67800 | Probabilistic Methods in Artificial Intelligence | Ex 2

Guy Lutsker 207029448

## Question 1 - Non-positive distributions

Let us compose the following MN:

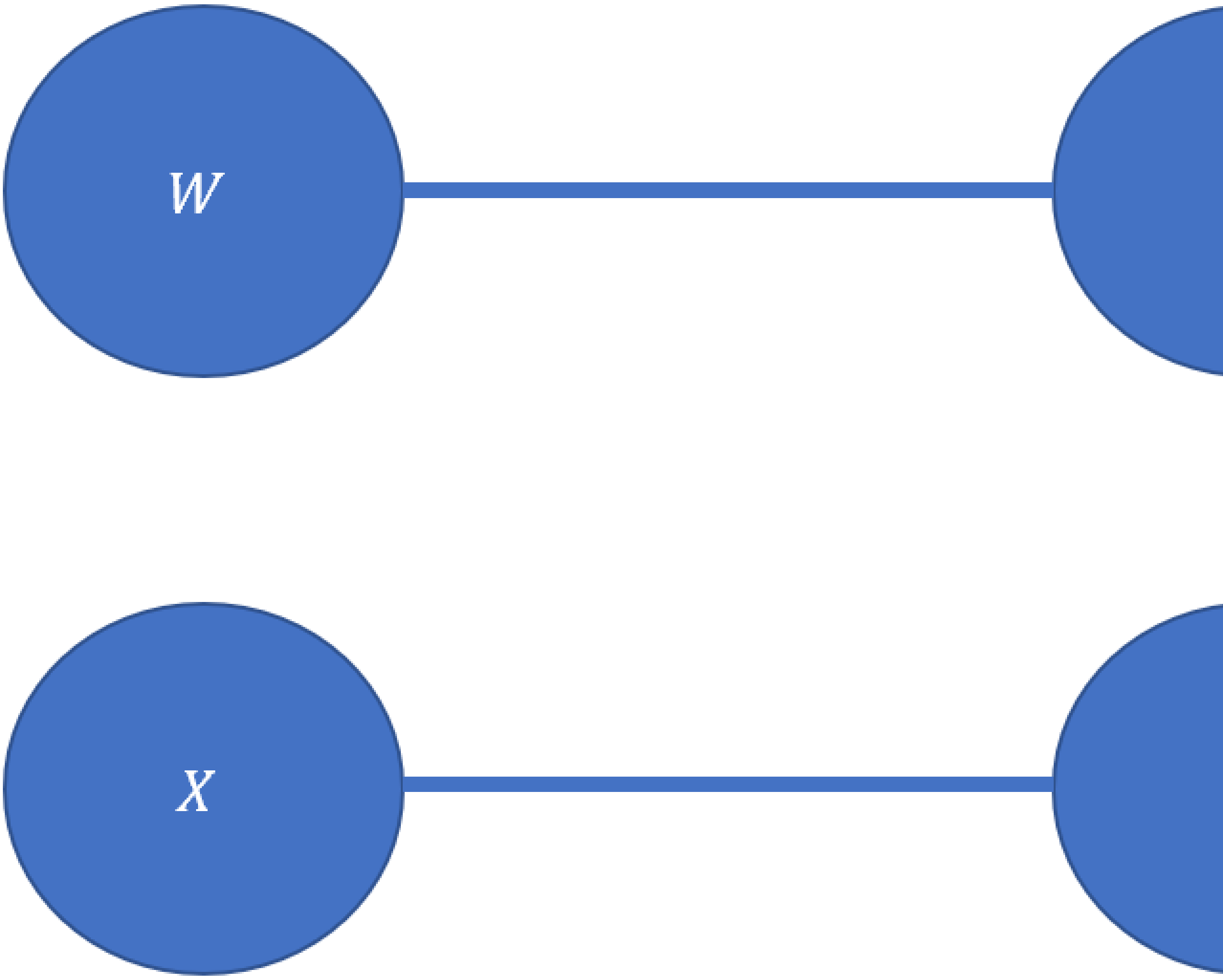


Figure 1: Graph of MN $\mathcal{H}$

As we can see the MN models the following independencies:

$$\mathcal{I}_{LM}(\mathcal{H}) = \{X \perp Z, W|Y \quad , \quad Y \perp Z, W|X, \quad Z \perp X, Y|W, \quad W \perp X, Y|Z\}$$

$$\mathcal{I}(\mathcal{H}) = \{X, Y \perp Z, W\}$$

As we can see $\mathcal{I}_{LM}(\mathcal{H}) \neq \mathcal{I}(\mathcal{H})$ and we would have to construct a distribution $p$ to model this. We shall define $p$ over 3 binary variables $X, Y, Z$ as follows:

$$p(X, Y, Z, W) = \mathbf{1}_{(\boldsymbol{X=Y=Z=W})}$$

Intuitively speaking, $p$ will be $\frac{1}{2}$ when $X = Y = Z = 0$ and when $X = Y = Z = 1$ , and all other cases will give 0 probability.

We need to prove that $p$ satisfies $\mathcal{I}_{LM}(\mathcal{H})$ and not $\mathcal{I}(\mathcal{H})$.

$p$ satisfies $\mathcal{I}_{LM}(\mathcal{H})$ since:

$$p(X=1, Z=1|Y=1) = \frac{p(X=1, Y=1, Z=1)}{p(Y=1)} = \frac{0.5}{0.5} = 1 = \frac{0.5}{0.5} \cdot \frac{0.5}{0.5} = \frac{p(X=1, Y=1)}{p(Y=1)} \cdot \frac{p(Z=1, Y=1)}{p(Y=1)} = p(X=1|$$

Which is identical to when $X = Y = Z = 0$.

also:

$$p(X=1, Z=0|Y=1) = \frac{p(X=1, Y=1, Z=0)}{p(Y=1)} = \frac{0}{0.5} = 0 = \frac{0.5}{0.5} \cdot \frac{0}{0.5} = \frac{p(X=1, Y=1)}{p(Y=1)} \cdot \frac{p(Z=0.Y=1)}{p(Y=1)} = p(X=1|$$

which is identical to $X = 0, Y = Z = 1$.

and:

$$p(X=1, Z=1|Y=0) = \frac{p(X=1, Y=0, Z=1)}{p(Y=0)} = \frac{0}{0.5} = 0 = \frac{0}{0.5} \cdot \frac{0}{0.5} = \frac{p(X=1, Y=0)}{p(Y=0)} \cdot \frac{p(Z=1, Y=0)}{p(Y=0)} = p(X=1|$$

$$\rightarrow p \models \mathcal{I}_{LM}(\mathcal{H})$$

But! if we will look at:

$$p(X=0, Z=0) = 0.5 \neq 0.25 = 0.5 \cdot 0.5 = p(X=0) \cdot p(Z=0)$$

$$\rightarrow p \not\models \mathcal{I}(\mathcal{H})$$

And so $p$ satisfies $\mathcal{I}_{LM}(\mathcal{H})$ and not $\mathcal{I}(\mathcal{H})$ as required.

# Question 2 - Markov Blankets

1. Since $\mathcal{H}$ is a P-map of $p$ we know that $MB_p(X) = ne(X)$ where $ne(X)$ is the set of neighbors of $X$. Proof:

   Since $MB_p(X)$ is the set of all variables which given them $X$ is independent of all other variables in the graph, we know that $ne(X) \subset MB_p(X)$ in any case.

   But since we also know that $\mathcal{H}$ is a P-map of $p$ we also know that all independencies that are present is $p$ are also present in $\mathcal{H}$. In addition we also know that $p$ is positive, and that means that $\mathcal{I}(p) = \mathcal{I}(\mathcal{H}) = \mathcal{I}_{local}(\mathcal{H})$ and This tells us that no node that is outside of $ne(X)$ could be in $MB_p(X)$, since the only independencies that $\mathcal{H}$ has are modeled by the neighbors, and that means that $ne(X) \supset MB_p(X)$ which all together brings us to $ne(X) = MB_p(X)$.

2. In this proof we can see that the underlying assumption here is that for each node there exists a unique MB and that why all edges are included in the I-MAP. and so the algorithm assumes the uniqueness of the MB while it might produce a different graph for the same distribution.

3. Closed under intersection:

   Let $U_1, U_2 \in \mathcal{U}(X)$, we need to show that $U_\cap = U_1 \cap U_2 \in \mathcal{U}(X)$.

   We know that $X_i \perp \mathcal{X}/(U_1 \cup \{X_i\})|U_1$ and in particular $X_i \perp U_2/U_\cap|U_1$ Since $U_2/U_\cap \subset \mathcal{X}/(U_1 \cup \{X_i\})$. And symmetrically we know that $X_i \perp U_1/U_\cap|U_2$.

   Notice that $U_j = U_j/U_\cap \cup U_\cap$ and so we can write that:

$$X_i \perp U_1/U_\cap|U_2/U_\cap, U_\cap$$

$$X_i| \perp U/U_\cap|U_1/U_\cap, U_\cap$$

   And from intersection we get that:

$$X_i \perp U_1/U_\cap, U_2/U_\cap|U_\cap$$

   We can split this term into 2 more easy to comprehend statements:

$$X_i \perp U_1/U_\cap|U_\cap$$

$$X_i \perp U_2/U_\cap|U_\cap$$

   We can also see that : $X_i \perp \mathcal{X}/(U_1 \cup \{X_i\})|U_1 = X_i \perp \mathcal{X}/(U_1 \cup \{X_i\})|U_1/U_\cap, U_\cap$

   And combining $X_i \perp \mathcal{X}/(U_1 \cup \{X_i\})|U_1/U_\cap, U_\cap$ and $X_i \perp U_1/U_\cap|U_\cap$ with the contraction lemma, we get that:

$$X_i \perp \mathcal{X}/(U_1 \cup \{X_i\})|U_1/U_\cap, U_\cap \ and \ X_i \perp U_1/U_\cap|U_\cap \overset{\text{Contraction}}{\rightarrow} X_i \perp \mathcal{X}/U_1|U_\cap$$

   And similarly:

$$X_i \perp \mathcal{X}/(U_2 \cup \{X_i\})|U_2/U_\cap, U_\cap \ and \ X_i \perp U_2/U_\cap|U_\cap \overset{\text{Contraction}}{\rightarrow} X_i \perp \mathcal{X}/U_2|U_\cap$$

   Which gets us to:

$$X_i \perp \mathcal{X}/(U_1 \cup U_2)|U_\cap$$

   In particular $U_\cap \subset (U_1 \cup U_2)$ and so we get that:

$$X_i \perp \mathcal{X}/(U_\cap/\{X_i\})|U_\cap$$

   Which means that :

$$U_1 \cup U_2 = U_\cap \in \mathcal{U}(X) \rightarrow \mathcal{U} \text{ is closed under intersection}$$

4. We know that if $Y \notin MB(X)$ :

$$X \perp Y|MB(X)$$

   In particular we can add any term $Z$ and it will hold that $X \perp Z|MB(X)$:

   And so we can write:

$$X \perp Y, \mathcal{X}/(\{X\} \cup \{Y\} \cup MB(X))|MB(X)$$

And from weak union we get that:

$$X \perp Y | MB(X), \mathcal{X}/(\{X\} \cup \{Y\} \cup MB(X))$$

$$= X \perp Y | \mathcal{X}/(\{X\} \cup \{Y\})$$

5. If we know that $X \perp Y | \mathcal{X}/(\{X\} \cup \{Y\})$ that means that we can say that:

$$p \models X \perp Y | \mathcal{X}/(\{X\})/\overbrace{(\boldsymbol{\mathcal{X}/\{X,Y\}})}^{\in \mathcal{U}}$$

But as marked, $(\mathcal{X}/\{X,Y\})$ is in $\mathcal{U}$ and $Y$ is not in $(\mathcal{X}/\{X,Y\})$ by construction, and so $Y$ cannot be in the intersection of all $\mathcal{U}$ possible which means that $Y \notin MB(X)$.

## Question 3 -Tree factorization

We know that in a MN we have that :

$$p(\mathcal{X}) = \frac{1}{\sum_{x_1,\dots x_n} \prod_i \phi_i(C_i)} \cdot \prod_i \phi_i(C_i)$$

We need to prove that :

$$p(\mathcal{X}) = \prod_{(i,j)\in \mathbf{E}_\mathcal{T}} \frac{p(X_i, X_j)}{p(X_i) \cdot p(X_j)} \cdot \prod_{j \in \mathcal{X}} p(X_j)$$

Let us consider a BN over the same graph of the MN. To have it be a well defined BN we need to add a direction to each edge.

The way we achieve such a construction is by choosing a vertex and basing it as the root of the tree such that all the edges will be facing away from the root. Choosing an arbitrary vertex as suggested will be suffice since it will practically prove the statement for any such directed tree. Let us examine this BN tree structure with the CPD's associated with it. We know that in this tree each vertex will have one parent at most, and so let us write the joint distribution over this BN:

$$p(\mathcal{X}) \overset{\text{BN}}{=} \prod_i p(X_i | pa(i)) \overset{\substack{\text{set } (X_1, \dots X_n) \text{ a topological order over} \\ \text{where } X_1 \text{ is the root we have chosen}}}{=} p(X_1) \cdot \prod_{i>1} p(X_i | pa(i)) \overset{\text{Def}}{=} p(X_1) \cdot \prod_{i>1} \frac{p(X_i, pa(i))}{p(pa(i))}$$

$$\overset{\text{Mult enumerator \& denominator by } p(X_i)}{=} p(X_1) \cdot \prod_{i>1} p(X_i | pa(i)) \overset{\text{Def}}{=} p(X_1) \cdot \prod_{i>1} \frac{p(X_i, pa(i)) \cdot p(X_i)}{p(pa(i)) \cdot p(X_i)}$$

$$= \overbrace{p(X_1) \cdot \prod_{i>1} p(X_i)}^{\prod_{i \in \mathcal{X}} p(X_i)} \cdot \prod_{i>1} \frac{p(X_i, pa(i))}{p(pa(i)) \cdot p(X_i)} = \prod_{i \in \mathcal{X}} p(X_i) \cdot \prod_{i>1} \frac{p(X_i, pa(i))}{p(pa(i)) \cdot p(X_i)}$$

In our tree only $X_1$ has no parents and all other verticies have 1 parent exacctly so we can write :

$$= \prod_{i \in \mathcal{X}} p(X_i) \cdot \prod_{(i,j) \in \mathbf{E}_\mathcal{T}} \frac{p(X_i, X_j)}{p(X_i) \cdot p(X_j)}$$

Now, notice that $p$ factorizes over the MN tree, and so it is an Imap of $p$. in addition also factorizes over the BN tree since it contains no immoralities, it is a perfect map and so we can say both graphs contain the same independencies, which is sufficient.

# Question 4 - Simplifying BN2O

Let us try to achieve the result by first trying and removing just one variable, the last one for convenience sake - $D_k$.

What we really need to do here is to sum out $D_k$ in the following manner:

$$p(s_i^0|D_1,...,D_{k-1}) = \sum_{D_k} p(s_i^0, D_k|D_1,...,D_{k-1}) = \sum_{D_k} p(s_i^0|D_1,...,D_k) \cdot p(D_k|D_1,...,D_{k-1}) \overset{D_i \text{ are idependant}}{=} \sum_{\boldsymbol{D_k}} \boldsymbol{p(s_i^0|D_1,...,D}$$

Recall our definition of: $p(s_i^0|pa_{S_i}^{\mathcal{G}}) = (1-\lambda_{i,0}) \cdot \prod_{D_j \in pa_{S_i}^{\mathcal{G}}} (1-\lambda_{i,j})^{D_j}$ .And so we can write:

$$\sum_{\boldsymbol{D_k}} \boldsymbol{(1-\lambda_{i,0}) \cdot \prod_{D_j \in pa_{S_i}^{\mathcal{G}}} (1-\lambda_{i,j})^{D_j}} \cdot p(D_k)$$

$$= \overbrace{(1-\lambda_{i,0}) \cdot \prod_{j=1}^{k-1} (1-\lambda_{i,j})^{D_j} \cdot p(D_k)}^{\text{Could be used for our reparametrization } \eta} \cdot \sum_{D_k} (1-\lambda_{i,k})^{D_k}$$

And now if we denote our new parameterization $\eta_{i,0} = 1 - [(1-\lambda_{i,0}) \cdot \prod_{j=1}^{k-1} (1-\lambda_{i,j})^{D_j} \cdot p(D_k)]$ we will still have noisy or as required.

This was done for one parameter $D_k$ but now it can be done for $D_{k-1}$ in the same manner, and then for $D_{k-2}, ...D_{l+1}$ as needed.

We can summarize the process and notice that the final argument we seek is:

$$p(s_i^0|D_1,...,D_l) = (1-\lambda_{i,0}) \cdot \prod_{j=1}^{l} (1-\lambda_{i,j})^{D_j} \cdot \prod_{t=l+1}^{k} \sum_{D_t} (1-\lambda_{i,t})^{D_t} \cdot p(D_t)$$