Testing hypotheses about sub- and super-critical spikes in multivariate statistical models

Alexei Onatski
joint with Iain Johnstone

Economics, Cambridge, and Statistics, Stanford

Hong Kong, January 2015

The null and alternative hypotheses

H_0: no spikes in the covariance or non-centrality

that is, $\Phi = 0$ in

0. SMD: $G \sim \Phi + GOE_p$
1. PCA: $H \sim W_p\left(n, \Sigma_0 + \Sigma_0^{1/2}\Phi\Sigma_0^{1/2}\right)$
2. REG$_0$: $H \sim W_p\left(n, \Sigma_0, n\Phi\right)$
3. SigDet: $H \sim W_p\left(n_1, \Sigma + \Sigma^{1/2}\Phi\Sigma^{1/2}\right), E \sim W_p\left(n_2, \Sigma\right)$
4. REG: $H \sim W_p\left(n_1, \Sigma, n_1\Phi\right), E \sim W_p\left(n_2, \Sigma\right)$
5. CCA: $H \sim W_p\left(n_1, \Sigma, n_1\Phi\right), E \sim W_p\left(n_2, \Sigma\right)$, Φ random

H_1: there are spikes

that is, $\Phi = \sum_{k=1}^{r} \theta_k \gamma_k \gamma_k'$ with some $\theta_k \neq 0$

The invariance

- Distributions of H and E are invariant w.r.t. a rich group of transformations, after reparametrization
- E.g. for REG, consider $B \in \mathcal{GL}(p)$, transformations

$$H \mapsto \tilde{H} = BHB', E \mapsto \tilde{E} = BEB',$$

  and reparametrization

$$\Sigma \mapsto \tilde{\Sigma} = B\Sigma B', \Phi \mapsto \tilde{\Phi} = \left(\tilde{\Sigma}^{-1/2} B \Sigma^{1/2}\right) \Phi \left(\tilde{\Sigma}^{-1/2} B \Sigma^{1/2}\right)^{-1}$$

- If Σ and eigenvectors γ_k of Φ are completely unknown, it is desirable to test H_0 against H_1 using the maximal invariant statistic, given by the roots of

$$\det\left(H/n_1 - \lambda E/n_2\right) = 0$$

Our approach

Study the double scaling asymptotic behavior of the likelihood ratio $\frac{p(\lambda,\Theta)}{p(\lambda,0)}$ under the null.

- The likelihood ratio (LR) is a sufficient statistic for $\Theta = diag\{\theta_1, ..., \theta_r\}$
- If the joint null asymptotic distribution of the LR and any statistic T is Gaussian, simply use Le Cam's 3rd lemma to get distribution of T under alternative $\Longrightarrow$ asymptotic power
- Neyman-Pearson Lemma $\Rightarrow$ best tests against point alternatives, power envelopes
- Can use the convergence of experiments theory to potentially obtain risk bounds for various statistical decision problems

It turns out that the LR behavior depends a lot on whether the spikes θ_k are sub- or super-critical.

Outline

Phase transition

- A brief review of some known results
- Asymptotic normality of the largest eigenvalues in the super-critical regime

Likelihood ratios below & above the transition

- LR convergence to a Gaussian process in the sub-critical regime
- Asymptotic power analysis in the sub-critical regime
- The Local Asymptotic Normality in the super-critical regime $\Longrightarrow$ the best tests for a super-critical spike depend only on λ_1

Review: Phase Transition for Largest Eigenvalue [of H]

Rank 1: $\Phi = h\gamma\gamma'$ $\exists$ Critical interval $I = [h_-, h_+] \ni 0$ s.t.:

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

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 $h \downarrow h_+$.

Recall: Bulk Distribution (Wachter)

Spectral density of limit $F(d\lambda) = \lim p^{-1} \sum_i \delta_{\lambda_i}$, where λ_i is the i-th largest eigenvalue of $(E/n_2)^{-1}(H/n_1)$:

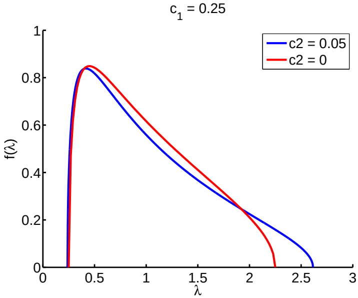
$$f(\lambda) = \frac{1-c_2}{2\pi} \frac{\sqrt{(b_+ - \lambda)(\lambda - b_-)}}{\lambda(c_1 + c_2\lambda)}$$

Let $r = \sqrt{c_1 + c_2 - c_1 c_2}$.

Support limits:

$$b_\pm = \left(\frac{1 \pm r}{1 - c_2}\right)^2 \to (1 \pm \sqrt{c_1})^2$$

as $c_2 \to 0$. [Marchenko-Pastur]



First Order Behavior of the Largest Eigenvalue

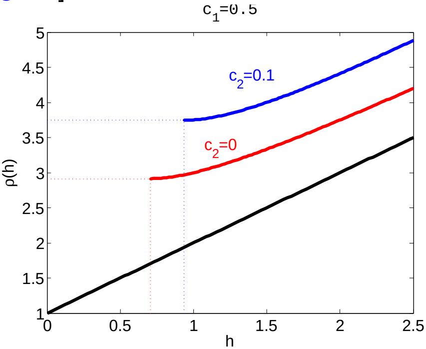
Upward Bias: for $h > h_+$:

$$\lambda_1 \xrightarrow{a.s.} \rho(h) = \frac{(h + c_1)(h+1)}{(1 - c_2)h - c_2} \xrightarrow{c_2 \to 0} \frac{h + c_1}{h}(h+1)$$

[Nadakuditi-Silverstein, 2010, for SigDet]

Location of threshold:

$$h_+ = \frac{r + c_2}{1 - c_2} \to \sqrt{c_1}$$



Phase transition and bias - parameter table

	b_+	h_+	$\rho(h)$
G: SMD	2	1	$h + 1/h$
L: PCA, REG_0	$(1+\sqrt{c})^2$	$\sqrt{c}$	$(1+h)\,\frac{c+h}{h}$
J: SigDet, REG	$\left(\frac{1+r}{1-c_2}\right)^2$	$\frac{c_2+r}{1-c_2}$	$(1+h)\frac{c_1+h}{(1-c_2)h-c_2}$

For CCA, we are concerned with eigenvalues of $S_{11}^{-1}S_{12}S_{22}^{-1}S_{21}$, which are functions of those of $(E/n_2)^{-1}(H/n_1)$. Also, our parameterization relates to $\Phi = \Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1/2}$ rather than to the random noncentrality Φ of H. To avoid confusion, we do not report CCA case in the table [see Bao, Hu, Pan, and Zhou, 2014, for this case].

Outline

Phase transition

- A brief review of some known results
- Asymptotic normality of the largest eigenvalues in the super-critical regime

Likelihood ratios below & above the transition

- LR convergence to a Gaussian process in the sub-critical regime
- Asymptotic power analysis in the sub-critical regime
- The Local Asymptotic Normality in the super-critical regime $\Longrightarrow$ the best tests for a super-critical spike depend only on λ_1

Gaussian limit for λ_1, SigDet

$$\rho\left(h; c_1, c_2\right) = \left(h + c_1\right)\left(h+1\right)/L(h), \qquad L(h) = (1-c_2)h - c_2$$

Actual centering: $\rho_p\left(h\right) = \rho\left(h; p/n_1, p/n_2\right)$

Theorem: For double scaling and $h > h_+$,

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

Structure of variance: $\tau^2(h) = r^2\omega(h)\rho'(h)$ with

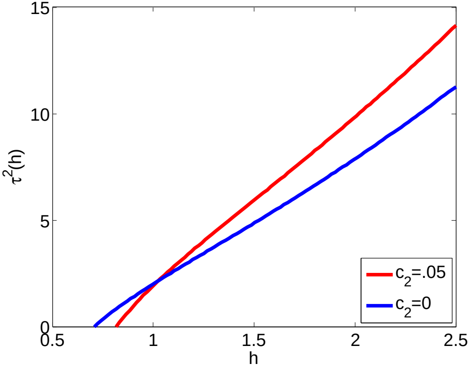
$r^2\omega\left(h\right) = 2r^2\left[h\left(h+1\right)/L\left(h\right)\right]^2$ scale factor in LAN

$\rho'\left(h\right) = \left(1-c_2\right)\left(h-h_-\right)\left(h-h_+\right)/L(h)^2$ zero at h_+

Properties of Variance $\tau^2(h; c_1, c_2)$

Variance inflation due to error d.f. (e.g. at $c_1 = 0.5$):

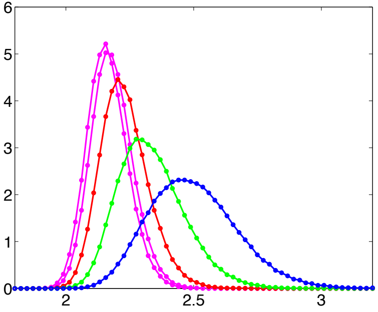
$$VI = \lim_{h\to\infty} \frac{\tau^2(h;c_1,c_2)}{\tau^2(h;c_1,0)} = \frac{r^2}{c_1(1-c_2)^3} = \begin{cases} 1.04 & c_2 = .01 \\ 1.22 & c_2 = .05 \\ 2.34 & c_2 = .20 \end{cases}$$

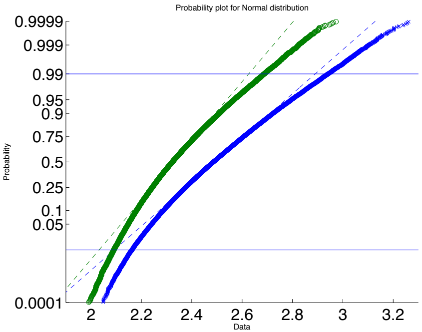


Numerical Illustration

$$p = 50, n_1 = 200 \qquad [\text{i.e. } c_1 = 0.25, c_2 = 0]$$

	subcritical	critical	supercritical
$h =$	0, 0.25,	$h_+ = 0.5$,	0.75, 1.





Asymptotic normality of the super-critical eigenvalue - parameter table

	b_+	$h_\pm$	$\rho(h)$	$\tau^2(h)$
SMD*	2	± 1	$h + 1/h$	$\rho'(h)$
PCA**, SigDet	$\left(\frac{1+r}{1-c_2}\right)^2$	$\frac{c_2 \pm r}{1-c_2}$	$(1+h)\frac{c_1+h}{L(h)}$	$r^2\omega(h)\rho'(h)$
REG_0^{**}, REG	–	–	–	$t^2\omega(h)\rho'(h)$

(*) $c_1 = 1, c_2 = 0$
(**) $c_1 = c, c_2 = 0$

$$\text{Where } \rho'(h) = (1-c_2)(h-h_-)(h-h_+)/L(h)^2,$$

$L(h) = (1-c_2)h - c_2,$ $\quad\omega(h) = 2\left[h(h+1)/L(h)\right]^2,$

$r^2 = c_1 + c_2 - c_1c_2,$ $\quad t^2 = c_1 + c_2 - c_1\frac{h^2-c_1}{(h+1)^2}$

Finite rank case

Suppose $h_1 > ... > h_m > h_+ > h_{m+1} > ... > h_k$

First order:

$$\lambda_i \xrightarrow{a.s.} \begin{cases} \rho\left(h_i; c_1, c_2\right) & h_i > h_+ \\ b_+ & h_i < h_+ \end{cases}$$

Second order, above the threshold:

Let $\lambda = (\lambda_1, ..., \lambda_m)$ $\qquad h = (h_1, ..., h_m)$
Then

$$\sqrt{p}\left(\lambda - \rho\left(h; p/n_1, p/n_2\right)\right) \xrightarrow{d} N_m\left(0, \mathrm{diag}\left(\tau^2\left(h\right)\right)\right)$$

- asymptotically independent.
- True for SMD, PCA, REG_0, REG, and SigDet. Remains to be established for CCA.

Outline

Phase transition

- A brief review of some known results
- Asymptotic normality of the largest eigenvalues in the super-critical regime

Likelihood ratios below & above the transition

- LR convergence to a Gaussian process in the sub-critical regime
- Asymptotic power analysis in the sub-critical regime
- The Local Asymptotic Normality in the super-critical regime $\Longrightarrow$ the best tests for a super-critical spike depend only on λ_1

Likelihood ratios below phase transition

For each of the six cases, define

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

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

$$\log L_{n.p}(\theta, \lambda) \Longrightarrow \mathcal{L}(\theta) \text{ in } \mathcal{C}(h_-, h_+),$$

a Gaussian process with

$$\begin{aligned} \mu(\theta) &= \frac{1}{4} \log\left[1 - \gamma^2(\theta)\right] \\ \Gamma(\theta_1, \theta_2) &= -\frac{1}{2} \log\left[1 - \gamma(\theta_1)\gamma(\theta_2)\right] \end{aligned}$$

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

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

Parameters in the six cases

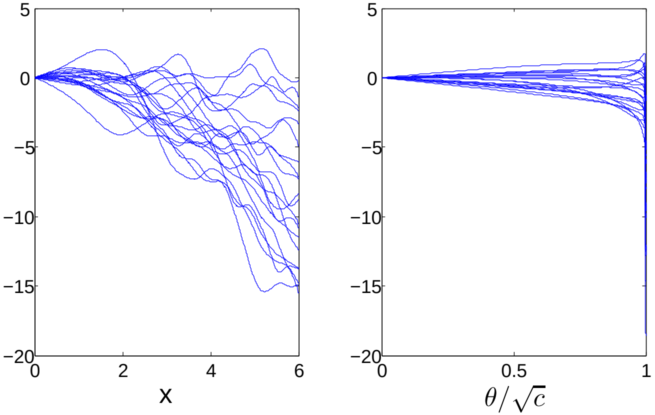
$$\mu(\theta) = \frac{1}{4}\log\left[1-\gamma^2(\theta)\right]$$
$$\Gamma(\theta_1,\theta_2) = -\frac{1}{2}\log\left[1-\gamma(\theta_1)\gamma(\theta_2)\right]$$

Cases	limit	$\gamma(\theta)$
G : SMD	$p \longrightarrow \infty$	θ
L: PCA, REG_0	$p/n \longrightarrow c$	$\theta/\sqrt{c}$
J: REG, SigDet, CCA	$p/n_1 \longrightarrow c_1$ $p/n_2 \longrightarrow c_2$	$r\theta/\left(c_1+c_2+c_2\theta\right)$

Numerical illustration: PCA, REG_0

$$\text{Let } x = \sqrt{-\log\left(1 - \theta^2/c\right)} \quad \Longrightarrow \quad x \to \infty \text{ as } \theta \to h_+ = \sqrt{c}$$

Here are 20 realizations of the limiting LR process,
under x- and θ-parameterization



Main tool - contour integral representation

Recall James' (1964) representation [Iain's Conclusion slide]

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

$\Longrightarrow$

$$L_{n,p}(\theta,\lambda) = \rho(\alpha;\Psi)\, {}_{\mathrm{p}}F_{\mathrm{q}}(a,b;c\Psi,\Lambda)\ ,$$

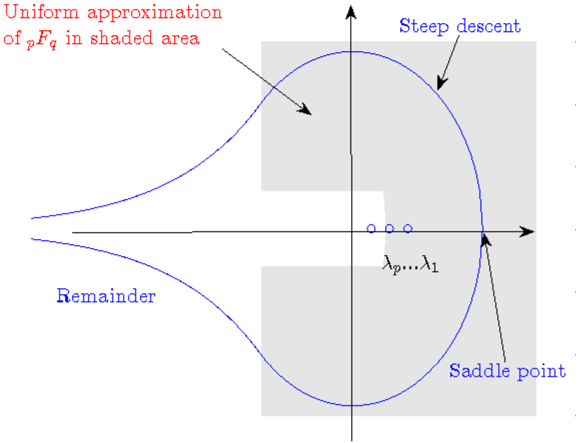
where $\Lambda = \mathrm{diag}(\lambda)$, $\Psi = \mathrm{diag}(\psi(\theta),0,\ldots,0)$, and $\alpha, a, b,$ and c depend on the case (in all cases, $\alpha, a, b \to \infty$).

Using the contour representation of ${}_{\mathrm{p}}F_{\mathrm{q}}(a,b;c\Psi,\Lambda)$, we get, for $m = p/2-1$,

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

where $c_m = \Gamma(m+1)\frac{(b)_m}{(a)_m}$.

Laplace approximation step



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

CLT step

The Laplace approximations imply that

$$L_{n,p}(\theta, \lambda) \overset{Asy}{\sim} \text{linear spectral statistic that depends on } \theta$$

- 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

Outline

Phase transition

- A brief review of some known results
- Asymptotic normality of the largest eigenvalues in the super-critical regime

Likelihood ratios below & above the transition

- LR convergence to a Gaussian process in the sub-critical regime
- Asymptotic power analysis in the sub-critical regime
- The Local Asymptotic Normality in the super-critical regime $\Longrightarrow$ the best tests for a super-critical spike depend only on λ_1

Computing power of tests of $H_0 : \theta = 0$

Let T_p be a test statistic for H_0 – likelihood ratio, corrected LR, other statistic...

Simply use Le Cam's 3rd lemma: If

$$\left(T_p, \log \frac{d\mathbb{P}_{p,\theta}}{d\mathbb{P}_{p,0}}\right) \overset{\mathbb{P}_{p,0}}{\Longrightarrow} N_{\dim(T)+1}\left[\begin{pmatrix} \mu_T \\ -\sigma^2/2 \end{pmatrix}, \begin{pmatrix} \Sigma & \tau \\ \tau & \sigma^2 \end{pmatrix}\right]$$

then

$$T_p \overset{\mathbb{P}_{p,\theta}}{\Longrightarrow} N_{\dim(T)}\left(\mu_T + \tau, \Sigma\right)$$

Asymptotic power envelopes

- By Neyman-Pearson lemma, the best test against point alternative $\theta = \bar{\theta}$ rejects the null when $T_p = \log \frac{d\mathbb{P}_{p,\bar{\theta}}}{d\mathbb{P}_{p,0}}$ is sufficiently large.
- By Le Cam's 3rd lemma,

$$\log \frac{d\mathbb{P}_{p,\bar{\theta}}}{d\mathbb{P}_{p,0}} \overset{\mathbb{P}_{p,\bar{\theta}}}{\Longrightarrow} N\left(-\frac{1}{4}\log\left[1-\gamma^2\left(\bar{\theta}\right)\right], -\frac{1}{2}\log\left[1-\gamma^2\left(\bar{\theta}\right)\right]\right)$$

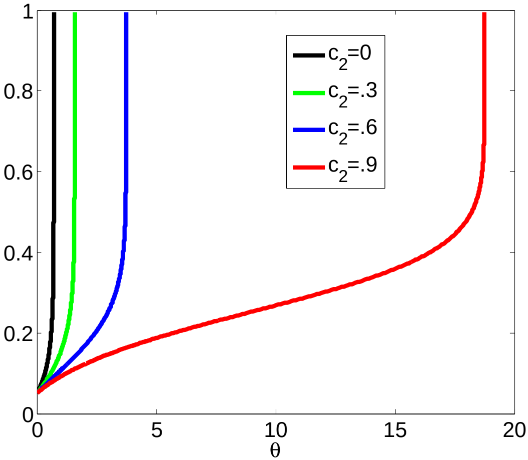
- Therefore, the asymptotic Power Envelope (PE) for one-sided alternative $\theta > 0$ is

$$\mathrm{PE}(\theta) = 1 - \Phi\left[\Phi^{-1}(1-\alpha) - \sqrt{-\frac{1}{2}\ln\left(1-\gamma^2(\theta)\right)}\right],$$

where α is the asymptotic side and Φ is the standard normal cdf

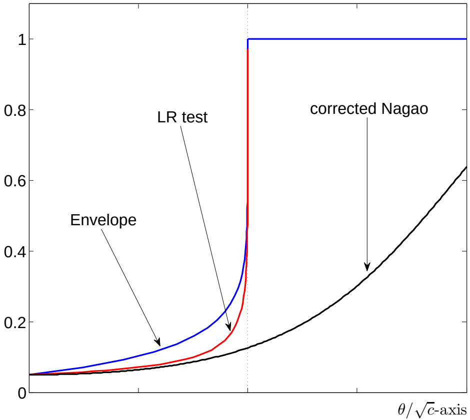
Numerical illustration: REG, SigDet

For $c_1 = 0.5$, we have $\theta_+ = \left(c_2 + \sqrt{0.5 + 0.5c_2}\right) / \left(1 - c_2\right)$ and power envelopes:



Two tests of $\Sigma = I$ (PCA)

- Test based on the corrected Nagao statistic
  $W = \frac{1}{p}\operatorname{tr}\left(\hat{\Sigma} - I\right)^2 - \frac{p}{n}\left[\frac{1}{p}\operatorname{tr}\hat{\Sigma}\right]^2 + \frac{p}{n}$ [Ledoit and Wolf, 02]
- The LR test based on the maximal invariant statistic



Outline

Phase transition

- A brief review of some known results
- Asymptotic normality of the largest eigenvalues in the super-critical regime

Likelihood ratios below & above the transition

- LR convergence to a Gaussian process in the sub-critical regime
- Asymptotic power analysis in the sub-critical regime
- The Local Asymptotic Normality in the super-critical regime $\Longrightarrow$ the best tests for a super-critical spike depend only on λ_1

Likelihood ratios above phase transition

For $\theta > h_+$, $\{\mathbb{P}_{p,\theta}\}, \{\mathbb{P}_{p,0}\}$ mutually singular as $p \to \infty$
Consider local alternatives $h = h_0 + g(h_0)\theta/\sqrt{p}$:

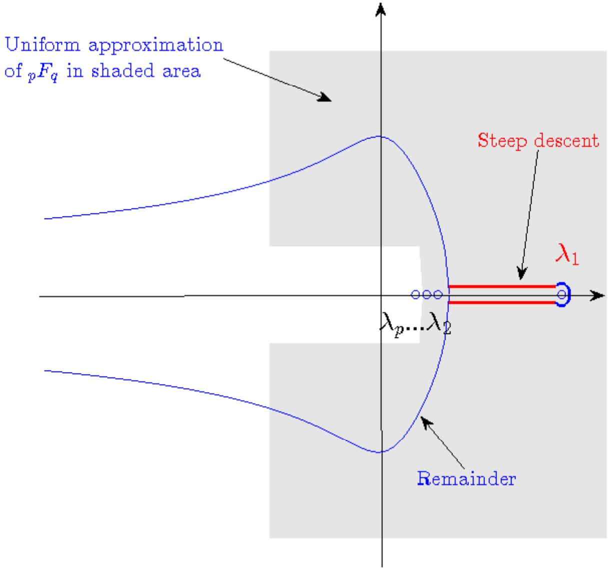
$$L_{n,p}(\theta, \lambda) = \frac{p\left(\lambda, h_0 + g(h_0)\theta/\sqrt{p}\right)}{p(\lambda, h_0)}$$

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

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

- likelihood ratio depends only on largest λ_1
- $g(h_0) = \begin{cases} r^2\omega(h_0) & \text{for PCA,SigDet} \\ t^2\omega(h_0) & \text{for REG}_0\text{,REG} \end{cases}$.

Laplace approximation step



Convergence of experiments

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

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

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

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

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

Illustration: LAN Confidence intervals for h

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

$\mathcal{E}_{h'}$ is a Gaussian shift experiment based on $\sqrt{p}\left[\lambda_1 - \rho\left(h'\right)\right]$,
$\Longrightarrow$ Approx. $100\left(1-\alpha\right)\%$ CI: $\left(\hat{h}^-, \hat{h}^+\right)$, by solving

$$\rho\left(\hat{h}^\pm\right) \mp z_{\alpha/2}\tau\left(\hat{h}^\pm\right)/\sqrt{p} = \lambda_1 .$$

Coverage probabilities, nominal 95% intervals

	LAN	Basic	Percentile	BCa
$c_2 = 0, n_1 = p = 100, h = 10$	98.3	61.3	97.5	67.8
$n_1 = n_2 = 100, p = 50, h = 15$	96.8	~ 0	~ 0	$\times$
$n_1 = n_2 = 100, p = 5, h = 10$	95.8	77.4	94.1	87.2
$n_1 = n_2 = 100, p = 2, h = 10$	95.3	77.3	91.2	89.7

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

Convergence of experiments below phase transition

Theorem: Consider PCA case with $\Sigma_0 = I$.

$$\begin{aligned} \mathcal{E}_{p,h} &= \left\{ (\lambda_1, ..., \lambda_p) \sim \mathbb{P}_{h,p}, h \in \left(0, \sqrt{c}\right) \right\} \\ \rightarrow \quad & \mathcal{E}_h \left\{ \{Y_j\}_{j=1}^{\infty}, Y_j \sim \mathrm{i.d.} N\left(h^j / \sqrt{2jc^j}, 1 \right), h \in \left(0, \sqrt{c}\right) \right\} \end{aligned}$$

a Gaussian sequence experiment with

$$\sqrt{2jc_p^j} Y_j \overset{Asy}{\sim} \sum_{i=1}^{p} \Gamma_j^{c_p}(\lambda_i) - \sqrt{c_p^j}\left(1 + (-1)^j\right)/2$$

Here $\Gamma_j^c(x)$ are shifted Chebyshev polynomials [Cabanal-Duvillard, 01; Kusalik et al, 07; Friesen et al, 13]

$$(-1)^j \Gamma_j^c(x) = c^{j/2} 2 \cos\left(j \arccos \frac{x - (1+c)}{2\sqrt{c}} \right) + a_j$$

with $a_1 = c$ and $a_j = (-c)^{j-1}(c-1)$ for $j > 1$.

Conclusion

James' (1964) representation:

$$p(\lambda;\Theta) = \rho(\alpha,\Psi)\ {}_pF_q(a,b;c\Psi,\Lambda)\ \pi(\lambda)\,\Delta(\lambda)$$

- powerful systematization for multivariate distributions
- leads to simple approximations in low rank cases via double scaling limit
- these approximations imply Local Asymptotic Normality of super-critical experiments
- asymptotic power envelopes in the sub-critical regime

THANK YOU!