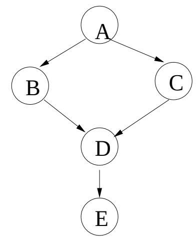
Probabilistic Reasoning in AI - Assignment 3

Due Monday, February 15, 2004, 10pm

1. [10 points] **Likelihood weighting**

   Consider the Bayes net below:

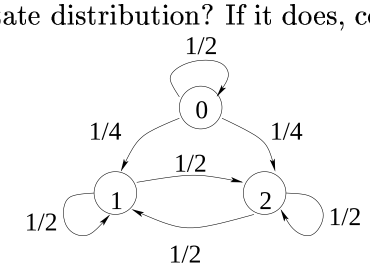
   Assume $P(A) = 0.7$, $P(B = 1|A = 1) = 0.8$, $P(B = 1|A = 0) = 0.2$, $P(C = 1|A = 1) = 0.6$, $P(C = 1|A = 0) = 0.3$, $P(E = 1|D = 1) = 0.8$, $P(E = 1|D = 0) = 0.2$, and assume that D is a deterministic "and" of its parents. Suppose we want to compute $P(A = 1, E = 1|D = 1, C = 1)$ using likelihood weighting. Generate two samples using this methods and show their weights. Will likelihood weighting encounter any problems in this network? Justify your answer.

2. [10 points] **Markov chain**

   Consider the Markov chain shown below. Compute the n-step transition probability matrix:

   $$p_{ij}^{(n)} = P(s_n = j|s_0 = i), \forall i, j \in \{0, 1, 2\}$$

   Does the chain have a steady-state distribution? If it does, compute it. Otherwise, justify why not.

3. [10 points] **Gibbs sampling** As promised in the lecture, show that:

   $$P(x_i|\text{MarkovBlanket}(X_i)) \propto P(x_i|\text{Parents}(X_i)) \prod_{j=1}^{k} P(Y_j|\text{Parents}(Y_j))$$

4. [25 points] **Gibbs sampling**

   Consider the Bayes net shown below:

   $$X \rightarrow Z \leftarrow Y$$

   (a) [10 points] Assume X and Y are uniformly distributed, and Z is the deterministic exclusive or of X and Y. Show that Gibbs sampling on this structure with evidence $Z = 1$ will estimate $P(X = 1|Z = 1)$ as either 1 or 0.

(b) [15 points] What happens if we make Z a slightly noisy exclusive or? E.g. Z is the exclusive or of X and Y with probability $1 - q$ and is chosen uniformly randomly with probability q. Explain this behavior in terms of what happens with the random walk generated by Gibbs sampling.

5. [10 points] **Parameter estimation in Bayes nets**

   Suppose we have a Bayes net in which the likelihood of any variable given its parents, $p(X_i|\text{Parents}(X_i))$ is given by a multinomial distribution. Show that in this case, the probability for any given assignment of values to the parents can be estimated separately from the other assignments (in other words, we can evaluate any row in the CPDs independently)

6. [20 points] **Maximum likelihood estimation**

   For discrete random variables, another distribution of interest is the *Poisson distribution*:

$$p(k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

   where $\lambda \in \Re$ is a parameter.

   (a) [5 points] Suppose that X is a random variable drawn from a Poisson distribution of unknown parameter λ. You observe n i.i.d. samples $x_1, \ldots x_n$ drawn from the distribution. Write the likelihood of parameter λ w.r.t. this sample.

   (b) [10 points] Derive the maximum likelihood estimate for λ. What is the sufficient statistic of the data in this case?

   (c) [5 points] Suppose you computed your MLE estimate, λ_n, based on the initial sample. You get a new sample, x_{n+1}. Write a formula that updates the old value, λ_n, to a new value, λ_{n+1}. Your formula for λ_{n+1} should be in terms of λ_n and x_{n+1}.

7. [15 points] **Learning Bayes net structure** (adapted from Russell & Norvig)

   Suppose we are given a set of data $D = \{\vec{d_1}, \ldots \vec{d_m}\}$ in which $d_j = \langle x_{j1} \ldots x_{jn}\rangle$ has values for all the random variables we are interested in, $X_1, \ldots X_n$. We want to use the log-likelihood scoring function to find a Bayes net B which maximizes the likelihood of the data. More precisely, we want to find a network B which maximizes $L(B|D) = \log \prod_{j=1}^{n} p(d_j|B)$.

   (a) [10 points] Consider two candidate networks, B_1 and B_2, which are identical except that B_2 has one extra arc. In both cases, the parameters in the CPDs have also been determined using maximum likelihood. What can you say about the relationship between $L(B_1|D)$ and $L(B_2|D)$?

   (b) [5 points] What are the consequences of your answer from (a) in terms of the learning algorithm?