Random matrices in statistics: testing in spiked models

Iain Johnstone

Statistics, Stanford

AustMS, Sydney, 1st October 2013

International Year of Statistics - 2013

www.statistics2013.org

- to promote public awareness of Statistics
- steered by professional societies [ASA, RSS, ISI, IBS, ...]
- $\geq$ 2000 participating organizations
  - 25 Australian sponsors at statistics2013.org/iyos/participants_countrysearch.cfm
- Videos, newsletters, website, events, conferences,

Outline

- Random matrices and covariances
  - Principal Components Analysis, examples
- Spiked covariance model
  - Phase transition
  - Weak signals
  - Strong signals

Random Matrix Theory (RMT)

Eigenvalues and vectors of large square random matrices:

$$\mathbf{A} v_j = \mu_j v_j \qquad \mathbf{A} = (A_{ij}) \quad n \times n.$$

Structured randomness:

- A_{ij} i.i.d. Hermitian, or [Wigner matrix]
- $\mathbf{A}$ invariant for $O(n), U(n)$ [GOE, GUE]

Interest in properties of eigenvalues:

- empirical distribution: $F_n(x) = n^{-1} \# \{ i : \lambda_i \leq x \}$
- extremes: $\lambda_{(1)} = \max \lambda_j$
- spacings ...

RMT: 'Wishart' case

Consider $\mathbf{X} = (X_{ij})$ $\quad n \times p \quad$ rectangular, i.i.d entries

Study eigenvalues λ_j of $\mathbf{X}^T\mathbf{X}$

Accessible because $\lambda_j = \mu_j^2$, with $\{\mu_j\}$ eigenvalues of

$$\mathbf{A} = \begin{pmatrix} \mathbf{0} & \mathbf{X} \\ \mathbf{X}^T & \mathbf{0} \end{pmatrix}$$

Link to statistics:

- $n^{-1}\mathbf{X}^T\mathbf{X}$ is (simple form) of covariance matrix.
- goal: helpful approximations based on $p/n \to \gamma > 0$

Covariance Matrices - Population

$X^T = (X_{,j}) \in \mathbb{R}^p$, a random (row) vector distributed as $\mathcal{P}$.

Population mean: $\mu = \mathbb{E}_{\mathcal{P}} X$

Pop. covariance matrix:

$$\Sigma = \mathbb{E}(X - \mu)(X - \mu)^T$$

$$= \mathbb{E}\, XX^T \qquad \text{if } \mu = 0.$$

$$\Sigma_{jj'} = \mathrm{Cov}(X_{,j}, X_{,j'})$$

In general $\Sigma\,(p \times p)$ has $O(p^2/2)$ parameters. Too many!

Simpler models: $\Sigma = \sigma^2 I_p$ 'white'

$\Sigma = \sigma^2(I_p + \sum_1^M h_\nu \mathbf{v}_\nu \mathbf{v}_\nu^T)$ low rank ('spiked')

Covariance Matrices - Sample

Data: $X_1^T, \ldots, X_n^T \in \mathbb{R}^p$ (or $\mathbb{C}^p$)
assumed to be independent draws from $X^T \sim \mathcal{P}_\Sigma$

Sample covariance matrix: $S = n^{-1} \sum_{i=1}^n X_i X_i^T$

Use observed S to estimate or test unknown Σ.

E.g. H_0: $\Sigma = I$ "null" hypothesis
H_A: $\Sigma = I + h\mathbf{v}\mathbf{v}^T$ "alternative" hypothesis

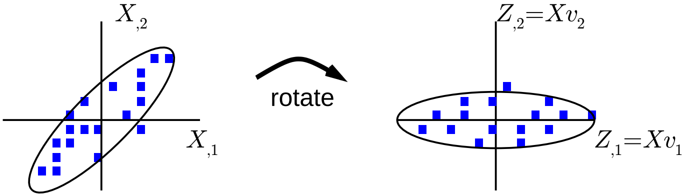
Link to RMT: $nS = \mathbf{X}^T \mathbf{X}$ using the $n \times p$ data matrix

$$\mathbf{X} = (X_{i,j}) = \begin{bmatrix} X_1^T \\ \vdots \\ X_n^T \end{bmatrix}$$

Principal Components Analysis

Statistical interpretation of eigenstructure: $S\mathbf{v}_j = \lambda_j \mathbf{v}_j$.

Goal: reduce dimensionality of data from p (large) to k (small):



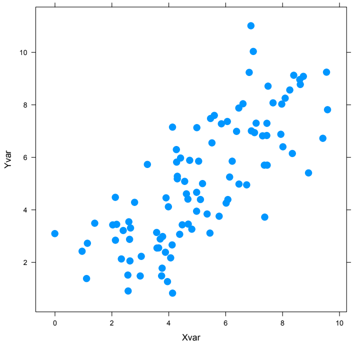
Interpret as directions $\mathbf{v}_j$ of maximum variance, with variances

$$\lambda_j = \max\{\mathbf{v}^T S\mathbf{v} : \mathbf{v}^T \mathbf{v}_{j'}, \|\mathbf{v}\| = 1\}$$
$$= \text{"principal component variances"}$$

Outline

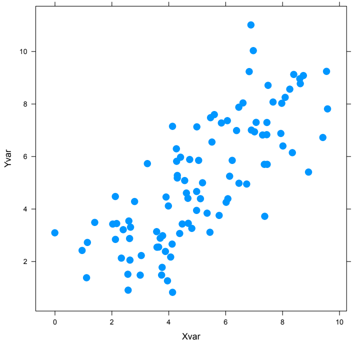
- Random matrices and covariances
  - Principal Components Analysis, two examples: genetics, finance
- Spiked covariance model
  - Phase transition
  - Weak signals
  - Strong signals

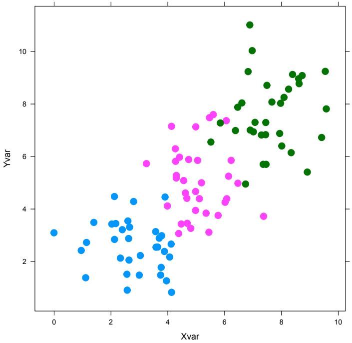
Example 1: PCA & population structure from genetic data



Gene (Y) vs. Phenotype (X) shows apparent correlation, but ...

Example 1: PCA & population structure from genetic data





Gene (Y) vs. Phenotype (X) shows apparent correlation, but ...
3 subpopulations — Within each population, no correlation exists!

Example 1: PCA & population structure from genetic data

Patterson et. al. (2006), Price et. al. (2006)

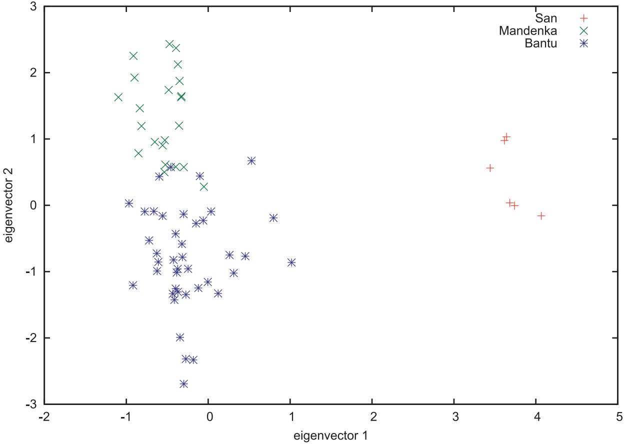
$$n = \#\text{individuals}, \qquad p = \#\text{markers (e.g. SNPs)}$$

$$X_{ij} = \text{(normalized) allele count,}$$
$$\text{case } i = 1, \ldots, n, \qquad \text{marker } j = 1, \ldots, p.$$
$$H = n \times \text{sample covariance matrix of } X_{ij}$$

- Eigenvalues of H: $\lambda_1 > \lambda_2 > \ldots > \lambda_{\min(n,p)}$
- How many λ_i are significant?
- Under H_0, distribution of λ_1 if $H \sim W_p(n, I)$?

Example 1: PCA & population structure from genetic data

- PPR (2006) example: 3 African populations, $n = 67, p = 993$
- Tracy-Widom theory $\Longrightarrow$ 2 "significant" eigenvectors,
  separates populations



Example 2: finance

Arbitrage Pricing Theory $\rightarrow$ a few *factors* “explain” returns

$$R = \sum_{\nu=1}^{M} b_\nu f_\nu + e$$

Example 2: finance

Arbitrage Pricing Theory $\rightarrow$ a few *factors* "explain" returns
Given $j = 1, \ldots, p$ securities, $\quad t = 1, \ldots, T$ observation times, (and $M = 1$),

$$R_{jt} = b_{j1} f_{1t} + e_{jt}.$$

Example 2: finance

Arbitrage Pricing Theory $\rightarrow$ a few *factors* “explain” returns.
Given $j = 1, \ldots, p$ securities, $t = 1, \ldots, T$ observation times,

$$R_{jt} = \sum_{\nu=1}^{M} b_{j\nu} f_{\nu t} + e_{jt}.$$

Example 2: finance

Arbitrage Pricing Theory $\rightarrow$ a few *factors* “explain” returns.
Given $j = 1, \ldots, p$ securities, $t = 1, \ldots, T$ observation times,

$$R_{jt} = \sum_{\nu=1}^{M} b_{j\nu} f_{\nu t} + e_{jt}.$$

Under Gaussian assumptions*, $\Sigma = \text{Cov}\,(R)$ has eigenvalues

$$(\ell_1 > \ell_2 = \cdots = \ell_M > \sigma_e^2, \ldots, \sigma_e^2).$$

$(^*)$ $b_{j\nu} \sim N(\beta, \sigma_b^2)$; $f_{\nu t} \sim N(0, \sigma_f^2)$; $e_{jt} \sim N(0, \sigma_e^2)$ all independent

Example 2: finance

Arbitrage Pricing Theory $\rightarrow$ a few *factors* "explain" returns.
Given $j = 1, \ldots, p$ securities, $t = 1, \ldots, T$ observation times,

$$R_{jt} = \sum_{\nu=1}^{M} b_{j\nu} f_{\nu t} + e_{jt}.$$

Under Gaussian assumptions*, $\Sigma = \text{Cov}\,(R)$ has eigenvalues

$$(\ell_1 > \ell_2 = \cdots = \ell_M > \sigma_e^2, \ldots, \sigma_e^2).$$

Form *sample* covariance matrix S

$$S_{jk} = T^{-1} \sum_t (R_{jt} - \bar{R}_{j\cdot})(R_{kt} - \bar{R}_{k\cdot})$$

Use (largest) sample eigenvalues $\hat{\ell}_i(S)$ to estimate ℓ_i.

$(^*)$ $b_{j\nu} \sim N(\beta, \sigma_b^2)$; $f_{\nu t} \sim N(0, \sigma_f^2)$; $e_{jt} \sim N(0, \sigma_e^2)$ all independent

Example 2: finance

S.J. Brown (1989) simulations, calibrated to NYSE data

4 factor model $\rightarrow \Sigma = \text{diag}(\ \ell_1, \ldots, \ell_4,\ \sigma_e^2, \ldots, \sigma_e^2)$

$$\ell_1 > \ell_2 = \ell_3 = \ell_4 > \sigma_e^2$$

Use $\hat{\ell}_i(S)$ to estimate $\ell_1, \ldots, \ell_4$.

Example 2: finance

S.J. Brown (1989) simulations, calibrated to NYSE data

4 factor model $\rightarrow \Sigma = \text{diag}(\ \ell_1, \ldots, \ell_4,\ \sigma_e^2, \ldots, \sigma_e^2)$

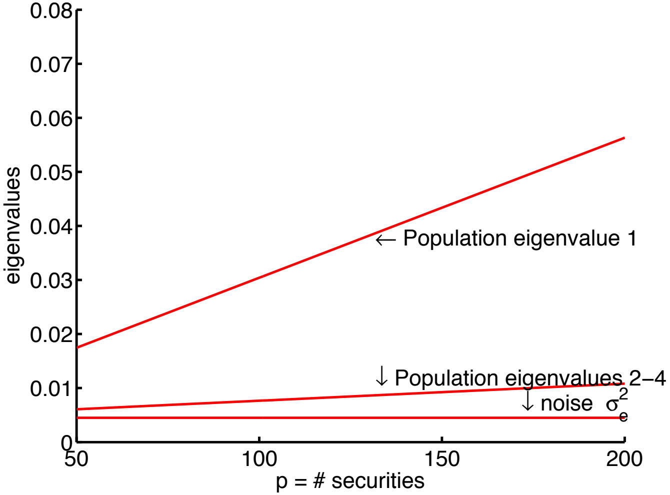
$$\ell_1 > \ell_2 = \ell_3 = \ell_4 > \sigma_e^2$$

Use $\hat{\ell}_i(S)$ to estimate $\ell_1, \ldots, \ell_4$.

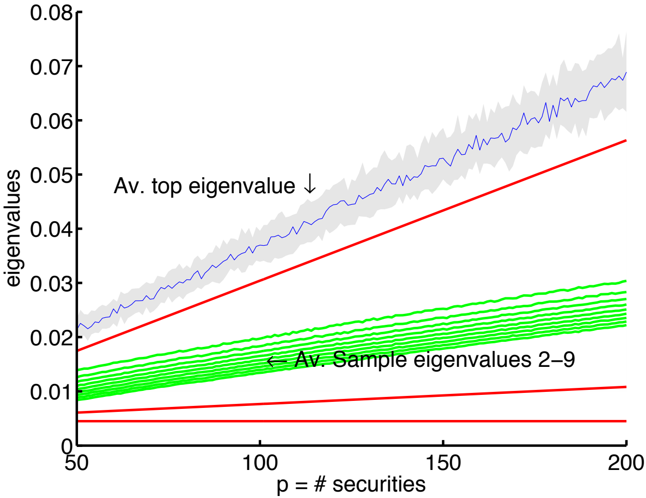
Empirical puzzle (Brown, 1989):
many sample eigenvalues swamp ℓ_2, ℓ_3, ℓ_4.

- Illustration: vary $p = 50(1)200$ $\quad (T = 80)$
- Plot theoretical $\ell_i(p)$ and simulated $\hat{\ell}_i(p)$ versus p.

Example 2: theoretical $\ell_i(p)$ values



Example 2: Brown(1989) plot



Example 2: finance

S.J. Brown (1989) simulations, calibrated to NYSE data

4 factor model* $\rightarrow \Sigma = \text{diag}(\ \ell_1, \ldots, \ell_4, \sigma_e^2, \ldots, \sigma_e^2)$

$$\ell_1 > \ell_2 = \ell_3 = \ell_4 > \sigma_e^2$$

Use $\hat{\ell}_i(S)$ to estimate $\ell_1, \ldots, \ell_4$.

Empirical puzzle (Brown, 1989):

many sample eigenvalues swamp ℓ_2, ℓ_3, ℓ_4.

- Illustration: vary $p = 50(1)200$
- Plot theoretical $\ell_i(p)$ and simulated $\hat{\ell}_i(p)$ versus p.

Explanation (Harding, 2008):

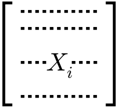
ℓ_2, ℓ_3, ℓ_4 are below a phase transition predicted by RMT.

Outline

- Random matrices and covariances
    - Principal Components Analysis, two examples
- Spiked covariance model
    - Phase transition
    - Weak signals
    - Strong signals

Spiked Covariance Model

- n (independent) observations on p-vectors: X_i
- correlation structure is "white + low rank":



$n \times p$

$$\Sigma = \mathrm{Cov}(X_i) = \sigma^2 I + \sum_{\nu=1}^{M} h_\nu \mathbf{v}_\nu \mathbf{v}_\nu^T$$

Interest in

- testing/estimating h_ν [today]
- determining M
- estimating $\mathbf{v}_\nu$

Some motivating models

1. Economics: X_i = vector of stocks (indices) at time i
   $\mathbf{v}_\nu$ = factor loadings, $f_{\nu i}$ factors, Z_i idiosyncratic terms.

2. ECG: X_i = ith heartbeat (p samples per cycle)
   $\mathbf{v}_\nu$ = may be sparse in *wavelet* basis.

3. Microarrays: X_i = expression of p genes in ith patient.
   $\mathbf{v}_\nu$ = may be sparse few genes involved in each factor.

4. Genetics: X_i = allele count at p SNPs in ith individual.

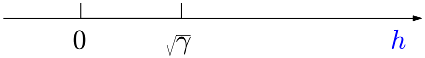
5. Sensors: X_i = observations at sensors
   $\mathbf{v}_\nu$ = cols. of steering matrix, $f_{\nu i}$ signals

6. Climate: X_i = measurements from global network at time i
   $\mathbf{v}_\nu$ = (empirical) orthogonal functions (EOF)

Outline

- Spiked Covariance model
  - Examples $\qquad \Sigma = I + h\,\mathbf{v}\mathbf{v}^T$
- Wishart eigenvalues and Phase Transition

$p/n \to \gamma$



- Weak Signals
  - Contiguity
- Strong Signals
  - Approximations to Power

p and n and all that

$p = \#$ variables/parameters
$n = \#$ of (independent?) observations

- $p = o(n)$ classical statistics
- $n = o(p)$ (nominally) high-dimensional data, sparsity

This talk:

p and n and all that

$p = \#$ variables/parameters
$n = \#$ of (independent?) observations

- $p = o(n)$ classical statistics
- $n = o(p)$ (nominally) high-dimensional data, sparsity

This talk:

- $p/n \to \gamma > 0$ less ambitious; important phenomena appear
  - for e.g. $p = 5, n = 20$, this limit *may* yield better approximation than p fixed, n large.

p and n and all that

$p = \#$ variables/parameters
$n = \#$ of (independent?) observations

- $p = o(n)$ classical statistics
- $n = o(p)$ (nominally) high-dimensional data, sparsity

This talk:

- $p/n \to \gamma > 0$ less ambitious; important phenomena appear
  - for e.g. $p = 5, n = 20$, this limit *may* yield better approximation than p fixed, n large.
- p, n *fixed* (strong signal asymptotics).

Wishart Distribution

$\mathbf{X} = (X_{i,j}) \qquad n \times p$

Rows $\quad X_i^T = (X_{i,j}) \overset{\text{indep}}{\sim} N_p(\mu, \Sigma)$

J. Wishart 1898-1956

Definition: sample covariance, unnormalized:

$$H = \mathbf{X}^T\mathbf{X} \qquad \sim W_p(n, \Sigma) \qquad \text{if } \mu = 0$$

p variables, n degrees of freedom. "Null case:" $\Sigma = I$

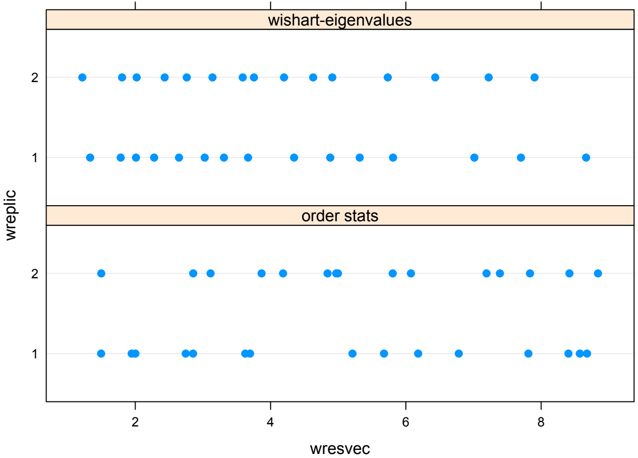
Eigenvalues of H: $\quad \lambda_1 > \lambda_2 > \cdots > \lambda_{n \wedge p} \geq 0$

Wishart Eigenvalues, Null case

2 draws of eigenvalues from $W_{15}(60, I)$

– Spreading of sample eigenvalues from 4 to [1,9].

2 draws of 15 independent $U(1, 9)$ variates – very different!



The Quarter Circle Law

Description of spreading phenomenon in null case:

Marčenko-Pastur, (67) For $H \sim W_p(n, I)$ $\quad p/n \to \gamma \leq 1$

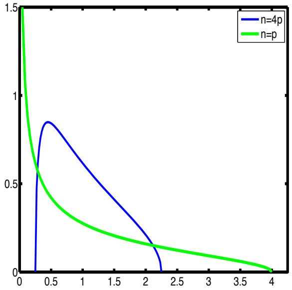
Empirical distribution function: for eigenvalues $\{n\lambda_j\}_{j=1}^p$ of H,

$$F_p(x) = p^{-1}\#\{\lambda_j \leq x\} \to F(x) = f(x)dx.$$

For $\Sigma = I$,

$$f^{MP}(x) = \frac{1}{2\pi\gamma x}\sqrt{(b_+ - x)(x - b_-)},$$

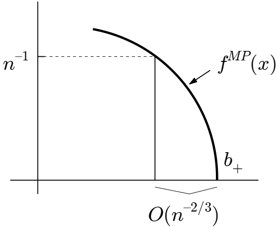
$$b_\pm = (1 \pm \sqrt{\gamma})^2.$$



Largest eigenvalue: Null case

Square root singularity:

$$f^{MP}(x) \sim c\sqrt{b_+ - x}, \qquad x \to b_+$$

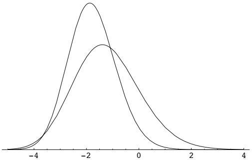


Heuristically,

$$\lambda_1 - b_+ = O_p(n^{-2/3}),$$

and

$$\frac{n^{2/3}\gamma^{1/6}}{(1+\sqrt{\gamma})^{4/3}}(\lambda_1 - b_+) \overset{\mathcal{D}}{\Rightarrow} TW_\beta,$$



the *Tracy-Widom* distributions [$\beta = 1$ for $\mathbb{R}$, $\beta = 2$ for $\mathbb{C}$.]

Largest eigenvalue: Non-null cases

Rank 1 for simplicity: $\Sigma = I + h\,\mathbf{v}\mathbf{v}^T$

For $0 \leq h < \sqrt{\gamma}$,

$$\frac{n^{2/3}\gamma^{1/6}}{(1+\sqrt{\gamma})^{4/3}}(\lambda_1 - (1+\sqrt{\gamma})^2) \stackrel{\mathcal{D}}{\Rightarrow} TW_\beta,$$

Limit does **not** depend on h.

"Fundamental asymptotic limit of sample eigenvalue based detection" (?)

	$\mathbb{R}$	$\mathbb{C}$
$h = 0$	J (01)	Johannson (00)
$h \in (0, \sqrt{\gamma})$	Féral-Péché (09)	Baik-Ben Arous-Péché (05)

Largest eigenvalue: Phase transition

Different rates, limit distributions:

$$\text{For } h < \sqrt{\gamma}: \quad n^{2/3}\left[\frac{\lambda_1 - \mu(\gamma)}{\sigma(\gamma)}\right] \overset{\mathcal{D}}{\Rightarrow} TW_\beta,$$

$$\text{For } h > \sqrt{\gamma}: \quad n^{1/2}\left[\frac{\lambda_1 - \rho(h,\gamma)}{\tau(h,\gamma)}\right] \overset{\mathcal{D}}{\Rightarrow} N(0,1)$$

Largest eigenvalue: Phase transition

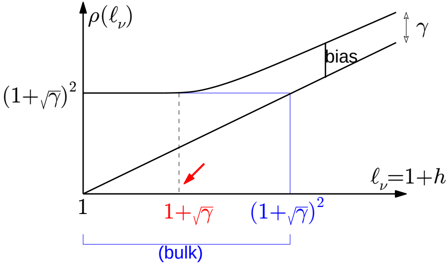
Different rates, limit distributions:

$$\text{For } h < \sqrt{\gamma}: \qquad n^{2/3}\left[\frac{\lambda_1 - \mu(\gamma)}{\sigma(\gamma)}\right] \overset{\mathcal{D}}{\Rightarrow} TW_\beta,$$

$$\text{For } h > \sqrt{\gamma}: \qquad n^{1/2}\left[\frac{\lambda_1 - \rho(h,\gamma)}{\tau(h,\gamma)}\right] \overset{\mathcal{D}}{\Rightarrow} N(0,1)$$

with

$$\rho(h,\gamma) = (1+h)\left(1+\frac{\gamma}{h}\right) \qquad \tau^2(h,\gamma) = 2(1+h)^2\left(1-\frac{\gamma}{h^2}\right)$$



Statistical physics lit, 94- Baik-Ben Arous-Peche(05) , Paul (07) Baik-Silverstein (06), Bloemendal-Virag (11) Mo (11) , Wang (12) Benaych-Georges-Guionnet-Maida (11)

Example:finance

How many factors are present in security returns? Use PCA??
S.J. Brown (1989) simulations, calibrated to NYSE data

4 factor model* $\rightarrow \Sigma = \mathrm{diag}(\ \ell_1, \ldots, \ell_4, \sigma_e^2, \ldots, \sigma_e^2)$

$$\ell_1 > \ell_2 = \ell_3 = \ell_4 > \sigma_e^2$$

Goal: Use $\hat{\ell}_i(S)$ to estimate $\ell_1, \ldots, \ell_4$.

Empirical puzzle (Brown, 1989):

many sample eigenvalues swamp ℓ_2, ℓ_3, ℓ_4.

Example:finance

How many factors are present in security returns? Use PCA??
S.J. Brown (1989) simulations, calibrated to NYSE data

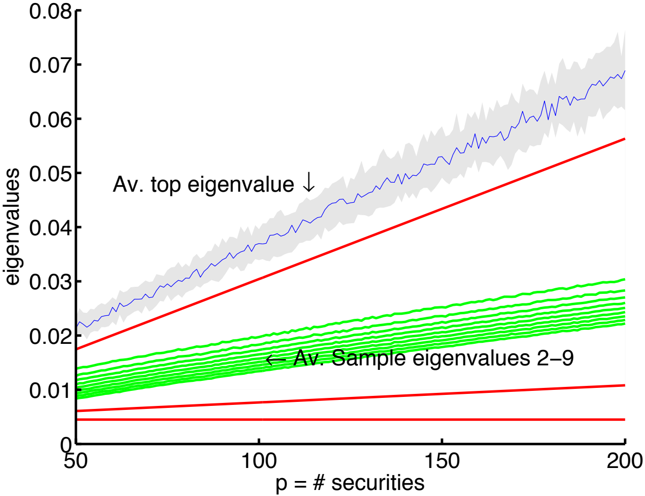
4 factor model* $\rightarrow \Sigma = \text{diag}(\ \ell_1, \ldots, \ell_4, \sigma_e^2, \ldots, \sigma_e^2)$
$\ell_1 > \ell_2 = \ell_3 = \ell_4 > \sigma_e^2$

Goal: Use $\hat{\ell}_i(S)$ to estimate $\ell_1, \ldots, \ell_4$.

Empirical puzzle (Brown, 1989):
many sample eigenvalues swamp ℓ_2, ℓ_3, ℓ_4.

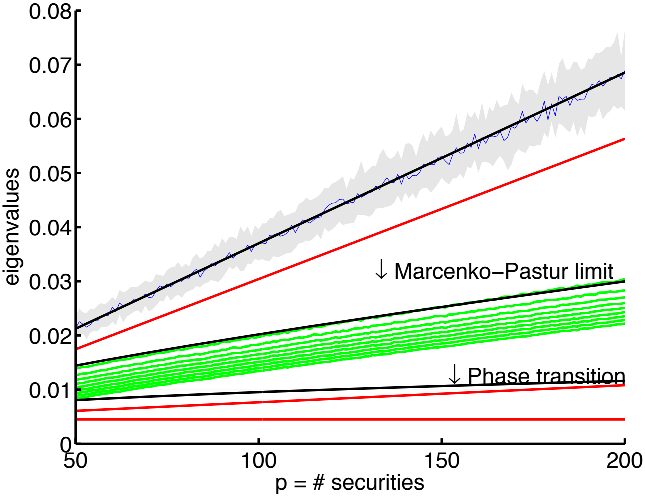
Explanation (Harding, 2008):
ℓ_2, ℓ_3, ℓ_4 are below the $1 + \sqrt{\gamma}$ phase transition.

Brown(1989) plot



Source: Harding(2008).

Marcenko-Pastur & phase transition

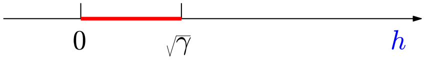


Source: Harding(2008).

Outline

- Spiked Covariance model
  - Examples $\Sigma = I + h\,\mathbf{v}\mathbf{v}^T$
- Phase Transition
- Weak Signals
  - Contiguity

$p/n \to \gamma$



- Strong Signals
  - Approximations to Power

Detecting weak signals?

For $h < \sqrt{\gamma}$, distribution of largest eigenvalue

$$\lambda_1 \approx \mu(\gamma) + n^{-2/3}\sigma(\gamma)\, TW_1$$

does not depend on h.

Onatski-Moreira-Hallin, AOS (2013):

- can detect $h < \sqrt{\gamma}$, with error
- use all eigenvalues
- contiguity ideas yield limit distributions for $h \in (0, \sqrt{\gamma})$.

Likelihood Ratio Test

$X_i \sim N_p(0, I + h\mathbf{v}\mathbf{v}^T), \quad H_0 : h = 0$ vs. $H_1 : h > 0$, $\mathbf{v}$ unspecified.

Invariant under rotations, so consider

$$p(\lambda; h) = \text{joint density of sample eigenvalues } \lambda = (\lambda_1, \ldots, \lambda_n).$$

Likelihood ratio test against fixed $h > 0$:

$$L(\lambda; h) = \frac{p(\lambda; h)}{p(\lambda; 0)}$$

Likelihood Ratio Test

$X_i \sim N_p(0, I + h\mathbf{v}\mathbf{v}^T), \quad H_0 : h = 0$ vs. $H_1 : h > 0$, $\mathbf{v}$ unspecified.

Invariant under rotations, so consider

$p(\lambda; h) =$ joint density of sample eigenvalues $\lambda = (\lambda_1, \ldots, \lambda_n)$.

$$= \frac{\gamma(n, p, \lambda)}{(1+h)^{n/2}} \int_{S(p)} e^{\frac{n}{2} \frac{h}{1+h} x_p' \Lambda x_p} \, (dx_p)$$

with $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_p)$.

Likelihood ratio test against fixed $h > 0$:

$$L(\lambda; h) = \frac{p(\lambda; h)}{p(\lambda; 0)} = \frac{1}{(1+h)^{n/2}} \int_{S(p)} e^{\frac{n}{2} \frac{h}{1+h} x_p' \Lambda x_p} \, (dx_p)$$

Asymptotic normality of likelihood ratio

Under H_0 ($h = 0$), for $0 \leq h \leq \bar{h} < \sqrt{\gamma}$, and $p/n \to \gamma$

$$\log L(h; \lambda) \Rightarrow \mathcal{L}(h; \lambda), \qquad \text{(O-M-H, 2013)}$$

a Gaussian process [by Bai-Silverstein CLT in RMT], with

Asymptotic normality of likelihood ratio

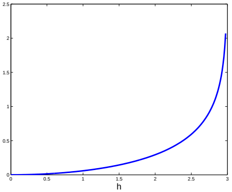
Under H_0 ($h = 0$), for $0 \le h \le \bar{h} < \sqrt{\gamma}$, and $p/n \to \gamma$

$$\log L(h;\lambda) \Rightarrow \mathcal{L}(h;\lambda), \qquad \text{(O-M-H, 2013)}$$

a Gaussian process [by Bai-Silverstein CLT in RMT], with

$$E\mathcal{L}(h;\lambda) = \tfrac{1}{4}\log\left(1 - \frac{h^2}{\gamma}\right)$$

$$\mathrm{Cov}\{\mathcal{L}(h_1;\lambda), \mathcal{L}(h_2;\lambda)\} = -\tfrac{1}{2}\log\left(1 - \frac{h_1 h_2}{\gamma}\right)$$



[if $h \ge H > \sqrt{\gamma}$, $\qquad L(h;\lambda) = O_p(e^{-n\delta})$.]

Detecting weak signals

Reparametrize: $\theta = \sqrt{-\log(1 - h^2/\gamma)}$ for $h < \sqrt{\gamma}$.

Seek optimal test of $H_0 : \theta = 0$ vs. $H_A : \theta = \theta_1 > 0$

Recap: Likelihood ratio: $L_n(\lambda; \theta_1) = p(\lambda; \theta_1)/p(\lambda; 0)$ satisfies

$$\log L_n \overset{P_{n,0}}{\Rightarrow} N(-\theta_1^2/4, \theta_1^2/2)$$
$$\overset{P_{n,\theta_1}}{\Rightarrow} N(+\theta_1^2/4, \theta_1^2/2) \qquad \text{(Contiguity!)}$$

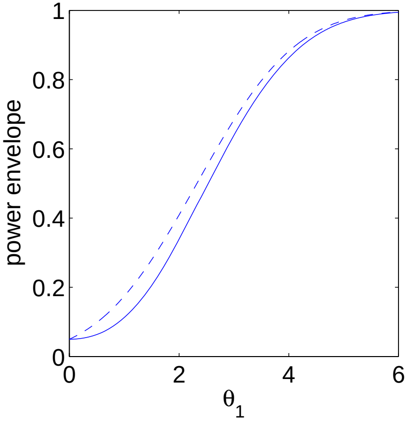
So for asymptotically optimal test,

$$\text{Reject} \quad \Leftrightarrow \quad \log L_n > C_{n,\alpha} = \theta_1 z_\alpha/\sqrt{2} - \theta_1^2/4$$

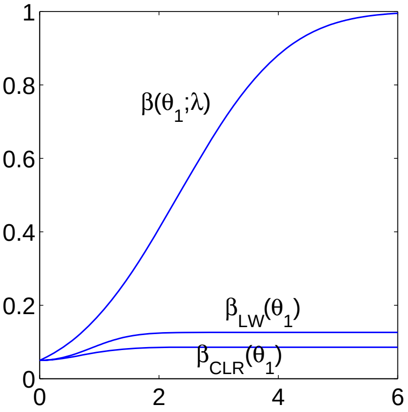
Asymptotic Power

$$\text{Compute 'Power' } \beta(\theta_1) = P_{\theta=\theta_1}(\text{Reject}) = \lim P_{n,\theta_1}(\log L_n > C_{n,\alpha})$$

$$\text{if} \qquad P_{\theta=0}(\text{Reject}) = \alpha$$



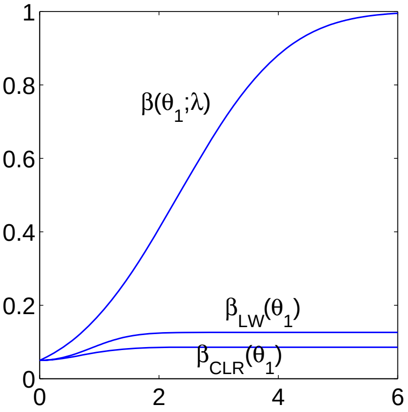
Asymptotic Power



$LW = p^{-1}\mathrm{tr}[(\hat{\Sigma} - I)^2] - \gamma_n[p^{-1}\mathrm{tr}\hat{\Sigma}]^2 + \gamma_n, \qquad \gamma_n = p/n, \hat{\Sigma} = H/n$ [Ledoit-Wolf]

$CLR = \mathrm{tr}\hat{\Sigma} - \log\det\hat{\Sigma} - p(1 - (1 - \gamma_n^{-1})\log(1 - \gamma_n))$ [Bai et. al.]

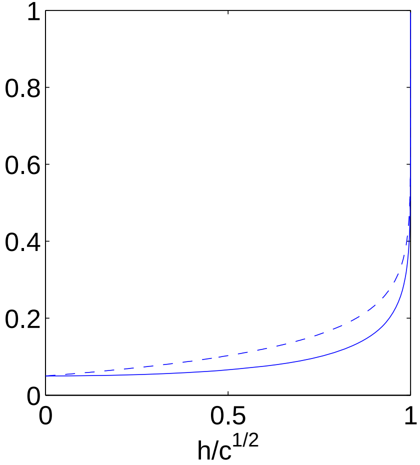
Asymptotic Power



Recall that $\theta = \sqrt{-\log(1 - h^2/\gamma)}$ ...

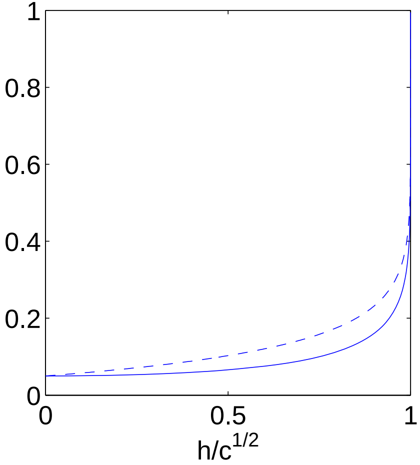
A dose of reality

In original parameter h, power is good only very close to $\sqrt{\gamma}$.



A dose of reality

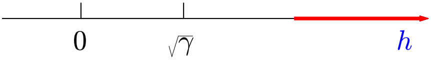
In original parameter h, power is good only very close to $\sqrt{\gamma}$.



.... "It is not done well; but you are surprised to find it done at all."

Outline

- Spiked Covariance model
  - Examples $\qquad \Sigma = I + h\,\mathbf{v}\mathbf{v}^T$
- Wishart eigenvalues and Phase Transition
- Weak Signals
  - Contiguity
- Strong Signals
  - Approximations to Power [with Boaz Nadler]



Strong signals

$$H \sim W_p(n, \sigma^2 I + h\mathbf{v}\mathbf{v}^T)$$

So far: $h < \sqrt{\gamma}$: $\lambda_1(H) \sim \mu_{TW} + \sigma_{TW} TW / n^{2/3}$

$h > \sqrt{\gamma}$: $\lambda_1(H) \sim N(\mu_{h,\gamma}, \sigma^2_{h,\gamma}/n)$

In strong signal regime: $h \gg \sqrt{\gamma}$,

Strong signals

$$H \sim W_p(n, \sigma^2 I + h \mathbf{v}\mathbf{v}^T)$$

So far:

$$h < \sqrt{\gamma}: \qquad \lambda_1(H) \sim \mu_{TW} + \sigma_{TW} TW / n^{2/3}$$

$$h > \sqrt{\gamma}: \qquad \lambda_1(H) \sim N(\mu_{h,\gamma}, \sigma^2_{h,\gamma}/n)$$

In strong signal regime: $h \gg \sqrt{\gamma}$, for rank one alternatives:

largest eigenvalue test is actually best test

e.g. p fixed, n large [so $\gamma = p/n \sim 0$]

$$\log L(h; \lambda) = \frac{n}{2}\left[\lambda_1 \frac{h}{h+1} - \log(1+h)\right](1 + o(1)).$$

Change perspective

$$H \sim W_p(n, \sigma^2 I + h\,\mathbf{v}\mathbf{v}^T)$$

Consider n, p fixed

Strong signal: h large $\Leftrightarrow$ σ^2 small

Goal: power approximation for "Roy's largest root test":

find $P_h(\lambda_1 > \lambda^{(\alpha)})$ where $P_0(\lambda_1 > \lambda^{(\alpha)}) = \alpha.$

Change perspective

$$H \sim W_p(n, \sigma^2 I + h\,\mathbf{v}\mathbf{v}^T)$$

Consider n, p fixed

Strong signal: h large $\Leftrightarrow$ σ^2 small

Goal: power approximation for "Roy's largest root test":

find $P_h(\lambda_1 > \lambda^{(\alpha)})$ where $P_0(\lambda_1 > \lambda^{(\alpha)}) = \alpha.$

An old open issue:

A.T. James (64): *'For numerical evaluation ... power series expansions of hypergeometric functions are of very limited value.'*

T.W. Anderson (84): *'No straightforward method exists for computing powers for Roy's statistic itself.'*

O'Brien and Shieh (92): *'To date, no acceptable method has been developed for transforming Roy's largest root test statistic to an F or χ^2 statistic.'*

Dozens of textbooks; G*Power3 software (07): power for linear statistics, not λ_1.

Small σ perturbation approach

Initial reductions: $\Sigma = I, \qquad v = e_1$

Suppose, at first deterministically

$$X_i = \begin{pmatrix} u_i \\ \mathbf{0} \end{pmatrix} + \sigma \begin{pmatrix} 0 \\ \boldsymbol{\xi}_i \end{pmatrix}$$

Then

$$H_\sigma = X^T X = \begin{bmatrix} z & \mathbf{0}^T \\ \mathbf{0} & 0_{m-1} \end{bmatrix} + \sigma\sqrt{z} \begin{bmatrix} 0 & b^T \\ b & 0_{m-1} \end{bmatrix} + \sigma^2 \begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & Z \end{bmatrix}$$

$$= A_0 + \sigma A_1 + \sigma^2 A_2$$

Small σ perturbation approach

Initial reductions: $\Sigma = I, \qquad v = e_1$

Suppose, at first deterministically

$$X_i = \begin{pmatrix} u_i \\ \mathbf{0} \end{pmatrix} + \sigma \begin{pmatrix} 0 \\ \boldsymbol{\xi}_i \end{pmatrix}$$

Then

$$H_\sigma = X^T X = \begin{bmatrix} z & \mathbf{0}^T \\ \mathbf{0} & 0_{m-1} \end{bmatrix} + \sigma\sqrt{z} \begin{bmatrix} 0 & b^T \\ b & 0_{m-1} \end{bmatrix} + \sigma^2 \begin{bmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & Z \end{bmatrix}$$

$$= A_0 + \sigma A_1 + \sigma^2 A_2$$

Stochastic assumptions: [to get $W_p(n, \sigma^2 I + h\, v v^T)$]

$$u_i \sim N(0, \sigma^2 + h) \qquad \boldsymbol{\xi}_i \sim N_{m-1}(0, I)$$

Single matrix result

Proposition:

SP: Assume $H \sim W_p(n, \sigma^2 I + hvv^T)$. Then

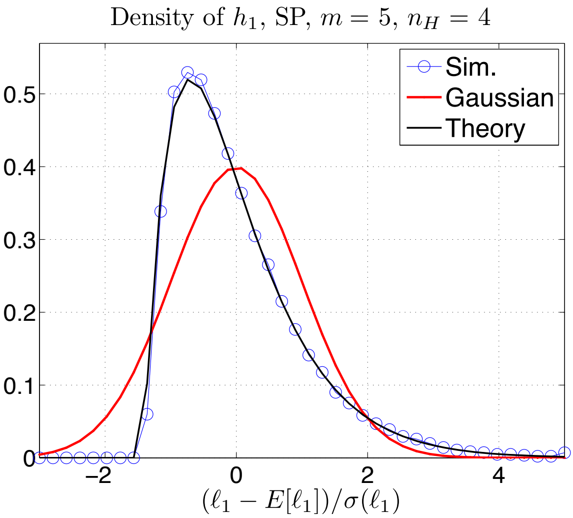
$$\lambda_1(H_\sigma) \sim V_0 + \sigma^2 V_2 + \sigma^4 V_4 + o_p(\sigma^4),$$

with

$$V_0 = (h + \sigma^2)\chi^2_n, \qquad V_2 = \chi^2_{p-1}, \qquad V_4 = (V_2/V_0)\chi^2_{n-1}$$

and each χ^2 is independent.

Example: Signal Detection, $h = 10$



[Classical] Multivariate Analysis

Single Wishart

- Principal Component analysis
- Factor analysis
- Multidimensional scaling

Double Wishart

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

[Classical] Multivariate Analysis

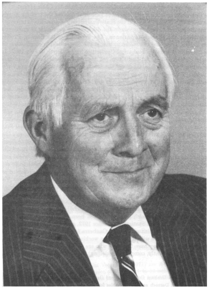
Single Wishart

- Principal Component analysis
- Factor analysis
- Multidimensional scaling

Double Wishart

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

A. T. James, 1924–2013



A. James (1964): Five-fold Way

Unified view of multivariate eigenvalue distributions:

single matrix,

Multivariate	Univariate Analog	Wisharts	Typical Application
${}_0F_0$	χ^2	$H \sim W_m(n, \Sigma + \Omega)$	Signal Detection Σ known
${}_0F_1$	non-central χ^2	$H \sim W_m(n, \Sigma, \Omega)$	Equality of Means Σ known

A. James (1964): Five-fold Way

Unified view of multivariate eigenvalue distributions:

two matrix,

Multivariate	Univariate Analog	Wisharts	Typical Application
${}_1F_0$	F	$H \sim W_m(n, \Sigma + \Omega)$ $E \sim W_m(n', \Sigma)$	Signal Detection Σ unknown
${}_1F_1$	non-central F	$H \sim W_m(n, \Sigma, \Omega)$ $E \sim W_m(n', \Sigma)$	Equality of Means Σ unknown

A. James (1964): Five-fold Way

Unified view of multivariate eigenvalue distributions:

canonical correlations

Multivariate	Univariate Analog	Wisharts	Typical Application
${}_2F_1$	$r^2/(1-r^2)$	$H \sim W_m(q, \Sigma, \Omega)$ $E \sim W_m(n-q, \Sigma)$	Canonical Correlations

A. James (1964): Five-fold Way

Unified view of multivariate eigenvalue distributions:

single matrix, two matrix, canonical correlations

Multivariate	Univariate Analog	Wisharts	Typical Application
${}_0F_0$	χ^2	$H \sim W_m(n, \Sigma + \Omega)$	Signal Detection Σ known
${}_0F_1$	non-central χ^2	$H \sim W_m(n, \Sigma, \Omega)$	Equality of Means Σ known
${}_1F_0$	F	$H \sim W_m(n, \Sigma + \Omega)$ $E \sim W_m(n', \Sigma)$	Signal Detection Σ unknown
${}_1F_1$	non-central F	$H \sim W_m(n, \Sigma, \Omega)$ $E \sim W_m(n', \Sigma)$	Equality of Means Σ unknown
${}_2F_1$	$r^2/(1 - r^2)$	$H \sim W_m(q, \Sigma, \Omega)$ $E \sim W_m(n - q, \Sigma)$	Canonical Correlations

Power, MANOVA example

Small σ perturbation approach extends to rank one alternatives in all James' cases. One example: ($_1F_1$ case, MANOVA)

Power, MANOVA example

Small σ perturbation approach extends to rank one alternatives in all James' cases. One example: (${}_1F_1$ case, MANOVA)

dim m	groups p	samples n_i	non-cent ω	power simulated	power approx	relative error
3	3	10	10	0.483	0.477	-.012
3	3	10	20	0.847	0.852	.006
3	3	10	40	0.995	0.996	.001
6	3	10	10	0.320	0.304	-.050
6	3	10	20	0.671	0.668	-.004
6	3	10	40	0.964	0.967	.003
10	6	20	10	0.208	0.136	-.346
10	6	20	20	0.520	0.442	-.150
10	6	20	40	0.932	0.912	-.021

$(\alpha = .05)$ (SE $\leq$ .0016).

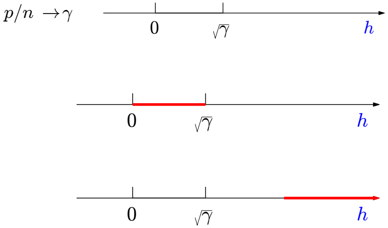
Approximation better for larger ω, smaller m, p, n and plausible power

Conclusion- I

- Spiked Covariance model
  - Examples,

$$\Sigma = I + h\,\mathbf{v}\mathbf{v}^T$$

- Phase Transition
- Weak Signals
  - Contiguity
- Strong Signals
  - Power Approximations



- Many extensions possible in other multivariate settings

Conclusion-II

McKinsey report 2011: projects excess demand for 140,000 - 190,000 "deep analytical positions"

Maths + Stats + Computing → good jobs

Conclusion-II

McKinsey report 2011: projects excess demand for 140,000 - 190,000 “deep analytical positions”

Maths + Stats + Computing → good jobs

THANK YOU!