# Eigenvalues and Condition Numbers of Random Matrices

by


Alan Edelman

M.S. Mathematics, Yale University (1984)
B.S. Mathematics, Yale University (1984)


Submitted to the Department of Mathematics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1989



Signature of Author____________________________

Department of Mathematics
May 1989

Certified by____________________________

Lloyd N. Trefethen
Associate Professor of Applied Mathematics
Thesis Supervisor

Accepted by____________________________

Daniel J. Kleitman
Chairman, Applied Mathematics Committee

Accepted by____________________________

Sigurdur Helgason
Chairman, Departmental Graduate Committee

**Eigenvalues and Condition Numbers of Random Matrices**

by

Alan Edelman

Submitted to the Department of Mathematics
on May 1989, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

## Abstract

This dissertation investigates the eigenvalues and condition numbers of real and complex random matrices with the goal of solving several open problems in numerical analysis that were raised in various forms by the work of von Neumann, Birkhoff, Smale, Demmel, and others. Roughly, we answer the question: What is the condition number of a random matrix? We also extend our techniques to solve an open problem in multivariate statistics communicated to us by Olkin.

We follow the literature and consider random matrices with elements from standard normal distributions, but we study more than real square matrices; we describe various distributions for rectangular or square, real or complex matrices. The condition number distributions that we derive for large matrices from a normal distribution model more general situations quite well.

The distribution of the condition number of a random matrix describes how many digits of numerical precision are lost due to ill conditioning when solving a random system of linear equations. It also describes how long an iterative method such as the conjugate gradient iteration would take to converge for a random system. In particular we show that a condition number bound used by von Neumann is correct for only 80% of real random matrices. We solve a problem posed by Smale by showing that the geometric mean of condition numbers of real square matrices grows like $4.65n$. (The arithmetic mean is infinite.) This leads to the conclusion that roughly $\log_{10} n + 0.7$ digits may be lost due to ill conditioning on average when solving large linear systems. The situation for random complex matrices is that $\log_{10} n + 0.4$ digits may be lost on average. Another application is that the conjugate gradient iteration is an $O(n^3)$ process on average. We also obtain an exact distribution that was estimated by Demmel.

When a sample is taken from a multivariate distribution, it is of interest to perform tests on the sample covariance matrix. We derive a recursion for the distribution and moments of the smallest eigenvalue of the covariance matrix when the elements are assumed to be independent standard normals, and we tabulate the expected values. In the language of multivariate statistics, we derive a recursion for the distributions and moments of the smallest eigenvalues of Wishart matrices.

Thesis Supervisor: Lloyd N. Trefethen
Title: Associate Professor of Applied Mathematics

**Acknowledgments**

Most special thanks go to my advisor, Nick Trefethen, for whom a proper thank you would be larger than this thesis. Many good advisors care about their students; Nick takes a personal interest that represents so much more. I am grateful to him for introducing me to numerical linear algebra, for inspiring me all along the way, for generously making available to me the best numerical analysis combination library and computing center in the Boston area, his office, and so much more. Nick made many useful suggestions that went into this thesis for which I will always be grateful, but it is his friendship that I have valued most.

Financial support has been provided by a Hertz Foundation Fellowship and an NSF Presidential Young Investigator award to my advisor. Technologically, this thesis would have been impossible without the software packages MATLAB and Mathematica, which made experiments, computations, and plotting so easy.

I have been most fortunate to have been able to discuss this work with Garrett Birkhoff, who suggested I look into the history of the subject and encouraged me to continue working in the field. I also must thank Bill Morton for his kind hospitality at the Oxford University Computing Laboratory where the early part of this research took place.

I would like to express additional thanks to Charlie Micchelli and IBM Yorktown Heights for introducing me to mathematical research and for encouraging me to pursue my graduate work, and also to Shizuo Kakutani, who was a great source of inspiration during my undergraduate career. I will never forget his saying that "mathematics is like climbing a mountain."

My MIT experience was made enjoyable by Alar Toomre, who one morning shipped me all one hundred roots of the one hundredth Laguerre polynomial as if I had requested them by mail order, and Gil Strang, who taught me about circulant matrices.

Last, but far from least, I wish to give thanks to my family, who did not force me to go to medical school but obtained another doctor in the family anyway. I would like to thank my mother, grandmother, and sisters for all their support, and my father, who started me early in a mathematical career by teaching me to connect the numbered dots, and who I wish could have seen where this would lead.

# Contents


**1 Foreword** **8**

**2 History and applications** **10**

2.1 Numerical analysis . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 10

2.2 Multivariate statistics . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 16

2.3 Nuclear physics . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 18

2.4 The Riemann zeta function . . . . . . . . . . . . . . . . . . . . . . . . . . . . 19

2.5 Graph theory . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 19

2.6 Contributions of this thesis . . . . . . . . . . . . . . . . . . . . . . . . . . . . 21

**3 Fundamental concepts** **30**

3.1 Random matrix notation and terminology . . . . . . . . . . . . . . . . . . . 30

3.2 Jacobians for matrix factorizations and Haar measure . . . . . . . . . . . . . 31

3.3 Joint element densities . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 34

3.4 Joint eigenvalue densities . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 36

3.5 Bidiagonalization of a Gaussian matrix . . . . . . . . . . . . . . . . . . . . . 38

**4 Smallest eigenvalue distribution for $W(m,n)$** **40**

4.1 A multivariate integral for the pdf of $\lambda_{\min}$ . . . . . . . . . . . . . . . . . . . . 40

4.2 The pdf of $\lambda_{\min}$ for $W(m,m)$ . . . . . . . . . . . . . . . . . . . . . . . . . . 41

4.3 The pdf of $\lambda_{\min}$ for $W(m,m+1)$ . . . . . . . . . . . . . . . . . . . . . . . . 43

4.4 The pdf of $\lambda_{\min}$ for $W(m,n)$ (recursion) . . . . . . . . . . . . . . . . . . . . 45

4.5 Proof of Theorem 4.3 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 47

4.6 The pdf of $\lambda_{\min}$ for $W(m, m+2)$ and $W(m, m+3)$ . . . . . 51
4.7 Summary . . . . . 52

**5 Other smallest eigenvalue results** **54**
5.1 Asymptotic smallest eigenvalue for $W(m, n)$ . . . . . 54
5.2 Asymptotic geometric means of $\lambda_{\min}$ for $W(m, n)$ . . . . . 59
5.3 Smallest eigenvalue results for $\tilde{W}(m, n)$ . . . . . 62

**6 Largest eigenvalue results** **64**
6.1 Asymptotic largest eigenvalue for $W(m, n)$ . . . . . 64
6.2 Asymptotic geometric means of $\lambda_{\max}$ for $W(m, n)$ . . . . . 66
6.3 Largest eigenvalue results for $\tilde{W}(m, n)$ . . . . . 68

**7 Condition numbers** **69**
7.1 Limiting condition number distributions, as $n \to \infty$ . . . . . 69
7.2 Exact expressions for $m = 2$ . . . . . 72
7.3 Demmel's condition number . . . . . 74
7.4 The tails of the condition number distributions . . . . . 77

**8 Moments of $\lambda_{\min}$ for $W(m, n)$** **80**
8.1 Computation of the moments . . . . . 80
8.2 Expected values when $n - m = 0, 1,$ and 3 . . . . . 82

**9 Random matrices and orthogonal polynomials** **84**
9.1 The random matrix associated with a weight function . . . . . 84
9.2 Characteristic polynomial . . . . . 86
9.3 The mode . . . . . 88

**A Mathematica programs** **90**

**B Table of expected values** **93**

**C Other uses of the programs** **98**

# List of Figures


2.1 The limiting density of $\kappa/n$ ($n \times n$ real matrices) . . . . . . . . . . . . . . . . . . . 25

2.2 The limiting density of $\kappa/n$ ($n \times n$ complex matrices) . . . . . . . . . . . . . . . . 26

2.3 The limiting density of $\log(\kappa/n)$ ($n \times n$ real matrices) . . . . . . . . . . . . . . . . 27

2.4 The limiting density of $\log(\kappa/n)$ ($n \times n$ complex matrices) . . . . . . . . . . . . . . 28

2.5 Experimental $\kappa$ for 50,000 matrices, $n = 75$ . . . . . . . . . . . . . . . . . . . . . . . 29

4.1 The density of $\lambda_{\min}$ for $W(3,n)$, $n = 3,4,5,\ldots,28$ . . . . . . . . . . . . . . . . . . 53

5.1 The density of $m\lambda_{\min}$ for $W(m,m)$, $m = 1,2,\infty$ . . . . . . . . . . . . . . . . . . . . . 56

5.2 The density of $\sqrt{m\lambda_{\min}}$ for $W(m,m)$, $m = 1,2,\infty$ . . . . . . . . . . . . . . . . . . 57

5.3 The density of $\lambda_{\min}/2m$ for $W(m,2m)$, $m = 3,9,15,21,27$ . . . . . . . . . . . . . . 58

6.1 The empirical cdf of $\lambda_{\max}/m$ for $W(m,m)$, $m = 100$ . . . . . . . . . . . . . . . . . 65

## Notation

| | |
|---|---|
| $m$ | number of rows in a matrix |
| $n$ | number of columns in a matrix ($m \leq n$ always) |
| $A^T$ | transpose of $A$ |
| $A^H$ | Hermitian transpose of $A$ |
| $\lambda_{\min}$ | smallest eigenvalue of a matrix |
| $\lambda_{\max}$ | largest eigenvalue of a matrix |
| $\kappa$ | condition number of a matrix |
| $i$ | $\sqrt{-1}$ (or an index) |
| iid | independent and identically distributed |
| cdf | cumulative density function (distribution function) |
| pdf | probability density function |
| $N(\mu, \sigma^2)$ | normal distribution with mean $\mu$ and variance $\sigma^2$ |
| $\tilde{N}(\mu, \sigma^2)$ | $x + yi,$ where $x$ and $y$ are iid $N(\mu, \sigma^2)$ |
| $G(m, n)$ | $m \times n$ matrix, where the $mn$ elements are iid $N(0, 1)$ |
| $\tilde{G}(m, n)$ | $m \times n$ matrix, where the $mn$ elements are iid $\tilde{N}(0, 1)$ |
| $W(m, n)$ | $m \times m$ matrix $AA^T,$ where $A$ is $G(m, n)$ |
| $\tilde{W}(m, n)$ | $m \times m$ matrix $AA^H,$ where $A$ is $\tilde{G}(m, n)$ |
| GOE | Gaussian orthogonal ensemble: $(A + A^T)/2,$ where $A$ is $G(m, n)$ |
| GUE | Gaussian unitary ensemble: $(A + A^H)/\sqrt{8},$ where $A$ is $\tilde{G}(m, n)$ |
| $\mathrm{tr}(A)$ | $\mathrm{trace}(A)$ |
| $\mathrm{etr}(A)$ | $\exp(\mathrm{tr}(A))$ |
| $\Gamma_m(a)$ | multivariate gamma function (p. 32) |
| $\tilde{\Gamma}_m(a)$ | complex multivariate gamma function (p. 33) |
| $\Delta$ | $\prod_{1 \leq i < j \leq m-1} \lvert x_i - x_j \rvert$ (p. 41) |
| $d\mu(x)$ | $xe^{-x/2}dx$ (p. 41) |
| $d\Omega$ | $d\mu(x_1) \ldots d\mu(x_{n-1})$ (p. 41) |
| $I^{\alpha}_{i,j}$ | $\int_{R_+^{m-1}} f(\lambda) \Delta d\Omega$ (p. 47) |

*He talks at random; sure the man is mad.*
— SHAKESPEARE, *I Henry VI V.iii.*

# Chapter 1

# Foreword

For three centuries, mathematicians searched for an exact formula for the roots of a general polynomial. Their efforts would necessarily be in vain, as was shown in the early nineteenth century when Ruffini (1813) and Abel (1827) independently demonstrated the unsolvability of quintics in terms of radicals. Recast in the language of linear algebra, we immediately conclude the lack of a simple expression for the eigenvalues of a general matrix, as they are the roots of an arbitrary polynomial.

What if we redefine the problem, replacing “general” with “random”? We are no longer interested in the eigenvalues of any one matrix; the distributions and expectations are what matter now. Can we escape the clutches of the theorem of Abel and Ruffini and obtain closed form solutions? Surprisingly, many times we can.

There are many applications of random eigenvalues. Our point of departure, however, was the desire to understand the ratio of the largest to smallest eigenvalue of $AA^T$ for a matrix $A$. This is the square of the *condition number* of $A$ which is often used to measure how many digits we might lose upon numerically solving linear systems of equations, or perhaps more importantly, how quickly an iterative method, such as the conjugate gradient iteration, will converge.

How does the condition number behave on average? This question has been asked in various forms by von Neumann et al. [49], Birkhoff and Gulati [9], Smale [42], Demmel [15], and many others. To capture a notion of "average," these authors chose a probability distribution for their random matrices. Answers usually have taken the form of an approximate bound on the condition number distribution or expectation. In this thesis, we settle many of the issues by obtaining exact formulas.

The next chapter outlines the history and applications of random eigenvalues in a variety of fields. At the conclusion of this survey (Section 2.6), we summarize the contributions of this thesis.

# Chapter 2

# History and applications

## 2.1 Numerical analysis

The limitations on solving large systems of equations are computer memory and speed. The speed of computation, however, is not only measured by clocking hardware; it also depends on numerical stability, and for iterative methods, on convergence rates. If we had a machine that performed Gaussian elimination on a random $n$ by $n$ matrix for $n = 1{,}000{,}000$ in 16-digit precision with a computation time of one year, after that year, would we have any correct digits at all? Would doubling the precision and a wait of two years suffice? What about iterative methods? How long do we need to wait to obtain convergence?

Such questions are not new. In 1946, von Neumann and his associates saw $n = 100$ on the horizon, and hoped to understand the numerical stability question mathematically. Today, perhaps the equivalent number is $n = 50{,}000$, but our understanding, though far better, is still imperfect.

We review here the work of von Neumann and his colleagues Bargmann, Goldstine, and Montgomery [49], for three reasons: to set the stage for random condition numbers, to illustrate historical lessons, and to fill a gap suggested by Wilkinson[57]:

> ...although their paper deservedly received widespread recognition very few have been able to give a concise sketch of the basis of their analysis.

Given the difficulty of extracting the various ideas from their work,[1] we will recast their ideas in modern notation, supplying equation and page references from [49] for the reader wishing to go back to the original source.

Goldstine and von Neumann's analysis [49c, 49d] has three components: the choice of algorithm, the rounding errors, and the implications. For their algorithm, they made the unfortunate choice of solving $Ax = b$ by forming $A^TA$ and solving the positive definite system $A^TAx = A^Tb$ (the "normal equations"). This would be achieved by inverting $A^TA$ and then forming $M = (A^TA)^{-1}A^T$ numerically. We know now that this algorithm unnecessarily squares the condition number $\kappa$, and thus is unstable. In fact, Goldstine and von Neumann recognized the importance of the condition number[2] (though they did not use this term), yet failed to observe that squaring it was unnecessary.[3]

Through a meticulous account of the numerical errors in Gaussian elimination, Goldstine and von Neumann bounded the matrix residual computed in fixed point arithmetic by[4]

$$\|AM - I\|_2 \leq 20\kappa^2 n^2 \beta^{-s}, \tag{2.1}$$

where $\beta$ is the base and $s$ is the number of digits. The factor $n^2$ arose from a worst case analysis of the accumulation of errors. In their later paper, they reduced the factor $20n^2$ to $200n$ by assuming that the errors accumulate statistically ([49d], Equation (9.19)). Thus, they obtained the estimate

$$\|AM - I\|_2 \lesssim 200\kappa^2 n \beta^{-s}. \tag{2.2}$$

In our summary so far, we have deliberately not introduced any random matrices; the only randomness in (2.2) is not in the matrix $A$, but in the assumption of statistical distributions of errors.

[1] Wilkinson referred to the paper's "indigestibility."

[2] "... the actual estimate of loss of precision ... depends not on $n$ only, but also on the ratio $l$ of the upper and lower absolute bounds of the matrix. ... It appears to be the 'figure of merit' expressing the difficulties caused by ... solving simultaneous equation systems ..." ([49a], p. 14).

[3] To their credit, Goldstine and von Neumann did remark, however, that their reasons for forming $A^TA$ "may not be absolutely and permanently valid" ([49c], p. 514). They were merely attempting to work with a model with which they could carry out the analysis. With this model, they avoided the issue of pivoting.

[4] This is a rough average of Equations $(7.5')$ and $(7.5'_I)$ ([49c], p. 551). See the comment after Equation (7.16) for the justification for taking an average.

How is $\kappa$ estimated in (2.2)? This is where von Neumann and his colleagues introduced the assumption of random matrices distributed with elements from independent normals. Various "rules of thumb"[5] are given for $\kappa$ when the matrices are so distributed: $n$ ([49a], p.14), $\sqrt{10}n$ ([49b], p.477), and $10n$ ([49c], p.555). We will show in Section 7.1 that the probabilities that $\kappa$ is less than $n$, $\sqrt{10}n$, and $10n$ converge to (approximately) 0.02, 0.44, and 0.80 respectively. Thus the estimates have the right order of magnitude, but even the final estimate of $10n$ will fail for 20% of random matrices.

To "reduce the probabilistic uncertainties to reasonably safe levels," in their published papers [49c, 49d], Goldstine and von Neumann made the assumptions that

$$\begin{aligned} \lambda_{\max} &\le 100n, \\ \lambda_{\min} &\ge 1/100n, \\ \kappa &\le 10n, \end{aligned} \tag{2.3}$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are the extreme eigenvalues of $A^TA$. ([49c], Equations ($7.14'_I$.a, b, and c), written in terms of the eigenvalues rather than the singular values.) The result for $\lambda_{\max}$ is rigorously proved to hold with probability greater than 0.99 ([49d], Equation (8.8)), but for the estimate of $\lambda_{\min}$ (which we will show fails for 20% of random matrices), they refer to a work of V. Bargmann ("Statistical distribution of proper values") which can not be located.[6]

Substituting (2.3) into (2.2), and taking ten digit precision, we obtain that the right hand side of (2.2) is less than unity if $n < 79$. Actually, Goldstine and von Neumann made a further approximation ([49d], Equation ($9.18''$)) and obtained as one of their "bottom lines" the value $n < 86$ ([49d], Equation (9.19.b)).

The principal direct methods today for solving linear equations are Gaussian elimination with partial pivoting and the QR factorization. Thanks to the work of Wilkinson, we no longer use (2.2) to estimate the errors, but rather we bound the relative error of the computed solution $\bar{x}$ by

$$\frac{\|x-\bar{x}\|_2}{\|x\|_2} \le \kappa f(n)\epsilon. \tag{2.4}$$

[5]Sometimes these estimates are referred to as an expectation and sometimes as a bound that holds with high, though unspecified, probability.

[6]In telephone communications during April of 1989, Bargmann and Goldstine informed me that there probably is no such document.

Here $\epsilon$ is floating-point machine epsilon, $\kappa$ is the condition number, and $f(n)$ includes the possibly combined effects of a polynomial and, for Gaussian elimination, the *growth factor.* Both factors are often dismissed as pessimistic and ignorable in practice.

In 1979, Birkhoff and Gulati [9] resurrected Goldstine and von Neumann's model, but incorporated Wilkinson's bound. They observed the orthogonal invariance of random matrices with independent normally distributed elements and performed several numerical experiments with random number generators and EISPACK. They further suggested that random matrices with normal elements are not ill-conditioned enough for testing Gaussian elimination and proposed a method that would be.

Once again drawing from Goldstine and von Neumann's model, Smale, in a 1984 seminar, asked for

$$L_n = E(\log \kappa),$$

the expected logarithm of the condition number of $n$ by $n$ matrices with normally distributed elements. This question is posed in [42] alongside other random algorithm questions (such as arise in linear programming) where the average tends to be far less costly than the worst case.

To understand the significance of $\log \kappa$, remember that the condition number can also be defined as the smallest number $\kappa$ (depending on $A$ only) such that for all $x$ and $\delta x$, if $Ax = b$ and $A(x + \delta x) = b + \delta b$, then

$$\frac{\|\delta x\|_2}{\|x\|_2} \leq \kappa \frac{\|\delta b\|_2}{\|b\|_2}. \tag{2.5}$$

As a consequence,

$$(\log \|\delta x\|_2 - \log \|x\|_2) - (\log \|\delta b\|_2 - \log \|b\|_2) \leq \log \kappa.$$

When taking the logarithm to the base 10, the equation above has the interpretation that the number of correct digits in $x$ can differ from the number of correct digits in $b$ by at most $\log \kappa$. Smale thus calls $\log \kappa$ the *loss of precision.* For another interpretation of $\log \kappa$, consider (2.4) with $f(n) \approx 1$. The logarithm of the condition number then bounds the exponent in the relative error, giving a kind of "Richter scale" where a relative error of size $\epsilon$ gives a loss of precision of 0, and every tenfold increase in the relative error increments this measure by 1.

In response to Smale's question, bounds for $L_n$ were obtained by Kostlan [29] and Ocneanu(see [42]), who found that the exponential of the expected logarithm to the base $e$ (i.e., the geometric mean) satisfies the following inequalities:

$$n^{2/3+o(1)} \leq \exp(L_n) \leq e n^{5/2}. \tag{2.6}$$

In Section 7.1, we will replace their bounds with $\exp(L_n) = 4.65n + o(n)$.

The gap between $2/3$ and $5/2$ in the exponents above represents a world of difference in terms of comparing the average behavior of the conjugate gradient iteration with that of Gaussian elimination. To see this, assume the condition number of a random matrix $A$ is order $n^\alpha$, and we apply conjugate gradient towards the linear system $A^T A x = A^T b$. The number of operations per iteration is given by the two matrix vector multiplications, hence it is order $n^2$, while the number of iterations is order of the condition number of $A$ which is order $n^\alpha$. Thus the total number of operations is order $n^{2+\alpha}$. If $\alpha$ were $2/3$, then we would know that conjugate gradient is better than Gaussian elimination for random matrices; the reverse would hold if $\alpha$ were $5/2$. Our conclusion that $\alpha = 1$ shows that both are $n^3$ operations, and the choice of method is more subtle; it depends on the various constants involved.

Though most often one hears about the condition number in the context of the sensitivity of solving linear systems, in fact every problem (as explained further in numerical analysis texts such as [25]) has a condition number that expresses the magnification of the relative output error given the relative input error. Problems are *ill-posed* when the condition number is infinite; they are *ill-conditioned* when the condition number is large. Demmel [15] explains that the ill-posed problems in a given class often form an *algebraic variety*, while the ill-conditioned problems form a *tubular neighborhood*. Though Demmel's model is very general, we will explain it for the case of real and complex $n$ by $n$ matrices, which can be identified with the Euclidean spaces $\mathbf{R}^{n^2}$ and $\mathbf{R}^{2n^2}$. The ill-posed matrices lie on a submanifold defined by $\det(A) = 0$. Since $\det(A)$ is a polynomial, algebraic geometers refer to this manifold as an algebraic variety. $A$ is ill-conditioned when it is, in a sense that can be made precise, close to this manifold. It naturally follows that ill-conditioned matrices lie in a tubular neighborhood around the manifold, i.e., the set of points within a certain distance of the manifold. Demmel found it convenient to define the condition number $\kappa_D = \|A\|_F \|A^{-1}\|_2$.

With his model, Demmel concluded that for complex matrices with normally distributed elements

$$\frac{(1-x^{-1})^{2n^2-2}}{2n^4x^2} \leq \text{Prob}(\kappa_D \geq x) \leq \frac{e^2n^5(1+n^2/x)^{2n^2-2}}{x^2}. \tag{2.7}$$

He also computed the large $x$ behavior for a fixed $n$:

$$\text{Prob}(\kappa_D \geq x) = \frac{n(n^2-1)}{x^2} + o(\frac{1}{x^2}). \tag{2.8}$$

The bounds given in (2.7) are not very tight. For example, for $x \approx n$ and $n$ large we get a lower bound growing like $e^{-2n}$ and an upper bound growing like $e^{2n}$. In Section 7.3, we will derive the exact result

$$P(\kappa_D \geq x) = 1 - (1 - n/x^2)^{n^2-1},$$

from which we can readily derive Demmel's asymptotic result (2.8) as a corollary.

When solving linear systems by Gaussian elimination with partial (or complete) pivoting, the term $f(n)$ in (2.4) is always factored as $g(n)p(n)$ where $g(n)$ is the *growth factor* and $p(n)$ is a polynomial. The worst case bound for $g(n)$ is $2^{n-1}$ for partial pivoting and is conjectured to be $n$ for complete pivoting, but what about the average $g(n)$? For partial pivoting, the growth factors that appear in practice are much smaller than $2^{n-1}$, but the this phenomenon represents an important gap in our understanding of partial pivoting. Recently, Trefethen and Schreiber [47] have studied this question and have given evidence that the growth factor is another example of a quantity whose average is far better than its worst case. Through careful experiments and the beginnings of a theory, they found that the average growth factors behave approximately like $n^{2/3}$ for partial pivoting and $n^{1/2}$ for complete pivoting.

## 2.2 Multivariate statistics

As important a role as the eigenvalues of $A^TA$ have in numerical analysis, they play an even more crucial role in multivariate statistics. Classical multivariate analysis considers random vectors $v \in \mathbf{R}^m$ such that for all (non-random) $w \in \mathbf{R}^m$, $w^Tv$ is a univariate normal. This is the $m$-variate normal distribution[7] with mean and covariance

$$\mu = E(v), \quad \Sigma = E((v-\mu)(v-\mu)^T).$$

Typically, $n$ samples from this distribution are collected as columns in an $m$ by $n$ matrix $A$. It is then natural to form the random matrix $W = AA^T$ which is so important that it has its own name, a *Wishart matrix*,[8] named for J. Wishart, who first computed the joint density of its elements in 1928 [59].

When $\mu = 0$, $W$ is said to have the *central Wishart distribution*, sometimes denoted $W_m(n, \Sigma)$. If $v_1, \ldots, v_n$ are samples from a multivariate normal distribution, then the *sample covariance matrix* $W$ is defined as $\frac{1}{n-1}\sum_{i=1}^n (v_i - \bar{v})(v_i - \bar{v})^T$, where $\bar{v} = \frac{1}{n}\sum_{i=1}^n v_i$. Clearly sample covariance matrices have the central Wishart distribution.

Eigenvalues of Wishart matrices were first studied as far back as 1939, when Fisher, Hsu, and Roy independently computed their joint distribution for $\Sigma = I$.[9] Since then, a large body of literature has emerged on the eigenvalues of Wishart matrices. The best introductions to the subject are [2], [36], [58]; we particularly recommend [36].

Following von Neumann's lead, we will be mostly concerned with matrices $A$ of elements that are independent standard normals. Equivalently, we are interested in the case in which the covariance matrix is the identity. We say $W$ is $W(m,n)$ if $W = AA^T$, where $A$ is an $m$ by $n$ matrix of independent standard normals.

Just as eigenvalues of a fixed matrix give much information about its underlying structure, eigenvalues of sample covariance matrices give information about the underlying distribution.

[7] In Galton's classic example, measurements are taken of the heights of fathers and sons. Each height is normally distributed, but there are correlations between the height of a father and that of his son.

[8] Though numerical analysts generally work with $A^TA$, we find it convenient here, and throughout the thesis, to consider $AA^T$.

[9] We derive this joint density in Section 3.4. Goldstine and von Neumann used this joint distribution to estimate $\lambda_{\max}$. See Equation (8.2) in [49].

One well-studied area is *principal components analysis*. In principal components analysis, one rewrites the samples $v_n$ in the basis of the eigenvectors of $\Sigma$. The components of the vector so formed are known as the principal components and their variances are the eigenvalues of $\Sigma$. Another important application of the condition number of a random matrix is the *sphericity test*, which evaluates the hypothesis that $\Sigma = \lambda I$, i.e., all the eigenvalues of $\Sigma$ are equal. See [36] for yet further applications.

In Chapter 4, we derive new results concerning the exact distribution of the smallest eigenvalue of a Wishart matrix from $W(m, n)$. These distributions can be used for tests of the null hypothesis that the covariance matrix is $I$. In combination with results on the largest eigenvalue, this can also be used for a sphericity test as can the distribution of Demmel's condition number described in Section 2.1.

Since we are focusing on the special case $\Sigma = I$, we are regretfully ignoring the beautiful theory of zonal polynomials, which are symmetric multivariate polynomials of the eigenvalues of a symmetric matrix. Their natural definition arises from group representation theory [27]. Hypergeometric functions of a matrix argument are then defined in terms of these zonal polynomials. The crowning achievement of this theory is that, with these functions, the distributions of the eigenvalues of Wishart matrices with arbitrary covariance $\Sigma$ take a particularly simple form. Unfortunately, despite this mathematical elegance, the zonal polynomials are notoriously difficult to compute.

## 2.3 Nuclear physics

The quantum theory predicts that the discrete energy levels of an atomic nucleus correspond to the eigenvalues of a Schrödinger operator. Unfortunately, for heavy atoms, the operator is too complex, so physicists have theorized that the Schrödinger operator could be replaced by a random Hermitian matrix. Its eigenvalues would then, it was hoped, correspond to the observed energy levels.

Though many probability distributions have been proposed, the most famous are the *Gaussian ensembles* proposed by Wigner[10] [54]. We will only mention the Gaussian orthogonal ensemble (GOE) and the Gaussian unitary ensemble (GUE). The Gaussian orthogonal ensemble can be defined as $(A + A^T)/2$, where the elements of $A$ are independent standard normals. It has the property of being invariant under orthogonal transformations. The GUE is a complex version that is invariant under unitary transformations.

Wigner derived the joint eigenvalue distribution for these matrices in 1962 [54]. We will give a cleaner derivation from the multivariate statistics literature in Section 3.3. He is also responsible for the famous "semi-circle" law, which states that the histogram for the eigenvalues of a large random matrix is roughly a semi-circle. To be more precise, let $A$ be a random matrix with independent standard normal elements. Let $M(x)$ denote the proportion of eigenvalues of $(A + A^T)/(2\sqrt{n})$ that are less than $x$. The semi-circle law states that as $n \to \infty$, $\frac{d}{dx}M(x)$ converges to $\frac{2}{\pi}(1 - x^2)^{1/2}$ on $[-1, 1]$, 0 elsewhere. In fact, Wigner proved this statement for random matrices that are more general than the GOE. Properly normalized, the conclusion applies to random symmetric matrices with independent elements taken from any distribution with finite moments, mean 0, and variance matching those of the GOE or GUE. Since Wigner, further progress has been made concerning the semi-circle law, including simpler proofs and strengthening of the convergence.[11] (One nice example is [19].)

To match experiments with theory and conversely, to make predictions from the theory, physicists have needed more information about the eigenvalues of the GOE and GUE. Specifically, they have been interested in the random variable giving the spacings between consecutive eigenvalues. When the eigenvalues are properly normalized, this random variable has the same

[10] Wigner and von Neumann were schoolmates in Budapest and colleagues at Princeton.
[11] There has even been an argument based on principles of statistical mechanics [35].

distribution no matter where the eigenvalue lies in the spectrum. More generally, there are physical reasons to study the $k$th eigenvalue larger than a given eigenvalue, known as the *$k$th nearest spacing*. Another important quantity is the *pair correlation function*, which gives the average number of eigenvalues in a fixed interval. The distributions of these random variables have been computed for the GOE and GUE in the limit $n \to \infty$. For further details, we refer the reader to [13], [35], and [39].[12]

## 2.4 The Riemann zeta function

Quite remarkably, there appears to be an intimate connection between the eigenvalues of a random matrix and the zeros of the Riemann zeta function. It has been conjectured that $\zeta(\frac{1}{2} + \lambda i)$ has real roots $\lambda$ because somehow the roots correspond to the eigenvalues of a Hermitian operator. Odlyzko [37] has tested this conjecture through an enormous computational effort; he computed 70 million zeros near zero number $10^{20}$ and several other large sets of zeros as well. He found that the eigenvalues of matrices from the GUE (which we will define in Section 3.1) modeled the zeros amazingly well. By plotting the theoretical correlation functions and the $k$th nearest spacing functions (briefly mentioned in the previous section) for the GUE against histograms of observed data from the zeros of zeta, he observed the plots were nearly identical.

## 2.5 Graph theory

Researchers in graph theory are finding interesting new problems by asking whether a random graph has a certain property. Sometimes, these problems can be transformed into a random matrix problem where the eigenvalues play an important role.

A graph can be described by its incidence matrix, $A_{ij}$, which is 1 if vertices $i$ and $j$ are connected by an edge, otherwise 0 . More generally, we can allow arbitrary positive integers for multiple edges and negative integers for directed graphs. If instead of a fixed graph, we take random graphs, we then have a distribution of random matrices.

[12] Some of this literature is very difficult to penetrate for the non-physicist. We hope this summary and the information in Chapter 3 will make it easier.

Wigner's original proof of his semi-circle law has an interesting interpretation in the graph theory context. Wigner's proof was based on calculating $\mathrm{tr}(A^k)$ for random matrices $A$ and observing that these traces, when normalized, converge. These limits give the $k$th moment of the eigenvalue distribution since $E(\lambda^k) = \mathrm{tr}(A^k)/n$. If $A$ is an incidence matrix of a graph, $\mathrm{tr}(A^k)$ counts the number of paths in the graph that return to the starting point. The $k$th moment of the eigenvalue then is related to the probability that a random walk of length $k$ will return to its starting point. Computation of the "semi-circle" (the functions actually are no longer semi-circles) for regular graphs was done in [33]. For an estimate of the second eigenvalue of a random regular graph, see [18]. Another interesting application can be found in [12], where the eigenvalues of a random graph are related to the important problem of graph bisection. See [11] for a survey of random graph theory.

## 2.6 Contributions of this thesis

This thesis settles many of the random eigenvalue and condition number questions described in this chapter by deriving the exact distributions and expectations in some cases and limiting formulas in other cases.

As an addition to the table of contents, we now itemize the results that we consider most important and where they can be found. We remark that starting with Chapter 4, all statements labeled as theorems represent, we believe, new results, while important results from the literature are labeled as propositions.

- Condition number distribution of large square real matrices with elements from a standard normal distribution: Theorem 7.1, p. 69. Figure 2.1, p. 25 plots the limiting distribution of $\kappa/n$ as $n \to \infty$. The average logarithm of the distribution is also given in Theorem 7.1, and Figure 2.3 on p. 27 plots the distribution of $\log \kappa/n$.

- Condition number distribution of large square complex matrices with elements from a complex standard normal distribution: Theorem 7.2, p. 70. Figure 2.2, p. 26 plots the limiting distribution of $\kappa/n$ as $n \to \infty$. The average logarithm of the distribution is also given in Theorem 7.2, and Figure 2.4 on p. 28 plots the distribution of $\log \kappa/n$.

- Condition number (degenerate) distribution of large real or complex rectangular matrices with elements from a real or complex standard normal distribution: Theorem 7.3, p. 71. The average logarithm of the distribution is also given in Theorem 7.3.

- Condition number distribution for Demmel's problem of complex matrices uniformly distributed on the sphere: Theorem 7.7, p. 76.

- Smallest eigenvalue distribution of the matrix $AA^T$, where $A$ is a square matrix with elements from a real standard normal distribution: Theorem 4.1, p. 41.

- Smallest eigenvalue distribution of the matrix $AA^T$, where $A$ is any $m$ by $n$ matrix with elements from a real standard normal distribution: Theorem 4.3, p. 45.

- Form of the moments of the smallest eigenvalue distribution of the matrix $AA^T$, where $A$ is an $m$ by $n$ matrix ($m$ and $n$ not both even) with elements from a real standard normal distribution: Theorem 8.1, p. 81.

- Expected smallest eigenvalues of the matrix $AA^T$, where $A$ is an $m$ by $n$ matrix ($m \leq 25$, $0 \leq n - m \leq 25$) with elements from a real standard normal distribution: Appendix B, p. 93.

- Average characteristic polynomials of various random matrices: Section 9.2, p. 86.

There are also other results about eigenvalues and condition number distributions not mentioned here.

We like to think that each of our descriptions of the eigenvalues of random matrices is a kind of solution to the question that we raised in the first two paragraphs of the Foreword. We believe, however, that our result on the average smallest eigenvalue of a Wishart matrix comes closest in spirit to the classical task of seeking constructible (with straight edge and compass) solutions to problems.

Here are two examples of our results. Let $A$ be an $m$ by $n$ matrix with independent elements from the standard normal distribution. Let $W = AA^T$. If $m = n = 25$, the expected value of the smallest eigenvalue of $W$ is

$$30285573653/1077824526597.$$

On the other hand, if $m = n = 27$, the expected value is $(p - q\sqrt{3})/r$, with

$$p = 931617797994681132254$$

$$q = 461090719632381299712$$

$$r = 511888460463800314 6675.$$

These are exact answers in what might be called simplest form. We show in Theorem 8.1 that so long as $m$ and $n$ are not both even, the expected value is always either in the rationals or the quadratic extension of the rationals formed by adjoining $\sqrt{m}$. Thus, they are roots of quadratics. Unfortunately, when $m$ and $n$ are both even, an arcsine of one of the parameters appears, and the expectation in this case is probably a transcendental number.

Table 2.1: Properties of the distributions of $\kappa/n$

| | Real | Complex |
|---|---|---|
| mode | 1.71 | 1.63 |
| median | 3.67 | 2.40 |
| geom. mean | 4.65 | 2.67 |
| arith. mean | $\infty$ | 3.54 |
| $P(\kappa/n) < 1$ | 0.02 | 0.02 |
| $P(\kappa/n) < 10$ | 0.80 | 0.96 |
| $P(\kappa/n) < 100$ | 0.98 | 0.9996 |

Our principal goal, however, was to settle the questions about condition numbers that were raised by von Neumann, Birkhoff, Smale, and Demmel. (See Section 2.1.) To this end, we derive the formula that

$$\lim_{n\to\infty} P(\kappa/n < x) = e^{-2/x-2/x^2}, \tag{2.9}$$

where the matrices have elements from a standard normal distribution. Figure 2.1 plots the density $\frac{d}{dx}e^{-2/x-2/x^2}$ with much useful information. The portion of the curve shown encloses an area of 86%, while the boundaries of the alternating shaded and unshaded regions mark off the deciles. A close inspection shows that the median is slightly less than 3.7. The solid dots in Figure 2.5.a show observed data from 25,000 random matrices with $n = 75$. We see that they agree very well, and predictions about the condition number distribution for large $n$ can be based on the plots.

Figure 2.2 plots the density and deciles for complex matrices with the same scale as the previous plot. This curve, however, encloses an area of 98%. Comparing the two curves, one notices the slower convergence of the tail of the real distribution. In fact, the real density converges to zero so slowly that it has an infinite mean, while the complex density has a finite mean.

Table 2.1 above lists some interesting properties of distributions of $\kappa/n$ for real and complex matrices.

The geometric mean result given in the table solves Smale's problem, while the $P(\kappa/n) < 10$ result for real matrices shows that von Neumann's estimate allows for 20% of random matrices to have larger condition numbers. We should mention that the geometric mean result can be computed from (2.9), but the result does not rigorously follow from this computation; more care needs to be taken because of the logarithm. In fact, we can extend Smale's problem by asking for the distribution of $\log \kappa/n$. We plot the limiting distributions and deciles as $n \to \infty$ for $\log \kappa/n$ for real matrices in Figure 2.3 and for complex matrices in Figure 2.4.

We further remark that many of the results in this thesis provide models for similar questions for other distributions. For example, Figure 2.5.b plots the same density as Figure 2.5.a, but the histogram is based on the condition numbers of 25,000 75 by 75 matrices from the uniform distribution $[-1, 1]$. Similarly, Figure 2.5.c is based on the condition numbers of 25,000 75 by 75 matrices from the discrete distribution $\{-1, 1\}$. Figure 2.5.d is based on the condition numbers of 25,000 75 by 75 matrices from the uniform distribution $[0, 1]$ which has non-zero mean. Notice that we no longer use the histogram of $\kappa/n$, but rather, $\frac{2}{\sqrt{3}}\kappa/n^{3/2}$. Thus these matrices are less well conditioned. See Section 7.1 for more about these matters.

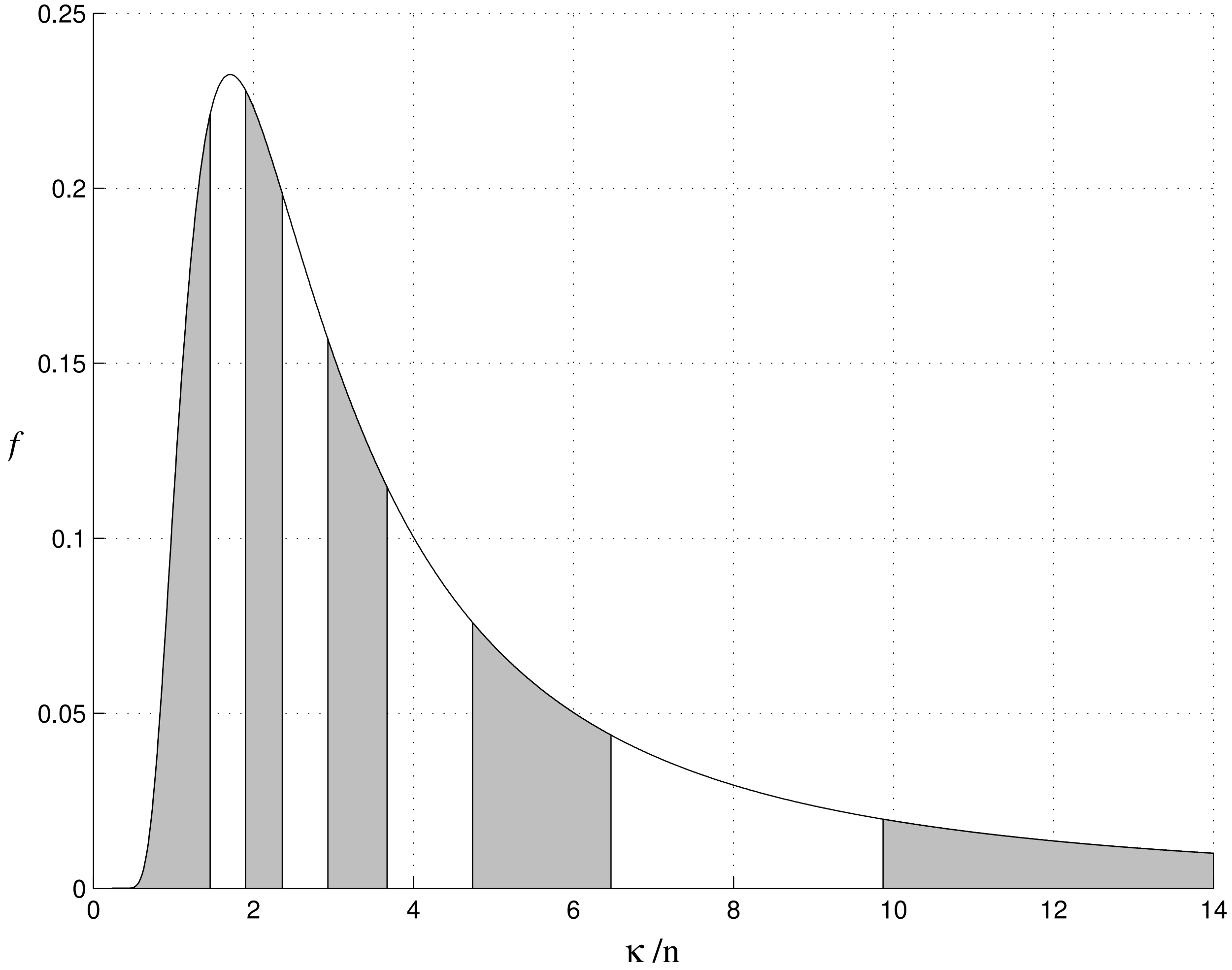


Figure 2.1: The limiting density of $\kappa/n$ ($n \times n$ real matrices)

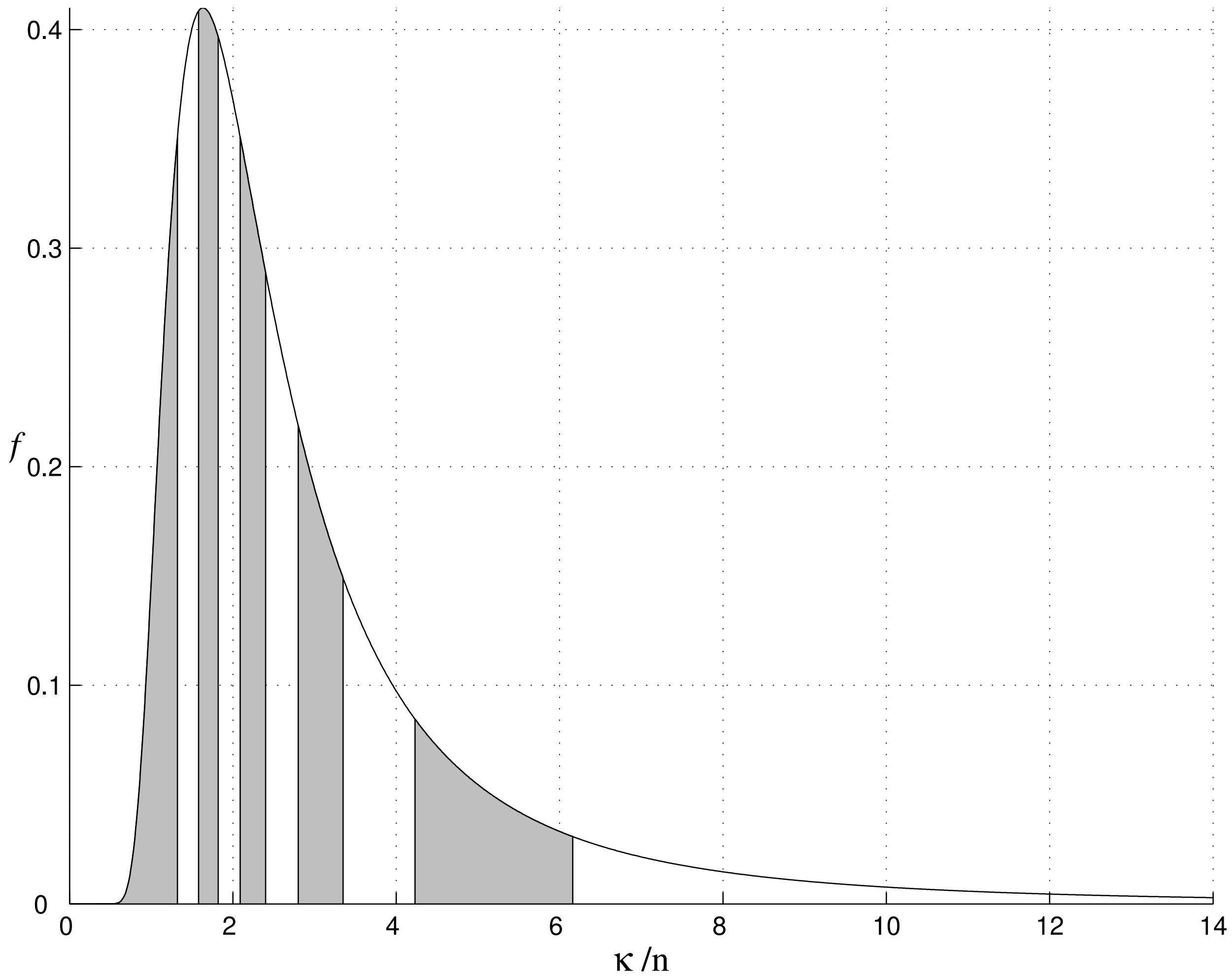


Figure 2.2: The limiting density of $\kappa/n$ ($n \times n$ complex matrices)

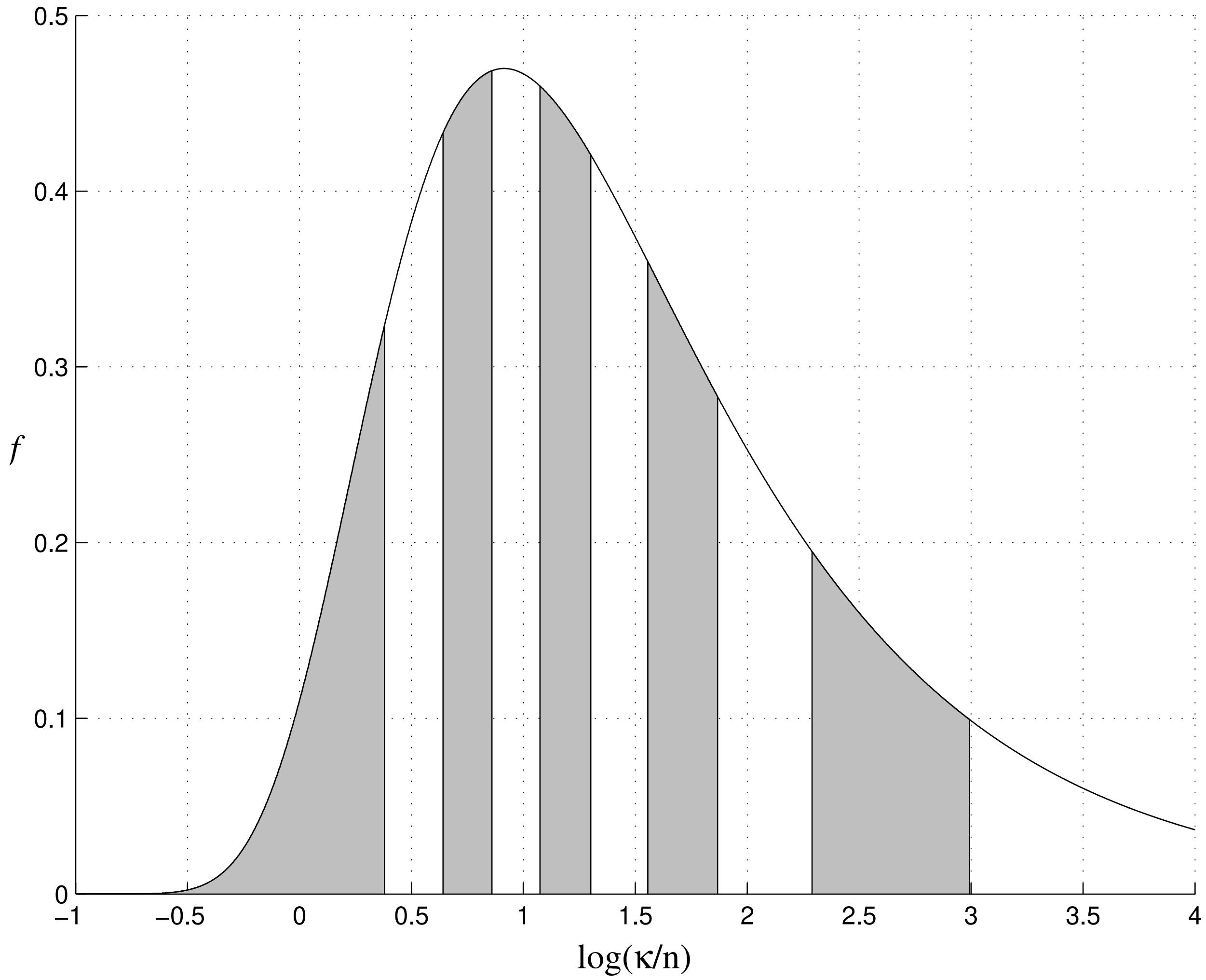


Figure 2.3: The limiting density of $\log(\kappa/n)$ ($n \times n$ real matrices)

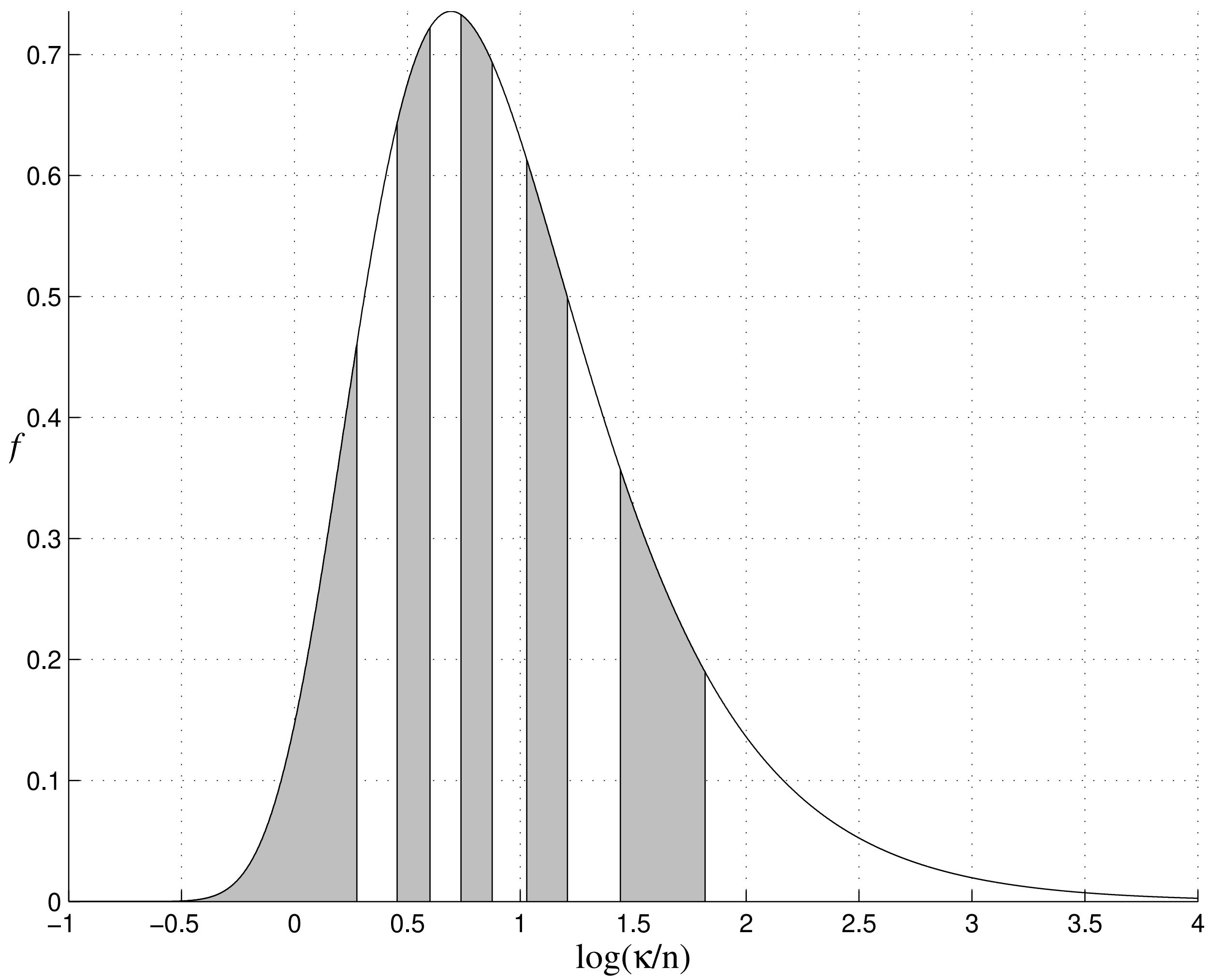


Figure 2.4: The limiting density of $\log(\kappa/n)$ ($n \times n$ complex matrices)

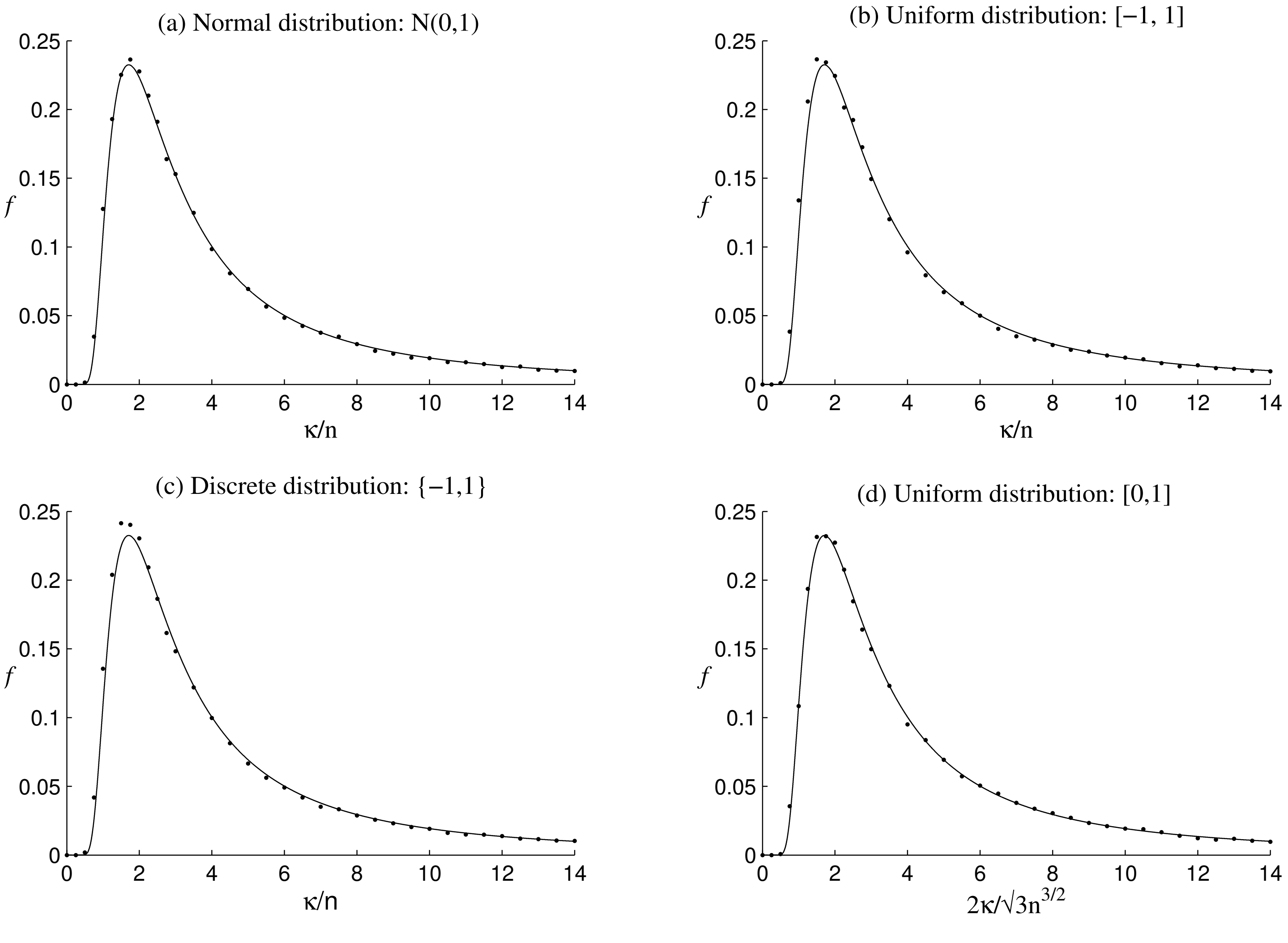


Figure 2.5: Experimental $\kappa$ for 50,000 matrices, $n = 75$

# Chapter 3

# Fundamental concepts

## 3.1 Random matrix notation and terminology

In elementary statistics, the most important distribution is $N(\mu, \sigma^2)$, the normal distribution with mean $\mu$ and variance $\sigma^2$. The normal distribution deserves its special place in part because of the central limit theorem, which, loosely interpreted, states that large sums of random variables often behave as if they were normally distributed.

In multivariate statistics, the most important distributions are derived from the normal. Again statisticians have discovered that multivariate generalizations of the normal distribution often suffice for their models. Following this principle, we will be primarily interested in the following real random matrices derived from the normal distribution.

**Gaussian ($G(m,n)$)** $m \times n$ random matrix with independent and identically distributed (**iid**) elements which are $N(0,1)$.

**Wishart ($W(m,n)$)** Symmetric $m \times m$ random matrix $AA^T$, where $A$ is $G(m,n)$.

**Gaussian Orthogonal Ensemble (GOE)** Symmetric $m \times m$ random matrix with iid elements that are $N(0, \frac{1}{2})$ on the upper-triangle and iid elements that are $N(0,1)$ on the diagonal. Equivalently, it is $(A + A^T)/2$, where $A$ is $G(m,m)$.

These random matrices have complex analogues which we will also study in depth. Let $\tilde{N}(\mu, \sigma^2)$ refer to the distribution of $x + iy$, where $x$ and $y$ are iid $N(\mu, \sigma^2)$.

**Complex Gaussian ($\tilde{G}(m,n)$)** $m \times n$ random matrix with iid elements which are $\tilde{N}(0,1)$.

**Complex Wishart ($\tilde{W}(m,n)$)** Hermitian $m \times m$ random matrix $AA^H$, where $A$ is $\tilde{G}(m,n)$.

**Gaussian Unitary Ensemble (GUE)** Hermitian $m \times m$ random matrix with iid elements that are $\tilde{N}(0, \frac{1}{4})$ on the upper-triangle and iid elements that are $N(0, \frac{1}{2})$ on the diagonal. Equivalently, it is $(A + A^H)/\sqrt{8}$, where $A$ is $\tilde{G}(m,m)$.

The arbitrary normalizations were chosen to most closely match the literature with which we were familiar.

## 3.2 Jacobians for matrix factorizations and Haar measure

Little known to many researchers in linear algebra is the fact that the familiar matrix factorizations, which can be viewed as changes of variables, have simple Jacobians. These Jacobians are used extensively in applications of random matrices in multivariate statistics and physics. As an example, consider the Cholesky factorization $A = LL^T$, where $A$ is symmetric positive definite and $L$ is lower triangular with positive diagonal elements. This factorization can be thought of as a change of variables from the $n(n+1)/2$ independent elements of $A$ to the $n(n+1)/2$ potentially non-zero elements of $L$. This change of variables is well defined since every symmetric positive definite matrix has one and only one Cholesky factorization. We recall that if we change variables from a vector $x$ to a vector $y$, the Jacobian is the determinant

$$\begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & & \vdots \\ \frac{\partial x_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}. \tag{3.1}$$

For a rectangular matrix $A$ with $mn$ independent elements, let

$$dA = \prod_{i,j} da_{ij}$$

denote the differential volume element. Following [36], we will abuse notation by using the same notation to denote the volume elements of more special matrices. For example, if $S$ is symmetric or $L$ is lower triangular, let

$$dS = \prod_{i \le j} ds_{ij}$$

and

$$dL = \prod_{i \le j} dl_{ij}.$$

In these cases the differential product includes only the $n(n+1)/2$ independent parameters. If $\Lambda$ is a diagonal matrix, then we let

$$d\Lambda = \prod_{i=1}^{n} d\lambda_{ii}.$$

For orthogonal matrices $Q$, $dQ$ will be a bit more complicated. Let $Q$ be an $m$ by $n$ matrix[1] with $QQ^T = I$. As a subspace of $\mathbf{R}^{mn}$, these matrices form a submanifold $V_{m,n}$ of dimension $mn - \frac{1}{2}m(m+1)$ called the *Stiefel manifold.* Let $dQ$ denote the volume element on $V_{m,n}$ induced by this embedding. It can be shown that the total volume of $V_{m,n}$ is

$$\int_{V_{m,n}} dQ = \frac{2^m \pi^{mn/2}}{\Gamma_m(n/2)}, \tag{3.2}$$

where the multivariate gamma function is defined by

$$\Gamma_m(a) = \pi^{m(m-1)/4} \prod_{i=1}^{m} \Gamma(a - (i-1)/2). \tag{3.3}$$

For $m = 1$, $dQ$ is the surface element of the sphere and (3.2) reduces to the surface volume of the sphere. When $m = n$, $dQ/\text{Volume}(V_{n,n})$ is often called the *(invariant) Haar measure.*

The most important property of Haar measure is its invariance under orthogonal transformation. This means

$$\int_{\mathcal{S}} dQ = \int_{Q\mathcal{S}} dQ,$$

where $\mathcal{S}$ is any (Lebesgue measurable) set of orthogonal matrices, $Q$ is any particular orthogonal matrix, and $Q\mathcal{S}$ is the set of products of $Q$ with elements of $\mathcal{S}$. The measure $dQ/\text{Volume}(V_{n,n})$ then defines a probability space on the orthogonal matrices which is invariant with respect to

[1]Here, and in the entire thesis, we always assume $m \le n$.

orthogonal transformations. For further details consult [36]. For information on the numerical generation of random orthogonal matrices distributed with Haar measure see [3] and [44].

For complex matrices, the volume elements are given in terms of the real and imaginary parts. Thus if $A$ is a complex rectangular matrix with $A = A_r + iA_c$, then $dA$=$dA_r dA_c$. If $H = H_r + iH_c$ is Hermitian, $dH = dH_r dH_c$. Since $H_c$ is an antisymmetric matrix, its volume element is the product of the $n(n-1)/2$ independent variables above the diagonal. If $L$ is complex lower triangular with real positive diagonals, the volume element has the same form as that for Hermitian matrices.

Unitary matrices are analogous to orthogonal matrices. Let $\tilde{Q}$ be an $m$ by $n$ matrix with $\tilde{Q}\tilde{Q}^H = I$. As a subpace of $\mathbf{R}^{2mn}$, these matrices form a submanifold $\tilde{V}_{m,n}$ of dimension $2mn - m^2$ with volume element $d\tilde{Q}$. It can be shown[2] that the total volume of $\tilde{V}_{m,n}$ is

$$\int_{\tilde{V}_{m,n}} d\tilde{Q} = \frac{2^m \pi^{mn}}{\tilde{\Gamma}_m(n)}, \tag{3.4}$$

where the complex multivariate gamma function  is defined by

$$\tilde{\Gamma}_m(a) = \pi^{m(m-1)/2} \prod_{i=1}^{m} \Gamma(a - i + 1). \tag{3.5}$$

Now that we have defined the differential volume elements, we need to know how to transform between them. The transformations are given in the following theorem.

**Theorem 3.1** *The changes of variables for the common factorizations have the following Jacobians*

$$\begin{array}{ll} \textit{Real} & \left\{ \begin{array}{lll} \textit{Cholesky} & S = LL^T & dS = 2^m \prod_{i=1}^m l_{ii}^{m+1-i} dL \\ \textit{LQ} & A = LQ & dA = \prod_{i=1}^m l_{ii}^{n-i} dL dQ \\ \textit{eigenvalue} & S = Q\Lambda Q^T & dS = \prod_{i<j} |\lambda_i - \lambda_j| d\Lambda dQ \end{array} \right. \\ \textit{Complex} & \left\{ \begin{array}{lll} \textit{Cholesky} & H = LL^H & dH = 2^m \prod_{i=1}^m l_{ii}^{2m-2i+1} dL \\ \textit{LQ} & A = L\tilde{Q} & dA = \prod_{i=1}^m l_{ii}^{2n-2i+1} dL d\tilde{Q} \\ \textit{eigenvalue} & H = \tilde{Q}\Lambda\tilde{Q}^H & dH = \prod_{i<j} (\lambda_i - \lambda_j)^2 d\Lambda d\tilde{Q} \end{array} \right. \end{array}$$

*Proof* The proof of the Cholesky factorization results is particularly easy because the Jaco-

[2]We computed this volume, but it must appear in the literature.

bian matrix is itself upper triangular. The LQ and eigenvalue decomposition results are more complicated. See [36] for a clear exposition on how to derive these results in the real case. The complex Cholesky factorization appears in [22]. We have not seen the complex LQ result in the literature and derived it for ourselves, but will not give the details here. The eigenvalue formulas for real and complex matrices appear in [35] and [54]. $\square$

The LQ formulas hold for rectangular matrices, so that if $A$ is $m$ by $n$, then $Q$ ($\tilde{Q}$) is $m$ by $n$ and $L$ is $m$ by $m$. To make the eigenvalue factorization unique, we assume that the diagonal elements of $L$ are ordered from greatest to least and that the first row of $Q$ ($\tilde{Q}$) is real and non-negative. Thus the eigenvector matrices $Q$ only occupy a proportion $2^{-m}$ of the volume of the Stiefel manifold, and the eigenvector matrices $\tilde{Q}$ occupy a submanifold of the unitary manifold. We are not concerned with sets of measure zero, such as the set of matrices with multiple eigenvalues.

## 3.3 Joint element densities

As already mentioned, the matrix distributions in this thesis are derived from the familiar univariate normal distribution, $N(\mu, \sigma^2)$, with probability density function (pdf)

$$(2\pi\sigma^2)^{-1/2}\exp(\frac{-(x-\mu)^2}{2\sigma^2}).$$

To construct the joint element densities for random matrices with independent elements, one simply takes the product of the densities of the independent elements. We note that $\mathrm{tr}(AA^T) = \sum_{i,j} a_{ij}^2$ and that $AA^T$ ($AA^H$) is unchanged when $A$ is multiplied by an orthogonal (unitary) matrix. Thus,

**Lemma 3.1** *The joint densities for the indicated random matrices are*

$$\begin{array}{ll} G(m,n) & (2\pi)^{-mn/2}\mathrm{etr}(-\frac{1}{2}AA^T) \\ GOE & 2^{-n/2}\pi^{-n(n+1)/4}\mathrm{etr}(-\frac{1}{2}A^2) \\ \tilde{G}(m,n) & (2\pi)^{-mn}\mathrm{etr}(-\frac{1}{2}AA^H) \\ GUE & 2^{n(n-1)/2}\pi^{-n^2/2}\mathrm{etr}(-A^2) \end{array}$$

*where* $\mathrm{etr}(A) = \exp(\mathrm{tr}(A))$*. Furthermore, these distributions are invariant under orthogonal (unitary) transformations.*

The elements of a Wishart matrix are generally not independent, and thus the joint distribution is more complicated. Following [36], we compute it in two steps using the LQ and Cholesky formulas in Theorem 3.1 and the joint element density for $G(m,n)$ given in Lemma 3.1. The following diagram outlines the calculation.

$$\begin{tikzcd} A \arrow[r, "A = LQ"] \arrow[dr, "W = AA^T"'] & L \arrow[d, "W = LL^T"] \\ & W \end{tikzcd}$$

The top (horizontal) arrow projects $A$ onto the space of upper triangular matrices. Integrating the $Q$ component over the Stiefel manifold gives that the joint density of $L$ is

$$\frac{2^{m-mn/2}}{\Gamma_m(n/2)}\mathrm{etr}(-\frac{1}{2}LL^T)\prod_{i=1}^{m} l_{ii}^{n-i}.$$

This is essentially the decomposition due to Bartlett [5] which states that the elements $l_{ij}$ are independent, $l_{ii}^2$ is $\chi^2_{n-i+1}$, and the off-diagonal elements are standard normals.

The right (vertical) arrow is a one-to-one map of $L$ onto the space of symmetric matrices. Via the Jacobian of the Cholesky factorization in Theorem 3.1, we can derive the joint density of the elements for a real Wishart matrix. The complex case is analogous. Thus,

**Lemma 3.2** *The joint density of the elements of a matrix from* $W(m,n)$ *is*

$$\frac{1}{2^{mn/2}\Gamma_m(n/2)}\mathrm{etr}(-\frac{1}{2}W)(\det W)^{(n-m-1)/2}, \tag{3.6}$$

*while for a matrix from $\tilde{W}(m,n)$, it is*

$$\frac{1}{2^{mn}\tilde{\Gamma}_m(n)}\operatorname{etr}(-\frac{1}{2}W)(\det W)^{n-m}. \tag{3.7}$$

A curious observation is that when $n = m$ in the complex case, or $n = m + 1$ in the real case, the $\det W$ term vanishes, and thus the off-diagonal parameters are independent aside from the constraint that the matrix be positive semi-definite.

## 3.4 Joint eigenvalue densities

Here we calculate the joint densities of eigenvalues for real and complex Wishart matrices and Gaussian ensembles. We believe this is the first time that all these calculations have been gathered in one place and given as corollaries of the same theorem.[3]

**Theorem 3.2** *Let the real symmetric, $m \times m$ matrix $S$ have a joint density function $f(S)$ which is invariant under orthogonal similarity transformations. Then the joint density of the $m$ eigenvalues of $S$, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m$ is*

$$\frac{\pi^{m^2/2}}{\Gamma_m(m/2)}f(\Lambda)\prod_{i<j}(\lambda_i - \lambda_j). \tag{3.8}$$

*If, instead, $S$ is Hermitian and $f(S)$ is invariant under unitary similarity transformations, then the joint density is*

$$\frac{\pi^{m(m-1)}}{\tilde{\Gamma}_m(m)}f(\Lambda)\prod_{i<j}(\lambda_i - \lambda_j)^2. \tag{3.9}$$

*Proof* The proof for real matrices follows from Theorem 3.1 by integrating the $Q$ component in the eigenvalue Jacobian, i.e., multiplying by the volume of $V_{n,n}$ and then dividing by $2^m$, since we assume the first row of $Q$ is non-negative. The complex proof is analogous, but we must divide by $(2\pi)^m$ for the arbitrary phases of the $m$ elements in the first row of $\tilde{Q}$. □

**Corollary 3.1** *The joint densities of the eigenvalues $\lambda_1 \geq \ldots \geq \lambda_m$ of the symmetric (Hermi-*

[3]The real case, however, is well covered in [2].

*tian) random matrices defined in Section 3.1 are*

$$W(m,n) \qquad \frac{\pi^{m^2}}{\Gamma_m(m/2)\Gamma_m(n/2)} \exp(-\frac{1}{2}\sum \lambda_i) \prod \lambda_i^{(n-m-1)/2} \prod_{i<j}(\lambda_i - \lambda_j) \tag{3.10}$$

$$GOE \qquad \frac{1}{2^{n/2}\prod_{i=1}^{n}\Gamma(i/2)} \exp(-\frac{1}{2}\sum \lambda_i^2) \prod_{i<j}(\lambda_i - \lambda_j) \tag{3.11}$$

$$\tilde{W}(m,n) \qquad \frac{2^{-mn}\pi^{m(m-1)}}{\tilde{\Gamma}_m(n)\tilde{\Gamma}_m(m)} \exp(-\frac{1}{2}\sum \lambda_i) \prod \lambda_i^{n-m} \prod_{i<j}(\lambda_i - \lambda_j)^2 \tag{3.12}$$

$$GUE \qquad \frac{2^{n(n-1)/2}}{\pi^{n/2}\prod_{i=1}^{n}\Gamma(i)} \exp(-\sum \lambda_i^2) \prod_{i<j}(\lambda_i - \lambda_j)^2, \tag{3.13}$$

*where the unlabeled sums and products run from $i = 1$ to $m$. For the Wishart cases, this is the joint distribution of the non-negative eigenvalues, while for the Gaussian ensembles, the eigenvalues may be anywhere on the real line.*

The joint density for the eigenvalues of real Wishart matrices was derived in three independent papers in 1939 by Fisher, Hsu, and Roy. James in 1960 computed the joint distribution for the more general case of arbitrary covariance matrices. See [27] for references and a survey of some of the related history in multivariate statistics. The equivalent calculations in physics were first performed by Wigner in 1962 [54].

## 3.5 Bidiagonalization of a Gaussian matrix

The standard method (due to Golub and Kahan) for computing the singular values of a general matrix $A$ is to first reduce it to bidiagonal form through orthogonal transformations. If $A$ is a random Gaussian matrix, the resulting random bidiagonal matrix takes a particularly simple form.

**Theorem 3.3** *If $A$ is $G(m,n)$, then $A$ is orthogonally similar to an $m$ by $n$ matrix*

$$\begin{pmatrix} x_n & & & & 0 & \cdots & 0 \\ y_{m-1} & x_{n-1} & & & \cdot & & \cdot \\ & \ddots & \ddots & & \vdots & & \vdots \\ & & y_1 & x_{n-(m-1)} & 0 & \cdots & 0 \end{pmatrix}, \tag{3.14}$$

*where $x_j^2$ and $y_j^2$ are distributed as $\chi^2$ variables with $j$ degrees of freedom (i.e. $\chi_i^2$).*[4] *The elements here are all non-negative and independent.*

The idea of the proof is to perform the standard Householder transformations on the random matrix. See [41] or [48] for details. Computing the singular values of a bidiagonal matrix is much less costly than computing them for a dense matrix ($O(n^2)$ operations in practice, rather than $O(n^3)$), so this theorem can be used to speed up random matrix experiments on larger matrices (though we did not find the need to do so in the course of this work).

We also have the complex analogue,

**Theorem 3.4** *If $A$ is $\tilde{G}(m,n)$, then $A$ is unitarily similar to an $m$ by $n$ matrix*

$$\begin{pmatrix} x_{2n} & & & & 0 & \cdots & 0 \\ y_{2(m-1)} & x_{2(n-1)} & & & \cdot & & \cdot \\ & \ddots & \ddots & & \vdots & & \vdots \\ & & y_2 & x_{2(n-(m-1))} & 0 & \cdots & 0 \end{pmatrix}, \tag{3.15}$$

*where the notation is as in Theorem 3.3.*

[4] Recall that the $\chi_j^2$ distribution is defined as the distribution of $j$ sums of squares of standard normals.

From these theorems, we can immediately obtain the interesting facts that the determinant of a matrix from $W(m,n)$ has the distribution $\chi_n^2 \chi_{n-1}^2 \cdots \chi_{n-m+1}^2$. (The notation refers to the distribution of the product of random variables with the indicated $\chi^2$ distributions.) For a matrix from $\tilde{W}(m,n)$, the determinant has the distribution $\chi_{2n}^2 \chi_{2(n-1)}^2 \cdots \chi_{2(n-(m-1))}^2$. Since the expected value of a random variable with distribution $\chi_k^2$ is $k$, we obtain that the expected value of the determinant is $\frac{n!}{(n-m)!}$ in the real case and $2^m \frac{n!}{(n-m)!}$ in the complex case.

Other matrix reductions of Gaussian matrices also have simple forms. We mentioned the Bartlett decomposition in Section 3.3 which gives the distribution of $Q$ and $R$ if the matrix $A$ is factored as an orthogonal matrix $Q$ times an upper triangular matrix $R$. This result had also been observed by Birkhoff and Gulati [9]. The standard method for computing the eigenvalues of a general symmetric matrix $A$ is to first reduce it to tridiagonal form through orthogonal similarity transformations. If we do this for a random symmetric matrix that is from the GOE or GUE, the resulting random tridiagonal matrix also takes a particularly simple form which is mentioned in [48].

# Chapter 4

# Smallest eigenvalue distribution for $W(m,n)$

The distributions of the extreme eigenvalues of Wishart matrices have been known in terms of zonal polynomials [30] and in terms of Pfaffians [31], but we have not seen any closed form expressions in the literature. In this chapter, we will integrate the joint eigenvalue distribution of real Wishart matrices to obtain the exact *pdf* (probability density function) of the smallest eigenvalue. For convenience, we rewrite formula (3.10) for the eigenvalues $\lambda_1 \geq \ldots \geq \lambda_m \geq 0$ as

$$K_{m,n} e^{-\frac{1}{2}\sum_{i=1}^m \lambda_i} \prod_{i=1}^m \lambda_i^{\frac{1}{2}(n-m-1)} \prod_{i<j} (\lambda_i - \lambda_j) d\lambda_1 \ldots d\lambda_m, \tag{4.1}$$

where

$$K_{m,n}^{-1} = \left(\frac{2^n}{\pi}\right)^{m/2} \prod_{i=1}^m \Gamma(\frac{n-i+1}{2})\Gamma(\frac{m-i+1}{2}). \tag{4.2}$$

## 4.1 A multivariate integral for the pdf of $\lambda_{\min}$

Using symmetry, it is easy to see by integrating (4.1) that the probability density function for the smallest eigenvalue $\lambda_{\min}$ is

$$f_{\lambda_{\min}}(\lambda) \quad = \quad \frac{K_{m,n}}{(m-1)!} \lambda^{(n-m-1)/2} e^{-\lambda/2}$$

$$\times \int_{R_\lambda} e^{-\sum_{i=1}^{m-1} \lambda_i/2} \prod_{1 \le i<j \le n-1} |\lambda_i - \lambda_j| \prod_{i=1}^{n-1} (\lambda_i - \lambda)\lambda_i^{(n-m-1)/2} d\lambda_i,$$

where $R_\lambda = \{(\lambda_1, \ldots, \lambda_{m-1}) : \lambda_i > \lambda\}$. Notice that we removed the arbitrary ordering on the $\lambda_i$ and correspondingly divided the density by $(m-1)!$. Now, performing the change of variables $x_i = \lambda_i - \lambda$, we obtain:

**Lemma 4.1** *The pdf of the smallest eigenvalue $\lambda_{\min}$ of a matrix from $W(m,n)$ is*

$$\frac{K_{m,n}}{(m-1)!} \lambda^{\frac{n-m-1}{2}} e^{-\lambda m/2} \int_{R_+^{m-1}} \prod_{i=1}^{m-1} (x_i + \lambda)^{\frac{n-m-1}{2}} \Delta d\Omega, \tag{4.3}$$

*where $\Delta = \prod_{1 \le i<j \le m-1} |x_i - x_j|$, $d\Omega = d\mu(x_1) \ldots d\mu(x_{n-1})$, $d\mu(x) = xe^{-x/2}dx$, and the integration takes place over the positive orthant, $R_+^{m-1} = \{(x_1, \ldots, x_{m-1}) : x_i \ge 0\}$.*

It is the integral in the lemma above that we wish to understand. We have not seen the pdf expressed in this form in the literature. However, as the rest of this chapter illustrates, this form is convenient for obtaining the exact pdf.

## 4.2 The pdf of $\lambda_{\min}$ for $W(m, m)$

The case $m = n$ is the case of interest to von Neumann, Birkhoff, and Smale, as described in Section 2.1. In this case, the integral in (4.3) turns out to satisfy the differential equation for the Tricomi function.

**Theorem 4.1**[1] *The pdf of the smallest eigenvalue of a matrix from $W(m,m)$ is*

$$f_{\lambda_{\min}}(\lambda) = \frac{m}{\sqrt{2\pi}} \Gamma\left(\frac{m+1}{2}\right) \lambda^{-1/2} e^{-\lambda m/2} U(\frac{m-1}{2}, -\frac{1}{2}, \lambda/2).$$

*When $a > 0$ and $b < 1$, the Tricomi function, $U(a,b,z)$, is the unique solution to Kummer's equation*

$$z\frac{d^2w}{dz^2} + (b-z)\frac{dw}{dz} - aw = 0, \tag{4.4}$$

[1] Throughout the remainder of this thesis, all statements labeled as theorems represent new results. Important results from the literature are labeled as propositions.

*satisfying* $U(a,b,0) = \Gamma(1-b)/\Gamma(1+a-b)$ *and* $U(a,b,\infty) = 0$.

*Proof* Let

$$w(\lambda) = \int_{R_+^{m-1}} \prod_{i=1}^{m-1} (x_i+\lambda)^{-1/2} \Delta d\Omega.$$

Our goal is to show that $w(\lambda/2)$ satisfies (4.4). We will need a little more notation. Let $\Delta = \delta\Delta_2$ where $\delta = \prod_{i=2}^{m-1} |x_1 - x_i|$ and $\Delta_2 = \prod_{2\le i<j\le m-1} |x_i - x_j|$. Further, let $f_j^{a,b} = x_j^a (x_j+\lambda)^b$ and $g_j = \prod_{i=j}^{m-1} (x_i+\lambda)^{-\frac{1}{2}}$. Lastly, let $d\Omega = d\mu(x_1)\ldots d\mu(x_{m-1})$ and $d\Omega_2 = d\mu(x_2)\ldots d\mu(x_{m-1})$. Below we express $w$,$w'$, and $w''$ using this notation. All the integrations are over $R_+^{m-1}$, and symmetry is used when possible.

$$\begin{aligned}
w &= \int g_1 \Delta d\Omega \\
w' &= -\frac{m-1}{2} \int f_1^{0,-\frac{3}{2}} g_2 \Delta d\Omega \\
w'' &= \frac{(m-1)(m-2)}{4} \int f_1^{0,-\frac{3}{2}} f_2^{0,-\frac{3}{2}} g_3 \Delta d\Omega + \frac{3}{4}(m-1) \int f_1^{0,-\frac{5}{2}} g_2 \Delta d\Omega.
\end{aligned}$$

Since $g_1 = (\lambda + x_1) f_1^{0,-\frac{3}{2}} g_2$, we have

$$\begin{aligned}
w &= \int x_1 f_1^{0,-\frac{3}{2}} g_2 \Delta d\Omega + \lambda \int f_1^{0,-\frac{3}{2}} g_2 \Delta d\Omega \\
&= -\frac{2\lambda}{m-1} w' + \int f_1^{1,-\frac{3}{2}} g_2 \Delta d\Omega \\
&= -\frac{2\lambda}{m-1} w' + \int f_1^{2,-\frac{3}{2}} g_2 e^{-x_1/2} \Delta dx_1 d\Omega_2 \\
&= -\frac{2\lambda}{m-1} w' - 2 \int f_1^{2,-\frac{3}{2}} g_2 \frac{d}{dx_1}\{e^{-x_1/2}\} \Delta dx_1 d\Omega_2 \\
&= -\frac{2\lambda}{m-1} w' + 2 \int \frac{d}{dx_1}\{f_1^{2,-\frac{3}{2}} \delta\} e^{-x_1/2} g_2 \Delta_2 dx_1 d\Omega_2.
\end{aligned}$$

The last line is the result of integration by parts. The differentiation gives three terms, so that

$$\begin{aligned}
w = -\frac{2\lambda}{m-1} w' + 4 \int f_1^{0,-\frac{3}{2}} g_2 \Delta d\Omega - 3 \int f_1^{1,-\frac{5}{2}} g_2 \Delta d\Omega + 2(m-2) \int \frac{x_1}{x_1 - x_2} f_1^{0,-\frac{3}{2}} g_2 \Delta d\Omega \\
= -\frac{(2\lambda + 8) w'}{m-1} - 3 \int f_1^{1,-\frac{5}{2}} g_2 \Delta d\Omega + 2(m-2) \int \frac{x_1}{x_1 - x_2} f_1^{0,-\frac{3}{2}} g_2 \Delta d\Omega. \qquad (4.5)
\end{aligned}$$

Investigating each of the above two integrals, we find

$$\int f_1^{1,-\frac{5}{2}} g_2 \Delta d\Omega = \int f_1^{0,-\frac{3}{2}} g_2 \Delta d\Omega - \lambda \int f_1^{0,-\frac{5}{2}} g_2 \Delta d\Omega, \tag{4.6}$$

and

$$\begin{aligned} \int \frac{x_1}{x_1 - x_2} f_1^{0,-\frac{3}{2}} g_2 \Delta d\Omega &= \int \frac{x_1(x_2+\lambda)}{x_1 - x_2} f_1^{0,-\frac{3}{2}} f_2^{0,-\frac{3}{2}} g_3 \Delta d\Omega \\ &= \lambda \int \frac{x_1}{x_1 - x_2} f_1^{0,-\frac{3}{2}} f_2^{0,-\frac{3}{2}} g_3 \Delta d\Omega, \end{aligned}$$

because $\frac{x_1 x_2}{x_1 - x_2}$ is anti-symmetric. We can use the identity $\frac{x_1}{x_1-x_2} + \frac{x_2}{x_2-x_1} = 1$ and symmetry to integrate this last expression. We obtain

$$\begin{aligned} \int \frac{x_1}{x_1 - x_2} f_1^{0,-\frac{3}{2}} g_2 \Delta\Omega &= \lambda \int \frac{x_1}{x_1 - x_2} f_1^{0,-\frac{3}{2}} f_2^{0,-\frac{3}{2}} g_3 \Delta\Omega \\ &= \frac{\lambda}{2} \int f_1^{0,-\frac{3}{2}} f_2^{0,-\frac{3}{2}} g_3 \Delta\Omega. \end{aligned} \tag{4.7}$$

We substitute (4.6) and (4.7) into (4.5), replacing the integrals with the expressions for $w'$ and $w''$, and finally rescale $z = \lambda/2$ to obtain equation (4.4). All we now need is $w(0)$ which has an integrand of the form (4.1) with $m$ and $n$ replaced by $m-1$ and $m+1$. Since (4.1) integrates to 1, we have $w(0) = K_{m-1,m+1}^{-1}(m-1)!$, and clearly $w(\infty) = 0$. The constant term in the pdf is then

$$\frac{K_{m,m}}{K_{m-1,m+1}} \frac{\Gamma(m/2+1)}{\Gamma(3/2)} = \frac{m}{2^{m-1/2}} \frac{\Gamma(m)}{\Gamma(m/2)} = \frac{m}{\sqrt{2\pi}} \Gamma\left(\frac{m+1}{2}\right),$$

and the theorem is proved. □

## 4.3 The pdf of $\lambda_{\min}$ for $W(m, m+1)$

As we remarked at the end of Section 3.2, if $n = m+1$ the elements of the Wishart matrix are independent aside from the constraint of positive definiteness. This leads to a particular simplicity in the smallest eigenvalue distribution.

**Theorem 4.2** *The pdf of the smallest eigenvalue of a matrix from* $W(m, m+1)$ *is* $f_{\lambda_{\min}}(\lambda) = \frac{m}{2}e^{-\lambda m/2}$*, i.e.,* $m\lambda_{\min}$ *is exponentially distributed.*

*Proof* When $n = m + 1$, the integral in (4.3) is independent of $\lambda$. In fact, other than the constant, the only term is $e^{-\lambda m/2}$. The constant must then be $m/2$ for the formula to be a pdf. When $m = 1$ this result is trivial, but it is surprising that it is true for all $m$. $\square$

## 4.4 The pdf of $\lambda_{\min}$ for $W(m,n)$ (recursion)

In the last two sections, we gave explicit formulas for the pdf of $\lambda_{\min}$ when $n-m$ is 0 or 1. In Section 4.6 we give an exact expression when $n-m$ is 2 or 3, but no exact expressions are known for other cases. Thus, we found it necessary to derive a recursion for the densities.

The following theorem is one of the principal contributions of this thesis. We state it in this section and prove it in the next.

**Theorem 4.3** *The pdf of the smallest eigenvalue of a matrix from $W(m,n)$ is $c_{m,n}\lambda^{(n-m-1)/2}e^{-\lambda m/2}g(\lambda)$, where*

$$g(\lambda) = \begin{cases} P_{m,n}(\lambda) & \text{if } n-m \text{ is odd,} \\ Q_{m,n}(\lambda)U(\lambda) + R_{m,n}(\lambda)U'(\lambda) & \text{if } n-m \text{ is even.} \end{cases}$$

*Here $U(\lambda)$ is the Tricomi function $U(\frac{m-1}{2}, -\frac{1}{2}, \frac{\lambda}{2})$, and $U'(\lambda)$ is its derivative, $-\frac{m-1}{4}U(\frac{m+1}{2}, \frac{1}{2}, \frac{\lambda}{2})$. The expressions $P_{m,n}(\lambda), Q_{m,n}(\lambda)$, and $R_{m,n}(\lambda)$ are polynomials with rational coefficients which are determined by recursions given below. The constant $c_{m,n}$ is*

$$c_{m,n} = \rho m 2^{m(m-n)/2}\frac{\Gamma((m+1)/2)}{\Gamma(1/2)}\prod_{j=1}^{n-m}\frac{\Gamma(j/2)}{\Gamma((m+j)/2)},$$

*where $\rho = 2^{m/2-1}$ if $n-m$ is odd and $2^{-1/2}$ if $n-m$ is even.*

One can verify by counting powers of $\pi$ and 2 that $c_{m,n}$ is rational when $n$ and $m$ have opposite parity. If $m$ and $n$ are even, $c_{m,n}2^{1/2}$ is rational, while if $m$ and $n$ are odd, $c_{m,n}(2\pi)^{1/2}$ is rational. The Tricomi function $U(a,b,z)$ is the confluent hypergeometric function described in Chapter 13 of [1] and Chapter 48 of [43].

The degree of $P_{m,n}$ is $\frac{1}{2}(m-1)(n-m-1)$, while the degrees of $Q_{m,n}(\lambda)$ and $R_{m,n}(\lambda)$ are at most $\frac{1}{2}(m-1)(n-m)$. Here are the recursion formulas for computing $P_{m,n}$, given $P_{m,n-2}$, or $Q_{m,n}$ and $R_{m,n}$, given $Q_{m,n-2}$ and $R_{m,n-2}$:

**Recursion for $\mathbf{P_{m,n}}, \mathbf{Q_{m,n}},$ and $\mathbf{R_{m,n}}$**

$\mathbf{n-m}$ **odd:**

$S_0 := P_{m,n-2}$

For $i = 1$ to $m-1$

$$S_i := (\lambda + n - i)S_{i-1} - \frac{2\lambda}{m-i}S'_{i-1} + \lambda(i-1)\frac{n-i-1}{m-i}S_{i-2}$$

$P_{m,n} := S_{m-1}$

$\mathbf{n-m}$ **even:**

$S_0 := \begin{pmatrix} Q_{m,n-2} \\ R_{m,n-2} \end{pmatrix}$

For $i = 1$ to $m-1$

$$S_i := (\lambda + n - i)S_{i-1} - \frac{2\lambda}{m-i}S'_{i-1} + \lambda(i-1)\frac{n-i-1}{m-i}S_{i-2}$$
$$-\frac{1}{m-i}\begin{pmatrix} 0 & \frac{m-1}{2} \\ 2\lambda & \lambda+1 \end{pmatrix} S_{i-1},$$

$\begin{pmatrix} Q_{m,n} \\ R_{m,n} \end{pmatrix} := S_{m-1}$

**Initial cases:**

$P_{m,m+1} = 1,\ Q_{m,m} = 1,$ and $R_{m,m} = 0.$

We found it very convenient to compute these polynomials (and also the moments of the distribution described in Chapter 8) using the symbolic package Mathematica [60]. The Mathematica programs appear in Appendix A.

## 4.5 Proof of Theorem 4.3

Theorem 4.3 is proved by combining the lemmas of this section. We begin by introducing some simplifying notation for the multivariate integrals that are of interest.

**Definition 4.1** *Let*

$$I_{i,j}^{\alpha} = \int_{R_+^{m-1}} f(\lambda) \Delta d\Omega,$$

*where the integrand* $f(\lambda)$ *is*

$$\underbrace{(x_1+\lambda)^{\alpha}\dots(x_i+\lambda)^{\alpha}}_{i \text{ terms}}\underbrace{(x_{i+1}+\lambda)^{\alpha-1}\dots(x_{i+j}+\lambda)^{\alpha-1}}_{j \text{ terms}}\underbrace{(x_{i+j+1}+\lambda)^{\alpha-2}\dots(x_m+\lambda)^{\alpha-2}}_{m-i-j-1 \text{ terms}}.$$

*Further define the operator*

$$I_{i,j}^{\alpha}[g] = \int_{R_+^{m-1}} f(\lambda) g \Delta d\Omega.$$

Notice that given a parameter $m$ (which we omit from the notation), $I_{i,j}^{\alpha}$ is a function of $\lambda$. We sometimes make this explicit by writing $I_{i,j}^{\alpha}(\lambda)$. The subscripts $i$ and $j$ and the implicit parameter $m-i-j-1$ give the number of terms of degree $\alpha$, $\alpha-1$, and $\alpha-2$ respectively. The total number of such terms is $(i)+(j)+(m-i-j-1) = m-1$.

In this notation, we can rewrite (4.3) as

$$f_{\lambda_{\min}}(\lambda) = k_{m,n} \lambda^{\alpha} e^{-\lambda m/2} I_{m-1,0}^{\alpha}(\lambda), \tag{4.8}$$

where $\alpha = \frac{n-m-1}{2}$ and $k_{m,n} = K_{m,n}/(m-1)!$.

It is already obvious from (4.3) or (4.8) that if $n-m$ is an odd integer, then $I_{m-1,0}^{\alpha}$ is a polynomial. This fact was observed using other methods in [30] and [35].

The integrals we are about to compute appear complicated, but the next lemma makes explicit how symmetry can be used to simplify them. Using symmetry is an old trick which was most recently popularized in this context in [4].

**Lemma 4.2** *We have*

$$I_{i,j}^{\alpha}[x_k] = \begin{cases} I_{i+1,j-1}^{\alpha} - \lambda I_{i,j}^{\alpha} & \text{if } i < k \le i+j, \\ I_{i,j+1}^{\alpha} - \lambda I_{i,j}^{\alpha} & \text{if } i+j < k < m. \end{cases} \tag{4.9}$$

*If the terms* $(x_k+\lambda)$ *and* $(x_l+\lambda)$ *have the same exponent in the integrals (i.e., both* $k$ *and* $l$ *fall within one of the closed intervals* $[1,i]$, $[i+1,i+j]$, *or* $[i+j+1,m-1]$*), then*

$$I^{\alpha}_{i,j}\left[\frac{x_kx_l}{x_k-x_l}\right] = 0, \tag{4.10}$$

$$I^{\alpha}_{i,j}\left[\frac{x_k}{x_k-x_l}\right] = \frac{1}{2}I^{\alpha}_{i,j}, \tag{4.11}$$

$$I^{\alpha}_{i,j}\left[\frac{x_k^2}{x_k-x_l}\right] = I^{\alpha}_{i,j}\left[x_k\right]. \tag{4.12}$$

*Proof* Equation (4.9) is little more than the statement $x_k=(x_k+\lambda)-\lambda$. The integral in (4.10) vanishes since the integrand is antisymmetric in $x_k$ and $x_l$. The integral in (4.11) follows from the identity $\frac{x_k}{x_k-x_l}+\frac{x_l}{x_l-x_k}=1$ and symmetry. The integral in (4.12) follows from (4.10) and the identity $\frac{x_k^2}{x_k-x_l}=x_k-\frac{x_kx_l}{x_k-x_l}$. □

In the next lemma, we give preliminary recurrence relations for computing $I^{\alpha}_{i,j}$. We say "preliminary" because we will shortly replace our first relation with a more efficient recurrence involving derivatives of the relevant polynomials.

**Lemma 4.3** *The integrals* $I^{\alpha}_{i,j}$ *satisfy the recurrence relations*

$$I^{\alpha}_{i,j} = (\lambda+2\alpha+j+2k+2)I^{\alpha}_{i-1,j+1}-\lambda(k+2(\alpha-1))I^{\alpha}_{i-1,j}+(i-1)\lambda I^{\alpha}_{i-2,j+2}, \tag{4.13}$$

$$I^{\alpha}_{0,j} = I^{\alpha-1}_{j,m-j-1}, \tag{4.14}$$

*where* $k=m-i-j-1$.

*Proof* The second equation is immediate from Definition 4.1. To prove (4.13), we begin by observing

$$I^{\alpha}_{i,j}=\lambda I^{\alpha}_{i-1,j+1}+I^{\alpha}_{i-1,j+1}[x_i],$$

by replacing $i$ with $i-1$ and $j$ with $j+1$ in (4.9). We choose the variable $x_i$ for an integration by parts in the integral $I^{\alpha}_{i-1,j+1}[x_i]$. The relevant terms for this variable have the form $-2(x_i+\lambda)^{\alpha-1}\prod_{i<l}|x_i-x_l|x_i^2\frac{d}{dx_i}e^{-x_i/2}$. Due to the factors of $x_i^2$ and $e^{-x_i/2}$, the integrand vanishes at the endpoints, and we need only compute $\frac{d}{dx_i}\left\{(x_i+\lambda)^{\alpha-1}\prod|x_i-x_l|x_i^2\right\}e^{-x_i/2}$. Since $\frac{d}{dx_i}|x_i-x_l|=$

$|x_i - x_l|/(x_i - x_l)$, we conclude

$$I^\alpha_{i,j} = \lambda I^\alpha_{i-1,j+1} + 2\left(2I^\alpha_{i-1,j+1} + (\alpha-1)I^\alpha_{i-1,j}[x_{i+j}] + \sum_{l\neq i} I^\alpha_{i-1,j+1}\left[\frac{x_i}{x_i - x_l}\right]\right),$$

$$= (\lambda + 2\alpha + 2)I^\alpha_{i-1,j+1} - 2\lambda(\alpha-1)I^\alpha_{i-1,j} + 2\sum_{l\neq i} I^\alpha_{i-1,j+1}\left[\frac{x_i}{x_i - x_l}\right]$$

obtaining the term for $I^\alpha_{i-1,j}[x_{i+j}]$ from (4.9).

Lastly, we calculate

$$2I^\alpha_{i-1,j+1}\left[\frac{x_i}{x_i - x_l}\right] = \begin{cases} \lambda I^\alpha_{i-2,j+2} & \text{if } 1 \leq l < i, \\ I^\alpha_{i-1,j+1} & \text{if } i < l \leq i+j, \\ 2I^\alpha_{i-1,j+1} - \lambda I^\alpha_{i-1,j} & \text{if } i+j < l < m. \end{cases}$$

These equations follow from Lemma 4.2. The first equation can be derived by pulling out a factor of $(x_l + \lambda)$ and using both (4.10) and (4.11). The second case is (4.11) exactly. To derive the last case, we pull out a factor of $x_i + \lambda$ to obtain $2I^\alpha_{i-1,j}\left[\frac{x^2_{i+j}}{x_{i+j}-x_l}\right] + 2\lambda I^\alpha_{i-1,j}\left[\frac{x_{i+j}}{x_{i+j}-x_l}\right]$. □

An algorithm can be based on (4.13) because it allows for reduction of the first index $i$ until we reach 0. Then (4.14) allows us to reduce $\alpha$ to $\alpha - 1$. Thus (4.13) and (4.14) give us the means to compute $I^\alpha_{i,j}$ by reducing it to the case of $I^1_{i,m-1-i}$ or $I^{\frac{1}{2}}_{i,m-1-i}$. In Lemma 4.6, we give expressions for these initial cases.

**Lemma 4.4** *If $i + j = m - 1$ (i.e., $m - i - j - 1$=0), then*

$$I^\alpha_{i,j} = (\lambda + 2\alpha + j + 2)I^\alpha_{i-1,j+1} - \frac{2\lambda}{j+1}\frac{d}{d\lambda}I^\alpha_{i-1,j+1} + \lambda(i-1)(1 + \frac{2\alpha}{j+1})I^\alpha_{i-2,j+2}.$$

*Proof* The property that $i + j = m - 1$ means (from Definition 4.1) that the integrand has exponents $\alpha$ and $\alpha - 1$, but no exponents $\alpha - 2$. Lemma 4.3 gives $I^\alpha_{i,j}$ in terms of $I^\alpha_{i-1,j+1}, I^\alpha_{i-1,j}$, and $I^\alpha_{i-2,j+2}$. Since $(i-1) + j \neq m - 1$, we need to express this term differently if we are to have a recursion with no exponents of $\alpha - 2$. We use the observation that when $m - i - j - 1 = 0$,

$$\frac{d}{d\lambda}I^\alpha_{i-1,j+1} = (i-1)\alpha I^\alpha_{i-2,j+2} + (j+1)(\alpha-1)I^\alpha_{i-1,j},$$

which comes from differentiating under the integral sign. Combining this with Lemma 4.3, the lemma is proved. $\square$

**Lemma 4.5** *Assume inductively that*

$$I^{\alpha}_{i_0,j_0} = A_i(\lambda)U(\lambda) + B_i(\lambda)U'(\lambda),$$

*for* $0 \le i_0 < i$ *and* $i_0 + j_0 = m - 1$, *where the* $A_{i_0}$ *and* $B_{i_0}$ *are polynomials, and* $U$ *and* $U'$ *are as in Theorem 4.3. If* $i + j = m - 1$, *then we have*

$$A_i = (\lambda + n - i)A_{i-1} - \frac{2\lambda}{m-i}A'_{i-1} + \lambda(i-1)\frac{n-i-1}{m-i}A_{i-2} - \frac{m-1}{2(m-i)}B_{i-1},$$

$$B_i = (\lambda + n - i)B_{i-1} - \frac{2\lambda}{m-i}B'_{i-1} + \lambda(i-1)\frac{n-i-1}{m-i}B_{i-2} - \frac{2\lambda A_{i-1} + (\lambda+1)B_{i-1}}{m-i}.$$

*Proof* This follows directly from Lemma 4.4 and the differential equation satisfied by $U$: $2\lambda U'' - (\lambda+1)U' - \frac{m-1}{2}U = 0$. (See 13.1.1 in [1]). $\square$

**Lemma 4.6** *For* $n - m$ *odd, the initial case* $\alpha = 1$ *is a multiple of a Laguerre polynomial,*

$$I^{1}_{i,m-1-i} = \frac{mi!}{2(m+1)!}L_i^{(m+2-i)}(-\lambda),$$

*and* $L_n^{(\alpha)}(-x) = \sum_{m=0}^{n}\binom{n+\alpha}{n-m}x^m$ *is the generalized Laguerre polynomial. For* $n - m$ *even, using the notation of Lemma 4.5, the initial case* $\alpha = 1/2$ *is a combination of Laguerre polynomials and Tricomi functions,*

$$\begin{aligned}
I^{\frac{1}{2}}_{i,m-1-i} &= A_iU + B_iU', \\
A_i &= \frac{\Gamma((m+1)/2)i!}{(m-1)!\sqrt{2\pi}}L_i^{(m+1-i)}(-\lambda), \\
B_i &= \frac{-2\lambda\Gamma((m+1)/2)i!}{(m-1)(m-1)!\sqrt{2\pi}}L_{i-1}^{(m+2-i)}(-\lambda),
\end{aligned}$$

*Proof* This is proved by induction. From (4.8) and Theorem 4.2,

$$I^{1}_{0,m-1} = I^{0}_{m-1,0} = mk^{-1}_{m,m+1}/2.$$

We then proceed by induction using Lemma 4.4 and two identities for Laguerre polynomials: $\lambda L_{n-2}^{(\alpha+2)}(-\lambda) + (\alpha + 1 + \lambda) L_{n-1}^{(\alpha+1)}(-\lambda) = n L_n^{(\alpha)}(-\lambda)$ and $\frac{d}{d\lambda} L_n^{(\alpha)}(-\lambda) = L_{n-1}^{(\alpha+1)}(-\lambda)$. This completes the odd case.

For $n - m$ is even, we note from (4.8) and Theorem 4.1

$$I_{0,m-1}^{1/2} = I_{m-1,0}^{-1/2} = \frac{m\Gamma(m)}{2^{m-1/2}\Gamma(m/2)} U.$$

Again, we continue by induction using Lemma 4.5 and the two identities above for Laguerre polynomials. □

## 4.6 The pdf of $\lambda_{\min}$ for $W(m, m+2)$ and $W(m, m+3)$

The pdf of $\lambda_{\min}$ when $n - m = 2$ and 3 follows from Lemma 4.6.

**Corollary 4.1** *The pdf of the smallest eigenvalue of a matrix from $W(m, m+2)$, is*

$$f_{\lambda_{\min}}(\lambda) = \frac{\Gamma(\frac{m+1}{2})}{\sqrt{2\pi}} \left[ L_{m-1}^{(2)}(-\lambda) U(\frac{m-1}{2}, -\frac{1}{2}, \frac{\lambda}{2}) + \frac{\lambda}{2} L_{m-2}^{(3)}(-\lambda) U(\frac{m+1}{2}, \frac{1}{2}, \frac{\lambda}{2}) \right].$$

*The pdf of the smallest eigenvalue of a matrix from $W(m, m+3)$, is*

$$f_{\lambda_{\min}}(\lambda) = \frac{1}{2(m+1)} L_{m-1}^{(3)}(-\lambda).$$

## 4.7 Summary

We list all the pdf's which we know exactly in Table 4.1. Other pdf's can be computed using the recursion formulas given in Section 4.4. In Appendix A, we provide Mathematica programs for such computations.

Table 4.1: The pdf of $\lambda_{\min}$ is given by $\lambda^{(n-m-1)/2}e^{-\lambda m/2}h(\lambda)$.

| $n-m$ | $h(\lambda)$ |
|---|---|
| 0 | $\frac{m}{2^{m-1/2}}\frac{\Gamma(m)}{\Gamma(m/2)}U(\frac{m-1}{2},-\frac{1}{2},\lambda/2)$ |
| 1 | $m/2$ |
| 2 | $\frac{\Gamma(\frac{m+1}{2})}{\sqrt{2\pi}}\left[L_{m-1}^{(2)}(-\lambda)U(\frac{m-1}{2},-\frac{1}{2},\frac{\lambda}{2})+\frac{\lambda}{2}L_{m-2}^{(3)}(-\lambda)U(\frac{m+1}{2},\frac{1}{2},\frac{\lambda}{2})\right]$ |
| 3 | $\frac{1}{2(m+1)}L_{m-1}^{(3)}(-\lambda)$ |

The pdf's for $m=3$ give a representative illustration of the the general case. Figure 4.1 plots the distribution functions for $m=3$ and $n=3,4,\ldots,28$. One might immediately notice that the distribution gets wider as $n$ gets larger. A special case of interest occurs when $m=n$. This case is unusual in that the pdf is infinite at 0, while for $n>m+1$ it vanishes at 0. Another unusual feature when $m=n$ is that the pdf monotonically decreasing, while for $n>m+1$ it is bell-shaped.

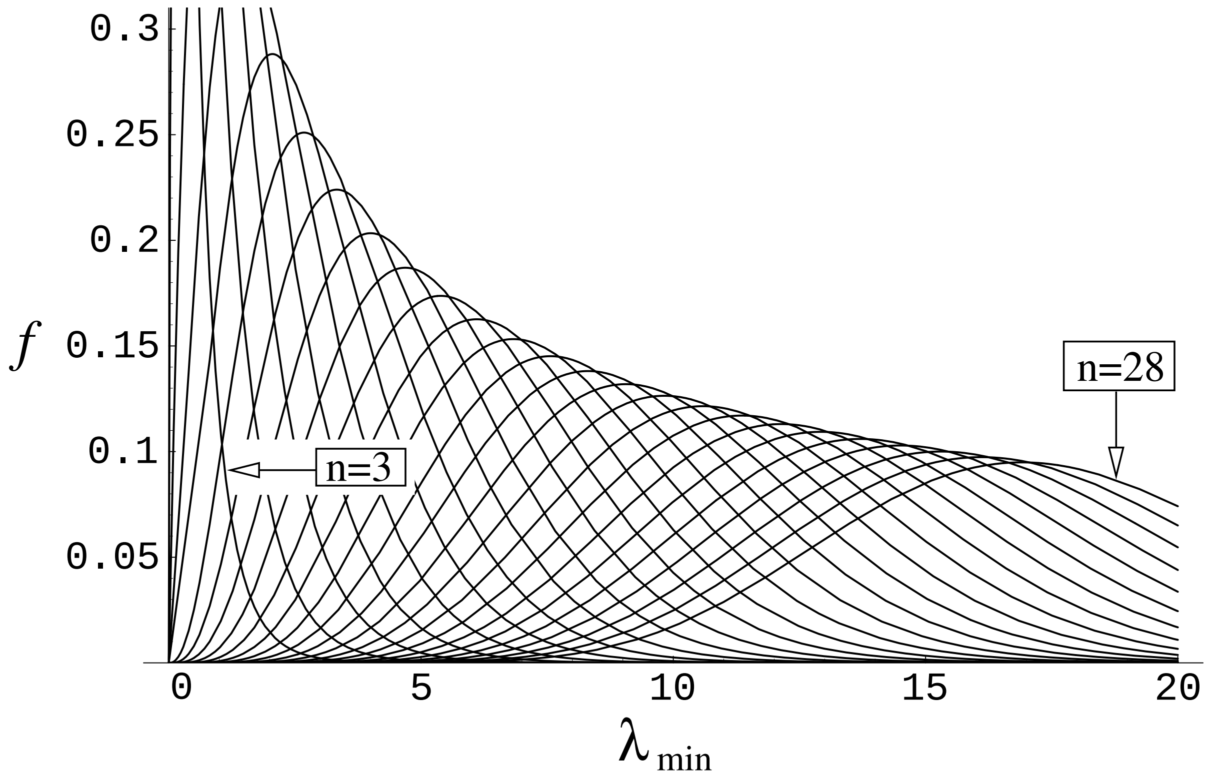


Figure 4.1: The density of $\lambda_{\min}$ for $W(3,n)$, $n = 3, 4, 5, \ldots, 28$

# Chapter 5

# Other smallest eigenvalue results

## 5.1 Asymptotic smallest eigenvalue for $W(m, n)$

In this section, we derive the limiting distribution of $m\lambda_{\min}$ when $m = n$ and $m \to \infty$. We further review a result by Silverstein [41] that gives the limiting distribution of $\lambda_{\min}/m$ when $m$ and $n$ go to infinity with a fixed ratio. The first distribution describes the behavior of the smallest singular value of a random square matrix from $G(m, m)$, while the second describes the behavior for a large rectangular random matrix from $G(m, n)$.

From Theorem 4.1, we can derive the following strong convergence result.

**Theorem 5.1** *The pdf of $m\lambda_{\min}$, where $\lambda_{\min}$ is the smallest eigenvalue of a matrix from $W(m, m)$ converges pointwise to*

$$f(x) = \frac{1+\sqrt{x}}{2\sqrt{x}} e^{-(x/2+\sqrt{x})},$$

*as $m \to \infty$.*

*Proof* From Theorem 4.1, the pdf of $x = m\lambda_{\min}$ is

$$f_m(x) = \sqrt{\frac{m}{2\pi}} \Gamma\left(\frac{m+1}{2}\right) x^{-1/2} e^{-x/2} U\left(\frac{m-1}{2}, -\frac{1}{2}, \frac{x}{2m}\right).$$

The limit of $f_m(x)$ as $m \to \infty$ follows from Stirling's formula and

$$\lim_{m\to\infty} 2\pi^{-1/2}\Gamma(\frac{m+2}{2})U(\frac{m-1}{2}, -\frac{1}{2}, \frac{x}{2m}) = (1+\sqrt{x})e^{-\sqrt{x}},$$

which is a variation of equation 13.3.3 in [1]. $\square$

The convergence given in Theorem 5.1 is very fast. Figure 5.1 plots $f_m(x)$, the density of $m\lambda_{\min}$ for $m = n$ with $m = 1, 2$, and $\infty$. Given the proximity of these curves it would be pointless to plot any of the intermediate curves! Figure 5.2 plots the density of $\sqrt{m\lambda_{\min}}$ also for $m = 1, 2$, and $\infty$. The density of $\sqrt{m\lambda_{\min}}$ has vanishing slope at the origin for all $m$.

When $m \neq n$, the smallest eigenvalue was described by Silverstein [41].

**Proposition 5.1** *If $m$ and $n$ tend to infinity in such a way that $m/n$ tends to a limit $y \in [0,1]$, then $\lambda_{\min}$ for $W(m,n)$ satisfies*[1]

$$\frac{1}{n}\lambda_{\min} \overset{a.s.}{\to} (1-\sqrt{y})^2 \ . \tag{5.1}$$

As an example, if we take $n = 2m$ we find that $\lambda_{\min}/n$ converges almost surely to $3/2-\sqrt{2} \approx 0.09$ as $n \to \infty$. Figure 5.3 plots the density of $n\lambda_{\min}$ for $W(m, 2m)$ for $m = 3, 9, 15, 21, 27$. When $m = 27$ the density is quite narrow.[2]

[1] Recall that "almost sure convergence" means that only a set of measure 0 of the sequences of growing matrices with $m/n$ converging to $y$ does not have this property.

[2] Equation (5.1) has been corrected on May 15, 2003 from the original version which read $\frac{1}{m}\lambda_{\min}$ as left hand side. We thank John Hudson of Nortel Networks for this correction.

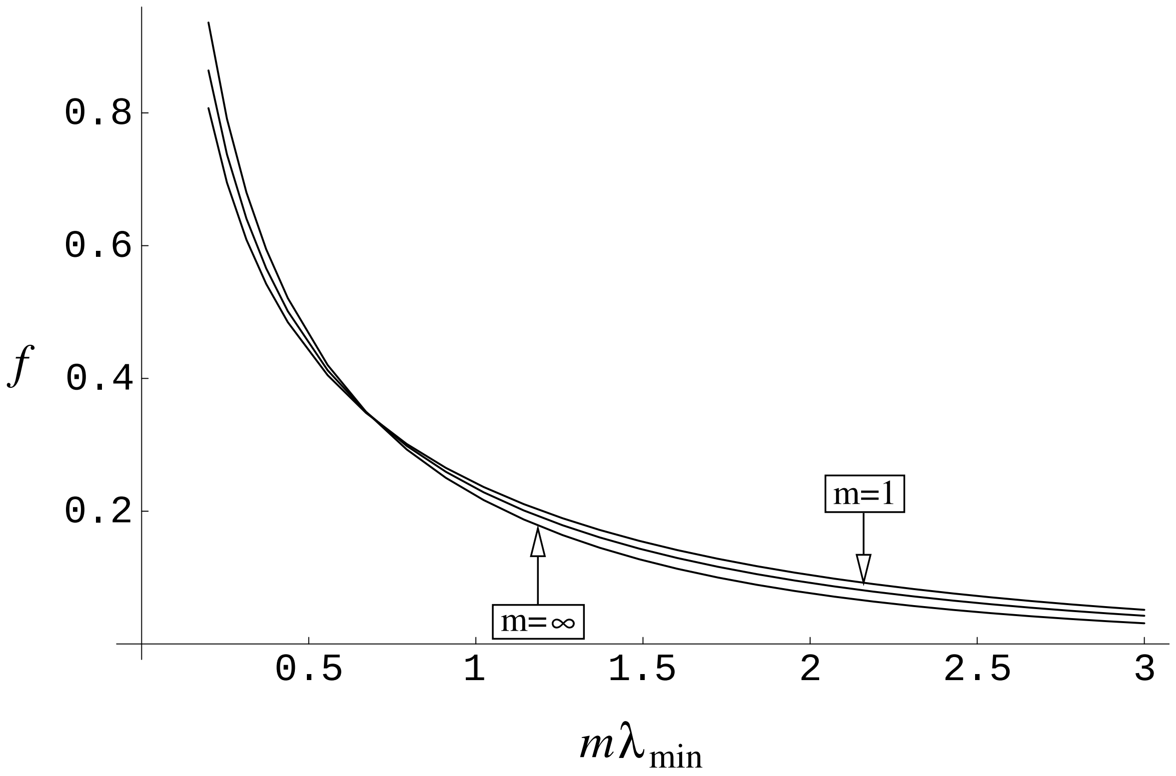


Figure 5.1: The density of $m\lambda_{\min}$ for $W(m,m)$, $m = 1, 2, \infty$

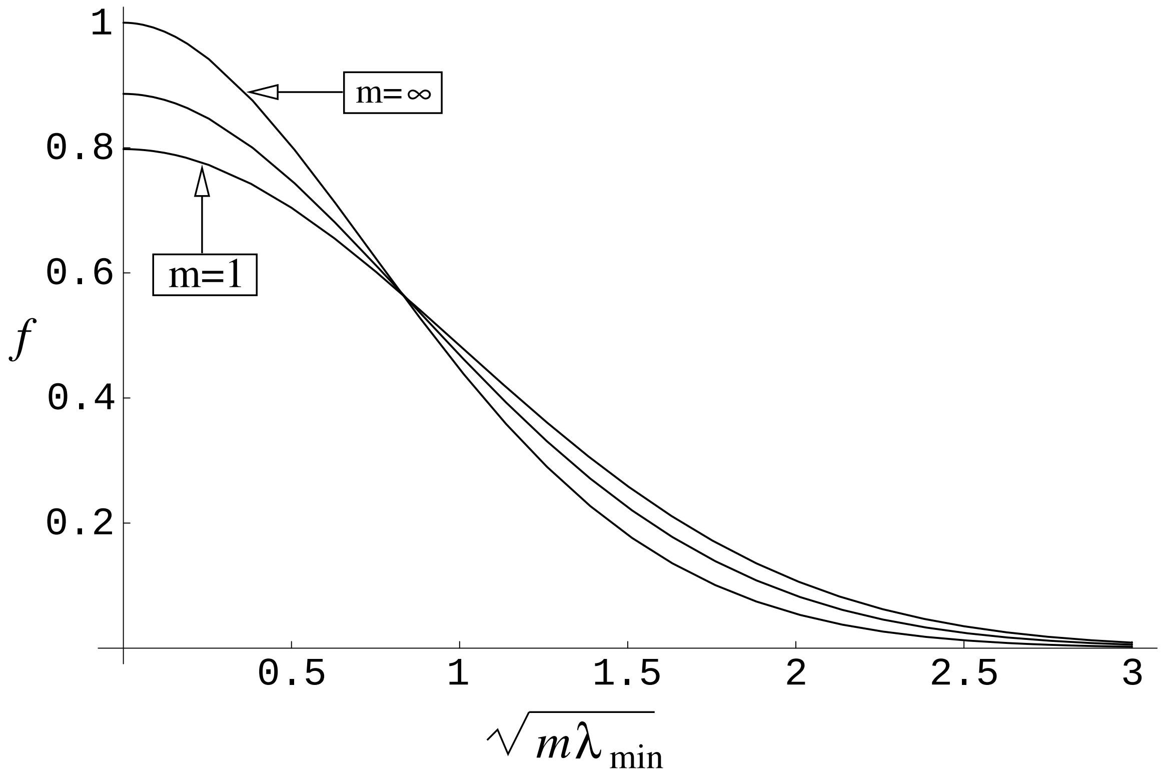


Figure 5.2: The density of $\sqrt{m\lambda_{\min}}$ for $W(m,m)$, $m = 1, 2, \infty$

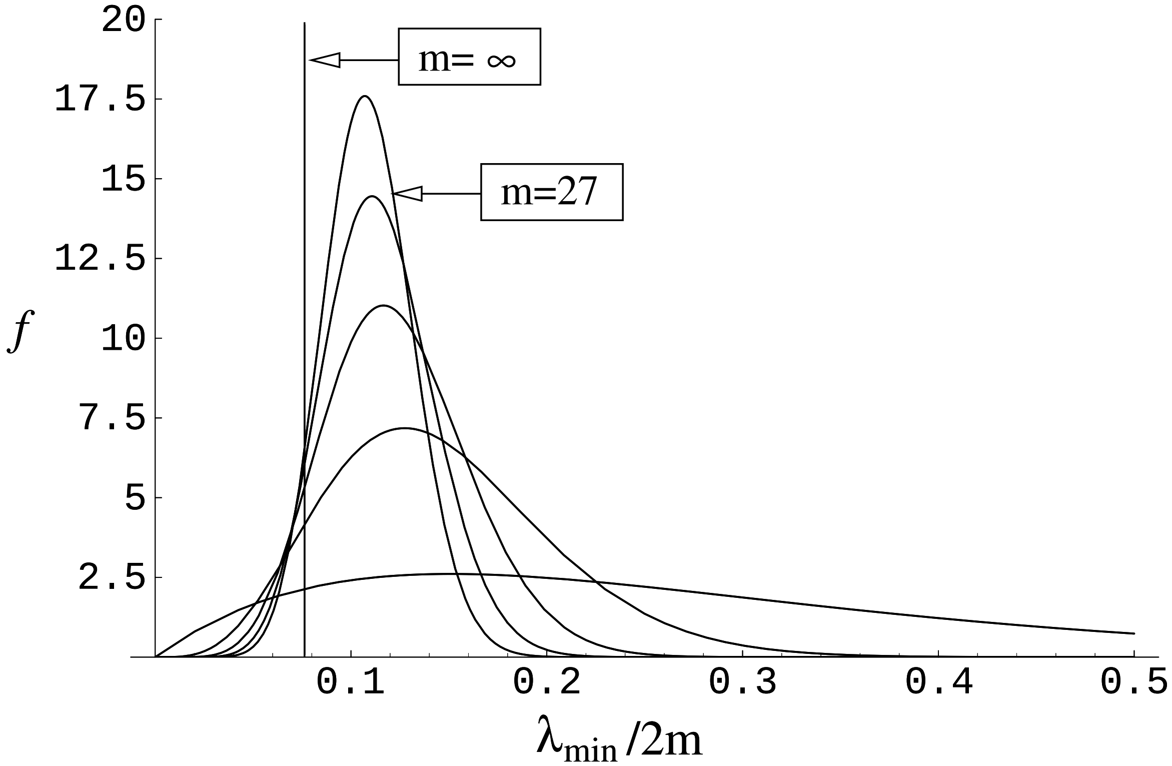


Figure 5.3: The density of $\lambda_{\min}/2m$ for $W(m, 2m)$, $m = 3, 9, 15, 21, 27$

## 5.2 Asymptotic geometric means of $\lambda_{\min}$ for $W(m,n)$

When $n = m$, inspection of the probability density given in Theorem 4.1 shows that $E(1/\lambda_{\min})$ is infinite for all $m$, though we will show in Section 8.2 that $E(\lambda_{\min}) \approx 0.69/m$ for $m$ large. In this section, we calculate $E(\log \lambda_{\min})$, whose exponential is the geometric mean of the distribution. This quantity enters into the evaluation of $E(\log \kappa)$ given in Chapter 7.

We begin with a technical result concerning the uniform integrability of the random variables $\log(\lambda_{\min})$. (See [7] p. 291 for the definition of uniform integrability.)

**Lemma 5.1** *The random variables $l_m = \log(\lambda_{\min})$ where $\lambda_{\min}$ is the smallest eigenvalue of a matrix from $W(m,m)$, m=1,2,..., are uniformly integrable.*

*Proof* The difficulties with the logarithm function occur near 0 and near $\infty$. The condition of uniform integrability then becomes, for all $\epsilon > 0$ there exists a positive number $\delta$ (which will be small) and a positive number $M$ (which will be large) such that for $m$ sufficiently large

$$\int \log x f_m(x)dx < \epsilon,$$

where the integration is over the two intervals $[0,\delta]$ and $[M,\infty]$. Here $f_m(x)$ is as in the proof of Theorem 5.1. For $x \leq 1$, $\log x f_m(x) \leq cx^{-1/2} \log x$ for $m$ sufficiently large. Since $cx^{-1/2} \log x$ is integrable, we can choose $\delta$ to make the integral from 0 to $\delta$ be smaller than $\epsilon/2$. The large $x$ behavior can be even more easily bounded. $\square$

**Theorem 5.2** *If $\lambda_{\min}$ is the smallest eigenvalue of a matrix from $W(m,m)$, then as $m \to \infty$*

$$E(\log(m\lambda_{\min})) \to -2\gamma - 2e^{1/2}\int_1^\infty \frac{e^{-1/2y^2}}{y+1}dy \approx -1.68788\ldots, \tag{5.2}$$

*where $\gamma \approx 0.5772$ is Euler's constant.*

*Proof* Using the notation of Theorem 5.1, the number we seek is

$$\lim_{m\to\infty}\int_0^\infty \log x f_m(x)dx.$$

From the previous lemma, we can interchange the integration and the limit since the functions

$\log x f_m(x)$ are uniformly integrable. We then obtain

$$\int_0^\infty \log x \frac{1+\sqrt{x}}{2\sqrt{x}} e^{-(x/2+\sqrt{x})} dx.$$

This integral can be manipulated into the integral given in (5.2) using equation 4.331.1 in [23], but we know of no simpler form. In this form, however, numerical integration is trivial.

We now turn to rectangular matrices. In the last section, we mentioned that if $m$ and $n$ tend to infinity in such a way that $m/n$ tends to a limit $y$, then $\frac{1}{m}\lambda_{\min} \overset{a.s.}{\to} (1-\sqrt{y})^2$. Thus we expect

**Theorem 5.3** *If $\lambda_{\min}$ is the smallest eigenvalue of a matrix satisfying the conditions of Proposition 5.1, then*

$$E(\log \lambda_{\min}/n) = \log(1-\sqrt{y})^2 + o(1).$$

*Proof* From Proposition 5.1, it is sufficient to verify that the variables $\lambda_{\min}/n$ are uniformly integrable. Let $\sigma = \lambda_{\min}/n$. All we must check is that for some $\epsilon$ and $r$ depending on $y$, $\int_0^\epsilon \log x f_\sigma(x)dx$ and $\int_r^\infty \log x f_\sigma(x)dx$ vanish as $m, n \to \infty$.

First we bound $f_{\lambda_{\min}}(\lambda)$.

$$\begin{aligned}
f_{\lambda_{\min}}(\lambda) &= K_{m,n}\lambda^{\frac{1}{2}(n-m-1)}e^{-\frac{\lambda}{2}}\int_{R_\lambda} e^{-\sum_{i=1}^{m-1}\frac{\lambda_i}{2}} \prod_{i<j}(\lambda_i-\lambda_j)\prod_{i=1}^{m-1}(\lambda_i-\lambda)\lambda_i^{\frac{1}{2}(n-m-1)}d\lambda_i \\
&\le K_{m,n}\lambda^{\frac{1}{2}(n-m-1)}e^{-\frac{\lambda}{2}}\int_{R_0} e^{-\sum_{i=1}^{m-1}\frac{\lambda_i}{2}} \prod_{i<j}(\lambda_i-\lambda_j)\prod_{i=1}^{m-1}\lambda_i^{\frac{1}{2}(n-m+1)}d\lambda_i \\
&= \frac{K_{m,n}}{K_{m-1,n+1}}\lambda^{\frac{1}{2}(n-m-1)}e^{-\lambda/2},
\end{aligned}$$

and from (4.2),

$$\frac{K_{m,n}}{K_{m-1,n+1}} = \frac{\pi^{\frac{1}{2}}2^{-\frac{1}{2}(n-m+1)}\Gamma(\frac{n+1}{2})}{\Gamma(m/2)\Gamma(\frac{n-m+1}{2})\Gamma(\frac{n-m+2}{2})}.$$

We then have, for $\epsilon < 1-y$,

$$0 \ge \int_0^\epsilon \log x f_\sigma(x)dx \ge \frac{K_{n,m}}{K_{n+1,m-1}} n^{\frac{1}{2}(n-m+1)}\int_0^\epsilon \log x \; x^{\frac{1}{2}(n-m-1)}e^{-nx/2}dx$$

$$\approx \left( \left( \frac{e\epsilon}{n(1-y)^2} \right)^{1-y} e^{-\epsilon} \right)^{n/2}.$$

Estimating the tail is much easier. Let $\tau$ have the distribution $\chi^2_{n+m-1}/n$. According to (3.14), we can define a probability space in which $\sigma \leq \tau$, and it is a straightforward asymptotic analysis to show that the tail vanishes as $n \to \infty$.

## 5.3 Smallest eigenvalue results for $\tilde{W}(m,n)$

Many of the results that were derived for $W(m,n)$ have straightforward analogues for $\tilde{W}(m,n)$ which we list here.

The joint distribution of the eigenvalues of $\tilde{W}(m,n)$ was given in 3.10. We rewrite the distribution in the convenient form

$$\tilde{K}_{m,n} e^{-\frac{1}{2}\sum_{i=1}^{m}\lambda_i} \prod_{i=1}^{m} \lambda_i^{n-m} \prod_{i<j} (\lambda_i - \lambda_j)^2 d\lambda_1 \ldots d\lambda_m, \tag{5.3}$$

where

$$\tilde{K}_{m,n}^{-1} = 2^{mn} \prod_{i=1}^{m} \Gamma(n-i+1)\Gamma(m-i+1) \ . \tag{5.4}$$

With an appropriate transformation and symmetry we obtain that the pdf of $\lambda_{\min}$ is

$$\frac{\tilde{K}_{m,n}}{(m-1)!} \lambda^{n-m} e^{-\lambda m/2} \int_{R_+^{m-1}} \prod_{i=1}^{m-1} (x_i + \lambda)^{n-m} \Delta^2 d\Omega, \tag{5.5}$$

where $\Delta = \prod_{1 \le i < j \le m-1} (x_i - x_j)$, $d\Omega = d\mu(x_1) \ldots d\mu(x_{n-1})$, $d\mu(x) = x^2 e^{-x/2}$, and the integration takes place over the positive orthant, $R_+^{m-1}$. Compare (4.3).

The variable $\lambda$ only appears within a polynomial factor in the integral in (5.5). Thus the complex version of Theorem 4.3 is in fact simpler than the real version.

**Theorem 5.4** *The pdf of the smallest eigenvalue of a matrix from $\tilde{W}(m,n)$ is $c_{m,n}\lambda^{n-m}e^{-\lambda m/2}P_{m,n}(\lambda)$, where $c_{m,n}$ is a constant and $P_{m,n}(\lambda)$ is a polynomial of degree $(n-m)(m-1)$.*

We have not worked out the recursion that these polynomials satisfy, but the initial case is trivial.

**Theorem 5.5** *The distribution of the smallest eigenvalue of a matrix from $W(m,m)$ is $f_{\lambda_{\min}}(\lambda) = \frac{m}{2}e^{-\lambda m/2}$, i.e., $m\lambda_{\min}$ is exponentially distributed.*

*Proof* The proof is exactly the same as that of Theorem 4.2. It is perhaps surprising that the case $m = n$ is so much simpler for random complex matrices than it is for real matrices. From this formula we immediately obtain

**Corollary 5.1** *The expected logarithm of the smallest eigenvalue of a matrix from* $\tilde{W}(m,m)$ *is* $E(\log(m\lambda_{\min})) = \log 2 - \gamma = 0.11593\ldots$.

*Proof* The appropriate integral is Equation 4.352 in [23].

We now turn to the case of large rectangular complex matrices. The result of Silverstein (Proposition 5.1) has a complex version:

**Proposition 5.2** *If* $m$ *and* $n$ *tend to infinity in such a way that* $m/n$ *tends to a limit* $y \in [0,1]$, *then* $\lambda_{\min}$ *for* $\tilde{W}(m,n)$ *satisfies*

$$\frac{1}{m}\lambda_{\min} \overset{a.s.}{\to} 2(1-\sqrt{y})^2 \ . \tag{5.6}$$

As in the real case, the expected logarithm does not misbehave, so we conclude:

**Theorem 5.6** *If* $A$ *is a matrix satisfying the conditions of Proposition 5.2, then*

$$E(\log \lambda_{\min}/n) = \log(1-\sqrt{y})^2 + o(1).$$

# Chapter 6

# Largest eigenvalue results

## 6.1 Asymptotic largest eigenvalue for $W(m,n)$

The largest eigenvalue of a Wishart matrix (or the largest singular value of a Gaussian matrix) is much easier to describe than the smallest eigenvalue. In particular, it is no longer necessary to distinguish between the cases $m = n$ and $m \neq n$. As we did in Proposition 5.1, we wish to take large matrices $W(m,n)$ for which $m/n$ tends to a limit $y \in [0,1]$. We cover the case $m = n$ by taking $y = 1$. We have another result from Silverstein [41],

**Proposition 6.1** *If $m$ and $n$ tend to infinity in such a way that $m/n$ tends to a limit $y \in [0,1]$, then $\lambda_{\max}$ for $W(m,n)$ satisfies*

$$\frac{1}{n}\lambda_{\max} \overset{a.s.}{\to} (1+\sqrt{y})^2 \ . \tag{6.1}$$

It is interesting to check Proposition 6.1 experimentally. Taking $y = 1$ ($m = n$, the proposition states that $\frac{1}{m}\lambda_{\max}$ converges in probability to 4. With $m = 100$, we computed $\lambda_{\max}/m$ for 1000 matrices. Figure 6.1 plots the empirical distribution function, which is quite close to a step function with step at 4.

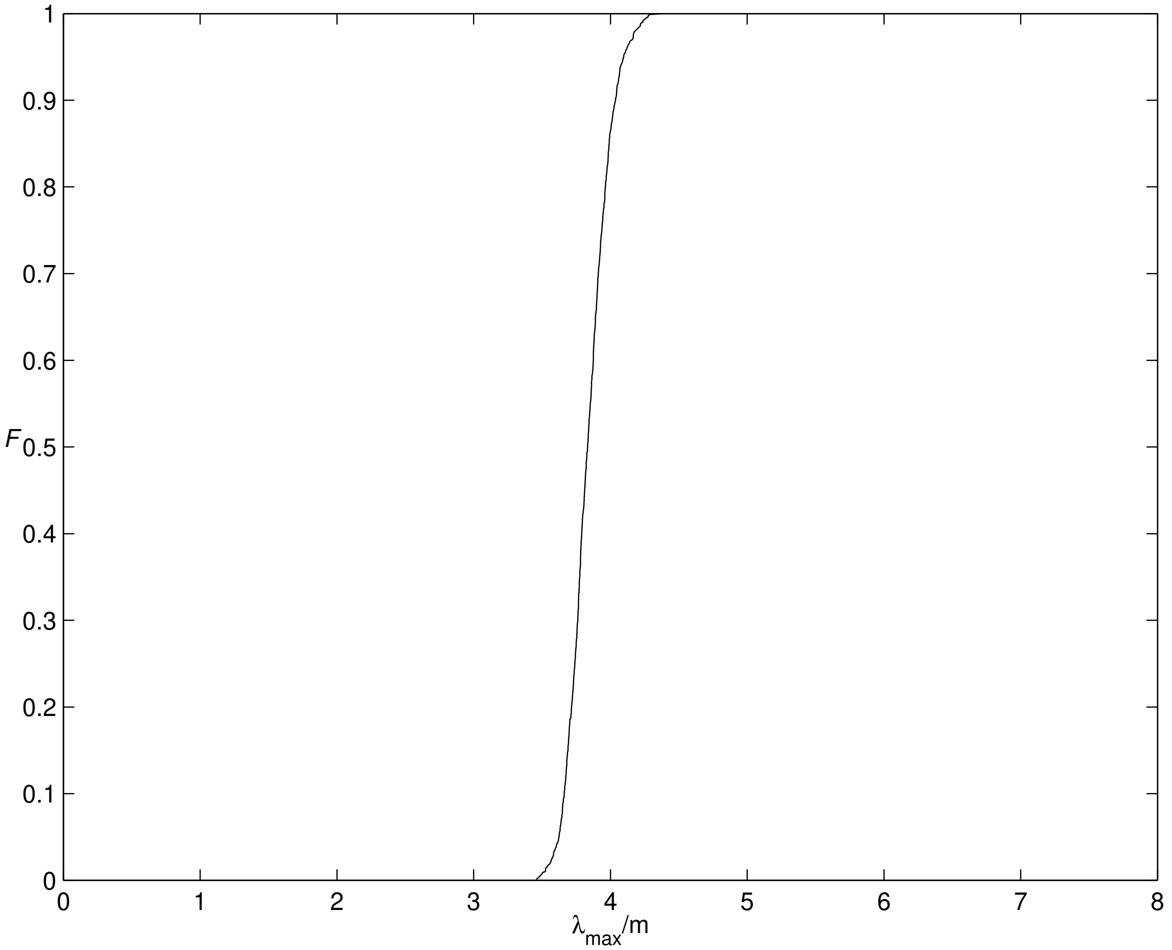


Figure 6.1: The empirical cdf of $\lambda_{\max}/m$ for $W(m, m)$, $m = 100$

## 6.2 Asymptotic geometric means of $\lambda_{\max}$ for $W(m,n)$

The next theorem gives the geometric mean of $\lambda_{\max}$ in an analogous manner to the way Theorem 5.2 obtains the geometric mean of $\lambda_{\min}$. This is, of course, twice the geometric mean of the largest singular value of the corresponding Gaussian matrix.

**Theorem 6.1** *If $m$ and $n$ tend to infinity in such a way that $m/n$ tends to a limit $y \in [0,1]$, then*

$$E(\log \lambda_{\max}/n) = \log(1+\sqrt{y})^2 + o(1).$$

*Proof* First we bound the pdf of max, $f_{\lambda_{\max}}(\lambda)$:

$$f_{\lambda_{\max}}(x) \leq \frac{K_{m,n}}{K_{m-1,n-1}} x^{\frac{1}{2}(n+m-3)} e^{-x/2} = \frac{\pi^{1/2} 2^{\frac{1-n-m}{2}}}{\Gamma(n/2)\Gamma(m/2)} x^{\frac{1}{2}(n+m-3)} e^{-x/2}. \tag{6.2}$$

This bound was derived for $m = n$ in [49] by manipulating (4.1). The same techniques work more generally.

Now let $\sigma$ denote $\lambda_{\max}/n$, and let $f_\sigma(x), F_\sigma(x)$ be the corresponding probability density function and cumulative density function. We break up $E(\log \sigma) = \int_0^\infty \log x f_\sigma(x) dx$ into three integrals:

$$\int_0^\epsilon + \int_\epsilon^r + \int_r^\infty$$

for values of $\epsilon$ and $r$ depending on $y$, but not $n$. By Lemma 4.1, the middle integral approaches $\log(1+\sqrt{y})^2$, and we proceed to show that the other integrals vanish in the limit.

**Step 1:**$\int_0^\epsilon$

Let $\tau$ be the random variable defined by $\frac{1}{n}(x_n^2 + y_{m-1}^2)$. Considering the first column of (3.14) we have $\|M\| = \|X\|^2 = \lambda_{\max} \geq x_n^2 + y_{m-1}^2$, i.e. $\sigma \geq \tau$. It follows that $F_\sigma(x) \leq F_\tau(x)$. Integrating by parts,

$$0 \geq \int_0^1 \log x f_\sigma(x) dx = -\int_0^1 \frac{F_\sigma(x)}{x} dx \geq -\int_0^1 \frac{F_\tau(x)}{x} dx = \int_0^1 \log x f_\tau(x) dx \ .$$

The terms $\log x F_\tau(x)$ and $\log x F_\sigma(x)$ produced by the integration by parts vanish as $x \to 0$. The former can be verified by using the fact that $\tau$ has the distribution $n^{-1}\chi^2_{n+m-1}$, and the latter follows from the former.

To complete the argument we take $m = m_n$, and let $k = n + m_n - 1$, so that $\tau$ has the distribution $\chi_k^2/n$, and $f_\tau(x) = \frac{(n/2)^{k/2}}{\Gamma(k/2)} x^{k/2-1} e^{-nx/2}$. Then,

$$0 \geq \int_0^\epsilon \log x f_\tau(x) dx \geq \frac{(n/2)^{k/2}}{\Gamma(k/2)} \int_0^\epsilon (\log x) x^{k/2-1} \approx (\frac{e\epsilon}{1+y})^{k/2}.$$

Here the $\approx$ indicates that only the exponential behavior is kept as $n \to \infty$. (Computing the asymptotics of this integral is routine but not obvious. A good reference is [6], chapter 6.) By choosing any $\epsilon < (1+y)/e$, we have the desired result.

**Step 2:** $\int_r^\infty$

For the singularity of the logarithm at $\infty$ we use Lemma 4.2, the fact that $f_\sigma(x) = n f_{\lambda_{\max}}(nx)$, and a standard asymptotic analysis.

For $r > 1 + y$,

$$\begin{aligned} \int_r^\infty f_\sigma(x) \log x dx \leq \int_r^\infty x f_\sigma(x) dx &= \int_{rn/2}^\infty (4x/n) f_{\lambda_{\max}}(2x) dx \\ &\leq \frac{\frac{2}{n}\pi^{1/2}}{\Gamma(n/2)\Gamma(m_n/2)} \int_{rn/2}^\infty x^{\frac{n+m_n-1}{2}} e^{-x} dx \\ &\approx (e^{-r}(er)^{1+y} y^{-y})^{n/2}. \end{aligned}$$

Here again, $\approx$ indicates that only the exponential behavior is kept as $n \to \infty$. By taking $r$ (depending on $y$) sufficiently large, we conclude step 2. $\square$

## 6.3 Largest eigenvalue results for $\tilde{W}(m,n)$

In Section 5.3, we obtained the analogues of the smallest eigenvalue results for complex matrices. Here we do the same for the largest eigenvalue results. Unlike the smallest eigenvalue case, there are no surprises here, and we omit the proofs.

**Proposition 6.2** *If $m$ and $n$ tend to infinity in such a way that $m/n$ tends to a limit $y$, then $\lambda_{\max}$ for $\tilde{W}(m,n)$ satisfies*

$$\frac{1}{m}\lambda_{\max} \stackrel{a.s.}{\to} 2(1+\sqrt{y})^2 \ . \tag{6.3}$$

**Theorem 6.2** *With the assumptions of Proposition 6.2,*

$$E(\log \lambda_{\max}/n) = \log 2(1+\sqrt{y})^2 + o(1).$$

# Chapter 7

# Condition numbers

## 7.1 Limiting condition number distributions, as $n \to \infty$

With the theory of the smallest and largest eigenvalue of Wishart matrices now developed, we can describe the limiting condition number distribution of matrices from $G(n,n)$ and $\tilde{G}(n,n)$. We consider the following theorem to be one of the principal contributions of this thesis:

**Theorem 7.1** *If $\kappa$ is the condition number of a matrix from $G(n,n)$, then the pdf of $\kappa/n$ converges pointwise to the pdf*

$$f(x) = \frac{2x+4}{x^3} e^{-2/x-2/x^2}$$

*as $n \to \infty$. Moreover,*

$$E(\log \kappa) = \log n + c + o(1)$$

*as $n \to \infty$, where $c \approx 1.537$.*

*Proof* From Proposition 6.1, we know $\frac{1}{n}\lambda_{\max} \overset{a.s.}{\to} 4$, and Theorem 5.1 gives the limiting distribution for $n\lambda_{\min}$. The distribution of the ratio of these quantities, $\kappa^2/n^2$, converges by a standard probability argument. The appropriate change of variables gives the limiting pdf of $\kappa/n$. The expected logarithm result follows from Theorem 5.2 and Theorem 6.1. Taking the exponential, we obtain $\exp(\log(\kappa)) = 4.65n + o(n)$.

**Corollary 7.1** *If $\kappa$ is the condition number of a matrix from $G(n,n)$, then*

$$\lim_{n\to\infty} P(\kappa/n < x) = e^{-2/x-2/x^2}.$$

Figure 2.1 plots the pdf given in Theorem 7.1. Figure 2.5.a is a histogram based on MATLAB computations of the condition numbers of 25,000 matrices with iid elements from a standard normal distribution with $n = 75$. The data support the claim that the formulas in Theorem 7.1 and Corollary 7.1 work very well for reasonably large values of $n$. Figure 2.5.b and Figure 2.5.c are histograms also based on the condition numbers of 25,000 matrices with $n = 75$ and iid elements from the uniform distribution $[-1,1]$ and the discrete distribution $\{-1,1\}$, respectively. The data suggest that Theorem 7.1 is robust under changes of distribution with mean 0. We have a partial proof that we hope to complete that this is so, but for now, we will state this as conjecture. We will also state as conjecture the distribution of the condition number when the mean is not 0.

**Conjecture 7.1** *If $\kappa$ is the condition number of a matrix with iid elements that have mean* 0 *and finite variance, then the pdf of $\kappa/n$ converges pointwise to the pdf given in Theorem 7.1. Furthermore, if $\kappa$ is the condition number of a matrix with iid elements that have mean $\mu$ and standard deviation $\sigma$, then the pdf of $\frac{2\sigma}{\mu}\kappa/n^{3/2}$ also converges pointwise to the pdf given in Theorem 7.1.*

The conjecture can be understood intuitively as follows. The first statement says that for large $n$, any distribution with mean 0 behaves like a normal distribution with mean 0. This is related to the central limit theorem. The second statement is based on the fact that the largest singular value of a random matrix with elements that have mean $\mu$ is $\mu n$, while the smallest singular value appears to only depend on the standard deviation, not the mean.

Figure 2.5.d illustrates what happens when we take the condition number of a matrix with iid elements, but a non-zero mean. Here the histogram is based on 25,000 samples of the random variable $\frac{2}{\sqrt{3}}\kappa/n^{3/2}$, where the matrices have iid elements from the uniform distribution $[0,1]$ and $n = 75$.

We turn now to the case of complex matrices.

**Theorem 7.2** *If $\kappa$ is the condition number of a matrix from $\tilde{G}(n,n)$, then the pdf of $\kappa/n$ converges pointwise to the pdf*

$$f(x) = \frac{8}{x^3} e^{-4/x^2}$$

*as $n \to \infty$. Moreover,*

$$E(\log \kappa) = \log n + \frac{1}{2}\gamma + \log 2 + o(1) \approx \log n + 0.982$$

*as $n \to \infty$.*

*Proof* The pdf follows from Theorems 5.5 and 6.2, while the expected logarithm follows from Theorems 5.6 and 6.2.

**Corollary 7.2** *If $\kappa$ is the condition number of a matrix from $\tilde{G}(n,n)$, then as $n \to \infty$*

$$P(\kappa/n < x) = e^{-4/x^2}.$$

The rectangular matrix result is

**Theorem 7.3** *If $\kappa$ is the condition number of a matrix from $G(m,n)$ or $\tilde{G}(m,n)$, where $m$ depends on $n$ in such a way that $\lim_{k\to\infty} m/n = y \in [0,1]$, then $\kappa$ converges almost surely to $\frac{1+\sqrt{y}}{1-\sqrt{y}}$ as $n \to \infty$. Moreover,*

$$E(\log \kappa) = \log \frac{1+\sqrt{y}}{1-\sqrt{y}} + o(1)$$

*as $n \to \infty$.*

The convergence follows trivially from Propositions 5.1, 5.2, 6.1, and 6.2. The expected logarithm of the condition number follows from Theorems 5.3 and 5.6 for $\lambda_{\min}$ and Theorems 6.1 and 6.2 for $\lambda_{\max}$.

## 7.2 Exact expressions for $m = 2$

It is possible to integrate (4.1) and (5.3) against the condition number to get the exact distributions of the condition numbers of real and complex $2 \times n$ matrices. We spare the reader the details and just give the results.

The pdf of the condition number of matrices from $G(2, n)$ is

$$f(x) = (n-1)2^{n-1} \frac{x^2 - 1}{(x^2+1)^n} x^{n-2} . \tag{7.1}$$

When the matrices are from $\tilde{G}(2, n)$, the pdf of the condition number is

$$f(x) = 2 \frac{\Gamma(2n)}{\Gamma(n)\Gamma(n-1)} \frac{x^{2n-3}(x^2-1)^2}{(x^2+1)^{2n}}. \tag{7.2}$$

We can use (7.1) and (7.2) to evaluate the integrals giving the expected condition numbers. The result is

**Theorem 7.4** *If $\kappa$ is the condition number of a matrix from $G(2, n)$, then*

$$E(\log \kappa) = \frac{1}{2}\sqrt{\pi} \frac{\Gamma(\frac{n-1}{2})}{\Gamma(n/2)}.$$

*If the matrix is from $\tilde{G}(2, n)$, then*

$$E(\log \kappa) = \log 2 + \frac{1}{2} - \sum_{k=2}^{n-1} \frac{1}{4^k} \binom{2k}{k} \frac{1}{k-1} .$$

We can also obtain the exact distribution for the smaller and the larger eigenvalues:

**Theorem 7.5** *If $\lambda_{\min}$ and $\lambda_{\max}$ are the extreme eigenvalues of a matrix from $W(2, n)$ and $\beta$ denotes $(n-1)/2$, then the pdf of $\lambda_{\min}$ is*

$$f_{\lambda_{\min}}(\lambda) = K_{n,2} e^{-\lambda} \left( 2\lambda^{\beta} e^{-\lambda/2} + 2^{\beta}(2\beta - \lambda)\Gamma(\beta, \lambda/2) \right)$$

*and the pdf of $\lambda_{\max}$ is*

$$f_{\lambda_{\max}}(\lambda) = K_{n,2} e^{-\lambda/2} \lambda^{\beta-1} \left( 2\lambda^{\beta} e^{-\lambda/2} - 2^{\beta}(2\beta - \lambda)\gamma(\beta, \lambda/2) \right) .$$

A similar result for $\tilde{W}(2, n)$ could be calculated.

## 7.3 Demmel's condition number

As mentioned in Section 2.1, Demmel [15] has been interested in the distribution of $\kappa_D = \|A\|_F \|A^{-1}\|_2$ when $A$ is $G(n,n)$ or $\tilde{G}(n,n)$. As mentioned in Section 2.2, this condition number has been used for a sphericity test in multivarite statistics. Demmel observed that the distributions of these random matrices, when considered as elements of $\mathbf{R}^{n^2}$ and $\mathbf{R}^{2n^2}$, spherically symmetric as are the condition numbers $\kappa_D$. Furthermore, the condition number is scale invariant, so there in no loss in generality in assuming that $A$ is distributed uniformly on the unit sphere $\|A\|_F = 1$. It remains to understand the distribution of $\|A^{-1}\|_2$, or equivalently, the distribution of the smallest singular value of $A$.

We use the same theory that we used to obtain the distribution of the smallest eigenvalue of Wishart Matrices. To make this concrete, let $W_D$ be the random matrix $AA^T$, where $A$ is a real matrix uniformly distributed on the unit sphere in $\mathbf{R}^{n^2}$. Let $\tilde{W}_D$ be $AA^H$, where $A$ is a complex matrix which can be thought of as uniformly distributed on the unit sphere in $\mathbf{R}^{2n^2}$.

**Lemma 7.1** *The joint density of the elements of a matrix from $W_D$ is proportional to*

$$\delta(\mathrm{tr}(W) - 1)(\det W)^{-1/2},$$

*while for $\tilde{W}_D$ it is proportional to*

$$\delta(\mathrm{tr}(\tilde{W}) - 1),$$

*where $\delta$ is the Dirac delta function.*

*Proof* Since $A$ is uniformly distributed on the sphere, its joint density is proportional to $\delta(\mathrm{tr}(AA^T) - 1)$ in the real case and $\delta(\mathrm{tr}(AA^H) - 1)$ in the complex case. We then perform the same two-step calculation indicated in the diagram before Lemma 3.2 and obtain the indicated results. In fact, it would be no more difficult to consider $m$ by $n$ matrices whose $mn$ elements are uniformly distributed on the unit sphere. □

From this lemma and Theorem 3.2, we conclude

**Proposition 7.1** *The joint density of the $n$ eigenvalues of $W$, $\lambda_1 \geq \ldots \geq \lambda_n$, $\sum_{i=1}^n \lambda_i = 1$, is proportional to $\prod_{i=1}^n \lambda_i^{-1/2} \prod_{i<j}(\lambda_i - \lambda_j)$ while the joint density of the eigenvalues of $\tilde{W}$ is proportional to $\prod_{i<j}(\lambda_i - \lambda_j)^2$.*

The above result for the real case is due to Bartlett (1951) and refered to in [40], where the integrals are evaluated in terms of two dimensional integrals. We will use an alternate method to obtain an expression for the pdf of $\lambda_{\min}$. Notice that $\lambda_{\min}$ cannot exceed $1/n$ since the sum of the $n$ eigenvalues is 1.

**Lemma 7.2** *The pdf of the smallest eigenvalue $\lambda_{\min}$ of $W$ is proportional to*

$$\lambda^{-1/2} \int_{S_\lambda^{n-1}} \Delta \prod_{i=1}^{n-1} \{(x_i + \lambda)^{-1/2} x_i\} ds, \tag{7.3}$$

*where $\Delta = \prod_{1 \le i < j \le n-1} |x_i - x_j|$, the integration takes place over the simplex $S_\lambda^{n-1} = \{(x_1, \ldots, x_{n-1}) : x_i \ge 0, \sum_{i=1}^{n-1} x_i = 1 - n\lambda\}$, and $ds$ is the volume element on this simplex. Similarly, the pdf of the smallest eigenvalue $\lambda_{\min}$ of $W$ is proportional to*

$$\int_{S_\lambda^{n-1}} \Delta^2 \prod_{i=1}^{n-1} x_i^2 ds. \tag{7.4}$$

We do not know how to perform the integral to obtain a simpler expression for the distribution for $W$, but for $\tilde{W}$, the integral is relatively easy:

**Theorem 7.6** *The pdf of the smallest eigenvalue of a complex matrix defined by $AA^H$, where $A$ is uniformly distributed on the sphere $\|A\|_F = 1$, is*

$$f_{\lambda_{\min}}(\lambda) = (n^3 - n)(1 - nx)^{n^2-2}.$$

*Proof* We make the change of variables $y_i = (1 - n\lambda)x_i$ and discover that the integral (7.4) becomes $(1 - n\lambda)^{n^2-2}$ times a quantity that is independent of $\lambda$. The $n^2 - 2$ factors of $(1 - n\lambda)$ have the following origins: $n^2 - 3n + 2$ from $\Delta^2$, $2n - 2$ from $\prod x_i^2$, and $n - 2$ from the volume element. Thus the pdf is a constant times $(1 - nx)^{n^2-2}$. Lastly, the constant is chosen so that the integral on $[0, 1/n]$ is unity. $\square$

**Corollary 7.3** *If $\lambda_{\min}$ is the smallest eigenvalue of a complex matrix defined by $AA^H$ where $A$ is uniformly distributed on the sphere $\|A\|_F = 1$, then*

$$P(\lambda_{\min} \ge x) = (1 - nx)^{n^2-1}, \quad 0 < x < 1/n.$$

We then obtain an exact formula for the distribution that Demmel estimated,

**Theorem 7.7** *If $\kappa_D$ is Demmel's condition number of a complex matrix $A$ that is uniformly distributed on the sphere $\|A\|_F = 1$, then*

$$P(\kappa_D \geq x) = 1 - (1 - n/x^2)^{n^2-1}, \quad x > \sqrt{n}.$$

*Proof* This follows from the previous corollary since $\kappa_D^2 = 1/\lambda_{\min}$. $\square$

## 7.4 The tails of the condition number distributions

In the previous sections, we described the behavior of the condition numbers but said little about the probability that a matrix with a large condition number may appear. Here we will approximate the condition numbers for square matrices in order to get a sense of the tails of the distributions.

There are four random condition numbers that we find interesting. Let $\kappa$ and $\tilde{\kappa}$ denote the 2-norm condition numbers of matrices from $G(n,n)$ and $\tilde{G}(n,n)$ respectively. Since we are only considering $n \times n$ matrices we omit the dependence on $n$ in the notation. The other two condition numbers are the Demmel condition numbers defined in Section 7.3. Let $\kappa_D$ and $\tilde{\kappa}_D$ denote the Demmel condition number in the real and complex cases as above. We chart the condition numbers and relate them to the eigenvalues of the corresponding Wishart matrix in the table below.

| $G(n,n)$ | $\tilde{G}(n,n)$ |
|---|---|
| $\kappa = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}}$ | $\tilde{\kappa} = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}}$ |
| $\kappa_D = \sqrt{\frac{\sum \lambda_i}{\lambda_{\min}}}$ | $\tilde{\kappa}_D = \sqrt{\frac{\sum \lambda_i}{\lambda_{\min}}}$ |

In the tables that follow, we consistently use the above ordering: real vs. complex in the columns and 2-norm vs. Demmel's norm in the rows.

The numbers in the table below are the values that the indicated expressions converge to almost surely as $n \to \infty$.

| | $W(n,n)$ | $\tilde{W}(n,n)$ |
|---|---|---|
| $\frac{1}{n}\lambda_{\max}$ | 4 | 8 |
| $\frac{1}{n^2}\sum_{i=1}^{n}\lambda_i$ | 1 | 2 |

The first row is (6.1) and (6.3). The second row is derived from the law of large numbers and the observation that the trace of a Wishart matrix has the $\chi^2_{n^2}$ distribution in the real case and the $\chi^2_{2n^2}$ in the complex case. Replacing these convergence results with equality, we define four approximate condition numbers:

| $G(n,n)$ | $\tilde{G}(n,n)$ |
|---|---|
| $\kappa' = \sqrt{\frac{4n}{\lambda_{\min}}}$ | $\tilde{\kappa}' = \sqrt{\frac{8n}{\lambda_{\min}}}$ |
| $\kappa'_D = \sqrt{\frac{n^2}{\lambda_{\min}}}$ | $\tilde{\kappa}'_D = \sqrt{\frac{2n^2}{\lambda_{\min}}}$ |

These approximate condition numbers are defined with the point of view that the numerators are merely normalizing factors. (See [56] p.191.)

Directly from the definition of these condition numbers we have the following justification of our approximation:

**Lemma 7.3** *As $n \to \infty$, $\kappa/\kappa'$, $\kappa_D/\kappa'_D$, $\tilde{\kappa}/\tilde{\kappa}'$, and $\tilde{\kappa}_D/\tilde{\kappa}'_D$ all converge almost surely to* 1.

The approximate condition numbers only depend on $\lambda_{\min}$. Thus it becomes necessary to investigate the probability that $\lambda_{\min}$ is small.

**Lemma 7.4** *As $\lambda \to 0$, $P(\lambda_{\min} < \lambda) \sim \sqrt{\lambda n}$ if $M$ has the distribution $W(n,n)$ and $P(\lambda_{\min} < \lambda) \sim \lambda n/2$ if $M$ has the distribution $\tilde{W}(n,n)$.*

*Proof* The real result comes from analyzing the formula given in Theorem 4.1. The complex result is trivial since $n\lambda_{\min}$ has the distribution $\chi^2_2$ according to Theorem 5.5.

**Theorem 7.8** *As* $x \to \infty$,

| $G(n,n)$ | $\tilde{G}(n,n)$ |
|---|---|
| $P(\kappa' > x) \sim 2n/x$ | $P(\tilde{\kappa}' > x) \sim 4n^2/x^2$ |
| $P(\kappa'_D > x) \sim n^{3/2}/x$ | $P(\tilde{\kappa}'_D > x) \sim n^3/x^2$ |

*Proof* Combine the small $\lambda$ behavior described in Lemma 7.4 with the definitions of our condition numbers. The results follow from the obvious change of variables.

In the complex Demmel case, we can compare our results with those known for the exact condition number. We have for all $n$, $P(\tilde{\kappa}_D > x) \sim (n^3 - n)/x^2$ as $x \to \infty$, while we have $P(\tilde{\kappa}'_D > x) \sim n^3/x^2$ as $x \to \infty$. The difference is negligible for all but very small $n$.

# Chapter 8

# Moments of $\lambda_{\min}$ for $W(m,n)$

## 8.1 Computation of the moments

From the form of the pdf of $\lambda_{\min}$ given in Theorem 4.3, we can obtain expected values and other moments of the distribution in closed form in terms of hypergeometric functions. Let

$$
\begin{aligned}
\rho &= \left(\frac{2}{m}\right)^{\beta+1}\Gamma(\beta+1),\\
f(z) &= {}_2F_1(\frac{m-1}{2}, z+1/2, m/2+z+3/2, 1-1/m),\\
g(z) &= {}_2F_1(\frac{m+1}{2}, z+1/2, m/2+z+3/2, 1-1/m).
\end{aligned}
$$

Then,

$$
\begin{aligned}
\int_0^\infty \lambda^\beta e^{-\lambda m/2} d\lambda &= \rho,\\
\int_0^\infty \lambda^\beta e^{-\lambda m/2} U(\frac{m-1}{2}, -\frac{1}{2}, \frac{\lambda}{2}) d\lambda &= \frac{\rho\Gamma(\beta+\frac{5}{2})}{\Gamma(\beta+\frac{m}{2}+2)} f(\beta+\frac{1}{2}),\\
\int_0^\infty \lambda^\beta e^{-\lambda m/2} \frac{d}{d\lambda} U(\frac{m-1}{2}, -\frac{1}{2}, \frac{\lambda}{2}) d\lambda &= -\frac{m-1}{4}\frac{\rho\Gamma(\beta+\frac{3}{2})}{\Gamma(\beta+\frac{m}{2}+2)} g(\beta+\frac{1}{2}).
\end{aligned}
$$

These integrals can by obtained by taking Laplace transforms and interchanging the order of integration. Since integrations of this form are documented in [23] (see formula 7.621.6), we merely list the results.

We derived a further set of formulas that allows us to reduce the hypergeometric functions

$f(z)$ and $g(z)$ into a simpler form.

**Lemma 8.1** *The hypergeometric functions defined above can be computed using the following formulas*

$$
\begin{aligned}
f(k) &= k(2-1/m)(k+(m+1)/2)f(k-1) - \frac{(k+(m-1)/2)(k+(m+1)/2)}{(k-1/2)(k+1)(1-1/m)}f(k-2),\\
g(k) &= (2k-1+(1-k)/m)(k+(m+1)/2)g(k-1) - \frac{(k+(m-1)/2)(k+(m+1)/2)}{k(k-1/2)(1-1/m)}g(k-2),\\
f(-1) &= \frac{m-1}{2}(1-1/m)^{-(m-1)/2}B((m-1)/2;3/2;1-1/m),\\
f(0) &= (m+1)\sqrt{m}/2m,\\
g(0) &= \frac{m+1}{2}(1-1/m)^{-(m+1)/2}B((m+1)/2;1/2;1-1/m),\\
g(1) &= \frac{m(m+3)}{m-1}(g(0)-(m+1)/2\sqrt{m}),
\end{aligned}
$$

*where $B$ denotes the incomplete Beta function, $B(\mu;\nu;x)$.*

*Proof* Though there are many relationships involving hypergeometric functions ([1] and [43]), we have not seen any of the two term recurrences listed in the literature and derived the formulas ourselves. The derivations of all these formulas are tedious, and we omit them.

If $m$ is odd, the two Beta functions can be expressed in the form $p+q\sqrt{m}$ where $p$ and $q$ are rational. (See formula 58:4:10 in [43]). Though we do not rewrite the formula here, it is used in the Mathematica programs in Appendix A.

We conclude from these formulas and Theorem 4.3 that

**Theorem 8.1** *If $m$ and $n$ are of opposite parity, then all the moments of the pdf of $\lambda_{\min}$ are rational. If $m$ and $n$ are odd, the moments take the form $p+q\sqrt{m}$ where $p$ and $q$ are rational.*

One might wonder what happens when $m$ and $n$ are even. In this case, the Beta function above can be expressed in terms that involve an arcsine (See 58:4:11 in [43]). Since Mathematica is capable of calculating the Beta function, we saw no reason to exploit this possibility.

## 8.2 Expected values when $n - m = 0, 1$, and $3$

In the special cases $n - m = 0, 1$, and $3$ (but not $2$), we have the exact expected values of the smallest eigenvalue of a matrix from $G(m, n)$. Table 8.1 lists these expected values and their asymptotics. The expected value integral for $n - m = 1$ is trivial from Theorem 4.2,

Table 8.1: $E(\lambda_{\min})$ for $n - m = 0, 1$, and $3$

| $n - m$ | $E(\lambda_{\min})$ | $m \to \infty$ |
|---|---|---|
| 0 | $\frac{8}{\sqrt{m}(m+1)(m+3)}\,{}_2F_1(\frac{m-1}{2}, \frac{3}{2}; \frac{m+5}{2}; 1 - \frac{1}{m})$ | $\sqrt{8}U(3/2, -1/2, 1/2)/m \approx 0.6886409/m$ |
| 1 | $2/m$ | $2/m$ |
| 3 | $\frac{8\binom{m+2}{3}}{m^3(m+1)}\,{}_2F_1(1-m, 3; 4; -2/m)$ | $(e^2 - 1)/m \approx 6.389056/m$ |

while when $n - m = 0$ or $3$, the integrals can be found in [23]. The asymptotics are most readily calculated from the integral forms for the hypergeometric functions. Particularly useful formulas are 60:3:3 (which has a typographical error in our edition) and 48:3:5 in [43]. We could use the same formulas to further derive exact expressions for all of the moments of these distributions, but in these tables, we content ourselves with merely the expected value. We know of no simple formula for the expected value integral when $n - m = 2$, and we believe that it may not be obtainable as a hypergeometric function of one variable.

Table 8.2, gives five correct digits for the expected value for $n = m$. We also performed a few MATLAB experiments with 1000 $n \times n$ matrices. We found typically that averaging the data from 1000 matrices would give at best two digits, which we included in the table.

Appendix B gives the expected values of $\lambda_{\min}$ for $m \le 25$ and $m - n \le 25$. We computed the exact expressions symbolically and then asked Mathematica to compute the numerical values to 22 decimal places. We then rounded the answers to ten decimal places.

The expected value for fixed $m$ is asymptotic to $n$ as $n \to \infty$. The explanation is the Law of Large Numbers. For large $n$, the matrix $\frac{1}{n}AA^T$ has diagonal elements which are almost surely 1 since these elements are the means of $n$ independent variables with mean value 1. Similarly

Table 8.2: $E(\lambda_{\min})$ when $m = n$

| $m$ | $E(\lambda_{\min})$ | experimental |
| ---: | --- | --- |
| 1 | 1.00000 | 1.01 |
| 2 | 0.42920 | 0.46 |
| 3 | 0.26795 | 0.25 |
| 4 | 0.19387 | 0.20 |
| 5 | 0.15164 | 0.15 |
| 6 | 0.12443 | 0.12 |
| 7 | 0.10547 | 0.10 |
| 8 | 0.091510 | 0.091 |
| 9 | 0.080804 | 0.080 |
| 10 | 0.072336 | 0.0723 |
| 100 | 0.0069209 | |
| 1000 | 0.00068899 | |
| 10000 | 0.000068868 | |

the off-diagonal elements are means of $n$ independent variables with mean value 0.

# Chapter 9

# Random matrices and orthogonal polynomials

We describe in this chapter some interesting relationships between random matrices and orthogonal polynomials. We show that Wishart matrices are associated with Laguerre polynomials, while Gaussian ensembles are associated with Hermite polynomials. The list would not be complete without the third set of classical polynomials, the Jacobi polynomials. They are also associated with eigenvalues of a class of random matrices which we will describe.

## 9.1 The random matrix associated with a weight function

We recall that if $w(x)$ is a non-negative function on a (possibly infinite) interval $[a, b]$, then we can define a sequence of orthogonal polynomials $p_n(x)$ of degree $n$ by the condition

$$\int_a^b p_k(x)p_n(x)w(x)dx = 0, \text{ if } k \neq n.$$

The three classical weight function classes are

| | $w(x)$ | $[a,b]$ | $p_n(x)$ |
|---|---|---|---|
| Laguerre | $x^\alpha e^{-x}$ | $x \in [0, \infty)$ | $L_n^{(\alpha)}(x)$ |
| Hermite | $e^{-x^2}$ | $x \in (-\infty, \infty)$ | $H_n(x)$ |
| Jacobi | $(1-x)^\alpha(1+x)^\beta$ | $x \in [-1, 1]$ | $P_n^{(\alpha,\beta)}(x)$. |

Through a linear change of variables, we may also define Laguerre, Hermite, and Jacobi polynomials on other intervals as well.

We are now ready to define the random matrix associated with a weight function. Let $w(x)$ be a weight function on an interval $[a, b]$, and let $\mathcal{S}$ be the set of symmetric matrices whose spectrum is inside this interval. For $S \in \mathcal{S}$, the scalar function of a matrix argument $f(S) = \det(w(S))$ is well-defined and is invariant under orthogonal similarity transformations of $S$. By normalizing $w$ so that

$$\int_{\mathcal{S}} f(S)dS = 1,$$

we can consider $\mathcal{S}$ to be a probability space. We then introduce

**Definition 9.1** *If $w(x)$ is a weight function on $[a, b]$, and $\mathcal{S}$ is the probability space of symmetric matrices whose eigenvalues are in $[a, b]$ with the probability measure described above, we say that the matrices in $\mathcal{S}$ have the random matrix distribution associated with $w(x)$.*

We immediately have the analogous definition for Hermitian matrices, which we will also use. In the following, we will not be concerned with whether $w(x)$ has been normalized, and we will allow for linear changes of variables as well.

**Theorem 9.1** *The real and complex Wishart matrix distributions are associated with (rescaled) Laguerre weight functions, while the Gaussian ensembles are associated with (rescaled) Hermite weight functions.*

*Proof* From Lemma 3.2, if $A$ is $W(m, n)$, then $2A$ is associated with the Laguerre weight function $x^{(n-m-1)/2}e^{-x}$, while if $A$ is $\tilde{W}(m, n)$ then $2A$ is associated with the Laguerre weight function $x^{n-m}e^{-x}$. From Lemma 3.1, if $A$ is from the $GOE$, then $\sqrt{2}A$ is associated with the Hermite weight function, while if $A$ is from the $GUE$, then it is associated with the Hermite weight function.

The only attempt at the kind of generality given in Definition 9.1 of which we are aware is the short article [32], which compares the random matrices associated with Hermite polynomials with those associated with Legendre polynomials. That Wishart matrices are associated with Laguerre polynomials appears to have never been observed.

## 9.2 Characteristic polynomial

One compact way to all the eigenvalues of a matrix is through its characteristic polynomial, $P_S(t) = \det(tI - S)$. For a random matrix $S$, we can ask for the average characteristic polynomial, $P(t) = E(P_S(t))$. The roots of this polynomial well deserve to be considered a typical set of eigenvalues for the random matrix at hand. So far as we know, the results in this section are new.

For random $n$ by $n$ Hermitian matrices associated with a weight function $w(x)$, the average characteristic polynomial is the nicest result it could be; it is the orthogonal polynomial of degree $n$ defined by $w(x)$:

**Theorem 9.2** *Let a random $n$ by $n$ Hermitian matrix $S$ be associated with a weight function $w$. Then the average characteristic polynomial is the monic orthogonal polynomial of degree $n$ defined by the weight function $w$.*

*Proof* This follows immediately from Theorem 3.2 (3.9), the fact that the characteristic polynomial is defined to be monic, and the following well-known formula from [46]:

**Proposition 9.1** *Let $w(x)$ be a weight function on an interval $[a,b]$. Then*

$$p_n(t) = \int_{[a,b]^n} \prod_{i=1}^{n}(t-\lambda_i)\prod_{i<j}(\lambda_i-\lambda_j)^2\prod_{i=1}^{n} w(\lambda_i)d\lambda_i$$

*defines orthogonal polynomials of degree $n$ with weight function $w$.*

From Theorems 9.1 and 9.2, we have

**Corollary 9.1** *The average characteristic polynomials for matrices from $\tilde{W}(m,n)$ and the GUE are*

$$\begin{array}{ll} \tilde{W}(m,n) & L_m^{n-m}(x/2) \\ GUE & H_n(x). \end{array}$$

For real symmetric matrices, we unfortunately do not have as general a theory as in Theorem 9.2, though we believe this would be a fruitful topic of further research. In particular, it would be very useful to understand

$$p_n^{(\omega)}(t) = \int_{[a,b]^n} \prod_{i=1}^{n}(t-\lambda_i) \prod_{i<j} |\lambda_i - \lambda_j|^{\omega} \prod_{i=1}^{n}(w(\lambda_i)d\lambda_i) \tag{9.1}$$

for arbitrary $\beta$, especially $\beta = 1$. It is possible, however, to compute $p_n^{(\omega)}(t)$ when the weight function $w(x)$ is one of the classical weight functions:

**Lemma 9.1** *The polynomial $p_n^{(\omega)}(t)$ defined in (9.1) is proportional to*

$$\begin{array}{ll} P_n^{(\alpha',\beta')}(t) & \text{if } w(x) = (1-x)^{\alpha}(1+x)^{\beta} \text{on } [-1,1], \\ L_n^{(\alpha')}(\frac{2}{\omega}t) & \text{if } w(x) = x^{\alpha}e^{-x} \text{on } [0,\infty], \\ H_n(\sqrt{\frac{2}{\omega}}t) & \text{if } w(x) = e^{-x^2} \text{on } [-\infty,\infty], \end{array}$$

*where $\alpha' = -1 + 2(\alpha+1)/\omega$ and $\beta' = -1 + 2(\beta+1)/\omega$.*

*Proof* The Jacobi result appears in [4]. The Laguerre and Hermite results do not appear in the literature, but can be derived from the Jacobi result using Equations (5.3.4) and (5.6.3) in [46]. We have derived an alternative proof of this lemma by verifying that in each case, the multivariate integral satisfies the differential equation for the polynomial. These proofs are similar in flavor to the proof we gave of Theorem 4.1, so we omit them. We suspect a yet simpler proof of this lemma might be obtained by taking the Laplace transform in $t$, because the classical orthogonal polynomials have simple Laplace transforms, though we have not pursued this idea.

We now consider another set of random matrices that arise in the multivariate analysis of variance (MANOVA). Let $A$ be a matrix from $W(m,n_1)$ and $B$ be from $W(m,n_2)$. Then we will say the matrix $C = A(A+B)^{-1}$ is $W(m,n_1,n_2)$. Since $C = (I+AB^{-1})^{-1}$, it is positive definite with all eigenvalues on the interval $[0,1]$. It is cleaner to work with the symmetric matrix $S = A^{-1/2}CA^{1/2}$, which has the same eigenvalues and is invariant under orthogonal similarity transformations. The joint density of the elements of the random matrix $S$ can be given as $c \det w(S)$ where $w(x) = x^{\frac{1}{2}(n_1-m-1)}(1-x)^{\frac{1}{2}(n_2-m-1)}$. (See [2], p. 529.) Thus these matrices are associated with the Jacobi weight function rescaled to the interval $[0,1]$.

**Theorem 9.3** *The average characteristic polynomials for matrices from $W(m,n)$, the GOE,* *and* $W(m,n_1,n_2)$ *are*

$$\begin{array}{ll} W(m,n) & L_m^{(n-m)}(x) \\ GOE & H_n(x) \\ W(m,n_1,n_2) & P_m^{(n_1-m,n_2-m)}(1-2x) \end{array}$$

*Proof* The proof follows from Lemma 9.1 and Theorem 9.1. $\square$

The following MATLAB instructions give a simple way to test these formulas:

```
rand('normal')
ph=zeros(1,8); pl=ph; pj=ph;
for j=1:10000, a=rand(7);s=.5*(a+a');ph=ph+poly(s);end
for j=1:10000, a=rand(7);s=a*a';pl=pl+poly(s);end
for j=1:10000, a=rand(7);s=a*a';b=rand(7);t=b*b';pj=pj+poly(s/(s+t));end
```

In the three cases below we list normalized results of the MATLAB experiment on the left and the coefficients of the true polynomials (normalized to be integers) on the right. The coefficients are written sequentially with that of the highest order term in the first row.

| *Hermite* | | *Laguerre* | | *Legendre* | |
|---|---|---|---|---|---|
| `ph` | $H_7$ | `pl` | $L_7$ | `pj` | $P_7$ |
| 128.0 | 128 | -1.0 | -1 | 3432.0 | 3432 |
| -3.5 | 0 | 49.0 | 49 | -11995.3 | -12012 |
| -1337.2 | -1344 | -880.8 | -882 | 16580.5 | 16632 |
| 27.4 | 0 | 7300.7 | 7350 | -11490.3 | -11550 |
| 3340.6 | 3360 | -28906.8 | -29400 | 4168.2 | 4200 |
| -25.8 | 0 | 51226.3 | 52920 | -748.6 | -756 |
| -1686.6 | -1680 | -33492.1 | -35280 | 55.5 | 56 |
| 38.8 | 0 | 4530.4 | 5040 | -1.0 | 1 |

## 9.3 The mode

Another notion of the most likely distribution of the eigenvalues is given by the set of numbers that maximize the joint density function. Unlike the characteristic polynomial viewpoint which we believe to be new, this has been noticed by the physicists for the cases of the GOE and the GUE.

We start with another proposition from [46].

**Proposition 9.2** *The set of $\lambda_i$ that maximize the expression*

$$T(\lambda_1, \ldots, \lambda_n) = \prod_{i=1}^{n} w(\lambda_i) \prod_{i<j} |\lambda_i - \lambda_j|,$$

*where $w(x)$ is a Jacobi, Laguerre, or Hermite weight function, is given by the roots of an orthogonal polynomial with respect to the same type of weight function.*

See ([46], p.140) for the details relating the weight function $w(x)$ with the orthogonal polynomial whose roots provide the maximum.

The conclusion to be reached is the following table listing the orthogonal polynomial whose roots maximize the joint density of the eigenvalues, when there is a maximum.

| | |
|---|---|
| $W(m,n)$ | $L_m^{(n-m-2)}(x)$ |
| $\tilde{W}(m,n)$ | $L_m^{(n-m-1)}(x/2)$ |
| GOE | $H_n(x)$ |
| GUE | $H_n(x)$. |

We find it curious that the mode and expected characteristic polynomial differ slightly in the Wishart case, but are exactly identical for the Gaussian ensembles. This makes us wonder whether the roots of Hermite polynomials may somehow be related to the behavior of the Riemann zeta function. (See Section 2.4.)

# Appendix A

# Mathematica programs

Our Mathematica programs for computing the distribution of the smallest eigenvalue of a Wishart matrix and the moments of the distribution follow. We have separated the even and odd cases, though they could in fact be merged into one routine with some "If" statements. Users should be careful not to try to calculate both cases at once with these programs.

```
(*************************************************************)
(************ Smallest Eigenvalue of Wishart Matrices ***********)
(************                n-m odd                 ***********)
(************              Alan Edelman   3/89       ***********)
(*************************************************************)

p[m_,n_]:=p[m,n]=If[n==m+1,1,
                   Block[{s},
                   (s[-1]=0; s[0] =p[m,n-2];
                   s[i_]:=s[i]=Expand[(l+n-i)s[i-1]-2l/(m-i)D[s[i-1],l]
                                     + l(i-1)((n-i-1)/(m-i))s[i-2]];
                   s[m-1])]]

c[m_,n_]:=m Product[Gamma[j/2]/Gamma[(j+m)/2],{j,2,n-m}]*
          2^(m(1+m-n)/2-1)

(* Calculate zth moment of W(m,n) *)
moment[m_,n_,z_]:=c[m,n]*CoefficientList[p[m,n],l].
    Table[(2/m)^(z+k+1+(n-m-1)/2)Gamma[z+k+1+(n-m-1)/2],{k,0,(m-1)(n-m-1)/2}]

(*************************************************************)
(************ Smallest Eigenvalue of Wishart Matrices ***********)
(************                n-m even                ***********)
(************              Alan Edelman   3/89       ***********)
(*************************************************************)

(* Here we compute an ordered pair of polynomials Q and R as
   described in Chapter 4; we denote the pair by q *)

(* n=m *)
q[m_,m_]:={1,0}

twobytwo[m_]={{0,(m-1)/2},{2l,l+1}}

(* n > m *)
q[m_,n_]:=q[m,n]=Block[{s},
       (s[-1]={0,0}; s[0]=q[m,n-2];
       s[i_]:=s[i]=Expand[ (l+n-i)s[i-1]-2l/(m-i)D[s[i-1],l]
         +l(i-1)((n-i-1)/(m-i))s[i-2]-twobytwo[m].s[i-1]/(m-i)];
       s[m-1])]

(* Calculate zth moment of  W(m,n) *)
moment[m_,m_,z_]:=Simplify[c[m,m]* Sqrt[(2/m)^(2z+1)] gr[m,z+m/2](z+1)f[m,z]]
moment[m_,n_,z_]:=(Print[q[m,n]];Simplify[c[m,n]*(
    CoefficientList[q[m,n][[1]],l].
    Table[Sqrt[(2/m)^(2z+2k+n-m+1)]gr[m,z+k+n/2](z+k+(n-m)/2+1)f[m,z+k+(n-m)/2],
       {k,0,(m-1)(n-m)/2,1}]
```

```
   -(m-1)/4*CoefficientList[q[m,n][[2]],l].
    Table [Sqrt[(2/m)^(2z+2k+n-m+1)]gr[m,z+k+n/2] g[m,z+k+(n-m)/2],
      {k,0,(m-1)(n-m)/2,1}])])

(* We use l on the next line to denote z+k+n/2 *)
gr[m_,l_]:=gr[m,l]=Gamma[l+1-m/2]Gamma[l+(-m+1)/2]/Gamma[m/2+l+(-m+3)/2]

c[m_,m_]:=m Gamma[(m+1)/2] / Sqrt[ 2Pi]
c[m_,n_]:=m Product[Gamma[j/2]/Gamma[(j+m)/2],{j,2,n-m}]*Sqrt[2^(m(m-n)-1)]

(* These are the f and g functions of section 5, when m is odd
   they are computed exactly, otherwise they are left in terms of Betas *)
   f[m_,-1]:=f[m,-1]=If[ EvenQ[m],
          (m-1)/2 (1-1/m)^(-(m-1)/2) Beta[1-1/m,(m-1)/2,3/2],
          Simplify[(m-1)/2 Sqrt[(1-1/m)^(-(m-1))] *
          Sum[(-1)^j((m-3)/2)!/j!/((m-3)/2-j)!
          (1-Sqrt[m^(-(2j+3))])/(j+3/2),{j,0,(m-3)/2}]]]
   f[m_,0] := f[m,0]= (m+1)*Sqrt[m]/2/m
   f[m_,k_]:=f[m,k]= Simplify[
          (k(2-1/m)(k+(m+1)/2)f[m,k-1]-(k+(m-1)/2)(k+(m+1)/2)f[m,k-2])/
          ((k-1/2)(1-1/m)(k+1))]
   g[m_,0]:=g[m,0]=If[EvenQ[m],
          (m+1)/2 (1-1/m)^(-((m+1)/2)) Beta[1-1/m,(m+1)/2,1/2],
          Simplify[(m+1)/2 Sqrt[(1-1/m)^(-(m+1))] *
          Sum[(-1)^j((m-1)/2)!/j!/((m-1)/2-j)!
          (1-Sqrt[m^(-(2j+1))])/(j+1/2),{j,0,(m-1)/2}]]]
   g[m_,1]:=g[m,1]=(g[m,0]-(m+1)/(2Sqrt[m]))m(m+3)/(m-1)
   g[m_,k_]:=g[m,k]= Simplify[
          ((2k-1+(1-k)/m)(k+(m+1)/2)g[m,k-1]-(k+(m-1)/2)(k+(m+1)/2)g[m,k-2])/
          (k(k-1/2)(1-1/m))]
```

# Appendix B

# Table of expected values

We tabulate $E(\lambda_{\min})$ for matrices from $W(m,n)$ for $2 \leq m \leq 25$ and $0 \leq n - m \leq 25$. The tables were computed to 22 significant places and then chopped to ten digits after the decimal point. Unlike standard numerical calculations, we had the expected values in an exact symbolic form, and thus it would have required little more work to compute 100 significant digits.

| n | m=2 | n | m=3 | n | m=4 |
|---|---|---|---|---|---|
| 2 | 0.42920 36732 | 3 | 0.26794 91924 | 4 | 0.19386 69492 |
| 3 | 1.00000 00000 | 4 | 0.66666 66666 | 5 | 0.50000 00000 |
| 4 | 1.64380 55098 | 5 | 1.14532 36690 | 6 | 0.88171 34535 |
| 5 | 2.33333 33333 | 6 | 1.67901 23456 | 7 | 1.31875 00000 |
| 6 | 3.05475 68872 | 7 | 2.25346 00587 | 8 | 1.79852 99628 |
| 7 | 3.80000 00000 | 8 | 2.85962 50571 | 9 | 2.31261 16071 |
| 8 | 4.56388 30351 | 9 | 3.49136 49631 | 10 | 2.85500 70562 |
| 9 | 5.34285 71428 | 10 | 4.14428 69481 | 11 | 3.42128 78999 |
| 10 | 6.13436 84145 | 11 | 4.81512 31731 | 12 | 4.00807 04176 |
| 11 | 6.93650 79365 | 12 | 5.50136 57834 | 13 | 4.61270 04251 |
| 12 | 7.74780 52559 | 13 | 6.20104 09819 | 14 | 5.23305 06344 |
| 13 | 8.56709 95671 | 14 | 6.91256 25619 | 15 | 5.86738 50439 |
| 14 | 9.39345 56939 | 15 | 7.63463 32629 | 16 | 6.51426 51192 |
| 15 | 10.22610 72261 | 16 | 8.36617 61781 | 17 | 7.17248 30335 |
| 16 | 11.06441 68149 | 17 | 9.10628 57451 | 18 | 7.84101 30065 |
| 17 | 11.90784 77078 | 18 | 9.85419 18999 | 19 | 8.51897 50858 |
| 18 | 12.75594 28659 | 19 | 10.60923 33212 | 20 | 9.20560 76928 |
| 19 | 13.60830 93377 | 20 | 11.37083 70985 | 21 | 9.90024 64731 |
| 20 | 14.46460 63584 | 21 | 12.13850 30368 | 22 | 10.60230 77679 |
| 21 | 15.32453 61449 | 22 | 12.91179 13689 | 23 | 11.31127 55311 |
| 22 | 16.18783 66763 | 23 | 13.69031 30120 | 24 | 12.02669 08511 |
| 23 | 17.05427 59614 | 24 | 14.47372 17549 | 25 | 12.74814 34726 |
| 24 | 17.92364 74344 | 25 | 15.26170 79269 | 26 | 13.47526 48690 |
| 25 | 18.79576 62205 | 26 | 16.05399 32198 | 27 | 14.20772 25338 |
| 26 | 19.67046 60775 | 27 | 16.85032 64164 | 28 | 14.94521 52378 |
| 27 | 20.54759 68694 | 28 | 17.65047 98383 | 29 | 15.68746 90600 |

| n | m=5 | n | m=6 | n | m=7 |
|---|---|---|---|---|---|
| 5 | 0.15163 83427 | 6 | 0.12443 41749 | 7 | 0.10547 25222 |
| 6 | 0.40000 00000 | 7 | 0.33333 33333 | 8 | 0.28571 42857 |
| 7 | 0.71765 48409 | 8 | 0.60542 71012 | 9 | 0.52371 54716 |
| 8 | 1.08825 60000 | 9 | 0.92736 29890 | 10 | 0.80839 56410 |
| 9 | 1.50106 56933 | 10 | 1.29000 25081 | 11 | 1.13193 23711 |
| 10 | 1.94856 39884 | 11 | 1.68673 30432 | 12 | 1.48851 80375 |
| 11 | 2.42523 57127 | 12 | 2.11257 47590 | 13 | 1.87368 43823 |
| 12 | 2.92689 13163 | 13 | 2.56366 14015 | 14 | 2.28389 92217 |
| 13 | 3.45025 85950 | 14 | 3.03691 77175 | 15 | 2.71630 95644 |
| 14 | 3.99272 34899 | 15 | 3.52984 87205 | 16 | 3.16856 99712 |
| 15 | 4.55215 82326 | 16 | 4.04039 64542 | 17 | 3.63872 35095 |
| 16 | 5.12680 32872 | 17 | 4.56683 94337 | 18 | 4.12511 66562 |
| 17 | 5.71518 38047 | 18 | 5.10772 01146 | 19 | 4.62633 69196 |
| 18 | 6.31604 89953 | 19 | 5.66179 13630 | 20 | 5.14116 61301 |
| 19 | 6.92832 71645 | 20 | 6.22797 61562 | 21 | 5.66854 48153 |
| 20 | 7.55109 17322 | 21 | 6.80533 67155 | 22 | 6.20754 45893 |
| 21 | 8.18353 51199 | 22 | 7.39305 04993 | 23 | 6.75734 64476 |
| 22 | 8.82494 83888 | 23 | 7.99039 12777 | 24 | 7.31722 34894 |
| 23 | 9.47470 51501 | 24 | 8.59671 40299 | 25 | 7.88652 70064 |
| 24 | 10.13224 87016 | 25 | 9.21144 27595 | 26 | 8.46467 51679 |
| 25 | 10.79708 16322 | 26 | 9.83406 05679 | 27 | 9.05114 37324 |
| 26 | 11.46875 73400 | 27 | 10.46410 14913 | 28 | 9.64545 83574 |
| 27 | 12.14687 30488 | 28 | 11.10114 37344 | 29 | 10.24718 81838 |
| 28 | 12.83106 40135 | 29 | 11.74480 40188 | 30 | 10.85594 04444 |
| 29 | 13.52099 86745 | 30 | 12.39473 28285 | 31 | 11.47135 59041 |
| 30 | 14.21637 45776 | 31 | 13.05061 03850 | 32 | 12.09310 49796 |

| n | m=8 | n | m=9 | n | m=10 |
|---|---|---|---|---|---|
| 8 | 0.09150 97954 | 9 | 0.08080 35714 | 10 | 0.07233 55653 |
| 9 | 0.25000 00000 | 10 | 0.22222 22222 | 11 | 0.20000 00000 |
| 10 | 0.46151 96969 | 11 | 0.41257 36025 | 12 | 0.37304 04473 |
| 11 | 0.71673 11774 | 12 | 0.64387 80726 | 13 | 0.58455 30642 |
| 12 | 1.00888 20636 | 13 | 0.91025 49030 | 14 | 0.82937 21482 |
| 13 | 1.33285 48256 | 14 | 1.20717 17360 | 15 | 1.10346 08797 |
| 14 | 1.68464 22272 | 15 | 1.53102 70970 | 16 | 1.40357 16270 |
| 15 | 2.06102 82827 | 16 | 1.87889 33531 | 17 | 1.72703 52508 |
| 16 | 2.45938 12163 | 17 | 2.24834 76088 | 18 | 2.07162 11316 |
| 17 | 2.87751 26503 | 18 | 2.63735 50957 | 19 | 2.43543 96181 |
| 18 | 3.31357 83268 | 19 | 3.04418 58566 | 20 | 2.81687 15791 |
| 19 | 3.76600 60646 | 20 | 3.46735 35262 | 21 | 3.21451 61001 |
| 20 | 4.23344 22226 | 21 | 3.90556 93055 | 22 | 3.62715 07720 |
| 21 | 4.71471 11079 | 22 | 4.35770 66875 | 23 | 4.05370 09721 |
| 22 | 5.20878 36624 | 23 | 4.82277 39753 | 24 | 4.49321 57239 |
| 23 | 5.71475 29409 | 24 | 5.29989 25622 | 25 | 4.94484 84626 |
| 24 | 6.23181 46503 | 25 | 5.78827 95487 | 26 | 5.40784 15261 |
| 25 | 6.75925 15224 | 26 | 6.28723 36745 | 27 | 5.88151 35139 |
| 26 | 7.29642 06311 | 27 | 6.79612 38175 | 28 | 6.36524 88867 |
| 27 | 7.84274 29995 | 28 | 7.31437 95075 | 29 | 6.85848 93333 |
| 28 | 8.39769 50102 | 29 | 7.84148 30349 | 30 | 7.36072 65504 |
| 29 | 8.96080 12483 | 30 | 8.37696 28383 | 31 | 7.87149 61581 |
| 30 | 9.53162 84953 | 31 | 8.92038 79254 | 32 | 8.39037 25418 |
| 31 | 10.10978 06538 | 32 | 9.47136 31341 | 33 | 8.91696 44497 |
| 32 | 10.69489 44335 | 33 | 10.02952 50854 | 34 | 9.45091 12170 |
| 33 | 11.28663 56620 | 34 | 10.59453 87070 | 35 | 9.99187 95082 |

| n | m=11 | n | m=12 | n | m=13 |
|---|---|---|---|---|---|
| 11 | 0.06547 13615 | 12 | 0.05979 53181 | 13 | 0.05502 38447 |
| 12 | 0.18181 81818 | 13 | 0.16666 66666 | 14 | 0.15384 61538 |
| 13 | 0.34043 73973 | 14 | 0.31308 57550 | 15 | 0.28980 93197 |
| 14 | 0.53529 06594 | 15 | 0.49372 04732 | 16 | 0.45816 49440 |
| 15 | 0.76180 39979 | 16 | 0.70449 10738 | 17 | 0.65524 97891 |
| 16 | 1.01636 14552 | 17 | 0.94214 07756 | 18 | 0.87811 53450 |
| 17 | 1.29602 32011 | 18 | 1.20399 77427 | 19 | 1.12432 37307 |
| 18 | 1.59835 00475 | 19 | 1.48782 67734 | 20 | 1.39182 06754 |
| 19 | 1.92128 60942 | 20 | 1.79172 97178 | 21 | 1.67885 00614 |
| 20 | 2.26307 62871 | 21 | 2.11407 51130 | 22 | 1.98389 33004 |
| 21 | 2.62220 63733 | 22 | 2.45344 65933 | 23 | 2.30562 46594 |
| 22 | 2.99735 79391 | 23 | 2.80860 39887 | 24 | 2.64287 73826 |
| 23 | 3.38737 39867 | 24 | 3.17845 33324 | 25 | 2.99461 74207 |
| 24 | 3.79123 20941 | 25 | 3.56202 33148 | 26 | 3.35992 26845 |
| 25 | 4.20802 31611 | 26 | 3.95844 65148 | 27 | 3.73796 64133 |
| 26 | 4.63693 43553 | 27 | 4.36694 42452 | 28 | 4.12800 36707 |
| 27 | 5.07723 52670 | 28 | 4.78681 41751 | 29 | 4.52936 02569 |
| 28 | 5.52826 65544 | 29 | 5.21742 01224 | 30 | 4.94142 35175 |
| 29 | 5.98943 05451 | 30 | 5.65818 35579 | 31 | 5.36363 46573 |
| 30 | 6.46018 33924 | 31 | 6.10857 64806 | 32 | 5.79548 22637 |
| 31 | 6.94002 84793 | 32 | 6.56811 53970 | 33 | 6.23649 68114 |
| 32 | 7.42851 08353 | 33 | 7.03635 62015 | 34 | 6.68624 59685 |
| 33 | 7.92521 23787 | 34 | 7.51288 97961 | 35 | 7.14433 05661 |
| 34 | 8.42974 78403 | 35 | 7.99733 83218 | 36 | 7.61038 11165 |
| 35 | 8.94176 12516 | 36 | 8.48935 18991 | 37 | 8.08405 47921 |
| 36 | 9.46092 29043 | 37 | 8.98860 57969 | 38 | 8.56503 27919 |

| n | m=14 | n | m=15 | n | m=16 |
|---|---|---|---|---|---|
| 14 | 0.05095 68663 | 15 | 0.04744 92130 | 16 | 0.04439 30078 |
| 15 | 0.14285 71428 | 16 | 0.13333 33333 | 17 | 0.12500 00000 |
| 16 | 0.26975 92161 | 17 | 0.25230 73055 | 18 | 0.23697 87425 |
| 17 | 0.42740 27205 | 18 | 0.40052 30533 | 19 | 0.37683 27319 |
| 18 | 0.61247 84415 | 19 | 0.57497 46818 | 20 | 0.54181 78817 |
| 19 | 0.82230 40167 | 20 | 0.77321 09593 | 21 | 0.72968 48225 |
| 20 | 1.05464 65296 | 21 | 0.99317 92498 | 22 | 0.93854 01424 |
| 21 | 1.30761 24117 | 22 | 1.23312 90389 | 23 | 1.16676 10162 |
| 22 | 1.57957 33412 | 23 | 1.49154 70620 | 24 | 1.41293 79112 |
| 23 | 1.86911 35462 | 24 | 1.76711 11141 | 25 | 1.67583 37845 |
| 24 | 2.17499 08255 | 25 | 2.05865 57950 | 26 | 1.95435 37743 |
| 25 | 2.49610 68584 | 26 | 2.36514 63283 | 27 | 2.24752 19758 |
| 26 | 2.83148 40691 | 27 | 2.68565 80779 | 28 | 2.55446 32132 |
| 27 | 3.18024 72650 | 28 | 3.01936 02250 | 29 | 2.87438 84636 |
| 28 | 3.54160 88442 | 29 | 3.36550 25602 | 30 | 3.20658 30189 |
| 29 | 3.91485 67254 | 30 | 3.72340 46604 | 31 | 3.55039 67491 |
| 30 | 4.29934 43910 | 31 | 4.09244 69233 | 32 | 3.90523 60065 |
| 31 | 4.69448 25949 | 32 | 4.47206 30702 | 33 | 4.27055 68309 |
| 32 | 5.09973 23984 | 33 | 4.86173 38242 | 34 | 4.64585 91999 |
| 33 | 5.51459 92771 | 34 | 5.26098 15408 | 35 | 5.03068 21280 |
| 34 | 5.93862 81007 | 35 | 5.66936 56163 | 36 | 5.42459 94618 |
| 35 | 6.37139 88288 | 36 | 6.08647 85381 | 37 | 5.82721 62514 |
| 36 | 6.81252 28008 | 37 | 6.51194 24684 | 38 | 6.23816 56018 |
| 37 | 7.26163 95212 | 38 | 6.94540 62739 | 39 | 6.65710 59279 |
| 38 | 7.71841 38601 | 39 | 7.38654 29325 | 40 | 7.08371 85498 |
| 39 | 8.18253 36047 | 40 | 7.83504 72574 | 41 | 7.51770 55783 |

| n | m=17 | n | m=18 | n | m=19 |
|---|---|---|---|---|---|
| 17 | 0.04170 64222 | 18 | 0.03932 62653 | 19 | 0.03720 29666 |
| 18 | 0.11764 70588 | 19 | 0.11111 11111 | 20 | 0.10526 31578 |
| 19 | 0.22340 78577 | 20 | 0.21130 84599 | 21 | 0.20045 33390 |
| 20 | 0.35579 46581 | 21 | 0.33698 61582 | 22 | 0.32006 99998 |
| 21 | 0.51229 09032 | 22 | 0.48582 66435 | 23 | 0.46197 06360 |
| 22 | 0.69082 44814 | 23 | 0.65591 42557 | 24 | 0.62437 83425 |
| 23 | 0.88964 31884 | 24 | 0.84562 25271 | 25 | 0.80577 90508 |
| 24 | 1.10723 85103 | 25 | 1.05354 55907 | 26 | 1.00485 93492 |
| 25 | 1.34229 46097 | 26 | 1.27845 24192 | 27 | 1.22046 51535 |
| 26 | 1.59365 20199 | 27 | 1.51925 43031 | 28 | 1.45157 23910 |
| 27 | 1.86028 06565 | 28 | 1.77498 06713 | 29 | 1.69726 52154 |
| 28 | 2.14125 90977 | 29 | 2.04476 05077 | 30 | 1.95671 92537 |
| 29 | 2.43575 82841 | 30 | 2.32780 77045 | 31 | 2.22918 83847 |
| 30 | 2.74302 84450 | 31 | 2.62340 92865 | 32 | 2.51399 40900 |
| 31 | 3.06238 84477 | 32 | 2.93091 57942 | 33 | 2.81051 67380 |
| 32 | 3.39321 70099 | 33 | 3.24973 33243 | 34 | 3.11818 83524 |
| 33 | 3.73494 53674 | 34 | 3.57931 68722 | 35 | 3.43648 65492 |
| 34 | 4.08705 11012 | 35 | 3.91916 47085 | 36 | 3.76492 94007 |
| 35 | 4.44905 28964 | 36 | 4.26881 35915 | 37 | 4.10307 10535 |
| 36 | 4.82050 60615 | 37 | 4.62783 46625 | 38 | 4.45049 79595 |
| 37 | 5.20099 86711 | 38 | 4.99582 99020 | 39 | 4.80682 56157 |
| 38 | 5.59014 82288 | 39 | 5.37242 90557 | 40 | 5.17169 57276 |
| 39 | 5.98759 87620 | 40 | 5.75728 69512 | 41 | 5.54477 37278 |
| 40 | 6.39301 82829 | 41 | 6.15008 11467 | 42 | 5.92574 65961 |
| 41 | 6.80609 65577 | 42 | 6.55050 98604 | 43 | 6.31432 09360 |
| 42 | 7.22654 31404 | 43 | 6.95829 01409 | 44 | 6.71022 12701 |

| n | m=20 |
|---|---|
| 20 | 0.03529 70957 |
| 21 | 0.10000 00000 |
| 22 | 0.19065 98061 |
| 23 | 0.30477 38068 |
| 24 | 0.44035 43346 |
| 25 | 0.59574 81050 |
| 26 | 0.76954 19025 |
| 27 | 0.96050 58427 |
| 28 | 1.16755 58113 |
| 29 | 1.38972 68838 |
| 30 | 1.62615 35922 |
| 31 | 1.87605 47494 |
| 32 | 2.13872 14618 |
| 33 | 2.41350 74618 |
| 34 | 2.69982 11795 |
| 35 | 2.99711 91535 |
| 36 | 3.30490 04921 |
| 37 | 3.62270 21724 |
| 38 | 3.95009 50184 |
| 39 | 4.28668 02354 |
| 40 | 4.63208 64042 |
| 41 | 4.98596 68596 |
| 42 | 5.34799 73929 |
| 43 | 5.71787 42283 |
| 44 | 6.09531 22337 |
| 45 | 6.48004 33323 |

| n | m=21 |
|---|---|
| 21 | 0.03357 68963 |
| 22 | 0.09523 80952 |
| 23 | 0.18177 92904 |
| 24 | 0.29087 51603 |
| 25 | 0.42067 56766 |
| 26 | 0.56963 81618 |
| 27 | 0.73644 02175 |
| 28 | 0.91992 76433 |
| 29 | 1.11908 01045 |
| 30 | 1.33298 69005 |
| 31 | 1.56082 90012 |
| 32 | 1.80186 52373 |
| 33 | 2.05542 13870 |
| 34 | 2.32088 13653 |
| 35 | 2.59767 99895 |
| 36 | 2.88529 69550 |
| 37 | 3.18325 17627 |
| 38 | 3.49109 94046 |
| 39 | 3.80842 66657 |
| 40 | 4.13484 89295 |
| 41 | 4.47000 74022 |
| 42 | 4.81356 66865 |
| 43 | 5.16521 26511 |
| 44 | 5.52465 05497 |
| 45 | 5.89160 33552 |
| 46 | 6.26581 02777 |

| n | m=22 |
|---|---|
| 22 | 0.03201 65067 |
| 23 | 0.09090 90909 |
| 24 | 0.17368 97257 |
| 25 | 0.27819 06321 |
| 26 | 0.40268 47041 |
| 27 | 0.54572 86244 |
| 28 | 0.70608 20529 |
| 29 | 0.88265 95627 |
| 30 | 1.07449 90916 |
| 31 | 1.28073 97419 |
| 32 | 1.50060 53503 |
| 33 | 1.73339 18668 |
| 34 | 1.97845 73796 |
| 35 | 2.23521 40576 |
| 36 | 2.50312 15255 |
| 37 | 2.78168 13413 |
| 38 | 3.07043 23372 |
| 39 | 3.36894 66512 |
| 40 | 3.67682 63181 |
| 41 | 3.99370 03198 |
| 42 | 4.31922 20172 |
| 43 | 4.65306 69012 |
| 44 | 4.99493 06136 |
| 45 | 5.34452 71970 |
| 46 | 5.70158 75419 |
| 47 | 6.06585 80022 |

| n | m=23 |
|---|---|
| 23 | 0.03059 46537 |
| 24 | 0.08695 65217 |
| 25 | 0.16628 98889 |
| 26 | 0.26656 75886 |
| 27 | 0.38617 27706 |
| 28 | 0.52375 17098 |
| 29 | 0.67813 84506 |
| 30 | 0.84831 02317 |
| 31 | 1.03335 83550 |
| 32 | 1.23246 77509 |
| 33 | 1.44490 18854 |
| 34 | 1.66999 11775 |
| 35 | 1.90712 38482 |
| 36 | 2.15573 85250 |
| 37 | 2.41531 81576 |
| 38 | 2.68538 49397 |
| 39 | 2.96549 60194 |
| 40 | 3.25523 98391 |
| 41 | 3.55423 29874 |
| 42 | 3.86211 74696 |
| 43 | 4.17855 83272 |
| 44 | 4.50324 15502 |
| 45 | 4.83587 22357 |
| 46 | 5.17617 29579 |
| 47 | 5.52388 23181 |
| 48 | 5.87875 36502 |

| n | m=24 |
|---|---|
| 24 | 0.02929 36778 |
| 25 | 0.08333 33333 |
| 26 | 0.15949 51310 |
| 27 | 0.25587 79818 |
| 28 | 0.37096 43285 |
| 29 | 0.50348 15168 |
| 30 | 0.65233 12262 |
| 31 | 0.81654 79465 |
| 32 | 0.99527 19581 |
| 33 | 1.18773 04413 |
| 34 | 1.39322 35540 |
| 35 | 1.61111 37618 |
| 36 | 1.84081 74121 |
| 37 | 2.08179 79246 |
| 38 | 2.33356 01865 |
| 39 | 2.59564 58696 |
| 40 | 2.86762 94706 |
| 41 | 3.14911 49284 |
| 42 | 3.43973 27097 |
| 43 | 3.73913 72790 |
| 44 | 4.04700 48883 |
| 45 | 4.36303 16351 |
| 46 | 4.68693 17478 |
| 47 | 5.01843 60643 |
| 48 | 5.35729 06770 |
| 49 | 5.70325 57219 |

| n | m=25 |
|---|---|
| 25 | 0.02809 87980 |
| 26 | 0.08000 00000 |
| 27 | 0.15323 41171 |
| 28 | 0.24601 35850 |
| 29 | 0.35691 06045 |
| 30 | 0.48472 61246 |
| 31 | 0.62842 35023 |
| 32 | 0.78708 96640 |
| 33 | 0.95990 99417 |
| 34 | 1.14615 05383 |
| 35 | 1.34514 56363 |
| 36 | 1.55628 75323 |
| 37 | 1.77901 88530 |
| 38 | 2.01282 62652 |
| 39 | 2.25723 52975 |
| 40 | 2.51180 60103 |
| 41 | 2.77612 93322 |
| 42 | 3.04982 39262 |
| 43 | 3.33253 34862 |
| 44 | 3.62392 43873 |
| 45 | 3.92368 36298 |
| 46 | 4.23151 70304 |
| 47 | 4.54714 76226 |
| 48 | 4.87031 42355 |
| 49 | 5.20077 02257 |
| 50 | 5.53828 23429 |

# Appendix C

# Other uses of the programs

We give here a few sample outputs from the Mathematica routines. For example, we find that q[3,15] is the ordered pair of polynomials

```
{43589145600 + 51412838400*l + 27622425600*l^2 + 9260697600*l^3 +
      2202076800*l^4 + 396506880*l^5 + 55883520*l^6 + 6216210*l^7 +
      540540*l^8 + 35970*l^9 + 1760*l^10 + 58*l^11 + l^12,
-29059430400*l - 24588748800*l^2 - 10218700800*l^3 - 2767564800*l^4 -
      545529600*l^5 - 82494720*l^6 - 9750510*l^7 - 893970*l^8 - 62370*l^9 -
      3190*l^10 - 110*l^11 - 2*l^12}
```

These correspond to the polynomials $Q_{3,15}$ and $R_{3,15}$ described in Theorem 2.1. The constant $c_{3,15}$ is obtained by typing c[3,15] and we find that $c_{3,15} = (2\pi)^{-1/2}/21576627072000$. From Theorem 4.3, we now know the distribution of the smallest eigenvalue for $W(3, 15)$. We could now readily plot the distribution or derive any quantity of interest from the distribution.

We obtained the first and second moments by typing m1=moment[3,15,1] and m2=moment[3,15,2]. We further numerically obtained the mean and variance to ten places by typing m=N[m1,10], v=N[m2-m1^2,10]. The results were

```
               485986370753
m1 = 14 - -------------------
          44079842304 Sqrt[3]

                16553532759625
m2 = 210 - -------------------
           66119763456 Sqrt[3]
```

```
m = 7.634633263
v = 7.168805888
```

One can also obtain very precise values for the percentage points of the distribution. This is particularly easy when $n - m$ is odd since Mathematica will perform the integration to obtain the cumulative density function. Then with the command FindRoot, Mathematica will find the percentage point to arbitrary precision.

# Bibliography


[1] M. Abramowitz and I.A. Stegun, eds., *Handbook of Mathematical Functions,* Dover Publications, New York, 1970.

[2] T.W. Anderson, *An Introduction to Multivariate Statistical Analysis,* John Wiley & Sons, New York, 1958.

[3] T.W. Anderson, I. Olkin, and L.G. Underhill, Generation of random orthogonal matrices, *SIAM J. Sci. Stat. Comp. 8* (1987), 625–629.

[4] K. Aomoto, Jacobi polynomials associated with Selberg integrals, *SIAM J. Math. Anal. 18* (1987), 545–549.

[5] M. S. Bartlett, On the theory of statistical regression, *Proc. R. Soc. Edinb. 53* (1933), 260–283.

[6] C.M. Bender and S.A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers,* McGraw-Hill, New York, 1978.

[7] P. Billingsley, *Probability and Measure,* John Wiley & Sons, New York, 1979.

[8] C. Bingham, An asymptotic expansion for the distribution of the eigenvalues of a 3 by 3 Wishart matrix, *Ann. Math. Stat. 43* (1972), 1498–1506.

[9] G. Birkhoff and S. Gulati, Isotropic distributions of test matrices, *J. Appl. Math. Physics. 30* (1979), 148–157.

[10] L. Blum and M. Shub, Evaluating rational functions: infinite precision is finite cost and tractable on average, *SIAM J. Comput. 15* (1986), 384–398.

[11] B. Bollobás, *Random Graphs,* Academic Press, London, 1985.

[12] R.B. Boppana, Eigenvalues and graph bisection: an average-case analysis, *Proc. 28th IEEE Symposium on Foundations of Computer Science* (1987), 280–285.

[13] T.A. Brody, J. Flores, J.P. French, P.A. Mello, A. Pandey, and S.S.M. Wong, Random-matrix physics: spectrum and strength fluctuations, *Rev. Modern Physics 53* (1981), 385–479.

[14] J.E. Cohen, H. Kesten, and C.M. Newman, eds., *Random Matrices and Their Applications (Contemporary Mathematics 50)*, American Mathematical Society, Providence, 1986.

[15] J. Demmel, The probability that a numerical analysis problem is difficult, *Mathematics of Computation 50* (1988), 449–480.

[16] A. Edelman, Eigenvalues and condition numbers of random matrices, *SIAM J. Matrix Anal. Appl. 9* (1988), 543–560.

[17] A. Edelman, The distribution and moments of the smallest eigenvalue of a random matrix of Wishart type, submitted to *SIAM J. Sci. Stat. Comp.*

[18] J. Friedman, J. Kahn, and E. Szemerédi, On the second eigenvalue of random regular graphs, *Proc. 21st ACM Symposium on Theory of Computing* (1989).

[19] Z. Füredi and J. Komlós, The eigenvalues of random symmetric matrices, *Combinatorica 1* (1981), 233-241.

[20] S. Geman, A limit theorem for the norm of random matrices, *Ann. Prob. 8* (1980), 252–261.

[21] G.H. Golub and C.F. van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, 1983.

[22] N.R. Goodman, Statistical analysis based on a certain multivariate complex Gaussian distribution, (An introduction), *Ann. Math. Stat. 34* (1963), 152–177.

[23] I.S. Gradshteyn and I.W. Ryzhik, *Table of Integrals, Series, and Products,* 4th ed., Academic Press, New York, 1965.

[24] R.D. Gupta and D.S.P. Richards, Hypergeometric functions of scalar matrix argument are expressible in terms of classical hypergeometric functions, *SIAM J. Math. Anal. 16* (1985), 852–858.

[25] P. Henrici, *Essentials of Numerical Analysis with Pocket Calculator Demonstrations,* John Wiley & Sons, New York, 1982.

[26] A.T. James, Normal multivariate analysis and the orthogonal group, *Ann. Math. Stat. 25* (1954), 40–75.

[27] A.T. James, Distributions of matrix variates and latent roots derived from normal samples, *Ann. Math. Stat. 35* (1964), 475–501.

[28] D. Jonsson, Some limit theorems for the eigenvalues of a sample covariance matrix, *J. Multivariate Anal. 12* (1982), 1–38.

[29] E. Kostlan, Numerical linear algebra and multivariant analysis, never published.

[30] P.R. Krishnaiah and T.C. Cheng, On the exact distributions of the smallest roots of the Wishart matrix using zonal polynomials, *Ann. I. Math. Stat. 23* (1971), 293–295.

[31] P.R. Krishnaiah and T.C. Cheng, On the exact distributions of the extreme roots of the Wishart and MANOVA matrices, *J. Multivariate Anal. 1* (1971), 108–116.

[32] H.S. Leff, Class of ensembles in the statistical theory of energy-level spectra, *J. Math. Phys. 5* (1964), 763–768.

[33] B.D. McKay, The expected eigenvalue distribution of a large regular graph, *Linear Algebra Appl. 40* (1981), 203–216.

[34] V.A. Marcenko and L.A. Pastur, Distributions of eigenvalues for some sets of random matrices, *Math. USSR-Sbornik 1* (1967), 457–483.

[35] M.L. Mehta, *Random Matrices and the Statistical Theory of Energy Levels,* Academic Press, New York, 1967.

[36] R.J. Muirhead, *Aspects of Multivariate Statistical Theory,* John Wiley & Sons, New York, 1982.

[37] A.M. Odlyzko, The $10^{20}$-th zero of the Riemann zeta function and 70 million of its neighbors, to be published.

[38] I. Olkin and S.N. Roy, On multivariate distribution theory, *Ann. Math. Stat. 25* (1954), 329–339.

[39] C.E. Porter, *Statistical Theories of Spectra: Fluctuations,* Academic Press, New York, 1965.

[40] F.J. Schuurmann, P.R. Krishnaiah, and A.K. Chattopadhyay, On the distributions of the ratios of the extreme roots to the trace of the Wishart matrix, *J. Multivariate Anal. 3* (1973), 445–453.

[41] J.W. Silverstein, The smallest eigenvalue of a large-dimensional Wishart matrix, *Ann. Prob. 13* (1985), 1364–1368.

[42] S. Smale, On the efficiency of algorithms of analysis, *Bull. Amer. Math. Soc., 13* (1985), 87–121.

[43] J. Spanier and K.B. Oldham, *An Atlas of Functions,* Hemisphere Publishing, Washington, 1987.

[44] G.W. Stewart, The efficient generation of random orthogonal matrices with an application to condition estimators, *SIAM J. Numer. Anal. 17* (1980), 403–409.

[45] T. Sugiyama, On the distribution of the largest latent root of the covariance matrix, *Annals Math. Stat. 38* (1967), 1148–1151.

[46] G. Szegő, *Orthogonal Polynomials,* American Mathematical Society, Providence, R.I., 1939.

[47] L.N. Trefethen and R.S. Schreiber, Average-case stability of Gaussian elimination, *SIAM J. Matrix Anal. Appl.*, to appear.

[48] H.F. Trotter, Eigenvalue distributions of large Hermitian matrices; Wigner's semi-circle law and a theorem of Kac, Murdock, and Szegő, *Advances in Math. 54* (1984), 67–82.

[49] J. von Neumann, *Collected Works, Vol. 5: Design of Computers, Theory of Automata and Numerical Analysis,* A.H. Taub, ed., Pergamon, New York, 1963. (There are four relevant papers with coauthors Bargmann, Goldstine, and Montgomery. See pages 1–32,421–572.)

[49a] H.H. Goldstine and J. von Neumann, On the principles of large scale computing machines, never published.

[49b] V. Bargmann, D. Montgomery, and J. von Neumann, Solution of linear systems of high order, never published.

[49c] H.H. Goldstine and J. von Neumann, Numerical inverting of matrices of high order, *Amer. Math. Soc. Bull. 53* (1947), 1021–1099.

[49d] H.H. Goldstine and J. von Neumann, Numerical inverting of matrices of high order, II, *Amer. Math. Soc. Proc. 2* (1951), 188–202.

[50] K.W. Wachter, The strong limits of random matrix spectra for sample matrices of independent elements, *Ann. Prob. 6* (1978), 1–18.

[51] N. Weiss, G.W. Wasilkowski, H. Woźniakowski, and M. Shub, Average condition number for solving linear equations, *Linear Algebra 83* (1986), 79–102.

[52] E.P. Wigner, Characteristic vectors of bordered matrices with infinite dimensions, *Ann. Math. 62* (1955), 548–564.

[53] E.P. Wigner, On the distribution of the roots of certain symmetric matrices, *Ann. Math. 67* (1958), 325–327.

[54] E.P. Wigner, Distribution laws for the roots of a random Hermitean matrix, (1962), first published in [39], p.446.

[55] E.P. Wigner, Random matrices in physics, *SIAM Review 9* (1967), 1–23.

[56] J.H. Wilkinson, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, 1965.

[57] J.H. Wilkinson, Error analysis revisted, *Inst. Math. Appl. 22* (1986), 192–200.

[58] S. Wilks, *Mathematical Statistics*, John Wiley & Sons, New York, 1967.

[59] J. Wishart, The generalized product moment distribution in samples from a normal multivariate population, *Biometrika 20A* (1928), 32–43.

[60] S. Wolfram, *Mathematica*, Addison-Wesley Publishing, Redwood City, CA, 1988.

# Index


Abel, N.H., 8
Almost sure convergence, 54
Associated orthogonal polynomials of a random matrix, 81–86

Bargmann, V., 12
  (*See also* von Neumann et al.)
Bartlett, M.S., 72
Bartlett decomposition, 35, 39
Beta function, incomplete, 78
Bidiagonalization of a Gaussian matrix, 38
Birkhoff, G., 2–3, 8, 23
  and Gulati, S., 13, 39

cdf, 7
Change of variables for matrix factorizations, 31, 33
Characteristic polynomial of a random matrix, 83–85
Chi-square distribution, 38
Cholesky factorization, 31
  change of variables for, 33
Complex Gaussian matrix
  condition number of, 69
Complex multivariate gamma function, 33
Complex random matrix
  definition of complex Gaussian, 31
  definition of complex Wishart, 31
  definition of Gaussian Unitary Ensemble, 31
Complex Wishart matrix, 31, 35–36, 60–61, 66, 83, 86
  smallest eigenvalue of, 60–61
  largest eigenvalue of, 66
Condition number, 11
  asymptotic distribution, 23–26, 67–69
  complex random matrix, 69
  definition, 8, 13
  Demmel's, 14, 71–76
  distribution of, 23–26, 67–76
    Demmel's, 14, 71–73
    the logarithm, 27–28
  geometric mean of, 23–24, 67, 69–70
  of a problem, 14
  tails of, 74–76
Conjugate gradient iteration, 2, 8, 14
Covariance matrix, 2, 16, 37

Deciles for the
  condition number, 23
  log of the condition number, 24
Demmel, J., 2, 8, 14–15, 17, 23, 71–76
Density function
  (*See item of interest*)
Determinant of a Gaussian matrix, 39
Differential volume element, 31–33
Discrete distribution, 24, 29, 68
Distribution
  (*See item of interest*)

Eigenvalue decomposition
  change of variables for, 33
EISPACK, 13
Error analysis of Gaussian elimination
  von Neumann estimate, 11
Expected value of
  logarithm
    (*See* Geometric mean – expected logarithm)
  smallest eigenvalue of a Wishart matrix, 22, 77–80, 90–94
Exponential distribution of smallest eigenvalues, 43, 60

Galton, F., 16
Gaussian elimination, 10, 12–14
  growth factor, 12, 15
  von Neumann error estimate, 11
  Wilkinson error bound, 12

Gaussian ensembles
  Gaussian Orthogonal Ensemble
    characteristic polynomial of, 83–85
    definition, 18, 30
    joint eigenvalue density, 36–37
    joint element density, 34–35
    mode of eigenvalue distribution, 85–86
    reduction to tridiagonal form, 39
  Gaussian Unitary Ensemble
    characteristic polynomial of, 83–85
    definition, 30
    joint eigenvalue density, 36–37
    joint element density, 34–35
    mode of eigenvalue distribution, 85–86
  in nuclear physics, 18–19
Gaussian matrices
  condition number of, 67–69
  definition, 30
  determinant of, 39
  factorizations of, 38
Geometric mean – expected logarithm
  of condition numbers, 13, 67, 69–70
  of largest eigenvalue
    of Complex Wishart matrices, 66
    of Wishart matrices, 62, 64–65
  of smallest eigenvalue
    of Complex Wishart matrices, 60–61
    of Wishart matrices, 58–59
GOE
  (*See* Gaussian Orthogonal Ensemble)
Goldstine, H.H.
  (*See* von Neumann et al.)
Golub, G.H., 38
Graph theory, 19
Growth factor, 12, 15
GUE
  (*See* Gaussian Unitary Ensemble)

Haar (invariant) measure, 32
  complex, 33
Hermite polynomial, 81–86
Hypergeometric functions, 77–78

iid, 7, 30–31, 68
Incidence matrix, 19
Incomplete Beta function, 78

Jacobi polynomial, 81–85
Jacobians for matrix factorizations, 31, 33
Joint eigenvalue densities of random matrices, 36–37
Joint element densities of random matrices, 34–35

Kahan, W., 38
Kakutani, S., 3
Kostlan, E., 13
Kummer's equation, 41

Laguerre polynomial, 49, 81–86
Largest eigenvalue of Wishart matrices
  asymptotics of, 62
  complex Wishart matrices, 66
  geometric mean of, 62, 64–65
Law of Large Numbers, 79
Legendre polynomial, 85
Loss of numerical precision, 2, 8, 10–11, 13
LQ decomposition
  change of variables for, 33
  of a Gaussian matrix, 35, 39

MANOVA, 84
Mathematica, 3, 45, 78–79, 87–89, 95–96
MATLAB, 3, 68, 79
Micchelli, C.A., 3
Mode of random matrix eigenvalues, 85–86
Moments of Wishart smallest eigenvalue distributions
  (*See* Smallest eigenvalue of Wishart matrices)
Morton, K.W., 3
Multivariate analysis
  (*See* Multivariate statistics)
Multivariate gamma function, 32
  complex, 33
Multivariate normal distribution, 16
Multivariate statistics
  random matrices in, 16–17, 30, 71

Normal distribution, 30, 34
Nuclear physics
  random matrices in, 18–19
Numerical analysis
  random matrices in, 10–15

Ocneanu, A., 13

Odlyzko, A.M., 19
Olkin, I., 2
Orthogonal invariance, 13, 18, 34
    of Haar measure, 32

pdf, 7
    (*See also item of interest*)
Pfaffian, 40
Principal components analysis, 16
Probability density function
    (*See item of interest*)

QR decomposition
    (*see* LQ decomposition)

Random matrix definitions, 30
Random matrix eigenvalues
    graph theory and, 19
    multivariate statistics and, 16–17
    nuclear physics and, 18–19
    numerical analysis and, 10–15
    orthogonal polynomials and, 81–86
    Riemann zeta function and, 19
Random matrix joint densities
    eigenvalues, 36–37
    elements, 33–34
Rectangular matrix, 31
    Complex, 61, 66, 69
    Gaussian
        condition number of, 68
        largest singular value of, 62, 64–65
        smallest singular value of, 54, 58–59
    LQ decomposition change of variables, 33
Recursion for smallest eigenvalue density of Wishart matrices, 44–49
Riemann zeta function, 19, 86
Ruffini, P., 8

Schreiber, R.S., 15
Semi-circle law, 18, 20
Silverstein, J.W., 54
Smale, S., 2, 8, 13, 23–24
Smallest eigenvalue of Wishart matrices
    asymptotics of, 53–57
    complex Wishart matrices, 60–61
    density of, 40–52, 55, 57
        $n-m=0$, 41–43, 51
        $n-m=1$, 43, 51
        $n-m=2$, 50–51
        $n-m=3$, 50–51
        recursion for, 44–49
        summary of, 51
    geometric mean of, 58–59
    moments of, 45, 77–80
Sphericity test, 17, 71
Stiefel manifold, 32
Strang, G., 3

Tails of condition number distributions, 23, 74–76
Toomre, A., 3
Trefethen, L.N., 3, 15
Tricomi function, 41, 44
    definition, 41

Uniform distribution, 24, 29, 68

Volume
    of Stiefel manifold, 32
    of complex Stiefel manifold, 33
Volume elements, 31–33
von Neumann, J. et al., 2, 8, 10–13, 16, 23–24

Wigner, E.P., 18, 37
    semi-circle law, 18, 20
Wilkinson, J.H., 10, 12-13
Wishart matrix
    characteristic polynomial of, 83–85
    complex
        (*See* Complex Wishart matrix)
    definition, 16, 30
    joint eigenvalue density, 36–37
    joint element density, 34–35
    largest eigenvalue of
        (*See* Largest eigenvalue of Wishart matrices)
    mode of eigenvalue distribution, 85–86
    moments of smallest eigenvalue
        (*See* Smallest eigenvalue of Wishart matrices)
    smallest eigenvalue of
        (*See* Smallest eigenvalue of Wishart matrices)

Zonal polynomials, 17, 40