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

By symmetry we can see that $P(Z = 1) = \frac{1}{2}$, *so*

$$\begin{aligned} P(X = x, Y = y \mid Z = 1) &= 2P(X = x)P(Y = y)P(Z = 1 \mid X = x, Y = y) \\ &= \frac{1}{2}\frac{1 + 2\mathbb{1}\{x = y\}}{4} = \frac{1 + 2\mathbb{1}\{x = y\}}{8}. \end{aligned}$$

That is, $3/8$ *if* $x = y$, $1/8$ *otherwise.*

Importantly, we see that $X \not\perp\!\!\!\perp Y \mid Z$ *even though* $X \perp\!\!\!\perp Y$.

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

$P(X = x, Z = z) = P(X = x, Z = z, Y = 0) + P(X = x, Z = z, Y = 1) = \frac{1}{4}$ *for any* x, z. *Hence* $P(X = x \mid Z = z) = \frac{1}{2}$, *and* $X \perp\!\!\!\perp Z$. *The joint independence clearly doesn't hold, by the definition given of* $P(Z \mid X, Y)$.

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

See the proofs in the lecture notes.

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.

Given z *with* $p(z) > 0$, *there exist* x', y' *such that* $p(x', y', z) > 0$; *divide both sides by* $p(x', y', z)$ *and treat* x', y' *as constants; then* p *factorizes as in Theorem 2.4(v). For the converse, it is easy to check that if Theorem 2.4(v) holds, plugging in its expression gives us the equality above.*

Alternatively for the first direction, if the equality holds we can integrate out x', y' *to obtain*

$$\begin{gathered} p(x, y, z) \int p(x', y', z)\, dx'\, dy' = \int p(x', y, z)\, dx' \cdot \int p(x', y', z)\, dy' \\ p(x, y, z) \cdot p(z) = p(y, z) \cdot p(x, z), \end{gathered}$$

which implies the conditional independence.

B: Core Questions

B1. Graphoids.

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

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

Property 3 is immediate from Theorem 2.4, since $X \perp\!\!\!\perp Y, W \mid Z$ implies that $p(x, y, w, z) = f(x, z)g(y, w, z)$, which implies $X \perp\!\!\!\perp W \mid Y, Z$. For property 4, the left hand sides suggest that

$$p(w \mid x, y, z) = p(w \mid y, z) \qquad p(y \mid x, z) = p(y \mid z),$$

so multiplying these together gives $p(w, y \mid x, z) = p(w, y \mid z)$ as required.

Show that, in general,

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

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

Choosing Y and Z to have a degenerate distribution and X correlated with them generally does the trick. A simple example: let $P(X = Y = Z) = 1$ with $P(X = 0) = P(X = 1) = \frac{1}{2}$. Then it is easy to check that $X \perp\!\!\!\perp Y \mid Z$ and $X \perp\!\!\!\perp Z \mid Y$ directly because the distribution is degenerate after we condition on either value. However, clearly $P(X = 0) = \frac{1}{2} \neq 1 = P(X = 0 \,|\, Y = Z = 0)$, for example.

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

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

We know from lectures that the independence holds if and only if $k_{12} = 0$, where $K = \Sigma^{-1}$ is the concentration matrix. The quantity above is the corresponding entry in the adjoint matrix for Σ, and is therefore the same as k_{12} up to -1 times the (positive) determinant of Σ. Hence these quantities are zero at the same time.

(b) Deduce that for jointly Gaussian random variables,

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

This follows immediately from plugging in $X_1 \perp\!\!\!\perp X_2$ ($\sigma_{12} = 0$) into the equation deduced above, and then noting that we obtain that either $X_1 \perp\!\!\!\perp X_3$ or $X_2 \perp\!\!\!\perp X_3$. Then one can notice that (in either case) Σ and hence $K = \Sigma^{-1}$ is block diagonal, and hence we respectively obtain that $X_1 \perp\!\!\!\perp X_2, X_3$ 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.]

We know that this is equivalent to $p(x, y, z) = p(x, z)p(y \mid z) = \alpha_x^z \beta_y^z$, which implies that $M^z = (p(x, y, z))_{x,y}$ is the outer product of two vectors α^z, β^z. Hence M^z has rank 1. For the converse, if M^z has rank 1 then it is the outer product of two vectors, so $M_{xy}^z = \alpha_x^z \beta_y^z$ and, from lectures (specifically the equivalence of characterizations (iii) and (v) in Theorem 2.4), this is equivalent to the original representation of π_{xyz}.

(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.

Note that $a, b \geq c$ in order for A and B to have rank c. The column span of AB^T is a subspace of that of A, so the rank is clearly at most c. On the other hand since B^T has full column rank c, by selecting the correct vector $x \in \mathbb{R}^b$ it is clear that $B^T x$ spans $\mathbb{R}^c$. Hence AB^T has the same column span as A, and therefore they both have 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].

We obtain

$$\pi_{xy+} = \pi_{xy0} + \pi_{xy1} = \alpha_x^0 \beta_y^0 + \alpha_x^1 \beta_y^1 = \begin{pmatrix} \alpha_x^0 & \alpha_x^1 \end{pmatrix} \begin{pmatrix} \beta_y^0 \\ \beta_y^1 \end{pmatrix}.$$

Hence the matrix π_{xy+} can be written as

$$(\pi_{xy+})_{xy} = \begin{pmatrix} \alpha^0 & \alpha^1 \end{pmatrix} \begin{pmatrix} \beta^0 & \beta^1 \end{pmatrix}^T = AB^T.$$

where $\alpha^0 = (\alpha^0_x)_x$ *and* $\beta^0 = (\beta^0_y)_y$ *are vectors. Then* $(\pi_{xy+})_{xy}$ *is a product of matrices* A *and* B *with 2 columns, and so by the previous part has rank 2 unless either* A *or* B *has rank 1. We know that* $X \perp\!\!\!\perp Y$ *means it must have rank 1.*

If A *is rank 1 this means that* α^0 *and* α^1 *are proportional, i.e.* $p(x \mid z = 0) = p(x \mid z = 1)$*, so* $X \perp\!\!\!\perp Z$*. Combining with* $X \perp\!\!\!\perp Y \mid Z$ *gives* $X \perp\!\!\!\perp Y, Z$*. If* B *is rank 1 then we obtain* $Y \perp\!\!\!\perp X, Z$ *similarly.*

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.

Note, for example, that if $B = D = 0$ *we always have* $C = 0$*. Hence* C *is a.s. constant on this set and thus trivially independent of* A*. Similar observations hold for other combinations of* B, D *and also for the other conditional independence.*

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

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

The key here is that for any pair of numbers in a factor (say A, B*), every combination of their values is possible. It therefore follows that the factor* ψ_{ab} *is non-zero for every* a, b*. But this means that* p *is non-zero for every combination of values* a, b, c, d*, which contradicts the restriction on the support of* p*.*

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.]
Using the definition,

$$\begin{aligned}
\sum_{Z: Z \subseteq M} (-1)^{|M \setminus Z|} \eta_Z &= \sum_{Z: Z \subseteq M} (-1)^{|M \setminus Z|} \sum_{W: W \subseteq Z} \zeta_W \\
&= \sum_{Z, W: W \subseteq Z \subseteq M} (-1)^{|M \setminus Z|} \zeta_W \\
&= \sum_{W \subseteq M} \sum_{Z': Z' \subseteq M \setminus W} (-1)^{|(M \setminus W) \setminus Z'|} \zeta_W \\
&= \sum_{W \subseteq M} (-1)^{|M \setminus W|} \zeta_W \sum_{Z': Z' \subseteq M \setminus W} (-1)^{|Z'|}
\end{aligned}$$

If $W = M$ then the inner sum is 1. Otherwise, by the hint, there as as many even as odd sized subsets of $M \setminus W$, so the sum is 0. The only non-zero term corresponds to ζ_M, which gives the result.

(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 $D \subseteq W$.
If $\log p$ can be written that way then the result is obvious. Conversely if p factorizes in that way then

$$\begin{aligned}
\log p(x_a, x_b, x_W) + \log p(1, 1, x_W) - \log p(1, x_b, x_W) - \log p(x_a, 1, x_W) &= 0 \\
\sum_{D \subseteq W} \{\lambda_{abD}(x_a, x_b, x_D) + \lambda_{abD}(1, 1, x_D) - \lambda_{abD}(1, x_b, x_D) - \lambda_{abD}(x_a, 1, x_D)\} &= 0 \\
\sum_{D \subseteq W} \lambda_{abD}(x_a, x_b, x_D) &= 0
\end{aligned}$$

Now we can apply (a), though a subtlety here is that the subsets over ζ are actually indexed by the vector x_W being away from its baseline level. Given a vector x_W define $M(x_W) = \{w \in W : x_W = 2\}$. Then setting $\eta_{M(x_W)} = \sum_{D \subseteq W} \lambda_{abD}(x_a, x_b, x_D)$ for each vector x_W, and noting that $\zeta_{D,x_W} = \lambda_{abD}(x_a, x_b, x_D) = 0$ if any $x_d = 1$ for $d \in D$, we obtain that for each $x_W \in \mathfrak{X}_W$:

$$\eta_{M(x_W)} = \sum_{D \subseteq W} \lambda_{abD}(x_a, x_b, x_D) = \sum_{D \subseteq M(x_W)} \lambda_{abD}(x_a, x_b, x_D).$$

Hence, if each $\eta_{M(x_W)} = 0$, by (a) so is each λ_{abD}.

(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 $\mathcal{G}$.

The previous part shows that $X_a \perp\!\!\!\perp X_b \mid X_W$ if and only if $\lambda_A = 0$ for $\{a, b\} \subseteq A$. Extending this to the entire pairwise property shows that this must hold for every missing edge. Hence $\lambda_A = 0$ unless every pair of vertices in A is joined by an edge in $\mathcal{G}$; that is, unless A is complete.

(d) Extend the result to arbitrary discrete variables.

This is essentially the same as (b), but one must do it for every other level of $\mathfrak{X}_V$.

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'*: $\mathbb{E}[f(X)g(Z) \mid Z] = g(Z)\mathbb{E}[f(X) \mid Z]$.]*

For CI $\implies$ A, first assume without loss of generality that $\mathbb{E}[f(X, Z) \mid Z] = 0$ (else replace f with $\tilde{f} = f - \mathbb{E}[f \mid Z]$). Then we have:

$$\begin{aligned}
\mathbb{E}[f(X, Z)g(Y, Z)] &= \mathbb{E}[\mathbb{E}[\mathbb{E}[f(X, Z)g(Y, Z) \mid Y, Z] \mid Z]] \\
&= \mathbb{E}[\mathbb{E}[\mathbb{E}[f(X, Z) \mid Y, Z]g(Y, Z) \mid Z]] && \textit{(g is Y, Z-measurable)} \\
&= \mathbb{E}[\mathbb{E}[\mathbb{E}[f(X, Z) \mid Z]g(Y, Z) \mid Z]] && \textit{(by conditional independence)} \\
&= 0 && \textit{(by assumption)}.
\end{aligned}$$

as required.

For $A \implies B$*, define* $h(Z) = \mathbb{E}[f(X,Z)g(Y,Z) \mid Z]$*, which we must show is almost surely zero. Then*

$$\begin{aligned} h(Z) &= h(Z)\mathbb{1}_{\{h(Z)\geq 0\}} + h(Z)\mathbb{1}_{\{h(Z)\leq 0\}} \\ &= \mathbb{E}[f(X,Z)g(Y,Z)\mathbb{1}_{\{h(Z)\geq 0\}} \mid Z] + \mathbb{E}[f(X,Z)g(Y,Z)\mathbb{1}_{\{h(Z)\leq 0\}} \mid Z] \\ &\equiv h^+(X) + h^-(Z). \end{aligned}$$

Note that $g(Y,Z)\mathbb{1}_{\{h(Z)\geq 0\}}$ *is smaller in magnitude than* g*, and hence is also integrable. Now using the factorization,*

$$\begin{aligned} \mathbb{E}h^+(X) &= \mathbb{E}[f(X,Z)g(Y,Z)\mathbb{1}_{\{h(Z)\geq 0\}}] \\ &= \mathbb{E}[\mathbb{E}[f(X,Z) \mid Z] \cdot \mathbb{E}[g(Y,Z)\mathbb{1}_{\{h(Z)\geq 0\}} \mid Z]] \\ &= 0. \end{aligned}$$

But since $h^+(Z) \geq 0$*, this means* $h^+(Z) = 0$ *almost surely. Similarly for* h^- *and therefore* h*.*

Finally consider $B \implies CI$*. Let* $g(Y,Z) = \mathbb{1}_{\{\mathbb{E}[f|Y,Z]\geq \mathbb{E}[f|Z]\}}$*. Then, using* B*,*

$$\mathbb{E}[f \mid Z]\mathbb{E}[g \mid Z] = \mathbb{E}[fg \mid Z] = \mathbb{E}[\mathbb{E}[f \mid Y,Z]g \mid Z] \geq \mathbb{E}[\mathbb{E}[f \mid Z]g \mid Z] = \mathbb{E}[f \mid Z]\mathbb{E}[g \mid Z].$$

This shows that

$$\mathbb{E}[(\mathbb{E}[f \mid Y,Z] - \mathbb{E}[f \mid Z])g \mid Z] = 0 \qquad a.s.$$

since the integrand is non-negative, this shows that it is zero almost surely; a similar approach with g *replaced by* $1 - g$ *gives us the result.*