Week 10

PCA under the Spiked covariance model

Refs: • Koch (2014). "Analysis of Multivariate and High-dim data."
• Paul & Aue (2014). "RMT in Statistics". Sect 4.1.2.
• Johnstone (2001). "On the distribution of the largest eigenvalue in principal component analysis."
• Anderson (2003) "Intro to multivariate stat. analysis".

Principal component analysis (PCA) is a fundamental tool in multivariate analysis.

Consider p-dimensional population $\mathbf{x}$ with covar $\Sigma = \mathrm{cov}(\mathbf{x})$.
Sample of size n $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$.
$\Rightarrow$ Sample covariance $\mathbf{S}_n$

PCA: Find orthogonal directions (successive) that maximally explain the variation in the data.

$$\lambda_j = \max\left\{ \frac{u'\mathbf{S}u}{u'u} : u \perp u_1, \dots, u_{j-1},\ j=1,2,\dots,\min(n,p) \right\}$$

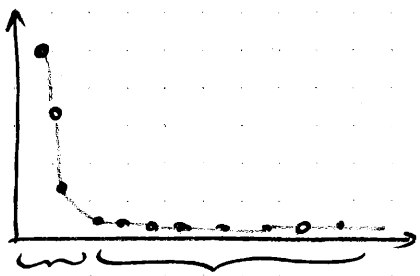
$$= \max\left\{ u'\mathbf{S}u : \|u\|=1,\ u\perp u_1, \dots, u_{j-1},\ j=1,\dots,\min(n,p) \right\}$$

Q: How many principal components should be retained?

- Screeplot
- Wachter plot

See Johnstone (2001); Fig 1 & 2.

Using screeplot, look for "elbow" or other break.



Eigenvalues seperate into two classes:

- Bulk
- Extreme.

We have seen that the bulk spectrum (eigenvalues) are well described by the Marchenko-Pastur distribution.

The extremes describe a "signal subspace" of higher variance from many noisy variables.

The null case ($\Sigma = I_p$) plays a fundamental role.

Geman (1980) showed (Assume $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$)

$$\lambda_1 \longrightarrow (1+\sqrt{y})^2 \qquad y = p/n.$$

This was later refined (Bai, Krishnaiah, Silverstein, Yin).

Theorem: Let $\mathbb{X} = (x_{ij})$ be a matrix with iid complex-valued entries $\mathbb{E}[x_{ij}] = 0$ $Var(x_{ij}) = 1$. $\forall_{ij}$. and $\mathbb{E}[x_{ij}^4] < \infty$. Set $\mathbb{x}_k = (x_{1k}, x_{2k}, \dots, x_{pk})'$ (k'th column) and sample covariance $S_n = \frac{1}{n} \sum_{k=1}^{n} \mathbb{x}_k \mathbb{x}_k^*$. Then if the eigenvalues of S_n are $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_p$ and $p/n \to y > 0$, we have

$$\lambda_1 \xrightarrow{as} b_y = (1 + \sqrt{y})^2$$

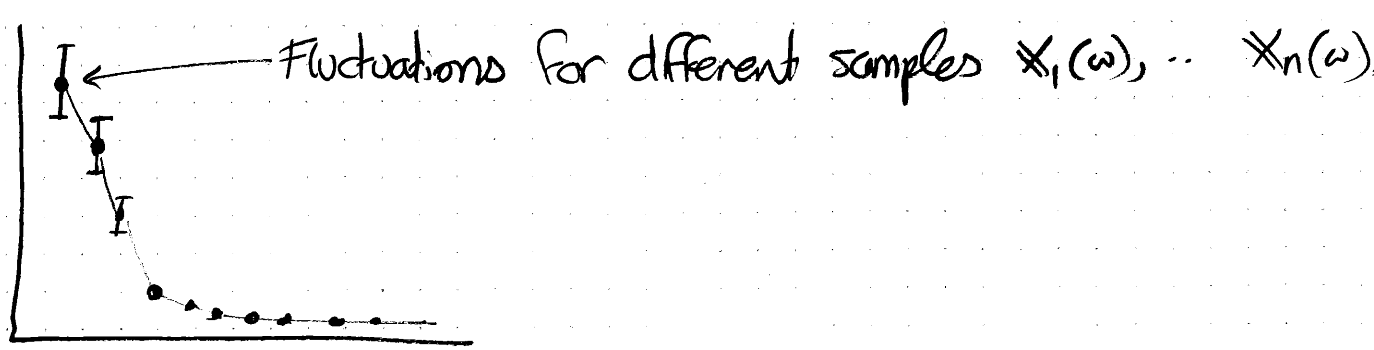
$$\lambda_{min} \xrightarrow{as} a_y = (1 - \sqrt{y})^2$$

where $\lambda_{min} = \lambda_p$ if $p \leq n$ and $\lambda_{min} = \lambda_{p-n+1}$ otherwise.

#

In other words, in the null case $(\Sigma = I_p)$ the smallest and largest eigenvalues of S_n are located near the right edge b_y and left edge a_y of the Marchenko-Pastur distribution.

Ian Johnstone (2001) further characterised the fluctuations of the largest eigenvalue λ_1.



Let $\mu_{np} = \frac{1}{n}\left[(n-1)^{1/2} + p^{1/2}\right]^2$

$\sigma_{np} = \left[(n-1)^{1/2} + p^{1/2}\right] \times \left[(n-1)^{-1/2} + p^{-1/2}\right]^{1/3}$

Notice for large p and n, $\mu_{np} \approx (1+\sqrt{y})^2$.

$\uparrow$ right edge MP dist.

<u>Theorem</u> (Johnstone):

$$\frac{\lambda_1 - \mu_{np}}{\sigma_{np}} \xrightarrow{d} F_1.$$

Here F_1 is a Tracy-Widom distribution of order 1.

The CDF is given by

$$F_1(s) = \exp\left(\frac{1}{2}\int_s^\infty [q(x) + (x-s)^2 q^2(x)]\,dx\right) \quad s \in \mathbb{R}.$$

and s solves the (Painlevé II) differential equation

$$q''(x) = x\,q(x) + 2q^3(x).$$

The CDF has no closed-form formula and can only be found numerically.

F_1 has mean 1.21 and sd=1.27.

When we look at a real-world data set we do not see the extreme eigenvalues clustering near the edges of the MP distribution.
This suggests that our null assumption ($\Sigma = I_p$) is incorrect and we should explore the <u>non-null case</u>.

Mathematically the non-null case can be described within the "spiked population model" as suggested by Johnstone (2001).

In its simplest form, we assume that the population covariance matrix Σ in the <u>spiked population model</u> has only m non-unit eigenvalues

$$\mathrm{spec}(\Sigma) = \{\underbrace{\alpha_1, \alpha_2, \cdots, \alpha_m}_{m \text{ spike eigenvalues}}, 1, 1 \cdots, 1\}.$$

Assume that $n \to \infty$, $p/n \to y > 0$.

As m is fixed, as $n, p \to \infty$, the ESD still converges to the MP distribution as the number of "non-spike" eigenvalues is overpowered by the $\underbrace{p-m}_{\to\infty}$ non-fixed eigenvalues.

However, the distribution of the extreme eigenvalues of S_n and the other $(m-2)$ are modified.
We will now look at this behaviour.

Limits of Spiked Sample eigenvalues.

We assume our observations $\mathbb{X}_i = \Sigma^{\frac{1}{2}} \mathbb{Y}_i \quad i = 1, 2, \cdots, n$
Where $\mathbb{Y}_i$ are i.i.d. p-dimensional vectors with mean zero and unit variance. and i.i.d. components.

i.e. $\quad \mathbb{Y}_i \sim N_p(0, I_p)$.

$$\Longrightarrow \quad \mathbb{X}_i \sim N_p(0, \Sigma) \quad \text{as} \quad \mathbb{X}_i = \sqrt{\Sigma}\, \mathbb{Y}_i$$

and Σ has structure

$$\Sigma = \begin{pmatrix} \Lambda & 0 \\ 0 & \mathbb{V}_p \end{pmatrix}$$

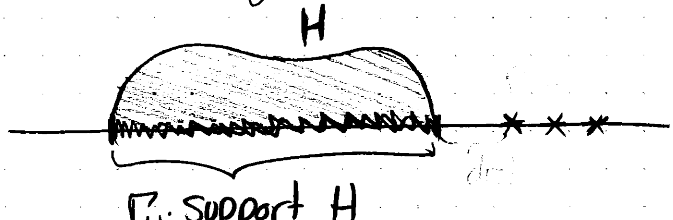
Assumptions

- Λ $m \times m$ matrix. Eigenvalues of Λ $\alpha_1 > \alpha_2 > \cdots > \alpha_K$ with multiplicity $m_1, m_2, \cdots m_K$ $(m = m_1 + \cdots + m_K)$
  J_j = set of m_j indexes of α_j in matrix Σ.
- ESD H_p of $\mathbb{V}_p$ converges to a nonrandom limiting distribution H.

- The sequence of the largest eigenvalue of Σ is bounded.
- The eigenvalues $\{\beta_{pj}\}$ of V_p are such that

$$\sup_j d(\beta_{pj}, \Gamma_H) = \varepsilon_p \to 0. \quad \boxed{\{\beta_{pj}\} \in \Gamma_H \text{ as } n \to \infty}$$

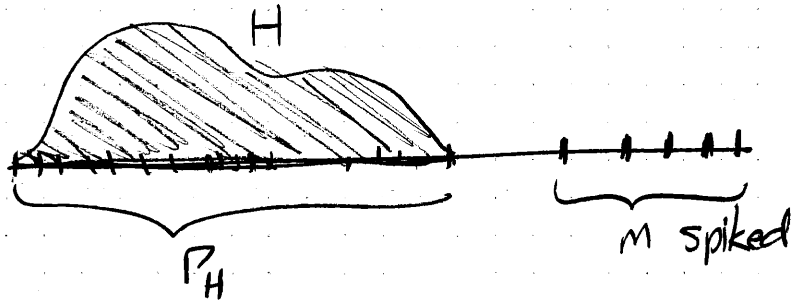
$d(x, A)$: distance of x to set A $\quad$ Γ_H: support of H.



<u>Def:</u> An eigenvalue α of Δ is called <u>generalised spike</u> or, <u>spike</u>, if $\alpha \notin \Gamma_H$.

We call this model the <u>generalised spike model</u>:

$\Rightarrow$ Eigenvalues of Σ are composed of a <u>main spectrum</u> made with the $\{\beta_{pj}\}$s and a <u>finite spectrum of m spike eigenvalues</u>.



We also need the technical conditions:

- $E[\mathbb{Y}_{ij}] = 0$, $E[|Y_{ij}|^2] = 1$, $E[|Y_{ij}|^4] < \infty$
- $p/n \to y > 0$, $p \wedge n \to \infty$.

We decompose our observations into blocks of size m and $p-m$:

$$\mathbb{X}_i = \begin{pmatrix} \mathbb{X}_{1i} \\ \mathbb{X}_{2i} \end{pmatrix} \qquad \mathbb{Y}_i = \begin{pmatrix} \mathbb{Y}_{1i} \\ \mathbb{Y}_{2i} \end{pmatrix}$$

Define the sample covariance matrix

$$S_n = \frac{1}{n}\sum_{k=1}^{n} \mathbb{X}_k \mathbb{X}_k^* = \frac{1}{n}\begin{pmatrix} \mathbb{X}_1\mathbb{X}_1^* & \mathbb{X}_1\mathbb{X}_2^* \\ \mathbb{X}_2\mathbb{X}_1^* & \mathbb{X}_2\mathbb{X}_2^* \end{pmatrix}$$

$$= \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

where $\mathbb{X}_1 = (\mathbb{X}_{11}, \dots, \mathbb{X}_{1n})$ $\qquad \mathbb{X}_2 = (\mathbb{X}_{21}, \dots, \mathbb{X}_{2n})$

We also define the same block decomposition for the $\mathbb{Y}_i$ vectors and data matrices $\mathbb{Y}_1$ and $\mathbb{Y}_2$, so that

$$\mathbb{X}_1 = \Lambda^{1/2}\mathbb{Y}_1 \qquad \mathbb{X}_2 = \Lambda^{1/2}\mathbb{Y}_2$$

An eigenvalue of $\mathbb{S}_n$ that is <u>not</u> an eigenvalue of $\mathbb{S}_{22}$ satisfies

λ_i s.t. $0 = |\lambda_i I_p - \mathbb{S}_n| = |\lambda_i I_{p-m} - \mathbb{S}_{22}| \cdot |\lambda_i I_m - \mathbb{K}_n(\lambda_i)|$

where $\mathbb{K}_n(r) = \mathbb{S}_{11} + \mathbb{S}_{12}(r I_{p-m} - \mathbb{S}_{22})^{-1}\mathbb{S}_{21}$

For large n it will eventually hold that

$$|\lambda_i I_{p-m} - \mathbb{S}_{22}| \neq 0.$$

and

$$|\lambda_i I_m - \mathbb{K}_n(\lambda_i)| = 0.$$

We now want to consider the limit of the random matrix $\mathbb{K}_n(r)$ with fixed dimension m. It holds that

$$\begin{aligned} \mathbb{K}_n(r) &= \mathbb{S}_{11} + \mathbb{S}_{12}(r I_{p-m} - \mathbb{S}_{22})^{-1}\mathbb{S}_{21} \\ &= \frac{1}{n}\mathbb{X}_1\mathbb{X}_1^* + \frac{1}{n}\mathbb{X}_1\mathbb{X}_2^*(r I_{p-m} - \mathbb{S}_{22})^{-1}\frac{1}{n}\mathbb{X}_2\mathbb{X}_1^* \\ &= \frac{1}{n}\mathbb{X}_1\left[I_n + \frac{1}{n}\mathbb{X}_2(r I_{p-m} - \mathbb{S}_{22})^{-1}\mathbb{X}_2\right]\mathbb{X}_1^* \end{aligned}$$

(Using identity: $r \neq 0$ not eigenvalue of A^*A

$$I_n + A(r I_{p-m} - A^*A)^{-1}A^* = r(r I_n - AA^*)^{-1} \quad)$$

$$\begin{aligned} &= \frac{r}{n}\mathbb{X}_1\left(r I_n - \frac{1}{n}\mathbb{X}_2^*\mathbb{X}_2\right)^{-1}\mathbb{X}_1^* \\ &= \frac{r}{n}\Lambda^{1/2}\mathbb{Y}_1\left(r I_n - \frac{1}{n}\mathbb{X}_2^*\mathbb{X}_2\right)^{-1}\mathbb{Y}_1^*\Lambda^{1/2} \end{aligned}$$

Since r is outside the support of LSD $F_{y,H}$ of $\mathbb{S}_{22}$ for large enough n, the (operator) norm of

$$\left(rI_n - \frac{1}{n}\mathbb{X}_2^*\mathbb{X}_2\right)^{-1}$$

is bounded.

By LLN, as $n \to \infty$,

$$\mathbb{K}_n(r) = \Lambda\left[r\, tr\left(rI_n - \frac{1}{n}\mathbb{X}_2^*\mathbb{X}_2\right)^{-1}\right] + o(1)$$

$$= -\Lambda \cdot r \cdot \underline{s}(r) + o(1).$$

$\underline{s}$ is Stieltjes transform of the LSD of $\frac{1}{n}\mathbb{X}_2^*\mathbb{X}_2$.

If for some subsequence $\{i\}$ of $\{1, 2, \dots, n\}$,

$$\lambda_i \to r$$

then $\mathbb{K}_n(\lambda_i) \to -\Lambda r \underline{s}(r)$

Therefore if r is an eigenvalue of $-\Lambda r \underline{s}(r)$, ie.

$$r = -\alpha_j r \underline{s}(r)$$

$$\iff \quad \underline{s}(r) = -1/\alpha_j$$

The function $\psi(\alpha) = \psi_{y,H}(\alpha) = \alpha + y\int \frac{t\alpha}{\alpha - t}\, dH(t)$

is the inverse of the function $x \mapsto -1/\underline{s}(x)$.

$\psi(\alpha)$ is well-defined for all $\alpha \notin \Gamma_H$

We have proved that if such a limit r exists then r is necessarily satisfying the equation

$$r = \psi(\alpha_j)$$

for some α_j.

Further it can be shown that $r = \psi(\alpha_j)$ is outside the support of the LSD $F_{y,H}$ if and only if $\psi'(\alpha_j) > 0$.

We have shown that if α_j is a spike eigenvalue such that $r = \psi(\alpha_j)$ is the limit for some subsequence of sample eigenvalues $\{\lambda_i\}$ then $\psi'(\alpha_j) > 0$.

This condition is also a sufficient condition.

<u>Theorem:</u> (1) For spiked eigenvalue α_j satisfying

$$\psi'(\alpha_j) > 0$$

there are m_j sample eigenvalues λ_i of $\mathbf{S}$ with $i \in J_j$ such that

$$\lambda_i \xrightarrow{a.s.} \psi_j = \psi(\alpha_j)$$

(2) For a spike eigenvalue α_j satisfying

$$\varphi'(\alpha_j) \leq 0.$$

there are m_j sample eigenvalues λ_i of S with $i \in J_j$ such that

$$\lambda_i \xrightarrow{a.s.} \gamma_j$$

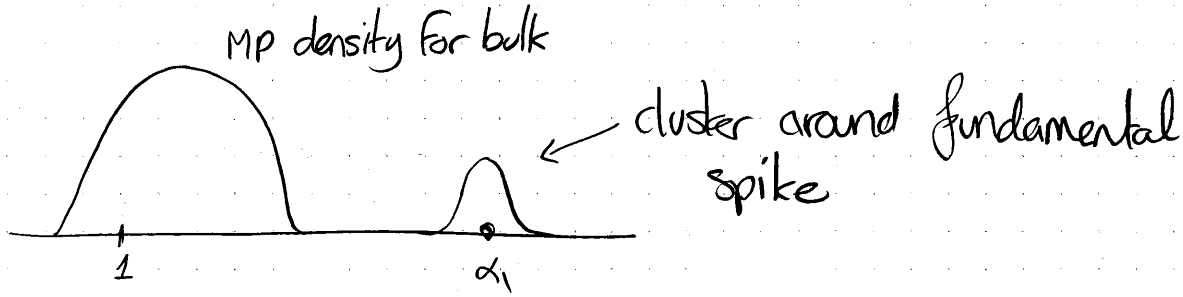
where γ_j is the γ'th quantile of $F_{y,H}$ with $\gamma = H(-\infty, \alpha_j]$ and H the LSD of V_p. #

Proof: See Johnstone (2001).

The point of the theorem is that the eigenvalues are separated into two groups:

- the eigenvalues with positive φ' can be called the <u>fundamental spikes</u>
- the eigenvalues with non-positive φ' can be called the <u>non-fundamental spikes</u>

A fundamental spike α_j is that for large enough n, exactly m_j sample eigenvalues will cluster in a neighbourhood of $\varphi_{y,H}(\alpha_j)$ which is outside the support of the LSD $F_{y,H}$.



These limits are considered as outliers compared to the bulk spectrum. We call them <u>spiked sample eigenvalues</u>.

The separation between the fundamental and non-fundamental spike eigenvalues depend not only on the base population spectral distribution H but also on the limiting ratio y.

Eg. Notice that when $y \to 0$,

$$\varphi_{y,H}(\alpha) = \alpha + y\int \frac{t\alpha}{\alpha - t}\, dH(t) \xrightarrow{y\to 0} \alpha.$$

so that $\varphi' \to 1$.

This means that for y small we have that any spike eigenvalue α_j is a fundamental spike and there will be m_j spike sample eigenvalues converging to $\varphi_{y,H}(\alpha_j)$.

When $p \ll n$, $\varphi_{y,H}(\alpha_j) \approx \alpha_j$

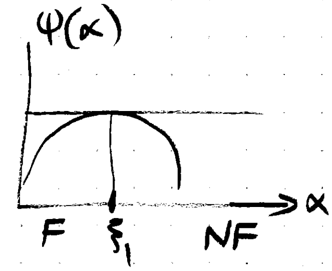
$\implies$ sample eigenvalues $\longrightarrow$ population eigenvalues

We have

$$\varphi'(\alpha) = 1 - y\int \frac{t^2}{(\alpha-t)^2}\, dH(t)$$

$$\varphi''(\alpha) = 2y\int \frac{t^2}{(\alpha-t)^3}\, dH(t)$$

If H has compact support $\Gamma_H = [\theta, \omega]$ then from the derivatives φ', φ'' we have:



- for $\alpha < \theta$, φ is concave and varies from $-\infty$ to $-\infty$ where $\varphi' = 0$ at a unique point denoted ξ_1

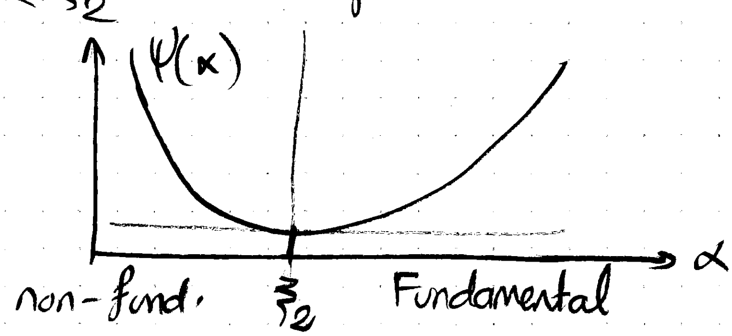
$\Rightarrow$ any spike $\alpha < \xi_1$ is fundamental

—— $\xi_1 \leq \alpha < \theta$ is non-fundamental.

- for $\alpha > \omega$, φ convex and varies from ∞ to ∞. where $\varphi' = 0$ at a unique point denoted ξ_2

$\Rightarrow$ any spike $\alpha > \xi_2$ fundamental.

—— $\omega < \alpha \leq \xi_2$ non-fundamental



Johnstone's spiked population model

$$\mathrm{spec}(\Sigma) = \{\alpha_1, \alpha_2, \cdots, \alpha_m, 1, \dots, 1\}.$$

$\mathbb{V}_p = I_{p-m}$ and PSD $H = \delta_1$. This gives

$$\psi(\alpha) = \alpha + \frac{y\alpha}{\alpha - 1}$$

$$\psi'(\alpha) = 1 - \frac{y}{(\alpha-1)^2}$$

In this case we see that ψ has:

- range $(-\infty, a_y] \cup [b_y, \infty)$
- $\psi(1-\sqrt{y}) = a_y \quad \psi(1+\sqrt{y}) = b_y$.
- $\psi'(\alpha) > 0 \iff |\alpha - 1| > \sqrt{y}$

  which means $\xi_1 = 1 + \sqrt{y} \quad \xi_2 = 1 - \sqrt{y}$.

The behaviour is given in the following corollary.

<u>Corollary:</u> When $\mathbb{V}_p = I_{p-m}$ we have:

(1) <u>large fundamental spikes:</u> for $\alpha_j > 1+\sqrt{y}$

$$\lambda_i \xrightarrow{a.s.} \alpha_j + \frac{y\alpha_j}{\alpha_j - 1} \qquad i \in J_j$$

(2) <u>large non-fundamental spikes:</u> for $1 < \alpha_j \leq 1+\sqrt{y}$

$$\lambda_i \xrightarrow{a.s.} (1+\sqrt{y})^2 \qquad i \in J_j$$

(3) <u>Small non-fundamental spikes:</u> for $1-\sqrt{y} \leq \alpha_j$ with $y<1$, or $\alpha_j < 1$ with $y \geq 1$.

$$\lambda_i \xrightarrow{a.s.} (1-\sqrt{y})^2 \qquad i \in J_j$$

(4) <u>small fundamental spikes:</u> for $\alpha_j < 1-\sqrt{y}$ with $y<1$,

$$\lambda_j \xrightarrow{a.s.} \alpha_j + \frac{y\alpha_j}{\alpha_j - 1} \qquad i \in J_j.$$