Lecture 2

Lecture date: Aug 31 *Scribe: Omid Etesami*

Definition 1 *Consider a finite set X, and a probability distribution over X^n such that for every $x \in X^n$,*

$$\mathbf{P}[x] = \frac{1}{Z} \prod_{a=1}^{m} \psi_a(x_{\partial a}),$$

where $\partial a \subseteq [n]$ and $\psi_a : X^{|\partial a|} \to \boldsymbol{R}_+$. Define the bipartite graph F with n variable nodes *on the left side and m* factor nodes *on the right side, where factor node a is connected to variable node i if $i \in \partial a$. We call F the* factor graph *of the probability distribution, and say that the distribution F*-factorizes.

Instead of thinking of the factor graph F as a bipartite graph, we can think of it as a hypergraph with n vertices and m hyperedges where the hyperedge corresponding to factor a consists of the set of vertices ∂a.

Definition 2 *For a factor graph F and three disjoint sets of variables A, B, and S, we say that S separates A from B in F if every path starting from $i \in A$ and ending in $i' \in B$ must intersect S.*

Recall that a path *in a factor graph is a sequence of variable nodes $i_1 = i, i_2, \ldots, i_r = i'$ where for all j it holds that $i_j, i_{j+1} \in \partial a$ for some factor node a. We say that this path starts at i and end at i'. This path intersects the set S if $i_j \in S$ for some j.*

Definition 3 *Consider a probability distribution over X^n and a hypergraph graph F on X. We say that the distribution is F*-Markov *if for all disjoint sets $A, B, S \subseteq [n]$ such that S separates A from B in F, we have*

$$\mathbf{P}[x_A, x_B | x_S] = \mathbf{P}[x_A | x_S]\mathbf{P}[x_B | x_S].$$

Theorem 4 *If a distribution F-factorizes then it is F-Markov.*

Proof: Exercise (1 point). □

The converse of the above theorem is not valid.

Proposition 5 *Consider the complete graph F with $|X| \geq 2$ and $n \geq 5$ variables nodes or $|X| \geq 10$ and $n \geq 3$. This graph consists of all factor nodes of a with $|\partial a| = 2$. Then there exist distributions that are F-Markov, but do not F-factorize.*

Proof: Let F be a complete graph on n vertices. We compare the dimension of all F-Markov distributions with the dimension of all F-factorable distributions.

Since F is complete, any distribution is F-Markov. Hence the dimension of F-Markov distributions on X^n is $|X|^n - 1$. On the other hand, for every edge a, the dimension of the set of all possible functions $\psi_a(\cdot)$ is $|X|^2 - 1$ up to the normalization of $\psi_a(\cdot)$. Thus the dimension of all F-factorable distributions is at most $(|X|^2 - 1)\binom{n}{2}$. For $|X| \geq 2$ and $n \geq 5$ we have $|X|^n - 1 > (|X|^2 - 1)\binom{n}{2}$. $\square$

Exercise 6 *Assume F is the complete k-hypergraph with n nodes and $\binom{n}{k}$ hyperedges. What's the smallest $n = n(k)$ for which F-Markov distributions are not necessarily F-factorable? (1–3 points)*

Definition 7 *A clique in a factor graph F is a collection of variable nodes $\{i_1, \ldots, i_k\}$ such that for every s, t there exists some ∂a containing both i_s and i_t.*

Definition 8 *The completion of F denoted by $\overline{F}$ is a hypergraph containing F, where for every clique $\{i_1, \ldots, i_k\} \in F$ we have added $\{i_1, \ldots, i_k\}$ as a hyperedge (factor node) in $\overline{F}$.*

Exercise 9 *Show that $\overline{\overline{F}} = \overline{F}$. (1 point)*

Theorem 10 (Clifford-Hammersley) *For every factor graph F, if a distribution on X^n is F-Markov then the distribution $\overline{F}$-factorizes, provided every element of X^n has a positive probability.*

Proof: Assume an F-Markov distribution on X^n. For every $x \in X^n$, define $g(x) = \log \mathbf{P}[x]$. (Since x has positive probability, $g(x)$ is finite.) Fix an element $\alpha \in X$, and for every $A \subseteq [n]$ define $g_A(x) = g(y)$ where $y_i = x_i$ for $i \in A$ and $y_i = \alpha$ for $i \notin A$.

Now for every $A \subseteq [n]$, define

$$\psi_A(x) = \sum_{B \subseteq A} (-1)^{|A|-|B|} g_B(x).$$

Claim 1: $\psi_A(x)$ depends only on x_A.
Claim 2: $g(x) = \sum_{A \subseteq [n]} \psi_A(x)$.
Claim 3: if A is not a clique in F, then $\psi_A(x) = 0$.

We will prove these claims shortly. Note that Claims 2 and 3 imply that

$$\mathbf{P}[x] = \exp(g(x)) = \prod_A \exp(\psi_A(x)),$$

where the product runs over all cliques A in F. So Claim 1 shows that this is indeed an $\overline{F}$-factorization of P proving the main statement of the theorem.

Claim 1 holds since by definition $\psi_A(x)$ only depends on $g_B(x)$ for $B \subseteq A$, and $g_B(x)$ only depends on x_B.
Claim 2 can be verified directly:

$$\sum_{A\subseteq[n]} \psi_A(x) = \sum_{A\subseteq[n]} \sum_{B\subseteq A} (-1)^{|A|-|B|} g_B(x) = \sum_{B\subseteq[n]} g_B(x) \sum_{A:B\subseteq A\subseteq[n]} (-1)^{|A-B|} = g_{[n]}(x) = g(x),$$

because

$$\sum_{A:B\subseteq A\subseteq[n]} (-1)^{|A-B|} = \begin{cases} 1 & \text{for B = [n];} \\ 0 & \text{otherwise.} \end{cases}$$

To prove *Claim 3*, assume $A \subseteq [n]$ is not a clique in F, that is, there exist $i, j \in A$ such that there exists no a for which $\{i, j\} \subseteq \partial a$. We write

$$\psi_A(x) = \sum_{B\subseteq A-\{i,j\}} (-1)^{|A-B|} \underbrace{(g_B(x) - g_{B\cup\{i\}}(x) - g_{B\cup\{j\}}(x) + g_{B\cup\{i,j\}}(x))}_{h_B(x)}.$$

We now show that $h_B(x) = 0$ for all $B \subseteq A - \{i, j\}$. Consider the event E that for $z \in X^n$ randomly drawn from the F-Markov distribution, we have

- $z_k = x_k$ for all $k \in B$; and
- $z_k = \alpha$ for all $k \in [n] - B - \{i, j\}$.

Since $[n] - \{i, j\}$ separates i from j, we know that under the condition that E occurs, the random variables z_i and z_j are independent. Therefore

$$\begin{aligned} \exp(h_B(x)) &= \frac{\mathbf{P}[z_i = \alpha, z_j = \alpha\ |E]\,\mathbf{P}[z_i = x_i, z_j = x_j|E]}{\mathbf{P}[z_i = \alpha, z_j = x_j|E]\,\mathbf{P}[z_i = x_i, z_j = \alpha\ |E]} \\ &= \frac{\mathbf{P}[z_i = \alpha\ |E]\,\mathbf{P}[z_j = \alpha\ |E]\,\mathbf{P}[z_i = x_i|E]\,\mathbf{P}[z_j = x_j|E]}{\mathbf{P}[z_i = \alpha\ |E]\,\mathbf{P}[z_j = x_j|E]\,\mathbf{P}[z_i = x_i|E]\,\mathbf{P}[z_j = \alpha\ |E]} \\ &= 1. \end{aligned}$$

This proves Claim 3. □

Exercise 11 *(1 point) Prove or disprove that a distribution is F-Markov iff it is $\overline{F}$-Markov.*

Exercise 12 *(1 point) Write the factor graph / factorization for the* $n \times n$ *Ising model. Is the factorization unique?*

Exercise 13 *(1 point) Write the factor graph / factorization for the SAT formula*

$$(x_1 \vee \overline{x}_2) \wedge (\overline{x}_1 \vee x_2) \wedge (x_1 \vee x_2) \wedge (\overline{x}_1 \vee \overline{x}_2).$$