① Var-approx.

$\underline{p(z|x) = ?}$, где x – наблюд.
z – скрытые

просто

$p(x,z) = \underline{p(x)} \cdot \underline{p(z|x)}$

$\log p(x) = \log p(x,z) - \log p(z|x) \Big| \mathbb{E}_{q(z)} \quad \pm \mathbb{E}[\log q(z)]$

$$\log p(x) = \int \log \frac{p(x,z)}{q(z)} q(z) dz + \int \log \frac{q(z)}{p(z|x)} q(z) dz$$

$$\log p(x) = \boxed{\overset{"}{\underline{L(q)}}} + \underbrace{KL(q(z) \overset{"}{\|} p(z|x))}$$

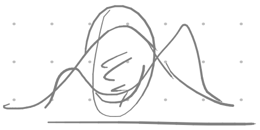
VLB
ELBO

$\rightarrow$ max $\Leftrightarrow$ $\rightarrow$ min

$\underline{q(z) = \prod_{i=1}^{M} q_i(z_i)}, \quad z_i \cap z_j \Rightarrow \varnothing \quad \Rightarrow$

$$\Rightarrow \boxed{\log q_j^*(z_j) = \mathbb{E}_{q^*(z_{-j})}\left[\log p(x,z)\right] + Const}$$

$KL(p(z|x) \| q(z)) \rightarrow \min \quad \Rightarrow \quad q_j^*(z_j) \propto \int p(x,z) dz_{-j}$



$\boxed{w^T \bar{x}}$
discrim model
$p(y|\bar{x})$

generative model
$p(\bar{x}, y)$

② Naive Bayes

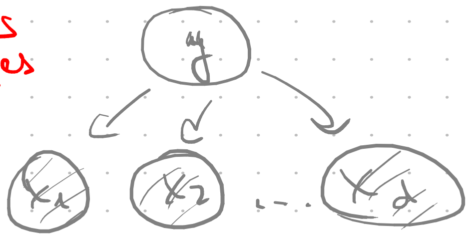
$\underline{D = \{(\bar{x}_{ni}, y_n)\}_{n=1}^{N}}$

$p(\bar{x}, y) = \underline{p(y)} \cdot \underline{\prod_{i=1}^{d} p(x_i|y)}$

naive assumption

$$p(\underline{\bar{x}}|y) = \prod_{i=1}^{d} p(x_i|y)$$

Idiot's Bayes



$y \in \{$спорт, политика, финансы$\}$

$\bar{x}$, x_i – слова?

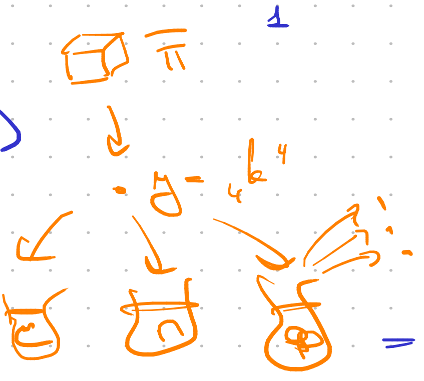
bag-of-words

multinomial / multivariate

$[\underset{1}{-} - - - - - - - -]\ \bar{x}$

$\pi_k = \underline{\log p(y=k)}$

$\theta_{wk} = \log p(x_i = w | y = k)$



$$p(D|\bar{\theta}, \bar{\pi}) = \prod_{n=1}^{N} p(y_n) \prod_{i=1}^{L_n} p(x_{ni}|y_n)$$

$$\log p(D|\bar{\theta}, \bar{\pi}) = \sum_n \left(\sum_k y_{nk} \left(\pi_k + \sum_{i=1}^{L_n} \theta_{w_{ni},k} \right) \right)$$

$$\hat{p}(y=k) \approx \frac{\#\{n: y_n=k\} + 1}{N + K}$$

Cromwell's rule

$$\hat{p}(x_i=w \mid y=k) \approx \frac{\#\{x_{ni}=w,\ y_n=k\} + 1}{\#\{cnd:\ y_n=k\} + |V|}$$

— Bag-of-words / naive assumption

— supervised learning / dataset $D=\{(\bar{x}_n, y_n)\}_{n=1}^{N}$

— $\forall$ document belongs to a topic

③ Clustering

$D=\{\bar{x}_n\}_{n=1}^{N}$ $\quad$ T – число тем $\quad$ X, Z

z_n – тема

$$p(D \mid \theta) = \prod_{n=1}^{N} \left(\sum_{t=1}^{T} p(z_n=t) \cdot \prod_{i=1}^{L_n} p(x_{ni} \mid z_n=t) \right)$$

$\pi_t = \log p(z_n=t)$, $\theta_{wt} = \log p(x_{ni}=w \mid z_n=t)$

$$p(D, Z \mid \bar{\pi}, \bar{\theta}) = \prod_{n=1}^{N} \prod_{t=1}^{T} e^{z_{nt}\left(\pi_t + \sum_{i=1}^{L_n} \theta_{x_{ni},t}\right)}$$

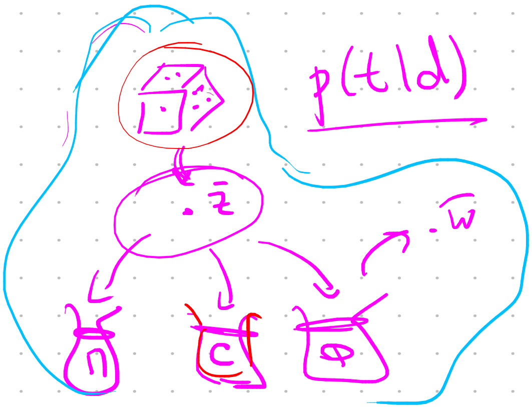
$$Q(\bar{\theta}, \bar{\theta}^{(m)}) = E_{Z \mid \theta^{(m)}}\left[\sum_{n,t} z_{nt}\left(\pi_t + \sum_{i=1}^{L_n} \theta_{x_{ni},t}\right)\right] =$$

$$= \sum_{n=1}^{N} \sum_{t=1}^{T} E[z_{nt}] \cdot \left(\pi_t + \sum_i \theta_{x_{ni},t}\right)$$

$$E[z_{nt}] = p(z_n=t \mid \bar{x}_n) = \frac{p(\bar{x}_n, z_n=t)}{p(\bar{x}_n)} = \frac{\pi_t + \sum_i \theta_{x_{ni},t}}{\sum_s\left(\pi_s + \sum_i \theta_{x_{ni},s}\right)}$$

④ Topic modeling

$D=\{\bar{x}_n\}$



$p(t \mid d)$

$$p(w \mid d) = \sum_{t=1}^{T} p(w, t \mid d) =$$

$$= \sum_{t=1}^{T} p(t \mid d) \cdot p(w \mid t)$$

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

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

$$\Theta = \begin{array}{cc} & d \\ t & \left(\bar{\theta}_{td}\right) \end{array} \qquad \Phi = \begin{array}{cc} & t \\ w & \left(\varphi_{wt}\right) \end{array}$$

$$T \times N \qquad\qquad V \times T$$

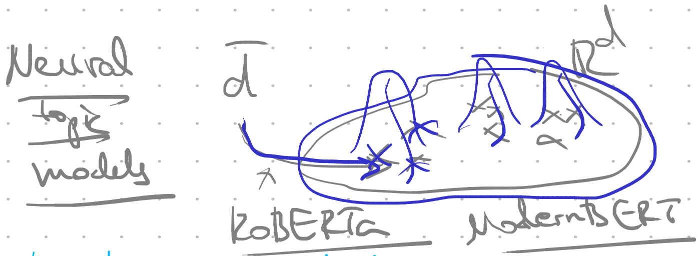
$$\Phi \cdot \Theta = \begin{array}{cc} & N \\ V & \left(\sum_{t=1}^{T} \theta_{td}\varphi_{wt}\right) \end{array} \approx X = \begin{array}{cc} & N \\ V & \left(\hat{p}(w|d) = \frac{n_{wd}}{n_d}\right) \end{array}$$

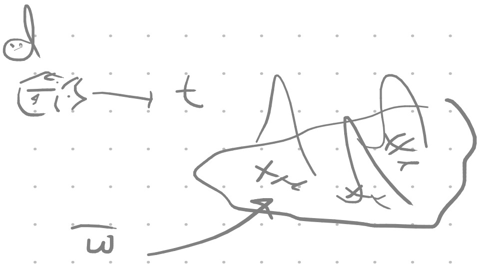
SVD

$$\boxed{T \ll V, N}$$

$$\Rightarrow$$

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



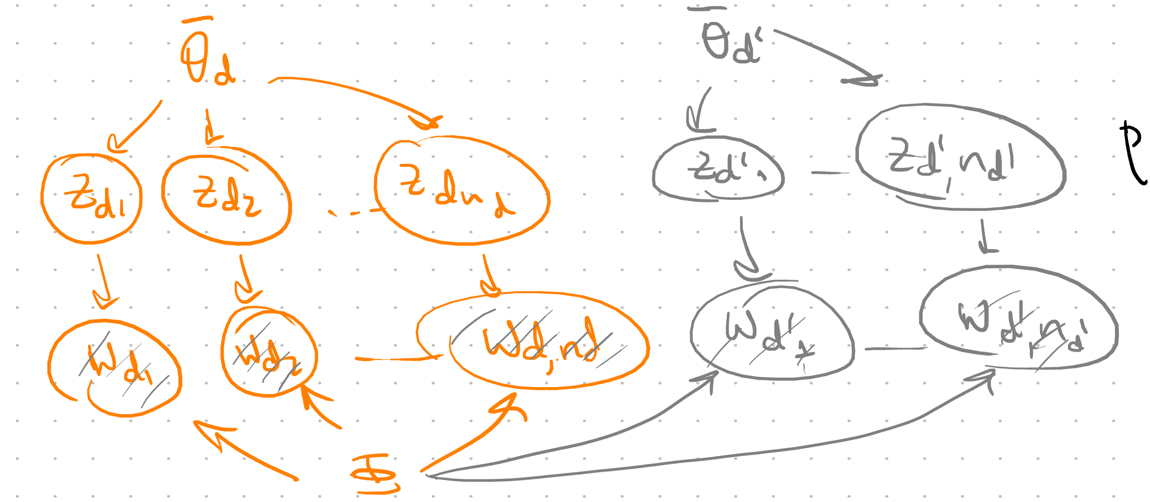


⑤ pLSI / LSA 1999/2000

prob. latent semantic indexing

latent semantic analysis

$$p(D|\Theta,\Phi) = \prod_{d=1}^{N} p(\bar{x}_d|\Theta,\Phi) = \prod_{d=1}^{N}\prod_{j=1}^{n_d}\left(\sum_{t=1}^{T}\theta_{td}\varphi_{w_{dj},t}\right) \xrightarrow{\Theta,\Phi} \max$$



$$p(w,z|\Theta,\Phi)$$

$z_{djt} = 1 \Leftrightarrow$ [в поз. j в док. d слово порождено из темы t] $\quad \bar{z}_{dj} = [\underset{1}{0} \_ 1 \_ \underset{T}{0}]$

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

$$Q(\Theta,\Phi,\Theta^{(m)},\Phi^{(m)}) = \mathbb{E}_{z|\Theta^{(m)},\Phi^{(m)}}\left[\log p(w,z|\Theta,\Phi)\right] =$$

$$= \sum_{d=1}^{N}\sum_{j=1}^{n_d}\sum_{t=1}^{T}\underline{\mathbb{E}_{(m)}[z_{djt}]}\cdot\left(\log\theta_{td} + \underline{\log\varphi_{w_{dj},t}}\right)$$

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

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

M-шаг: $\theta^{(m+1)}_{td} = \frac{\mathbb{E}[\# t \text{ в } d]}{n_d} = \frac{\sum_w n^{(m)}_{dwt}}{n_d} = \frac{n^{(m)}_{d*t} + 1}{n_{d**} + T}$

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

Regularization — $X \approx \underline{\Phi\cdot\Theta} = (\Phi\cdot A)(A^{-1}\cdot\Theta)$

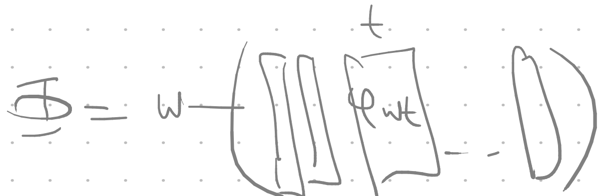
⑥ ARTM — additive regular. of topic models / Константин Воронцов

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

$\text{ReLU}(x) = \max(x,0)$

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

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

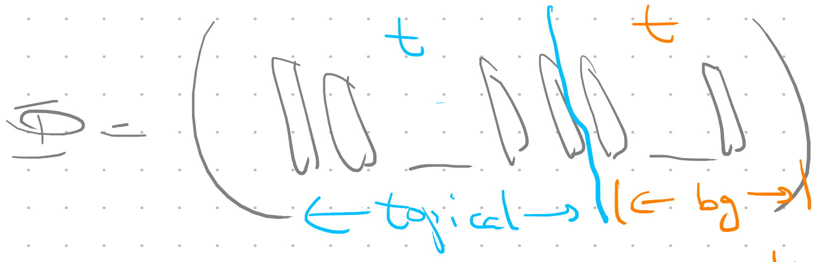


1) $\bar{\varphi}_{*t} \neq \text{Unif}$

$\underline{KL(\bar{\varphi}\,\|\,\text{Unif})}$, $\underline{KL(\text{Unif}\,\|\,\bar{\varphi})} \to \max$

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

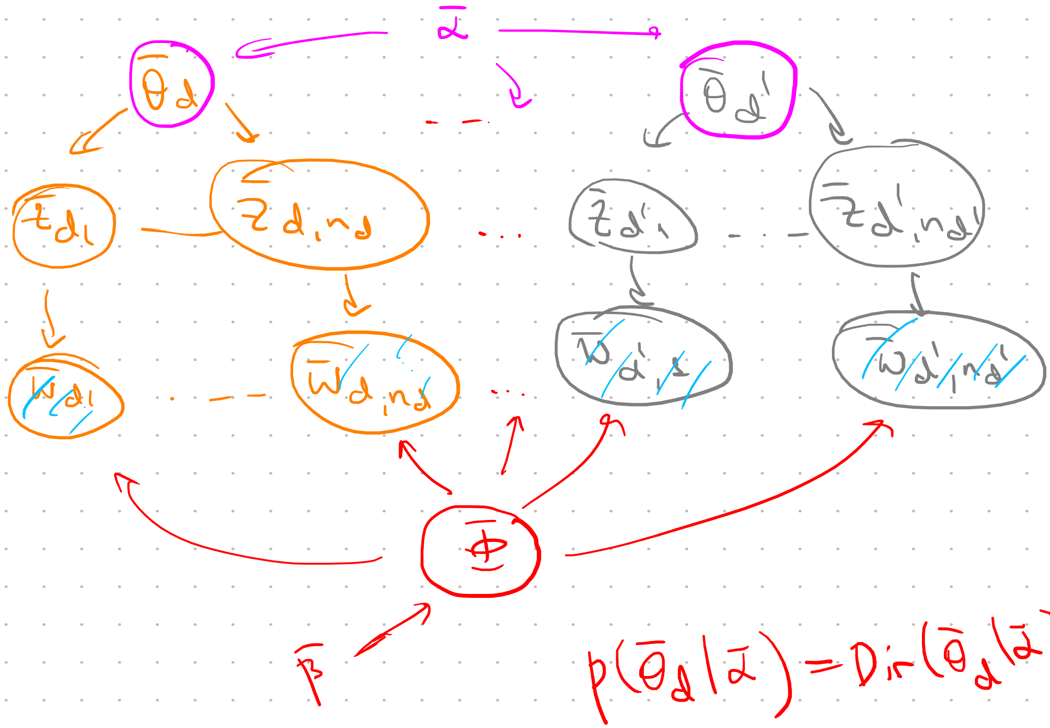
3)



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

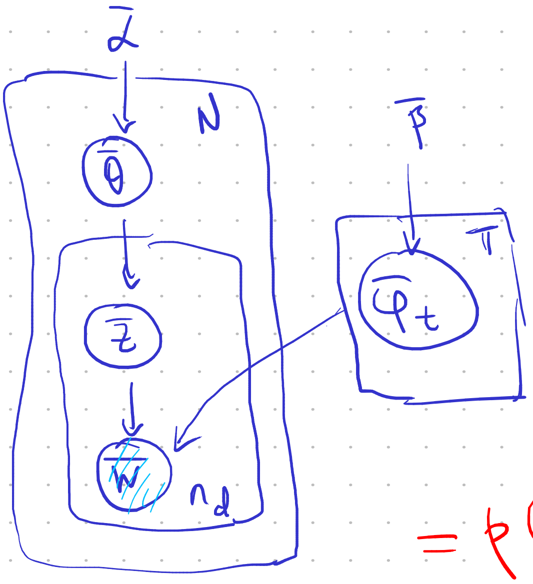
$KL(\bar{\varphi}_{*t} \| Unif) \to \min$

⑦ LDA – Latent Dirichlet Allocation



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

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



$$p(\Theta, \Phi | W, \bar{\alpha}, \bar{\beta}) \propto$$

$$\propto p(\Theta, \Phi | \bar{\alpha}, \bar{\beta})\, p(W|\Theta, \Phi) =$$

$$= p(\Theta|\bar{\alpha})\, p(\Phi|\bar{\beta}) \cdot \sum_{Z} p(W, Z | \Theta, \Phi)$$

$$p(W, Z, \Theta, \Phi | \bar{\alpha}, \bar{\beta}) = (\quad) \cdot p(Z|\Theta) \cdot p(X|Z, \Phi) =$$

$$= \left(\prod_{t=1}^{T} \overbrace{p(\bar{\varphi}_t|\bar{\beta})}^{Dir}\right) \cdot \prod_{d=1}^{N}\left(\overbrace{p(\bar{\theta}_d|\bar{\alpha})}^{Dir} \cdot \prod_{j=1}^{n_d} \overbrace{p(\bar{z}_{d_j}|\bar{\theta}_d)}^{Mult(\bar{\theta})} \cdot \overbrace{p(\bar{w}_{d_j}|z_{d_j}, \Phi)}^{Mult(\bar{\varphi}_t)}\right)$$

⑧ Variational approx. in LDA

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

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

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

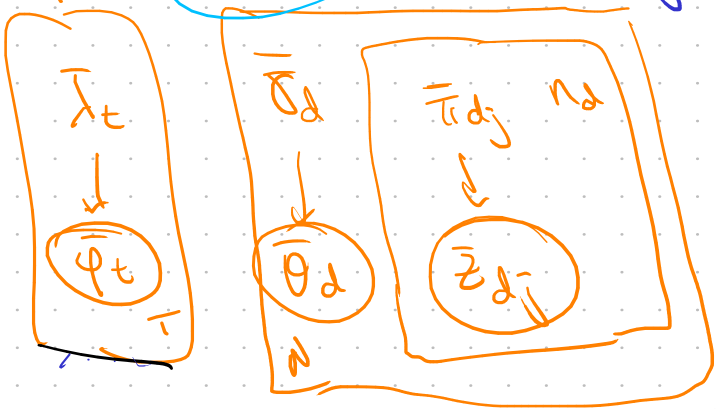
$$q(Z,\Theta,\Phi) = q(Z)\cdot q(\Theta,\Phi) =$$

$$= \left(\prod_{t=1}^{T}\underbrace{q_t(\bar\varphi_t|\bar\lambda_t)}_{\text{'' }Dir(\bar\varphi_t|\bar\lambda_t)}\right)\cdot\left(\prod_{d=1}^{N}\underbrace{q_d(\bar\theta_d|\bar\gamma_d)}_{\text{'' }Dir(\bar\theta_d|\bar\gamma_d)}\right)\times$$

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

$$= q(\ldots|\Lambda,\Gamma,\Pi)$$

$$\times\ \prod_{d=1}^{N}\prod_{j=1}^{n_d}\underbrace{q_{dj}(\bar z_{dj}|\bar\pi_{dj})}_{\text{'' }Mult(\bar z_{dj}|\bar\pi_{dj})}$$



$$p(\bar\theta_d|\bar\alpha)\cdot p(\bar z_d|\bar\theta_d)\cdot p(\bar w_d|\bar z_d,\Phi)$$

⑨ Φ fixed.

$$q(\Theta,Z|\Gamma,\Pi) \approx p(\Theta,Z|W,\Phi,\bar\alpha,\bar\beta)$$

$$L(q_d) = \mathbb{E}_{q_d}\left[\log\frac{p(\bar w_d,\bar z_d,\bar\theta_d|\Phi,\bar\alpha)}{q(\bar z_d,\bar\theta_d|\bar\gamma_d,\Pi_d)}\right]$$

$$q(\bar\theta_d|\bar\gamma_d)\cdot q(\bar z_d|\Pi_d)$$

$$L(q_d) = \underbrace{\mathbb{E}_q[\log p(\bar\theta_d|\bar\alpha)]}_{Dir}^{\boxed{I}} + \underbrace{\mathbb{E}_q[\log p(\bar z_d|\bar\theta_d)]}_{Mult}^{\boxed{II}} + \underbrace{\mathbb{E}_q[\log p(\bar w_d|\bar z_d, \Phi)]}_{Mult}^{\boxed{III}}$$

$$- \underbrace{\mathbb{E}_q[\log q(\bar\theta_d|\bar\gamma_d)]}_{Dir}^{\boxed{IV}} - \underbrace{\mathbb{E}_q[\log q(\bar z_d|\Pi_d)]}_{Mult}^{\boxed{V}}$$

$$Dir(\bar x|\bar\alpha) = \frac{1}{\underbrace{B(\bar\alpha)}} \prod x_i^{\alpha_i - 1} \qquad " \; \frac{\prod\Gamma(\alpha_t)}{\Gamma(\sum\alpha_t)}$$

$$Mult(\bar x|\bar\pi) = \prod \pi_k^{x_k}$$

$$\boxed{I} \quad \mathbb{E}_q[\log p(\bar\theta_d|\bar\alpha)] = \mathbb{E}_q\Big[-\log B(\bar\alpha) + \sum_{t=1}^{T}(\alpha_t - 1)\log\theta_{td}\Big] =$$

$$= \log\Gamma\Big(\sum\alpha_t\Big) - \sum_t \log\Gamma(\alpha_t) + \sum_{t=1}^{T}(\alpha_t - 1)\cdot\mathbb{E}_q[\log\theta_{td}] = (*)$$

$$\mathbb{E}_{Dir(\bar x|\bar\alpha)}[\log x_i] = ?$$

$$h(\bar x) = 1$$

$$\bar\eta = \bar\alpha - \bar 1$$

$$\boxed{T(\bar x) = \log \bar x}$$

$$\boxed{a(\bar\eta) = \log B(\bar\alpha)}$$

$$p = \underline{Dir(\bar x|\bar\alpha)} = e^{\sum_i(\alpha_i - 1)\log x_i - \log B(\bar\alpha)}$$

$$\mathbb{E}_p[T(\bar x)] = \nabla_{\bar\eta}\, a(\bar\eta)$$

$$\mathbb{E}_{Dir(\bar x|\bar\alpha)}[\log x_i] = \frac{\partial \log B(\bar\alpha)}{\partial\alpha_i} =$$

$$= \boxed{\frac{\partial\log\Gamma(\alpha_i)}{\partial\alpha_i}} - \frac{\partial\log\Gamma(\sum\alpha_j)}{\partial\alpha_i}$$

digamma function

$$\psi(x) = \frac{d\log\Gamma(x)}{dx}$$

$$\boxed{\mathbb{E}_{Dir(\bar x|\bar\alpha)}[\log\alpha_i] = \psi(\alpha_i) - \psi\Big(\sum_j\alpha_j\Big)}$$

$$\mathbb{E}_q[\log p(\bar\theta_d|\bar\alpha)] = \log\Gamma\Big(\sum_t\alpha_t\Big) - \sum_{t=1}^{T}\log\Gamma(\alpha_t) +$$

$$\boxed{I_*} \qquad + \sum_{t=1}^{T}(\alpha_t - 1)\cdot\Big(\psi(\gamma_{td}) - \psi\Big(\sum_{s=1}^{T}\gamma_{sd}\Big)\Big)$$

$$\boxed{II}\quad E_q[\log p(\bar z_d|\bar\theta_d)] = E_q\Big[\sum_{j=1}^{n_d} \log p(z_{dj}|\bar\theta_d)\Big] =$$

$$= E_q\Big[\sum_{j=1}^{n_d}\sum_{t=1}^{T} [z_{dj}=t]\cdot \log\theta_{td}\Big] =$$

$$q_d(\bar z, \bar\theta) = q_d(\bar z)\, q_d(\bar\theta)$$

$$= \sum_{j=1}^{n_d}\sum_{t=1}^{T} E_q[z_{dj}=t]\cdot E_q[\log\theta_{td}]$$

$$E_q[\log p(\bar z_d|\bar\theta_d)] = \sum_{j=1}^{n_d}\sum_{t=1}^{T} \pi_{djt}\cdot\Big(\psi(\gamma_{td}) - \psi\Big(\sum_{s=1}^{T}\gamma_{sd}\Big)\Big)$$

$$\boxed{III}\quad E_q[\log p(\bar w_d|\bar z_d, \Phi)] = E_q\Big[\sum_{j=1}^{n_d}\sum_{t=1}^{T}\sum_{v=1}^{V} [z_{dj}=t][w_{dj}=v]\cdot\log\varphi_{tv}\Big]$$

$$= \sum_{j=1}^{n_d}\sum_{t=1}^{T}\sum_{v=1}^{V} [w_{dj}=v]\cdot\pi_{djt}\cdot\log\varphi_{tv}$$

$$\boxed{IV}\quad E_q[\log q(\bar\theta_d|\bar\gamma_d)] = \log\Gamma\Big(\sum_s\gamma_{sd}\Big) - \sum_s\log\Gamma(\gamma_{sd}) +$$

$$+ \sum_{t=1}^{T}(\gamma_{td}-1)\Big(\psi(\gamma_{td}) - \sum_{s=1}^{T}\psi(\gamma_{sd})\Big)$$

$$\boxed{V}\quad E_q[\log q(\bar z_d|\bar\pi_d)] = E_q\Big[\sum_{j=1}^{n_d}\sum_{t=1}^{T}[z_{dj}=t]\cdot\log\pi_{djt}\Big]$$

$$= \sum_{j=1}^{n_d}\sum_{t=1}^{T}\pi_{djt}\log\pi_{djt}$$

$$L(\bar\gamma_d, \Pi_d) = I + II + III - IV - V \xrightarrow{\bar\gamma_d,\, \Pi_d} \max$$

при усл. $\sum_t \pi_{djt} = 1 \quad \forall j$

$\pi_{djt} \geq 0$