성긴 스파이크 공분산 모형

김성민

서울대학교
통계학과, 베이즈통계 연구실

2023. 02. 28

목차

CONTENTS

표본 공분산

- 표본 공분산이 고차원($\frac{p}{n} \to c \in (0, \infty]$) 인 경우에 좋지 않은 추정량이다.
  - 가장 큰 표본 고유값이 일치성을 가지지 않음. El Karoui [2008]
  - 표본 고유벡터가 일치성을 가지지 않음. Johnstone and Lu [2009]
  - $p > n$ 인 경우에는 표본 공분산의 역행렬이 존재하지 않음.

Fixed p, large n

- 표본 공분산 (S_p) 의 고유값이 Σ_p 고유값의 좋은 추정량이다..
- l_i : S_p 의 고유값 ($l_1 > l_2 > \cdots$)
- λ_i : Σ_p 의 고유값 ($\lambda_1 > \lambda_2 > \cdots$).

$$\sqrt{n}(l_i - \lambda_i) \xrightarrow{d} N(0, \lambda_i^2)$$

where X_i are normally distributed.

Large n, large p

- 가장 단순한 경우인 $\Sigma_p = I_p$ 를 고려하자.
- X_i 가 iid 이고 유한의 4차 모먼트를 가질 때, 만약 $\dfrac{p}{n} \to \gamma$이면 다음이 성립한다.
$$l_1 \to (1 + \sqrt{\gamma})^2 \quad \text{a.s.}$$
- 만약 $n = p$ 이면 l_1 는 4로 수렴한다.

- Paul [2007] 는 표본 공분산의 고유벡터에 대해 집중.
- $\Sigma_p = diag(l_1, l_2, \ldots, l_M, 1, \ldots, 1)$
- $p/n \to \gamma \in (0,1)$ 이면 표본 고유벡터가 일치성을 가지지 않음..
  - If $l_v > 1 + \sqrt{\gamma}$,

$$|< p_v, e_v >| \to \sqrt{\left(1 - \frac{\gamma}{(l_v - 1)^2}\right) \Big/ \left(1 + \frac{\gamma}{l_v - 1}\right)}.$$

  - If $l_v \le 1 + \sqrt{\gamma}$,

$$|< p_v, e_v >| \to 0.$$

표본 공분산

- 공분산 추정을 위해서는 성김, 밴디드와 같은 구조적 가정 필요.
    - 공분산 공간에 대한 제한이 없는 경우의 (베이즈) 최대최소 수렴 속도 보임. Lee and Lee [2018]
    - 고차원에서는 비제약하에서 적절한 추론 불가.

스파이크 공분산

- 스파이크 공분산은 다음과 같이 정의된다.

$$\Sigma = V \Lambda V^T + \sigma^2 I_p$$

  - $\Lambda = diag(\lambda_1, \ldots, \lambda_r)$
  - V는 $V^T V = I_r$ 을 만족하는 $p \times r$ 행렬.
- r개의 가장 큰 고유값은 $\lambda_i + \sigma^2$, $i = 1, \ldots, r$ 이다.

CONTENTS

스파이크 공분산 모형

스파이크 공분산 모형

$$X_i \sim N(0, \Sigma) \quad \Sigma = V\Lambda V^T + \sigma^2 I_p \quad \text{for } i = 1, \ldots, n.$$
$$\Lambda = diag(\lambda_1, \ldots, \lambda_r), \; V^T V = I_r$$

- 스파이크 공분산 모형은 다음과 같은 잠재 요인 모형으로 생각할 수 있다.

$$X_i = V\Lambda^{1/2} W_i + Z_i, \quad W_i \sim N_r(0, I), \; Z_i \sim N_p(0, I).$$

성김 조건

$$X_i = V\Lambda^{1/2}W_i + Z_i, \quad W_i \sim N_r(0, I), \; Z_i \sim N_p(0, I).$$

- Paul [2007] 에서 고차원에서 샘플 고유벡터가 일치성을 가지지 않음을 보임.
- 고유벡터에 성김 조건을 부여.
  - V의 각 행마다 성김 조건(약한 l_q 제한) 부여 - Cai et al. [2013]
  - 고유벡터의 0이 아닌 원소의 개수 조건 부여 - Gao and Zhou [2015]
  - V의 몇몇 행을 0 벡터로 제한 - Xie [2021]

추정 목표

- Principal subspace 의 추정.

$$\left\|\hat{V}\hat{V}^T - VV^T\right\| \to 0$$

- 공분산 추정.

$$\left\|\hat{\Sigma} - \Sigma\right\| \to 0$$

CONTENTS

선행 연구1

- Pati, D., Bhattacharya, A., Pillai, N. S., and Dunson, D. (2014).

  Posterior contraction in sparse bayesian factor models for massive covariance matrices.
  The Annals of Statistics, 42(3):1102–1130

$$\begin{aligned} X_i &= V\Lambda^{1/2}W_i + Z_i \\ &= V\Lambda^{1/2}U^T W_i + Z_i, \quad \text{where } U^T U = I \\ &= BW_i + Z_i, \quad W_i \sim N_r(0, I),\ Z_i \sim N_p(0, \sigma^2 I). \end{aligned}$$

- 요인 차원 r에 다음과 같은 사전 분포 부여.

$$\pi_r(r > j) \leq \exp(-Cj), \quad \pi_r(r = r_{0n}) \geq \exp(-Cs_n r_{0n} \log n).$$

- Loading matrix B 의 각 원소에 Shrinkage prior.
- 공분산에 대한 사후수렴속도 보임.

선행 연구2

- Gao, C. and Zhou, H. H. (2015). Rate-optimal posterior contraction for sparse pca.
  The Annals of Statistics, pages 785–818

$$X_i = V\Lambda^{1/2}W_i + Z_i \quad W_i \sim N_r(0, I),\ Z_i \sim N_p(0, \sigma^2 I).$$

- 직교 행렬($V\Lambda^{1/2}$)에 대한 사전분포 제안.
- Sparsity 와 Rank 동시에 추정.

선행 연구3

- Xie, F., Cape, J., Priebe, C. E., and Xu, Y. (2022). Bayesian sparse spiked covariance model with a continuous matrix shrinkage prior. *Bayesian Analysis*, 17(4):1193–1217

$$\begin{aligned} X_i &= V\Lambda^{1/2}W_i + Z_i \\ &= V\Lambda^{1/2}U^T W_i + Z_i, \quad \text{where } U^T U = I \\ &= BW_i + Z_i, \quad W_i \sim N_r(0, I),\ Z_i \sim N_p(0, \sigma^2 I). \end{aligned}$$

- 요인 차원 r 이 알려져 있음을 가정.

선행 연구3

- B에 Matrix Spike-and-Slab Lasso 사전분포 제안.

$$(B_{1\cdot}, \ldots, B_{p\cdot}) \overset{iid}{\sim} (1-\theta)\prod_{k=1}^{r}\psi_r(b_{jk}|\lambda+\lambda_0) + \theta\prod_{k=1}^{r}\psi_1(b_{jk}|\lambda)$$

$$\lambda_0 \sim IG(1/p^2, 1)$$

$$\theta \sim Beta(1, p^{1+\kappa})$$

- $\psi_\alpha(x|\lambda)$는 다음과 같은 double Gamma 분포의 pdf이다.

$$\psi_\alpha(x|\lambda) = \frac{\lambda^{1/\alpha}}{2\Gamma(1/\alpha)}|x|^{1/\alpha-1}\exp(-\lambda|x|)$$

선행 연구

- Pati et al. [2014] : Sub-optimal (r이 증가할 때)
- Gao and Zhou [2015] : Improvement of Pati et al. [2014], Contraction rate of principal subspace.
- Xie et al. [2022] : Optimal , Known r.

CONTENTS

스파이크 고유값

- 공분산의 고유값에 대한 추정 : Marchenko-Pastur Equation
- 스파이크 공분산의 고유값에 대한 추정
  - $\Sigma = diag(l_1, l_2, \ldots, l_M, 1, \ldots, 1)$: Paul [2007]
  - $\Sigma = \begin{pmatrix} T & 0 \\ 0 & I_p \end{pmatrix}$: Bai and Yao [2008]
  - $\Sigma = \begin{pmatrix} T & 0 \\ 0 & V_p \end{pmatrix}$: Bai and Yao [2012]
  - Generalized spiked covariance : Dey and Lee [2019]

$$\Sigma = Udiag(l_1, l_2, \ldots, l_m, \beta_1, \ldots, \beta_{p-m})U^T$$

스파이크 고유값 추정

- Dey and Lee [2019] 의 정리1 에서 스파이크 고유값에 대한 추정값 제안.

$$|f_F(d_k) - \lambda_k| \xrightarrow{p} 0.$$

where $f_F(d_k) \approx d_k \left(1 + \frac{\gamma}{p-m} \sum_{i=m+1}^{p} \frac{d_i}{d_k - d_i}\right)^{-1}$

  - λ_k, d_k : 실제 및 표본 고유값.
  - m : Generalized spiked 의 개수
  - $\frac{p}{n} \to \gamma$

스펙트럼 분포 정의

- Σ_p가 고유값 $(\lambda_1, \ldots, \lambda_p)$를 가진다고 하자.
- **population spectral distribution** H_p 를 다음과 같이 정의한다.

$$dH_p(x) = \frac{1}{p}\sum_{i=1}^{p}\delta_{\lambda_i}(x).$$

- 마찬가지로, **empirical spectral distribution** F_p 를 다음과 같이 정의한다.

$$dF_p(x) = \frac{1}{p}\sum_{i=1}^{p}\delta_{l_i}(x).$$

- **Zero-excluded empirical spectral distribution** 는 다음과 같이 정의된다.

$$dF_n(x) = \frac{1}{n \wedge p}\sum_{i=1}^{n \wedge p}\delta_{l_i}(x).$$

스펙트럼 분포의 예

- $dH_p = (1 - \frac{1}{p})\delta_1 + \frac{1}{p}\delta_2.$
  - Σ_p가 (p-1)개의 고유값 1, 1개의 고유값 2를 가짐.
- H_p weakly converges to H_∞, with $dH_\infty = \delta_1$.

MP 분포

Zero-excluded MP distribution

$\gamma > 0$ 이 주어졌을 때, Zero-excluded MP 분포는 다음과 같은 밀도함수를 가진다.

$$f_\gamma(x;\sigma^2) = \frac{1}{2\pi\sigma^2}\frac{1}{x(\gamma \wedge 1)}\sqrt{(x-\sigma^2 h_-)(\sigma^2 h_+ - x)} \cdot 1\{\sigma^2 h_- < x < \sigma^2 h_+\},$$

- $\frac{p}{n} \to \gamma \in (0, \infty)$ 일 때, 적절한 조건하에서 다음이 성립한다. (Götze and Tikhomirov [2004])

$$\mathbb{E}[\sup_x |F_n(x) - F_\gamma(x)|] = O(n^{-1/2}).$$

스파이크 고유값 개수 추정

- Ke et al. [2021]가 Bulk eigenvalue matching analysis (BEMA) 제안.
- 다음의 두 단계로 스파이크 고유값 개수 추정
  1. Bulk eigenvalue 를 이용하여 MP(Marchenko-Pastur) 밀도함수 추정.
  2. 추정된 MP 밀도함수의 토대를 벗어나는 고유값 확인.

CONTENTS

성긴 스파이크 공분산 모형

스파이크 공분산 모형

$$X_i \sim N(0, \Sigma) \quad \Sigma = V \Lambda V^T + \sigma^2 I_p \quad \text{for } i = 1, \ldots, n.$$

$$\Lambda = diag(\lambda_1, \ldots, \lambda_r), \ V^T V = I_r$$

- 직교정규 행렬 (Orthonormal matrix) V 의 사전분포
- 요인의 차원 r의 사전분포
- 고유값 $(\lambda_1, \ldots, \lambda_r)$의 사전분포

목표

- 고유값과 차원 r 에 대한 적절한 사전분포 : Random matrix theory 기반
- Optimal : 공분산, principal subspace, 고유값
- Efficient computation.

참고문헌 I

Bai, Z. and Yao, J. (2012). On sample eigenvalues in a generalized spiked population model. *Journal of Multivariate Analysis*, 106:167–177.

Bai, Z. and Yao, J.-f. (2008). Central limit theorems for eigenvalues in a spiked population model. In *Annales de l'IHP Probabilités et statistiques*, volume 44, pages 447–474.

Cai, T. T., Ma, Z., and Wu, Y. (2013). Sparse pca: Optimal rates and adaptive estimation. *The Annals of Statistics*, 41(6):3074.

Dey, R. and Lee, S. (2019). Asymptotic properties of principal component analysis and shrinkage-bias adjustment under the generalized spiked population model. *Journal of multivariate analysis*, 173:145–164.

El Karoui, N. (2008). Spectrum estimation for large dimensional covariance matrices using random matrix theory. *The Annals of Statistics*, 36(6):2757–2790.

참고문헌 II

Gao, C. and Zhou, H. H. (2015). Rate-optimal posterior contraction for sparse pca. *The Annals of Statistics*, pages 785–818.

Götze, F. and Tikhomirov, A. (2004). Rate of convergence in probability to the marchenko-pastur law. *Bernoulli*, 10(3):503–548.

Johnstone, I. M. and Lu, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *Journal of the American Statistical Association*, 104(486):682–693.

Ke, Z. T., Ma, Y., and Lin, X. (2021). Estimation of the number of spiked eigenvalues in a covariance matrix by bulk eigenvalue matching analysis. *Journal of the American Statistical Association*, pages 1–19.

Lee, K. and Lee, J. (2018). Optimal bayesian minimax rates for unconstrained large covariance matrices. *Bayesian Analysis*, 13(4):1215–1233.

참고문헌 III

Pati, D., Bhattacharya, A., Pillai, N. S., and Dunson, D. (2014). Posterior contraction in sparse bayesian factor models for massive covariance matrices. *The Annals of Statistics*, 42(3):1102–1130.

Paul, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, pages 1617–1642.

Xie, F. (2021). Euclidean representation of low-rank matrices and its statistical applications. *arXiv preprint arXiv:2103.04220*.

Xie, F., Cape, J., Priebe, C. E., and Xu, Y. (2022). Bayesian sparse spiked covariance model with a continuous matrix shrinkage prior. *Bayesian Analysis*, 17(4):1193–1217.