

# 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:

$$\begin{aligned}\ln p(X | Y) &= I(X = x^1)I(Y = y^1) \ln p_{x^1|y^1} + I(X = x^2)I(Y = y^1) \ln p_{x^2|y^1} \\ &\quad + I(X = x^1)I(Y = y^2) \ln p_{x^1|y^2} + I(X = x^2)I(Y = y^2) \ln p_{x^2|y^2}\end{aligned}$$

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

- **First form:**

$$\begin{aligned}\ln p(X | Y) &= \theta^T s(X, Y) - A(\theta) \\ &= \begin{pmatrix} \ln p_{x^1|y^1} \\ \ln p_{x^2|y^1} \\ \ln p_{x^1|y^2} \\ \ln p_{x^2|y^2} \end{pmatrix}^T \begin{pmatrix} I(X = x^1)I(Y = y^1) \\ I(X = x^2)I(Y = y^1) \\ I(X = x^1)I(Y = y^2) \\ I(X = x^2)I(Y = y^2) \end{pmatrix} - 0 \\ &= \begin{pmatrix} \theta_{11} \\ \theta_{21} \\ \theta_{12} \\ \theta_{22} \end{pmatrix}^T \begin{pmatrix} I(X = x^1)I(Y = y^1) \\ I(X = x^2)I(Y = y^1) \\ I(X = x^1)I(Y = y^2) \\ I(X = x^2)I(Y = y^2) \end{pmatrix} - 0\end{aligned}$$

- **Second form:**

$$\begin{aligned}\ln p(X | Y) &= \theta(Y)^T s(X) - A(Y) \\ &= \begin{pmatrix} I(Y = y^1) \ln p_{x^1|y^1} + I(Y = y^2) \ln p_{x^1|y^2} \\ I(Y = y^1) \ln p_{x^2|y^1} + I(Y = y^2) \ln p_{x^2|y^2} \end{pmatrix}^T \begin{pmatrix} I(X = x^1) \\ I(X = x^2) \end{pmatrix} - 0 \\ &= \begin{pmatrix} m_1^Y \cdot \theta_{11} + m_2^Y \cdot \theta_{12} \\ m_1^Y \cdot \theta_{21} + m_2^Y \cdot \theta_{22} \end{pmatrix}^T \begin{pmatrix} I(X = x^1) \\ I(X = x^2) \end{pmatrix} - 0\end{aligned}$$

- **Third form:**

$$\begin{aligned}\ln p(X | Y) &= \theta(X)^T s(Y) - A(X) \\ &= \begin{pmatrix} I(X = x^1) \ln p_{x^1|y^1} + I(X = x^2) \ln p_{x^2|y^1} \\ I(X = x^1) \ln p_{x^1|y^2} + I(X = x^2) \ln p_{x^2|y^2} \end{pmatrix}^T \begin{pmatrix} I(Y = y^1) \\ I(Y = y^2) \end{pmatrix} - 0 \\ &= \begin{pmatrix} m_1^X \cdot \theta_{11} + m_2^X \cdot \theta_{21} \\ m_1^X \cdot \theta_{12} + m_2^X \cdot \theta_{22} \end{pmatrix}^T \begin{pmatrix} I(Y = y^1) \\ I(Y = y^2) \end{pmatrix} - 0\end{aligned}$$

## 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):**

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

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

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

- **From moment to natural parameters: (matrix representation)**

$$\begin{aligned}
\ln p(X \mid \mathbf{Y}) &= \theta^T s(X, \mathbf{Y}) - A(\theta) + h(\mathbf{Y}) \\
&= \begin{pmatrix} -1(2\sigma^2)^{-1} & = & \theta_{-1} \\ \beta'(\sigma^2)^{-1} & = & \theta_\beta \\ -\beta'\beta^T(2\sigma^2)^{-1} & = & \theta_{\beta\beta^T} \end{pmatrix}^T \begin{pmatrix} XX^T & = & \mathbb{E}(XX^T) \\ Y'X^T & = & \mathbb{E}(YX) \\ Y'Y^T & = & \mathbb{E}(YY^T) \end{pmatrix} \\
&- \left( \frac{\beta_0^2}{2\sigma^2} + \ln \sigma \right) + \frac{1}{\ln \sqrt{2\mu_{X|Y}}}
\end{aligned}$$

where

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_r \end{pmatrix} \quad XX^T = \begin{pmatrix} X_1 \\ \vdots \\ X_r \end{pmatrix} (X_1 \quad \cdots \quad X_r) = \begin{pmatrix} X_1X_1 & \cdots & X_1X_r \\ X_2X_1 & \cdots & X_2X_r \\ \vdots & & \vdots \\ X_rX_1 & \cdots & X_rX_r \end{pmatrix}$$

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \quad Y' = \begin{pmatrix} 1 \\ Y_1 \\ \vdots \\ Y_n \end{pmatrix} \quad Y'Y^T = \begin{pmatrix} Y_1 & \cdots & Y_n \\ Y_1Y_1 & \cdots & Y_1Y_n \\ Y_2Y_1 & \cdots & Y_2Y_n \\ \vdots & & \vdots \\ Y_nY_1 & \cdots & Y_nY_n \end{pmatrix}$$

$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} \quad \beta' = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} \quad \beta'\beta^T = \begin{pmatrix} \beta_0\beta_1 & \cdots & \beta_0\beta_n \\ \beta_1\beta_1 & \cdots & \beta_1\beta_n \\ \vdots & & \vdots \\ \beta_n\beta_1 & \cdots & \beta_n\beta_n \end{pmatrix}$$

$$Y'X^T = \begin{pmatrix} 1 \\ Y_1 \\ \vdots \\ Y_n \end{pmatrix} (X_1 \quad \cdots \quad X_r) = \begin{pmatrix} X_1 & X_2 & \cdots & X_r \\ Y_1X_1 & Y_1X_2 & \cdots & Y_1X_r \\ \vdots & & & \vdots \\ Y_nX_1 & Y_nX_2 & \cdots & Y_nX_r \end{pmatrix}$$

\* FIRST STEP:

$$\begin{aligned}
\mu_X &= E(X) = E(Y^T X)_{[1,r]} \\
\mu_Y &= E(Y) = E(Y Y^T)_{[1,n]} \\
\Sigma_{XX} &= E(X X^T) - E(X)^2 = E(X X^T) - E(Y X)_{[1,r]}^T E(Y X)_{[1,r]} \\
\Sigma_{YY} &= E(Y Y^T) - E(Y)^2 = E(Y Y^T) - E(Y Y^T)_{[n,1]} E(Y Y^T)_{[1,n]} \\
\Sigma_{XY} &= E(Y X)^T - E(X) E(Y) = E(Y X)^T - E(Y X)_{[1,r]} E(Y Y^T)_{[n,1]} \\
\Sigma_{YX} &= E(Y X) - E(Y) E(X) = E(Y X) - E(Y Y^T)_{[n,1]} E(Y X)_{[1,r]}
\end{aligned}$$

\* SECOND STEP (Theorem 7.4 in page 253, Koller & Friedman):

$$\begin{aligned}
\beta_0 &= \mu_X - \Sigma_{XY} \Sigma_{YY}^{-1} \mu_Y \\
\beta &= \Sigma_{XY} \Sigma_{YY}^{-1} \\
\sigma^2 &= \Sigma_{XX} - \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{YX}
\end{aligned}$$

All natural parameters  $\theta$  can now be calculated considering these equations.

– **From natural to moment parameters:** Via inference.

• **Second form:**

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

• **Third form:**

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

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



## 4 A base distribution given a binary parent

Let  $X$  be any base distribution variable, and let  $Y$  be a binary variable. The log-conditional probability of the child-node  $X$  given its binary parent-node  $Y$  is expressed as follows:

$$\begin{aligned}\ln p(X | Y) &= I(Y = y^1) \ln p_{X|y^1} + I(Y = y^2) \ln p_{X|y^2} \\ &= I(Y = y^1) \left( \theta_{X1} \cdot s(X) - A(\theta_{X1}) \right) + I(Y = y^2) \left( \theta_{X2} \cdot s(X) - A(\theta_{X2}) \right) \\ &= I(Y = y^1) \cdot \theta_{X1} \cdot s(X) - I(Y = y^1) \cdot A(\theta_{X1}) + I(Y = y^2) \cdot \theta_{X2} \cdot s(X) - I(Y = y^2) \cdot A(\theta_{X2})\end{aligned}$$

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

- **First form:**

$$\begin{aligned}\ln p(X | Y) &= \theta^T s(X, Y) - A(\theta) \\ &= \begin{pmatrix} -A(\theta_{X1}) \\ -A(\theta_{X2}) \\ \theta_{X1} \\ \theta_{X2} \end{pmatrix}^T \begin{pmatrix} I(Y = y^1) \\ I(Y = y^2) \\ s(X) \cdot I(Y = y^1) \\ s(X) \cdot I(Y = y^2) \end{pmatrix} - 0\end{aligned}$$

- **Second form:**

$$\begin{aligned}\ln p(X | Y) &= \theta(Y)^T s(X) - A(Y) \\ &= \begin{pmatrix} I(Y = y^1) \\ I(Y = y^2) \\ I(Y = y^1) \cdot \theta_{X1} \\ I(Y = y^2) \cdot \theta_{X2} \end{pmatrix}^T \begin{pmatrix} -A(\theta_{X1}) \\ -A(\theta_{X2}) \\ s(X) \\ s(X) \end{pmatrix} - 0 \\ &= \begin{pmatrix} m_1^Y \\ m_2^Y \\ m_1^Y \cdot \theta_{X1} \\ m_2^Y \cdot \theta_{X2} \end{pmatrix}^T \begin{pmatrix} -A(\theta_{X1}) \\ -A(\theta_{X2}) \\ s(X) \\ s(X) \end{pmatrix} - 0\end{aligned}$$

- **Third form:**

$$\begin{aligned}\ln p(X | Y) &= \theta(X)^T s(Y) - A(X) \\ &= \begin{pmatrix} -A(\theta_{X1}) \\ -A(\theta_{X2}) \\ s(X) \cdot \theta_{X1} \\ s(X) \cdot \theta_{X2} \end{pmatrix}^T \begin{pmatrix} I(Y = y^1) \\ I(Y = y^2) \\ I(Y = y^1) \\ I(Y = y^2) \end{pmatrix} - 0\end{aligned}$$

## 5 A base distribution given a set of multinomial parents

Let  $X$  be any base distribution, 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:

$$\begin{aligned} \ln p(X | \mathbf{Y}) &= \sum_{l=1}^q I(\mathbf{Y} = \mathbf{y}^l) \cdot \ln p_{X|\mathbf{y}^l} \\ &= \sum_{l=1}^q I(\mathbf{Y} = \mathbf{y}^l) \cdot \left( \theta_{Xl} \cdot s(X) \cdot A(\theta_{Xl}) \right) \\ &= \sum_{l=1}^q I(\mathbf{Y} = \mathbf{y}^l) \cdot \theta_{Xl} \cdot s(X) - I(\mathbf{Y} = \mathbf{y}^l) \cdot A(\theta_{Xl}) \end{aligned}$$

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

- **First form:**

$$\begin{aligned} \ln p(X | \mathbf{Y}) &= \theta^T s(X, \mathbf{Y}) - A(\theta) \\ &= \begin{pmatrix} -A(\theta_{X1}) \\ \vdots \\ -A(\theta_{Xq}) \\ \theta_{X1} \\ \vdots \\ \theta_{Xq} \end{pmatrix}^T \begin{pmatrix} I(\mathbf{Y} = \mathbf{y}^1) \\ \vdots \\ I(\mathbf{Y} = \mathbf{y}^q) \\ s(X) \cdot I(\mathbf{Y} = \mathbf{y}^1) \\ \vdots \\ s(X) \cdot I(\mathbf{Y} = \mathbf{y}^q) \end{pmatrix} - 0 \end{aligned}$$

- **Second form:**

$$\begin{aligned}
\ln p(X \mid \mathbf{Y}) &= \theta(\mathbf{Y})^T s(X) - A(\mathbf{Y}) \\
&= \begin{pmatrix} I(\mathbf{Y} = \mathbf{y}^1) \\ \vdots \\ I(\mathbf{Y} = \mathbf{y}^q) \\ I(\mathbf{Y} = \mathbf{y}^1) \cdot \theta_{X1} \\ \vdots \\ I(\mathbf{Y} = \mathbf{y}^q) \cdot \theta_{Xq} \end{pmatrix}^T \begin{pmatrix} -A(\theta_{X1}) \\ \vdots \\ -A(\theta_{Xq}) \\ s(X) \\ \vdots \\ s(X) \end{pmatrix} - 0 \\
&= \begin{pmatrix} \mathbf{m}_1^{\mathbf{Y}} \\ \vdots \\ \mathbf{m}_q^{\mathbf{Y}} \\ \mathbf{m}_1^{\mathbf{Y}} \cdot \theta_{X1} \\ \vdots \\ \mathbf{m}_q^{\mathbf{Y}} \cdot \theta_{Xq} \end{pmatrix}^T \begin{pmatrix} -A(\theta_{X1}) \\ \vdots \\ -A(\theta_{Xq}) \\ s(X) \\ \vdots \\ s(X) \end{pmatrix} - 0
\end{aligned}$$

• **Third form:**

$$\begin{aligned}
\ln p(X \mid \mathbf{Y}) &= \theta(X)^T s(\mathbf{Y}) - A(X) \\
&= \begin{pmatrix} -A(\theta_{X1}) \\ \vdots \\ -A(\theta_{Xq}) \\ s(X) \cdot \theta_{X1} \\ \vdots \\ s(X) \cdot \theta_{Xq} \end{pmatrix}^T \begin{pmatrix} I(\mathbf{Y} = \mathbf{y}^1) \\ \vdots \\ I(\mathbf{Y} = \mathbf{y}^q) \\ I(\mathbf{Y} = \mathbf{y}^1) \\ \vdots \\ I(\mathbf{Y} = \mathbf{y}^q) \end{pmatrix} - 0
\end{aligned}$$

$$\ln p(X \mid \mathbf{Y}) = \theta(X, \mathbf{Y}')^T s(Y_i) - A(X) \text{ such that } \mathbf{Y}' = \mathbf{Y} \setminus Y_i$$

$$= \begin{pmatrix} -A(\theta_{X1}) \\ \vdots \\ -A(\theta_{Xq}) \\ s(X) \cdot \mathbf{m}_1^{\mathbf{Y}'} \cdot \theta'_{X1} + \dots + s(X) \cdot \mathbf{m}_1^{\mathbf{Y}'} \cdot \theta'_{X1} \\ \vdots \\ s(X) \cdot \mathbf{m}_{q'}^{\mathbf{Y}'} \cdot \theta'_{Xq'} + \dots + s(X) \cdot \mathbf{m}_{q'}^{\mathbf{Y}'} \cdot \theta'_{Xq'} \end{pmatrix}^T \begin{pmatrix} I(Y_i = y_i^1) \\ \vdots \\ I(Y_i = y_i^{r_i}) \\ I(Y_i = y_i^1) \\ \vdots \\ I(Y_i = y_i^{r_i}) \end{pmatrix} - 0$$

## 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., $1 \leq j \leq k$
$Y$	One parent variable
$\mathbf{Y}$	Set of parent variables
$n$	Number of parent variables
$i$	Index over parent variables, i.e., $1 \leq i \leq n$
$r_i$	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., $1 \leq l \leq q$
$\mathbf{y}^l$	The $l^{th}$ configuration of a multinomial parent set $\mathbf{Y}$
$\theta_{jl}$	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
$\theta_{Xl}$	Equal to $\ln p_{X \mathbf{y}^l}$ , denoting the log-conditional probability of a base distribution variable $X$ given the $l^{th}$ parent configuration
$p$	Probability distribution
$m$	Expected sufficient statistics
$s$	Sufficient statistics