Factor Graphs

Michael Gutmann

Probabilistic Modelling and Reasoning (INFR11134)
School of Informatics, University of Edinburgh

Spring semester 2018
Recap

- Undirected and directed graphical models have complementary properties
- Both encode and (visually) represent statistical independencies (I-maps)
- Graphs tell us how probability density/mass functions factorise
- For directed graphs with parent sets $\text{pa}_i$

$$p(x_1, \ldots, x_d) = \prod_{i=1}^{d} p(x_i | \text{pa}_i)$$

- For undirected graphs with maximal clique sets $\mathcal{X}_c$

$$p(x_1, \ldots, x_d) = \frac{1}{Z} \prod_{c} \phi_c(\mathcal{X}_c)$$
1. What are factor graphs?

2. Advantages over directed or undirected graphs?
1. What are factor graphs?
   - Definition
   - Visualising Gibbs distributions
   - Visualising factors that are conditionals

2. Advantages over directed or undirected graphs?
Definition of factor graphs

- A factor graph represents the factorisation of an arbitrary function
- Example: \( h(x_1, x_2, x_3, x_4) = f_A(x_1, x_2, x_3)f_B(x_3, x_4)f_C(x_4) \)

- Two types of nodes: factor and variable nodes
- Convention: squares for factors, circles for variables (other conventions are used too)
- Edge between variable \( x \) and factor \( f \) \( \Longleftrightarrow \) \( x \) is an argument of \( f \)
- We can also use directed edges (to indicate conditionals)
Visualising Gibbs distributions as factor graphs

- Example: \( p(x_1, x_2, x_3, x_4) = \frac{1}{Z} \phi_1(x_1, x_2, x_3)\phi_2(x_3, x_4)\phi_3(x_4) \)

- General case: \( p(x_1, \ldots, x_d) \propto \prod_c \phi_c(X_c) \)
  - Factor node for all \( \phi_c \)
  - For all factors \( \phi_c \):
    - draw an undirected edge between \( \phi_c \) and all \( x_i \in X_c \).

- Can visualise any undirected graphical model as a factor graph.
More informative than undirected graphs

- Mapping from Gibbs distribution to undirected graph is many to one but one-to-one for factor graphs.
- Example

\[
p_A(x_1, x_2, x_3) \propto \phi_1(x_1, x_2)\phi_2(x_2, x_3)\phi_3(x_3, x_1) \\
p_B(x_1, x_2, x_3) \propto \phi(x_1, x_2, x_3)
\]
Visualising factors that are conditionals

▶ For \( p(x_1, x_2, x_3) = p(x_1)p(x_2)p(x_3|x_1, x_2) \), we may want to include the information that \( x_3 \) is conditioned on \( x_1, x_2 \)

▶ Use arrows as in directed graphs.

▶ Can visualise any directed graphical model as a factor graph.
Let \( p(x_1, x_2, x_3) = p(x_1, x_2)p(x_3 | x_1, x_2) \).

Visualise the conditioning for \( p(x_3 | x_1, x_2) \) but don’t impose ordering on \( x_1 \) and \( x_2 \).
1. What are factor graphs?

2. Advantages over directed or undirected graphs?
   - Computational advantages
   - Statistical advantages
Importance of factorisation

- Factorisation was central in the development so far.
- But directed and undirected graphs are not able to fully represent arbitrary factorisations of pdfs/pmfs.

For example, same graph for

\[
p(x_1, x_2, x_3) \propto \phi_1(x_1, x_2)\phi_2(x_2, x_3)\phi_3(x_3, x_1)
\]

\[
p(x_1, x_2, x_3) \propto \phi(x_1, x_2, x_3)
\]

- We should expect that being able to better represent the factorisation has advantages.
Example of computational advantages

Assume binary random variables $x_i$.

- Same undirected graph but

  \[ p(x_1, \ldots, x_d) \propto \phi(x_1, \ldots, x_d) \]  
  has $2^d$ free parameters,

  \[ p(x_1, \ldots, x_d) \propto \prod_{i<j} \phi_{ij}(x_i, x_j) \]  
  has \( \binom{d}{2} 2^2 \) free parameters.

- The difference matters for learning and inference when the number of variables is large.
Example of statistical advantages

- Let $x_1$ and $x_2$ be two inputs
- $x_1$ controls a variable $y_1$
- $x_2$ controls $y_2$
- Variables $y_1$ and $y_2$ influence each other.

- Model: $p(y_1, y_2, x_1, x_2) = p(y_1, y_2|x_1, x_2)p(x_1)p(x_2)$

- Choose $p(y_1, y_2|x_1, x_2)$ such that $p(y_1, y_2, x_1, x_2)$ satisfies
  - $x_1 \perp \perp x_2$ (independence between control variables)
  - $x_1 \perp \perp y_2 | y_1, x_2$ ($y_2$ is only influenced by $y_1$ and $x_2$)
  - $x_2 \perp \perp y_1 | y_2, x_1$ ($y_1$ is only influenced by $y_2$ and $x_1$)
Example of statistical advantages

▶ Independencies are satisfied if \( p(y_1, y_2|x_1, x_2) \) factorises as

\[
p(y_1, y_2|x_1, x_2) = p(y_1|x_1)p(y_2|x_2)\phi(y_1, y_2)n(x_1, x_2)
\]

where \( n(x_1, x_2) \) ensures normalisation of \( p(y_1, y_2|x_1, x_2) \)

\[
n(x_1, x_2) = \left( \int p(y_1|x_1)p(y_2|x_2)\phi(y_1, y_2)dy_1dy_2 \right)^{-1}
\]

(see tutorials)

▶ Directed and undirected graphs cannot represent the independencies induced by factorisation of \( p(y_1, y_2|x_1, x_2) \)

(see tutorials).

▶ Chain graphs (see Barber, Section 4.3, not covered in lecture) and factor graphs can represent them.
Example of statistical advantages

(not examinable)

- Overall model:

\[ p(y_1, y_2, x_1, x_2) = \frac{p(y_1, y_2 | x_1, x_2)}{p(y_1 | x_1) p(y_2 | x_2) \phi(y_1, y_2) n(x_1, x_2)} p(x_1) p(x_2) \]

- Factor graph (Note: directed edges to \( y_1, y_2 \) for all factors involved in the conditional)

- Independencies can be found from separation rules for factor graphs (see Barber, Section 4.4.1, and original paper “Extending Factor Graphs so as to Unify Directed and Undirected Graphical Models” by B. Frey, UAI 2003).
1. What are factor graphs?
   - Definition
   - Visualising Gibbs distributions
   - Visualising factors that are conditionals

2. Advantages over directed or undirected graphs?
   - Computational advantages
   - Statistical advantages