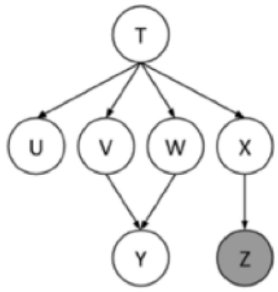
Variable elimination

Using the same Bayes Net (shown below), we want to compute $P(Y \mid +z)$. All variables have binary domains. Assume we run variable elimination to compute the answer to this query, with the following variable elimination ordering: X, T, U, V, W.



$$2^6 = 64$$

$$P(T, U, V, W, Y, X, +z) = \prod_i^{\text{all nodes}} P(\text{Node}_i \mid \text{Parents}[\text{Node}_i])$$

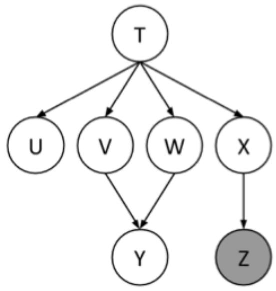
$$= P(Y \mid V)\, P(Y \mid W)\, P(U \mid T)\, P(V \mid T)\, P(W \mid T)\, P(+z \mid X)\, P(X \mid T)$$

To get $P(A \mid +b)$ we need to eliminate all other variables, $X_1, X_2, \ldots \neq A, B$

For each i:
- Join all factors involving X_i
- Sum out X_i

1 Variable Elimination

Using the same Bayes Net (shown below), we want to compute $P(Y \mid +z)$. All variables have binary domains. Assume we run variable elimination to compute the answer to this query, with the following variable elimination ordering: X, T, U, V, W.



Complete the following description of the factors generated in this process:

After inserting evidence, we have the following factors to start out with:

$$P(T), P(U|T), P(V|T), P(W|T), P(X|T), P(Y|V, W), P(+z|X)$$

(a) When eliminating X we generate a new factor f_1 as follows, which leaves us with the factors:

$$P(T, X, +z) = P(+z|X)\,P(X|T)\,P(T)$$

$\Updownarrow$ divide by P(T) on both sides

$$P(X, +z \mid T) = P(+z|X)\,P(X|T)$$

$$\Updownarrow$$

$$\sum_{x_i} P(x_i, +z \mid T) = \sum_{x_i} (+z \mid x_i)\,P(x_i \mid T)$$

$$\Updownarrow$$

$$f_1(+z|T) = P(+z \mid T) = \sum_{x_i} (+z \mid x_i)\,P(x_i \mid T)$$

(a) When eliminating X we generate a new factor f_1 as follows, which leaves us with the factors:

$$f_1(+z|T) = \sum_x P(x|T)P(+z|x) \qquad P(T), P(U|T), P(V|T), P(W|T), P(Y|V,W), f_1(+z|T)$$

(b) When eliminating T we generate a new factor f_2 as follows, which leaves us with the factors:

① Join:

$$P(T)\,P(U|T)\,P(V|T)\,P(W|T)\,f_1(+z|T)$$

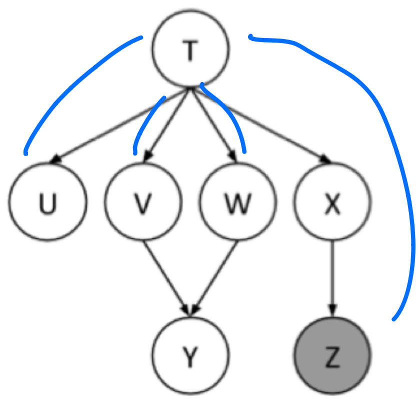
$$\parallel$$

$$P(U,V,W,+z|T)\,P(T) = P(\text{joint})$$

by cond. indep. properties of the graph

② Sum out:

$$f_2(U,V,W,+z) = \sum_t P(t)\,P(U|t)\,P(V|t)\,P(W|t)\,f_1(+z|t)$$



factors remaining:

$$f_2(U,V,W,+z),\ P(Y|V,W)$$

(c) When eliminating U we generate a new factor f_3 as follows, which leaves us with the factors:

$$f_3(V,W,+z) = \sum_u f_2(u,V,W,+z) \qquad f_3(V,W,+z),\ P(Y|V,W)$$

(d) When eliminating V we generate a new factor f_4 as follows, which leaves us with the factors:

$$f_4(W,Y,+z) = \sum_v f_3(v,W,+z)\,P(Y|v,W) \qquad f_4(W,Y,+z)$$

(e) When eliminating W we generate a new factor f_5 as follows, which leaves us with the factors:

$$f_5(Y,+z) = \sum_w f_4(w,Y,+z) \qquad f_5(Y,+z)$$

(f) How would you obtain $P(Y \mid +z)$ from the factors left above:

$$P(Y \mid +z) = \frac{f_5(Y, +z)}{P(+z)} \xrightarrow{1} \frac{f_5(Y, +z)}{\sum_y f_5(y, +z)}$$

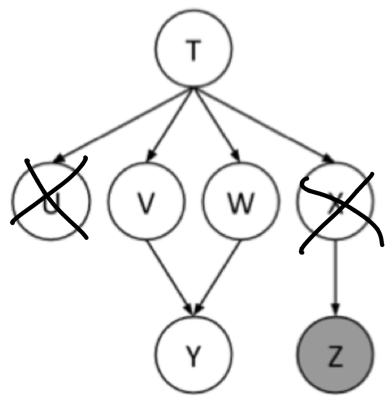
don't have!!

(g) What is the size of the largest factor that gets generated during the above process?

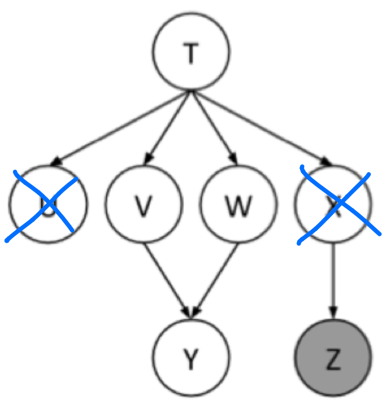
f_2, 2^3 rows

(m) Does there exist a better elimination ordering (one which generates smaller largest factors)?

T right away $\rightarrow 2^4$



U, X, T, V, W $\rightarrow 2^2$



$\rightarrow$ $f: (V, W, +z) \rightarrow 2^2$

Sampling

Prior	Rejection	Likelihood Weighting

Prior → Discard samples inconsistent with evidence AFTER generating

Rejection → Discard samples inconsistent with evidence WHILE generating

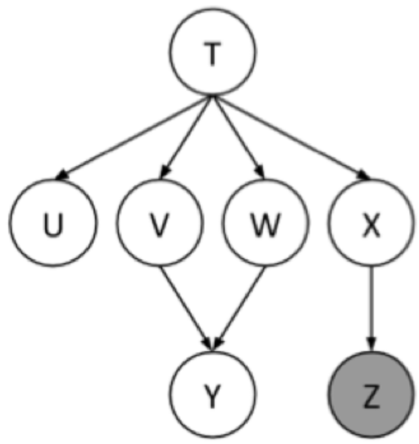
Likelihood Weighting

Trying to be less wasteful.

→ Enforce evidence

└ PROBLEMATIC

→ To counteract, we use <u>weights</u>



w = "probability of evidence given sampled variables"

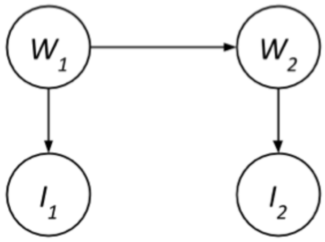
$$= \prod_{\text{evidence var } e} P(e \mid parents[e])$$

Gibbs Sampling

→ Iterative random resampling

2 Sampling and Dynamic Bayes Nets

We would like to analyze people's ice cream eating habits on sunny and rainy days Suppose we consider the weather, along with a person's ice-cream eating, over the span of two days. We'll have four random variables: W_1 and W_2 stand for the weather on days 1 and 2, which can either be rainy R or sunny S, and the variables I_1 and I_2 represent whether or not the person ate ice cream on days 1 and 2, and take values T (for truly eating ice cream) or F. We can model this as the following Bayes Net with these probabilities.



W_1	$P(W_1)$
S	0.6
R	0.4

W_1	W_2	$P(W_2 \mid W_1)$
S	S	0.7
S	R	0.3
R	S	0.5
R	R	0.5

W	I	$P(I \mid W)$
S	T	0.9
S	F	0.1
R	T	0.2
R	F	0.8

Suppose we produce the following samples of (W_1, I_1, W_2, I_2) from the ice-cream model:

~~R, F, R, F~~ ~~R, F, R, F~~ ~~S, F, S, T~~ ~~S, T, S, T~~ S, T, R, F
~~R, F, R, T~~ ~~S, T, S, T~~ ~~S, T, S, T~~ S, T, R, F ~~R, F, S, T~~

1. What is $\widehat{P}(W_2 = \mathtt{R})$, the probability that sampling assigns to the event $W_2 = \mathtt{R}$?

$$\hat{P}(W_2 = R) = \frac{5}{10} = 0.5$$

2. Cross off samples above which are rejected by rejection sampling if we're computing $P(W_2 \mid I_1 = \mathtt{T}, I_2 = \mathtt{F})$.

Suppose we produce the following samples of (W_1, I_1, W_2, I_2) from the ice-cream model:

R, F, R, F R, F, R, F S, F, S, T S, T, S, T S, T, R, F
R, F, R, T S, T, S, T S, T, S, T S, T, R, F R, F, S, T

Rejection sampling seems to be wasting a lot of effort, so we decide to switch to likelihood weighting. Assume we generate the following six samples given the evidence $I_1 = \mathtt{T}$ and $I_2 = \mathtt{F}$:

$$(W_1, I_1, W_2, I_2) = \Big\{(\mathtt{S},\mathtt{T},\mathtt{R},\mathtt{F}), (\mathtt{R},\mathtt{T},\mathtt{R},\mathtt{F}), (\mathtt{S},\mathtt{T},\mathtt{R},\mathtt{F}), (\mathtt{S},\mathtt{T},\mathtt{S},\mathtt{F}), (\mathtt{S},\mathtt{T},\mathtt{S},\mathtt{F}), (\mathtt{R},\mathtt{T},\mathtt{S},\mathtt{F})\Big\}$$

3. What is the weight of the first sample (S, T, R, F) above?

w = "probability of evidence given sampled variables"

$$= P(I_1 = T, I_2 = F \mid W_1 = S, W_2 = R)$$

$$= P(I_1 = T \mid W_1 = S, W_2 = R)\, P(I_2 = F \mid W_1 = S, W_2 = R)$$

$= P(I_1 = T \mid W_1 = S)\ P(I_2 = F \mid W_2 = R) = (0.9)(0.8) = 0{,}72$

$= \prod_{\text{evidence var } e} P(e \mid parents[e])$ (GENERAL FORMULA)

4. Use likelihood weighting to estimate $P(W_2 | I_1 = \mathrm{T}, I_2 = \mathrm{F})$.

The sample weights are given by

(W_1, I_1, W_2, I_2)	w	(W_1, I_1, W_2, I_2)	w
S, T, R, F	0.72	S, T, S, F	0.09
R, T, R, F	0.16	S, T, S, F	0.09
S, T, R, F	0.72	R, T, S, F	0.02

$$P(W_2 = R \mid I_1 = T, I_2 = F) = \frac{0.72 + 0.16 + 0.72}{0.72 + 0.16 + 0.72 + 0.09 + 0.09 + 0.02}$$

$$= 0{,}889$$

$$P(W_2 = S \mid I_1 = T, I_2 = F) = 1 - 0.889 = 0{,}111$$