Practice Problems for CSC 412/2506 Midterm

1. Let $p(k)$ be a one-dimensional discrete distribution that we wish to approximate, with support on nonnegative integers. One way to fit an approximating distribution $q(k)$ is to minimize the Kullback-Leibler divergence:

   $$KL(p||q) = \sum_{k=0}^{\infty} p(k) \log \frac{p(k)}{q(k)}$$

   Show that when $q(k)$ is a Poisson distribution, this KL divergence is minimized by setting λ to the mean of $p(k)$.

   $$q(k) = \frac{\lambda^k \exp(-\lambda)}{k!}$$

   Solution:

   $$\frac{\partial KL}{\partial \lambda} = 0 \Rightarrow \lambda = E[p(k)]$$

2. Recall that the definition of an exponential family model is:

   $$f(x|\eta) = h(x) g(\eta) \exp(\eta^\top T(x))$$

   where:

   $$\begin{aligned} \eta & \quad \text{are the parameters} \\ T(x) & \quad \text{are the sufficient statistics} \\ h(x) & \quad \text{is the base measure} \\ g(\eta) & \quad \text{is the normalizing constant} \end{aligned}$$

   Consider the univariate Gaussian, with mean μ and precision $\lambda = 1/\sigma^2$ is:

   $$p(D|\mu, \lambda = \prod_{i=1}^{N} \left(\frac{\lambda}{2\pi}\right)^{1/2} \exp(-\frac{\lambda}{2}(x_i - \mu)^2)$$

   What are η and $T(x)$ for this distribution when it is represented in exponential family form?

   Solution:

   $$p(D|\mu, \lambda = (2\pi)^{-N/2}[\lambda^{1/2} \exp(-\frac{\lambda}{2}\mu^2)]^N \exp[\mu\lambda \sum_i x_i - \lambda/2 \sum_i x_i^2]$$

   $\eta = [\mu\lambda \quad ; \quad -\lambda/2]$

   $T(x) = [\sum_i x_i \quad ; \quad \sum_i x_i^2]$

3. Consider the DAG in Figure 1. List all variables that are independent of A given evidence on B. You can use the "Bayes ball" algorithm, the d-separation criterion, or the method of converting to an undirected graph (all should give the same result).

   Solution:

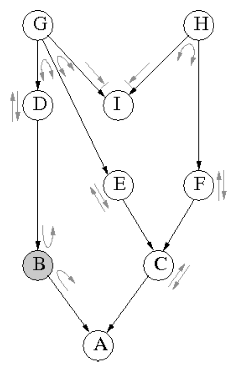

Figure 1: A directed graphical model.

4. Variable Elimination: Murphy 20.1

   Solution: See Figure 2.

   (a). The largest intermediate term has size 3 (we connect 1,2,3 and 4,5,6).

   (b). The largest maxclique has size 3.

   (c). The largest intermediate term has size 4 (we connect 2,3,4,5).

   (d). The largest maxclique has size 4.

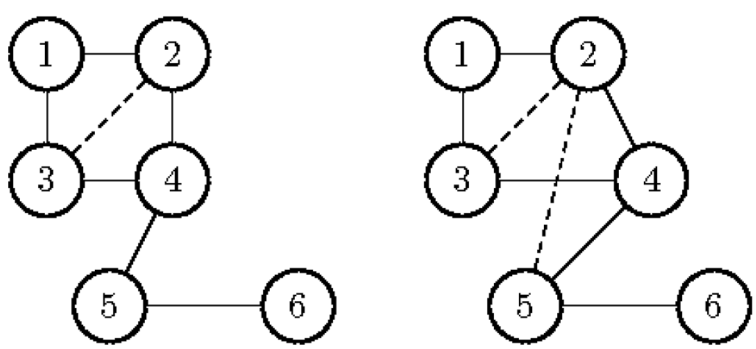

Figure 2: MRF with filled-in edges for two different orderings.

5. Consider a tree-structured factor graph over discrete variables, and suppose we wish to evaluate the joint distribution $[(x_a, x_b)$ associated with two variables x_a and x_b that do not belong to a common factor. Define a procedure for using the sum-product algirhtm to evaluate this joint distribution in which one of the variables is successively clamped to each of its allowed values.

Solution:

We start by using the product and sum rules to write

$$p(x_a, x_b) = p(x_b|x_a)p(x_a) = \sum_{\mathbf{x}_{\setminus ab}} p(\mathbf{x})$$

where $\mathbf{x}_{\setminus ab}$ denote the set of all all variables in the graph except x_a and x_b. We can use the sum-product algorithm to first evaluate $p(x_a)$, by marginalizing over all other variables (including x_b). Next we successively fix x_a at all its allowed values and for each value, we use the sum-product algorithm to evaluate $p(x_b|x_a)$, by marginalizing over all variables except x_b and x_a, the latter of which will only appear in the formulae at its current, fixed value. Finally, we use the above equation to evaluate the joint distribution $p(x_a, x_b)$.