Approximate Inference in Bayes Nets

- Often the Bayes net is not solvable by Variable Elimination: under any ordering of the variables we end up with a factor that is too large to compute (or store).
- Since we are trying to compute a probability (which only predicts the likelihood of an event occurring) it is natural to consider approximating answer.

Sampling Techniques

- **Direct Sampling** from the **prior** distribution.
- Every Bayes net specifies the probability of every atomic event:
  - Each atomic event is a particular assignment of values to all of the variables in the Bayes nets.
  - Let $V_1, \ldots, V_n$ be the variables in the Bayes net.
  - Let $d_1, \ldots, d_n$ be values for these variables (d_i is the value variable V_i takes).
  - The Bayes net specifies that

$$\Pr(V_1 = d_1, V_2 = d_2, \cdots, V_n = d_n) = \prod_{i=1}^{n} \Pr\big(V_i = d_i \mid ParVals(V_i)\big)$$

  where $ParVals(V_i)$ is the set of assignments $V_k = d_k$ for each $V_k \in Par(V_i)$

Sampling Techniques

- So we want to sample atomic events in such a way that the probability we select event **e** is equal to Pr(**e**)

1. Select an unselected variable V_i such that all parents of V_i in the Bayes Net have already been selected.
2. Let $[P_1, P_2, \ldots, P_k]$ be the parents of V_i in the Bayes net. Let $[b_1, \ldots, b_k]$ be the values that have already been selected for these parents ($P_i = b_i$).
3. Set V_i to the value $d \in Dom[V_i]$ with probability

$$Pr(V_i = d \mid P_1 = b_1, P_2 = b_2, \ldots, P_k = b_k)$$

Sampling Techniques

- Note that the probabilities
  $$\Pr(V_i = d \mid P_1=b_1, P_2=b_2, \ldots, P_k=b_k)$$
  are specified in V_i's CPT in the Bayes net.

- Each variable is given a value by a separate random selection so the probability one obtains a particular atomic event (a setting of all of the variables) via this algorithm is as specified by the Bayes Net.

$$\Pr(\mathrm{e}=[V_1=d_1, V_2=d_2, \cdots, V_n=d_n]) = \prod_{i=1}^{n} \Pr\big(V_i = d_i \mid ParVals(V_i)\big)$$

Sampling Techniques

- Say we want to evaluate $\Pr(V_1 = d_3)$
- We select **N** random samples of atomic events via this method
- Then we compute the proportion of these N events in which $V_1 = d_3$
- This proportion

  (Number of Events where $V_1 = d_3$)/**N**

  is an estimate of $\Pr(V_1 = d_3)$.
- The estimate gets better as **N** gets larger, and by the law of large numbers as **N** approaches infinity the estimate converges (becomes closer and closer) to the exact $\Pr(V_1 = d_3)$

Sampling Techniques

- If we want to compute a conditional probability like $\Pr(V_1 = d_3 \mid V_4 = d_1)$, then we can
  - Discard all atomic events in which $V_4 \neq d_1$
  - This gives a new smaller set of **N'** sampled atomic events.
  - From those **N'** we compute the proportion in which $V_1 = d_3$
  - This proportion
    (Number of Events where $V_1 = d_3$ from the remaining samples**)/N'**
    is an estimate of $\Pr(V_1 = d_3 \mid V_4 = d_1)$
  - This is called **Rejection Sampling**

Sampling Techniques

- **Problem,** almost all samples might be rejected if $V_4 = d_1$ has very low probability.
- The accuracy of the estimate depends on the size of **N'** (the samples that remain after rejection).
- So if very few are left our estimate is not good.
- E.g., if $\Pr(V_4 = d_1) = 0.0000001$, then if we generate $1/0.0000001 = 10{,}000{,}000$ samples we expect to reject 9,999,999 of them. In that case our estimate of $\Pr(V_1 = d_3 \mid V_4 = d_1)$ will be 1 or 0! (Either our sole remaining sample has $V_1 = d_3$ or it doesn't).
- In most cases we want to compute **posterior** probabilities, i.e., probabilities conditioned on the **evidence**. So this is a major problem.

Sampling Techniques

- **Likelihood Weighting** tries to address this issue.
- Force all samples to be compatible with the conditioning event.
- Don't select a value for a variable whose value is specified in the evidence that we are conditioning on.
- Weigh each sample by its probability—some samples count more than others in computing the estimate.

Sampling Techniques

1. Set **w = 1,** let the evidence be a set of variables whose values are already given.
2. **While there are unselected variables**
   1. Select an unselected variable V_i such that all parents of V_i in the Bayes Net have already been selected.
   2. Let $[P_1, P_2, \ldots, P_k]$ be the parents of V_i in the Bayes net. Let $[b_1, \ldots, b_k]$ be the values that have already been selected for these parents ($P_i = b_i$).
   3. **If** V_i's value is specified in the **evidence** and d is the value specified then
      w = w * $Pr(V_i = d \mid P_1=b_1, P_2=b_2, \ldots, P_k=b_k)$
   4. **Else** set V_i to the value $d \in Dom[V_i]$ with probability
      $Pr(V_i = d \mid P_1=b_1, P_2=b_2, \ldots, P_k=b_k)$

Sampling Techniques

- If we want to compute a conditional probability like $\Pr(V_1 = d_3 \mid V_4 = d_1)$, then we can
  - Generate a collection **N** of likelihood weighted samples using the evidence $V_4 = d_1$
  - Each sample (atomic event) **e** has a weight **w.**
  - We compute the sum of the weights of the samples in **N** in $V_1 = d_3$ and divide this by the sum of the weights of all samples in **N.**
  - This number

    (Sum of weights of samples in **N** where $V_1 = d_3$)/(sum of weights of samples in **N**)

    is an estimate of $\Pr(V_1 = d_3 \mid V_4 = d_1)$

Sampling Techniques

- **Problem**, many samples might have very low weight. Some might even have zero weight.
  - Zero weight occurs when we have selected the parents of an evidence variable in such a way that
    $Pr(V_i = d \mid P_1=b_1, P_2=b_2, \ldots, P_k=b_k)$
    is zero (this is multiplied into the sample weight).
- The accuracy of the estimate increases as the total weight of the samples increases, so if each sample has very low weight, we may need a very large number of samples.