Introduction to Random Matrix Theory: Theory and Applications Part II

Boaz Nadler

Department of Computer Science and Applied Mathematics
The Weizmann Institute of Science

May 2013

Part II: Modern Random Matrix Theory

Part II: Modern Random Matrix Theory

Starting point:

[Eugene Wigner, 1955, 1958]
[Mehta and Gaudin, 1960]

random matrix models to study quantum phenomena.

Example: Energy levels of atom = eigenvalues of Hermitian operator

$$H\psi_j = E_j\psi_j$$

Low energy states can be tackled analytically. For higher energy levels, analysis becomes intractable.

Part II: Modern Random Matrix Theory

Starting point:

[Eugene Wigner, 1955, 1958]
[Mehta and Gaudin, 1960]

random matrix models to study quantum phenomena.

Example: Energy levels of atom = eigenvalues of Hermitian operator

$$H\psi_j = E_j\psi_j$$

Low energy states can be tackled analytically. For higher energy levels, analysis becomes intractable.

Wigner replaced operator H by finite large random $m \times m$ Hermitian matrix H_m.

Wigner Matrix

Wigner Matrix [Hermite/Gaussian ensemble]:

- $m \times m$ Hermitian matrix W,

Wigner Matrix

Wigner Matrix [Hermite/Gaussian ensemble]:

- $m \times m$ Hermitian matrix W,

- $W_{i,j}$ independent (except symmetry constraint)

$$\mathbb{E}[W_{i,j}] = 0 \quad \mathbb{E}[W_{i,j}^2] = \begin{cases} \sigma^2 & i \neq j \\ 2\sigma^2 & i = j \text{ (but } \sigma^2 \text{ also ok)} \end{cases}$$

Assume $\mathbb{E}[W_{i,j}^k] < \infty$ (weaking possible)

Wigner Matrix

Wigner Matrix [Hermite/Gaussian ensemble]:

- $m \times m$ Hermitian matrix W,
- $W_{i,j}$ independent (except symmetry constraint)

$$\mathbb{E}[W_{i,j}] = 0 \quad \mathbb{E}[W_{i,j}^2] = \begin{cases} \sigma^2 & i \neq j \\ 2\sigma^2 & i = j \text{ (but } \sigma^2 \text{ also ok)} \end{cases}$$

Assume $\mathbb{E}[W_{i,j}^k] < \infty$ (weaking possible)

If $\sigma^2 = 1/m$, then this is a *standard Wigner Matrix*.

Empirical Spectral Measure

Empirical Spectral Distribution (ESD):

$$\boxed{F_m(t) = \frac{1}{m}\#\{i \,|\, \ell_i(W) \leq t\}}$$

Empirical Spectral Measure

Empirical Spectral Distribution (ESD):

$$\boxed{F_m(t) = \frac{1}{m}\#\{i \,|\, \ell_i(W) \leq t\}}$$

Question: *Does F_m converge to limiting F as $m \to \infty$?*

Wigner's Semi-Circle Law

Theorem: As $m \to \infty$, for standard Wigner matrix,

$$dF_m(x) \to \frac{1}{4\pi}\sqrt{4 - x^2}dx \qquad x \in [-2, 2]$$

Wigner's Semi-Circle Law

Theorem: As $m \to \infty$, for standard Wigner matrix,

$$dF_m(x) \to \frac{1}{4\pi}\sqrt{4 - x^2}dx \qquad x \in [-2, 2]$$

Numerical Illustration:

```
X = randn(m,m);
```

Wigner's Semi-Circle Law

Theorem: As $m \to \infty$, for standard Wigner matrix,

$$dF_m(x) \to \frac{1}{4\pi}\sqrt{4 - x^2}dx \qquad x \in [-2, 2]$$

Numerical Illustration:

```
X = randn(m,m);
W = 1/sqrt(2*m) (X + X') ;
```

Wigner's Semi-Circle Law

Theorem: As $m \to \infty$, for standard Wigner matrix,

$$dF_m(x) \to \frac{1}{4\pi}\sqrt{4 - x^2}dx \qquad x \in [-2, 2]$$

Numerical Illustration:

```
X = randn(m,m);
W = 1/sqrt(2*m) (X + X') ;
L = eig(W);
```

Wigner's Semi-Circle Law

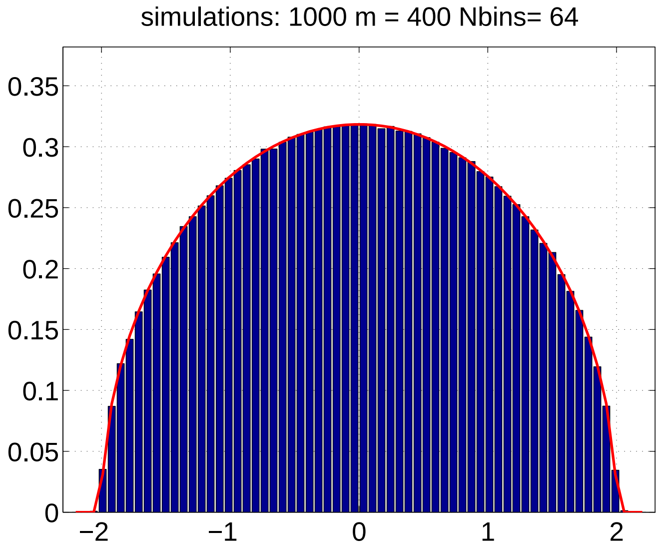
Theorem: As $m \to \infty$, for standard Wigner matrix,

$$dF_m(x) \to \frac{1}{4\pi}\sqrt{4 - x^2}dx \qquad x \in [-2, 2]$$

Numerical Illustration:

```
X = randn(m,m);
W = 1/sqrt(2*m) (X + X') ;
L = eig(W);
histL = hist(L,x);
```

Wigner's Semi-Circle Law



Wigner's Semi-Circle Law

Wigner's original proof $=$ method of moments.

Instead of studying empirical spectral distribution F_m, study its *moments*.

$$\int x^k dF_m(x)$$

Wigner's Semi-Circle Law

Wigner's original proof = method of moments.

Instead of studying empirical spectral distribution F_m, study its *moments*.

$$\int x^k dF_m(x)$$

Prove that moments converge to some limits.

Wigner's Semi-Circle Law

Wigner's original proof = method of moments.

Instead of studying empirical spectral distribution F_m, study its *moments*.

$$\int x^k dF_m(x)$$

Prove that moments converge to some limits.

Find distribution corresponding to these limiting moments.

Method of Moments / Wigner Matrix

Simple Observation:

$$\beta_{m,k} = \int x^k dF_m(x) = \frac{1}{m} \sum \ell_i^k = \frac{1}{m} Tr[W^k].$$

Simple Observation:

$$\beta_{m,k} = \int x^k dF_m(x) = \frac{1}{m}\sum \ell_i^k = \frac{1}{m}Tr[W^k].$$

Since W_{ij} are independent zero mean, many entries in W^k have zero mean as well !

Example:

$$\frac{1}{m}Tr[W] = \frac{1}{m}\sum_{i=1}^{m} W_{ii}$$

Hence

$$\mathbb{E}\,Tr[W] = 0.$$

Example:

$$\frac{1}{m} Tr[W] = \frac{1}{m} \sum_{i=1}^{m} W_{ii}$$

Hence

$$\mathbb{E}\, Tr[W] = 0.$$

Next, $(Tr[W])^2 = \sum_{i,j} W_{ii} W_{jj}$.

Method of Moments / Wigner Matrix

Example:

$$\frac{1}{m} Tr[W] = \frac{1}{m} \sum_{i=1}^{m} W_{ii}$$

Hence

$$\mathbb{E}\, Tr[W] = 0.$$

Next, $(Tr[W])^2 = \sum_{i,j} W_{ii} W_{jj}$.
Hence,

$$\mathbb{E}[(\tfrac{1}{m} Tr[W])^2] = 1/m^2 \sum_i W_{ii}^2 = \tfrac{1}{m^2} \cdot m \cdot \tfrac{2}{m} = \tfrac{1}{m^2}$$

Conclusion: First moment $\beta_{m,1}$ has zero mean, and variance tends to zero as $m \to \infty$. Thus

$$\beta_{m,1} \to 0 \quad as \quad m \to \infty$$

Second Moment:

$$Tr[W^2] = \sum_i (W^2)_{ii} = \sum_i \sum_j W_{ij}^2$$

Hence,

$$\mathbb{E}\left[\frac{1}{m} Tr[W^2]\right] = \frac{1}{m} \cdot m^2 \frac{1}{m} = 1.$$

Second Moment:

$$Tr[W^2] = \sum_i (W^2)_{ii} = \sum_i \sum_j W_{ij}^2$$

Hence,

$$\mathbb{E}\left[\frac{1}{m} Tr[W^2]\right] = \frac{1}{m} \cdot m^2 \frac{1}{m} = 1.$$

More complicated calculation,

$$Var[\frac{1}{m} Tr[W^2]] \to 0 \quad as \quad m \to \infty$$

Conclusion:

$$\beta_{m,2} \to 1$$

General Moments:

$$(W^k)_{11} = \sum_{i_1,\ldots,i_{k-1}} W_{1,i_1} W_{i_1 i_2} \cdots W_{i_{k-1}1}$$

General Moments:

$$(W^k)_{11} = \sum_{i_1,\ldots,i_{k-1}} W_{1,i_1} W_{i_1 i_2} \cdots W_{i_{k-1}1}$$

For *odd* $k = 2r + 1$, at least one term appears only once,

$$\mathbb{E}[Tr(W^{2r+1})] = 0.$$

General Moments:

$$(W^k)_{11} = \sum_{i_1,\dots,i_{k-1}} W_{1,i_1} W_{i_1 i_2} \cdots W_{i_{k-1}1}$$

For *odd* $k = 2r + 1$, at least one term appears only once,

$$\mathbb{E}[Tr(W^{2r+1})] = 0.$$

For *even* $k = 2r$, each term must appear even number of times for non-zero mean.

If each term appears exactly twice, mean is $(1/m)^{k/2}$
Exact number of such terms is a combinatorial graph enumeration problem.

Summary:

$\frac{1}{m}\mathbb{E}tr[W^{2k+1}] = 0$

$\frac{1}{m}tr[W^{2k}] \to \textit{Catalan Numbers}$

Method of Moments / Wigner Matrix

Summary:

$\frac{1}{m}\mathbb{E}tr[W^{2k+1}] = 0$

$\frac{1}{m}tr[W^{2k}] \to$ *Catalan Numbers*

Distribution with these moments is *semi-circle law*.

Universality

Limiting Distribution depends only on underlying moments

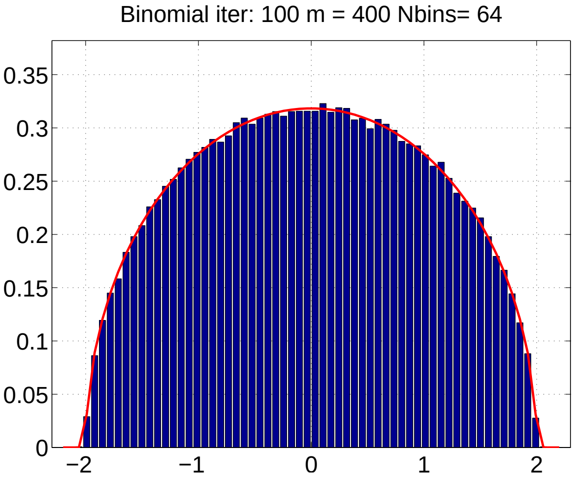
Let W be a standardized Wigner matrix whose entries come from *any* distribution with finite moments,

Then as $m \to \infty$ eigenvalues of W converge to semi-circle law.

Example: Matrix with entries ± 1

```
X = randn(m,m);
W = sign(X+X');
W = W / sqrt(m);
L = eig(W);
```

Universality / Semi-Circle Law



Method of Moments

Applicable to other more complicated random matrix models.

Mathematical Foundations:

Method of Moments

Applicable to other more complicated random matrix models.

Mathematical Foundations:

Lemma 1: *Moments uniquely determine a distribution.*

Method of Moments

Applicable to other more complicated random matrix models.

Mathematical Foundations:

Lemma 1: *Moments uniquely determine a distribution.*
Precisely, F - a distribution, with finite moments $\beta_k = \int x^k dF(x)$
If Carleman condition holds

$$\sum_k \beta_{2k}^{-1/2k} = \infty$$

then F is uniquely determined by $\{\beta_k\}$

Method of Moments

Method of Moments

Lemma 2: under some technical conditions,
Convergence of moments $\rightarrow$ convergence of distributions.

The Stieltjes Transform

Very useful tool in RMT.

Definition: $f(x)$ probability density function.

$$S_f(z) = \int_{-\infty}^{\infty} \frac{1}{x-z} f(x)dx \qquad z \in \mathbb{C}^+$$

where $\mathbb{C}^+ = \{z = x + \mathbf{i}y \,|\, y > 0\}$

The Stieltjes Transform

Very useful tool in RMT.

Definition: $f(x)$ probability density function.

$$S_f(z) = \int_{-\infty}^{\infty} \frac{1}{x-z} f(x)dx \qquad z \in \mathbb{C}^+$$

where $\mathbb{C}^+ = \{z = x + \mathbf{i}y \,|\, y > 0\}$

Inversion Formula:

$$F(a,b) = \frac{1}{\pi} \lim_{v \to 0^+} \int_a^b Im(S(u + \mathbf{i}v))du$$

The Stieltjes Transform

Lemma: *convergence of Stieltjes transforms* $\rightarrow$ *convergence of distributions*

The Stieltjes Transform

Lemma: *convergence of Stieltjes transforms $\rightarrow$ convergence of distributions*

Precisely:

- F_n sequence of probability distributions,

- $s_n(z)$ their Stieltjes transforms.

If $s_n(z) \rightarrow s(z)$ for all $z \in \mathbb{C}^+$ and $\lim_{v \to \infty} vs(v) = -1$,
then $F_n \rightarrow F$.

The Stieltjes Transform

Matrix Inversion Lemma:

$$Tr(A^{-1}) = \sum_{k=1}^{n} \frac{1}{a_{kk} - \alpha_k^T A_k^{-1} \alpha_k}$$

where α_k is k-th row of A with k-th element removed
A_k is k-th minor of A (remove k-th row and column).

Matrix Inversion Lemma:

$$Tr(A^{-1}) = \sum_{k=1}^{n} \frac{1}{a_{kk} - \alpha_k^T A_k^{-1} \alpha_k}$$

where α_k is k-th row of A with k-th element removed
A_k is k-th minor of A (remove k-th row and column).

Take $A = W_m - zI_m$. Then

$$\frac{1}{m} Tr(A^{-1}) = \frac{1}{m} \sum_{j=1}^{m} \frac{1}{\ell_j - z} = s_m(z).$$

The Stieltjes Transform

α_k = vector of size $(m-1)$ with i.i.d. zero mean entries, unit variance.

α_k *independent* of A_k and hence of A_k^{-1}.

The Stieltjes Transform

α_k = vector of size $(m-1)$ with i.i.d. zero mean entries, unit variance.

α_k *independent* of A_k and hence of A_k^{-1}.

Quadratic form $\alpha_k^T A_k^{-1} \alpha_k$ concentrates around mean value, which is $\frac{1}{m} Tr(W_m^{(k)} - zI_m^{(k)}) \approx s_m(z)$.

$a_{kk} \approx -z$.

The Stieltjes Transform

α_k = vector of size $(m-1)$ with i.i.d. zero mean entries, unit variance.

α_k *independent* of A_k and hence of A_k^{-1}.

Quadratic form $\alpha_k^T A_k^{-1} \alpha_k$ concentrates around mean value, which is $\frac{1}{m} Tr(W_m^{(k)} - zI_m^{(k)}) \approx s_m(z)$.

$a_{kk} \approx -z$.

Overall:

$$s_m(z) \approx \frac{-1}{z + s_m(z)}$$

The Stieltjes Transform

α_k = vector of size $(m-1)$ with i.i.d. zero mean entries, unit variance.

α_k *independent* of A_k and hence of A_k^{-1}.

Quadratic form $\alpha_k^T A_k^{-1} \alpha_k$ concentrates around mean value, which is $\frac{1}{m} Tr(W_m^{(k)} - zI_m^{(k)}) \approx s_m(z)$.

$a_{kk} \approx -z$.

Overall:

$$s_m(z) \approx \frac{-1}{z + s_m(z)}$$

Now invert - precisely the semi-circle law.

The Quarter Circle Law

[Marchenko and Pastur, 1965]

- Let X be $m \times n$ random matrix.
- X_{ij} zero mean, unit variance.
- Look at sample covariance matrix

$$S_n = \frac{1}{n} X X^T$$

- its eigenvalues $\ell_1, \ldots, \ell_m$.

The Quarter Circle Law

[Marchenko and Pastur, 1965]

- Let X be $m \times n$ random matrix.
- X_{ij} zero mean, unit variance.
- Look at sample covariance matrix

$$S_n = \frac{1}{n} X X^T$$

- its eigenvalues $\ell_1, \ldots, \ell_m$.

For fixed dimension m, as $n \to \infty$,

$$S_n \to \Sigma = I_m \quad \to \quad \ell_i \to 1$$

The Quarter Circle Law

[Marchenko and Pastur, 1965]

- Let X be $m \times n$ random matrix.
- X_{ij} zero mean, unit variance.
- Look at sample covariance matrix

$$S_n = \frac{1}{n} X X^T$$

- its eigenvalues $\ell_1, \ldots, \ell_m$.

For fixed dimension m, as $n \to \infty$,

$$S_n \to \Sigma = I_m \quad \to \quad \ell_i \to 1$$

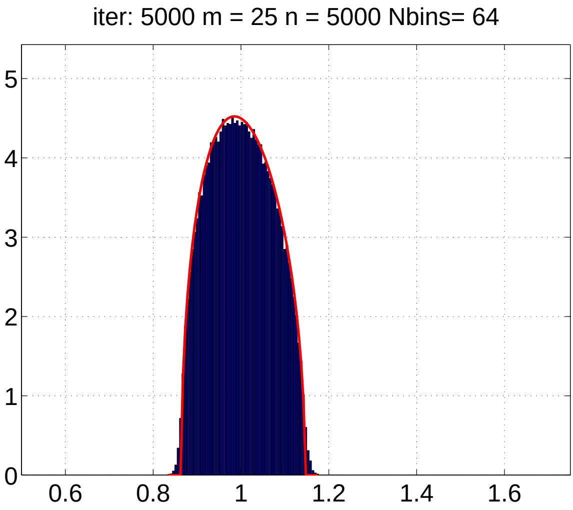
Question: *Eigenvalue spread as* $m, n \to \infty$ *?*

Simulation: Eigenvalue Spread

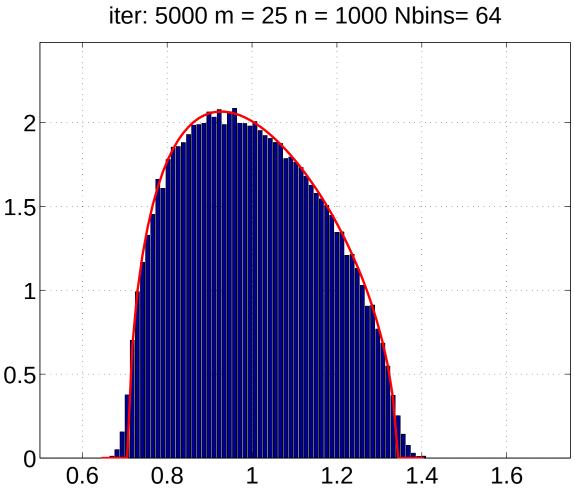
Example:

```
X = randn(m,n);
S = 1/n X X' ;
L = eig(S) ;
```

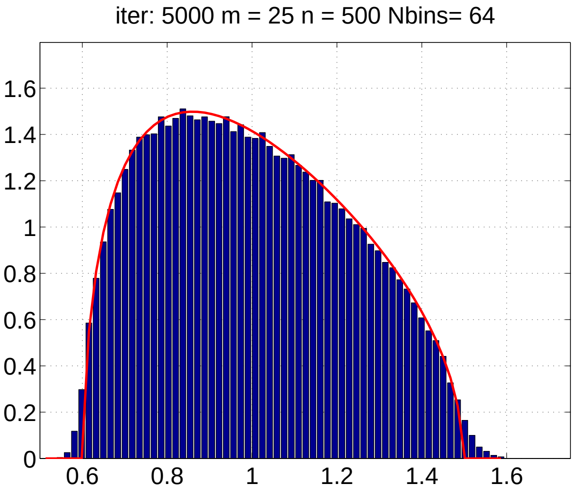
Simulation: Eigenvalue Spread



Simulation: Eigenvalue Spread



Simulation: Eigenvalue Spread



Spread of Sample Eigenvalues

Let $\{\ell_i\}_{i=1}^m$ be the eigenvalues of a random symmetric matrix H.

Empirical Spectral Distribution Function:

$$F_m(t) = \frac{1}{m}\#\{\ell_i \leq t\}$$

The Quarter-Circle Law

[Marchenko & Pastur, 1967]

Let $H \sim W_m(n, \Sigma)$.

Theorem: For $\Sigma = I$, as $m, n \to \infty$ with $m/n \to c$, $(c < 1)$ let ℓ_i be sample eigenvalues of H/m, then

$$f_{MP}(t) = \frac{1}{2\pi ct}\sqrt{(b-t)(t-a)} \quad t \in [a, b]$$

where $a = (1 - \sqrt{c})^2$, $b = (1 + \sqrt{c})^2$

The Quarter-Circle Law

[Marchenko & Pastur, 1967]

Let $H \sim W_m(n, \Sigma)$.

Theorem: For $\Sigma = I$, as $m, n \to \infty$ with $m/n \to c$, $(c < 1)$ let ℓ_i be sample eigenvalues of H/m, then

$$f_{MP}(t) = \frac{1}{2\pi ct}\sqrt{(b-t)(t-a)} \quad t \in [a, b]$$

where $a = (1-\sqrt{c})^2$, $b = (1+\sqrt{c})^2$

If $c > 1$, then $a = 0$, and there are $m - n$ sample eigenvalues exactly at zero.

Extensions

Non-identity Covariance Matrices: $X \sim W_m(n, \Sigma_m)$ where Σ_m eigenvalue density converges to limiting distribution.

Extensions

Non-identity Covariance Matrices: $X \sim W_m(n, \Sigma_m)$ where Σ_m eigenvalue density converges to limiting distribution.

Then, as $n, m \to \infty$, empirical density of $1/nXX'$ converges to limit. Can describe its Stieltjes transform in terms of that of limiting Σ.

Extensions

Non-identity Covariance Matrices: $X \sim W_m(n, \Sigma_m)$ where Σ_m eigenvalue density converges to limiting distribution.

Then, as $n, m \to \infty$, empirical density of $1/nXX'$ converges to limit. Can describe its Stieltjes transform in terms of that of limiting Σ.

Double Wishart Matrices:

Let $A \sim W_m(n_1, \Sigma)$ and let $B \sim W_m(n_2, \Sigma)$.
As $m, n_1, n_2 \to \infty$, eigenvalue distribution of $A^{-1}B$ converges to limiting Marchenko-Pastur type distribution.

Remark: Eigenvalues of $A^{-1}B$ same as those of $(\Sigma^{-1}A)^{-1}(\Sigma^{-1}B)$, so result independent of Σ.
Can choose $\Sigma = I$.

Largest Eigenvalue of Random Matrices

Let $X_{m \times n}$ be random $X_{ij} \sim \mathcal{N}(0,1)$.

Marchenko-Pastur law: limiting spectral density of eigenvalues of $1/nXX'$.

Largest Eigenvalue of Random Matrices

Let $X_{m\times n}$ be random $X_{ij} \sim \mathcal{N}(0,1)$.

Marchenko-Pastur law: limiting spectral density of eigenvalues of $1/nXX'$.

However, how large can the largest eigenvalue be ?

Largest Eigenvalue of Random Matrices

Let $X_{m\times n}$ be random $X_{ij} \sim \mathcal{N}(0,1)$.

Marchenko-Pastur law: limiting spectral density of eigenvalues of $1/nXX'$.

However, how large can the largest eigenvalue be ?

$$\text{Question: Bound on } \ell_1 = \|1/mXX^T\|_2 \text{ ?}$$

Largest Eigenvalue of Random Matrices

Let $X_{m \times n}$ be random $X_{ij} \sim \mathcal{N}(0, 1)$.

Marchenko-Pastur law: limiting spectral density of eigenvalues of $1/nXX'$.

However, how large can the largest eigenvalue be ?

$$\text{Question: Bound on } \ell_1 = \|1/mXX^T\|_2 \text{ ?}$$
$$\text{Is it true that } \ell_1 \to (1 + \sqrt{m/n})^2 \text{ ?}$$

Largest Eigenvalue of Random Matrices

[Wachter, 1978]

$\ell_1 \to (1+\sqrt{m/n})^2$ with probability 1

Largest Eigenvalue of Random Matrices

[Wachter, 1978]

$\ell_1 \to (1 + \sqrt{m/n})^2$ with probability 1

[Gordon 1985]

non-asymptotic concentration of measure bounds for largest eigenvalue of random matrices.

many follow-up works on this topic ...

[complex case, Johansson, 00']
[real case, Johnstone, 01']

Single Wishart Case:

$$\Pr\left[\ell_1 < \mu_{nm} + \sigma_{nm} s\right] \to F_\beta(s)$$

where F_β is the Tracy-Widom distribution of order β.

Universality
under mild moment conditions, largest eigenvalue converges to TW distribution

[Soshnikov, Tao & Vu, etc]

Tracy-Widom Distribution

[Tracy & Widom, 1994, 1996]

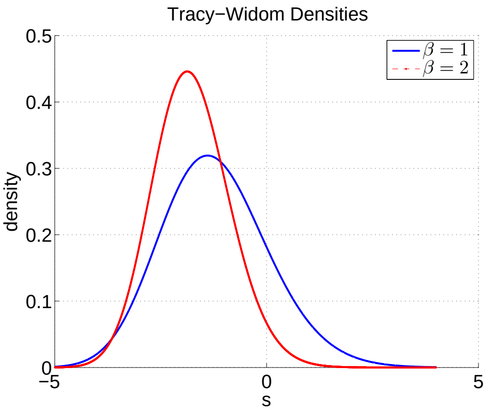
$$
\begin{aligned}
F_2(s) &= \exp\left(-\int_s^\infty (x-s)^2 q(x)dx\right) \\
F_1(s)^2 &= F_2(s)\exp\left(-\int_s^\infty q(x)dx\right)
\end{aligned}
$$

where $q(x)$ is the solution to Painleve-II non-linear 2nd order differential equation

$$
q'' = sq + 2q^3, \quad q(s) \sim Ai(s) \text{ as } s \to \infty
$$

q and F_β are somewhat tricky to compute. Several packages. Folkmar Bornemann (by request) - matlab code.

Tracy-Widom Distributions



Tracy-Widom Distributions

Large x asymptotics:

$$F_2(x) = 1 - \frac{e^{-\frac{4}{3}x^{3/2}}}{16\pi x^{3/2}}\left(1 + O(x^{-3/2}\right)$$

$$F_1(x) = 1 - \frac{e^{-\frac{2}{3}x^{3/2}}}{4\sqrt{\pi}x^{3/2}}\left(1 + O(x^{-3/2}\right)$$

Second Order Accuracy

[complex case, El-Karoui, 07]
[real case, Ma, Johnstone and Ma]

For Wishart matrices,

with careful choice of centering and scaling parameters, as $n, m \to \infty$ with $n/m \to \gamma$,

$$|\Pr[\ell_1 < \mu_{nm} + \sigma_{nm} s] - F_\beta(s)| \le Ce^{-cs} m^{-2/3}$$

Second Order Accuracy

[complex case, El-Karoui, 07]
[real case, Ma, Johnstone and Ma]

For Wishart matrices,

with careful choice of centering and scaling parameters, as $n, m \to \infty$ with $n/m \to \gamma$,

$$|\Pr[\ell_1 < \mu_{nm} + \sigma_{nm} s] - F_\beta(s)| \leq Ce^{-cs} m^{-2/3}$$

For real valued data

$$\mu_{nm} = \left(\sqrt{n - \frac{1}{2}} + \sqrt{m - \frac{1}{2}} \right)^2$$

$$\sigma_{nm} = \sqrt{\mu_{nm}} \left(\frac{1}{\sqrt{n - \frac{1}{2}}} + \frac{1}{\sqrt{m - \frac{1}{2}}} \right)^{1/3}$$

Largest Eigenvalue Fluctuations

for ℓ_1/m fluctuations are $O(1/m^{2/3})$, much *smaller* than standard $1/\sqrt{m}$ fluctuations for averages of many random variables.

TW is a new "universal" distribution - attractor in terms of largest eigenvalue due to universality.

Distribution of the Largest Eigenvalue Divided By Trace

In various settings, interest is in following ratio

$$U = \frac{\ell_1}{\frac{1}{p}\sum_{j=1}^{p} \ell_j} > \textit{threshold}$$

Distribution of the Largest Eigenvalue Divided By Trace

In various settings, interest is in following ratio

$$U = \frac{\ell_1}{\frac{1}{p}\sum_{j=1}^{p} \ell_j} > \textit{threshold}$$

This random variable plays a role in several different applications:

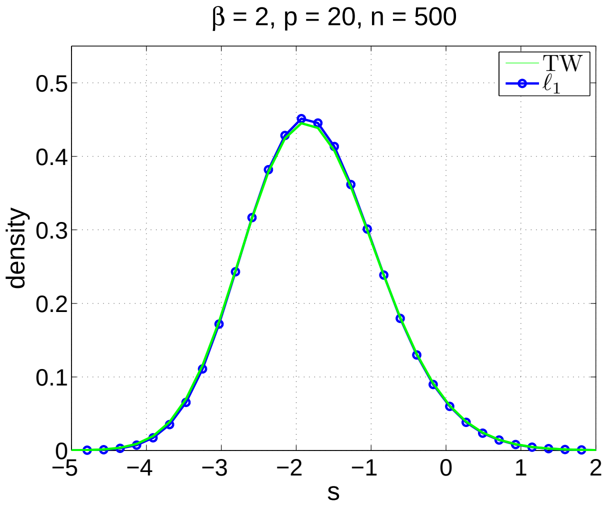
- Signal Detection [Besson & Scharf 06', Kritchman & N. 08, Bianchi et al. 09']
- Two-way models of interaction [Johnson & Graybill, 72']
- Models for Quantum Information Channels.

Ratio of Largest Eigenvalue to Trace

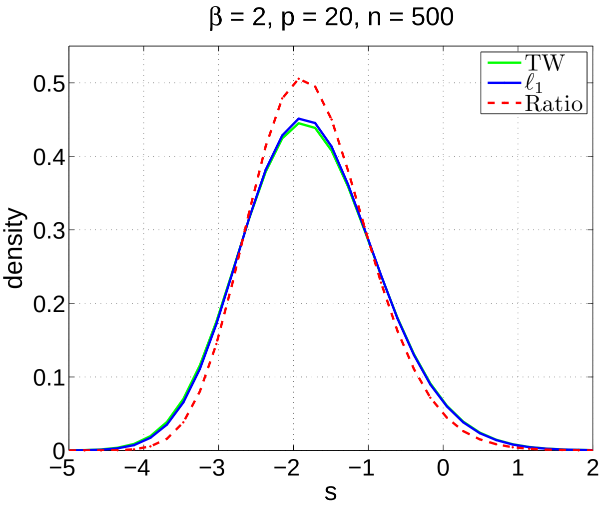
In principle, as $p, n \to \infty$, [Nechita 08', Bianchi et al. 09']

$$\Pr\left[\frac{U - \mu_{n,p}}{\sigma_{n,p}} < s\right] \to TW_\beta(s)$$

Ratio Distribution



Ratio Distribution



Ratio of Largest Eigenvalue to Trace

Theorem: As, $p, n \to \infty$,

$$\Pr\left[\frac{U - \mu_{n,p}}{\sigma_{n,p}} < s\right] \approx F_\beta(s) - \frac{1}{\beta np}\left(\frac{\mu_{n,p}}{\sigma_{n,p}}\right)^2 F''_\beta(s)$$

[N., J. Mult. Anal., 2011]

Ratio of Largest Eigenvalue to Trace

Theorem: As, $p, n \to \infty$,

$$\Pr\left[\frac{U - \mu_{n,p}}{\sigma_{n,p}} < s\right] \approx F_\beta(s) - \frac{1}{\beta np}\left(\frac{\mu_{n,p}}{\sigma_{n,p}}\right)^2 F''_\beta(s)$$

[N., J. Mult. Anal., 2011]

This distribution also relevant to study performance of various detection methods

[N. IEEE Tran. Sig. Proc. 10']

[N. Penna and Garello Int. Conf. Comm. 11']

Part III:
Signal Bearing Matrices

Spiked Covariance Models

Consider model whereby

$$\Sigma = diag(\lambda_1, \lambda_2, \ldots, \lambda_k, 0, \ldots, 0) + \sigma^2 I_m$$

Spiked covariance with k spikes.

Observe n vectors $\mathbf{x}_i \in \{\mathbb{R}, \mathbb{C}\}^m$ from this model.

Question: What happens to largest sample eigenvalues and eigenvectors as $n, m \to \infty$, with k, λ_j fixed ?

Phase Transition

[complex case, Ben-Arous, Baik, Peche]
[real case, Baik and Silverstein]

Theorem: For spike model with k spikes, as $n, m \to \infty$ with $m/n \to c$, for $j = 1, \ldots, k$,

$$\ell_j \to \begin{cases} (\lambda_j + \sigma^2)\left(1 + \frac{m-k}{n}\frac{\sigma^2}{\lambda_j}\right) & \lambda_j > \sigma^2\sqrt{m/n} \\ \sigma^2(1 + \sqrt{m/n})^2 & \lambda_j < \sigma^2\sqrt{m/n} \end{cases}$$

Phenomena known as *retarded learning* in statistical physics.

[D. Paul 07', Nadler 08']

Theorem: As $m, n \to \infty$

$$R^2(m/n) = |\langle \mathbf{v}_{\text{PCA}}, \mathbf{v} \rangle|^2 = \begin{cases} 0 & \text{if } \lambda < \sigma^2 \sqrt{m/n} \\ \dfrac{\frac{n\lambda^2}{p\sigma^4} - 1}{\frac{n\lambda^2}{p\sigma^4} + \frac{\lambda}{\sigma^2}} & \text{if } \lambda > \sigma^2 \sqrt{m/n} \end{cases}$$

In statistical physics:

[Hoyle and Rattray, Reimann & al, Biehl, Watson]

Phase Transition / Eigenvectors

[D. Paul 07', Nadler 08']

Theorem: As $m, n \to \infty$

$$R^2(m/n) = |\langle \mathbf{v}_{\text{PCA}}, \mathbf{v} \rangle|^2 = \begin{cases} 0 & \text{if } \lambda < \sigma^2\sqrt{m/n} \\ \dfrac{\frac{n\lambda^2}{p\sigma^4} - 1}{\frac{n\lambda^2}{p\sigma^4} + \frac{\lambda}{\sigma^2}} & \text{if } \lambda > \sigma^2\sqrt{m/n} \end{cases}$$

In statistical physics:

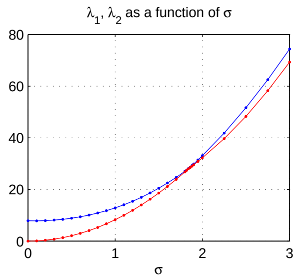
[Hoyle and Rattray, Reimann & al, Biehl, Watson]

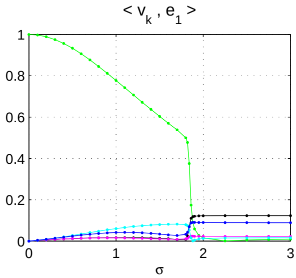
Asymptotic $\sqrt{n}$-Gaussian fluctuations [Paul, 07]

$$\sqrt{n}(\ell_1 - \mathbb{E}[\ell_1]) \sim \mathcal{N}(0, \sigma^2(\lambda_1))$$

Phase Transition for finite p as function of σ

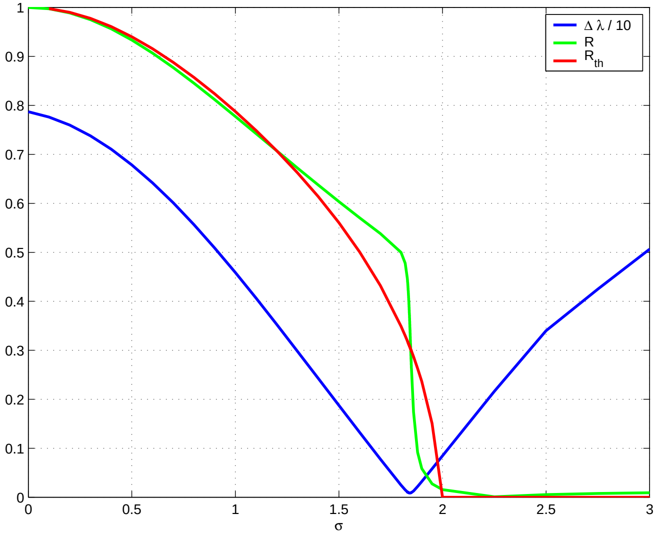
First, a "thought experiment": Take training set $\{\mathbf{x}^\nu\}$ with finite m, n and start increasing σ. What should be the expected behavior of $R = \langle \mathbf{v}_{\text{PCA}}, \mathbf{v} \rangle$ and of ℓ_1 ?





$$\lambda \sim \kappa^2 + \sigma^2(1 + m/n) \qquad\qquad R \sim 1 - \sigma^2/\kappa^2 m/n$$
$$n = 50, m = 200, \kappa^2 = 7.87$$

Phase Transition as function of σ



Tools in Random Matrix Theory

Moment Methods

Stieltjes Transform

Concentration of Measure

Many others...

Part IV:

Signal Detection / Cognitive Radio Application

Signal Detection Model

$\{\mathbf{x}_i\}_{i=1}^n$ are n i.i.d. observations from

$$\mathbf{x} = \sqrt{\lambda} s(t)\mathbf{h} + \sigma \xi(t)$$

Signal Detection as Hypothesis Testing

Given $\{\mathbf{x}_i\}$ determine between

$$\mathcal{H}_0 : \text{no signal, } \lambda = 0 \quad vs. \mathcal{H}_1 : \text{signal present } \lambda > 0$$

Signal Detection as Hypothesis Testing

Given $\{\mathbf{x}_i\}$ determine between

$$\mathcal{H}_0 : \text{no signal, } \lambda = 0 \quad vs. \mathcal{H}_1 : \text{signal present } \lambda > 0$$

$\mathcal{H}_0 =$ *null hypothesis.*

$\mathcal{H}_1 =$ *alternative*

Test Statistics

Test statistic: A function $f(\mathbf{x}_1, \ldots, \mathbf{x}_n) \to \mathbb{R}$ such that

$$f(\mathbf{x}_1, \ldots, \mathbf{x}_n) \leq t \quad \to \quad \text{accept } \mathcal{H}_0$$

Test statistic: A function $f(\mathbf{x}_1, \ldots, \mathbf{x}_n) \to \mathbb{R}$ such that

$$f(\mathbf{x}_1, \ldots, \mathbf{x}_n) \leq t \quad \to \quad \text{accept } \mathcal{H}_0$$

$$f(\mathbf{x}_1, \ldots, \mathbf{x}_n) > t \quad \to \quad \text{reject } \mathcal{H}_0$$

False Alarm and Power

Two common quantities of interest:

False Alarm and Power

Two common quantities of interest:

False Alarm Rate:

$$P_{FA} = \Pr[f(\mathbf{x}_1, \ldots, \mathbf{x}_n) > t \,|\, \mathcal{H}_0 \text{ true}]$$

False Alarm and Power

Two common quantities of interest:

False Alarm Rate:

$$P_{FA} = \Pr[f(\mathbf{x}_1, \ldots, \mathbf{x}_n) > t \,|\, \mathcal{H}_0 \text{ true}]$$

Detection Power:

$$P_D = \Pr[f(\mathbf{x}_1, \ldots, \mathbf{x}_n) > t \,|\, \mathcal{H}_1 \text{ true}]$$

False Alarm and Power

Two common quantities of interest:

False Alarm Rate:

$$P_{FA} = \Pr[f(\mathbf{x}_1, \ldots, \mathbf{x}_n) > t \,|\, \mathcal{H}_0 \text{ true}]$$

Detection Power:

$$P_D = \Pr[f(\mathbf{x}_1, \ldots, \mathbf{x}_n) > t \,|\, \mathcal{H}_1 \text{ true}]$$

Typically, choose threshold $t = t(\alpha)$ such that $P_{FA} \approx \alpha$ for some $\alpha \ll 1$.

Optimal Tests

A test is optimal at level α, if

$$P_D[f] \geq P_D[\text{any other test}]$$

with same false alarm rate α.

Optimal Tests

A test is optimal at level α, if

$$P_D[f] \geq P_D[\text{any other test}]$$

with same false alarm rate α.

Neyman-Pearson Lemma: If both $\mathcal{H}_0$ and $\mathcal{H}_1$ are simple hypothesis, then optimal test statistic is *likelihood ratio test*

$$\frac{p(\mathbf{x}_1, \ldots, \mathbf{x}_n \,|\mathcal{H}_1)}{p(\mathbf{x}_1, \ldots, \mathbf{x}_n \,|\mathcal{H}_0)}$$

If σ is known, then $\mathcal{H}_0$ is simple.

However, $\mathcal{H}_1$ is *not simple*.

Neyman Pearson Lemma not directly applicable

Test Statistics for Signal Detection / Cognitive Radio

If σ is known, then $\mathcal{H}_0$ is simple.

However, $\mathcal{H}_1$ is *not simple*.

Neyman Pearson Lemma not directly applicable

In practice: Many test statistics were proposed, all based on the eigenvalues of the sample covariance matrix S_n.

Energy Detection:	$Tr(S_n)$	known σ
Largest Eigenvalue:	ℓ_1	known σ
Max/Min Ratio:	ℓ_1/ℓ_p	unknown σ
GLRT:	$\ell_1/Tr(S_n)$	unknown σ
etc. etc.		

Optimal Test for Signal Detection ?

Given so many tests, which one to choose ?

Mathematical Principle: Look at simplest possible setting.

Optimal Test for Signal Detection ?

Given so many tests, which one to choose ?

Mathematical Principle: Look at simplest possible setting.

σ - noise level known.

Optimal Test for Signal Detection ?

Given so many tests, which one to choose ?

Mathematical Principle: Look at simplest possible setting.

σ - noise level known.

λ - signal strength, known.

Optimal Test for Signal Detection ?

Given so many tests, which one to choose ?

Mathematical Principle: Look at simplest possible setting.

σ - noise level known.

λ - signal strength, known.

$$\mathcal{H}_0 \text{ is simple}$$
$$\mathcal{H}_1 \text{ is still not simple}$$

Optimal Test for Signal Detection ?

Given so many tests, which one to choose ?

Mathematical Principle: Look at simplest possible setting.

σ - noise level known.

λ - signal strength, known.

$\mathcal{H}_0$ is simple

$\mathcal{H}_1$ is still not simple

but

if we consider test statistics based

only

on sample eigenvalues then it is !

Which Test Statistic to use ?

Consider the case of two (nearly) *simple* hypothesis

$$\mathcal{H}_0 : \mathbf{\Sigma} = \mathbf{I} \quad \textit{vs.} \quad \mathcal{H}_1 : \mathbf{W}'\mathbf{\Sigma}\mathbf{W} = \mathbf{I} + \textit{diag}(\lambda, 0, \ldots, 0)$$

with λ - known. What is unknown is the basis which makes $\mathbf{\Sigma}$ diagonal in $\mathcal{H}_1$.
Suppose we use only eigenvalues $\{\ell_j\}$ of H as a test statistic.

Which Test Statistic to use ?

Consider the case of two (nearly) *simple* hypothesis

$$\mathcal{H}_0 : \mathbf{\Sigma} = \mathbf{I} \quad vs. \quad \mathcal{H}_1 : \mathbf{W}'\mathbf{\Sigma}\mathbf{W} = \mathbf{I} + diag(\lambda, 0, \ldots, 0)$$

with λ - known. What is unknown is the basis which makes $\mathbf{\Sigma}$ diagonal in $\mathcal{H}_1$.
Suppose we use only eigenvalues $\{\ell_j\}$ of H as a test statistic.

Neyman-Pearson: optimal method is likelihood ratio test

$$\frac{p(\ell_1, \ldots, \ell_m | \mathcal{H}_1)}{p(\ell_1, \ldots, \ell_m | \mathcal{H}_0)} \gtrless C(\alpha)$$

Which Test Statistic ?

From multivariate analysis (Muirhead 78')

$$p(\ell_1,\ldots,\ell_m|\Sigma) = C_{n,m}\prod \ell_i^{(n-m-1)/2}\prod_{i<j}(\ell_i-\ell_j)\ {}_0F_0(-\frac{1}{2}nL,\Sigma^{-1})$$

${}_0F_0$ - hypergeometric function with matrix argument.
Key point: asymptotically in sample size n, for dimension m fixed,

$$\log\left(\frac{p(\ell_1,\ldots,\ell_m|\mathcal{H}_1)}{p(\ell_1,\ldots,\ell_m|\mathcal{H}_0)}\right) \approx n(\ell_1-h(\lambda)) + O(\sum c_{1j}/(\ell_1-\ell_j))$$

Asymptotically, as $n\to\infty$, should only look at *largest* eigenvalue !

Which Test Statistic ?

From multivariate analysis (Muirhead 78')

$$p(\ell_1, \ldots, \ell_m | \Sigma) = C_{n,m} \prod \ell_i^{(n-m-1)/2} \prod_{i<j} (\ell_i - \ell_j)\ {}_0F_0(-\frac{1}{2} nL, \Sigma^{-1})$$

${}_0F_0$ - hypergeometric function with matrix argument.
Key point: asymptotically in sample size n, for dimension m fixed,

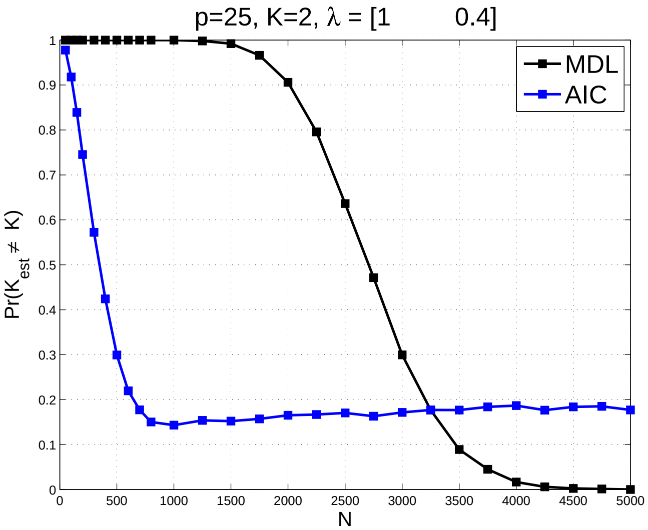
$$\log \left(\frac{p(\ell_1, \ldots, \ell_m | \mathcal{H}_1)}{p(\ell_1, \ldots, \ell_m | \mathcal{H}_0)} \right) \approx n(\ell_1 - h(\lambda)) + O(\sum c_{1j}/(\ell_1 - \ell_j))$$

Asymptotically, as $n \to \infty$, should only look at *largest* eigenvalue !

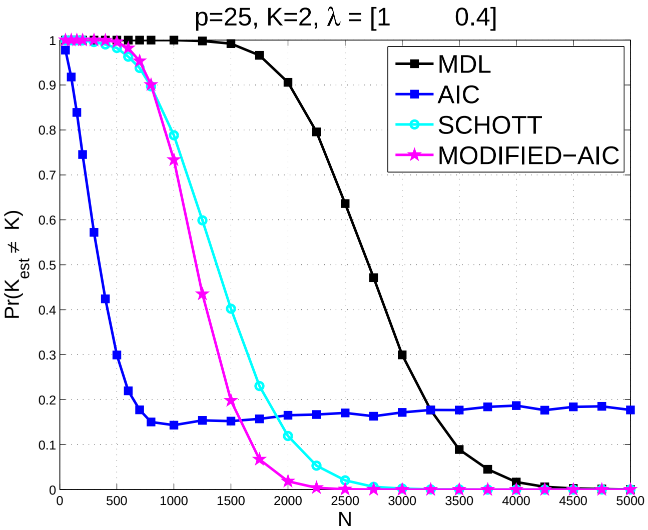
Roy's Largest Root Test

[Kritchman & N., IEEE-TSP, 09']

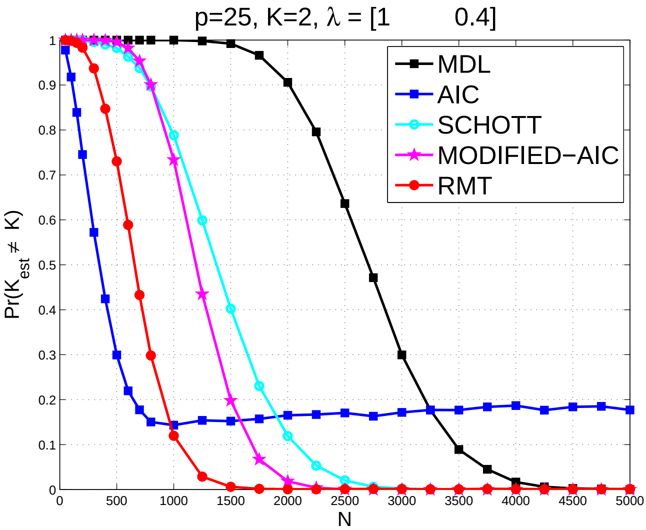
Using Roy's Largest Root Test



Using Roy's Largest Root Test



Using Roy's Largest Root Test



Properties of Roy's Test

Detection Power of Roy's test

via Random Matrix Theory

Properties of Roy's Test

Detection Power of Roy's test

via Random Matrix Theory

Via largest root test, asymptotically for m, n large only
$\lambda/\sigma^2 > \sqrt{m/n}$ can be detected (with probability one) !

Via other tests, can detect (but not with probability one) weaker signals

[Onatsky et. al.]

Final word of caution: Roy's largest root test is asymptotically optimal when $n \to \infty$. Not so for very weak signals.

What we did not cover

A lot !

What we did not cover

A lot !

- Free Probability (Voiculescu)
- Determinantal Processes
- Kernel random matrices
- Random matrices with heavy tailed distributions
- Random matrices with dependent (correlated) entries
- Random Graphs
- Other transforms
- Concentration of Measure (non-asymptotic finite sample bounds)
- relation to statistical physics
- Rate of convergence
- Linear spectral statistics, Central Limit Theorems - etc etc.

Some (Recent) References

much more material can be found in:

Multivariate Statistics:

- A. T. James, Distributions of matrix variates and latent roots derived from normal samples, *Ann. Math. Statist.*, vol. 35, 475-501, 1964.
- T.W. Anderson, *An introduction to multivariate statistical analysis*, Wiley, 2003.
- R.J. Muirhead, *Aspects of Multivariate Statistical Theory*, 2005.

Random Matrices / Recent Books:

- G. Anderson, A. Guionnet, O. Zeitouni, *An Introduction to Random Matrices*, Cambridge, 2009.
- A. Tulino, S. Verdu, *Random Matrix Methods and Wireless Communications*, 2011.
- Z. D. Bai, J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, Springer, 2009.

Some (Recent) References

Largest Eigenvalue Distribution:

- I.M. Johnstone, On the distribution of the largest eigenvalue in principal component analysis, *Ann. Stat.*, **29**:295–327, 2001.
- I.M. Johnstone, Approximate Null Distribution of the Largest Root in Multivariate Analysis, *Ann. Applied Stat*, 2009
- N. El Karoui. A rate of convergence result for the largest eigenvalue of complex white Wishart matrices. *Annals of Probability*, 2006.
- B. Nadler, On the distribution of the ratio of the largest eigenvalue to the trace of a Wishart matrix, *J. Mult. Anal.*, 2010.

Some (Recent) References

Largest Eigenvalue Signal + Noise:

- J. Baik, G. Ben-Arous, S. Peche, Phase transition of the largest eigenvalue for non-null complex sample covariance matrices, *Ann. Prob.*, 2005.

- J. Baik, J. W. Silverstein, Eigenvalues of large sample covariance matrices of spiked population models, *J. Mult. Anal.*, 2006.

- D. Paul, Asymptotics of sample eigenstructure for a large dimensional spiked covariance model, *Stat. Sinica*, 2007.

- B. Nadler, Finite sample approximation results for PCA, *Ann. Stat.*, 2008.

- R.R. Nadakuditi and J. W. Silverstein. Fundamental Limit of Sample Generalized Eigenvalue Based Detection of Signals in Noise Using Relatively Few Signal-Bearing and Noise-Only Samples. IEEE Journal of Selected Topics in Signal Processing, 2010.

- I. M. Johnstone, B. Nadler, Tech. Report, Stanford University, 2011.

The End

Obrigado pela sua atencao !

more material at

www.wisdom.weizmann.ac.il/$\sim$nadler