6 - meeting

$\langle \exp \phi_{kt} \rangle$ is constant.

how to obtain $f(\Phi_{fk})$ (use hint probably)

- $p(\phi_{fk})$ normal → will contribute $-\frac{1}{2}\phi_{fk}^2$
  - $p(X_{ft} \mid \ldots)$ normal

→ only considering Φ_{fk} (param. of interest)

- Laplace mean should be mode
  variance should be $-H^{-1}$

from BP-NMF paper

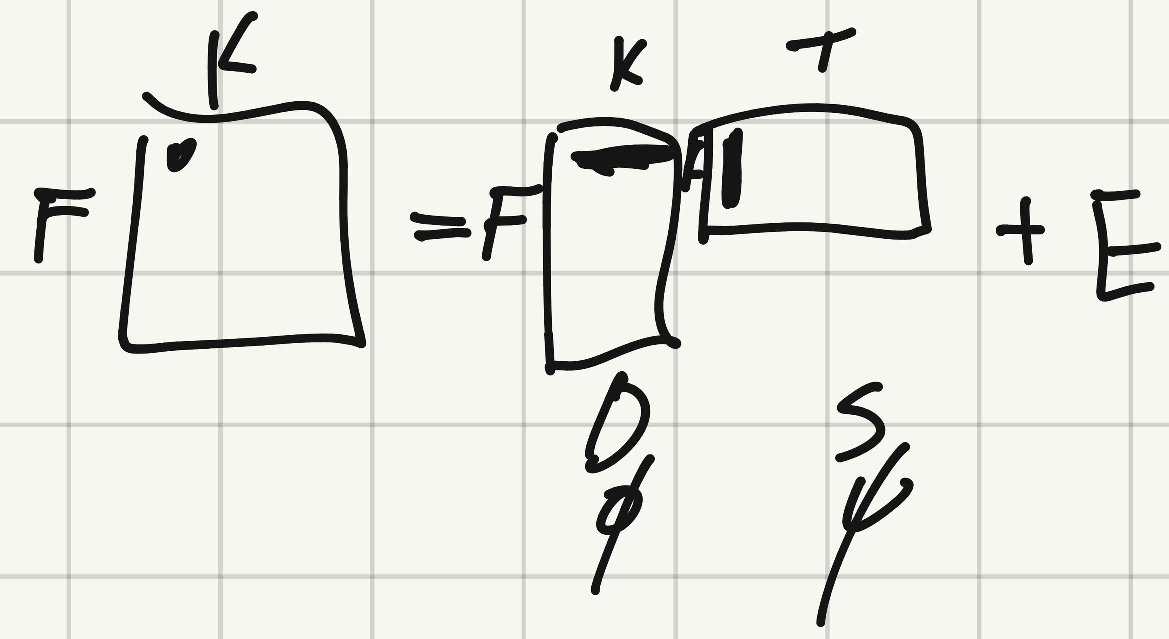
- $f(\Phi_{fk}) \propto a\exp(b\phi) - \frac{1}{2}\phi^2$
  - mean - not closed probably
  - variance - probably closed

Another paper (referenced in BP-NMF paper)

Bayesian non parametric matrix factorization for recorded music.

- θ_ℓ → shrink to 0

1. Problem Setting



- $$\underset{F\times T}{X} = \underset{F\times K}{D}\,(\underset{K\times T}{S} \odot Z) + \underset{F\times T}{E}$$

- $$\begin{aligned}
x_t &= D(s_t \odot z_t) + \varepsilon_t \\
\log d_k &\sim \mathcal{N}(0, I_F) \quad \rightarrow \phi \\
s_t &\sim \Gamma(\alpha, \beta) \quad \rightarrow \psi \\
z_k &\sim \mathrm{Bern}(\pi_k) \\
\pi_k &\sim \mathrm{Beta}\left(\frac{a_0}{K}, \frac{b_0(K-1)}{K}\right) \\
\varepsilon_t &\sim \mathcal{N}(0, \gamma_\varepsilon^{-1} I_F) \\
\gamma_\varepsilon &\sim \Gamma(c_0, d_0)
\end{aligned}$$

2. Laplace Approximation Variational Inference

$$\Theta = \{D, S, Z, \pi, \gamma_\varepsilon\} \rightarrow \{\phi, \psi, Z, \pi, \gamma_\varepsilon\}$$

- $\phi_{fk} = \log(D_{fk})$
- $\psi_{kt} = \log(S_{kt})$

Variational distribution:

$$q(\Theta) = q(\gamma_\varepsilon)\left[\prod_{k=1}^{K} q(\pi_k)\left(\prod_{f=1}^{F} q(\phi_{fk})\right)\prod_{t=1}^{T}\left[q(\psi_{kt})\,q(z_{kt})\right]\right]$$

where

$$\begin{aligned}
q(\phi_{fk}) &= \mathcal{N}\left(\mu_{fk}^{(\phi)}, \frac{1}{\gamma_{fk}^{(\phi)}}\right) \\
q(\psi_{kt}) &= \mathcal{N}\left(\mu_{kt}^{(\psi)}, \frac{1}{\gamma_{kt}^{(\psi)}}\right) \\
q(z_{kt}) &= \mathrm{Bern}(p_{kt}^{(z)}) \\
q(\pi_k) &= \mathrm{Beta}(\alpha_k^{(\pi)}, \beta_k^{(\pi)}) \\
q(\gamma_\varepsilon) &= \Gamma(\alpha^{(\varepsilon)}, \beta^{(\varepsilon)})
\end{aligned}$$

- $$\begin{aligned}
\mathrm{ELBO} &= \mathbb{E}_q[\log P(X, \Theta)] - \mathbb{E}_q[\log q(\Theta)] \\
\log P(X) &\geq \mathrm{ELBO} \\
&= \mathbb{E}_q[\log P(X \mid \phi, \psi, z, \pi, \gamma_\varepsilon)] + \mathbb{E}_q[\log P(\phi)] \\
&\quad + \mathbb{E}_q[\log P(\psi)] + \mathbb{E}_q[\log P(z \mid \pi)] + \mathbb{E}_q[\log P(\pi)] \\
&\quad + \mathbb{E}_q[\log P(\gamma_\varepsilon)] + \underbrace{\mathbb{H}_q[q]}_{\text{entropy}}
\end{aligned}$$

2.1 Update ϕ

$$= P(X \mid \phi, \psi, z, \gamma_\varepsilon) P(\phi) \cdots$$

- $$q(\phi_{fk}) \propto \exp\{\mathbb{E}_{-\phi_{fk}}[\log P(X, \Theta)]\}$$

$$\propto \exp\{\mathbb{E}_{-\phi_{fk}}[\log P(x_f \mid \phi_f, \psi, z, \gamma_\varepsilon)] + \mathbb{E}_{-\phi_{fk}}[\log P(\phi_{fk})]\}$$

normal (0,1)

$$= \exp\{\langle \log P(x_f \mid \phi_f, \psi, z, \gamma_\varepsilon)\rangle_{-\phi_{fk}} + \log P(\phi_{fk})\}$$

$$= \exp\{f(\phi_{fk})\}$$

→ need to compute to apply laplace approx.

$$\langle \log P(x_f \mid \phi_f, \psi, z, \gamma_\varepsilon)\rangle = \sum_{t=1}^{T} \langle \log P(X_{ft} \mid \phi_f, \psi_t, z_t, \gamma_\varepsilon)\rangle$$

where $P(X_{ft} \mid \phi_f, \psi_t, z_t, \gamma_\varepsilon)$ can be written as exp. family (gaussian).

now,

- $$X_{ft} \sim N\left(\sum_{j=1}^{K}[D_{fj}(S_{jt} \times z_{jt})], \frac{1}{\gamma_\varepsilon}\right)$$

$$= N\left(\sum_{j=1}^{K}[\exp(\phi_{fj})\exp(\psi_{jt}) \cdot z_{jt}], \frac{1}{\gamma_\varepsilon}\right)$$

refer to (basics & likelihood from AMSI → Exp. family in Zotero), we know the general normal exp. family form. Substituting, we get

- $$\eta = \begin{bmatrix} \frac{\mu}{\sigma^2} \\ -\frac{1}{2\sigma^2} \end{bmatrix} = \begin{bmatrix} \gamma_\varepsilon \cdot \sum_{j=1}^{K} \exp(\phi_{fj}) \cdot \exp(\psi_{jt}) \cdot z_{jt} \\ -\frac{1}{2}\gamma_\varepsilon \end{bmatrix}$$

- $$A(\eta) = \frac{\mu^2}{2\sigma^2} + \log\sigma = -\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\log(-2\eta_2)$$

$$= -\frac{1}{2}\left[\gamma_\varepsilon\left(\sum_{j=1}^{K}\exp(\phi_{fj})\exp(\psi_{jt}) \cdot z_{jt}\right)^2 + \log(\gamma_\varepsilon)\right]$$

- $$T(X_{ft}) = \begin{bmatrix} X_{ft} \\ X_{ft}^2 \end{bmatrix}$$

now,

$$\langle \log P(X_{ft} \mid \phi_f, \psi_t, z_t, \gamma_\varepsilon)\rangle_{-\phi_{fk}} \overset{(\text{wrt. } \phi_{fk})}{\propto} \langle \eta^T\rangle T(X_{ft}) - \langle A(\eta)\rangle$$

$$= \langle \gamma_\varepsilon\rangle\left[\sum_{j=1}^{K}\langle \exp(\phi_{fj})\rangle\langle \exp(\psi_{jt})\rangle\langle z_{jt}\rangle\right] X_{ft} - \frac{1}{2}\langle\gamma_\varepsilon\rangle X_{ft}^2$$

$$- \frac{1}{2}\left[\langle\gamma_\varepsilon\rangle\left\langle\left(\sum_{j=1}^{K}\exp(\phi_{fj})\exp(\psi_{jt}) z_{jt}\right)^2\right\rangle + \langle \log(\gamma_\varepsilon)\rangle\right]$$

$$= \langle \gamma_\varepsilon \rangle \Big[\sum_{j=1}^{K} \langle \exp(\phi_{fj}) \rangle \langle \exp(\psi_{jt}) \rangle \langle z_{jt} \rangle\Big] X_{ft} - \cancel{\tfrac{1}{2}\langle \gamma_\varepsilon \rangle X_{ft}^2} \text{ (no } \phi_{fk})$$

$$- \frac{1}{2}\Big[\langle \gamma_\varepsilon \rangle \Big\langle \Big(\sum_{j=1}^{K} \exp(\phi_{fj}) \cdot \exp(\psi_{jt}) \cdot z_{jt}\Big)^2 \Big\rangle + \cancel{\langle \log(\gamma_\varepsilon) \rangle}\Big]$$

$$\overset{(\phi_{fk})}{\propto} \langle \gamma_\varepsilon \rangle X_{ft} \Big[\exp(\phi_{fk}) \langle \exp(\psi_{kt}) \rangle \langle z_{kt} \rangle + \cancel{\sum_{j \neq k} \langle \exp(\phi_{fj}) \rangle \langle \exp(\psi_{jt}) \rangle \langle z_{jt} \rangle}^{\text{(no } \phi_{fk})}\Big]$$

$$- \frac{1}{2} \langle \gamma_\varepsilon \rangle \Big[\ \cdots\ \Big]$$

$P(X_{tf})$ $\quad P(X_t)$

- $$\Big(\sum_{j=1}^{K} \exp(\phi_{fj}) \exp(\psi_{jt})\, z_{jt}\Big)^2$$

$$= \Big(\exp(\phi_{fk}) \exp(\psi_{kt})\, z_{kt} + \underbrace{\sum_{j \neq k} \exp(\phi_{fj}) \exp(\psi_{jt})\, z_{jt}}\Big)^2$$

$$\overset{(\phi_{fk})}{\propto} \exp(2\phi_{fk}) \exp(2\psi_{kt})\, z_{kt}^2 + 2\exp(\phi_{fk}) \exp(\psi_{kt})\, z_{kt} \times \sum_{j \neq k} \boxed{}$$

$$\Rightarrow \Big\langle \Big(\sum\ \Big)^2 \Big\rangle$$

$$\propto \exp(2\phi_{fk}) \langle \exp(2\psi_{kt}) \rangle \langle z_{kt}^2 \rangle$$

$$+ 2\exp(\phi_{fk}) \langle \exp(\psi_{kt}) \rangle \langle z_{kt} \rangle \Big\langle \sum_{j \neq k} \exp(\phi_{fj}) \exp(\psi_{jt})\, z_{jt} \Big\rangle$$

now, put $\sum_{t=1}^{T}$ for $\log P(x_f \mid \cdots)$

$z_i \sim \mathrm{Bern}(p_i)$

$\Rightarrow z_i^2 \sim \mathrm{Bern}(p_i)$

$$\propto \langle \gamma_\varepsilon \rangle \exp(\phi_{fk}) \sum_{t=1}^{T} X_{ft} \langle \exp(\psi_{kt}) \rangle \langle z_{kt} \rangle$$

$$- \frac{1}{2}\langle \gamma_\varepsilon \rangle \Big[\exp(2\phi_{fk}) \sum_{t=1}^{T} \langle \exp(2\psi_{kt}) \rangle \langle z_{kt}^2 \rangle \quad = \langle z_{kt} \rangle$$

$$+ 2\exp(\phi_{fk}) \sum_{t=1}^{T} \langle \exp(\psi_{kt}) \rangle \langle z_{kt} \rangle \Big\langle \sum_{j \neq k} \exp(\phi_{fj}) \exp(\psi_{jt})\, z_{jt} \Big\rangle\Big]$$

$$= \langle \gamma_\varepsilon \rangle \exp(\phi_{fk}) \sum_{t=1}^{T} \langle \exp(\psi_{kt}) \rangle \langle z_{kt} \rangle \Big\langle X_{ft} - \sum_{j \neq k} \exp(\phi_{fj}) \exp(\psi_{jt})\, z_{jt} \Big\rangle$$

$$- \frac{1}{2}\langle \gamma_\varepsilon \rangle \cdot \exp(2\phi_{fk}) \sum_{t=1}^{T} \langle \exp(2\psi_{kt}) \rangle \langle z_{kt} \rangle$$

$$\Rightarrow f(\phi_{fk}) \propto \langle \gamma_\varepsilon \rangle \exp(\phi_{fk}) \boxed{\sum_{t=1}^{T} \langle \exp(\psi_{kt}) \rangle \langle z_{kt} \rangle \langle \hat{X}_{ft}^{-k} \rangle}$$

$$- \frac{1}{2}\langle \gamma_\varepsilon \rangle \cdot \exp(2\phi_{fk}) \sum_{t=1}^{T} \langle \exp(2\psi_{kt}) \rangle \langle z_{kt} \rangle - \frac{1}{2}\phi_{fk}^2$$

from $\log P(\phi_{fk})$ $\sim$ Normal

$$\langle \gamma_\varepsilon \rangle \Big[\sum_{j=1}^{K} \langle \exp(\phi_{fj})\rangle \langle \exp(\psi_{jt})\rangle \langle z_{jt}\rangle \Big] X_{ft} - \cancel{\tfrac{1}{2}\langle\gamma_\varepsilon\rangle X_{ft}^2} \quad \text{(no } \phi_{fk}\text{)}$$

$$-\frac{1}{2}\Big[\langle \gamma_\varepsilon \rangle \Big\langle \sum_{j=1}^{K} \exp(\phi_{fj}) \cdot \exp(\psi_{jt}) \cdot z_{jt} \Big\rangle^2 + \langle \log(\gamma_\varepsilon)\rangle \Big]$$

(ϕ_{fk})

$$\propto \langle \gamma_\varepsilon \rangle X_{ft} \Big[\exp(\phi_{fk}) \langle \exp(\psi_{kt})\rangle \langle z_{kt}\rangle + \cancel{\sum_{j\neq k} \langle \exp(\phi_{fj})\rangle \langle \exp(\psi_{jt})\rangle \langle z_{jt}\rangle}\Big] \quad \text{(no } \phi_{fk}\text{)}$$

$$-\frac{1}{2}\langle \gamma_\varepsilon \rangle \Big[\big(\exp(\phi_{fk}) \langle \exp(\psi_{kt})\rangle \langle z_{kt}\rangle\big)^2 + 2\exp(\phi_{fk}) \langle \exp(\psi_{kt})\rangle \langle z_{kt}\rangle \times \sum_{j\neq k} \boxed{\cdots} \Big]$$

put $\sum_{t=1}^{T}$ for $\log p(\phi_{fk}|\cdots)$

$$\propto \langle \gamma_\varepsilon \rangle \exp(\phi_{fk}) \sum_{t=1}^{T} \langle \exp(\psi_{kt})\rangle \langle z_{kt}\rangle X_{ft}$$

different

$$-\frac{1}{2}\langle \gamma_\varepsilon \rangle \exp(2\phi_{fk}) \sum_{t=1}^{T} \langle \exp(\psi_{kt})\rangle^2 \langle z_{kt}\rangle^2 + 2\exp(\phi_{fk}) \times \cdots\cdots$$

$$\Rightarrow f(\phi_{fk}) \propto \langle \gamma_\varepsilon \rangle \exp(\phi_{fk}) \times \sum_{t=1}^{T} \langle \exp(\psi_{kt})\rangle \langle z_{kt}\rangle \langle \hat{X}_{ft}^{-k}\rangle$$

correct derivation

$$-\frac{1}{2}\langle \gamma_\varepsilon \rangle \exp(2\phi_{fk}) \times \sum_{t=1}^{T} \langle \exp(2\psi_{kt})\rangle \langle z_{kt}\rangle$$

$$-\frac{1}{2}\phi_{fk}^2 \quad (\text{from } \log p(\phi_{fk}))$$

should I use the sum.

now, $$\frac{\partial f}{\partial \phi_{fk}} = \langle \gamma_\varepsilon \rangle \Big[\exp(\phi_{fk}) \sum_{t=1}^{T} \langle \exp(\psi_{kt})\rangle \langle z_{kt}\rangle \langle \hat{X}_{ft}^{-k}\rangle - \exp(2\phi_{fk}) \sum_{t=1}^{T} \langle \exp(2\psi_{kt})\rangle \langle z_{kt}\rangle \Big] - \phi_{fk}$$

Newton Raphson

$\overset{\text{set}}{=} 0$ to find $\hat{\phi}_{fk}$ (no closed form)

$$\frac{\partial^2 f}{\partial \phi_{fk}^2} = \langle \gamma_\varepsilon \rangle \Big[\exp(\phi_{fk}) \sum_{t=1}^{T} \langle \exp(\psi_{kt})\rangle \langle z_{kt}\rangle \langle \hat{X}_{ft}^{-k}\rangle - 2\exp(2\phi_{fk}) \sum_{t=1}^{T} \langle \exp(2\psi_{kt})\rangle \langle z_{kt}\rangle \Big] - 1$$

$\therefore\ q(\phi_{fk}) \approx N(\mu_{fk}^{(\phi)}, \gamma_{fk}^{(\phi)})$ with (using Laplace approx.)

$\mu_{fk}^{(\phi)} = \hat{\phi}_{fk}$, $\gamma_{fk}^{(\phi)} = -\dfrac{\partial^2 f(\hat{\phi}_{fk})}{\partial \phi_{fk}^2}$ updates.

2.2. Update ψ

- $q(\psi_{kt}) \propto \exp\{\langle \log P(X, \Theta)\rangle_{-\psi_{kt}}\}$

$$\propto \exp\{\langle \log P(x_t \mid \phi, \psi_t, z_t, r_\varepsilon)\rangle + \langle \log P(\psi_{kt})\rangle\}$$

$$= \exp\{f(\psi_{kt})\}$$

here, S_{kt} is Gamma, so $\psi_{kt} \overset{\log S_{kt}}{=}$ is log-gamma distributed.

now, Gamma has pdf: $\frac{\beta^\alpha}{\Gamma(\alpha)} S_{kt}^{\alpha-1} \cdot \exp(-\beta S_{kt})$ but $S_{kt} = \exp(\psi_{kt})$

$$\Rightarrow p(\psi_{kt}) = \frac{\beta^\alpha}{\Gamma(\alpha)} \exp(\psi_{kt})^{\alpha \cancel{-1}} \cdot \exp(-\beta \exp(\psi_{kt}))$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \exp\{\alpha\psi_{kt} - \beta\exp(\psi_{kt})\}$$

$$\Rightarrow \langle \log p(\psi_{kt})\rangle \propto \alpha\psi_{kt} - \beta\exp(\psi_{kt})$$

and $\langle \log P(x_t \mid \phi, \psi_t, z_t, r_\varepsilon)\rangle$ (same form as before, swapping appropriately)

$$\propto \langle r_\varepsilon\rangle \exp(\psi_{kt}) \sum_{f=1}^{F} \langle \exp(\phi_{fk})\rangle \langle z_{kt}\rangle \langle \hat{X}_{ft}^{-k}\rangle$$

$$- \frac{1}{2}\langle r_\varepsilon\rangle \exp(2\psi_{kt}) \sum_{f=1}^{F} \langle \exp(2\phi_{fk})\rangle \langle z_{kt}\rangle$$

$$\Rightarrow f(\psi_{kt}) \propto \langle r_\varepsilon\rangle \exp(\psi_{kt}) \sum_{f=1}^{F} \langle \exp(\phi_{fk})\rangle \langle z_{kt}\rangle \langle \hat{X}_{ft}^{-k}\rangle$$

$$- \frac{1}{2}\langle r_\varepsilon\rangle \exp(2\psi_{kt}) \sum_{f=1}^{F} \langle \exp(2\phi_{fk})\rangle \langle z_{kt}\rangle$$

$$+ \alpha\psi_{kt} - \beta\exp(\psi_{kt})$$

⋮

similar update as before, do $\frac{\partial f}{\partial \psi_{kt}}$, $\frac{\partial^2 f}{\partial \psi_{kt}^2}$

2.3 Update Z

- $q(Z_{kt}) \propto \exp\{\langle \log P(X, \Theta)\rangle_{-Z_{kt}}\}$

$$\propto \exp\{\langle \log P(x_t \mid \phi, \psi_t, Z_t, \gamma_\epsilon)\rangle + \langle \log P(Z_{kt} \mid \pi_k)\rangle\}$$

Since Z_{kt} is Bernoulli, can explicitly compute $p_1 \propto q(Z_{kt}=1)$ and $p_0 \propto q(Z_{kt}=0)$

- if $Z_{kt}=1$,

from $\langle \log P(x_t \mid \phi, \psi_t, Z_t, \gamma_\epsilon)\rangle$ (wrt. Z_{kt})

$$\propto \langle \gamma_\epsilon\rangle \sum_{f=1}^{F} \langle \exp(\psi_{kt})\rangle \langle \exp(\phi_{fk})\rangle \cancelto{1}{\langle Z_{kt}\rangle} \langle \hat{X}_{ft}^{-k}\rangle$$

$$-\frac{1}{2}\langle \gamma_\epsilon\rangle \sum_{f=1}^{F} \langle \exp(2\psi_{kt})\rangle \langle \exp(2\phi_{fk})\rangle \cancelto{1}{\langle Z_{kt}\rangle}$$

$\propto \langle \gamma_\epsilon\rangle \sum_{f=1}^{F} \{\langle \exp(\phi_{fk})\rangle \cdots$ and $\langle \log P(Z_{kt} \mid \pi_k)\rangle \propto \langle \log \pi_k\rangle$

- if $Z_{kt}=0$, $\log q(Z_{kt}) \propto p_0 = \langle \log(1-\pi_k)\rangle$

where $\langle \log \pi_k\rangle = \psi(\alpha_k^{(\pi)}) - \psi(\alpha_k^{(\pi)} + \beta_k^{(\pi)})$

($\psi(\cdot)$ is digamma function) (* fact for $\mathbb{E}[\log(\text{Beta})]$)

so $\langle \log(1-\pi_k)\rangle = \psi(\beta_k^{(\pi)}) - \psi(\alpha_k^{(\pi)} + \beta_k^{(\pi)})$

$$\Rightarrow \rho_{kt}^{(z)} \Leftarrow \frac{\exp(p_1)}{\sum_{i \in \{0,1\}} \exp(p_i)} \quad \text{update rule}$$

$= \exp(p_0) + \exp(p_1)$

2.4 π, 2.5 γ_ϵ have closed form due to conjugacy with

Beta-Bernoulli normal-gamma