Relative perturbation bounds for empirical covariance operators

Moritz Jirak, Universität Wien
(and Martin Wahl, Humboldt-Universität zu Berlin)

October 14, 2022

PCA in high dimensions

Let $X, X_1, \ldots, X_n$ be i.i.d. centered random variables taking values in a p-dimensional Hilbert $\mathcal{H}$ space with (empirical) covariance operator

$$\Sigma = \mathbb{E} X \otimes X \qquad \text{and} \qquad \hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} X_i \otimes X_i$$

$(\lambda_j)_{j=1}^{p}$: non-increasing sequence of eigenvalues of Σ
$(u_j)_{j=1}^{p}$: sequence of eigenvectors of Σ
$(P_j)_{j=1}^{p}$: sequence of spectral projectors of Σ, $P_j = u_j \otimes u_j$

Challenges: p increases in n (the same order as n) or even $p = \infty$

PCA in high dimensions

Let $X, X_1, \ldots, X_n$ be i.i.d. centered random variables taking values in a p-dimensional Hilbert $\mathcal{H}$ space with (empirical) covariance operator

$$\Sigma = \mathbb{E} X \otimes X \qquad \text{and} \qquad \hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} X_i \otimes X_i$$

$(\hat{\lambda}_j)_{j=1}^{p}$: non-increasing sequence of eigenvalues of $\hat{\Sigma}$
$(\hat{u}_j)_{j=1}^{p}$: sequence of eigenvectors of $\hat{\Sigma}$
$(\hat{P}_j)_{j=1}^{p}$: sequence of spectral projectors of $\hat{\Sigma}$, $\hat{P}_j = \hat{u}_j \otimes \hat{u}_j$

Challenges: p increases in n (the same order as n) or even $p = \infty$

How close are $\hat{\lambda}_j, \hat{P}_j$ to their population counterparts λ_j, P_j?

High-dimensional phenomena in the spiked model

Theorem 1 (Baik & Silverstein '06, Paul '07, Nadler '08, etc.)

Suppose that X is Gaussian and that

- $\Sigma = \mathrm{diag}(\lambda_1, 1, \ldots, 1) \in \mathbb{R}^{p \times p}$ *with* $\lambda_1 > 1$ *fixed*

Then, as $p/n \to \gamma > 0$, *almost surely,*

$$\hat{\lambda}_1 \to \begin{cases} \lambda_1 + \gamma \frac{\lambda_1}{\lambda_1 - 1} & \text{if } \frac{\gamma}{(\lambda_1 - 1)^2} < 1 \\ (1 + \sqrt{\gamma})^2 & \text{otherwise} \end{cases}$$

$$\|\hat{P}_1 - P_1\|_2^2 \to \begin{cases} c_1 \frac{\gamma \lambda_1}{(\lambda_1 - 1)^2} & \text{if } \frac{\gamma}{(\lambda_1 - 1)^2} < 1 \\ 2 & \text{otherwise} \end{cases}$$

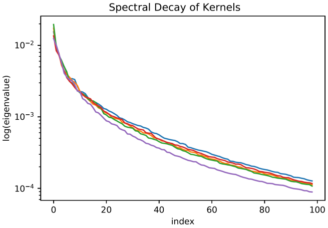
Related results hold for more complicated spiked models.

- Extensions to general eigenvalue settings?
- Extensions to more general distributional settings?

Covariance operators in infinite dimensions

- Key feature in functional data analysis and kernel-based learning: **spectral decay** of Σ
- polynomial decay: $\lambda_j = j^{-\alpha}$, $j \geq 1$
- exponential decay: $\lambda_j = e^{-\alpha j}$, $j \geq 1$
- theory less developed:

$$\frac{\hat{\lambda}_j - \lambda_j}{\lambda_j} \xrightarrow{w} \mathcal{N}(0,1), \quad \text{which } j\,?$$



Liang & Rakhlin '20: decay on MNIST

- S. Fischer and I. Steinwart. "Sobolev norm learning rates for regularized least-squares algorithms". In: *J. Mach. Learn. Res.* ()
- P. L. Bartlett et al. "Benign overfitting in linear regression". In: *Proc. Natl. Acad. Sci. USA* ()
- P. Hall and J.L. Horowitz (Feb. 2007). "Methodology and convergence rates for functional linear regression". In: *The Annals of Statistics* 35.1, pp. 70–91

Functional regression

- Given $(X_k)_{k\in\mathbb{Z}}$, $(Y_k)_{k\in\mathbb{Z}}$, consider

$$X_k = \Phi(Y_k) + \epsilon_k, \quad k \in \mathbb{Z},$$

where Φ is an unknown linear operator, and $(\epsilon_k)_{k\in\mathbb{Z}}$ is a noise sequence.

- A common estimator for Φ is (with sample size n)

$$\widehat{\Phi}^b(\cdot) = \sum_{j=1}^{b} \frac{1}{n} \sum_{k=1}^{n} \frac{\langle Y_k, \widehat{u}_j^y \rangle X_k}{\widehat{\lambda}_j^y} \langle \widehat{u}_j^y, \cdot \rangle, \quad b = b_n \to \infty.$$

Optimal choice of b_n (depends on $(\lambda_j^y)_{j\in\mathbb{N}}$) leads to minimax rates, but requires good control of $\widehat{\lambda}_j^y$ and $\widehat{u}_j^y$ for $j \le b_n$.

- P. Hall and J.L. Horowitz (Feb. 2007). "Methodology and convergence rates for functional linear regression". In: *The Annals of Statistics* 35.1, pp. 70–91

Functional AR(1)

- If $Y_k = X_{k-1}$, functional regression becomes the functional AR(1) model

$$X_k = \Phi(X_{k-1}) + \epsilon_k, \quad k \in \mathbb{Z}.$$

- More generally, we can consider AR(q) in $\mathcal{H}$ processes

$$X_k = \sum_{i=1}^{q} \Phi_i(X_{k-i}) + \epsilon_k, \quad k \in \mathbb{Z},$$

where Φ_j are unknown linear operators.

- Can even let $q = \infty$.
- In all those cases, estimation crucially depends on $\hat{u}_j$, $\hat{\lambda}_j$.

Classical math tools for thinking about spectral methods

Weyl bound

We have $|\hat{\lambda}_j - \lambda_j| \le \|E\|_\infty$ with $\|\cdot\|_\infty$ operator norm and $E = \hat{\Sigma} - \Sigma$

Davis-Kahan $\sin\Theta$ bound

We have

$$\|\hat{P}_j - P_j\|_2 \le \frac{2\sqrt{2}\|E\|_\infty}{g_j}$$

with spectral gap $g_j = \min(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1})$ and HS norm $\|\cdot\|_2$

- **applied to** kernel PCA (Blanchard et al. '05), functional PCA (Horváth & Kokoszka '12), sparse PCA (Vu & Lei '13), robust PCA (Minsker & Wei '17), distributed PCA (Fan et al. '19)

Classical math tools for thinking about spectral methods

Definition 2

The reduced resolvent of Σ at λ_j is defined by $R_j = \sum_{k \neq j} \frac{1}{\lambda_k - \lambda_j} P_k$

Linear perturbation expansion

If

$$\gamma_j := \frac{\|E\|_\infty}{g_j} < 1/2$$

then

$$\|\hat{P}_j - P_j + R_j E P_j + P_j E R_j\|_2 \leq \frac{4\gamma_j^2}{1 - 2\gamma_j}$$

More generally $\hat{\lambda}_j$, $\hat{P}_j$ admit a Taylor series in E provided that $\gamma_j < 1/2$

- T. Hsing and R. Eubank (2015). *Theoretical foundations of functional data analysis*. John Wiley & Sons

Relative idea for thinking about spectral methods

Relative $\sin\Theta$ bound (J. & W.)

We have

$$\|\hat{P}_j - P_j\|_2 \leq C\|(|R_j|^{1/2} + g_j^{-1/2}P_j)E(|R_j|^{1/2} + g_j^{-1/2}P_j)\|_\infty$$

for some absolute constant $C > 0$.

Previous work and different approach:

- A. Mas and F. Ruymgaart. "High-dimensional principal projections". In: *Complex Anal. Oper. Theory* ()

Relative idea for thinking about spectral methods

Let $\mathcal{J} = \{1, \ldots, J\}$ (write j if $\mathcal{J} = \{j\}$). We write

$$P_{\mathcal{J}} = \sum_{j \in \mathcal{J}} P_j, \qquad P_{\mathcal{J}^c} = \sum_{k \in \mathcal{J}^c} P_k, \qquad R_{\mathcal{J}^c} = \sum_{k \in \mathcal{J}^c} \frac{1}{\lambda_k - \lambda_j} P_k.$$

$$\delta_{\mathcal{J}} = \delta_{\mathcal{J}}(E) := \left\| \left(|R_{\mathcal{J}^c}|^{1/2} + g_J^{-1/2} P_{\mathcal{J}} \right) E \left(|R_{\mathcal{J}^c}|^{1/2} + g_J^{-1/2} P_{\mathcal{J}} \right) \right\|_{\infty}.$$

Moreover, for a Hilbert-Schmidt operator A on $\mathcal{H}$ we define

$$L_{\mathcal{J}} A = \sum_{j \in \mathcal{J}} \sum_{k \in \mathcal{J}^c} \frac{1}{\lambda_j - \lambda_k} (P_k A P_j + P_j A P_k).$$

Relative idea for thinking about spectral methods

Theorem 3 (J. & W.)

We have

$$\|P_{\mathcal{J}} - \hat{P}_{\mathcal{J}}\|_2^2 \leq 32 \min(|\mathcal{J}|, |\mathcal{J}^c|)\delta_{\mathcal{J}}^2 \qquad (1)$$

and

$$\|\hat{P}_{\mathcal{J}} - P_{\mathcal{J}} - L_{\mathcal{J}} E\|_2^2 \leq 48 \min(|\mathcal{J}|, |\mathcal{J}^c|)^2 \delta_{\mathcal{J}}^4. \qquad (2)$$

Possible to replace $\delta_{\mathcal{J}}$ by $\min(\delta_{\mathcal{J}}, \delta_{\mathcal{J}^c})$.

- Eigenvalues, and eigenvectors?
- Control of γ_j and $\delta_{\mathcal{J}}$?

Effective versus relative rank setting

The effective rank (Koltchinkii & Lounici '17) and the relative rank (J. & W. '18) are defined by

$$\mathrm{e}_j(\Sigma) = \frac{\mathrm{tr}(\Sigma)}{g_j} \qquad \text{literature } \mathrm{r}\,!$$

$$\mathrm{r}_j(\Sigma) = \sum_{k \neq j} \frac{\lambda_k}{|\lambda_j - \lambda_k|} + \frac{\lambda_j}{g_j}$$

While the effective rank grows reciprocally with the gap, the relative rank remains largely unaffected

upper bounds	$\lambda_j = j^{-\alpha}$	$\lambda_j = e^{-\alpha j}$
$\mathrm{r}_j(\Sigma)$	$j \log j$	j
$\mathrm{e}_j(\Sigma)$	$j^{\alpha+1}$	$e^{\alpha j}$

Note on convexity

- Convexity condition: There is a convex function

$$\lambda : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}, \quad \text{such that} \quad \lambda(j) = \lambda_j, \tag{3}$$

  at least for j large enough.

- Exploiting the convexity, it follows that

$$\mathrm{r}_j(\Sigma) \leq C_1 \sum_{k \neq j} \frac{\lambda_k}{|\lambda_j - \lambda_k|} \leq C_2 j \log j \quad \text{and} \quad \sum_{k \neq j} \frac{\lambda_k \lambda_j}{(\lambda_j - \lambda_k)^2} \leq C j^2,$$

  where C is a constant which only depends on $\mathrm{tr}(\Sigma)$.

- The convexity condition is quite general, valid in particular for polynomial and exponential decay of eigenvalues.

- H. Cardot, A. Mas, and P. Sarda (2007). "CLT in functional linear regression models". In: *Probab. Theory Related Fields*

Relative bound for eigenvalues

$$\text{Key:} \quad \bar{\eta}_{kl} = \frac{\langle u_k, E u_l \rangle}{\sqrt{\lambda_k \lambda_l}} = \frac{\langle u_k, (\hat{\Sigma} - \Sigma) u_l \rangle}{\sqrt{\lambda_k \lambda_l}}, \quad k, l \geq 1.$$

Theorem 4 (J. & W.)

Let $j \geq 1$. Suppose that λ_j is a simple eigenvalue, meaning that $\lambda_j \neq \lambda_k$ for all $k \neq j$. Let $x > 0$ be such that $|\bar{\eta}_{kl}| \leq x$ for all $k, l \geq 1$. Suppose that

$$\mathrm{r}_j(\Sigma) = \sum_{k \neq j} \frac{\lambda_k}{|\lambda_j - \lambda_k|} + \frac{\lambda_j}{g_j} \leq 1/(3x). \tag{4}$$

Then we have

$$|\hat{\lambda}_j - \lambda_j - \lambda_j \bar{\eta}_{jj}| / \lambda_j \leq C x^2 \mathrm{r}_j(\Sigma). \tag{5}$$

Relative bound for eigenvectors

Theorem 5 (J. & W.)

Let $j \geq 1$. Suppose that λ_j is a simple eigenvalue. Let $x > 0$ be such that $|\bar{\eta}_{kl}| \leq x$ for all $k, l \geq 1$. Suppose that Condition (4) holds. Then we have

$$\left\| \hat{u}_j - u_j - \sum_{k \neq j} \frac{\sqrt{\lambda_j \lambda_k}}{\lambda_j - \lambda_k} \bar{\eta}_{jk} u_k \right\| \leq C x^2 \mathrm{r}_j(\Sigma) \sqrt{\sum_{k \neq j} \frac{\lambda_j \lambda_k}{(\lambda_j - \lambda_k)^2}} \tag{6}$$

and

$$\left| \| \hat{u}_j - u_j \|^2 - \sum_{k \neq j} \frac{\lambda_j \lambda_k}{(\lambda_j - \lambda_k)^2} \bar{\eta}_{jk}^2 \right| \leq C x^3 \mathrm{r}_j(\Sigma) \sum_{k \neq j} \frac{\lambda_j \lambda_k}{(\lambda_j - \lambda_k)^2}. \tag{7}$$

In (6) and (7), the sign of u_j is chosen such that $\langle \hat{u}_j, u_j \rangle > 0$.

Effective versus relative rank setting

Write $X = \sum_{j\geq 1} \lambda_j^{1/2} u_j \eta_j$ with Karhunen-Loéve coefficients $\eta_1, \eta_2, \ldots$

Setting 1

For some $q > 4$ we have $\sup_{j\geq 1} \mathbb{E}|\eta_j|^q \lesssim 1$

- $\gamma_j < 1/2$ w.h.p. if $\frac{1}{\sqrt{n}} \mathrm{e}_j(\Sigma) \lesssim 1$
- $\delta_j < 1/2$ w.h.p. if $\frac{1}{\sqrt{n}} \mathrm{r}_j(\Sigma) \lesssim 1$
- Control $\{|\bar{\eta}_{kl}| \leq x\}$ w.h.p., $x \approx n^{-1/2}$ (essentially).

- S. V. Nagaev (1979). "Large deviations of sums of independent random variables". In: *Ann. Probab.*
- U. Einmahl and D. Li. "Characterization of LIL behavior in Banach space". In: *Trans. Amer. Math. Soc.* ()

High-dimensional phenomena under spectral decay

Theorem 6 (J. & W.)

Let $X = X^{(n)}$ *be a sequence on r.v. in Setting 1 with covariances* $\Sigma = \Sigma^{(n)}$*. If*

$$\frac{1}{\sqrt{n}} \mathrm{r}_j(\Sigma) \to 0 \quad \textit{as} \quad n \to \infty \tag{8}$$

then

$$g_j^{-1}(\hat{\lambda}_j - \lambda_j) \xrightarrow{\mathbb{P}} 0 \tag{9}$$

$$\|\hat{P}_j - P_j\|_2 \xrightarrow{\mathbb{P}} 0 \tag{10}$$

$$(\sqrt{n}(\hat{\lambda}_j - \lambda_j/\lambda_j) \quad \textit{is tight.} \tag{11}$$

Moreover, for $j = 1$ *there is a sequence of r.v.* $X = X^{(n)}$ *in Setting 1 with covariance operators* $\Sigma^{(n)}$ *such that* (8), (9), (10) *and* (11) *are equivalent.*

Example for Setting 1

$$\text{Key:} \quad \bar{\eta}_{kl} = \frac{\langle u_k, E u_l \rangle}{\sqrt{\lambda_k \lambda_l}} = \frac{\langle u_k, (\hat{\Sigma} - \Sigma) u_l \rangle}{\sqrt{\lambda_k \lambda_l}}, \quad k, l \geq 1.$$

This can be written as a sum of (i.i.d.) random variables

$$\sqrt{n}\bar{\eta}_{kl} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\eta_{ik}\eta_{il} - \mathbb{E}\eta_{ik}\eta_{il} \right).$$

Union bound and (standard) concentration inequalities provide control of ($p = \dim(\mathcal{H})$)

$$\mathbb{P}\Big(\max_{1 \leq k, l \leq p} |\bar{\eta}_{kl}| \geq x \Big) \leq \sum_{1 \leq i, j \leq p} \mathbb{P}\Big(|\bar{\eta}_{kl}| \leq x \Big)$$

Example for Setting 1

For example, Fuk-Nagaev inequality yields

$$\mathbb{P}\Big(\max_{1\le k,l\le d} |\bar{\eta}_{kl}| \ge \frac{C\sqrt{\log n}}{\sqrt{n}} \Big) \lesssim p^2 n^{1-q/4}.$$

Using a more refined argument one can drastically reduce dependence on p here. Hence, if

$$\frac{\sqrt{\log n}}{\sqrt{n}} \max_{1\le j\le J} \mathrm{r}_j(\Sigma) \lesssim 1,$$

we get (for instance)

$$|\hat{\lambda}_j - \lambda_j - \lambda_j \bar{\eta}_{jj}|/\lambda_j \lesssim \frac{\sqrt{\log n}}{\sqrt{n}}, \quad 1 \le j \le J,$$

with high probability. No spatial dependence assumption on $(\eta_j)_{j\ge 1}$ here!

Example for Setting 1: weak dependence

Let $(\epsilon_i)_{i\in\mathbb{Z}}$ be i.i.d. For f taking values in $\mathcal{H}$, consider the Bernoulli-shift sequence

$$X_i = f\big(\epsilon_i, \epsilon_{i-1}, \dots\big), \quad i \in \mathbb{N},$$

Recall $X_i = \sum_{j\geq 1} \sqrt{\lambda}_j u_j \eta_{ij}$, where $\eta_{ij} = \lambda_j^{-1/2}\langle X_i, u_j\rangle$. ϵ_0' independent copy of ϵ_0, independent of $(\epsilon_i)_{i\in\mathbb{Z}}$. Coupling X_i' of X_i defined as

$$X_i' = f\big(\epsilon_i, \dots, \epsilon_1, \epsilon_0', \epsilon_{-1}, \dots\big), \quad i \in \mathbb{N}.$$

For $j \geq 1$, let $\eta_{ij}' = \lambda_j^{-1/2}\langle X_i', u_j\rangle$. Coupling distance

$$\theta_{iq} = \sup_{j\geq 1} \mathbb{E}^{1/p}\big|\eta_{ij} - \eta_{ij}'\big|^q. \tag{12}$$

- I. A. Ibragimov (1966). "On the accuracy of approximation by the normal distribution of distribution functions of sums of independent random variables". In: *Teor. Verojatnost. i Primenen* 11, pp. 632–655
- W. B. Wu (Jan. 2011). "Asymptotic theory for stationary processes". In: *Statistics and its Interface* 4, pp. 207–226

Corollary 7 (J. & W.)

Suppose we are in Setting 1 with $q \geq 16$. If (3) *holds, then*

$$\mathbb{E}\|\hat{P}_j - P_j\|_\infty^2 \leq \mathbb{E}\|\hat{P}_j - P_j\|_2^2 \leq Cj^2/n, \quad 1 \leq j \leq C\sqrt{n}(\log n)^{-5/2}.$$

- This result is (up to log terms) optimal in the case where $\lambda_j = j^{-\alpha-1}$, $\alpha > 0$ in a certain sense.

- This result is (up to log terms) optimal in the case where $\lambda_j = Cj^{-\alpha-1}$, $\alpha > 0$ in a certain sense.
- For such a polynomial decay, given that $\sup_{j\geq 1} \mathbb{E}|\eta_j|^{2q} \leq q!C^q$ for all $q \geq 1$, it has been shown that for any $j \geq 1$ (exists also information theoretic bound)

$$\mathbb{E}\|\hat{P}_j - P_j\|_\infty^2 \geq c(j^2/n) \wedge 1.$$

- We obtain the optimal bound for almost the whole range (up to the factor $(\log n)^{-5/2}$) where the trivial bound 2 does not apply. Moreover, only require mild conditions.
- Note: The stochastic behaviour of the scores $(\eta_j)_{j\geq 1}$ in terms of their dependence structure is irrelevant for the optimal algebraic structure conditions. In other words, this result **cannot** be improved assuming that $(\eta_j)_{j\geq 1}$ are independent.

Effective versus relative rank setting

Write $X = \sum_{j\geq 1} \lambda_j^{1/2} u_j \eta_j$ with Karhunen-Loéve coefficients $\eta_1, \eta_2, \ldots$

Sub-gaussian setting

$\eta_1, \eta_2, \ldots$ are independent and sub-Gaussian, i.e. $\sup_{j\geq 1} \|\eta_j\|_{\psi_2} \lesssim 1$

- $\gamma_j < 1/2$ w.h.p. if $\frac{\lambda_j}{g_j} \frac{\mathrm{e}_j(\Sigma)}{n} \lesssim 1$
- $\delta_j < 1/2$ w.h.p. if $\frac{\lambda_j}{g_j} \frac{\mathrm{r}_j(\Sigma)}{n} \lesssim 1$

- V. Koltchinskii and K. Lounici. "Normal approximation and concentration of spectral projectors of sample covariance". In: *Ann. Statist.* ()
- V. Koltchinskii. "Asymptotically efficient estimation of smooth functionals of covariance operators". In: *J. Eur. Math. Soc.* ()

Effective versus relative rank setting

Setting 2

(i) For some $q > 4$ we have $\sup_{j \geq 1} \mathbb{E}|\eta_j|^q \lesssim 1$

(ii) For some $m \geq 4$ we have $\mathbb{E}\eta_{i_1}\eta_{i_2}\ldots\eta_{i_m} = 0$ whenever one of the indices $i_1, \ldots, i_m \geq 1$ occurs only once

- $\gamma_j < 1/2$ w.h.p. if $\frac{\lambda_j}{g_j}\frac{\mathrm{e}_j(\Sigma)}{n} \lesssim 1$
- $\delta_j < 1/2$ w.h.p. if $\frac{\lambda_j}{g_j}\frac{\mathrm{r}_j(\Sigma)}{n} \lesssim 1$

regimes	$\lambda_j = j^{-\alpha}$	$\lambda_j = e^{-\alpha j}$
relative	$j^2 \log j \lesssim n$	$j \lesssim n$
effective	$j^{2+\alpha} \lesssim n$	$j \lesssim \log n$

High-dimensional phenomena under spectral decay

Theorem 8 (J. & W.)

Assume Setting 2 with $m = 4$.

- *If* $\frac{\lambda_j}{g_j}\frac{\mathrm{r}_j(\Sigma)}{n} \to 0$ *then* $g_j^{-1}(\hat{\lambda}_j - \lambda_j) \xrightarrow{\mathbb{P}} 0$ *and* $\|\hat{P}_j - P_j\|_2 \xrightarrow{\mathbb{P}} 0$

- Current work: If $n^{\epsilon}\frac{\lambda_j}{g_j}\frac{\mathrm{r}_j(\Sigma)}{n} \to 0$, $\epsilon > 0$ arbitrarily small, then limit theorems and much more are possible (higher order expansions).
- Subject to appropriate Assumptions, replace n^{ϵ} with something weaker ($\log^q n$, more structure).
- Can extend everything to: longrun covariance operator, autocovariance operators, robust empirical covariance operators.

Spectral decay versus spiked models

bound	$\lambda_j = j^{-\alpha-1}$	$\lambda_j = e^{-\alpha j}$
relative regime	$j^2 \log j \lesssim n$	$j \lesssim n$
$\lvert\hat{\lambda}_j - \lambda_j\rvert/\lambda_j$	$\frac{1}{\sqrt{n}} + \frac{j \log j}{n}$	$\frac{1}{\sqrt{n}} + \frac{j}{n}$
$\lVert \hat{P}_j - P_j \rVert_2$	$\frac{j}{\sqrt{n}}$	$\frac{1}{\sqrt{n}}$
effective regime	$j^{2+2\alpha} \lesssim n$	$j \lesssim \log n$

PCA and RMT	$\lambda_1 > 1 = \cdots = 1,\ \frac{p}{n} \to \gamma$	general (λ_j)
phase transition	$\frac{\gamma}{(\lambda_1 - 1)^2} < 1$	$\frac{\lambda_1}{g_1} \frac{\mathrm{r}_1(\Sigma)}{n} \lesssim 1$
eigenvalue bias	$\gamma \frac{\lambda_1}{\lambda_1 - 1}$	$\frac{\lambda_1}{n} \mathrm{r}_1(\Sigma)$

But remember: phase transitions can already occur for $\frac{\mathrm{r}_1(\Sigma)}{\sqrt{n}} \geq c$!

Quantitative limit theorems in high dimensions

- $T_j = n\|\hat{P}_j - P_j\|_2^2$ (results actually apply to $\mathcal{J}$.)
- $S_j = \|L_j(Z)\|_2^2 = \|R_j Z P_j + P_j Z R_j\|_2^2$
- Z Gaussian r.v. with $\mathrm{cov}(Z) = \mathrm{cov}(X \otimes X)$
- **uniform metric:**

$$\mathrm{U}\big(T_j, S_j\big) = \sup_{x \in \mathbb{R}} |\mathbb{P}(T_j \leq x) - \mathbb{P}(S_j \leq x)|$$

- A. Naumov, V. Spokoiny, and V. Ulyanov. "Bootstrap confidence sets for spectral projectors of sample covariance". In: *Probab. Theory Related Fields* ()
- V. Koltchinskii and K. Lounici. "Normal approximation and concentration of spectral projectors of sample covariance". In: *Ann. Statist.* ()

Quantitative limit theorems in high dimensions

Theorem 9 (J. & W.)

Assume Setting 2 with $m = 4$.

(i) *If* $\lambda_k = k^{-\alpha}$, $k \geq 1$, *then*

$$\mathrm{U}(n\|\hat{P}_j - P_j\|_2^2, \|L_j(Z)\|_2^2) \lesssim_{p,\alpha} \frac{j}{\sqrt{n}}(\log j \log n)^{3/2}$$

(ii) *If* $\lambda_k = e^{-\alpha k}$, $k \geq 1$, *then*

$$\mathrm{U}(n\|\hat{P}_j - P_j\|_2^2, \|L_j(Z)\|_2^2) \lesssim_{p,\alpha} \left(\frac{j^3 \log^3 n}{n}\right)^{1/2}$$

- A. Naumov, V. Spokoiny, and V. Ulyanov. "Bootstrap confidence sets for spectral projectors of sample covariance". In: *Probab. Theory Related Fields* ()
- V. Koltchinskii and K. Lounici. "Normal approximation and concentration of spectral projectors of sample covariance". In: *Ann. Statist.* ()

Bootstrap approximations in high dimensions

Popular and powerful method are multiplier methods, which we also employ here. Let (w_i) be an i.i.d. sequence with

$$\mathbb{E}w_i^2 = 1, \quad \mathbb{E}w_i^{2q} < \infty. \tag{13}$$

Our bootstrap method is quite simple and given below.

Algorithm 1.1 (Bootstrap)

Given (X_i) and (w_i), construct the sequence $(X_i)^* = (w_i X_i)$. Treat (X_i^*) as new sample, compute correspondingly:

- $\hat{\Sigma}^*$ and $\hat{P}_j^* = P_j(\hat{\Sigma}^*)$ bootstrapped versions
- $T_j^* = \frac{1}{2}\|\hat{P}_j^* - \hat{P}_j\|_2^2$ and $T_j = \|\hat{P}_j - P_j\|_2^2$

- A. Naumov, V. Spokoiny, and V. Ulyanov (2019). "Bootstrap confidence sets for spectral projectors of sample covariance". In: *Probab. Theory Related Fields* 174.3-4, pp. 1091–1132

Bootstrap approximations in high dimensions

- $\hat{\Sigma}^*$ and $\hat{P}_j^* = P_j(\hat{\Sigma}^*)$ bootstrapped versions
- $T_j^* = \frac{1}{2}\|\hat{P}_j^* - \hat{P}_j\|_2^2$ and $T_j = \|\hat{P}_j - P_j\|_2^2$

Theorem 10 (J. & W.)

Assume Setting 2 with $m = 8$.

(i) *If the eigenvalues decay polynomially with $\alpha \geq 2$, then w.h.p.*

$$\mathrm{U}\big(\mathcal{L}(T_j^*|X_1,\dots,X_n),\mathcal{L}(T_j)\big) \lesssim_{p,\alpha} n^{-1/8}(\log j)^{3/4} + \frac{j}{\sqrt{n}}(\log j \log n)^{3/2}$$

(ii) *If the eigenvalues decay exponentially, then w.h.p.*

$$\mathrm{U}\big(\mathcal{L}(T_j^*|X_1,\dots,X_n),\mathcal{L}(T_j)\big) \lesssim_{p,\alpha} n^{-1/8} + \Big(\frac{j^3 \log^3 n}{n}\Big)^{1/2}$$

Pervasive Factor models

- Recall $\mathcal{J} = \{1, \dots, J\}$, where we assume $J \geq 6$.
- Literature (approximate, pervasive) factor models: assumes the first J eigenvalues diverge at rate $\asymp d$ (with $d = \dim \mathcal{H}$), all remaining eigenvalues are bounded.
- We assume that there exist constants $0 < c \leq C < \infty$, such that

$$\lambda_1 \leq C\lambda_J, \qquad \lambda_J - \lambda_{J+1} \geq c\lambda_J, \qquad \frac{\mathrm{tr}_{\mathcal{J}^c}(\Sigma)}{\lambda_1} \leq C. \tag{14}$$

- Observe that this implies

$$\frac{\mathrm{tr}(\Sigma)}{\lambda_1} \asymp J,$$

which is the desired feature of pervasive factor models.

- J. Bai (2003). "Inferential theory for factor models of large dimensions". In: *Econometrica* 71.1, pp. 135–171
- James H. Stock and Mark W. Watson (2002). "Forecasting using principal components from a large number of predictors". In: *J. Amer. Statist. Assoc.* 97.460, pp. 1167–1179

Theorem 11 (J. & W.)

Assume Setting 2 with $m = 8$. Then for $j \in \mathcal{J}$

(i)

$$\mathrm{U}\Big(n\|\hat{P}_j - P_j\|_2^2, \|L_j Z\|_2^2\Big) \lesssim_p \left(\frac{J^6}{n}\right)^{1/2}\Big((\log n)^{3/2} + J^{5/2}\Big).$$

(ii) *W.h.p. for $j \in \mathcal{J}$*

$$\mathrm{U}\big(\mathcal{L}(T_j^*|X_1, \ldots, X_n), \mathcal{L}(T_j)\big) \lesssim_p \left(\frac{J^3}{n}\right)^{1/5} + \left(\frac{J^6 \log^3 n}{n}\right)^{1/2} + \left(\frac{Jp \log n}{n}\right)^{1/2} + n^{(6-q)/12}.$$

Take-home message

- While the size of $\|\hat{\Sigma} - \Sigma\|_\infty$ is closely linked to the effective rank, the behavior of $\hat{\lambda}_j, \hat{P}_j, \hat{u}_j$ is closely linked to the so-called relative ranks
- Can use this to obtain limit theorems, concentration inequalities for $\hat{\lambda}_j, \hat{P}_j, \hat{u}_j$.

Thank you for your attention!

Corollary 12

Suppose we are in the i.i.d. setting with $p \geq 16$*. If* (3) *holds, then*

$$\mathbb{E}\|\hat{P}_j - P_j\|_\infty^2 \leq \mathbb{E}\|\hat{P}_j - P_j\|_2^2 \leq Cj^2/n, \quad 1 \leq j \leq C\sqrt{n}(\log n)^{-5/2}.$$

- This result is (up to log terms) optimal in the case where $\lambda_j = Cj^{-\alpha-1}$, $\alpha > 0$ in a certain sense.

- This result is (up to log terms) optimal in the case where $\lambda_j = Cj^{-\alpha-1}$, $\alpha > 0$ in a certain sense.
- For such a polynomial decay, given that $\sup_{j\geq 1} \mathbb{E}|\eta_j|^{2p} \leq p! C^p$ for all $p \geq 1$, it has been shown that for any $j \geq 1$ (exists also information theoretic bound)

$$\mathbb{E}\|\hat{P}_j - P_j\|_\infty^2 \geq c(j^2/n) \wedge 1.$$

- This result is (up to log terms) optimal in the case where $\lambda_j = Cj^{-\alpha-1}$, $\alpha > 0$ in a certain sense.
- For such a polynomial decay, given that $\sup_{j\geq 1} \mathbb{E}|\eta_j|^{2p} \leq p!C^p$ for all $p \geq 1$, it has been shown that for any $j \geq 1$ (exists also information theoretic bound)

$$\mathbb{E}\|\hat{P}_j - P_j\|_\infty^2 \geq c(j^2/n) \wedge 1.$$

- We obtain the optimal bound for almost the whole range (up to the factor $(\log n)^{-5/2}$) where the trivial bound 2 does not apply. Moreover, only require mild conditions.

- This result is (up to log terms) optimal in the case where $\lambda_j = Cj^{-\alpha-1}$, $\alpha > 0$ in a certain sense.
- For such a polynomial decay, given that $\sup_{j\geq 1} \mathbb{E}|\eta_j|^{2p} \leq p! C^p$ for all $p \geq 1$, it has been shown that for any $j \geq 1$ (exists also information theoretic bound)

$$\mathbb{E}\|\hat{P}_j - P_j\|_\infty^2 \geq c(j^2/n) \wedge 1.$$

- We obtain the optimal bound for almost the whole range (up to the factor $(\log n)^{-5/2}$) where the trivial bound 2 does not apply. Moreover, only require mild conditions.
- Note: The stochastic behaviour of the scores $(\eta_j)_{j\geq 1}$ in terms of their dependence structure is irrelevant for the optimal algebraic structure conditions. In other words, this result **cannot** be improved assuming that $(\eta_j)_{j\geq 1}$ are independent.

- Another prominent example ist the spiked covariance model.

- Another prominent example ist the spiked covariance model.
- Signature feature: Several eigenvalues are larger than the remaining. Typically one is interested in recovering these leading eigenvalues and their associated eigenvectors (spiked part), since these explain most variation of the data.

- Another prominent example ist the spiked covariance model.
- Signature feature: Several eigenvalues are larger than the remaining. Typically one is interested in recovering these leading eigenvalues and their associated eigenvectors (spiked part), since these explain most variation of the data.
- One possible way to define the model: Assume $\mathcal{H} = \mathbb{R}^d$, let $f_1, \dots, f_d$ ($\|f_j\| = 1$) be orthogonal vectors and A be a covariance matrix such that

$$C_A^{-1} \leq \lambda_d(A) \leq \lambda_1(A) \leq C_A.$$

- Another prominent example ist the spiked covariance model.
- Signature feature: Several eigenvalues are larger than the remaining. Typically one is interested in recovering these leading eigenvalues and their associated eigenvectors (spiked part), since these explain most variation of the data.
- One possible way to define the model: Assume $\mathcal{H} = \mathbb{R}^d$, let $f_1, \ldots, f_d$ ($\|f_j\| = 1$) be orthogonal vectors and A be a covariance matrix such that

$$C_A^{-1} \leq \lambda_d(A) \leq \lambda_1(A) \leq C_A.$$

- For a sequence of weights $\omega_1, \ldots, \omega_d$, consider the spiked covariance model

$$\Sigma = F + A = \sum_{k=1}^{d} \omega_k^2 f_k f_k^\top + A,$$

where F denotes the 'spiked parts'.

- Generating probabilistic model: $F_1, \ldots, F_d$ is a martingale difference sequence with $\mathbb{E}F_k^2 = 1$ (factor loadings).

- Generating probabilistic model: $F_1, \ldots, F_d$ is a martingale difference sequence with $\mathbb{E}F_k^2 = 1$ (factor loadings).
- Similarly: $Y = (Y_1, \ldots, Y_d)^\top$ is a random vector where $Y_1, \ldots, Y_d$ form a martingale difference sequence.

- Generating probabilistic model: $F_1, \ldots, F_d$ is a martingale difference sequence with $\mathbb{E}F_k^2 = 1$ (factor loadings).
- Similarly: $Y = (Y_1, \ldots, Y_d)^\top$ is a random vector where $Y_1, \ldots, Y_d$ form a martingale difference sequence.
- F and Y are mutually uncorrelated, that is, all cross correlations are zero.

- Generating probabilistic model: $F_1, \ldots, F_d$ is a martingale difference sequence with $\mathbb{E}F_k^2 = 1$ (factor loadings).
- Similarly: $Y = (Y_1, \ldots, Y_d)^\top$ is a random vector where $Y_1, \ldots, Y_d$ form a martingale difference sequence.
- F and Y are mutually uncorrelated, that is, all cross correlations are zero.
- Idiosyncratic error ϵ and canonical factor model are defined as

$$X = \sum_{k=1}^{d} \omega_k F_k f_k + \epsilon, \quad \epsilon = A^{1/2} Y.$$

- Generating probabilistic model: $F_1, \ldots, F_d$ is a martingale difference sequence with $\mathbb{E}F_k^2 = 1$ (factor loadings).
- Similarly: $Y = (Y_1, \ldots, Y_d)^\top$ is a random vector where $Y_1, \ldots, Y_d$ form a martingale difference sequence.
- F and Y are mutually uncorrelated, that is, all cross correlations are zero.
- Idiosyncratic error ϵ and canonical factor model are defined as

$$X = \sum_{k=1}^{d} \omega_k F_k f_k + \epsilon, \quad \epsilon = A^{1/2} Y.$$

- Obviously, X has covariance matrix Σ.

- Apply results: Need to control moments of $\eta_j = \lambda_j^{-1/2} \langle X, u_j \rangle$.

- Apply results: Need to control moments of $\eta_j = \lambda_j^{-1/2}\langle X, u_j\rangle$.
- We have a connection between moments of η_j and F, Y via the following result.

- Apply results: Need to control moments of $\eta_j = \lambda_j^{-1/2}\langle X, u_j\rangle$.
- We have a connection between moments of η_j and F, Y via the following result.

Proposition 3.1

For $p \geq 2$, suppose that

$$\mathbb{E}|F_k|^p \leq C_F, \quad \mathbb{E}|Y_k|^p \leq C_Y$$

for all $k = 1, \ldots, d$. Then the conditions above imply $\mathbb{E}\eta_j = 0$ and

$$\max_{j\geq 1} \mathbb{E}|\eta_j|^p \leq C_\eta,$$

where C_η only depends on C_F, C_Y and C_A.

- Apply results: Need to control moments of $\eta_j = \lambda_j^{-1/2}\langle X, u_j \rangle$.
- We have a connection between moments of η_j and F, Y via the following result.

Proposition 3.1

For $p \geq 2$, suppose that

$$\mathbb{E}|F_k|^p \leq C_F, \quad \mathbb{E}|Y_k|^p \leq C_Y$$

for all $k = 1, \ldots, d$. Then the conditions above imply $\mathbb{E}\eta_j = 0$ and

$$\max_{j \geq 1} \mathbb{E}|\eta_j|^p \leq C_\eta,$$

where C_η only depends on C_F, C_Y and C_A.

Now get CLT and concentration inequalities.

- Consider the standard time series model for financial data

$$X_i = v_i \epsilon_i, \quad i \in \mathbb{N},$$

where $(\epsilon_i)_{i\in\mathbb{N}} \in \mathcal{H}$ are i.i.d. random variables, $(v_i^2)_{i\in\mathbb{N}} \in \mathbb{R}$ is a stationary, ergodic sequence that exhibits long memory and is independent of $(\epsilon_i)_{i\in\mathbb{N}}$.

- Consider the standard time series model for financial data

$$X_i = v_i \epsilon_i, \quad i \in \mathbb{N},$$

where $(\epsilon_i)_{i\in\mathbb{N}} \in \mathcal{H}$ are i.i.d. random variables, $(v_i^2)_{i\in\mathbb{N}} \in \mathbb{R}$ is a stationary, ergodic sequence that exhibits long memory and is independent of $(\epsilon_i)_{i\in\mathbb{N}}$.

- To be more precise, assume

$$b_n^{-1} \sum_{i=1}^{n} \left(v_i^2 - \mathbb{E}v_i^2\right) \xrightarrow{d} W_b,$$

where $b_n = n^b L(n)$ for $b \in (1/2, 1)$ and some slowly varying function $L(x)$. W_b is a nondegenerate random variable. Well-known example: W_b corresponds to a fractional Brownian motion, hence a normal distribution.

- What is a natural operator of interest here?

- What is a natural operator of interest here?
- Note: $(X_i)_{i\in\mathbb{N}}$ for a martingale sequence ($X_i = \epsilon_i v_i$).

- What is a natural operator of interest here?
- Note: $(X_i)_{i \in \mathbb{N}}$ for a martingale sequence ($X_i = \epsilon_i v_i$).
- Subject to some additional regularity assumptions, the Martingale-CLT implies

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \xrightarrow{d} \mathcal{N}(0, \Sigma), \; \Sigma = \mathbb{E} X_i \otimes X_i = \mathbb{E} v_0^2 \Sigma_\epsilon,$$

where Σ_ϵ is the covariance operator of ϵ.

- What is a natural operator of interest here?
- Note: $(X_i)_{i\in\mathbb{N}}$ for a martingale sequence $(X_i = \epsilon_i v_i)$.
- Subject to some additional regularity assumptions, the Martingale-CLT implies

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i \xrightarrow{d} \mathcal{N}\big(0,\Sigma\big),\ \Sigma = \mathbb{E}X_i \otimes X_i = \mathbb{E}v_0^2\Sigma_\epsilon,$$

where Σ_ϵ is the covariance operator of ϵ.

- So the 'standard' covariance operator Σ and the empirical counterpart $\hat{\Sigma}$ are still of high interest, where we recall

$$\hat{\Sigma} = \hat{\Sigma}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \otimes X_i.$$

Need some assumptions: Suppose that $\mathbb{E}\|\epsilon_i\|^2 < \infty$, let Σ_ϵ be the covariance operator of ϵ_i, and

$$\epsilon_i = \sum_{j \geq 1} \sqrt{\lambda_j^\epsilon} u_j \eta_{ij}^\epsilon, \quad i \in \mathbb{N}.$$

Need some assumptions: Suppose that $\mathbb{E}\|\epsilon_i\|^2 < \infty$, let Σ_ϵ be the covariance operator of ϵ_i, and

$$\epsilon_i = \sum_{j\geq 1} \sqrt{\lambda_j^\epsilon} u_j \eta_{ij}^\epsilon, \quad i \in \mathbb{N}.$$

Assumption 4.1

Suppose that for $p \geq 4$

$$\mathbb{E}\eta_{ij}^\epsilon = 0, \quad \mathbb{E}|\eta_{ij}^\epsilon|^p \leq C_\epsilon, \quad \mathbb{E}v_i^2 = 1, \quad \mathbb{E}|v_i|^p \leq C_v,$$

for all $i, j \geq 1$. Moreover, we assume that

$$\mathbb{E}\Big|\sum_{i=1}^n \left(v_i^2 - \mathbb{E}v_i^2\right)\Big|^2 \leq C_v b_n^2 \qquad b_n^{-1} \sum_{i=1}^n \left(v_i^2 - \mathbb{E}v_i^2\right) \xrightarrow{w} W_b.$$

- We now consider a triangular array of $X_1^{(n)}, \ldots, X_n^{(n)} \in \mathcal{H}^n$ with covariance operator $\Sigma^{(n)}$, $n = 1, 2, \ldots$, satisfying our previous long-memory setting.

- We now consider a triangular array of $X_1^{(n)}, \dots, X_n^{(n)} \in \mathcal{H}^n$ with covariance operator $\Sigma^{(n)}$, $n = 1, 2, \dots$, satisfying our previous long-memory setting.
- Notation: everything gets an $^{(n)}$, for instance $\lambda_j^{(n)}$, and so on.

Theorem 4.2

Fix $j_0 \geq 1$. Suppose that $\lambda_j^{(n)}$, $1 \leq j \leq j_0$ are simple for all $n \geq 1$ and Assumption 4.1 holds with C_ϵ, C_v independent of n. If

$$b_n n^{-1} \max_{1 \leq j \leq j_0} \sum_{k \neq j} \frac{\lambda_k^{(n)}}{|\lambda_j^{(n)} - \lambda_k^{(n)}|} \to 0 \quad \text{as} \quad n \to \infty,$$

and $\lambda_{j_0}^{(n)} \leq \lambda_{i_0}^{(n)}/2$ for some fixed $i_0 > j_0$, then

$$n b_n^{-1} \left(\frac{\hat{\lambda}_1^{(n)} - \lambda_1^{(n)}}{\lambda_1^{(n)}}, \ldots, \frac{\hat{\lambda}_{j_0}^{(n)} - \lambda_{j_0}^{(n)}}{\lambda_{j_0}^{(n)}} \right)^\top \xrightarrow{d} (W_b, \ldots, W_b)^\top.$$

- As one would expect, the long-memory behaviour transfers to the fluctuations of the empirical eigenvalues.

- As one would expect, the long-memory behaviour transfers to the fluctuations of the empirical eigenvalues.
- Somewhat surprising: all converge towards the identical limit. In particular, we have

$$nb_n^{-1} \max_{1 \le j,k \le j_0} \left| \frac{\hat{\lambda}_j^{(n)} - \lambda_j^{(n)}}{\lambda_j^{(n)}} - \frac{\hat{\lambda}_k^{(n)} - \lambda_k^{(n)}}{\lambda_k^{(n)}} \right| \xrightarrow{\mathbb{P}} 0,$$

a result of the dominating nature of the long-memory component.

- As one would expect, the long-memory behaviour transfers to the fluctuations of the empirical eigenvalues.
- Somewhat surprising: all converge towards the identical limit. In particular, we have

$$nb_n^{-1} \max_{1\le j,k\le j_0} \left| \frac{\hat{\lambda}_j^{(n)} - \lambda_j^{(n)}}{\lambda_j^{(n)}} - \frac{\hat{\lambda}_k^{(n)} - \lambda_k^{(n)}}{\lambda_k^{(n)}} \right| \xrightarrow{\mathbb{P}} 0,$$

  a result of the dominating nature of the long-memory component.
- Note the different normalization

$$b_n n^{-1} \max_{1\le j\le j_0} \sum_{k\ne j} \frac{\lambda_k^{(n)}}{|\lambda_j^{(n)} - \lambda_k^{(n)}|} \to 0 \quad \text{as} \quad n \to \infty.$$

- As one would expect, the long-memory behaviour transfers to the fluctuations of the empirical eigenvalues.
- Somewhat surprising: all converge towards the identical limit. In particular, we have

$$nb_n^{-1} \max_{1\leq j,k\leq j_0} \left| \frac{\hat{\lambda}_j^{(n)} - \lambda_j^{(n)}}{\lambda_j^{(n)}} - \frac{\hat{\lambda}_k^{(n)} - \lambda_k^{(n)}}{\lambda_k^{(n)}} \right| \xrightarrow{\mathbb{P}} 0,$$

  a result of the dominating nature of the long-memory component.
- Note the different normalization

$$b_n n^{-1} \max_{1\leq j\leq j_0} \sum_{k\neq j} \frac{\lambda_k^{(n)}}{|\lambda_j^{(n)} - \lambda_k^{(n)}|} \to 0 \quad \text{as} \quad n \to \infty.$$

- Most likely an artifact in this case.

Consider the case where $d = \dim \mathcal{H} < \infty$.

Theorem 4.3

In the previous setting, suppose that $\lambda_1^{(n)}$ is simple for all $n \geq 1$. If $d = d_n$ and the sequence $d_n \big(b_n n^{-1/2}\big)^{-p/2}$ is bounded,

$$\frac{1}{\sqrt{n}} \sum_{k>1} \frac{\lambda_k^{(n)}}{\lambda_1^{(n)} - \lambda_k^{(n)}} \to 0 \quad \text{as} \quad n \to \infty,$$

then the sequence $n b_n^{-1} (\hat{\lambda}_1^{(n)} - \lambda_1^{(n)}) / \lambda_1^{(n)}$ is tight.

Consider the case where $d = \dim \mathcal{H} < \infty$.

Theorem 4.3

In the previous setting, suppose that $\lambda_1^{(n)}$ is simple for all $n \geq 1$. If $d = d_n$ and the sequence $d_n \left(b_n n^{-1/2}\right)^{-p/2}$ is bounded,

$$\frac{1}{\sqrt{n}} \sum_{k>1} \frac{\lambda_k^{(n)}}{\lambda_1^{(n)} - \lambda_k^{(n)}} \to 0 \quad \text{as} \quad n \to \infty,$$

then the sequence $n b_n^{-1} (\hat{\lambda}_1^{(n)} - \lambda_1^{(n)}) / \lambda_1^{(n)}$ is tight.

- Convinced we can actually now prove a limit theorem.