Statistical Inference and Random Matrices

N.S. Witte

Institute of Fundamental Sciences
Massey University
New Zealand

5-12-2017

Joint work with Peter Forrester

6th Wellington Workshop in Probability
and Mathematical Statistics
4-6 December 2017

Applications

Historical Origins:

- Integrals over Classical Groups $U(N), O(N), Sp(2N)$ Hurwitz 1897, Haar 1933
- Mathematical Statistics: Wishart 1928, James 1954-64, Constantine 1964, Mathai 1997
- Quantisation of Classically Chaotic Systems: Wigner 1955, 1958

Applications

Historical Origins:

- Integrals over Classical Groups $U(N), O(N), Sp(2N)$ Hurwitz 1897, Haar 1933
- Mathematical Statistics: Wishart 1928, James 1954-64, Constantine 1964, Mathai 1997
- Quantisation of Classically Chaotic Systems: Wigner 1955, 1958

Contemporary Applications:

- Principal Component Analysis, sample covariance matrices, Wishart matrices, null and non-null SCM
- mathematical finance, cross correlations of financial data, sample correlation matrices,
- Polynuclear growth models, random permutations, last passage percolation, queuing models,
- Biogeographic pattern of species nested-ness, ordered binary presence-absence matrices,
- distribution of mutation fitness effects across species, Fisher's geometrical model,
- complex networks modeled by random graphs, e.g. adjacency matrices
- data analysis and statistical learning
- Stable signal recovery from incomplete and inaccurate measurements
- Compressed sensing, best k-term approximation, n-widths
- Wireless communication, antenna networks,
- quantum entanglement
- quantum chaos, semi-classical approximation
- quantum transport in mesoscopic systems

E.g. **G**aussian **O**rthogonal **E**nsembles of Random Matrices: **GOE** aka Gaussian Wigner matrix
i.i.d random variables

$$x_{j,j} \sim N[0,1]$$
$$x_{j,k} \sim N[0, \tfrac{1}{\sqrt{2}}]$$

Construct $n \times n$ real, symmetric matrix $X = (x_{j,k})_{j,k=1}^{n}$

Joint p.d.f for the elements

$$P(X) = \frac{1}{C_n} \prod_{j=1}^{n} e^{-\frac{1}{2} x_{j,j}^2} \prod_{1 \le j < k \le n} e^{-x_{j,k}^2} = \frac{1}{C_n} \prod_{1 \le j,k \le n} e^{-\frac{1}{2} x_{j,k}^2} = \frac{1}{C_n} e^{-\frac{1}{2} \mathrm{Tr}(X^2)}$$

Invariance under orthogonal transformations $X \mapsto OXO^{\dagger}$, $OO^{\dagger} = I$

- $n \times n$ Real Symmetric Matrices $X = (x_{j,k})_{1\leq j,k\leq n}$
  Eigenvalue Analysis $\lambda_1, \ldots, \lambda_n$
  $$X = O\Lambda O^{\dagger}$$
  $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, Orthogonal Eigenvectors $O = (O_1, \ldots, O_n)$, $OO^{\dagger} = I$
  Volume form
  $$(dX) = \wedge_{j=1}^{n} dx_{j,j} \wedge_{1\leq j<k\leq n} dx_{j,k}$$
  Change of $\frac{1}{2}n(n+1)$ variables $\{x_{j,k}\} \mapsto \{\lambda_j, O_j\}$ Jacobian
  $$(dX) = (O^{\dagger} dO) \prod_{1\leq j<k\leq n} |\lambda_j - \lambda_k| \wedge_{j=1}^{n} d\lambda_j$$
- $n \times m$ Real Matrices X
  Singular Value Decomposition $\sigma_1, \ldots, \sigma_m$
  $$X = O\Sigma P^{\dagger}$$
  $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_m) \in \mathbb{R}^{n\times m}$, Orthogonal $O \in \mathbb{R}^{n\times n}$, $P \in \mathbb{R}^{m\times m}$
- Gram-Schmidt orthogonalisation, QR, LU, Cholesky, Hessenberg Decompositions

Joint p.d.f for eigenvalues

$$P(\lambda) = \frac{1}{C_n} \prod_{j=1}^{n} e^{-\frac{1}{2}\lambda_j^2} \prod_{1 \le j < k \le n} |\lambda_j - \lambda_k|^1, \quad \lambda_j \in \mathbb{R}$$

N.B.

- repulsion parameter, Dyson index, $\beta = 1$ in Stieltjes picture, "Log-gas"
- Hermite weight, one of the classical orthogonal polynomial weights
- normalisation is Selberg integral

$X \in \mathbb{F}^{n\times p}$ $(\mathbb{F} = \mathbb{R}, \mathbb{C})$ with

$$X = \{x_k^{(j)}\}_{\substack{j=1,\dots,n \\ k=1,\dots,p}} ,$$

with

$$p = \#\text{ of variables,}$$
$$n = \#\text{ of data points}$$

$p \times p$ Covariance matrix

$$A = X^\dagger X = \left(\sum_{j=1}^{n} x_{k_1}^{(j)} x_{k_2}^{(j)}\right)_{\substack{k_1=1,\dots,p \\ k_2=1,\dots,p}}$$

A is a Wishart matrix if $x_{j,k}$ are i.i.d random variables drawn from $N[0,1]$

Joint eigenvalue p.d.f.

$$\frac{1}{C}\prod_{k=1}^{p}\lambda_k^{\beta a/2} e^{-\beta\lambda_k/2} \prod_{1\le j<k\le p} |\lambda_j - \lambda_k|^\beta, \lambda_k \in [0,\infty)$$

i.e. Laguerre weight, $a = n - p + 1 - 2/\beta$, $n \ge p$

$$\beta = \begin{cases} 1 & \text{Real}, \mathbb{R} \\ 2 & \text{Complex}, \mathbb{C} \end{cases}$$

Translation

Single-Wishart Null hypothesis $\mu = 0, \Sigma = I$ p − variate degrees of freedom $= n$	Laguerre LβE $e^{-\frac{\beta}{2}\lambda}\lambda^{\frac{\beta}{2}(n-p+1-2/\beta)}$
Double-Wishart Null hypothesis $\mu_1 = 0, \Sigma_1 = I,\ \mu_2 = 0, \Sigma_2 = I$ $p - 1^{st}$ variate $q - 2^{nd}$ variate degrees of freedom $= n$	Jacobi JβE $(1-\lambda)^{\frac{\beta}{2}(q-p+1-2/\beta)}(1+\lambda)^{\frac{\beta}{2}(n-q-p+1-2/\beta)}$

Global Properties of the spectra: Empirical Spectral Distribution

$n \to \infty$ **Wigner semi-circle Law** for the *global density of eigenvalues*

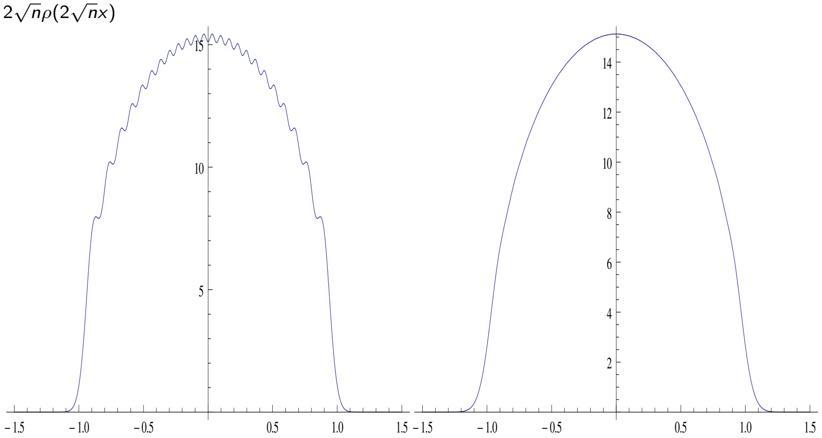
$$\rho(\lambda) = \frac{1}{\pi}\sqrt{2n - \lambda^2}$$

$n, p \to \infty$ **Marchenko-Pastur Law** for eigenvalues of $X^\dagger X$

$$\rho(\lambda) = \frac{1}{2\pi\lambda}\sqrt{(\lambda - nx_-)(nx_+ - \lambda)}$$

where

$$x_\pm = \left(c^{-1/2} \pm 1\right)^2, \qquad c = \frac{p}{n} \leq 1$$


$2\sqrt{n}\rho(2\sqrt{n}x)$
5
10
15
2
4
6
8
10
12
14
−1.5
−1.0
−0.5
0.5
1.0
1.5
−1.5
−1.0
−0.5
0.5
1.0
1.5

Prop. 2.2 and Lemma 4.1 of Haagerup and Thorbjørnsen [2012]

Theorem

The eigenvalue density $\rho(x)$ for the GUE satisfies the third order, homogeneous ordinary differential equation

$$\rho''' + (4n - x^2)\rho' + x\rho = 0$$

subject to certain boundary conditions, for fixed n as $x \to \pm\infty$.

W & Forrester [2013]

Theorem

The eigenvalue density $\rho(x)$ for the GOE satisfies the fifth order, linear homogeneous ordinary differential equation

$$-4\rho^{(V)} + 5(x^2 - 4n + 2)\rho''' - 6x\rho'' + [-x^4 + (8n-4)x^2 - 16n^2 + 16n + 2]\rho' + x(x^2 - 4n + 2)\rho = 0$$

again subject to certain boundary conditions, for fixed n as $x \to \pm\infty$.

The many ways to look at these problems

Statistic on $\mathrm{Spec}(X) = \{\lambda_1, \dots \lambda_n\}$	Regime
density $\rho(\lambda)$	Global spectrum
m-point correlation functions $\rho_m(\lambda_1, \dots, \lambda_m)$	
Linear Spectral Statistics, $\sum_{i=1}^{n} f(\lambda_i)$	Hypothesis tests, Distribution theory
Extreme Eigenvalues, $\lambda_{max}, \lambda_{min}$	Large deviations, Spectrum edge
Eigenvalue Spacings, $\lambda_{i+1} - \lambda_i$	Bulk or Edge spectrum
Spectral Gaps $\forall j, \lambda_j \notin J \subset \mathrm{Spec}(X)$	
Condition Numbers, $\frac{\lambda_{max}}{\lambda_{min}}$	
Determinants, Characteristic Polynomials $\prod_{i=1}^{n}(\zeta - \lambda_i)$	

Tools one can use

Approach	Primary Object
Moment Methods	
Concentration Inequalities	
Large Deviation Theory	Potential problems and equilibrium measures
Free Probability	Stieltjes transform
Loop Equations	Stieltjes transform, Resolvents
Hypergeometric Functions of Matrix Argument	Zonal and Jack Polynomials
Orthogonal/Bi-orthogonal Polynomials	Riemann-Hilbert asymptotics
Integrable Systems, Painlevé equations	Gap probabilities, Characteristic polynomial

Gap probability i.e. probability of no eigenvalues of $n \times n$ GUE in (t, ∞) denoted by

$$E_{2,n}(0; (t, \infty))$$

Shift and scale t as

$$t = \sqrt{2n} + \frac{s}{\sqrt{2}n^{1/6}}$$

Take limit $n \to \infty$ of Gap probability

$$\lim_{n \to \infty} \mathbb{P}[\frac{\lambda_{max} - \sqrt{2n}}{\sqrt{2}n^{1/6}} \leq s] = F_2(s)$$

Tracy-Widom [1994] Fredholm determinant

$$F_2(s) = \det(1 - \mathbb{K}_2)_{L^2(s,\infty)}$$

where the integral operator $\mathbb{K}_2$ has the kernel, the *Airy Kernel*

$$K_2(x, y) = \frac{\mathrm{Ai}(x)\mathrm{Ai}'(y) - \mathrm{Ai}(y)\mathrm{Ai}'(x)}{x - y}$$

The P_{II} transcendent $q(t;\alpha)$ then satisfies the standard form of the second Painlevé equation

$$\frac{d^2}{dt^2}q = 2q^3 + tq + \alpha, \quad \alpha \in \mathbb{C}$$

Gap probability, i.e. Tracy-Widom distribution $F_2(s)$, is

$$E_2^{\mathrm{Soft\ Edge}}(0;(s,\infty)) = \exp\left(-\int_s^\infty dt\ (t-s)q(t)^2\right)$$

where $q(t)$ is the $\alpha = 0$ solution for P_{II} with the boundary condition

$$q(t) \underset{t\to\infty}{\sim} \mathrm{Ai}(t)$$

Hastings and McLeod solution, see Hastings, S. P. and McLeod, J. B.
A boundary value problem associated with the second Painlevé transcendent and the Korteweg-Vries equation. Arch. Rational Mech. Anal., 1980, **73**(1), 31-51

Forrester & W [2012], Tails

$$\log E_\beta^{\mathrm{Soft\ Edge}}(n;(s,\infty)) \underset{\substack{s\to-\infty \\ n\ll|s|}}{\sim} -\frac{\beta|s|^3}{24} + \frac{\sqrt{2}}{3}|s|^{3/2}\left(\beta n + \frac{\beta}{2} - 1\right) + \left[\frac{\beta}{2}n^2 + \left(\frac{\beta}{2}-1\right)n + \frac{1}{6}\left(1 - \frac{2}{\beta}\left(1-\frac{\beta}{2}\right)^2\right)\right]\log|s|^{-3/4}$$

Extension of central limit theorems and the Gaussian distribution

Conjecture

As the rank of the random matrix ensembles $n \to \infty$, with or without a similar scaling of other parameters, the ensembles

- *have well-defined limits,*
- *these limits define new distributions which are insensitive to details of the finite model other than their symmetry class, β*
- *are characterised by the solutions of integrable dynamical systems, e.g. of the integrable hierarchies such as the Toda lattice, K-dV or K-P systems, or more precisely by Painlevé type equations.*

Proven Cases:

- "Four Moments Theorems" in Tao, T., Vu, V., *Random covariance matrices: universality of local statistics of eigenvalues.* Ann. Probab., 2012, **40**, 1285–1315.
- "Riemann-Hilbert Approach" in Deift, P., Gioev, D. *Random Matrix Theory: Invariant Ensembles and Universality.* Amer. Math. Soc., 2009.

Baik, J., Ben Arous, G. and Péché, S.
Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices.
Ann. Probab., 2005, **33**(5), 1643–1697

$\beta = 2$, λ_1 is the largest eigenvalue of the sample covariance matrix, $\gamma^2 = \frac{n}{p}$, $\ell_1 \geq \ell_2$
Population covariance matrix

$$\Sigma = \begin{pmatrix} \ell_1 & & \\ & \ell_2 & \\ & & I_{p-2} \end{pmatrix}$$

As $p, n \to \infty$ either

$$\mathbb{P}\left([\lambda_1 - (1+\gamma^{-1})^2)] \frac{\gamma}{(1+\gamma)^{4/3}} n^{2/3} \leq x \right) \to \begin{cases} F_2(x), & 0 < \ell_1, \ell_2 < 1+\gamma^{-1} \\ F_1^2(x), & 0 < \ell_2 < 1+\gamma^{-1} = \ell_1 \\ F(x) & \ell_1 = \ell_2 = 1+\gamma^{-1} \end{cases}$$

or

$$\mathbb{P}\left([\lambda_1 - \ell_1(1 + \frac{\gamma^{-2}}{\ell_1 - 1})] \frac{n^{1/2}}{\ell_1\sqrt{1-\gamma^{-2}/(\ell_1-1)^2}} \leq x \right) \to \begin{cases} G_1(x), & \ell_1 > 1+\gamma^{-1}, \ell_1 > \ell_2 \\ G_2(x), & \ell_1 = \ell_2 > 1+\gamma^{-1} \end{cases}$$

As $n \to \infty$ and $p \ll n$ PCA works: Sample CM $\to$ Population CM

As $n, p \to \infty$ and $p = \mathrm{O}(n)$???

Issue: How close are the eigenvalues of the sample PCA to the population PCA?

> *"Even though for n finite there is no phase transition ($n \to \infty$) as a function of n or some other parameter the eigenvector of the sample PCA (e.g associated with eigenvalue λ_1) may exhibit a sharp loss of tracking suddenly losing its relation to the eigenvector of the population PCA."*

Nadler, B. *Finite sample approximation results for principal component analysis: a matrix perturbation approach.* Ann. Statist. 36 (2008), no. 6, 27912817.

Johnstone, I.M.
On the distribution of the largest eigenvalue in principal components analysis
Ann. Statist., 2001, **29**(2), 295–327

A model:

$$x_i = \mu + Au_i + \sigma z_i, \qquad i = 1, \dots, n$$

where

p	number of variables
M	number of spikes
x_i	observation p-vector
μ	p-vector of means
A	$p \times M$ factor loading matrix
u_i	M-vector of random factors
z_i	p-vector of white noise

Population Covariance Matrix

$$\Sigma = \sum_{j=1}^{M} \ell_j^2 q_j q_j^T + \sigma^2 I_p$$

where

Φ	$M \times M$ covariance matrix of u_i
$\ell_j, q_j, j = 1, \dots, M$	eigenvalues/vectors of $A\Phi A^T$

Ma, Z. *Sparse principal component analysis and iterative thresholding.*
Ann. Statist. 41 (2013), no. 2, 772801.

P-I
$$\frac{d^2y}{dx^2} = 6y^2 + x$$

P-II
$$\frac{d^2y}{dx^2} = 2y^3 + xy + \nu$$

P-III'
$$\frac{d^2y}{dx^2} = \frac{1}{y}\left(\frac{dy}{dx}\right)^2 - \frac{1}{x}\frac{dy}{dx} + \frac{y^2}{4x^2}(\gamma y + \alpha) + \frac{\beta}{4x} + \frac{\delta}{4y}, \quad \gamma = 4, \delta = -4$$

P-IV
$$\frac{d^2y}{dx^2} = \frac{1}{2y}\left(\frac{dy}{dx}\right)^2 + \frac{3}{2}y^3 + 4xy^2 + 2(x^2 - \alpha)y + \frac{\beta}{y}$$

P-V
$$\begin{aligned}\frac{d^2y}{dx^2} &= \left\{\frac{1}{2y} + \frac{1}{y-1}\right\}\left(\frac{dy}{dx}\right)^2 - \frac{1}{x}\frac{dy}{dx} \\ &\quad + \frac{(y-1)^2}{x^2}\left\{\alpha y + \frac{\beta}{y}\right\} + \frac{\gamma y}{x} + \frac{\delta y(y+1)}{y-1}\end{aligned}, \qquad \delta = -1/2$$

P-VI
$$\begin{aligned}\frac{d^2y}{dx^2} &= \frac{1}{2}\left\{\frac{1}{y} + \frac{1}{y-1} + \frac{1}{y-x}\right\}\left(\frac{dy}{dx}\right)^2 - \left\{\frac{1}{x} + \frac{1}{x-1} + \frac{1}{y-x}\right\}\frac{dy}{dx} \\ &\quad + \frac{y(y-1)(y-x)}{x^2(x-1)^2}\left\{\alpha + \frac{\beta x}{y^2} + \frac{\gamma(x-1)}{(y-1)^2} + \frac{\delta x(x-1)}{(y-x)^2}\right\}\end{aligned}$$

Classical Solution	P	Affine Weyl Group
—	$\mathrm{P_I}$	
Airy: $\mathrm{Ai}(x), \mathrm{Bi}(x)$	$\mathrm{P_{II}}$	A_1
Bessel: $I_\nu(x), K_\nu(x)$	$\mathrm{P_{III}}$	B_2
Hermite-Weber: $D_\nu(x)$	$\mathrm{P_{IV}}$	A_2
Confluent Hypergeometric: ${}_1F_1(a, c; x)$	$\mathrm{P_V}$	A_3
Gauss Hypergeometric: ${}_2F_1(a, b; c; x)$	$\mathrm{P_{VI}}$	D_4

- Muirhead, R. J., *Aspects of Multivariate Statistical Theory*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons Inc., 1982,
- Bai, Z. and Silverstein, J. W., *Spectral Analysis of Large Dimensional Random Matrices*, 2nd Edition, Springer, New York, 2010,
- Mehta, M. L., *Random Matrices*, Pure and Applied Mathematics (Amsterdam), Vol. 142, 3rd Edition, Elsevier/Academic Press, Amsterdam, 2004,
- Anderson, G. W. and Guionnet, A. and Zeitouni, O., *An Introduction to Random Matrices*, Cambridge University Press, Cambridge, 2010,
- Forrester, P. J., *Log Gases and Random Matrices*, Princeton University Press, 2010
- Akemann, G. and Baik, J. and Di Francesco, P., *Handbook on Random Matrix Theory*, Oxford University Press, 2011,
- Couillet, R. and Debbah, M. , *Random Matrix Methods for Wireless Communications*, Cambridge University Press, 2011

- Johnstone, I. M., *High Dimensional Statistical Inference and Random Matrices*, International Congress of Mathematicians. Vol. I, Eur. Math. Soc., Zürich, 2006, pp. 307-333
- Paul, D. and Aue, A, *Random matrix theory in statistics: A review* J. Statist. Plann. Inference, **150**, 1–29, 2014