Introduction to Random-Matrix Theory

by

Alan J. Izenman

Introduction

Random-matrix theory gained attention during the 1950s due to work by Eugene Wigner in mathematical physics. Specifically, Wigner wished to describe the general properties of the energy levels (or of their spacings) of highly excited states of heavy nuclei as measured in nuclear reactions (Wigner, 1957). Such a complex nuclear system is represented by an Hermitian operator $\mathbf{H}$ (called the *Hamiltonian*) living in an infinite-dimensional Hilbert space governed by physical laws. Unfortunately, in any specific case, $\mathbf{H}$ is unknown. Moreover, even if it were known, it would be much too complicated to write down and, even if we could write it down, no computer would be able to solve its eigenequation $\mathbf{Hv} = \lambda\mathbf{v}$ (the so-called Schrödinger equation of the physical system), where λ and $\mathbf{v}$ are an eigenvalue-eigenvector pair corresponding to $\mathbf{H}$.

Wigner argued that we should instead regard a specific Hamiltonian $\mathbf{H}$ as behaving like a large random matrix that is a member of a large class (or ensemble) of Hamiltonians, all of which would have similar general properties as the $\mathbf{H}$ in question (Wigner, 1955). The energy levels (represented by the eigenvalues of $\mathbf{H}$) of the physical system could then be approximated by the eigenvalues of a large random matrix. Furthermore, the spacings between energy levels of heavy nuclei could be more easily modelled by the spacings between successive eigenvalues of a random $(n \times n)$-matrix as $n \to \infty$.

Since the 1960s, Wigner and his colleagues, including Freeman Dyson and Madan Lal Mehta, worked on random-matrix theory and developed it to the point that it became a very powerful tool in mathematical physics (see Mehta, 2004). Dyson, in a series of papers in 1962, introduced a classification of three types of random-matrix ensembles based upon the property of time-reversal invariance. The matrices corresponding to these three types of random-matrix ensembles have elements that are complex (not time-reversal invariant), real (time-reversal invariant), or self-dual quaternion (time-reversal invariant, but with a restriction).

During the last decade or so, we have seen more interest paid to random-matrix theory. One of the most important early discoveries in random-matrix theory was its connection to quantum chaos (Bohigas, Giannoni, and Schmit, 1984), which led to a random-matrix theory of quantum transport (Beenakker,

1997). Random-matrix theory has since become a major tool in many fields, including number theory and combinatorics, wireless communications (Tulino and Verdú, 2004), and in multivariate statistical analysis and principal components analysis (Johnstone, 2001). A common element in these types of situations is that random-matrix theory has been used as an indirect method for solving complicated problems arising from physical or mathematical systems.

Much of the recent work on random matrices has tried to establish the so-called *universality conjecture*. This conjecture states that local behavior (i.e., fluctuation properties) of the eigenvalues of large random matrices have limits that are independent of the probability distribution on the matrix ensembles. Because this conjecture is not true in general, researchers have focused instead on showing that it is true within specific families of probability distributions, which in turn may depend upon the types of random-matrix ensembles considered.

Nomenclature

We will need the following terminology. An *ensemble* of random matrices is a family (or collection) of random matrices together with a probability density p that shows how likely it is that any member of the family can be observed.

Wigner and Dyson were most interested in approximating $\mathbf{H}$ by an ensemble of finite, large, $(n \times n)$ Hermitian matrices $\mathbf{H}_n$ whose probability density has the following form,

$$p(\mathbf{H}_n) \propto e^{-\beta \mathrm{tr}[V(\mathbf{H}_n)]}, \tag{1}$$

where V is some function of $\mathbf{H}_n$, such as a finite polynomial function of $\mathbf{H}_n$, where the highest power is even and its coefficient positive, and where the constant of proportionality depends only on n. For example, a possible choice of V could be

$$V(\mathbf{H}_n) = a\mathbf{H}_n^2 + b\mathbf{H}_n + c, \tag{2}$$

where a, b, and c are real numbers and $a > 0$. The entries of $\mathbf{H}_n = (H_{ij})$ can be real ($\beta = 1$), complex ($\beta = 2$), or real-quaternion ($\beta = 4$). If $V(\mathbf{H}_n) \propto \mathbf{H}_n^2$, then $\mathrm{tr}(\mathbf{H}_n^2) = \sum_i \sum_j H_{ij}^2$, and (1) reduces to a Gaussian ensemble. For example, in the case of a real, symmetric (2×2) random matrix,

$$\mathbf{H}_n = \begin{pmatrix} H_{11} & H_{12} \\ H_{12} & H_{22} \end{pmatrix}, \tag{3}$$

with independent elements, we have that $\mathrm{tr}(\mathbf{H}_n^2) = H_{11}^2 + H_{22}^2 + 2H_{12}^2$, so that each element is an independent Gaussian random variate and the variance of each off-diagonal element is one-half that of the diagonal elements.

We define a "time-reversal" transformation as

$$\mathbf{H}_n \rightarrow \mathbf{U}\mathbf{H}_n\mathbf{U}^{-1}, \tag{4}$$

where $\mathbf{U}$ is orthogonal ($\beta = 1$), unitary ($\beta = 2$), or symplectic ($\beta = 4$). (A symplectic matrix is a unitary matrix with real-quaternion elements.) *Time-reversal invariance* means that the time-reversal transformation leaves $p(\mathbf{H}_n)$ invariant (Porter and Rosenzweig, 1960).

Gaussian Ensembles

With time reversal as a motivation, Freeman Dyson introduced the classification of three types of Gaussian ensembles, *Gaussian orthogonal ensemble*, *Gaussian unitary ensemble*, and *Gaussian symplectic ensemble*. We define these ensembles as follows:

Gaussian Orthogonal Ensemble (GOE): $\mathbf{A}$ is an $(n \times n)$-matrix whose entries are each iid as $\mathcal{N}(0,1)$. A symmetric $(n \times n)$-matrix $\mathbf{H}_n$ is formed by setting $\mathbf{H}_n = (\mathbf{A} + \mathbf{A}^\tau)/2$, where $\mathbf{A}^\tau$ is the transpose of the matrix $\mathbf{A}$. The diagonal entries of $\mathbf{H}_n$ are distributed as iid $\mathcal{N}(0,1)$ and the off-diagonal entries are iid (subject to being symmetric) as $\mathcal{N}(0, \frac{1}{2})$.

Gaussian Unitary Ensemble (GUE): $\mathbf{A}$ is an $(n \times n)$-matrix whose entries are each complex-valued and iid as $\mathcal{N}^C(0,1)$. A Hermitian $(n \times n)$-matrix $\mathbf{H}_n$ is formed by setting $\mathbf{H}_n = (\mathbf{A} + \mathbf{A}^*)/2$, where $\mathbf{A}^*$ is the Hermitian transpose of the complex matrix $\mathbf{A}$. The diagonal entries of $\mathbf{H}_n$ are distributed as iid $\mathcal{N}(0,1)$ and the off-diagonal entries are iid (subject to being Hermitian) as $\mathcal{N}_2(0, \frac{1}{2})$. In other words, $\mathbf{H}_n = (H_{\ell m})$, where $H_{\ell m} = U_{\ell m} + iV_{\ell m}$, and $U_{\ell m}, V_{\ell m} \sim \mathcal{N}(0, \frac{1}{2})$, $1 \le \ell < m \le n$, and $H_{mm} \sim \mathcal{N}(0,1)$, $1 \le m \le n$.

Gaussian Symplectic Ensemble (GSE): $\mathbf{A}$ is an $(n \times n)$-matrix whose entries are each real-quarternian and iid as $\mathcal{N}^Q(0,1)$. A self-dual $(n \times n)$-matrix is formed by setting $\mathbf{H}_n = (\mathbf{A} + \mathbf{A}^D)/2$, where $\mathbf{A}^D$ denotes the dual transpose of the quarternian matrix $\mathbf{A}$. The diagonal entries of $\mathbf{H}_n$ are distributed as iid $\mathcal{N}(0,1)$ and the off-diagonal entries are iid (subject to being self-dual) as $\mathcal{N}_4(0, \frac{1}{2})$. In other words, if $\mathbf{H}_n = (H_{\ell m})$, then, $H_{\ell m} = U_{\ell m} + iV_{\ell m} + jW_{\ell m} + kZ_{\ell m}$, where $U_{\ell m}, V_{\ell m}, W_{\ell m}, Z_{\ell m} \sim \mathcal{N}(0, \frac{1}{2})$, for $1 \le \ell < m \le n$, and $H_{mm} \sim \mathcal{N}(0,1)$, $1 \le m \le n$.

For the GOE, $\mathbf{U}$ in (4) is orthogonal with real entries; for the GUE, $\mathbf{U}$ is unitary with complex entries; and for the GSE, $\mathbf{U}$ is symplectic with self-dual quaternion entries. See Table 1. It can be shown that the GOE is for systems that are time-reversal invariant, the GUE for systems that are not time-reversal invariant, and GSE for systems that are time-reversal invariant, but do not have spin-rotational symmetry. From a quantum-mechanical view, time-reversal invariance is the most realistic property, and, hence, the GOE is the most natural ensemble. If all three types of ensembles are time-reversal invariant and all elements of $\mathbf{H}_n$ are statistically independent, then the form of $p(\mathbf{H}_n)$ is automatically restricted to have $V(\mathbf{H}_n) = a\mathbf{H}_n^2 + b\mathbf{H}_n + c$, with a, b, and c real and $a > 0$ (Porter and Rosenzweig, 1960).

Table 1: *Dyson's classification of Gaussian ensembles. The Hermitian matrix* $\mathbf{H}_n = (H_{ij})$ *and its matrix of eigenvectors* $\mathbf{U}$ *are classified by the parameter* $\beta \in \{1, 2, 4\}$*, depending upon the presence or absence of time-reversal invariance (TRI) and spin-rotational symmetry (SRS). NA means 'not appropriate.' (Adapted from Table 1 in Beenakker, 1996.)*

β	Ensemble	TRI	SRS	H_{ij}	$\mathbf{U}$
1	GOE	Yes	Yes	real	orthogonal
2	GUE	No	NA	complex	unitary
4	GSE	Yes	No	real-quaternion	symplectic

Spectrum of Random Matrices

Of particular interest is the stochastic behavior of the *bulk* and the *extremes* (or *edges*) of the spectrum of large random matrices. The bulk deals with most of the eigenvalues of a given matrix and the extremes refer to the largest and smallest of those eigenvalues. Note that regardless of which ensemble we study, the eigenvalues of $\mathbf{H}_n$ are all real and can, therefore, be rank-ordered. The extremes (and especially the smallest eigenvalue) of the spectrum are important in determining the stability and invertability of a square matrix. Recent work has shown that there are differences between the statistics of the bulk of the spectrum and those of the extreme eigenvalues at the edge of the spectrum.

One of the main features of all three random matrix ensembles is the idea of *repulsion*, that any two (correlated) eigenvalues obtained from a GOE (or GUE or GSE) matrix are unlikely to be close together; that is, the probability that adjacent eigenvalues are close together is small, and the probability quickly goes to zero as a power of the distance between them. Hence, the spacings distribution precludes near-zero spacings. This property is related to certain aspects of quantum chaos (Kiecherbauer, Marklof and Soshnikov, 2001).

Bulk of the Spectrum

Gaussian Case: The Wigner Matrix

Wigner originally studied a real symmetric $(n \times n)$-matrix $\mathbf{H}_n = (H_{ij})$ where the diagonal entries were each 0 and the off-diagonal entries (subject to the symmetric constraint) were independently ± 1 with probability $\frac{1}{2}$. He later realized that his results for this matrix would continue to hold more generally.

Let $\mathbf{A}$ be an $(n \times n)$-matrix filled with independent and identically distributed standard Gaussian deviates,

$$A_{ij} \sim \mathcal{N}(0, 1), \quad i, j = 1, 2, \ldots, n, \tag{5}$$

where we assume n is large. Then,

$$\mathbf{H}_n = \frac{1}{2}(\mathbf{A} + \mathbf{A}^\tau) \tag{6}$$

is a symmetric (i.e., $\mathbf{H}_n = \mathbf{H}_n^\tau$) random $(n \times n)$-matrix, also known as a *Wigner matrix*, where the ijth entry is given by

$$H_{ij} \sim \mathcal{N}(0, \sigma_{ij}), \quad \sigma_{ij} = \frac{1}{2}(1 + \delta_{ij}), \tag{7}$$

$\delta_{ij} = 1$ if $i = j$ and 0 otherwise. Thus, the diagonal entries of a Wigner matrix are distributed as $\mathcal{N}(0, 1)$, while the off-diagonal entries are distributed as $\mathcal{N}(0, \frac{1}{2})$. The Wigner matrix (7) is a member of the *Gaussian Orthogonal Ensemble*. Studying the behavior of Wigner matrices forms a large part of random-matrix theory.

Remark: The Gaussian assumption for the Wigner matrix is not necessary, unless it is required to be a member of the GOE. In general, it could be replaced by a distribution having the same independence properties and same variance as the Gaussian.

Finite n: Exact Distribution

Let $\lambda_1 > \lambda_2 > \cdots > \lambda_n$ be the ordered eigenvalues of $\mathbf{H}_n$ and let $\mathbf{U}$ be the matrix of associated eigenvectors. Then, $\mathbf{H}_n = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\tau$, where $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \ldots, \lambda_n\}$. Now, $\text{tr}[V(\mathbf{H}_n)] = \sum_{j=1}^n V(\lambda_j)$ depends only on the eigenvalues. Thus, the distribution $p(\mathbf{H}_n)$ in (1) is independent of the eigenvectors, which can be viewed as being uniformly distributed over the members of each matrix ensemble. The exact joint probability distribution of the eigenvalues of an $(n \times n)$ Wigner matrix is then found by multiplying $p(\mathbf{H}_n)$ by the Jacobian of the transformation from the matrix to its eigenvalues and eigenvectors.

The exact distribution of the eigenvalues, therefore, has the form,

$$p(\lambda_1, \ldots, \lambda_n) = c_n \prod_{j=1}^n [w(\lambda_j)]^{1/2} \prod_{1 \le j < k \le n} |\lambda_j - \lambda_k|, \tag{8}$$

where

$$w(\lambda) = e^{-\lambda^2}, \quad \lambda \in \Re, \tag{9}$$

is the weight function for the Hermite family of orthogonal polynomials (Abromowitz and Stegun, 1970, Table 22.2) and c_n is a normalizing constant dependent upon n.

For general β-Gaussian-Hermite ensembles, where $\beta = 1$ (GOE), 2 (GUE), or 4 (GSE), the joint probability density of the eigenvalues of $\mathbf{H}_n$ is given by

$$p_\beta(\lambda_1, \cdots, \lambda_n) = c'_{n,\beta} \prod_{j=1}^n [w_\beta(\lambda_j)]^{1/2} \prod_{1 \le j < k \le n} |\lambda_j - \lambda_k|^\beta, \tag{10}$$

where

$$w_\beta(\lambda) = e^{-\beta\lambda^2}, \tag{11}$$

and $c'_{n,\beta}$ is a normalizing constant,

$$c'_{n,\beta} = (2\pi)^{-n/2}\beta^{n/2+\beta n(n-1)/4}\prod_{i=1}^{n}\frac{\Gamma(1+\frac{\beta}{2})}{\Gamma(1+\frac{\beta j}{2})}, \tag{12}$$

dependent upon n and β (Mehta, 2004, p. 58).

Large n: Wigner's Semi-Circle Law

Next, consider the limiting probability distribution of the eigenvalues of a Wigner matrix $\mathbf{H}_n$ as n tends to infinity. We define the *empirical distribution function* of the eigenvalues, $\lambda_1, \lambda_2, \ldots, \lambda_n$, of $\mathbf{H}_n$ as

$$\frac{1}{n}\#\{i:\ \lambda_i \le \lambda\} = \frac{1}{n}\sum_{i=1}^{n} I_{[\lambda_i \le \lambda]}, \tag{13}$$

where I_A is the indicator function of the event A ($I_A = 1$ if A is true and 0 otherwise) and $\#\{\cdot\}$ denotes the number of elements in the set indicated. Wigner's result says that the empirical distribution (13) of the eigenvalues of $\mathbf{H}_n$ converges a.s. to a nonrandom limiting distribution $G(\lambda)$,

$$\frac{1}{n}\#\{i:\ \lambda_i \le \lambda\} \overset{a.s.}{\to} G(\lambda), \quad n \to \infty, \tag{14}$$

with density,

$$g(x) = \frac{1}{2\pi}\sqrt{4-\lambda^2}, \quad |\lambda| \le 2, \tag{15}$$

and zero for $|\lambda| > 2$. This limiting density is a semi-circle with radius 2 (Wigner, 1955, 1958).

Remark: For (14) and (15) to hold, we only need the existence of second moments for the off-diagonal entries; we do not need such a moment requirement for the diagonal entries.

Remark: Wigner's semi-circle law still holds if the Gaussian assumption (7) in the definition of Wigner's matrix $\mathbf{H}_n$ is replaced by any symmetric distribution (discrete or continuous) with mean zero, finite variance σ^2, and finite higher moments. Under symmetry, all odd moments are zero.

Figure 1 shows an illustration of the convergence to Wigner's semi-circle law for a single $(n \times n)$ Wigner matrix. We sampled $n = 1,000$ (left panel) and $n = 25,000$ (right panel) iid standard Gaussian deviates, computed $\mathbf{A}$ and then $\mathbf{H}_n$, and found the eigenvalues of $\mathbf{H}_n$. The histograms of the eigenvalues of $\mathbf{H}_n$ for both cases are given in Figure 1.

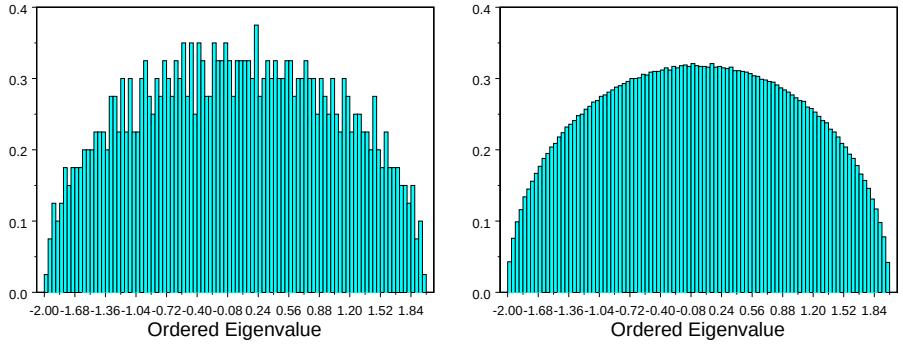

Figure 1: *Illustration of the convergence to Wigner's Semi-Circle Law. Normalized histograms of the eigenvalues from a single* $(n \times n)$ *Wigner matrix. Left panel:* $n = 1,000$. *Right panel:* $n = 25,000$. *For each* n, *there are 100 bins.*

Single Wishart Matrix

In multivariate statistical analysis, we are often interested in a random r-vector $\mathbf{X}$ that is distributed with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, where

$$\boldsymbol{\mu} = \mathrm{E}\{\mathbf{X}\}, \quad \boldsymbol{\Sigma} = \mathrm{E}\{(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^\tau\}. \tag{16}$$

We may sometimes need to assume that $\mathbf{X}$ is also Gaussian. Many topics in multivariate analysis, (e.g., principal component analysis, factor analysis, and multidimensional scaling) deal with the study of functions of $\boldsymbol{\Sigma}$, such as its eigenvalues and associated eigenvectors.

Typically, $\boldsymbol{\Sigma}$ is unknown, and so has to be estimated using a sample of data. Given a set of independent random r-vectors, $\mathbf{X}_i, i = 1, 2, \ldots, n$, drawn from the same underlying distribution as $\mathbf{X}$, the usual estimate of $\boldsymbol{\Sigma}$ is given by

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^\tau = \frac{1}{n}\mathcal{X}_c\mathcal{X}_c^\tau, \tag{17}$$

where the sample mean vector, $\bar{\mathbf{X}} = n^{-1}\sum_{i=1}^{n}\mathbf{X}_i$, is an estimator of the population mean vector $\boldsymbol{\mu}$. In (17), $\mathcal{X} = (\mathbf{X}_1, \cdots, \mathbf{X}_n)$ and $\mathcal{X}_c = \mathcal{X}(\mathbf{I}_n - n^{-1}\mathbf{J}_n)$, where $\mathbf{J}_n = \mathbf{1}_n\mathbf{1}_n^\tau$ and $\mathbf{1}_n$ is an n-vector of 1s. We can make $\widehat{\boldsymbol{\Sigma}}$ an unbiased estimator of $\boldsymbol{\Sigma}$ by replacing $1/n$ by $1/(n-1)$ in the averaging operation in (17). We are interested in studying the behavior of the eigenvalues of $\widehat{\boldsymbol{\Sigma}}$ as estimators of the population eigenvalues of $\boldsymbol{\Sigma}$. Here, we consider the distribution of the eigenvalues of $\widehat{\boldsymbol{\Sigma}}$ under different assumptions on the number of variables r and the number of observations n.

A Distributional Result

We will need the following result below. Let $\mathbf{A}$ be an $(r \times r)$ positive-definite matrix with density function $p(\mathbf{A})$. The joint density of the eigenvalues $\lambda_1 >$

$\lambda_2 > \cdots > \lambda_r$ of $\mathbf{A}$ is given by (Muirhead, 1982, Theorem 3.2.17)

$$p(\lambda_1, \ldots, \lambda_r) = \frac{\pi^{r^2/2}}{\Gamma_r(r/2)} \prod_{1 \le j < k \le r} |\lambda_j - \lambda_k| \int_{\mathcal{O}(r)} p(\mathbf{Q}\mathbf{L}\mathbf{Q}^\tau)(d\mathbf{Q}), \qquad (18)$$

where $(d\mathbf{Q})$ is the Haar invariant measure on the set $\mathcal{O}(r)$ of $(r \times r)$ orthogonal matrices, normalized so that $\int_{\mathcal{O}(r)}(d\mathbf{Q}) = 1$. In (18), the function Γ_r is a multivariate gamma function defined by (Muirhead, 1982, Section 2.1.2)

$$\Gamma_r(x) = \pi^{r(r-1)/4} \prod_{j=1}^{r} \Gamma\left(x - \frac{j-1}{2}\right), \quad \mathrm{Re}(x) > \frac{r-1}{2}. \qquad (19)$$

The product in (18) involving the pairwise differences of eigenvalues is the Jacobian term, and is the determinant of the *Vandermonde matrix*,

$$\mathbf{V}_r = \begin{pmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{r-1} \\ 1 & \lambda_2 & \lambda_2^2 & \cdots & \lambda_2^{r-1} \\ 1 & \lambda_3 & \lambda_3^2 & \cdots & \lambda_3^{r-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_n & \lambda_n^2 & \cdots & \lambda_n^{r-1} \end{pmatrix}; \qquad (20)$$

the determinant of (20) is also known as the *Vandermonde determinant* (see, e.g., Bellman, 1960, p. 193).

Fixed r, Finite n, and $r < n$: Exact Distributions

Prior to the public availability of high-speed computation and large data storage facilities, the number of variables r was kept reasonably small and the number of observations n, though larger than r, was still small by modern standards. Distribution theory was either exact (fixed r and finite n) or asymptotic with a fixed r and $n \to \infty$. These are the two cases we deal with first.

Without loss of generality, suppose $\boldsymbol{\mu} = \mathbf{0}$. Suppose also that $\mathbf{X}_i \overset{iid}{\sim} \mathcal{N}_r(\mathbf{0}, \boldsymbol{\Sigma})$, $i = 1, 2, \ldots, n$, and set $\mathcal{X} = (\mathbf{X}_1, \cdots, \mathbf{X}_n)$. Then,

$$n\widehat{\boldsymbol{\Sigma}} = \mathbf{S} = \mathcal{X}\mathcal{X}^\tau \sim \mathcal{W}_r(n, \boldsymbol{\Sigma}) \qquad (21)$$

(Wishart, 1928). If $\boldsymbol{\Sigma} = \mathbf{I}_r$, then this is the so-called real *white Wishart* distribution.

When $\mathbf{S} \sim \mathcal{W}_r(n, \boldsymbol{\Sigma})$, we substitute the form of the Wishart density,

$$p(\mathbf{S}|n, \boldsymbol{\Sigma}) = c_{r,n}|\boldsymbol{\Sigma}|^{-n/2}|\mathbf{S}|^{(n-r-1)/2} e^{-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S})}, \qquad (22)$$

where

$$c_{r,n}^{-1} = 2^{nr/2}\pi^{r(r-1)/4} \prod_{i=1}^{r} \Gamma\left(\frac{n-i+1}{2}\right), \qquad (23)$$

into (18) to obtain the exact joint distribution of the eigenvalues, $\lambda_1 > \lambda_2 > \cdots > \lambda_r$, of $\mathbf{S}$ (James, 1964),

$$p(\lambda_1, \ldots, \lambda_r) = c_{r,n} \prod_{j=1}^{r} \lambda_j^{(n-r-1)/2} \prod_{1 \le j < k \le r} |\lambda_j - \lambda_k| \int_{\mathcal{O}(r)} e^{-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{Q}\mathbf{L}\mathbf{Q}^\tau)} (d\mathbf{Q}), \tag{24}$$

where $\mathbf{L} = \mathrm{diag}\{\lambda_1, \cdots, \lambda_r\}$, $|\mathbf{L}| = \prod_{j=1}^{r} \lambda_j$, $(d\mathbf{Q})$ is the Haar invariant measure on the set $\mathcal{O}(r)$ of $(r \times r)$ orthogonal matrices, normalized so that $\int_{\mathcal{O}(r)} (d\mathbf{Q}) = 1$, and $c_{r,n}$ is the normalization constant,

$$c_{r,n} = \frac{\pi^{r^2/2}}{2^{nr/2} |\boldsymbol{\Sigma}|^{n/2} \Gamma_r(r/2) \Gamma_r(n/2)}. \tag{25}$$

The integral in (24) over the orthogonal group $\mathcal{O}(r)$ is difficult to evaluate in the case of general $\boldsymbol{\Sigma}$. Some efforts in this direction have been made using infinite series expansions in zonal polynomials, but these have not yielded practical results.

If $\boldsymbol{\mu} \neq \mathbf{0}$, then the sample covariance matrix $\mathbf{S} = n^{-1}\mathcal{X}_c\mathcal{X}_c^\tau$ has the Wishart distribution $\mathcal{W}_r(n-1, n^{-1}\boldsymbol{\Sigma})$. In this case, the previous results, (24) and (25), for the eigenvalue density can be modified by substituting $n-1$ for n and $n\lambda_i$ for λ_i.

In the white Wishart case (i.e., $\boldsymbol{\Sigma} = \mathbf{I}_r$), the integral over the orthogonal group $\mathcal{O}(r)$ is easily evaluated to be $e^{-\frac{1}{2}\sum_j \lambda_j}$. The resulting density (24) reduces to

$$p(\lambda_1, \ldots, \lambda_r) = c_{r,n} \prod_{j=1}^{r} [w_{n,r}(\lambda_j)]^{1/2} \prod_{1 \le j < k \le r} |\lambda_j - \lambda_k|, \tag{26}$$

where

$$w_{n,r}(\lambda) = \lambda^{n-r-1} e^{-\lambda}, \quad \lambda \in [0, \infty), \quad n > r, \tag{27}$$

is the weight function for a generalized Laguerre family of orthogonal polynomials (Abromowitz and Stegun, 1970, Table 22.2), and $c_{r,n}$ is a normalizing constant dependent upon r and n. For a proof, see Anderson (1984, Section 13.3). The second product in (26) involving the pairwise differences of eigenvalues is, as before, the Jacobian term, and is the determinant of the *Vandermonde matrix* (10). The eigenvalue density (26) was found independently and simultaneously by Fisher, Girshick, Hsu, and Roy in 1939, and in 1951 independently by Mood.

For general β-Wishart-Laguerre ensembles, where $\beta = 1$, 2, or 4 for real, complex, or quaternian Gaussian entries of $\mathcal{X}$, respectively, the joint probability density of the eigenvalues of $\mathbf{S}$ is given by

$$p_\beta(\lambda_1, \cdots, \lambda_r) = c'_{n,r,\beta} \prod_{j=1}^{r} [w_{n,r,\beta}(\lambda_j)]^{1/2} \prod_{1 \le j < k \le r} |\lambda_j - \lambda_k|^\beta, \tag{28}$$

where

$$w_{n,r,\beta}(\lambda) = \lambda^{\beta(n-r+1)-2} e^{-\beta\lambda}, \quad \lambda \in [0, \infty), \tag{29}$$

and $c'_{n,r,\beta}$ is a normalizing constant,

$$c'_{n,r,\beta} = 2^{-\beta nr/2} \prod_{i=1}^{n} \frac{\Gamma(1 + \frac{\beta}{2})}{\Gamma(1 + \frac{\beta i}{2})\Gamma(\frac{\beta}{2}(r - n + i))}, \tag{30}$$

which is dependent upon n, r, and β.

In the case of general $\mathbf{\Sigma}$, when the population eigenvalues are not all equal, the exact joint distribution of the sample eigenvalues is known (James, 1960) but is extremely complicated, involving zonal polynomials (i.e., power-series expansions in hypergeometric functions). For large n, the zonal polynomial series converges very slowly, and so the results have very limited practical value (James, 1964).

Fixed r, Large n

For fixed r and large n, and $\mathbf{X} \sim \mathcal{N}_r(\boldsymbol{\mu}, \mathbf{\Sigma})$, the sample eigenvalues, $\widehat{\lambda}_j$, $j = 1, 2, \ldots, r$, of $\widehat{\mathbf{\Sigma}}$ are jointly asymptotically independently distributed according to

$$\sqrt{n}(\widehat{\lambda}_j - \lambda_j) \overset{\mathcal{D}}{\Rightarrow} \mathcal{N}(0, 2\lambda_j^2), \quad \text{as } n \to \infty, \ j = 1, 2, \ldots, r, \tag{31}$$

where the $\{\lambda_j\}$ are the distinct eigenvalues of $\mathbf{\Sigma}$ (Anderson, 1963). This result shows that the jth sample eigenvalue, $\widehat{\lambda}_j$ is a consistent estimator of the jth population eigenvalue, λ_j, $j = 1, 2, \ldots, r$.

Large r, Large n: The Marčenko–Pastur's Quarter-Circle Law

One of the most important results of random-matrix theory for use in multivariate analysis is the Marčenko–Pastur Law, which is an analogue of Wigner's Semi-Circle Law. The Marčenko–Pastur Law gives the limiting distribution of the eigenvalues of a sample covariance matrix (as the size of the matrix grows without bound) in the null case when $\mathbf{S} \sim \mathcal{W}_r(n, \mathbf{I}_r)$. In the null case, all the eigenvalues of the population covariance matrix $\mathbf{I}_r$ are equal to one. Although $\widehat{\mathbf{\Sigma}}$ is a good approximation to $\mathbf{\Sigma}$ for fixed r and large n, that does not hold when r and n are both large.

If we let $r \to \infty$ and $n \to \infty$ in such a way that the *matrix aspect ratio* converges to a non-zero constant, i.e., $r/n \to \gamma \in (0, \infty)$, then the empirical distribution of the eigenvalues, $\widehat{\lambda}_i, i = 1, 2, \ldots, r$, of $\widehat{\mathbf{\Sigma}}$ follows the Marčenko–Pastur Law:

$$\frac{1}{r} \#\{i : \widehat{\lambda}_i \leq x\} \overset{a.s.}{\to} G(x), \tag{32}$$

where $G(x)$ has density $g(x) = G'(x)$ and

$$g(x) = \frac{1}{2\pi\gamma x} \sqrt{(b_+ - x)(x - b_-)}\, I_{[b_-, b_+]}(x), \quad b_\pm = (1 \pm \sqrt{\gamma})^2, \tag{33}$$

where $I_{[b_-,b_+]}(x)$ is the indicator function that is equal to 1 for $b_- < x < b_+$ and 0 otherwise (Marčenko and Pastur, 1967). This is the so-called *Quarter-Circle Law*. Note that the limiting density only depends upon γ. If $\gamma \in (0,1)$, then $r < n$; in this case, the spectra of $\mathcal{X}\mathcal{X}^\tau$ and $\mathcal{X}^\tau\mathcal{X}$ differ by $n-r$ zero eigenvalues, and so there is an additional point mass at the origin $x = 0$ with weight $1-\gamma$. Except for this point mass at the origin, the results given here for the nonzero eigenvalues hold regardless of whether r or n is larger.

The largest sample eigenvalue, $\widehat{\lambda}_1$, converges a.s. to the right-hand support point $b_+ = (1+\sqrt{\gamma})^2$, and the smallest sample eigenvalue, $\widehat{\lambda}_r$ converges a.s. to the left-hand support point $b_- = (1-\sqrt{\gamma})^2$. See also Geman (1980), who first proved the convergence result for the largest sample eigenvalue. However, $\widehat{\lambda}_1$ is not a consistent estimator of λ_1. For example, in the case when $r = n$ (i.e., $\gamma = 1$) and $\mathbf{\Sigma} = \mathbf{I}_r$, the largest population eigenvalue $\lambda_1 = 1$, while the largest sample eigenvalue, $\widehat{\lambda}_1$, of $\mathbf{S}$ converges to the value 4. If $r > n$, then $\widehat{\lambda}_{n+1} = \cdots = \widehat{\lambda}_r = 0$. The results (32) and (33), unfortunately, remained obscure for a while (see, e.g., Wachter, 1978, who derived similar results apparently unaware of the Marčenko-Pastur paper).

A visual representation of the Marčenko–Pastur Law is given in Figure 2, where we have separated the values of γ by $\gamma \leq 1$ (left panel) and $\gamma \geq 1$ (right panel). We see that even though all the population eigenvalues are equal to 1, the spread of the sample eigenvalues varies directly with the ratio $\gamma = r/n$: the larger the ratio (i.e., the bigger r is relative to n), the more spread out are the sample eigenvalues. For example, when $\gamma = 1/4$, the density is supported on the interval $[\frac{1}{4}, \frac{9}{4}]$, when $\gamma = 1$ (i.e., $n = r$), the density is supported on $[0,4]$, and when $\gamma = 4$, the density is supported on $[1,9]$.

The a.s. result is fascinating, but for statistical inference purposes we would also like to have some insight into second-order information (i.e., variability) about the bulk of the sample eigenvalues, especially the largest eigenvalue, which is of importance in principal component analysis.

Two Wishart Matrices

Fixed r, Finite n: Exact Distribution

Real case. Let $\mathbf{X}_i \sim \mathcal{N}_r(\boldsymbol{\mu}, \mathbf{\Sigma})$, $i = 1, 2, \ldots, m$, and let $\mathcal{X} = (\mathbf{X}_1, \cdots, \mathbf{X}_m)$ be an $(r \times m)$-matrix. Then, $\mathbf{A} = \mathcal{X}\mathcal{X}^\tau \sim \mathcal{W}_r(m, \mathbf{\Sigma})$. Suppose we have another $(r \times r)$-matrix $\mathbf{B} \sim \mathcal{W}_r(n, \mathbf{\Sigma})$ that is independent of $\mathbf{A}$. Because we are interested in the eigenvalues of $\mathbf{B}^{-1}\mathbf{A}$ and because the distribution of those eigenvalues does not depend upon $\mathbf{\Sigma}$, without loss of generality, we can take $\mathbf{\Sigma} = \mathbf{I}_r$.

So, suppose we have two independent white Wishart matrices, $\mathbf{A} \sim \mathcal{W}_r(m, \mathbf{I}_r)$ and $\mathbf{B} \sim \mathcal{W}_r(n, \mathbf{I}_r)$. If $m, n \geq r$, then both $\mathbf{A}$ and $\mathbf{B}$ are invertable as is also their sum $\mathbf{A} + \mathbf{B}$. We are interested in solving (for λ) the following generalized eigenvalue problem,

$$|\mathbf{A} - \lambda(\mathbf{A} + \mathbf{B})| = 0. \tag{34}$$

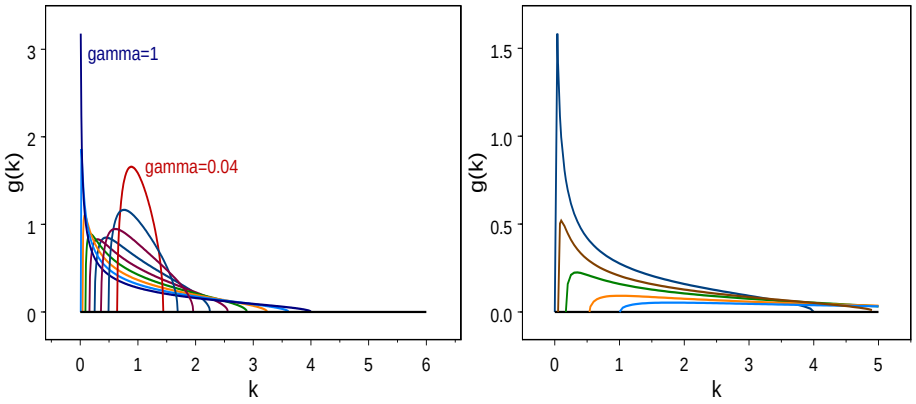

Figure 2: *Density of eigenvalues from the Marčenko-Pastur Law. Left panel:* $\gamma = 0.04, 0.09, 0.16, 0.25, 0.36, 0.49, 0.64, 0.81, 1$ *(i.e.,* $r \leq n$*). Right panel:* $\gamma = 1, 1.5, 2, 3, 4$ *(i.e.,* $r \geq n$*).*

That is, we are interested in the eigenvalues of $(\mathbf{A} + \mathbf{B})^{-1}\mathbf{A}$. Because $\mathbf{B}$ is positive definite, $0 < \lambda < 1$. The eigenequation (34) can be reexpressed as

$$|\mathbf{A} - \theta\mathbf{B}| = 0, \tag{35}$$

and, in this form, we are interested in the eigenvalues of $\mathbf{B}^{-1}\mathbf{A}$. The eigenvalues λ and θ are related by $\lambda = \theta/(1+\theta)$ or $\theta = \lambda/(1-\lambda)$.

The exact joint distribution of the eigenvalues of the generalized eigenequation (34) is given by

$$p(\lambda_1, \ldots, \lambda_r) = c_{m,n,r} \prod_i [w_{a,b}(\lambda_i)]^{1/2} \prod_{i<j} |\lambda_i - \lambda_j|, \tag{36}$$

where

$$w_{a,b}(\lambda) = \lambda^a (1-\lambda)^b, \quad a = m - r - 1, \; b = n - r - 1, \tag{37}$$

is a weight function for the Jacobi family of orthogonal polynomials (Abromowitz and Stegun, 1970, Table 22.2), and c is a normalizing constant that depends upon m, n, and r.

The general β-form of (36) is given by

$$p_\beta(\lambda_1, \ldots, \lambda_r) = c_{m,n,r,\beta} \prod_i [w_{r,m,n,\beta}(\lambda_i)]^{1/2} \prod_{i<j} |\lambda_i - \lambda_j|^\beta, \tag{38}$$

where

$$w_{r,m,n,\beta}(\lambda) = \lambda^a (1-\lambda)^b, \quad a = \beta(m - r + 1) - 2, \; b = \beta(n - r + 1) - 2, \tag{39}$$

and $c_{m,n,r,\beta}$ is a normalizing constant that depends upon m, n, r, and β. As before, $\beta = 1$ for the real case, $\beta = 2$ for the complex case, and $\beta = 4$ for the quaternion case.

Setting $\lambda_i = \theta_i/(1+\theta_i)$, the joint distribution of the eigenvalues of the generalized eigenequation (35) is given by

$$p(\theta_1, \ldots, \theta_r) = c_{m,n,r} \prod_i [w_{a,b}(\theta_i)]^{1/2} \prod_{i<j} |\theta_i - \theta_j|, \tag{40}$$

where

$$w_{a,b}(\theta) = \theta^a (1+\theta)^b, \quad a = m-r-1, \quad b = m+n. \tag{41}$$

Proofs of these results can be found in Anderson (1984, Section 13.2).

If we carry out a change of variable in (40) and (41) by setting $\lambda = (1+x)/2$, we obtain the *Jacobi orthogonal ensemble*,

$$p(x_1, \cdots, x_r) = c_{m,n,r} \prod_i [w_{a,b}(x_i)]^{1/2} \prod_{i<j} |x_i - x_j|, \tag{42}$$

where

$$w_{a,b}(x) = (1-x)^a (1+x)^b, \quad a = m-r-1, \quad b = n-r-1, \tag{43}$$

is the weight function for the Jacobi family of orthogonal polynomials (Abromowitz and Stegun, 1970, Table 22.2).

Complex case. We write $\mathbf{X} \sim \mathcal{N}_r^C(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, for a complex-valued random r-vector having a multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. This complex-valued random vector $\mathbf{X}$ has $2r$ real components given by $(\mathrm{Re}(\mathbf{X}), \mathrm{Im}(\mathbf{X}))^\tau$ with distribution (Brillinger, 1975, Section 4.2)

$$\mathcal{N}_{2r}\left(\begin{pmatrix} \mathrm{Re}(\boldsymbol{\mu}) \\ \mathrm{Im}(\boldsymbol{\mu}) \end{pmatrix}, \frac{1}{2} \begin{pmatrix} \mathrm{Re}(\boldsymbol{\Sigma}) & -\mathrm{Im}(\boldsymbol{\Sigma}) \\ \mathrm{Im}(\boldsymbol{\Sigma}) & \mathrm{Re}(\boldsymbol{\Sigma}) \end{pmatrix} \right). \tag{44}$$

If $\mathbf{X}_i \sim \mathcal{N}_r^C(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $i = 1, 2, \ldots, n$, then $\mathcal{X}\mathcal{X}^* = \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^* \sim \mathcal{W}_r^C(n, \boldsymbol{\Sigma})$, where $\mathcal{X}^*$ denotes the complex-conjugate transpose of $\mathcal{X} = (\mathbf{X}_1, \cdots, \mathbf{X}_n)$.

Suppose now that we have two independent complex-Wishart $(r \times r)$ random matrices, $\mathbf{A} \sim \mathcal{W}_r^C(m, \mathbf{I}_r)$ and $\mathbf{B} \sim \mathcal{W}_r^C(n, \mathbf{I}_r)$. Then, the exact joint density of the eigenvalues $1 \geq \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r \geq 0$ of $(\mathbf{A}+\mathbf{B})^{-1}\mathbf{A}$ is given by

$$p(\lambda_1, \cdots, \lambda_r) = c_{m,n,r} \prod_i w(\lambda_i) \prod_{i<j} |\lambda_i - \lambda_j|^2. \tag{45}$$

where

$$w_{a,b}(\lambda) = \lambda^a (1-\lambda)^b, \quad a = m-r, \; b = n-r. \tag{46}$$

Application to Multivariate Regression

The eigenproblem (35) is of interest in multivariate reduced-rank regression, which includes as special cases canonical variate and correlation analysis, and

Table 2: *Families of orthogonal polynomials and their weight functions $w(x)$.*

Case	$w(x)$	Interval	OrthoPoly
Gaussian	e^{-x^2}	$(-\infty, \infty)$	Hermite
Wishart	$x^a e^{-x}$	$[0, \infty)$	Laguerre
Two Wisharts	$x^a(1-x)^b$	$(0, 1)$	Jacobi

linear discriminant analysis. Suppose $\mathcal{X}_c$ is an $(r \times n)$-matrix and $\mathcal{Y}_c$ is an $(s \times n)$-matrix, where the subscript c indicates that both $\mathcal{X}$ and $\mathcal{Y}$ are centered (by subtracting out row means from each row), and where we assume $s \leq r$.

Set $\mathbf{S}_{XX} = \mathcal{X}_c\mathcal{X}_c^\tau$, $\mathbf{S}_{YY} = \mathcal{Y}_c\mathcal{Y}_c^\tau$, and $\mathbf{S}_{XY} = \mathcal{X}_c\mathcal{Y}_c^\tau = \mathbf{S}_{YX}^\tau$. Then set $\mathbf{A} = \mathbf{S}_{YX}\mathbf{S}_{XX}^{-1}\mathbf{S}_{XY}$ to be the variation due to the multivariate regression and $\mathbf{B} = \mathbf{S}_{YY} - \mathbf{S}_{YX}\mathbf{S}_{XX}^{-1}\mathbf{S}_{XY}$ to be the residual variation, so that $\mathbf{A} + \mathbf{B} = \mathbf{S}_{YY}$. We have that $\mathbf{A} \sim \mathcal{W}_s(r, \mathbf{\Sigma}_{YY})$ and $\mathbf{B} \sim \mathcal{W}_s(n-r-1, \mathbf{\Sigma}_{YY})$. The eigenequation (34), thus, boils down to the following:

$$|\mathbf{S}_{YX}\mathbf{S}_{XX}^{-1}\mathbf{S}_{XY} - \lambda\mathbf{S}_{YY}| = 0, \tag{47}$$

so that we are interested in the eigenvalues of $\mathbf{S}_{YY}^{-1}\mathbf{S}_{YX}\mathbf{S}_{XX}^{-1}\mathbf{S}_{XY}$, or of its symmetric version, $\mathbf{S}_{YY}^{-1/2}\mathbf{S}_{YX}\mathbf{S}_{XX}^{-1}\mathbf{S}_{XY}\mathbf{S}_{YY}^{-1/2}$. The joint distribution of the eigenvalues of the generalized eigenequation (34) is given by

$$p(\lambda_1, \ldots, \lambda_s) = c_{n,r,s} \prod_{i=1}^{s} [w_{a,b}(\lambda_i)]^{1/2} \prod_{i<j}^{s} |\lambda_i - \lambda_j|, \tag{48}$$

where

$$w_{a,b}(\lambda) = \lambda^a(1-\lambda)^b, \tag{49}$$

$$a = r - s - 1, \quad b = n - r - s - 2, \tag{50}$$

and c is a normalizing constant that depends upon n, r, and s (Anderson, 1984, Section 13.4).

Edges of the Spectrum

Perhaps the most exciting results to have been derived from random-matrix theory are the Tracy–Widom laws for the distribution of the appropriately-normalized largest eigenvalue of a random matrix.

Largest Eigenvalue: Gaussian Ensembles

For the GUE, Tracy–Widom (1994) showed that the largest eigenvalue has the limiting distribution,

$$\lim_{n\to\infty} P\left\{\frac{\widehat{\lambda}_1 - 2\sqrt{n}}{n^{1/6}} \leq t\right\} = F_2(t), \tag{51}$$

Table 3: *Percentage points of the Tracy–Widom distributions for $\beta = 1, 2, 4$. Tabulated is the value of x such that $F_\beta(x) = P\{W_\beta < x\} = p$. (Adapted from Table 1 of Bejan, 2004.)*

β	0.005	0.025	0.05	0.95	0.975	0.99	0.995	0.999
1	–4.1505	–3.5166	–3.1808	0.9703	1.4538	2.0234	2.4224	3.2724
2	–3.9139	–3.4428	–3.1945	–0.2325	0.0915	0.4776	0.7462	1.3141
4	–4.0531	–3.6608	–3.4556	–1.0904	–0.8405	–0.5447	–0.3400	0.0906

where F_2 is the *Tracy–Widom law of order 2* with distribution function,

$$F_2(t) = \exp\left(-\int_t^\infty (x-t)[q(x)]^2 dx\right), \tag{52}$$

and q uniquely solves the Painlevé II ordinary differential equation,

$$q''(x) = xq(x) + 2[q(x)]^3, \quad q(x) \sim \mathrm{Ai}(x) \text{ as } x \to \infty \tag{53}$$

for all x, where $\mathrm{Ai}(x)$ is the Airy function,

$$Ai''(x) = x \cdot Ai(x), \tag{54}$$

with the boundary condition,

$$Ai(x) \sim \frac{e^{-\frac{2}{3}x^{3/2}}}{2\sqrt{\pi}x^{1/4}} \text{ as } x \to \infty. \tag{55}$$

In (53), $q(x) \sim \mathrm{Ai}(x)$ as $x \to \infty$ means that $\lim_{x\to\infty} \frac{q(x)}{\mathrm{Ai}(x)} = 1$ (Tracy and Widom, 1996). Table 2 gives the percentage points of the Tracy–Widom distributions for $\beta = 1, 2, 4$.

Largest Eigenvalue: Wishart–Laguerre Ensembles

Fill an $(r \times n)$-matrix $\mathcal{X}$ with iid $\mathcal{N}(0,1)$ deviates. Then, $\mathbf{S} = \mathcal{X}\mathcal{X}^\tau$ has the white Wishart distribution $\mathcal{W}_r(n, \mathbf{I}_r)$ and represents the null case (i.e., $\mathbf{\Sigma} = \mathbf{I}_r$). The eigenvalues of $\mathbf{S}$ are real and nonnegative. Denote the largest eigenvalue of $\mathbf{S}$ by $\widehat{\lambda}_1$.

Fixed r, Finite n: Exact Distribution

The exact distribution of $\widehat{\lambda}_1$ in the null case was found by Constantine (1963) and is expressed as an infinite expansion in zonal polynomials (i.e., hypergeometric functions of two matrix arguments). See Muirhead (1982, Chapter 7) for a detailed exposition of zonal polynomials. Unfortunately, such a series representation, which converges very slowly, is impractical for numerical computation and statistical usage.

Large r, Large n: The Tracy–Widom Laws

The development of random-matrix theory has provided us with the following useful results concerning the limiting distribution of the largest eigenvalue when the dimensions n and r of the matrix $\mathcal{X}$ are both very large.

Real case. Let

$$\mu_{nr} = (\sqrt{n-1} + \sqrt{r})^2, \tag{56}$$

and

$$\sigma_{nr} = (\sqrt{n-1} + \sqrt{r}) \left(\frac{1}{\sqrt{n-1}} + \frac{1}{\sqrt{r}} \right)^{1/3} \tag{57}$$

be centering and scaling factors, respectively. Suppose both r and n are large and that $r/n \to \gamma \in (0, \infty)$. Johnstone (2001, 2006) showed that under the null model,

$$\frac{\widehat{\lambda}_1 - \mu_{nr}}{\sigma_{nr}} \overset{\mathcal{D}}{\Rightarrow} W_1 \sim F_1, \tag{58}$$

where F_1 is the *Tracy–Widom law of order 1* and has distribution function,

$$F_1(t) = [F_2(t)]^{1/2} \exp\left(-\frac{1}{2} \int_t^\infty q(x) dx \right). \tag{59}$$

From (58), we write $TW_1(n, r)$ for the Tracy–Widom F_1 distribution of $\mu_{nr} + \sigma_{nr} W_1$, which can be used to approximate the distribution of $\widehat{\lambda}_1$. Johnstone showed that the asymptotic distribution result (58) is still useful for n and r as small as 10.

Remark: The assumption $r/n \to \gamma \in (0, \infty)$ has been extended to include $\gamma = 0$ and $\gamma = \infty$ (El Karoui, 2003). The extension to $\gamma = \infty$ is important in that it allows applications where $p \gg n$.

Remark: If the Gaussian assumption on the elements of $\mathcal{X}$ is replaced by an assumption that $n - r = O(r^{1/3})$ and an assumption that the elements of the matrix $\mathcal{X}$ are symmetrically distributed with finite even moments and sufficiently light tails (i.e., a subGaussian distribution), then the Tracy–Widom Law (58) still holds (Soshnikov, 2002; Péché, 200x).

Tracy and Widom (2000) and Johnstone (2001) show that F_1 has the following properties: (1) the F_1 distribution does not depend upon any parameters and, therefore, its role can be viewed in the same light as the role of the standard Gaussian distribution in the central limit theorem, (2) the F_1 density function is not symmetric, but is unimodal with mean approximately -1.21, standard deviation approximately 1.27, and different decay rates as $x \to -\infty$ or $x \to +\infty$, (3) the standard deviation, σ_{nr}, increases with n as $n^{1/2}$, (4) approximately 83% of the distribution is less than μ_{nr}, (5) approximately 95% of the distribution is less than $\mu_{nr} + \sigma_{nr}$, and (6) approximately 99% of the distribution is less than $\mu_{nr} + 2\sigma_{nr}$.

The limiting distribution F_1 was discovered by Tracy and Widom to be one of a family of distributions, F_β, where $\beta = 1$ (real case), 2 (complex case), and

4 (real-quaternion case), Computation in S-PLUS and MATLAB of the Tracy-Widom distributions is slow and cubic spline approximations are preferred (Bejan, 2005).

Complex case. The asymptotic distribution of the largest eigenvalue of a symmetric $(r \times r)$ complex Wishart matrix was actually found (Johansson, 2000) before the real-case result given by Johnstone. Johansson showed that, for a Hermitian (e.g., complex Wishart) matrix,

$$\frac{\widehat{\lambda}_1 - \mu_{nr}}{\sigma_{nr}} \overset{\mathcal{D}}{\Rightarrow} W_2 \sim F_2, \tag{60}$$

where F_2 is the *Tracy-Widom law of order 2* given by (52) and

$$\mu_{nr} = (\sqrt{n} + \sqrt{r})^2 \tag{61}$$

and

$$\sigma_{nr} = (\sqrt{n} + \sqrt{r}) \left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{r}} \right)^{1/3}. \tag{62}$$

In other words,

$$P\left\{ \frac{\widehat{\lambda}_1 - \mu_{nr}}{\sigma_{nr}} \leq x \right\} \rightarrow F_2(x), \quad \text{as } r \rightarrow \infty, \ x \in \Re. \tag{63}$$

The real case with limiting distribution F_1 was found by Johnstone (2001) by using an independent approach with a different construction than was used in the complex case.

Remark: Ramírez, Rider, and Virág (2008) extended the Tracy–Widom Laws to all $\beta > 0$.

Largest Eigenvalue: Two Wishart Matrices

The problem of approximating the distribution of either the largest eigenvalue λ of $(\mathbf{A} + \mathbf{B})^{-1}\mathbf{A}$ or the largest eigenvalue θ of $\mathbf{B}^{-1}\mathbf{A}$ when $\mathbf{A}$ and $\mathbf{B}$ are both $(r \times r)$-matrices and m, n, and r are large has been studied in detail by Johnstone (2008). Johnstone shows that with appropriate centering and scaling of the logit transform of λ, the Tracy-Widom laws F_1 and F_2 continue to hold, F_1 for the real case and F_2 for the complex case, as in the single-Wishart case above.

Large r, Large m and n: The Tracy–Widom Laws

Real case. Suppose $\mathbf{A} \sim \mathcal{W}_r(m, \mathbf{I}_r)$ and $\mathbf{B} \sim \mathcal{W}_r(n, \mathbf{I}_r)$ are independent white Wishart matrices. Assume that $n \geq r$. Then, $\mathbf{A}$ is positive definite. If λ_{1r} is the largest eigenvalue of $(\mathbf{A} + \mathbf{B})^{-1}\mathbf{A}$, then, $0 < \lambda_{1r} < 1$. The following results

have been proved for r even only, but empirical results indicate that they may also hold for general r.

Let

$$\mu_r = 2 \log \tan \left(\frac{\phi + \gamma}{2} \right) \tag{64}$$

and

$$\sigma_r^3 = \frac{16}{(m+n-1)^2} \frac{1}{\sin^2(\phi+\gamma) \sin\phi \sin\gamma} \tag{65}$$

be centering and scaling factors, respectively, for

$$W_r = \text{logit}(\lambda_{1r}) = \log \left(\frac{\lambda_{1r}}{1-\lambda_{1r}} \right), \tag{66}$$

where the angle parameters, γ and ϕ, are defined by

$$\sin^2 \left(\frac{\gamma}{2} \right) = \frac{\min(r,m) - 1/2}{m+n-1} \tag{67}$$

$$\sin^2 \left(\frac{\phi}{2} \right) = \frac{\max(r,m) - 1/2}{m+n-1}, \tag{68}$$

respectively. If $m = m_r$, $n = n_r \to \infty$ as $r \to \infty$ in such a way that $\lim_{r\to\infty} \min(r,m)/(m+n) > 0$ and $r/n \to \xi < 1$, then,

$$\frac{W_r - \mu_r}{\sigma_r} \overset{\mathcal{D}}{\Rightarrow} Z_1 \sim F_1, \tag{69}$$

where F_1 is the Tracy-Widom law of order 1 given by (59).

A more precise convergence result can be made: As $m_r, n_r \to \infty$ as $r \to \infty$ through the even integers, then there exists a constant $C > 0$ depending upon (ϕ, γ) such that for large x,

$$\left| P\left\{ \frac{W_r - \mu_r}{\sigma_r} \le x \right\} - F_1(x) \right| \le C r^{-2/3} e^{-x/2}. \tag{70}$$

Complex case. Suppose that we have two independent complex-Wishart random matrices, $\mathbf{A} \sim \mathcal{W}_s^C(m, \mathbf{I}_s)$ and $\mathbf{B} \sim \mathcal{W}_s^C(n, \mathbf{I}_s)$. We are interested in the distribution of the largest eigenvalue of $(\mathbf{A}+\mathbf{B})^{-1}\mathbf{A}$.

Let

$$W^C = \text{logit}(\lambda_r^C) = \log \left(\frac{\lambda_r^C}{1-\lambda_r^C} \right). \tag{71}$$

Assume that $m_r, n_r \to \infty$ as $r \to \infty$ in the same way as for the real case. Defining μ_r^C and σ_r^C as appropriate centering and scaling constants, Johnstone (2008) showed that

$$\frac{W_r^C - \mu_r^C}{\sigma_r^C} \overset{\mathcal{D}}{\Rightarrow} Z_2 \sim F_2, \tag{72}$$

where F_2 is the Tracy-Widom distribution of order 2 given by (52). The centering and scaling constants are given by

$$\mu^C = \frac{\tau_N^{-1} u_N + \tau_{N-1}^{-1} u_{N-1}}{\tau_N^{-1} + \tau_{N-1}^{-1}} \tag{73}$$

$$\frac{1}{\sigma^C} = \frac{1}{4}(\tau_N^{-1} + \tau_{N-1}^{-1}), \tag{74}$$

where

$$u_N = 2 \log \tan \left(\frac{\phi_N + \gamma_N}{2} \right) \tag{75}$$

$$\tau_N^3 = \frac{16}{(2N + \alpha + \beta + 1)^2} \frac{1}{\sin^2(\phi_N + \gamma_N) \sin \phi_N \sin \gamma_N}, \tag{76}$$

$$\sin^2 \left(\frac{\gamma_N}{2} \right) = \frac{N + 1/2}{2N + \alpha + \beta + 1} \tag{77}$$

$$\sin^2 \left(\frac{\phi_N}{2} \right) = \frac{N + \beta + 1/2}{2N + \alpha + \beta + 1}, \tag{78}$$

and

$$N = \min(m, r), \quad \alpha = n - r, \quad \beta = |m - r|. \tag{79}$$

Moreover, for large enough x, there exists a constant C depending upon (ϕ, γ) such that

$$\left| P \left\{ \frac{W_r^C - \mu_r^C}{\sigma_r^C} \le x \right\} - F_2(x) \right| \le C r^{-2/3} e^{-x/2}. \tag{80}$$

Applications of the Tracy–Widom Laws

The Tracy-Widom Laws turn out to be the limiting distributions for many different problems. Examples of diverse problems for which the Tracy-Widom laws hold are:

1. The length of the longest increasing subsequence of a random permutation of n objects as $n \to \infty$.

Consider a permutation π of the first n integers $\{1, 2, \ldots, n\}$. We can write π as $\{\pi_1, \pi_2, \ldots, \pi_n\}$. Then, π has an increasing subsequence $\ell_n(\pi)$ of length k if there exist indices $1 \le i_i < i_2 < \cdots < i_k \le n$ such that $\pi_{i_1} < \pi_{i_2} < \cdots < \pi_{i_k}$. (By switching the directions of the inequality signs, a similar definition can be given for a decreasing subsequence of length k.)

Assume the $n!$ permutations each of length n are equally likely. For each such permutation, find the length $L_n = \ell_n(\pi)$ of the longest increasing subsequence. For example, let $n = 8$ and consider the permutation $\pi = \{2, 5, 1, 3, 4, 8, 6, 7\}$; the length of the longest increasing subsequence is $L_8 = 5$, given by $\{2, 3, 4, 6, 7\}$

and $\{1, 3, 4, 6, 7\}$. Thus, L_n is a random variable, but the actual subsequence may not be unique. The main questions are: what is the distribution of L_n, and what is its mean, $\mathrm{E}\{L_n\}$, and its variance, $\mathrm{var}\{L_n\}$, for large n?

It has been shown that $\lim_{n\to\infty} \frac{1}{\sqrt{n}}\mathrm{E}\{L_n\}$ exists (Ulam, 1961; Hammersley, 1972) and has the value 2 (Vershik and Kerov, 1977). Thus, for large n, $\mathrm{E}(L_n) \sim 2\sqrt{n}$. Recent work has also showed that

$$\lim_{n\to\infty} \frac{\mathrm{E}\{L_n\} - 2\sqrt{n}}{n^{1/6}} \;\sim\; -1.711 \tag{81}$$

and

$$\lim_{N\to\infty} \frac{\mathrm{var}\{L_n\}}{n^{1/3}} \;\sim\; 0.902, \tag{82}$$

where the limiting constants on the rhs of (81) and (82) were obtained by Baik, Deift, and Johansson (1999). Furthermore, the large-n distribution of L_n has been shown (Baik, Deift, and Johansson, 1999) to be

$$P\left\{\frac{L_n - 2\sqrt{n}}{n^{1/6}} \le x\right\} \;\to\; F_2(x), \quad \text{as } n \to \infty, \quad x \in \Re, \tag{83}$$

where F_2 is the Tracy–Widom law. Thus, for large n, the distribution of the length of the longest increasing subsequence of a permutation π of $\{1, 2, \ldots, n\}$, appropriately centered and scaled, is identical to the distribution of the largest eigenvalue of a random GUE matrix.

2. Patience-Sorting card game.

Patience sorting is a one-person card game in Britain (called *solitaire* in the United States). As explained by Aldous and Diaconis (1999), the game starts out with a "deck" of cards labelled $1, 2, 3, \ldots, n$, and then the deck is shuffled and a card is drawn from the top of the deck and placed into a "pile" according to the following rule. A newly turned-up card can be placed on a card at the top of an existing pile only if its value is lower than that of the top card; otherwise, the new card starts another pile to the right of all existing piles. For example, a 3 can be placed on top of a 5, but a 6 starts a new pile. The game continues until all cards are dealt and placed into piles. The objective is to complete the game with as few piles as possible.

Consider the eight cards in the same order as above: 2, 5, 1, 3, 4, 8, 6, 7. Start the first pile with the 2; the 5 is bigger than the 2, so the 5 starts a new pile; the 1 goes on top of the 2, and the 3 goes on top of the 5; the 4 starts a new pile; the 8 starts a new pile; the 6 goes on top of the 8; and the 7 starts a new pile. The piles are as follows:

1	**3**		**6**	
2	5	**4**	8	**7**

This "greedy" strategy, which places each card on top of the most-leftwise pile possible, is the optimal strategy. In this example, we obtained five piles, which

is the same result we obtained above for the longest increasing subsequence of the given permutation. In general, the number of piles using the optimal greedy strategy will always equal the length of the longest increasing subsequence.

Note that the top cards in the resulting piles (shown above in boldface) will not necessarily be in permutation order and, hence, will not be an increasing subsequence. For example, consider the sequence 8, 6, 1, 3, 4, 7, 5, 2. The piles are:

$$\begin{array}{cccc} \mathbf{1} & & & \\ 6 & \mathbf{2} & & \mathbf{5} \\ 8 & 3 & \mathbf{4} & 7 \end{array}$$

Clearly, the boldface numbers 1, 2, 4, 5 do not form an increasing subsequence of the original sequence. However, the number of piles (4) is equal to the length (4) of the subsequences 1, 3, 4, 5 or 1, 3, 4, 7.

This correspondence between patience sorting and the length of the longest increasing subsequence yields a very efficient algorithm for computing $L_n = \ell_n(\pi)$.

3. The spacings of consecutive zeroes of the Riemann zeta function.

Riemann's zeta function is defined by adding up inverse powers of the positive integers:

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s} = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \cdots \tag{84}$$

If we set $s = 2$, for example, we have that $\zeta(2) = \pi^2/6$, a result discovered by Leonhard Euler. Euler showed that the zeta function could also be reexpressed as a product over the prime numbers:

$$\zeta(s) = \prod_{p} (1 - p^{-s})^{-1} = \frac{1}{(1 - \frac{1}{2^s})(1 - \frac{1}{3^s})(1 - \frac{1}{5^s})(1 - \frac{1}{7^s}) \cdots} \tag{85}$$

This is known as the *Euler product formula*. Thus, the zeta function has properties that are intimately connected to the distribution of prime numbers. However, more is true. Riemann showed that the zeta function could be written as a product over its zeroes in the complex plane:

$$\zeta(s) = f(s) \left(1 - \frac{s}{\rho_1}\right) \left(1 - \frac{s}{\rho_2}\right) \left(1 - \frac{s}{\rho_3}\right) \cdots, \tag{86}$$

where $\rho_1, \rho_2, \ldots$ are the complex numbers for which $\zeta(s) = 0$, and $f(s)$ is a simple "fudge factor."

The zeta function has "trivial" zeroes at the negative even integers (i.e., at $s = -2, -4, -6, \ldots$). Riemann showed that all the nontrivial zeroes can be found in an infinite "critical" strip lying above and below the unit interval $0 < Re(s) < 1$ in the complex plane. Moreover, the zeroes are symmetrically located: each zero lying above the unit interval has a mirror-image zero lying below the unit interval. The Riemann Hypothesis says that all the zeroes of the

zeta function lie on the vertical line through $Re(s) = \frac{1}{2}$, in which case a zeta zero can be written as $\rho_n = \frac{1}{2} \pm i\gamma_n$, for an ordered sequence of real, positive constants $\{\gamma_n\}$, where $\gamma_n \sim 2\pi n/(\log\ n)$ as $n \to \infty$.

Attempts have been made to check the Riemann Hypothesis using numerical methods. Andrew Odlyzko developed extremely efficient algorithms for computing zillions of zeroes in the critical strip, and showed that they all satisfy the Riemann Hypothesis (Odlyzko, 1992). Odlyzko further showed that the spacings between consecutive zeroes of the zeta function behave, statistically, like the spacings between consecutive eigenvalues of large, random matrices from the GUE (Odlyzko, 1987). In other words, the zeroes of the zeta function can be viewed as having a "spectral" interpretation, which agrees with the belief (by physicists) that the zeroes of the zeta function can be interpreted as energy levels in some quantum chaos system. In fact, Hilbert and Polya had conjectured that the Riemann Hypothesis is true precisely because the zeroes of the zeta function correspond to eigenvalues of a positive linear (Hermitian) operator. For special cases of the zeta function that have been proved, the statistical properties of the eigenvalue spacings and the spacings of the zeroes of the zeta function turn out to be identical.

Other applications for which the Tracy–Widom laws hold include the following: the positions at any given intermediate time of N independent 1–D Brownian motion paths with time in $[0, 1]$ that start and end at the same point but do not intersect at any intermediate time; the last-passage time of a certain last-passage percolation model; and the height fluctuations of a certain random growth model. Some of these problems and other fascinating examples (such as the distances between parked cars in London, waiting times for buses in Cuernavaca, Mexico, and an airline boarding problem) are explained in detail by Deift (2006).

Each of these quantities behaves statistically like either the eigenvalues or the spacings between consecutive eigenvalues of a random $(n \times n)$-matrix as $n \to \infty$. In fact, the Tracy-Widom laws have been found to be of such great importance in random-matrix theory that they have been said to play a similar role as that of the Gaussian distribution in classical statistical theory (see, e.g., Diaconis, 2003).

Software

MATLAB 7.0 software for computing Tracy–Widom distributions and simulation of random matrices can be found at Momar Dieng's website,
`math.arizona.edu/~momar/research.htm`
There is also N. Raj Rao's `RMTool`, a publicly available MATLAB symbolic toolbox, which is used to compute the limiting spectral density of a large class of random matrices and can be downloaded from the website
`www.mit.edu/~raj/rmtool`

S-Plus software for computing Marčenko–Pastur "semi-circle"-type laws, cumulative distribution functions, quantiles, Wishart-matrix simulations, eigenvalues of a white Wishart matrix, Tracy–Widom distributions, quantiles, simulations, and cubic-spline interpolations, can be found at Andrei Bejan's website, `www.vitrum.md/andrew/MScWrwck/codes.txt`

Bibliographical Notes

The classic book in this area is Mehta (2004), now in its third edition. Although it deals at great length with Gaussian ensembles, it does not mention Laguerre ensembles and Wishart matrices and their important roles in mathematical statistics. The book by Porter (1965) is a collection of all the important papers published on random matrix theory prior to 1965. Another recent monograph is Guionnet (2008).

An excellent historical account of random-matrix theory can be found in Forrester, Snaith, and Verbaarschot (2003), which is acually a Preface to an special issue of the *Journal of Physics* on random-matrix theory. For a more technical review of random-matrix theory, see Bai (1999). See also Edelman and Rao (2005) for an excellent survey of the field.

References

Anderson, T.W. (1963). Asymptotic theory for principal component analysis, *Annals of Mathematical Statistics*, **34**, 122–148.

Bai, Z.D. (1999). Methodologies in spectral analysis of large dimensional random matrices, a review, *Statistica Sinica*, **9**, 611–677.

Baik, J., Buckingham, R., and DiFranco, J. (2008). Asymptotics of Tracy–Widom distributions and the total integral of a Painlevé II function, `arXiv:0704.3636v1`.

Baik, J., Deift, P., and Johansson, K. (1999). On the distribution of the length of the longest increasing subsequence of random permutations, *Journal of the American Mathematical Society*, **12**, 1119–1178.

Bejan, A.I. (200x). Largest eigenvalues and sample covariance matrices. Tracy–Widom and Painleve II: computational aspects and realization in S-Plus with applications.

Bellman, R. (1960). *Introduction to Matrix Analysis*, New York: McGraw-Hill Book Co.

Bohigas, O., Giannoni, M.J., and Schmit, C. (1984). Characterizations of chaotic quantum spectra and universality of level fluctuation laws, *Physics Reviews Letters*, **52**, 1.

Constantine, A.G. (1963). Some noncentral distribution problems in multivariate analysis, *Annals of Mathematical Statistics*, **34**, 1270–1285.

Deift, P. (2006). Universality for mathematical and physical systems, `arXiv:math-ph/0603038v2 19 May 2006`.

Diaconis, P. (2003). Patterns in eigenvalues: The 70th Josiah Willard Gibbs Lecture, *Bulletin of the American Mathematical Society*, **40**, 155–178.

Edelman, A. and Rao, N.R. (2005). Random matrix theory, *Acta Numerica*, 1–65.

El Karoui, N. (2003). On the largest eigenvalue of Wishart matrices with identity covariance when n, p and p/n tend to infinity. `arXiv:math.ST/0309355`.

El Karoui, N. (200x). Tracy–Widom limit for the largest eigenvalue of a large class of complex Wishart matrices, `arXiv:math.PR/0503109`.

Forrester, P.J., Snaith, N.C., and Verbaarschot, J.J.M. (2003). Developments in random matrix theory, *Journal of Physics A. Special Edition on Random Matrix Theory*, `arXiv:cond-mat/0303207 v1 11 Mar 2003`.

Guionnet, A. (2008). *On Random Matrices*. Lecture Notes in Mathematics, New York: Springer.

Hammersley, J. (1972). A few seedlings of research, *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, **1** (L.M. LeCam, J. Neyman, and E.L. Scott, eds.), Berkeley, CA: University of California Press, pp. 345–394.

James, A.T. (1960). The distribution of the latent roots of the covariance matrix, *The Annals of Mathematical Statistics*, **31**, 151–158.

James, A.T. (1964). Distributions of matrix variates and latent roots derived from normal samples, *The Annals of Mathematical Statistics*, **35**, 475–501.

Johnstone, I.M. (2001). On the distribution of the largest eigenvalue in principal components analysis, *The Annals of Statistics*, **29**, 295–327.

Johnstone, I.M. (2006). High dimensional statistical inference and random matrices. `arXiv:math.ST/0611589`.

Kiecherbauer, T., Marklof, J., and Soshnikov, A. (2001). Random matrices and quantum chaos, *Proceedings of the National Academy of Sciences*, **98**, 10531-10532.

Marcenko, V.A, and Pastur, L.A. (1967). Distributions of eigenvalues of some sets of random matrices, *Math. USSR-Sb*, **1**, 507–536.

Mehta, M.L. (2004). *Random Matrices, 3rd Edition*, Pure and Applied Mathematics (Amsterdam), 142, Amsterdam, Netherlands: Elsevier/Academic Press.

Muirhead, R.J. (1978). Latent roots and matrix variates: a review of some asymptotic results, *Annals of Statistics*, **6**, 5–33.

Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*, New York: Wiley.

Odlyzko, A.M. (1987). On the distribution of spacings between zeros of the zeta function, *Mathematical Computation*, **48**, 273–308.

Péché, S. (200x). Universality results for largest eigenvalues of some sample covariance matrices, `arXiv:0705.1701`.

Porter, C.E. (1965). *Statistical Theory of Spectra: Fluctuations*, New York: Academic Press.

Porter, C.E. and Rosenzweig, N. (1960). Statistical properties of atomic and nuclear spectra, *Annals of the Acad. Sci. Fennicae, Serie A VI Physica*, **44**, 1.

Ramírez, J., Ride, B., and Virág, B. (2008). Beta ensembles, stochastic Airy spectrum, and a diffusion, `arXiv:math/0607331v3`.

Soshnikov, A. (2002). A note on universality of the distribution of the largest eigenvalue in certain sample covariance matrices, *Journal of Statistical Physics*, **108**, 1033–1056.

Tracy, C.A. and Widom, H. (1996). On orthogonal and symplectic matrix ensembles, *Communications in Mathematical Physics*, **177**, 727–754.

Tulino, A.M. and Verdú, S. (2004). Random matrix theory and wireless communications, *Foundations and Trends in Communications and Information Theory*, **1**, 1–182.

Wachter, K.W. (1978). The strong limits of random matrix spectra for sample matrices of independent elements, *The Annals of Probability*, **6**, 1–18.

Wigner, E.P. (1955). Characteristic vectors of bordered matrices with infinite dimensions, *Annals of Mathematics*, **62**, 548–564.

Last revised: 24 December 2008.