① Hidden Markov models (HMM)

$a_{ij} = p(x_{t+1}=j \mid x_t=i)$ $\quad$ A

Markov chains

$p(x_1 = i) = \pi_i$

hidden state: $f \to a \to n' \to e\!: \to m \to ə$

observable

NER – named entity recognition

B-DATE → DATE → DATE → DATE → B-PER → PER → BG → B-EVENT → EVENT → BG

11 марта 843 г. импер-ца Феодора учредила „Торжество Православия" .

DATE $\quad$ PERSON $\quad$ BG $\quad$ EVENT

$X_t \in \{1,\dots,n\}$ $\qquad$ $y_t \in \{1,\dots,m\}$

HMM

$$\begin{tikzcd} \pi \arrow[r] & x_1 \arrow[r] \arrow[d] & x_2 \arrow[r] \arrow[d] & \dots \arrow[r] & x_t \arrow[r] \arrow[d] & x_{t+1} \arrow[r] \arrow[d] & \dots \arrow[r] & x_T \arrow[d] \\ & y_1 & y_2 & \dots & y_t & y_{t+1} & & y_T \end{tikzcd}$$

$X = x_1 - x_T$

$Y = y_1 - y_T$

$$p(X,Y) = p(x_1)\,p(y_1|x_1)\,p(x_2|x_1)\,p(y_2|x_2)\dots p(y_t|x_t)\,p(x_{t+1}|x_t)\dots p(y_T|x_T)$$

$$p(y_t \mid X, Y_{-t}) = p(y_t \mid x_t)$$
$$p(x_{t+1} \mid x_1 \dots x_t) = p(x_{t+1} \mid x_t)$$

предположения

$$p(Y,X) = p(x_1)\,p(y_1|x_1)\,p(x_2|x_1,\cancel{y_1})\,p(y_2|\cancel{x_1},x_2,\cancel{y_1})\,p(x_3|x_1,x_2,y_1,y_2) \dots$$
$$\dots\ p(y_T \mid X, Y_{-T})$$

π : $\quad \pi_i = p(x_1 = i)$ $\qquad$ $n \times 1$

A : $\quad a_{ij} = p(x_{t+1}=j \mid x_t = i)$ $\qquad$ $n \times n$

B : $\quad b_i(k) = p(y_t = k \mid x_t = i)$ $\qquad$ $n \times m$

$\theta = (\pi, A, B)$

знач. $y_1 - y_T$

$\mathcal{D} = \{D = d_1 \dots d_T\}$

знач. $x_1 - x_T$

$Q = q_1 - q_T$

$$p(D,Q|\bar\theta) = \pi_{q_1} b_{q_1}(d_1)\, a_{q_1 q_2} b_{q_2}(d_2) \cdots a_{q_{T-1} q_T} b_{q_T}(d_T)$$

$$\mathcal{D} = \{(d_1^{(s)} \dots d_T^{(s)})\}_{s=1}^{S}$$

$$p(\mathcal{D}|\bar\theta) = \prod_{s=1}^{S} p(D^{(s)}|\bar\theta) = \prod_{s=1}^{S}\left(\sum_{Q} p(D^{(s)},Q|\bar\theta)\right) \xrightarrow{\bar\theta} \max$$

auxiliary 1) $p(D|\bar\theta) = \sum_Q p(D,Q|\bar\theta) = ?$

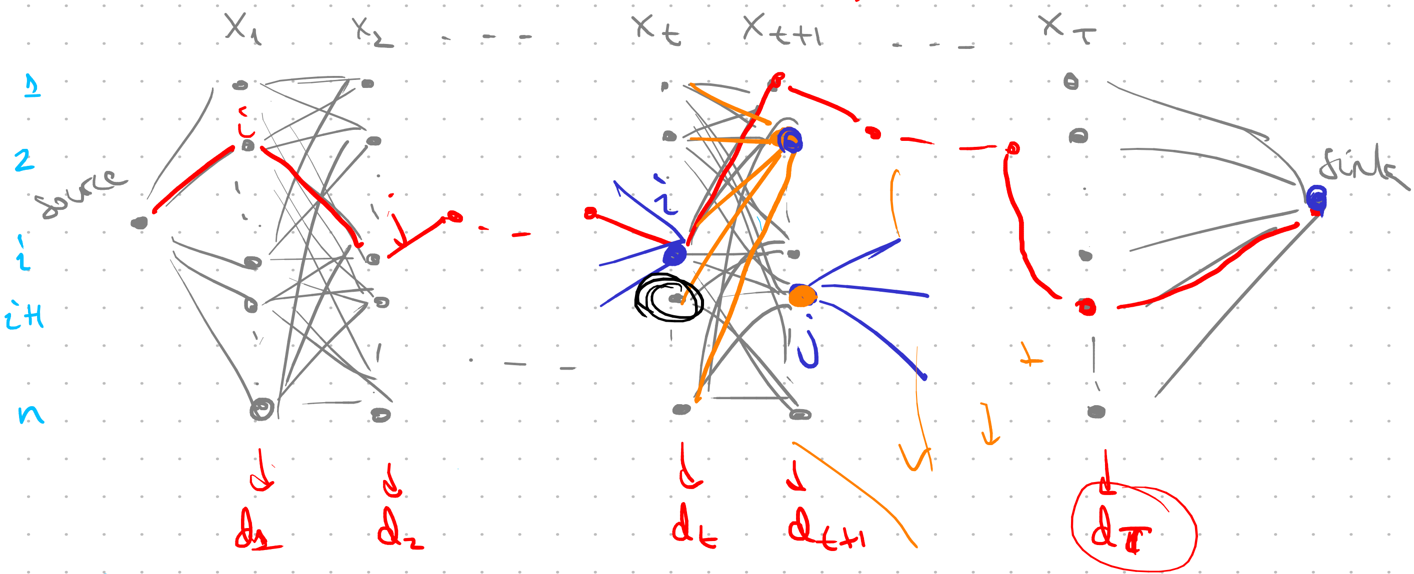
inference 2) $\bar\theta = (\pi, A, B)$: $\quad p(Q|D,\bar\theta) \xrightarrow{Q} \max$

learning 3) $p(\mathcal{D}|\bar\theta) \xrightarrow{\bar\theta} \max$

② $p(D|\bar\theta) = ?$

$$\sum_Q p(D,Q|\bar\theta) =$$

$$= \sum_Q \pi_{q_1} b_{q_1}(d_1)\, a_{q_1 q_2} \dots b_{q_T}(d_T)$$



$$\boxed{\alpha_t(i) = p(d_1 d_2 \dots d_t, q_t = i|\bar\theta)} = \sum_{q_1 \dots q_{t-1}} p(d_1 \dots d_t, q_1 \dots q_t = i|\bar\theta)$$

$$\alpha_1(i) = p(d_1, q_1 = i|\bar\theta) = \pi_i \cdot b_i(d_1)$$

$$\alpha_{t+1}(i) = p(d_1 \dots d_t, d_{t+1}, q_{t+1} = i|\bar\theta) = \sum_{j=1}^{n} p(d_1 \dots d_t, d_{t+1}, q_t = j, q_{t+1} = i|\bar\theta)$$

$$= \sum_j \underbrace{p(d_1 \dots d_t, q_t = j|\bar\theta)}_{\alpha_t(j)} \cdot \underbrace{p(q_{t+1} = i|q_t = j, \cancel{d_1 \dots d_t}, \bar\theta)}_{a_{ji}}\, \underbrace{p(d_{t+1}|q_{t+1} = i, \cancel{q_t = j}, \cancel{d_1 \dots d_t}, \bar\theta)}_{b_i(d_{t+1})}$$

$$\boxed{\alpha_{t+1}(i) = \sum_{j=1}^{n} a_{ji}\, b_i(d_{t+1})\, \alpha_t(j)}$$

$$\boxed{p(D|\bar\theta) = p(d_1 \dots d_T|\bar\theta) = \sum_{i=1}^{n} \alpha_T(i)}$$

$$\beta_t(i) = p(d_{t+1}, \dots, d_T \mid q_t = i, \bar\theta) \qquad \beta_T(i) = 1$$

$$\beta_t(i) = \sum_j p(d_{t+1} \dots d_T, q_{t+1} = j \mid q_t = i, \bar\theta) =$$

$$= \sum_j \underbrace{p(q_{t+1} = j \mid q_t = i, \bar\theta)}_{a_{ij}} \underbrace{p(d_{t+1} \mid q_{t+1} = j, \cancel{q_t = i}, \bar\theta)}_{b_j(d_{t+1})} \cdot \overbrace{p(d_{t+2} \dots d_T \mid q_{t+1} = j, \cancel{q_t = i}, \cancel{d_{t+1}}, \bar\theta)}^{\beta_{t+1}(j)}$$

$$\boxed{\beta_t(i) = \sum_{j=1}^{n} a_{ij}\, b_j(d_{t+1})\, \beta_{t+1}(j)}$$

$$p(D \mid \bar\theta) = \sum_i p(q_1 = i \mid \bar\theta)\, p(d_1 \mid q_1 = i, \bar\theta)\, p(d_2 \dots d_T \mid q_1 = i, \cancel{d_1}, \bar\theta)$$

$$\boxed{p(D \mid \bar\theta) = \sum_{i=1}^{n} \pi_i\, b_i(d_1) \cdot \beta_1(i)}$$

③ Inference

$$p(Q \mid D, \bar\theta) \xrightarrow{Q} \max$$

$$p(q_t \mid D, \bar\theta) \xrightarrow{q_t} \max \qquad \boxed{p(q_t = i \mid D, \bar\theta) = \gamma_t(i)}$$

$$\gamma_t(i) = \frac{p(q_t = i, d_1 \dots d_t \mid \bar\theta)}{p(D \mid \bar\theta)} = \frac{\overbrace{p(q_t = i, d_1 \dots d_t \mid \bar\theta)}^{\alpha_t(i)}\, \overbrace{p(d_{t+1} \dots d_T \mid q_t = i, \cancel{d_1 \dots d_t}, \bar\theta)}^{\beta_t(i)}}{p(D \mid \bar\theta)}$$

$$\boxed{\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{p(D \mid \bar\theta)} \propto \alpha_t(i)\beta_t(i)}$$

Viterbi algorithm

$$\boxed{\delta_t(i) = \max_{q_1 \dots q_{t-1}} p(q_1 \dots q_{t-1}, q_t = i, d_1 \dots d_t \mid \bar\theta)} \qquad \delta_1(i) = \alpha_1(i) = \pi_i b_i(d_1)$$

$$\delta_{t+1}(i) = \max_j \max_{q_1 \dots q_{t-1}} p(q_1 \dots q_{t-1}, q_t = j, q_{t+1} = i, d_1 \dots d_{t+1} \mid \bar\theta)$$

$$\delta_{t+1}(i) = \max_{j=1}^{n} \left(\delta_t(j) \cdot a_{ji} \cdot b_i(d_{t+1}) \right)$$

④ Learning $\quad p(\mathcal{D}|\bar\theta) \xrightarrow[\pi, A, B]{} \max$

$$p(\mathcal{D}|\bar\theta) = \prod_{s=1}^{S} \sum_{Q} p(D^{(s)}, Q|\bar\theta) \xrightarrow{\bar\theta} \max$$

$$\bar\theta \to \max$$

$$Q(\bar\theta, \bar\theta^{(k)}) = E_{Q^{(s)} \sim p(Q^{(s)}|D^{(s)}, \bar\theta^{(k)})}\left[\log \prod_{s=1}^{S} p(D^{(s)}, Q^{(s)}|\bar\theta)\right] =$$

$$= E_{Q^{(s)}}\left[\sum_{s=1}^{S}\left(\log \pi_{q_1^{(s)}} + \log b_{q_1^{(s)}}(d_1^{(s)}) + \log a_{q_1^{(s)} q_2^{(s)}} + \dots + \log a_{q_{T-1}^{(s)} q_T^{(s)}} + \log b_{q_T^{(s)}}(d_T^{(s)})\right)\right]$$

$$= \sum_{s=1}^{S} \sum_{q_1^{(s)} \dots q_T^{(s)}} p(q_1^{(s)} \dots q_T^{(s)}|\bar\theta^{(k)}, D^{(s)}) \cdot \left(\log \pi_{q_1^{(s)}} + \dots + \log b_{q_T^{(s)}}(d_T^{(s)})\right)$$

$$\log \pi_i: \quad \sum_{s} \sum_{q_2^{(s)} \dots q_T^{(s)}} p(q_1^{(s)}=i, q_2^{(s)} \dots q_T^{(s)}|\bar\theta^{(k)}, D^{(s)}) = \sum_{s} \underbrace{p(q_1^{(s)}=i|D^{(s)}, \bar\theta^{(k)})}_{=\gamma_1^{(s)}(i)}$$

$$\log \pi_i: \quad \sum_{s} \gamma_1^{(s)}(i) \qquad \boxed{\pi_i^{(k+1)} = \frac{\sum_s \gamma_1^{(s)}(i) + 1}{S + n}}$$

$$\log a_{ij}: \quad \sum_{s} \sum_{Q^{(s)}} p(Q^{(s)}|D^{(s)}, \bar\theta^{(k)}) \cdot \sum_{t=1}^{T-1} [q_t^{(s)}=i, q_{t+1}^{(s)}=j] =$$

$$= \sum_{s} \sum_{t=1}^{T-1} \xi_t^{(s)}(i,j)$$

$$\xi_t(i,j) = p(q_t=i, q_{t+1}=j|\bar\theta, D) = \frac{p(d_1 \dots d_t, q_t=i, q_{t+1}=j, d_{t+1} \dots d_T|\bar\theta)}{p(D|\bar\theta)} \propto$$

$$\propto p(d_1 \dots d_t, q_t=i|\theta)\, p(q_{t+1}=j|q_t=i, \bar\theta)\, p(d_{t+1}|q_{t+1}=j, \bar\theta) \cdot$$

$$\cdot\, p(d_{t+2} \dots d_T|q_{t+1}=j, \theta)$$

$$\boxed{\xi_t(i,j) \propto \alpha_t(i)\, a_{ij}\, b_j(d_{t+1})\, \beta_{t+1}(j)}$$

$$a_{ij}^{(k+1)} = \frac{\sum_s \sum_{t=1}^{T-1} \xi_t^{(s)}(i,j) + 1}{\sum_s \sum_{t=1}^{T-1} \gamma_t^{(s)}(i) + n}$$

$$A = \begin{pmatrix} a_{ii} \\ \dots \; a_{ii} \; \dots \end{pmatrix} \; i$$

$$(1 - a_{ii}) \cdot 1 + a_{ii}(1 - a_{ii}) \cdot 2 + a_{ii}^2 (1 - a_{ii}) \cdot 3$$

$\log b_i(\ell)$: $\quad \mathbb{E}_{\bar\Theta^{(m)}}\left[\#\ q_t^{(s)}=i \text{ u } \underline{d_t^{(s)}=\ell}\right] =$

$$= \sum_{s=1}^{S} \sum_{t:d_t^{(s)}=\ell} \gamma_t^{(s)}(i)$$

Baum-Welch algorithm

Gaussian mixture model

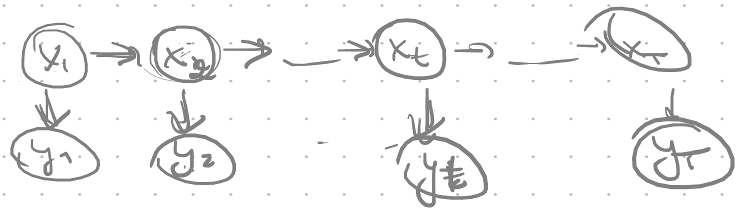
HMM - GMM

$\mathbb{R}^{40}$ - MFCC $\rightsquigarrow \mathbb{R}^d$ - DBN

$$\boxed{b_i(\ell)^{(k+1)} = \frac{\sum_s \sum_{t:d_t^{(s)}=\ell} \gamma_t^{(s)}(i) + 1}{\sum_s \sum_t \gamma_t^{(s)}(i) + m}}$$

⑤ Directed graphical models / Bayesian belief networks

Judea Pearl causality



$$p(x,y) = p(x_1)p(y_1|x_1)p(x_2|x_1)p(y_2|x_2) \dots p(y_T|x_T)$$

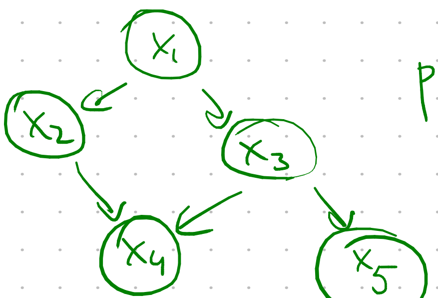
$x_1 \dots x_{T-1}, y_1 \dots y_{T-1}$

$$p(x_1,x_2,\dots x_n) = \prod_{i=1}^{n} p(x_i|\dots)$$

$\quad f(x_1 \dots x_n)$

$$= p(x_1)p(x_2|x_1)p(x_3|\cancel{x_1},x_2)\dots p(x_n|x_1 \cancel{\dots x_{n-1}})$$

$$\boxed{p(x_1,\dots,x_n) = \prod_{i=1}^{n} p(x_i|\mathrm{par}(x_i))}$$



$$p(x_1 \dots x_5) = p(x_1)p(x_2|x_1)p(x_3|x_1)p(x_4|x_2,x_3)p(x_5|x_3)$$

$p(D|\bar\theta) \to \max$

$\underline{p(\bar\theta|D)} \to \max$

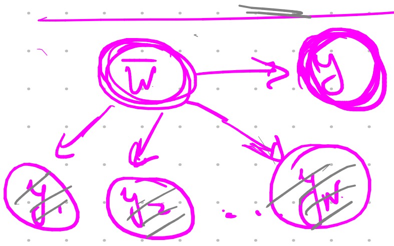
$$p(\bar x|D) = \int \underline{p(\bar x|\bar\theta)p(\bar\theta|D)}\,d\theta$$

$$p(\bar\theta, D) = \overbrace{p(\bar\theta)}^{\text{prior}}\ \overbrace{p(D|\bar\theta)}^{\text{likelihood}}$$

1) Lin. regr. $\quad \bar w \quad \underline{p(\bar w)}, \ \underline{p(D|\bar w) = \prod_n p(y_n|\bar w)}$

$$p(y, \bar w, D) = p(\bar w)\,p(D|\bar w)\cdot p(y|\bar w)$$

$$p(\bar w, D) = p(\bar w)\cdot \underline{\prod_n p(y_n|\bar w)}$$



marginal distribution

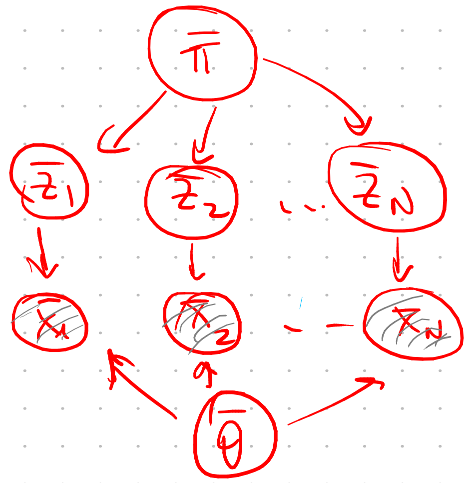
$$\underline{p(y|D)} = \int p(y|\bar w)\,\underline{p(\bar w|D)}\,d\bar w$$

$\underline{p(\bar w|D) = ?}$

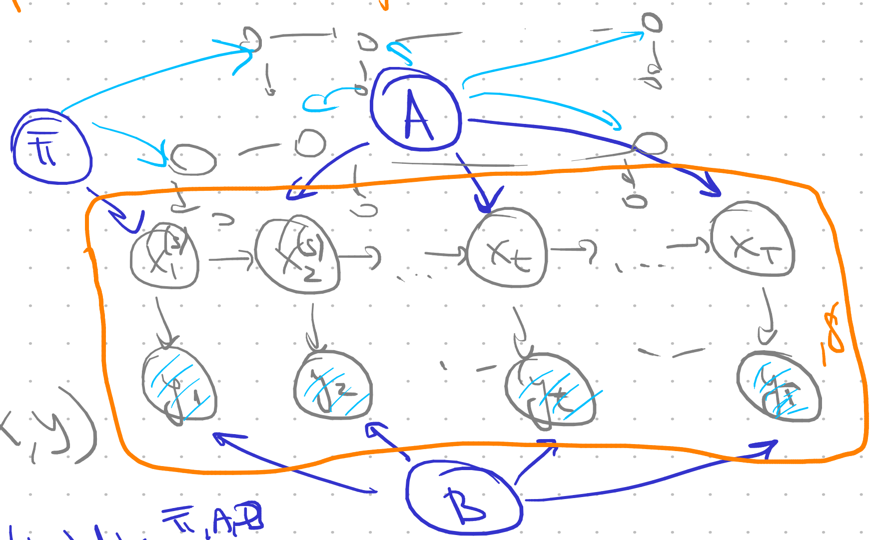
Marginalization

2) Clustering

$$p(\bar\pi, \bar\theta = \{\bar\mu_k, \Sigma_k\}_{k=1}^{K}, Z, X) = p(\bar\pi)\,p(\bar\theta) \cdot \prod_{n=1}^{N} p(\bar z_n|\bar\pi)\,p(\bar x_n|\bar z_n, \bar\theta)$$

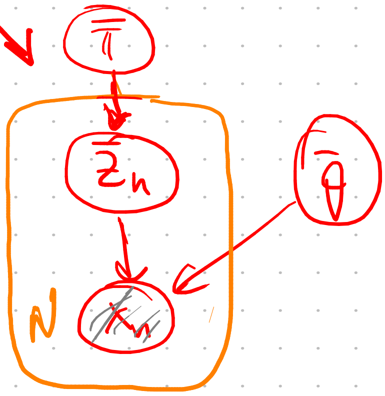


$$p(\bar\pi, \bar\theta|X) = \int p(\bar\pi, \bar\theta, z|X)\,dz$$



3) HMM $\quad p(\bar\pi, A, B, X, Y)$

$$p(\bar\pi, A, B|y) = \int p(\bar\pi, A, B, X|y)\,dX \xrightarrow{\bar\pi, A, B} \max$$



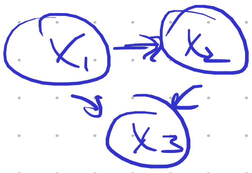
⑥ d-раздельность / d-separability

1) $n=1$ $\quad x_1 \quad p(x_1) = p(x_1)$

2) $n=2$ $\quad x_1 \quad x_2 \quad p(x_1, x_2) = p(x_1)\,p(x_2)$ $\quad x_1 \leftarrow x_2$

$x_1 \to x_2 \quad p(x_1, x_2) = p(x_1)\,p(x_2|x_1) = p(x_2)\,p(x_1|x_2)$

3) $n=3$ $\quad x_1 \quad x_2 \quad x_3$



$p(x_1)\,p(x_2|x_1)\,p(x_3|x_1, x_2)$

$x_1 \quad x_2 \to x_3$

$p(x_1)\,p(x_2)\,p(x_3|x_2)$

Последовательная связь $\quad x_1 \to x_2 \to x_3$

$$p(x_1, x_2, x_3) = p(x_1)\,p(x_2|x_1)\,p(x_3|x_2)$$

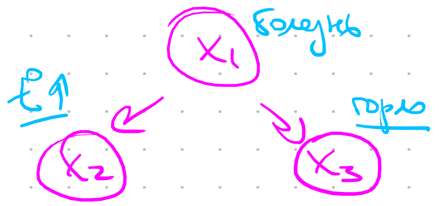
?

X_1 и X_3 условно независимы при усл. X_2

$$p(x_1, x_3 | x_2) \overset{?}{=} p(x_1|x_2)p(x_3|x_2)$$

$$\frac{p(x_1,x_2,x_3)}{p(x_2)} = \frac{p(x_1)p(x_2|x_1)p(x_3|x_2)}{p(x_2)}$$

Расходящаяся связь

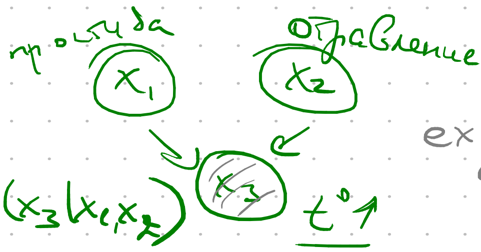


$$p(x_1)p(x_2,x_3|x_1)$$

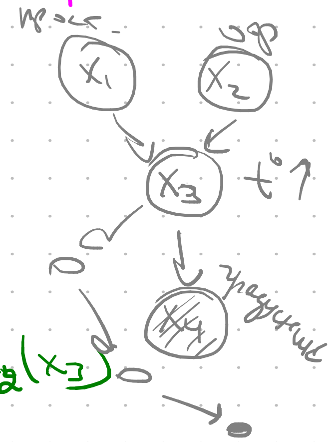
$$p(x_1,x_2,x_3) = p(x_1)p(x_2|x_1)p(x_3|x_1)$$

X_2 и X_3 усл. нез. при X_1 $\quad p(x_2,x_3|x_1) \overset{?}{=} p(x_2|x_1)p(x_3|x_1)$

Сходящаяся связь



explaining away

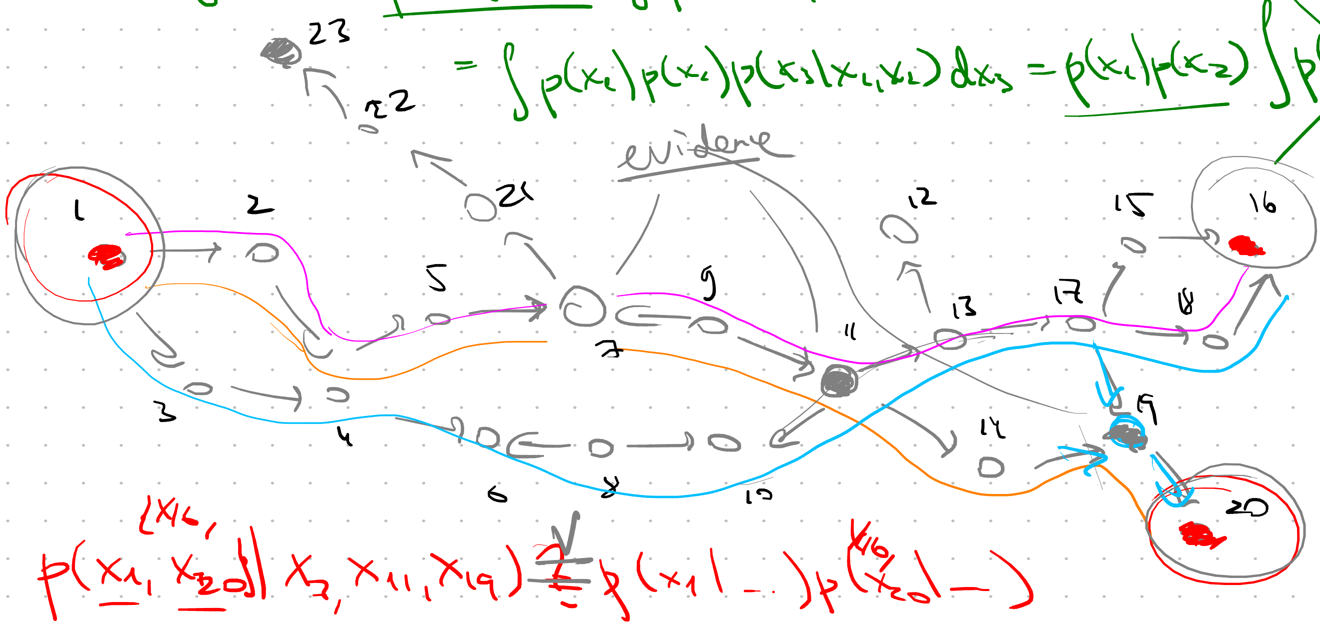


$$p(x_1,x_2,x_3) = p(x_1)p(x_2)p(x_3|x_1,x_2)$$

~~X_1, X_2 усл. нез. при x_3~~ $\quad$ ~~$p(x_1,x_2|x_3) = p(x_1|x_3)p(x_2|x_3)$~~

X_1 и X_2 незав: $\quad p(x_1,x_2) = \int p(x_1,x_2,x_3)dx_3 =$

$$= \int p(x_1)p(x_2)p(x_3|x_1,x_2)dx_3 = p(x_1)p(x_2)\cancel{\int p(x_3|\ldots)dx_3}$$



$$p(x_1, x_{20} | \{x_{16}, x_7, x_{11}, x_{19}) \overset{?}{=} p(x_1|\ldots)p(x_{20}|\ldots)$$