1 Probabilistic Graphical Models

Recall that we can represent joint probability distributions with directed acyclic graphs (DAGs). Let G be a DAG with vertices $X_1, ..., X_k$. If P is a (joint) distribution for $X_1, ..., X_k$ with (joint) probability mass function p, we say that G represents P if

$$p(x_1, \cdots, x_k) = \prod_{i=1}^{k} P(X_i = x_i | \text{pa}(X_i)), \tag{1}$$

where $\text{pa}(X_i)$ denotes the parent nodes of X_i. (Recall that in a DAG, node Z is a parent of node X iff there is a directed edge going out of Z into X.)

Consider the following DAG

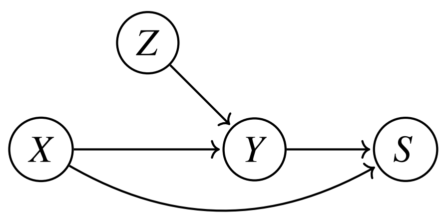

Figure 1: G, a DAG

(a) Write down the joint factorization of $P_{S,X,Y,Z}(s, x, y, z)$ implied by the DAG G shown in Figure 1.

(b) Is $S \perp Z \mid Y$?

(c) Is $S \perp X \mid Y$?

2 PGMs: Sleeping in Class

In this question, you'll be reasoning about a Dynamic Bayesian Network (DBN), a form of a Probabilistic Graphical Model.

Your favorite discussion section TA wants to know if their students are getting enough sleep. Each day, the TA observes the students in their section, noting if they fall asleep in class or have red eyes. The TA makes the following conclusions:

1. The prior probability of getting enough sleep, S, with no observations, is 0.7.
2. The probability of getting enough sleep on night t is 0.8 given that the student got enough sleep the previous night, and 0.3 if not.
3. The probability of having red eyes R is 0.2 if the student got enough sleep, and 0.7 if not.
4. The probability of sleeping in class C is 0.1 if the student got enough sleep, and 0.3 if not.

(a) Formulate this information as a dynamic Bayesian network that the professor could use to filter or predict from a sequence of observations. If you were to reformulate this network as a hidden Markov model instead (that has only a single observation variable), how would you do so? Give a high-level description (probability tables for the HMM formulation are not necessary.)

(b) Consider the following evidence values at timesteps 1, 2, and 3:

(a) e_1 = not red eyes, not sleeping in class

(b) e_2 = red eyes, not sleeping in class

(c) e_3 = red eyes, sleeping in class

Compute state estimates for timesteps t at 1, 2, and 3; that is, calculate $P(S_t|e_{1:t})$. Assume a prior on $P(s_0)$ that is consistent with the prior in the previous part; that is, $P(s_0) = 0.7$.

(c) Compute smoothing estimates $P(S_t|e_{1:3})$ for each timestep, using the same evidence as the previous part.

(d) Compare, in plain English, the filtered estimates you computed for timesteps 1 and 2 with the smoothed estimates. How do the two analyses differ?

