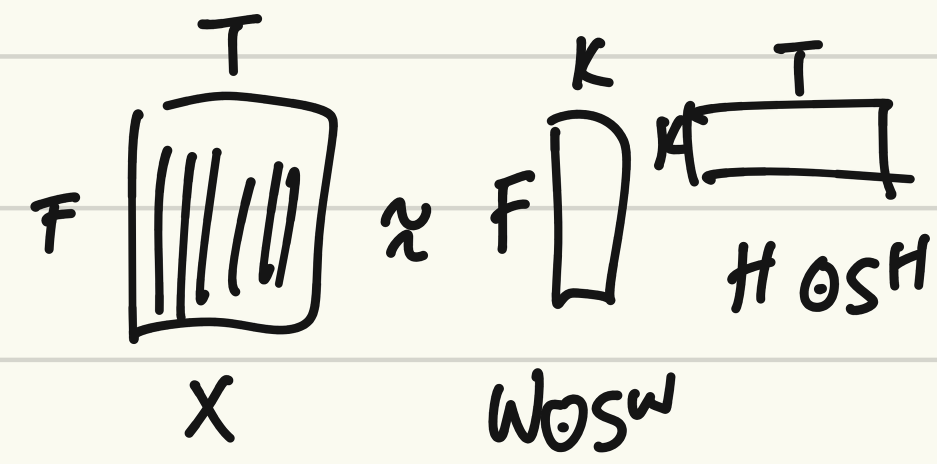
Laplace VI with extension (another sparsity matrix, Indep. to W)

**Set-up**




$$X = (W \odot S^W)(H \odot S^H) + E$$

$$x_t = (W \odot S^W)(H_t \odot S^H_t) + \varepsilon_t$$

$$\log w_k \sim N(0, I_F)$$

$$S^W_{fk} \sim \mathrm{Bern}(\pi^W_k)$$

$$\pi^W_k \sim \mathrm{Beta}(a_0^W/K,\ b_0^W(K-1)/K)$$

$$H_t \sim \Gamma(\alpha, \beta)$$

$$S^H_{kt} \sim \mathrm{Bern}(\pi^H_k)$$

$$\pi^H_k \sim \mathrm{Beta}\left(\frac{a_0^H}{K},\ \frac{b_0^H(K-1)}{K}\right)$$

$$\varepsilon_t \sim N(0, \gamma_\varepsilon^{-1} I_F)$$

$$\gamma_\varepsilon \sim \Gamma(c_0, d_0)$$

**Laplace approx. VI.**

$$\theta = \{W, S^W, \pi^W, H, S^H, \pi^H, \gamma_\varepsilon\} \to \{\phi, S^W, \pi^W, \psi, S^H, \pi^H, \gamma_\varepsilon\}$$

- $\phi_{fk} = \log(W_{fk})$
- $\psi_{kt} = \log(H_{kt})$

**variational distribution**

$$q(\theta) = q(\gamma_\varepsilon) \cdot \prod_k \left[ \left(q(\pi^W_k) q(\pi^H_k)\right) \cdot \prod_f \left(q(\phi_{fk}) q(S^H_{fk})\right) \cdot \prod_t \left(q(\psi_{kt}) q(S^W_{kt})\right)\right]$$

where

$$q(\phi_{fk}) = N(\mu^\phi_{fk},\ 1/\gamma^\phi_{fk})$$

$$q(\psi_{kt}) = N(\mu^\psi_{kt},\ 1/\gamma^\psi_{kt})$$

$$q(S^W_{fk}) = \mathrm{Bern}(p^W_{fk})$$

$$q(S^H_{kt}) = \mathrm{Bern}(p^H_{kt})$$

$$q(\pi^W_k) = \mathrm{Beta}(\alpha^W_k, \beta^W_k)$$

$$q(\pi^H_k) = \mathrm{Beta}(\alpha^W_k, \beta^W_k)$$

$$q(\gamma_\varepsilon) = \Gamma(\alpha^\varepsilon, \beta^\varepsilon)$$

**Updates**

**$\phi$**

$p(\phi_{fk}) \cdots p(X \mid \phi, \ldots)$

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

$$\propto \exp\{\langle \log p(x_f \mid \phi_f, S^W_f, \psi, S^H, \gamma_\varepsilon)\rangle + \langle \log p(\phi_{fk})\rangle\}$$

$N(0,1)$

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

compute to apply laplace approx

$$\bullet\ \langle \log p(x_f \mid \phi_f, S_f^H, \psi, S^W, r_\varepsilon)\rangle = \sum_t \langle \log p(X_{ft} \mid \phi_f, S_f^W, \psi_t, S_t^H, r_\varepsilon)\rangle$$

we can write $p(X_{ft} \mid \phi_f, S_f^W, \psi_t, S_t^H, r_\varepsilon)$ in exp family form (normal)

$$\bullet\ X_{ft} \sim N\Big(\sum_k W_{fk} \cdot S_{fk}^W \cdot H_{kt} \cdot S_{kt}^H,\ \frac{1}{r_\varepsilon}\Big)$$
$$= N\Big(\sum_k \exp(\phi_{fk})\, S_{fk}^W \cdot \exp(\psi_{kt}) \cdot S_{kt}^H,\ 1/r_\varepsilon\Big)$$

now,

$$\bullet\ \eta = \begin{bmatrix} \frac{\mu}{\sigma^2} \\ -\frac{1}{2\sigma^2} \end{bmatrix} = \begin{bmatrix} r_\varepsilon \sum_{j=1}^K \exp(\phi_{fj}) S_{fj}^W \exp(\psi_{jt}) S_{jt}^H \\ -\frac{1}{2} r_\varepsilon \end{bmatrix}$$

$$\bullet\ A(\eta) = \frac{\mu^2}{2\sigma^2} + \log \sigma = -\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\log(-2\eta_2)$$
$$= -\frac{1}{2}\Big[ r_\varepsilon \Big(\sum_{j=1}^K \exp(\phi_{fj}) S_{fj}^W \exp(\psi_{jt}) S_{jt}^H\Big)^2 + \log(r_\varepsilon)\Big]$$

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

So, $\langle \log p(X_{ft} \mid \cdots)\rangle_{\phi_{fk}} \overset{(\phi_{fk})}{\propto} \langle \eta^T \rangle T(X_{ft}) - \langle A(\eta)\rangle$

$$= \Big\langle r_\varepsilon \sum_{j=1}^K \exp(\phi_{fj}) S_{fj}^W \exp(\psi_{jt}) S_{jt}^H \Big\rangle X_{ft} - \cancel{\frac{1}{2}\langle r_\varepsilon \rangle X_{ft}^2}$$
$$- \frac{1}{2}\Big[\Big\langle r_\varepsilon \Big(\sum_{j=1}^K \exp(\phi_{fj}) S_{fj}^W \exp(\psi_{jt}) S_{jt}^H\Big)^2 + \cancel{\log r_\varepsilon}\Big\rangle\Big]$$

no $\phi_{fk}$

$$\propto \langle r_\varepsilon \rangle X_{ft}\Big[\exp(\phi_{fk})\langle S_{fk}^W\rangle \langle \exp(\psi_{kt})\rangle \langle S_{kt}^H \rangle + \cancel{\sum_{j\neq k} \langle \exp(\phi_{fj}) S_{fj}^W \exp(\psi_{jt}) S_{jt}^H \rangle}\Big]$$
$$- \frac{1}{2}\langle r_\varepsilon \rangle \Big[\Big\langle \underbrace{\Big(\sum_{j=1}^K \exp(\phi_{fj}) S_{fj}^W \exp(\psi_{jt}) S_{jt}^H\Big)^2}\Big\rangle\Big]$$

$$\Big(\exp(\phi_{fk}) S_{fk}^W \exp(\psi_{kt}) S_{kt}^H + \sum_{j\neq k} \exp(\phi_{fj}) S_{fj}^W \exp(\psi_{jt}) S_{jt}^H\Big)^2$$
$$\propto \exp(2\phi_{fk})\exp(2\psi_{kt})(S_{fk}^W)^2 (S_{kt}^H)^2 + 2\exp(\phi_{fk})\exp(\psi_{kt}) S_{fk}^W S_{kt}^H \sum_{j\neq k} \boxed{\cdots}$$

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

$$\propto \langle r_\varepsilon \rangle \cdot \exp(\phi_{fk}) \langle S_{fk}^W \rangle \sum_{t=1}^T X_{ft} \langle \exp(\psi_{kt})\rangle \langle S_{kt}^H \rangle$$

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

$$- \frac{1}{2}\langle r_\varepsilon \rangle \Big[\exp(2\phi_{fk})\langle S_{fk}^W \rangle \sum_{t=1}^T \langle \exp(2\psi_{kt})\rangle \langle S_{kt}^H \rangle$$
$$+ 2\exp(\phi_{fk})\langle S_{fk}^W \rangle \sum_{t=1}^T \langle \exp(\psi_{kt}) \rangle \langle S_{kt}^H \rangle \Big\langle \sum_{j\neq k} \exp(\phi_{fj}) S_{fj}^W \exp(\psi_{jt}) S_{jt}^H \Big\rangle \Big]$$

$$\propto \langle r_\varepsilon\rangle \cdot \exp(\phi_{fk})\langle S^W_{fk}\rangle \sum_{t=1}^T X_{ft}\langle \exp(\psi_{kt})\rangle\langle S^H_{kt}\rangle$$

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

$$-\frac{1}{2}\langle r_\varepsilon\rangle \Big[ \exp(2\phi_{fk})\langle S^W_{fk}\rangle \sum_{t=1}^T \langle \exp(2\psi_{kt})\rangle\langle S^H_{kt}\rangle$$

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

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

$$-\frac{1}{2}\langle r_\varepsilon\rangle \exp(2\phi_{fk})\langle S^W_{fk}\rangle \sum_{t=1}^T \langle \exp(2\psi_{kt})\rangle\langle S^H_{kt}\rangle$$

$\Rightarrow f(\phi_{fk})$

$$\propto \langle r_\varepsilon\rangle \exp(\phi_{fk})\langle S^W_{fk}\rangle \sum_{t=1}^T \langle \exp\psi_{kt}\rangle\langle S^H_{kt}\rangle \langle \hat{X}^{-k}_{ft}\rangle$$

$$-\frac{1}{2}\langle r_\varepsilon\rangle \exp(2\phi_{fk})\langle S^W_{fk}\rangle \sum_{t=1}^T \langle \exp(2\psi_{kt})\rangle\langle S^H_{kt}\rangle \underbrace{-\frac{1}{2}\phi_{fk}^2}_{\text{from } \log p(\phi_{fk})}$$

now,

$$\frac{\partial f}{\partial \phi_{fk}} = \langle r_\varepsilon\rangle \Big[\exp(\phi_{fk})\langle S^W_{fk}\rangle \sum_{t=1}^T \langle \exp\psi_{kt}\rangle\langle S^H_{kt}\rangle\langle \hat{X}^{-k}_{ft}\rangle$$

$$- \exp(2\phi_{fk})\langle S^W_{fk}\rangle \sum_{t=1}^T \langle \exp(2\psi_{kt})\rangle\langle S^H_{kt}\rangle\Big]$$

$$-\phi_{fk}$$

use optimization algorithm

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

$$\frac{\partial^2 f}{\partial \phi_{fk}^2} = \langle r_\varepsilon\rangle \Big[\exp(\phi_{fk})\langle S^W_{fk}\rangle \sum_t \langle \exp(\psi_{kt})\rangle\langle S^H_{kt}\rangle\langle \hat{X}^{-k}_{ft}\rangle$$

$$-2\cdot \exp(2\phi_{fk})\sum_t \langle \exp 2\psi_{kt}\rangle\langle S^H_{kt}\rangle\Big]$$

$$-1$$

$\therefore$ $q(\phi_{fk}) \approx N(\mu^\phi_{fk}, r^\phi_{fk})$ with

$$\mu^\phi_{fk} = \hat{\phi}_{fk}, \quad r^\phi_{fk} = -\frac{\partial^2 f(\hat{\phi}_{fk})}{\partial \phi_{fk}^2}$$ by Laplace Approx.

$\psi$: • $q(\psi_{kt}) \propto \exp\{\langle \log P(X,\theta)_{-\psi_{kt}}\rangle$

$\propto \exp\{\langle \log p(x_t \mid \phi, S^H, \psi_t, S^H_t, r_\epsilon)\rangle + \langle \log P(\psi_{kt})\rangle\}$

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

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

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

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

and $\langle \log P(x_t \mid \cdots)\rangle$ has similar form as before, with terms swapped appropriately ($f$ and $t$)

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

$$-\frac{1}{2}\langle r_\epsilon\rangle \exp(2\psi_{kt})\langle S^H_{kt}\rangle \sum_{f=1}^F \langle \exp(2\phi_{fk})\rangle\langle S^W_{fk}\rangle$$

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

$\vdots$

very similar update as before

$S^H_{kt}$ • $q(S^H_{kt}) \propto \exp\{\langle \log P(X,\theta)\rangle_{-S^H_{kt}}\}$

$\propto \exp\{\langle \log P(x_t \mid \phi, S^W, \psi_t, S^H_t, r_\epsilon)\rangle$

$+ \langle \log P(S^H_{kt} \mid \pi^H_k)\rangle\}$

Bernoulli

Bernoulli$(\pi^H_k)$

when $S^H_{kt}=1$, $\log q(S^H_{kt}) \propto p_1 := \langle \log \pi^H_k\rangle + \langle r_\epsilon\rangle \sum_f \{\langle \exp\phi_{fk}\rangle\langle\exp\psi_{kt}\rangle\langle S^W_{fk}\rangle \cancel{\langle S^H_{kt}\rangle}^{1} \langle \hat{X}^{-k}_{ft}\rangle$

$-\frac{1}{2}\langle \exp 2\psi_{kt}\rangle\langle \exp 2\phi_{fk}\rangle\langle S^W_{fk}\rangle\cancel{\langle S^H_{kt}\rangle}^{1}\}$

$S^H_{kt}=0$, $\log q(S^H_{kt}) \propto p_0 := \langle \log(1-\pi^H_k)\rangle$

where $\langle \log \pi^H_k\rangle = \psi(\alpha^{\pi^H}_k) - \psi(\alpha^{\pi^H}_k + \beta^{\pi^H}_k)$

$\langle \log(1-\pi^H_k)\rangle = \psi(\beta^{\pi^H}_k) - \psi(\alpha^{\pi^H}_k + \beta^{\pi^H}_k)$

• $\psi(\cdot)$ is the digamma function.

$$\Rightarrow p^H_{kt} \leftarrow \frac{\exp(p_1)}{\exp(p_0) + \exp(p_1)}$$

$S^W_{fk}$: (similar to above)

• $\log(S^W_{fk}) \propto p_1 := \langle \log \pi^W_k\rangle + \langle r_\epsilon\rangle \sum_t \{\cdots\}$ with $S^W_{fk}=1$

$p_0 := \langle \log(1-\pi^W_k)\rangle$

$$\Rightarrow p^W_{fk} \leftarrow \frac{e^{p_1}}{e^{p_0}+e^{p_1}}$$

$\pi_k^H$ $\cdot$ $q(\pi_k) \propto \exp\{\langle \log P(X,\theta)\rangle_{-\pi_k}\}$

$$\propto \exp\{\langle \log P(S_k^H \mid \pi_k^H)\rangle + \langle \log \pi_k^H \rangle$$

Bern, Beta

$$= \exp\{\langle \textstyle\sum_t \log P(S_{kt}^H \mid \pi_k^H)\rangle + \log \pi_k^H\}$$

$$= (\pi_k^H)^{\sum_{t=1}^T \langle S_{kt}^H\rangle + \frac{a_0^H}{K} - 1} \cdot (1-\pi_k^H)^{T-\sum_{t=1}^T \langle S_{kt}^H\rangle + \frac{b_0^H(K-1)}{K} - 1}$$

$$\Rightarrow \quad a_k^{\pi^H} \leftarrow \frac{a_0^H}{K} + \sum_{t=1}^T \langle S_{kt}^H\rangle$$

$$b_k^{(\pi)} \leftarrow \frac{b_0^H(K-1)}{K} + T - \sum_{t=1}^T \langle S_{kt}^H\rangle$$

$\pi_k^W$ $\cdot$ Similar but $\sum_{f=1}^F \langle S_{fk}^W\rangle$

$$\frac{1}{\gamma_\varepsilon} = \sigma^2$$

$$\gamma_\varepsilon = \sigma^{-2}$$

$\gamma_\varepsilon$ $\cdot$ $q(\gamma_\varepsilon) \propto \exp\{\langle \log P(X,\theta)\rangle_{-\gamma_\varepsilon}\}$

normal

$$\propto \exp\{\langle \log P(X \mid \phi, \psi, S^W, S^H, \gamma_\varepsilon)\rangle + \langle \log P(\gamma_\varepsilon)\rangle\}$$

ma, am

$$= \exp\{\langle \textstyle\sum_{f,t} \log P(X_{ft} \mid \phi_f, \psi_t, S_f^W, S_t^H, \gamma_\varepsilon)\rangle + \langle \log P(\gamma_\varepsilon)\rangle\}$$

$$\propto \exp\{\textstyle\sum_{f,t} (\log \gamma_\varepsilon^{\frac{1}{2}} - \frac{1}{2}\gamma_\varepsilon \langle x_{ft} - \sum_k (W_{fk} \odot S_{fk}^W)(H_{kt} \odot S_{kt}^H)\rangle$$

$$+ c_0 \log \gamma_\varepsilon - d_0 \gamma_\varepsilon\}$$

$$= \exp\{(\log \gamma_\varepsilon)(c_0 + F \times T) - \gamma_\varepsilon(d_0 + \tfrac{1}{2}\textstyle\sum_t \|x_t - \langle (W \odot S^W)(H_t \odot S_t^H)\rangle\|_2^2)$$

$$= \Gamma\left(c_0 + \tfrac{1}{2} F \times T,\ d_0 + \tfrac{1}{2}\sum_{t=1}^T \|x_t - \langle (W \odot S^W)(H_t \odot S_t^H)\rangle\|_2^2\right.$$