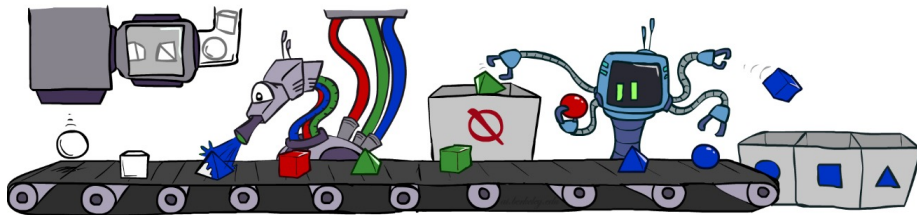
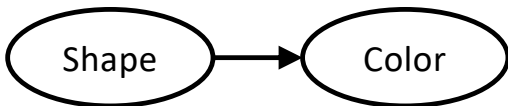


Likelihood Weighting

Problem with rejection sampling:

- If evidence is unlikely, rejects lots of samples
- Evidence not exploited as you sample
- Consider $P(\text{Shape} | \text{blue})$

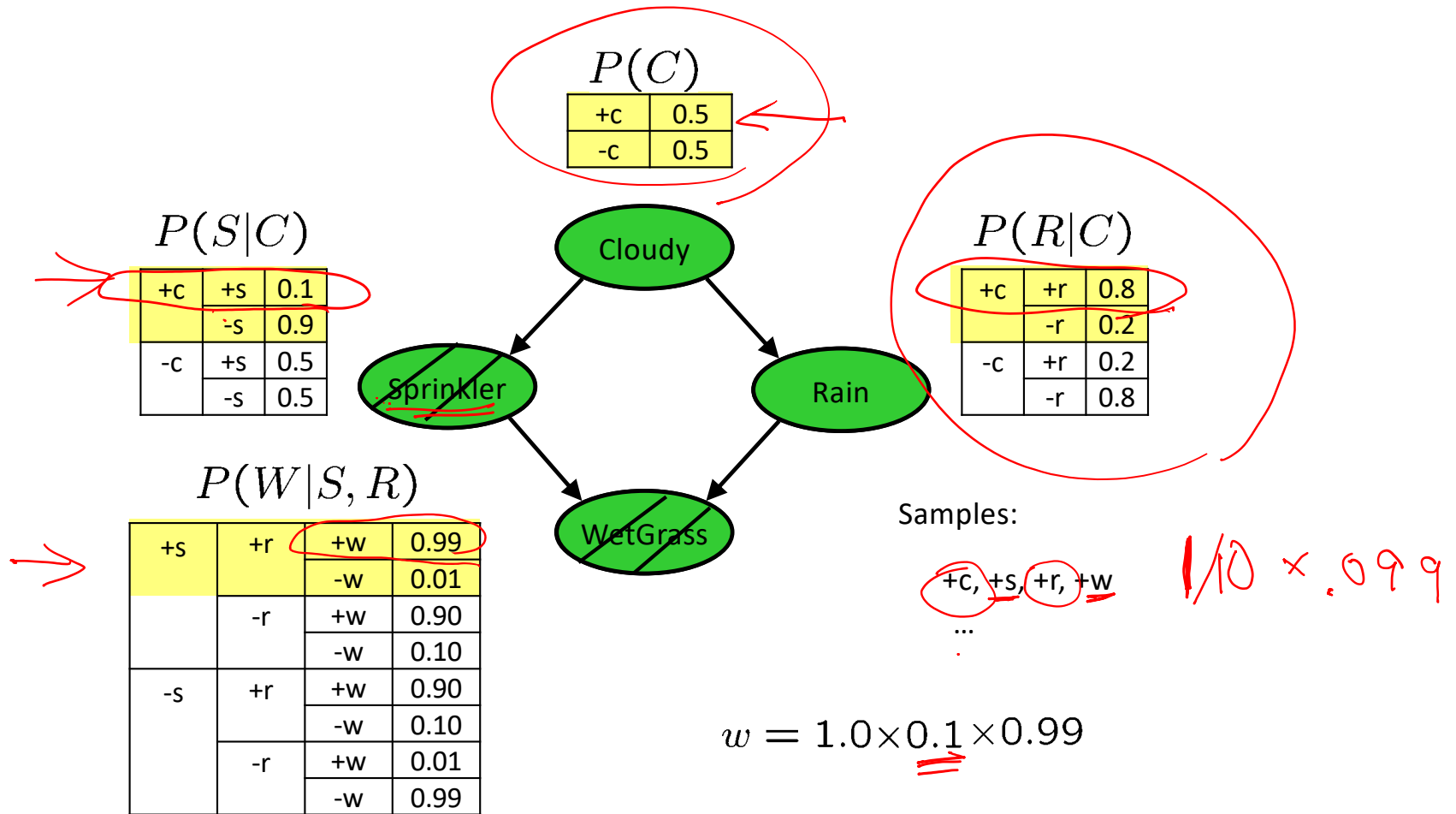
~~pyramid, green~~
~~pyramid, red~~
 sphere, blue
~~cube, red~~
~~sphere, green~~



- Idea: fix evidence variables and sample the rest

- Problem: sample distribution not consistent!
- Solution: weight by probability of evidence given parents

Likelihood Weighting



Likelihood Weighting

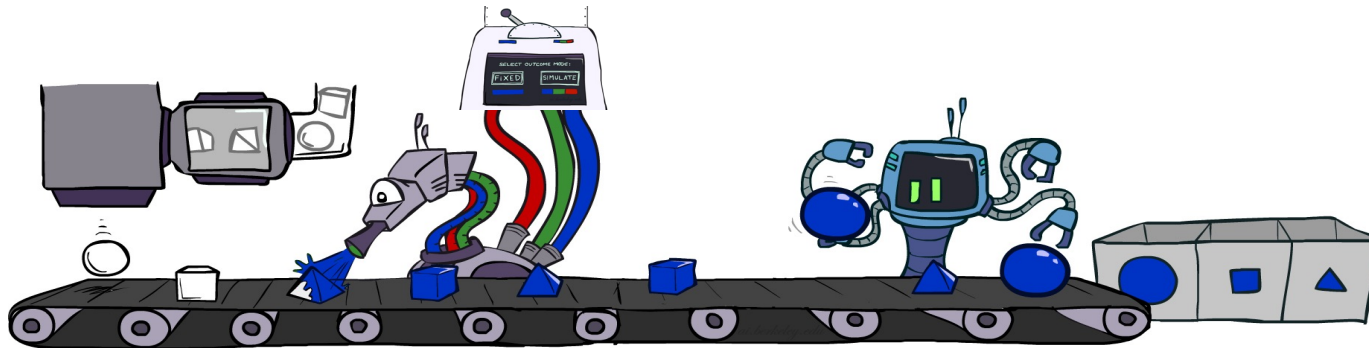
IN: evidence instantiation

$w = 1.0$

for $i=1, 2, \dots, n$

- if X_i is an evidence variable
 - $X_i =$ observation x_i for X_i
 - Set $w = w * P(x_i | \text{Parents}(X_i))$
- else
 - Sample x_i from $P(X_i | \text{Parents}(X_i))$ ←

return $(x_1, x_2, \dots, x_n), w$



Likelihood Weighting

No evidence:
Prior Sampling

Input: no evidence

for $i=1, 2, \dots, n$

- Sample x_i from $P(X_i \mid \text{Parents}(X_i))$

return (x_1, x_2, \dots, x_n)

Some evidence:

Likelihood Weighted Sampling

Input: evidence instantiation

$w = 1.0$

for $i=1, 2, \dots, n$

if X_i is an evidence variable

- $X_i =$ observation x_i for X_i
- Set $w = w * P(x_i \mid \text{Parents}(X_i))$

else

- Sample x_i from $P(X_i \mid \text{Parents}(X_i))$

return $(x_1, x_2, \dots, x_n), w$

All evidence:

Likelihood Weighted

Input: evidence instantiation

$w = 1.0$

for $i=1, 2, \dots, n$

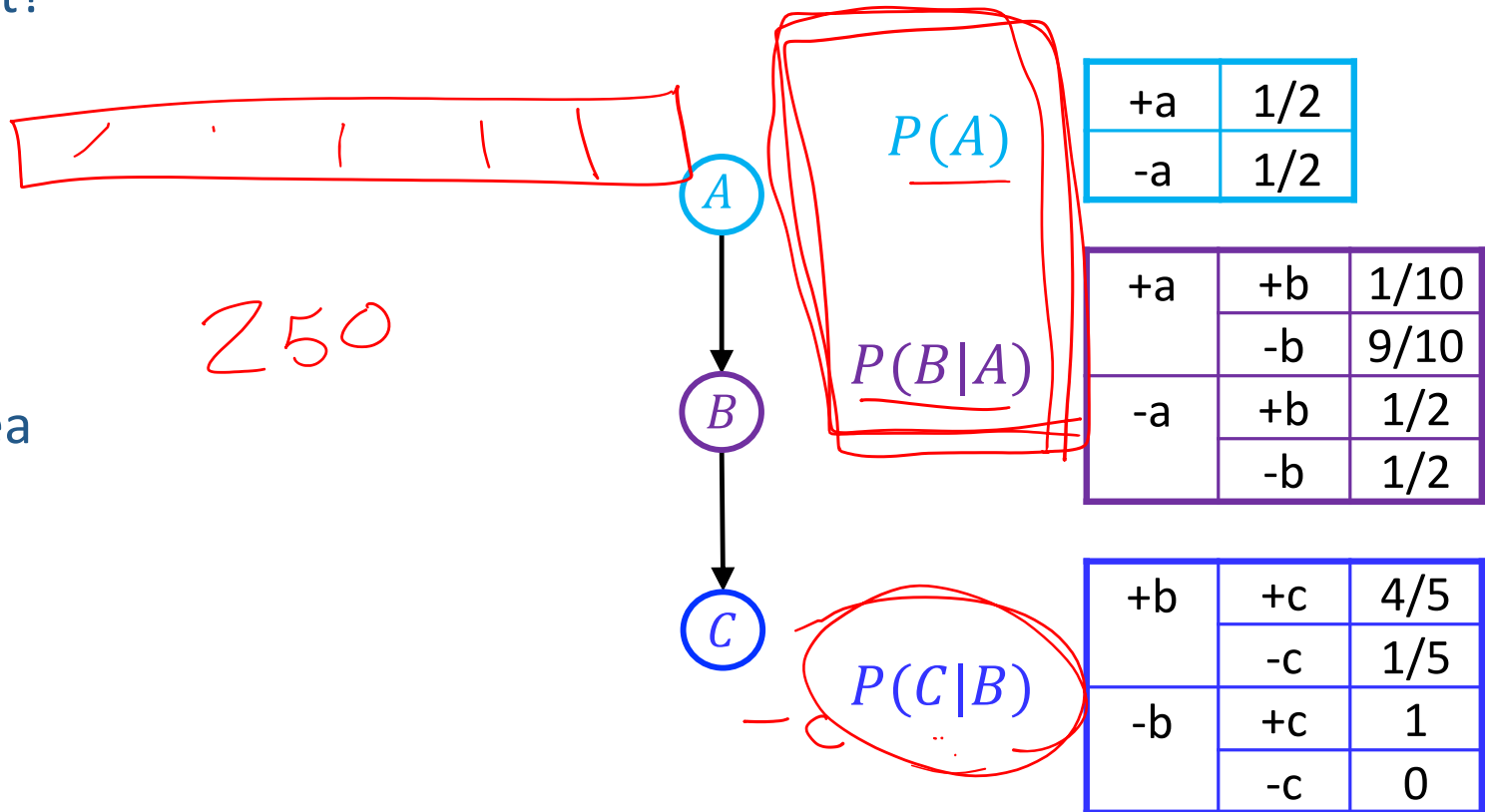
- Set $w = w * P(x_i \mid \text{Parents}(X_i))$

return w

Remember Poll 2

How many $\{-a, +b, -c\}$ samples out of $N=1000$ should we expect?

- A. 1
- B. 50**
- C. 125
- D. 200
- E. I have no idea



Likelihood Weighting

How many $\{-a, +b, -c\}$ samples out of $N=1000$ should we expect?



$P(A)$

+a	1/2
-a	1/2

$P(B|A)$

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

$P(C|B)$

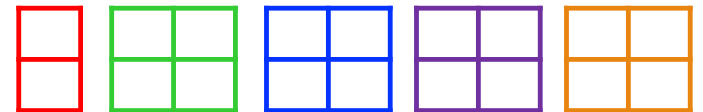
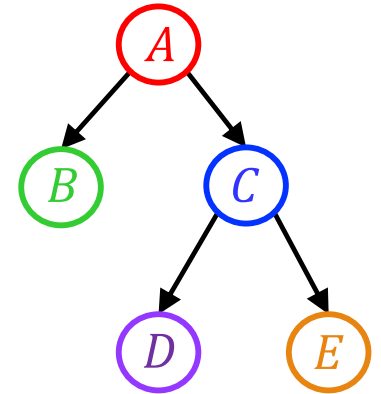
+b	+c	4/5
	-c	1/5
-b	+c	1
	-c	0

Likelihood Weighting Distribution

Consistency of likelihood weighted sampling distribution

Joint from Bayes nets

$$P(A, B, C, D, E) = P(A) P(B|A) P(C|A) P(D|C) P(E|C)$$



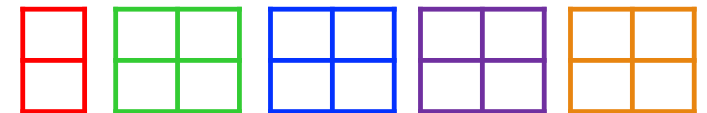
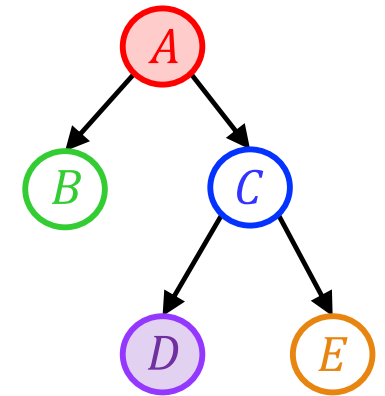
Likelihood Weighting Distribution

Consistency of likelihood weighted sampling distribution

Evidence: $+a$, $-d$

Joint from Bayes nets

$$P(A, B, C, D, E) = P(+a) P(B|+a) P(C|+a) P(-d|C) P(E|C)$$



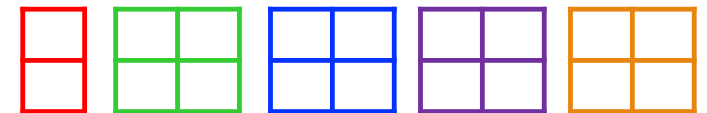
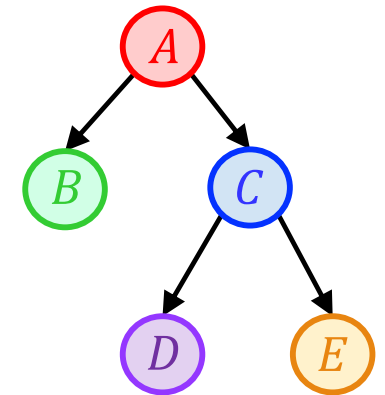
Likelihood Weighting Distribution

Consistency of likelihood weighted sampling distribution

Evidence: $+a$, $+b$, $-c$, $-d$, $+e$

Joint from Bayes nets

$$P(A, B, C, D, E) = P(+a) P(+b|+a) P(-c|+a) P(-d|-c) P(+e|-c)$$



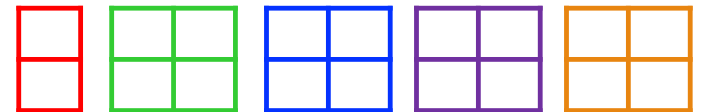
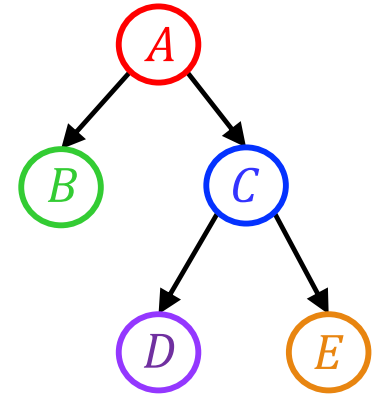
Likelihood Weighting Distribution

Consistency of likelihood weighted sampling distribution

Evidence: None

Joint from Bayes nets

$$P(A, B, C, D, E) = P(A) P(B|A) P(C|A) P(D|C) P(E|C)$$



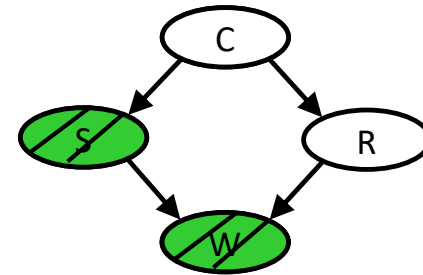
Likelihood Weighting

Sampling distribution if z sampled and e fixed evidence

$$\rightarrow \underline{S_{WS}(z, e)} = \prod_{i=1}^l P(z_i | \text{Parents}(Z_i))$$

Now, samples have weights

$$\underline{w(z, e)} = \prod_{i=1}^m P(e_i | \text{Parents}(E_i))$$



Together, weighted sampling distribution is consistent

$$\begin{aligned} S_{WS}(z, e) \cdot w(z, e) &= \prod_{i=1}^l P(z_i | \text{Parents}(z_i)) \prod_{i=1}^m P(e_i | \text{Parents}(e_i)) \\ &= P(z, e) \end{aligned}$$

Poll 4

Given a fixed query, two identical samples from likelihood weighted sampling will have the same exact weights.

- A. True
- B. False
- C. It depends
- D. I don't know



Poll 5

What does the following likelihood weighted value approximate?

$$\text{weight}_{(+a,-b,+c)} \cdot \frac{N(+a,-b,+c)}{N}$$

- A. $P(+a, -b, +c)$
- B. $P(+a, -b \mid +c)$
- C. I'm not sure

Practice Likelihood Weighted Sampling

For each variable:

If evidence variable: you know the value, and multiply the weight of the sample by $P(\text{evidence} | \text{parents})$

Else: Use a random number generator to pick a number between 0 and 100, divide random number by 100 to get a probability. Check the relevant probability distribution to determine what value is selected.

Continue selecting values for each variable until you have a single sample

Record the sample, repeat the process for 9 more samples

The probability for a given set of variables equals

$$(\text{COUNT OF SAMPLES} * \text{WEIGHT}) / (\text{TOTAL COUNT OF SAMPLES})$$

Likelihood Weighting

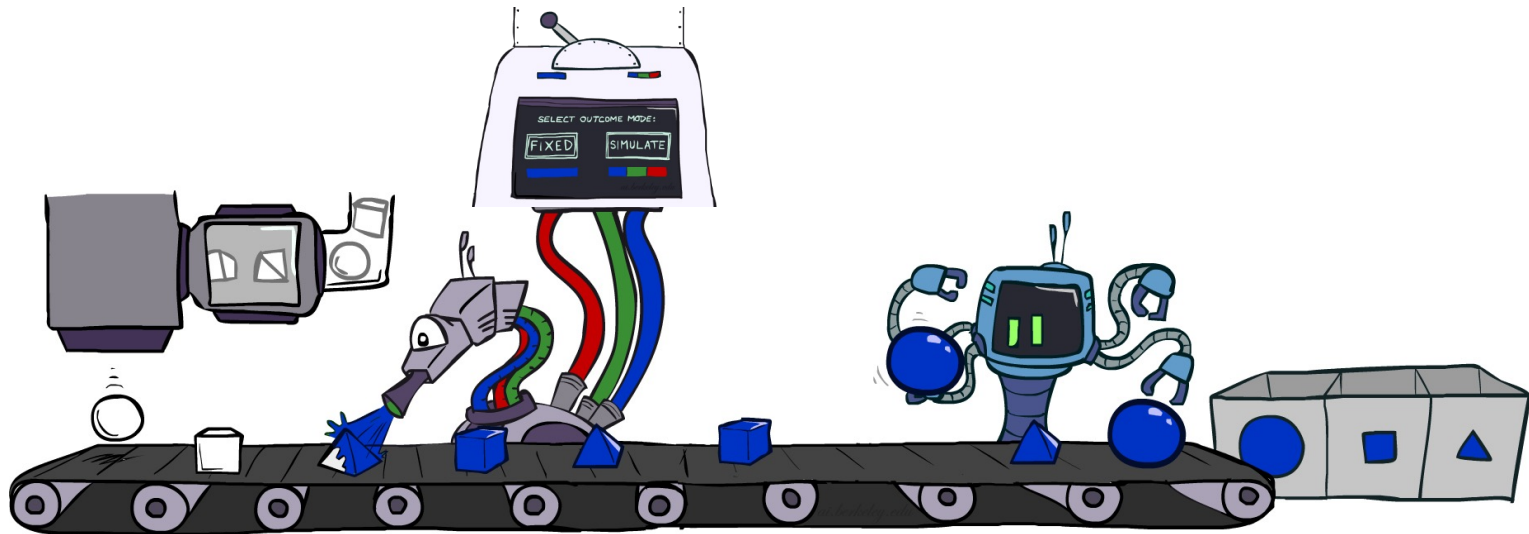
Likelihood weighting is good

- We have taken evidence into account as we generate the sample
- E.g. here, W 's value will get picked based on the evidence values of S , R
- More of our samples will reflect the state of the world suggested by the evidence

Likelihood weighting doesn't solve all our problems

- Evidence influences the choice of downstream variables, but not upstream ones (C isn't more likely to get a value matching the evidence)

We would like to consider evidence when we sample every variable



Likelihood Weighting

Likelihood weighting is good

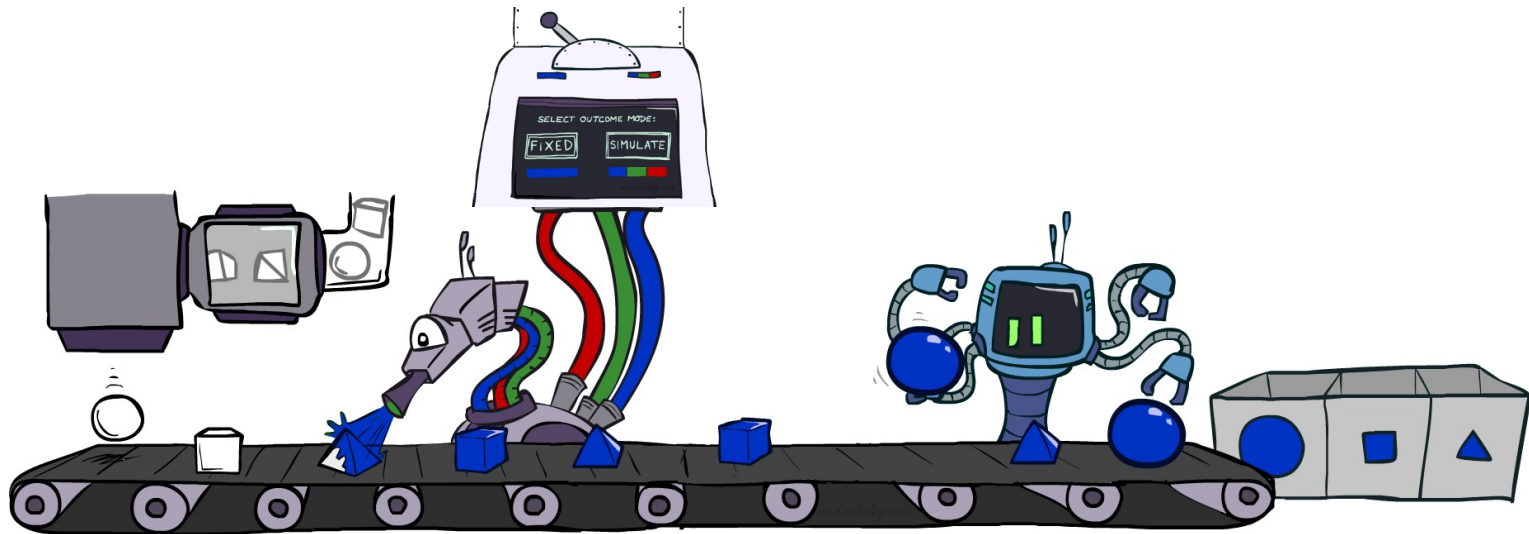
- We have taken evidence into account as we generate the sample
- E.g. here, W 's value will get picked based on the evidence values of S , R
- More of our samples will reflect the state of the world suggested by the evidence

Likelihood weighting doesn't solve all our problems

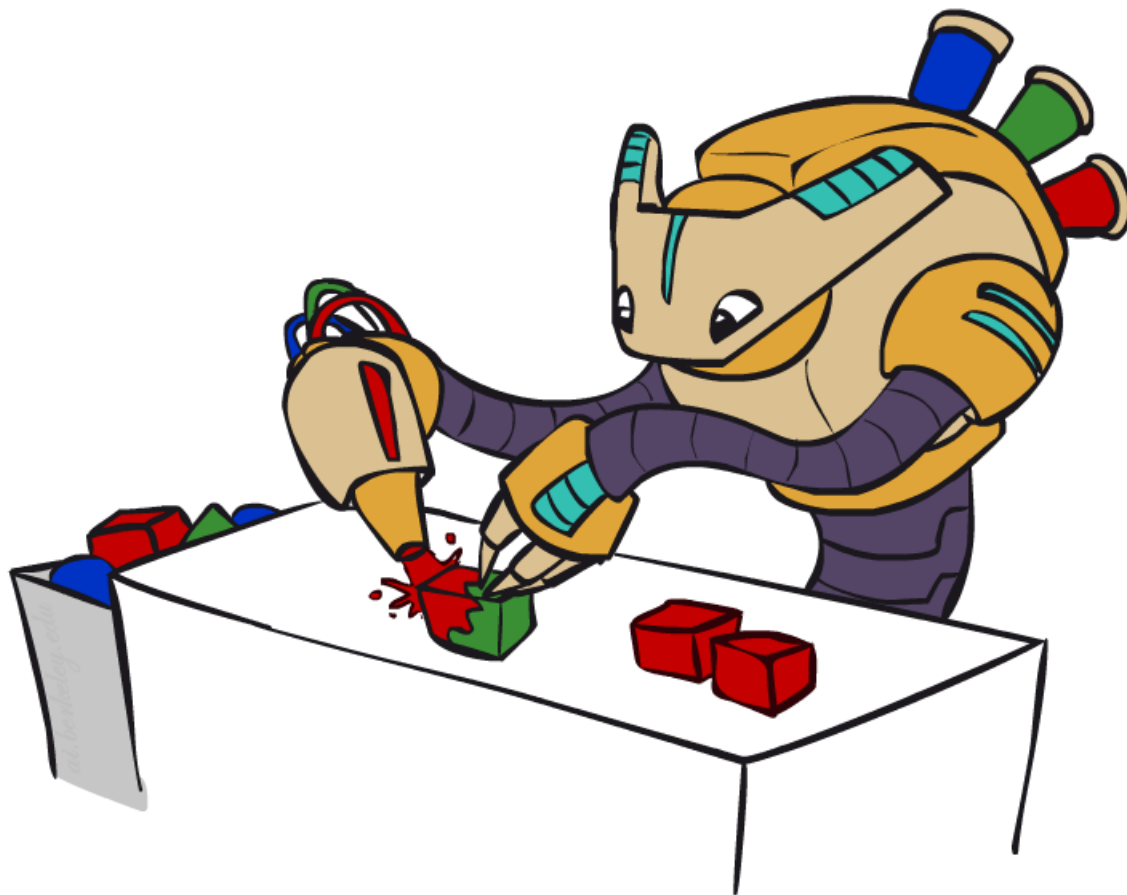
- Evidence influences the choice of downstream variables, but not upstream ones (C isn't more likely to get a value matching the evidence)

We would like to consider evidence when we sample every variable

→ Gibbs sampling



Gibbs Sampling



Gibbs Sampling

Iteration Procedure: keep track of a full instantiation x_1, x_2, \dots, x_n .

1. Start with an arbitrary instantiation consistent with the evidence.
2. Sample one variable at a time, conditioned on all the rest, but keep evidence fixed.
3. Keep repeating this for a long time.

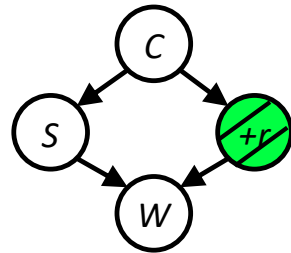
When done, keep last values of variables as 1 sample.

Repeat iteration for each sample.

Gibbs Sampling Example: $P(S \mid +r)$

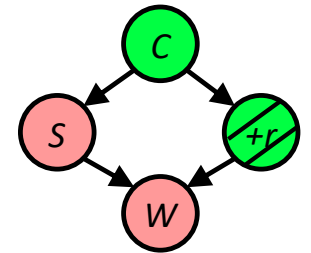
Step 1: Fix evidence

- $R = +r$



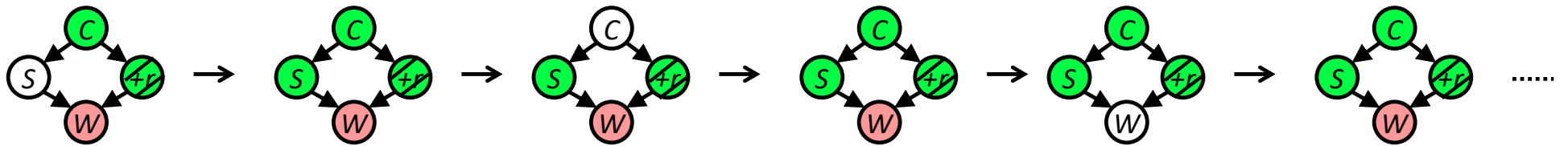
Step 2: Initialize other variables

- Randomly



Steps 3: Repeat

- Choose a non-evidence variable X
- Resample X from $P(X \mid \text{all other variables})$



Sample from $P(S \mid +c, -w, +r)$

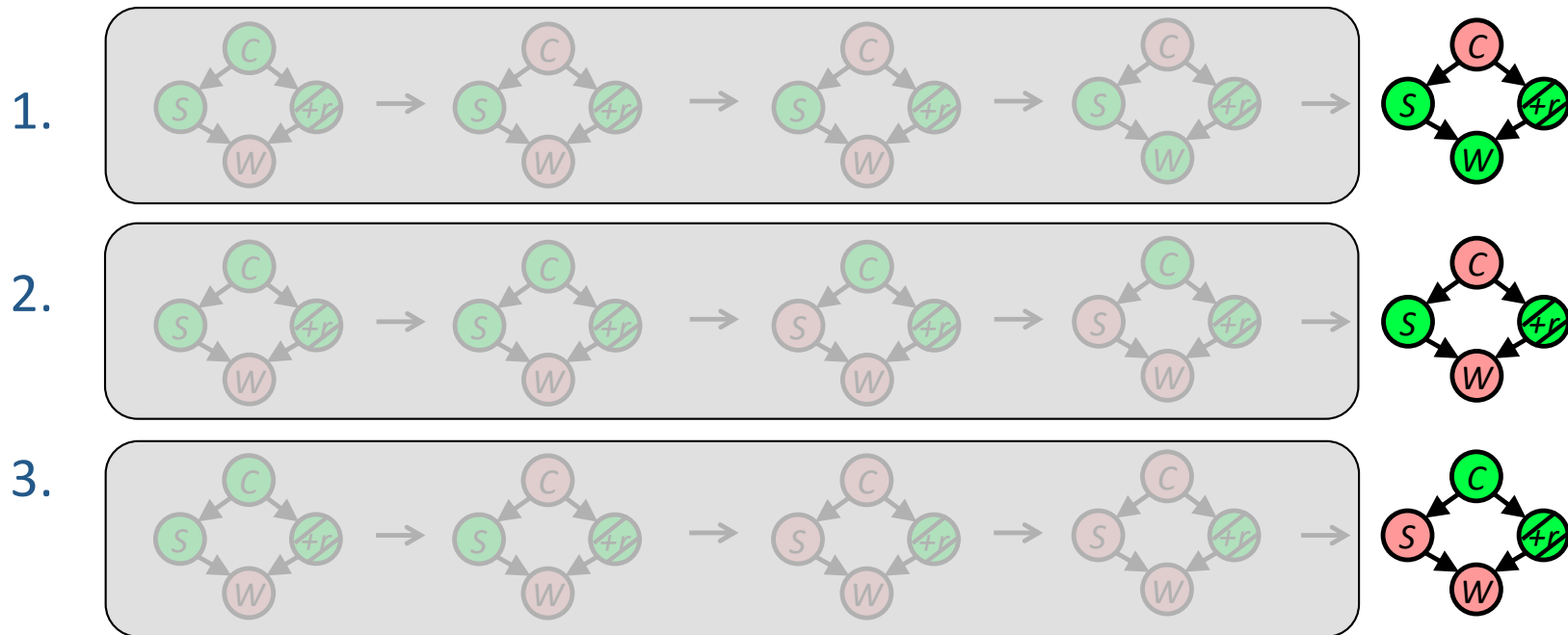
Sample from $P(C \mid +s, -w, +r)$

Sample from $P(W \mid +s, +c, +r)$

Gibbs Sampling Example: $P(S \mid +r)$

Each iteration here is repeated 5 times.

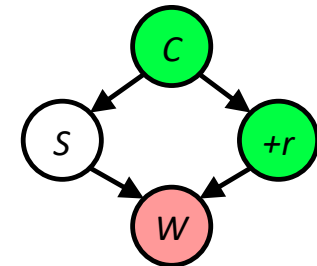
Keep only the last sample from each iteration:



Efficient Resampling of One Variable

Sample from $P(S \mid +c, +r, -w)$

$$\begin{aligned} P(S \mid +c, +r, -w) &= \frac{P(S, +c, +r, -w)}{P(+c, +r, -w)} \\ &= \frac{P(S, +c, +r, -w)}{\sum_s P(s, +c, +r, -w)} \\ &= \frac{P(+c)P(S \mid +c)P(+r \mid +c)P(-w \mid S, +r)}{\sum_s P(+c)P(s \mid +c)P(+r \mid +c)P(-w \mid s, +r)} \\ &= \frac{P(+c)P(S \mid +c)P(+r \mid +c)P(-w \mid S, +r)}{P(+c)P(+r \mid +c) \sum_s P(s \mid +c)P(-w \mid s, +r)} \\ &= \frac{P(S \mid +c)P(-w \mid S, +r)}{\sum_s P(s \mid +c)P(-w \mid s, +r)} \end{aligned}$$



Many things cancel out – only CPTs with S remain!

More generally: only CPTs that have resampled variable need to be considered, and joined together

Practice Gibbs Sampling

Procedure: keep track of a full instantiation x_1, x_2, \dots, x_n .

1. Start with an arbitrary instantiation consistent with the evidence.
2. Sample one variable at a time, conditioned on all the rest, but keep evidence fixed.
3. Iterate to repeat this for a long time.
4. Keep only the last set of variables as ONE sample
5. Repeat with new sample

Property: in the limit of repeating this infinitely many times the resulting sample is coming from the correct distribution

Gibbs Sampling

Property: in the limit of repeating this infinitely many times the resulting sample is coming from the correct distribution

Rationale: both upstream and downstream variables condition on evidence.

In contrast: likelihood weighting only conditions on upstream evidence, and hence weights obtained in likelihood weighting can sometimes be very small. Sum of weights over all samples is indicative of how many “effective” samples were obtained, so want high weight.

Further Reading on Gibbs Sampling

Gibbs sampling produces sample from the query distribution $P(Q | e)$ in limit of re-sampling infinitely often

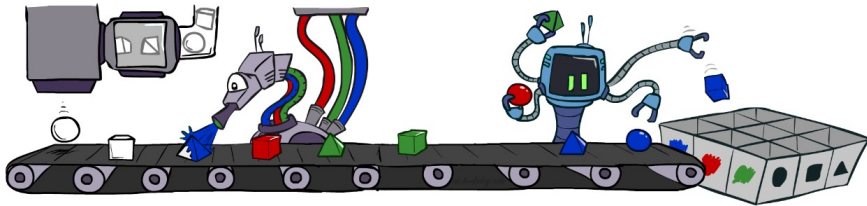
Gibbs sampling is a special case of more general methods called Markov chain Monte Carlo (MCMC) methods

- Metropolis-Hastings is one of the more famous MCMC methods (in fact, Gibbs sampling is a special case of Metropolis-Hastings)

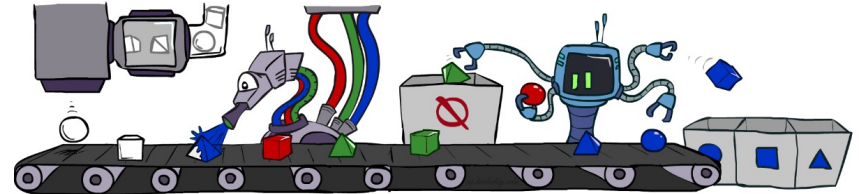
You may read about Monte Carlo methods – they're just sampling

Bayes' Net Sampling Summary

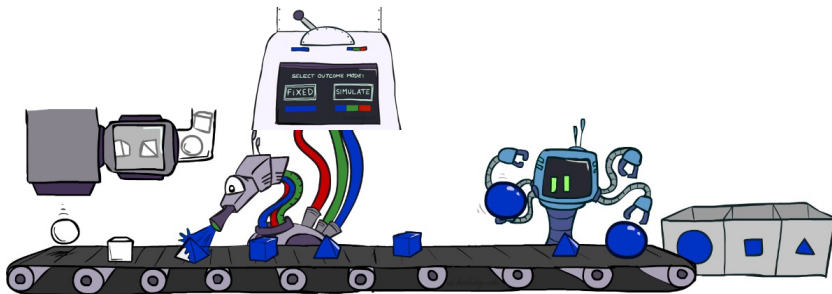
Prior Sampling $P(Q, E)$



Rejection Sampling $P(Q | e)$



Likelihood Weighting $P(Q, e)$



Gibbs Sampling $P(Q | e)$

