Motivating Example

$X_i$ = yield in cell $i$

Bayes net: unnatural

$X_i \perp \text{non-neighbors} | \text{neighbors}$

$X_i \perp X_8 | X_5$

if $S$ separates $A$ from $B$ in $G$

$\text{sep}(A, B | S)$
Markov Properties for Undirected Graphical Model

Undirected graphical models are probability distributions that satisfy a set of conditional independence properties with respect to a dependence graph $\mathcal{G}$. Formally:

- Let $\mathcal{G} = (V, E)$ be a graph with nodes $V = \{1, \ldots, n\}$.
- For $A, B, S \subseteq V$, say that $S$ separates $A$ from $B$ if all paths from $A$ to $B$ in $\mathcal{G}$ go through $S$, written $\text{sep}_\mathcal{G}(A, B | S)$.

The joint distribution of random variables $X_1, \ldots, X_n$ satisfies the global Markov property with respect to $\mathcal{G}$ if

$$\text{sep}_\mathcal{G}(A, B | S) \implies X_A \perp X_B | X_S$$ (G)

What form of distribution $p(x_1, \ldots, x_n)$ has this property?

Warmup: Characterization of Conditional Independence

Recall the definition of conditional independence

$$p(x, y | z) = \frac{p(x,y,z)}{p(z)} \quad \text{iff} \quad p(x,y) = p(x) p(y | x)$$

for some $p_1, p_2, p_3, p_4$.

Today we’ll use two other properties of conditional independence:

1. $X \perp Y | Z \iff p(x, y, z) = p(x) p(y | x) p(z) \text{ for some } \phi_1, \phi_2$
2. $X \perp (Y, W) | Z \implies X \perp Y | Z$

Proofs: exercise

Note: (1) says that conditional independence holds iff the joint distribution factorizes in a certain way, which is very important.
Motivation

Markov Random Fields

A Markov random field is a probability distribution that factorizes over a set of "cliques" \( C \):

\[
p(x) = \frac{1}{Z} \prod_{c \in C} \phi_c(x_c), \quad Z = \sum_{x \in \mathcal{X}} \prod_{c \in C} \phi_c(x_c)
\]

- Each \( c \subseteq V = \{1, \ldots, n\} \) is a set of indices, or "clique".
- The function \( \phi_c \) is a non-negative factor or potential. It only depends on \( x_i \) for \( i \in c \).
- We say it has scope \( c \) and define \( \text{Scope}(\phi_c) := c \)
- \( Z \) is the normalizing constant or "partition function"

Dependence Graph

The dependence graph \( G = (V, E) \) of the MRF \( p(x) = \frac{1}{Z} \prod_{c \in C} \phi_c(x_c) \) is the graph where nodes \( i \) and \( j \) are connected by an edge if they appear together in some factor:

\[
V = \{1, \ldots, n\}, \quad E = \{(i, j) : i \in c \text{ and } j \in c \text{ for some } c \in C\}
\]

With this definition, every \( c \in C \) is a clique (fully connected set) in \( G \).

Concrete Example

\[
p(x_1, x_2, x_3) = \frac{1}{Z} \phi_{12}(x_1, x_2) \phi_{23}(x_2, x_3)
\]

\[
\begin{array}{ccc}
  x_1 & x_2 & x_3 \\
  0 & 0 & 1 \\
  0 & 1 & 0 \\
  1 & 0 & 2 \\
  1 & 2 & 1 \\
\end{array}
\]

\[
\phi_{12} = \begin{cases} 1 & \text{if } x_1 = x_2 \\ 0 & \text{otherwise} \end{cases}
\]

\[
\phi_{23} = \begin{cases} 1 & \text{if } x_2 = x_3 \\ 0 & \text{otherwise} \end{cases}
\]

Factorization and Markov Properties

\[
\phi_c = \text{factor or potential} \quad c = \text{Scope}(\phi_c) = \text{clique}_c
\]
Factorization

Let $G$ be a graph. A distribution $p(x)$ factorizes with respect to $G$ if

$$p(x) = \frac{1}{Z} \prod_{c \in \mathcal{C}} \phi_c(x_c), \quad \mathcal{C} = \text{cliques}(G) \quad (F)$$

In other words, it is an MRF with dependence graph $G$.

As in Bayes nets, there is a close relationship between factorization and Markov properties obtained from graph separation.

Markov Properties

The global Markov property $(G)$, the local Markov Property $(L)$ and pairwise Markov property $(P)$ are three different properties of a distribution that hold relative to a graph $G$.

$$\text{sep}_G(A, B|S) \implies X_A \perp X_B | X_S \quad (G)$$

$$i \in V \implies X_i \perp X_{V \setminus \text{nb}(i) \cup \{i\}} | X_{\text{nb}(i)} \quad (L)$$

$$(i, j) \notin E \implies X_i \perp X_j | X_{V \setminus \{i, j\}} \quad (P)$$

Above, $\text{nb}(i)$ is the set of neighbors of node $i$ in $G$.

Claim: $(G) \implies (L) \implies (P)$

It’s easy to see $(G) \implies (L)$ and $(G) \implies (P)$ by taking the appropriate choices of $A$, $B$, $S$. We leave $(L) \implies (P)$ as an exercise.

Markov Property Examples
Factorization Implies Markov

Like in Bayes nets, factorization implies conditional independencies (Markov properties).

**Claim:** \( F \Rightarrow (G) \Rightarrow (L) \Rightarrow (P) \)

**Proof** ("by example"): We only need to show \( (F) \Rightarrow (G) \).

Assume \( \text{sep}(A,B \mid S) \), more factors to "A" or "B" sides, use simple CI defn.

\[
p(x) = \frac{1}{Z} \prod_{c \in \mathcal{C}} (x_c) \prod_{x \in \mathcal{V}} f(x) g(x)
\]

\[
x_i \perp x_j, x_k \mid x_s
\]

First, remove \( S \) from \( G \). The resulting graph is disconnected and has no paths from \( A \) to \( B \).

- Let \( \bar{A} \) be the union of all connected components containing a node from \( A \)
- Let \( \bar{B} = V \setminus \bar{A} \)

Then each \( c \in \mathcal{C} \) is a subset of either \( \bar{A} \cup S \) or \( \bar{B} \cup S \)

- Let \( \mathcal{C}_A \) be the cliques contained in \( \bar{A} \cup S \)
- Let \( \mathcal{C}_B \) be the cliques contained in \( \bar{B} \cup S \)
Then
\[ p(x) = \prod_{c \in C_A} \phi_c(x_c) \prod_{c \in C_B} \phi_c(x_c) = h(x_A, x_S)k(x_B, x_S) \]
\[ \implies X_A \perp X_B | X_S \]
\[ \iff (X_A, X_{\tilde{A}}, A) \perp (X_B, X_{\tilde{B}}, B) | X_S \]
\[ \implies X_A \perp X_B | X_S \]

Markov Implies Factorization: Hammersley-Clifford Theorem

There is a famous partial converse. For a positive distribution, (P) \( \Rightarrow \) (F), which implies all the conditions are equivalent:

**Theorem (Hammersley-Clifford).** If \( p(x) > 0 \) for all \( x \), then

\[ (F) \iff (G) \iff (L) \iff (P). \]

The theorem holds for a very general class of distributions, e.g., ones with continuous, discrete, or both types of random variables.