

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:

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

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

- **First form:**

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

- **Second form:**

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

- **Third form:**

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

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 $\mathbf{Y} = \{Y_1, \dots, Y_n\}$ denote the set of parents of X , such that all of them are multinomial. Each parent Y_i , $1 \leq i \leq n$, 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 $\{Y_1 = y_1^v, \dots, Y_i = y_i^v, \dots, Y_n = y_n^v\}$ such that y_i^v denotes a potential value of variable Y_i such that $1 \leq v \leq r_i$. Let $q = r_1 \times \dots \times r_n$ 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:

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

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

- **First form:**

$$\ln p(X | \mathbf{Y}) = \theta^T s(X, \mathbf{Y}) - A(\theta)$$

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

- **Second form:**

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

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:**

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

$$\begin{aligned}
\ln p(X \mid \mathbf{Y}) &= \theta(X, \mathbf{Y}')^T s(Y_i) - A(X) \quad \text{such that } \mathbf{Y}' = \mathbf{Y} \setminus Y_i \\
&= \begin{pmatrix} m_1^X I(\mathbf{Y}' = \mathbf{y}'^1) \ln p_{x^1|\mathbf{y}'^1} + \dots + m_k^X I(\mathbf{Y}' = \mathbf{y}'^1) \ln p_{x^k|\mathbf{y}'^1} \\ \vdots \\ m_1^X I(\mathbf{Y}' = \mathbf{y}'^{q'}) \ln p_{x^1|\mathbf{y}'^{q'}} + \dots + m_k^X I(\mathbf{Y}' = \mathbf{y}'^{q'}) \ln p_{x^k|\mathbf{y}'^{q'}} \end{pmatrix}^T \begin{pmatrix} I(Y_i = y_i^1) \\ \vdots \\ I(Y_i = y_i^{r_i}) \end{pmatrix} - 0 \\
&= \begin{pmatrix} m_1^X \cdot \mathbf{m}_1^{\mathbf{Y}'} \cdot \theta'_{11} + \dots + m_k^X \cdot \mathbf{m}_1^{\mathbf{Y}'} \cdot \theta'_{k1} \\ \vdots \\ m_1^X \cdot \mathbf{m}_{q'}^{\mathbf{Y}'} \cdot \theta'_{1q'} + \dots + m_k^X \cdot \mathbf{m}_{q'}^{\mathbf{Y}'} \cdot \theta'_{kq'} \end{pmatrix}^T \begin{pmatrix} I(Y_i = y_i^1) \\ \vdots \\ I(Y_i = y_i^{r_i}) \end{pmatrix} - 0
\end{aligned}$$

where $\mathbf{m}_1^{\mathbf{Y}'} = I(\mathbf{Y}' = \mathbf{y}'^1) = I(Y_1 = y_1^1) \cdot \dots \cdot I(Y_{i-1} = y_{i-1}^1) \cdot I(Y_{i+1} = y_{i+1}^1) \cdot \dots \cdot I(Y_n = y_n^1)$ denotes the expected sufficient statistics for the first configuration of the parent set $\mathbf{Y}' = \mathbf{Y} \setminus Y_i$, and $\mathbf{m}_{q'}^{\mathbf{Y}'} = I(\mathbf{Y}' = \mathbf{y}'^{q'}) = I(Y_1 = y_1^{q'}) \cdot \dots \cdot I(Y_{i-1} = y_{i-1}^{q'}) \cdot I(Y_{i+1} = y_{i+1}^{q'}) \cdot \dots \cdot I(Y_n = y_n^{q'})$ denotes the expected sufficient statistics for the last configuration of the parent set \mathbf{Y}' , with $q' = q/r_i$ denotes the total number of configurations of the parent set \mathbf{Y}' .

3 A normal child given a set of normal parents

Let X be a normal variable and $\mathbf{Y} = \{Y_1, \dots, Y_n\}$ 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:

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

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

- **First form - Joint suff. stat. (Maxim. Likelihood):**

$$\ln p(X | \mathbf{Y}) = \theta^T s(X, \mathbf{Y}) - A(\theta) + h(\mathbf{Y})$$

$$= \begin{pmatrix} \frac{-1}{2\sigma^2} & = & \theta_{-1} \\ \frac{-\beta_1^2}{2\sigma^2} & = & \theta_{1^2} \\ \vdots & & \\ \frac{-\beta_n^2}{2\sigma^2} & = & \theta_{n^2} \\ \frac{\beta_0}{\sigma^2} & = & \theta_0 \\ \frac{\beta_1}{\sigma^2} & = & \theta_1 \\ \vdots & & \\ \frac{\beta_n}{\sigma^2} & = & \theta_n \\ \frac{-\beta_0\beta_1}{\sigma^2} & = & \theta_{01} \\ \vdots & & \\ \frac{-\beta_0\beta_n}{\sigma^2} & = & \theta_{0n} \\ \frac{-\beta_1\beta_2}{\sigma^2} & = & \theta_{12} \\ \vdots & & \\ \frac{-\beta_1\beta_n}{\sigma^2} & = & \theta_{1n} \\ \vdots & & \\ \frac{-\beta_{n-1}\beta_n}{\sigma^2} & = & \theta_{n-1n} \end{pmatrix}^T \begin{pmatrix} X^2 & = & m_{X^2} \\ Y_1^2 & = & m_{Y_1^2} \\ \vdots & & \\ Y_n^2 & = & m_{Y_n^2} \\ X & = & m_X \\ XY_1 & = & m_{XY_1} \\ \vdots & & \\ XY_n & = & m_{XY_n} \\ Y_1 & = & m_{Y_1} \\ \vdots & & \\ Y_n & = & m_{Y_n} \\ Y_1 Y_2 & = & m_{Y_1 Y_2} \\ \vdots & & \\ Y_1 Y_n & = & m_{Y_1 Y_n} \\ \vdots & & \\ Y_{n-1} Y_n & = & m_{Y_{n-1} Y_n} \end{pmatrix} - \left(\frac{\beta_0^2}{2\sigma^2} + \ln \sigma \right) + \frac{1}{\ln \sqrt{2\mu_{X|Y}}}$$

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

- From moment to natural parameters:

* FIRST STEP:

$$\begin{aligned}
\mu_X &= m_X \\
\mu_Y &= m_Y \\
\Sigma_{XX} &= m_{X^2} - m_X^2 \\
\Sigma_{YY} &= m_{Y^2} - m_Y^2 \\
\Sigma_{XY} &= m_{XY} - m_X m_Y \\
\Sigma_{YX} &= m_{YX} - m_Y m_X
\end{aligned}$$

* SECOND STEP:

$$\begin{aligned}
\beta_0 &= \mu_X - \Sigma_{XY} \Sigma_{YY}^{-1} \mu_Y \\
&= m_X + \frac{m_X m_Y^2 - m_{XY} m_Y}{m_{Y^2} - m_Y^2} \\
\beta &= \Sigma_{XY} \Sigma_{YY}^{-1} \\
&= \frac{m_{XY} - m_X m_Y}{m_{Y^2} - m_Y^2} \\
\sigma^2 &= \Sigma_{XX} - \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{YX} \\
&= m_{X^2} - m_X^2 + \frac{m_Y m_X (m_{XY} + m_{YX} - m_X m_Y) - m_{XY} m_{YX}}{m_{Y^2} - m_Y^2}
\end{aligned}$$

• Second form:

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

• Third form:

$$\ln p(X \mid \mathbf{Y}) = \theta(X)^T s(\mathbf{Y}) - A(\theta(X)) + h(\mathbf{Y})$$

$$\begin{aligned}
&= \begin{pmatrix} -\frac{\beta_1^2}{2\sigma^2} \\ \dots \\ -\frac{\beta_n^2}{2\sigma^2} \\ \frac{\beta_1(X-\beta_0)}{\sigma^2} \\ \dots \\ \frac{\beta_n(X-\beta_0)}{\sigma^2} \\ -\frac{\beta_1\beta_2}{\sigma^2} \\ \dots \\ -\frac{\beta_1\beta_n}{\sigma^2} \\ \dots \\ -\frac{\beta_{n-1}\beta_n}{\sigma^2} \end{pmatrix}^T \begin{pmatrix} Y_1^2 \\ \dots \\ Y_n^2 \\ Y_1 \\ \dots \\ Y_n \\ Y_1Y_2 \\ \dots \\ Y_1Y_n \\ \dots \\ Y_{n-1}Y_n \end{pmatrix} - \left(\frac{(X-\beta_0)^2}{\sigma^2} + \ln \sigma \right) + \frac{1}{\ln \sqrt{2\mu_{X|Y}}} \\
&= \begin{pmatrix} \theta_{1^2} \\ \dots \\ \theta_{n^2} \\ \theta_1 m^X + \theta_{01} \\ \dots \\ \theta_n m^X + \theta_{0n} \\ \theta_{12} \\ \dots \\ \theta_{1n} \\ \dots \\ \theta_{n-1n} \end{pmatrix}^T \begin{pmatrix} Y_1^2 \\ \dots \\ Y_n^2 \\ Y_1 \\ \dots \\ Y_n \\ Y_1Y_2 \\ \dots \\ Y_1Y_n \\ \dots \\ Y_{n-1}Y_n \end{pmatrix} - \left(\frac{X^2 - 2X\beta_0 + \beta_0^2}{\sigma^2} + \ln \sigma \right) + \frac{1}{\ln \sqrt{2\mu_{X|Y}}} \\
&= \begin{pmatrix} \theta_{1^2} \\ \dots \\ \theta_{n^2} \\ \theta_1 m^X + \theta_{01} \\ \dots \\ \theta_n m^X + \theta_{0n} \\ \theta_{12} \\ \dots \\ \theta_{1n} \\ \dots \\ \theta_{n-1n} \end{pmatrix}^T \begin{pmatrix} Y_1^2 \\ \dots \\ Y_n^2 \\ Y_1 \\ \dots \\ Y_n \\ Y_1Y_2 \\ \dots \\ Y_1Y_n \\ \dots \\ Y_{n-1}Y_n \end{pmatrix} - \left((-2\beta_{-1}m^{X^2} - 2m^X\beta_0 - \frac{1}{2}\beta_0^2\beta_{-1}^{-1}) + \frac{\ln(2\theta_{-1})}{2} \right) \\
&+ \ln \frac{1}{\sqrt{2(\theta_0 + \sum_i^n \theta_i m^{Y_i})}}
\end{aligned}$$

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 $\ln p_{x^j \mathbf{y}^l}$, 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