EE-613: Lab 2

Fall 2019

1. **Party Animal**.
   The party animal problem takes into account the following variables:
   - P: been to party
   - H: Got a headache
   - D: Demotivated at work
   - U: Underperform at work
   - A: Boss angry

   The distribution over these variables factorizes as:

   $$p(P, H, D, U, A) = p(P)p(D)p(U|P, D)p(H|P)p(A|U)$$

   The specifications of the individual CPD tables are:

   $$\begin{array}{ll} p(U = tr|P = tr, D = tr) = 0.999 & p(U = tr|P = fa, D = tr) = 0.9 \\ p(U = tr|P = tr, D = fa) = 0.9 & p(U = tr|P = fa, D = fa) = 0.001 \\ p(H = tr|P = tr) = 0.9 & p(H = tr|P = fa) = 0.2 \\ p(A = tr|U = tr) = 0.95 & p(A = tr|U = fa) = 0.5 \\ p(P = tr) = 0.2 & p(D = tr) = 0.4 \end{array}$$

   (a) Draw the graphical model corresponding to this problem.

   (b) Given that the boss is angry and that the worker has a headache, what is the probability that the worker has been to the party ?

2. **Asbestos**.
   There is a synergistic relationship between Asbestos (A) exposure, Smoking (S) and Cancer (C). A model describing this relationship is given by

   $$p(A, S, C) = p(C|A, S)p(A)p(S)$$

   (a) Is $A \perp\!\!\!\perp S$? Is $A \perp\!\!\!\perp S|C$?

   (b) How could you adjust the model to account for the fact that people who work in the building industry have a higher likelihood to also be smokers and also a higher likelihood to be exposed to Asbestos ?

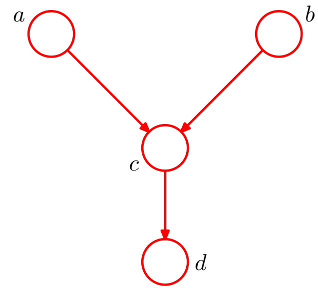

Figure 1: Graphical model. Exploring the conditional independence properties.

3. **Conditional independence**.
In the course, we have studied the explaining away effect of the 'head-to-head' graphical model. Here we show that it still holds when the observed variable is a descendent of the head-to-head configuration. Consider the directed graphical model shown in Fig. 1, in which none of the variables is observed. Show that $a \perp\!\!\!\perp b$. Suppose now that we observe d. Show that in general $a \not\perp\!\!\!\perp b|d$.

4. **Chest Clinic Network.**[1]
This network concerns the diagnosis of lung disease (tuberculosis, lung cancer, or both, or neither). In this model, a visit to Asia is assumed to increase the probability of tuberculosis. The involved variables are:
   - x: positive X-ray
   - d: Dyspnea (shortness of breath)
   - e: either tuberculosis or lung cancer
   - t: Tuberculosis
   - l: Lung cancer
   - b: Bronchitis
   - a: visit to Asia
   - s: Smoker

   According to this model, the distribution factorizes as:

$$p(x, d, e, t, l, b, a, s) = p(a)p(s)p(t|a)p(l|s)p(b|s)p(d|b, e)p(e|t, l)p(x|e)$$

   (a) Draw the graphical model associated with this problem.

   (b) State whether the following conditional independence relationships are true of false, and explain why.
   - tuberculosis $\perp\!\!\!\perp$ smoking | shortness of breath
   - lung cancer $\perp\!\!\!\perp$ bronchitis | smoking
   - visit to Asia $\perp\!\!\!\perp$ smoking | lung cancer

[1] S. Lauritzen and D. Spiegelhalter. Local computation with probabilities on graphical structures and their application to expert systems. *Jl of Royal Statistical Society B,* 1988.

- visit to Asia $\perp\!\!\!\perp$ smoking | lung cancer, shortness of breath

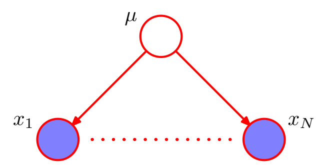

Figure 2: Inference of a Gaussian mean.

5. **Estimation of a Gaussian mean.** Consider the graphical model of Fig. 2, characterized by the following generative process:

$$\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\mu}|\boldsymbol{\mu}_0, \Sigma_0) \quad (1)$$
$$\mathbf{x}_i \sim \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \Sigma), \forall i \quad (2)$$

and let us denote $\mathcal{X} = \{\mathbf{x}_i, i = 1 \dots N\}$, and $\Theta = (\boldsymbol{\mu}_0, \Sigma_0, \Sigma)$.

(a) Draw the graphical model using the plate notation.

(b) Using d-separation, state whether $\mathbf{x}_i \perp\!\!\!\perp \mathbf{x}_j \mid \boldsymbol{\mu}$ holds true or not. Similarly, do we have $\mathbf{x}_i \perp\!\!\!\perp \mathbf{x}_j$ in general?

(c) Without computation and exploiting the course, state what is the type of the distribution of $p(\boldsymbol{\mu}|\mathcal{X})$?

(d) Suppose now that the $\mathbf{x}$ are one dimensionnal, so that $\boldsymbol{\mu}_0 = \mu_0, \Sigma_0 = \sigma_0, \Sigma = \sigma$. Compute analytically the parameters defining the distribution $p(\mu|\mathcal{X})$.

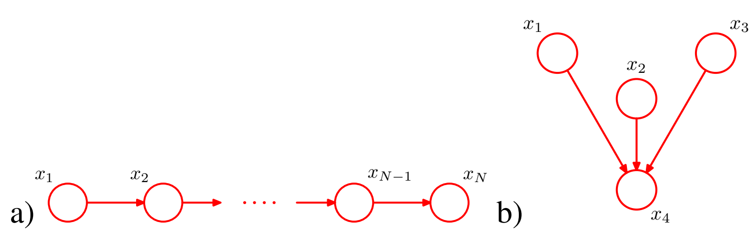

Figure 3: Converting a directed graphical model to an undirected one.

6. **Converting directed graph to undirected graphs.** Consider the graphical models of Fig. 3. For each of them:

(a) Write down the corresponding pdf.

(b) Propose an undirected model that has a similar factorisation, and provide the corresponding graph, potential function, and normalization constant. What can you notice in the case of the second example? Do we keep the same conditional independence statements ?