① Var. approx. in LDA $\quad p(Z,\Theta,\Phi \mid W,\alpha,\beta) \approx q(Z,\Theta,\Phi)$

$$q(Z,\Theta,\overset{\text{fixed}}{\Phi}) = \Big(\prod_t \underbrace{q_t(\bar\varphi_t \mid \lambda_t)}_{\text{Dir}}\Big)\Big(\prod_d \underbrace{q_d(\bar\theta_d \mid \bar\gamma_d)}_{\text{Dir}}\Big)\Big(\prod_d \prod_j \underbrace{q_{dj}(\bar z_{dj} \mid \bar\pi_{dj})}_{\text{mult}}\Big)$$

$$L(q_d) = E_q[\log p(\bar\theta_d \mid \bar\alpha)] = \log\Gamma\Big(\sum \alpha_t\Big) - \sum \log\Gamma(\alpha_t) + \underline{\sum_t (\alpha_t - 1)\Big(\psi(\gamma_{td}) - \psi\Big(\sum_s \gamma_{sd}\Big)\Big)}$$

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

$$+ E_q[\log p(\bar w_d \mid \bar z_d, \Phi)] = \sum_j \sum_t \sum_{v=1}^{V} [w_{dj} = v] \cdot \pi_{djt} \log \varphi_{tv}$$

$$- E_q[\log q(\bar\theta_d \mid \bar\gamma_d)] = \underline{\log\Gamma\Big(\sum_s \gamma_{sd}\Big) - \sum_s \log\Gamma(\gamma_{sd})} + \underline{\sum_t (\gamma_{td} - 1)\Big(\psi(\gamma_{td}) - \psi\Big(\sum_s \gamma_{sd}\Big)\Big)}$$

$$- E_q[\log q(\bar z_d \mid \bar\pi_d)] = \sum_j \sum_t \pi_{djt} \log \pi_{djt}$$

$\gamma_d, \Pi_d \to \max, \quad \underline{\sum_t \pi_{djt} = 1}, \quad \pi_{djt} \geq 0$

$$L(\bar\pi_{dj}) = \sum_t \pi_{djt}\Big(\psi(\gamma_{td}) - \psi\Big(\sum_s \gamma_{sd}\Big)\Big) + \sum_t \sum_v [w_{dj} = v]\, \pi_{djt} \log \varphi_{tv} - \sum_t \underline{\pi_{djt} \log \pi_{djt}} + \mu_{dj}\underline{\Big(\sum_s \pi_{djs} - 1\Big)}$$

$$\frac{\partial L}{\partial \pi_{djt}} = \psi(\gamma_{td}) - \psi\Big(\sum_s \gamma_{sd}\Big) + \log \varphi_{t,w_{dj}} - \log \pi_{djt} - 1 + \mu_{dj} = 0$$

$$\underline{\pi_{djt}} = e^{\psi(\gamma_{td}) - \psi(\sum_s \gamma_{sd}) + \log \varphi_{t,w_{dj}} - 1 + \mu_{dj}}$$

$$\boxed{\pi^*_{djt} \propto e^{\psi(\gamma_{td}) - \psi(\sum_s \gamma_{sd}) + \log \varphi_{t,w_{dj}}}}$$

$$L(\tilde{\gamma}_d) = \sum_t (\alpha_t - 1)\left(\psi(\gamma_{td}) - \psi\left(\sum_s \gamma_{sd}\right)\right) +$$

$$+ \sum_j \sum_t \pi_{djt} (\,\cdot\,) - \sum_t (\gamma_t - 1)(\,\cdot\,)$$

$$- \log\Gamma\left(\sum_t \gamma_{td}\right) + \sum_t \log\Gamma(\gamma_{td}) =$$

$$= \sum_t \left(\psi(\gamma_{td}) - \psi\left(\sum_s \gamma_{sd}\right)\right) \cdot \left(\alpha_t + \sum_j \pi_{djt} - \gamma_t\right) -$$

$$- \log\Gamma\left(\sum_t \gamma_{td}\right) + \sum_t \log\Gamma(\gamma_{td})$$

$$\frac{\partial L}{\partial \gamma_{jt}} = -\psi'\left(\sum_s \gamma_{sd}\right) \cdot \sum_s \left(\alpha_s + \sum_j \pi_{djs} - \gamma_s\right) +$$

$$+ \psi'(\gamma_{td})\left(\alpha_t + \sum_j \pi_{djt} - \gamma_t\right)$$

$$- \left(\cancel{\psi(\gamma_{td})} - \cancel{\psi\left(\sum_s \gamma_{sd}\right)}\right) - \cancel{\psi\left(\sum_s \gamma_{sd}\right)} + \cancel{\psi(\gamma_{td})}$$

$$= 0$$

$\forall t$

$$\psi'(\gamma_{td})\left(\alpha_t + \sum_j \pi_{djt} - \gamma_t\right) = \psi'\left(\sum_s \gamma_{sd}\right) \sum_s \left(\alpha_s + \sum_j \pi_{djs} - \gamma_s\right)$$

$$\boxed{\gamma_{td}^* = \alpha_t + \sum_{j=1}^{n_d} \pi_{djt}}$$

EM-algorithm w.r.t Φ:

- $\underline{\text{E-step}}$: $\Phi^{(m)}$ fixed
  - repeat until convergence:

    $$\gamma_{td} := \alpha_t + \sum_{j=1}^{n_d} \pi_{djt}$$

    $$\pi_{djt} \propto e^{\psi(\gamma_{td}) - \psi(\sum_s \gamma_{sd}) + \log \varphi^{(m)}_{t, w_{dj}}}$$

- $\underline{\text{M-step}}$:

  $$\varphi_{tv} \propto \sum_{d=1}^{N} \sum_{j=1}^{n_d} [w_{dj} = v] \cdot \pi_{djt}$$

② Φ not fixed

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

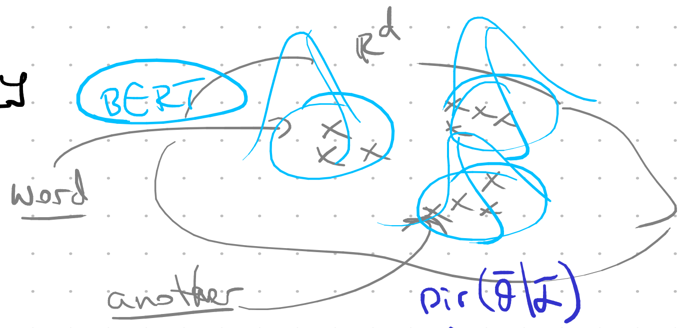
$$= \sum_{d=1}^{N} L_d(\bar{\gamma}_d, \Pi_d) + \sum_t \mathbb{E}_q[\log p(\bar{\varphi}_t\mid\bar{\beta})] - \sum_t \mathbb{E}_q[\log q(\bar{\varphi}_t\mid\bar{\lambda}_t)]$$

$$\boxed{\lambda^*_{tv} = \beta_v + \sum_{d=1}^{N}\sum_{j=1}^{n_d}[w_{dj}=v]\,\pi_{djt}}$$

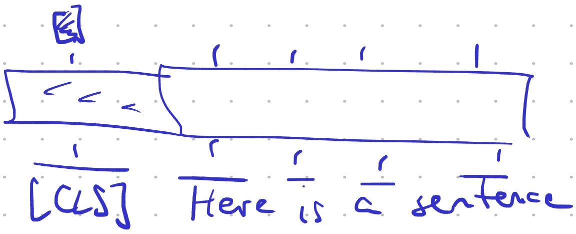
$-\pi$
$-\gamma$
$-\lambda$

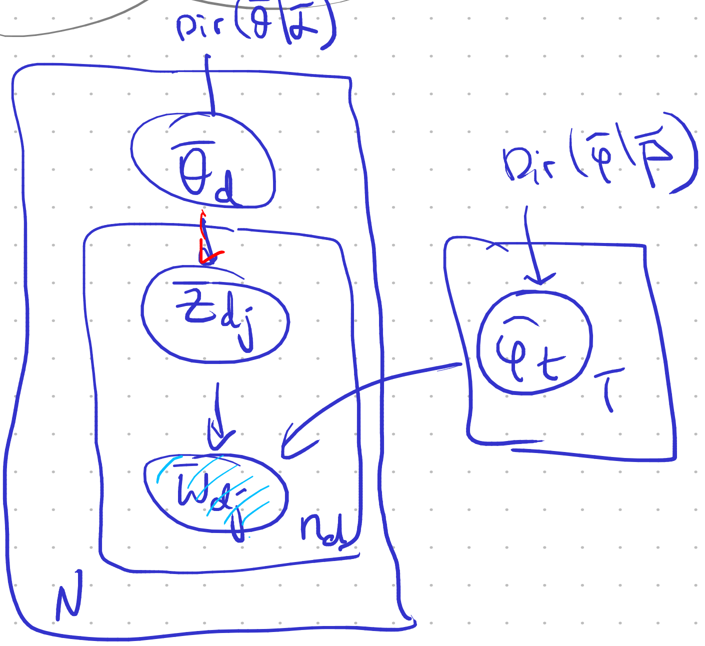
③ Neural topic modeling

word2vec



RoBERTa





④ Gibbs sampling for LDA

$$p(\underbrace{\Theta,\Phi,Z}\mid W,\bar{\alpha},\bar{\beta})$$

$$p(\bar{\theta}_d\mid\Theta_{-d},\Phi,Z,W,\bar{\alpha},\bar{\beta}) \propto$$

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

$$\propto p(\bar{\theta}_d\mid\bar{\alpha})\cdot\prod_{j=1}^{n_d} p(\bar{z}_{dj}\mid\bar{\theta}_d) = \mathrm{Dir}(\bar{\theta}_d\mid\bar{\alpha})\cdot\prod_j \mathrm{Mult}(\bar{z}_{dj}\mid\bar{\theta}_d)$$

$$\propto \prod_t \theta_{td}^{\alpha_t-1}\cdot\prod_{j=1}^{n_d}\prod_{t=1}^{T}\theta_{td}^{z_{djt}}$$

$$p(\bar{\theta}_d\mid\ldots) = \mathrm{Dir}\left(\bar{\theta}_d\mid\bar{\alpha}+\sum_{j=1}^{n_d}\bar{z}_{dj}\right)$$

$$\bar\varphi_t \sim p(\bar\varphi_t|\ldots) \propto \mathrm{Dir}(\bar\varphi_t|\bar\beta)\cdot \prod_{d=1}^{N}\prod_{j=1}^{n_d}\prod_{t=1}^{T} \mathrm{Mult}(\bar w_{dj}|\bar\varphi_t)^{z_{djt}}$$

$$\propto \prod_v \varphi_{tv}^{\beta_v-1} \cdot \prod_d \prod_j \prod_t \cdot \prod_v \varphi_{tv}^{z_{djt}\cdot w_{djv}}$$

$$p(\bar\varphi_t|-) = \mathrm{Dir}(\bar\varphi_t|\bar\beta + \sum_d \sum_j \sum_t \bar z_{dj}\odot \bar w_{dj})$$

$$\bar z_{dj} \sim p(\bar z_{dj}|z_{-dj}, \underbrace{\Theta, \Phi}, --)$$

$$p(z_{djt}=1|\bar\theta_d, \Phi, \bar w_{dj}) = \frac{p(z_{djt}=1, \bar w_{dj}|-)}{p(\bar w_{dj}|-)} = \frac{\theta_{dt}\varphi_{t,w_{dj}}}{\sum_s \theta_{ds}\varphi_{s,w_{dj}}}$$

5 Collapsed Gibbs sampling

$$\underbrace{\bar\theta_1 - \bar\theta_N}, \underbrace{\bar\varphi_1 - \bar\varphi_T}, \bar z_{11} - - \bar z_{N,n_N}$$

$$\bar z_{dj} \sim p(\bar z_{dj}|z_{-dj}, W, \bar\alpha, \bar\beta) \propto p(z, W|\bar\alpha, \bar\beta) = \overbrace{p(z|\bar\alpha)}^{\Theta}\overbrace{p(W|z, \bar\beta)}^{\Phi}$$

$$p(z|\bar\alpha) = \int p(z|\Theta)\, p(\Theta|\bar\alpha)\, d\Theta =$$

$$= \int \left(\prod_{d=1}^{N} p(\bar\theta_d|\bar\alpha)\right)\cdot\left(\prod_d\prod_j p(\bar z_{dj}|\bar\theta_d)\right) d\Theta =$$

$$= \int\left(\prod_{d=1}^{N}\frac{1}{B(\bar\alpha)}\cdot\prod_t \theta_{td}^{\alpha_t-1}\right)\cdot\left(\prod_{d=1}^{N}\prod_{j=1}^{n_d}\prod_{t=1}^{T}\theta_{td}^{z_{djt}}\right) d\Theta =$$

$$= \frac{1}{B(\bar\alpha)^N}\cdot\int \prod_{d=1}^{N} \underbrace{\prod_{t=1}^{T} \theta_{td}^{\alpha_t + \sum_j z_{djt} - 1}}_{\propto\, \mathrm{Dir}(\bar\theta_d|\bar\alpha + \sum_j \bar z_{dj})} d\Theta = \frac{1}{B(\alpha)^N}\prod_{d=1}^{N} B(\bar\alpha + \sum_j \bar z_{dj})$$

$\leftarrow B(--)$

$$\boxed{p(z|\bar\alpha) = \prod_{d=1}^{N}\frac{B(\bar\alpha + \sum_j \bar z_{dj})}{B(\bar\alpha)}}$$

$\leftarrow \bar\sigma_d$

$$p(W|Z,\bar\beta)=\int\left(\prod_{t=1}^{T}\frac{1}{B(\bar\beta)}\cdot\prod_{v=1}^{V}\varphi_{tv}^{\beta_v-1}\right)\left(\prod_d\prod_j\prod_t\prod_v \varphi_{tv}^{z_{djt}\cdot w_{djv}}\right)d\Phi$$

$$=\frac{1}{B(\bar\beta)^T}\int\prod_{t=1}^{T}\left(\prod_{v=1}^{V}\varphi_{tv}^{\beta_v+\sum_d\sum_j z_{djt}\cdot w_{djv}-1}\right)d\Phi$$

$$\boxed{p(W|Z,\bar\beta)=\prod_{t=1}^{T}\frac{B(\bar\beta+\underbrace{\sum_d\sum_j z_{djt}\cdot\bar w_{dj}}_{\bar\delta_t})}{B(\bar\beta)}}$$

$$p(Z,W|\bar\alpha,\bar\beta)=\left(\prod_{d=1}^{N}\frac{B(\bar\alpha+\bar\sigma_d)}{B(\bar\alpha)}\right)\cdot\left(\prod_{t=1}^{T}\frac{B(\bar\beta+\bar\delta_t)}{B(\bar\beta)}\right)$$

$$p(z_{djt}|Z_{-dj},W,\bar\alpha,\bar\beta)\propto p(Z,W|\bar\alpha,\bar\beta)\propto$$

$$\propto B(\bar\alpha+\bar\sigma_d)\cdot B(\bar\beta+\bar\delta_t)=\frac{\prod_s\Gamma(\alpha_s+\sigma_{ds})}{\Gamma\left(\sum_s(\alpha_s+\sigma_{ds})\right)}\cdot\frac{\prod\Gamma(\beta_v+\delta_{tv})}{\Gamma\left(\sum_v(\beta_v+\delta_{tv})\right)}$$

$$\sigma_{ds}=\sum_{j=1}^{n_d}z_{djs}$$

$$\sigma_{ds}^{-dj}=\begin{cases}\sigma_{ds}-1, & \text{если } z_{djs}=1\\ \sigma_{ds}\end{cases}$$

$$\Gamma(\ldots)\cdot(\ldots)$$

$$\prod_s\Gamma(\alpha_s+\sigma_{ds})=\prod_{s\neq t}\Gamma(\alpha_s+\sigma_{ds}^{-dj})\cdot\Gamma(\alpha_t+\sigma_{dt}^{-dj}+1)=$$

$$=\left(\cancel{\prod_s\Gamma(\alpha_s+\sigma_{ds}^{-dj})}\right)\cdot(\alpha_t+\sigma_{dt}^{-dj})$$

не зависит от z_{djt}

$\cancel{\Gamma\left(\sum_s(\alpha_s+\sigma_{ds})\right)}$ – не зав. от z_{djt}

$$\delta_{tv}^{-dj} = \begin{cases} \delta_{tv}-1, & \text{если } z_{djt}=1, w_{djv}=1 \\ \delta_{tv} \end{cases}$$

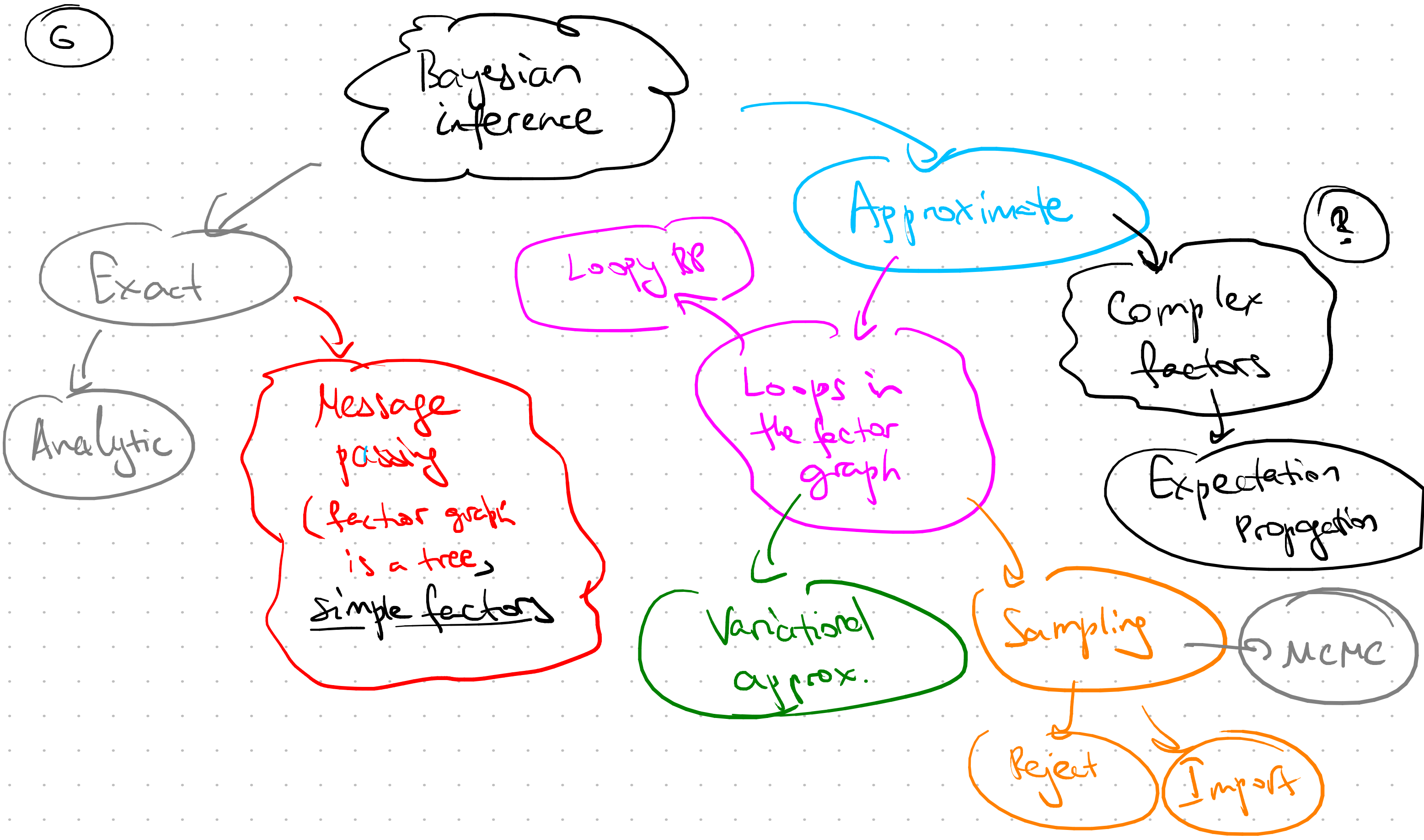
$$\prod_v \Gamma(\beta_v+\delta_{tv}) = \prod_v \Gamma(\beta_v+\delta_{tv}^{-dj}) \cdot (\beta_{w_{dj}} + \delta_{t,w_{dj}})$$

$$\Gamma\left(\sum_v(\beta_v+\delta_{tv})\right) = \Gamma\left(\sum_v(\beta_v+\delta_{tv}^{-dj})+1\right) = \Gamma\left(\sum_v(\_\_)\right)\cdot\left(\sum_v(\beta_v+\delta_{tv}^{-dj})\right)$$

$$\boxed{p(z_{djt}|\_\_\_) \propto (\alpha_t + \gamma_{dt}^{-dj}) \cdot \frac{\beta_{w_{dj}}+\delta_{t,w_{dj}}^{-dj}}{\sum_v(\beta_v+\delta_{tv}^{-dj})}}$$

$$\theta_{td} = \frac{\alpha_t + \sum_j z_{djt}}{\sum_s(\alpha_s+\sum_j z_{djs})}, \quad \varphi_{tv} = \frac{\beta_v + \sum_d\sum_j z_{djt}[w_{djv}=1]}{\sum_u\left(\beta_u+\sum_d\sum_j z_{djt}[w_{dju}=1]\right)}$$

6



7 Moment matching

$$q(\bar{x}) \approx p(\bar{x})$$

$$= h(\bar{x})\cdot e^{-\bar{\eta}^T \bar{u}(\bar{x}) + a(\bar{\eta})}$$

$$KL(p\|q) \xrightarrow[\bar{\eta}]{} \min$$

$$KL(p\|q) = Const - \int p(\bar{x})\cdot\left(\cancel{\log h(\bar{x})} - \bar{\eta}^T u(\bar{x}) + a(\bar{\eta})\right) d\bar{x} =$$

$$\int p \log \frac{p}{q} dx \quad = Const + \bar{\eta}^T \int p(\bar{x}) u(\bar{x}) d\bar{x} - a(\bar{\eta}) \quad \xrightarrow[\bar{\eta}]{} \min$$

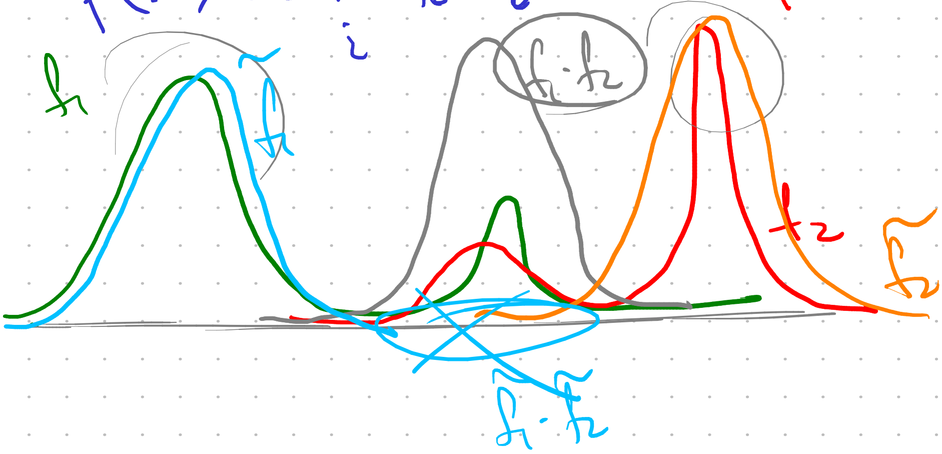
$$\nabla_{\bar{\eta}}(KL) = \mathbb{E}_{p(x)}[\bar{u}(x)] - \underbrace{\nabla_{\bar{\eta}} a(\bar{\eta})}_{= \mathbb{E}_{q(\bar{x})}[\bar{u}(\bar{x})]} = 0$$

$$KL(p\|q) \to \min \iff \boxed{\mathbb{E}_{q(\bar{x})}[\bar{u}(\bar{x})] = \mathbb{E}_{p(x)}[\bar{u}(\bar{x})]}$$

$$\begin{pmatrix} \tau \\ \mu\tau \end{pmatrix}$$

8 Expectation Propagation

$$p(\bar{x}) \propto \prod_i f_i(\bar{x}_i) \approx q(\bar{x}) = \frac{1}{Z}\prod_i \tilde{f}_i(\bar{x})$$



$$KL(p\|q) = KL\left(\frac{1}{Z'}\prod_{i=1}^{n} f_i(\bar{x}_i) \,\Big\|\, \frac{1}{Z}\prod_{i=1}^{n} \tilde{f}_i(\bar{x}_i)\right) \xrightarrow[\tilde{f}_i]{} \min$$

EP - init $\tilde{f}_j^{(0)}(\bar{x}_j) \approx f_j(\bar{x}_j)$

- for $j = 1, 2, \ldots, n, 1, 2, \ldots, n$
  - find $\tilde{f}_j(\bar{x}_j)$:

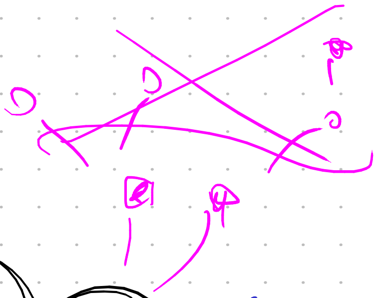
$$\tilde{f}_j'(\bar{x}_j) = \underset{\tilde{f}_j}{\operatorname{argmin}}\, KL\left(\frac{1}{Z_j} f_j(\bar{x}_j)\prod_{i\neq j}\tilde{f}_i(\bar{x}_i) \,\Big\|\, \overbrace{\tilde{f}_j(\bar{x}_j)\prod_{i\neq j}\tilde{f}_i(\bar{x}_i)}^{q(\bar{x})}\right)$$

1) $q^{-j}(\bar{x}) = \prod_{i \neq j} \hat{f}_i(\bar{x}_i) = \dfrac{q(\bar{x})}{\hat{f}_j(\bar{x})}$

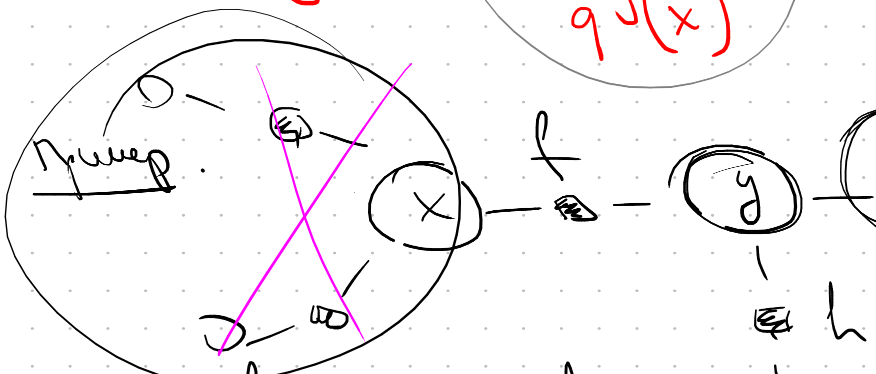
moment matching

2) $q^{new}(\bar{x}) = \arg\min KL\left(\frac{1}{Z_j} f_j(\bar{x}_j) \cdot q^{-j}(\bar{x}) \,\|\, q^{new}(\bar{x})\right)$

3) $\hat{f}_j(\bar{x}) = \dfrac{q^{new}(\bar{x})}{q^{-j}(\bar{x})}$



Пример.



$p(x_1 \ldots x_4) = f(x,y)\, g(y,z) \cdot h(y,t)$

$\tilde{q}(x_1 \ldots x_4) = \hat{f}(x,y)\, \hat{g}(y,z)\, \hat{h}(y,t) = \tilde{f}_x(x)\, \tilde{f}_y(y)\, \tilde{g}_y(y)\, \tilde{g}_z(z)\, \tilde{h}_y(y)\, \tilde{h}_t(t)$

$\tilde{g}(y,z) = \tilde{g}_y \cdot \tilde{g}_z$

$q^{-g}(x,y,z,t) = \tilde{f}_x \tilde{f}_y \tilde{h}_y \tilde{h}_t$

$q^{new} = \arg\min KL\left(g(y,z) \cdot q^{-g} \,\|\, q^{new}(x,y,z,t)\right)$

$KL(p \,\|\, q = \prod q_i) = \int p \log p \, d\bar{x} - \sum_i \int p \cdot \log q_i \, d\bar{x}$ → max

$q_i(\bar{x}_i) = \int p(\bar{x})\, d\bar{x}_{-i}$

$q^{new}_x(x) = \int g(y,z)\, \tilde{f}_x \tilde{f}_y \tilde{h}_y \tilde{h}_t \, dy\, dz\, dt \propto \tilde{f}_x(x)$

$q^{new}_t(t) = \ldots \propto \tilde{h}_t(t)$

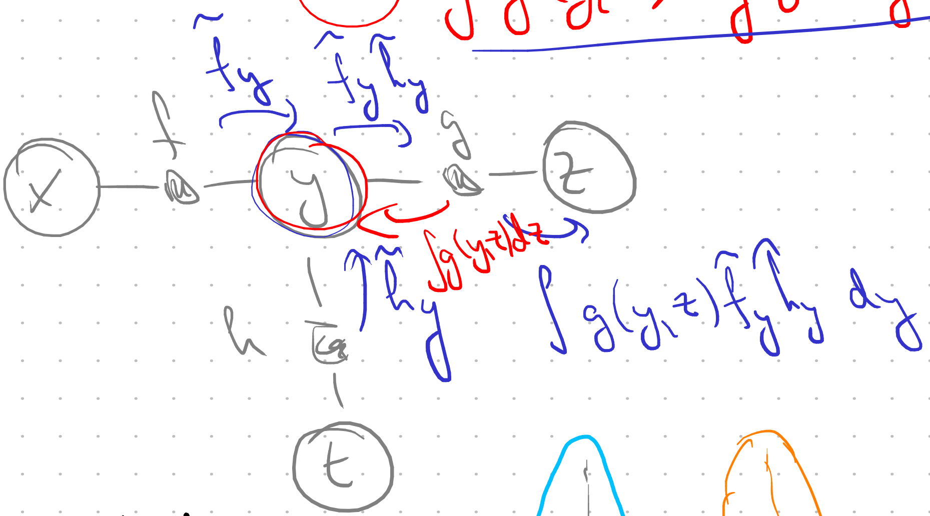
$$q_y^{new}(y) = \int g(y,z)\, \tilde f_x \tilde f_y \hat h_y \hat h_t \, dx\,dz\,dt \propto$$

$$\propto \tilde f_y \cdot \hat h_y \cdot \int g(y,z)\,dz$$

$$q_z^{new}(z) = \int g(y,z)\, \tilde f_x \tilde f_y \hat h_y \hat h_t \, dx\,dy\,dt \propto$$

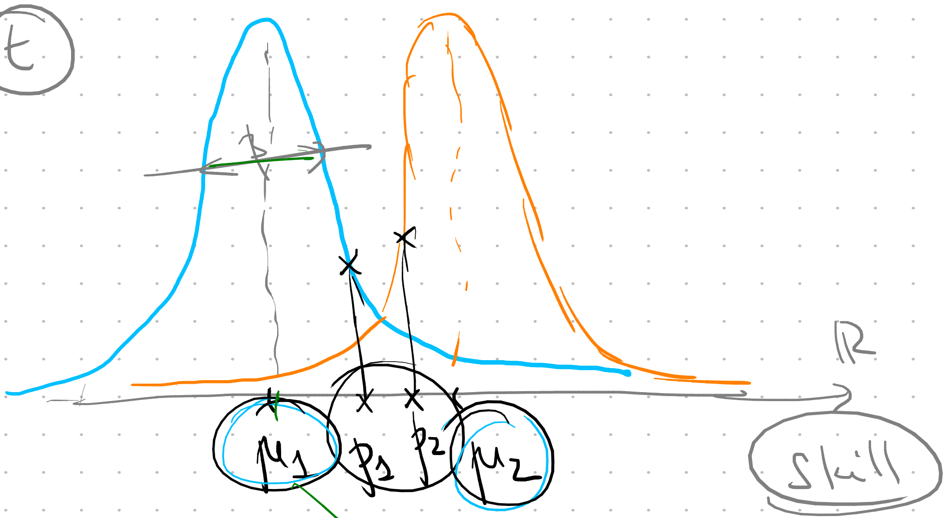
≈

$$\propto \int g(y,z)\, \tilde f_y(y)\, \hat h_y(y)\, dy$$



(9) TrueSkill

Elo rating

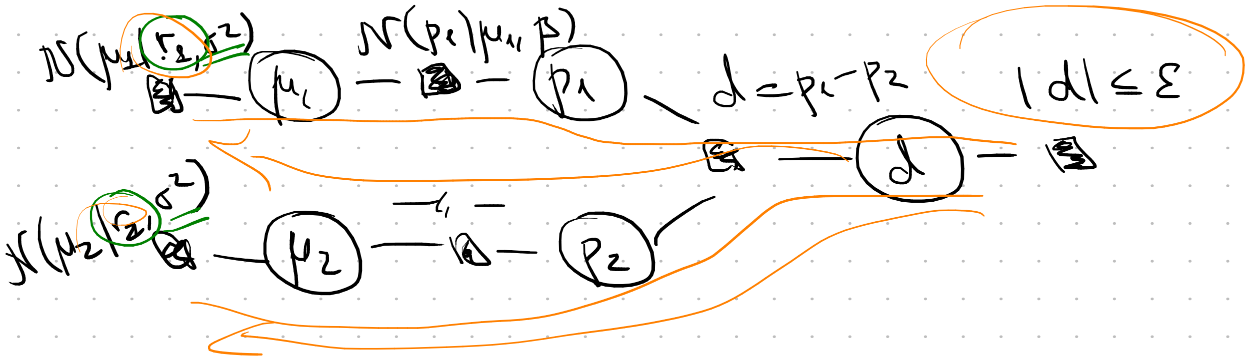


$$\begin{cases} p_1 > p_2 + \varepsilon & 1 \\ |p_1 - p_2| \le \varepsilon & 1/2 \\ p_1 + \varepsilon < p_2 & 0 \end{cases}$$

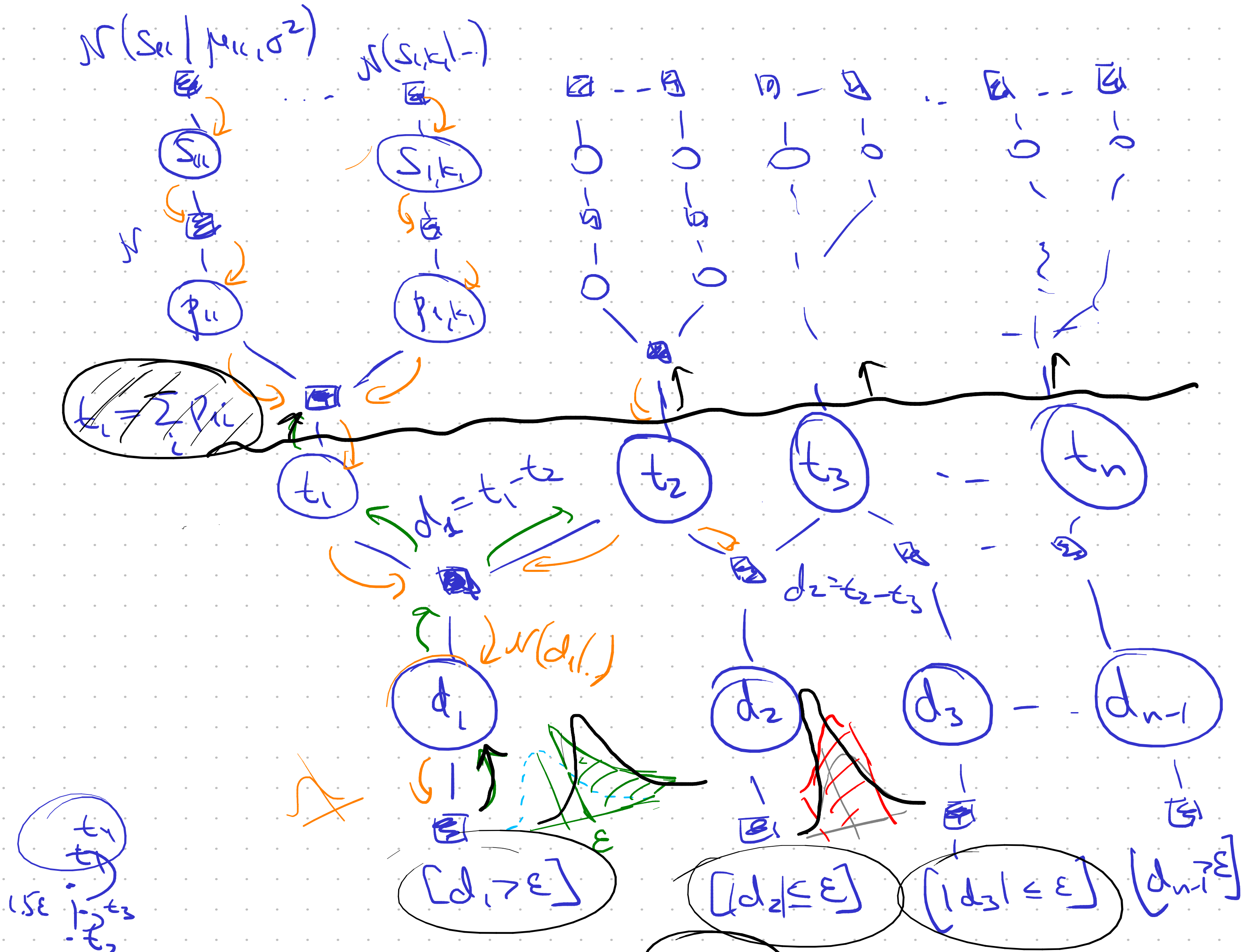
$$p(\mu_1 \mid \text{ничья } 1 \text{ и } 2) \propto p(\mu_1) \cdot p(\text{ничья } 1 \text{ и } 2 \mid \mu_1, \mu_2) =$$

$$= p(\mu_1) \cdot \int_{|p_1 - p_2| \le \varepsilon} \overbrace{p(p_1 \mid \mu_1)}^{\mathcal{N}(p_1 \mid \mu_1, \beta)} \cdot \overbrace{p(p_2 \mid \mu_2)}^{\mathcal{N}(p_2 \mid \mu_2, \beta)} \, dp_1\, dp_2$$

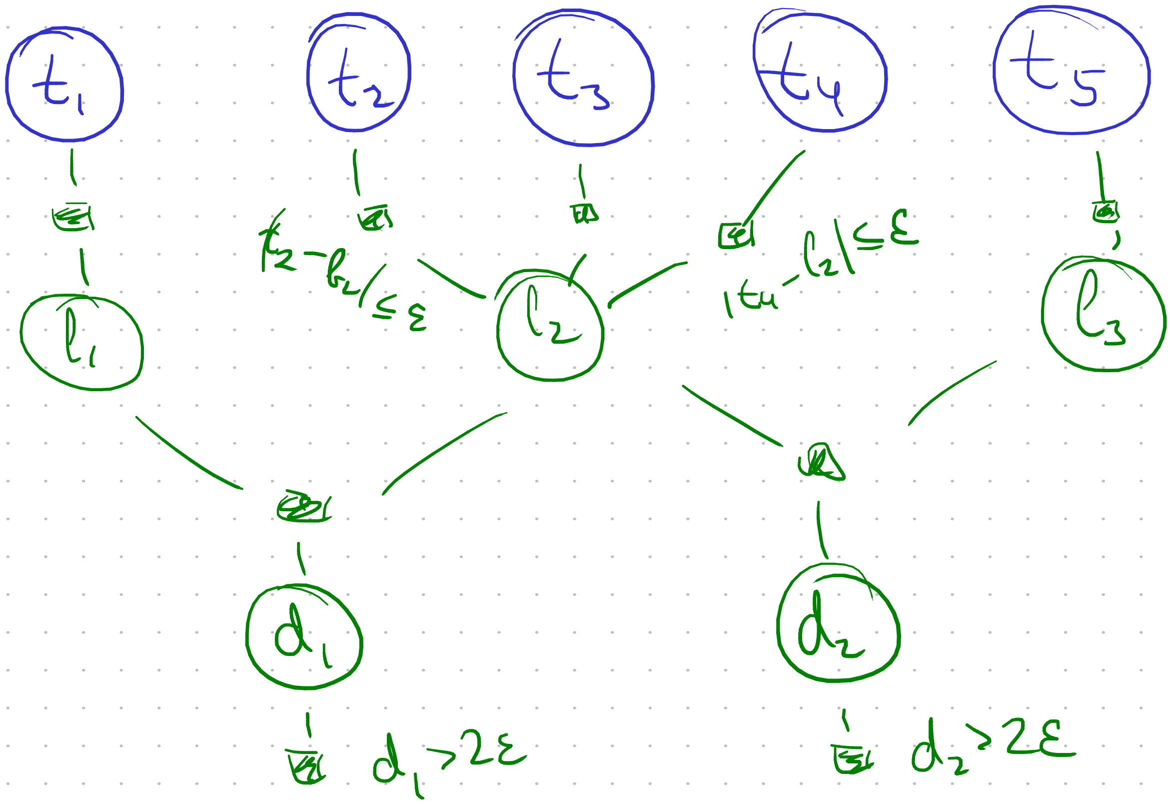
$$\| \quad \mathcal{N}(\mu_1 \mid .., \sigma^2)$$

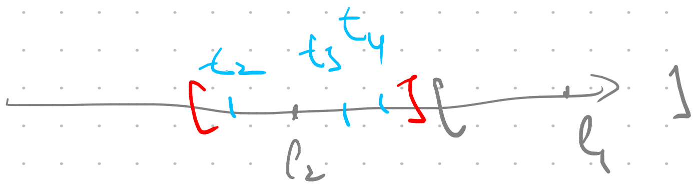


t_1 $\quad \underline{\mu_{11}} - \underline{\mu_{1,k_1}}$
t_2
$\vdots$
t_n $\quad \underline{\mu_{n1}} - \underline{\mu_{n,k_n}}$



- $t_1 > t_2 + \varepsilon$ $\qquad t_4 \ge t_1$
- $|t_2 - t_3| \le \varepsilon$ $\qquad |t_4 - t_2| \le 2\varepsilon$?
- $|t_3 - t_4| \le \varepsilon$


t_1
t_2
t_3
t_4
t_5
$|t_2 - l_2| \leq \varepsilon$
$|t_4 - l_2| \leq \varepsilon$
l_1
l_2
l_3
d_1
d_2
$d_1 > 2\varepsilon$
$d_2 > 2\varepsilon$


t_2
t_3
t_4
l_2
l_1