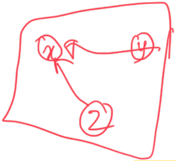
Can All Distributions be Represented as BNs?

$x \perp\!\!\!\perp y, x \perp\!\!\!\perp z, y \perp\!\!\!\perp z, x \not\perp\!\!\!\perp \{x, y\}$

$x = y \oplus z$ $y \perp\!\!\!\perp z$



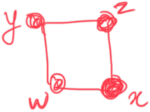
$x \not\perp\!\!\!\perp y$ $y \perp\!\!\!\perp z$

$x \not\perp\!\!\!\perp z$

cannot be perfectly represented by a

Can All Distributions be Represented as BNs?

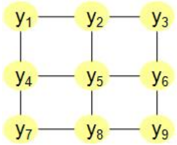
Symmetric dependencies: $\boxed{x \perp\!\!\!\perp y | z, w \qquad z \perp\!\!\!\perp w | x, y}$



x, y, z, w

x, w, y, z

Undirected graphical models (Markov Random Fields)

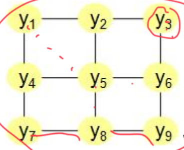


- Graph G: arbitrary undirected graph
- Useful when variables interact symmetrically, no natural parent-child relationship
- Example: labeling pixels of an image.
- Potentials $\psi_C(\mathbf{y}_C)$ defined on arbitrary cliques C of G.
- $\psi_C(\mathbf{y}_C)$: Any arbitrary non-negative value, cannot be interpreted as probability.
- Probability distribution

$$\Pr(y_1 \ldots y_n) = \frac{1}{\boxed{Z}} \prod_{C \in G} \psi_C(\mathbf{y}_C)$$

where $Z = \sum_{\mathbf{y}'} \prod_{C \in G} \psi_C(\mathbf{y}'_C)$ (partition function)

Example



$y_i = 1$ (part of foreground), 0 otherwise.

- Node potentials
  - $\psi_1(0) = 4$, $\psi_1(1) = 1$
  - $\psi_2(0) = 2$, $\psi_2(1) = 3$
  - ....
  - $\psi_9(0) = 1$, $\psi_9(1) = 1$

clique size is 1
$\psi_c(y_c)$ $c = 1, 2, \ldots, 9$
non-maximal cliques

- Edge potentials: Same for all edges
  - $\psi(0,0) = 5$, $\psi(1,1) = 5$, $\psi(1,0) = 1$, $\psi(0,1) = 1$
- Probability: $\Pr(y_1 \ldots y_9) \propto \prod_{k=1}^{9} \psi_k(y_k) \prod_{(i,j)\in E(G)} \psi(y_i, y_j)$

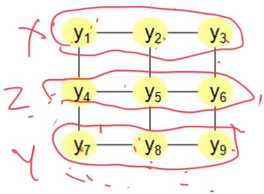
$Z = 2^9$ 0 0 0 0 0 0 0 0 0 $\psi_1(0)\psi_2(0)\psi_3(0) \ldots \psi(0,0)^{11}$

1 1 . . . 1

Conditional independencies (CIs) in an undirected graphical model

Let $V = \{y_1, \ldots, y_n\}$.
Let distribution P be represented by an undirected graphical model G. If Z separates X and Y in G, then $\boxed{X \perp\!\!\!\perp Y|Z}$ in P.
The set of all such CIs are called Global-CI of the UGM.
Example:

1. $y_1 \perp\!\!\!\perp y_3, y_5 y_6, y_7, y_8, y_9 | y_2, y_4$
2. $y_1 \perp\!\!\!\perp y_3 | y_2, y_4, y_5, y_6, y_7, y_8, y_9$
3. $y_1, y_2, y_3 \perp\!\!\!\perp y_7, y_8, y_9 | y_4, y_5, y_6$



Factorization implies Global-CI

Theorem

Let G be a undirected graph over $V = x_1, \ldots, x_n$ nodes and $P(x_1, \ldots, x_n)$ be a distribution. If P is represented by G that is, if it can be factorized as per the cliques of G, then P will also satisfy the global-CIs of G

$$\text{Factorize}(P, G) \implies \text{Global-CI}(P, G)$$

Any subset W, Y, Z:

$$P(W, Y, Z) = \prod_{c \in (W \cup Z)} \psi_c(x_c) \prod_{c \in (Y \cup Z)} \psi_c(x_c) \prod_{c \in (W,Z)} \psi_c(x_c) \prod_{c \in (W,Y,Z)} \psi_c(x_c)$$

$$\boxed{\prod_{c \in (W \cup Z)} \psi_c(x_c)} \; \boxed{\prod_{c \in (Y \cup Z)} \psi_c(x_c)}$$

$$f(W, Z)\, g(Z, Y)$$

Z separates W & Y in G

$$\Rightarrow Y \perp\!\!\!\perp W \mid Z.$$

Factorization implies Global-CI (Proof)

Available as proof of Theorem 4.1 in KF book.

Global-CI does not imply factorization (*)

(Taken from example 4.4 of KF book)
But global-CI does not imply factorization. Consider a distribution over 4 binary variables: $P(x_1, x_2, x_3, x_4)$
Let G be

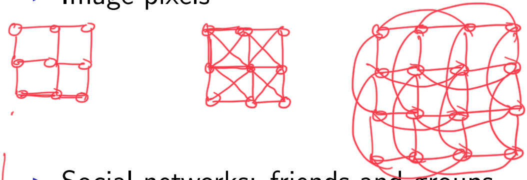
$$\begin{tikzcd} x_1 \arrow[r, dash] \arrow[d, dash] & x_2 \arrow[d, dash] \\ x_4 \arrow[r, dash] & x_3 \end{tikzcd}$$

Let $P(x_1, x_2, x_3, x_4) = 1/8$ when x_1, x_2, x_3, x_4 takes values from this set $=\{0000,1000,1100,1110,1111,0111,0011,0001\}$. In all other cases it is zero. One can painfully check that all four globals CIs in the graph: e.g. $x_1 \perp\!\!\!\perp \{x_3\}|x_2, x_4$ etc hold in the graph.
Now let us look at factorization. The factors correspond to the edges in $\psi(x_1, x_2)$. Each of the four possible assignment of each factor will get a positive value. But that cannot represent the zero probability for cases like $x_1, x_2, x_3, x_4 = 0101$.

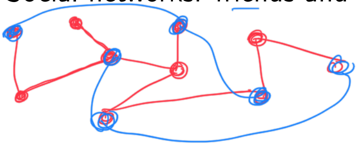
Drawing an undirected graphical model (UGM)

Two methods:

- Starting from factors: Connect together all variables that you want to ~~connect~~ put together in a factor
  - Image pixels



  - Social networks: friends and groups



  - Language models from n-gram scores $P(x_1, x_2 \ldots x_n) \propto \prod_{i=1}^{n-2} \psi(x_i, x_{i+1}, x_{i+2})$
- Starting from CIs: Simple methods do not work..

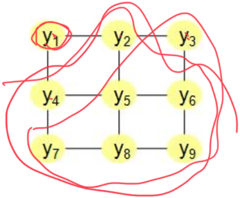
Other Conditional independencies (CIs) in an undirected graphical model (UGM)

Let $V = \{y_1, \ldots, y_n\}$.

1. Local CI: $y_i \perp\!\!\!\perp V - ne(y_i) - \{y_i\} | ne(y_i)$
2. Pairwise CI: $y_i \perp\!\!\!\perp y_j | V - \{y_i, y_j\}$ if edge (y_i, y_j) does not exist.
3. Global CI: $X \perp\!\!\!\perp Y | Z$ if Z separates X and Y in the graph.

Equivalent when the distribution $P(x)$ is positive, that is
$P(x) > 0, \quad \forall x$

1. $y_1 \perp\!\!\!\perp y_3, y_5 y_6, y_7, y_8, y_9 | y_2, y_4$ Node level CI
2. $y_1 \perp\!\!\!\perp y_3 | y_2, y_4, y_5, y_6, y_7, y_8, y_9$
3. $y_1, y_2, y_3 \perp\!\!\!\perp y_7, y_8, y_9 | y_4, y_5, y_6$



Local-CI does not imply Global-CI (*)

Let G be a undirected graph over $V = x_1, \ldots, x_n$ nodes and $P(x_1, \ldots, x_n)$ be a distribution. If P satisfies Global-CIs of G, then P will also satisfy the local-CIs of G but the reverse is not always true. We will show this with an example.

Consider a distribution over 5 binary variables: $P(x_1, \ldots, x_5)$ where $x_1 = x_2$, $x_4 = x_5$ and $x_3 = x_2\ AND\ x_4$.

Let G be

$$x_1 \text{ --- } x_2 \text{ --- } x_3 \text{ --- } x_4 \text{ --- } x_5$$

All 5 local CIs in the graph: e.g. $x_1 \perp\!\!\!\perp \{x_3, x_4, x_5\} | x_2$ etc hold in the graph.

However, the global CI: $x_2 \perp\!\!\!\perp x_4 | x_3$ does not hold.

Relationship between Local-CI and Pairwise-CI (*)

Let G be a undirected graph over $V = x_1, \ldots, x_n$ nodes and $P(x_1, \ldots, x_n)$ be a distribution. If P satisfies Local-CIs of G, then P will also satisfy the pairwise-CIs of G but the reverse is not always true. We will show this with an example.

Consider a distribution over 3 binary variables: $P(x_1, x_2, x_3)$ where $x_1 = x_2 = x_3$. That is, $P(x_1, x_2, x_3) = 1/2$ when all three are equal and 0 otherwise.

Let G be

$$x_1 \text{———} x_2 \qquad x_3$$

All 2 pairwise CIs in the graph: e.g. $x_1 \perp\!\!\!\perp \{x_3\}|x_2$ and $x_2 \perp\!\!\!\perp \{x_3\}|x_1$ hold in the graph.

DOUBT

However, the local CI: $x_1 \perp\!\!\!\perp x_3$ does not hold.

Factorization and CIs

Theorem

(Hammerseley Clifford Theorem) If a positive distribution $P(x_1, \ldots, x_n)$ confirms to the pairwise CIs of a UDGM G, then it can be factorized as per the cliques C of G as

$$P(x_1, \ldots, x_n) \propto \prod_{C \in G} \psi_C(\mathbf{y}_C)$$

Proof.

Theorem 4.8 of KF book (partially) □

Summary

Let P be a distribution and H be an undirected graph of the same set of nodes.

$\text{Factorize}(P, H) \implies \text{Global-CI}(P, H) \implies \text{Local-CI}(P, H) \implies \text{Pairwise-CI}(P, H)$

But only for positive distributions

$\text{Pairwise-CI}(P, H) \implies \text{Factorize}(P, H)$

Constructing an UGM from a positive distribution

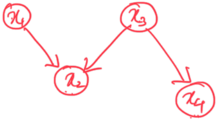
Given a positive distribution $P(x_1, \ldots, x_n)$ to which we can ask any CI of the form "Is $X \perp\!\!\!\perp Y | Z$?" and get a yes, no answer.

Goal: Draw a minimal, correct UGM G to represent P. Two options: Let V denote the set of all n variables.

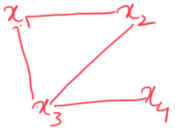
1. **Using Pairwise CI:** For each pair of vertices x_i, x_j, if $x_i \not\perp\!\!\!\perp x_j | V - \{x_i, x_j\}$ in P, add an edge between x_i and x_j in G.
2. **Using Local CI:** For each vector x_i, find the smallest subset U s.t. $x_i \perp\!\!\!\perp V - U - \{x_i\} | U$ in P. Make U the neighbors of x_i in P. S

Constructing a UGM from a positive distribution (Examples)

Hidden distribution



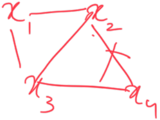
Pairwise CIs.



$x_1 \perp\!\!\!\perp x_2 \mid x_3, x_4$ No

$x_1 \perp\!\!\!\perp x_3 \mid x_2, x_4$

Local CI



$x_1 \perp\!\!\!\perp x_3, x_4 \mid x_2$? ✗

$x_1 \perp\!\!\!\perp x_3 \mid x_2, x_4$ ✗

$x_1 \perp\!\!\!\perp x_4 \mid x_2, x_3$ ✓

$x_2 \perp\!\!\!\perp x_3 \mid x_1, x_4$