① Другой взгляд на EM-алгоритм

observ latent params

$p(X|\bar\theta) \longrightarrow \max$ $\qquad$ $\underbrace{p(X|\bar\theta)}_{\text{сложное}} = \int \underbrace{p(X, Z|\bar\theta)}_{\text{простое}}\, dZ$

$$p(X, Z|\bar\theta) = \underline{p(X|\bar\theta)} \cdot p(Z|X, \bar\theta)$$

$$\log p(X|\bar\theta) = \log p(X, Z|\bar\theta) - \log p(Z|X, \bar\theta)$$

$\mathbb{E}_{Z \sim p(Z|X, \bar\theta^{(m)})}$ $\qquad$ $\int \log p \, q \, d\bar x$ $\qquad$ $\int \log \frac{q}{p} q \, dx$

$$\log p(X|\bar\theta) = \mathbb{E}_Z\left[\log p(X, Z|\bar\theta) - \log p(Z|X, \bar\theta)\right] =$$

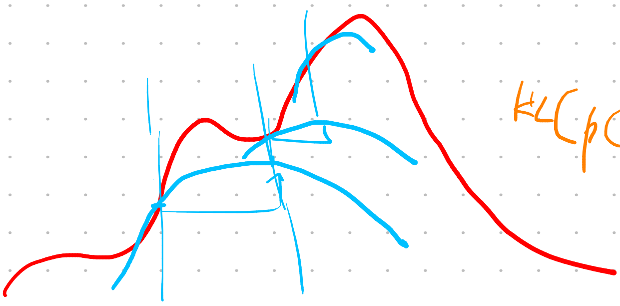
$$= \int \log p(X, Z|\bar\theta)\, p(Z|X, \bar\theta^{(m)})\, dZ - \int \log p(Z|X, \bar\theta)\, p(Z|X, \bar\theta^{(m)})\, dZ$$

$\bar\theta$ max

$$\pm \int \log p(Z|X, \bar\theta^{(m)})\, p(Z|X, \bar\theta^{(m)})\, dZ$$

$$= \int \log p(X, Z|\bar\theta)\, p(Z|X, \bar\theta^{(m)})\, dZ + \int \log \frac{p(Z|X, \bar\theta^{(m)})}{p(Z|X, \bar\theta)}\, p(Z|X, \bar\theta^{(m)})\, dZ$$

$$\underbrace{- \int \log p(Z|X, \bar\theta^{(m)})\, p(Z|X, \bar\theta^{(m)})\, dZ}_{\text{Const}}$$



$$KL\left(p(Z|X, \bar\theta^{(m)}) \,\|\, p(Z|X, \bar\theta)\right) \geq 0$$

$$= 0 \text{ if } \bar\theta = \bar\theta^{(m)}$$

$$\log p(X|\bar\theta) \geq \underbrace{\int \log p(X, Z|\bar\theta)\, p(Z|X, \bar\theta^{(m)})\, dZ}_{Q(\bar\theta, \bar\theta^{(m)})} + \text{Const}$$

$(=)$ if $\bar\theta = \bar\theta^{(m)}$ $\qquad$ $Q(\bar\theta, \bar\theta^{(m)}) \xrightarrow[\bar\theta]{} \max$

② Variational approximation

observables latent + params posterior

$$p(X, Z) \qquad p(Z|X) = ?$$

$D \qquad \bar\theta, Z$

$$p(X, Z) = p(X)\, p(Z|X)$$

$$\underbrace{\log p(X)}_{\text{Const}} = \log p(X, Z) - \log \underbrace{p(Z|X)}_{\text{сложное}}$$

$\mathbb{E}_{q(Z)}$

$$\log p(X) = \mathbb{E}_Z\left[\log p(X, Z)\right] - \mathbb{E}_Z\left[\log p(Z|X)\right] \pm \mathbb{E}_{q(Z)}\left[\log q(Z)\right]$$

$$\log p(X) = \int \log \frac{p(X, Z)}{q(Z)}\, q(Z)\, dZ + \int \log \frac{q(Z)}{p(Z|X)}\, q(Z)\, dZ$$

$$\text{Const} = \log p(X) = \mathcal{L}(q) + KL\left(q(Z) \,\|\, p(Z|X)\right)$$

$\log p(x)$

$\updownarrow \; KL(q(z) \| p(z|x)) \xrightarrow{q} \min$

$\Updownarrow$

$\updownarrow \; L(q) \xrightarrow{q} \max$

$\underline{q(z) \approx p(z|x)}$

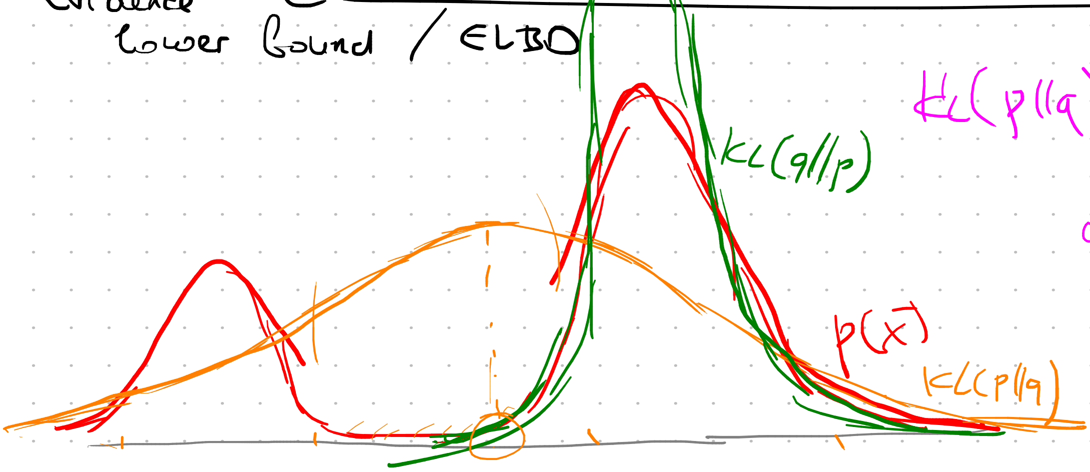
Variational lower bound

Evidence lower bound / ELBO

$$\boxed{L(q) = \int \log p(x, z)\, q(z)\, dz - \int \log q(z)\, q(z)\, dz}$$

$KL(p\|q) = \int \log \frac{p}{q}\, p\, d\bar{x}$

$q(\bar{x}) = \mathcal{N}(\bar{x} | \ldots)$



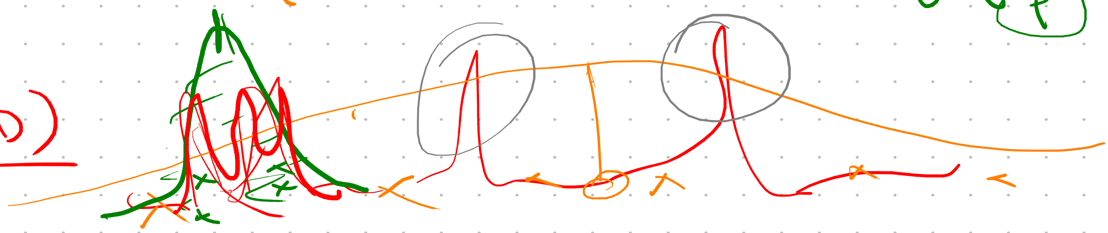
$KL(p\|q) \to \min$

$\underline{KL(q\|p) \to \min}$

$KL(p\|q) = \int \log \frac{p}{q}\, p\, d\bar{x}$

$KL(q\|p) = \int \log \frac{q}{p}\, q\, d\bar{x}$

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



$$KL(p\|q) = \int p \log \frac{p}{q}\, d\bar{x} = \int p \log p\, d\bar{x} - \underbrace{\int p \log q\, d\bar{x}}_{\xrightarrow[q]{} \max}$$

$\underline{\xrightarrow[q]{} \min}$

$p = p_{data}$

D

$$\int p \log q\, d\bar{x} = E_p[\log q] = \frac{1}{N} \sum_{n=1}^{N} \log q(\bar{x}_n) \xrightarrow[q]{} \max$$

③ $\underline{\text{Независимые подмножества } Z}$

$\boxed{q(\bar\theta)} \approx \boxed{p(\bar\theta|D)}$

$$\underline{q(Z) = \prod_{i=1}^{M} q_i(Z_i)} \quad \text{где} \quad Z_i \subseteq Z, \quad Z_i \cap Z_j = \varnothing$$

$$L(q) = \int \log \frac{p(X,Z)}{\prod_i q_i(z_i)} \prod_i q_i(z_i)\, dz_1 dz_2 \dots dz_M =$$

$$= \int \prod_i q_i \cdot \log p(X,Z)\, dZ - \sum_{i=1}^{M} \underbrace{\int \log q_i(z_i) \cdot \prod_j q_j(z_j)\, dz_1 \dots dz_M}$$

$$\int \log q_i \prod q_j \, dZ = \left(\int q_i \log q_i \, dz_i\right) \cdot \prod_{j \neq i} \underbrace{\left(\int q_j \, dz_j\right)}_{=1}$$

$$= \int \log p(X,Z) \cdot \prod_{i=1}^{M} q_i(z_i)\, dZ - \underline{\sum_{i=1}^{M} \int q_i(z_i) \log q_i(z_i)\, dz_i}$$

$$L(q) \xrightarrow[q_j]{} \max$$

$$L(q_j) = \text{Const} - \int q_j \log q_j \, dz_j + \int q_j(z_j) \left[\int \log p(X,Z) \prod_{i \neq j} q_i \, dz_i\right] dz_j$$

$$= \underbrace{\text{Const} - \int q_j(z_j) \log \frac{q_j(z_j)}{e^{\int \dots}}\, dz_j}_{\xrightarrow[q_j]{} \min} \xrightarrow[q_j(z_j)]{} \max$$

$$= KL(q_j \,\|\, \tilde{p}(z_j))$$

простое

$$\boxed{q_j^*(z_j) = \tilde{p}(z_j), \text{ где } \log \tilde{p}(z_j) = \mathbb{E}_{q_{-j}^*(z_{-j})}\left[\log p(X,Z)\right] + \text{Const}}$$

(4) Вар. прибл. для двум. гауссиана

$$p(\bar z) = \mathcal{N}(\bar z \mid \bar\mu, \Lambda^{-1}) = \frac{\sqrt{\det \Lambda}}{2\pi} e^{-\frac{1}{2}(\bar z - \bar\mu)^T \Lambda (\bar z - \bar\mu)}$$

$$\boxed{p(z_1, z_2) \approx q_1(z_1)\, q_2(z_2)}$$

$$\begin{cases} \log q_1^*(z_1) = \mathbb{E}_{q_2^*(z_2)}\left[\log p(\bar z)\right] + \text{Const} \\ \log q_2^*(z_2) = \mathbb{E}_{q_1^*(z_1)}\left[\log p(\bar z)\right] + \text{Const} \end{cases}$$

$$\log q_1^*(z_1) = \mathbb{E}_{q_2^*}\left[-\log 2\pi + \frac{1}{2}\log\det\Lambda - \frac{1}{2}(\bar{z}-\bar{\mu})^T \Lambda (\bar{z}-\bar{\mu})\right] + \text{const}$$

$$= \mathbb{E}_{q_2^*}\left[-\frac{1}{2}(\bar{z}-\bar{\mu})^T \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{12} & \lambda_{22} \end{pmatrix} (\bar{z}-\bar{\mu})\right] + \text{const} =$$

$$= \mathbb{E}_{q_2^*}\left[-\frac{1}{2}\left((z_1-\mu_1)^2\lambda_{11} + 2(z_1-\mu_1)(z_2-\mu_2)\lambda_{12} + \underbrace{(z_2-\mu_2)^2\lambda_{22}}_{+\,\text{const}}\right)\right]$$

$$= -\frac{1}{2}\mathbb{E}_{q_2^*(z_2)}\left[\lambda_{11}z_1^2 - 2\lambda_{11}\mu_1 z_1 + 2\lambda_{12}(z_2-\mu_2)\cdot z_1\right] + \text{const}$$

$$= -\frac{1}{2}\left(\lambda_{11}z_1^2 - 2\left(\lambda_{11}\mu_1 - \lambda_{12}\left(\mathbb{E}_{q_2^*}[z_2]-\mu_2\right)\right)z_1\right) + \text{const}$$

$$q_1^*(z_1) = \mathcal{N}\left(z_1 \,\middle|\, \mu_1 - \frac{\lambda_{12}}{\lambda_{11}}\left(\mathbb{E}_{q_2^*}[z_2]-\mu_2\right),\ \lambda_{11}\right)$$

$$\log q_2^*(z_2) = \mathbb{E}_{q_1^*}\left[-\frac{1}{2}\left(\cancel{(z_1-\mu_1)^2\lambda_{11}} + 2(z_1-\mu_1)(z_2-\cancel{\mu_2})\lambda_{12} + (z_2-\mu_2)^2\lambda_{22}\right)\right] + \text{const}$$

$$q_2^*(z_2) = \mathcal{N}\left(z_2 \,\middle|\, \mu_2 - \frac{\lambda_{12}}{\lambda_{22}}\left(\mathbb{E}_{q_1^*}[z_1]-\mu_1\right),\ \lambda_{22}\right)$$

$$\begin{cases} m_1 = \mu_1 - \dfrac{\lambda_{12}}{\lambda_{11}}(m_2-\mu_2) \\ m_2 = \mu_2 - \dfrac{\lambda_{12}}{\lambda_{22}}(m_1-\mu_1) \end{cases} \Rightarrow \begin{cases} m_1 = \mu_1 \\ m_2 = \mu_2 \end{cases}$$

$$KL(p\|q) \to \min$$

$$= \text{const} - \underline{\int p(z_1,z_2)\left(\log q_1(z_1) + \log q_2(z_2)\right)dz_1 dz_2}$$

$$q = \prod q_i(z_i) \Rightarrow KL(p(\bar{z})\|q_1\ldots q_M) \to \min$$

$$\text{const} - \sum_{i=1}^{M} \underbrace{\int p(\bar{z}) \log q_i(z_i)\, dz_1 \ldots dz_M}_{\to\max}$$

$$\boxed{q_i^*(z_i) = \int p(z)\, dz_{-i} = p(\bar{z}_i)}$$

5 Обучение гауссиана

$$p(x|\mu,\tau) = \sqrt{\frac{\tau}{2\pi}}\, e^{-\frac{\tau}{2}(x-\mu)^2}$$

$D = \{x_1, \dots, x_N\}$

$$\mathrm{Gam}(\tau|\alpha_0,\beta_0)\cdot\mathcal{N}(\mu|\mu_0,\lambda_0\tau)$$

$$p(\underbrace{\mu,\tau}_{z}|\underbrace{D}_{X}) \propto p(\mu,\tau)\cdot p(D|\mu,\tau) = p(\underbrace{\mu,\tau}_{z},\underbrace{x_1,\dots,x_N}_{X})$$

$$p(\mu,\tau|D) \approx q(\mu,\tau) = q_\mu(\mu)\cdot q_\tau(\tau)$$

$$\log q_\mu^*(\mu) = \mathbb{E}_{q_\tau^*}\left[\log p(\mu,\tau,x_1\dots x_N)\right] + \mathrm{Const}$$

$$\log p(\tau) + \log p(\mu|\tau) + \log p(x_1\dots x_N|\mu,\tau) =$$

$$= \mathrm{Const} + (\alpha_0-1)\log\tau - \beta_0\tau + \frac{1}{2}\log\tau - \frac{\lambda_0\tau}{2}(\mu-\mu_0)^2 +$$

$$+ \sum_{n=1}^{N}\left(\frac{1}{2}\log\tau - \frac{\tau}{2}(x_n-\mu)^2\right)$$

$$\log q_\mu^*(\mu) = \mathrm{Const} + \mathbb{E}_{q_\tau^*}\left[-\frac{\lambda_0\tau}{2}(\mu-\mu_0)^2 - \frac{\tau}{2}\sum_n(x_n-\mu)^2\right] =$$

$$= \mathrm{Const} - \frac{\mathbb{E}[\tau]}{2}\left(\lambda_0\mu^2 - 2\lambda_0\mu\mu_0 + N\mu^2 - 2\mu\cdot\sum x_n\right) =$$

$$= \mathrm{Const} - \frac{\mathbb{E}[\tau](\lambda_0+N)}{2}\left(\mu - \frac{\lambda_0\mu_0+\sum x_n}{\lambda_0+N}\right)^2$$

$$q_\mu^*(\mu) = \mathcal{N}\left(\mu \,\middle|\, \frac{\lambda_0\mu_0+\sum x_n}{\lambda_0+N}, \left((\lambda_0+N)\cdot\mathbb{E}_{q_\tau^*}[\tau]\right)^{-2}\right)$$

$$\log q_\tau^*(\tau) = \mathrm{Const} + \mathbb{E}_{q_\mu^*}\Big[\left(\alpha_0 - 1 + \frac{N+1}{2}\right)\log\tau - \Big(\beta_0 + \frac{\lambda_0}{2}(\mu-\mu_0)^2$$
$$+ \frac{1}{2}\sum_n(x_n-\mu)^2\Big)\tau\Big]$$

$$q_\tau^*(\tau) = \mathrm{Gam}\left(\tau \,\middle|\, \alpha_0 + \frac{N+1}{2}, \beta_0 + \frac{1}{2}\mathbb{E}_{q_\mu^*}\left[\lambda_0(\mu-\mu_0)^2 + \frac{1}{2}\sum_n(x_n-\mu)^2\right]\right)$$

Non-informative prior: $\alpha_0 = \beta_0 = \mu_0 = \lambda_0 = 0$

$$\begin{cases} q^*_\mu = \mathcal{N}\left(\mu \mid \frac{1}{N}\sum x_n, (N \cdot \mathbb{E}[\tau])\right) \\ q^*_\tau = \text{Gam}\left(\tau \mid \frac{N+1}{2}, \frac{1}{2}\mathbb{E}\left[\sum_n (x_n-\mu)^2\right]\right) \end{cases}$$

$$\mathbb{E}_{\text{Gam}(\alpha,\beta)}[\tau] = \frac{\alpha}{\beta}$$

$$\mathbb{E}_\tau[\tau] = \frac{N+1}{\mathbb{E}_\mu\left[\sum_n (x_n-\mu)^2\right]}$$

$$= \mathbb{E}_\mu\left[\sum_n x_n^2 - 2\sum x_n \cdot \mu + N \cdot \mu^2\right] =$$

$$= \sum_n x_n^2 - 2\,\mathbb{E}[\mu] \sum x_n + N \cdot \left(\mathbb{E}[\mu]^2 + \overbrace{\text{Var}[\mu]}^{\sigma^2}\right)$$

$$\mathbb{E}[\mu] = \frac{1}{N}\sum x_n$$

$$\mathbb{E}_\tau[\tau] = \frac{N+1}{\sum x_n^2 - \cancel{2}\sum x_n \cdot \frac{1}{N}\sum x_n + \cancel{\frac{N}{N^2}\left(\sum x_n\right)^2} + \cancel{N} \cdot \frac{1}{\cancel{N} \cdot \mathbb{E}[\tau]}}$$

$$\frac{N+\cancel{1}}{\mathbb{E}[\tau]} = \sum x_n^2 - \frac{1}{N}\left(\sum x_n\right)^2 + \cancel{\frac{1}{\mathbb{E}[\tau]}}$$

$$\mathbb{E}[\tau] = \frac{N}{\sum x_n^2 - \frac{1}{N}\left(\sum x_n\right)^2} = \frac{N^2}{N \cdot \sum x_n^2 - \left(\sum x_n\right)^2}$$

$$\sum_n (x_n - \bar{x})^2 = \sum_n x_n^2 - 2 \cdot \sum x_n \cdot \frac{1}{N} \cdot \sum x_n + \frac{N}{N^2}\left(\sum x_n\right)^2 =$$

$$= \sum x_n^2 - \frac{1}{N}\left(\sum x_n\right)^2$$

$$\mathbb{E}[\tau] = \left(\frac{1}{N}\sum_n (x_n - \bar{x})^2\right)^{-1}$$

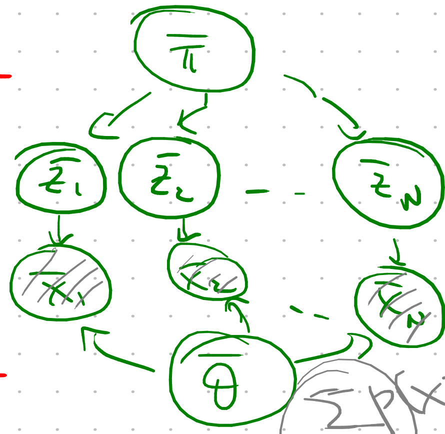
$$q^*_\mu(\mu) = \mathcal{N}\left(\mu \mid \bar{x}, \frac{1}{N}\sum_n (x_n - \bar{x})^2\right)$$

$$q^*_\tau(\tau) = \text{Gam}\left(\tau \mid \frac{N+1}{2}, \frac{1}{2}\,\frac{N+1}{N}\sum_n (x_n - \bar{x})^2\right)$$

6 Обучение смеси гауссианов

$X = \{\bar{x}_1 \ldots \bar{x}_N\}$

$Z = \{\bar{z}_1, \ldots, \bar{z}_N\}$ $\quad \bar{z}_n = (0 \ldots 1 \ldots 0)$ $\quad z_{nk} = 1 \Leftrightarrow \bar{x}_n \in C_k$



$$p(Z|\bar{\pi}) = \prod_{n=1}^{N} \prod_{k=1}^{K} \pi_k^{z_{nk}}$$

$$p(X|Z, \bar{\mu}, \bar{\Lambda}) = \prod_{n=1}^{N} \prod_{k=1}^{K} \left(\mathcal{N}(\bar{x}_n | \bar{\mu}_k, \Lambda_k) \right)^{z_{nk}}$$

$$\sum_{Z} p(X, Z|\bar{\pi}, \bar{\mu}, \bar{\Lambda})$$

$$p(X, Z|\bar{\pi}, \bar{\mu}, \bar{\Lambda}) \qquad p(\bar{\pi}, \bar{\mu}, \bar{\Lambda}|X) \propto \underline{p(\bar{\pi}, \bar{\mu}, \bar{\Lambda})} \cdot \boxed{p(X|\bar{\pi}, \bar{\mu}, \bar{\Lambda})}$$

$$x_n \in \mathbb{R}, \quad p(x_n|\mu_k, \lambda_k) = \sqrt{\frac{\lambda_k}{2\pi}} e^{-\frac{\lambda_k}{2}(x_n - \mu_k)^2}$$

$$p(\bar{\pi}, \bar{\mu}, \bar{\lambda}) = \underbrace{Dir(\bar{\pi}|\alpha_0)}_{\propto \pi_1^{\alpha_0 - 1} \ldots \pi_K^{\alpha_0 - 1}} \cdot \prod_{k=1}^{K} Gam(\lambda_k | a_0, b_0) \cdot \mathcal{N}(\mu_k | m_{0k}, \beta_{0k} \cdot \lambda_k)$$

единств. предположение!

$$p(Z, \bar{\pi}, \bar{\mu}, \bar{\lambda}|X) \approx q(Z, \bar{\pi}, \bar{\mu}, \bar{\lambda}) = q(Z) \cdot q(\bar{\pi}, \bar{\mu}, \bar{\lambda})$$

$$\underline{\log q^*(Z)} = Const + \mathbb{E}_{\pi, \mu, \lambda}\left[\log p(X, Z, \bar{\pi}, \bar{\mu}, \bar{\lambda})\right] =$$

$$= Const + \mathbb{E}\Big[\log p(\bar{\pi}) + \sum_k \log p(\lambda_k, \mu_k) + \underline{\sum_n \log p(\bar{z}_n|\bar{\pi})} +$$

$$+ \sum_n \sum_k z_{nk} \cdot \log p(x_n|\mu_k, \lambda_k)\Big]$$

$$= Const + \mathbb{E}\Big[\underbrace{\sum_n \sum_k z_{nk} \log \pi_k}_{} + \sum_n \sum_k z_{nk} \log p(x_n|\mu_k, \lambda_k)\Big] =$$

$$= \underline{Const} + \underbrace{\sum_n \sum_k z_{nk}}_{} \left(\mathbb{E}_{\pi}[\log \pi_k] + \mathbb{E}_{\mu, \lambda}[\log p(x_n|\mu_k, \lambda_k)] \right)$$

$$\underbrace{q^*(Z) = \prod_{n=1}^{N} \prod_{k=1}^{K} r_{nk}^{z_{nk}}}_{}, \text{ где } \quad r_{nk} \propto e^{\mathbb{E}_{\pi}[\log \pi_k] + \mathbb{E}_{\mu, \lambda}[\log p(x_n|\mu_k, \lambda_k)]}$$

$$\mathbb{E}_{\mu, \lambda}[\log p(x_n|\mu_k, \lambda_k)] = \underbrace{\frac{1}{2}\log \lambda_k - \frac{\lambda_k}{2}(x_n - \mu_k)^2}_{}$$

$$\underbrace{\log q^*(\bar\pi, \bar\mu, \bar\lambda)} = \text{Const} + \log p(\bar\pi) + \sum_k \log p(\mu_k, \lambda_k) + \mathbb{E}_z[\log p(Z|\bar\pi)] + \mathbb{E}_z[\log p(x|Z, \bar\mu, \bar\lambda)]$$

$$q^*(\bar\pi, \bar\mu, \bar\lambda) = q^*(\bar\pi) \cdot q^*(\bar\mu, \bar\lambda)$$

$$\log q^*(\bar\pi) = \text{Const} + \underline{(\alpha_0 - 1)} \cdot \sum_{k=1}^{K} \log \pi_k + \mathbb{E}_z\Big[\sum_n \sum_k z_{nk} \log \pi_k\Big] =$$

$$= \text{Const} + \sum_{k=1}^{K} \log \pi_k \cdot \Big(\alpha_0 + \sum_n \mathbb{E}_z[z_{nk}] - 1\Big)$$

$$q^*(\bar\pi) = \text{Dir}\Big(\ldots, \alpha_0 + \sum_n \mathbb{E}_z[z_{nk}], \ldots\Big)$$

$$= \text{Dir}\Big(\ldots, \alpha_0 + \sum_n z_{nk}, \ldots\Big)$$

$$\log \underline{q^*(\bar\mu, \bar\lambda)} = \text{Const} + \underbrace{\sum_k} \log p(\mu_k, \lambda_k) + \mathbb{E}_z\Big[\sum_n \underbrace{\sum_k} z_{nk} \log p(x_n | \mu_k, \lambda_k)\Big]$$

$$\shortparallel$$

$$\prod_k q^*(\mu_k, \lambda_k)$$

$$\log q^*(\mu_k, \lambda_k) = \text{Const} + \underline{(a_0 - 1)\log \lambda_k} - b_0 \lambda_k + \frac{1}{2}\underline{\log \lambda_k} - \frac{\lambda_k \beta_0}{2}(\mu_k - m_{0k})^2$$

$$z_{nk} =$$

$$+ \sum_n \mathbb{E}[z_{nk}] \Big(\frac{1}{2}\underline{\log \lambda_k} - \frac{\lambda_k}{2}(x_n - \mu_k)^2\Big)$$

$$q^*(\mu_k, \lambda_k) = \text{Gam}(\lambda_k | a, b) \cdot \mathcal{N}(\mu_k | m_k, \beta \cdot \lambda_k)$$

$$a = a_0 + \frac{1}{2} \sum_n z_{nk}, \qquad \beta = \beta_0 + \sum_n z_{nk}$$

$$m_k = \frac{\beta_0 m_{0k} + \sum_n z_{nk} x_n}{\beta_0 + \sum_n z_{nk}}, \qquad b = b_0 + \frac{1}{2}\Big(\beta_0 m_{0k}^2 + \sum_n z_{nk} x_n^2 - \beta \cdot m_k^2\Big)$$

$$q(\bar\pi, \bar\mu, \bar\lambda) \rightsquigarrow z_{nk}?$$

$$\mathbb{E}_{Dir(\bar{\pi}|\bar{\alpha})}[\log \bar{\pi}_k]$$

$$Dir(\bar{\pi}|\bar{\alpha}) = \frac{1}{B(\bar{\alpha})} \cdot \pi_1^{\alpha_1 - 1} \cdots \pi_k^{\alpha_k - 1} = e^{\sum_k (\alpha_k - 1) \log \bar{\pi}_k - \log B(\bar{\alpha})}$$

$$p(\bar{\pi}|\bar{\eta}) = h(\bar{\pi}) \cdot e^{\bar{\eta}^T \bar{t}(\bar{\pi}) - a(\bar{\eta})}$$

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

$$\bar{t}(\bar{\pi}) = \log \bar{\pi} = \begin{pmatrix} \log \pi_1 \\ \vdots \\ \log \pi_K \end{pmatrix}$$

$$\boxed{\mathbb{E}_p[\bar{t}(\bar{\pi})] = \nabla_{\bar{\eta}} a(\bar{\eta})}$$

$$B(\bar{\alpha}) = \frac{\prod \Gamma(\alpha_k)}{\Gamma(\sum \alpha_k)}$$

$$\mathbb{E}_{Dir(\bar{\pi}|\bar{\alpha})}[\log \bar{\pi}_k] = \frac{\partial \log B(\bar{\alpha})}{\partial \alpha_k} = \frac{d \log \Gamma(\alpha_k)}{d \alpha_k} - \frac{d \log \Gamma(\sum \alpha_k)}{d \alpha_k}$$

$$\frac{d \log \Gamma(x)}{dx} = \Psi(x) \quad \text{digamma function}$$

$$\log \varrho_{nk} = \text{Const} + \psi\left(\alpha_0 + \sum_n \varrho_{nk}\right) - \psi\left(\sum_\ell \left(\alpha_0 + \sum_n \varrho_{n\ell}\right)\right)$$

$+ \ldots$