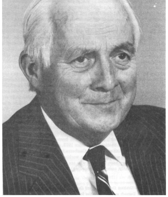
Low rank structure in highly multivariate models

Iain Johnstone

Statistics & Biomedical Data Science, Stanford

ISI Bengalaru, Feb 2 2017

Theme



High dimensional data

with low dimensional structure

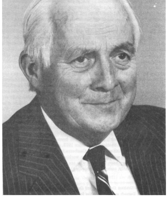
revives old classification of multivariate methods

James' classification suggests how

phenomena seen in high-dimensional PCA

occur widely in multivariate settings

Theme



High dimensional data

with low dimensional structure

revives old classification of multivariate methods

James' classification suggests how

phenomena seen in high-dimensional PCA

occur widely in multivariate settings

joint with Alexei ONATSKI, Prathapa DHARMAWANSA

Outline

- High-d Phenomena in PCA - Overview
  - Spiked model for low-d structure, phase transition, testing

Outline

- High-d Phenomena in PCA - Overview
  - Spiked model for low-d structure, phase transition, testing
- James' family of multivariate methods
  - examples, extension to high-d/low-d structure

Outline

- High-d Phenomena in PCA - Overview
  - Spiked model for low-d structure, phase transition, testing
- James' family of multivariate methods
  - examples, extension to high-d/low-d structure
- Likelihood ratio testing in spiked models
  - Gaussian process limits

Outline

- High-d Phenomena in PCA - Overview
  - Spiked model for low-d structure, phase transition, testing
- James' family of multivariate methods
  - examples, extension to high-d/low-d structure
- Likelihood ratio testing in spiked models
  - Gaussian process limits
- Common threads
  - James' hypergeometric functions, approximations

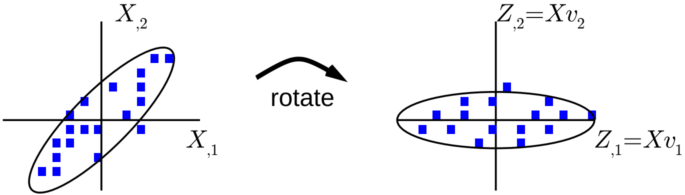
Principal Components Analysis (PCA, KLT, EOF, POD, EVD...)

Sample covariance matrix $S = n^{-1} \sum (X_i - \bar{X})(X_i - \bar{X})'$

Eigenstructure: $S\mathbf{v}_j = \lambda_j \mathbf{v}_j$.

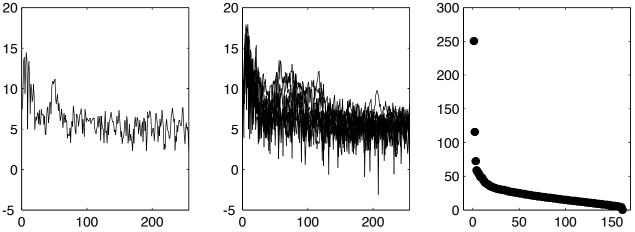
Reduce dimensionality: p (large) to k (small):

Directions $\mathbf{v}_j$ of maximum variance λ_j, $1 \leq j \leq k$,



$\Leftrightarrow$ Best low rank k approximation to data.

Scree plot and spikes



[Buja-Hastie-Tibshirani, 95; J, 01]

- scree plot of ordered eigenvalues
- here, $p = 256, n = 162$ i.e. $\boxed{p \propto n}$
- some (how many?) sample eigenvalues emerge from "bulk"
- which population eigenvalues differ from H_0, e.g. $\Sigma = I$?

Spiked Covariance Model

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

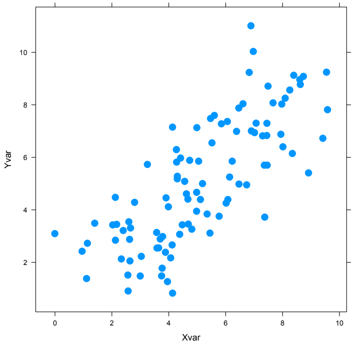
$$\begin{bmatrix} \cdots\cdots\cdots \\ \cdots\cdots\cdots \\ \cdots X_i \cdots \\ \cdots\cdots\cdots \end{bmatrix}_{n \times p}$$

$$\Sigma = \mathrm{Cov}(X_i) = \sigma^2 \Sigma_0 + \sum_{\nu=1}^{M} h_\nu \gamma_\nu \gamma_\nu^T \qquad [J, 01]$$

Interest in

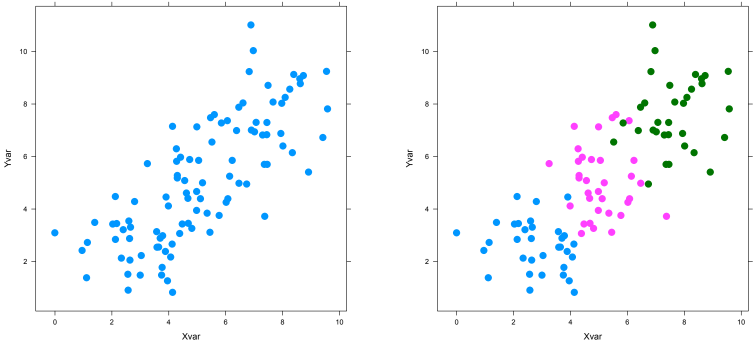
- testing/estimating h_ν [today]
- determining M
- estimating γ_ν, Σ

Example: PCA & population structure from genetic data



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

Example: PCA & population structure from genetic data



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

Example: PCA & population structure from genetic data

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

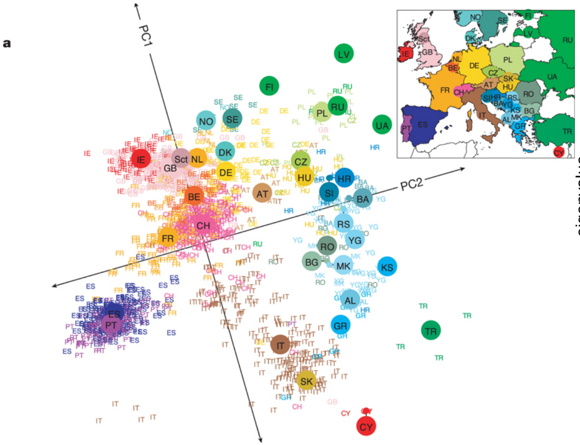
$$n = \#\text{individuals}, \qquad p = \#\text{markers (e.g. SNPs)}$$

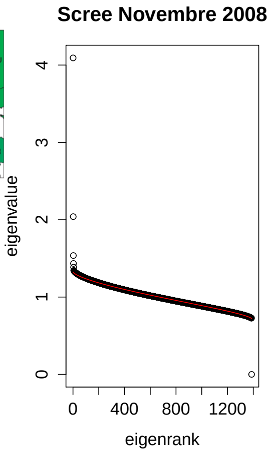
$$X_{ij} = \text{(normalized) allele count},$$
$$\text{case } i = 1, \ldots, n, \qquad \text{marker } j = 1, \ldots, p.$$
$$H = n \times \text{sample covariance matrix of } X_{ij}$$

- 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 $H \sim W_p(n, I)$?

Novembre et. al. 2008

“Genes mirror geography within Europe”, *Nature*





Some other examples

1. Economics: X_i = vector of stocks (indices) at time i
   γ_ν = factor loadings,
2. Virology: X_i = mutations at p amino acid sites in sample i
   γ_ν = sectors of functional significance.
3. ECG: X_i = ith heartbeat (p samples per cycle)
   γ_ν = may be sparse in wavelet basis.
4. Sensors: X_i = observations at sensors
   γ_ν = cols. of steering matrix,
5. Climate: X_i = measurements from global network at time i
   γ_ν = (empirical) orthogonal functions (EOF)

Recap: PCA for rank one spiked model

$$X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(0, \Sigma_p) \qquad\qquad \Sigma_p = I_p + h\gamma\gamma'$$

$$\Rightarrow \quad nH = X'X = \sum_1^n X_i X_i' \sim W_p(n, \Sigma)$$ Wishart distribution

Double scaling: $p = p_n, \quad p/n \to c_1 > 0.$

Recap: PCA for rank one spiked model

$$X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(0, \Sigma_p) \qquad \Sigma_p = I_p + h\gamma\gamma'$$

$$\Rightarrow \quad nH = X'X = \sum_1^n X_i X_i' \sim W_p(n, \Sigma)$$ Wishart distribution

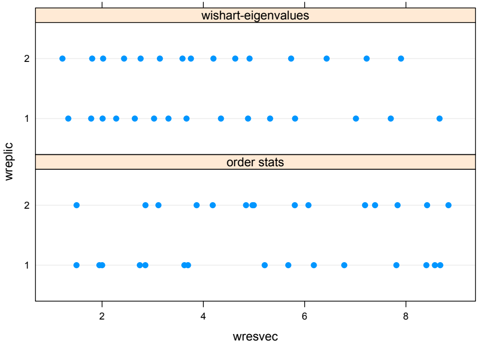
Double scaling: $p = p_n, \quad p/n \to c_1 > 0.$

$\boxed{\Sigma = I,}$ $p = 15 \ n = 60$

Sample eigenvalues:
$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$.

vs.

$U_1 \ldots U_{15} \overset{ind}{\sim} U(0, 9)$



Marchenko-Pastur Law

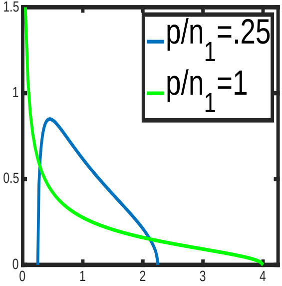
Let $nH \sim W_p(n, I)$ $\qquad p/n \to c_1 \leq 1$

Empirical d.f. of eigenvalues $\{\lambda_j\}_{j=1}^p$ of H,

$$p^{-1}\#\{\lambda_j \leq x\} \to \int_{-\infty}^x f^{MP}$$

$$f^{MP}(x) = \frac{1}{2\pi c_1 x}\sqrt{(b_+ - x)(x - b_-)},$$

$$b_\pm = (1 \pm \sqrt{c_1})^2.$$

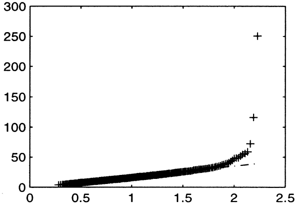


Q-Q plots against Marchenko-Pastur law

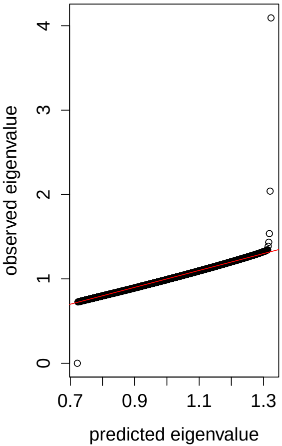
– Wachter's (1976) version of scree-plot

– bulk matches Marchenko-Pastur, + some spikes

Phoneme data [J,01]



QQPlot Novembre 2008

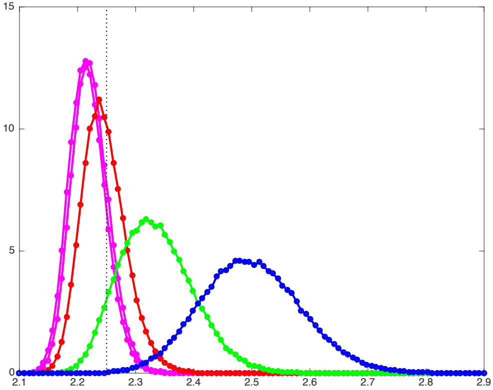


Largest Eigenvalue λ_1: Numerical illustration

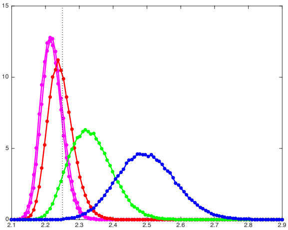
$$p = 200, n = 800 \qquad [\text{i.e. } c_1 = p/n = 0.25]$$

subcritical critical supercritical

Spike $h =$ $0,\ 0.25,$ $h_+ = 0.5,$ $0.75,\ 1.$



Largest eigenvalue: BBP Phase transition



Different rates, limit distributions:

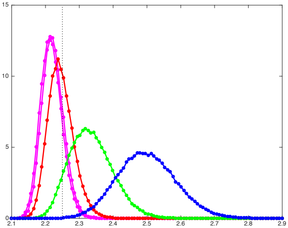
$$\Sigma = I_p + h\,\gamma\gamma', \qquad p/n \to c_1$$

$$\text{For } h < \sqrt{c_1}: \qquad \lambda_1 \approx \mu(c_1) + \frac{\sigma(c_1)}{n^{2/3}} TW_\beta,$$

$$\text{For } h > \sqrt{c_1}: \qquad \lambda_1 \approx \rho(h, c_1) + \frac{\tau(h, c_1)}{n^{1/2}} N(0,1)$$

[$X_i \in \mathbb{C}$: Baik-Ben Arous-Peché (05);
$X_i \in \mathbb{R}$: $h = \sqrt{c_1}$: Bloemendal-Virag (13)]

Inference below Phase Transition



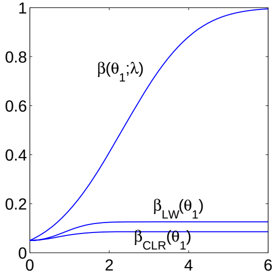
For $h < \sqrt{c_1}$:

$$\lambda_1 \approx \mu(c_1) + \frac{\sigma(c_1)}{n^{2/3}} TW_\beta,$$

- **Largest** eigenvalue λ_1 carries **no** information
- but ... **Can** build informative test using **all** eigenvalues:

[Onatski-Moreira-Hallin, 13]:

Power for $\theta_1 = \sqrt{-\log(1 - h^2/c_1)}$.

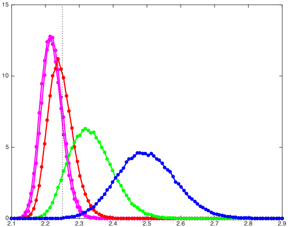


Inference above Phase Transition

For $h > \sqrt{c_1}$:

$$\lambda_1 \approx \rho(h, c_1) + \frac{\tau(h, c_1)}{n^{1/2}} N(0,1)$$

$$[\,\rho(h, c_1) > 1 + h\,]$$



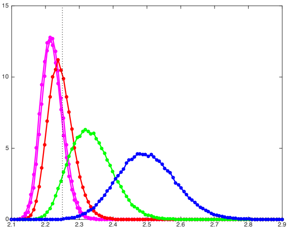
- Gaussian limits for λ_1 with $\rho(h) \nearrow$ in h
- $\Rightarrow$ can distinguish $h_0 \neq h_1$
- optimality of λ_1, confidence intervals for h?

Inference above Phase Transition

For $h > \sqrt{c_1}$:

$$\lambda_1 \quad \approx \quad \rho(h, c_1) \quad + \quad \frac{\tau(h, c_1)}{n^{1/2}} N(0,1)$$

$$[\, \rho(h, c_1) > 1 + h \,]$$



- Gaussian limits for λ_1 with $\rho(h) \nearrow$ in h
- $\Rightarrow$ can distinguish $h_0 \neq h_1$
- optimality of λ_1, confidence intervals for h?

Other phenomena: inconsistency of sample eigenvectors, estimation of Σ, ...

Outline

- High-d Phenomena in PCA
- James' family of multivariate methods
  - examples: MANOVA, regression, CCA ...
  - extension to high-d/low-d structure
- Likelihood ratio testing in spiked models
- Common threads

Basic equation of classical multivariate statistics

$$\det(H - x E) = 0$$

with $p \times p$ matrices

$$n_1 H = \sum_{\nu=1}^{n_1} \mathbf{x}_\nu \mathbf{x}_\nu' \qquad \text{‘hypothesis’ SS}$$

$$n_2 E = \sum_{\nu=1}^{n_2} \mathbf{z}_\nu \mathbf{z}_\nu' \qquad \text{‘error’ SS}$$

(Invariant) methods use generalized eigenvalues $\{x_i\}_{i=1}^{p}$

$\Leftrightarrow$ eigenvalues of "F-ratio" $E^{-1}H$.

Topics with $E^{-1}H$ in (> 20) textbooks

- Canonical correlation analysis
- Discriminant analysis
- Factor analysis*
- Multidimensional scaling*
- Multivariate Analysis of Variance – MANOVA
- Multivarate regression analysis
- Principal Component analysis*
- Signal detection (equality of covariance matrices)

* use limiting form $\det(H - xI) = 0$ with $E = I_p$, $(n_2 \to \infty)$

Principal Components Analysis (PCA) $[{}_0F_0]$

Data

$$X = [\mathbf{x}_1 \cdots \mathbf{x}_{n_1}] \qquad p \times n_1$$

Covariance structure:

$$\Sigma = \mathrm{Cov}(\mathbf{x}_\nu) = \Sigma_0 \ + \Phi$$

Low rank:

$$\Phi = \sum_{k=1}^{r} \theta_k \gamma_k \gamma_k'$$

Sample covariance matrix:

$$n_1 H = XX'$$

Eigenvalues:

$$\det(H - \lambda_i I) = 0$$

Regression - Known Covariance (REG$_0$) $[{}_0F_1]$

p-variate response, $\nu = 1, \ldots, n_1$

$$\mathbf{y}_\nu = B'\mathbf{x}_\nu + \mathbf{z}_\nu, \qquad \Sigma_0 = \mathrm{Cov}(\mathbf{z}_\nu) \ \text{known}$$
$$H_0 : CB = 0 \qquad C = \text{contrast matrix}$$

Sums of squares matrix:

$$n_1 H = Y P_H Y' \qquad n_1 \text{ hypothesis d.f.}$$

Eigenvalues:

$$\det(H - \lambda_i I) = 0.$$

Low rank: noncentrality (e.g. MANOVA), $M = \mathbb{E}Y = B'X$

$$\Phi = \Sigma_0^{-1} M M' / n_1 = \Sigma_0^{-1} \sum_{k=1}^{r} \theta_k \gamma_k \gamma_k'$$

Matrix Denoising (also REG_0) $[{}_0F_1]$

$$Y = M + Z \qquad Z \sim N(0, \sigma^2 I_{n_1} \otimes I_p)$$

Low rank mean:

$$M = \sum_{k=1}^{r} \sqrt{n_1 \theta_k} \gamma_k \psi_k'$$

SVD of Y uses eigenvalues of

$$n_1 H = YY'$$

Noncentrality matrix: $\Phi = MM'/n_1 = \sum_{k=1}^{r} \theta_k \gamma_k \gamma_k'$

[e.g. Cai-Candès-Shen (10), Shabalin-Nobel (13), Rao Nadakuditi (14), Josse-Sardy (16) , Donoho-Gavish(16)]

Regression - Multiple Response (REG) $[{}_1F_1]$

$$\mathbf{y}_\nu = B'\mathbf{x}_\nu + \mathbf{z}_\nu, \qquad \Sigma = \mathrm{Cov}(\mathbf{z}_\nu) \ \text{ unknown}$$

Sums of squares matrices:

$$n_1 H = Y P_H Y' \qquad n_1 \text{ hypothesis d.f.}$$
$$n_2 E = Y P_E Y' \qquad n_2 \text{ error d.f.}$$

Eigenvalues:

$$\det(H - \tilde{\lambda}_i E) = 0 \qquad \text{multivariate } F$$

Low rank: noncentrality

$$\Phi = \Sigma^{-1} M M' / n_1 = \Sigma^{-1} \sum_{k=1}^{r} \theta_k \gamma_k \gamma_k'$$

Signal Detection (SigDet) $[{}_1F_0]$

Data: $\mathbf{x}_\nu = \sum_1^r \sqrt{\theta}_k u_{\nu,k} \gamma_k + \mathbf{z}_\nu$

$$u_{\nu,k} \overset{ind}{\sim} (0,1), \quad \mathrm{Cov}(\mathbf{z}_\nu) = \Sigma$$

Low rank: test $H_0 : \theta = 0$ in

$$\mathrm{Cov}(\mathbf{x}_\nu) = \Phi + \Sigma, \qquad \Phi = \sum_1^r \theta_k \gamma_k \gamma_k'$$

Two samples: $n_1 H = \sum_1^{n_1} \mathbf{x}_\nu \mathbf{x}_\nu' \qquad n_2 E = \sum_{n_1}^{n_1+n_2} \mathbf{z}_\nu \mathbf{z}_\nu'$

Eigenvalues: $\det(H - \tilde{\lambda}_i E) = 0$

Canonical Correlation Analysis (CCA) $[{}_2F_1]$

$$\mathbf{x}_\nu \in \mathbb{R}^p \qquad \mathbf{y}_\nu \in \mathbb{R}^{n_1} \qquad \nu = 1, \ldots, n_1 + n_2 + 1$$

Seek maximally correlated $a'\mathbf{x}_\nu, \; b'\mathbf{y}_\nu$

$$\mathrm{Cov}\begin{pmatrix}\mathbf{x}_\nu \\ \mathbf{y}_\nu\end{pmatrix} = \begin{pmatrix}\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22}\end{pmatrix} \qquad \text{sample} \begin{pmatrix}S_{11} & S_{12} \\ S_{21} & S_{22}\end{pmatrix}$$

Eigenvalues:

$$\det(S_{11}^{-1}S_{12}S_{22}^{-1}S_{21} - \lambda_i I_p) = 0$$

Low rank:

$$\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2} = \Phi^{1/2} = \sum_1^r \sqrt{\theta}_k \gamma_k \eta_k' \overset{e.g.}{=} \begin{bmatrix}\mathrm{diag}(\sqrt{\theta}_1, \ldots, \sqrt{\theta}_r) & 0 \\ 0 & 0\end{bmatrix}$$

Gaussian assumptions

Assume $\mathbf{x}_\nu$, resp $\mathbf{z}_\nu$ are Gaussian (⇒ likelihood ratios)

Why eigenvalues?

Group structure $\implies$ (λ_i) are maximal invariants.

O.K. for low rank alternatives if subspaces are unknown.

Gaussian assumptions

Assume $\mathbf{x}_\nu$, resp $\mathbf{z}_\nu$ are Gaussian $\qquad (\Rightarrow$ likelihood ratios)

Why eigenvalues?

Group structure $\implies (\lambda_i)$ are maximal invariants.

O.K. for low rank alternatives if subspaces are unknown.

[**Wishart definition:** If $\underset{n\times p}{Z} \sim N(M, I \otimes \Sigma)$ is a normal data matrix, then

$$H = Z'Z = \sum_1^n \mathbf{z}_\nu \mathbf{z}_\nu' \sim W_p(n, \Sigma, \Omega),$$

with degrees of freedom n , and non-centrality $\Omega = \Sigma^{-1/2} M' M \Sigma^{-1/2}$]

James' Five Fold Way – and ${}_pF_q$-space

PCA $[{}_0F_0]$

$$n_1 H \sim W_p(n_1, \Sigma_0 + \Phi)$$

REG$_0$ $[{}_0F_1]$

$$n_1 H \sim W_p(n_1, \Sigma_0, n_1 \Phi)$$

SigDet $[{}_1F_0]$

$$n_1 H \sim W_p(n_1, \Sigma + \Phi)$$
$$n_2 E \sim W_p(n_2, \Sigma)$$

REG $[{}_1F_1]$

$$n_1 H \sim W_p(n_1, \Sigma, n_1 \Phi)$$
$$n_2 E \sim W_p(n_2, \Sigma)$$

CCA $[{}_2F_1]$

$$n_1 H \sim W_p(n_1, I - \Phi, \Omega(\Phi))$$
$$n_2 E \sim W_p(n_2, I - \Phi)$$

$\Omega(\Phi)$ random

Symmetric Matrix Denoising (SMD)

$$G = \Phi + Z \qquad \qquad \Phi, Z \ \ \text{symmetric } p \times p$$

$$\sqrt{p} Z_{ij} \overset{ind}{\sim} N(0,1)$$

Symmetric Matrix Denoising (SMD)

$$G = \Phi + Z$$

Φ, Z symmetric $p \times p$

$$\sqrt{p} Z_{ij} \overset{ind}{\sim} \begin{cases} N(0,1) & i \neq j \\ N(0,2) & i = j \end{cases}$$

GOE_p

Symmetric Matrix Denoising (SMD)

$$G = \Phi + Z \qquad \Phi, Z \ \text{symmetric } p \times p$$

$$\sqrt{p} Z_{ij} \stackrel{ind}{\sim} \begin{cases} N(0,1) & i \neq j \\ N(0,2) & i = j \end{cases} \qquad \text{GOE}_p$$

Low rank: $\Phi = \sum_1^r \theta_k \gamma_k \gamma_k'$

Eigenvalues: $\det(G - \lambda_i I_p) = 0$

Symmetric Matrix Denoising (SMD)

$$G = \Phi + Z \qquad\qquad \Phi, Z \quad \text{symmetric } p \times p$$

$$\sqrt{p} Z_{ij} \overset{ind}{\sim} \begin{cases} N(0,1) & i \neq j \\ N(0,2) & i = j \end{cases} \qquad\qquad \mathrm{GOE}_p$$

Low rank: $\Phi = \sum_1^r \theta_k \gamma_k \gamma_k'$

Eigenvalues: $\det(G - \lambda_i I_p) = 0$

Limiting Case: $H_n \sim W_p(n_1, \Sigma_{n_1}) \qquad \Sigma_{n_1} = I_p + \Phi/\sqrt{n_1}.$

For p fixed, PCA $\xrightarrow{n_1 \to \infty}$ SMD:

$$[\ \sqrt{n_1}(H_{n_1}/n_1 - I_p) \overset{\mathcal{D}}{\Rightarrow} \Phi + \sqrt{p} Z, \qquad Z \sim \mathrm{GOE}_p\ ]$$

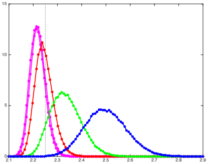
SMD as the limiting "simple" case

$$
\begin{tikzcd}
 & & G = \Phi + Z & & \\
W_p(n_1, \Sigma_0 + n^{-1/2}\Phi) \arrow[urr, "n_1 \to \infty"] & & & & W_p(n_1, \Sigma_0, n^{1/2}\Phi) \arrow[ull, "n_1 \to \infty"'] \\
\begin{matrix} W_p(n_1, \Sigma + n_1^{-1/2}\Phi) \\ W_p(n_2, \Sigma) \end{matrix} \arrow[u, "n_2 \to \infty"] & & & & \begin{matrix} W_p(n_1, \Sigma, n_1^{1/2}\Phi) \\ W_p(n_2, \Sigma) \end{matrix} \arrow[u, "n_2 \to \infty"'] \\
 & & \begin{matrix} W_p(n_1, I - \Phi_n, \Omega(\Phi_n)) \\ W_p(n_2, I - \Phi_n) \\ \Phi_n = n_2^{-1/2}\Phi \end{matrix} \arrow[uuu, "n_2 \to \infty"] & &
\end{tikzcd}
$$

Outline

- High-d Phenomena in PCA
- James' family of multivariate methods
- Likelihood ratio testing in spiked models
  - Gaussian process limits below/above phase transitions
  - confidence limits for spikes
- Common threads

Phase Transition for λ_1 [of H, $E^{-1}H$]



Rank 1: $\Phi = h\gamma\gamma'$ In the 6 cases:

$\exists$ Critical interval $I = [h_-, h_+] \ni 0$ s.t.:

$$h \in I^0, \qquad p^{2/3}(\lambda_1 - b_+) \to \sigma TW$$
$$h \notin I, \qquad p^{1/2}(\lambda_1 - \rho(h)) \to N\left(0, \tau^2(h)\right)$$

$b_+ =$ upper endpoint of spectral distribution ('bulk')

Below h_+: $p^{2/3}$ rate
λ_1 carries no information about h

Above h_+: $p^{1/2}$ rate
$\rho(h) > h$ biased up, $\tau^2(h) \downarrow 0$ as $h \downarrow h_+$.

[many authors]

Below PT: Convergence of Likelihood Ratios

Likelihood ratios below phase transition

For each case, set $\quad p(\lambda; h) = \text{joint density of } \lambda = (\lambda_1, \ldots, \lambda_p).$

$$L_{n.p}\left(h, \lambda\right) := p\left(\lambda; h\right) / p\left(\lambda; 0\right)$$

Likelihood ratios below phase transition

For each case, set $\quad p(\lambda; h) =$ joint density of $\lambda = (\lambda_1, \ldots, \lambda_p)$.

$$L_{n.p}(h, \lambda) := p(\lambda; h) / p(\lambda; 0)$$

Theorem: Under the null ($H_0 : h = 0$),

$$\log L_{n.p}(h, \lambda) \xrightarrow{\mathcal{D}} \mathcal{L}(h) \quad \text{in} \quad \mathcal{C}(h_-, h_+),$$

a Gaussian process with

Likelihood ratios below phase transition

For each case, set $\quad p(\lambda; h) = \text{joint density of } \lambda = (\lambda_1, \ldots, \lambda_p).$

$$L_{n.p}(h, \lambda) := p(\lambda; h) / p(\lambda; 0)$$

Theorem: Under the null $(H_0 : h = 0)$,

$$\log L_{n.p}(h, \lambda) \xrightarrow{\mathcal{D}} \mathcal{L}(h) \quad \text{in} \quad \mathcal{C}(h_-, h_+),$$

a Gaussian process with

$$\begin{aligned} \mu(h) &= \tfrac{1}{4} \log \left[1 - \gamma^2(h)\right] \\ \Gamma(h_1, h_2) &= -\tfrac{1}{2} \log \left[1 - \gamma(h_1)\gamma(h_2)\right] \end{aligned}$$

In particular, $\quad \mu(h) = -\frac{1}{2}\Gamma(h, h)$

$$\Longrightarrow \{\mathbb{P}_{p,h}\}, \{\mathbb{P}_{p,0}\} \text{ mutually contiguous as } p \to \infty$$

Parameters in the six cases

$$\begin{aligned}\mu(h) &= \tfrac{1}{4}\log\left[1-\gamma^2(h)\right] \\ \Gamma(h_1,h_2) &= -\tfrac{1}{2}\log\left[1-\gamma(h_1)\gamma(h_2)\right]\end{aligned}$$

Cases	limit	$\gamma(h)$	h_+
G : SMD	$p \to \infty$	h	1
L: PCA, REG_0	$p/n_1 \to c_1$	$h/\sqrt{c_1}$	$\sqrt{c_1}$
J: REG, SigDet, CCA	$p/n_1 \to c_1$ $p/n_2 \to c_2$	$rh/(c_1+c_2+c_2 h)$	$(r+c_2)/(1-c_1)$

$$[\ r = \sqrt{c_1+c_2-c_1c_2}\ ]$$

Asymptotic power envelopes

- Neyman-Pearson lemma: best test against point alternative $h = \bar{h}$ rejects $H_0 : h = 0$ for $T_p = \log \frac{d\mathbb{P}_{p,\bar{h}}}{d\mathbb{P}_{p,0}}$ large.
- Contiguity + Le Cam's 3rd lemma $\Rightarrow$ law under alternative $\bar{h}$, with $\sigma^2(h) = \Gamma(h, h)$:

$$\log \frac{d\mathbb{P}_{p,\bar{h}}}{d\mathbb{P}_{p,0}} \quad \xrightarrow{\mathbb{P}_{p,\bar{h}}} \quad N\left(\tfrac{1}{2}\sigma^2(\bar{h}), \sigma^2(\bar{h})\right)$$

- $\Rightarrow$ asymptotic Power Envelope (PE) for one-sided $h > 0$:

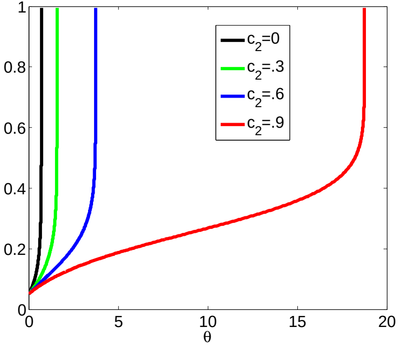
$$\mathrm{PE}(h) = 1 - \Phi\left[\Phi^{-1}(1-\alpha) - \sigma(h)\right],$$

[α = size, Φ = standard normal cdf]

Numerical illustration: REG, SigDet

For e.g. $c_1 = 0.5$, $\quad h_+ = \left(c_2 + \sqrt{0.5 + 0.5c_2}\right) / (1 - c_2)$

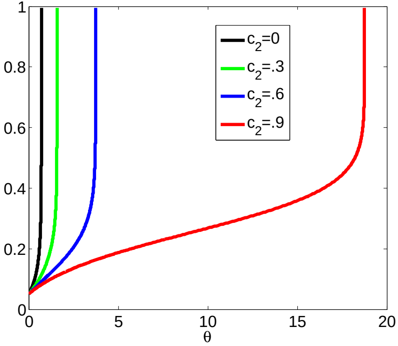
$\Rightarrow$ power envelopes:



Numerical illustration: REG, SigDet

For e.g. $c_1 = 0.5$, $\qquad h_+ = \left(c_2 + \sqrt{0.5 + 0.5c_2}\right) / (1 - c_2)$

$\Rightarrow$ power envelopes:



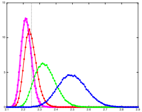
Testing below phase transition: (i.e. weak signals)

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

[Samuel Johnson]

Above PT: Local Asymptotic Normality

Upward bias in λ_1

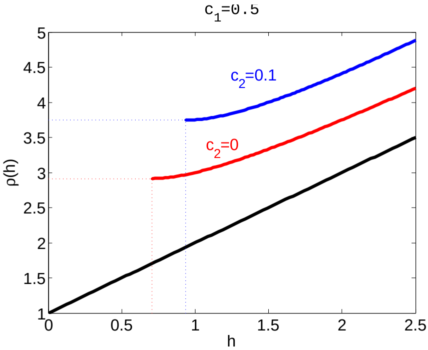


First order - limiting bias:

$$\lambda_1 \overset{a.s.}{\rightarrow} \rho(h; c_1, c_2) > h + 1$$

Bias larger for even **small** $c_2 > 0$!

$$[c_i = p/n_i]$$



$$\left[\rho(h) = \frac{(h + c_1)(h+1)}{(1-c_2)h - c_2} \overset{c_2 \to 0}{\rightarrow} \frac{h + c_1}{h}(h+1) \qquad h_+ = \frac{r + c_2}{1 - c_2} \rightarrow \sqrt{c_1}\right]$$

Gaussian limit for λ_1

For $p/n_i \to c_i$ and $h > h_+$, [all cases exc. CCA]

$$\sqrt{p}[\,\lambda_1 - \rho_p(h)\,] \xrightarrow{\mathcal{D}} N(\,0, \tau^2(h)\,).$$

Structure of variance:

$$\tau^2(h) = \alpha_1(h)\rho'(h)$$

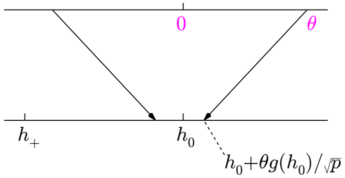
$$\rho'(h) = \alpha_2(h)(h - h_-)(h - h_+)$$

[SMD, PCA: Paul (07), Onatski, Benaych-Georges-Maida (11),
SigDet, REG: DJO (14), Non-Gaussian: Wang-Yao (15)]

Likelihood ratios above phase transition

For $h > h_+$, use local alternatives

$$h = h_0 + g(h_0)\theta/\sqrt{p}:$$

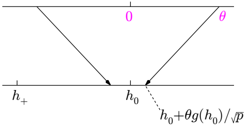


Theorem: (Quadratic approx). If $c_p = (p/n_1, p/n_2) \to (c_1, c_2)$,

$$\log L_{n,p}(\theta, \lambda) = \theta\sqrt{p}\,[\lambda_1 - \rho(h_0, c_p)] - \tfrac{1}{2}\theta^2\tau^2(h_0) + o_{\mathrm{P}}(1).$$

- likelihood ratio depends only on largest λ_1
- all cases except CCA, explicit $g(h), \rho(h), \tau(h)$.

Convergence of experiments



$$\log L_{n,p}(\theta,\lambda) = \theta\sqrt{p}\,[\lambda_1 - \rho(h,c_p)] - \tfrac{1}{2}\theta^2\tau^2(h) + o_{\mathrm{P}}(1).$$

- Convergence to Gaussian limit – shift experiment in θ – depending on $\rho(h)$ and $\tau(h)$:

$$\begin{aligned}
\mathcal{E}_{p,h} &= \left\{(\lambda_1,\dots,\lambda_p) \sim \mathbb{P}_{h+\theta g(h)/\sqrt{p},p}, \quad \theta \in \mathbb{R}\right\} \\
\to \quad \mathcal{E}_h &= \left\{ \quad Y \sim N\left(\theta\tau^2(h), \tau^2(h)\right), \quad \theta \in \mathbb{R}\right\}
\end{aligned}$$

with $$Y \overset{Asy}{\sim} \sqrt{p}\,[\lambda_1 - \rho(h,c_p)]$$

- best tests in supercritical regime use λ_1 in rank one case.

Illustration: LAN Confidence intervals for h

$$\begin{aligned}\text{Lik. Ratio C.I.} &= \{h' : H_0 : h = h' \text{ does not reject in } \mathcal{E}_{p,h'}\}\\ &\approx \{h' : H_0 : \theta = 0 \text{ does not reject in } \mathcal{E}_{h'}\}\end{aligned}$$

$\Rightarrow$ Approx. $100(1-\alpha)\%$ CI: $(\hat{h}^-, \hat{h}^+)$, by solving

$$\rho(\hat{h}^\pm) \mp z_\alpha \tau(\hat{h}^\pm)/\sqrt{p} = \lambda_{1p}.$$

Illustration: LAN Confidence intervals for h

$$\text{Lik. Ratio C.I.} = \{h' : H_0 : h = h' \text{ does not reject in } \mathcal{E}_{p,h'}\}$$
$$\approx \{h' : H_0 : \theta = 0 \text{ does not reject in } \mathcal{E}_{h'}\}$$

$\Rightarrow$ Approx. $100(1-\alpha)\%$ CI: $(\hat{h}^-, \hat{h}^+)$, by solving

$$\rho(\hat{h}^\pm) \mp z_\alpha \tau(\hat{h}^\pm)/\sqrt{p} = \lambda_{1p}.$$

Coverage probabilities, nominal 95% intervals

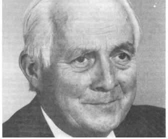
	LAN	Basic	Percentile	BCa
$c_2 = 0, n_1 = p = 100, h + 1 = 10$	94.5	83.6	91.5	86.3
$n_1 = n_2 = 100, p = 50, h = 15$	94.0	~ 0	~ 0	$\times$
$n_1 = n_2 = 100, p = 10, h = 10$	95.1	85.5	91.5	92.4
$n_1 = n_2 = 100, p = 2, h = 10$	94.3	89.2	93.6	92.4

[1000 *reps*, 2SE $\approx$ 1.4%]

Outline

- High-d Phenomena in PCA
- James' family of multivariate methods
- Likelihood ratio testing in spiked models
- Common threads
  - joint densities: James' hypergeometric functions
  - integral formula and approximations

Ann. Math. Stat. (1964)



DISTRIBUTIONS OF MATRIX VARIATES AND LATENT ROOTS DERIVED FROM NORMAL SAMPLES[1]

By Alan T. James

Yale University

1. Summary. The paper is largely expository, but some new results are included to round out the paper and bring it up to date.

The following distributions are quoted in Section 7.

1. Type ${}_0F_0$, exponential: (i) χ^2, (ii) Wishart, (iii) latent roots of the covariance matrix.

2. Type ${}_1F_0$, binomial series: (i) variance ratio, F, (ii) latent roots with unequal population covariance matrices.

3. Type ${}_0F_1$, Bessel: (i) noncentral χ^2, (ii) noncentral Wishart, (iii) noncentral means with known covariance.

4. Type ${}_1F_1$, confluent hypergeometric: (i) noncentral F, (ii) noncentral multivariate F, (iii) noncentral latent roots.

5. Type ${}_2F_1$, Gaussian hypergeometric: (i) multiple correlation coefficient, (ii) canonical correlation coefficients.

Common structure

SMD

PCA REG$_0$

SigDet REG

CCA

$\Lambda = \mathrm{diag}(\lambda_i)$ eigenvalues of G, H or $E^{-1}H$.

$\Phi = \Gamma\Theta\Gamma'$ low rank alternative, $\Theta = \mathrm{diag}(\theta_1, \ldots, \theta_r)$.

Common structure

SMD

PCA REG$_0$

SigDet REG

CCA

$\Lambda = \mathrm{diag}(\lambda_i)$ eigenvalues of G, H or $E^{-1}H$.

$\Phi = \Gamma\Theta\Gamma'$ low rank alternative, $\Theta = \mathrm{diag}(\theta_1, \ldots, \theta_r)$.

Joint density of eigenvalues in the six cases: with $\Psi = \Psi(\Theta)$,

$$p(\Lambda; \Theta) \propto \rho(\Psi)^{\alpha} \cdot {}_{\mathrm{p}}F_{\mathrm{q}}(a, b; c\Psi, \Lambda)\, \pi(\Lambda)\Delta(\Lambda)$$

(after James, (64))

The real win for James(64): large p

$$p(\Lambda;\Theta) \propto \rho(\Psi)^{\alpha} \cdot {}_{\mathrm{p}}F_{\mathrm{q}}(a, b; c\Psi, \Lambda)\, \pi(\Lambda)\Delta(\Lambda)$$

- $\pi(\Lambda)\Delta(\Lambda)$
  - null hypothesis distributions: $\Theta = 0 \Rightarrow \Psi = 0$ and $\boxed{{}_{\mathrm{p}}F_{\mathrm{q}} = 1.}$
  - large $(p, n_i) \Rightarrow$ 'bulk' laws of RMT
- $\rho(\Psi)^{\alpha} \cdot {}_{\mathrm{p}}F_{\mathrm{q}}(a, b; c\Psi, \Lambda)$
  - finite rank departure from null in $\Psi = \Psi(\Theta)$
  - large $(p, n_i) \Rightarrow$ seek informative **approximation**

Null Hypothesis: Links to RMT

$$H_0: \ p_0(\Lambda) = \pi(\Lambda)\Delta(\Lambda) \qquad F_p(\lambda) = p^{-1}\#\{i : \lambda_i \le \lambda\} \xrightarrow{a.s.} F(\lambda)$$

weight	$\pi(\lambda)$	Spectral Law $F(\lambda)$
Gaussian [SMD]	$e^{-\lambda^2/2}$	Semi-circle
Laguerre [PCA REG_0]	$\lambda^\alpha e^{-\lambda/2}$	Marcenko-Pastur
Jacobi [SigDet REG CCA]	$\lambda^a(1-\lambda)^b$	Wachter

Alternatives: matrix hypergeometric functions

Scalar: $${}_{\mathrm{p}}F_{\mathrm{q}}(a, b; x) = \sum_{k=0}^{\infty} \frac{(a_1)_k \cdots (a_{\mathrm{p}})_k}{(b_1)_k \cdots (a_{\mathrm{q}})_k} \frac{x^k}{k!}$$

Alternatives: matrix hypergeometric functions

Scalar: $${}_{\mathrm{p}}F_{\mathrm{q}}(a,b;x) = \sum_{k=0}^{\infty} \frac{(a_1)_k \cdots (a_{\mathrm{p}})_k}{(b_1)_k \cdots (a_{\mathrm{q}})_k} \frac{x^k}{k!}$$

Single matrix argument: S symmetric, usually diagonal

$${}_{\mathrm{p}}F_{\mathrm{q}}(a,b;S) = \sum_{k=0}^{\infty} \sum_{\kappa} \frac{(a_1)_\kappa \cdots (a_{\mathrm{p}})_\kappa}{(b_1)_\kappa \cdots (a_{\mathrm{q}})_\kappa} \frac{C_\kappa(S)}{k!}$$

$${}_0F_0(S) = e^{\mathrm{tr} S}, \qquad {}_1F_0(a,S) = |I - S|^{-a}$$

Alternatives: matrix hypergeometric functions

Scalar:
$$_{\mathrm{p}}F_{\mathrm{q}}(a, b; x) = \sum_{k=0}^{\infty} \frac{(a_1)_k \cdots (a_{\mathrm{p}})_k}{(b_1)_k \cdots (a_{\mathrm{q}})_k} \frac{x^k}{k!}$$

Single matrix argument: S symmetric, usually diagonal

$$_{\mathrm{p}}F_{\mathrm{q}}(a, b; S) = \sum_{k=0}^{\infty} \sum_{\kappa} \frac{(a_1)_\kappa \cdots (a_{\mathrm{p}})_\kappa}{(b_1)_\kappa \cdots (a_{\mathrm{q}})_\kappa} \frac{C_\kappa(S)}{k!}$$

$$_0F_0(S) = e^{\mathrm{tr} S}, \qquad {}_1F_0(a, S) = |I - S|^{-a}$$

Two matrix arguments: S, T symmetric

$$_{\mathrm{p}}F_{\mathrm{q}}(a, b; S, T) = \int_{O(p)} {}_{\mathrm{p}}F_{\mathrm{q}}(a, b; SUTU')(dU)$$

Joint Density - parameter table

$$p(\Lambda;\Theta) \;\propto\; \rho(\Psi)^{\alpha}\,{}_{\mathrm{p}}F_{\mathrm{q}}(a,b;c\Psi,\Lambda)\;\pi(\Lambda)\Delta(\Lambda)$$

		$2a$	$2b$	2α	$\Psi(\theta)$	$\rho(\Psi)$
SMD	${}_0F_0$	.	.	p	θ	$e^{-\frac{1}{2}\mathrm{tr}\Psi^2}$
PCA	${}_0F_0$	.	.	n_1	$\theta/(1+\theta)$	$\lvert I-\Psi\rvert$
SigDet	${}_1F_0$	n	.	n_1	$\theta/(1+\theta)$	$\lvert I-\Psi\rvert$
REG_0	${}_0F_1$	.	n_1	n_1	θ	$e^{-\mathrm{tr}\Psi}$
REG	${}_1F_1$	n	n_1	n_1	θ	$e^{-\mathrm{tr}\Psi}$
CCA	${}_2F_1$	n,n	n_1	n	$\theta/I(\theta)$	$\lvert I-\Psi\rvert$

$$n = n_1 + n_2, \qquad c = \left(\frac{n_1}{2}\right)^{\mathrm{q}+1}\left(\frac{2}{n_2}\right)^{\mathrm{p}}$$

Joint Density - parameter table

$$p(\Lambda;\Theta) \;\propto\; \rho(\Psi)^{\alpha}\,{}_{\mathrm{p}}F_{\mathrm{q}}(a,b;c\Psi,\Lambda)\;\pi(\Lambda)\Delta(\Lambda)$$

		$2a$	$2b$	2α	$\Psi(\theta)$	$\rho(\Psi)$
SMD	${}_0F_0$	.	.	p	θ	$e^{-\frac{1}{2}\mathrm{tr}\Psi^2}$
PCA	${}_0F_0$	.	.	n_1	$\theta/(1+\theta)$	$\lvert I-\Psi\rvert$
SigDet	${}_1F_0$	n	.	n_1	$\theta/(1+\theta)$	$\lvert I-\Psi\rvert$
REG_0	${}_0F_1$	.	n_1	n_1	θ	$e^{-\mathrm{tr}\Psi}$
REG	${}_1F_1$	n	n_1	n_1	θ	$e^{-\mathrm{tr}\Psi}$
CCA	${}_2F_1$	n, n	n_1	n	$\theta/I(\theta)$	$\lvert I-\Psi\rvert$

Going beyond James:

- Ψ is low rank [often still 1]
- high dimension: $p/n_1 \to c_1 > 0, \quad p/n_2 \to c_2 \in [0,1)$

Unifying Tool: An Integral to reduce dimension

Assume $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_r)$, $\quad m = \frac{r}{2} - 1$, and

$\Psi = \mathrm{diag}(\psi, 0, \ldots, 0)$ has rank one. Then [DJ, 14]

$$\boxed{{}_pF_q(a, b; \Psi, \Lambda) = \frac{c_m}{\psi^m} \frac{1}{2\pi i} \int_K {}_pF_q(a - m, b - m; \psi s) \prod_{i=1}^{r} (s - \lambda_i)^{-1/2} ds}$$

Unifying Tool: An Integral to reduce dimension

Assume $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_r)$, $\quad m = \frac{r}{2} - 1,$ and

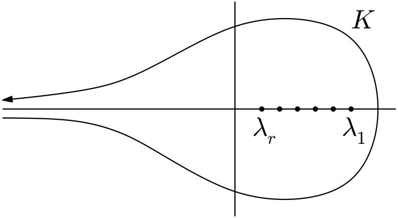
$\Psi = \mathrm{diag}(\psi, 0, \ldots, 0)$ has rank one. Then [DJ, 14]

$$\boxed{{}_{\mathrm{p}}F_{\mathrm{q}}(a, b; \Psi, \Lambda) = \frac{c_m}{\psi^m} \frac{1}{2\pi i} \int_K {}_{\mathrm{p}}F_{\mathrm{q}}(a - m, b - m; \psi s) \prod_{i=1}^{r} (s - \lambda_i)^{-1/2} ds}$$

Univariate integral and univariate ${}_{\mathrm{p}}F_{\mathrm{q}}$!

$$c_m = \frac{\Gamma(m+1)}{\rho_m(a-m, b-m)}$$

$$\rho_k(a, b) = \frac{(a_1)_k \cdots (a_{\mathrm{p}})_k}{(b_1)_k \cdots (a_{\mathrm{q}})_k}$$



rank$(\Psi) = r$: r-fold contour integral Passemier-McKay-Chen (14).

Use of the integral

Study likelihood ratio (LR)

$$\frac{p(\Lambda, \Theta)}{p(\Lambda, 0)} = \rho(\alpha; \Psi) \cdot {}_{\mathrm{p}}F_{\mathrm{q}}(a, b; c\Psi, \Lambda) = \text{integral}$$

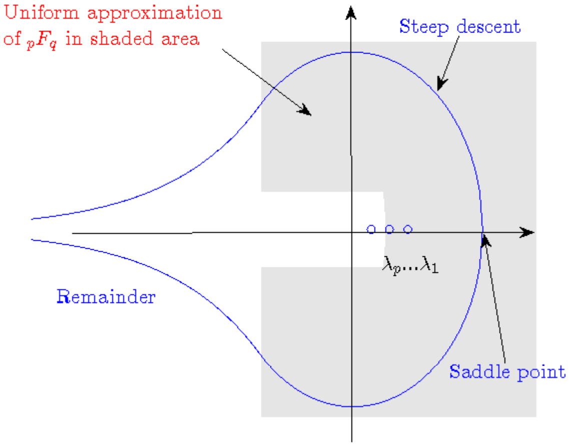
under $H_0 : \Theta = 0$, in double scaling limit, $\quad p/n_i \to c_i$.

- uniform Laplace approximations to integral
- apply functional eigenvalue CLT to result

Laplace approximation step - below PT

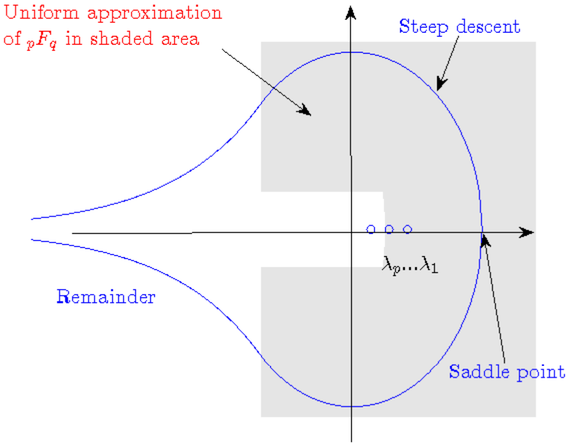
For each of the six cases, approximate

$$L_{n,p}(\theta,\Lambda) = \frac{\rho(\alpha;\Psi)\,c_m}{\psi^m 2\pi \mathrm{i}} \int_K {}_{\mathrm{p}}F_{\mathrm{q}}\left(a-m, b-m; c\psi s\right) \prod_{i=1}^{p} (s-\lambda_i)^{-\frac{1}{2}}\,\mathrm{d}s,$$



Laplace approximation step - below PT

$_0F_0$ and $_1F_0$ have explicit form
Uniform approximation for $_0F_1$ follows from Olver (1954)
For $_1F_1$, derive from Pochhammer's representation
For $_2F_1$, extend point-wise analysis of Paris (2013)



CLT step - below PT

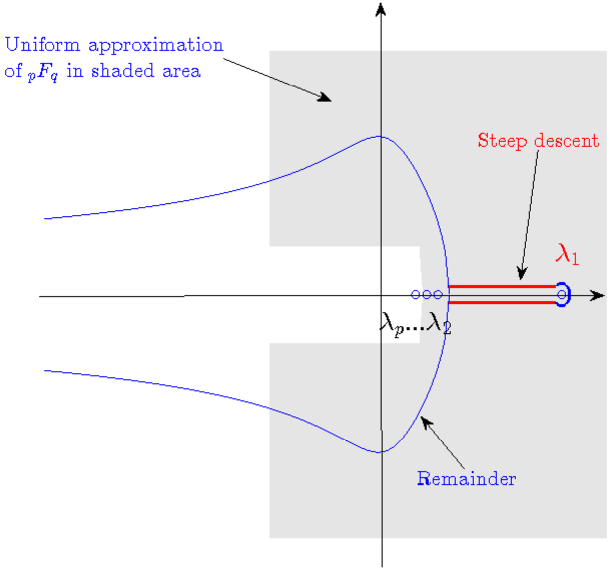
The Laplace approximations imply that

$$\log L_{n,p}(\theta, \Lambda) \overset{Asy}{\sim} \Delta_p(\theta) = \sum_i \left[h_\theta(\lambda_i) - \int h_\theta dF \right]$$

$$h_\theta(\lambda) = \log[z_0(\theta) - \lambda]$$

- linear statistic involves all eigenvalues
- Use CLTs from Bai and Silverstein (2004), Zheng (2012), and Young and Pan (2012) to obtain weak convergence of the finite dimensional distributions
- Establish tightness

Laplace approximation step - above PT



Conclusion

James' representation using ${}_pF_q(a, b; c\Psi, \Lambda)$ yields

- simple approximations in large p, low rank cases, e.g.:
- Local Asymptotic Normality of super-critical experiments
- asymptotic power envelopes in the sub-critical regime

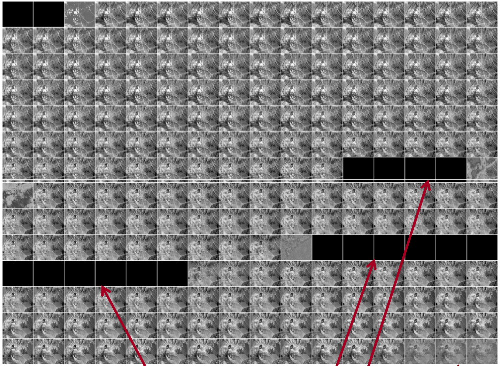
THANK YOU!

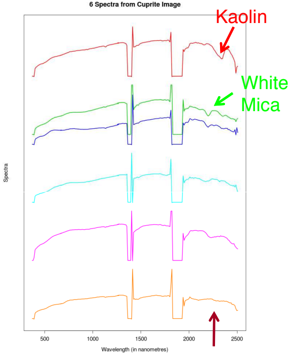
[Current versions: *arXiv* 1509.07269, 1411.3875]

Limiting case example

Hyperspectral image example: Cuprite, Nevada

(224 AVIRIS images, [370, 2507] nm, ~ 9.6 nm apart, 614 x 512 (= 314,368) pixels, atmospherically corrected)





Water absorption bands

Noisy bands

Most mineral diagnostic information for minerals in [2000, 2500]

Of Fables, Fairy Tales and Statistical Theory

FABLE: "2. A short story devised to convey some useful lesson"
FAIRY TALE: "any of various short tales having folkloric elements and featuring fantastic or magical events or characters."
...deriving from long traditions of oral storytelling