1 Introduction

- Independence: $x$ and $y$ are random variables (RV) that are marginally independent. Thus we have:
  $$P(x, y) = P(x)P(y)$$

- Conditional Independence: $x$, $y$, and $z$ are RVs where $x$ and $y$ are conditionally independence given $z$.
  Formally, we have: $x \perp y | z \rightarrow P(x, y | z) = P(x | z)P(y | z)$ or $P(x | y, z) = P(x | z)$

Graphical Model (GM) Visualization

Directed Graphical Model (DGM) is also known as Bayesain networks. We use graph $G(V, E)$ to represent any DGM where $V = [n] = \{1, 2, \cdots n\}$ is the set of nodes in the graph and $E = \{(s, t) : s, t \in V\}$. Given a DGM our goal is compute the joint distribution of all the variables, $P(x_1:n)$. We can assume each $x_i$ is a binary RV for the sake of simplicity, $x_i = \{0, 1\}$. Computing the joint distribution can be done by using a lookup table. Each row of this table corresponds to one configuration of the RVs. We have $2^n$ possible configuration of the RVs. Thus, the size of the table is $2^n$ that is not scalable to large values of $n$.

Goal: Have an intermediate table to show the same distribution that is scalable for large values of $n$

Topological ordering

For every graph we have the following ordering:

- $Pa(i)$: is the parent nodes of $i$-th node.
- We order each node such that the parents of each node have occurred before in the left to right ordering

We define a function, $f_i(x_i, Pa(i))$, for node $i$ and the set of parents corresponding to node $i$. The main property for this function is that it should behave as local probability function. Thus, we have:

- $f_i(x_i, Pa(i)) \geq 0$
- $\sum_{x_i} f_i(x_i, Pa(i)) = 1$
2 Example:

We consider a simple example where we have six RVs. Figure 1 illustrates one valid topological ordering of these RVs. We can compute the joint model of these six RVs as follows:

\[
p(x_{1:6}) = \prod_{i=2}^{6} p(x_i | x_{1:i-1})p(x_1) = p(x_1)p(x_2 | x_1)p(x_3 | x_1)p(x_4 | x_2)p(x_5 | x_3)p(x_6 | x_2, x_5)
\]

Given a topological ordering we can detect which RVs are independent and which RVs are conditional independent.

Main Rule:

The \(i\)-th RV is conditional independent from all the previous RVs in a topological ordering given all the parents of \(i\): \(x_i \perp x_{\text{prev}(i)} \cup \text{Pa}(i) | x_{\text{pa}(i)}\)

Example:

We can apply the above rule to the first example and we have the following

- \(x_1 \perp \emptyset | \emptyset\)
- \(x_2 \perp \emptyset | x_1\)
- \(x_3 \perp x_2 | x_1\)
- \(x_4 \perp (x_1, x_3) | x_2\)
- \(x_5 \perp (x_1, x_2, x_4) | x_3\)
- \(x_6 \perp (x_1, x_3, x_4) \cup \{x_2, x_5\}\)

3 General GM rules:

To simplify the rules, we assume that we have three nodes in a graph. We can easily extend these rules to more complicated ordering. Let \(x, y, \) and \(z\) be the RVs. Thus, we have three possible topology for these three nodes. These topologies are shown below:
<table>
<thead>
<tr>
<th>Model Name</th>
<th>Chain</th>
<th>Tent</th>
<th>Colliader (V-structure)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structure</td>
<td><img src="image1" alt="Diagram" /></td>
<td><img src="image2" alt="Diagram" /></td>
<td><img src="image3" alt="Diagram" /></td>
</tr>
<tr>
<td></td>
<td>$p(x, y, z) = p(x)p(y</td>
<td>x)p(z</td>
<td>y)$</td>
</tr>
<tr>
<td>Conditional</td>
<td>$p(x, z</td>
<td>y) = p(x</td>
<td>y)p(z</td>
</tr>
</tbody>
</table>

## 4 Bayes’s Ball Algorithm

We want to know if the following statement is true or not, $x_A \perp x_B|x_C$?

Using the rules mentioned above, we have the following algorithm that is referred to as the Bayes’s ball algorithm. We allow a ball to move from A to B if the ball reaches B then the above statement is false and if not the above statement is true.

- **Example:** In the case of two RVs that are connected by an edge. These two RVs are dependent as an edge in graphical model indicates an dependence.

- **Example:** In the case of long chain. Any two nodes are dependent as there is a path between them.

- **Example:** In the chain model, given the RV $y$ then the ball from $x$ can not reach $z$. Because, when $y$ is given, it is similar to the case that $y$ is blocked. Thus, $x$ and $z$ are independent given $y$.

## 5 Example of Graphical Structure for Existing Models

**Markov Model:**

We consider that we have a Markov Chain (MC) of size $n$. The join distribution of MC is computed as follows:

$$p(x_1, x_2, \cdots x_n) = \prod_{i=2}^{n} p(x_i|p_{i-1}) p(x_1)$$
Gaussian Generative Model:

We have $n$ RVs that are iid from $N(\mu, \sigma^2)$.

$x_i \sim N(\mu, \sigma^2)$

Bayesian Gaussian Generative Model:

In the Bayesian model, we have prior over the $\mu$ and $\sigma$.

$$x_i \sim N(\mu, \sigma^2)$$
$$\mu \sim f$$
$$\sigma \sim g$$

where $f$ and $g$ can be any prior distribution functions over the parameters.

Linear Regression Model:

We have studied linear regression in previous classes, we assume the following model:

$$y_i = x_i^T \beta + \epsilon_i$$
$$\epsilon_i \sim N(0, \sigma^2)$$
$$\beta \sim N(0, \tau)$$
Laten Variable Mixture Model:

In mixture model, we have \( k \)-component that observed that are generated.

\[
(x_i|z_i, k) \sim N(\mu_k, \Sigma_k) \\
z_i \sim \text{Model}(\pi) \\
\mu_k \sim f \\
\Sigma_k \sim g
\]

where \( f \) and \( g \) can be any prior distribution functions over the parameters.

Probabilistic PCA:

\[
z_i \sim N(0, I) \\
(x_i|z_i) = wz_i + \epsilon_i \\
\epsilon_i \sim N(0, \sigma^2)
\]
6 Application of Graphical Model to Imputation and Phasing

**Genotype Imputation:** In genotype imputation problem, we have the genotypes for certain positions of genome. However, certain SNPs (loci) genotypes are missing (not collected) for certain individuals. Typical solution to genotype imputation is to use a training dataset where the genotypes for all loci are collected for a group of individuals. Then, we use the correlation between the alleles for two adjacent SNPs. After learning this correlation, we can predict (impute) the missing genotypes.

The observed genotype:

\[
\begin{array}{cccccccc}
0 & 1 & ? & 0 & 0 & 0 & 1 & 1 \\
\end{array}
\]

where ? indicates the missing genotype. We use an existing dataset where genotypes of all individuals are collected. Thus, we assume that we have the following graphical model:

We can compute the joint distribution:

\[
p(z_{1:n}) = p(z_1) \prod_{i=2}^{n} p(z_i|z_{i-1})
\]

To predict the missing genotypes for SNP3, we use the conditional distribution:

\[
p(z_3|z_1, z_2, z_4:n)
\]

**Phasing:** Each individuals have two haplotypes. One haplotype encodes the paternal genetic information and another haplotype encodes the maternal genetic information. Given the genotypes can we reconstruct the two haplotypes.
where $z^p_i$ and $z^m_i$ are the genotypes of the $i$-th SNP for paternal and maternal, respectively. In this model, $z^p_i$ and $z^m_i$ are hidden RVs and $x_i$ are observed genotypes.