CS281A/Stat241A Homework Assignment 3 (due March 15, 2007)

1. **(Exponential family)**
   Consider the following models:

$$\left\{p(x) = \frac{\lambda^x e^{-\lambda}}{x!} : \lambda \geq 0\right\}, \qquad \text{where } x \text{ is a non-negative integer.} \tag{1}$$

$$\left\{p(x) = \sum_{i=1}^{l} x_i \pi_i : \pi_i \geq 0, \sum_i \pi_i = 1\right\} \qquad \text{where } x_i \in \{0,1\}, \sum_i x_i = 1. \tag{2}$$

$$\left\{p(x) = \frac{\Gamma\left(\sum_{i=1}^{l} \alpha_i\right)}{\prod_{i=1}^{l} \Gamma(\alpha_i)} \prod_{i=1}^{l} x_i^{\alpha_i - 1} : \alpha_i \geq 0\right\} \qquad \text{where } x_i \geq 0, \sum_{i=1}^{l} x_i = 1. \tag{3}$$

   For each of these parameterized classes of distributions,

   (a) Show that it is in the exponential family.
   Specify the natural parameter, sufficient statistic, log normalization and reference measure.

   (b) If p_η is the distribution with natural parameter η, what is the KL-divergence $D(p_\eta \parallel p_{\eta'})$ between p_η and $p_{\eta'}$?

   (c) Specify a conjugate prior in the exponential family. Use the natural parameterization.
   (Note: For (3), it is fine to express the log normalization as an integral.)

2. **(Estimation in the exponential family)**
   On the course website, there is a dataset hw3-1.data with n lines and one x^i on each line. We wish to estimate the parameter of the model (1).

   (a) For each $i = 1, \ldots, n$, use the initial subsequence $x^1, \ldots, x^i$ of data from the website to calculate a maximum likelihood estimate $\hat{\eta}_{ML,i}$ of the natural parameter.
   Plot $\hat{\eta}_{ML,i}$ versus i.
   Plot $D(p_{\hat{\eta}_{ML,n}} \parallel p_{\hat{\eta}_{ML,i}})$ versus i.

   (b) Using the conjugate prior from Question 1, choose a value for the conjugate parameters and, for each $i = 1, \ldots, n$, use the initial subsequence $x^1, \ldots, x^i$ of data from the website to compute the mode $\hat{\eta}_i$ of the posterior distribution of the natural parameter.
   Plot $\hat{\eta}_i$ versus i.
   Plot $D(p_{\hat{\eta}_{ML,n}} \parallel p_{\hat{\eta}_i})$ versus i.

   Repeat (a) and (b) for the dataset hw3-2.data, with model (2). For (b), use conjugate parameters so that the prior, expressed in the moment parameterization, is non-uniform.

3. **(Chordal Graphs)** For each graph of Figure 1, give one example of each of the following (or demonstrate why such an example does not exist):

   (a) A chordless cycle.

   (b) A decomposition. (As defined in lectures: If the graph is complete, a decomposition is the graph itself, otherwise a decomposition is a three-way partition of the graph's vertices, $V = A \cup B \cup C$ for which B separates A from C, together with decompositions of the two subgraphs that remain when one of A and C is removed).

   (c) A simplicial vertex sequence.

   (d) A junction tree.

   (e) An elimination ordering for which the reconstituted graph returned by the UndirectedGraphEliminate algorithm is the same as this graph.

   (f) An orientation for the edges so that the resulting directed graph is acyclic and has this graph as its moral graph.

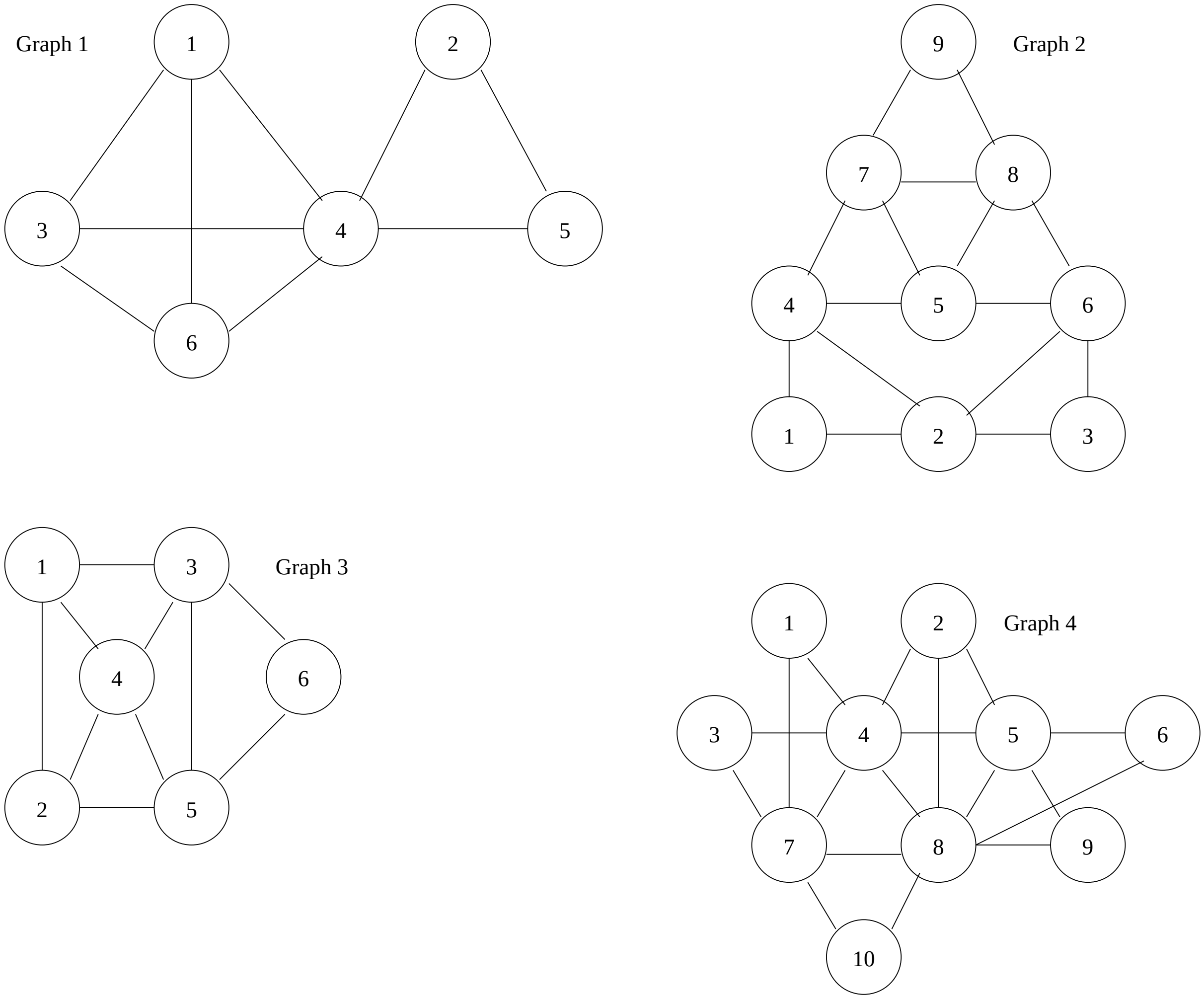

Figure 1: Some undirected graphs.

4. **(Simplicial vertices and junction trees)** We have seen that an undirected graph is recursively simplicial if and only if it has a junction tree. The notes on the website gave a proof via the implications:

$$\text{chordal} \Longrightarrow \cdots \Longrightarrow \text{recursively simplicial} \Longrightarrow \text{has junction tree} \Longrightarrow \text{chordal}.$$

In this question, we consider a direct argument for the implication

$$\text{has junction tree} \Longrightarrow \text{recursively simplicial}.$$

Give (as pseudocode) an algorithm that takes as input a graph G and a junction tree for G and returns a simplicial vertex sequence. Show that your algorithm is correct.