Eigenvalues and Variance Components

Iain Johnstone

Statistics and Biomedical Data Science, Stanford

Paris, 14 November, 2017

Joint with Mark Blows & Zhou Fan

Motivating example for our work

- p quantitative phenotypic traits, measured across a population
- Mean $\mu \in \mathbb{R}^p$, covariance matrix $\Sigma \in \mathbb{R}^{p \times p}$

Q: Suppose natural selection acts on one of the traits. What will be the distribution of this trait in the next generation? On the distribution of other traits?

Motivating example for our work

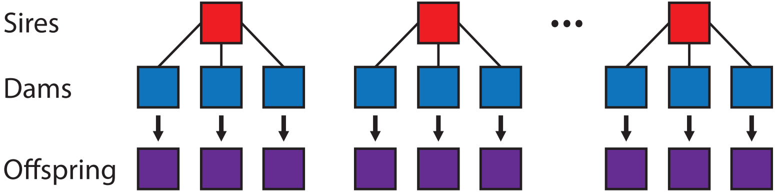
- p quantitative phenotypic traits, measured across a population
- Mean $\mu \in \mathbb{R}^p$, covariance matrix $\Sigma \in \mathbb{R}^{p \times p}$

Q: Suppose natural selection acts on one of the traits. What will be the distribution of this trait in the next generation? On the distribution of other traits?

A: Depends on the "additive genetic component" of Σ

Can estimate genetic covariance using variance components models (Fisher 1918)

Half-sib experiment



One-way layout: $n = IJ$ individuals, I groups of size J

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \in \mathbb{R}^p \qquad \text{(as row vector)}$$

$$\alpha_i \overset{iid}{\sim} \mathcal{N}(0, \Sigma_A), \qquad \text{group effect}$$

$$\varepsilon_{ij} \overset{iid}{\sim} \mathcal{N}(0, \Sigma_E), \qquad \text{residual error}$$

Additive genetic covariance: $G \approx 4\Sigma_A$

[vs. phenotypic covariance: $\mathrm{Cov}(Y) = \Sigma_A + \Sigma_E$]

Understanding G for high-dimensional trait sets

Multivariate breeder's equation (Lande and Arnold 1983):

$$\Delta\mu = G\beta$$

- Leading principal components of G indicate directions of strongest evolutionary response
- Rank of G indicates effective dimensionality of evolution
- Sparsity of G indicates extent of genetic correlations among traits

MANOVA estimator for Σ_A

Between-group and within-group "sums of squares":

$$SSA = J \sum_i (\bar{Y}_i - \bar{Y})^T (\bar{Y}_i - \bar{Y}) \qquad SSE = \sum_{i,j} (Y_{ij} - \bar{Y}_i)^T (Y_{ij} - \bar{Y}_i)$$
$$= Y^T P_A Y \qquad\qquad = Y^T P_E Y$$

Standard MANOVA estimator:

$$\hat{\Sigma}_A = \frac{1}{J}\left(\frac{1}{I-1} SSA - \frac{1}{n-I} SSE\right) = Y^T B Y,$$

for $\qquad B = \tau_A P_A - \tau_E P_E. \qquad$ NOT positive definite!

Estimation of genetic covariance Σ_A quite different from 'phenotypic covariance' $n^{-1} Y^T Y$ (estimates $\Sigma_A + \Sigma_E$).

Agenda

- Variance component models and quadratic estimators
- Bulk eigenvalue distributions
  - role of free probability
- Extreme eigenvalue distributions
- Spiked models: effect on $\hat{\Sigma}_A$

Theory is asymptotic ($p, I \to \infty$ proportionally and J fixed), using random matrix theory.

Simulations assess accuracy in finite samples.

Multivariate variance component models

One way design $\quad Y_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad$ is an example of:

$$Y = X\beta + U_1\alpha_1 + \ldots + U_{k-1}\alpha_{k-1} + \epsilon$$

$X\beta$ – fixed effects model

U_r – fixed incidence matrices

α_r – $l_r \times p$ random effects, rows $\sim N(0, \Sigma_r)$

ϵ – $n \times p$ residual errors, rows $\sim N(0, \Sigma_k)$

Covariance components: $\quad \Sigma_1, \ldots, \Sigma_k$

Some designs in quantitative genetics

[Examples, mixed model $Y = X\beta + U_1\alpha_1 + \ldots + U_{k-1}\alpha_{k-1} + \epsilon$]

[e.g. Lynch & Walsh, 1998]

Balanced half-sib	Balanced one way	I, J
Unbalanced half-sib	Unbalanced one way	I, n_i
Full-sib half-sib Monozygotic twin half-sib	Nested two-way	I, J_i, n_{ij}
	Balanced nested multi-way	$I, J, \ldots, n$
Comstock-Robinson	Replicated crossed two-way	I, J, K, L
all cases:		$I \propto p$

Quadratic Estimators of Variance Component Matrices

In the mixed model, consider estimators of Σ_r of form:

$$\hat{\Sigma} = Y^T B Y$$

- (M)ANOVA estimators:

equate Sum of Squares to expectations & solve, e.g. (1-way)

$$B = \tau_A P_A - \tau_E P_E$$

- MINQUE [minimum norm quadratic unbiased estimator]

e.g. unbalanced one way design, with 'prior parameter' $\rho \geq 0$,

$$B = \tau_{\rho A} \Pi_{\rho A} - \tau_{\rho E} P_E$$

Quadratic Estimators - Structure

Insert data

$$Y = U_1\alpha_1 + \cdots + U_k\alpha_k$$

(after setting $X\beta = 0, \epsilon = U_k\alpha_k$)

into estimator

$$\hat{\Sigma}_r = Y^T BY = \sum_{r,s=1}^{k} \alpha_r^T U_r^T BU_s\alpha_s$$

Write $\alpha_s = G_s\Sigma_s^{1/2}$, G_s i.i.d Gaussian matrix; $B_{r,s} = U_r^T BU_s$.

$$\hat{\Sigma}_r = \sum_{r,s} \Sigma_r^{1/2} G_r^T B_{r,s} G_s \Sigma_s^{1/2}$$

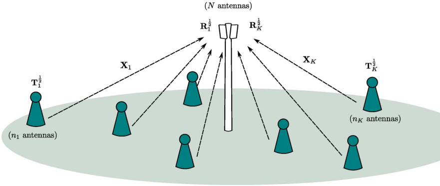
"Doubly correlated", general B.

Models in Wireless Communications

- MIMO-Multiple Access Channels

$$S + \sum_{r=1}^{k} R_r^{1/2} G_r^* T_r G_r R_r^{1/2}$$

$R_r, T_r \succ 0$ — receiver, transmitter covariances.



Couillet-Debbah-Silverstein 2011

- Frequency selective MIMO Moustakas-Simon 2007, Dupuy-Loubaton 2011

$$\sum_{r,s=1}^{k} R_r^{1/2} G_r^* T_r^{1/2} T_s^{1/2} G_s R_s^{1/2}$$

- Variance components: $B = (B_{r,s})$ symmetric, not necess $\succ 0$

$$\sum_{r,s} \Sigma_r^{1/2} G_r^T B_{r,s} G_s \Sigma_s^{1/2}$$

Bulk eigenvalue distribution

Special case: no common effects

Y has n rows $\stackrel{\text{iid}}{\sim} N_p(0, \Sigma)$; $\qquad \gamma = p/n \to \gamma_\infty$

$$\hat{\Sigma} = n^{-1} Y^T Y \qquad\qquad (B = n^{-1} I_n)$$

Empirical eigenvalue distribution: $\quad \mu_{\hat{\Sigma}} = p^{-1} \sum_{i=1}^{p} \delta_{\lambda_i(\hat{\Sigma})}$

Theorem [Marčenko-Pastur] There are deterministic measures μ_{0n} s.t.

$$\mu_{\hat{\Sigma}} - \mu_{0n} \to 0 \qquad \text{a.s.}$$

μ_{0n} has Stieltjes transform $m_0(z) = \int (\lambda - z)^{-1} \mu_{0n}(d\lambda)$ given by

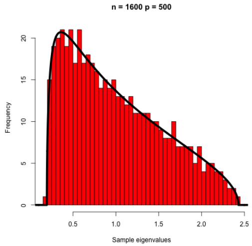
$$m_0(z) = \frac{1}{p} \operatorname{Tr} \left[(1 - \gamma - \gamma z m_0(z)) \Sigma - z \operatorname{Id} \right]^{-1},$$

(with unique solution $m_0(z) \in \mathbb{C}^+$)

Special^2 case: M-P quarter circle law

When $\Sigma = \mathrm{Id}$, $\quad \mu_0 = \mu_{0n}$ has the quarter circle density

$$f(x) = \frac{\sqrt{(E_+ - x)(x - E_-)}}{2\pi x}, \qquad E_\pm = (1 \pm \sqrt{p/n})^2$$



For general Σ, density $f(x)$ can have multiple intervals of support.

General case for $\mu_{\hat{\Sigma}} = p^{-1} \sum_{i=1}^{p} \delta_{\lambda_i(\hat{\Sigma})}$

Assume

- $n, p, I_1, \ldots, I_k \to \infty$ proportionally,
- $\hat{\Sigma} = Y^T B Y$
- $\|\Sigma_r\|, \ \|U_r\| \le O(1), \quad \|B\| \le O(1/n)$.

General case for $\mu_{\hat{\Sigma}} = p^{-1} \sum_{i=1}^{p} \delta_{\lambda_i(\hat{\Sigma})}$

Assume

- $n, p, l_1, \ldots, l_k \to \infty$ proportionally,
- $\hat{\Sigma} = Y^T B Y$
- $\|\Sigma_r\|, \ \|U_r\| \leq O(1), \quad \|B\| \leq O(1/n)$.

Theorem [Fan J 2016] There are deterministic measures μ_{0n} s.t.

$$\mu_{\hat{\Sigma}} - \mu_{0n} \to 0 \qquad \text{a.s.}$$

μ_{0n} has Stieltjes transform $m_0(z) = \int (\lambda - z)^{-1} \mu_{0n}(d\lambda)$ given by

$$m_0(z) = -\frac{1}{p} \operatorname{Tr} \left[\left(z \operatorname{Id} + \sum_{r=1}^{k} b_r(z) \Sigma_r \right] ^{-1} \right.,$$

for $a_1(z), \ldots, a_k(z), \ b_1(z), \ldots, b_k(z)$ the unique solution to a certain system of $2k$ equations.

Particular cases

Notes:

1. μ_0 depends on $n, p, I_1, \ldots, I_{k-1}, \Sigma_1, \ldots, \Sigma_k$
2. Recovers Marchenko-Pastur theorem for $k = 1$, $\hat{\Sigma} = n^{-1} Y^T Y$

System of equations for balanced one-way layout, $k = 2$, for $\hat{\Sigma}_A$:

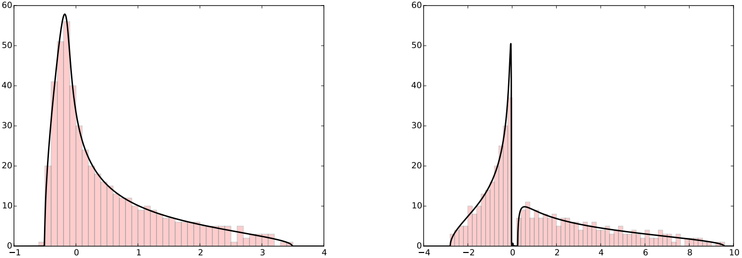
$$
\begin{aligned}
a_A(z) &= -(1/I)\,\mathrm{Tr}[(z\,\mathrm{Id} + b_A(z)\Sigma_A + b_E(z)\Sigma_E)^{-1}\Sigma_A] \\
a_E(z) &= -(1/n)\,\mathrm{Tr}[(z\,\mathrm{Id} + b_A(z)\Sigma_A + b_E(z)\Sigma_E)^{-1}\Sigma_E] \\
b_A(z) &= -\frac{1}{1 + a_A(z) + a_E(z)} \\
b_E(z) &= \frac{J-1}{J(J-1-a_E(z))} - \frac{1}{J + Ja_A(z) + Ja_E(z)} \\
m_0(z) &= -(1/p)\,\mathrm{Tr}[(z\,\mathrm{Id} + b_A(z)\Sigma_A + b_E(z)\Sigma_E)^{-1}]
\end{aligned}
$$

Easily solvable by iteration to compute $\frac{1}{\pi}\Im m_0(z)$ near real-axis.

Computing the deterministic equivalent

Theorem (Fan, J. (cont'd))

To compute $m_0(z)$, the preceding system of equations may be solved by initializing (b_1, b_2) arbitrarily, then iteratively updating (a_1, a_2) and (b_1, b_2) until convergence.



Black curves: $\pi^{-1}\Im m_0(z)$ for $\Im z = 10^{-4}$.
Right: $\Sigma_A = \Sigma_E = \mathrm{Id}$. $I = 200$ groups of size $J = 2$, $p = 500$.

Balanced nested/crossed designs

Lattice subspace structure:

$$S_r = \bigoplus_{r' \preceq r} \mathring{S}_{r'},$$

Mean (Sum of) Squares MS_r for $\mathring{S}_r$

$$\mathbb{E}\, MS_t = \sum_{r \succeq t} c_r \Sigma_r$$

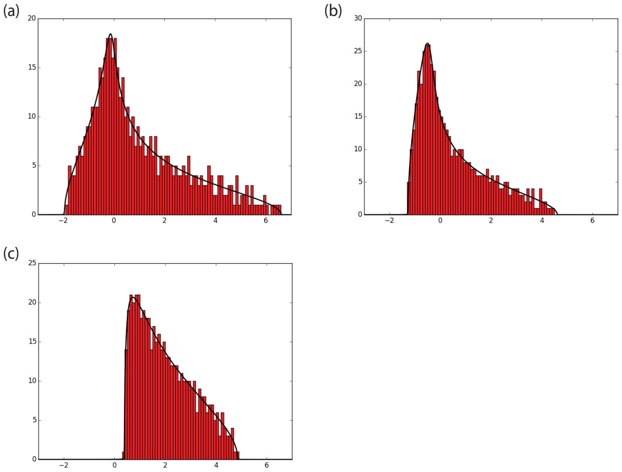
Möbius inversion: $\hat{\Sigma}_t = Y^T B_t Y$

$$\begin{tikzcd}[row sep=small] S_0 \arrow[d, dash] \\ S_1 \\ \vdots \\ S_r \arrow[d, dash] \\ S_{r+1} \\ \vdots \\ S_k \end{tikzcd} \qquad \begin{tikzcd}[row sep=small, column sep=small] & S_0 \arrow[d, dash] & \\ & S_1 \arrow[dl, dash] \arrow[dr, dash] & \\ S_2 \arrow[dr, dash] & & S_3 \arrow[dl, dash] \\ & S_4 \arrow[d, dash] & \\ & S_5 & \end{tikzcd}$$

In equation system for $m_0(z)$ for $\hat{\Sigma}_t$,

- compute $b_r(z)$ as rational function of $a_s(z)$
- included in software (forthcoming)

Two-way nested example



$$n = 2400, \quad I_1 = 400, \quad I_2 = 2, \quad I_3 = 3, \quad p = 500$$

$$\Sigma_1 = \mathrm{Id}_p, \qquad \Sigma_2 = 0.5\,\mathrm{Id}_p, \qquad \Sigma_3 = 2\,\mathrm{Id}_p,$$

$$\text{(a)}: \ \hat{\Sigma}_1, \qquad \text{(b)}: \ \hat{\Sigma}_2, \qquad \text{(c)}: \ \hat{\Sigma}_3,$$

About proof of bulk convergence

- Via rectangular free probability
- In classical probability, if scalar variables $G \perp\!\!\!\perp H|A$, then

$$\mathbb{E}[e^{isG}|H, A] = \mathbb{E}[e^{isG}|A]$$

- use operator-valued free probability for eigenvalues of

$$W = Y^T BY = \sum_{r,s=1}^{k} H_r^T G_r^T B_{r,s} G_s H_s,$$

via block matrix embedding,

Free probability correspondence

All the familiar probability notions have analogues:

(commutative) probability	non-commutative probability
scalar r.v.s X, Y	random matrices A, B
sample space	operator (W^*-)algebra
expectation $\mathbb{E}$	trace τ
moments $\mathbb{E}X^k$	moments $\tau(A^k) = \sum_i \lambda_i^k(A)$
Fourier transform	Stieltjes/Cauchy transform
(asy) independence	(asy) freeness
sub σ-field $\mathcal{H}$	sub (W^*-)algebra $\mathcal{B}$
conditional expectation	$\mathcal{B}$-valued expectation operator
conditional indep. $\vert\mathcal{H}$	freeness w. amalgamation over $\mathcal{B}$
(conditional) Fourier transform	(operator valued) Cauchy transform

Voiculescu 1991, 1995, Speicher 1998, Nica, Shlyakhtenko and Speicher 2002, Benaych-Georges 2009, Hiai and Petz 2000, Speicher and Vargas 2012,

Proof ideas: in one-way case, Σ_A, Σ_E

Since $Y = U\alpha + \epsilon$,

$$\hat{\Sigma}_A = (U\alpha + \varepsilon)^T B (U\alpha + \varepsilon) = \sum_{r,s \in \{A,E\}} \Sigma_r^{1/2} G_r^T F_{rs} G_s \Sigma_s^{1/2},$$

where $F_{AA} = U^T BU, \quad F_{AE} = U^T B, \quad F_{EA} = BU, \quad F_{EE} = B.$

These matrices are rectangular, of dimensions p, I, and $n = IJ$.

Set $N = p + I + n$ and study equivalent $N \times N$ square matrices by zero-padding [Benaych-Georges '09]:

Σ_A, Σ_E	G_A^T	G_E^T
G_A	F_{AA}	F_{AE}
G_E	F_{EA}	F_{EE}

Proof ideas

The padded versions of G_A, G_E are invariant in law under

$$G_A \mapsto O_A^T G_A O_0, \quad O = \begin{pmatrix} O_0 & & \\ & O_A & \\ & & O_E \end{pmatrix},$$

where O_0, O_A, O_E are independent Haar-orthogonal.

Proof ideas

The padded versions of G_A, G_E are invariant in law under

$$G_A \mapsto O_A^T G_A O_0, \quad O = \begin{pmatrix} O_0 & & \\ & O_A & \\ & & O_E \end{pmatrix},$$

where O_0, O_A, O_E are independent Haar-orthogonal.

Approximate

$$\hat{\Sigma}_A = \sum_{r,s \in \{A,E\}} \Sigma_r^{1/2} G_r^T F_{rs} G_s \Sigma_s^{1/2}$$

by

$$w = \sum_{r,s \in \{A,E\}} h_r^* g_r^* f_{rs} g_s h_s,$$

where w, h_r, g_r, f_{rs} are elements of a von Neumann algebra $\mathcal{A}$ with tracial state τ, such that:

Proof ideas

$$w = \sum_{r,s\in\{A,E\}} h_r^* g_r^* f_{rs} g_s h_s \quad \Leftrightarrow \quad \hat{\Sigma}_A = \sum_{r,s\in\{A,E\}} \Sigma_r^{1/2} G_r^T F_{rs} G_s \Sigma_s^{1/2},$$

there is approximation of moments:

τ-moments in $\mathcal{A}$		$N^{-1}\,\mathrm{Tr}$-moments in $\mathbb{R}^{N\times N}$	
f_{rs}	$=$	F_{rs},	$r, s \in \{A, E\}$

Proof ideas

$$w = \sum_{r,s \in \{A,E\}} h_r^* g_r^* f_{rs} g_s h_s \quad \Leftrightarrow \quad \hat{\Sigma}_A = \sum_{r,s \in \{A,E\}} \Sigma_r^{1/2} G_r^T F_{rs} G_s \Sigma_s^{1/2},$$

there is approximation of moments:

τ-moments in $\mathcal{A}$		N^{-1} Tr-moments in $\mathbb{R}^{N \times N}$	
f_{rs}	$=$	F_{rs},	$r, s \in \{A, E\}$
h_A, h_E	$=$	$\Sigma_A^{1/2}, \Sigma_E^{1/2}$	

Proof ideas

$$w = \sum_{r,s\in\{A,E\}} h_r^* g_r^* f_{rs} g_s h_s \quad \Leftrightarrow \quad \hat{\Sigma}_A = \sum_{r,s\in\{A,E\}} \Sigma_r^{1/2} G_r^T F_{rs} G_s \Sigma_s^{1/2},$$

there is approximation of moments:

τ-moments in $\mathcal{A}$		N^{-1} Tr-moments in $\mathbb{R}^{N\times N}$		
f_{rs}	$=$	F_{rs},	$r, s \in \{A, E\}$	
h_A, h_E	$=$	$\Sigma_A^{1/2}, \Sigma_E^{1/2}$		
g_A, g_E	$\sim$	G_A, G_E	as	$I \propto p \to \infty$

Proof ideas

$$w = \sum_{r,s\in\{A,E\}} h_r^* g_r^* f_{rs} g_s h_s \quad \Leftrightarrow \quad \hat{\Sigma}_A = \sum_{r,s\in\{A,E\}} \Sigma_r^{1/2} G_r^T F_{rs} G_s \Sigma_s^{1/2},$$

there is approximation of moments:

τ-moments in $\mathcal{A}$		N^{-1} Tr-moments in $\mathbb{R}^{N\times N}$		
f_{rs}	$=$	F_{rs},		$r,s\in\{A,E\}$
h_A, h_E	$=$	$\Sigma_A^{1/2}, \Sigma_E^{1/2}$		
g_A, g_E	$\sim$	G_A, G_E	as	$I \propto p \to \infty$

and $\{f_{rs}\}, \{h_A, h_E\}, \{g_A, g_E\}$ are conditionally free over a 3-dimensional sub-algebra of projections $\mathcal{D} = \langle p_0, p_A, p_E\rangle$.

This free model is N-dependent. [Speicher, Vargas '12]

Proof ideas

Asymptotic freeness of random matrices: [Voiculescu '91], [Dykema '95], [Hiai and Petz '00], [Collins and Śniady '06], ...

Freeness for rectangular matrices: [Benaych-Georges '09]

Theorem (Fan, J)

If random matrix families $\{A_i\}$ and $\{B_j\}$ are independent and invariant under conjugation by block-orthogonal matrices O, then they are asymptotically (in N) well-approximated by families $\{a_i\}$ and $\{b_j\}$ in $(\mathcal{A}, \tau)$ which are conditionally free.

Aside: (conditional) freeness and cumulants

For $a \in \mathcal{A}$, its Stieltjes transform may be defined as

$$m_0(z) = \tau((a - z)^{-1}) = \sum_{l \geq 0} z^{-(l+1)} \tau(a^l),$$

where $\tau(a^l)$ are τ-moments of a, and satisfies

$$m_0(z^{-1} + \mathcal{R}(z)) = -z,$$

for the $\mathcal{R}$-transform of a,

$$\mathcal{R}(z) = \sum_{l \geq 1} z^{l-1} \kappa_l(a).$$

If a, b are free, then $\kappa_l(a + b) = \kappa_l(a) + \kappa_l(b)$. Analogous results hold for conditional freeness. [Voiculescu '95, Speicher '98]

Proof ideas

$w \in \mathcal{A}$ approximating $\hat{\Sigma}_A$ is a deterministic operator—it remains to characterize its law under τ.

We show that its Stieltjes transform m_0 satisfies the preceding fixed-point equations, via a cumulant-based computation in $\mathcal{A}$.

Computation passes between (conditional) Cauchy and $\mathcal{R}$-transforms with respect to different sub-algebras, using tools and ideas from [Nica, Shlyakhtenko, Speicher '02] and [Speicher, Vargas '12].

Role of conditional freeness

Classical: B, H independent $\Leftrightarrow$ mixed joint cumulants vanish.

If $B \perp\!\!\!\perp H|\mathcal{D}$ and then

$$\mathbb{E}[e^{itB}|\sigma(H, \mathcal{D})] = \mathbb{E}[e^{itB}|\mathcal{D}]$$

Freeness: b, h are $\mathcal{D}$-free ("free with amalgamation over $\mathcal{D}$") iff all mixed $\mathcal{D}$-cumulants vanish: e.g.:

$$\kappa_5^{\mathcal{D}}(b, h, b, b, h) = 0$$

Consequence: If b, h are $\mathcal{D}$-free, and $\mathcal{H} = \langle h, \mathcal{D} \rangle$, then

$$\kappa_\ell^{\mathcal{H}}(bh, \ldots, bh, b) = \kappa_\ell^{\mathcal{D}}(bF^{\mathcal{D}}(h), \ldots, bF^{\mathcal{D}}(h), b)$$

$\Rightarrow$ express $\mathcal{R}^{\mathcal{H}}$ in terms of $\mathcal{R}^{\mathcal{D}}$. [Nica-Shlyakhtenko-Speicher 2002]

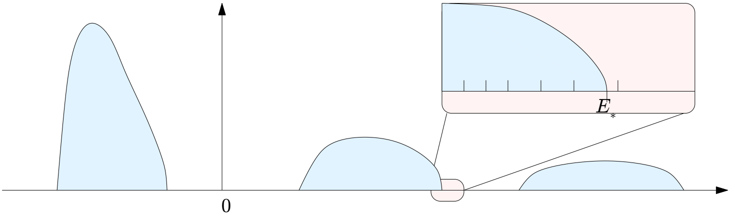
Extreme eigenvalue distribution

Tracy-Widom at each edge

$X = (X_{\alpha i})$ $M \times N$, $\quad X_{\alpha i} \sim (0, N^{-1})$; $\quad T$ symmetric

$\hat{\Sigma} = X^T T X \qquad$ (and, if $T \geq 0$, $\tilde{\Sigma} = T^{1/2} X X^T T^{1/2}$)

For $M/N \to d > 0$ and a regular right (or left) edge E_*:



Theorem [Fan, J 17] If $\lambda_{\max}, \lambda_{\min}$ are extreme eigenvalues near E_*,

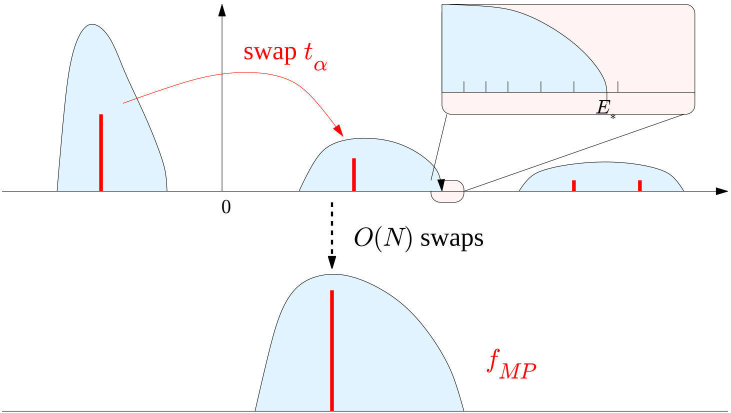
$$(\gamma N)^{2/3}(\lambda_{\max} - E_*) \xrightarrow{L} \mathrm{TW}_1 \qquad \text{(right edge)}$$

$$(\gamma N)^{2/3}(E_* - \lambda_{\min}) \xrightarrow{L} \mathrm{TW}_1 \qquad \text{(left edge)}$$

[Extends Lee – Schnelli, 2016: $T > 0$, largest eigenvalue only]

Lindeberg swapping

Swap eigenvalues (t_α) of T one-at-a-time between bulks:



$O(N)$ swaps with careful scaling doesn't change limit of $\lambda_{\max}$!
Then build on resolvent comparison approach, Lee - Schnelli, 2016.

Largest eigenvalue under "global null hypothesis"

Back to var. components model: $Y = X\beta + U_1\alpha_r + \cdots + U_k\alpha_k$

$$H_0 : \Sigma_r = c_r^2 \,\mathrm{Id}, \quad r = 1, \ldots, k \qquad \text{"sphericity"}$$

$$\hat{\Sigma} = Y^T BY, \qquad [BX = 0]$$

Note: B has $+$ve *and* $-$ve eigenvalues, even in the limit.

Corollary (Fan J 17)

Assume $n, p \to \infty$ proportionally, H_0 holds, $\|B\| \asymp 1/n$, and B satisfies certain regularity conditions. Then

$$(\lambda_{\max}(\hat{\Sigma}) - \mu_{np})/\sigma_{np} \xrightarrow{L} TW_1,$$

where $\mu_{np}, \sigma_{np} = (\kappa p)^{-2/3}$ are functions of $p, \lambda_1(B), \ldots, \lambda_n(B)$.

Formulas for center μ_{np} and scale σ_{np}

Let $t_1, \ldots, t_M$ be eigenvalues of $T = (pc_r c_s U_r^T B U_s)$,
with $M = l_1 + \cdots + l_k$.

$$z(m) = -\frac{1}{m} + \frac{1}{p} \sum_{\alpha=1}^{M} \frac{t_\alpha}{1 + t_\alpha m}$$

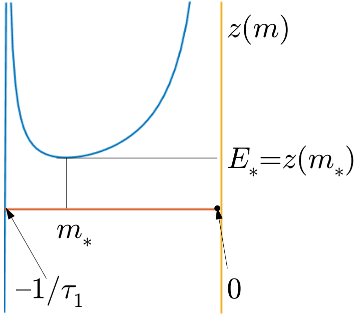
Silverstein-Choi (95)

m_* : solves $z'(m) = 0$

$\mu_{np} = E_+ = z(m_*)$

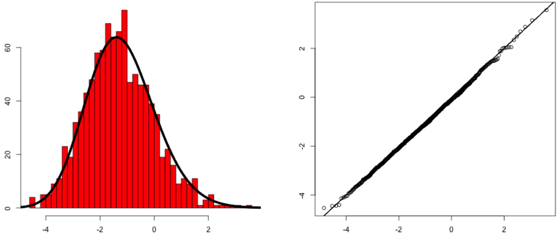
$\sigma_{np} = [z''(m_*)/(2p^2)]^{1/3}$

El Karoui 2007,
Hachem-Hardy-Najim 2016



Largest eigenvalue: balanced one-way example

$$H_0: \ \Sigma_A = 0, \quad \Sigma_E = \mathrm{Id}, \qquad B_A = \tau_A P_A - \tau_E P_E$$



Histogram & QQ-plot of scaled $\lambda_{\max}(\hat{\Sigma}_A)$ $\mu_{np} = 0.91, \sigma_{np} = 0.012$

$I = 400, \quad J = 4, \quad p = 500$

- Conditions ok if $\lambda_{\max}(B) \asymp 1/n$ with multiplicity $\asymp n$
- For c_r unknown, can use $\hat{c}_r^2 = p^{-1}\mathrm{Tr}(\hat{\Sigma}_r)$.

Approximation accuracy in finite samples

	F_1	$n = p$ $J = 2$	$J = 5$	$J = 10$	$n = 4 \times p$ $J = 2$	$J = 5$	$J = 10$
$p = 20$	0.90	0.938	0.944	0.953	0.932	0.937	0.940
	0.95	0.971	0.974	0.978	0.968	0.970	0.972
	0.99	0.995	0.995	0.995	0.993	0.994	0.995
$p = 100$	0.90	0.926	0.934	0.931	0.923	0.919	0.918
	0.95	0.963	0.969	0.968	0.962	0.961	0.961
	0.99	0.992	0.995	0.995	0.993	0.993	0.994
$p = 500$	0.90	0.922	0.917	0.918	0.913	0.909	0.916
	0.95	0.961	0.958	0.959	0.956	0.954	0.959
	0.99	0.992	0.992	0.991	0.993	0.992	0.994

Empirical CDF values at the theoretical 0.90, 0.95, and 0.99 quantiles of the F_1 law. (Standard errors 0.001–0.003.)

The Tracy-Widom test is slightly conservative in practice.

Spiked Variance Component models

The spiked covariance model

Assume [J '01]:

$$\Sigma_A = \sigma_A^2 \,\mathrm{Id} + \text{finite number of spikes}$$
$$\Sigma_E = \sigma_E^2 \,\mathrm{Id} + \text{finite number of spikes}$$

- For an eigenvalue of $\hat{\Sigma}_A$ close to the bulk, how do we tell if it's really caused by a spike?
- **When there are spikes, where do corresponding outlier eigenvalues of $\hat{\Sigma}_A$ appear in the spectrum?**

Locations of outliers

Let

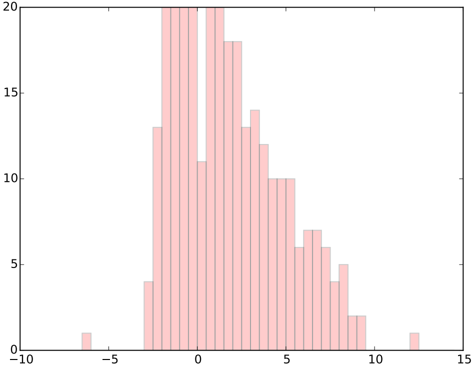
$$\Sigma_A = \sigma_A^2 \operatorname{Id} + V_A \Theta_A V_A^T, \qquad \Sigma_E = \sigma_E^2 \operatorname{Id} + V_E \Theta_E V_E^T,$$

where Θ_A, Θ_E (diagonal) contain the spike values and columns of V_A, V_E contain the corresponding eigenvectors.

Set

$$\Theta = \begin{pmatrix} \Theta_A & 0 \\ 0 & \Theta_E \end{pmatrix}, \qquad S = \begin{pmatrix} \operatorname{Id} & V_A^T V_E \\ V_E^T V_A & \operatorname{Id} \end{pmatrix}.$$

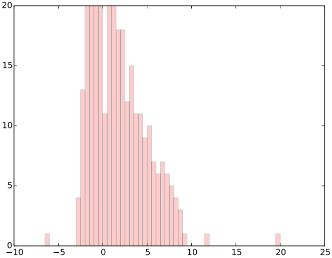
Aliasing from the error covariance



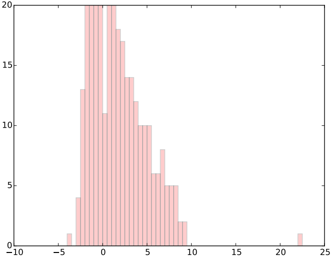
Eigenvalues of $\hat{\Sigma}_A$

$\Sigma_A = \mathrm{Id}$, but $\Sigma_E = \mathrm{Id} +$ spike at 25
$I = 200$ groups of size $J = 2$, $p = 500$ traits.

Dependence on eigenvector alignment



Eigenvalues of $\hat{\Sigma}_A$



Eigenvalues of $\hat{\Sigma}_A$

$\Sigma_A = \mathrm{Id} +$ spike at 15, $\Sigma_E = \mathrm{Id} +$ spike at 25.

Left: Spikes orthogonal. Right: Spikes aligned.

Locations of outliers

Theorem (Fan & Yi Sun, informal version)

Suppose $p, I \to \infty$, and $\sigma_A^2, \sigma_E^2, \Theta, S$ are fixed. Let $\hat{\Sigma} = c_1 MSA + c_2 MSE$ with deterministic equivalent measure μ_0. For each root $\lambda \in \mathbb{R} \setminus \operatorname{supp}(\mu_0)$ of

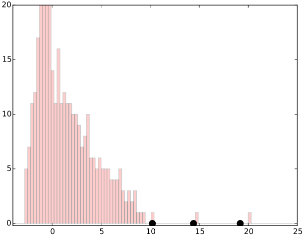
$$\det\left(\mathrm{Id} + S\Theta \begin{pmatrix} t_1(\lambda; c_1, c_2)\,\mathrm{Id} & 0 \\ 0 & t_2(\lambda; c_1, c_2)\,\mathrm{Id} \end{pmatrix}\right) = 0,$$

an eigenvalue of $\hat{\Sigma}$ converges to λ. The remaining eigenvalues of $\hat{\Sigma}$ converge to $\operatorname{supp}(\mu_0)$.

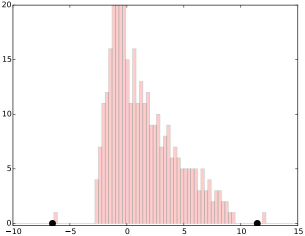
Here, t_1 and t_2 are two (explicit) analytic functions of λ, c_1, c_2.

$\to$ Algorithm to estimate Θ from loci of outlier eigenvalues of $\hat{\Sigma}(c_1, c_2)$ as (c_1, c_2) vary.

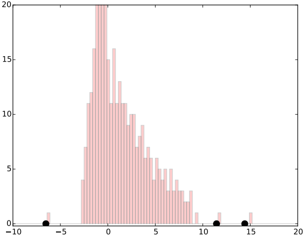
Locations of outliers in $\hat{\Sigma}_A$

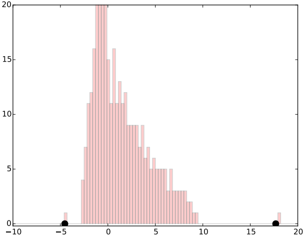


$\Sigma_A = \mathrm{Id} + [2, 5, 10, 15], \Sigma_E = \mathrm{Id},$



$\Sigma_A = \mathrm{Id}, \Sigma_E = \mathrm{Id} + 25$





$\Sigma_A = \mathrm{Id} + 10, \Sigma_E = \mathrm{Id} + 25;$ left: orthogonal, right: aligned

Black dots indicate theoretical predictions for outlier locations.

References

Z Fan, I Johnstone. Eigenvalue distributions of variance components estimators in high-dimensional random effects models. *arXiv:1607.02201*

Z Fan, I Johnstone. Tracy-Widom at each edge of real covariance estimators. *arXiv:1707.02352*

M Blows, Z Fan, E Hine, I Johnstone, Y Sun. Spiked covariances and eigenvalue estimation in high-dimensional random effects models. *In preparation.*

THANK YOU!