CS188 Spring 2014 Section 8: Inference and Sampling

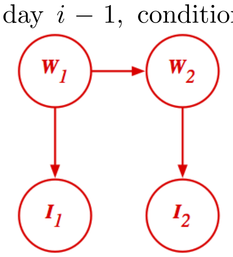
1 Sampling and Dynamic Bayes Nets

Many people would prefer to eat ice cream on a sunny day than on a rainy day. We can model this situation with a Bayesian network. 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`.

(a) What conditional independence relationships might be reasonable to assume among these variables?

This is very much subjective, and intended as a prompt for discussion. Question (b) gives one example of reasonable assumptions.

(b) Suppose you decide to assume for modeling purposes that the weather on each day might depend on the previous day's weather, but that ice cream consumption on day i is independent of ice cream consumption on day $i - 1$, conditioned on the weather on day i. Draw a Bayes net encoding these independence assumptions.



Suppose you are given the following conditional probability distributions:

$W_1 = \texttt{S}$	$W_1 = \texttt{R}$
.6	.4

(a) $P(W_1)$

	$W_2 = \texttt{S}$	$W_2 = \texttt{R}$
$W_1 = \texttt{S}$	.7	.3
$W_1 = \texttt{R}$	.5	.5

(b) $P(W_2|W_1)$

	$I = \texttt{T}$	$I = \texttt{F}$
$W = \texttt{S}$	.9	.1
$W = \texttt{R}$	.2	.8

(c) $P(I|W)$

Together with the graph you drew above, these distributions fully specify a joint probability distribution for the four variables. Now we want to do inference, and in particular we'll try approximate inference through sampling. Suppose we sample from the prior to 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~~

(c) What is $\widehat{P}(W_2 = \texttt{R})$, the probability that sampling assigns to the event $W_2 = \texttt{R}$?
Number of samples in which $W_2 = \texttt{R}$: 5. Total number of samples: 10. Answer $5/10 = 0.5$.

(d) Cross off samples rejected by rejection sampling if we're computing $\mathbb{P}(W_2|I_1 = \texttt{T}, I_2 = \texttt{F})$

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 $I_1 = \texttt{T}$ and $I_2 = \texttt{F}$:

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

(e) What is the weight of the first sample (S, T, R, F) above?
The weight given to a sample in likelihood weighting is

$$\prod_{\text{Evidence variables } e} \Pr(e|\text{Parents}(e)).$$

In this case, the evidence is $I_1 = \texttt{T}, I_2 = \texttt{F}$. The weight of the first sample is therefore

$$w = \Pr(I_1 = \texttt{T}|W_1 = \texttt{S}) \cdot \Pr(I_2 = \texttt{F}|W_2 = \texttt{R}) = 0.9 \cdot 0.8 = 0.72$$

(f) Use likelihood weighting to estimate $\widehat{P}(W_2|I_1 = \texttt{T}, I_2 = \texttt{F})$
The sample weights are given by

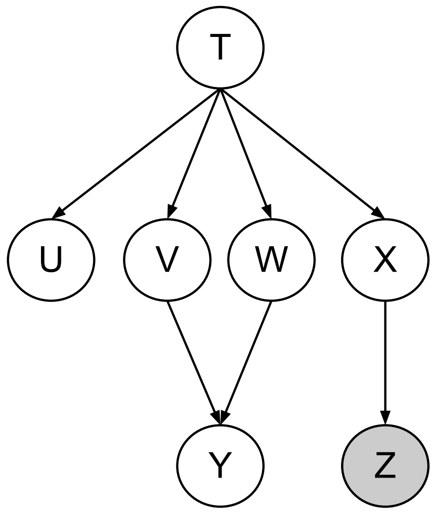
(W_1, I_1, W_2, I_2)	w	(W_1, I_1, W_2, I_2)	w
S, T, R, F	0.72	S, T, S, F	0.09
R, T, R, F	0.16	S, T, S, F	0.09
S, T, R, F	0.72	R, T, S, F	0.02

To compute the probabilities, we thus normalize the weights and find

$$\widehat{P}(W_2 = \mathtt{R} | I_1 = \mathtt{T}, I_2 = \mathtt{F}) = \frac{0.72 + 0.16 + 0.72}{0.72 + 0.16 + 0.72 + 0.09 + 0.09 + 0.02} = 0.889$$
$$\widehat{P}(W_2 = \mathtt{S} | I_1 = \mathtt{T}, I_2 = \mathtt{F}) = 1 - 0.889 = 0.111.$$

2 Variable Elimination

For the Bayes' net below, we are given the query $P(Y \mid +z)$. 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:

$$f_1(T,+z) = \sum_x P(x|T)P(+z|x)$$

(b) This leaves us with the factors:

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

(c) When eliminating T we generate a new factor f_2 as follows:

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

(d) This leaves us with the factors:

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

(e) When eliminating U we generate a new factor f_3 as follows:

$$f_3(V, W, +z) = \sum_u f_2(u, V, W, +z)$$

(f) This leaves us with the factors:

$$P(Y|V, W), f_3(V, W, +z)$$

(g) When eliminating V we generate a new factor f_4 as follows:

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

(h) This leaves us with the factors:

$$f_4(W, Y, +z)$$

(i) When eliminating W we generate a new factor f_5 as follows:

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

(j) This leaves us with the factors:

$$f_5(Y, +z)$$

(k) How would you obtain $P(Y \mid +z)$ from the factors left above:
Simply renormalize $f_5(Y, +z)$ to obtain $P(Y \mid +z)$. Concretely,

$$P(y \mid +z) = \frac{f_5(y, +z)}{\sum_{y'} f_5(y', +z)}$$

(l) What is the size of the largest factor that gets generated during the above process?
$f_2(U, V, W, +z)$, of size 3.

(m) Does there exist a better elimination ordering (one which generates smaller largest factors)?
Yes, elimination ordering of X, U, T, V, W generates only factors of size 2.