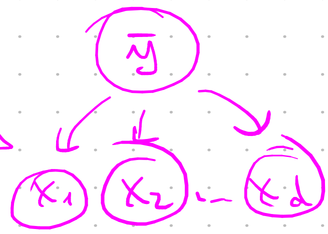
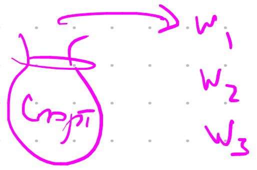
① NB

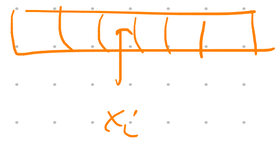
$$D = \{(\bar{x}, \bar{y})\} \quad \bar{y} = (0 \ldots 1 \ldots 0), \quad \bar{x} = (x_1 \ldots x_d)$$

$$p(\bar{x}, \bar{y}) = \underbrace{p(\bar{y})}_{p(C_k)} \cdot \prod_{i=1}^{d} p(x_i \mid \bar{y})$$



$$p(d, C=k) = \underbrace{p(C=k)}_{\theta_k} \cdot \prod_{i=1}^{|d|} \underbrace{p(w_i \mid C=k)}$$





$\theta_{imk} = p(x_i = m \mid y = k)$

$$\theta_{mk} = p(w_i = m \mid C = k)$$

$k = 1 \ldots K$

$m = 1 \ldots V$

$i = 1 \ldots d$

$$\theta_k = p(C=k) \approx \frac{\#\{d \in C_k\} + 1}{|D| + K}$$

$$\theta_{mk} \approx \frac{\#\{w_i = m,\ d \in C_k\} + 1}{\#\{\text{cuvinte în } C_k\} + |V|}$$

Bayesian smoothing

Cromwell's rule

- bag of words
- supervised learning - labeled data
- $\forall$ doc belongs to a topic

② Clustering

$$p(D \mid \bar{\theta}) = \prod p(d \mid \bar{\theta}) = \prod_{d} \sum_{t=1}^{T} p(d, t \mid \bar{\theta}) =$$

EM algorithm

$$= \prod_{d} \sum_{t=1}^{T} \underbrace{p(t)}_{\theta_t} \cdot \prod_{w \in d} \underbrace{p(w \mid t)}_{\theta_{wt}}$$

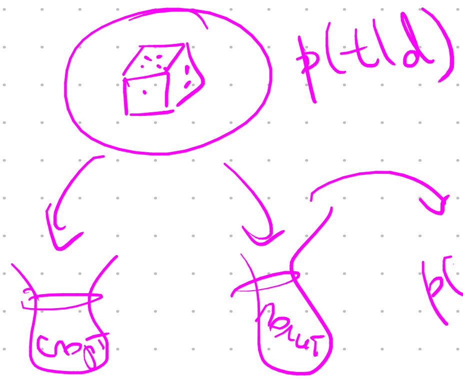
③ Topic modeling

$t : 1 \ldots T$

$w : 1 \ldots V$

$d : 1 \ldots N$

$$X = \begin{array}{c|c} & d \\ \hline w & n_{wd} = \#\{w \in d\} \end{array}_{V \times N}, \quad \frac{n_{wd}}{n_d}$$



$$p(t \mid d) = \theta_{td}$$

$$p(w \mid t) = \varphi_{wt}$$

$$p(w \mid d) = \sum_{t=1}^{T} p(w, t \mid d) = \sum_{t=1}^{T} p(t \mid d)\, p(w \mid t) = \sum \theta_{td} \varphi_{wt}$$

$$p(w|d) = \sum_{t=1}^{T} \theta_{td} \varphi_{wt}$$

low-rank decomposition

$$\Theta = \begin{array}{c|c} & d \\ \hline t & \begin{pmatrix} \theta_{td} \end{pmatrix} \end{array} \quad T \times N$$

$$\Phi = \begin{array}{c|c} & t \\ \hline w & \begin{pmatrix} \varphi_{wt} \end{pmatrix} \end{array} \quad V \times T$$

$$\Phi \cdot \Theta = \begin{array}{c|c} & d \\ \hline w & \begin{pmatrix} \sum_t \varphi_{wt}\theta_{td} \end{pmatrix} \end{array} = \begin{array}{c|c} & d \\ \hline w & \begin{pmatrix} p(w|d) \end{pmatrix} \end{array} \approx X$$

$(V \times T) \cdot (T \times N)$ $\qquad V \times N$

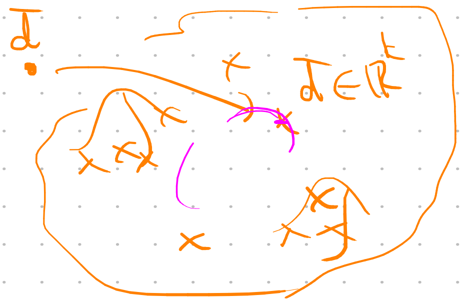
$T \ll V, N$

prob. latent semantic indexing

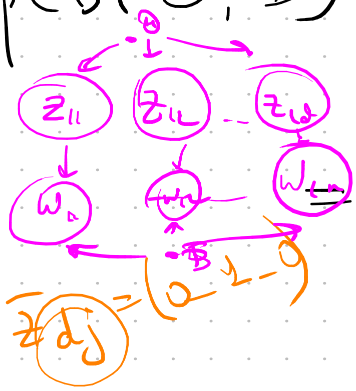
④ pLSI / LSA

Latent semantic analysis

Neural topic models



$$p(D|\Theta, \Phi) = \prod_{d=1}^{N} p(d|\Theta, \Phi) =$$



$$= \prod_{d} \prod_{j=1}^{n_d} \left(\sum_{t=1}^{T} \theta_{td} \cdot \varphi_{w_{dj},t} \right) \qquad \Theta, \Phi \longrightarrow \max$$

$z_{djt} = 1 \iff$ позиция j в док. d порожд у темы t

$$p(D, z|\Theta, \Phi) = \prod_{d=1}^{N} \prod_{j=1}^{n_d} \prod_{t=1}^{T} (\theta_{td} \varphi_{w_{dj},t})^{z_{djt}}$$

EM-algorithm:

$$Q(\Theta, \Phi, \Theta^{(m)}, \Phi^{(m)}) = E_{z|\Theta^{(m)}, \Phi^{(m)}} \left[\log p(D, z|\Theta, \Phi) \right] =$$

$$= \sum_{d} \sum_{j} \sum_{t} \underbrace{E[z_{djt}]} \cdot \left(\log \theta_{td} + \log \varphi_{w_{dj},t} \right)$$

E-шаг:

$$E[z_{djt}] = p(t \mid d, w_{dj}) = \frac{p(t,d,w_{dj})}{p(d,w_{dj})} = \frac{\theta^{(m)}_{td} \cdot \varphi^{(m)}_{w_{dj},t}}{\sum_{s=1}^{T} \theta^{(m)}_{sd} \varphi^{(m)}_{w_{dj},s}}$$

$$n^{(m)}_{dwt} = E[\#\ \text{слов } w \text{ в } d \text{ из } t] = n_{dw} \cdot p(t \mid d, w)$$

M-шаг:

$$\theta^{(m+1)}_{td} = \frac{E[\# t \text{ в } d]}{n_d} = \frac{\sum_{w=1}^{V} n^{(m)}_{dwt}}{n_d} = \frac{n^{(m)}_{d*t} + 1}{n_{d**} + T}$$

$$\varphi^{(m+1)}_{wt} = \frac{E[\# w \text{ в } t]}{E[\# \text{ слов в } t]} = \frac{\sum_{d=1}^{N} n^{(m)}_{dwt}}{\sum_d \sum_w n^{(m)}_{dwt}} = \frac{n^{(m)}_{*wt} + 1}{n^{(m)}_{**t} + V}$$

$$X' \approx \Phi \cdot \Theta = (\Phi \cdot A)\underbrace{\ }_{T \times T}(A^{-1} \cdot \Theta)$$

5. ARTM — additive regularization of topic models

Константин Воронцов

$$\log p(D \mid \Theta, \Phi) = \sum_{d=1}^{N} \sum_{w=1}^{V} n_{dw} \cdot \log\left(\sum_{t=1}^{T} \theta_{td} \varphi_{wt} \right) + R(\Theta, \Phi)$$

$$n_{*wt} = \mathrm{ReLU}\left(\sum_{d=1}^{N} n_{dwt} + \varphi_{wt} \cdot \frac{\partial R}{\partial \varphi_{wt}} \right)$$

$$n_{d*t} = \mathrm{ReLU}\left(\sum_{w=1}^{V} n_{dwt} + \theta_{td} \frac{\partial R}{\partial \theta_{td}} \right)$$

$$\Phi = \begin{array}{cc} & t \\ w & \left(\begin{array}{c|c|c} & \varphi_{wt} & \end{array} \right) \end{array}$$

$$\bar{\varphi}_{*t} \not\sim \mathrm{Unif}$$

$$KL(\bar{\varphi} \| Un) \to \max$$

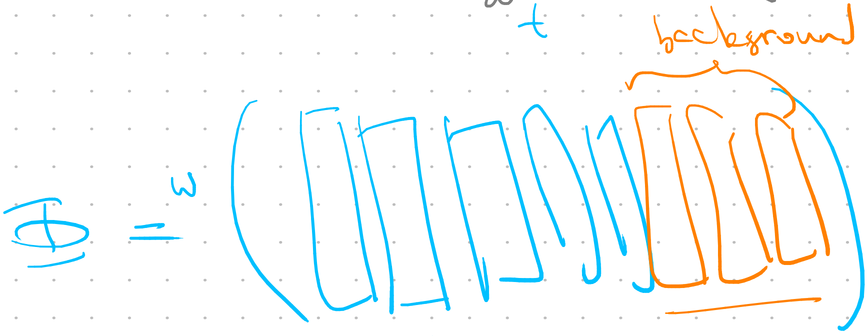
$$KL(Un \| \bar{\varphi}) \to \max$$

$\forall t, s$

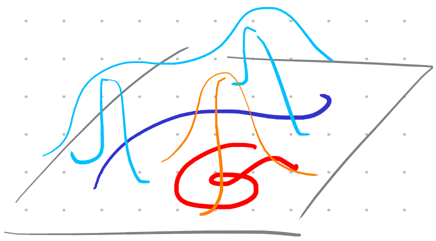
$$KL(\bar{\varphi}_{*t} \| \bar{\varphi}_{*s}) \to \max$$

$$KL(\text{Unif} \,\|\, \varphi) = \sum_w \frac{1}{|W|} \cdot \log \frac{1/|W|}{\varphi_{wt}} \to \max \qquad \boxed{\sum \log \varphi_{wt} \to \min}$$

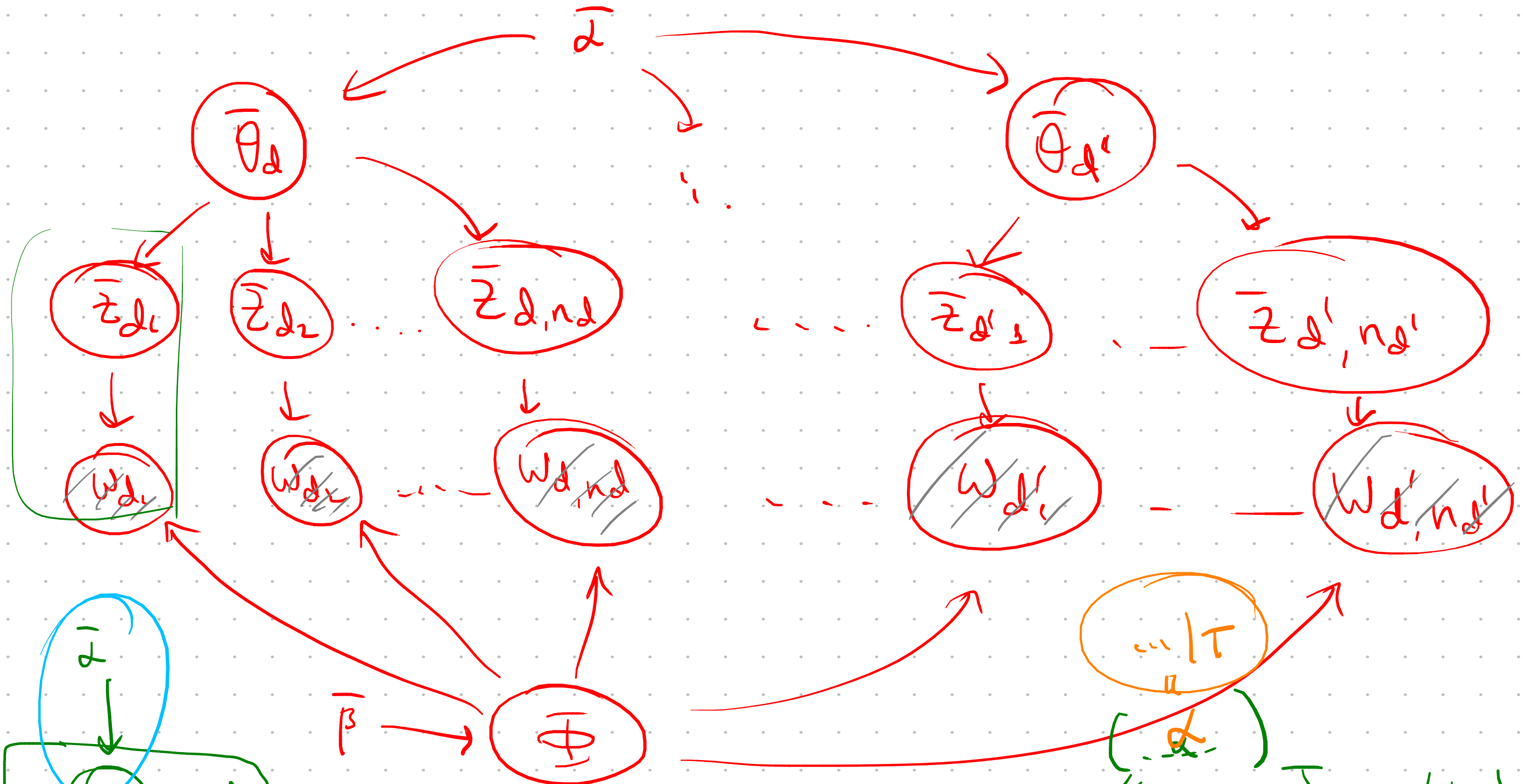
$$KL(\varphi \,\|\, \text{Unif}) = \sum_w \varphi_{wt} \log \varphi_{wt} / \text{const} \to \max \qquad \boxed{\text{Ent} \to \min}$$

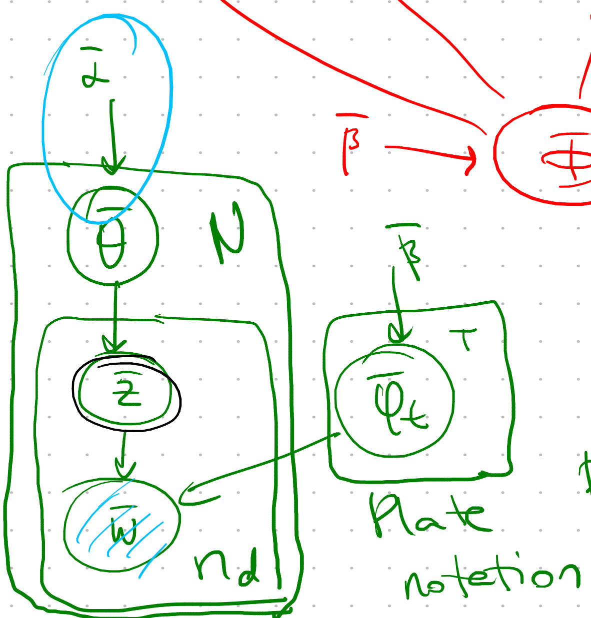


$KL(\varphi \,\|\, \bar{\varphi}) \to \max$ $\quad$ $KL(\varphi \,\|\, \bar{\varphi}) \to \min$



(6) LDA – Latent Dirichlet Allocation





$$p(\bar{\theta}_d | \bar{\alpha}) = \text{Dir}(\bar{\theta}_d | \bar{\alpha}) = \frac{1}{B(\bar{\alpha})} \prod_{t=1}^{T} \theta_{td}^{\alpha_t - 1}$$

$$p(\bar{\varphi}_t | \bar{\beta}) = \text{Dir}(\bar{\varphi}_t | \bar{\beta}) = \frac{1}{B(\bar{\beta})} \cdot \prod_{w=1}^{V} \varphi_{wt}^{\beta_w - 1}$$

$$p(\Theta, \Phi \mid X, \bar\alpha, \bar\beta) \longrightarrow \max \qquad \text{MAP?}$$

$$\propto \; p(X, \Theta, \Phi \mid \bar\alpha, \bar\beta) = \sum_{Z} \underline{p(W, Z, \Theta, \Phi \mid \bar\alpha, \bar\beta)}$$

$$p(W, Z, \Theta, \Phi \mid \bar\alpha, \bar\beta) = p(\Phi \mid \bar\beta) \cdot p(\Theta \mid \bar\alpha) \cdot p(Z \mid \Theta)\, p(X \mid Z, \bar\Phi)$$

$$= \left[\prod_{t=1}^{T} \underbrace{p(\bar\varphi_t \mid \bar\beta)}_{\text{Dir}}\right] \cdot \left[\prod_{j=1}^{M} \left[\underbrace{p(\bar\theta_j \mid \bar\alpha)}_{\text{Dir}} \cdot \prod_{n=1}^{N_j} \underbrace{p(\bar z_{jn} \mid \bar\theta_j)}_{\text{Mult}(\bar\theta_j)}\, \underbrace{p(w_{jn} \mid \bar z_{jn}, \bar\Phi)}_{\text{Mult}(\Phi_{\bar z_{jn}})}\right]\right]$$

7 Variational approximation

$$p(Z, \Theta, \Phi \mid W, \alpha, \beta) \approx q(Z, \Theta, \Phi)$$

$$\log p(W \mid \bar\alpha, \bar\beta) = \log \boxed{\int_{\Phi}\int_{\Theta}\sum_{Z}} \; p(Z, \Theta, \Phi, W \mid \alpha, \beta) \cdot \frac{\boxed{q(Z, \Theta, \Phi)}}{q(Z, \Theta, \Phi)}\, d\Theta\, d\Phi$$

$$\mathbb{E}_q[\;\cdot\;]$$

$$\geq \int_{\Phi}\int_{\Theta}\sum_{Z} q(Z, \Theta, \Phi) \log \frac{p(Z, \Theta, \Phi, W \mid \alpha, \beta)}{q(Z, \Theta, \Phi)}\, d\Theta\, d\Phi =$$

$$\cancel{= -\,KL(q \,\|\, p\,-\,-)}$$

$$\pm\, q(Z, \Theta, \Phi)\, \mathbb{E}_q$$

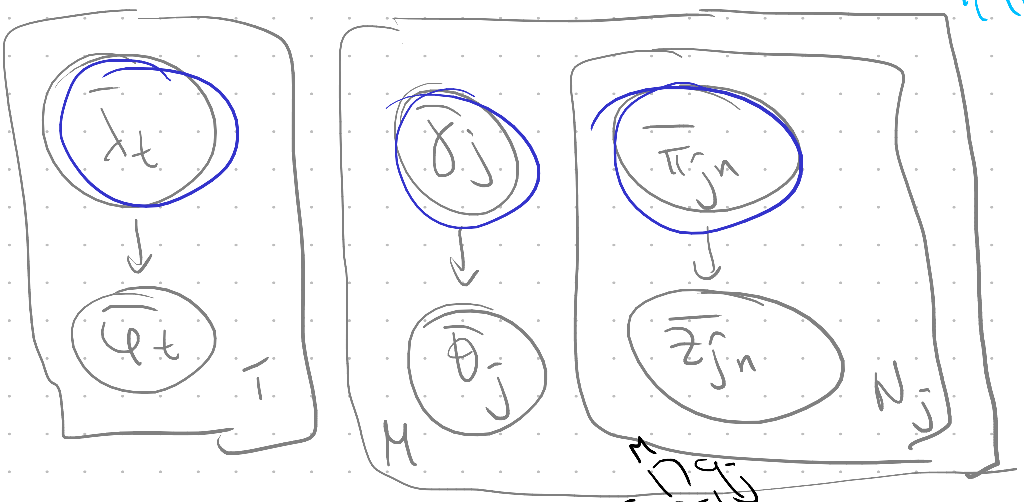
$$\log p(W \mid \bar\alpha, \bar\beta) = \log p(\;\_\_\;) - \log p\underbrace{(Z, \Theta, \Phi \mid W, \alpha, \beta)}$$

$$= \underbrace{\mathbb{E}_q\left[\log \frac{p(Z, \Theta, \Phi, W \mid \alpha, \beta)}{q(Z, \Theta, \Phi)}\right]}_{\to \max} + KL\big(\underbrace{q \,\|\,}_{\to \min}\;\big)$$

$$L(q) = \mathbb{E}_q\left[\log p(Z,\Theta,\Phi,W\mid\alpha,\beta)\right] - \mathbb{E}_q\left[q(Z,\Theta,\Phi)\right]$$

$$q(Z,\Theta,\Phi) = q(Z)\cdot q(\Theta)\cdot q(\Phi) =$$

$$= \left(\prod_{t=1}^{T} \underbrace{q_t(\bar\varphi_t\mid\bar\lambda_t)}_{=\,\mathrm{Dir}(\bar\varphi_t\mid\bar\lambda_t)}\right)\left(\prod_{j=1}^{M} \underbrace{q_j(\bar\theta_j\mid\bar\gamma_j)}_{=\,\mathrm{Dir}(\bar\theta_j\mid\bar\gamma_j)}\right)\left(\prod_{j=1}^{M}\prod_{n=1}^{N_j} \underbrace{q_{jn}(\bar z_{jn}\mid\bar\pi_{jn})}_{=\,\mathrm{Mult}(\bar z_{jn}\mid\bar\pi_{jn})}\right)$$



$$q(\Phi,\Theta,Z\mid\Lambda,\Gamma,\Pi)$$

$$= \prod_{j=1}^{M} p_j$$

(8) Φ fixed. $\quad q(\Theta,Z\mid\Gamma,\Pi) \approx p(Z,\Theta\mid W,\Phi,\alpha,\beta)$

$$L(q_j) = \mathbb{E}_{q_j}\left[\log \frac{p(\bar w_j,\bar z_j,\bar\theta_j\mid\Phi,\alpha,\beta)}{q_j(\bar z_j,\bar\theta_j\mid\bar\gamma_j,\Pi_j)}\right] =$$

$$= \underbrace{\mathbb{E}_{q_j}\left[\log p(\bar\theta_j\mid\bar\alpha)\right]}_{\boxed{\mathrm{I}}} + \underbrace{\mathbb{E}_{q_j}\left[\log p(\bar z_{jn}\mid\bar\theta_j)\right]}_{\boxed{\mathrm{II}}} + \underbrace{\mathbb{E}_{q_j}\left[\log p(\bar w_j\mid\bar z_j,\Phi)\right]}_{\boxed{\mathrm{III}}}$$

$$- \underbrace{\mathbb{E}_{q_j}\left[\log q_j(\bar\theta_j\mid\bar\gamma_j)\right]}_{\boxed{\mathrm{IV}}} - \underbrace{\mathbb{E}_{q_j}\left[\log q_j(\bar z_{jn}\mid\Pi_j)\right]}_{\boxed{\mathrm{V}}}$$

$\boxed{\mathrm{I}}$ $$\mathbb{E}_q\left[\log p(\bar\theta_j\mid\bar\alpha)\right] = \mathbb{E}_{q_j}\left[-\underbrace{\log B(\bar\alpha)}_{} + \sum_{t=1}^{T}\underbrace{(\alpha_t-1)}_{}\underbrace{\log\theta_{jt}}_{}\right] =$$

$$B(\bar\alpha) = \frac{\prod\Gamma(\alpha_t)}{\Gamma(\sum_t \alpha_t)}$$

$$= \log\Gamma\left(\sum\alpha_t\right) - \sum_t \log\Gamma(\alpha_t) + \sum_t(\alpha_t-1)\,\mathbb{E}_{q_j}\left[\log\theta_{jt}\right]$$

$$\boxed{E_{Dir(\bar{x}|\bar{\beta})}[\log x_i] = \psi(\beta_i) - \psi(\Sigma\beta_i)}$$

$$\mathrm{I} \quad E_{q_j}[\log p(\bar{\theta}_j|\bar{\alpha})] = \log\Gamma(\Sigma\alpha_t) - \sum_t \log\Gamma(\alpha_t) + \sum_t(\alpha_t - 1)\left(\Psi(\gamma_{jt}) - \psi\left(\sum_s \gamma_{js}\right)\right)$$

$$\mathrm{II} \quad E_{q_j}[\log p(\bar{z}_j|\bar{\theta}_j)] = E_{q_j}\left[\sum_{n=1}^{N_j} \log p(\bar{z}_{nj}|\bar{\theta}_j)\right] =$$

$$= E_{q_j}\left[\sum_{n=1}^{N}\sum_{t=1}^{T}[z_{jn}=t]\cdot\log\theta_{jt}\right] =$$

$\psi(\_) - \psi(\_)$

$q_j(\bar{z}_{jn})\cdot q_j(\bar{\theta}_j)$

$$= \sum_n\sum_t \underbrace{E_{q_j}[z_{jn}=t]}_{\pi_{jnt}}\cdot \underbrace{E_{q_j}[\log\theta_{jt}]}_{\psi(\_)-\psi(\_)}$$

$$\mathrm{II} \quad E_{q_j}[\log p(\bar{z}_j|\bar{\theta}_j)] = \sum_{n=1}^{N_j}\sum_{t=1}^{T}\pi_{jnt}\left(\psi(\gamma_{jt}) - \psi\left(\sum_s\gamma_{js}\right)\right)$$

$$\mathrm{III} \quad E_{q_j}[\log p(\bar{w}_j|\bar{z}_j, \Phi)] = E_{q_j}\left[\sum_{n=1}^{N_j}\sum_{t=1}^{T}\sum_{v=1}^{V}\underbrace{[z_{jn}=t]}_{\pi_{jnt}}[w_{jn}=v]\log\varphi_{tv}\right]$$

$$= \sum_{n=1}^{N}\sum_{t=1}^{T}\sum_{v=1}^{V}[w_{jn}=v]\cdot\pi_{jnt}\cdot\log\varphi_{tv}$$

$$\mathrm{IV} \quad E_{q_j}[\log q(\bar{\theta}_j|\bar{\gamma}_j)] = -\sum_s\log\Gamma(\gamma_{js}) + \log\Gamma\left(\sum_s\gamma_{js}\right)$$

$$\sum_t(\gamma_{jt}-1)\left(\psi(\gamma_{jt}) - \psi\left(\sum_s\gamma_{js}\right)\right)$$

$$\mathrm{V} \quad E_{q_j}[\log q_j(\bar{z}_j|\bar{\pi}_j)] = E_q\left[\sum_{n=1}^{N_j}\sum_{t=1}^{T}[z_{jn}=t]\cdot\log\pi_{jnt}\right] =$$

$$= \sum_{n=1}^{N}\sum_{t=1}^{T}\pi_{jnt}\cdot\log\pi_{jnt}$$

$$L(\bar\gamma_j, \bar\pi_j) = \bar{\text{I}} + \bar{\text{II}} + \bar{\text{III}} - \bar{\text{IV}} - \bar{\text{V}} \xrightarrow{\bar\gamma_j, \bar\pi_j} \max$$

при условии $\forall j, n$ $\sum_t \pi_{jnt} = 1$, $\pi_{jnt} \geq 0$

$$L(\bar\pi_j) = \sum_n \sum_t \pi_{jnt}\left(\psi(\gamma_{jt}) - \psi\left(\sum_s \gamma_{js}\right)\right) + $$

$$+ \sum_n \sum_t \sum_v [w_{jn} = v] \pi_{jnt} \log \varphi_{tv} - \sum_n \sum_t \pi_{jnt} \log \pi_{jnt}$$

$$= \sum_n \sum_t \left[\underbrace{\pi_{jnt}\left(\psi(\gamma_{jt}) - \psi\left(\sum_s \gamma_{js}\right) + \sum_v [w_{jn}=v] \log \varphi_{tv} - \log \pi_{jnt}\right)}_{\xrightarrow{\bar\pi_{jn}} \max} \right] + \lambda_{jn}\left(\sum_s \pi_{jns} - 1\right)$$

$\forall t$

$$\frac{\partial L}{\partial \pi_{jnt}} = \underbrace{\psi(\gamma_{jt}) - \psi\left(\sum_s \gamma_{js}\right) + \log \varphi_{t, w_{jn}}} - \log \pi_{jnt} - 1 + \lambda_{jn} = 0$$

$\forall t$ $\log \pi_{jnt} = \lambda_{jn} - 1 + \ldots$

$$\pi_{jnt} = e^{(\lambda_{jn} - 1) + (\ldots)} \propto e^{(\ldots)}$$

$$\boxed{\pi^*_{jnt} \propto e^{\psi(\gamma_{jt}) - \psi\left(\sum_s \gamma_{js}\right) + \log \varphi_{t, w_{jn}}}}$$

$$L(\bar\gamma_j) = \sum_t (\alpha_t - 1)\left(\psi(\gamma_{jt}) - \psi\left(\sum_s \gamma_{js}\right)\right) +$$

$$+ \sum_n \sum_t \pi_{jnt}(\psi \ldots - \psi \ldots)$$

$$- \sum_s \log \Gamma(\gamma_{js}) + \log \Gamma\left(\sum_s \gamma_{js}\right) - \sum_t \gamma_{jt}(\psi(\ldots) - \psi \ldots)$$

$$= \sum_t \left(\psi(\gamma_{jt}) - \psi\left(\sum_s \gamma_{js}\right)\right)\left(\alpha_t + \sum_n \pi_{jnt} - \gamma_{jt}\right) - \log \Gamma(\Sigma) + \Sigma \log \Gamma(\ldots)$$

$$\frac{\partial L}{\partial \gamma_{jt}} = -\psi'\left(\sum_s \gamma_{js}\right) \cdot \sum_{s=1}^{T}\left(\alpha_s + \sum_n \pi_{jns} - \gamma_{js}\right)$$

$$+ \psi'(\gamma_{jt})\left(\alpha_t + \sum_n \pi_{jnt} - \gamma_{jt}\right)$$

$$- \left(\cancel{\psi(\gamma_{jt}) - \psi\left(\sum_s \gamma_{js}\right)}\right)$$

$$+ \cancel{\psi(\gamma_{jt}) - \psi\left(\sum_s \gamma_{js}\right)} = 0$$

$\forall t$

$$\psi'(\gamma_{jt})\left(\alpha_t + \sum_n \pi_{jnt} - \gamma_{jt}\right) = \psi'\left(\sum_s \gamma_{js}\right) \cdot \sum_{s=1}^{T}\left(\alpha_s + \sum_n \pi_{jns} - \gamma_{js}\right)$$

$$\boxed{\gamma_{jt}^* = \alpha_t + \sum_{n=1}^{N_j} \pi_{jnt}}$$

EM-algorithm w.r.t. Φ:

- E-war: $KL(q_j \| p_j) \to \min \quad \forall j \quad \Phi^{(m)}$
  - repeat until convergence

$$\gamma_{jt} := \alpha_t + \sum_{n=1}^{N_j} \pi_{jnt}$$

$$\pi_{jnt} := \frac{1}{Z} e^{\psi(\gamma_{jt}) - \psi\left(\sum_s \gamma_{js}\right) + \log \varphi^{(m)}_{t, w_{jn}}}$$

- M-war:

$$\varphi_{tv} \propto \sum_{j=1}^{M} \sum_{n=1}^{N_j} [w_{jn} = v]\, \pi_{jnt}$$

5 Φ not fixed

$$q(Z, \Theta, \Phi \mid r, n, \lambda) \approx p(Z, \Theta, \Phi \mid W, \bar{\alpha}, \bar{\beta})$$

$$h(q) = \sum_{j=1}^{M} L(q_j, \pi_j) + \sum_t E_q[\log p(\bar{\varphi}_t \mid \bar{\beta})]$$

$$- \sum_t E_q[\log q(\bar{\varphi}_t \mid \bar{\lambda}_t)]$$

$$\boxed{\lambda^*_{tv} = \beta_v + \sum_{j=1}^{M} \sum_{n=1}^{N_j} [w_{jn} = v] \pi_{jnt}}$$