Sparse factor models with IBP prior

Presented by Ilsang Ohn

January 10, 2017

Pati et al. (2014): Setup

Let $\boldsymbol{\Sigma}_{0n}$ be a true sequence of covariance matrices. We observe

$$\mathbf{x}^{(n)} = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \overset{iid}{\sim} \mathrm{N}_{p_n}(\mathbf{0}, \boldsymbol{\Sigma}_{0n})$$

where $\boldsymbol{\Sigma}_{0n}$ is assumed to be of the form

$$\boldsymbol{\Sigma}_{0n} = \boldsymbol{\Lambda}_{0n}\boldsymbol{\Lambda}_{0n}^{\top} + \boldsymbol{\Omega}_{0n}, \quad \boldsymbol{\Lambda}_{0n} \in \mathbb{R}^{p_n \times k_{0n}}, \; k_{0n} < p_n, \; \boldsymbol{\Omega}_{0n} = \sigma_{0n}^2 \mathbf{I}_{p_n} \qquad \text{(A0)}$$

i.e., we assume the data are generated from a factor model

$$\mathbf{x}_i = \boldsymbol{\Lambda}_{0n}\mathbf{f}_i + \epsilon_i, \quad \mathbf{f}_i \sim \mathrm{N}_{k_{0n}}(0, \mathbf{I}_{k_{0n}}), \; \epsilon_i \sim \mathrm{N}_{p_n}(0, \sigma_{0n}^2 \mathbf{I}_{p_n})$$

We model the data as

$$\mathbf{x}_i \overset{iid}{\sim} \mathrm{N}_{p_n}(\mathbf{0}, \boldsymbol{\Sigma}_n), \quad \boldsymbol{\Sigma}_n = \boldsymbol{\Lambda}_n\boldsymbol{\Lambda}_n^{\top} + \sigma_n^2 \mathbf{I}_{p_n} \qquad (1)$$

Pati et al. (2014): Setup

ASSUMPTION. There exist sequences of positive real numbers c_n, s_n with $c_n \lesssim s_n$ (i.e., $\exists C > 0$ s.t. $c_n \leq C s_n$) such that:

(A1) Each column of $\boldsymbol{\Lambda}_{0n}$ belongs to $l_0[s_n; p_n] := \{\boldsymbol{\lambda} \in \mathbb{R}^{p_n} : |\mathrm{supp}(\boldsymbol{\lambda})| \leq s_n\}$. i.e. a loading vector corresponding to each factor is s_n-sparse.

(A2) $\exists\ \sigma_0^{(1)}$ s.t. $\sigma_0^{(1)} \leq \sigma_{0n}^2 \leq c_n$

(A3) $\left\| \frac{1}{c_n} \boldsymbol{\Lambda}_{0n}^\top \boldsymbol{\Lambda}_{0n} - \mathbf{I}_{k_{0n}} \right\|_2 = o\left(k_{0n} \sqrt{\frac{\log k_{0n}}{n}} \right)$

(A4) $\lim_{n \to \infty} c_n k_{0n}^{3/2} \sqrt{\frac{s_n \log p_n}{n}} \sqrt{\log n} = 0;\ k_{0n}^{3/2} \sqrt{\frac{s_n \log p_n}{n}} (\log n)^{3/2} = O(1)$

Let $\mathcal{C}_{0n}$ be the clss of covariance matrices satisfying (A0)-(A4)

Pati et al. (2014): Prior

- For the residual variance σ^2,

$$\sigma^2 \sim \mathrm{Gamma}(a, b)$$

- Prior distribution on the number of factors k, π_k is assumed to be satisfy

$$\pi_k(k > \tilde{k}) \leq \exp(-C\tilde{k}) \tag{2}$$

for all $\tilde{k} \geq \tilde{k}_0$ for some $\tilde{k}_0$ and

$$\pi_k(k = k_{0n}) \geq \exp(-Cs_n k_{0n} \log n) \tag{3}$$

For example $\mathrm{Poisson}(1)$ satisfies (2) and (3) if $\log k_{0n} \leq s_n \log n$.

Pati et al. (2014): Prior

For the factor loadings λ_{jh}, $j = 1, \ldots, p_n$, $h = 1, \ldots, k$,

- (PL1)

$$\begin{aligned} \lambda_{jh} | k, \gamma &\sim (1-\gamma)\delta_0 + \gamma g(\cdot) \\ \gamma &\sim \mathrm{Beta}(1, \kappa k_{0n} p_n + 1) \end{aligned}$$

  where g is an absolutely continuous density with exponential tails or heavier, e.g., standard double exponential density.

- (PS) for $\tilde{\boldsymbol{\lambda}} = \mathrm{vec}(\boldsymbol{\Lambda}_n) \in \mathbb{R}^{p_n k}$,

$$\begin{aligned} \tilde{\lambda}_j | k, \tau, \psi_j &\sim \mathrm{DE}(\tau \psi_j) \\ (\psi_1, \ldots, \psi_{p_n k - 1}) &\sim \mathrm{Dir}\left(\frac{\alpha}{p_n k}, \ldots, \frac{\alpha}{p_n k} \right) \\ \psi_{p_n k} &= 1 - \sum_{j=1}^{p_n k - 1} \psi_j \\ \tau &\sim \mathrm{Exp}(1/2) \end{aligned}$$

  where $\mathrm{DE}(\psi)$ denotes the double-exponential density with a density $f(x) = \frac{1}{2\psi} e^{-|x|/\psi}$.

Pati et al. (2014): Posterior contraction rate

THEOREM. Suppose $\boldsymbol{\Sigma}_{0n} \in \mathcal{C}_{0n}$ with $s_n k_{0n} \gtrsim \log p_n$, and model (1) is fitted with a prior distribution on the number of factors satisfying (2) and (3). Assume independent priors $\Pi(\sigma^2) = \mathrm{Gamma}(a, b)$ on the residual variances and $\Pi(\boldsymbol{\Lambda}|k)$=(PL1) (or (PS)) on the loadings. Then for any constant $M > 0$,

$$\lim_{n\to\infty} \mathbb{E}_{\Sigma_{0n}} \Pi_n \left(\|\Sigma_n - \Sigma_{0n}\|_2 > M\epsilon_n \middle| \mathbf{x}^{(n)} \right) = 0$$

where

$$\epsilon_n = c_n k_{0n}^{3/2} \sqrt{\frac{s_n \log p_n}{n}} \sqrt{\log n}.$$

REMARK. The posterior contraction rate obtained in the above is equal to the minimax rate up to a $\sqrt{\log n}$ term.

Knowles and Ghahramani (2011)

- Infinite factor model:

$$\mathbf{x}_i \overset{iid}{\sim} \mathrm{N}_p(\mathbf{0}, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Omega}$$

  where $\boldsymbol{\Lambda} \in \mathbb{R}^{p\times\infty}$ and $\boldsymbol{\Omega} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_p^2)$.

- IBP prior:

$$\begin{aligned}
\lambda_{jh}|\gamma_{jh}, \tau_h &\sim (1-\gamma_{jh})\delta_0 + \gamma_{jh}\mathrm{N}(0, \tau_h^{-1}) \\
\gamma_{jh}|\alpha &\sim \mathrm{IBP}(\alpha) \\
\alpha &\sim \mathrm{Gamma}(a_\alpha, b_\alpha) \\
\tau_h &\sim \mathrm{Gamma}(a_\tau, b_\tau) \\
(\sigma_j^2)^{-1} &\sim \mathrm{Gamma}(a_\sigma, b_\sigma),
\end{aligned}$$

- Let K^+ denote the effective factor dimension of $[\boldsymbol{\Gamma}]$, that is the largest index K^+ so that $\gamma_{jh} = 0$ for all $k > K^+$, $j = 1, \cdots, p$. Then for fixed α

$$K^+ \sim \mathrm{Poisson}\left(\alpha \sum_{j=1}^{p} \frac{1}{j}\right)$$

  which safisfies (2) and (3).

Knowles and Ghahramani (2011): Inference

Use Gibbs sampling, but with MH steps for sampling new factors.

Sample γ_{jh} For $j = 1, \ldots, p$, $h = 1, \ldots, K^+$,

- The ratio of posterior probabilities for γ_{jh} being 1 or 0 is given by

$$\frac{P(\gamma_{jh} = 1 | \mathbf{X}, -)}{P(\gamma_{jh} = 0 | \mathbf{X}, -)} = \frac{P(\gamma_{jh} = 1 | -)}{P(\gamma_{jh} = 0 | -)} \frac{P(\mathbf{X} | \gamma_{jh} = 1, -)}{P(\mathbf{X} | \gamma_{jh} = 0, -)} = \frac{m_{-j,h}}{p - m_{-j,h}} \sqrt{\frac{\tau_h}{\tau^*_{jh}}} \exp\left(\frac{1}{2} \tau^*_{jh} (\mu^*_{jh})^2\right),$$

where $m_{-j,h} = \sum_{l \neq j} \gamma_{lh}$ and

$$\tau^*_{jh} = \frac{1}{\sigma^2_j} \sum_{i=1}^{n} f^2_{ih} + \tau_h, \quad \mu^*_{jh} = \frac{1}{\sigma^2_j (\tau^*_{jh})^2} \sum_{i=1}^{n} f_{ih} (x_{ij} - \boldsymbol{\lambda}^\top_{j,-h} \mathbf{f}_{i,-h})$$

Sample λ_{jh} For $j = 1, \ldots, p$, $h = 1, \ldots, K^+$,

- If $\gamma_{jh} = 1$, sample $\lambda_{jh} \sim \mathrm{N}(\mu^*_{jh}, (\tau^*_{jh})^{-1})$.

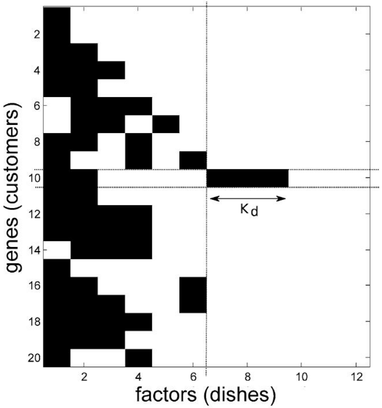
Knowles and Ghahramani (2011): Inference

Adding new factor (Sample κ_j, update K^+) For $j = 1, \ldots, p$,

- Let κ_j be the number of columns of $\mathbf{\Gamma} = (\gamma_{jh})$ which contain 1 only in row j, that is, the number of features which are active only for dimension j.
- The new κ_j^* and the corresponding $1 \times \kappa_j^*$ loading vector $\boldsymbol{\lambda}^*$ are proposed by a MH step with proposal distribution

$$
\begin{aligned}
J(\kappa_j^*) &= (1-\pi)\mathrm{Poisson}\,(\phi\alpha) + \pi I(\kappa_j^* = 1) \\
J(\boldsymbol{\lambda}^* | \kappa_j^*) &= \mathrm{N}_{\kappa_j^*}(0, \tau_h^{-1}\mathbf{I}_{\kappa_j^*}).
\end{aligned}
$$

where π, ϕ are tuned parameters.



Knowles and Ghahramani (2011): Inference

Sample α

- Sample the IBP parameter α from

$$p(\alpha|-) = \mathrm{Gamma}\left(a_\alpha + K^+, b_\alpha + \sum_{j=1}^{p} \frac{1}{j}\right),$$

where K_+ is the number of nonzero columns of $\mathbf{Z}$.

Sample $\mathbf{f}_n$ For $i = 1, \ldots, n$,

- Sample the latent factors $\mathbf{f}_i$ from

$$\begin{aligned} p(\mathbf{f}_i|-) &\propto p(\mathbf{x}_i|\mathbf{f}_i, -)p(\mathbf{f}_i) \\ &= \mathrm{N}((\mathbf{\Lambda}^\top\mathbf{\Omega}^{-1}\mathbf{\Lambda} + \mathbf{I})^{-1}\mathbf{\Lambda}^\top\mathbf{\Omega}^{-1}\mathbf{x}_i, (\mathbf{\Lambda}^\top\mathbf{\Omega}^{-1}\mathbf{\Lambda} + \mathbf{I})^{-1}). \end{aligned}$$

Sample τ_h For $h = 1, \ldots, K^+$,

- Sample the factor precisions τ_h from

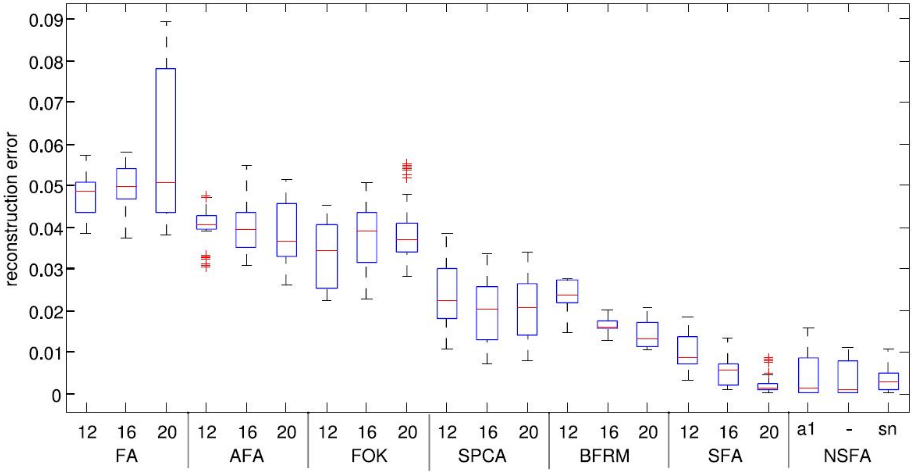
$$p(\tau_h|-) = \mathrm{Gamma}\left(a_\tau + \frac{m_h}{2}, b_\tau + \sum_{j=1}^{p} \lambda_{jh}^2\right)$$

Sample σ_j^2 For $j = 1, \ldots, p$,

- Sample the noise variances σ_j^2 from

$$p((\sigma_j^2)^{-1}|-) = \mathrm{Gamma}\left(a_\sigma + \frac{n}{2}, b_\sigma + \sum_{i=1}^{n}(x_{ij} - \boldsymbol{\lambda}_j^\top\mathbf{f}_i)^2\right)$$

Knowles and Ghahramani (2011): Simulation



Ročková and George. (2017)

- Infinite factor model:

$$\mathbf{x}_i \stackrel{iid}{\sim} \mathrm{N}_p(\mathbf{0}, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Omega}$$

  where $\boldsymbol{\Lambda} \in \mathbb{R}^{p\times\infty}$ and $\boldsymbol{\Omega} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_p^2)$.

- Spike and slab Lasso prior:

$$\begin{aligned} \lambda_{jh}|\gamma_{jh} &\sim (1-\gamma_{jh})\mathrm{DE}(\psi_{0k}) + \gamma_{jh}\mathrm{DE}(\psi_1) \\ \gamma_{jh} &\sim \mathrm{IBP}(\alpha) \\ (\sigma_j^2)^{-1} &\sim \mathrm{Gamma}(a_\sigma, b_\sigma), \end{aligned}$$

  with $\psi_1 \gg \psi_{0k}$. $\mathrm{DE}(\psi)$ denotes the double-exponential density with a density $f(x) = \frac{1}{2\psi}e^{-|x|/\psi}$.

- ψ_{0k}, ψ_1 and α are fixed (e.g., $\psi_{0k} = 1/20$, $\psi_1 = 1000$, $\alpha = 1/p$).
- If $\psi_{0k} \to 0$, we obtain a point mass mixture prior.

Appendix

LEMMA 1. If $\boldsymbol{\Sigma}_{0n} \in \mathcal{C}_{0n}$, then for $\eta_n = \sqrt{s_n k_{0n}/n}$,

$$\Pi_n(\|\boldsymbol{\Sigma}_n - \boldsymbol{\Sigma}_{0n}\| \leq \eta_n) \geq \exp(-C s_n k_{0n} \log n)$$

LEMMA 2. Let

$$e_n = s_n k_{0n} \log n, \quad t_n = C e_n^2, \quad \delta_n = \epsilon_n/(e_n t_n), \quad \delta_n' = \delta_n/(p_n e_n)$$

and

$$W_n = \{|\mathrm{supp}_{\delta_n'}(\boldsymbol{\Lambda}_n) \leq H e_n, \|\boldsymbol{\Lambda}_n\|_1 \leq t_n, \sigma^2 \leq t_n)\}, \quad V_n = \{k \leq C e_n\}$$

Then there exist the event $A_n \in \sigma(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ with $\mathbb{P}_{\boldsymbol{\Sigma}_{0n}}(A_n) \to 1$ and constants $H, C > 0$ such that

$$\lim_{n\to\infty} \mathbb{E}_{\boldsymbol{\Sigma}_{0n}}[\Pi_n(W_n^c \cap V_n | \mathbf{x}^{(n)}) \mathbf{1}_{A_n}] = 0$$

$$\lim_{n\to\infty} \mathbb{E}_{\boldsymbol{\Sigma}_{0n}}[\Pi_n(V_n^c | \mathbf{x}^{(n)}) \mathbf{1}_{A_n}] = 0$$