

# 비모수 인자 모형

김성민

서울대학교  
통계학과, 베이지통계 연구실

2024. 08. 07

# 목차

- 1 참고문헌
- 2 인자모형
- 3 비모수 선형인자모형
- 4 식별성
- 5 결과

# 목차

- 1 참고문헌
- 2 인자모형
- 3 비모수 선형인자모형
- 4 식별성
- 5 결과

- Kaiser, H. F. (1958). [The varimax criterion for analytic rotation in factor analysis.](#)  
*Psychometrika*, 23(3):187–200
- Poworoznek, E., Ferrari, F., and Dunson, D. (2021). [Efficiently resolving rotational ambiguity in bayesian matrix sampling with matching.](#)  
*arXiv preprint arXiv:2107.13783*
- Xu, M., Herring, A. H., and Dunson, D. B. (2023). [Identifiable and interpretable nonparametric factor analysis.](#)  
*arXiv preprint arXiv:2311.08254*

# 목차

- 1 참고문헌
- 2 인자모형
- 3 비모수 선형인자모형
- 4 식별성
- 5 결과

# 인자모형(Factor Model)

## 인자모형 정의

$x_1, \dots, x_n \in \mathbb{R}^p$  를 고려하자. 다음과 같은 모형을 생각할 수 있다.

$$x_i = g(\eta_i) + \epsilon_i, \quad i = 1, \dots, n$$

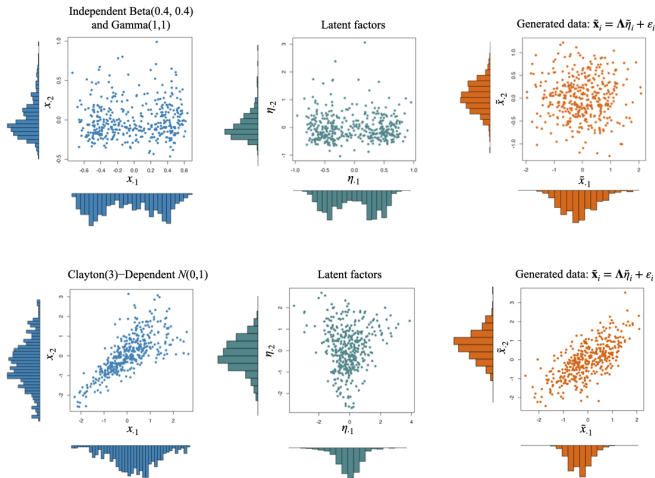
여기서  $\eta_i \in \mathbb{R}^K$ ,  $K \leq n$  는 저차원 잠재 요인이고,  $g : \mathbb{R}^K \rightarrow \mathbb{R}^p$  사상(mapping) 함수이다.  $\epsilon_i$  는  $p$ 차원 잔차이다.

$$x_i = \Lambda \eta_i + \epsilon_i, \quad \eta_i \sim N(0, I_k), \quad \epsilon_i \sim N(0, \Sigma), \quad i = 1, \dots, n$$

- $\Lambda : p \times K$  인자 로딩 행렬
- $\Sigma$  : 잔차의 공분산 행렬 (일반적으로 대각행렬 가정)

$$\implies x_i \sim N(0, \Lambda \Lambda^T + \Sigma)$$

# 정규 선형 인자모형의 한계



- 비선형 관계 설명 불가.
- 비정규 데이터 설명 불가.



- 비선형 인자 모형 ( $g$ 가 비선형)

$$x_i = g(\eta_i) + \epsilon_i, \quad i = 1, \dots, n$$

- 식별성 문제 ( $g$ 와  $\eta_i$ )
- 해석의 어려움
- $g : \mathbb{R}^k \rightarrow \mathbb{R}^p$  추정을 위해 많은 데이터가 필요
- Distributional shift :  $\eta_i$ 에 가정한 분포와 경험적 누적분포가 달라서 모형의 성능이 떨어짐.

# 정규 선형 인자모형의 대안

- 선형 인자모형에서 잠재요인을 비모수적으로 모델링

$$\begin{aligned}x_i &= \Lambda \eta_i + \epsilon_i, \quad i = 1, \dots, n \\ \eta_{ik} &= g_k(u_{ik}), \quad k = 1, \dots, K.\end{aligned}$$

- $u_{i1}, \dots, u_{iK}$  의 일부가 동일하도록 하여 요인간의 의존성을 부여 (비선형 관계 설명)

# 목차

- 1 참고문헌
- 2 인자모형
- 3 비모수 선형인자모형**
- 4 식별성
- 5 결과

# 비모수 선형인자모형(Nonparametric Linear Factor Model)

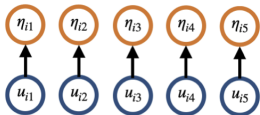
$$\begin{aligned}x_i &= \Lambda \eta_i + \epsilon_i, \quad \epsilon_i \sim N(0, \Sigma), \quad i = 1, \dots, n, \\ \eta_{ik} &= g_k(u_{il_k}), \quad k = 1, \dots, K, \\ u_{il} &\stackrel{iid}{\sim} U(0, 1), \quad l = 1, \dots, m, \quad m \leq K.\end{aligned}$$

- $g_k : [0, 1]$ 에서  $\mathbb{R}$ 로의 증가함수
- $\eta_{ik} : \text{잠재위치(latent location)} u_{ik}$ 를 잠재사상(latent mapping)  $g_h$ 를 통해 변환한 잠재요인(latent factor)

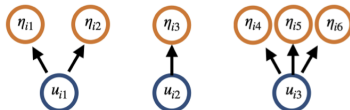
# 독립성분분석(Independent Component Analysis)

$$x_i = \Lambda \eta_i + \epsilon_i$$

- $\eta_i$ 가 서로 독립이고 비정규분포를 따르면 ICA 모형이라고 함.
- 비선형 관계를 설명하지 못함
- NIFTY (nonparametric linear factor analysis) 모형을 통해서 비선형 관계 설명 가능



(a) ICA models.



(b) NIFTY models.

# Distributional Shift (정규 선형인자모형)

$$x_i = \Lambda \eta_i + \epsilon_i, \quad i = 1, \dots, n$$

- $\eta_i$ 의 경험적 누적분포함수가  $N(0, I_p)$ 와 다름
- $\eta_i$ ,  $\Gamma$ ,  $\Sigma$ 에 대한 추정이 달라짐.
- $x_i$ 의 경험적 공분산 :  $\Lambda \text{cov}(\eta_i, i = 1, \dots, n) \Lambda^T + \Sigma$
- $x_{i1} \sim \text{Beta}(0.4, 0.4)$ ,  $x_{i2} \sim \text{Gamma}(1, 1)$  인 경우를 고려하자.

+0.14	-0.01
-0.01	+0.95

(a) Empirical covariance of  $\mathbf{x}_i$  in the data.

+0.35	+0.49
+0.49	+2.07

(b) Posterior estimation via  $\hat{\Lambda} \hat{\Lambda}^T + \hat{\Sigma}$ .

+0.11	+0.06
+0.06	+0.98

(c) Posterior estimation via  $\hat{\Lambda} \text{cov}(\hat{\eta}_i, i = 1, \dots, N) \hat{\Lambda}^T + \hat{\Sigma}$ .

# Distributional Shift (비모수 선형인자모형)

- $u$ 의 분포가 독립인 균등분포에서 벗어나 데이터에 치우쳐짐.
- $u$ 의 사전분포에 제약을 부여하여 분포가 균등분포가 되게 해야 함.

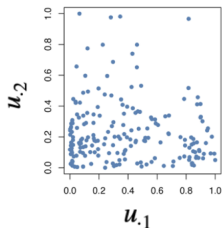
$$\Pi_{u, I}(u_{1I}, \dots, u_{nI}) = \prod_{i=1}^n 1(u_{iI} \in [0, 1]) \exp(-\nu \mathcal{W}_2(U_I, U))$$

- $\mathcal{W}_2(U_I, U)$  는  $u, I$ 의 경험적 누적분포와  $U(0, 1)$  사이의 와서스타인-2 거리이다.

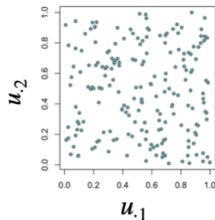
$$\mathcal{W}_2(F, G) = \left( \int_0^1 |F^{-1}(z) - G^{-1}(z)|^2 dz \right)^{1/2}$$

- $\nu, n \rightarrow \infty$  이면  $u$ 의 사전분포가 독립인 균등분포로 수렴함이 알려져 있다.

# Distributional Shift (비모수 선형인자모형)



(a) One posterior sample of the latent locations without the constraint relaxation term.



(b) One posterior sample of the latent locations with the constraint relaxation term ( $\nu = 1000$ ).



# 목차

- 1 참고문헌
- 2 인자모형
- 3 비모수 선형인자모형
- 4 식별성**
- 5 결과

# 식별성(Identifiability)

$$\begin{aligned}x_i &= \Lambda \eta_i + \epsilon_i, \quad \epsilon_i \sim N(0, \Sigma), \quad i = 1, \dots, n, \\ \eta_{ik} &= g_k(u_{il_k}), \quad k = 1, \dots, K, \\ u_{il} &\stackrel{iid}{\sim} U(0, 1), \quad l = 1, \dots, m, \quad m \leq K.\end{aligned}$$

- Latent locations :  $u_{il}$
- Residual covariance :  $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_p^2)$
- Loadings and factors :  $\Lambda$  and  $g$ 
  - General Identifiability

$$\mathbb{P}(x|\Lambda, g) = \mathbb{P}(x|\Lambda', g') \text{ iff } \Lambda = \Lambda' R \text{ and } g = R^T g'$$

for some rotation matrix  $R \in \mathbb{R}^{K \times K}$

- Strict Identifiability

$$\mathbb{P}(x|\Lambda, g) = \mathbb{P}(x|\Lambda', g') \text{ iff } \Lambda = \Lambda' \text{ and } g = g'$$

## 가정1

각 인자  $u_{ij}$ 에 대해서 다음을 만족하는  $j$ 가 존재한다.

$$\mathbb{E}[x_{ij}|u_i] = \lambda_{jh}g_h(u_{ij}), \quad \forall i.$$

- $u_{ij}$ 의 식별성 증명을 위해 필요

## 가정2

인자 로딩 행렬  $\Gamma$ 는 full column rank 이고, 각 열의 노름이 고정되어 있다.

- 양의 실수  $a$ 에 대해서  $(\lambda_{jh}, g_h)$ 와  $(a\lambda_{jh}, \frac{1}{a}g_h)$  동일한 모형을 표현함을 방지

## 가정3

가정 1에서  $\mathbb{E}[x_{ij}|u_i] = \lambda_{jh}g_h(u_{il})$  가 만족한다면,

$$x_{ij}|u_i \sim N(\lambda_{jh}g_h(u_{il}), \sigma_j^2)$$

이 성립함을 알 수 있다. 이 때,  $\sigma_j^2$ 이 알려져 있음을 가정한다.

- $x_i = \eta_i + \epsilon = (\eta_i + \epsilon/2) + \epsilon/2$  처럼 다양한 방식으로 표현됨을 방지

- 가정 1,2,3 하에서  $u_{ij}$ ,  $\Sigma$  의 식별성과  $\Lambda, g$  의 일반(general) 식별성을 증명할 수 있다.

$$\mathbb{P}(x|\Lambda, g) = \mathbb{P}(x|\Lambda', g') \text{ iff } \Lambda = \Lambda' \text{ and } g = g'$$

- 강한(strict) 식별성을 위해 추가 조건(혹은 작업)이 필요하다. (Poworoznek et al. [2021])
  - 로딩행렬에 하삼각행렬과 같은 제약을 부여하여 샘플링
  - $\Lambda$ 의 특정 성분을 0으로 선택
  - 사후표본을 사후처리

- Xu et al. [2023]가 제안한 사후처리-사후표본

---

**Algorithm 1:** post-processing the posterior samples to solve ambiguity

---

**Input:**  $(\Lambda^{(1)}, \mathbf{g}^{(1)}, \Sigma^{(1)}), \dots, (\Lambda^{(M)}, \mathbf{g}^{(M)}, \Sigma^{(M)})$

**for**  $m = 1, \dots, M$  **do**

**for**  $k = 1, \dots, K$  **do**

        Orthogonalize the  $k$ th partition  $\Lambda^{(m)k}$  and tackle the label and  
        sign switching with MatchAlign algorithm (Poworoznek et al., 2021);

        Obtain rotation matrix  $\mathbf{R}^{(m)k}$  and  $\Lambda_h^{(m)k} \leftarrow \Lambda^{(m)k} \mathbf{R}^{(m)k}$ ;

        Rotate  $\mathbf{g}^k$  and obtain  $\mathbf{g}^{(m)k} \leftarrow (\mathbf{R}^{(m)k})^T \mathbf{g}^{(m)k}$ .

**for**  $h = 1, \dots, H$  **do**

$\Lambda_h^{(m)} \leftarrow \Lambda_h^{(m)} / \|\Lambda_h^{(m)}\|_2$ ;

$\mathbf{g}^{(m)} \leftarrow \mathbf{g}^{(m)} \|\Lambda_h^{(m)}\|_2$ ;

---

- Poworoznek et al. [2021]에서 제안한 MatchAlign 알고리즘 이용

---

**Algorithm 1** MatchAlign algorithm to solve rotational ambiguity in matrix valued parameters.

---

**Input:**  $\{\Lambda^{(t)} : t = 1, \dots, T\}$

1. **for**  $t$  in  $1 : T$  **do**

    | Orthogonalize  $\Lambda^{(t)}$  using Varimax and output  $\tilde{\Lambda}^{(t)}$

**end**

2. Choose a pivot  $\Lambda^P$  from  $\{\tilde{\Lambda}^{(t)} : t = 1, \dots, T\}$

3. **for**  $t$  in  $1 : T$  **do**

**for**  $j$  in  $1 : k$  **do**

        | Compute normed differences between  $c_j^{(t)}$  and  $\Lambda^P$  and  $-\Lambda^P$  columns

        | Retain the  $j^{th}$  column having minimum norm value

        | Drop the matched column and its negative from the pivot

**end**

    | Reorder and re-sign

**end**

---

- VariMax 처리 후에 각  $\Lambda$ 에 대해 적절한  $S = \text{diag}(s_1, \dots, s_K)$ ,  $s_i \in \{-1, 1\}$ 과 치환행렬  $Q$  선택

$$\arg \min_{Q, S} \left\| \Lambda Q S - \Lambda^P \right\|_F$$

- Pivot  $\Lambda^P$ 는 사후표본 중에서 하나 선택.



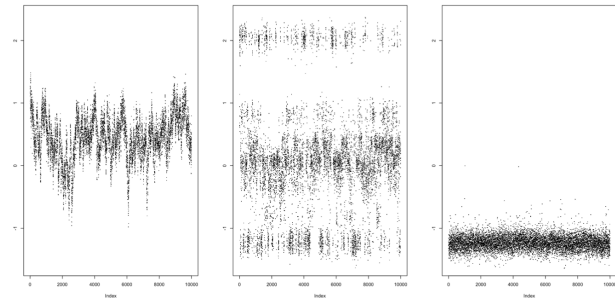
- 'Variance is maximized' 의 약자로 Kaiser [1958] 가 제안
- 다음을 만족하는 회전행렬  $R \in O_p$ 를 찾는 것이 목표이다.

$$R_{\text{VariMax}} = \arg \max_R \sum_{j=1}^k \left[ \frac{1}{p} \sum_{i=1}^p (\Lambda R)_{ij}^4 - \left( \frac{1}{p} \sum_{i=1}^p (\Lambda R)_{ij}^2 \right)^2 \right]$$

- $R_{\text{VariMax}}$ 은  $\Lambda$ 를 회전변환하여 각 열의 분산의 합이 가장 크게 만든다.

# MatchAlign 알고리즘 적용 결과

- 왼쪽은  $\Lambda$  한 원소에 대한 Traceplot 이고, 중앙은 Varimax 적용 후, 오른쪽은 MatchAlign 적용 후의 결과이다.



# 목차

- 1 참고문헌
- 2 인자모형
- 3 비모수 선형인자모형
- 4 식별성
- 5 결과

$$x_i = \Lambda \eta_i + \epsilon_i, \quad \epsilon_i \sim N(0, \Sigma), \quad i = 1, \dots, n,$$

$$\eta_{ik} = g_k(u_{il_k}), \quad k = 1, \dots, K,$$

$$\Pi_{u_{\cdot l}}(u_l) = \prod_{i=1}^n 1(u_{il} \in [0, 1]) \exp(-\nu \mathcal{W}_2(U_l, U)), \quad l = 1, \dots, m$$

$$\lambda_{jk} \sim N(0, \tau \gamma_{jk} \sigma_j^2)$$

$$\gamma_{jh} \sim C^+(0, 1)$$

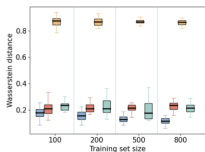
$$\tau \sim C^+(0, 1)$$

- $\Lambda$ 에 horseshoe 사전분포 부여
- 충분히 큰 수의 잠재인자 개수  $K$ 를 선택한 후에 수축사전 분포를 통해 의미가 없는 요인 제거.

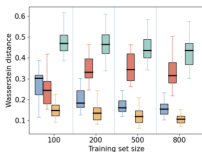
- 다음의 4 가지 잠재인자 모형 사용.
  - NITFY : 논문에서 제안한 모형
  - GP-LVM (gaussian process latent variable model) : 비선형 인자모형으로 회귀 함수에 GP 사전분포 부여
  - VAE (variational autoencoder) : 비선형 인자모형으로 DNN 이용
  - PPCA (probabilistic pca) : 정규 선형 인자모형

# 결과

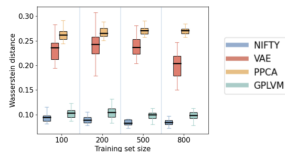
- 주어진 데이터셋의 분포를 얼마나 잘 학습하나를 out-of-sample 를 통해 확인



(a) 2-dimensional data generated from independent Gamma and Beta distribution.



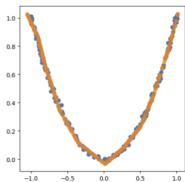
(b) 20-dimensional data generated from a Gaussian linear factor.



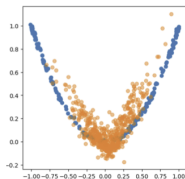
(c) 10-dimensional data generated from two latent curves.

- NIFTY 가 대부분 데이터에서 가장 좋은 성능 지님.
- PPCA 는 정규 선형 인자모형을 통해서 생성된 데이터에서 가장 좋은 성능 보임.
- GPLVM 과 VAE 는 NIFTY 다음의 성능을 보이나, curve 데이터에서는 VAE 성능 안 좋음.

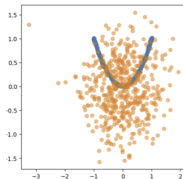
- Curve 데이터에서 NIFTY와 GPLVM 이 좋은 성능 가짐



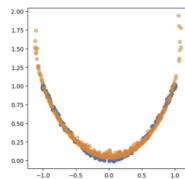
(a) NIFTY



(b) VAE



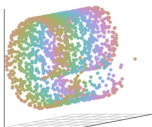
(c) PPCA



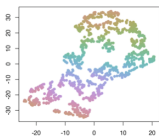
(d) GPLVM

# 차원 축소 결과

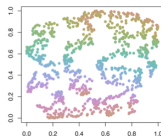
- Latent location  $u_{il}$ 는 저차원 단위 큐브에서의 좌표로 생각할 수 있다.
- $u_{il}$ 를 통한 차원 축소 결과는 아래와 같다.
- ( $u_{il}$ 이 아닌  $g(u_{il})$ ?)



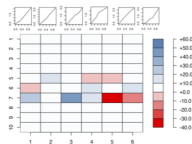
(a) The three-dimensional Swiss roll generated from uniform variables  $u_i$  and  $v_i$ .



(b) 2D visualization from diffusion maps.



(c) 2D visualization from NIFTY latent locations.



(d) Loading matrix and latent mappings.



- $u$ 에 제약이 있는 사전분포?
  - $n$ 이 충분히 크지 않으면 distributional shift 방지 불가
  - 와서스타인이 포함된 부분이  $n, \nu$ 가 커지면 거의 무시되면서도, distributional shift를 억제할 수 있나?
- NIFTY가 커버할 수 있는 데이터 분포?
  - $\Lambda$  : 데이터를 사상시킬 초평면(hyperplane) 결정
  - $\eta$  : 초평면에서의 데이터 분포 결정
- $\Lambda$ 의 식별성 문제로 인해서 수렴속도에 영향을 미침(Multimodality)
  - $\eta \sim F$ 로 모델링하면  $\Lambda$ 를 하삼각행렬로 제한하면 괜찮을수도?

# References I

- Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3):187–200.
- Poworoznek, E., Ferrari, F., and Dunson, D. (2021). Efficiently resolving rotational ambiguity in bayesian matrix sampling with matching. *arXiv preprint arXiv:2107.13783*.
- Xu, M., Herring, A. H., and Dunson, D. B. (2023). Identifiable and interpretable nonparametric factor analysis. *arXiv preprint arXiv:2311.08254*.