Graphical Models: Worksheet 1 MT 2023

Warm-up and Optional questions will not be marked, but solutions will be provided.

A: Warm-Up

A1. Conditional Independence

Let X, Y be independent random variables taking the value 0 with probability $\frac{1}{2}$, and 1 otherwise. Now let Z be a random variable with conditional distribution

$$P(Z = 1 \mid X = x, Y = y) = \begin{cases} \frac{3}{4} & \text{if } x = y \\ \frac{1}{4} & \text{if } x \neq y \end{cases}$$

with $P(Z = 0 \mid X = x, Y = y) = 1 - P(Z = 1 \mid X = x, Y = y)$.

(a) Find $P(X, Y \mid Z = 1)$.

(b) Show that $X \perp\!\!\!\perp Z$ and $Y \perp\!\!\!\perp Z$ but that $X, Y \not\!\perp\!\!\!\perp Z$.

A2. Prove that properties (ii) and (iv) of Theorem 2.4 are equivalent to (i), (iii) and (v).

A3. Show that $X \perp\!\!\!\perp Y \mid Z$ is equivalent to

$$p(x, y, z) \cdot p(x', y', z) = p(x', y, z) \cdot p(x, y', z)$$

for Z-almost all x, x', y, y', z.

B: Core Questions

B1. Graphoids.

By completing the implications of the theorem from lectures, show that

$$X \perp\!\!\!\perp Y, W \mid Z \iff X \perp\!\!\!\perp W \mid Z \text{ and } X \perp\!\!\!\perp Y \mid W, Z.$$

Show that, in general,

$$X \perp\!\!\!\perp Y \mid Z \text{ and } X \perp\!\!\!\perp Z \mid Y \quad \not\Longrightarrow \quad X \perp\!\!\!\perp Y, Z.$$

[Hint: $X \perp\!\!\!\perp Y \mid Y$ for any X, Y.]

B2. Some Strange Independences.

(a) Let (X_1, X_2, X_3) follow a multivariate Gaussian distribution with covariance matrix Σ. Show that $X_1 \perp\!\!\!\perp X_2 \mid X_3$ if and only if

$$\sigma_{33}\sigma_{12} - \sigma_{13}\sigma_{23} = 0.$$

(b) Deduce that for jointly Gaussian random variables,

$$X_1 \perp\!\!\!\perp X_2 \mid X_3 \text{ and } X_1 \perp\!\!\!\perp X_2 \quad \Longleftrightarrow \quad X_1 \perp\!\!\!\perp X_2, X_3 \text{ or } X_2 \perp\!\!\!\perp X_1, X_3.$$

(c) Let X, Y be discrete random variables. Show that $X \perp\!\!\!\perp Y \mid Z = z$ if and only if the matrix $M^z = (\pi_{xyz})_{x,y}$, where $\pi_{xyz} = P(X = x, Y = y, Z = z)$, has rank one. *[Hint: Recall that a matrix M has rank one if and only if it can be written as $M = \alpha\beta^T$ for vectors α, β.]*

(d) Let A, B be $a \times c$ and $b \times c$ matrices each of rank c. Show that AB^T also has rank c.

(e) Hence, or otherwise, show that for binary Z and finite discrete X, Y we have

$$X \perp\!\!\!\perp Y \mid Z \text{ and } X \perp\!\!\!\perp Y \quad \Longleftrightarrow \quad X \perp\!\!\!\perp Y, Z \text{ or } Y \perp\!\!\!\perp X, Z.$$

[Hint: show that if $X \perp\!\!\!\perp Y \mid Z$, then $(\pi_{xy+})_{xy}$ can be written as a product of matrices of rank 2].

B3. Factorization and Conditional Independence.

Consider four binary variables A, B, C, D; let the support (i.e. the set of combinations whose probability is > 0) of these variables be:

$$\begin{aligned}(a, b, c, d) = &(0, 0, 0, 0) \qquad (1, 0, 0, 0) \qquad (1, 1, 0, 0) \qquad (1, 1, 1, 0)\\ &(1, 1, 1, 1) \qquad (0, 1, 1, 1) \qquad (0, 0, 1, 1) \qquad (0, 0, 0, 1).\end{aligned}$$

(a) Show that $A \perp\!\!\!\perp C \mid B, D$ and $B \perp\!\!\!\perp D \mid A, C$ for any distribution with support in this set.

(b) Show that we cannot write the joint distribution in the form

$$P(A = a, B = b, C = c, D = d) = \psi_{ab}(a, b) \cdot \psi_{bc}(b, c) \cdot \psi_{cd}(c, d) \cdot \psi_{da}(d, a).$$

C: Optional

C1. Möbius Inversion

(a) Let $(\zeta_M)_{M \subseteq V}$ be a vector indexed by subsets, and let

$$\eta_M = \sum_{Z \subseteq M} \zeta_Z, \qquad \forall M \subseteq V.$$

Show the *Möbius inversion formula*:

$$\zeta_M = \sum_{Z \subseteq M} (-1)^{|M \setminus Z|} \eta_Z, \qquad \forall M \subseteq V.$$

Deduce that $\eta_M = 0$ for all M if and only if $\zeta_M = 0$ for all M.
[Hint: any non-empty set A has the same number of even-sized subsets as odd-sized subsets.]

(b) Now let X_V be binary variables with joint distribution

$$\log p(x_V) = \sum_{A \subseteq V} \lambda_A(x_A)$$

using the identifiability constraints from lectures. Let $a, b \in V$ and $W = V \setminus \{a, b\}$. By considering

$$\log p(x_a, x_b, x_W) + \log p(1, 1, x_W) - \log p(1, x_b, x_W) - \log p(x_a, 1, x_W)$$

or otherwise, show that the joint distribution factorizes as $p(x_V) = f(x_a, x_W)g(x_b, x_W)$ if and only if $\lambda_{abD} = 0$ for all $D \subseteq W$.

(c) Deduce that a positive distribution $p(x_V)$ on binary variables is Markov with respect to an undirected graph $\mathcal{G}$ if and only if $\lambda_A = 0$ whenever A is not a complete set of vertices in $\mathcal{G}$.

(d) Extend the result to arbitrary discrete variables.

C2. Conditional Expectation. [Involves some measure theory, though could be 'proved' without knowing it]. Given an integrable random variable X and two other random variables Y, Z, we say that X is *conditionally independent* of Y given Z if for any integrable $f(X)$ we have

$$\mathbb{E}[f(X) \mid Y, Z] = \mathbb{E}[f(X) \mid Z] \qquad \text{(a.s.)}.$$

Equivalently,

$$\mathbb{E}[f(X, Z) \mid Y, Z] = \mathbb{E}[f(X, Z) \mid Z] \qquad \text{(a.s.)}.$$

[Why is this equivalent?]

Consider the following alternative statements.

A. $\mathbb{E}[f(X, Z)g(Y, Z)] = \mathbb{E}[\mathbb{E}[f(X, Z) \mid Z]\mathbb{E}[g(Y, Z) \mid Z]]$ for all integrable f, g.
B. $\mathbb{E}[f(X, Z)g(Y, Z) \mid Z] = \mathbb{E}[f(X, Z) \mid Z]\mathbb{E}[g(Y, Z) \mid Z]$ a.s. for all integrable f, g.

Show that A and B are equivalent to one another, and also to the definition of conditional independence.

[Hint: you will need the tower property*: $\mathbb{E}[X \mid Y] = \mathbb{E}[\mathbb{E}[X \mid Y, Z] \mid Y]$ holds for any Y, Z and integrable X, and* 'taking out what is known'*: $\mathbb{E}[f(X)g(Z) \mid Z] = g(Z)\mathbb{E}[f(X) \mid Z]$.]*