Non white sample covariance matrices.

S. Péché ,
Université Grenoble 1,
joint work with D. Féral, Uni. Bordeaux, O. Ledoit, Uni. Zurich

03/12/2009, Institut Henri Poincaré

Plan

- I. Eigenvalues/eigenvectors of sample covariance matrices: problem and motivations.
- II. Eigenvalues: global behavior, extreme eigenvalues.
- III. Eigenvectors: the white case and the non white case.
- IV. Conclusion.

Model

We consider sample covariance matrices:

$$M_N(\Sigma) = \frac{1}{p}\Sigma^{1/2} X X^* \Sigma^{1/2}$$

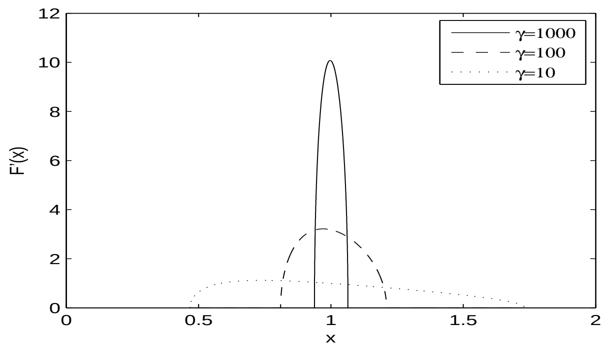
where

- X is a $N \times p$ random matrix s.t. the entries X_{ij} are i.i.d. complex (or real) random variables with distribution μ, $\int x d\mu(x) = 0, \int |x|^2 d\mu(x) = 1$.
- $p = p(N)$ with $p/N \to \gamma \in (0, \infty)$ as $N \to \infty$;
- Σ is a $N \times N$ Hermitian deterministic (or random) matrix, $\Sigma > 0$ with bounded spectral radius. Σ is independent of X.

What can be said about the spectrum (eigenvalues and eigenvectors) as $N \to \infty$?

Motivations I.

Statistics Knowing $M_N(\Sigma)$ what can be said about Σ?
-if N is fixed and $p \to \infty$: $M_N(\Sigma)$ good estimator of Σ;
-in high dimension (genetics, finance, ...)?
Understand e.g. the behavior of PCA in such a setting.



Density of the eigenvalues of $M_N(\Sigma)$ when $\Sigma = Id$.
Dispersion of the eigenvalues: $M_N(\Sigma)$ is NOT a good estimator of Σ (smallest and largest eigenvalues e.g.)

Motivations II.

Communication theory "CDMA" : received signal $r = \sum_{k=1}^{K} b_k s_k + w$,
with K number of users, $s_k \in \mathbb{C}^N$ the signature
$b_k \in \mathbb{C}$, $\mathbb{E} b_k = 0$, $\mathbb{E}|b_k|^2 = p_k$ transmitted signal,
and $w \in \mathbb{C}^N$ a Gaussian white noise with i.i.d. $\mathcal{N}(0, \sigma^2)$ components.

One has to decode/estimate the signal b_k. A measure of the performance of the communication channel is the so-called "SIR" (Signal to Interference Ratio): linear receiver $\hat{x}_1 = c_1^* r$

$$SIR = \frac{|C_1^* s_1|^2 p_1}{|c_1|^2 \sigma^2 + \sum_{i \geq 2} |c_1^* s_i|^2 p_i}.$$

$\Longrightarrow$ as $N, K \to \infty$, $K/N \to \gamma$, the SIR depends on the eigenvalues AND the eigenvectors of SDS^* where $S = [s_2, \ldots, s_K]$ is the signature matrix (random) and $D = \mathrm{diag}(p_2, \ldots, p_N)$.

Eigenvalues.

The eigenvalues I

We denote by $\pi_1 \geq \pi_2 \geq \cdots \geq \pi_N$ the eigenvalues of Σ and suppose that

$$\rho_N(\Sigma) := \frac{1}{N} \sum_{i=1}^{N} \delta_{\pi_i} \overset{a.s.}{\rightarrow} H,$$

where H is a probability measure.

Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ be the eigenvalues of $M_N(\Sigma)$; $\mu_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i}$.

Theorem Marchenko-Pastur (67)
A.s. $\lim_{N\to\infty} \mu_N = \rho_{MP}$, where the Stieltjes transform of ρ_{MP} given by

$$\forall\, z \in \mathbb{C}, \Im(z) > 0, \quad m_\rho(z) := \int \frac{1}{\lambda - z} d\rho_{MP}(\lambda),$$

satisfies $m_\rho(z) = \int_{-\infty}^{+\infty} \left\{\tau\left[1 - \gamma^{-1} - \gamma^{-1} z\, m_\rho(z)\right] - z\right\}^{-1} dH(\tau).$

The eigenvalues II

If $\Sigma = Id$, one knows explicitly the density of the Marchenko-Pastur distribution

$$\gamma \geq 1, \quad \frac{d\rho_{MP}}{du} = \frac{\gamma}{2\pi u}\sqrt{(u_+ - u)(u - u_-)}1_{[u_-,u_+](u)},$$

$$\text{with } u_\pm = (1 \pm \frac{1}{\sqrt{\gamma}})^2.$$

Valid for both complex and real random matrices.

For general H, the relationship between ρ_{MP} and H is not "simple", determining H from ρ_{MP} is not easy. El Karoui (2008) gives a consistent estimator (using convex approximation).

Assume that H has been estimated, can we improve our knowledge of Σ?

The "usual" behavior of largest eigenvalues

Assume that $\Sigma = Id$.

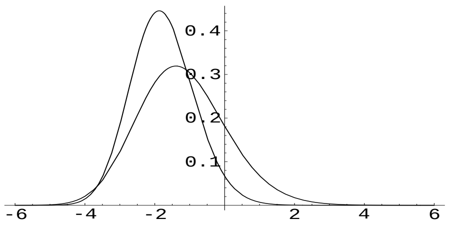
Theorem Johnstone (2001) Johansson (2000), El Karoui (2005): $\mu = \mathcal{N}(0,1)$

Soshnikov (2001) Péché (2007) : μ non Gaussian symmetric distribution with sub-Gaussian tails

$$\exists C > 0,\ \forall k > 0,\ \int |x|^{2k} d\mu(x) \leq (Ck)^k.$$

λ_1 largest value of $M_N(Id) = M_N$, $u_+^N = (1 + \frac{1}{\sqrt{\gamma_N}})^2$.

$$\lim_{N\to\infty} \mathbb{P}\Big(\frac{N^{2/3}}{\gamma_N^{-1/2}\,(u_+^N)^{2/3}}\,(\lambda_1(M_N(\Sigma)) - u_+^N) \leq x\Big) = F_{2(1)}^{TW}(x),$$ Tracy Widom distribution.



A “slight” perturbation of the true covariance

Let $\Sigma = \text{diag}(\pi_1, \pi_2, \ldots, \pi_r, 1, \ldots, 1)$, $\pi_i \geq \pi_{i+1} > 1$, $i \leq r-1$, r independent of N.
Σ is a finite rank perturbation of the identity matrix: $H = \delta_1$.

What is the impact of the π_i's on the spectrum?
The global behavior of the spectrum is unchanged but the largest eigenvalues are impacted.

Studied by : Baik-Ben Arous-Péché (2005) μ complex $\mathcal{N}(0,1)$; Bai-Yao (2008) and Féral-Péché (2008) for more general ensembles.
Baik-Silverstein (2006): a.s. limit of the largest and smallest eigenvalues for very general ensembles.
El Karoui (2007): X Gaussian Σ finite rank perturbation of a deterministic $\Sigma_0 \neq I_N$.

Phase Transition

We set

$$w_c := 1 + \frac{1}{\sqrt{\gamma}}, \quad \tau(\pi_1) = \pi_1 \left(1 + \frac{\gamma^{-1}}{\pi_1 - 1}\right), \quad \sigma(\pi_1) = \pi_1 \sqrt{1 - \gamma^{-1}/(\pi_1 - 1)^2}.$$

- (F.-P.) If $\pi_1 < w_c$, and μ is symmetric with sub-gaussian tails then $\lim_{N,p\to\infty} \mathbb{P}\left(\frac{N^{2/3}}{\gamma_N^{-1/2} \left(u_+^N\right)^{2/3}} (\lambda_1(M_N(\Sigma)) - u_+^N) \leq x\right) = F_{2(1)}^{\mathrm{TW}}(x)$ Tracy-Widom distribution (complex or real). Same as if $\Sigma = I_N$.

- (Bai-Yao) If $\pi_1 = \ldots = \pi_k > w_c$ and $\pi_{k+1} < \pi_1$ and $\mathbb{E}|X_{ij}|^4 < \infty$ then

$$\lim_{N,p\to\infty} \mathbb{P}\left(\frac{\sqrt{N}}{\sigma(\pi_1)} (\lambda_1(M_N(\Sigma)) - \tau(\pi_1)) \leq x\right) = G_k(x),$$

  where G_k is the distribution of the largest eigenvalue of the GUE $H = (H_{ij})_{i,j=1}^k$ with i.i.d. complex $\mathcal{N}(0, \sigma(\mu))$ entries.

Remarks

- "Spikes" in the true covariance can be detected if they are large enough.
- Actually the "true" conjecture assumes that first point should hold true provided that $\mathbb{E}|X_{ij}|^4 < \infty$ only.
- If $\pi_1 = \ldots = \pi_k = w_c$ and $\pi_{k+1} < w_c$ the limiting distribution of $\lambda_1(M_N(\Sigma))$ is also determined (in particular $\lambda_1(M_N(\Sigma)) \to u_+^N$).
- The asymptotic fluctuations of the smallest eigenvalues is expected to exhibit the same behavior (Baik-Silverstein (2006)).
- The proof of these results relies on the explicit computation of the distribution of the largest eigenvalues (Gaussian case). The extension to other ensembles is based on the moment approach due to Soshnikov (Féral-Péché) and via the resolvent and Central Limit Theorem (Bai-Yao, Baik-Silverstein).
- No result for non Gaussian μ if $H \neq \delta_1$.

Eigenvectors: the white case.

Gaussian sample

Suppose that $\Sigma = Id$ and X_{ij} i.i.d. $\mathcal{N}(0,1)$ complex or real.
$M_N = M_N(Id)$ is a so-called "white Wishart matrix".
Let (U, D) be a diagonalization of M_N: $M_N = UDU^*$ with $U \in \mathbb{U}(N)$ and D a real diagonal matrix.

U is Haar distributed.

Proof: Gram-Schmidt+ rotationnal invariance of the Gaussian distribution.

Conjecture: if $\Sigma = Id$ and if X has non-Gaussian entries with $\mathbb{E}|X_{ij}|^4 < \infty$, the matrix of eigenvectors of M_N shall "asymptotically be Haar distributed".
Idea: neither direction is preferred.

Question: how to define "asymptotically Haar distributed"?

Non Gaussian matrices I.

Silverstein's idea ('95): U is asymptotically Haar distributed if, given an arbitrary vector $x \in \mathbb{S}^{N-1} = \{x \in \mathbb{R}^N, |x| = 1\}$, $y = Ux$ is asymptotically uniformly distributed on the unit sphere. Or setting

$$Y_N(t) := \sqrt{\frac{N}{2}} \sum_{i=1}^{[Nt]} (|y_i|^2 - 1/N),$$

$Y_N(t)$ shall converge in distribution to a Brownian bridge if y is uniformly distributed ($y = Z/|Z|^2$ with Z Gaussian).

Consider instead $X_N(t) = Y_N(F^N(t)) = \sqrt{\frac{N}{2}} \left(F_1^N(t) - F_N(t)\right)$ with $F^N(t) = \dfrac{1}{N} \displaystyle\sum_{i=1}^{N} 1_{\lambda_i \leq t}$ cumulative distribution function (c.d.f.) of the spectral measure of $M_N(\Sigma)$ and

$$F_1^N(t) = \frac{1}{N} \sum_{i=1}^{N} |y_i|^2 1_{\lambda_i \leq t}, \text{ with } y = U^* x$$

also a c.d.f. (but combining the eigenvectors).

Non Gaussian matrices II.

Let

$$G_N(t) = \sqrt{N}\left(F_1^N(t) - F_*^N(t)\right)$$

where F_*^N is the c.d.f. of ρ_{MP} when $\gamma \to p/N$ and $H \to \rho_N(\Sigma)$ spectral measure of Σ).
Here $G_N \simeq X_N$ and should be close to $B(F(t))$ if B is a Brownian bridge.
Let also g be analytic on $[u_-, u_+]$.

Theorem Bai-Miao-Pan (2007)
Assume also that $\mathbb{E}|X_{ij}|^4 = 2$ and $x^*(\Sigma - zI)^{-1}x \to \int \frac{1}{\lambda - z} dH(\lambda)$. Then as $N \to \infty$,

$$\int g(x) dG_N(x) \to \text{ a Gaussian random variable (centered and with known variance).}$$

Remark: extension to non-white matrices but with the additionnal assumption on $x^*(\Sigma - zI)^{-1}x$.

A few explanations

- the finiteness of $\mathbb{E}|X_{ij}|^4$ "ensures" that the largest eigenvalues have the same asymptotic behavior as for a Gaussian sample (conjecture). If this moment is not finite, the eigenvectors associated to the largest eigenvalues are actually determined by the largest entries of X (Biroli-Bouchaud-Potters (2007), Auffinger-Ben Arous-Péché (2009)).
  The fact that the fourth moment needs to equal that of a Gaussian random variable was proved by Silverstein ('81).
  One needs a certain proximity with the Gaussian distribution!

- the assumption on x ensures that the projection of x on the eigenvectors of $M_N(\Sigma)$ does not see the lack of rotationnal invariance.
  It also ensures that $F_1^N(t) \to F(t)$ if F is the c.d.f of the Marchenko-Pastur distribution ρ_{MP}.

Eigenvectors: the non-white case.

Preliminary remarks

- Even for a Gaussian sample,the distribution of the eigenvectors is unknown if $\Sigma \neq Id$.
- It is NOT expected that the matrix of eigenvectors is Haar distributed.
- Only known result due to D. Paul (2006):

$$\Sigma = \text{diag}(\pi_1, 1, \ldots, 1) \text{ with } \pi_1 > 1 + 1/\sqrt{\gamma}.$$

Let u_1 (resp. e_1) be the normalized eigenvector of $M_N(\Sigma)$ (resp. of Σ) associated to λ_1 (resp. π_1):

$$\lim_{N\to\infty} |< u_1, e_1 >| = \sqrt{\frac{1-\gamma/(\pi_1-1)^2}{1+\gamma/(\pi_1-1)}} \text{ a.s .}$$

Idea: perturbation of the eigenvector associated to π_1 (the largest eigenvalue of Σ) by a random matrix.

Another approach (Ledoit-Péché (2009))

The idea is to study functionals:

$$\theta_N(g) := \frac{1}{N}\mathrm{Tr}\left(g(\Sigma)(M_N(\Sigma) - zI)^{-1}\right),$$

with $z \in \mathbb{C}^+ = \{z \in \mathbb{C}, \Im z > 0\}$,
g is a regular function (bounded with a finite number of discontinuities or analytic),
$g(\Sigma) = V\mathrm{diag}(g(\pi_1), \ldots, g(\pi_N))V^*$ if V is the matrix of eigenvectors of Σ.

Aim : understand how the eigenvectors of $M_N(\Sigma)$ project onto those of Σ.

Remarks:
-if $g \equiv 1$, then θ is just the Stieltjes transform of μ_N.
-If $\Sigma \propto Id$ useless. We thus concentrate on the case where $H \neq \delta_1$.

A theoretical result

Assume that the support of H is included in $[a_1, a_2]$ with $a_1 > 0$ and

$$\mathbb{E}|X_{ij}|^{12} < \infty \text{ independent of } N \text{ and } p.$$

Theorem: Ledoit-Péché (2009)
Let g be a bounded function with a finite number of discontinuities on $[a_1, a_2]$. Then $\theta_N(g) \to \theta(g)$ a.s. as $N \to \infty$ where

$$\forall z \in \mathbb{C}^+, \; \Theta^g(z) = \int_{-\infty}^{+\infty} \left\{\tau \left[1 - \gamma^{-1} - \gamma^{-1} z m_\rho(z)\right] - z\right\}^{-1} g(\tau) dH(\tau).$$

Remark: the same kernel

$$\left\{\tau \left[1 - \gamma^{-1} - \gamma^{-1} z m_\rho(z)\right] - z\right\}^{-1}$$

arises as in the Marchenko-Pastur theorem.

Corrolary 1.

Question: How much do the eigenvectors of $M_N(\Sigma)$ deviate from those of Σ?

We set $g = 1_{(-\infty,\tau)}$ and $\Phi_N(\lambda,\tau) = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\sum_{j=1}^{N} |u_i^* v_j|^2\, 1_{[\lambda_i,+\infty)}(\lambda) \times 1_{[\tau_j,+\infty)}(\tau)$.

Let v_j be the normalized eigenvector of Σ associated to π_j. The average of $N|u_i^* v_j|^2$ bearing on the eigenvectors associated to sample eigenvalues (resp. eigenvalues of the true covariance) in the interval $[\underline{\lambda}, \overline{\lambda}]$ (resp. $[\underline{\tau}, \overline{\tau}]$) is:

$$\frac{\Phi_N(\overline{\lambda},\overline{\tau}) - \Phi_N(\overline{\lambda},\underline{\tau}) - \Phi_N(\underline{\lambda},\overline{\tau}) + \Phi_N(\underline{\lambda},\underline{\tau})}{[F_N(\overline{\lambda}) - F_N(\underline{\lambda})] \times [H_N(\overline{\tau}) - H_N(\underline{\tau})]},$$

if F_N (resp. H_N) is the c.d.f. of $M_N(\Sigma)$ (resp. Σ).

If one can choose $\underline{\lambda}$, $\overline{\lambda}$ and $\underline{\tau}$, $\overline{\tau}$ arbitrarily close, then one gets precise information!

Corrolary 1.

Theorem: $\Phi_N(\lambda,\tau) \xrightarrow{a.s.} \Phi(\lambda,\tau)$ at any point of continuity of Φ. And $\forall(\lambda,\tau) \in \mathbb{R}^2$, $\Phi(\lambda,\tau) = \int_{-\infty}^{\lambda}\int_{-\infty}^{\tau} \varphi(l,t)\, dH(t)\, d\rho_{MP}(l)$, where

$$\varphi(l,t) = \begin{cases} \dfrac{\gamma^{-1} l t}{(at-l)^2 + b^2 t^2}, \quad 1 - \dfrac{1}{\gamma} - \dfrac{l\,\breve{m}_\rho(l)}{\gamma} =: a + ib, & \text{if } l > 0 \\ \dfrac{1}{(1-\gamma)[1 + \breve{m}_{\underline{\rho}}(0)\, t]} & \text{if } l = 0 \text{ and } \gamma < 1 \\ 0 & \text{otherwise} \end{cases}$$

Here $\breve{m}_\rho(0) = \lim_{z\to 0} m_\rho(z)$ and $m_{\underline{\rho}}$ is the limiting Stieltjes transform of $X^*\Sigma X/N$.

Thus in principle one can obtain precise information on the eigenvectors (but this assumes that one knows the c.d.f. of H_N).

Corrolary 2.

Question: how does $M_N(\Sigma)$ differ from Σ and how can we improve the initial estimator of Σ given by $M_N(\Sigma)$?

We get a better estimator by choosing $g(x) = x$.

One seeks an estimator of Σ of the kind UD_NU^*, D_N diagonal i.e. an estimator which has the same eigenvectors as $M_N(\Sigma)$.
The best estimator (Frobenius norm) is

$$\widetilde{D}_N = \text{diag}(\widetilde{d}_1, \ldots, \widetilde{d}_N) \quad \text{where} \quad \forall i = 1, \ldots, N \quad \widetilde{d}_i = u_i^* \, \Sigma_N \, u_i.$$

Can we say a few things on the $\tilde{d}_i$'s:
yes asymptotically by choosing $g(x) = x$.

Corrolary 2.

We set

$$\forall x \in \mathbb{R}, \quad \Delta_N(x) = \frac{1}{N}\sum_{i=1}^{N} \widetilde{d}_i \, 1_{[\lambda_i,+\infty)}(x) = \frac{1}{N}\sum_{i=1}^{N} u_i^* \Sigma_N u_i \times 1_{[\lambda_i,+\infty)}(x).$$

Then one has

$$\forall i = 1, \ldots, N \qquad \widetilde{d}_i = \lim_{\varepsilon \to 0^+} \frac{\Delta_N(\lambda_i + \varepsilon) - \Delta_N(\lambda_i - \varepsilon)}{F_N(\lambda_i + \varepsilon) - F_N(\lambda_i - \varepsilon)}.$$

Theorem: For all $x \neq 0$, $\Delta_N(x) \to \Delta(x)$. Moreover $\Delta(x) = \int_{-\infty}^{x} \delta(\lambda)\, dF(\lambda)$, with

$$\forall \lambda \in \mathbb{R}, \qquad \delta(\lambda) = \begin{cases} \frac{\lambda}{\left|1-\gamma^{-1}-\gamma^{-1}\lambda\, \breve{m}_\rho(\lambda)\right|^2} & \text{if } \lambda > 0 \\ \dfrac{\gamma}{(1-\gamma)\, \breve{m}_{\underline{\rho}}(0)} & \text{if } \lambda = 0 \text{ and } \gamma < 1 \\ 0 & \text{otherwise.} \end{cases}$$

An improved estimator

We consider the "improved" estimator $\widetilde{S}_N := UD'U^*$, where

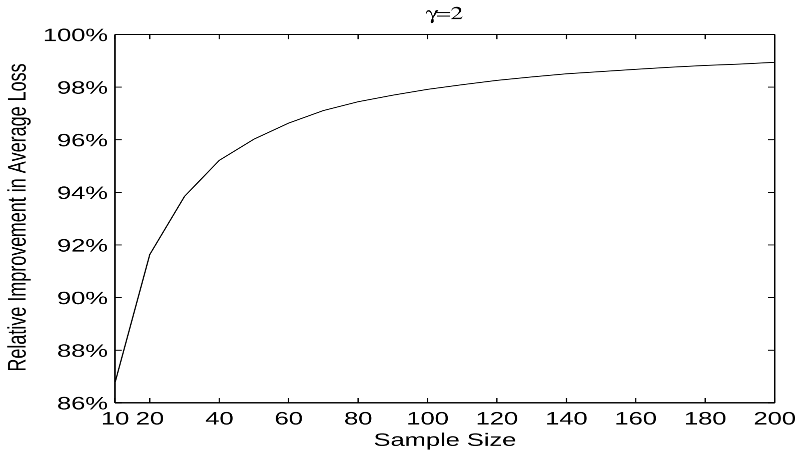
$$D'_i = \lambda_i / |1 - \gamma^{-1} - \gamma^{-1}\lambda_i \breve{m}_\rho(\lambda_i)|^2.$$

We ran 10,000 simulations with $\rho_N(\Sigma) = 0.2\delta_1 + 0.4\delta_3 + 0.4\delta_{10}, \quad \gamma = 2$ and increasing the number of variables p from 5 to 100. For each simulation, we calculate the "Percentage Relative Improvement in Average Loss" (PRIAL):
if M is an estimator of Σ_N and if $|A|_F^2 = \mathrm{Tr} AA^*$ (Frobenius norm),

$$PRIAL(M) = 100 \times \left[1 - \frac{\mathbb{E}\left\| M - U_N \widetilde{D}_N U_N^* \right\|_F^2}{\mathbb{E}\left\| M_N(\Sigma) - U_N \widetilde{D}_N U_N^* \right\|_F^2} \right].$$

Simulations

Even for small sizes, $p = 40$, the PRIAL is 95%.



Remarks and conclusion

Eigenvalues

-Using the techniques introduced by Tao-Vu (2009), the universality results can surely be improved for largest and smallest eigenvalues (condition number).

Eigenvectors

-$\theta_N(g)$ is a new tool that allows to study the average behavior of the eigenvectors: for instance we cannot recover D. Paul's result for the eigenvector associated to the largest eigenvalue separating from the bulk.

-in general we cannot say anything on the eigenvectors associated to extreme eigenvalues: average behavior of the eigenvectors.

-for the moment theoretical results only: one has to define first appropriate estimators for $\breve{m}_\rho, H_N \ldots$