Undirected graphical models

- A *potential function* is a non-negative function.
- We can define a joint density by a *normalized product of potential functions.* For example, we could define the BURGLARY density as follows:

  $$p_{EBAJM}(e, b, a, j, m) = \frac{1}{Z}\psi_E(e) \cdot \psi_B(b) \cdot \psi_{AEB}(a, e, b) \cdot \psi_{JA}(j, a) \cdot \psi_{MA}(m, a)$$

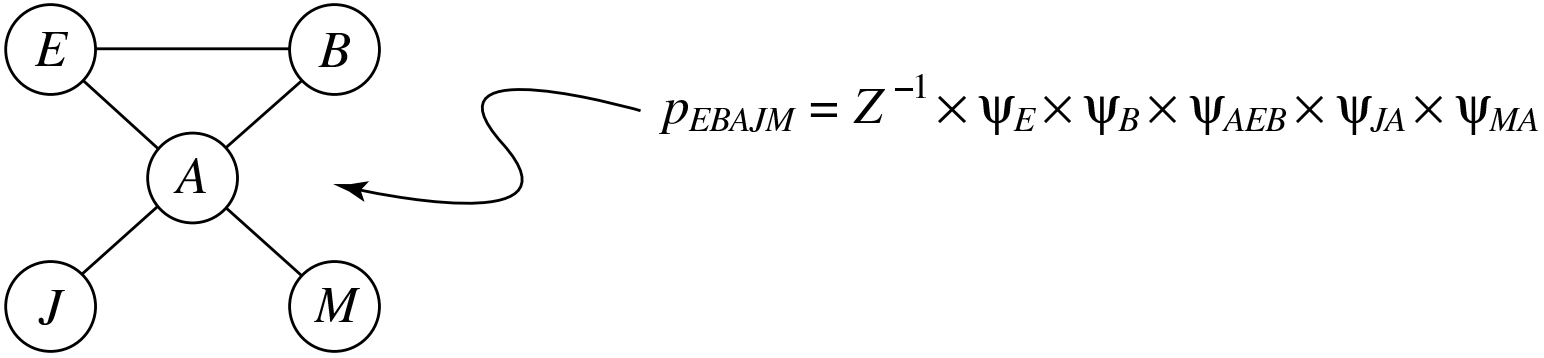
  where each ψ function is a potential and

  $$Z = \sum_e \sum_b \sum_a \sum_j \sum_m \psi_E(e) \cdot \psi_B(b) \cdot \psi_{AEB}(a, e, b) \cdot \psi_{JA}(j, a) \cdot \psi_{MA}(m, a)$$

  is the normalization constant (a.k.a. *partition function*).
- In general the potentials do not have a probabilistic interpretation, but they are interpretable: values with higher potential are more probable. The potentials trade-off with each other via the partition function.
- Multivariate Gaussians can be represented in this way.

Conditional independence in undirected graphical models

- For a density $p = \frac{1}{Z} \prod_i \psi_i$ we define an undirected graph G as follows:
  - Each variable of p becomes a node of G.
  - For each potential ψ_i we place a *clique* over its arguments in G.



- This is called an *undirected graphical model* (a.k.a. *Markov random field*).
- Then $X \perp\!\!\!\perp Y \mid \{Z_1, \ldots, Z_k\}$ if X is *separated* from Y by $Z_1, \ldots, Z_k$, i.e., if when $Z_1, \ldots, Z_k$ are removed there is no path between X and Y.

The Hammersley-Clifford Theorem

- When p is *strictly positive*, the connection between conditional independence and factorization is much stronger.
- Let G be an undirected graph over a set of random variables $\{X_1, \ldots, X_k\}$.
- Let $\mathcal{P}_1$ be the set of positive densities over $\{X_1, \ldots, X_k\}$ that are of the form

$$p = \frac{1}{Z} \prod_C \psi_C$$

  where each ψ_C is a potential over a clique of G.
- Let $\mathcal{P}_2$ be the set of positive densities with the conditional independencies encoded by graph separation in G.
- Then $\mathcal{P}_1 = \mathcal{P}_2$.

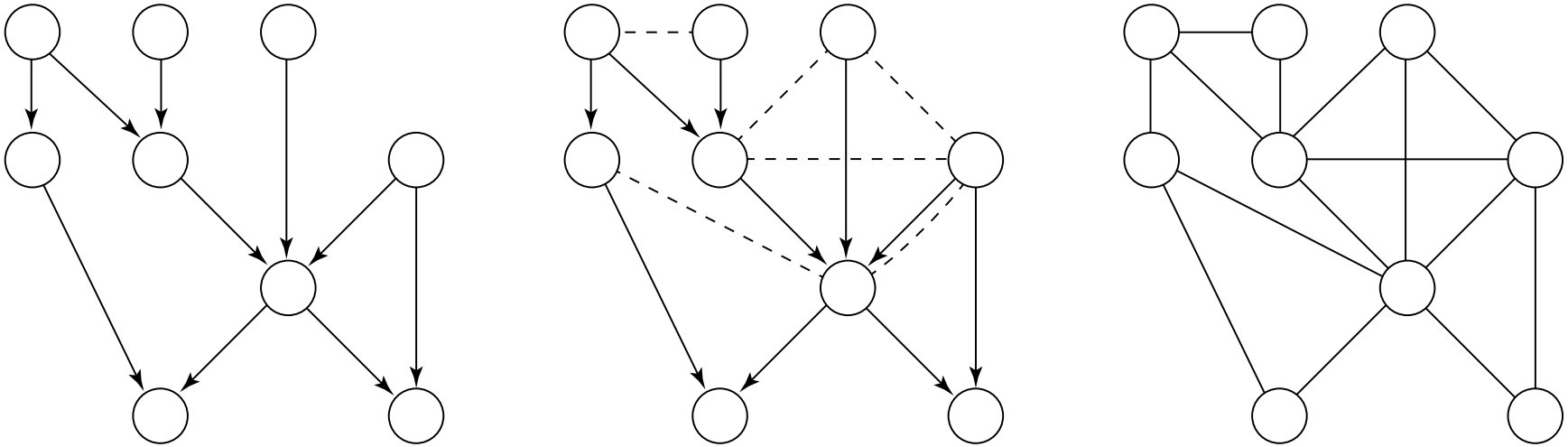
Comparing directed and undirected graphical models

- Specifying an undirected graphical model is easy (normalized product of potentials), but the factors don't have probabilistic interpretations. Specifying a directed graphical model is harder (we must choose an ordering of the variables), but the factors are marginal and conditional densities.
- Determining independence in undirected models is easy (graph separation), and in directed models it is hard (d-separation).
- Directed and undirected models are different languages: there are densities with independence properties that can be described only by directed models; the same is true for undirected models.
- In spite of this, inference in a directed model usually starts by converting it into an undirected graphical model with *fewer* conditional independencies.

Moralization

- Because the factors of a Bayesian network are marginal and conditional densities, they are also potential functions.
- Thus, a directed factorization is also an undirected factorization (with $Z = 1$). Each clique consists of a variable and its parents in the Bayes net.
- We can transform a Bayesian network into a Markov random field by placing a clique over each family of the Bayesian network and dropping the directed edges.
- This process is called *moralization* because we *marry* (or connect) the variable's parents and then drop the edge directions.

Example of moralization



All of the conditional independencies represented by the undirected model are also present in the directed model, but the reverse is not true: *the directed model has conditional independencies that are not represented by the undirected model.*