Largest eigenvalues and eigenvectors in multivariate analysis

Iain Johnstone, Statistics, Stanford

imj@stanford.edu

Yale, May, 2009
What’s this?
Eigenvalues: Theme

▲ Many problems of classical multivariate statistics involve eigenvalues $x_1 > x_2 > \ldots > x_p$

▲ Asymptotics in $p \Rightarrow$ useful information, [even $p$ small]

▲ Illustrate for null distributions for $x_1$
Many problems of classical multivariate statistics involve eigenvalues $x_1 > x_2 > \ldots > x_p$

Asymptotics in $p \Rightarrow$ useful information, [espec. $p$ small]

Illustrate for null distributions for $x_1$

\[ y = X \beta + \epsilon \quad \epsilon \sim N(0, I_n \otimes \Sigma_p). \]

Tests of $H_0 : \beta = 0$ use roots of \[ \det[A + x_i(A + B)] = 0. \]

$A = SS_H = \hat{\beta}^T X^T X \hat{\beta}$

$B = SS_E = y^T (I - P_X)y$

“Hypothesis”

“Error”
Double Wishart Setting

\[ A \sim W_p(n_1, I) \]
\[ B \sim W_p(n_2, I) \]

2 independent Wisharts, \( p \leq n_1, n_2 \)

“null hypothesis” setting

Common feature: \( \text{roots} := (x_i)_{i=1}^p \) of generalized eigenproblem:

\[ \det[x(A + B) - A] = 0 \]

Single Wishart

- Principal Component analysis
- Factor analysis
- Multidimensional scaling

Double Wishart

- Canonical correlation analysis
- Multivariate Analysis of Variance (MANOVA)
- Multivariate regression analysis
- Discriminant analysis
- Tests of equality of covariance matrices
Joint density of eigenvalues

Single Wishart: \( \det[A - x_i I] = 0. \)

Double Wishart: \( \det[A - x_i (A + B)] = 0. \)

In each case, (Fisher, Girshick, Hsu, Mood, Roy, (1939)):

\[
f(x_1, \ldots, x_p) = c \prod_{i} w^{1/2}(x_i) \prod_{i<j} (x_i - x_j) \quad x_1 \geq \ldots \geq x_p
\]

Single Wishart: \( w(x) = x^{n-p} e^{-x}, \quad \text{(Laguerre)} \)

Double Wishart: \( w(x) = x^{p-q-1}(1 - x)^{n-p-q-1}. \quad \text{(Jacobi)} \)
I. Setting
   1. Single & Double Wishart, examples

II. Largest Eigenvalue: Limiting Law ($H_0$)
   1. Gaussian
   2. Laguerre/Single W.
   3. Jacobi/Double W.

III. Largest Eigenvalue: Concentration
   1. Single W.
   2. Double W.

IV. Largest Eigenvector(s)
Symmetric Gaussian matrices

\( A \) – symmetric \( N \times N \) matrix drawn from (\( \beta = 1 \) for GOE)

\[
f(A) = c \exp \left\{ -\frac{\beta}{2} \text{tr}(A^T A) \right\}.
\]

[for GUE: complex Hermitian]. Largest eigenvalue:

\[
\theta_N = \lambda_{max}(A)
\]

Centering and scaling constants:

\[
\mu_N = \sqrt{2N}, \quad \sigma_N = 2^{-1/2} N^{-1/6}
\]

Tracy-Widom limit: TW (1994, 1996) For \( \beta = 1, 2, 4 \) (\( \mathbb{R}, \mathbb{C}, \mathbb{Q} \)),

\[
\frac{\theta_N - \mu_N}{\sigma_N} \overset{D}{\Rightarrow} F_\beta.
\]
Tracy-Widom distributions (1994,96)

\[ q'' = sq + 2q^3 \quad \text{(Painlevé II)} \]
\[ q(s) \sim \text{Ai}(s) \text{ as } s \to \infty \]

\[ F_2(s) = e^{-\int_s^\infty (x-s)^2 q(x)dx} \]
\[ F_1(s)^2 = F_2(s)e^{-\int_s^\infty q(x)dx}. \]

**Right tail decay:**
\[ 1 - F_\beta(s) \approx e^{-(2/3) s^{3/2}}. \]

**Rate of convergence:** First order \( N^{-1/3} \), Choup (2006,07)

**“Second-order”**: J + Ma (2008) For \( \beta = 1, 2, \) & \( \mu_N \overset{(\beta=1)}{=} \sqrt{2N-1}, \)

\[ |P\{\lambda_{max}(A) \leq \mu_N + \sigma_N s\} - F_\beta(s)| \leq CN^{-2/3}e^{-s/2}. \]
Better approximation in right tail: location of turning point of Airy function

additional $O(N^{-4/3})$ mean correction included
Approximations at $N = 10$
Outline

I. Setting
   1. Single & Double Wishart, examples

II. Largest Eigenvalue: Limiting Law ($H_0$)
   1. Gaussian
   2. Laguerre/Single W.
   3. Jacobi/Double W.

III. Largest Eigenvalue: Concentration
   1. Single W.
   2. Double W.

IV. Largest Eigenvector(s)
For \{\text{real, complex}\}, \{\text{single, double}\} Wishart matrices, \(\frac{n}{p} \to \gamma\), \(\text{or} \ (n_1/p, n_2/p) \to (\gamma_1, \gamma_2)\), then

\[
P\{n\hat{\ell}_1 \leq \mu_{np} + \sigma_{np}s|H_0\} \to F_\beta(s)
\]

— “universality” in RMT (Deift, 06 ICM)

**Single Wishart:** \(\beta = 2\) (Johansson, 00), \(\beta = 1\) (J, 01)

“Microarray case” \(n, p \to \infty, \ n/p \to 0, \infty\). (El Karoui, 03)

**Non Gaussian:** (Soshnikov, 02, 06; S. + Fyodorov, 06, Péché, 07)

\(\uparrow\) subGaussian: \(\Rightarrow\) TW limit; heavy tails \(\Rightarrow\) Poisson
Second order accuracy

For \{real, complex\}, \{single, double\} Wishart matrices, if \(n/p \to \gamma\), [or \((n_1/p, n_2/p) \to (\gamma_1, \gamma_2)\),] then

\[
|P\{n\hat{\ell}_1 \leq \mu_{np} + \sigma_{np}s|H_0\} - F_\beta(s)| \leq Ce^{-cs}p^{-2/3}.
\]

**Single Wishart**  Complex: El Karoui (2006).

**Real:** Ma (2008, poster here)

\[
\mu_{np} = \left(\sqrt{n-\frac{1}{2}} + \sqrt{p-\frac{1}{2}}\right)^2
\]
\[
\sigma_{np} = \left(\sqrt{n-\frac{1}{2}} + \sqrt{p-\frac{1}{2}}\right)^1 \left(\frac{1}{\sqrt{n-\frac{1}{2}}} + \frac{1}{\sqrt{p-\frac{1}{2}}}\right)^{1/3}.
\]
Beyond the “Null Hypothesis”

Classical RMT ensembles (e.g. $W_p(n, I)$)

$\leftrightarrow$ “null hypothesis”, symmetry, no structure.

For $W_p(n, \Sigma)$: For what conditions on $\Sigma$ does

$$P\{\hat{\ell}_1 \leq \mu_{np}(\Sigma) + \sigma_{np}(\Sigma)s\} \rightarrow F_\beta(s)$$

Some answers:

▲ sufficiently many $\ell_k(\Sigma)$ accumulate near $\ell_1(\Sigma)$ (data in $\mathbb{C}$)

El Karoui, 2007

▲ small number of (not too!) isolated $\ell_i(\Sigma)$


Harding (2008, Economics Letters)
Outline

I. Setting
   1. Single & Double Wishart, examples

II. Largest Eigenvalue: Limiting Law ($H_0$)
   1. Gaussian
   2. Laguerre/Single W.
   3. Jacobi/Double W.

III. Largest Eigenvalue: Concentration
   1. Single W.
   2. Double W.

IV. Largest Eigenvector(s)
Roy’s Greatest Root Test: Package Output

SAS: (From Gledhill et. al.: NOAA Fisheries Reef Fish Video Surveys.)

Table 18. Multivariate statistics and F approximations for the CANDISC procedure ran with all mincount and habitat data.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
<th>F Value</th>
<th>Num DF</th>
<th>Den DF</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilks' Lambda</td>
<td>0.58109196</td>
<td>1.15</td>
<td>50</td>
<td>454.87</td>
<td>0.2325</td>
</tr>
<tr>
<td>Pillai's Trace</td>
<td>0.50115530</td>
<td>1.15</td>
<td>50</td>
<td>515</td>
<td>0.2342</td>
</tr>
<tr>
<td>Hotelling-Lawley Trace</td>
<td>0.59036501</td>
<td>1.15</td>
<td>50</td>
<td>305.82</td>
<td>0.2355</td>
</tr>
<tr>
<td>Roy's Greatest Root</td>
<td>0.26759820</td>
<td>2.76</td>
<td>10</td>
<td>103</td>
<td>0.0047</td>
</tr>
</tbody>
</table>

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

R: (car, heplots packages, Fox et. al.)

Multivariate Tests:

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Df</th>
<th>test stat</th>
<th>approx F</th>
<th>num Df</th>
<th>den Df</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pillai</td>
<td>5.0000</td>
<td>0.417938</td>
<td>1.845226</td>
<td>15.0000</td>
<td>171.0000</td>
<td>0.0320861</td>
</tr>
<tr>
<td>Wilks</td>
<td>5.0000</td>
<td>0.623582</td>
<td>1.893613</td>
<td>15.0000</td>
<td>152.2322</td>
<td>0.0276949</td>
</tr>
<tr>
<td>Hotelling-Lawley</td>
<td>5.0000</td>
<td>0.538651</td>
<td>1.927175</td>
<td>15.0000</td>
<td>161.0000</td>
<td>0.0239619</td>
</tr>
<tr>
<td>Roy</td>
<td>5.0000</td>
<td>0.384649</td>
<td>4.384997</td>
<td>5.0000</td>
<td>57.0000</td>
<td>0.0019053</td>
</tr>
</tbody>
</table>

Signif. codes: 0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
The Tracy-Widom approximation

Let \( W = \text{logit}(x_1) = \log(x_1/(1 - x_1)) \).

Result: [J,08] As \( p \propto n_1, n_2 \rightarrow \infty \), (and with \( O(p^{-2/3}) \) error):

\[
\frac{W - \mu_p}{\sigma_p} \Rightarrow W_\infty \sim F_1.
\]

In other words:

\[
x_1 \approx e^{\mu + \sigma W_\infty} - 1 + e^{\mu + \sigma W_\infty}.
\]

Percentiles \( f_\alpha \):

\[
\begin{align*}
f_{.90} &= 0.4501 \\
f_{.95} &= 0.9793 \\
f_{.99} &= 2.0234
\end{align*}
\]
Centering and Scaling Constants

- Set $N = n_1 + n_2 - 1$

- Define $\gamma, \phi$ from

\[
\sin^2(\frac{\gamma}{2}) = (p - \frac{1}{2})/N, \quad \sin^2(\frac{\phi}{2}) = (n_1 - \frac{1}{2})/N
\]

- Then $\mu$ and $\sigma$ are given by

\[
\mu_p = 2 \log \tan \left( \frac{\phi + \gamma}{2} \right), \quad \sigma_p^3 = \frac{16}{N^2} \frac{1}{\sin^2(\phi + \gamma) \sin \phi \sin \gamma}.
\]

- (given a table of $F_1$), straightforward to code:

```plaintext
papptw, qapptw, rapptw.
```
Accuracy of approximate $\alpha^{th}$ percentile

Approximate percentile using Tracy-Widom percentile $f_{\alpha}$:

$$x_\alpha = x_{\alpha}^{TW}(p, n_1, n_2) = e^{\mu_p + f_{\alpha} \sigma_p} / (1 + e^{\mu_p + f_{\alpha} \sigma_p}).$$

William Chen’s tables: (2003, 2004a, 2002, 2004b) $\Rightarrow$ can compare

$\Delta$ ’Exact’ $x_{\alpha}(p, n_1, n_2)$ with

$\Delta$ T-W approx $x_{\alpha}^{TW}(p, n_1, n_2)$.

Relative error: $r = (x_{\alpha}^{TW} / x_{\alpha}) - 1$.

Tracy-Widom based on $p \rightarrow \infty$, but try $p = 2$ as $n_1, n_2$ vary
p = 2 at 95th percentile

Contours of relative error \( r = (\theta_{\alpha}^{TW}/\theta_{\alpha}) - 1 \)

\( \log_{10} n \)

\( m \)
$p = 4$ at 90th percentile

90th %tile, $S=4$, Contours of relative error $r = \left(\frac{\theta^{TW}}{\theta^{\alpha}}\right)^{-1}$
Recap, for Double Wishart

**T-W approximation to null distribution of Roy’s largest root:**

▲ Conventional percentiles: generally accurate to $< 10\%$ relative error

▲ [Rough $p-$value assessments: qualitatively o.k. over many orders of magnitude]

▲ Similar $O(N^{-2/3})$ approximations for Wishart and Gaussian.

**Recommend:** Tracy-Widom approximation replace $F$—lower bound in default package printouts
Software: **RMTstat**

In development, for R, MATLAB (Z. Ma, P. Perry, M. Shahram, IMJ)

[Preliminary R version now at CRAN]

▲ Tracy-Widom distribution: `ptw`, `qtw`, `dtw`, `rtw`
   (Prähofer-Spohn tables + interpolation)

▲ Sampling from \{Gaussian, Laguerre, Jabobi\}, \{Unitary, Orthogonal\} ensembles

▲ TW approximation to extreme eigenvalues
  ▲ null distributions
  ▲ 'spiked' models

▲ common tests based on largest root
Outline

I. Setting
   1. Single & Double Wishart, examples

II. Largest Eigenvalue: Limiting Law ($H_0$)
   1. Gaussian
   2. Laguerre/Single W.
   3. Jacobi/Double W.

III. Largest Eigenvalue: Concentration
   1. Single W.
   2. Double W.

IV. Largest Eigenvector(s)
Wishart: Concentration for $l_1$

$A \sim W_p(n, I), \quad A = X^TX \quad X \sim N(0, I_n \otimes I_p)$

**Theorem** (e.g. Davidson-Szarek, 2001)  
With $l_1(A) = \sigma^2_1(X)$,

$$P\{\sigma_1(X) > E\sigma_1(X) + \sqrt{nt}\} \leq e^{-nt^2/2},$$

and

$$E\sigma_1(X) \leq \sqrt{n} + \sqrt{p}. $$

from Standard Result: If $X \sim N_m(0, I)$ and $f : \mathbb{R}^m \rightarrow \mathbb{R}$ is Lipschitz-1 (implied by $|\nabla f| \leq 1$), then

$$P\{f(X) > Ef(X) + s\} \leq e^{-s^2/2}. $$
Rescale $X \sim N(0, I_n \otimes I_p)/\sqrt{n}$ so that $E\sigma_1(X) \approx 1 + \sqrt{\gamma}$.

Tracy-Widom limit suggests for $t \approx sn^{-2/3}$,

$$P\{\sigma_1 \geq E\sigma_1 + ct\} \approx e^{-\left(2/3\right)nt^{3/2}}.$$

Lipschitz concentration of measure $\Rightarrow \forall t > 0$

$$P\{\sigma_1 \geq E\sigma_1 + t\} \leq e^{-nt^2/2}.$$

▲ Not optimal (though simple) for ‘small’ deviations $t \ll 1$

(recent: Majumdar-Vargassola)

▲ Effective for ’large’ deviations $t > 1$
Motivations:

▲ analog of Davidson-Szarek
▲ arises in Candes-Tao sparse linear regression, \( Y = A\beta + \epsilon \),
restricted orthogonality constants \( \theta_{S,S'} \), and analogs \( \gamma_{S,S'} \):

\[
\langle A_T v, A_T' v' \rangle \leq \theta_{S,S'} \|v\| \|v'\| \quad \left( \leq \gamma_{S,S'} \|A_T v\| \|A_T' v'\| \right).
\]

Use CCA form: \( X \sim N(0, I_n \otimes I_p) \perp \perp Y \sim N(0, I_n \otimes I_q) \)

\[
R_{p,q;n}(X, Y) = \max \{ \text{Corr}(Xu, Yv) : \|u\| = \|v\| = 1 \}
\]

[ not Lipschitz in \( X, Y \)!]
$R_{max}$ as singular value

Use SVD: \[ X = U_X D_X V_X^T, \quad Y = U_Y D_Y V_Y^T \]

\[ \Rightarrow R(X, Y) = \sigma_1(U_X^T U_Y), \quad \text{with} \]

$U_X \in \mathbb{V}_{n,p}$ – Stiefel manifold of orthonormal $p-$frames, etc

Claim

\[ P\{\sigma_1 \geq E\sigma_1 + t\} \leq e^{-(n-2)t^2/8}. \]

\[ \Rightarrow \text{other double Wishart cases (regression, MANOVA, etc).} \]
Concentration for Riemannian manifolds

Theorem (Ledoux, 1992) Assume:

1. manifold \((X, g)\) compact, connected, smooth Riemannian manifold, dimension \(\geq 2\)
2. volume \(d\mu\) normalized Riemannian volume element
3. Lipschitz \(F\) 1-Lipschitz on \(X\), i.e. \(|\nabla F| \leq 1\)
4. curvature \(c(X) = \inf \{ \text{Ric}(\nabla f, \nabla f) : |\nabla f| = 1 \}\)

Ricci curvature

Then

\[ \mu\{ F \geq \int F d\mu + t \} \leq e^{-ct^2/2}. \]

Need to interpret (1) - (4) for \(F = \sigma_1(U_X^T U_Y) = R(X, Y)\).
Outline

I. Setting
   1. Single & Double Wishart, examples

II. Largest Eigenvalue: Limiting Law ($H_0$)
   1. Gaussian
   2. Laguerre/Single W.
   3. Jacobi/Double W.

III. Largest Eigenvalue: Concentration
   1. Single W.
   2. Double W.

IV. Largest Eigenvector(s)
Estimation of Eigenvectors

\[ S \sim W_p(n, \Sigma), \quad \Sigma = \sigma^2 I + \sum_{\nu=1}^{M} \lambda_{\nu} \theta_{\nu} \theta_{\nu}^T \]

Estimation and inference for \( \theta_{\nu} \)?

Classical: \( p \) fixed, \( n \) large: \( \sqrt{n}(\hat{\theta}_{\nu} - \theta_{\nu}) \to N_p(0, \Gamma_{\nu}) \)

BUT: inconsistency when \( p/n \to \gamma > 0 \):

\[
\langle \hat{\theta}_{\nu}, \theta_{\nu} \rangle \to \begin{cases} 
0 & \lambda_{\nu} \in [0, \sqrt{\gamma}] \\
\frac{1 - \gamma/\lambda_{\nu}^2}{1 + \gamma/\lambda_{\nu}} & \lambda_{\nu} > \sqrt{\gamma}
\end{cases}
\]

Reimann, v.d.Broeck, Bex, Hoyle, Rattray; Paul, Baik, Silverstein
Assume \( \exists \) a basis with sparse representation:
\[
\theta \in \Theta_q(C) : \quad \text{e.g.} \quad |\theta_{\nu,(\mu)}| \leq C|\mu|^{-1/q} \quad q < 2
\]

\( \implies \) near sharp upper & lower bounds for minimax risk:
\[
\inf_{\hat{\theta}} \sup_{\theta_{\nu} \in \Theta_q(C)} E\|\hat{\theta}_{\nu} - \theta_{\nu}\|^2.
\]

Goal: Approximate by “signal in Gaussian noise” model

Lemma \((M = 1)\) Let \( \hat{C} = \langle \hat{\theta}, \theta \rangle \) and \( \hat{\theta}^\perp = \hat{\theta} - \hat{C}'\theta \). Then
\[
\hat{\theta} = \hat{C}'\theta + \hat{SU}
\]

\( U = \hat{\theta}^\perp / \|\hat{\theta}^\perp\| \) is uniform on ”\( S^{p-2} \)” \( \Rightarrow \) nearly Gaussian.

move from eigenvectors to sparse mean estimation.
THANK YOU!


Webpage: [www-stat.stanford.edu/~imj](http://www-stat.stanford.edu/~imj)