성긴 요인 모형

김성민

서울대학교
통계학과, 베이즈통계 연구실

2024. 12. 09

목차

목차

공분산

- 다변량 통계에서 공분산 추정은 변수들간의 상관성을 이해하는데 중요하다.
- 표본 공분산

$$S = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})^T$$

  은 대표적인 공분산 추정량이다.
- Johnstone [2001] 와Johnstone and Lu [2009] 는 다음과 같은 고차원 가정하에서 표본 공분산의 가장 큰 고유값과 고유벡터가 불일치성을 가짐을 보였다.

$$\frac{p}{n} \to c > 0.$$

공분산의 수축 추정량

- 수축 추정량
  - Stein [1975], Haff [1980], Ledoit and Wolf [2004] 와Bodnar et al. [2014] 는 다으.
  - Karoui [2008], Ledoit and Wolf [2012] and Lam [2016] suggested orthogonal invariants shrinkage estimator based on Random matrix theory.
- 수축 통계량은 다음과 같은 조건하에서 얻어진다.
  - 고유값이 $n \to \infty$ 임에도 발산하지 않는다.
- 고유값이 발산하는 고차원 공분산을 추정하기 위해서 다음과 같은 추가적은 가정이 필요하다.

구조적 가정

- 성긴 공분산 (or 정밀도 행렬)
  - Frequentist : Bickel and Levina [2008a] Cai and Zhou [2012] Cai et al. [2016]
  - Bayesian : Banerjee and Ghosal [2015] Lee et al. [2022] Lee and Lee [2023]
- 밴디드 공분산 (or 정밀도 행렬)
  - Frequentist : Bickel and Levina [2008b] Cai et al. [2010] Bien et al. [2016]
  - Bayesian : Banerjee and Ghosal [2014] Lee et al. [2023]
- 성김 혹은 밴디드 가정은 공분산에 대한 사전 지식이 필요하다.

스파이크 공분산

- 스파이크 공분산은 Johnstone [2001] 에 의해 처음으로 제안되었다.
- 다음과 같은 고유분해를 고려하자.

$$\Sigma = \Gamma \Lambda \Gamma^T, \quad \Lambda = diag(\lambda_1, \cdots, \lambda_p)$$

- 스파이크 공분산 Σ 는 k개의 큰 고유값과 비교적 작은 나머지 고유값으로 이루어져 있다.

$$\lambda_1 > \cdots > \lambda_k >> \lambda_{k+1} > \cdots > \lambda_p.$$

목표

- 고차원 가정하에서, 역위샤트 사전분포를 사용하면 고유값 추정량이 퍼지는 경향성을 가진다.
- Berger et al. [2020] 는 다음과 같은 수축역위샤트(Shrinkage Inverse Wishart) 사전분포를 제안하였다.

$$\pi(\Sigma|a, H) \propto \frac{1}{\left|\prod_{i<j}(\lambda_i - \lambda_j)\right|} |\Sigma|^{-a} etr(-\frac{1}{2}\Sigma^{-1}H),$$

  여기서 $\lambda_1, \cdots, \lambda_p$는 Σ의 고유값이다.
- $\left|\prod_{i<j}(\lambda_i - \lambda_j)\right|$ 가 고유값이 퍼지는 현상을 억제한다.
- 모의실험에서 SIW 가 좋은 성능을 보임을 확인하였지만, 이론적인 근거가 제시되지는 않았다.

모형

$$X_i \overset{iid}{\sim} N(0, \Sigma), \quad \forall i = 1, \cdots, n$$

$$\pi(\Sigma) \propto \frac{\prod_{i=1}^{p} \lambda_i^{-a_i}}{\left|\prod_{i<j}(\lambda_i - \lambda_j)\right|} etr(-\frac{1}{2}\Sigma^{-1}H)$$

- 다음과 같은 Σ의 고유분해를 고려하자.:

$$\Sigma = \Gamma\Lambda\Gamma^T$$

  여기서 $\Lambda = diag(\lambda_1, \cdots, \lambda_p)$ 이고 Γ 는 $p \times p$ 고유벡터 행렬이다.
- Σ에서 (Λ, Γ)로 변환의 자코비안은 다음과 같다.

$$\left|\frac{\partial\Sigma}{\partial(\Gamma, \Lambda)}\right| = \left|\prod_{i<j}(\lambda_i - \lambda_j)\right|.$$

가정

고유구조의 사후수렴 속도를 얻기 위해서 다음과 같은 가정들을 고려하자.

1. $n/p \to 0$.
2. 상수 $c_0, C_0 > 0$가 존재하여 참 공분산의 고유값이 다음을 만족한다.
$$\lambda_{0,1} > \cdots > \lambda_{0,k} > C_0 > \lambda_{0,k+1} > \cdots > \lambda_{0,p} > c_0.$$
3. k 개의 큰 고유값이 상수 $\delta_0 > 0$에 의해서 잘 구분된다.:
$$\frac{\lambda_{0,j} - \lambda_{0,j+1}}{\lambda_{0,j}} \geq \delta_0, \quad \forall j = 1, \cdots, k.$$
4. k 개의 큰 고유값에 대해서, $d_j = \dfrac{p}{n\lambda_{0,j}}$ 가 유계이다.

목차

사후수렴속도

Theorem

Assume the assumptions on slide 10 (A1 $\sim$ A4). If $a_1, \cdots, a_k$ are constant, $a_{k+1}, \cdots, a_n \approx n$, $a_{n+1}, \cdots, a_p \succ p$ and $H = hI_p$ with $h \prec n$, then the posterior convergence rates of eigenvalues satisfy

$$\mathbb{E}\Big[\frac{\lambda_i - \lambda_{0,i}}{\lambda_{0,i}}\Big|\mathbb{X}\Big] = O\Big(\frac{p}{n\lambda_{0,i}}\Big) + O\big(n^{-\frac{1}{2}+\delta}\big),$$

for $i = 1, \cdots, k$ and any small $\delta > 0$.

- This prior achieves the same convergence rate as the Inverse Wishart prior.

Theorem

If $2a_i - 4 = \dfrac{n\hat{\lambda}_n}{\hat{\lambda}_i - \hat{\lambda}_n}, \quad i = 1, \cdots, k$ *where* $\hat{\lambda}_i$ *is* i*th eigenvalue of sample covariance* $S = \dfrac{1}{n}X^T X$*, then the posterior convergence rates of eigenvalues satisfy*

$$\mathbb{E}\Big[\frac{\lambda_i - \lambda_{0,i}}{\lambda_{0,i}}\Big|\mathbb{X}\Big] = O\Big(\frac{1}{\lambda_{0,i}}\sqrt{\frac{p}{n}}\Big) + O(n^{-\frac{1}{2}+\delta}),$$

for $i = 1, \cdots, k$ *and any small* $\delta > 0$.

- If the condition $\lambda_{0,i} \lesssim \dfrac{p}{\sqrt{n}}$ holds, then the convergence rate is faster than the prior described in Theorem1.
- The rate is equal to the shrinkage eigenvalues proposed by Wang and Fan [2017].

$$\hat{\lambda}_j^S = \hat{\lambda}_j - \frac{p}{np - nk - pk}\Big(tr(S) - \sum_{i=1}^{k}\hat{\lambda}_i\Big)$$

- ξ_i : Eigenvector corresponding to λ_i, the ith eigenvalue of the Σ.
- $\xi_{0,i}$: Eigenvector corresponding to $\lambda_{0,i}$, the ith eigenvalue of the true covariance.

Theorem

Under assumption on Theorem 2, the posterior convergence rates of eigenvectors satisfy

$$\mathbb{E}\left[1 - \left|\xi_{0,i}^T \xi_i\right|^2 \middle| \mathbb{X}\right] = O\left(\frac{p}{n\lambda_{0,i}}\right),$$

for $i = 1, \cdots, k$.

목차

근사 요인 모형

- 다음과 같은 모형을 고려하자.

$$Y_i = B\eta_i + \epsilon_i,\ \epsilon_i \sim N(0, \Sigma_u), \quad i = 1, \cdots, n$$

  - η_i : $N(0, I_K)$를 따르는 잠재 요인(latent factor).
  - $B \in \mathbb{R}^{p\times K}$: 잠재요인 η_i와 관측 변수 Y_i를 연관시키는 로딩 행렬.
  - ϵ_i : 특이 오차
  - Σ_e : 특이 공분산

요인 모형

- 다음과 같은 모형을 고려하자.

$$
\begin{aligned}
Y_i &\sim N(0, BB^T + \Sigma_u) \\
BB^T + \Sigma_u &= \Gamma\Lambda\Gamma^T + \Sigma_u \\
(\Gamma, \Lambda) &\sim \prod_{i=1}^{p} \lambda_i^{-a_i} etr(-\frac{1}{2}\Gamma\Lambda^{-1}\Gamma^T H) \\
\Sigma_u &\sim \pi(\Sigma_u)
\end{aligned}
$$

  - $\Gamma \in \mathbb{R}^{p\times k}$: i번째 열이 BB^T의 i번째 고유벡터
  - $\Lambda \in \mathbb{R}^{k\times k}$: 대각성분이 BB^T의 k개의 고유값인 대각행렬
  - Σ_e : 특이 행렬

사후분포

$$\pi(\Gamma, \Lambda, \Sigma_u | Y)$$
$$\propto P(Y|\Gamma, \Lambda, \Sigma_u)\pi(\Gamma, \Lambda)\pi(\Sigma_u)$$
$$\propto \left|\Gamma\Lambda\Gamma^T + \Sigma_u\right|^{-\frac{n}{2}} \cdot \prod_{i=1}^{k} \lambda_i^{-a_i} \cdot etr(-\frac{n}{2}(\Gamma\Lambda\Gamma^T + \Sigma_u)^{-1}S - \frac{1}{2}\Gamma\Lambda^{-1}\Gamma^T H) \cdot \pi(\Sigma_u)$$

- $\bar{\Gamma} = Q\Gamma, \quad (S = nQWQ^T)$
- $\Sigma_u^* = Q^T\Sigma_u Q$

$$\pi(\bar{\Gamma}, \Lambda, \Sigma_u^* | Y)$$
$$\propto \left|\bar{\Gamma}\Lambda\bar{\Gamma}^T + \Sigma_u^*\right|^{-\frac{n}{2}} \cdot \prod_{i=1}^{k} \lambda_i^{-a_i} \cdot etr(-\frac{n}{2}(\bar{\Gamma}\Lambda\bar{\Gamma}^T + \Sigma_u^*)^{-1}W - \frac{h}{2}\bar{\Gamma}\Lambda^{-1}\bar{\Gamma}^T) \cdot \pi(\Sigma_u)$$

샘플링

- 깁스 추출법을 이용

1. Sampling Λ : Inverse Wishart 에서 샘플링 후 Rejection Sampling (or MH)

$$q(\Lambda|\cdot) \propto \prod_{i=1}^{k} \lambda_i^{-n/2-a_i} etr(-\frac{1}{2}\Lambda^{-1}(n\bar{\Gamma}^T W \bar{\Gamma} + hI_k))$$

2. Sampling $\bar{\Gamma}$: Matrix Bingham distribution 에서 샘플링 후 Rejection Sampling (or MH)

$$q(\bar{\Gamma}|\cdot) \propto etr(-\frac{1}{2}\Lambda^{-1}\bar{\Gamma}^T(nW + hI_p)\bar{\Gamma})$$

3. Sampling Σ_u : $\pi(\Sigma_u)$ 에서 샘플링 후 Rejection Sampling (or MH)

Matrix Bingham 분포의 샘플링

$$P(X) \propto etr(BX^TAX)I(X \in O(p \times k))$$

- $X = \{Nz, X_{[,-1]}\}, \quad z \in \mathcal{S}_{p-k+1}$
- N :Orthonormal basis for the null space of $X_{[,-1]}$

$$\begin{aligned} P(z|X[,-1]) &\propto \exp\left(b_{1,1}z^TN^TANz\right)I(z \in \mathcal{S}_{p-k+1}) \\ &\propto \exp\left(z^T\tilde{A}z\right)I(z \in \mathcal{S}_{p-k+1}) \end{aligned}$$

- 위의 분포는 vector Bingham distribution 이므로 샘플링 가능
- 각 열에 대해서 샘플링 반복

목차

모형

- 다음과 같은 모형을 고려하자

$$
\begin{aligned}
Y_i &\sim N(0, BB^T + \sigma^2 I_p) \\
BB^T + \sigma^2 I_p &= \Gamma\Lambda\Gamma^T + \sigma^2 I_p \\
(\Gamma, \Lambda | R) &\sim \prod_{i=1}^{p} \lambda_i^{-a_i} etr(-\frac{1}{2}\Gamma\Lambda^{-1}\Gamma^T H) I(\Gamma = \Gamma \otimes R) \\
R &\sim \prod_{i,j} Ber(R_{ij}; \theta) I(supp(R_{.i}) \geq k) \\
\sigma^2 &\sim \pi(\sigma^2)
\end{aligned}
$$

  - $\Gamma \in \mathbb{R}^{p\times k}$: i번째 열이 BB^T의 i번째 고유벡터
  - $R \in \mathbb{R}^{p\times k}$: Γ의 성김에 대한 이진 행렬
  - $\Lambda \in \mathbb{R}^{k\times k}$: 대각성분이 BB^T의 k개의 고유값인 대각행렬

샘플링

$$\pi(\Gamma_{\cdot i}, R_{\cdot i} | R_{-i}, \Gamma_{-i}, \cdot) \propto \exp\left(-\frac{1}{2}\Gamma_{\cdot i}^T B \Gamma_{\cdot i}\right) I(\Gamma_{\cdot i} = \Gamma_{\cdot i} \otimes R_{\cdot i}) \times \prod_j Ber(R_{ji}; \theta) I(supp(R_{\cdot i}) \geq k)$$

- 따라서, 먼저 $\pi(R_{\cdot i} | R_{-i}, \Gamma_{-i}, \cdot)$ 에서 샘플링

  $\pi(R_{\cdot i} | R_{-i}, \Gamma_{-i}, \cdot)$

  $\propto \int_{\mathcal{S}_{p-1}} \exp\left(-\frac{1}{2}\Gamma_{\cdot i}^T B \Gamma_{\cdot i}\right) I(\Gamma_{\cdot i} = \Gamma_{\cdot i} \otimes R_{\cdot i}) d\Gamma_{\cdot i} \cdot \prod_j Ber(R_{ji}; \theta) I(supp(R_{\cdot i}) \geq k)$

  $= \int_{\mathcal{S}_{|R_{\cdot i}|-1}} \exp\left(-\frac{1}{2}\Gamma_{r_i i}^T B_{r_i, r_i} \Gamma_{r_i i}\right) d\Gamma_{r_i i} \cdot \prod_j Ber(R_{ji}; \theta) I(supp(R_{\cdot i}) \geq k)$

- 이후에 $\pi(\Gamma_{\cdot i} | R, \Gamma_{-i}, \cdot)$ 에서 샘플링

$$\pi(\Gamma_{\cdot i} | R, \Gamma_{-i}, \cdot) \propto \exp\left(-\frac{1}{2}\Gamma_{\cdot i}^T B \Gamma_{\cdot i}\right) I(\Gamma_{\cdot i} = \Gamma_{\cdot i} \otimes R_{\cdot i})$$

Hypergeometric function of matrix argument

$$ {}_1F_1\left(\frac{1}{2}; \frac{p}{2}; Z\right) = \int_{\mathcal{S}_{p-1}} \exp\left(x^T Z x\right) dx, \quad Z : \text{diagonal matrix} $$

- Bingham 분포의 상수텀이 Hypergeometric function 형태가 된다.
- 샘플링을 위해서는 Hypergeometric function 처리가 필요

한계점

$$\pi(\Gamma_{\cdot i}|R, \Gamma_{-i}, \cdot) \propto \exp\left(-\frac{1}{2}\Gamma_{\cdot i}^T B \Gamma_{\cdot i}\right) I(\Gamma_{\cdot i} = \Gamma_{\cdot i} \otimes R_{\cdot i})$$

- 만약에 $supp(R_{\cdot i}) \geq k$ 가 아니라면 Γ가 직교행렬이 되도록 하는 $\Gamma_{\cdot i}$ 가 존재하지 않을 수 있다.
- 하지만 해당 가정은 $\Gamma_i = e_i$ 인 경우를 설명하지 못한다.
- $supp(R_{\cdot i}) < k$ 인 경우에는 $R_{\cdot i}$ 의 가능한 집합이 Γ_{-i} 에 의존한다.
  - $o_i = \{j : R_{ji} = 1\}$
  - $\Gamma_{o_i,-i}$ 의 rank 가 $\min(n(o_i), k)$보다 작아야 한다. $(null(\Gamma_{o_i,-i}) \neq \phi)$

References I

Sayantan Banerjee and Subhashis Ghosal. 2014. Posterior convergence rates for estimating large precision matrices using graphical models. *Electronic Journal of Statistics* 8 (2014), 2111–2137.

Sayantan Banerjee and Subhashis Ghosal. 2015. Bayesian structure learning in graphical models. *Journal of Multivariate Analysis* 136 (2015), 147–162.

James O Berger, Dongchu Sun, and Chengyuan Song. 2020. Bayesian analysis of the covariance matrix of a multivariate normal distribution with a new class of priors. *The Annals of statistics* 48, 4 (2020), 2381–2403.

Peter J Bickel and Elizaveta Levina. 2008a. Covariance Regularization by Thresholding. *The Annals of Statistics* (2008), 2577–2604.

Peter J Bickel and Elizaveta Levina. 2008b. Regularized Estimation of Large Covariance Matrices. *The Annals of Statistics* (2008), 199–227.

Jacob Bien, Florentina Bunea, and Luo Xiao. 2016. Convex banding of the covariance matrix. *J. Amer. Statist. Assoc.* 111, 514 (2016), 834–845.

References II

Taras Bodnar, Arjun K Gupta, and Nestor Parolya. 2014. On the strong convergence of the optimal linear shrinkage estimator for large dimensional covariance matrix. *Journal of Multivariate Analysis* 132 (2014), 215–228.

T Tony Cai, Zhao Ren, and Harrison H Zhou. 2016. Estimating structured high-dimensional covariance and precision matrices: Optimal rates and adaptive estimation. *Electronic Journal of Statistics* 10 (2016), 1–59.

T Tony Cai, Cun-Hui Zhang, and Harrison H Zhou. 2010. OPTIMAL RATES OF CONVERGENCE FOR COVARIANCE MATRIX ESTIMATION. *The Annals of Statistics* 38, 4 (2010), 2118–2144.

T Tony Cai and Harrison H Zhou. 2012. OPTIMAL RATES OF CONVERGENCE FOR SPARSE COVARIANCE MATRIX ESTIMATION. *The Annals of Statistics* (2012), 2389–2420.

LR Haff. 1980. Empirical Bayes estimation of the multivariate normal covariance matrix. *The Annals of Statistics* 8, 3 (1980), 586–597.

Iain M Johnstone. 2001. On the distribution of the largest eigenvalue in principal components analysis. *The Annals of statistics* 29, 2 (2001), 295–327.

References III

Iain M Johnstone and Arthur Yu Lu. 2009. On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.* 104, 486 (2009), 682–693.

Noureddine El Karoui. 2008. Spectrum Estimation for Large Dimensional Covariance Matrices Using Random Matrix Theory. *The Annals of Statistics* (2008), 2757–2790.

Clifford Lam. 2016. Nonparametric eigenvalue-regularized precision or covariance matrix estimator. *Annals of Statistics* 44, 3 (2016), 928–953.

Olivier Ledoit and Michael Wolf. 2004. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis* 88, 2 (2004), 365–411.

Olivier Ledoit and Michael Wolf. 2012. Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics* 40, 2 (2012), 1024–1060.

References IV

Kyoungjae Lee, Seongil Jo, and Jaeyong Lee. 2022. The beta-mixture shrinkage prior for sparse covariances with near-minimax posterior convergence rate. *Journal of Multivariate Analysis* 192 (2022), 105067.

Kwangmin Lee and Jaeyong Lee. 2023. Post-processed posteriors for sparse covariances. *Journal of Econometrics* 236, 1 (2023), 105475.

Kwangmin Lee, Kyoungjae Lee, and Jaeyong Lee. 2023. Post-processed posteriors for banded covariances. *Bayesian Analysis* 18, 3 (2023), 1017–1040.

Charles Stein. 1975. Estimation of a covariance matrix. In *39th Annual Meeting IMS, Atlanta, GA, 1975*.

Weichen Wang and Jianqing Fan. 2017. Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *Annals of statistics* 45, 3 (2017), 1342.