Computation, statistics and random matrix theory

Edgar Dobriban

Statistics, Stanford

Outline

Computation: SPECTRODE

Statistics: PCA

Broadening the reach of RMT

Supplement

slides available at github.com/dobriban

Computation, statistics and random matrix theory

- Computation, statistics and random matrix theory intersect fruitfully
  - RMT leads to new computational problem: compute limit ESD of general covariance matrices
  - New method SPECTRODE
  - Enables solving challenging problem in theoretical statistics (PCA)
- Bigger picture: Broadening the reach of RMT (software, teaching)

Computation: SPECTRODE

Statistics: PCA

Broadening the reach of RMT

Supplement

Covariance matrices

- Understanding the correlations between sets of observations (e.g., age, height, weight, disease, genetic markers) is key
- Covariance matrix is fundamental tool
- Random vector x, $\Sigma = \mathrm{Cov}\left[x\right] = \mathbb{E}(x - \mathbb{E}x)(x - \mathbb{E}x)^\top$
- n samples x_i, arrange into $n \times p$ matrix X
- Sample covariance matrix $\widehat{\Sigma} = n^{-1} X_c^\top X_c$, $X_c = X - 1_n \cdot \bar{X}$
- Goal: Understand $\hat{\Sigma}$ vs Σ

RMT models for covariance matrices

- $X = Z_{n\times p}\Sigma^{1/2}$
  - $Z_{n\times p}$ has iid standardized entries
  - Σ unobserved $p \times p$ covariance matrix: $\mathrm{Cov}\left[x_i, x_i\right] = \Sigma$
- High dimension: $n, p \to \infty$, $p/n \to \gamma > 0$ (wlog $p/n = \gamma$)
- $l_1 \geq l_2 \geq \ldots \geq l_p$ eigenvalues of Σ. Distribution $H_p = p^{-1}\sum_i \delta_{l_i}$
- $H_p \Rightarrow H$
- Sample eigenvalues λ_i of $\widehat{\Sigma}$ inconsistent estimates: $\lambda_i \nrightarrow l_i$
- "Bulk" distribution of λ_i converges to limiting ESD $F_{\gamma,H}$
- Marchenko-Pastur map $H \to F_{\gamma,H}$

MP map example: white noise $\Sigma = I_p$

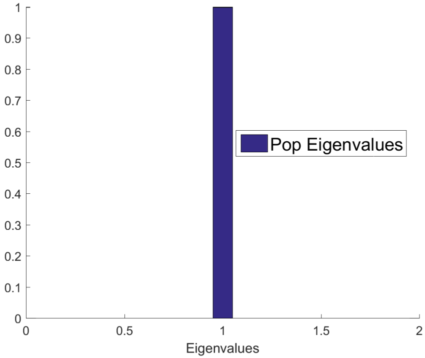


Figure : Eigenvalues $H = \delta_1$ of an identity covariance matrix $\Sigma = I_p$.

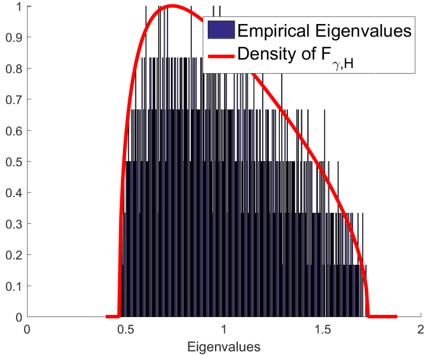


Figure : Marchenko-Pastur density: $g(x) = \sqrt{(\gamma_+ - x)(x - \gamma_-)}/\ (2\pi\gamma x)$, $x \in [\gamma_-, \gamma_+]$, $\gamma_\pm = (1 \pm \sqrt{\gamma})^2$, $\gamma = 1/10$.

MP map example: Autoregressive model, order 1, (AR-1)

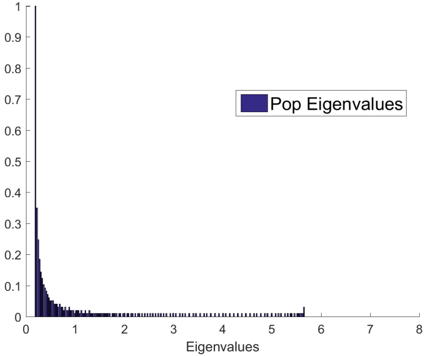


Figure : Eigenvalues H of an AR-1 covariance matrix Σ with $\Sigma_{ij} = \rho^{|i-j|}$ ($p = 400$; $\rho = 0.7$).

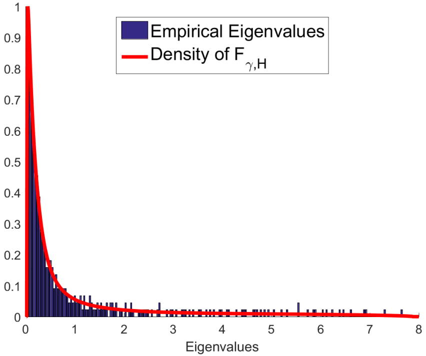


Figure : Eigenvalues of a sample covariance matrix $\widehat{\Sigma}$ with $n = 800$ samples.

Computation

- How to compute the limit ESD?
- Previous proposals
  - "Successive approximation" (Marchenko and Pastur, 1967)
  - Polynomial method (Rao & Edelman, 2008)
  - Fixed-pont algorithm (Couillet et al., 2011)
- Dobriban (2015) developed carefully motivated new method SPECTRODE
  1. a new algorithmic idea
  2. with proven properties
  3. empirically much faster than previous proposals

SPECTRODE computes limit ESD $F_{\gamma,H}$

SPECTRODE: Input and Output
Input:
$H \leftarrow$ population spectrum
$\gamma \leftarrow$ aspect ratio
$\varepsilon \leftarrow$ precision control
Output:
$\hat{l}_k, \hat{u}_k \leftarrow$ endpoints of intervals in the support
$\hat{f}(x) \leftarrow$ density of limit ESD $F_{\gamma,H}$
$\hat{s}(x) \leftarrow$ Stieltjes transform of limit ESD $F_{\gamma,H}$

Computation: the idea

- Stieltjes transform $s(z) = \int \frac{1}{x-z} dF(x)$
- $\underline{s}(z)$ ST of $(1-\gamma)F_{\gamma,H} + \gamma\delta_0$
- MP/Silverstein equation defines inverse ST, for $z \in \mathbb{C}^+$

$$z = -\frac{1}{\underline{s}(z)} + \gamma \int \frac{t}{1 + t\underline{s}(z)} dH(t)$$

- Find edges: ST increasing for $x \in \mathbb{R}$ outside the support of $F_{\gamma,H}$
  - has increasing inverse there (Silverstein and Choi, 1995)
- Run ODE derived from Silv eq to find smooth density within support
  - Starting point using fixed-pont algoritm

SPECTRODE: "Comb" model

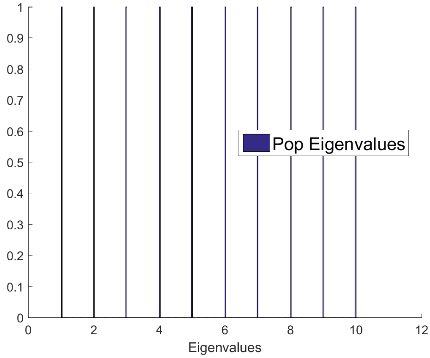


Figure : "Comb" model
$H = 10^{-1} \sum_{i=1}^{10} \delta_{l_i}$, with $l_i = i$.

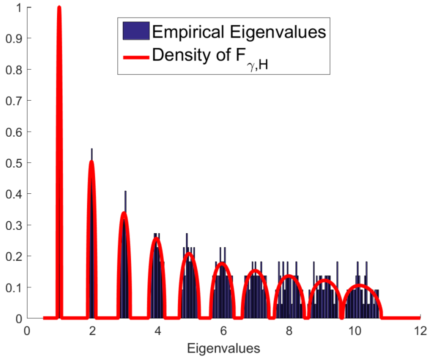


Figure : Eigenvalues of $\widehat{\Sigma}$ with $n = 800$ samples, and density. Density computed with SPECTRODE

SPECTRODE: a "universal" MP calculator

- A lot of theoretical results in this area
- Many (all?) depend on $F_{\gamma,H}$
- SPECTRODE useful for many problems, see Dobriban (2015):
  - Examples of limit spectra (previous slides)
  - Quantiles, moments and contour integrals of MP distribution
  - Principal component analysis (next)
  - Spiked models
  - Teaching (next)
- Matlab and R software at github.com/dobriban
  - With documentation and examples

There's more: Z. Che (2016) Correlated Wigner Mxes

For each $N \in \mathbb{N}$, we consider an array of centered real random variables $(x_{ij})_{1 \le i \le j \le N}$. We assume that there is a four dimensional tensor $\xi = \xi^{(N)}$ such that

$$\mathbb{E}\left[x_{ij} x_{kl}\right] = \xi_{ijkl} \,. \tag{2.1}$$

We assume that the (x_{ij}) are K-dependent for some constant $K > 0$

For each N define a matrix-valued map $\Xi : \mathbb{C}^{N \times N} \to \mathbb{C}^{N \times N}$ through

$$(\Xi(M))_{ik} := \frac{1}{N} \sum_{j,l} \xi_{ijkl} M_{jl} \,, \forall i, k \in \mathbb{N}_N \,. \tag{2.5}$$

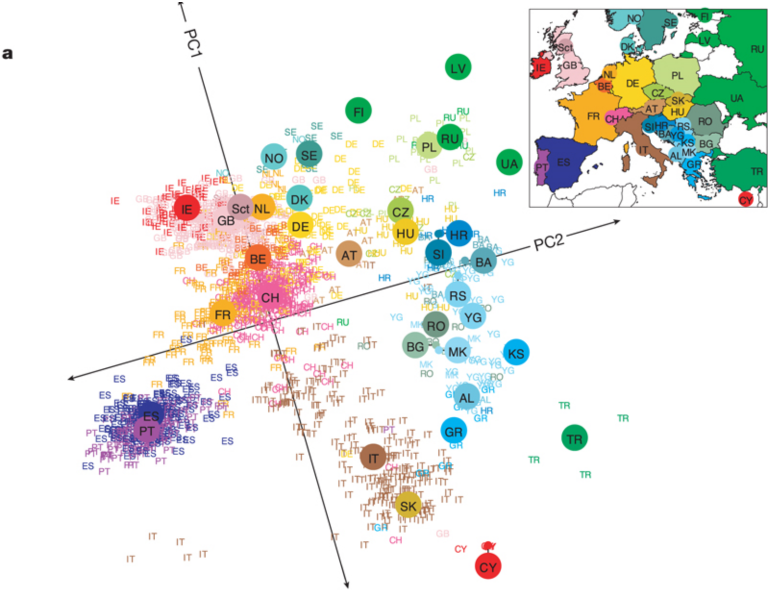
For each $N \in \mathbb{N}$ and $z \in \mathbb{C}^+$ we consider the equation

$$M(-z - \Xi(M)) = I \,. \tag{2.6}$$

PCA

- Principal component analysis (PCA) is a widely used method for dimension reduction
- X an $n \times p$ matrix.
  - n samples from centered p-dimensional population
- PCs: linear combinations Xu_i of features that explain a lot of variance
- u_i - eigenvectors of sample covariance matrix $\widehat{\Sigma} = n^{-1}X^\top X$
- Corresponding eigenvalue λ_i is variance of PC

Genes mirror geography within Europe – Novembre et al. (2008)



PCA Review: spiked covariance model

- (Signal) spike $t = l_1$ fixed
- (Noise) distribution of $l_2, ..., l_p$ converges to H in Restricted KS sense
  - $H_p \Rightarrow_{RKS} H$, if $H_p \Rightarrow_{KS} H$, and $\max Support(H_p) \to \max Support(H)$
- Top sample eigenvalue λ_1 "pushed upward" from l_1
- BBP phase transition (Baik et al., 2005; Benaych-Georges and Nadakuditi, 2011):
  - Above phase transition (PT): if l_1 large, λ_1 separates from MP map "bulk"
  - Below PT: else λ_1 does not separate

Spiked model: AR-1 example

- Population covariance matrix

$$\Sigma = \begin{bmatrix} t & 0^\top \\ 0 & M \end{bmatrix}$$

- Spike t
- M is a $p \times p$ AR(1) covariance matrix $M_{ij} = \rho^{|i-j|}$. $\rho = 0.5$
- Sample $n = 500$ Gaussian variates of $p = 250$ dimensions
- $mean(trace(M)) = 1$
- null: $t = 1$, alternative: larger t

AR-1 Example - Above PT

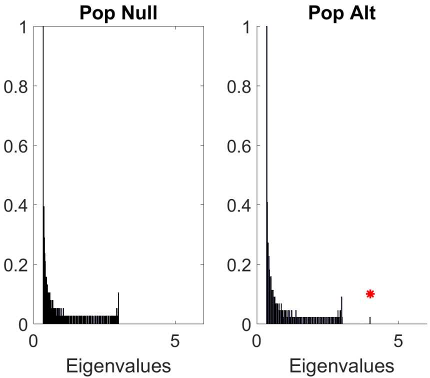


Figure : Eigenvalues of Σ. Null: $t = 1$. Alternative: $t = 4$.

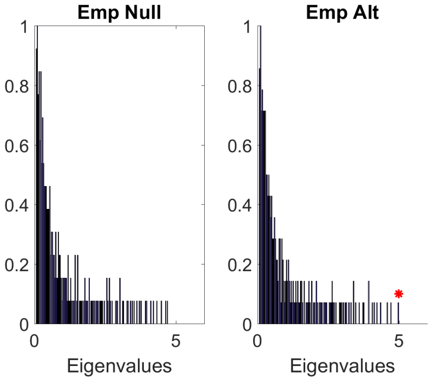


Figure : Eigenvalues of $\widehat{\Sigma}$. Null and alternative.

AR-1 Example - Below PT

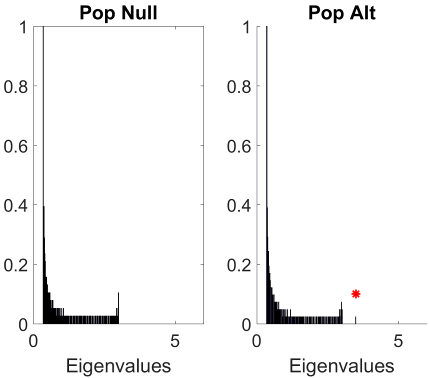


Figure : Eigenvalues of Σ. Null: $t = 1$. Alternative: $t = 3.5$.

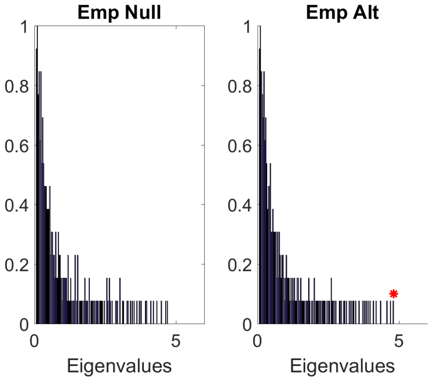


Figure : Eigenvalues of $\widehat{\Sigma}$. Null and alternative.

Statistical implications

- Below PT, top eigenvalue test based on λ_1 has trivial power
- ... Despite its near-universal use
- ... Despite its asy optimality in low dim, p fixed (Anderson, 1963)
- Can we detect PCs below the phase transition? If so, how?
- Onatski et al. (2013, 2014) (OMH) the standard spiked model of Johnstone (2001)

$$H_0 : \Sigma_p = I_p, \text{ vs}$$

$$H_1 : \Sigma_p = I_p + \sum_{j=1}^{r} (l_j - 1) v_j v_j^\top, v_j \text{ unknown orthonormal}$$

- Optimal test is equivalent to a linear spectral statistic (LSS)

$$\mathrm{tr}(\varphi(\widehat{\Sigma}))$$

Onatski, Moreira, Hallin (2013)

The Annals of Statistics
2013, Vol. 41, No. 3, 1204–1231
DOI: 10.1214/13-AOS1100

ASYMPTOTIC POWER OF SPHERICITY TESTS FOR HIGH-DIMENSIONAL DATA

BY ALEXEI ONATSKI, MARCELO J. MOREIRA[1] AND MARC HALLIN[2]

University of Cambridge, FGV/EPGE, and Université libre de Bruxelles and Princeton University

This paper studies the asymptotic power of tests of sphericity against perturbations in a single unknown direction as both the dimensionality of the data and the number of observations go to infinity. We establish the convergence, under the null hypothesis and contiguous alternatives, of the log ratio of the joint densities of the sample covariance eigenvalues to a Gaussian process indexed by the norm of the perturbation. When the perturbation norm is larger than the *phase transition threshold* studied in Baik, Ben Arous and Péché [*Ann. Probab.* **33** (2005) 1643–1697] the limiting process is degenerate, and discrimination between the null and the alternative is asymptotically certain. When the norm is below the threshold, the limiting process is nondegenerate, and the joint eigenvalue densities under the null and alternative hypotheses are mutually contiguous. Using the asymptotic theory of statistical experiments, we obtain asymptotic power envelopes and derive the asymptotic power for various sphericity tests in the contiguity region. In particular, we show that the asymptotic power of the Tracy–Widom-type tests is trivial (i.e., equals the asymptotic size), whereas that of the eigenvalue-based likelihood ratio test is strictly larger than the size, and close to the power envelope.

Onatski, Moreira, Hallin (2014)

The Annals of Statistics
2014, Vol. 42, No. 1, 225–254
DOI: 10.1214/13-AOS1181

SIGNAL DETECTION IN HIGH DIMENSION: THE MULTISPIKED CASE

BY ALEXEI ONATSKI[1], MARCELO J. MOREIRA[2] AND MARC HALLIN[3]

University of Cambridge, FGV/EPGE and Université libre de Bruxelles and Princeton University

This paper applies Le Cam's asymptotic theory of statistical experiments to the signal detection problem in high dimension. We consider the problem of testing the null hypothesis of sphericity of a high-dimensional covariance matrix against an alternative of (unspecified) multiple symmetry-breaking directions (*multispiked* alternatives). Simple analytical expressions for the Gaussian asymptotic power envelope and the asymptotic powers of previously proposed tests are derived. Those asymptotic powers remain valid for non-Gaussian data satisfying mild moment restrictions. They appear to lie very substantially below the Gaussian power envelope, at least for small values of the number of symmetry-breaking directions. In contrast, the asymptotic power of Gaussian likelihood ratio tests based on the eigenvalues of the sample covariance matrix are shown to be very close to the envelope. Although based on Gaussian likelihoods, those tests remain valid under non-Gaussian densities satisfying mild moment conditions. The results of this paper extend to the case of multispiked alternatives and possibly non-Gaussian densities, the findings of an earlier study [*Ann. Statist.* **41** (2013) 1204–1231] of the single-spiked case. The methods we are using here, however, are entirely new, as the Laplace approximation methods considered in the single-spiked context do not extend to the multispiked case.

Our results

- OMH is one example where we can calculate optimal tests
- In this talk, we find optimal tests for local alternatives generally
- Our construction is mathematically precise and clean
- It can derive OMH plus much else
- Relies crucially on SPECTRODE for computation

Local alternatives model

- Recall $X = Z_{n\times p}\Sigma^{1/2}$ and $H_p = p^{-1}\sum_{i=1}^{p}\delta_{l_i}$ the spectrum of Σ
- Model bulk H perturbed by spikes G_0 (vs G_1)
- Local alternatives model:

$$H_{p,0} : H_p = \left(1 - \frac{h}{p}\right) H + \frac{h}{p} G_0, \text{ vs}$$
$$H_{p,1} : H_p = \left(1 - \frac{h}{p}\right) H + \frac{h}{p} G_1.$$

- h is number of spikes (fixed)
- E.g., standard spiked model: $H = G_0 = \delta_1,\ G_1 = \delta_t$
- Allows correlations, flexible data modelling

Optimal tests in local alternatives model

- Given (H, h, G_0, G_1, γ) we derive a linear spectral statistic $T = \mathrm{tr}\{\varphi(\widehat{\Sigma})\}$.
- Gives the asymptotically best test for $H_{p,0}$ against $H_{p,1}$

Mean-variance problem

- There are mean and variance parameters $\mu_\varphi, \sigma_\varphi^2$ s.t for some constants c_p
  - Under $H_{p,0}$, $\mathrm{tr}(\varphi(\widehat{\Sigma})) - c_p \Rightarrow \mathcal{N}(0, \sigma_\varphi^2)$
  - Under $H_{p,1}$, $\mathrm{tr}(\varphi(\widehat{\Sigma})) - c_p \Rightarrow \mathcal{N}(\mu_\varphi, \sigma_\varphi^2)$.
- With $\langle f, g\rangle = \int_{\mathcal{I}} f(x)g(x)dx$

$$\mu_\varphi = -h\langle \varphi', \Delta\rangle \quad \text{and}$$
$$\sigma_\varphi^2 = \langle \varphi', K\varphi'\rangle.$$

- Find optimal LSS φ, maximizing the efficacy

$$\max_\varphi \frac{\mu_\varphi}{\sigma_\varphi}$$

Main result: Finding the optimal LSS

Theorem (Dobriban (2016))

Two cases for testing (H, G_0) vs (H, G_1) in the local alternatives model:

1. *If $\Delta \in \mathrm{Im}(K)$, the optimal linear spectral statistics φ are given by a Fredholm integral equation:*

$$K(\varphi') = -\eta\Delta,$$

   where $\eta > 0$ is any constant.

2. *On the other hand, if $\Delta \notin \mathrm{Im}(K)$, then the maximal efficacy is $+\infty$.*

Computing the optimal LSS

- Solve $Kg = -\Delta$, i.e., $\int k(x,y)g(y)dy = -\Delta(x)$,

$$k(x,y) = \frac{1}{2\pi^2}\log\left(1 + 4\frac{\Im(\underline{s}(x))\,\Im(\underline{s}(y))}{|\underline{s}(x) - \underline{s}(y)|^2}\right)$$

  and $\underline{s}(x)$ is Stieltjes transform of $(1-\gamma)F_{\gamma,H} + \gamma\delta_0$
- Use SPECTRODE to compute $\underline{s}(x)$, k, Δ. Discretize to linear equation

Optimal LSS example: AR-1

- population covariance matrix

$$\Sigma = \begin{bmatrix} t & 0^\top \\ 0 & M \end{bmatrix}$$

- Spike t
- $M_{ij} = \rho^{|i-j|}$
- $H = spec(M)$
- Test "pure AR" vs "spiked AR"

$$H_{p,0} : H_p = \left(1 - \frac{1}{p}\right) H + \frac{1}{p}\delta_1, \text{ vs}$$
$$H_{p,1} : H_p = \left(1 - \frac{1}{p}\right) H + \frac{1}{p}\delta_t.$$

Optimal LSS example: AR-1

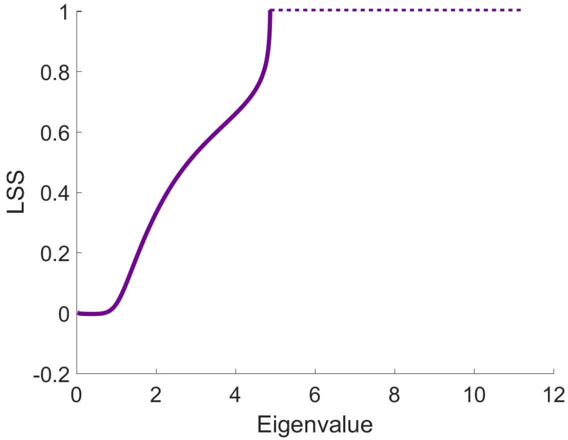


Figure : New optimal LSS $\varphi(x)$ in AR-1 model: $\gamma = 0.5, \rho = 0.5$, $t = 3.5$ below PT.

Probability background - CLT for LSS

Theorem. [Bai and Silverstein (2004) CLT]: Let $X = Z_{n\times p}\Sigma^{1/2}$ and Z_{ij} iid real standardized with $\mathbb{E}Z_{ij}^4 = 3$. If $H_p \Rightarrow H$, for φ analytic on a compact interval $\mathcal{I}$ including all supports of ESDs, we have

$$\mathrm{tr}(\varphi(\widehat{\Sigma})) - p\int_{\mathcal{I}} \varphi(x)dF_{\gamma,H_p}(x) \Rightarrow \mathcal{N}(m_\varphi, \sigma_\varphi^2)$$

- $\sigma_\varphi^2 = \int_{\mathcal{I}\times\mathcal{I}} \varphi'(x)\varphi'(y)k(x,y)dx\,dy = \langle \varphi', K\varphi'\rangle$, where k is a covariance kernel, and K is the associated operator
- $\underline{s}(x)$ is the limit Stieltjes transform of $(1-\gamma)F_{\gamma,H} + \gamma\delta_0$ as $z \to x \in \mathbb{R}$

$$k(x,y) = \frac{1}{2\pi^2}\log\left(1 + 4\frac{\Im(\underline{s}(x))\,\Im(\underline{s}(y))}{|\underline{s}(x) - \underline{s}(y)|^2}\right)$$

Software

- MP area of RMT (general non-null spectra)
  - A lot of theory; deep insights
  - Relatively few applications
- Why?
  - Pracitioners cannot see relevance
  - Hard to understand
  - No easy way to try it
- Developing software can broaden the reach of RMT
  - "the development of computational tools will go a long way in making the results of RMT accessible to professional statisticians" (Paul & Aue, 2014)

Software: EigenEdge

- Developing publicly available software for working with large random matrices
  - Matlab and R at github.com/dobriban. Documentation and examples
- Focus: Covariance matrices, MP distributions, statistical methods
- Complements existing tools: RMTStat (Johnstone, Perry, ...), RMTool (Nadakuditi & Edelman)

Software: EigenEdge components

- SPECTRODE: compute limit ESD
  - also: quantiles, moments, contour integrals
- general spiked models
  - forward/inverse maps for spikes and cosines
- linear spectral statistics
  - mean, variance, optimal LSS
- statistical methods (in progress)
  - PCA: shrinkage of PC scores
  - tests for sphericity (future)
- ...

Teaching

- Computational courses in RMT+statistics broaden the reach
- David Donoho's course "High-Dimensional Stats and RMT", 2016
- Rigorous experiments for phenomena beyond the reach of theory
- e.g., limit of eigenvalue quantile process (gen. BB)
- SPECTRODE a key tool

Summary

- Computation: New algorithm SPECTRODE
- Statistics: Optimal testing for principal components in high dimensions
- Software: Broaden the reach of RMT

- Thanks
  - Support: NSF, HHMI
  - Discussion: David Donoho, Iain Johnstone

Using the CLT

Goal: get an expression for the mean

- In the local alternatives model:

$$\text{under } H_0 : \mathrm{tr}(\varphi(\widehat{\Sigma})) - p\int_{\mathcal{I}} \varphi(x) dF_{\gamma,H_{p,0}} \Rightarrow \mathcal{N}(m_\varphi, \sigma_\varphi^2), \text{ while}$$

$$\text{under } H_1 : \mathrm{tr}(\varphi(\widehat{\Sigma})) - p\int_{\mathcal{I}} \varphi(x) dF_{\gamma,H_{p,1}} \Rightarrow \mathcal{N}(m_\varphi, \sigma_\varphi^2).$$

- In the limit, test $\mathcal{N}(0, \sigma_\varphi^2)$ vs $\mathcal{N}(\mu_\varphi, \mu_\varphi^2)$, where

$$\mu_\varphi = \lim_{p\to\infty} \int_{\mathcal{I}} \varphi(x) d\left[p(F_{\gamma,H_{p,1}} - F_{\gamma,H_{p,0}})\right],$$

Key new object: Weak derivative of MP map

- The weak derivative of MP map F_γ is the signed measure

$$\delta\mathcal{F}_\gamma(H, G) = \lim_{\varepsilon \to 0} \frac{F_{\gamma,(1-\varepsilon)H+\varepsilon G} - F_{\gamma,H}}{\varepsilon}.$$

- so $p[F_{\gamma,H_{p,1}} - F_{\gamma,H_{p,0}}] \Rightarrow h \cdot \Delta$, where $\Delta = \delta\mathcal{F}_\gamma(H, G_1) - \delta\mathcal{F}_\gamma(H, G_0)$
- integrate by parts

$$\mu_\varphi = h \cdot \langle \varphi, d\Delta \rangle = -h \cdot \langle \varphi', \Delta \rangle.$$

Key new object: Weak derivative of MP map

- Nice properties:
  - Has a density within support of F_γ
  - Spike is above PT iff isolated point mass in $\delta\mathcal{F}_\gamma$ — a new perspective on phase transitions in spiked models

Weak derivative: Example

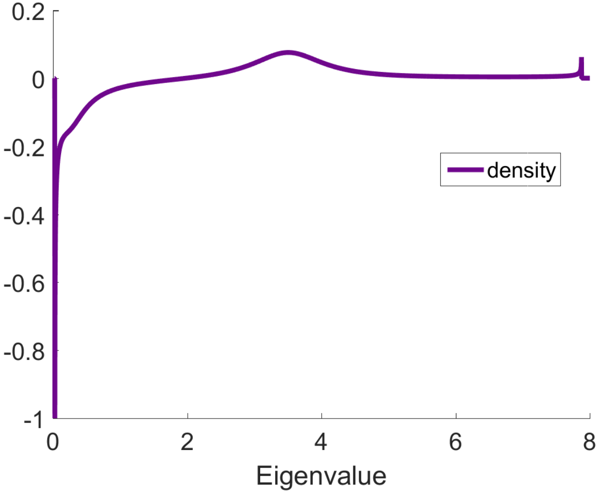


Figure : Density of weak derivative $\delta\mathcal{F}_\gamma(H, G)$ in AR-1 example $H = spec(\Sigma)$, $G = \delta_{3.5}$ below PT;

References I

Z. Bai and J. W. Silverstein. CLT for linear spectral statistics of large-dimensional sample covariance matrices. *The Annals of Probability*, 32(1A):553–605, 2004.

J. Baik, G. Ben Arous, and S. Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Annals of Probability*, 33(5):1643–1697, 2005.

F. Benaych-Georges and R. R. Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics*, 227(1):494–521, 2011.

R. Couillet, M. Debbah, and J. W. Silverstein. A deterministic equivalent for the analysis of correlated mimo multiple access channels. *IEEE Trans. Inform. Theory*, 57(6):3493–3514, 2011.

E. Dobriban. Efficient computation of limit spectra of sample covariance matrices. *Random Matrices: Theory and Applications*, 04(04): 1550019, 2015.

E. Dobriban. Sharp detection in pca under correlations: all eigenvalues matter. *to appear in The Annals of Statistics*, 2016.

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

V. A. Marchenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mat. Sb.*, 114(4):507–536, 1967.

J. Novembre, T. Johnson, K. Bryc, Z. Kutalik, A. R. Boyko, A. Auton, A. Indap, K. S. King, S. Bergmann, M. R. Nelson, et al. Genes mirror geography within europe. *Nature*, 456(7218):98–101, 2008.

A. Onatski, M. J. Moreira, and M. Hallin. Asymptotic power of sphericity tests for high-dimensional data. *The Annals of Statistics*, 41 (3):1204–1231, 2013.

A. Onatski, M. J. Moreira, and M. Hallin. Signal detection in high dimension: The multispiked case. *The Annals of Statistics*, 42(1): 225–254, 2014.

J. W. Silverstein and S.-I. Choi. Analysis of the limiting spectral distribution of large dimensional random matrices. *J. Multivariate Anal.*, 54(2):295–309, 1995.