$$\textcircled{A} \longrightarrow \textcircled{B}$$

$$\theta = \begin{cases} \theta^A_a, & a \in Val(A) \\ \theta^B_{b|a}, & a \in Val(A),\ b \in Val(B) \end{cases}$$

$p(a,b) = p(a)p(b|a)$

$\log p(a,b) = \log p(a) + \log p(b|a)$

Data

$x^{(n)} = (a^{(n)}, b^{(n)})$

$$\log p_\theta(\overset{a}{1}, \overset{b}{0}) = \log \theta^A_1 + \log \theta^B_{0|1}$$

$$= \sum_{a \in Val(A)} \mathbb{I}[a=1] \log \theta^A_a$$

$$+ \sum_a \sum_b \mathbb{I}[a=1, b=0] \log \theta^B_{b|a}$$

$$\log p_\theta(a^{(n)}, b^{(n)}) = \sum_{a \in Val(A)} \mathbb{I}[a=a^{(n)}] \log \theta^A_a + \sum_a \sum_b \mathbb{I}[a=a^{(n)}, b=b^{(n)}] \log \theta^B_{b|a}$$

$$\mathcal{L}(\theta | x^{(1:N)}) = \frac{1}{N} \sum_n \log p_\theta(a^{(n)}, b^{(n)})$$

$$= \frac{1}{N} \sum_n \left( \ldots \text{RHS} \ldots \right)$$

$$= \sum_a \frac{\#(A=a)}{N} \log \theta^A_a + \sum_a \sum_b \frac{\#(A=a, B=b)}{N} \log \theta^B_{b|a}$$

$$\max_{\theta^A_\cdot, \theta^B_{\cdot|\cdot}} \mathcal{L}(\theta | x^{(1:N)}) = \max_{\theta^A_\cdot, \theta^B_{\cdot|\cdot}} \left( \ldots \text{RHS} \ldots \right)$$

$$= \max_{\theta^A_\cdot} \sum_a \frac{\#(A=a)}{N} \log \theta^A_a$$

— multinomial estimation

constraint: sum to one

$$+ \sum_a \max_{\theta^B_{\cdot|a}} \sum_b \frac{\#(A=a, B=b)}{N} \log \theta^B_{b|a}$$

Result:

$$\hat{\theta}^A_a = \frac{\#(A=a)}{N}$$

$$\hat{\theta}^B_{b|a} = \frac{\#(A=a, B=b)}{\#(A=a)}$$