10-708 Probabilistic Graphical Models

Recitation 1: Bayesian Networks & Undirected Graphical Models

January 27, 2017 *Maruan Al-Shedivat*

1 D-separation in BNs & UGMs

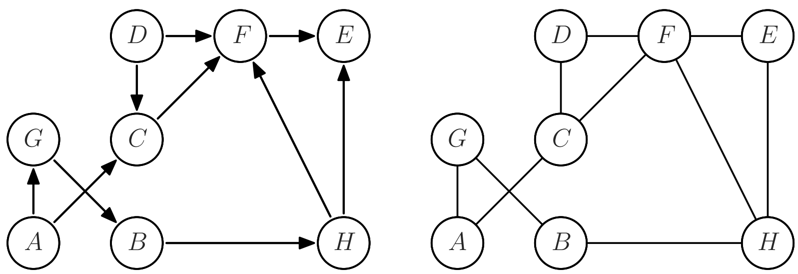

Figure 1: BN and its skeleton.

(a) Find a set that D-separates: (i) A and F, (ii) C and H.

(b) Is this true: $A \perp\!\!\!\perp H \mid G, E$?

(c) Same questions as in (a), (b), but for an undirected version of the graph.

(d) Construct a *moralized* version of the the graph. Does the moralized version help to answer (a) & (b)?

2 I-Equivalence

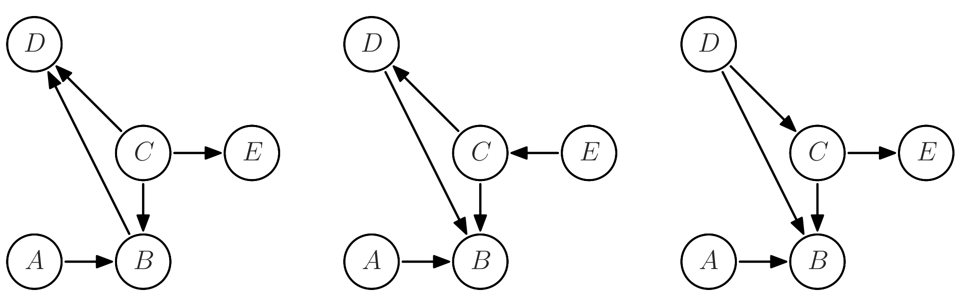

Figure 2: I-equivalent or not?

Are the graphs I-equivalent? (*Hint:* Consider their skeletons and sets of immoralities[1].)

[1] **Definition** (Immorality). *A v-structure* $X \to Z \leftarrow Y$ *is called an* immorality *if there is no direct edge between* X *and* Y*.*

3 Minimal I-maps

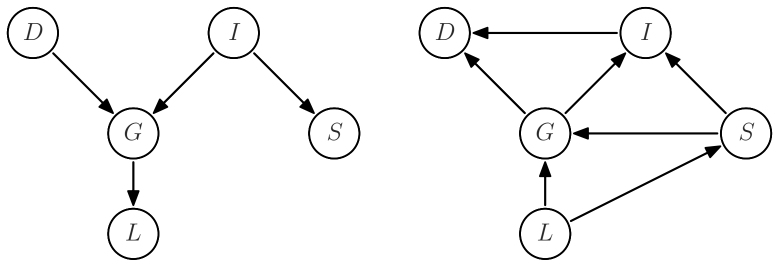

Figure 3: *Left:* P-map of a distribution. *Right:* a minimal I-map or not?

You are given a graph G that corresponds to a distribution P (i.e., think of it as a perfect map of the distribution that encodes all the independencies).

(a) Show that the other graphs is a minimal I-maps of that distribution[2].

(b) Construct another minimal I-map from the following ordering of the nodes: L, D, S, I, G. To construct a minimal I-map, you could use the following steps:

- Fix an ordering of the nodes, $X_1, \ldots, X_n$.
- Now, examine each variable X_i in the specified order, $i = 1, \ldots, n$. For each X_i pick some minimal subset $\mathbf{U} \subseteq \{X_1, \ldots, X_{i-1}\}$ that satisfies $(X_i \perp\!\!\!\perp \{X_1, \ldots, X_{i-1}\} - \mathbf{U} \mid \mathbf{U})$. Make the nodes of this subset $\mathbf{U}$ to be parents of X_i.

(c) Does the given algorithm construct a minimal I-map? Check that the graph constructed in (b) is a minimal I-map.

4 Markov random fields

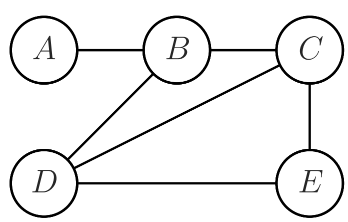

Figure 4: Markov random field.

(a) Write out the canonical representation of the Gibbs distribution in the exponential form for the undirected graph given in Figure 4.

[2]**Definition** (Minimal I-map). *A graph $\mathcal{K}$ is called a minimal I-map of for a set of independencies $\mathcal{I}$ if it is an I-map for $\mathcal{I}$, and if the remoal of even a single edge from $\mathcal{K}$ renders it not an I-map.*

(b) In this question, you will design a very simple undirected graphical model for denoising the output of a black-and-white printer with a defective cartridge head. Our old printer is defective in two ways. Sometimes when the correct output is black it outputs white, and vice versa. Furthermore, sometimes the cartridge gets stuck and has difficulty putting down any ink for an entire row of pixels. We represent the unobserved true output for an $m \times n$ image as $Z \in \{0,1\}^{m\times n}$, where 1 denotes black ink and 0 denotes no ink. Furthermore, we have unobserved $Y \in \{0,1\}^m$, where 1 denotes that on this row of pixels the cartridge got stuck and thus puts down less ink. Finally, we have observed data $X \in \{0,1\}^{m\times n}$. We parameterize the distribution as follows:

$$\psi_\alpha(Z_{ij}) = \alpha^{\{Z_{ij}=1\}} \tag{1}$$

$$\psi_\beta(Z_{ij}, Z_{i,j+1}) = \beta^{\{Z_{ij}=Z_{i,j+1}\}} \tag{2}$$

$$\psi_\gamma(Z_{ij}, Z_{i+1,j}) = \gamma^{\{Z_{ij}=Z_{i+1,j}\}} \tag{3}$$

$$\psi_\theta(Z_{ij}, X_{ij}) = \theta^{\{X_{ij}=Z_{ij}\}} \tag{4}$$

$$\psi_\mu(Y_i, X_{ij}) = \mu^{\{X_{ij}=1 \wedge Y_i=1\}} \tag{5}$$

$$\tag{6}$$

- Draw the corresponding graphical model for $m = 3$, $n = 3$. Also, for each of the parameters $(\alpha, \beta, \gamma, \mu)$, based on your intuition, state whether each is probably < 1 or > 1.
- Suppose we formulate the denoising problem as a CRF where we condition on $\mathbf{X}$. Write out the partition function that corresponds to $P(\mathbf{Y}, \mathbf{Z}|\mathbf{X}; \alpha, \beta, \gamma, \theta, \mu)$.

5* Bayesian reasoning as an extension to propositional calculus

Frequentist statistics views probabilities only as long-run frequencies of repeatable events, in contrast to Bayesian theory, in which probabilities may describe degrees of belief or states of partial knowledge. However, why does Bayesian calculus of belief has to follow the well known probability rules, i.e., where do these fundamental "sum" and "product" rules come from? It turns out that it is possible to define a set of axioms and show that Bayesian reasoning is a natural extension of the classical deductive logic.

In this exercise, we ask you to derive/prove product rule and sum rules for the degree of belief from a set of initial statements (axioms). To avoid confusion between the "degree of belief" and "probability," here we associate the former with a quantity called "plausibility." Again, the goal is to show that the sum and product rules for plausibility follow from the five statements given below.

(R1) $(A \mid X)$, the plausibility of A given X, is a single real number. There exists a real number T such that $(A \mid X) \leq T$ for every X and A.

(R2) Plausibility assignments are compatible with the propositional calculus:

  (1) If A is equivalent to A' then $(A \mid X) = (A' \mid X)$ for any consistent X.

  (2) If A is tautology them $(A \mid X) = T$.

  (3) $(A \mid B, CX) = (A \mid (B \wedge C), X)$.

  (4) If X is consistent and $(\neg A \mid X) < T$, then A, X si also consistent.

(R3) There exists a nonincreasing function S_0 such that $(\neg A \mid X) = S_0(A \mid X)$ for all A and consistent X.

(R4) There exists a nonempty set of real numbers P_0 with the following two properties:

  – P_0 is a dense subset of (F, T). That is, for every pair of real numbers a, b such that $F \leq a < b \leq T$, there exists some $c \in P_0$ such that $a < c < b$.

  – For every $y_1, y_2, y_3 \in P_0$ there exists some consistent X with a basis of at least three atomic propositions—call them A_1, A_2, A_3—s.t. $(A_1 \mid X) = y_1$, $(A_2 \mid A_1 X) = y_2$, $(A_3 \mid A_1, A_2, X) = y_3$.

(R5) There exists a continuous function $F : [F, T]^2 \mapsto [F, T]$, strictly increasing in both arguments on $(F, T]^2$, such that $(A \wedge B \mid X) = F((A \mid B, X), (B \mid X))$ for any A, B and consistent X.

Definitions that you should use:

Definition 1. *We say that X is consistent if there is no proposition A for which $(A \mid X) = T$ and $(\neg A \mid X) = T$.*

Definition 2. *Tautology is a proposition that is true regardless of truth or falsity of its atomic propositions.*

Remark. This problem is beyond the scope of the class and we suggest you work on it only if you are interested in the question of consistency of Bayesian inference. To better understand the problem and the historical context, you may refer to the following paper:

Kevin S. Van Horn, "Constructing a logic of plausible inference: a guide to Cox's theorem", in *International Journal of Approximate Inference 34 (2003) 3–24.*