1 A binary child given a binary parent

Let X and Y be two binary variables. The log-conditional probability of the child-node X given its parent-node Y is expressed as follows:

$$\begin{aligned}\ln p(X|Y) = I(X=0)I(Y=0)\ln p_{x|y} + I(X=1)I(Y=0)\ln p_{\bar{x}|y} \\ +I(X=0)I(Y=1)\ln p_{x|\bar{y}} + I(X=1)I(Y=1)\ln p_{\bar{x}|\bar{y}}\end{aligned}$$

This conditional probability can be expressed in different exponential forms as follows:

- **First form**:

$$\begin{aligned}\ln p(X \mid Y) &= \theta^T s(X,Y) - A(\theta) \\ &= \begin{pmatrix}\ln p_{x|y} \\ \ln p_{\bar{x}|y} \\ \ln p_{x|\bar{y}} \\ \ln p_{\bar{x}|\bar{y}}\end{pmatrix}^T \begin{pmatrix} I(X=0)I(Y=0) \\ I(X=1)I(Y=0) \\ I(X=0)I(Y=1) \\ I(X=1)I(Y=1)\end{pmatrix} - 0 \\ &= \begin{pmatrix}\theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3\end{pmatrix}^T \begin{pmatrix} I(X=0)I(Y=0) \\ I(X=1)I(Y=0) \\ I(X=0)I(Y=1) \\ I(X=1)I(Y=1)\end{pmatrix} - 0\end{aligned}$$

- **Second form**:

$$\begin{aligned}\ln p(X \mid Y) &= \theta(Y)^T s(X) - A(Y) \\ &= \begin{pmatrix} I(Y=0)\ln p_{x|y} + I(Y=1)\ln p_{x|\bar{y}} \\ I(Y=0)\ln p_{\bar{x}|y} + I(Y=1)\ln p_{\bar{x}|\bar{y}}\end{pmatrix}^T \begin{pmatrix} I(X=0) \\ I(X=1)\end{pmatrix} - 0 \\ &= \begin{pmatrix} m_0^Y \cdot \theta_0 + m_1^Y \cdot \theta_2 \\ m_0^Y \cdot \theta_1 + m_1^Y \cdot \theta_3\end{pmatrix}^T \begin{pmatrix} I(X=0) \\ I(X=1)\end{pmatrix} - 0\end{aligned}$$

- **Third form**:

$$\begin{aligned}\ln p(X \mid Y) &= \theta(X)^T s(Y) - A(X) \\ &= \begin{pmatrix} I(X=0)\ln p_{x|y} + I(X=1)\ln p_{\bar{x}|y} \\ I(X=0)\ln p_{x|\bar{y}} + I(X=1)\ln p_{\bar{x}|\bar{y}}\end{pmatrix}^T \begin{pmatrix} I(Y=0) \\ I(Y=1)\end{pmatrix} - 0 \\ &= \begin{pmatrix} m_0^X \cdot \theta_0 + m_1^X \cdot \theta_1 \\ m_0^X \cdot \theta_2 + m_1^X \cdot \theta_3\end{pmatrix}^T \begin{pmatrix} I(Y=0) \\ I(Y=1)\end{pmatrix} - 0\end{aligned}$$

2 A multinomial child given a set of multinomial parents

Let X be a multinomial variable with k possible values such that $k \geq 2$, and let $\mathbf{Y} = \{Y_1, \ldots, Y_n\}$ denote the set of parents of X, such that all of them are multinomial. Each parent Y_i, $1 \geq i \geq n$, has r_i possible values or states such that $r_i \geq 2$. A parental configuration for the child-node X is then a set of n elements $\{Y_1 = y_1^v, \ldots, Y_i = y_i^v, \ldots, Y_n = y_n^v\}$ such that y_i^v denotes a potential value of variable Y_i such that $1 \leq v \leq r_i$. Let $q = r_1 \times \ldots \times r_n$ denote the total number of parental configurations, and let $\mathbf{y}^l$ denote the l^{th} parental configuration such that $1 \leq l \leq q$.

The log-conditional probability of the child-node X given its parent-nodes $\mathbf{Y}$ can be expressed as follows:

$$\ln p(X \mid \mathbf{Y}) = \sum_{j=1}^{k} \sum_{l=1}^{q} I(X = x^j) I(\mathbf{Y} = \mathbf{y}^l) \ln p_{x^j|\mathbf{y}^l}$$

Similarly the above log-conditional probability can be expressed in the following exponential forms:

- **First form**:

$$\begin{aligned}
\ln p(X \mid \mathbf{Y}) &= \theta^T s(X, \mathbf{Y}) - A(\theta) \\
&= \begin{pmatrix} \ln p_{x^1|\mathbf{y}^1} \\ \vdots \\ \ln p_{x^1|\mathbf{y}^q} \\ \vdots \\ \ln p_{x^k|\mathbf{y}^1} \\ \vdots \\ \ln p_{x^k|\mathbf{y}^q} \end{pmatrix}^T \begin{pmatrix} I(X = x^1) I(\mathbf{Y} = \mathbf{y}^1) \\ \vdots \\ I(X = x^1) I(\mathbf{Y} = \mathbf{y}^q) \\ \vdots \\ I(X = x^k) I(\mathbf{Y} = \mathbf{y}^1) \\ \vdots \\ I(X = x^k) I(\mathbf{Y} = \mathbf{y}^q) \end{pmatrix} - 0 \\
&= \begin{pmatrix} \theta_{11} \\ \vdots \\ \theta_{1q} \\ \vdots \\ \theta_{k1} \\ \vdots \\ \theta_{kq} \end{pmatrix}^T \begin{pmatrix} I(X = x^1) I(\mathbf{Y} = \mathbf{y}^1) \\ \vdots \\ I(X = x^1) I(\mathbf{Y} = \mathbf{y}^q) \\ \vdots \\ I(X = x^k) I(\mathbf{Y} = \mathbf{y}^1) \\ \vdots \\ I(X = x^k) I(\mathbf{Y} = \mathbf{y}^q) \end{pmatrix} - 0
\end{aligned}$$

- **Second form**:

$$
\begin{aligned}
\ln p(X \mid \mathbf{Y}) &= \theta(\mathbf{Y})^T s(X) - A(\mathbf{Y}) \\
&= \begin{pmatrix} I(\mathbf{Y}=\mathbf{y}^1)\ln p_{x^1|\mathbf{y}^1} + \ldots + I(\mathbf{Y}=\mathbf{y}^q)\ln p_{x^1|\mathbf{y}^q} \\ \vdots \\ I(\mathbf{Y}=\mathbf{y}^1)\ln p_{x^k|\mathbf{y}^1} + \ldots + I(\mathbf{Y}=\mathbf{y}^q)\ln p_{x^k|\mathbf{y}^q} \end{pmatrix}^T \begin{pmatrix} I(X=x^1) \\ \vdots \\ I(X=x^k) \end{pmatrix} - 0 \\
&= \begin{pmatrix} \mathbf{m}_1^{\mathbf{Y}} \cdot \theta_{11} + m_q^{\mathbf{Y}} \cdot \theta_{1q} \\ \vdots \\ \mathbf{m}_1^{\mathbf{Y}} \cdot \theta_{k1} + m_q^{\mathbf{Y}} \cdot \theta_{kq} \end{pmatrix}^T \begin{pmatrix} I(X=x^1) \\ \vdots \\ I(X=x^k) \end{pmatrix} - 0
\end{aligned}
$$

such that $\mathbf{m}_1^{\mathbf{Y}} = \prod_{i=1}^n I(Y_i = y_i^1) = \prod_{i=1}^n m_1^{Y_i}$ denotes the expected sufficient statistics for the first parental configuration, and $\mathbf{m}_q^{\mathbf{Y}} = \prod_{i=1}^n I(Y_i = y_i^{r_i}) = \prod_{i=1}^n m_{r_i}^{Y_i}$ denotes the expected sufficient statistics for the last parental configuration.

- **Third form**:

$$
\begin{aligned}
\ln p(X \mid \mathbf{Y}) &= \theta(X)^T s(\mathbf{Y}) - A(X) \\
&= \begin{pmatrix} I(X=x^1)\ln p_{x^1|\mathbf{y}^1} + \ldots + I(X=x^k)\ln p_{x^k|\mathbf{y}^1} \\ \vdots \\ I(X=x^1)\ln p_{x^1|\mathbf{y}^q} + \ldots + I(X=x^k)\ln p_{x^k|\mathbf{y}^q} \end{pmatrix}^T \begin{pmatrix} I(\mathbf{Y}=\mathbf{y}^1) \\ \vdots \\ I(\mathbf{Y}=\mathbf{y}^q) \end{pmatrix} - 0 \\
&= \begin{pmatrix} m_1^X \cdot \theta_{11} + \ldots + m_k^X \cdot \theta_{k1} \\ \vdots \\ m_1^X \cdot \theta_{1q} + \ldots + m_k^X \cdot \theta_{kq} \end{pmatrix}^T \begin{pmatrix} I(\mathbf{Y}=\mathbf{y}^1) \\ \vdots \\ I(\mathbf{Y}=\mathbf{y}^q) \end{pmatrix} - 0
\end{aligned}
$$

3 A normal child given a set of normal parents

Let X be a normal variable and $\mathbf{Y} = \{Y_1, \ldots, Y_n\}$ denote the set of parents of X, such that all of them are normal.

The log-conditional probability of X given its parents $\mathbf{Y}$ can be expressed as follows:

$$\ln p(X|Y_1, \ldots, Y_n) = \ln \left(\frac{1}{\sigma\sqrt{2(\beta_0 + \sum_i^n \beta_i Y_i)}} \mathrm{e}^{-\frac{(y-(\beta_0+\sum_i^n \beta_i Y_i))^2}{2\sigma^2}} \right)$$

Similarly the above log-conditional probability can be expressed in the following exponential forms:

- **First form**:

$$\begin{aligned}
\ln p(X \mid \mathbf{Y}) &= \theta^T s(X, \mathbf{Y}) - A(\theta) + h(\mathbf{Y}) \\
&= \begin{pmatrix}
\frac{-1}{2\sigma^2} & = & \theta_{\text{-}1} \\
\frac{-\beta_1^2}{2\sigma^2} & = & \theta_{1^2} \\
\vdots & & \\
\frac{-\beta_n^2}{2\sigma^2} & = & \theta_{n^2} \\
\frac{\beta_0}{\sigma^2} & = & \theta_0 \\
\frac{\beta_1}{\sigma^2} & = & \theta_1 \\
\vdots & & \\
\frac{\beta_n}{\sigma^2} & = & \theta_n \\
\frac{-\beta_0\beta_1}{\sigma^2} & = & \theta_{01} \\
\vdots & & \\
\frac{-\beta_0\beta_n}{\sigma^2} & = & \theta_{0n} \\
\frac{-\beta_1\beta_2}{\sigma^2} & = & \theta_{12} \\
\vdots & & \\
\frac{-\beta_1\beta_n}{\sigma^2} & = & \theta_{1n} \\
\vdots & & \\
\frac{-\beta_{n-1}\beta_n}{\sigma^2} & = & \theta_{n\text{-}1n}
\end{pmatrix}^T
\begin{pmatrix}
X^2 \\ Y^2 \\ \vdots \\ Y_n^2 \\ X \\ XY_1 \\ \vdots \\ XY_n \\ Y_1 \\ \vdots \\ Y_n \\ Y_1Y_2 \\ \vdots \\ Y_1Y_n \\ \vdots \\ Y_{n-1}Y_n
\end{pmatrix}
- \left(\frac{\beta_0^2}{2\sigma^2} + \ln\sigma \right) + \frac{1}{\ln\sqrt{2\mu_{X|Y}}}
\end{aligned}$$

- **Second form**:

$$
\begin{aligned}
\ln p(X \mid \mathbf{Y}) &= \theta(\mathbf{Y})^T s(X) - A\big(\theta(\mathbf{Y})\big) + h(\mathbf{Y}) \\
&= \begin{pmatrix} \frac{\mu_{X|Y}}{\sigma^2} \\ \frac{-1}{2\sigma^2} \end{pmatrix}^T \begin{pmatrix} X \\ X^2 \end{pmatrix} - \left(\frac{\mu_{X|Y}^2}{2\sigma^2} + \ln\sigma \right) + \ln \frac{1}{\sqrt{2\mu_{X|Y}}} \\
&= \begin{pmatrix} \theta_0 + \theta_i m_0^{Y_i} \\ \theta_{\text{-}1} \end{pmatrix}^T \begin{pmatrix} X \\ X^2 \end{pmatrix} - \left(\frac{\mu_{X|Y}^2}{2\sigma^2} + \ln\sigma \right) + \ln \frac{1}{\sqrt{2\mu_{X|Y}}}
\end{aligned}
$$

where $\mu_{X|Y} = \beta_0 + \sum_i^n \beta_i Y_i$

- **Third form**:

$$
\begin{aligned}
\ln p(X \mid \mathbf{Y}) &= \theta(X)^T s(\mathbf{Y}) - A\big(\theta(X)\big) + h(\mathbf{Y}) \\
&= \begin{pmatrix} -\frac{\beta_1^2}{2\sigma^2} \\ \cdots \\ -\frac{\beta_n^2}{2\sigma^2} \\ \frac{\beta_1(X-\beta_0)}{\sigma^2} \\ \cdots \\ \frac{\beta_n(X-\beta_0)}{\sigma^2} \\ -\frac{\beta_1\beta_2}{\sigma^2} \\ \cdots \\ -\frac{\beta_1\beta_n}{\sigma^2} \\ \cdots \\ -\frac{\beta_{n-1}\beta_n}{\sigma^2} \end{pmatrix}^T \begin{pmatrix} Y_1^2 \\ \cdots \\ Y_n^2 \\ Y_1 \\ \cdots \\ Y_n \\ Y_1Y_2 \\ \cdots \\ Y_1Y_n \\ \cdots \\ Y_{n-1}Y_n \end{pmatrix} - \left(\frac{(X-\beta_0)^2}{\sigma^2} + \ln\sigma \right) + \frac{1}{\ln\sqrt{2\mu_{X|Y}}} \\
&= \begin{pmatrix} \theta_{1^2} \\ \cdots \\ \theta_{n^2} \\ \theta_1 m_0^X + \theta_{01} \\ \cdots \\ \theta_n m_0^X + \theta_{0n} \\ \theta_{12} \\ \cdots \\ \theta_{1n} \\ \cdots \\ \theta_{n\text{-}1n} \end{pmatrix}^T \begin{pmatrix} Y_1^2 \\ \cdots \\ Y_n^2 \\ Y_1 \\ \cdots \\ Y_n \\ Y_1Y_2 \\ \cdots \\ Y_1Y_n \\ \cdots \\ Y_{n-1}Y_n \end{pmatrix} - \left(\frac{(X-\beta_0)^2}{\sigma^2} + \ln\sigma \right) + \frac{1}{\ln\sqrt{2\mu_{X|Y}}}
\end{aligned}
$$

Notations

The list below presents a summary of the used notations:

X	Child variable
k	Range of possible values of a multinomial variable X
j	Index over X values, i.e., $1 \geq j \geq k$
Y	One parent variable
$\mathbf{Y}$	Set of parent variables
n	Number of parent variables
i	Index over parent variables, i.e., $1 \geq i \geq n$
r_i	Range of possible values of a multinomial variable Y_i
q	Total number of configurations of a multinomial parent set $\mathbf{Y}$
l	Index over the possible parental configuration values, i.e., $1 \geq l \geq q$
$\mathbf{y}^l$	The l^{th} configuration of a multinomial parent set $\mathbf{Y}$
θ_{jl}	Equal to $\ln p_{x^j \mid \mathbf{y}^l}$, denoting the log-conditional probability of X in its state j given the l^{th} parent configuration
p	Probability distribution
m	Expected sufficient statistics
s	Sufficient statistics