1 Undirected Graphical Model (Markov Random fields)

UGMs define CI relationships via simple graph separation as follows: for sets of nodes A, B, and C, we say $x_A \perp_G x_B | x_C$ iff C separates A from B in the graph G. This means that, when we remove all the nodes in C, if there are no paths connecting any node in A to any node in B, then the CI property holds. For example, in following figure, we have that $\{1, 2\} \perp \{6, 7\} | \{3, 4, 5\}$

![UGM Diagram](image.png)

The smallest set of nodes that renders a node t conditionally independent of all the other nodes in the graph is called t’s Markov blanket; we denote this by $mb(t)$. Formally, the Markov blanket satisfies the following property:

$$t \perp V \setminus cl(t) | mb(t)$$

where $cl(t) \triangleq mb(t) \cup \{t\}$ is the closure of node t. One can show that, in a UGM, a node’s Markov blanket is its set of immediate neighbors. For example, in Figure 1, we have $mb(5) = \{2, 3, 4, 6\}$

Let’s say we have a DGM: $Z_1 \rightarrow Z_2 \rightarrow Z_3$.

$$P(Z_{1:3}) = \prod_{i=2}^3 P(Z_i | Z_{i-1})P(Z1) = (P(Z_1)P(Z_2 | Z_1))P(Z3 | Z2)$$

Let’s say we have an UGM: $Z_1 - Z_2 - Z_3$.

$$P(Z_{1:3}) = \frac{1}{Z} \Psi_{1,2}(Z_1, Z_2)\Psi_{1,3}(Z_1, Z_3)$$

$$\Psi_{1,2}(Z_1, Z_2) = \Psi_1(Z_1)\Psi_2(Z_2)$$

Now we add a hidden layer $X_1, X_2, X_3$, then we have:

$$P(Z_{1:3}|X_{1:3}) = \frac{1}{P(X_{1:3})}(P(X_1|Z_1)P(Z_1))\prod_{i=2}^3 P(Z_i | Z_{i-1})P(X_i | Z_i)$$
2 Trees

\[
P(X) = \frac{1}{Z} \prod_{(s,t) \in E} \Psi_{s,t}(x_s, x_t) \prod_{s \in V} \Psi_s(x_s)
\]

Backward-forward

\[
\alpha_t(\alpha) = \sum_K \alpha_{t-1}(K) \Psi_{t-1}(K, \alpha) \Psi_t(K) \\
\beta_{t-1}(\alpha) = \sum_K \beta_t(K) \Psi_t(K) \beta_{t-1}(\alpha, K)
\]

\[
P(X_t = K) \propto \alpha_t(K) \beta_t(K)
\]

In forward, messages send from leafs to root. In backward, messages send from root to leafs.

\[
bel_t^{(-)}(X_t) = P(X_t|\text{Data below } t) \propto \prod_{C \in \text{child}(t)} m_{C \rightarrow t}^{(-)}(X_t) \Psi_t(X_t)
\]

\[
m_{s \rightarrow t}^{(-)}(X_t) = \sum_{X_s} \Psi_{s,t}(X_s, X_t)(\prod_{C \in \text{child}(s)} m_{C \rightarrow s}^{(-)}(X_s) \Psi_s(X_s))
\]

This happen all the way to the root. Once reach the root:

\[
bel_r^{(-)}(X_r) = P(X_r|\text{Data below root}) = P(X_r|\text{Data}) \propto \Psi_r(X_r) \prod_{C \in \text{child}(r)} m_{C \rightarrow r}^{(-)}(X_r)
\]

\[
bel_s^{(-)}(X_s) = P(X_s|\text{Data}) \propto bel_s^{(-)}(X_s)(\prod_{t \in \text{pa}(s)} m_{t \rightarrow s}^{(+)}(X_s))
\]

Note:

\[
bel_s^{(-)}(X_s) \propto \Psi_s(X_s) \prod_{t \in \text{neighbors}(s)} m_{t \rightarrow s}(X_s)
\]

\[
m_{s \rightarrow t}(X_t) = \sum_{X_s} \Psi_s(X_s) \Psi_{s,t}(X_s, X_t) \prod_{u \in \text{neighbors}(s) \setminus t} m_{u \rightarrow s}(X_s)
\]

3 Markov Chain Monte Carlo (MCMC)

3.1 Monte Carlo methods

Reject on Sampling

Given target distribution \( p(x) = \frac{\tilde{p}(x)}{Z} \), proposal distribution \( q(x) \), \( Mq(x) \geq (\tilde{p})(x), x \sim q(x) \), and \( u \sim \text{unif}(0, 1) \). If \( fu \leq \frac{\tilde{p}(x)}{Mq(x)} \) we accept \( x \), else reject \( x \).

Importance Sampling

\[
I = E[f(X)] = \int \partial x f(x) p(x) = \int \partial x f(x) \frac{p(x)}{q(x)} q(x)
\]

\[
\hat{I} = \frac{1}{n} \sum_{i=1}^{n} f(x_i)
\]
\[ x_i \sim p(x) \]
\[ \hat{I}_{IS} = \sum_{i=1}^{n} f(x_i) \left( \frac{p(x_i)}{q(x_i)} \right) \]
\[ x_i \sim q(x) \]

### 3.2 MCMC

We set up a Markov Chain \( X_1, X_2, X_3, \ldots, X_t \).

\[ A_{ij} = p(X_{t+1} = j | X_t = i) \]
\[ \pi_i = p(X_1 = i) \]

**Gibbs Sampling**

Initialize \( x^{(0)} \sim q(x) \)

for iteration \( i = 1, 2, \ldots \) do

\[ x_1^{(i)} \sim p(X_1 = x_1 | X_2 = x_2^{(i-1)}, X_3 = x_3^{(i-1)}, \ldots, X_D = x_D^{(i-1)}) \]
\[ x_2^{(i)} \sim p(X_2 = x_2 | X_1 = x_1^{(i)}, X_3 = x_3^{(i-1)}, \ldots, X_D = x_D^{(i-1)}) \]

\[ \vdots \]

\[ x_D^{(i)} \sim p(X_D = x_D | X_1 = x_1^{(i)}, X_2 = x_2^{(i)}, \ldots, X_{D-1} = x_{D-1}^{(i)}) \]

end for