Topics in random matrices:
Theory and applications to probability and statistics

Termeh Kousha

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements for the degree of
Doctorate of Philosophy in Mathematics [1]

Department of Mathematics and Statistics
Faculty of Science
University of Ottawa

[1]The Ph.D. program is a joint program with Carleton University, administered by the Ottawa-Carleton Institute of Mathematics and Statistics

Abstract

In this thesis, we discuss some topics in random matrix theory which have applications to probability, statistics and quantum information theory.

In Chapter 2, by relying on the spectral properties of an associated adjacency matrix, we find the distribution of the maximum of a Dyck path and show that it has the same distribution function as the unsigned Brownian excursion which was first derived in 1976 by Kennedy [38]. We obtain a large and moderate deviation principle for the law of the maximum of a random Dyck path. Our result extends the results of Chung [10], Kennedy [38] and Khorunzhiy and Marckert [39]. This gave rise to the paper, [41].

In Chapter 3, we discuss a method of sampling called the Gibbs-slice sampler. This method is based on Neal's slice sampling [48] combined with Gibbs sampling.

In Chapter 4, we discuss several examples which have applications in physics and quantum information theory.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisors Dr. Benoît Collins, Dr. David Handelman and Dr. David McDonald for their continuous support of my PhD study and research, for their patience, motivation, enthusiasm and immense knowledge. They were always available and always willing to help. This dissertation could not have been finished without them.

I owe my deepest gratitude to Dr. Victor LeBlanc, for whom a proper "thank you" would be larger than this thesis. I am grateful to him for all the encouragements, support and his positivity throughout my studies.

It is an honor for me to thank Dr. Erhard Neher, who has made available his support in a numerous of ways. I would like to thank him for all the guidance, encouragement, inspiration and help.

I have been most fortunate to have been able to discuss my work with Dr. Rafal Kulik, Dr. Gilles Lamothe and Dr. Ion Nechita. Many thanks also go to all my friends and colleagues for their support and help.

Last but far from least, I thank my family, Mom, Dad, Homa, Gerhard, Neda and Alireza, for their constant support and encouragement. I would like to extend my heartfelt thanks to Masoud, who has always been there for me. My degree, life and this thesis would not have been possible without these amazing people.

Dedication

To Mom, Dad, and Homa.
For their unconditional love, never-ending support and the sense of security they have given when I wanted it most.

Contents

List of Figures

Chapter 1

Introduction

Random matrix theory first gained attention in the 1950s in nuclear physics [68]. It was introduced by Eugene Wigner to model the spectra of heavy atoms. He used random matrices to describe the general properties of the energy levels of highly excited states of heavy nuclei as measured in nuclear reactions. In solid-state physics, random matrices model the behaviour of large disordered Hamiltonians in the mean field approximation.

Since then, random matrix theory has found uses in a wide variety of problems in mathematics, physics and statistics. Wigner, Dyson and Mehta used random matrices theory as a very powerful tool in mathematical physics [43].

In quantum chaos, it has emerged that the statistical properties of many quantum systems can be modeled by random matrices. Bohigas, Giannoni and Schmit [6] discovered the connection between random matrix theory and quantum transport. His conjecture asserts that the spectral statistics of quantum systems whose classical counterparts exhibit chaotic behaviour are described by random matrix theory. Random matrix theory has also found applications to quantum gravity in two dimensions, mesoscopic physics, and more. Since quantum mechanics deals with the noncommutative algebras, the random objects of study are presented by matrices.

Via random interactions, generic loss of coherence of a fixed central system coupled to a quantum-chaotic environment is represented by a random matrix ensemble.

In multivariate statistics, random matrices were introduced by John Wishart [69], for statistical analysis of large samples. Since then, random matrix theory become a major tool in wireless communications (Tulino and Verdu, [63]) and in multivariate statistical analysis (Johnstone, [36]).

In numerical analysis, random matrix theory was used by von Neumann and Goldstine [67], to describe computation errors in operations such as matrix multiplication. Other applications can be found in random tilings, operator algebras, free probability, quantitative finance in econophysics, and current research even suggests it could have applications in improving web search engines.

One of the interesting topics in random matrix theory is asymptotic distribution of eigenvalues of random matrices. Wigner discussed the general conditions of validity for his famous semicircle law for the distribution of eigenvalues of random matrices. Let $A = [a_{ij}]_{N\times N}$ be a standard self-adjoint Gaussian matrix. By Wigner's semicircle law [51], if A_N is a self-adjoint Gaussian $N \times N$-random matrix, then A_N converges, for $N \to \infty$, in distribution towards a semicircular element s,

$$A_N \xrightarrow{d} s,$$

i.e.,

$$\lim_{N\to\infty} \mathrm{tr} \otimes E(A_N^k) = \frac{1}{2\pi}\int_{-2}^{2} t^k\sqrt{4-t^2}dt \quad \forall k \in \mathbb{N},$$

where tr denotes a normalized trace and

$$(\mathrm{tr} \otimes E)(A) = E[\mathrm{tr}(A)] = \frac{1}{N}\sum_{i=1}^{N} E[a_{ii}].$$

For each $k \in \mathbb{N}$, let $m(k)$ denote the mth moment of $A_{N\times N}$. It was shown [51] that for all $k \in \mathbb{N}$ we have,

$$m(k) := \frac{1}{2\pi}\int_{-2}^{2} t^k\sqrt{4-t^2}dt = \begin{cases} 0, & \text{if } k \text{ is odd;} \\ C_p, & \text{if } k \text{ is even, } k = 2p, \end{cases}$$

where C_p is the pth Catalan number, i.e.,

$$C_p = \frac{1}{p+1}\binom{2p}{p}.$$

Catalan number, C_p also gives the number of Dyck paths in $2p$ steps. A Dyck path of length $2p$ is a path in $\mathbb{Z}^2$ starting from $(0,0)$ and ending at $(2p,0)$ with increments of the form $(1,+1)$ or $(1,-1)$ and staying above the real axis. For any integer $N>0$, define the set of Dyck paths with length of $2N$ as follows,

$$D_{2N} = \{S := (S_i)_{0\le i\le 2N} : S_0 = S_{2N} = 0, S_{i+1} = S_i \pm 1, S_i \ge 0 \ \forall\ i \in [0, 2N-1]\}.$$

The first motivation for this thesis was the construction of a deterministic matrix that has the same moments as a standard Gaussian random matrix $A_{N\times N}$. Let $T_{N\times N}$ be the adjacency matrix associated to a simple random walk with N states. In Chapter 2, after appropriately weighting the eigenvalues of matrix $T_{N\times N}$, we show that $T_{N\times N}$ has the same moments as the semi-circle random variable. This yields a representation of Catalan numbers. Let $D_{2N,n} = \{S : S \in D_{2N}, \max_{1\le i\le 2N} S_i < n\}$. Note that if $N < n$ then $D_{2N,n} = D_{2N}$. We show

$$|D_{2N,n}| = \sum_{s=1}^{n} \left(\frac{2}{n+1}\right) \sin^2\left(\frac{\pi s}{n+1}\right) \left(2\cos\left(\frac{\pi s}{n+1}\right)\right)^{2N},$$

where $|x|$ is the cardinality of x. This is based on fundamental observation that

$$|D_{2N,n}| = (T^{2N})_{11},$$

which follows directly from the definition of $D_{2N,n}$ and properties of adjacency graphs. For positive integers $n > 0$ and $0 < N \le n-1$, the Nth Catalan number satisfies $C_N = |D_{2N}| = |D_{2N,n}|$.

Applying this, we find the distribution of the maximum of a Dyck path for the case where the length of the Dyck path is proportional to the square root of the height. Also we considered two rare event cases and we find a large and moderate deviation principles for these case. This gave rise to the following submitted preprint, [41].

An interesting question that arises in the random matrices theory or quantum information theory, is whether one can generate random vectors, or random matrices from a given distribution function. When the target distribution comes from a standard parametric family, plenty of software exists to generate random variables. Chapter 3 introduces a method of sampling in order to be able to sample random vectors, or random matrices, from difficult target distributions, e.g., when f can be calculated, at least up to a proportionally constant, but f cannot be sampled. We discuss a method of sampling called Gibbs-slice sampler. This method is based on Neal's slice sampling [48] which is combined by the idea of Gibbs sampling to sample large random matrices from a distribution function which does not come from a standard parametric family.

In Chapter 4, we apply this method to several examples having applications in quantum information theory. All cases discussed in Chapter 4 are unimodal distribution functions. In the last example we use the idea of the Gibbs-slice sampling to generate hermitian positive definite matrices with operator norm of max 1. We also expand this to the case where simulated matrices have a fixed unit trace, which are known as density matrices. Density matrices are used in quantum theory to describe the statistical state of a quantum system. These matrices were first introduced by von Neumann in 1927 and are useful for describing and performing calculations with a mixed state, which is a statistical ensemble of several quantum states.

A future goal would be to extend the idea underlying the Gibbs-slice sampler to more complicated samplings of random matrices; example would include those with non-unimodal target density functions, or those with disconnected horizontal slices, as well as sampling from convex sets of random matrices with more complicated properties where no method is currently available.

Chapter 2

Asymptotic behaviour of a random Dyck path

2.1 Introduction

Let $W(t),\ 0 \leq t < \infty$ denote a standard Brownian motion and define
$\tau_1 = \sup\{t < 1 : W(t) = 0\}$ and $\tau_2 = \inf\{t > 1 : W(t) = 0\}$. Define the process $W_1(s),\ 0 \leq s \leq 1$ by setting

$$W_1(s) = \frac{|W(s\tau_2 + (1-s)\tau_1)|}{(\tau_2 - \tau_1)^{\frac{1}{2}}}.$$

The process W_1 is known as the *unsigned, scaled Brownian excursion process* [38]. Brownian excursion process is actually a Brownian motion conditioned to be positive and tied down at 0 and 1, i.e., it is a Brownian bridge process conditioned to be positive. These definitions require care, since they are derived by conditioning on events of probability zero. Therefore, it is known that they can precisely defined as limits, see for instance [15].

Chung [10] and Kennedy [38] derived the distribution of the maximum of the unsigned scaled Brownian excursion. They also proved that the maximum of Brownian excursion and the range of Brownian bridge have the same distribution. Vervaat [64]

showed that Brownian excursion is equal in distribution to Brownian bridge with the origin placed at its absolute minimum, explaining the results of Chung [10] and Kennedy [38].

For a path in the lattice $\mathbb{Z}^2$, the *NE-SE path* is a path which starts at $(0,0)$ and makes steps either of the form $(1,1)$ (north-east steps) or of the form $(1,-1)$ (south-east steps). A Dyck path is a NE-SE path which ends on the $x-$axis and never goes below it. It is known that after normalization, a Dyck path converges to W_1 in distribution [37]. For any integer $N > 0$, define the set of Dyck paths with length of $2N$,

$$D_{2N} = \{S := (S_i)_{0\leq i\leq 2N} : S_0 = S_{2N} = 0, S_{i+1} = S_i \pm 1, S_i \geq 0 \ \forall \ i \in [0, 2N-1]\}.$$

Dyck paths play a significant role in combinatorics. Dyck paths can be counted by the Catalan numbers. Khorunzhiy and Marckert [39] showed that for any $\lambda > 0$,

$$E\left(\exp(\lambda(2N^{-\frac{1}{2}} \max_{1\leq i\leq 2N} S_i))\right)$$

converges and coincides with the moment generating functions of the maximum of the normalized Brownian excursion on $[0,1]$.

In this chapter, by relying on the spectral properties of an associated adjacency matrix, we find the distribution of the maximum of a Dyck path and show that it has the same distribution function as the unsigned Brownian excursion; this was first derived in 1976 by Kennedy [38]. We also obtain large and moderate deviation principles for the law of the maximum of a random Dyck path, which gave rise to the following submitted preprint, [41].

This chapter is organized as follows. We begin with basic definitions and theorems in Section 2.

In Section 3, we consider an adjacency matrix of a simple random walk, $T_{n\times n}$, and by applying weights to its eigenvalues, we show that it has the same moments as the semi-circle random variable. This yields us to a representation of a Catalan number as follows. Let $D_{2N,n} = \{S : S \in D_{2N}, \max_{1\leq i\leq 2N} S_i < n\}$. If $N < n$ then $D_{2N,n} = D_{2N}$.

We show

$$|D_{2N,n}| = \sum_{s=1}^{n} \left(\frac{2}{n+1}\right) \sin^2\left(\frac{\pi s}{n+1}\right) \left(2\cos\left(\frac{\pi s}{n+1}\right)\right)^{2N},$$

where $|x|$ is the cardinality of x. This derives from the fundamental observation that

$$|D_{2N,n}| = (T^{2N})_{11},$$

this follows directly from the definition of $D_{2N,n}$ and properties of adjacency graphs. For positive integers $n > 0$ and $0 < N \leq n-1$, the Nth Catalan number satisfies $C_N = |D_{2N}| = |D_{2N,n}|$.

In Section 4, by applying this representation of Catalan number, we find the distribution of the maximum of a of the Dyck path for the case where the length of the Dyck path is proportional to the square root of the height. Let $\mathbb{P}_N$ be the uniform distribution on D_{2N}. Then

$$\frac{|D_{2N,n}|}{C_N} = \mathbb{P}_N(\text{max height of the Dyck paths in } 2N \text{ steps } < n).$$

Let $[x]$ be the largest integer less than or equal to x. The main theorem in this section is:

Theorem 2.1.1. *Let $N = [tn^2]$ where t is any positive number. Then*

$$\lim_{n\to\infty} \frac{|D_{2N,n}|}{C_N} = f(t),$$

where $f(t) = 4\sqrt{\pi}t^{\frac{3}{2}} \sum_{s=1}^{\infty} s^2\pi^2 \exp\left(-ts^2\pi^2\right)$.

Let K denote the function used by Kennedy [38] and Chung [10] as the distribution of the maximum of the unsigned scaled Brownian excursion,

$$K(x) := \mathbb{P}\left(\max_{s\in[0,1]} W_1(s) \leq x\right) = 1 - 2\sum_{s=1}^{\infty} \left(4x^2s^2 - 1\right) \exp\left(-2x^2s^2\right), \quad \text{for } x > 0.$$

We show that for every $x > 0$, $f(x) = K(x)$, where f is the function defined in Theorem 2.1.1. We also consider two other cases wherein the length of the Dyck path

is greater than or less than the square root of its height.

$$\lim_{n\to\infty} \frac{|D_{2N,n}|}{C_N} = \begin{cases} 0, & N \gg n^2; \\ 1, & n \ll N \ll n^2. \end{cases}$$

In Section 5, we discuss two rare events, and we find moderate and large deviation principles for the law of the maximum of a random Dyck path for those cases. We have two main theorems.

Theorem 2.1.2. *Let N and n be positive integers satisfying $N \gg n^2$. Then*

$$\lim_{N\to\infty} \frac{(n+1)^2}{N} \log \mathbb{P}_N(\textit{max height of the Dyck Path with length } \ 2N < n) \longrightarrow -\pi^2.$$

Theorem 2.1.3. *Let N be any positive integer and $x > 0$,*

- *If $n \ll N \ll n^2$, then*

  $$\lim_{N\to\infty} \frac{N}{2n^2} \log \mathbb{P}_N \left(\textit{max height of a Dyck path with length } 2N > xn\right) \to -x^2.$$

- *If $n \sim 2N$, then*

  $$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{P}_N \left(\textit{max height of a Dyck path with length } 2N > xn\right) \to h(x),$$

  for $0 < x \leq \frac{1}{2}^-$ where

  $$h(x) = -(1+2x)\log(1+2x) - (1-2x)\log(1-2x).$$

  Otherwise,

  $$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{P}_N \left(\textit{max height of a Dyck path with length } 2N > xn\right) \to -\infty.$$

Remark 1. *Note that $h(x)$ is the entropy function of the Bernoulli distribution. Let $p = 1 + 2x$ and $q = 1 - 2x$. Then for all $0 \leq x \leq \frac{1}{2}$, $0 \leq p \leq 1$ and $p = 1 - q$. So we can rewrite $h(x)$ in the previous theorem as follows,*

$$H(p, q) = -p \log p - q \log q.$$

In general, entropy is a measure of unexpectedness. For example when a fair coin is flipped, the outcome is either heads or tails, and there is no way to predict which. So the fair coin has maximum entropy. If X is a discrete random variable with the following distribution,

$$\mathbb{P}(X = x_k) = p_k \quad \text{for } k = 1, 2, \ldots,$$

then we define the entropy of X as follows,

$$H(X) = -\sum_{k \geq 1} p_k \log p_k.$$

Remark 2. *In Theorem* 2.1.3 *if we assume $x = 1$, we have for $n \ll N \ll n^2$*

$$\lim_{N \to \infty} \frac{N}{n^2} \log \mathbb{P}_N \left(\textit{max height of a Dyck path with length } 2N > n \right) \to -2.$$

2.2 Definitions

Noncommutative probability space

Definition 1. *[51] A* ***noncommutative probability space*** $(\mathcal{A}, \varphi)$ *consists of a unital algebra* $\mathcal{A}$ *over* $\mathbb{C}$ *and a unital linear functional*

$$\varphi : \mathcal{A} \to \mathbb{C}, \qquad \varphi(1_{\mathcal{A}}) = 1.$$

The elements $a \in \mathcal{A}$ *are called non-commutative random variables in* $(\mathcal{A}, \varphi)$*. We suppose that* $\mathcal{A}$ *is a* $*$*-algebra, i.e., that* $\mathcal{A}$ *is also endowed with antilinear* $*$*-operation* $\mathcal{A} \ni a \to a^* \in \mathcal{A}$*, such that* $(a^*)^* = a$ *and* $(ab)^* = b^*a^*$ *for all* $a, b \in \mathcal{A}$*. If we have*

$$\varphi(a^*a) \geq 0, \ \forall \ a \in \mathcal{A},$$

then we say that the functional φ *is positive and we will call* $(\mathcal{A}, \varphi)$ *a* $*$***-probability space***.

Definition 2. *For a random variable* $a \in \mathcal{A}$ *let* a^* *denote the conjugate transpose of* a*, then we say* a *is*

- *a* ***selfadjoint*** *random variable if* $a = a^*$*;*
- *a* ***unitary*** *random variable if* $aa^* = a^*a = 1$*;*
- *a* ***normal*** *random variable if* $aa^* = a^*a$*.*

Definition 3. *[51] Let* $(\mathcal{A}, \varphi)$ *be a* $*$*-probability space, and let* a *be a selfadjoint element* $\mathcal{A}$*. The* ***moments*** *of* a *are the numbers* $\varphi(a^k), \ k \geq 0$*.*

Definition 4. *[51] Let* $(\mathcal{A}, \varphi)$ *be a* $*$*-probability space, let* x *be a selfadjoint element of* $\mathcal{A}$ *and let* r *be a positive number. If* x *has density equal to*

$$\frac{2}{\pi r^2}\sqrt{r^2 - t^2}dt$$

on the interval $[-r, r]$*, then we will say that* x *is a* ***semicircular element*** *of radius* r*. The semicircular elements of radius 2 are called* ***standard semicircular*** *(they are normalized with respect to variance).*

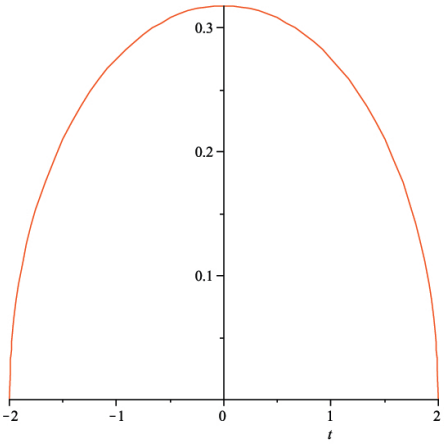


Figure 2.1: Distribution of a semicircular element of radius 2

Definition 5. *[51] Let* $(\mathcal{A}, \varphi)$ *be a noncommutative probability space and let* I *be a fixed index set.*

1. *Unital subalgebras* $(\mathcal{A}_i)_{i\in I}$ *are called* ***tensor independent****, if the subalgebras* $\mathcal{A}_i$ *commute (i.e.* $ab = ba$ *for all* $a \in \mathcal{A}_i$ *and all* $b \in \mathcal{A}_j$ *and all* $i, j \in I$ *with* $i \neq j$*) and if* φ *factorizes as:*

$$\varphi\left(\Pi_{j\in J} a_j\right) = \Pi_{j\in J}\varphi\left(a_j\right)$$

   for all finite subsets $J \subset I$ *and all* $a_j \in \mathcal{A}_j$ $(j \in J)$*.*

2. *For each* $i \in I$*, let* $\mathcal{A}_i \subset \mathcal{A}$ *be a unital subalgebras* $(\mathcal{A}_i)_{i\in I}$ *are called* ***freely independent****, if*

$$\varphi(a_1 \dots a_k) = 0$$

   whenever we have the following:

- k *is a positive integer;*
- $a_j \in \mathcal{A}_{i(j)}$ $(i(j) \in I)$ *for all* $j = 1, \ldots, k$*;*
- $\varphi(a_j) = 0$ *for all* $j = 1, \ldots, k$*;*
- *and neighbouring elements are from different subalgebras, i.e.*

$$i(1) \neq i(2), i(2) \neq i(3), \ldots, i(k-1) \neq i(k).$$

Definition 6. *A* ***self-adjoint matrix*** *(or hermitian) is a square matrix with complex entries, and is equal to its own conjugate transpose; that is, the element in the* i*th row and the* j*th column is equal to the complex conjugate of the element in the* j*th row and the* i*th column: for all indices* i *and* j*,* $a_{i,j} = \overline{a_{j,i}}$*. The conjugate transpose of matrix* A *is denoted* A^**. A unitary matrix is an* n *by* n *matrix* U *satisfying*

$$U^*U = UU^* = I_n,$$

where I_n *is the identity matrix. This says that a matrix* U *is unitary if and only if it has an inverse which is equal to its conjugate transpose.*

Definition 7. *[51] Let* $(\mathcal{A}_\mathcal{N}, \varphi_N)$ $(N \in \mathbb{N})$ *and* $(\mathcal{A}, \varphi)$ *be non-commutative probability spaces and consider random variables* $a_N \in \mathcal{A}_\mathcal{N}$ *for each* $N \in \mathbb{N}$ *and* $a \in \mathcal{A}$*. We say that* a_N ***converges in distribution*** *towards* a *for* $N \to \infty$*, and denote this by*

$$a_N \xrightarrow{d} a,$$

if

$$\lim_{N\to\infty} \varphi_N(a_N^n) = \varphi(a^n) \quad \forall n \in \mathbb{N}.$$

Definition 8. *Let* μ *be a probability measure on* $\mathbb{R}$ *with moments*

$$m_n := \int_\mathbb{R} t^n d\mu(t).$$

We say that μ *is* ***determined by its moments****, if* μ *is the only probability measure on* $\mathbb{R}$ *with these moments.*

Theorem 2.2.1. *(**Classical central limit theorem[51]**) Let $(\mathcal{A}, \varphi)$ be a $*$-probability space and $a_1, a_2, \ldots \in \mathcal{A}$ a sequence of independent and identically distributed selfadjoint random variables. Furthermore, assume that all variables are centred, $\varphi(a_r) = 0$ $(r \in \mathbb{N})$, and denote by $\sigma^2 := \varphi(a_r^2)$ the common variance. Then*

$$\frac{a_1 + \ldots + a_N}{\sqrt{N}} \xrightarrow{d} x,$$

where x is a normally distributed random variable of variance σ^2, i.e.,

$$\lim_{N\to\infty} \varphi\left(\left(\frac{a_1 + \ldots + a_N}{\sqrt{N}}\right)^n\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} t^n e^{-t^2/2\sigma^2} dt \quad \forall n \in \mathbb{N}.$$

Theorem 2.2.2. *(**Free central limit theorem [51]**) Let $(\mathcal{A}, \varphi)$ be a $*$-probability space and $a_1, a_2, \ldots \in \mathcal{A}$ a sequence of freely independent and identically distributed selfadjoint random variables. Furthermore, assume that all variables are centered, $\varphi(a_r) = 0$ $(r \in \mathbb{N})$, and denote by $\sigma^2 := \varphi(a_r^2)$ the common variance. Then we have*

$$\frac{a_1 + \ldots + a_N}{\sqrt{N}} \xrightarrow{d} s,$$

where s is a semicircular element of variance σ^2.

Random matrices, their moments and Wigner's semicircle law

Random matrices are matrices whose entries are classical random variables.

Definition 9. *[51] A $*$-probability space $N \times N$ **random matrices** is given by $(M_N(L^{\infty-}(\Omega, P)), tr \otimes E)$, where (Ω, P) is a classical probability space,*

$$L^{\infty-}(\Omega, P) := \bigcap_{1\le p<\infty} L^p(\Omega, P),$$

$M_N(\mathcal{A})$ denotes the algebra of $N \times N$ matrices with entries from $\mathcal{A}$, E denotes the expectation with respect to P and tr denotes the normalized trace on $M_N(\mathbb{C})$, i.e.,

$$tr(a) = \frac{1}{N} \sum_{i=1}^{N} \alpha_{ii} \quad for \quad a = (\alpha_{ij})_{i,j=1}^{N} \in M_N(\mathbb{C}).$$

More concretely, this means elements in the probability space are of the form

$$A = (\alpha_{ij})_{i,j=1}^{N}, \quad \textit{with} \quad \alpha_{ij} \in (L^{\infty-}(\Omega, P))$$

and

$$(tr \otimes E)(A) = E[tr(A)] = \frac{1}{N}\sum_{i=1}^{N} E[a_{ii}].$$

Definition 10. *[51] A* ***self-adjoint Gaussian random matrix*** *is a $N \times N$ random matrix $A = (\alpha_{ij})$ with $A^* = A$ such that the entries α_{ij} $(i, j = 1, \ldots, N)$ form a complex Gaussian family which is determined by the covariance*

$$E[\alpha_{ij}\alpha_{kl}] = \frac{1}{N}\delta_{il}\delta_{jk} \qquad (i, j, k, l = 1, \ldots N).$$

Theorem 2.2.3. *(**Wigner's semicircle law [51]**) For each $N \in \mathbb{N}$, let A_N be a selfadjoint Gaussian $N \times N$-random matrix. Then A_N converges, for $N \to \infty$, in distribution towards a semicircular element s,*

$$A_N \xrightarrow{d} s,$$

i.e.,

$$\lim_{N\to\infty} tr \otimes E(A_N^k) = \frac{1}{2\pi}\int_{-2}^{2} t^k\sqrt{4-t^2}dt \quad a.s. \quad \forall k \in \mathbb{N}.$$

Lemma 2.2.1. *[51] For all $k \in \mathbb{N}$ we have,*

$$s(k) := \frac{1}{2\pi}\int_{-2}^{2} t^k\sqrt{4-t^2}dt = \begin{cases} 0, & \textit{if } k \textit{ is odd;} \\ C_p, & \textit{if } k \textit{ is even, } k = 2p, \end{cases}$$

where C_p is the pth Catalan number, i.e.,

$$C_p = \frac{1}{p+1}\binom{2p}{p}.$$

Proof: The function odd, so obviously $s(k) = 0$, $\forall$ odd k. Assume $k = 2p$ and apply the following change of variable,

$$t = 2\cos\theta, dt = -2\sin\theta d\theta,$$

so we get,

$$\begin{aligned}
\frac{1}{2\pi}\int_{-2}^{2} t^k\sqrt{4-t^2}dt &= -\frac{1}{2\pi}\int_{\pi}^{0} 2^{2p+2}\cos^{2p}\theta\sin^2\theta d\theta \\
&= \frac{1}{2\pi}4^{p+1}\left(\int_0^{\pi}\cos^{2p}d\theta - \int_0^{\pi}\cos^{2(p+1)}d\theta\right) \\
&= \frac{1}{2\pi}4^{p+1}\left(\frac{\pi}{4^p}\binom{2p}{p} - \frac{\pi}{4^{p+1}}\binom{2(p+1)}{p+1}\right) \\
&= \frac{1}{2}\left(4\binom{2p}{p} - \frac{2(2p+1)}{p+1}\binom{2p}{p}\right) \\
&= \frac{1}{p+1}\binom{2p}{p}.
\end{aligned}$$

So from Lemma 2.2.1 and Wigner's law, it follows that if A_N is a selfadjoint Gaussian $N \times N$-random matrix, then

$$\lim_{N\to\infty} \mathrm{tr}\otimes E(A_N^k) = \begin{cases} 0, & \text{if } k \text{ is odd;} \\ C_p, & \text{if } k \text{ is even, } k = 2p. \end{cases}$$

This means that the odd moments of a Gaussian random matrix are zero and the even moments converge to the corresponding Catalan number.

2.3 Constructing a matrix with the same moments as the semi-circle random variable

We shall start with some basic definitions. A *path* is a non-empty graph $P = (V, E)$ of the form $V = \{x_0, x_1, \cdots, x_k\}$ and $E = \{x_0x_1, x_1x_2, \cdots, x_{k-1}x_k\}$ where x_i are distinct. Consider a path P_n with n vectors, as the picture shown below,

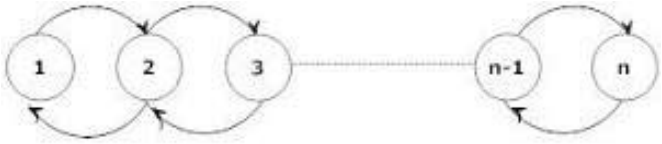


Figure 2.2: A path P_n with n vectors

The adjacency matrix $T = (t_{ij})_{n\times n}$ is an n by n matrix with 1 in t_{ij} if i is connected to j and 0 for the others, i.e.,

$$T := \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ 0 & 1 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}$$

Lemma 2.3.1. *The eigenvalues of T are $\{2\cos(\frac{\pi k}{n+1}); 1 \leq k \leq n\}$ and the eigenvector*

for the lth eigenvalue is $\begin{pmatrix} \sin(\frac{\pi}{n+1}l) \\ \sin(\frac{2\pi}{n+1}l) \\ \vdots \\ \sin(\frac{n\pi}{n+1}l) \end{pmatrix}$, *that is, for each* $1 \leq l \leq n$, *we have*

$$\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ 0 & 1 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \sin(\frac{\pi}{n+1}l) \\ \sin(\frac{2\pi}{n+1}l) \\ \vdots \\ \vdots \\ \sin(\frac{n\pi}{n+1}l) \end{pmatrix} = 2\cos\left(\frac{\pi l}{n+1}\right) \begin{pmatrix} \sin(\frac{\pi}{n+1}l) \\ \sin(\frac{2\pi}{n+1}l) \\ \vdots \\ \vdots \\ \sin(\frac{n\pi}{n+1}l) \end{pmatrix}$$

Proof: We have

$$\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ 0 & 1 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \sin(\frac{\pi}{n+1}l) \\ \sin(\frac{2\pi}{n+1}l) \\ \vdots \\ \vdots \\ \sin(\frac{n\pi}{n+1}l) \end{pmatrix} = \begin{pmatrix} \sin(\frac{2\pi}{n+1}l) \\ \sin(\frac{\pi}{n+1}l) + \sin(\frac{3\pi}{n+1}l) \\ \vdots \\ \sin(\frac{(n-2)\pi}{n+1}l) + \sin(\frac{n\pi}{n+1}l) \\ \sin(\frac{(n-1)\pi}{n+1}l) \end{pmatrix}.$$

We also have, for all x and y,

$$2\sin(x)\cos(y) = \sin(x+y) - \sin(x-y).$$

The result follows, since the right hand side is equal to

$$2\cos\left(\frac{\pi l}{n+1}\right)\begin{pmatrix}\sin(\frac{\pi}{n+1}l)\\ \sin(\frac{2\pi}{n+1}l)\\ \vdots\\ \sin(\frac{(n-1)\pi}{n+1}l)\\ \sin(\frac{n\pi}{n+1}l)\end{pmatrix} = \begin{pmatrix}2\sin(\frac{\pi}{n+1}l)\cos\left(\frac{\pi l}{n+1}\right)\\ 2\sin(\frac{2\pi}{n+1}l)\cos\left(\frac{\pi l}{n+1}\right)\\ \vdots\\ 2\sin(\frac{(n-1)\pi}{n+1}l)\cos\left(\frac{\pi l}{n+1}\right)\\ 2\sin(\frac{n\pi}{n+1}l)\cos\left(\frac{\pi l}{n+1}\right)\end{pmatrix}$$

$$= \begin{pmatrix}\sin(\frac{2\pi}{n+1}l)\\ \sin(\frac{\pi}{n+1}l)+\sin(\frac{3\pi}{n+1}l)\\ \vdots\\ \sin(\frac{(n-2)\pi}{n+1}l)+\sin(\frac{n\pi}{n+1}l)\\ \sin(\frac{(n-1)\pi}{n+1}l)\end{pmatrix}.$$

In this section, after weighting the eigenvalues of the matrix T appropriately, we show that this matrix has the same moments as the semi-circle random variable. Note that the kth power of the first entry of T, $(T^k)_{11}$, denotes the number of paths starting from 1, return to 1 in k steps. Since we jump one step at a time, starting at any state and returning to it requires an even number of steps. It is known that for $1 \leq k \leq 2(n-1)$, this number is equal to $C_{\frac{k}{2}}$, where $C_k = \frac{1}{k+1}\binom{2k}{k}$ is the kth Catalan number. Therefore

$$(T^k)_{11} = \begin{cases} 0, & \text{if } k \text{ is odd;} \\ C_{\frac{k}{2}}, & \text{if } k \text{ is even.} \end{cases}$$

Lemma 2.3.2. *Let A be an $n \times n$ hermitian matrix. Let $e_1 \geq \cdots \geq e_n$ be its real eigenvalues. Then there exist $\lambda_1, \cdots, \lambda_n \geq 0$ such that $\sum_{i=1}^n \lambda_i = 1$, and for all $k \in \mathbb{N}$*

$$\left(A^k\right)_{11} = \sum_{i=1}^n \lambda_i e_i^k.$$

Proof: By the spectral theorem, there exists a unitary matrix U such that $A =$

UDU^*, where $D = \begin{pmatrix} e_1 & 0 & \cdots & 0 \\ 0 & e_2 & 0 & \vdots \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & e_n \end{pmatrix}$.

Therefore,

$$A^k = UD^kU^* = U \begin{pmatrix} e_1^k & 0 & \cdots & 0 \\ 0 & e_2^k & 0 & \vdots \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & e_n^k \end{pmatrix} U^*.$$

We write $\left(A^k\right)_{11} = \text{trace}\left[\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 \end{pmatrix} A^k\right]$,

since,

$$\text{trace}\left[\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 \end{pmatrix} A^k\right] = \text{trace}\begin{pmatrix} a_{11}^k & a_{12}^k & \cdots & a_{1n}^k \\ 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 \end{pmatrix} = a_{11}^k = \left(A^k\right)_{11}.$$

Therefore,

$$\begin{aligned} \left(A^k\right)_{11} &= \text{trace}\left[\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 \end{pmatrix} U \begin{pmatrix} e_1^k & 0 & \cdots & 0 \\ 0 & e_2^k & 0 & \vdots \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & e_n^k \end{pmatrix} U^*\right], \\ &= \text{trace}\left[U^* \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 \end{pmatrix} U \begin{pmatrix} e_1^k & 0 & \cdots & 0 \\ 0 & e_2^k & 0 & \vdots \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & e_n^k \end{pmatrix}\right]. \end{aligned}$$

Let $P = (p_{ij}) = U^* \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 \end{pmatrix} U$. Then

$$\begin{aligned} \left(A^k\right)_{11} &= \text{trace}\left[P \begin{pmatrix} e_1^k & 0 & \cdots & 0 \\ 0 & e_2^k & 0 & \vdots \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & e_n^k \end{pmatrix}\right], \\ &= \text{trace}\left[\begin{pmatrix} p_{11}e_1^k & p_{12}e_2^k & \cdots & p_{1n}e_n^k \\ p_{21}e_1^k & p_{22}e_2^k & \cdots & p_{2n}e_n^k \\ \vdots & \vdots & \ddots & \vdots \\ P_{n1}e_1^k & p_{2n}e_2^k & \cdots & p_{nn}e_n^k \end{pmatrix}\right], \\ &= \sum_{i=1}^{n} p_{ii}e_i^k. \end{aligned}$$

Now for each $0 \leq i \leq n$, let $\lambda_i = p_{ii}$. Therefore, we have

$$\begin{aligned} \sum_{i=1}^{n} \lambda_i &= \sum_{i=1}^{n} p_{ii} \\ &= \text{trace}[P] \\ &= 1, \end{aligned}$$

since P is a rank one idempotent. Moreover P is positive semidefinite; since P is unitarily equivalent to a rank one projector, it is a rank one projector and therefore is positive semidefinite.

Remark 3. *We can rewrite the matrix P as follows,*

$$\begin{aligned} P &= U^* \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 \end{pmatrix} U \\ &= U^* \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix} U \\ &= \begin{pmatrix} \overline{U_{11}} \\ \vdots \\ \vdots \\ \overline{U_{1n}} \end{pmatrix} \begin{pmatrix} U_{11} & \cdots & \cdots & U_{1n} \end{pmatrix}. \end{aligned}$$

Therefore, $\lambda_i = U_{1i}\overline{U_{1i}}$.

Let V be the matrix whose columns are eigenvectors of T, i.e.,

$$V = (v_{ij})_{n\times n} = \begin{pmatrix} \sin(\frac{\pi}{n+1}) & \sin(\frac{2\pi}{n+1}) & \cdots & \sin(\frac{n\pi}{n+1}) \\ \sin(\frac{2\pi}{n+1}) & \cdots & \cdots & \sin(\frac{2n\pi}{n+1}) \\ \vdots & \cdots & \cdots & \vdots \\ \sin(\frac{n\pi}{n+1}) & \cdots & \cdots & \sin(\frac{n^2\pi}{n+1}) \end{pmatrix}$$

Lemma 2.3.3. *Let $V' = \delta_n V$, where $\delta_n = \sqrt{\frac{2}{n+1}}$. Then V' is unitary and symmetric.*

Proof: The symmetry of V' follows from the symmetry of V. To show that V' is unitary we just need to show

$$(V')(V')^T = I_n,$$

since the entries of V' are real. By symmetry we have $(V')(V')^T = (V')^2$, therefore we have to show

$$\left[\sqrt{\frac{2}{n+1}}\begin{pmatrix} \sin(\frac{\pi}{n+1}) & \sin(\frac{2\pi}{n+1}) & \cdots & \sin(\frac{n\pi}{n+1}) \\ \sin(\frac{2\pi}{n+1}) & \cdots & \cdots & \sin(\frac{2n\pi}{n+1}) \\ \vdots & \cdots & \cdots & \vdots \\ \sin(\frac{n\pi}{n+1}) & \cdots & \cdots & \sin(\frac{n^2\pi}{n+1}) \end{pmatrix}\right]^2 = I_n.$$

So we have to show for all $1 \leq j, z \leq n$

$$\sum_{s=1}^{n} \sin^2\left(\frac{js\pi}{n+1}\right) = \frac{n+1}{2}, \quad \text{and}$$

$$\sum_{s=1}^{n} \sin\left(\frac{js\pi}{n+1}\right)\sin\left(\frac{zs\pi}{n+1}\right) = 0.$$

Using the fact that, $\sin^2(\theta) = \frac{1}{2}(1-\cos(2\theta))$ we get,

$$\sum_{s=1}^{n} \sin^2\left(\frac{js\pi}{n+1}\right) = \frac{1}{2}\left(n - \sum_{s=1}^{n}\cos\left(\frac{2js\pi}{n+1}\right)\right).$$

We also know that $\cos(\theta)$ is a real part of $e^{i\theta}$. We have,

$$\begin{aligned} \sum_{s=1}^{n}\left(e^{\frac{2j\pi i}{n+1}}\right)^s &= e^{\frac{2j\pi i}{n+1}} + \left(e^{\frac{2j\pi i}{n+1}}\right)^2 + \cdots + \left(e^{\frac{2j\pi i}{n+1}}\right)^n \\ &= \frac{1-e^{\frac{2j\pi i}{n+1}}}{e^{\frac{2j\pi i}{n+1}}-1} \\ &= -1. \end{aligned}$$

Note that $\sum_{s=1}^{n}\left(e^{\frac{2j\pi i}{n+1}}\right)^s$ is a geometric series and $\left(e^{\frac{2j\pi i}{n+1}}\right)^{n+1} = e^{2j\pi i} = 1$. Thus, $\sum_{s=1}^{n}\cos\left(\frac{2js\pi}{n+1}\right) = -1$. Therefore, for all $1 \leq j, z \leq n$

$$\sum_{s=1}^{n} \sin^2\left(\frac{sj\pi}{n+1}\right) = \frac{1}{2}(n-(-1)) = \frac{n+1}{2},$$

and

$$\begin{aligned} \sum_{s=1}^{n}\sin\left(\frac{js\pi}{n+1}\right)\sin\left(\frac{zs\pi}{n+1}\right) &= \frac{1}{2}\left(\sum_{s=1}^{n}\cos\left(\frac{(j-z)s\pi}{n+1}\right) - \cos\left(\frac{(j+z)s\pi}{n+1}\right)\right) \\ &= \frac{1}{2}(-1-(-1)) \\ &= 0. \end{aligned}$$

as desired.

Remark 4. *V' is both unitary and symmetric, so we have*

$$V'^{*} = V \;\Rightarrow\; V'^{\,2} = I_n \Rightarrow V' = V'^{-1}.$$

Going back to the matrix T, since T is a symmetric matrix with real entries, we can apply the Lemma 2.3.2 to prove the following Theorem.

Theorem 2.3.1. *Let T be the following symmetric matrix;*

$$T := \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ 0 & 1 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}.$$

There exist weights, $\lambda_1, \cdots, \lambda_n \geq 0$, for the eigenvalues, such that $(T)_{11}$ has the same moment as a semicircular random variable.

Proof: By Lemma 2.3.2, there exist $\lambda_1, \cdots, \lambda_n \geq 0$ such that $\sum_{s=1}^{n} \lambda_s = 1$, and for all $k \in \mathbb{N}$

$$\left(T^k\right)_{11} = \sum_{s=1}^{n} \lambda_s e_s^k,$$

where $e_s = 2\cos(\frac{s\pi}{n+1})$. We also know that

$$(T^k)_{11} = \begin{cases} 0, & \text{k is odd;} \\ C_{\frac{k}{2}}, & \text{k is even.} \end{cases}$$

So we get

$$\sum_{s=1}^{n} \lambda_s e_s^k = \begin{cases} 0, & k \text{ is odd;} \\ C_{\frac{k}{2}}, & k \text{ is even.} \end{cases}$$

Remark 5. *For $1 \leq k \leq 2(n-1)$, we have*

$$\sum_{s=1}^{n} \lambda_s \left(2\cos(\frac{s\pi}{n+1})\right)^k = \begin{cases} 0, & k \text{ is odd;} \\ C_{\frac{k}{2}}, & k \text{ is even.} \end{cases}$$

From Remarks 1 and 2 we have,

$$\lambda_s = \overline{V'}_{1s} V'_{1s} = V'^2_{1s} = \delta_n^2 \sin^2\left(\frac{s\pi}{n+1}\right).$$

All together we have,

$$\begin{aligned} \sum_{s=1}^{n} \lambda_s e_s^k &= \sum_{s=1}^{n} \lambda_s \left(2\cos\left(\frac{s\pi}{n+1}\right)\right)^k \\ &= \sum_{s=1}^{n} \delta_n^2 \sin^2\left(\frac{s\pi}{n+1}\right)\left(2\cos\left(\frac{s\pi}{n+1}\right)\right)^k \\ &= \begin{cases} 0, & k \text{ is odd;} \\ C_{\frac{k}{2}}, & k \text{ is even.} \end{cases} \end{aligned}$$

Remark 6. *Since $\sum_{s=1}^{n} \lambda_s = 1$, we have*

$$\sum_{s=1}^{n} \delta_n^2 \sin^2\left(\frac{s\pi}{n+1}\right) = 1 \;\Rightarrow\; \delta_n^2 = \frac{1}{\sum_{s=1}^{n} \sin^2(\frac{s\pi}{n+1})}.$$

So, $\delta_n^2 \to 0$ as $n \to \infty$.

Remark 7. *For all $n > 0$, we have*

$$\delta^2 = \frac{2}{n+1}.$$

This follows from $\sum_{s=1}^{n} \sin^2\left(\frac{s\pi}{n+1}\right) = \frac{n+1}{2}$ and $\delta_n^2 = \frac{1}{\sum_{s=1}^{n} \sin^2\left(\frac{s\pi}{n+1}\right)}$.

Remark 8. *Lemma 2.3.2 not only holds for hermitian matrices, but also for symmetric real matrices and hermitian quaternionic matrices. In fact, it holds for Euclidean Jordan algebras, for which the reader, is referred to [20].*

For convenience of the reader we recall some of the concepts used in the following lemma. A vector space V over $\mathbb{R}$ or $\mathbb{C}$ is said to be a *Jordan algebra* if for all x and y in V, we have

$$xy = yx \qquad \text{and} \qquad x(x^2y) = x^2(xy).$$

An *Euclidean Jordan algebra* is a real Jordan algebra V satisfying

$$x^2 + y^2 = 0 \;\Rightarrow\; x = 0 = y \;\text{ for all }\; x, y \in V.$$

Every Euclidean Jordan algebra is a direct product of simple Euclidean Jordan algebras. A Euclidean Jordan algebra is said to be simple if it is not the direct product of two (non-trivial) Euclidean Jordan algebras. The simple Euclidean Jordan algebras are symmetric real matrices of arbitrary size (including $\mathbb{R}$), the hermitian complex and quaternionic matrices, the so-called spin factor of rank 2 and the exceptional 27-dimensional Euclidean Jordan algebra.

An *idempotent* of a Euclidean Jordan algebra V is an element $c \in V$ satisfies $c^2 = c$. Two idempotent c_1 and c_2 are orthogonal of $c_1c_2 = 0$. An idempotent is primitive if

$$\{v \in V : cv = v\} = \mathbb{R}_c.$$

A *Jordan frame* is a system $(c_1, \cdots, c_n)$ of pairwise orthogonal primitive idempotents such that $c_1 + \cdots c_n = e$, the identity element of V. Any two Jordan frames are conjugate under the automorphism group of V. In particular, they have the same size, known as the rank of V.

An $x \in V$ has a so-called *minimal decomposition*, i.e., there exists a Jordan frame $(d_1, \cdots, d_n)$ of V and unique nonnegative real numbers $\epsilon_1, \cdots, \epsilon_n$, called the eigenvalues of x, such that $x = \epsilon_1 d_1 + \cdots \epsilon_n d_n$.

A Jordan frame $(c_1, \cdots, c_n)$ of V induces a Peirce decomposition of V,

$$V = \bigoplus_{1 \le i \le j \le n} V_{ij}, \qquad (\star)$$

where the V_{ij} are the Peirce spaces of V. For example,

$$V_{11} = \{v \in V : c_1 v = v\} = \mathbb{R}_{c_1}.$$

The Peirce-11-component of $v \in V$ is the component of v with respect to the Peirce decomposition $(\star)$, which is given by $v_{11} = P(c_1)v$ for $P(x)y = 2x(xy) - x^2 y$. So $P(c_1)v = 2c_1(c_1)v - c_1 v$.

Lemma 2.3.4 (Lemma 2.3.2 for Jordan algebras)**.** *Let V be a Euclidean Jordan algebras of rank n and let $(c_1, \cdots, c_n)$ be a Jordan frame (complete system of orthogonal primitive idempotents). Let $x \in V$ and let $\epsilon_1, \cdots, \epsilon_n$ be its eigenvalues, counting multiplicities. For $y \in V$, we denote y_{11} the Peirce-11-component of y with respect to the Jordan frame $(c_1, \cdots, c_n)$. Then there exist $\lambda_1, \cdots, \lambda_r \geq 0$ satisfying*

(i) $\sum_i \lambda_i = 1$, *and*

(ii) *for all* $\xi \in \mathbb{N}$, $(x^\xi)_{11} = \left(\sum_{i=1}^r \lambda_i \epsilon_i^\xi\right) c_1$.

Proof: We know there exists another Jordan frame (minimal decomposition), $(d_1, \cdots, d_n)$ such that $x = \sum_{i=1}^n \epsilon_i d_i$, whence $x^\xi = \sum_{i=1}^n \epsilon_i^\xi d_i$ and $(x^\xi)_{11} = \sum_{i=1}^n \epsilon_i^\xi P(c_1) d_i$, where $P(.)$ is the quadratic operator of V. Since d_i lies in the boundary of the symmetric cone of V, we set $P(c_1)d_i = \lambda_i c_1$ with $\lambda_i \geq 0$. Also

$$\begin{aligned}
\left(\sum_{i=1}^n \lambda_i\right) c_1 &= \sum_{i=1}^n P(c_1) d_i \\
&= P(c_1) \sum_{i=1}^n d_i \\
&= P(c_1) e \\
&= c_1^2 \\
&= c_1,
\end{aligned}$$

whence $\sum_{i=1}^n \lambda_i = 1$.

2.4 Asymptotic behaviour of a Dyck path

Recall that, for every integer $n \geq 0$ we will denote by C_n the nth Catalan number,

$$C_n := \frac{1}{n+1}\binom{2n}{n} = \frac{(2n)!}{n!(n+1)!},$$

with the convention that $C_0 = 1$. We first give the definitions relevant to Dyck paths. A NE–SE path is a path in $\mathbb{Z}^2$ which starts at $(0,0)$ and makes steps either of the form $(1,1)$ (north-east steps) or of the form $(1,-1)$ south-east steps).

Definition 11. *A Dyck path is a NE-SE path γ which ends on the $x-$axis. That is, all the lattice points visited by γ are of the form (i,j) with $j \geq 0$, and the last of them is the form $(k,0)$ [51].*

In the next section, we will discuss connections between Dyck path and Catalan numbers. For any integer $N > 0$, define the set of Dyck paths with length of $2N$ as follows,

$$D_{2N} = \{S := (S_i)_{0\leq i\leq 2N} : S_0 = S_{2N} = 0, S_{i+1} = S_i \pm 1, S_i \geq 0 \ \forall\ i \in [0, 2N-1]\}.$$

It is well known that $C_N = |D_{2N}|$ where $|x|$ gives the cardinality of x.
Let $D_{2N,n} = \{S : S \in D_{2N}, \max_{1\leq i\leq 2N} S_i < n\}$. Note that if $N < n$ then $D_{2N,n} = D_{2N}$.

Lemma 2.4.1. *For positive integers n and N, we have*

$$|D_{2N,n}| = \sum_{s=1}^{n} \left(\frac{2}{n+1}\right) \sin^2\left(\frac{\pi s}{n+1}\right) \left(2\cos\left(\frac{\pi s}{n+1}\right)\right)^{2N}. \tag{2.4.1}$$

Proof: It follows directly from the definition of $D_{2N,n}$ and from properties of adjacency graphs that

$$|D_{2N,n}| = (T^{2N})_{11}.$$

So by Theorem 2.3.1 we have the result.

Corollary 2.4.1. *For positive integers $n > 0$ and $0 < N \leq n-1$, the Nth Catalan number satisfies $C_N = |D_{2N}| = |D_{2N,n}|$. Therefore,*

$$C_{N,n} = \sum_{s=1}^{n} \left(\frac{2}{n+1}\right) \sin^2\left(\frac{\pi s}{n+1}\right) \left(2\cos\left(\frac{\pi s}{n+1}\right)\right)^{2N}.$$

Lemma 2.4.2. *Asymptotically, the Catalan numbers behave as*

$$C_n = \frac{4^n}{n^{\frac{3}{2}}\sqrt{\pi}}\left(1 + o(1)\right).$$

Proof: From Stirling's formula,

$$\frac{n!}{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n} \to 1.$$

Thus

$$\begin{aligned} C_n &= \frac{(2n)!}{n!(n+1)!} \\ &\sim \frac{\sqrt{4\pi n}\left(\frac{2n}{e}\right)^{2n}}{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n \sqrt{2\pi(n+1)}\left(\frac{n+1}{e}\right)^{n+1}} \\ &= \frac{4^n}{\sqrt{\pi}(n+1)^{\frac{3}{2}}}\left[e\left(\frac{n}{n+1}\right)^n\right]. \end{aligned}$$

We know that for sufficiently large n, $\left(1+\frac{1}{n}\right)^n = \left(\frac{n+1}{n}\right)^n \to e$, so

$$\begin{aligned} C_n &\sim \frac{4^n}{\sqrt{\pi}(n+1)^{\frac{3}{2}}}\left[\left(\frac{n+1}{n}\right)^n \left(\frac{n}{n+1}\right)^n\right] \\ &\sim \frac{4^n}{\sqrt{\pi}(n+1)^{\frac{3}{2}}} \\ &\sim \frac{4^n}{\sqrt{\pi}(n)^{\frac{3}{2}}}. \end{aligned}$$

Let $D_{t,n,\kappa}$ denote the number of Dyck paths in tn^κ steps with maximum height less than n, where t is an even constant. Therefore,

$$D_{t,n,\kappa} = \sum_{s=1}^{n} \left(\frac{2}{n+1}\right) \sin^2\left(\frac{\pi s}{n+1}\right) \left(2\cos\left(\frac{\pi s}{n+1}\right)\right)^{2tn^\kappa}.$$

Let $m(n) \sim tn^{\kappa}$. In this section, we will discuss the situations where $\kappa < 2$, $\kappa = 2$ *and* $\kappa > 2$. Note that, in the whole section we assume that $t(n)^{\kappa}$ is an integer, otherwise we let $m = 2[\frac{tn^{\kappa}}{2}]$.

The tn^{κ}th Catalan number gives the number of Dyck paths of length tn^{κ} and $D_{t,n,\kappa}$ gives the number of the number of Dyck paths of length tn^{κ} with the maximum height less than n. Let $\mathbb{P}_n$ be the uniform distribution, then we have

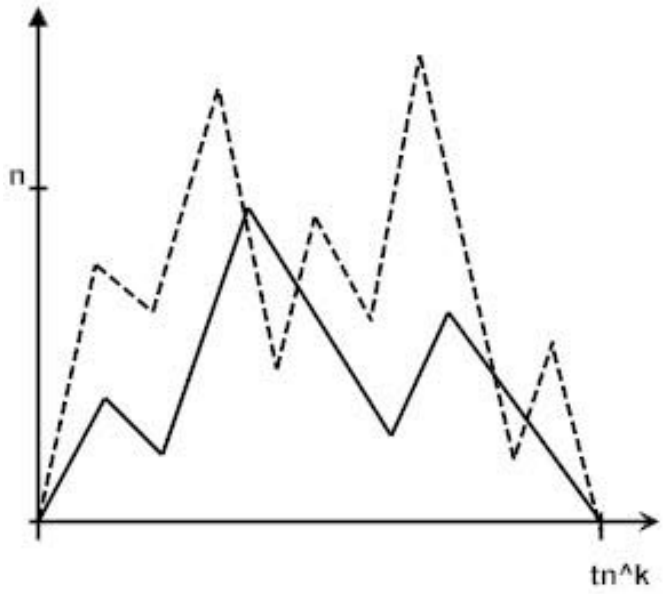


Figure 2.3: Dyck path of length tn^{κ} with maximum height of n

$$\frac{D_{t,n,\kappa}}{C_{tn^{\kappa}}} = \mathbb{P}_n(\max_n \text{height of Dyck paths in } 2tn^{\kappa}\text{steps} < n).$$

2.4.1 Results where $\kappa = 2$

We show that in this case, $\lim_{n\to\infty} \frac{D_{t,n,2}}{C_{tn^2}} = f(t)$, where f is a function that only depends on t. First we start by stating the dominated convergence theorem.

We recall that Lebesgue's dominated convergence theorem [11] states that if $(X, \mathcal{A}, \mu)$ is a measure space, g is an $[0, \infty)$-valued integrable function on X, and f and

$f_1,\ f_2, \ldots,$ are $(-\infty, \infty)$-valued $\mathcal{A}$-measurable functions on X such that the relations

$$f(x) = \lim_n f_n(x)$$

and

$$|f_n(x)| \leq g(x),\ n = 1,\ 2, \ldots$$

hold at almost every x in X, then f and $f_1,\ f_2, \ldots,$ are integrable, and $\int f d\mu = \lim_n \int f_n d\mu$.

Lemma 2.4.3. *For all* $0 \leq x \leq \frac{\pi}{2}$, *we have*

$$\cos(x) \leq \exp\left(-\frac{x^2}{2}\right).$$

Proof: Let $f(x) = \cos(x)\exp\left(\frac{x^2}{2}\right)$, where $0 \leq x \leq \frac{\pi}{2}$. Then $f(x)$ is a decreasing function on this interval we have

$$\begin{aligned} f'(x) &= -\sin(x)\exp\left(\frac{x^2}{2}\right) + \cos(x)x\exp\left(\frac{x^2}{2}\right) \\ &= \exp\left(\frac{x^2}{2}\right)(-\sin(x) + x\cos(x)). \end{aligned}$$

So $f'(x) = 0$ yields $x = \tan(x)$; on the interval $[0, \frac{\pi}{2}]$ forces $x = 0, \pi$. Therefore $f(x)$ is decreasing and obtains its maximum at zero. We have $f(0) = 1$ and $f(\frac{\pi}{2}) = 0$, and $f(x) \leq 1$ on the interval.

Now let $x = \frac{\pi s}{n+1}$, so for all $0 \leq s \leq \frac{\sqrt{n+1}}{(\ln n)^{\frac{1}{4}}}$, we have

$$\cos\left(\frac{\pi s}{n+1}\right) \leq \exp\left(-\frac{\pi^2 s^2}{2(n+1)^2}\right).$$

Let $g(s) = \exp(-t\pi^2 s^2)$, then g is a real-valued function and we have $\sum_{s=1}^{n} g(s) < \infty$ and also $\lim_{n\to\infty} g(s) \to 0$. So we have by the dominated convergence theorem,

$$\sum_{s=1}^{n} \cos\left(\frac{\pi s}{n+1}\right)^{2tn^2} < \sum_{s=1}^{n} \exp\left(-t\pi^2 s^2\right) < \infty.$$

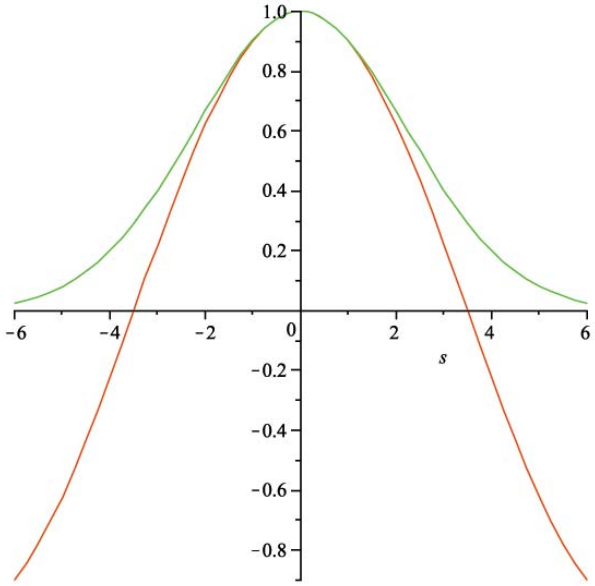


Figure 2.4: $n = 6$, s varies between 0 and $\frac{n+1}{2}$ the graph of $e^{-\frac{s^2}{2}}$ is on top of the graph of $\cos(\frac{\pi s}{n+1})$.

Lemma 2.4.4. *For all fixed s with $0 \leq s < \frac{\sqrt{n+1}}{(\ln n)^{\frac{1}{4}}}$,*

$$\lim_{n\to\infty} \cos\left(\frac{\pi s}{n+1}\right)^{2tn^2} \to g(s); \quad \textit{and,}$$

$$\lim_{n\to\infty} (n+1)^2 \sin^2\left(\frac{\pi s}{n+1}\right) \to \pi^2 s^2,$$

where $g(s) = \exp(-t\pi^2 s^2)$. Hence for all fixed $1 \leq s < \frac{\sqrt{n+1}}{(\ln n)^{\frac{1}{4}}}$, $(n+1)^2 \sin^2\left(\frac{\pi s}{n+1}\right) \cos\left(\frac{\pi s}{n+1}\right)^{2tn^2}$ converges pointwise to $s^2\pi^2 g(s)$.

Proof: Recall that the Taylor series for $\sin(x)$ and $\cos(x)$ are

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots$$

$$\cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots.$$

So $\sin^2\left(\frac{\pi s}{n+1}\right)$ and $\cos\left(\frac{\pi s}{n+1}\right)$ can be presented as follows,

$$\sin^2\left(\frac{\pi s}{n+1}\right) = \left(\frac{s\pi}{(n+1)} - \frac{s^3\pi^3}{3!(n+1)^3} + \frac{s^5\pi^5}{5!(n+1)^5} - \cdots\right)^2$$

and

$$\cos\left(\frac{\pi s}{n+1}\right) = 1 - \frac{1}{2}\frac{s^2\pi^2}{(n+1)^2} + \frac{s^4\pi^4}{4!(n+1)^4} - \cdots$$

So

$$\cos\left(\frac{\pi s}{n+1}\right)^{2tn^2} = \exp\left[2tn^2 \ln\left(1 - \frac{1}{2}\frac{s^2\pi^2}{(n+1)^2} + O\left(\frac{s^4}{n^4}\right)\right)\right].$$

But we also have that

$$\ln(1+x) = x + O(x^2),$$

so

$$\ln\left(1 - \frac{1}{2}\frac{s^2\pi^2}{(n+1)^2} + O\left(\frac{s^4}{n^4}\right)\right) = -\frac{1}{2}\frac{s^2\pi^2}{(n+1)^2} + O\left(\frac{s^4}{n^4}\right).$$

Thus,

$$\begin{aligned}\lim_{n\to\infty} \cos\left(\frac{\pi s}{n+1}\right)^{2tn^2} &= \lim_{n\to\infty} \exp\left[2tn^2\left(-\frac{1}{2}\frac{s^2\pi^2}{(n+1)^2} + O\left(\frac{s^4}{n^4}\right)\right)\right] \\ &= g(s).\end{aligned}$$

Moreover,

$$\begin{aligned}\lim_{n\to\infty} (n+1)^2 \sin^2\left(\frac{\pi s}{n+1}\right) &= \lim_{n\to\infty} (n+1)^2 \left(\frac{s\pi}{(n+1)} - \frac{s^3\pi^3}{3!(n+1)^3} - \cdots\right)^2 \\ &= \pi^2 s^2 + \lim_{n\to\infty} (n+1)^2 \left(-\frac{s^3\pi^3}{3!(n+1)^3} - \cdots\right)^2 \\ &= \pi^2 s^2.\end{aligned}$$

Note that $O(\frac{S^4}{n^4}) < O(\frac{1}{(n+1)^2 \ln n})$ for all $1 \le s \le \frac{\sqrt{n+1}}{(\ln n)^{\frac{1}{4}}}$. So $(n+1)^2 O(\frac{s^4}{n^4})$ goes to zero as later goes to infty for all $1 \le s < \frac{\sqrt{n+1}}{(\ln n)^{\frac{1}{4}}}$.

Lemma 2.4.5. *For all* $\frac{\sqrt{n+1}}{(\ln n)^{\frac{1}{4}}} \le s \le \frac{n}{2}$, *we have*

$$\lim_{n\to\infty} (n+1)^2 \sin^2\left(\frac{\pi s}{n+1}\right) \cos\left(\frac{\pi s}{n+1}\right)^{2tn^2} \to 0$$

exponentially fast. So,

$$\lim_{n\to 0} (n+1)^2 \sum_{s=\frac{\sqrt{n+1}}{(\ln n)^{\frac{1}{4}}}}^{\frac{n}{2}} \sin^2\left(\frac{\pi s}{n+1}\right) \cos\left(\frac{\pi s}{n+1}\right)^{2tn^2} \to 0.$$

Proof: For all $\frac{\sqrt{n+1}}{(\ln n)^{\frac{1}{4}}} \leq s \leq \frac{n}{2}$ we have,

$$\begin{aligned}(n+1)^2 \sin^2\left(\frac{\pi s}{n+1}\right)\cos\left(\frac{\pi s}{n+1}\right)^{2tn^2} &< (n+1)^2 \exp\left(-\frac{\pi^2 s^2}{(n+1)^2}2tn^2\right)\\ &< (n+1)^2 \exp\left(-\frac{\pi^2(n+1)}{\sqrt{\ln n}}\right)\\ &\to 0.\end{aligned}$$

Moreover we have,

$$\begin{aligned}(n+1)^2 \sum_{s=\frac{\sqrt{n+1}}{(\ln n)^{\frac{1}{4}}}}^{\frac{n}{2}} \sin^2\left(\frac{\pi s}{n+1}\right)\cos\left(\frac{\pi s}{n+1}\right)^{2tn^2} &< \frac{n}{2}(n+1)^2 \exp\left(-\frac{\pi^2(n+1)}{\sqrt{\ln n}}\right)\\ &< n^3 \exp\left(-\frac{\pi^2(n+1)}{\sqrt{\ln n}}\right).\end{aligned}$$

So,

$$\lim_{n\to 0}\ (n+1)^2 \sum_{s=\frac{\sqrt{n+1}}{(\ln n)^{\frac{1}{4}}}}^{\frac{n}{2}} \sin^2\left(\frac{\pi s}{n+1}\right)\cos\left(\frac{\pi s}{n+1}\right)^{2tn^2} \to 0.$$

For positive integers n, N and s define,

$$G_{N,n}(s) = \sin^2\left(\frac{\pi s}{n+1}\right)\cos^{2N}\left(\frac{\pi s}{n+1}\right).$$

For fixed n we have $\sin\left(\frac{\pi}{n+1}\right) = \sin\left(\frac{n\pi}{n+1}\right)$ and $\cos\left(\frac{\pi}{n+1}\right) = -\cos\left(\frac{n\pi}{n+1}\right)$. So by symmetry we have $G_{N,n}(1) = G_{N,n}(n)$ and for even n, $\sum_{s=1}^{n} G_{N,n}(s) = 2\sum_{s=1}^{\frac{n}{2}} G_{N,n}(s)$. Note that, without loss of generality for large value of n, we can assume n is even. Since if n is odd we have the sum is equal to $2\sum_{s=1}^{\frac{n-1}{2}} G_{N,n}(s) + \sin^2\left(\frac{n\pi}{n+1}\right)\left(2\cos\left(\frac{n\pi}{n+1}\right)\right)^{2N}$ and as $n \to \infty$ the last term will go to zero.

Theorem 2.4.1. *Let $\kappa = 2$; then we have*

$$\lim_{n\to\infty} \frac{D_{t,n,2}}{C_{tn^2}} \longrightarrow f(t),$$

where

$$f(t) = 4\sqrt{\pi}t^{\frac{3}{2}} \sum_{s=1}^{\infty} s^2\pi^2 \exp\left(-ts^2\pi^2\right).$$

Proof: We have,

$$\begin{aligned} D_{t,n,2} &= \sum_{s=1}^{n} \frac{2}{n+1} \sin^2\left(\frac{\pi s}{n+1}\right)\left(2\cos\left(\frac{\pi s}{n+1}\right)\right)^{2tn^2} \\ &= 2\sum_{s=1}^{\frac{n}{2}} \frac{2}{n+1} \sin^2\left(\frac{\pi s}{n+1}\right)\left(2\cos\left(\frac{\pi s}{n+1}\right)\right)^{2tn^2} \\ &= \frac{4^{tn^2+1}}{(n+1)^3} \sum_{s=1}^{\frac{n}{2}} (n+1)^2 \sin^2\left(\frac{\pi s}{n+1}\right)\left(\cos\left(\frac{\pi s}{n+1}\right)\right)^{2tn^2} \end{aligned}$$

On the other hand, we have from Lemma 2.4.2 that asymptotically, Catalan number grow as $C_n \sim \frac{4^n}{\sqrt{\pi}n^{\frac{3}{2}}}$. So

$$C_{tn^2} \sim \frac{4^{tn^2}}{\sqrt{\pi}t^{\frac{3}{2}}n^3}.$$

We have from Lemma 2.4.4 that for all $1 \leq s \leq \frac{\sqrt{n+1}}{(\ln n)^{\frac{1}{4}}}$, $(n+1)^2 \sin^2\left(\frac{\pi s}{n+1}\right)\cos\left(\frac{\pi s}{n+1}\right)^{2tn^2}$ converges pointwise to $s^2\pi^2 g(s)$. Moreover, we have for all $s \geq 1$,

$$\sin^2\left(\frac{\pi s}{n+1}\right) \leq \frac{\pi^2 s^2}{(n+1)^2}.$$

So for all $1 \leq s \leq \frac{\sqrt{n+1}}{(\ln n)^{\frac{1}{4}}}$,

$$(n+1)^2 \sin^2\left(\frac{\pi s}{n+1}\right)\cos\left(\frac{\pi s}{n+1}\right)^{2tn^2} \leq (n+1)^2 \frac{\pi^2 s^2}{(n+1)^2} g(s) \leq \pi^2 s^2 g(s),$$

where $\sum_{s=1}^{\infty} s^2\pi^2 g(s) < \frac{1}{4t^{\frac{3}{2}}\sqrt{\pi}}$. So by the dominated convergence theorem, we have

$$\lim_{n\to\infty} \sum_{s=1}^{\frac{\sqrt{n+1}}{(\ln n)^{\frac{1}{4}}}} (n+1)^2 \sin^2\left(\frac{\pi s}{n+1}\right)\cos\left(\frac{\pi s}{n+1}\right)^{2tn^2} \to \sum_{s=1}^{\infty} s^2\pi^2 g(s).$$

By Lemma 2.4.3 and Lemma 2.4.5, we get

$$\begin{aligned}
\lim_{n\to\infty}\frac{D_{t,n,\kappa}}{C_{tn^2}} &= \lim_{n\to\infty}\frac{\frac{4^{tn^2+1}}{(n+1)^3}\sum_{s=1}^{\frac{n}{2}}(n+1)^2\sin^2\left(\frac{\pi s}{n+1}\right)\left(\cos\left(\frac{\pi s}{n+1}\right)\right)^{2tn^2}}{\frac{4^{tn^2}}{\sqrt{\pi}t^{\frac{3}{2}}n^3}} \\
&= \lim_{n\to\infty}\frac{\frac{4^{tn^2+1}}{(n+1)^3}\left(\sum_{s=1}^{\frac{\sqrt{n+1}}{(\ln n)^{\frac{1}{4}}}}+\sum_{\frac{\sqrt{n+1}}{(\ln n)^{\frac{1}{4}}}}^{\frac{n}{2}}\right)(n+1)^2\sin^2\left(\frac{\pi s}{n+1}\right)\left(\cos\left(\frac{\pi s}{n+1}\right)\right)^{2tn^2}}{\frac{4^{tn^2}}{\sqrt{\pi}t^{\frac{3}{2}}n^3}} \\
&= \lim_{n\to\infty}\frac{4^{tn^2}\left(\frac{4}{(n+1)^3}\right)}{\frac{4^{tn^2}}{\sqrt{\pi}t^{\frac{3}{2}}n^3}}\sum_{s=1}^{\infty}s^2\pi^2\exp\left(-ts^2\pi^2\right) \\
&= \lim_{n\to\infty}\frac{4^{tn^2}\left(\frac{4}{(n+1)^3}\right)}{\frac{4^{tn^2}}{\sqrt{\pi}t^{\frac{3}{2}}n^3}}\lim_{n\to\infty}\sum_{s=1}^{\frac{n}{2}}(n+1)^2\sin^2\left(\frac{\pi s}{n+1}\right)\left(\cos\left(\frac{\pi s}{n+1}\right)\right)^{2tn^2} \\
&= \lim_{n\to\infty}\frac{4^{tn^2}\left(\frac{4}{(n+1)^3}\right)}{\frac{4^{tn^2}}{\sqrt{\pi}t^{\frac{3}{2}}n^3}}\sum_{s=1}^{\infty}s^2\pi^2\exp\left(-ts^2\pi^2\right) \\
&= 4\sqrt{\pi}t^{\frac{3}{2}}\sum_{s=1}^{\infty}s^2\pi^2\exp\left(-ts^2\pi^2\right).
\end{aligned}$$

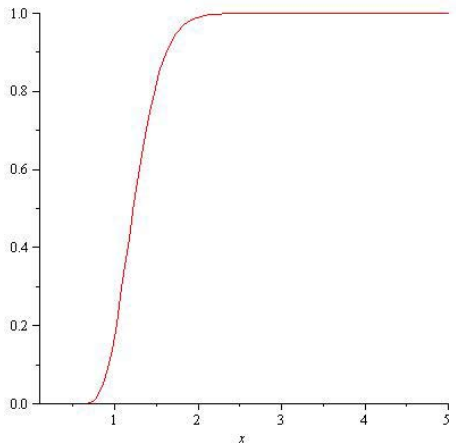


Figure 2.5: $f(x),\ x \in [0,4]$

Let $f(t) = 4\sqrt{\pi}t^{\frac{3}{2}}\sum_{s=1}^{\infty}s^2\pi^2\exp\left(-ts^2\pi^2\right)$. Let x be a positive number such that

$x^2 = \frac{1}{2t}$, equivalently we have $t = \frac{1}{2x^2}$. We can reformulate our results as follows,

$$\begin{aligned} f(t) &= 4\sqrt{\pi}t^{\frac{3}{2}} \sum_{s=1}^{\infty} s^2\pi^2 \exp\left(-ts^2\pi^2\right). \\ &= 4\sqrt{\pi}(2)^{-\frac{3}{2}}x^{-3} \sum_{s=1}^{\infty} s^2\pi^2 \exp\left(-\frac{\pi^2 s^2}{2x^2}\right). \end{aligned}$$

For $n > 1$, let $f(x, n)$ be the truncated version of f, i.e.,

$$f(x,n) = 4\sqrt{\pi}(2)^{-\frac{3}{2}}x^{-3} \sum_{s=1}^{n} s^2\pi^2 \exp\left(-\frac{\pi^2 s^2}{2x^2}\right).$$

Brownian bridges and Brownian excursions

A *one-dimensional Brownian motion* or a *Wiener process* is a real-valued process $W_t,\ t \geq 0$ that has the following properties [16]:

1. W_t has independent increments, i.e., if $t_0 < t_1 < \ldots < t_n$, then $W(t_0),\ W(t_1) - W(t_0), \ldots, W(t_n) - W(t_{n-1})$ are independent.

2. The increments $W(s+t) - W(s)$ have a normal distribution with mean 0 and variance t, i.e., if $s, t \geq 0$, then

$$P(W(s+t) - W(s) \in A) = \int_A (2\pi t)^{-\frac{1}{2}} \exp(\frac{-x^2}{2t})dx.$$

From [16], two consequences follow immediately:

1. **Translation Invariance:** $\{W_t - W_0, t \geq 0\}$ are independent of W_0 and have the same distribution as a Brownian motion with $W_0 = 0$.

2. **The Brownian Scaling Relation:** For any $t > 0$,

$$\{W_{st}, s \geq 0\} =^d \{t^{\frac{1}{2}} W_s, s \geq 0\};$$

that is, the two families of random variables have the same finite-dimensional distributions.

A *Brownian bridge*, W_t° is a continuous-time stochastic process W_t defined on $[0,1]$, whose probability distribution is the conditional probability distribution of a Wiener process $W_0 = W_1 = 0$. The expected value of the Brownian bridge is zero with variance $t(1-t)$. If W_t is a standard Brownian motion (i.e., W_t is normally distributed with expected value 0 and variance t), then

$$W_t^\circ = W_t - tW_1$$

is a Brownian bridge.

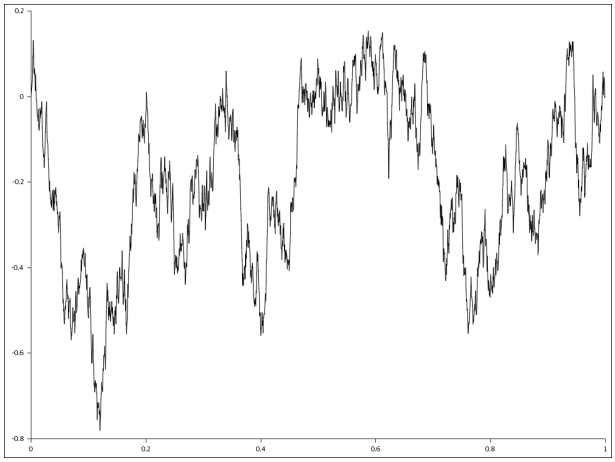


Figure 2.6: Example of a Brownian bridge.

Let $C[0,1]$ denote the collection of all continuous functions $f : [0,1] \to R$. Let $X_1,\ X_2,\ \cdots$ be a sequence of i.i.d. random variables defined on a probability space $(\Omega, \mathcal{B}, \mathbb{P})$ where $\mathcal{B} := \mathcal{B}(C[0,1])$ is the Borel σ-algebra on $C[0,1]$. Assume X_i are integer-valued and $E(X_i) = 0,\ \mathbb{E}(X_i^2) = 1$. Form the random walk $\{S_n : n \geq 0\}$ by defining $S_0 = 0$ and $S_n = X_1 + \cdots + X_n,\ n \geq 1$. Let $D \equiv D[0,1]$ be the function space of all real-valued, right continuous functions on $[0,1]$ with left-hand limits. For all $\omega \in \Omega$, $n \geq 1$ and $t \in (0,1]$, define

$$\varphi_n(t,\omega) := \frac{1}{\sqrt{n}} \sum_{i-1}^{n} \left[S_{i-1}(\omega) + n\left(t - \frac{i-1}{n}\right) X_i(\omega) \right] 1_{\left(\frac{i-1}{n}, \frac{i}{n}\right]}(t)$$

Also define $\varphi_n(0,\omega) := 0$. Note that $\varphi_n := \{\varphi_n(t); t \in [0,1]\}$ is merely the linear interpolation of the normalized random walk $\{\frac{S_1}{\sqrt{n}}, \cdots, \frac{S_n}{\sqrt{n}}\}$.

Theorem 2.4.2. ***Donsker's Theorem*** *[5] We have $\varphi_n \Rightarrow W$, where W denotes the standard Brownian motion and $\Rightarrow$ means weak convergence (or convergence in distribution).*

Let $W(t),\ 0 \le t < \infty$ denote a standard Brownian motion and $\tau_1 = \sup\{t < 1 : W(t) = 0\},\ \tau_2 = \inf\{t > 1 : W(t) = 0\}$. Define the process $W_1(s),\ 0 \le s \le 1$ by setting

$$W_1(s) = \frac{|W(s\tau_2 + (1-s)\tau_1)|}{(\tau_2 - \tau_1)^{\frac{1}{2}}}.$$

The process W_1 is known as the *unsigned normalized Brownian excursion process* [34]. Define the hitting time T to be the $\min\{n \ge 1 : S_n = 0\}$ ($+\infty$ if no such n exists). Let $\mathbb{P}_n^\circ = \mathbb{P}\left[\frac{S_1}{\sigma\sqrt{n}} = x_1, \cdots, \frac{S_n}{\sigma\sqrt{n}} = x_n | T = n\right]$. Kaigh [37] proved the following theorem.

Theorem 2.4.3. *[37] The sequence of probability measures $\{\mathbb{P}_n^\circ\}$ on D converges weakly to a probability measure W_1 which assigns probability one to continuous sample paths. The weak limit W_1 corresponds to the Brownian excursion process on $[0,1]$.*

The distribution of the maximum of the unsigned scaled Brownian excursion process and the modification of that process was derived in 1976 by Kennedy [38].

Theorem 2.4.4. *[38] For every $x > 0$,*

$$\mathbb{P}\left(\sup_{1\le s\le n} W_1(s) \le x\right) = 1 - 2\sum_{n=1}^{\infty}\left(4x^2n^2 - 1\right)\exp\left(-2x^2n^2\right).$$

Chung [10] reported the same results derived by a different argument. It was shown in [38, 10] that the maximum of a Brownian excursion has the same distribution as the range of the Brownian bridge, i.e.,

$$\sup_{0\le s\le 1} W_1(s) \stackrel{d}{=} \sup_{0\le s\le 1} W^\circ(s) - \inf_{0\le s\le 1} W^\circ(s).$$

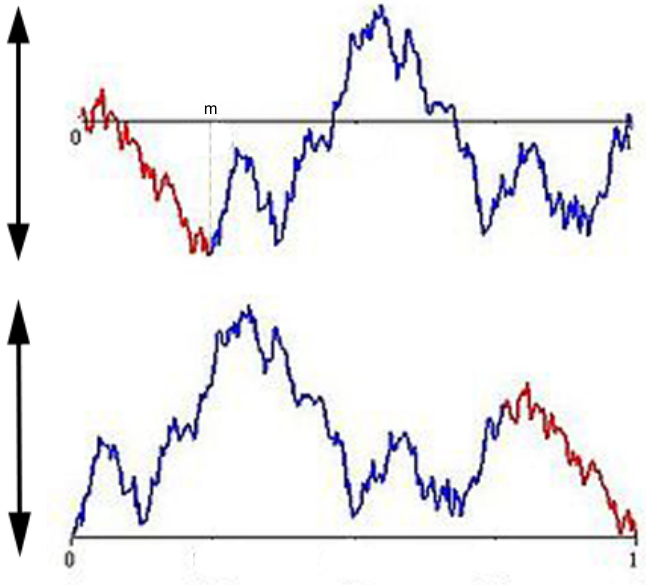


Figure 2.7: Range of the Brownian bridge and the maximum of Brownian excursion have the same distribution.

Dyck paths and Brownian excursions

Recall that a Dyck path of length $2n$, where n is a nonnegative integer, is a map $y : [0, 2n] \to \{x \in \mathbb{R} : x \geq 0\}$, with $y(0) = y(2n) = 0$ and $|y(i) - y(i+1)| = 1$ for $i \in \mathbb{N} \cup \{0\},\ i < 2n$. The values $y(s)$ are called the heights of the paths. For $0 \leq s \leq 2n$, let $\tilde{Y}_n(s)$ denote the height of a random Dyck path on length $2n$. It is well known, see [8], that the average maximal height $\tilde{h}(n) \sim \sqrt{\pi n}$ as $n \to \infty$. [54] Let

$$D_{2n} = \{\mathbf{S} = (S_i)_{0 \leq i \leq n} : S_0 = S_n = 0, S_{i+1} = S_i \pm 1 \text{ for any } i \in [0, n-1]\}$$

be a Dyck path chosen uniformly from the set of Dyck paths with $2n$ steps. Khorunzhiy and Marckert [39] showed that, for any $\lambda > 0$ the sequence

$$\left\{\mathbb{E}\left(\exp(\lambda(2n)^{-\frac{1}{2}}\max D_{2n})\right)\right\}$$

converges, and thus is bounded uniformly in n.

Theorem 2.4.5. *[39] For any $\lambda > 0$ we have*

$$\mathbb{E}_{2n}^{(d)}\left(\exp(\lambda(2n)^{-\frac{1}{2}}\max \boldsymbol{S})\right) \to \mathbb{E}\left(\exp(\lambda \max_{t\in[0,1]} e(t))\right)$$

where $(e(t), t \in [0,1])$ is the normalized Brownian excursion and $\mathbb{E}_{2n}^{(d)}$ are the expectations with respect to the uniform distribution. In particular, for any $\lambda > 0$,

$$\sup_n \mathbb{E}_{2n}^{(d)}\left(\exp(\lambda(2n)^{-\frac{1}{2}}\max \boldsymbol{S})\right) < +\infty.$$

Using the computation of Chung [10] or Kennedy [38], we observe that the right hand side of the last inequality is finite for every $\lambda > 0$,

$$\mathbb{P}\left(\max_{t\in[0,1]} e(t) \le x\right) = 1 - 2\sum_{s=1}^{\infty}\left(4x^2s^2 - 1\right)\exp\left(-2x^2s^2\right), \quad \text{for } x > 0.$$

Corollary 2.4.2. *For every $x > 0$, let*

$$K(x) = \mathbb{P}\left(\sup_{1\le s\le n} W_1(s) \le x\right) = 1 - 2\sum_{n=1}^{\infty}\left(4x^2n^2 - 1\right)\exp\left(-2x^2n^2\right).$$

be the function determined by Kennedy [38] and Chung [10] as distribution of the maximum of the unsigned scaled Brownian excursion. If

$$f(t) = \lim_{n\to\infty} \mathbb{P}_n(\max D_{2tn^2} \le n),$$

then,

$$f(t) = 4\sqrt{\pi}t^{\frac{3}{2}}\sum_{s=1}^{\infty} s^2\pi^2\exp\left(-ts^2\pi^2\right).$$

Let $t = \frac{1}{2x^2}$ *and define* $T(x)$ *as follows,*

$$T(x) = \sqrt{2\pi}x^{-3}\sum_{n=1}^{\infty} n^2\pi^2 \exp\left(-\frac{n^2\pi^2}{2x^2}\right).$$

Then for all $x > 0$*,*

$$T(x) = K(x). \qquad (*)$$

Proof: For $x > 0$ let,

$$F(x) = \sum_{n=1}^{\infty} \exp\left(-\frac{n^2\pi^2}{2x^2}\right).$$

This function is analytic in the open disk of radius $\sqrt{\frac{\pi}{2}}$. In that disk we can calculate the derivative term by term. Thus,

$$F'(x) = x^{-3}\sum_{n=1}^{\infty} n^2\pi^2 \exp\left(-\frac{n^2\pi^2}{2x^2}\right).$$

Set

$$G(x) = x\sum_{n=1}^{\infty} \exp(-2x^2n^2);$$

this is also analytic in the open disk of radius $\sqrt{\frac{\pi}{2}}$. Therefore,

$$G'(x) = -\sum_{s=1}^{\infty}(4x^2n^2 - 1)\exp(-2x^2n^2).$$

Thus (*) is equivalent to

$$\sqrt{2\pi}F'(x) = 1 + 2G'(x),$$

that is

$$\sqrt{2\pi}F(x) = x + 2G(x) - 2k. \qquad (**)$$

If we calculate $F\left(\frac{\pi}{2x}\right)$ we obtain,

$$\begin{aligned} F\left(\frac{\pi}{2x}\right) &= \sum_{n=1}^{\infty}\exp(-2x^2n^2) = \frac{G(x)}{x}, \quad \text{so} \\ F(x) &= G\left(\frac{\pi}{2x}\right)\frac{2x}{\pi}. \end{aligned}$$

From (**),

$$\begin{aligned} G\left(\frac{\pi}{2x}\right)\frac{2x}{\pi}\sqrt{2\pi} &= x + 2G(x) - 2k, \quad \text{so} \\ G(x) &= \sqrt{\frac{2}{\pi}}xG\left(\frac{\pi}{2x}\right) - \frac{x}{2} + k \\ &= \sqrt{\frac{2}{\pi}}x\left(\sqrt{\frac{2}{\pi}}\frac{\pi}{2x}G(x) - \frac{\pi}{4x} + k\right) + k - \frac{x}{2} \\ &= G(x) - \frac{\sqrt{\pi}}{2\sqrt{2}} + k\sqrt{\frac{2}{\pi}}x + k - \frac{x}{2} \end{aligned}$$

whence $k = \frac{\sqrt{\pi}}{2\sqrt{2}}$. Hence,

$$G(x) = \sqrt{\frac{2}{\pi}}xG\left(\frac{\pi}{2x}\right) - \frac{x}{2} + \frac{\sqrt{\pi}}{2\sqrt{2}}.$$

So proving (*) is now equivalent to proving that for $x > 0$,

$$\begin{aligned} x\sum_{n=1}^{\infty}\exp(-2x^2n^2) &= \sqrt{\frac{2}{\pi}}x\frac{\pi}{2x}\sum_{n=1}^{\infty}\exp\left(-\frac{n^2\pi^2}{2x^2}\right) - \frac{x}{2} + \frac{1}{2}\sqrt{\frac{\pi}{2}} \\ &= \sqrt{\frac{\pi}{2}}\sum_{n=1}^{\infty}\exp\left(-\frac{n^2\pi^2}{2x^2}\right) - \frac{x}{2} + \frac{1}{2}\sqrt{\frac{\pi}{2}}. \end{aligned}$$

Thus

$$\sum_{n=1}^{\infty}\exp(-2x^2n^2) = \sqrt{\frac{\pi}{2}}\frac{1}{x}\sum_{n=1}^{\infty}\exp\left(-\frac{n^2\pi^2}{2x^2}\right) - \frac{1}{2} + \frac{1}{2x}\sqrt{\frac{\pi}{2}}.$$

Let $t = 2x^2$;

$$\sum_{n=1}^{\infty}\exp(-tn^2) = \sqrt{\frac{\pi}{t}}\sum_{n=1}^{\infty}\exp\left(-\frac{n^2\pi^2}{t}\right) - \frac{1}{2} + \frac{1}{2}\sqrt{\frac{\pi}{t}}.$$

Take the Laplace transformation $\int_0^{\infty}\cdot e^{-st}dt$ of both sides. Then,

$$\int_0^{\infty}\sum_{n=1}^{\infty}e^{-tn^2}e^{-st}dt = \int_0^{\infty}\sqrt{\frac{\pi}{t}}\sum_{n=1}^{\infty}e^{-\frac{n^2\pi^2}{t}}e^{-st}dt - \int_0^{\infty}\frac{1}{2}e^{-st}dt + \int_0^{\infty}\frac{1}{2}\sqrt{\frac{\pi}{t}}e^{-st}dt.$$

Convergence allows us to interchange summand and integral. Then

$$\sum_{n=1}^{\infty}\int_0^{\infty} e^{-tn^2}e^{-st}dt = \sum_{n=1}^{\infty}\int_0^{\infty}\sqrt{\frac{\pi}{t}}e^{-\frac{n^2\pi^2}{t}}e^{-st}dt - \int_0^{\infty}\frac{1}{2}e^{-st}dt + \int_0^{\infty}\frac{1}{2}\sqrt{\frac{\pi}{t}}e^{-st}dt.$$

From the Laplace transform table in [2],

$$\sum_{n=1}^{\infty}\frac{1}{n^2+s} = \frac{\pi}{\sqrt{s}}\sum_{n=1}^{\infty}e^{-n\pi\sqrt{s}} - \frac{1}{2s} + \frac{\pi}{2\sqrt{s}} \qquad (***)$$

i.e.,

$$\sum_{n=1}^{\infty}\frac{1}{n^2+s^2} = \frac{\pi}{s}\sum_{n=1}^{\infty}e^{-n\pi s} - \frac{1}{2s^2} + \frac{\pi}{2s}$$

We also have

$$\begin{aligned}\sum_{n=1}^{\infty}\frac{1}{n^2+s^2} &= \frac{\pi}{2s}\coth(s\pi) - \frac{1}{2s^2}\\ &= \frac{\pi}{2s}\left(\frac{1+e^{-\pi s}}{1-e^{-\pi s}}\right) - \frac{1}{2s^2}.\end{aligned}$$

So proving (*) is equivalent to showing the left hand side of (***) is equal to $\frac{\pi}{2s}\left(\frac{1+e^{-\pi s}}{1-e^{-\pi s}}\right) - \frac{1}{2s^2}$. We have,

$$\begin{aligned}\frac{\pi}{s}\sum_{n=1}^{\infty}e^{-n\pi s} - \frac{1}{2s^2} + \frac{\pi}{2s} &= \frac{\pi}{s}\left(\frac{e^{-\pi s}}{1-e^{-\pi s}}\right) - \frac{1}{2s^2} + \frac{\pi}{2s}\\ &= \frac{\pi}{2s}\left(1 + \frac{2e^{-\pi s}}{1-e^{-\pi s}}\right) - \frac{1}{2s^2}\\ &= \frac{\pi}{2s}\left(\frac{1+e^{-\pi s}}{1-e^{-\pi s}}\right) - \frac{1}{2s^2},\end{aligned}$$

as desired.

A different proof of Corollary 2.4.2 can be achieved using the Poisson summation formula. For an appropriate function f, the Poisson summation formula may be stated as:

$$\sum_{n=-\infty}^{\infty} f(t+nT) = \frac{1}{T}\sum_{k=-\infty}^{\infty}\hat{f}\left(\frac{k}{T}\right)\exp\left(2\pi i\frac{k}{T}t\right),$$

where $\hat{f}$ is the Fourier transform of f.

Proof: (**Alternative proof.**)

Let $f(x) = \exp(-2x^2)$, so $\hat{f} = \sqrt{\frac{\pi}{2}} \exp\left(-\frac{k^2\pi^2}{2}\right)$. Therefore, by the Poisson summation formula with $t = 0$ and $\frac{1}{x}$ we have,

$$\sum_{n=-\infty}^{\infty} f(nx) = \frac{1}{x} \sum_{k=-\infty}^{\infty} \hat{f}\left(\frac{k}{x}\right).$$

Therefore,

$$\sum_{n=-\infty}^{\infty} \exp(-2n^2x^2) = \frac{1}{x} \sum_{k=-\infty}^{\infty} \sqrt{\frac{\pi}{2}} \exp\left(\frac{-\pi^2k^2}{2x^2}\right).$$

Separating positive and negative indices and multiplying by x yields

$$x\left(1 + 2\sum_{n=1}^{\infty} \exp(-2n^2x^2)\right) = 1 + 2\sum_{k=1}^{\infty} \sqrt{\frac{\pi}{2}} \exp\left(\frac{-\pi^2k^2}{2x^2}\right).$$

Now by taking the derivative with respect to x,

$$1 - 2\sum_{n=1}^{\infty} \exp(-2n^2x^2)(4n^2x^2 - 1) = \sqrt{2\pi}\frac{\pi^2k^2}{x^3} \sum_{k=1}^{\infty} \exp\left(\frac{-\pi^2k^2}{2x^2}\right).$$

This completes the proof.

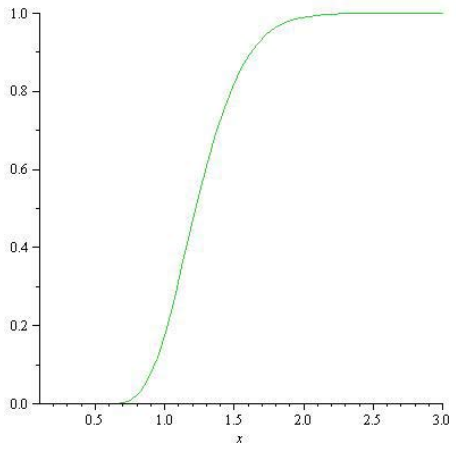


Figure 2.8: The diagram of $K(x) = T(x)$

The reason that we did the previous scaling was that the height in Kennedy's function was $x\sqrt{N}$, so we have

$$\begin{aligned} N &= 2tn^2 \\ x\sqrt{N} &= n. \end{aligned}$$

This yields $x^2 = \frac{1}{2t}$, or equivalently $t = \frac{1}{2x^2}$.

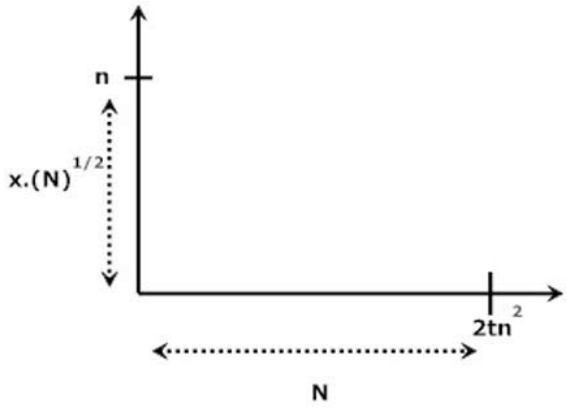


Figure 2.9: Rescaling of f(x)

2.4.2 Results where $\kappa > 2$

The goal of this section is to show if $\kappa > 2$,

$$\lim_{n\to\infty} \frac{D_{t,n,\kappa}}{C_{tn^\kappa}} \longrightarrow 0.$$

We need to prove the following proposition first.

Proposition 2.4.1. *For sufficiently large m, let*

$$G_{m,s} = \frac{2}{n+1} \sin^2\left(\frac{\pi s}{n+1}\right) \cos^{2m}\left(\frac{\pi s}{n+1}\right)$$

where $m \sim n^\kappa$, and $\kappa > 2$. We have

$$G_{m,1} = G_{m,n}$$

and for all $1 < s < n$, we have

$$\frac{G_{m,s}}{G_{m,n}} < \frac{1}{n}, \qquad so \qquad \lim_{n\to\infty} \frac{G_{m,s}}{G_{m,n}} \longrightarrow 0.$$

Proof: For fixed n, we have

$$\sin\left(\frac{\pi}{n+1}\right) = \sin\left(\frac{\pi n}{n+1}\right)$$

$$\cos\left(\frac{\pi}{n+1}\right) = -\cos\left(\frac{\pi n}{n+1}\right),$$

so consequently we get $G_{m,1} = G_{m,n}$.

Now, for all $1 \le s \le n$, define $\alpha_n \equiv \alpha_n(s) = \sin^2\left(\frac{\pi s}{n+1}\right)$, and $\beta_n = \sin^2\left(\frac{\pi}{n+1}\right)$. If $s = \frac{n+1}{2}$, then it is nothing left to prove. Otherwise, we have

$$\begin{aligned}
\ln\left(\frac{\cos^{2m}\left(\frac{\pi s}{n+1}\right)}{\cos^{2m}\left(\frac{\pi}{n+1}\right)}\right) &= m\left(\ln(1-\alpha_n) - \ln(1-\beta_n)\right) \\
&= -m(\alpha_n - \beta_n + \frac{\alpha_n^2 - \beta_n^2}{2} + \frac{\alpha_n^3 - \beta_n^3}{6} + \cdots) \\
&= -m(\alpha_n - \beta_n)(1 + \frac{\alpha_n + \beta_n}{2} + \frac{\alpha_n^2 + \alpha_n\beta_n + \beta_n^2}{6} + \cdots) \\
&\le -m(\alpha_n - \beta_n).
\end{aligned}$$

Therefore, we have

$$\frac{\cos^{2m}\left(\frac{\pi s}{n+1}\right)}{\cos^{2m}\left(\frac{\pi}{n+1}\right)} \leq e^{-m(\alpha_n-\beta_n)}.$$

Thus,

$$\begin{aligned}\frac{G_{m,s}}{G_{m,n}} &\leq \frac{\alpha_n}{\beta_n}e^{-m(\alpha_n-\beta_n)} \\ &= e^{-m(\alpha_n-\beta_n)+\ln\alpha_n-\ln\beta_n}.\end{aligned}$$

Now, consider two cases where $1 \leq s \leq \sqrt{n}$ and $\sqrt{n} \leq s \leq n-\sqrt{n}$. Note that because of the symmetry, s can be replaced by $n-s$, so that dealing with the interval $n-\sqrt{n} \leq s \leq n$ is the same as dealing with $\sqrt{n} \leq s \leq n-\sqrt{n}$.
If $1 \leq s \leq \sqrt{n}$, then we have

$$\frac{\alpha_n}{\beta_n} = \frac{\sin^2\left(\frac{\pi s}{n+1}\right)}{\sin^2\left(\frac{\pi}{n+1}\right)} \leq \pi^2 s^2\left(1+O\left(\frac{1}{n}\right)\right),$$

and

$$\alpha_n - \beta_n \geq \frac{\pi^2 s^2}{(n+1)^2}\left(1-O\left(\frac{1}{n}\right)\right).$$

i.e.,

$$e^{-(n+1)^2(\alpha_n-\beta_n)} \leq e^{-s^2\pi^2\left(1-O\left(\frac{1}{n}\right)\right)}.$$

Thus,

$$\frac{\alpha_n}{\beta_n}e^{-m(\alpha_n-\beta_n)} \leq \pi^2 s^2\left(1+O\left(\frac{1}{s^2}\right)\right)e^{-s^2\pi^2\left(1-O\left(\frac{1}{s^2}\right)\right)}.e^{-(m-(n+1)^2)(\alpha_n-\beta_n)} \to 0,$$

since $e^{-(m-(n+1)^2)(\alpha_n-\beta_n)} \to 0$ exponentially fast for $m > (1+\delta)n^2$.
If $\sqrt{n} \leq s \leq n-\sqrt{n}$, then we have $\alpha_n \geq \frac{1}{n}\left(1-O\left(\frac{1}{n^2}\right)\right)$. So

$$\alpha_n - \beta_n \geq \frac{1}{n}\left(1-O\left(\frac{1}{n^2}\right)\right) - \frac{\pi^2}{(n+1)^2}.$$

This is because $\sin(x) \leq x$ for all x. Moreover, $-\ln\alpha_n \geq 0$, since $\alpha_n \leq 1$. Also,

$\ln \beta_n \sim -2 \ln\left(\frac{n+1}{\pi}\right)$. Therefore, we have

$$\begin{aligned} 3(n+1)\ln\left(\frac{n+1}{\pi}\right)(\alpha_n-\beta_n)-\ln\alpha_n+\ln\beta_n &\geq 3(n+1)\ln\left(\frac{n+1}{\pi}\right)\left(1-O\left(\frac{1}{n^2}\right)\right) \\ &\quad -\frac{3\pi^2}{n+1}\ln\left(\frac{n+1}{\pi}\right)-2\ln\left(\frac{n+1}{\pi}\right) \\ &\geq \ln\left(\frac{n+1}{\pi}\right)\left(3-O\left(\frac{1}{n^2}\right)-\frac{3\pi^2}{n+1}-2\right) \\ &= \ln\left(\frac{n+1}{\pi}\right)\left(1-\left(O\left(\frac{1}{n^2}\right)+\frac{3\pi^2}{n+1}\right)\right) \\ &> 0, \quad \text{for large } n. \end{aligned}$$

Therefore,

$$3(n+1)\ln\left(\frac{n+1}{\pi}\right)(\alpha_n-\beta_n) > \ln\alpha_n-\ln\beta_n.$$

So we get,

$$\begin{aligned} e^{-m(\alpha_n-\beta_n)+\ln\alpha_n-\ln\beta_n} &\leq e^{-m(\alpha_n-\beta_n)+3(n+1)\ln\left(\frac{n+1}{\pi}\right)(\alpha_n-\beta_n)} \\ &= e^{-\left(m-3(n+1)\ln\left(\frac{n+1}{\pi}\right)\right)(\alpha_n-\beta_n)} \\ &\to 0, \end{aligned}$$

as $n \to \infty$, since $\alpha_n > \beta_n$ and $m > n^2$.

Theorem 2.4.6. *Let $m \sim tn^\kappa$ with $\kappa > 2$. We have*

$$\lim_{n\to\infty} \frac{D_{t,n,\kappa}}{C_{tn^\kappa}} \longrightarrow 0.$$

Proof: From Proposition 2.4.1,

$$\begin{aligned} D_{t,n,\kappa} &= \sum_{s=1}^{n}\left(\frac{2}{n+1}\right)\sin^2\left(\frac{\pi s}{n+1}\right)\left(2\cos\left(\frac{\pi s}{n+1}\right)\right)^{2tn^\kappa} \\ &= \frac{2}{n+1}4^{tn^\kappa}\left(2\sin^2\left(\frac{\pi}{n+1}\right)\cos\left(\frac{\pi}{n+1}\right)^{2tn^\kappa}+A\right), \end{aligned}$$

where the $\lim_{n\to\infty} A \to 0$ exponentially fast.

$$\begin{aligned}\lim_{n\to\infty}\frac{D_{t,n,\kappa}}{C_{tn^\kappa}} &= \lim_{n\to\infty}\frac{\frac{2}{n+1}4^{tn^\kappa}\left(2\sin^2\left(\frac{\pi}{n+1}\right)\cos\left(\frac{\pi}{n+1}\right)^{2tn^\kappa}+A\right)}{\frac{4^{tn^\kappa}}{\sqrt{\pi}t^{\frac{3}{2}}n^{\frac{3\kappa}{2}}}} \\ &\leq \lim_{n\to\infty}\frac{2n^{\kappa\frac{3}{2}}t^{\frac{3}{2}}\sqrt{\pi}}{(n+1)}\left(2\left(\frac{\pi}{n+1}\right)^2\exp\left(-tn^\kappa\frac{\pi^2}{(n+1)^2}+O(\frac{1}{n^4})\right)+A\right),\end{aligned}$$

which goes to zero as $n\to\infty$ exponentially fast, since $\kappa > 2$.

2.4.3 Results where $\kappa < 2$

In this section, we want to show that when $\kappa < 2$,

$$\lim_{n\to\infty}\frac{D_{t,n,\kappa}}{C_{tn^\kappa}} \to 1.$$

Lemma 2.4.6. *Let $f(x) = \sin^2(x)\cos^{2m}(x)$ for all m. Then $f(x)$ is a nonnegative, real-valued function, and reaches its maximum value at $x = \pm\arctan(\frac{1}{\sqrt{m}})$,*

$$f\left(\pm\arctan(\frac{1}{\sqrt{m}})\right) = \frac{1}{(1+\frac{1}{m})^{m+1}m},$$

and reaches its minimum value at $x = 0$, $f(0) = 0$.

Proof: Obviously $f(x) \geq 0$, since $\sin(x)\cos^m(x)^2 \geq 0$. Thus,

$$\min f(x) = f(0) = 0.$$

Also,

$$\begin{aligned}f'(x) &= 2\sin(x)\cos^{2m+1}(x) - (2m)\cos^{2m-1}\sin^3(x) \\ &= 2\sin(x)\cos^{2m-1}(x)\left[\cos^2(x) - m\sin^2(x)\right],\end{aligned}$$

which equals zero only if $x = 0,\ \pm\arctan(\frac{1}{\sqrt{m}})$. Thus

$$\max f(x) = f\left(\pm\arctan(\frac{1}{\sqrt{m}})\right) = \frac{1}{(1+\frac{1}{m})^{m+1}m}.$$

This behaves asymptotically as $\frac{1}{(em)}$.

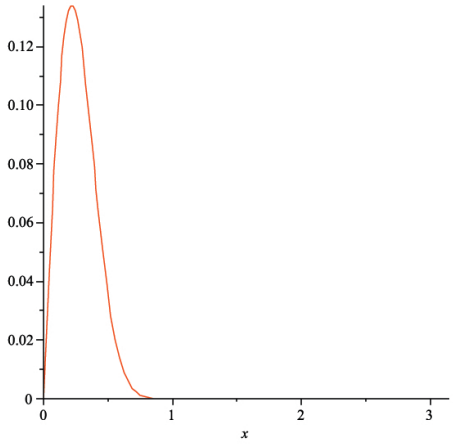


Figure 2.10: $f(x) = \sin^2(x)\cos^{2m}(x),\ x = 0..\pi,\ m = 10$

Lemma 2.4.7. *Let* $f(x) = \sin^2(x)\cos^{2m}(x)$, *where* $m \sim tn^{\kappa}$ *and* $\kappa < 2$. *As* $n \to \infty$,

$$\left| \sum_{s=0}^{n+1} \frac{\pi}{n+1} \sin^2\left(\frac{\pi s}{n+1}\right) \cos^{2m}\left(\frac{\pi s}{n+1}\right) - \int_0^{\pi} f(x)dx \right| \to 0.$$

Proof: We know for a real-valued function on the interval $[a, b]$, we have

$$\int_a^b f(x)\ dx = \sum_{k=1}^{n} f(x_k)\Delta x_k + \text{error},$$

where $a < x < b$ and the error is the difference between the upper and lower estimate, which is less than

$$\sum_{k=1}^{n} \left| \frac{f(x_{k+1}) - f(x_k)}{n} \right|.$$

Thus,

$$\begin{aligned}\left|\sum_{s=0}^{n+1}\frac{\pi}{n+1}\sin^2\left(\frac{\pi s}{n+1}\right)\cos^{2m}\left(\frac{\pi s}{n+1}\right)-\int_0^\pi f(x)dx\right| &< \sum_{k=0}^{n}\left|\frac{f(x_{k+1})-f(x_k)}{n+1}\right| \\ &< \left|\frac{f(x_{n+1})-f(x_0)}{n+1}\right| \\ &< \left|\frac{f_{\max}-f_{\min}}{n+1}\right| \\ &= \frac{f_{\max}}{n+1} \\ &= \frac{1}{me(n+1)}.\end{aligned}$$

As $m \sim tn^\kappa$, the last goes to zero very fast.

We also need to show that the error term is much smaller than the value of integral. From Lemma 2.4.7 we get,

$$\begin{aligned}\lim_{n\to\infty}\sum_{s=0}^{n+1}\frac{\pi}{n+1}\sin^2\left(\frac{\pi s}{n+1}\right)\cos^{2m}\left(\frac{\pi s}{n+1}\right) &= \int_0^\pi \sin^2(x)\cos^{2m}(x)dx \\ &= \frac{\pi(2m-1)!!}{(2m+2)!!},\end{aligned}$$

where the double factorial for a positive integer m is defined by

$$m!! = \begin{cases} m(m-2)\ldots 5.3.1, & m>0 \text{ is odd;} \\ m(m-2)\ldots 6.4.2, & m>0 \text{ is even;} \\ 1, & m=0. \end{cases}$$

Lemma 2.4.8. *We have,*

$$\lim_{n\to\infty}\frac{(2n-1)!!}{(2n)!!}\sqrt{\pi n}=1.$$

Proof: We have

$$\begin{aligned}(2n)!! &= (2n)(2n-2)(2n-4)\ldots 2 \\ &= 2^n[n(n-1)(n-2)\ldots 1] \\ &= 2^n n!\end{aligned}$$

On the other hand we have,

$$\begin{aligned}(2n-1)!!(2n)!! &= [(2n-1)(2n-3)\dots 1][(2n)(2n-2)\dots 2] \\ &= [(2n)(2n-1)(2n-2)\dots(2)(1)] \\ &= (2n)!\end{aligned}$$

Hence,

$$\begin{aligned}(2n-1)!! &= \frac{(2n)!}{(2n)!!} = \frac{(2n)!}{2^n n!} \qquad \text{so,} \\ \frac{(2n-1)!!}{(2n)!!} &= \frac{(2n)!}{(2^n n!)^2}.\end{aligned}$$

The Stirling approximation yields, $n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$, we have

$$\begin{aligned}\frac{(2n)!}{(2^n n!)^2} &\sim \frac{\sqrt{4\pi n}\left(\frac{2n}{e}\right)^{2n}}{4^n 2\pi n\left(\frac{n}{e}\right)^{2n}} \\ &= \frac{1}{\sqrt{\pi n}}\,.\end{aligned}$$

Hence,

$$\begin{aligned}\int_0^\pi \sin^2(x)\cos^{2m}(x)dx &= \frac{\pi(2m-1)!!}{(2m+2)!!} \\ &= \frac{\pi(2m-1)!!}{(2m+2)(2m)!!} \\ &\sim \frac{\pi}{2(m+1)}\frac{1}{\sqrt{\pi m}} \\ &\sim \frac{\sqrt{\pi}}{2(m+1)\sqrt{m}}.\end{aligned}$$

Assuming $m \sim tn^\kappa$, the following inequality holds if only $\kappa < 2$,

$$\frac{1}{me(n+1)} < \frac{\sqrt{\pi}}{2(m+1)\sqrt{m}}.$$

This shows that the error term is much smaller than the value of the integral.

Theorem 2.4.7. *If $\kappa < 2$, then*

$$\lim_{n\to\infty} \frac{D_{t,n,\kappa}}{C_{tn^\kappa}} \longrightarrow 1$$

where $D_{t,n,\kappa} = \sum_{s=1}^{n} \left(\frac{2}{n+1}\right) \sin^2\left(\frac{\pi s}{n+1}\right) \left(2\cos\left(\frac{\pi s}{n+1}\right)\right)^{2tn^\kappa}$ *and* C_{tn^κ} *is the* (tn^κ) *th Catalan number.*

Proof: Let $m \sim tn^\kappa$; then from Lemma 2.4.7 and Lemma 2.4.8 we have that for sufficiently large n,

$$\begin{aligned}
\sum_{s=1}^{n} \frac{\pi}{n+1} \sin^2\left(\frac{\pi s}{n+1}\right) \cos^{2m}\left(\frac{\pi s}{n+1}\right) &= \sum_{s=0}^{n+1} \frac{\pi}{n+1} \sin^2\left(\frac{\pi s}{n+1}\right) \cos^{2m}\left(\frac{\pi s}{n+1}\right) \\
&= \int_0^\pi \sin^2(x)\cos^{2m}(x)dx \\
&= \frac{\pi(2m-1)!!}{(2m+2)!!} \\
&= \frac{\pi(2m-1)!!}{(2m+2)(2m)!!} \\
&\sim \frac{\pi}{2(m+1)} \frac{1}{\sqrt{\pi m}} \\
&= \frac{\sqrt{\pi}}{2\left(t^{\frac{3}{2}} n^{\frac{3\kappa}{2}} + t^{\frac{1}{2}} n^{\frac{\kappa}{2}}\right)}.
\end{aligned}$$

Thus,

$$\begin{aligned}
\lim_{n\to\infty} \frac{D_{t,n,\kappa}}{C_{tn^\kappa}} &= \lim_{n\to\infty} \frac{4^{tn^\kappa} \sum_{s=0}^{n+1} \frac{2}{n+1} \sin^2\left(\frac{\pi s}{n+1}\right) \cos^{2m}\left(\frac{\pi s}{n+1}\right)}{\frac{4^{tn^\kappa}}{\sqrt{\pi} t^{\frac{3}{2}} n^{\frac{3\kappa}{2}}}} \\
&= \lim_{n\to\infty} \frac{4^{tn^\kappa} \frac{2}{\pi} \sum_{s=0}^{n+1} \frac{\pi}{n+1} \sin^2\left(\frac{\pi s}{n+1}\right) \cos^{2m}\left(\frac{\pi s}{n+1}\right)}{\frac{4^{tn^\kappa}}{\sqrt{\pi} t^{\frac{3}{2}} n^{\frac{3\kappa}{2}}}} \\
&= \lim_{n\to\infty} \frac{2}{\sqrt{\pi}} t^{\frac{3}{2}} n^{\frac{3\kappa}{2}} \left[\frac{\sqrt{\pi}}{2\left(t^{\frac{3}{2}} n^{\frac{3\kappa}{2}} + t^{\frac{1}{2}} n^{\frac{\kappa}{2}}\right)}\right] \\
&= \lim_{n\to\infty} \left[\frac{t^{\frac{3}{2}} n^{\frac{3\kappa}{2}}}{\left(t^{\frac{3}{2}} n^{\frac{3\kappa}{2}} + t^{\frac{1}{2}} n^{\frac{\kappa}{2}}\right)}\right] \\
&= 1.
\end{aligned}$$

2.5 Large and moderate deviation principles for the law of the maximum of a random Dyck path

Large deviation theory deals with the decay of the probability of increasingly unlikely events. Let $X_1, X_2, \cdots$ be independent identically distributed random variables with mean μ satisfying

$$\varphi(\lambda) := \log \mathbb{E}e^{\lambda X} < \infty,$$

and assume S_n is their partial sums, i.e.,

$$S_n := \sum_{i=1}^{n} X_i.$$

The probability of large deviation events goes to zero exponentially fast by an order n. That is, the S_n goes beyond its average, μ by more than nx, i.e., $\{S_n - \mu n \geq nx\}$ decays exponentially for $x > 0$. Cramér's theorem tells exactly how fast it goes to zero.

Theorem 2.5.1 (Cramér's theorem)**.** *For any $x > \mu$ we have,*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left\{\frac{1}{n} S_n \geq x\right\} = -\varphi^*(x),$$

where $\varphi^*(x)$ given by

$$\varphi^*(x) := \sup_{\lambda\in\mathbb{R}} \{\lambda x - \varphi(\lambda)\}$$

is the Legendre transform of φ.

The central limit theorem declares that for large enough n, the distribution of S_n is approximately normal with mean μ and variance σ^2. For real z,

$$\lim_{n\to\infty} \mathbb{P}(\sqrt{n}(S_n - \mu) \leq z) = \Phi(z/\sigma),$$

where $\Phi(x)$ is the distribution of the standard normal law. Therefore, it tells us how fast the S_n goes beyond its average, μ, by more than $\sqrt{n}x$, i.e.,

$$\lim_{n\to\infty} \mathbb{P}(\sqrt{n}(S_n - \mu) \geq z) = 1 - \Phi(z/\sigma).$$

Now consider sequences a_n, where $\sqrt{n} \ll a_n \ll n$. Neither the central limit theorem nor Cramér's theorem tells us how fast the events of type $\{S_n - \mu n \geq x a_n\}$ decay. The moderate deviation principle deals with these events.

Theorem 2.5.2 (Moderate deviation principle)**.** *[47] Under the same assumptions as in Theorem* 2.5.1*, if $\sqrt{n} \ll a_n \ll n$ we have, for all $x > 0$,*

$$\lim_{n\to\infty} \frac{n}{a_n^2} \log \mathbb{P}\left\{S_n - \mu n \geq x a_n\right\} = -\frac{x^2}{2\sigma^2}.$$

2.5.1 Moderate deviation principle when $\kappa > 2$

In this section, we find the moderate deviation principle for the law of the maximum of a random Dyck path in the case where $\kappa > 2$.

Lemma 2.5.1. *If $\kappa > 2$,*

$$\lim_{n\to\infty} \left(\frac{1}{2tn^\kappa}\right)^{\frac{\kappa-2}{\kappa}} \log \sum_{s=1}^{n} \sin^2\left(\frac{\pi s}{n+1}\right) \cos^{2tn^\kappa}\left(\frac{\pi s}{n+1}\right) \longrightarrow -\frac{-\pi^2 (2t)^{\frac{2}{\kappa}}}{2}.$$

Proof: From Proposition 2.4.1 we have,

$$\sum_{s=1}^{n} \sin^2\left(\frac{\pi s}{n+1}\right) \cos^{2tn^\kappa}\left(\frac{\pi s}{n+1}\right) < n \sin^2\left(\frac{\pi}{n+1}\right) \cos^{2tn^\kappa}\left(\frac{\pi}{n+1}\right).$$

Now, apply the logarithm to both sides

$$\log \sum_{s=1}^{n} \sin^2\left(\frac{\pi s}{n+1}\right) \cos^{2tn^\kappa}\left(\frac{\pi s}{n+1}\right) < \log n + \log \sin^2\left(\frac{\pi}{n+1}\right) \cos^{2tn^\kappa}\left(\frac{\pi}{n+1}\right).$$

So

$$0 < \log \sum_{s=1}^{n} \sin^2\left(\frac{\pi s}{n+1}\right) \cos^{2tn^\kappa}\left(\frac{\pi s}{n+1}\right) - \log \sin^2\left(\frac{\pi}{n+1}\right) \cos^{2tn^\kappa}\left(\frac{\pi}{n+1}\right) < \log n.$$

Divide both sides by $\left(\frac{1}{2tn^\kappa}\right)^{\frac{\kappa-2}{\kappa}}$ and $n \to \infty$; we obtain

$$0 \leq \lim_{n\to\infty} \left(\tfrac{1}{2tn^\kappa}\right)^{\frac{\kappa-2}{\kappa}} \left(\log \textstyle\sum_{s=1}^{n} \sin^2\left(\tfrac{\pi s}{n+1}\right) \cos^{2tn^\kappa}\left(\tfrac{\pi s}{n+1}\right) - \log \sin^2\left(\tfrac{\pi}{n+1}\right) \cos^{2tn^\kappa}\left(\tfrac{\pi}{n+1}\right)\right) \leq 0.$$

Therefore,

$$\lim_{n\to\infty} \left(\frac{1}{2tn^\kappa}\right)^{\frac{\kappa-2}{\kappa}} \log \sum_{s=1}^{n} \sin^2\left(\frac{\pi s}{n+1}\right) \cos^{2tn^\kappa}\left(\frac{\pi s}{n+1}\right)$$

$$= \lim_{n\to\infty} \left(\frac{1}{2tn^\kappa}\right)^{\frac{\kappa-2}{\kappa}} \log \sin^2\left(\frac{\pi}{n+1}\right) \cos^{2tn^\kappa}\left(\frac{\pi}{n+1}\right).$$

We know that for large values of n, we have $\cos(\frac{\pi}{n+1}) \sim \exp\left(-\frac{\pi^2}{2(n+1)^2}\right)$; thus

$$\lim_{n\to\infty} \left(\tfrac{1}{2tn^\kappa}\right)^{\frac{\kappa-2}{\kappa}} \log \sin^2\left(\tfrac{\pi}{n+1}\right) \cos^{2tn^\kappa}\left(\tfrac{\pi}{n+1}\right) =$$

$$\lim_{n\to\infty} \left(\tfrac{1}{2tn^\kappa}\right)^{\frac{\kappa-2}{\kappa}} \left(\log \sin^2\left(\tfrac{\pi}{n+1}\right) + \log \cos^{2tn^\kappa}\left(\tfrac{\pi}{n+1}\right)\right) =$$

$$\lim_{n\to\infty} \left(\tfrac{1}{2tn^\kappa}\right)^{\frac{\kappa-2}{\kappa}} \left(\log\left(\exp\left(\tfrac{-\pi^2 2tn^\kappa}{2(n+1)^2}\right)\right)\right) = -\tfrac{\pi^2 (2t)^{\frac{2}{\kappa}}}{2}.$$

Lemma 2.5.2. *If $\kappa > 2$, then*

$$\lim_{n\to\infty} \left(\frac{1}{2tn^\kappa}\right)^{\frac{\kappa-2}{\kappa}} \log\left(\frac{D_{t,n,\kappa}}{C_{tn^\kappa}}\right) = -\frac{\pi^2 (2t)^{\frac{2}{\kappa}}}{2}.$$

Proof: We have

$$\begin{aligned}
\lim_{n\to\infty} \left(\frac{1}{2tn^\kappa}\right)^{\frac{\kappa-2}{\kappa}} \log\left(\frac{D_{t,n,\kappa}}{C_{tn^\kappa}}\right) &= \lim_{n\to\infty} \left(\frac{1}{2tn^\kappa}\right)^{\frac{\kappa-2}{\kappa}} \log\left(\frac{\frac{2}{n+1} 4^{tn^\kappa} \sum_{s=1}^{n} \sin^2\left(\frac{\pi s}{n+1}\right) \cos\left(\frac{\pi}{n+1}\right)^{2tn^\kappa}}{\frac{4^{tn^\kappa}}{\sqrt{\pi} t^{\frac{3}{2}} n^{\frac{3\kappa}{2}}}}\right) \\
&= \lim_{n\to\infty} \left(\frac{1}{2tn^\kappa}\right)^{\frac{\kappa-2}{\kappa}} \log(2\sqrt{\pi} t^{\frac{3}{2}} n^{\frac{3\kappa}{2}-1}) \\
&+ \lim_{n\to\infty} \left(\frac{1}{2tn^\kappa}\right)^{\frac{\kappa-2}{\kappa}} \log\left(\sum_{s=1}^{n} \sin^2\left(\frac{\pi s}{n+1}\right) \cos\left(\frac{\pi}{n+1}\right)^{2tn^\kappa}\right).
\end{aligned}$$

The first term in the last equality goes to zero, so by lemma 2.5.1 we have,

$$\lim_{n\to\infty}\left(\frac{1}{2tn^{\kappa}}\right)^{\frac{\kappa-2}{\kappa}}\log\left(\frac{D_{t,n,\kappa}}{C_{tn^{\kappa}}}\right)=-\frac{\pi^2(2t)^{\frac{2}{\kappa}}}{2}.$$

Theorem 2.5.3. *Let $\kappa > 2$, $N = 2tn^{\kappa}$ and $x^{\kappa} = \frac{1}{2t}$. Moreover assume D_N is a Dyck path of length N. Then*

$$\lim_{N\to\infty}\frac{1}{N^{\frac{\kappa-2}{\kappa}}}\log\mathbb{P}[\max_{N} D_N \le x.N^{\frac{1}{\kappa}}]=-\frac{\pi^2}{2x^2}$$

Proof: The tn^{κ}th Catalan number gives the number of Dyck paths of length tn^{κ} and $D_{t,n,\kappa}$ gives the number of Dyck paths of length tn^{κ} with maximum height n. So,

$$\mathbb{P}[\max_{n}\ \text{height of Dyck paths in } 2tn^{\kappa} \text{ steps} \le n]=\frac{D_{t,n,\kappa}}{C_{tn^{\kappa}}}.$$

By lemmas 2.5.1 and 2.5.2,

$$\lim_{N\to\infty}\frac{1}{N^{\frac{\kappa-2}{\kappa}}}\log\mathbb{P}[\max_{n} D_N \le x.N^{\frac{1}{\kappa}}]\longrightarrow-\frac{\pi^2}{2x^2}.$$

Lemma 2.5.3. *Assume $a_n \gg n^2$; then*

$$\lim_{n\to\infty}\frac{n^2}{2a_n}\log\frac{D_{a_n,n}}{C_n}=-\frac{\pi^2}{2}.$$

Proof: We have,

$$\frac{D_{a_n,n}}{C_n}=\frac{\sum_{s=1}^{n}\frac{2}{n+1}\sin^2\left(\frac{\pi s}{n+1}\right)4^{a_n}\cos^{2a_n}\left(\frac{\pi s}{n+1}\right)}{\frac{4^{a_n}}{\sqrt{\pi}a_n^{\frac{3}{2}}}}.$$

From Proposition 2.4.1,

$$\begin{aligned}\sum_{s=1}^{n}\frac{2}{n+1}\sin^2\left(\frac{\pi s}{n+1}\right)4^{a_n}\cos^{2a_n}\left(\frac{\pi s}{n+1}\right) &= 2\frac{2}{n+1}\sin^2\left(\frac{\pi}{n+1}\right)4^{a_n}\cos^{2a_n}\left(\frac{\pi}{n+1}\right)\\ &\sim \frac{4}{n+1}\left(\frac{\pi}{n+1}\right)^2 4^{a_n}\exp\left(\frac{-\pi^2}{2(n+1)^2}\right)^{2a_n}\end{aligned}$$

Thus,

$$\begin{aligned}\frac{D_{a_n,n}}{C_n} &= \frac{4}{n+1}\sqrt{\pi}a_n{}^{\frac{3}{2}}\left(\frac{\pi}{n+1}\right)^2\exp\left(\frac{-\pi^2}{2(n+1)^2}\right)^{2a_n};\qquad \text{therefore}\\ \log\left(\frac{D_{a_n,n}}{C_n}\right) &= \log\left(\frac{4}{(n+1)^3}\pi^{\frac{3}{2}}a_n{}^{\frac{3}{2}}\right)-\frac{\pi^2 2a_n}{2(n+1)^2}.\end{aligned}$$

Now divide both sides by $\frac{n^2}{2a_n}$ and let n tend toward infinity; then

$$\lim_{n\to\infty} \frac{n^2}{2a_n} \log \frac{D_{a_n,n}}{C_n} = -\frac{\pi^2}{2}.$$

Corollary 2.5.1. *Let* $b_n = \sqrt{\frac{2a_n}{n}}$*; then* $\sqrt{n} \ll b_n \ll n$ *and*

$$\lim_{n\to\infty} \frac{n}{{b_n}^2} \log P(\max_n \textit{Dyck Path} < n) = -\frac{\pi^2}{2}.$$

2.5.2 Moderate deviation principle when $\kappa < 2$

In this section we find the moderate deviation principle for the law of the maximum of a random Dyck path in the case where $\kappa < 2$. The main goal is to show if $\kappa < 2$,

$$\lim_{N\to\infty} \frac{1}{2N^{\frac{2}{\kappa}-1}} \log P\left(\text{max of Dyck path with length } 2N > xN^{\frac{1}{\kappa}}\right) \longrightarrow -x^2$$

for $x > 0$ and an even number N. We start by finding a lower bound and an upper bound for the number of Dyck paths which hit $xN^{\frac{1}{\kappa}}$ at least once in $2N$ steps. Then we calculate the probability and show the logarithm of these two bounds approach the same value asymptomatically.

Let $r = xN^{\frac{1}{\kappa}}$. To find a lower bound, assume that a is the first time that we hit $r = xN^{\frac{1}{\kappa}}$ and assume $c = 2N - b$ is the last time we hit r and between a and c we never go below r. (Of course we have to assume a, b are both positive and are both less or equal to N. It is also obvious that this yields a lower bound, as some constraints are imposed, hence fewer Dyck paths are counted.)

To calculate the number of Dyck paths with this property, we have to calculate separately the number of paths in three different intervals, $(0, a), (a, c)$ and $(c, 2N)$ and multiply them together.

For the first interval, we calculate the number of paths starting at the origin $(0, 0)$ and ending at (a, r), never going below the x-axis.

Lemma 2.5.4. *The number of paths starting at the origin* $(0, 0)$ *ending at* (a, r)*, and never going below* x*-axis is equal to* $\binom{a}{\frac{a+r}{2}} \frac{r+1}{\frac{a+r}{2}+1}$.

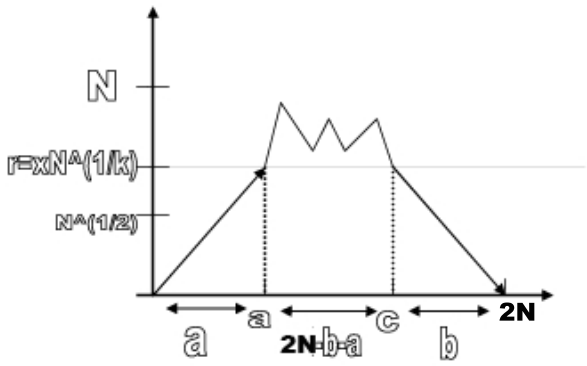


Figure 2.11: Finding a lower bound

Proof: We know from [21] (p. 75), the number of all paths from the origin to the point (a, r) is given by

$\binom{a}{\frac{a+r}{2}}$ and this is the number of good paths plus the number of bad paths.

Good paths are those which never go below x-axis and we want to calculate their cardinality. We instead calculate the number of bad paths, those which at some point go below x-axis. For calculating this, we use the *reflection principle*. Take the first point where the path hits the line $y = -1$ and reflect the rest of the path vertically through the line.

The reflected path ends at $y = -r - 2$. Conversely any path which ends at $-r - 2$ starting at the origin must cross $y = -1$ at some point. This gives a bijection between the set of all bad paths and the set of all paths which start at the origin and stop at $(a, -r - 2)$. The second has exactly $\binom{a}{\frac{a+r}{2}+1}$ elements, since out of a steps we need exactly to take $\frac{a+r}{2} + 1$ steps downward and $\frac{a+r}{2} - 1$ upward. So we get

$$\#\text{ good paths} = \binom{a}{\frac{a+r}{2}} - \binom{a}{\frac{a+r}{2}+1}.$$

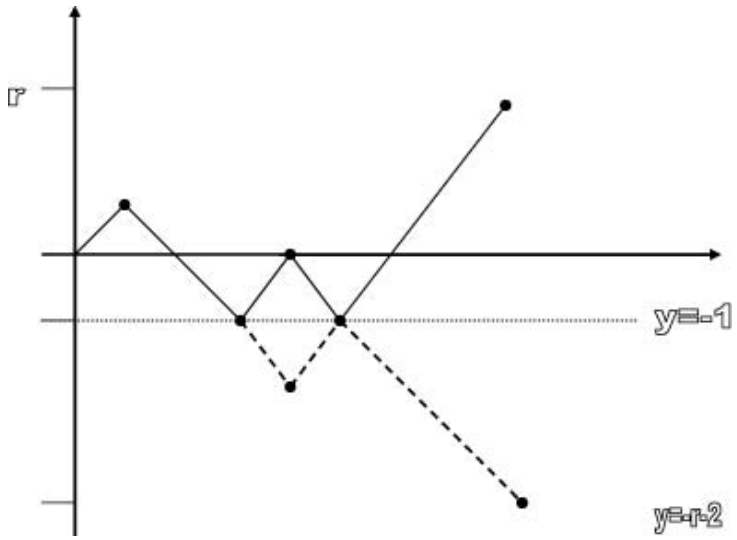


Figure 2.12: Bijection between the set of bad paths and the set of all paths which start at the origin and end at $(a, -r-2)$

We have,

$$\begin{aligned}
\binom{a}{\frac{a+r}{2}} - \binom{a}{\frac{a+r}{2}+1} &= \frac{a!}{\left(\frac{a+r}{2}\right)!\left(\frac{a-r}{2}\right)!} - \frac{a!}{\left(\frac{a+r}{2}+1\right)!\left(\frac{a-r}{2}-1\right)!} \\
&= \frac{a!}{\left(\frac{a+r}{2}\right)!\left(\frac{a-r}{2}-1\right)!}\left(\frac{1}{\frac{a-r}{2}} - \frac{1}{\frac{a+r}{2}+1}\right) \\
&= \frac{a!}{\left(\frac{a+r}{2}\right)!\left(\frac{a-r}{2}-1\right)!}\left(\frac{r+1}{\left(\frac{a-r}{2}\right)\left(\frac{a+r}{2}+1\right)}\right) \\
&= \frac{a!}{\left(\frac{a+r}{2}\right)!\left(\frac{a-r}{2}\right)!}\left(\frac{r+1}{\frac{a+r}{2}+1}\right) \\
&= \binom{a}{\frac{a+r}{2}}\frac{r+1}{\frac{a+r}{2}+1}.
\end{aligned}$$

as desired.

From the same argument, it follows that the number of paths which start at (c, r), end at $(2N, 0)$, and never go below the x-axis is equal to

$$\binom{b}{\frac{b+r}{2}}\frac{r+1}{\frac{b+r}{2}+1}.$$

Lemma 2.5.5. *We have*

$$\log\left(\binom{a}{\frac{a+r}{2}}\right)\frac{r+1}{\frac{a+r}{2}+1} = \log\left(\frac{r+1}{\frac{a+r}{2}+1}\right) + a\log 2 - \frac{r^2}{a} + o(1),$$

where $r = xN^{\frac{1}{\kappa}}$.

Proof: By Stirling's formula,

$$\begin{aligned}\binom{a}{\frac{a+r}{2}} &= \frac{a^a}{\left(\frac{a+r}{2}\right)^{\frac{a+r}{2}}\left(\frac{a-r}{2}\right)^{\frac{a-r}{2}}}(1+o(1)) \\ &= \frac{a^a\ 2^a}{(a+r)^{\frac{a+r}{2}}(a-r)^{\frac{a-r}{2}}}(1+o(1)).\end{aligned}$$

Hence,

$$\begin{aligned}\log\left(\binom{a}{\frac{a+r}{2}}\right) &= a\log a + a\log 2 - \left(\frac{a+r}{2}\right)\log a\left(1+\frac{r}{a}\right) - \left(\frac{a-r}{2}\right)\log a\left(1-\frac{r}{a}\right) + o(1) \\ &= a\log a + a\log 2 - a\log a - \left(\frac{a+r}{2}\right)\left(\frac{r}{a}\right) - \left(\frac{a-r}{2}\right)\left(-\frac{r}{a}\right) + o(1) \\ &= a\log 2 - \frac{r^2}{2a} - \frac{r^2}{2a} + o(1) \\ &= a\log 2 - \frac{r^2}{a} + o(1).\end{aligned}$$

Therefore,

$$\log\left(\binom{a}{\frac{a+r}{2}}\right)\frac{r+1}{\frac{a+r}{2}+1} = \log\left(\frac{r+1}{\frac{a+r}{2}+1}\right) + a\log 2 - \frac{r^2}{a} + o(1).$$

To calculate the second interval, the number of paths which start at (a, r), end at (c, r), and never go below r, notice that this number is exactly equal to the number of Dyck paths in $2N - a - b$ steps. We know this number is equal to $C_{\frac{2N-b-a}{2}}$, where C_n is the nth Catalan number. Asymptotically,

$$\begin{aligned}C_{\frac{2N-b-a}{2}} &= \frac{4^{\frac{2N-b-a}{2}}}{\sqrt{\pi}\left(\frac{2N-b-a}{2}\right)^{\frac{3}{2}}}\left(1+O\left(\frac{1}{N}\right)\right) \\ &= \frac{2^{2N-b-a}}{\sqrt{\pi}\left(\frac{2N-b-a}{2}\right)^{\frac{3}{2}}}\left(1+O\left(\frac{1}{N}\right)\right)\end{aligned}$$

Note that in the case where $a = b = N$ this part will disappear.

Thus we obtain a lower bound for the number of Dyck paths with maximum greater than $xN^{\frac{1}{\kappa}}$,

$$\binom{a}{\frac{a+r}{2}} \frac{r+1}{\frac{a+r}{2}+1} \frac{2^{2N-b-a}}{\sqrt{\pi}\left(\frac{2N-b-a}{2}\right)^{\frac{3}{2}}} \binom{b}{\frac{b+r}{2}} \frac{r+1}{\frac{b+r}{2}+1} \qquad (*).$$

Lemma 2.5.6. *For $\kappa < 2$ we have,*

$$\lim_{N\to\infty} \frac{1}{2N^{\frac{2}{\kappa}-1}} \log P\left(\max \text{ of Dyck path with length } 2N > xN^{\frac{1}{\kappa}}\right) \geq -x^2.$$

Proof: By (*),

$$P\left(\max \text{ of Dyck path with length } 2N > xN^{\frac{1}{\kappa}}\right) \geq$$

$$\frac{\binom{a}{\frac{a+r}{2}} \binom{r+1}{\frac{a+r}{2}+1} \left(\frac{2^{2N-b-a}}{\sqrt{\pi}\left(\frac{2N-b-a}{2}\right)^{\frac{3}{2}}}\right) \binom{b}{\frac{b+r}{2}} \binom{r+1}{\frac{b+r}{2}+1}}{2^{2N}}.$$

By Lemma 2.5.5,

$$\log \frac{\binom{a}{\frac{a+r}{2}} \binom{r+1}{\frac{a+r}{2}+1} \left(\frac{2^{2N-b-a}}{\sqrt{\pi}\left(\frac{2N-b-a}{2}\right)^{\frac{3}{2}}}\right) \binom{b}{\frac{b+r}{2}} \binom{r+1}{\frac{b+r}{2}+1}}{2^{2N}} =$$

$$\log\left(\frac{r+1}{\frac{a+r}{2}+1}\right) + a\log 2 - \frac{r^2}{a} + (2N-b-a)\log 2 - \log\sqrt{\pi}\left(\frac{2N-b-a}{2}\right)^{\frac{3}{2}} + \log\left(\frac{r+1}{\frac{b+r}{2}+1}\right) +$$

$$b\log 2 - \frac{r^2}{b} - 2N\log 2 + O\left(\frac{1}{N}\right) =$$

$$\log\left(\frac{r+1}{\frac{a+r}{2}+1}\right) + \log\left(\frac{r+1}{\frac{b+r}{2}+1}\right) - r^2\left(\frac{1}{a}+\frac{1}{b}\right) - \log\sqrt{\pi}\left(\frac{2N-b-a}{2}\right)^{\frac{3}{2}} + O\left(\frac{1}{N}\right).$$

Setting $a = b = N$ the last becomes

$$\log\left(\frac{r+1}{\frac{\frac{N}{2}+r}{2}+1}\right) + \log\left(\frac{r+1}{\frac{\frac{N}{2}+r}{2}+1}\right) - r^2\left(\frac{2}{N}\right) + O\left(\frac{1}{N}\right).$$

Hence,

$$\lim_{N\to\infty} \frac{1}{2N^{\frac{2}{\kappa}-1}} \log P\left(\max \text{ of Dyck path with length } 2N > xN^{\frac{1}{\kappa}}\right) \geq -x^2,$$

since $r = xN^{\frac{1}{\kappa}}$.

To find an upper bound we consider all the paths starting at the origin that hit $xN^{\frac{1}{\kappa}}$ at least once, and end at any point. Therefore we need to take $N + xN^{\frac{1}{\kappa}}$ upwards steps out of $2N$. So the number of such paths equals $\binom{2N}{N+xN^{\frac{1}{\kappa}}}$.

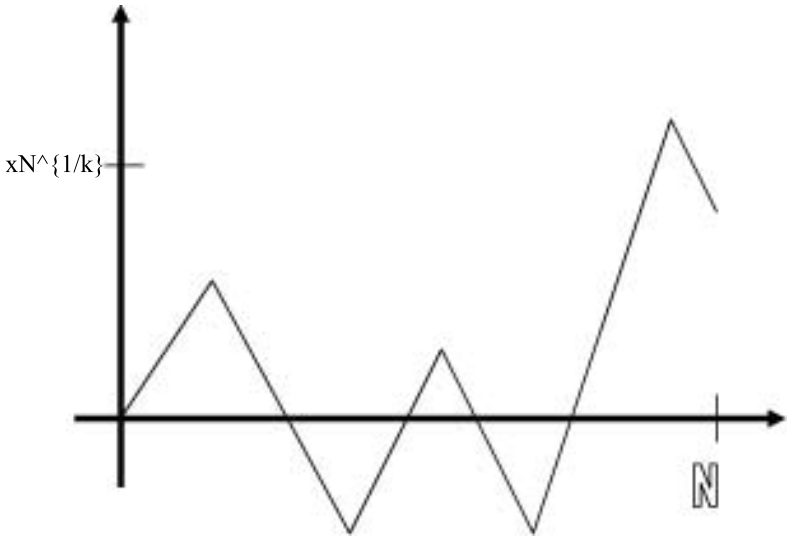


Figure 2.13: Finding an upper bound

Lemma 2.5.7. *Let $\kappa < 2$; then*

$$\lim_{N\to\infty} \frac{1}{2N^{\frac{2}{\kappa}-1}} \log P\left(\max \textit{ of Dyck path with length } 2N > xN^{\frac{1}{\kappa}}\right) \leq -x^2,$$

Proof: We know that

$$P\left(\max \text{ of Dyck path with length } 2N > xN^{\frac{1}{\kappa}}\right) \leq \frac{\binom{2N}{N+xN^{\frac{1}{\kappa}}}}{2^{2N}}.$$

We expand the log of the left hand side.

$$\begin{aligned}
\log \frac{\binom{2N}{N+xN^{\frac{1}{\kappa}}}}{2^{2N}} &= \log \frac{2N^{2N}}{\left(N+xN^{\frac{1}{\kappa}}\right)^{N+xN^{\frac{1}{\kappa}}}\left(N-xN^{\frac{1}{\kappa}}\right)^{N-xN^{\frac{1}{\kappa}}}} - 2N\log 2 + O\left(\frac{1}{N}\right) \\
&= 2N\log 2N - \left(N+xN^{\frac{1}{\kappa}}\right)\log\left(N(1+xN^{\frac{1}{\kappa}-1})\right) \\
&\quad - \left(N-xN^{\frac{1}{\kappa}}\right)\log\left(N(1-xN^{\frac{1}{\kappa}-1})\right) - 2N\log 2 + O\left(\frac{1}{N}\right)
\end{aligned}$$

We know $\log(1-xN^{\frac{1}{\kappa}-1}) \sim -xN^{\frac{1}{\kappa}-1}$ as $xN^{\frac{1}{\kappa}-1} \to 0$. Thus,

$$\begin{aligned}
\log \frac{\binom{2N}{N+xN^{\frac{1}{\kappa}}}}{2^{2N}} &= 2N\log 2N - 2N\log N - \left(N+xN^{\frac{1}{\kappa}}\right)(xN^{\frac{1}{\kappa}-1})\left(N-xN^{\frac{1}{\kappa}}\right)(-xN^{\frac{1}{\kappa}-1}) \\
&\quad -2N\log 2 + O(\frac{1}{N}) \\
&= 2N\log 2N - 2N\log N - x^2N^{\frac{2}{\kappa}-1} - x^2N^{\frac{2}{\kappa}-1} - 2N\log 2 + O\left(\frac{1}{N}\right) \\
&= -2x^2N^{\frac{2}{\kappa}-1} + O\left(\frac{1}{N}\right)
\end{aligned}$$

Thus,

$$\log P\left(\max \text{ of Dyck path with length } 2N > xN^{\frac{1}{\kappa}}\right) \leq -2x^2N^{\frac{2}{\kappa}-1} + O\left(\frac{1}{N}\right),$$

Therefore,

$$\lim_{N\to\infty} \frac{1}{2N^{\frac{2}{\kappa}-1}} \log P\left(\max \text{ of Dyck path with length } N > xN^{\frac{1}{\kappa}}\right) \leq -x^2,$$

as desired.

Theorem 2.5.4. *If $\kappa < 2$; then*

$$\lim_{N\to\infty} \frac{1}{2N^{\frac{2}{\kappa}-1}} \log P\left(\max \textit{ of Dyck path with length } 2N > xN^{\frac{1}{\kappa}}\right) = -x^2,$$

for any $x > 0$ and even number N.

Proof: By Lemma 2.5.6,

$$\lim_{N\to\infty} \frac{1}{2N^{\frac{2}{\kappa}-1}} \log P\left(\max \text{ of Dyck path with length } 2N > xN^{\frac{1}{\kappa}}\right) \geq -x^2.$$

From Lemma 2.5.7 we have,

$$\lim_{N\to\infty} \frac{1}{2N^{\frac{2}{\kappa}-1}} \log P\left(\max \text{ of Dyck path with length } 2N > xN^{\frac{1}{\kappa}}\right) \leq -x^2,$$

and the result follows.

Corollary 2.5.2. *Assume $\sqrt{N} \ll a_N \ll N$, then we have*

$$\lim_{N\to\infty} \frac{N}{(a_N)^2} \log P\left(\max \textit{ of Dyck path with length } 2N > a_N\right) \to -2.$$

2.5.3 Large deviation principle

Here we describe the large deviation principle in two cases.

Theorem 2.5.5 (Large deviation principle for a random Dyck path, when $N \gg n^2$)**.** *Let n and N be positive integers. For a fixed height n,*

$$\lim_{N\to\infty} \frac{(n+1)^2}{N} \log P(\max_n \textit{Dyck Path with length } 2N < n) \longrightarrow -\pi^2.$$

Proof: We have,

$$P(\max_n \text{Dyck Path with length } 2N < n) = \frac{D_{N,n}}{C_N} = \frac{\sum_{s=1}^{n} \frac{2}{n+1} \sin^2\left(\frac{\pi s}{n+1}\right) 4^N \cos^{2N}\left(\frac{\pi s}{n+1}\right)}{\frac{4^N}{\sqrt{\pi} N^{\frac{3}{2}}}}.$$

From Proposition 2.4.1,

$$\begin{aligned}\sum_{s=1}^{n} \frac{2}{n+1} \sin^2\left(\frac{\pi s}{n+1}\right) 4^N \cos^{2N}\left(\frac{\pi s}{n+1}\right) &= 2\frac{2}{n+1} \sin^2\left(\frac{\pi}{n+1}\right) 4^N \cos^{2N}\left(\frac{\pi}{n+1}\right) \\ &= \frac{4^{N+1}}{n+1}\left(\frac{\pi}{n+1}\right)^2 \exp\left(\frac{-\pi^2}{2(n+1)^2}\right)^{2N} + o(1)\end{aligned}$$

Thus,

$$\begin{aligned}\frac{D_{N,n}}{C_N} &= \frac{4}{n+1}\sqrt{\pi} N^{\frac{3}{2}} \left(\frac{\pi}{n+1}\right)^2 \exp\left(\frac{-\pi^2}{2(n+1)^2}\right)^{2N} + o(1), \quad \text{therefore} \\ \log\left(\frac{D_{N,n}}{C_n}\right) &= \log\left(\frac{4}{(n+1)^3} \pi^{\frac{3}{2}} N^{\frac{3}{2}}\right) - \frac{\pi^2 2N}{2(n+1)^2} + o(1).\end{aligned}$$

So for n fixed,

$$\lim_{N\to\infty}\frac{1}{N}\log\left(\frac{D_{N,n}}{C_n}\right)=-\frac{\pi^2}{(n+1)^2},$$

as desired.

Theorem 2.5.6 (Large deviation principle for random Dyck path, when $n\sim N$)**.** *Let N be an even positive integer and $x>0$,*

- *For $0<x<\frac{1}{2}$*

$$\lim_{N\to\infty}\frac{1}{N}\log P\,(\max \text{ of Dyck path with length } 2N>x2N)=h(x),$$

  where $h(x)=-(1+2x)\log(1+2x)-(1-2x)\log(1-2x)$;

- *For $x\geq\frac{1}{2}$*

$$\lim_{N\to\infty}\frac{1}{N}\log P\,(\max \text{ of Dyck path with length } 2N>x2N)=-2\log 2.$$

Proof: Suppose $0<x<\frac{1}{2}$ by the argument of Lemma 2.5.7, we have

$$P\,(\max \text{ of Dyck path with length } 2N>x2N)=\frac{\binom{2N}{N+xN}}{2^{2N}}.$$

By Stirling's formula,

$$\binom{2N}{N+xN}\sim\frac{(2N)^{2N}}{(N+x2N)^{N+x2N}\,(N-x2N)^{N-x2N}}$$

We have,

$$\begin{aligned}
\log\left(\begin{array}{c}2N\\ N+2xN\end{array}\right) &\sim \log\left(\frac{(2N)^{2N}}{(N+x2N)^{N+x2N}\,(N-x2N)^{N-x2N}}\right)\\
&= 2N\log 2N-(N+x2N)\log N\,(1+2x)-(N-2xN)\log N\,(1-2x)\\
&= 2N\log 2N-2N\log N-N\,(1+2x)\log(1+2x)-N\,(1-2x)\log(1-2x)\\
&= 2N\log 2-N\,(1+2x)\log(1+2x)-N\,(1-2x)\log(1-2x)
\end{aligned}$$

Therefore for $0 < x < \frac{1}{2}$ we have,

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\log\left(\frac{\binom{2N}{1+x2N}}{2^{2N}}\right) &= \lim_{N\to\infty}\frac{1}{N}\left(-N(1+2x)\log(1+2x)-N(1-2x)\log(1-2x)\right)\\ &= -(1+2x)\log(1+2x)-(1-2x)\log(1-2x).\end{aligned}$$

For the case where $x \geq \frac{1}{2}$, we have

$$\begin{aligned}P\left(\max \text{ of Dyck path with length } 2N > x2N\right) &= P\left(\max \text{ of Dyck path with length } 2N = N\right)\\ &= \frac{1}{C_N}.\end{aligned}$$

The maximum of a Dyck path cannot be greater than half the number of steps and in the case of equality, there is only path with this property. Thus,

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\log P\left(\max \text{ of Dyck path with length } N > xN\right) &= \lim_{N\to\infty}\frac{1}{N}\left(\log\left(\frac{1}{4^{\frac{N}{2}}}\right)+o(1)\right)\\ &= -2\log 2.\end{aligned}$$

Note that,

$$\lim_{x\to\frac{1}{2}}\left(-(1+2x)\log(1+2x)-(1-2x)\log(1-2x)\right)\to -2\log 2,$$

since by l'Hôpital's rule, $\lim_{x\to\frac{1}{2}}(1-2x)\log(1-2x)=0$.

Remark 9. *Note that $h(x)$ is the entropy function of the Bernoulli distribution. Let $p = 1+2x$ and $q = 1-2x$. Then for all $0 \leq x \leq \frac{1}{2}$, $0 \leq p \leq 1$ and $p = 1-q$. So we can rewrite $h(x)$ in the previous theorem,*

$$H(p,q) = -p\log p - q\log q.$$

In general, entropy is a measure of unexpectedness. For example, when a fair coin is flipped the outcome is either heads or tails and there is no way to predict which. So fair coin toss has maximum entropy. If X is a discrete random variable with the following distribution, for $k = 1, 2, \ldots$, $\mathbb{P}(X = x_k) = p_k$, then we define the entropy of X,

$$H(X) = -\sum_{k\geq 1} p_k \log p_k.$$

Chapter 3

Simulation of high dimensional random matrices

3.1 Introduction

An interesting question that arises in the theory of random matrices (or quantum information theory), is whether one can generate random vectors, or random matrices from a given distribution function. When the target distribution comes from a standard parametric family, like the uniform, normal, etc, plenty of software exist to generate random variables. For example the method of the inverse cumulative distribution function will generate a random variable X from familiar distributions. For any continuous distribution function F, if $U \sim \text{Unif}(0,1)$, then $X = F^{-1}(U) = \inf\{x : F(x) \geq U\}$ has cumulative distribution function equal to F. If F^{-1} is available, then this strategy is probably the simplest option.

For instance, for generating $X \sim \text{Unif}(a,b)$, draw $U \sim \text{Unif}(0,1)$; then let $X = a + (b-a)U$. For the exponential distribution $\exp(\lambda)$, draw $U \sim \text{Unif}(0,1)$; then $X = -\frac{\log U}{\lambda}$.

In order to apply this method, one needs to find the inverse of the distribution func-

tion, F^{-1}. This is not always possible; in many cases, such as, the t distribution, the distribution function does not even have an explicit form.

For sampling from difficult target distributions, e.g., when f can be calculated, at least up to a proportionally constant, but f can not be sampled, we have to use other simulation approaches. Monte Carlo (MC) methods play a key role in simulating this kind of densities. A variety of techniques can be applied: the rejection method, the Gibbs sampler, and the one-dimensional Metropolis-Hastings method, for example. For these methods, one needs to find a proper proposal density function. This proposal density must be selected carefully as otherwise the rejection ratio will be very high. Sometimes it is not easy to find a very good proposal density.

Slice sampling, which is based on auxiliary variables in sampling, has been developed for improving the MC methods. In this method no proposal density is involved.

The idea of using auxiliary variables to improve MC methods was first introduced by Swendsen and Wang [60]. Edwards and Sokal [19] generalized the Swendsen-Wang method and Damien and Walker [13] provided a black-box algorithm for sampling from truncated probability density functions.

Neal [48] introduced a single auxiliary variable slice sampling method for univariate and multivariate distribution sampling. One can describe Neal's method of slice sampling from a unimodal density function proportional to f with three major steps. First, sample y uniformly from $(0, f(x_0))$, where x_0 is the current point that defines the vertical slice. Next, define the horizontal slice $S := \{x : f(x) > y\}$. The last step is to sample uniformly from an interval I which contains the S, and reject the point if it does not fall in $S \cap I$. Neal showed that the Markov chain constructed from this algorithm leaves the target distribution invariant [48].

Mira and Tierney [46] also discussed properties of slice sampling and proved that the slice sampling algorithm manages better than Metropolis-Hastings algorithm in terms of asymptotic variance in the central limit theorem. Roberts and Rosenthal [58] proved that the simple slice sampler is stochastically monotone and based on

this property, they found quantitative bounds on the convergence of slice sampling for certain classes of probability distributions. They showed that under some weak conditions slice sampling is geometrically ergodic.

The contribution here is to apply ideas involved in Neal's slice sampling to random matrices simulated according to some given densities which do not follow a familiar parametric distribution.

This chapter is organized as follows. In Section 2, we discuss the Rejection method, the Metropolis-Hastings method, and the Gibbs sampler. We also study the disadvantages of these methods.

In Section 3, we study Neal's slice sampling and its properties.

In Section 4, we discuss the method of Gibbs-slice sampling; that is based on Neal's slice sampling method and can be used to simulate multi-dimensional random vectors or large random matrices from a given unimodal density function. Then we discuss the correctness of the algorithm.

3.2 Simulating Random Matrices using Statistical Methods

Our goal is to simulate random matrices with unfamiliar distribution. In general, we would like to sample a random matrix $A_{N\times N}$ according to a density proportional to

$$A_{N\times N} = \{a_{ij}\}_{N\times N} \sim \exp\left(-\mathrm{Tr}P(A)\right)$$

where $P(A)$ can be any polynomial and Tr denotes the non-normalized trace. We are also interested in generating pair of random matrices, A and B. For example, we show how to generate a pair of self-adjoint N-dimensional random matrices $A = \{a_{ij}\}$ and $B = \{b_{ij}\}$ according to the probability

$$f(A, B) \sim e^{-\frac{N}{2}\mathrm{Tr}(A^2+B^2+\gamma AB)},$$

where $|\gamma| < 2$ is a real number. We first try to solve these questions using two different methods of sampling.

- The rejection method
- Markov Chain Monte Carlo (MCMC) slice sampler.

First we discuss the rejection method and show why is not always efficient. Then we discuss other methods that are more efficient and can be used more generally.

3.2.1 The Rejection Method

To use the rejection method, we must be able to calculate $f(x)$ at least up to a proportionality constant. Then we use rejection sampling to obtain a random draw from exactly the target distribution. With this method, we also need another density function, such as g, which is easier to sample from and calculate with. Let $e(\cdot)$ denote an envelope, having the following property,

$$e(x) = \frac{g(x)}{\alpha}, \quad \text{for all } x \text{ for which } f(x) > 0,$$

for a given constant $\alpha \leq 1$.

Ther rejection sampling proceeds as follows:

(1) Sample $Y \sim g$.

(2) Sample $U \sim \text{Unif}(0, 1)$.

(3) Reject Y if $U > \frac{f(Y)}{e(Y)}$. In this case, we do not record the value of Y as an element in the target random sample. Instead return to step 1.

(4) Otherwise, maintain the value of Y. Set $X = Y$, and consider X to be an element of the target random sample. Return to step 1 until a sample of desired size has accumulated [26].

With no approximation involved, the draw using this algorithm constitutes an i.i.d. sample from the target density f. The proof is easy:

$$\begin{aligned}
F_X(t) &= P(X \leq t) \\
&= P\left(Y \leq t | U \leq \frac{f(Y)}{e(Y)}\right) \\
&= \frac{P\left(Y \leq t, U \leq \frac{f(Y)}{e(Y)}\right)}{P\left(U \leq \frac{f(Y)}{e(Y)}\right)} \\
&= \frac{\int_0^t \left[\frac{\alpha f(y)}{g(y)}\right] g(y)dy}{\int_{-\infty}^{+\infty} \left(\int_0^{\frac{f(y)}{e(y)}} 1du\right) g(y)d(y)} \\
&= \frac{\alpha \int_0^t f(y)dy}{\int_{-\infty}^{+\infty} \frac{\alpha f(y)}{g(y)} g(y)dy} \\
&= \frac{\alpha \int_0^t f(y)dy}{\alpha} \\
&= \int_0^t f(y)dy
\end{aligned}$$

Using this algorithm we can now solve our problem. Assume $A = \{a_{ij}\}$ and $B = \{b_{ij}\}$ constitute a pair of $N \times N$ self-adjoint random matrices such that the entries on

the main diagonal have distribution Normal$(0, \frac{1}{c})$ and the entries outside the main diagonal have distribution Normal$(0, \frac{1}{2c})$ for some c in $\mathbb{R}$. Let g_1 denote the density of a matrix A, and let g_2 denote the density of a matrix B, such that the former is a joint density of all entries:

$$\begin{aligned} g_1 &\sim \prod_{i<j} \exp\left(-\frac{a_{ij}^2}{2\frac{1}{2c}}\right) \times \prod_i \exp\left(-\frac{a_{ij}^2}{2\frac{1}{c}}\right) \\ &\sim \exp\left(-c \cdot \sum_{i<j} a_{ij}^2 - c \cdot \frac{\sum_i a_{ii}^2}{2}\right) \\ &= \exp\left(-\frac{c}{2} \cdot \left(2\sum_{i<j} a_{ij}^2 + \sum_i a_{ii}^2\right)\right) \\ &= \exp\left(-\frac{c}{2} \cdot \mathrm{Tr}(A^2)\right). \end{aligned}$$

Since

$$\begin{aligned} \mathrm{Tr}(A^2) &= \sum_i \sum_j a_{ij} a_{ji} \\ &= \sum_i a_{ii}^2 + 2\sum_{i<j} a_{ij}^2, \qquad \text{therefore} \\ g_1 &\sim \exp\left(-\frac{c}{2} \cdot \mathrm{Tr}(A^2)\right) \\ g_2 &\sim \exp\left(-\frac{c}{2} \cdot \mathrm{Tr}(B^2)\right). \end{aligned}$$

Define

$$g \sim \exp\left(-\frac{c}{2} \cdot \mathrm{Tr}(A^2 + B^2)\right).$$

We have to find c, such that $g > f$, i.e.,

$$c \cdot \mathrm{Tr}(A^2 + B^2) \leq \mathrm{Tr}(A^2 + B^2 + \gamma AB);$$

This is equivalent to,

$$\begin{aligned} \mathrm{Tr}(A^2(1-c) + B^2(1-c) + \gamma AB) &> 0, \qquad \text{that is,} \\ \mathrm{Tr}(A^2 + B^2 + \frac{\gamma}{1-c} AB) &> 0. \end{aligned}$$

For each ϵ we have,

$$\begin{aligned} \mathrm{Tr}\left((A+\epsilon B)^2+(\epsilon A+B)^2\right) &> 0, \qquad \text{that is,} \\ \mathrm{Tr}\left((1+\epsilon^2)(A^2+B^2)+4\epsilon AB\right) &= (1+\epsilon^2)\mathrm{Tr}\left(A^2+B^2+\frac{4\epsilon}{1+\epsilon^2}AB\right) \\ &> 0. \end{aligned}$$

So it is enough to assume $\frac{\gamma}{1-c}=\frac{4\epsilon}{1+\epsilon^2}$. The right hand side of the equality is minimized when $\epsilon=1$, i.e., $c=1-\frac{\gamma}{2}$, and we get

$$\mathrm{Tr}(A^2+B^2+\frac{\gamma}{1-c}AB)>0,$$

where $\left|\frac{\gamma}{1-c}\right|\leq 2$, since $\left|\frac{4\epsilon}{1+\epsilon^2}\right|\leq 2$.
The disadvantage of the rejection method is that it is difficult to find a good envelope which is easy to calculate and at the same time approximates $f(x)$ so that the rejection rate is low. A bad envelope will lead to a high rejection rate and hence a long execution time. Another disadvantage arises if the acceptance region becomes very small as N increases, so we reject almost all the time.

3.2.2 MC Methods

The Markov chain method that is described in this section can be used to generate a draw from a distribution that approximates f. MC methods are better known than the rejection method, possibly owing to their iterative nature and flexility.

Definition and properties of Markov chains

Consider a sequence of random variables $\{X_n : n\in T\}$, where T is a countable time set that satisfies

$$P(X_{n+1}=x|X_1=x_1,X_2=x_2,\ldots,X_n=x_n)=P(X_{n+1}=x|X_n=x_n).$$

A collection of $X = (X_0, X_1, \ldots)$ of random variables, with each X_i taking values in a state space, $\mathfrak{X}$, are assumed measurable individually with respect to some given σ-field $\mathbb{B}(\mathfrak{X})$. A Markov chain $\{X_n\}$ is completely determined by its one-step transition probabilities:

$$P(x, A) = P(X_n \in A | X_{n-1} = x),$$

which are well-defined for appropriate initial points x and for any set $A \in \mathbb{B}(\mathfrak{X})$. The *occupation time*, η_A, is the number of visits by X_n to $A \in \mathbb{B}(\mathfrak{X})$ after time zero, and is given by

$$\eta_A = \sum_{n=1}^{\infty} \mathbb{I}\{X_n \in A\}.$$

Definition 12. *A set A is transient if there exists $M < \infty$ such that*

$$E_x[\eta_A] \leq M \quad \textit{for all } x \in A,$$

and is recurrent if

$$E_x[\eta_A] = \infty \quad \textit{for all } x \in A.$$

Definition 13. *A Markov chain $\{X_n\}$ is said to be irreducible if for each i and j,*

$$P(X_n = j | X_0 = i) = P_{ij}^n > 0,$$

i.e., if it is possible to get from one state to the other one in a finite number of steps. If for each point $x \in \mathfrak{X}$, and for every open set $O \in \mathbb{B}(\mathfrak{X})$ containing x (i.e., for every neighborhood of x), we have

$$\sum_n P^n(y, O) > 0, \quad y \in \mathfrak{X},$$

Then the chain is called open set irreducible [44].

Let $\mathfrak{F}$ be a σ-algebra on $\mathfrak{X}$. For any x in a countable state space $\mathfrak{X}$, let

$$d(x) = \gcd\{n \geq 1 : P^n(x, x) > 0\}.$$

An irreducible Markov chain is *aperiodic* if $d(x) = 1$ for all $x \in \mathfrak{X}$.
The probability distribution $\pi(\cdot)$ on $(\mathfrak{X}, \mathfrak{F})$ is *stationary* with respect to the chain, if

$$\int_{\mathfrak{X}} \pi(dx)P(x, A) = \pi(A) \quad \text{for all} \quad A \in \mathfrak{F}.$$

Markov chains with stationary distributions form the basis of MC algorithms. For the algorithm to be valid, it is crucial that the chain converge in distribution. If the state space is countable, and the Markov chain is aperiodic and also irreducible, then it is known that convergence to the stationary distribution is guaranteed from all starting states.

Theorem 3.2.1 (Ergodic Theorem)**.** *If $X_1,\ X_2, \cdots$ are realizations from an irreducible aperiodic Markov chain with stationary distribution π, then X_n converges in distribution to the distribution given by π, and for any function h,*

$$\frac{1}{n}\sum_{t=1}^{n} h(X_t) \to E_\pi\{h(X)\}$$

almost surely as $n \to \infty$, provided $E_\pi\{|h(X)|\}$ exists. [26]

On the other hand, classical irreducibility is not possible when the state space is uncountable. A weaker property is ϕ-irreducibility, having positive probability of reaching every subset A with $\phi(A) > 0$ from every state $x \in \mathfrak{X}$, for some non-trivial measure ϕ. It is known that a ϕ-irreducible, aperiodic Markov chain with stationary probability distribution $\pi(\cdot)$, must still converge to $\pi(\cdot)$ from every starting point [58, 59, 61].

ϕ-Irreducibility

The idea underlying irreducibility is that all parts of the space can be reached by a Markov chain, no matter what the starting point. For any set A in $\mathbb{B}(\mathfrak{X})$ define the

return time probabilities as follows.

$$\begin{aligned} L(x, A) &= P_x(\tau_A < \infty) \\ &= P_x(X \text{ ever enters } A), \end{aligned}$$

where $\tau_A = \min\{n \geq 1 : X_n \in A\}$ [44].
We call $X = \{X_n\}$ is *ϕ-irreducible* if there exists a measure ϕ on $\mathbb{B}(\mathfrak{X})$ such that whenever $\phi(A) > 0$, we have $L(x, A) > 0$ for all $x \in \mathfrak{X}$. [44]

Harris Recurrence

Harris recurrence is a concept introduced by Harris in 1956 [29]. A Markov chain with stationary probability distribution and which is ϕ-irreducible and aperiodic, will converge to its stationary distribution from almost any starting point. Harris recurrence allows us to replace almost all by all; this is potentially important when running MC algorithms [59].
Consider a Markov chain $\{X_n\}$ with transition probabilities $P(x, \cdot)$, on a state space $\mathfrak{X}$ with σ-algebra $\mathfrak{F}$. Let $P^n(x, \cdot)$ be the n-step transition kernel, and for $A \in \mathfrak{F}$, let $\tau_A = \inf\{n \geq 1 : X_n \in A\}$ be the first return to A [59].

Definition 14. *A Markov chain with stationary distribution $\pi(\cdot)$ is Harris recurrent if for all $A \in \mathfrak{X}$ with $\pi(A) > 0$ and all $x \in \mathfrak{X}$,*

$$L(x, A) = 1.$$

This definition means that for all $A \in \mathfrak{X}$ with $\pi(A) > 0$, the probability that the chain visits A infinitely many times for all $x \in \mathfrak{X}$ is 1.

Theorem 3.2.2. *[59] For a ϕ-irreducible, aperiodic Markov chain with stationary probability distribution $\pi(\cdot)$, the following are equivalent:*

(i) *The chain is Harris recurrent.*

(ii) *For all $A \in \mathfrak{X}$ with $\pi(A) > 0$ and all $x \in \mathfrak{X}$, we have $P(X_n \in A \text{ i.o.} | X_0 = x) = 1$.*
(i.o., infinitely often, meaning infinitely many different times n)

(iii) *For all $x \in \mathfrak{X}$, and all $A \in \mathfrak{F}$ with $\pi(A) = 1$, $P(\tau_A < \infty | X_0 = x) = 1$.*

(iv) *For all $x \in \mathfrak{X}$, and all $A \in \mathfrak{F}$ with $\pi(A) = 0$, $P(X_n \in A \text{ for all } n | X_0 = x) = 0$.*

3.2.3 Simulation with Monte Carlo method

When a target density f can be evaluated but not easily sampled, we use the MC method. The MC sampling strategy is to construct an irreducible aperiodic Markov chain for which the stationary distribution equals the target distribution f. A very general method for constructing a Markov chain is the Metropolis-Hastings algorithm [30]. The Metropolis-Hastings algorithm can draw samples from any probability distribution $f(x)$, requiring only that a function proportional to the density can be calculated at x. We start at $t = 0$ with selection $X^{(0)}$ drawn randomly from a proposal distribution g, such that $f(X^{(0)}) > 0$. Given $X^{(t)}$, the algorithm generates $X^{(t+1)}$ as follows:

(1) Sample a candidate value X^* from proposal distribution $g(\cdot|X^{(t)})$.

(2) Compute the Metropolis-Hasting ratio $R(X^{(t)}, X^*)$, where

$$R(u, v) = \frac{f(v)g(u|v)}{f(u)g(v|u)}.$$

(3) Sample a value for $X^{(t+1)}$ according to the following:

$$X^{(t+1)} = \begin{cases} X^* & \text{with probability } \min\left(R(X^{(t)}, X^*), 1\right) ; \\ X^{(t)} & \text{otherwise.} \end{cases}$$

In this step, each time we generate U_t from Uniform[0,1], if $U_t \leq \min\left(R(X^{(t)}, X^*), 1\right)$, then we let $X^{(t+1)} = X^*$; otherwise we let $X^{(t+1)} = X^{(t)}$

(4) Increase t and return to step 1.[26]

Gibbs Sampling

The Gibbs sampling, described by two brothers , Stuart and Donald Geman [25], is an MC method which allows us to sample form multidimensional target distributions. The goal is to construct a Markov chain whose stationary distribution equals the target distribution f by updating component by component. Let $X = (X_1, \cdots, X_p)$ and let $X_{-i} = (X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_p)$. Suppose that the conditional density of $X_i|X_{-i} = x_{-i}$, denoted $f(x_i|x_{-i})$, is easily sampled for $i = 1, \cdots, p$. Then from a starting value x^0, a Gibbs sampler procedure can be described as follows for iterate t: [26]

(1) Choose an ordering of the components of x^t.

(2) For each i in the ordering chosen above, sample $X^*|x^t_{-i} \sim f(x_i|x^t_{-i})$.

(3) Once step 2 has been completed for each component of X in the selected order set $X^{t+1} = X^*$.

It is known that a Markov chain constructed from the Gibbs sampler is ϕ-irreducible and aperiodic, where ϕ is Lebesgue measure, e.g., see [9, 42, 56].

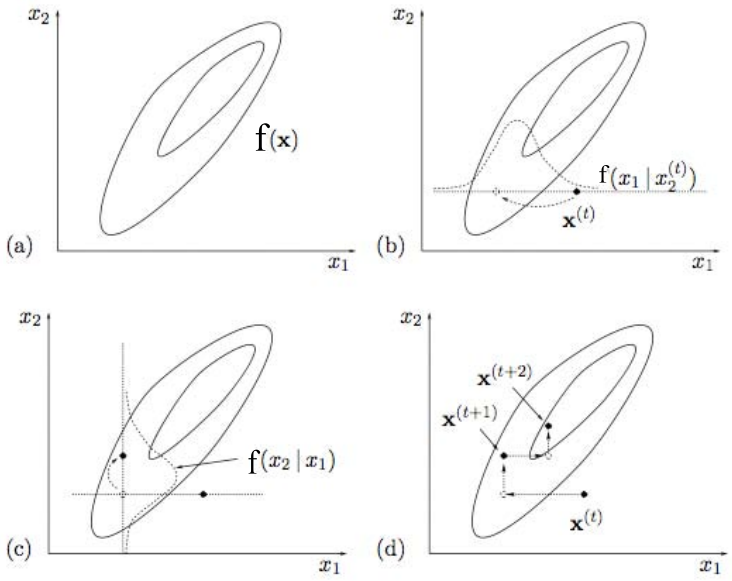


Figure 3.1: Gibbs Sampler

3.3 Slice sampling

Methods such as Gibbs sampling and the Metropolis algorithm are the most commonly used MC methods for sampling from many different complex multivariate distributions. However, to perform Gibbs sampling, one needs to know how to sample from all the required conditional distributions. Also to implement the Metropolis algorithm, we must find an appropriate proposal distribution to achieve efficient sampling. Even in one-dimensional cases, it is not always easy to find a suitable proposal density. In this section, we use another MC method which is easier than Gibbs sampling and more efficient than the Metropolis updates.

Slice sampling is a type of MC algorithm used to draw samples from a statistical distribution. It is a means of sampling uniformly from the region under a graph proportional to the density function [48].

3.3.1 History of slice sampling

Neal [48] developed an efficient algorithm for sampling from complex distributions using an auxiliary variable method. The idea of this method was first introduced by two physicists, Trotter and Tukey [62]. Edwards and Sokal [19] improved on the idea, and in 1993 Besag and Green [1] improved the efficiency of the simulation process. As a special case of the auxiliary variable method, where a single auxiliary variable is used, slice sampling was first studied by Roberts and Rosenthal [57] in 1999, who proved that the algorithm is geometrically ergodic under some weak conditions and found quantitative convergence bounds on the variation distance from stationary distribution. Mira and Tierney [46] established conditions to guarantee the Markov chain constructed by these algorithm converges to the target distribution. They also provided a sufficient condition for uniform ergodicity.

3.3.2 The idea behind slice sampling

Assume we wish to sample from a distribution for a variable, X, whose unimodal density, π, is proportional to some function $f(x)$, i.e., $\pi(x) = cf(x)$, where $c = \frac{1}{\int f(x)dx}$. Moreover, assume that it is not possible to sample directly from f, although we are able to compute f up to a constant. The idea is to introduce the auxiliary variable Y such that the joint density function of X and Y, $\pi(x, y)$, is proportional to $f(x, y) = \mathbb{I}_{\{f(x)>y\}}$, i.e., $\pi(x, y) = cf(x, y)$. It can be seen that the conditional density of X given Y is uniform on the region $\{x : f(x) > y\}$, which is called the horizontal slice. By Bayes' rule, we have

$$f(x|y) \propto \mathbb{I}_{\{f(x)>y\}} \quad \text{and} \quad f(y|x) \propto \mathbb{I}_{\{f(x)>y\}}.$$

So the conditional density of Y given X is also uniform on the region $\{y : f(x) > y\}$ and this is called the vertical slice. These slices, construct an irreducible, aperiodic Markov chain $\{X_i, Y_i\}$ which by the ergodic Theorem will converge to its stationary

distribution, $\pi(x, y)$. By integrating out the auxiliary variable Y, we see that the marginal density of X is

$$\int_{-\infty}^{\infty} \pi(x, y)dy = c \int_{-\infty}^{\infty} \mathbb{I}_{\{f(x)>y\}}dy = c \int_{0}^{f(x)} \mathbb{I}dy = cf(x) = \pi(x).$$

as desired.

For the simple variable slice sampling method discussed here, replace the current value x_0 with a new value x_1, as follows where f is considered to be unimodal.

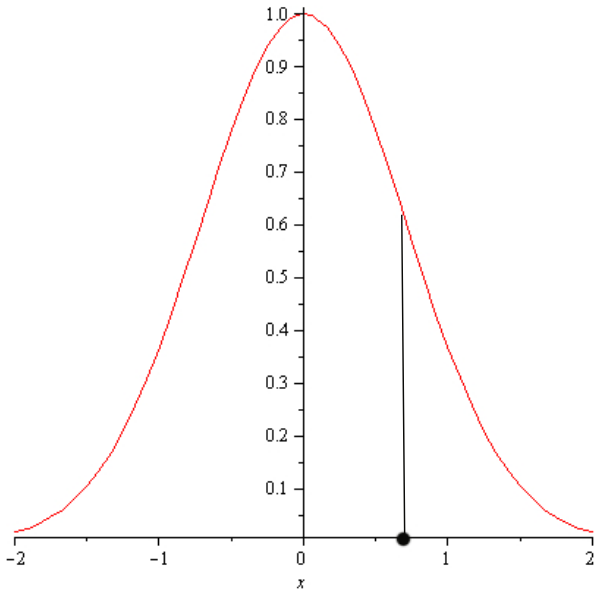


Figure 3.2: Step 1, horizontal slice.

- **Step 1**. Calculate $f(x_0)$, and then draw a real value y uniformly from $(0, f(x_0))$.
- **Step 2**. For a horizontal slice $S = \{x : y < f(x)\}$, find an interval $I = (L, R)$ around x_0 that contains all of the slice. It is not always easy to do that. Methods for finding I will be discussed later.
- **Step 3**. Randomly draw a new point x_1, uniformly within the interval $I \cap S$. For this purpose, we sample the new point uniformly from I until the point lies

in S. However, this method could be very inefficient if S is much smaller than I. If S is an interval, we may shrink I each time that a point drawn is not in S.

- **Step 4**. Repeat the algorithm as many times as needed.

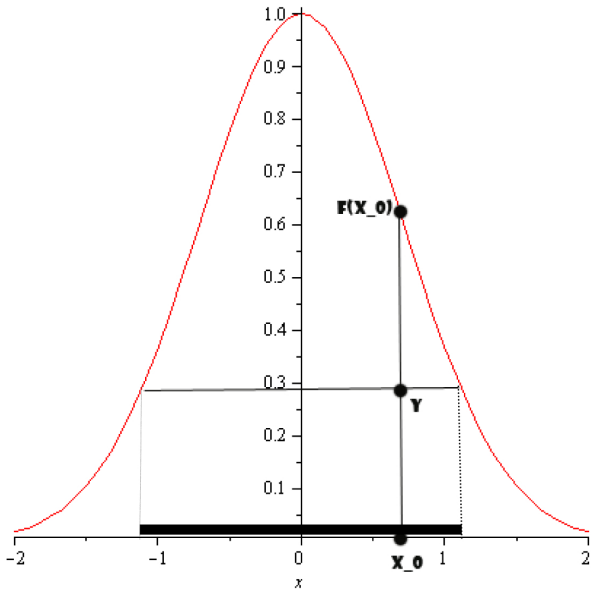


Figure 3.3: Step 2, finding the interval I.

Detailed balance property

A probability distribution $\pi(x)$ of a Markov chain with state space $\mathfrak{X}$ and transition kernel $K(x, x')$ for all $x, x' \in \mathfrak{X}$, is said to be stationary (invariant), if each step in the chain leaves $\pi(x)$ invariant, i.e.,

$$\pi(x) = \sum_{x'} K(x', x)\pi(x').$$

A sufficient but not necessary condition for the stationary distribution $\pi(x)$ to be invariant is *detailed balance property*:

$$K(x', x)\pi(x') = K(x, x')\pi(x).$$

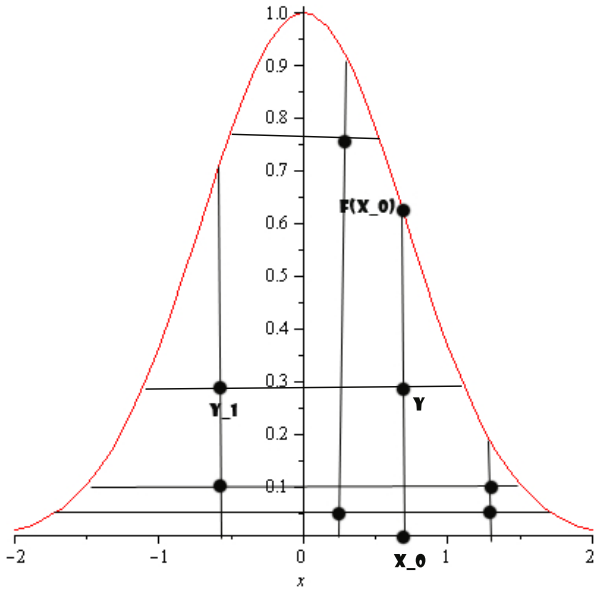


Figure 3.4: Step 3, vertical slice.

Lemma 3.3.1. *A transition kernel satisfying the detailed balance property will leave $\pi(x)$ invariant, i.e., π is the stationary distribution of the chain.*

Proof: The transition kernel satisfies the balance property, therefore

$$\begin{aligned}\sum_{x'} K(x',x)\pi(x') &= \sum_{x'} K(x,x')\pi(x) \\ &= \pi(x)\underbrace{\sum_{x'} K(x,x')}_{1} \\ &= \pi(x).\end{aligned}$$

Remark 10. *If the transition kernel is symmetric, then the detailed balance property holds. If π is the uniform distribution, then it is sufficient to show:*

$$K(x,x') = K(x',x) \quad \forall\ x,x' \in \mathfrak{X}.$$

This yields the basic idea of slice sampling.

Finding an appropriate interval

In step 2 of the algorithm, at each iteration we are supposed to find an interval $I = (L, R)$ from which we draw the next point.

Ideally, I is the smallest interval that contains S. So the best candidates are $L = \inf(S)$ and $R = \sup(S)$. However, this is not executable all the time. Even in the one-dimensional case, if f is a nonstandard complex density, we are not always able to find inf or sup of our horizontal slice.

If the range of x is bounded, then we can assume I is the range. Then we sample from region $I \cap S$ using the rejection method, i.e., we sample x from I and we reject if x is not S. This is not a good choice for I, since in many cases I is much smaller than the range of S and leads us to frequent rejections, thereby lowering the efficiency of the algorithm.

For this purpose, we can use one of two methods discussed by Neal [48]. In Neal's method, we do not necessarily find I such that $I \subseteq S$.

1. **Stepping out procedure:** Here, we need an estimate, w, for the typical size of slice around the current point x_0. In order to position w randomly around x_0, we pick u uniformly from $[0, 1]$ and let $L \leftarrow x_0 - w * u$ and $R \leftarrow L + w$. So we have an interval of length w positioned randomly around x_0. Now we expand our interval in steps of size w from left by letting $L \leftarrow L + w$ until $f(L) < y$ where y is the vertical level defining the slice. We expand the right end of the interval in the same way, i.e., $R \leftarrow R + W$ until $f(R) < y$. At this stage both end points are outside the slice [48].

2. **Doubling procedure:** This can extend the interval faster than stepping out, so is more efficient, particularly when the initial estimate of the interval, w, is very small. We begin by picking an initial interval of size w containing the current point x_0. Let y be the vertical level defining the slice. Pick u uniformly from $[0, 1]$ and let $L \leftarrow x_0 - w * u$ and $R \leftarrow L + w$. Then we have to check whether

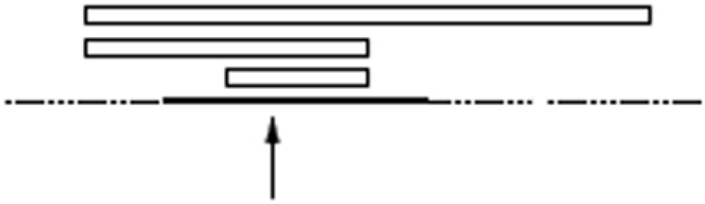


Figure 3.5: The initial interval is doubled twice, until the interval contains the whole slice.

I contains the whole slice. For this purpose, we need to check whether $y > f(L)$ and $y > f(R)$. If not, we double our interval. For the sake of correctness of the algorithm, we are not supposed to double the interval by expanding both sides equally. Instead only one side is doubled at random regardless of whether that side is already outside the slice. We repeat until both sides are outside the slice.

The correctness of the doubling procedure can be complicated in cases where the conditional distribution for the variable being updated is not known to be unimodal, or more generally when the slice contains more than one interval. For the correctness of the algorithm, we have to check the balance property, thereby, whether if the initial interval found by doubling procedure starting from x_0 includes x_1, then the same interval would be found starting from x_1. The example discussed in [48], where in case 1, starting from x_0 by doubling the initial interval twice, you may be able to be in state x_1. However, in case 2, where the starting point is different, no doubling is needed and so the final interval could not have been obtained from other points in other slices. For these cases Neal [48] found a procedure called an acceptance test. This procedure rejects a point when one of the intervals found by doubling from x_0 has both ends

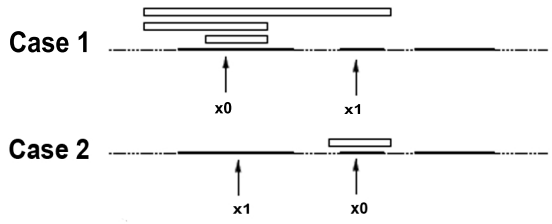


Figure 3.6: Neal's example about the correctness of doubling in the case where I may contain more than one interval.

outside the slice, but does not contain the current point. So for each scheme, Neal defines the set A as the acceptable successor states, via:

$$A = \{x \in S \cap I \text{ and } P(\text{select } I| \text{ At state } x) = P(\text{select } I| \text{ state } x_0)\}.$$

Remark 11. *In the examples discussed in this thesis, certain shortcuts can be used. In Examples 1,2 and 3, we can show that horizontal slices obtained in each iteration are convex. So we only need to double our initial interval as many time as needed, to make sure both sides are outside the interval; this is faster than the doubling procedure.*

Shrinkage procedure

The shrinkage technique was introduced by Neal [48] in order to improve the efficiency of the slice sampler. Assume $I = (L, R)$ is an interval which has been found with one of the methods discussed. The next step is to draw a new point uniformly from this interval, and reject if the point is not in $S \cap I$ until the new point lies in the acceptance set. The idea of this technique is to shrink the initial interval I each time we reject. Each time we reject a draw, x, we can replace the L or R with this point and shrink the our interval. By doing this the number of rejections decreases dramatically; therefore, it makes the algorithm much faster. If the acceptance set is a tiny portion of I, using this procedure is crucial.

Burn-in procedure

When we apply the slice sampler procedure, it is necessary to wait until the Markov chain converges to the invariant target distribution, and only then sample from the resulting distribution. The slice sampler generates dependent random variables and the first sequences do not replicate the target distribution, i.e., the later samples are closer to the stationary distribution. Thus we need to throw out some of the samples before collecting statistics. This is called the burn-in phase of the slice sampling. We need to discard samples obtained during the burn-in phase to guarantee the convergence of the Markov chain. If the number of samples in this phase is large enough, then the law of large numbers for dependent chains implies that one could just use a single sufficiently large Markov chain [53]. This implies it is not necessary to use different chains to ensure that convergence occurs. Discarding more terms in the beginning will lead us to a more accurate result, but will slow the procedure.

3.4 Gibbs-slice sampling

In this section, we discuss the Gibbs-slice sampling, which is based on Neal's single slice sampler, in order to be able to simulate random vectors or more generally large random matrices from a given unimodal density function. For example, assume our goal is to sample d-dimensional random vectors $(X^1, X^2, \cdots, X^d)$ from a density function, π, proportional to $f(x^1, \cdots, x^d)$. The idea is that in each iteration, we fix all entries except one, x^i, and then by applying one dimensional slice sampler, we sample x^i from $f(x^i|x^1, \cdots, x^{i-1}, x^{i+1}, \cdots, x^d)$.
From a starting point $(x_0^1, x_0^2, \cdots, x_0^d)$, the Gibbs-slice sampling procedure can be described as follows, where f assumed to be unimodal.

1. Calculate $t = f(x_0^1, x_0^2, \cdots, x_0^d)$. Draw y uniformly from the $[0, t]$.

2. Let w be the typical size of slice around the x_0^1. Pick u uniformly from $[0,1]$ and let $L \leftarrow x_0^1 - w * u$ and $R \leftarrow L + w$. Check whether or not $y > f(L)$ and $y > f(R)$. If not, double the interval $I = [L, R]$, until I contains the whole slice $S = \{x : y < f(x, x_0^2, \cdots, x_0^d)\}$.

3. Draw u uniformly from I. Reject u if $f(u, x_0^2, \cdots, x_0^d) < f(x_0^1, x_0^2, \cdots, x_0^d)$ and draw another u. Repeat until u is accepted. when u is accepted, let $u = x_1^1$.
   In order to accelerate the algorithm, in the case where we reject u, we can apply the shrinkage method to reduce the interval we are sampling from.

4. Fix x_1^1 and go back to Step 1 to sample x_1^2 from $f(x_1^2|x_1^1, x_0^3, \cdots, x_0^d)$. Repeat this for all d entries to get a new vector $(x_1^1, x_1^2, \cdots, x_1^d)$.

5. Repeat this algorithm as many time as needed to make sure the Markov chain constructed by this method converges to its steady state.

3.4.1 Convergence and correctness of the algorithm

For the slice sampler to be correct, we require that each iteration leave the desired distribution invariant. We therefore need to show that the selections in steps 2 and 3 will leave the joint distribution of X and Y invariant. Suppose that $f : \mathbb{R}^d \to [0, \infty)$ is a density function, i.e., a nonnegative measurable function which is not almost everywhere zero with respect to d-dimensional Lebesgue measure. This algorithm gives rise to a Markov chain, $\{X_t\}_{t=0}^m$, $X_t \in \mathbb{R}^d$, which has ν_f as a stationary distribution, where

$$\nu_f(A) = \frac{\int_A f(x)dx}{\int_{\mathbb{R}^d} f(x)dx} \qquad A \subseteq \mathbb{R}^d.$$

This is easy to see. Note that the Markov chain $\{X_t, Y_t\}_{t=0}^m$ is a Gibbs sampling on the distribution with density f with respect to Lebesgue measure on the set $\{(x, y)|f(x) >$

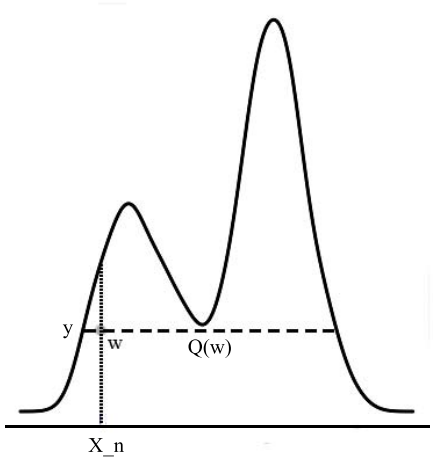


Figure 3.7: Slice sampler, nth iteration

$y\}$. To see that $\{X_t\}$ is also a Markov chain, note that for any $A \subseteq \mathbb{R}^d$, we have

$$P(X_{n+1} \in A | X_n = x_n, \cdots, X_0 = x_0) = \frac{P(X_{n+1} \in A, X_n = x_n, , \cdots, X_0 = x_0)}{P(X_n = x_n, , \cdots, X_0 = x_0)}$$

$$\begin{aligned}
&= \frac{\int_y P(X_{n+1} \in A, Y_n \in dy, \ X_n = x_n, \cdots, X_0 = x_0)}{P(X_n = x_n, \ , X_0 = x_0)} \\
&= \frac{\int_y P(X_{n+1} \in A, Y_n \in dy | \ X_n = x_n, \cdots, X_0 = x_0) P(X_n = x_n, \cdots, X_0 = x_0)}{P(X_n = x_n, \ \cdots, X_0 = x_0)} \\
&= \int_y P(X_{n+1} \in A, Y_n \in dy | \ X_n = x_n, \ \cdots, X_0 = x_0) \\
&= \int_{y=0}^{f(x)} \frac{dy}{f(x_n)} \int_{x_{n+1} \in A} \mathbb{I}\{f(x_{n+1} > y)\} dx_{n+1} \\
&= \int_y P(X_{n+1} \in A, Y_n \in dy | \ X_n = x_n) \\
&= P(X_{n+1} \in A | \ X_n = x_n).
\end{aligned}$$

The joint density for (x, y) is given by $\frac{\mathbb{I}_{0<y<f(x)}}{\int_{\mathbb{R}^d} f(x)dx}$. Therefore, the marginal density for x is equal to $\nu(A)$ for $A \subseteq \mathbb{R}^d$.

Let $L(y) = \{x \in \mathbb{R}^d; f(x) \geq y\}$ and $Q(y) = m\{L(y)\}$, where m is d-dimensional

Lebesgue measure. So the algorithm proceeds by alternately updating

$$Y_{n+1} \sim \text{Uniform}\{[0, f(X_n)]\},$$

and

$$X_{n+1} \sim \text{Uniform}\{L(Y_{n+1})\}.$$

We have,

$$P\{f(X_{n+1}) > z | f(X_n) = y\} = \frac{1}{y}\int_0^y \frac{Q(z)}{Q(w)} dw.$$

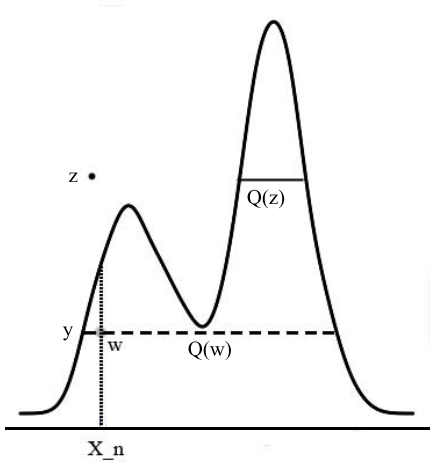


Figure 3.8: $Q(z)$ when $z > w$

Theorem 3.4.1. *A transition kernel in the method of Gibbs-slice sampler will leave* $\pi(x)$ *invariant under the full transition* $(X_0^1, X_0^2, \cdots, X_0^d, Y_0) \to (X_1^1, X_1^2, \cdots, X_1^d, Y_1)$*, where* Y *is the vertical coordinate of the slice sampler, i.e.,* π *is the stationary distribution of the chain.*

Proof: For simplicity, we show that π remains invariant under the transition $(X_0^1, X_0^2, Y_0) \to (X_1^1, X_1^2, Y_1)$,

This argument can similarly be extended to the higher dimensional cases.

Let $Q(X^1, X^2, Y)$ be any continuous function. First we show

$$E_\pi Q(X_0^1, X_0^2, Y_0) = E_\pi Q(X_1^1, X_0^2, Y_0).$$

That is, the measure will remain invariant under the transition $(X_0^1, X_0^2, Y_0) \to (X_1^1, X_0^2, Y_0)$, i.e., upgrading X^1 when X^2 is fixed. Given that we are at (x_0^1, x_0^2, y_0) we move to (x_1^1, x_0^2, y_0) as follows:

$$y_0 \sim \text{Uniform}[0, f(x_0^1, x_0^2)],$$

$$x_1^1 \sim \text{Uniform}[\{z|f(z, x_0^2) > y_0\}].$$

Let λ be the 3-dimensional Lebesgue measure. We have

$$\begin{aligned}
E_\pi Q(X_1^1, X_0^2, Y_0) &= c \int_{y_0} \int_{x_0^2} \int_{x_1^1} \int_{x_0^1} \mathbb{I}\{y_0 \le f(x_0^1, x_0^2)\} dx_0^1 Q(x_1^1, x_0^2, y_0) \\
&\qquad \frac{\mathbb{I}\{y_0 \le f(x_1^1, x_0^2)\}}{\lambda\{z|f(z, x_0^2) > y_0\}} dx_1^1 dx_0^2 dy_0 \\
&= c \int_{y_0} \int_{x_0^2} \int_{x_1^1} \underbrace{\left[\int_{x_0^1} \frac{\mathbb{I}\{y_0 \le f(x_0^1, x_0^2)\}}{\lambda\{z|f(z, x_0^2) > y_0\}} dx_0^1 \right]}_{1} Q(x_1^1, x_0^2, y_0) \\
&\qquad \mathbb{I}\{y_0 \le f(x_1^1, x_0^2) dx_1^1 dx_0^2 dy_0 \\
&= c \int_{y_0} \int_{x_0^2} \int_{x_1^1} Q(x_1^1, x_0^2, y_0) \mathbb{I}\{y_0 \le f(x_1^1, x_0^2)\} dx_1^1 dx_0^2 dy_0 \\
&= c \int_{y_0} \int_{x_0^2} \int_{x_0^1} Q(x_0^1, x_0^2, y_0) \mathbb{I}\{y_0 \le f(x_0^1, x_0^2)\} dx_0^1 x_0^2 dy_0 \\
&= E_\pi Q(X_0^1, X_0^2, Y_0).
\end{aligned}$$

The last equality follows from the fact that x_1^1 is a dummy variable and can be replaced by another dummy variable x_0^1.

Similarly, we have

$$E_\pi Q(X_0^1, X_0^2, Y_0) = E_\pi Q(X_0^1, X_1^2, Y_0).$$

Next, we have to show that

$$E_\pi Q(X_0^1, X_0^2, Y_0) = E_\pi Q(X_0^1, X_0^2, Y_1).$$

We have,

$$\begin{aligned}
E_\pi Q(X_0^1, X_0^2, Y_1) &= c\int_{x_0^1}\int_{x_1^2}\int_{y_1}\int_{y_0} \mathbb{I}\{y_0 \le f(x_0^1, x_0^2)\} dy_0 Q(x_0^1, x_0^2, y^1) \\
&\qquad \frac{\mathbb{I}\{y_1 \le f(x_0^1, x_0^2)\}}{f(x_0^1, x_0^2)} dx_0^1 dx_0^2 dy_1 \\
&= c\int_{x_0^1}\int_{x_1^2}\int_{y_1} \underbrace{\left[\int_{y_0} \frac{\mathbb{I}\{y_0 \le f(x_0^1, x_0^2)\}}{f(x_0^1, x_0^2)}\right]}_{1} dy_0 Q(x_0^1, x_0^2, y^1) \\
&\qquad \mathbb{I}\{y_1 \le f(x_0^1, x_0^2)\} dx_0^1 dx_0^2 dy_1 \\
&= c\int_{x_0^1}\int_{x_1^2}\int_{y_1} Q(x_0^1, x_0^2, y_1)\mathbb{I}\{y_1 \le f(x_0^1, x_0^2)\} dx_0^1 dx_0^2 dy_1 \\
&= c\int_{x_0^1}\int_{x_1^2}\int_{y_0} Q(x_0^1, x_0^2, y_0)\mathbb{I}\{y_0 \le f(x_0^1, x_0^2)\} dx_0^1 dx_0^2 dy_0 \\
&= E_\pi Q(X_0^1, X_0^2, Y_0).
\end{aligned}$$

The measure π is invariant under transitions $(X_0^1, X_0^2, Y_0) \to (X_1^1, X_0^2, Y_0)$, $(X_0^1, X_0^2, Y_0) \to (X_0^1, X_1^2, Y_0)$ and $(X_0^1, X_0^2, Y_0) \to (X_0^1, X_0^2, Y_1)$. So it remains invariant under $(X_0^1, X_0^2, Y_0) \to (X_1^1, X_1^2, Y_1)$.

Remark 12. *It is also known that a ϕ-irreducible aperiodic Markov chain with stationary distribution will converge to its stationary distribution from almost all starting points (e.g., see [33, 44, 52, 61]). Because of the nature of the algorithm, it is vital to replace "from almost all starting points" by "from all starting points". Harris recurrence will allow us to do so.*

Remark 13. *It is known that a Markov chain constructed from the Gibbs sampler is ϕ-irreducible and aperiodic, where ϕ is Lebesgue measure, e.g., see [9, 42, 56]. Also, by [61] the Markov chain induced by slice sampling is λ-irreducible and aperiodic.*

Theorem 3.4.2. *The Markov chain $\{X_t\}_{t=0}^m$ constructed by the Gibbs-slice sampling method is Harris recurrent.*

Before proving Theorem 3.4.2, we state the following lemma.

Lemma 3.4.1. *[59] Let $(i_1, i_2, \cdots, i_n)$ be the sequence of coordinate directions. Assume that each of the d directions appears at least once in the sequence $(i_1, i_2, \cdots, i_n)$. Let $M(x, .)$ be the kernel conditional on moving, i.e., on $X_n \neq X_{n+1}$ and $\lambda(.)$ be the d-dimensional Lebesgue measure. Then $M_{i_1} M_{i_2} \cdots M_{i_n}$ is absolutely continuous with respect to Lebesgue measure, i.e., if $A \in \mathfrak{F}$ with $\lambda(A) = 0$, then $(M_{i_1} M_{i_2} \cdots M_{i_n})(x, A) = 0$.*

Proof: (**Proof of Theorem 3.4.2**)

From Theorem 3.2.2, showing a λ-irreducible Markov chain is Harris recurrent is equivalent to showing, for all $x \in \mathfrak{X}$ and all $A \in \mathfrak{F}$ with $\nu_f(A) = 0$, that

$$P(X_n \in A \text{ for all } n | X_0 = x) = 0.$$

Define E_n as an event such that by time n at least one of the entries of the random vector has not been changed, i.e., all the chain has not yet moved in each direction. We have constructed the chain so that from any initial state x, with probability 1 the chain will eventually move at least once in each coordinate; therefore, for all $x \in \mathfrak{X}$ we have

$$\lim_{n \to \infty} P[E_n | X_0 = x] = 0.$$

Now, let $\nu_f(A) = 0$. We have,

$$\nu_f(A) \propto \int_A f(x) \lambda(dx), \quad A \in \mathfrak{F}.$$

For all $x \in \mathfrak{X}$, $f(x) > 0$, therefore, $\nu_f(A) = 0$ implies that $\lambda(A) = 0$. Note that if A has Lebesgue measure 0, by Lemma 3.4.1, we have

$$P[X_n \in A \cap E_n^c | X_0 = x] = 0.$$

Therefore,

$$\begin{aligned} P[X_n \in A | X_0 = x] &= P[X_n \in A \cap E_n | X_0 = x] + P[X_n \in A \cap E_n^c | X_0 = x] \\ &= P[X_n \in A \cap E_n | X_0 = x] \\ &< P[E_n | X_0 = x]. \end{aligned}$$

Hence,

$$P[X_n \in A,\ \forall\ n\ |X_0 = x] \leq \lim_{n\to\infty} P[X_n \in A|X_0 = x] \leq \lim_{n\to\infty} P[E_n|X_0 = x] = 0,$$

as desired.

Chapter 4

Examples of Gibbs-slice sampling

4.1 Introduction

In this chapter, we use the method of Gibbs-slice sampling to simulate large random matrices according to some given densities which do not follow a familiar parametric distribution which are inspired by quantum information theory.

Since quantum mechanics deal with noncommutative algebras, the random objects of study are presented by matrices. Via random interactions, generic loss of coherence of a fixed central system coupled to a quantum-chaotic environment is represented by a random matrix ensemble. Therefore, quantum information theory provides a wealth of random matrices problems.

This chapter is organized as follows. In Section 2 we use this method to simulate selfadjoint $N \times N$ random matrices, or a pair of random matrices, with the density proportional to $f(A) \sim \exp(-\mathrm{Tr}(P(A)))$ where is P is a given polynomial.

In Section 3, we expand the idea of Gibbs-slice sampler, introducing the advanced rotating Gibbs-slice sampling, in order to generate random unitary matrices with certain distributions.

In Section 4, we use the idea of Gibbs-slice sampler to generate hermitian positive

definite matrices with operator norm of max 1. We also expand this to the case where simulated matrices have a fixed unit trace, which are known as density matrices.
A density matrix, hermitian positive semidefinite matrix of trace one, describes the statistical state of a quantum system. This was first introduced by von Neumann in 1927 and is useful for describing and performing calculations with a mixed state, which is a statistical ensemble of several quantum states. Nechita [49], showed that there is a strong connection between these random density matrices and the Wishart ensemble of random matrix theory.

4.2 Simulating self-adjoint random matrices with the density proportional to $f(A) \sim \exp(-\mathbf{Tr}(P(A)))$

In this section, our goal is to use the idea of the Gibbs-slice sampler to generate random matrices with complex distributions. We start by generating self-adjoint random matrices with $A_{N\times N} = \{a_{ij}\}_{N\times N}$ with distribution proportional to $f(A) \sim \exp(-\mathrm{Tr}P(A))dA$, where P is a polynomial and Tr is an unnormalized trace.

Remark 14. *$x \mapsto x^4$ and $x \mapsto x^2$ are both convex functions, so by Jensen's trace inequality [28], $Tr(x^2)$ and $Tr(x^4)$ are also convex functions. So $f(x) \sim \exp(-Tr(x^4))$ and $f(x) \sim \exp(-Tr(x^4))$ are both concave, which implies that the horizontal slices are convex sets in the next 3 examples.*

Example 1

Assume we want to generate a symmetric matrix $A_{N\times N} = \{a_{ij}\}_{N\times N}$ with distribution proportional to $f(A) \sim \exp(-\mathrm{Tr}(c \cdot A^4))$. The proper scaling for $f(A)$ to be a density function is $c = \frac{N}{4}$, see [14, p.129] for instance. Therefore,

$$f(A) \sim \exp(-\frac{N}{4}\mathrm{Tr}(A^4)).$$

Suppose $A_0 = \{a_{ij}\}_{N\times N}$ is an initial matrix where for all i, j, $a_{ij} = 0.1$.
The first step is to calculate $f(A_0)$. In order to calculate the i, jth entry of the matrix, we fix all other entries and use the single variable slice sampling method. Since we are assuming that the matrix A is symmetric, we just need to simulate the entries on the upper triangle and diagonal of the matrix, $(\frac{N^2+N}{2})$ entries. We draw y uniformly from $[0, f(A_0)]$, which is our vertical slice. In each iteration, we need to find a proper interval I. In order to find the interval $I = (R, L)$ such that I contains $S = \{x : f(x) > y\}$.. Let $R = \frac{1}{\sqrt{N}} + A_0[i, j]$ and $L = A_0[i, j] - \frac{1}{\sqrt{N}}$ be our initial values. So we need both $f(R)$ and $f(L)$ be less than $f(A_0[i, j])$. We use doubling and

shrinking procedure as many times as needed to make sure the interval contains the slice S. Next we draw u uniformly from interval I.

Since the interval also contains points which are not in S, we have to make sure that u has been drawn from $S \cap I$. In order to do that, we have to calculate $f(u)$ to check $f(u) > f(A_0[i,j])$. If that is the case, we let both $A_{i,j}$ and $A_{j,i}$ be equal to u, in order to keep the matrix A symmetric. We increment i and j and repeat the procedure. Otherwise we reject u, shrink the interval using the shrinkage technique, and draw another value and repeat the algorithm of the same i and j. At the end, we will end up with a new symmetric matrix. Since we are interested in the empirical distribution we calculate and store the eigenvalues of the simulated matrix each time.

We go back to step one, and repeat the algorithm as many times as needed to make sure our Markov chain converges.

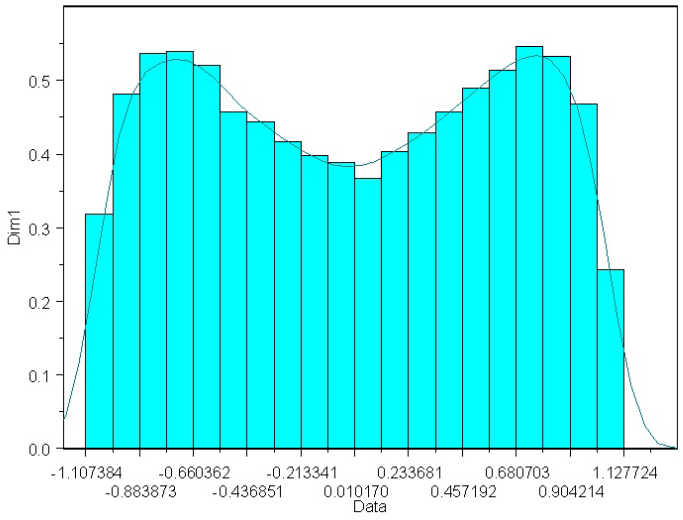


Figure 4.1: The histogram of the distribution of eigenvalues of simulated Matrix $A_{100\times100}$ using slice sampling with distribution proportional to $\exp\left(-\frac{100}{4}\mathrm{Tr}(A^4)\right)$.

Example 2

Let $A = \{a_{ij}\}_{N\times N}$ be a self-adjoint (or symmetric) random matrix, i.e., $a_{ij} = \overline{a_{ji}}$ for all $1 \leq i, j \leq N$. Moreover, for $1 \leq i \leq j \leq N$, let a_{ij} be i.i.d. random variables, such that

- $a_{ij} \sim \text{Normal}(0, \frac{1}{2N})$, when $i \neq j$,
- $a_{ij} \sim \text{Normal}(0, \frac{1}{N})$, when $i = j$.

$A = \{a_{ij}\}_{N\times N}$ is called a standard self adjoint Gaussian matrix. By Wigner's semicircle law [51], if A_N is a self-adjoint Gaussian $N \times N$-random matrix, then A_N converges in distribution, for $N \to \infty$, towards a semicircular element s,

$$\begin{aligned} A_N &\xrightarrow{d} s, \qquad \text{that is,} \\ \lim_{N\to\infty} \text{tr} \otimes E(A_N^k) &= \frac{1}{2\pi}\int_{-2}^{2} t^k\sqrt{4-t^2}dt \quad a.s. \quad \forall k \in \mathbb{N}. \end{aligned}$$

Let g be the joint density of the entries of $A_{N\times N}$. We have,

$$\begin{aligned} g &\sim \prod_{i<j} \exp\left(-\frac{a_{ij}^2}{2\frac{1}{2N}}\right) \times \prod_i \exp\left(-\frac{a_{ii}^2}{2\frac{1}{N}}\right) \\ &\sim \exp\left(-N\sum_{i<j} a_{ij}^2 - N\frac{\sum_i a_{ii}^2}{2}\right) \\ &= \exp\left(-\frac{N}{2}\left(2\sum_{i<j} a_{ij}^2 + \sum_i a_{ii}^2\right)\right) \\ &= \exp\left(-\frac{N}{2}\text{Tr }(A^2)\right). \end{aligned}$$

Since

$$\begin{aligned} \text{Tr } A^2 &= \sum_i \sum_j a_{ij}a_{ji} \\ &= \sum_i a_{ii}^2 + 2\sum_{i<j} a_{ij}^2. \end{aligned}$$

In this example, our aim is to simulate a pair of symmetric $N \times N$ random matrices $A,\ B$, where

$$(A, B) \sim \exp\left(-\frac{N}{2}\mathrm{Tr}(A^2 + B^2 + \gamma AB)\right),$$

where $|\gamma| \leq 2$. We use the same argument as in Example 1. The algorithm is very similar to that discussed in example 1.

Now consider the asymptotic behaviour of the empirical eigenvalue distribution of A and B. Let $(\lambda_A)_i$ and $(\lambda_B)_i$ be the eigenvalues of A and B respectively. Define

$$F_{A_{N\times N}}(x) = \frac{1}{N}\#\{(\lambda_A)_i : (\lambda_A)_i \leq x\}, \quad F_{B_{N\times N}}(x) = \frac{1}{N}\#\{(\lambda_B)_i : (\lambda_B)_i \leq x\},$$

where both are random distribution functions.

Lemma 4.2.1. *If (A, B) is a pair of symmetric $N \times N$ random matrices such that*

$$(A, B) \sim f(A, B) = \exp\left(-\frac{N}{2}\,Tr(A^2 + B^2 + \gamma AB)\right),$$

where $|\gamma| < 2$. Then both $F_{A_{N\times N}}$ and $F_{B_{N\times N}}$ converge in distribution to $\frac{s\sqrt{2}}{\sqrt{2-\gamma}}$, where s is semicircular element.

Proof: There exist symmetric matrices $X_{N\times N}$ and $Y_{N\times N}$ such that A and B can be written as a linear combination of them, i.e., for some real a, b, c, d,

$$A = aX + bY \quad \text{and} \quad B = cX + dY.$$

This is easy to see; consider the system of equation with $eh \neq fg$,

$$\begin{pmatrix} e & f \\ g & h \end{pmatrix}\begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} X \\ Y \end{pmatrix},$$

where X and Y are symmetric matrices. Define a, b, c, d via

$$\begin{pmatrix} e & f \\ g & h \end{pmatrix}^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Therefore,

$$\begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix},$$

where $ad - bc \neq 0$. So we substitute,

$$\begin{aligned} \mathrm{Tr}(A^2 + B^2 + \gamma AB) &= \mathrm{Tr}((aX + bY)^2 + (cX + dY)^2 + \gamma(aX + bY)(cX + dY)) \\ &= \mathrm{Tr}((a^2 + c^2 + \gamma ac)X^2 + (b^2 + d^2 + \gamma bd)Y^2 \\ &\quad +(2ab + 2cd + \gamma ad + \gamma bc)XY) \end{aligned}$$

By choosing $a = -c$ and $b = d$, we have $2ab + 2cd + \gamma ad + \gamma bc = 0$. Therefore,

$$\mathrm{Tr}(A^2 + B^2 + \gamma AB) = \mathrm{Tr}\left((2 - \gamma)a^2X^2 + (2 + \gamma)b^2Y^2\right).$$

Let $b = \sqrt{\frac{2-\gamma}{2+\gamma}}a$, where $|\gamma| < 2$. We have,

$$\exp\left(-\frac{N}{2}\mathrm{Tr}(A^2 + B^2 + \gamma AB)\right) = \exp\left(-\frac{N}{2}\mathrm{Tr}(a^2(2 - \gamma)(X^2 + Y^2))\right).$$

By Wigner's Theorem [51], the marginal density of $(a\sqrt{2-\gamma})X$ and $(a\sqrt{2-\gamma})Y$ converges in distribution to Wigner's semicircle law. So we have,

$$F_{A_{N\times N}} \xrightarrow{d} \frac{s}{a\sqrt{2-\gamma}}, \quad \text{and} \quad F_{B_{N\times N}} \xrightarrow{d} \frac{s}{a\sqrt{2-\gamma}},$$

where s is a semicircular element and a is a nonzero real number. When $\gamma = 0$ we know the scalar should be 1, so we get $a = \frac{1}{\sqrt{2}}$. Note that we do not need to worry about the Jacobian determinant of the transformation, since we have

$$f(A, B)dAdB = f(X, Y)|J_{X,Y}|dXdY \sim f(X, Y)dXdY,$$

where $|J_{X,Y}| \neq 0$.

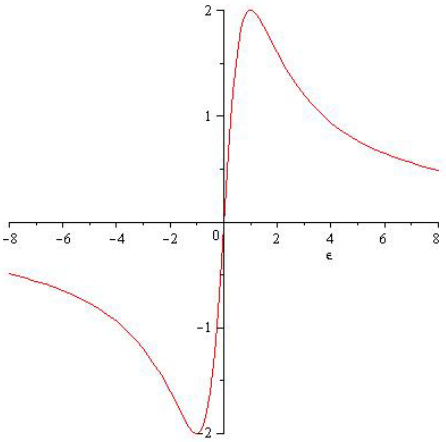


Figure 4.2: The histogram of $f(\epsilon) = \frac{4\epsilon}{1+\epsilon^2}$. The range of the above function is between -2 and 2.

Remark 15. *In order to have* $Tr(A^2 + B^2 + \gamma AB) > 0$, $|\gamma| < 2$. *Since we know, for all* ϵ,

$$\begin{aligned} Tr\left((\epsilon A + B) + (A + \epsilon B)^2\right) &> 0, \qquad \textit{that is} \\ Tr\left((1+\epsilon^2)(A^2 + B^2) + 4\epsilon AB\right) &= (1+\epsilon^2)\, Tr\left(A^2 + B^2 + \frac{4\epsilon}{1+\epsilon^2} AB\right). \end{aligned}$$

We know that $|\frac{4\epsilon}{1+\epsilon^2}| \leq 2$.
Moreover, $\exp\left(-\frac{N}{2} Tr(A^2 + B^2 + \gamma AB)\right)$ *is a density function only if we assume* $|\gamma| \leq 2$. *If we take* $\gamma > 2$, *then*

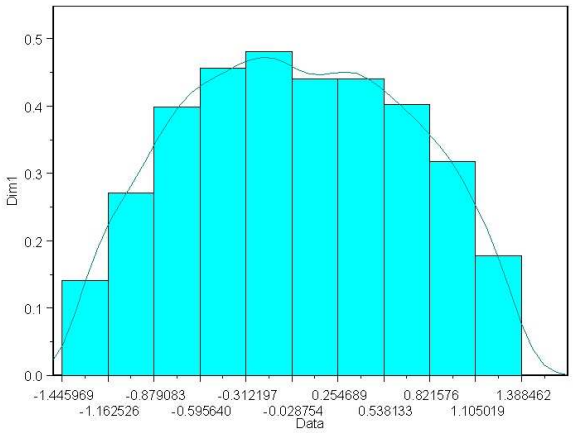
$$\begin{aligned} \exp\left(-\frac{N}{2} Tr(A^2 + B^2 + \gamma AB)\right) &= \exp\left(-\frac{N}{2} Tr(A^2 + B^2 + 2AB + (\gamma - 2)AB)\right) \\ &= \exp\left(-\frac{N}{2} Tr\left((A+B)^2 + (\gamma - 2)AB\right)\right). \end{aligned}$$

Let $a_+ = \frac{A+B}{\sqrt{2}}$ *and* $a_- = \frac{A-B}{\sqrt{2}}$. *The Jacobian determinant of this linear transformation*

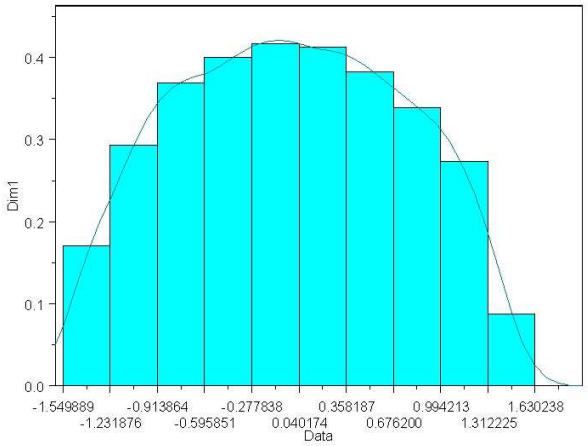
is 1. We have,

$$\begin{aligned} Tr\left((A+B)^2+(\gamma-2)AB\right) &= Tr\left(2a_+^2+(\gamma-2)\left(\frac{1}{2}(a_+^2-a_-^2)\right)\right) \\ &= Tr\left((2+\frac{1}{2}(\gamma-2))a_+^2-\frac{1}{2}(\gamma-2)a_-^2\right) \\ &= Tr\left((1+\frac{1}{2}\gamma)a_+^2-\frac{1}{2}(\gamma-2)a_-^2\right). \end{aligned}$$

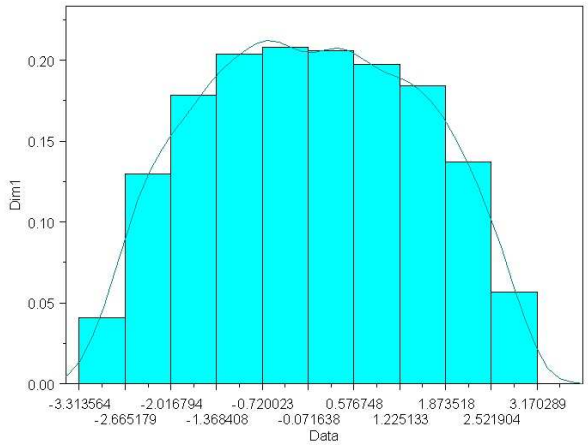
This cannot be a density function, since $\frac{1}{2}(\gamma-2)>0$.



Histogram of eigenvalues of A, where N=40 and gamma=0.5



Histogram of eigenvalues of A, where N=40 and gamma=1



Histogram of eigenvalues of A, where N=40 and gamma=1.5

Example 3

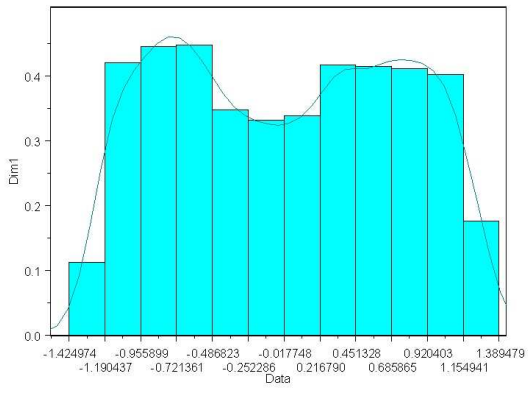
In this example, with the same algorithm as Example 2, we want to simulate pairs of symmetric $N \times N$ random matrices $A,\ B$, with

$$(A, B) \sim \exp\left(-\frac{N}{4}\mathrm{Tr}(A^4 + B^4 + \gamma AB)\right),$$

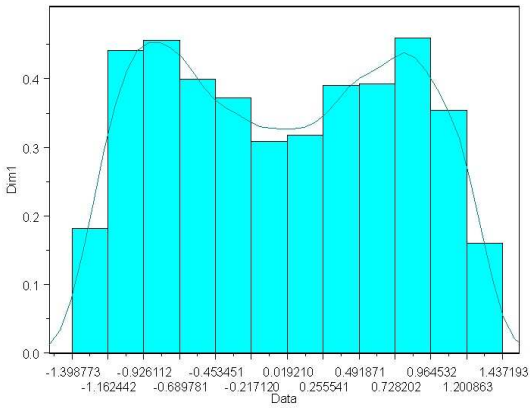
where $\gamma > 0$.

Conjecture for the marginal densities of A and B

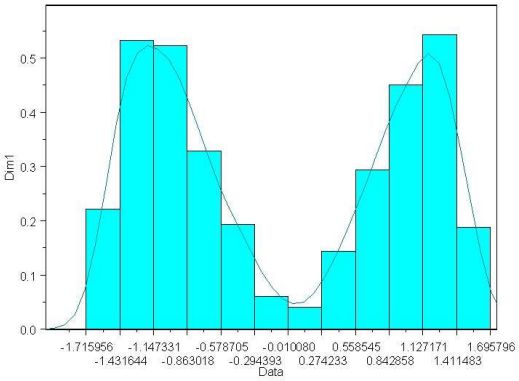
By looking at the histograms of empirical distribution of A and B, we observe that for small γs limiting distribution of A and B has connected spectrum. However, for the larger value of γ, almost $\gamma > 6$, there will be spectral gap centered at zero, i.e., the threshold has 2 connected component.



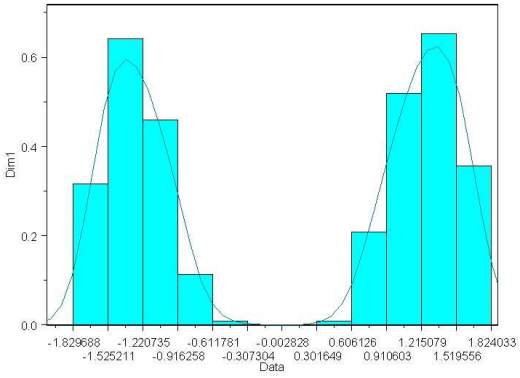
Gamma=0.1



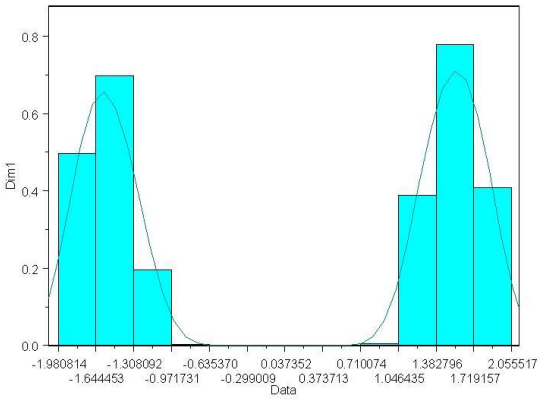
Gamma=1



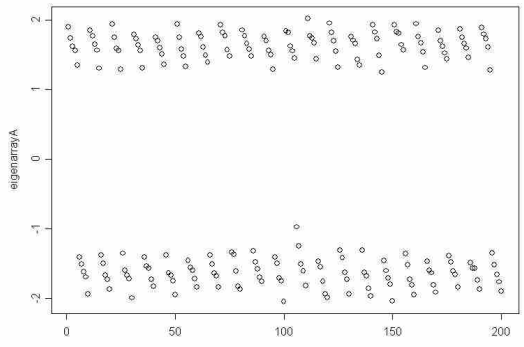
Gamma=5



Gamma=7

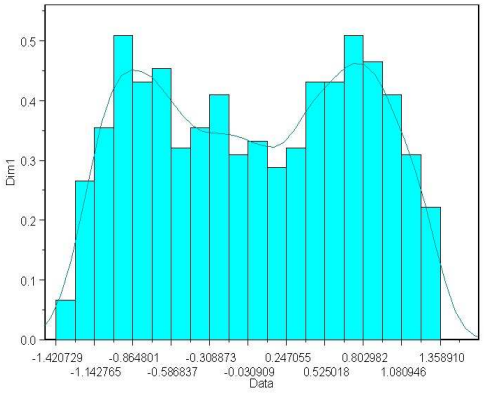


Gamma=10

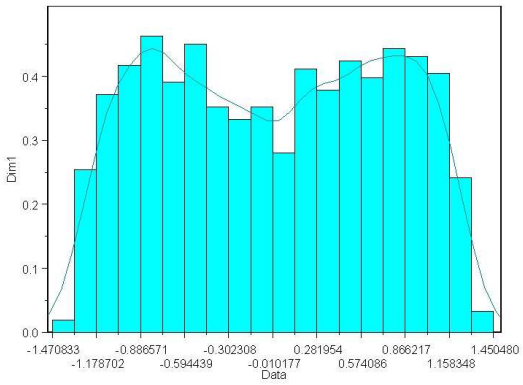


Plot of eigenvalues for Gamma=10

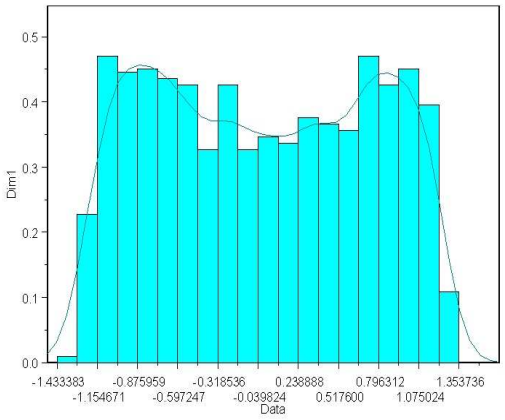
For a fixed gamma=0.1, below are four diagrams of different values of N. The histograms are apparently independent of the value of N; this shows that the density function should converge.



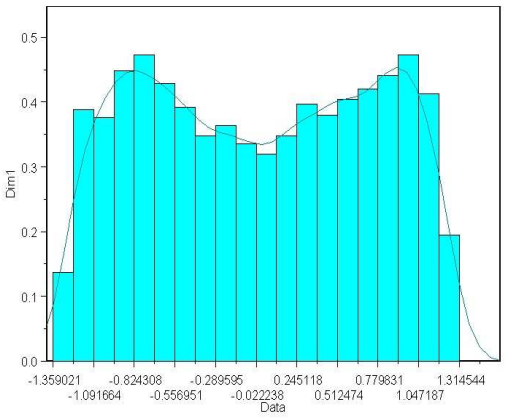
N=10



N=15

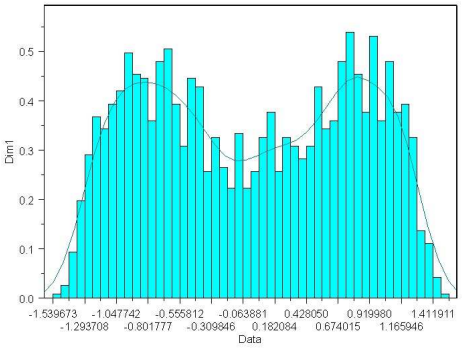


N=20

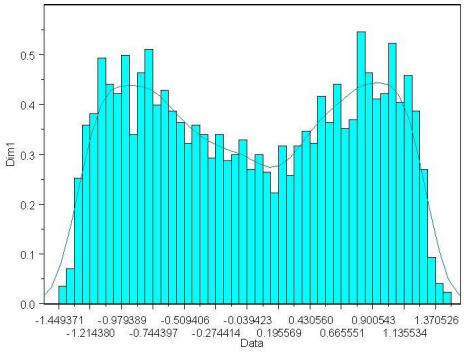


N=25

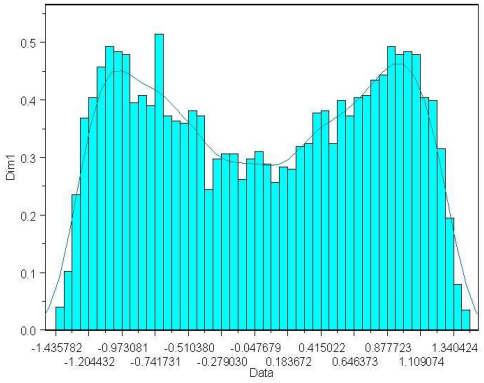
Gamma=2



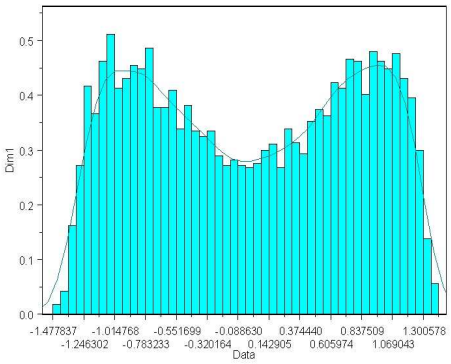
N=10



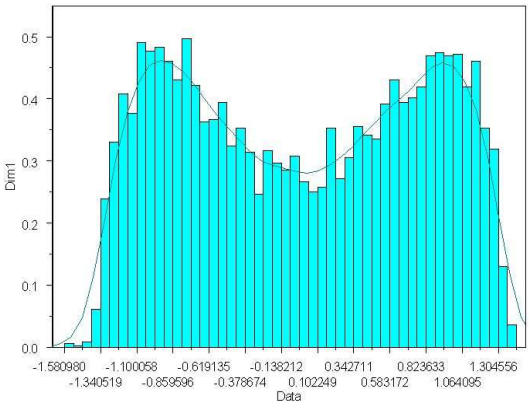
N=15



N=20



N=25



N=30

4.3 Generating random unitary matrices with certain distribution using advanced rotating slice sampling

In this section we are trying to generate random unitary matrices from a given distribution via a new method of slice sampling.

An $N \times N$ unitary matrix $U = (u_{jk})$ satisfies the relation $U^*U = UU^* = I$, i.e.,

$$\sum_{k=1}^{N} u_{jk}^* u_{kl} = \sum_{k=1}^{N} \overline{u}_{kj} u_{kl} = \delta_{jl} \quad \text{and} \quad \sum_{k=1}^{N} u_{jk} u_{kl}^* = \sum_{k=1}^{N} \overline{u}_{jk} u_{lk} = \delta_{jl},$$

where U^* is the conjugate transpose of U, i.e., $u_{jk}^* = \overline{u}_{kj}$. This implies that the set of columns and the set of rows of a unitary matrix are both an orthonormal base of $\mathbb{C}^N$. Let $U(N)$ be the set of unitary matrices of size N. With respect to the unique invariant measure under group multiplication, Haar measure, $U(N)$ is a probability space. The matrix equations imply that the matrix elements are statistically correlated. The Generation of unitary random matrices has applications in quantum mechanics and physics [45].

In this section, we introduce an algorithm to generate unitary random matrices from a given density function $f(x)$ with respect to Haar measure. The algorithm is essentially based on the invariant properties of Haar measure. So we start with giving the definition of Haar measure.

4.3.1 Haar measure and invariance

Definition 15. *A Borel measure μ on a topological group is said to be left invariant if $\mu(gE) = \mu(E)$ for each $g \in G$ and each Borel set E. Similarly, a Borel measure ν is said to be right invariant if $\nu(Eg) = \nu E$. A left and right invariant measure on a compact group is called Haar measure. That is to say, for a group G a nonzero*

measure $\mu : G \to [0, \infty)$ *is a Haar measure, if for all* $S \subseteq G$ *and* $g \in G$*:*

$$\mu(gS) = \mu(Sg) = \mu(S),$$

where

$$\mu(S) := \int_{g \in S} d\mu(g).$$

It is known that Haar measure exists on every compact topological group (in particular, on unitary and orthogonal groups) and is unique up to a constant. Haar measure normalized to one is a natural choice for a probability measure on a compact group because, being invariant under group multiplication, any region $U(N)$ carries the same weight in a group average.

Generating a matrix in the Ginibre ensemble is the first step toward generating random unitary matrices.

Ginibre ensemble

The space of matrices for the Ginibre ensemble is $GL(N, \mathbb{C})$, the set of all the invertible $N \times N$ complex matrices $Z = (z_{jk})$, where z_{jk}s are i.i.d. standard normal complex random variables [45]. Therefore the probability density function of z_{jk} is

$$f(z_{jk}) = \frac{1}{\pi} e^{-|z_{jk}|^2}.$$

Since the matrix entries are independent, the joint probability density function can be given as follows,

$$f(Z) = \frac{1}{\pi^{N^2}} \prod_{j,k=1}^{N} e^{-|z_{jk}|^2} = \frac{1}{\pi^{N^2}} \exp\left(- \sum_{j,k=1}^{N} |z_{jk}|^2 \right) = \frac{1}{\pi^{N^2}} \exp(-\text{Tr } Z^* Z).$$

For any $U \in U(N)$ we have,

$$f(UZ) = \frac{1}{\pi^{N^2}} \exp(-TrZ^*U^*UZ) = \frac{1}{\pi^{N^2}} \exp(-\text{Tr } Z^*Z) = f(Z).$$

The QR decomposition

One of the most common ways to simulate a Haar (uniform) distribution on unitary matrices is to apply the QR decomposition. Let A be a real $n \times m$ matrix. It is known that A can decomposed as a product of an $n \times m$ orthogonal matrix and an $m \times m$ upper triangular matrix. The easiest way to compute this is to use the Gram-Schmidt orthonormalization process. For an arbitrary complex $N \times N$ matrix Z of full rank, applying the Gram-Schmidt orthonormalization to its columns yields a unitary matrix Q. It was shown in [17] that if Z belongs to the Ginibre ensemble, i.e., entries of Z are i.i.d. standard complex normal variables, then Q is distributed with respect to Haar measure. Any matrix $Z \in GL(N, \mathbb{C})$ can be decomposed as

$$Z = QR,$$

where $Q \in U(N)$ and R is an upper-triangle and invertible, i.e., $ZR^{-1} = Q$.
To simulate a unitary matrix it is tempting to first produce a matrix in the Ginibre ensemble and then apply the Gram-Schmidt algorithm in order to obtain a QR decomposition.

Why the QR decomposition does not work

The QR decomposition exists for every $Z \in GL(N, \mathbb{C})$; however, it is not unique. This is easy to see, for any diagonal matrix $D \in U(N)$, we have,

$$QR = (QD)(D^*R) = Q'R',$$

where $Q' = QD$ is unitary and $R' = D^*R$ is upper-triangular. Therefore $Z = Q'R'$ is also a valid QR decomposition of Z. The QR decomposition does not guarantee nonnegative diagonal entries in R. So Q is not distributed with a Haar measure [45]. In order to generate matrices distributed with a Haar measure, we need this decomposition to be unique.

Uniqueness can be achieved by demanding that R must have positive diagonal entries. We can modify the Gram-Schmidt algorithm to create a random unitary matrix with distribution given by Haar measure by defining $R' = D^{-1}R$ and $Q' = QD$ where

$$D = \begin{pmatrix} \frac{r_{11}}{|r_{11}|} & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \frac{r_{NN}}{|r_{NN}|} \end{pmatrix},$$

and r_{ii}s are the diagonal elements of R. The diagonal elements of R' are always real and strictly positive. Therefore, $R' \in T(N)$, where $T(N)$ is a set of all invertible $N \times N$ upper-triangle complex matrices with positive diagonal matrices. The identity matrix, I, is the only matrix which belongs to $T(N) \bigcap U(N)$. Therefore the matrix Q' is distributed with respect to Haar measure [17, p 234].

4.3.2 Method of rotating slice sampling for generating random unitary matrices from a given distribution

In this section we are trying to generate random matrices from a given distribution using a new method of slice sampling with respect to Haar measure.

An example is to sample random unitary matrices for which the density is $f \sim \exp(-N\mathrm{Tr}(A + A^*))$, with respect to Haar measure, where A is a unitary matrix and A^* is the conjugate transpose of A. The algorithm is as follows.

We start the algorithm by generating an $N \times N$ matrix $Z \in GL(N, \mathbb{C})$. Then we apply the QR decomposition and as was explained before, we calculate Q'. Therefore, $A_0 = Q'$ will be our initial matrix. In unitary matrices all the entries are mutually dependent. So, unlike previous QR method we simulate the whole matrix instead of sampling each entry separately.

Next, we calculate $f(A_0)$ and sample y uniformly from $[0, f(A_0)]$. So now we have a horizontal slice. Then we draw θ uniformly from $[-\pi, \pi]$ and construct matrix R_1 as

follows:

$$R_1 = \begin{pmatrix} \cos(\theta) & \sin(\theta) & 0 & \cdots & 0 \\ -\sin(\theta) & \cos(\theta) & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & \cdots & 0 & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix},$$

and we calculate $M = A_0 R_1$. Next we have to check whether $f(M) > y$. If not, we draw another θ and repeat until the condition is satisfied. Then we replace A with our new matrix M and again draw $\theta \sim \text{uniform}[-\pi, \pi]$ and this time we calculate,

$$R_2 = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & \cos(\theta) & \sin(\theta) & \cdots & 0 \\ 0 & -\sin(\theta) & \cos(\theta) & \cdots & 0 \\ 0 & \cdots & 0 & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}.$$

Again, we calculate $M = A_0 R_2$ and as before to repeat until the condition $f(M) > y$ is satisfied.

We repeat this procedure $N - 1$ times, i.e., for all $1 \leq i \leq N - 1$, at each step, we draw randomly $\theta \sim \text{uniform}[-\pi, \pi]$ and calculate

$$R_i = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & 0 & \cdots & \cdots & \vdots \\ \vdots & 0 & \cos(\theta) & \sin(\theta) & \cdots & \vdots \\ \vdots & \vdots & -\sin(\theta) & \cos(\theta) & \cdots & \vdots \\ \vdots & \vdots & 0 & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & 0 & 1 \end{pmatrix}.$$

We calculate $M = A_0 R_i$, and check if the condition $f(M) > y$ is satisfied. Otherwise draw another θ and repeat until the condition is satisfied. So A_1 will be achieved by rotating A_0 in $N - 1$ random directions. We calculate and store eigenvalues of A_1

and start over with A_1 is the initial matrix. We repeat the algorithm as many times as needed to make sure the Markov chain constructed by this algorithm, $\{A_i\}$, will converge to its stationary distribution.

Remark 16. *Other methods for decomposing matrices include the singular value decomposition (SVD) and the polar decomposition. Let $Z \in GL(N, \mathbb{C})$, where Z is almost surely invertible. Then $Q = Z(Z^*Z)^{-\frac{1}{2}}$ is orthogonal. Since,*

$$\begin{aligned} QQ^* &= Z(Z^*Z)^{-\frac{1}{2}}(Z(Z^*Z)^{-\frac{1}{2}})^* \\ &= Z(Z^*Z)^{-\frac{1}{2}}((Z^*Z)^{-\frac{1}{2}})^*Z^* \\ &= Z(Z^*Z)^{-1}Z^* \\ &= ZZ^{-1}(Z^*)^{-1}Z^* \\ &= I, \end{aligned}$$

which implies that Q is unitary. The polar decomposition of a complex matrix A is $A = UP$, where U is unitary and P is positive definite. The matrix P can be uniquely achieved by $P = \sqrt{AA^*}$ and if A is invertible, we get $U = AP^{-1}$.

Remark 17. *We are constructing an irreducible recurrent Markov chain. Also in each step R_i is an unitary matrix; therefore, AR_i remains unitary.*

Theorem 4.3.1 (Convergence of the algorithm)**.** *The transition kernel acts on the slice in the algorithm, preserves the uniform measure on the slice.*

Proof: In order to this, we have to show the detailed balance property is satisfied in each step. Assume A^n and A^{n+1} are matrices simulated in steps nth and $(n+1)$th. Since we are choosing uniformly from the slice, it is sufficient to show

$$P(A^{n+1} = b|\ A^n = a) = P(\ A^{n+1} = a|\ A^n = b).$$

Algorithm 1 Algorithm for simulating unitary matrix from $f \sim \exp(-N\mathrm{Tr}(A+A^*))$

Take $Z \in GL(N, \mathbb{C})$

Apply QR, SVD or polar decomposition to get an initial unitary matrix A

Calculate $f(A)$

while $k < m$ **do**

  $y \sim \mathrm{Uniform}[0, f(A)]$

  for all $1 \leq i \leq n-1$ **do**

    $\theta \sim \mathrm{Uniform}[-\pi, \pi]$

    $M \leftarrow A \times R_i$

    if $f(M) > y$ **then**

      Accept M

    else

      Reject, draw another θ and calculate R_i again

    end if

  end for

  $k \leftarrow k + 1$

end while

According to the algorithm , $A^{n+1} \leftarrow A^n R_i(\theta)$, where θ is drawn uniformly from $[-\pi, \pi]$. We have,

$$\begin{aligned} P(A^{n+1} = b|\ A^n = a) &= P(aR_i(\theta) = b) \\ &= P(a = bR_i(-\theta)) \\ &= P(A^{n+1} = a|\ A^n = b). \end{aligned}$$

Note that for each i, R_i is an invertible matrix and $R_i^{-1}(\theta) = R_i(-\theta)$.

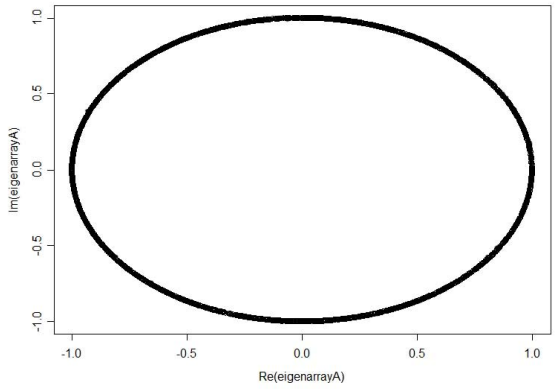


Figure 4.3: The histogram of the distribution of complex and real eigenvalues of simulated Matrix $A_{100\times100}$ with distribution proportional to $f \sim \exp(-N\mathrm{Tr}(A + A^*))$. The eigenvalues lie on the unit circle.

Eigenvalues 1 and -1

If $A_{N\times N} \in O(N)$, then either $\det(A) = \pm 1$. So a random simulated matrix $A_{N\times N}$ in the previous algorithm will have $\det(A) = -1$ with probability $\frac{1}{2}$. Assume N is even

and $\det(A) = -1$, we have:

$$\begin{aligned}
\det(A - I) &= \det(-(I - A)) \\
&= (-1)^N \det(I - A) \\
&= \det(A^{-1}) \det(I - A) \\
&= \det(A^{-1}) \det(A - I)(-1)^N \\
&= \det(A^{-1}) \det(A - I) \\
&= -\det(A - I).
\end{aligned}$$

Which implies that $\det(A - I) = 0$, i.e., 1 is an eigenvalue of A. With exactly the same argument we can show that under these assumptions -1 will be an eigenvalue of A. This shows that in the case where N is even, with probability $\frac{1}{2}$ the simulated matrix will have at least two real eigenvalues. In the case where N is odd, A has at least one real eigenvalue which should be 1 or -1. So we have the following corollary.

Corollary 4.3.1. *Let A be random matrix simulated from $O(N)$, where $O(N)$ is the set of all $N \times N$ orthogonal matrices. Then the probability of A having a real eigenvalue, 1 or -1, is at least 0.75.*

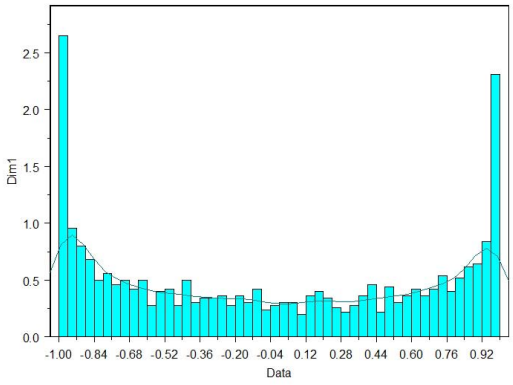


Figure 4.4: The histogram of the distribution of the real part of eigenvalues of a simulated matrix $A_{100\times100}$ with distribution proportional to $f \sim \exp(-N\mathrm{Tr}(A + A^*))$. There are point masses at $x = 1$ and $x = -1$.

4.4 Generating hermitian positive definite matrices with the operator norm of max 1

Definition 16. *Let V and W be finite dimensional normed vector spaces. The operator norm of a linear operator $S : V \to V'$ is defined via*

$$||S|| = \sup_{||v||=1} ||S(v)||.$$

For any two operators T and S we have,

$$||TS|| \leq ||T||||S||.$$

If $A \in \mathbb{M}(\mathbb{C}^n)$, define $T_A : \mathbb{C}^n \to \mathbb{C}^n$, $T_A(v) = Av$. We have

$$\begin{aligned} ||A|| &:= ||T_A|| \\ &= \sqrt{\text{largest eigenvalue of } A^*A}, \end{aligned}$$

where A^* is the transpose of A. If S_N^+ denotes the set of all positive definite matrices, then S_N^+ is a proper cone (i.e., closed and convex).
In this section our goal is to generate random matrices uniformly from

$$\mathbb{P}^+ = \{A \in \mathbb{M}_n(\mathbb{C}) |\ A = A^*,\ A > 0, ||A|| \leq 1\},$$

where A^* is the conjugate transpose of A, $A \geq 0$ means A is positive definite, and by $||A||$, we mean the operator norm of A.

Remark 18. *Let $A_{n\times n}$ be a hermitian matrix. Then the following statements are equivalent:*

- *The matrix A is positive definite.*
- *All eigenvalues of A are strictly positive.*

Lemma 4.4.1. *If $A \in \mathbb{P}^+$, then all of the eigenvalues of A lie between zero and one. This is equivalent to saying that, the matrices A and $I - A$ both have no negative eigenvalues.*

Proof: Let $A \in \mathbb{P}^+$ and let $\lambda_{max} > 0$ be the largest eigenvalue of A. Then λ_{max}^2 is the largest eigenvalue of A^2, so $\sqrt{\lambda_{max}^2} \leq 1$. So, we have that for all $1 \leq i \leq n,\ 0 < \lambda_i < 1$, where the λ_i are the eigenvalues of A.
If λ is an eigenvalue of $I - A$, then we have:

$$|(I - A) - \lambda I| = |A - (1 - \lambda)I| = 0.$$

So we have $0 < 1 - \lambda < 1$, i.e., $0 < \lambda < 1$.

4.4.1 Method of simulation matrices from $\mathbb{P}^+$

The idea is to start the algorithm by taking any matrix from the set $\mathbb{P}^+$. We may take as an initial matrix, $A^{(0)}$, any diagonal matrix whose diagonal entries are between zero and one. Select

$$A^{(0)}_{N\times N} = \begin{pmatrix} \frac{1}{2} & 0 & \dots & \dots & 0 \\ 0 & \frac{1}{2} & 0 & \dots & 0 \\ \vdots & 0 & \ddots & 0 & 0 \\ \vdots & 0 & 0 & \ddots & 0 \\ 0 & \dots & \dots & 0 & \frac{1}{2} \end{pmatrix}_{N\times N} .$$

Next, we simulate the entries on the diagonal. Note that since the matrix is hermitian, the diagonal entries are real. For simulating the first entry a_{11}, we fix all other entries and then pick uniformly u from $[-\epsilon, +\epsilon]$, where ϵ is a randomly drawn from $(0, 1)$. We replace a_{11} with $a_{11} + u$. Now we need to check whether this replacement should be accepted. For this, we need to calculate all the eigenvalues of the new matrix $A^{(1)}$, where a_{11} has been replaced by $a_{11} + u$. If all the eigenvalues of the matrix $A^{(1)}$ and

$I - A^{(1)}$ are positive, the replacement is accepted. Otherwise, we have to reject it and pick another u; then we repeat the process for the first entry as necessary as that the entry will be accepted. We iterate the process for all n diagonal entries.
Next is to simulate the complex off-diagonal entries. The matrix must be hermitian, so we just need to simulate the entries on the upper triangle of the matrix, i.e., $a_{ij},\ i > j$, and replace the corresponding a_{ji} with the conjugate of that. In order to simulate the a_{ij}, draw u_1 and u_2 uniformly from $[-\epsilon, +\epsilon]$. Replace a_{ij} with $a_{ij} + (u_1 + iu_2)$ and a_{ji} with $a_{ji} + (u_1 - iu_2)$. We have to calculate the eigenvalues of $A^{(1)}$ and $A^{(1)} - I$ to see if we can accept this replacement. Again, if all the eigenvalues are positive we accept, otherwise we reject and draw another u_1 and u_2 and repeat so that eventually the new entry is accepted. We repeat this for all the $a_{ij},\ i > j$.
At the end, we have a new matrix $A^{(1)} \in \mathbb{P}^+$. We repeat the algorithm this time by taking the $A^{(1)}$ as the initial matrix to construct $A^{(2)}$. We repeat this as many times as needed, m times for instance, to make sure the Markov chain constructed by this method, $\{A^{(m)}\}$ converges to its steady state.

Let π be the stationary distribution of the Markov chain induced by this method and let λ denote the 1-dimensional Lebesgue measure. It is easy to see that the chain $\{A^m\}$ is $\lambda-$irreducible and aperiodic: Let $D \in \mathbb{P}^+$ be a set with non zero Lebesgue measure, we have to show that

$$P(A^{n+1} \in D | A^n = x) > 0, \quad \text{for } x \in \mathbb{P}^+.$$

For any $x \in \mathbb{P}^+$, there is an $\epsilon \in [0, 1]$ such that for $u \in [-\epsilon, \epsilon]$, $f(x|u) > 0$. Therefore,

$$\begin{aligned} P(A^{n+1} \in D | u < \epsilon) &> 0, \qquad \text{i.e.,} \\ P(A^{n+1} \in D | A^n = x) &> P(A^{n+1} \in D | u < \epsilon) P(u < \epsilon | A^n = x) \\ &> 0. \end{aligned}$$

So, by Ergodic theorem the chain converges to π as $n \to \infty$.

Algorithm 2 Algorithm for simulating $A = (a_{ij})_{n\times n}$ from $\mathbb{P}^+$

while $k < m$ **do**
  for all $1 \leq i \leq n$ **do**
    $u \sim \text{Uniform}[-\epsilon, \epsilon]$
    $a_{ii} \leftarrow a_{ii} + u$
    if $A \in \mathbb{P}^+$ **then**
      Accept a_{ii}
    else
      Reject a_{ii}, draw another u and repeat!
    end if
  end for
  for all $i < j$ **do**
    $d1 \sim \text{Uniform}[-\epsilon, \epsilon]$
    $d2 \sim \text{Uniform}[-\epsilon, \epsilon]$
    $a_{ij} \leftarrow a_{ij} + (d1 + id2)$
    $a_{ji} \leftarrow a_{ji} + (d1 - id2)$
    if $A \in \mathbb{P}^+$ **then**
      Accept a_{ij}
    else
      Reject a_{ij}, draw another $d1$ and $d2$ and repeat!
    end if
  end for
  $k \leftarrow k + 1$
end while

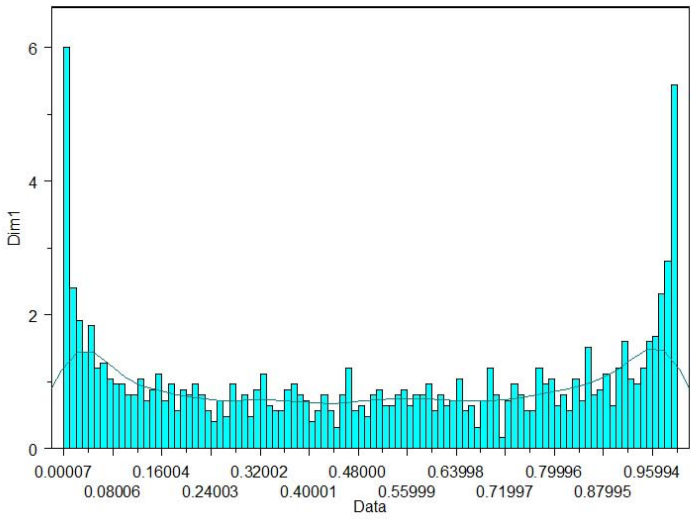


Figure 4.5: Histogram of eigenvalues of simulated matrix $A_{25\times 25} \in \mathbb{P}^+$.

Theorem 4.4.1 (Convergence of the algorithm)**.** *The transition kernel acts on the slice in the algorithm, preserves the uniform measure on the slice.*

Proof: In order to show this, we have to show the detailed balance property is satisfied in each step. Assume $A^n = \{a^n_{i,j}\}$ and $A^{n+1} = \{a^{n+1}_{i,j}\}$ are matrices simulated in steps nth and $(n+1)$th. Since we are choosing uniformly from the slice, it is sufficient to show:

$$P(A^{n+1} = b|\ A^n = a) = P(\ A^{n+1} = a|\ A^n = b).$$

According to the algorithm, for the entries on the diagonal we have, $a^{n+1}_{i,i} \to a^n_{i,i} + u$ where $u \in \text{Unif}[-\epsilon, +\epsilon]$ and $\epsilon \in \text{Unif}[0, 1]$. Therefore,

$$\begin{aligned} P(a^{n+1}_{i,i} = b|\ a^n_{i,i} = a) &= P(a + u = b | u \in \text{Unif}[-\epsilon, +\epsilon]) \\ &= P(a = b + u | u \in \text{Unif}[-\epsilon, +\epsilon]) \\ &= P(a^{n+1}_{i,i} = a|\ a^n_{i,i} = b). \end{aligned}$$

With the same argument we can show that detailed balance property also holds for the complex off-diagonal entries. This completes the proof.

4.4.2 Justifying the correctness of simulation with theoretical results

There is also another, easier method to generate random matrices from the $\mathbb{P}^+$, using random projectors and Jacobi unitary ensembles. We start by giving the relevant definitions.

Definition 17. *A projector is a square matrix P such that $P^2 = P$. If P is also hermitian, then P is an orthogonal projector.*

Definition 18. *[23] Let α and β be two positive real numbers. Consider the probability distribution on $M_n(\mathbb{C})_{sa}$ given by*

$$(Z_n^{\alpha,\beta})^{-1} \det(1-M)^{\alpha} \det(M)^{\beta} \mathbb{I}_{0 \leq M \leq 1} dM$$

where $Z_n^{\alpha,\beta}$ is a normalized constant. This probability measure is called Jacobi unitary ensemble of parameter (n, α, β).

Remark 19. *The probability measure the Jacobi unitary ensemble with parameter $(n, 0, 0)$ is the uniform distribution on $[0, 1]$.*

It was proved by Collins [12] that a random matrix $\pi_n \tilde{\pi}_n \pi_n$ is distributed according to a Jacobi ensemble of parameters $(q_n, n - q_n - \tilde{q}_n, \tilde{q}_n - q_n)$, where π_n and $\tilde{\pi}_n$ are random orthogonal projections of rank $\tilde{q}_n$ and $\tilde{q}_n$.
Let A be an $n \times n$ diagonal matrix, where the first $\left[\frac{n}{2}\right]$ diagonal entries of that are 1

and the rest are 0. Without loss of generality, we can assume n is even.

$$A = \begin{pmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & \vdots & \cdots & \cdots & 0 \\ 0 & 0 & 1 & \cdots & \cdots & \vdots \\ 0 & \cdots & \cdots & 0 & \cdots & \vdots \\ \vdots & \cdots & \cdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & \ldots & \cdots & 0 \end{pmatrix}_{n \times n}.$$

Clearly, $A \in \mathbb{P}^+$ and $\text{trace}(A) = \frac{n}{2}$ (i.e., A is of rank $\frac{n}{2}$). However, A is deterministic. Our goal is to generate random matrices form the set $\mathbb{P}^+$.

Lemma 4.4.2. *Let P_n be a random orthogonal projector of rank $\frac{n}{2}$. Let the random matrix $B_n \in M_n(\mathbb{C})_{sa}$ be $B_n = P_n A P_n$. Then we have, $B_n \in \mathbb{P}^+$.*

Proof: Note that both A and P_n are positive, hermitian matrices with operator norm less or equal to 1 of rank $\frac{n}{2}$. Moreover, by [12] the random matrix $P_n A P_n$ is distributed according to a Jacobi ensemble of parameter $(\frac{n}{2}, 0, 0)$, which by Remark 19 is the uniform distribution on $[0, 1]$.

Free projectors and arcsine distribution law

According to Nica, Speicher and Voiculescu [50]:

> 'Free probability theory is a line of research which parallels aspects of classical probability, in a non-commutative context where tensor products are replaced by free products, and independent random variables are replaced by free random variables. It grew out from attempts to solve some longstanding problems about von Neumann algebras of free groups.'

Definition 19. *[51] Let $(\mathcal{A}, \varphi)$ be a non-commutative probability space and let I be a fixed index set. For each $i \in I$, let $\mathcal{A}_i \subset \mathcal{A}$ be a unital subalgebras $(\mathcal{A}_i)_{i\in I}$ are called* ***freely independent****, if*

$$\varphi(a_1 \dots a_k) = 0$$

whenever we have the following:

- *k is a positive integer;*
- *$a_j \in \mathcal{A}_{i(j)}$ $(i(j) \in I)$ for all $j = 1, \dots, k$;*
- *$\varphi(a_j) = 0$ for all $j = 1, \dots, k$;*
- *and neighboring elements are from different subalgebras, i.e.,*

$$i(1) \neq i(2), i(2) \neq i(3), \dots, i(k-1) \neq i(k).$$

Definition 20. *Let A be an hermitian matrix with eigenvalues $\lambda_1, \cdots, \lambda_n$. Let μ denote the empirical distribution on the set of its eigenvalues:*

$$\mu(A) = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i}.$$

Let π_n and $\tilde{\pi}_n$ be orthogonal projectors of $M_n(\mathbb{C})$ of rank q_n and $\tilde{q}_n$ such that $q_n \sim \alpha n$ and $\tilde{q}_n \sim \beta n$ as $n \to \infty$. It has been shown in [12] that π_n and $\tilde{\pi}_n$ are asymptotically

free. Therefore, by [65] $\pi_n\tilde{\pi}_n\pi_n$ has an empirical eigenvalues distribution converging toward $\mu_1 \boxtimes \mu_2$, where μ_1 is the probability

$$(1-\alpha)\delta_0 + \alpha\delta_1$$

and μ_2 is the probability

$$(1-\beta)\delta_0 + \beta\delta_1.$$

Let

$$r_\pm = \alpha + \beta - 2\alpha\beta \pm \sqrt{4\alpha\beta(1-\alpha)(1-\beta)}.$$

Using the standard S-transform argument ([66], example 3.6.7), we have [12]

$$\mu_1 \boxtimes \mu_2 = [1-\min(\alpha,\beta)]\delta_0 + [\max(\alpha+\beta-1,0)]\delta_1 + \frac{\sqrt{(r_+-x)(x-r_-)}}{2\pi x(1-x)}\mathbb{I}_{[r_-,r_+]}dx.$$

Lemma 4.4.3. *The empirical distribution of the eigenvalues of the random matrix B_n constructed in Lemma* 4.4.2 *converges to the arcsine law.*

Proof: Note that $B_n = U_nAU_n$, where U_n and A_n are both of rank $\frac{n}{2}$, which implies $\alpha = \beta = \frac{1}{2}$. Therefore, B_n has an empirical distribution converging to $\mu \boxtimes \mu$, where μ is the probability $\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$. Having $\alpha = \beta = \frac{1}{2}$ gives us $r_+ = 1$ and $r_- = 0$. So we have,

$$\begin{aligned}\mu \boxtimes \mu &= [1-\frac{1}{2}]\delta_0 + \frac{\sqrt{(1-x)x}}{2\pi x(1-x)}\mathbb{I}_{[0,1]}dx \\ &= \frac{1}{2}\delta_0 + \frac{1}{2\pi\sqrt{x(1-x)}}\mathbb{I}_{[0,1]}dx.\end{aligned}$$

where $\mathbb{I}_n$ is the identity matrix.

4.4.3 Method of simulation matrices from $\mathbb{P}^+$ with a unit trace

Let

$$\mathbb{P}_1^+ = \{A_{n,n} \in \mathbb{P}^+ | \mathrm{Tr}(A) = 1\}.$$

Our goal is to generate uniformly random matrices from $\mathbb{P}_1^+$.

The algorithm is almost the same as the algorithm for generating matrices from the set $\mathbb{P}^+$. The only difference is that the trace stays equal to 1 during the process. For this purpose, we start from an arbitrary matrix form the compact set $\mathbb{P}_1^+$. For instance let $A^0 = \frac{1}{n}\mathbb{I}_n$. Clearly $A^0 \in \mathbb{P}_1^+$. The restriction for the trace only affects the entries on the main diagonal. So for off-diagonal entries, the algorithm will be the same as the one defined for $\mathbb{P}^+$.

As before, we generate the entries on the main diagonal separately. In order to simulate the first entry of the main diagonal, a_{11}, as before we draw u uniformly from the $[-\epsilon, \epsilon]$. However, this time, we also need to update a_{22} at the same time, to keep the trace equal to 1. Therefore, we replace a_{11} with $a_{11} + u$, and replace a_{22} with $a_{22} - u$. Since the trace will remain 1, in each iteration we only need to check the eigenvalues of A to make sure they are positive. If all the eigenvalues are positive, the replacement is accepted. Otherwise we reject and we draw another u.

Theorem 4.4.2 (Convergence of the algorithm)**.** *The transition kernel acts on the slice in the algorithm, preserves the uniform measure on the slice.*

Proof: As before, in order to show this, we have to show the detailed balance property is satisfied in each step. The algorithm is the same as the one we used for sampling from $\mathbb{P}^+$ for the off-diagonal entries. So only we have to show the detailed balance property is satisfied for diagonal entries where, $a_{i,i}$, $2 \leq i \leq n$.

Assume $A^n = \{a_{i,j}^n\}$ and $A^{n+1} = \{a_{i,j}^{n+1}\}$ are matrices simulated in steps nth and $(n+1)$th. According to the algorithm, for the entries on the diagonal we have, for $2 \leq i \leq n$, $a_{i,i}^{n+1} \to (a_{i,i}^n - u) + v$ where $u, v \in \text{Unif}[-\epsilon, +\epsilon]$ and $\epsilon \in \text{Unif}[0, 1]$. Therefore,

$$\begin{aligned} P(a_{i,i}^{n+1} = b |\ a_{i,i}^n = a) &= P(a - u + v = b | u, v \in \text{Unif}[-\epsilon, +\epsilon]) \\ &= P(a = b - u + v | u, v \in \text{Unif}[-\epsilon, +\epsilon]) \\ &= P(a_{i,i}^{n+1} = a |\ a_{i,i}^n = b). \end{aligned}$$

Algorithm 3 Algorithm for simulating diagonal entries of $A = (a_{ij})_{n\times n}$ from $\mathbb{P}_1^+$

while $k < m$ **do**
  for all $1 \leq i \leq n-1$ **do**
    $u \sim \text{Uniform}[-\epsilon, \epsilon]$
    $a_{i,i} \leftarrow a_{ii} + u$
    $a_{i+1,i+1} \leftarrow a_{i+1,i+1} - u$
    if $A \in \mathbb{P}_1^+$ **then**
      Accept a_{ii}
    else
      Reject a_{ii} and draw another u and repeat!
    end if
  end for
  $k \leftarrow k + 1$
end while

The Wishart ensemble

Let $A_{n\times k}$ complex matrix whose entries are i.i.d. Normal$(0, 1)$ random variables. Then AA^* is called a Wishart matrix with parameters n and k. Wishart matrices are positive and have n positive eigenvalues.

It is known ([43]) that if W is a Wishart matrix with parameters n and k, then the distribution of the eigenvalues $(\lambda_1, \cdots, \lambda_n)$ has a density with respect to the Lebesgue measure on $\mathbb{R}_+^n$ given by

$$\phi_{n,k} = C_{n,k} \exp\left(-\sum_{i=1}^{n} \lambda_i\right) \prod_{i=1}^{n} \lambda_i^{k-n} \Delta(\lambda)^2,$$

where

$$C_{n,k}^{\omega} = \left(\prod_{j=0}^{n-1} \Gamma(n+1-j)\Gamma(k-j)\right)^{-1}$$

and

$$\Delta(\lambda) = \prod_{1 \leq i < j \leq n} (\lambda_i - \lambda_j).$$

When we have a Wishart matrix of parameter 1, ($W = AA^*,\ A_{n\times n}$), $\prod_{i=1}^{n} \lambda_i^{k-n}$ will disappear, since $n = k$, and we have that W has density function with respect to Lebesgue measure on positive definite matrices of the form $\exp(-\mathrm{Tr}(W))$. Hence, when we divide W by its trace, we induce the uniform measure on the compact set of positive definite matrices with unit trace [49]. Therefore, matrices simulated from the $\mathbb{P}_1^+$ behave like a normalized Wishart matrices.

Behaviour of the largest eigenvalue and justifying the correctness of simulation with theoretical results

A very interesting topic in random matrix theory is the study of the largest eigenvalue. Since $A \in \mathbb{P}_1^+$ behaves like a Wishart matrix, the largest eigenvalue of a matrix A behaves also like the largest eigenvalue in the Wishart case.
Consider a sequence of random Wishart matrices $(W_n)_n$ with parameters n and $k(n)$ and let $\lambda_{\max}(W_n)$ be the largest eigenvalue of W_n. The following theorem discuss the convergence of the largest eigenvalue of a Wishart matrix.

Theorem 4.4.3. *[49] Let $c \in (0, \infty)$ and $k(n)_n$ be a sequence of integers such that $\lim_{n\to\infty} \frac{k(n)}{n} = c$. Then we have almost surely*

$$\lim_{n\to\infty} \frac{1}{n} \lambda_{max}(W_n) = (\sqrt{c} + 1)^2.$$

Back to our case, simulating matrices from $\mathbb{P}_1^+$, having a unit trace implies that the eigenvalues behave like $\frac{k}{n}$, where k is a constant. Moreover, we have a flat density on square random matrices, so $c = 1$. So by Theorem 4.4.3, we have that the largest eigenvalue of the simulated matrix from $\mathbb{P}_1^+$ cannot exceed $\frac{4}{n}$.

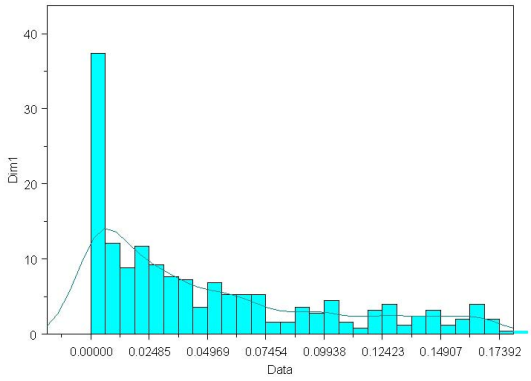


Figure 4.6: Histogram of eigenvalues of simulated matrix $A_{20\times 20} \in \mathbb{P}_1^+$ which behaves like Wishart matrix. The largest eigenvalue is almost $.20 \sim \frac{4}{n}$.

Appendix A

Splus codes for Algorithms

Splus code for Rejection Method

```
n<-20  # n is the dimension of the Matrix
n2<-n*n
gamma<-0.5
c<-1+gamma/2
A<-matrix(0,n,n)
B<-matrix(0,n,n)
t<-1
alpha<-(1/sqrt(n)^n)*(1/sqrt(2*pi))*(1/sqrt(2))^(n^2-n)
eigenA<-rep(0,1)
eigenB<-rep(0,1)
eigenAB<-rep(0,1)
tracepoly<-rep(0,1)
f<-rep(0,1)
g1<-rep(0,1)
g2<-rep(0,1)
g<-rep(0,1)
traceA2<-rep(0,1)
traceB2<-rep(0,1)
counter<-0

#Rejection method
while(t<2){
 counter<-counter+1
 a<-array(sqrt(1/n)*rnorm(n*n,0,sqrt(1/2)))
 b<-array(sqrt(1/n)*rnorm(n*n,0,sqrt(1/2)))
 A<-matrix(a,ncol=n)
 B<-matrix(b,ncol=n)
 #Make A and B symmetric
 for (l in 1:n-1){
  for (p in (l+1):n){
   A[l,p]=A[p,l]
   B[l,p]=B[p,l]    }}
    # We have different variance for the diagonal entries
    for(h in 1:n){
     A[h,h]<-sqrt(1/n)*rnorm(1,0,1)
     B[h,h]<-sqrt(1/n)*rnorm(1,0,1)}
     #create function f
     eigenA2<-Re(eigen(A%*%A)$value)
     eigenB2<-Re(eigen(B%*%B)$value)
     eigenAB<-Re(eigen(A%*%B)$value)
     tracepoly<-(sum(eigenA2)+sum(eigenB2)+gamma*sum(eigenAB))
     f<-exp(-tracepoly)
```

```
  #create function g1 and g2

  traceA2<-sum(eigenA2)
  traceB2<-sum(eigenB2)

  g1<-exp(-(c/2)*traceA2)
  g2<-exp(-(c/2)*traceB2)
  g<-g1*g2
  #create an envelope
  e<-g/alpha
  #Rejection method
   u<-runif(1,0,1)
  if(u<f/e){resultA<-A; resultB<-B; t<-2}
 }
# We store the eigenvalues to find the empirical distribution
eigenresultA<-eigen(resultA)$value
eigenresultB<-eigen(resultB)$value
```

Simulating a symmetric n by n random matrix A~ Exp (-n/4 trace(A^4))

```
# n is the dim of the matrix and m is
the number of iterations
n<-4
m<-100
# initial matrix we start from
A0<-matrix (.01,n,n)
count<-0  count2<-0   count3<-0
k<-1
y<-0
l1<-0
l2<-0

eigenA0<-matrix (0,1,n)
eigenarrayA0<-array(0,n*m)

# {Main program, we repeat the algorithm for m times}

 while(k<m+1){ #{while loop }

#{each time we fix all other entries, except i,j th  entry and simulate A[i,j]}
    for(i in 1:n){   #{for loop 1}
       for(j in 1:n){   #{for loop 2}

    #{Since the matrix is symmetric, we  just need to simulate the upper triangle}
    if(j>=i){  #{if loop }

     A0[j,i]<-A0[i,j]

     #{calculating the trace of matrix in order to calculate f}
     A04<-A0%*%A0%*%A0%*%A0
     traceA04<-0
      for(ii in 1:n)
       traceA04<-traceA04+A04[ii,ii]
      #{we define function f here and each time needed we call the function}
      f<-function(A0) {return(exp(-(traceA04)))}
      t<-f(A0)
      check<-1 #{a and b are the values that we need them for doubling procedure}
```

```
 a<-1/sqrt(n)
 b<-1/sqrt(n)

#{drawing uniformly from the vertical slice}
 u1<-runif(1,0,t)
 r<-A0[i,j]
  check2<-1

 #{ we are trying to find the right interval I, in this loop we want to make sure the right bound is
     good,  "count2" counts the number of times we use doubling procedure for the right
      bound}
 while(check2>0){

 A0[i,j]<-r+a
 A0[j,i]<-r+a
 A04<-A0%*%A0%*%A0%*%A0
 traceA04<-0

  for(ii in 1:n)
   traceA04<-traceA04+A04[ii,ii]

  f1<-f(A0)
  if(u1>f1) check2<-0 else {a<-a*2; count2<-count2+1}
  }

 A0[i,j]<-r
 check3<-1

#{In this loop we want to make sure the left bound is good,
 "count3" counts the number of times we use doubling procedure for the left bound}
 while(check3>0){

  A0[i,j]<-r-b
  A0[j,i]<-r-b
  A04<-A0%*%A0%*%A0%*%A0
  traceA04<-0

  for(ii in 1:n)
   traceA04<-traceA04+A04[ii,ii]

  f2<-f(A0)
  if(u1>f2) check3<-0 else {b<-b*2; count3<-count3+1}
 } #{Interval has been found; we sample "u"  uniformly form the interval , we check to see if we
 can u  is drawn from the slice, otherwise we have to reject u and draw another sample.
 Command "count" counts the number of times we reject}
 while{check>0}
```

```
      u<-runif(1,r-b,r+a)
      A0[i,j]<-u
      A0[j,i]<-u
      A04<-A0%*%A0%*%A0%*%A0
      traceA04<-0
      for(ii in 1:n)

       traceA04<-traceA04+A04[ii,ii]
      fu<-f(A0)
      if(fu>u1){check<-0;A0[i,j]<-u ; A0[j,i]<-u} else { count<-count+1}
   #{shrinking method}
   if(u>r){right<-u; l1<-l1+1} else{left<-u;l2<-l2+1}
   }

      } #{end of if loop}
   }#{end of for loop2}
 } #{end of for loop 1}

   #{each time we calculate and store the eigenvalues of the simulated matrix and we keep them in an
   array "eigenarray so that later we can make the histogram of them}
  if((k/s)==floor(k/s))
   {eigenA0<-eigen(A0)$value;
    for(h in 1:n){eigenarrayA0[y+h]<-eigenA0[h]; }
    y<-y+n}
   k<-k+1

}
```

Simulating a pair of symmetric n by n random matrices (A,B)~ Exp(-n/2 trace(A^2+B^2+cAB))

```
n<-30 #{dim of the matrix}
m<-100#{number of time we run the program}
c<-0.5 # {the constant in the density function}
v1<-array(rnorm(n*n,0,1))
v2<-array(rnorm(n*n,0,1))
A0<-matrix(v1,n,n)
B0<-matrix(v2,n,n)
D0<-matrix(0,n,n)
count<-0
count2<-0
count3<-0
countb<-0
count2b<-0
count3b<-0
k<-1
y<-0
s<-1 #{skipping every s  eigenvalues}
burn<-200  #burn-in phase
eigenA0<-matrix(0,1,n)
eigenB0<-matrix(0,1,n)
eigenD0<-matrix(0,1,n)
eigenarrayA0<-array(0,n*m/s)
eigenarrayB0<-array(0,n*m/s)
eigenarrayD0<-array(0,n*m/s)    # D is the correlation between A and B
eigenarrayA<-array(0,n*m/s-burn)#eigenvalues of matrix A after burn-in phase
eigenarrayB<-array(0,n*m/s-burn)#eigenvalues of matrix B after burn in phase
eigenarrayD<-array(0,n*m/s-burn)
l1<-0   , l2<-0, l1b<-0, l2b<-0
#Main program
while(k<m+1){ # number of iteration
 # simulating matrix A
 # each time we fix all other entries and simulate one.

 for(i in 1:n){
 for(j in 1:n){
  # the matrices should be symmetric, so we only need to simulate the upper
triangle
  if(j>=i){
   a<-6/sqrt(n)  #initial guess for doubling procedure
   b<-6/sqrt(n)
   A0[j,i]<-A0[i,j]

   A02<-A0%*%A0
   B02<-B0%*%B0
   A0B0<-A0%*%B0

   f<-function(A0,B0) {return((exp(-
    (n/2)*(sum(diag(A02))+c*sum(diag(A0B0))+sum(diag(B02)))))}
   t<-f(A0,B0)
   check<-1
   #vertical slice
   u1<-runif(1,0,t)
   r<-A0[i,j]
   check2<-1
```

```
   #finding proper interval for that contain the  horizontal slice
   #doubling and shrinking procedure for right end
   while(check2>0){
     A0[i,j]<-r+a
     A0[j,i]<-r+a
     A02<-A0%*%A0
     A0B0<-A0%*%B0

     f1<-f(A0,B0)
     if(u1>f1) check2<-0 else {a<-a*2; count2<-count2+1}

   }

   A0[i,j]<-r
   check3<-1
   #doubling and shrinking  procedure for left end
   while(check3>0){
    A0[i,j]<-r-b
    A0[j,i]<-r-b
    A02<-A0%*%A0
    A0B0<-A0%*%B0
    f2<-f(A0,B0)
    if(u1>f2) check3<-0 else {b<-b*2; count3<-count3+1}
   }

   left<-r-b
   right<-r+a
   # We draw u uniformly from the interval we find, and use the rejection method
     to make sure u is drawn from the horizontal slice
   while(check>0){
    u<-runif(1,left,right)
    A0[i,j]<-u
    A0[j,i]<-u
    A02<-A0%*%A0
    A0B0<-A0%*%B0
    fu<-f(A0,B0)
    if(fu>u1){check<-0;A0[i,j]<-u ; A0[j,i]<-u} else { count<-count+1}
    #each time we reject, we shrink the interval
    if(u>r){right<-u; l1<-l1+1} else{left<-u;l2<-l2+1}

 }}}}
# At this point the matrix A is simulated, we fix A and start simulating matrix B

 A02<-A0%*%A0
#simulating matrix B
 for(i in 1:n){
 for(j in 1:n){

  if(j>=i){

  a<-6/sqrt(n)
  b<-6/sqrt(n)

  B0[j,i]<-B0[i,j]

  B02<-B0%*%B0
  A0B0<-A0%*%B0
```

```
  t<-f(A0,B0)
  check<-1
  u1<-runif(1,0,t)
  r<-B0[i,j]
  check2<-1

   while(check2>0){

    B0[i,j]<-r+a
    B0[j,i]<-r+a
    B02<-B0%*%B0
    A0B0<-A0%*%B0

    f1<-f(A0,B0)
    if(u1>f1) check2<-0 else {a<-a*2; count2b<-count2b+1}

   }

   B0[i,j]<-r

   check3<-1

   while(check3>0){

    B0[i,j]<-r-b
    B0[j,i]<-r-b
    B02<-B0%*%B0
    A0B0<-A0%*%B0

    f2<-f(A0,B0)
    if(u1>f2) check3<-0 else {b<-b*2; count3b<-count3b+1}
   }

  left<-r-b
  right<-r+a

  while(check>0){
   u<-runif(1,left,right)
   B0[i,j]<-u
   B0[j,i]<-u
   B02<-B0%*%B0
   A0B0<-A0%*%B0

   fu<-f(A0,B0)
   if(fu>u1){check<-0;B0[i,j]<-u ; B0[j,i]<-u} else { countb<-countb+1}
   if(u>r){right<-u; l1b<-l1b+1} else{left<-u;l2b<-l2b+1}
 }}}}
 #correlation between A and B
 #each time we also calculate AB-BA and store its eigenvalues
 D<-A0%*%B0-B0%*%A0
 #finding and storing eigenvalues
 if((k/s)==floor(k/s))
 {eigenA0<-eigen(A0)$value; eigenB0<-eigen(B0)$value; eigenD0<-eigen(D0)$value
 for(h in 1:n){eigenarrayA0[y+h]<-eigenA0[h];eigenarrayB0[y+h]<-eigenB0[h];
eigenarrayD0[y+h]<-eigenD0[h]}
 y<-y+n}
```

```
 k<-k+1
# burn-in phase: we throw out the first 100 or 200 samples before
collecting the statistics

 pp<-m*n
 burn1<-burn+1
 for( i in burn1:pp){
  eigenarrayA[i-burn]<-eigenarrayA0[i]
  eigenarrayB[i-burn]<-eigenarrayB0[i]
  eigenarrayD[i-burn]<-eigenarrayD0[i]
 }
}
```

```
n<-100     % n is the size of matrix
m<-100     % m is number of times we want to repeat the algorithm
G<-matrix(0,n,n)
G2<-matrix(0,n,n)
rpart<-array(rnorm(n,0,1))
ipart<-array(rnorm(n,0,1))
C<-outer(rpart, ipart*1i, "+") % generate random complex normal matrix
B<-qr(C)
A1<-qr.Q(B,complete=TRUE)   }   Using QR decomposition to get a Unitary matrix
R<-qr.R(B,complete=TRUE)    }
r1<-diag(R)
for(i in 1:n) {G[i,i]<-(r1[i]/Mod(r1[i]))}  % R should have positive diagonal entries
A<-A1%*%G   %A will be the initial matrix to start with
```

Figure A.1: Splus code for QR decomposition

```
f<-function(A) {return(exp(-n*(sum(diag(A+t(A))))))}
k<-1
p<-0
eigenA<-matrix(0,1,n)
eigenarrayA<-array(0,n*m)
count<-0
count2<-0
while(k<m+1){
y<-runif(1,0,f(A))

for(i in 1:n-1){ % in this loop for n-1 times we multiply our matrix with the ith rotati
matrix
l<-1
while(l<2){
theta<-runif(1,-180,180) % draw randomly
R<-diag(n)
options(digits=20)
R[i,i]<-cos(theta*pi/180)
R[i+1,i+1]<-cos(theta*pi/180)
R[i+1,i]<-sin(theta*pi/180)
R[i,i+1]<--sin(theta*pi/180)
M2<-A%*%R
B<-qr(M2)
M1<-qr.Q(B,complete=TRUE)
MR<-qr.R(B,complete=TRUE)
r2<-diag(MR)
for(i in 1:n) {G2[i,i]<-(r2[i]/Mod(r2[i]))}
M<-M1%*%G2
f1<-f(M)
if(f1>y) {l<-2;A<-M} else count<-count+1
}
count2<-count2+1
}
eigenA<-eigen(A)$value
for(h in 1:n){eigenarrayA[p+h]<-eigenA[h] }  % storing the eigenvalue each time
p<-p+n
k<-k+1

}
```

Constructing the ith Rotating matrix

Figure A.2: Splus code for simulating unitary matrix from $f \sim \exp(-n\mathrm{Tr}(A+A^*))$.

```
# simulating entries on the diagonal
while (check3>0){
 u<-runif(1,-0.05/ss,0.05/ss)
 d<-A[i,i]
 A[i,i]<-d+u
eigenA<-eigen(A)$value
eigenI<-eigen(Id-A)$value
check4<-1
for(p in 1:n){
 if(Re(eigenA[p])<0) check4<-0
 if(Re(eigenI[p]<0)) check4<-0}
if(check4>0) {check3<-0} else {count3<-count3+1; A[i,i]<-d}
if (count3>10) {ss<-10; count4<-count4+1}
if(count3>30)  {ss<-100; count3<-0}
}#end while
# Simulating entries on the diagonal
 for(i in 1:n){
 for(j in 1:n){
  if(j>i){
   check<-1
  while(check>0){
   uni1<-runif(1,-0.05/s,0.05/s)
   uni2<-runif(1,-0.05/s,0.05/s)
   u1<-outer(uni1, uni2*1i, "+")
   u2<-outer(uni1, uni2*1i, "-")
  d1<-A[i,j]
  d2<-A[j,i]
  A[i,j]<-u1+d1
  A[j,i]<-u2+d2
  eigenA<-eigen(A)$value
  eigenI<-eigen(Id-A)$value
  check2<-1
  for(p in 1:n){
   if(Re(eigenA[p])<0) check2<-0
   if(Re(eigenI[p]<0)) check2<-0}
if(check2>0) {check<-0} else {count<-count+1; A[i,j]<-d1; A[j,i]<-d2}
if (count>10) {s<-10; count2<-count2+1}
if(count>30){s<-100; count<-0}
}#end while
```

Here, we calculate the eigenvlaues of A and I-A. We reject the new point is any of the eigenvalues is negative.

This command is for the cases where we reach the boundary of the set. If this cases, we make our movements 10 or 100 times shorter.

Here, we are generating complex number uni1, by drawing picking u1 and u2 uniformly. We add uni to the ij th entry of the matrix and its conjugate to ji th entry.

Here, we calculate the eigenvlaues of A and I-A. We reject the new point is any of the eigenvalues is negative.

Figure A.3: Splus code for simulation matrices from $\mathbb{P}^+$.

```
while(check3>0){

u<-runif(1,-0.05/ss,0.05/ss)
d1<-A[i,i]
d2<-A[i+1,i+1]
A[i,i]<-d1+u
A[i+1,i+1]<-d2-u

eigenA<-eigen(A)$value

check4<-1

for(p in 1:n){
if(Re(eigenA[p])<0) check4<-0


if(check4>0) {check3<-0} else {count3<-count3+1;A[i,i]<-d1;A[i+1,i+1]<-d2}


}#end while
```

Figure A.4: Splus code for simulating entries of diagonal of $A_{n\times n}$ from $\mathbb{P}_1^+$.

Bibliography

[1] Besag, J.; Green, P. (1993) *Spatial statistics and Bayesian computation.* J. Roy. Statist. Soc. Ser. B 55, no. 1, 2537.

[2] Beyer, William H. (1991) *CRC Standard Mathematical Tables and Formulae* 25th edn, CRC.

[3] Biane, P. (1998) *Free probability for probabilists.* MSRI Preprint 1998-040, Mathematical Sciences Research Institute.

[4] Billingsley, P. (1986) *Probability and Measure.* 3rd edn, Wiley, NewYork.

[5] Billingsley, P. (1990) *Convergence of Probability measures.* 2nd edn, Wiley, NewYork.

[6] Bohigas, O.; Giannoni, M.J.; Schmit, Schmit (1984). *Characterization of Chaotic Quantum Spectra and Universality of Level Fluctuation Laws.* Phys. Rev. Lett. 52: 14.

[7] Bremaud, Pierre. (1998) *Markov Chains. Gibbs Fields, Monte Carlo Simulation, and Queues.* Texts in Applied Mathematics , Vol. 31.

[8] Bruijn, N.G; Knuth,D.; Rice, O. (1972) *The average height of planted plane trees, in: Graph Theory and Computing*, pp. 1522, Academic Press, NewYork.

[9] Casella, G.; Robert, P. (2004) *Monte Carlo statistical methods.* Second edition. Springer Texts in Statistics. Springer-Verlag, New York.

[10] Chung, K. (1975) *Maxima in Brownian excursions.* Bull. Amer. Math. Soc. 81 , 742–745.

[11] Cohn, Donald L. (1980) *Measure theory.* Birkhäuser, Boston, Mass., . ix+373 pp. ISBN: 3-7643-3003-1 (Reviewer: Dieter Hoffmann) 28-01

[12] Collins, B. (2005) *Product of random projections, Jacobi ensembles and universality problems arising from free probability.* Probab. Theory Related Fields 133 , no. 3, 315344.

[13] Damien, P.; Walker, S. G. (2001) *Sampling truncated normal, beta, and gamma densities.* J. Comput. Graph. Statist. 10, no. 2, 206215

[14] Deift, P. (2000) *Orthogonal polynomials and random matrices: a Riemann-Hilbert approach.* Courant Lecture Notes in Mathematics, 3. New York University, Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI.

[15] Durrett, R; Iglehart, D. (1977) *Functionals of Brownian meander and Brownian excursion.* Ann. Probability 5 , no. 1, 130135.

[16] Durrett, R. (1995) *Probability: Theory and Examples.* 2nd ed, Duxbury Press, Belmont, Calif.

[17] Eaton, Morris L. (1983) *Multivariate statistics. A vector space approach.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York.

[18] Edelman, A. ; Rao, N. (2005) *Random matrix theory.* (English summary) Acta Numer. 14 , 233–297.

[19] Edwards, R. G.; Sokal, A. D. (1988) *Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm.* Phys. Rev. D (3) 38 , no. 6, 20092012.

[20] Faraut, J; Kornyi, A. (1994) *Analysis on symmetric cones.* Oxford Mathematical Monographs. Oxford Science Publications.

[21] Feller, W. (1968) *An introduction to probability theory and its applications. Vol. I.* Third edition John Wiley and Sons, Inc., New York-London-Sydney.

[22] Feller, W. (1971) *An introduction to probability theory and its applications. Vol. II.* Second edition John Wiley and Sons, Inc., New York-London-Sydney.

[23] Forrester, P. (2010) *Los-gases and Random Matrices* London Mathematical Society Monographs Series, 34. Princeton University Press, Princeton, NJ.

[24] Freedman, D. (1971) *Brownian motion and diffusion.* Holden-Day, San Francisco, Calif.-Cambridge-Amsterdam. xii+231 pp. (Reviewer: W. E. Pruitt) 60J65

[25] Geman, S.; Geman, D. (1984) *Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images.* IEEE Transactions on Pattern Analysis and Machine Intelligence 6 (6): 721741

[26] Givens, Geof H.; Hoeting, Jennifer A. (2005) *Computational statistics.* Wiley Series in Probability and Statistics.

[27] Goodman, F.; De la Harpe, P.; Jones,V. (1989) *Coxeter graphs and towers of algebras*, MSRI Publications 14, Springer Verlag.

[28] Hansen, F.; Pedersen, G. K. (2003) *Jensen's operator inequality.* Bull. London Math. Soc. 35, no. 4, 553564.

[29] Harris, T. E. (1956) *The existence of stationary measures for certain Markov processes.* Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 19541955, vol. II, pp. 113124. University of California Press, Berkeley and Los Angeles.

[30] Hastings, W. K. (1970) *Monte Carlo sampling methods using Markov chains and their applications* Biometrika, vol. 57, no. 1, pp. 97–109.

[31] Hiai, F.; Petz, D. (2000) *The semicircle law, free random variables and entropy.* (English summary) Mathematical Surveys and Monographs, 77. American Mathematical Society, Providence, RI.

[32] Hida, T. (1980) *Brownian motion.* Springer-Verlag, Applications of Math. vol. 11.

[33] Hobert, J.; Jones, G.; Presnell, B.; Rosenthal, Jeffrey S. (2002) *On the applicability of regenerative simulation in Markov chain Monte Carlo.* (English summary) Biometrika 89, no. 4, 731743.

[34] Itô, K; McKean, H. (1965) *Diffusion Processes and Their Sample Paths.* Jr. Academic Press, New York; Springer, Berlin, 1965.

[35] Janson, S. (2007) *Brownian excursion area, Wright's constants in graph enumeration, and other Brownian areas.* (English summary) Probab. Surv. 4, 80–145 (electronic). 60J65.

[36] Johnstone, I. M. (2001) *On the distribution of the largest eigenvalue in principal components analysis.* Ann. Statist. 29 295327.

[37] Kaigh, W. D. (1976) *An invariance principle for random walk conditioned by a late return to zero.* Ann. Probability 4, no. 1, 115–121.

[38] Kennedy, Douglas P. (1976) *The distribution of the maximum Brownian excursion.* J. Appl. Probability 13 , no. 2, 371–376.

[39] Khorunzhiy, O; Marckert, J. F. (2009) *Uniform bounds for exponential moments of maximum of Dyck paths.* Elect. Comm. in Probab. 14 (2009), 327333.

[40] Knight, Frank B. (1981) *Essentials of Brownian Motion and Diffusion.* A. M. S. Surveys, Vol. 18, Amer. Math. Soc., Providence RI.

[41] Kousha, T. (2012) *Asymptotic behavior and moderate deviation principle for the maximum of a Dyck path.* Statistics and Probability Letters 82 pp. 340-347

[42] Lin, S.; Thompson, E.; Wijsman, E. (1993) *Achieving irreducibility of the Markov chain Monte Carlo method applied to pedigree data.* (English summary) IMA J. Math. Appl. Med. Biol. 10, no. 1, 117. 92D10

[43] Mehta, M. (2004) *Random marices* Elsevier/Academic Press.

[44] Meyn, S.; Tweedie, Richard L. (2009) *Markov chains and stochastic stability.* Second edition. With a prologue by Peter W. Glynn. Cambridge University Press, Cambridge.

[45] Mezzadri, F. (1997) *How to Generate Random Matrices from the Classical Compact Groups* Notices of the AMS, Volume 54, Issue 5, pp. 592-604.

[46] Mira, A.; Tierney, L. (2002) *Efficiency and convergence properties of slice samplers.* (English summary) Scand. J. Statist. 29, no. 1, 112.

[47] Morters, P. (2008) *Large deviation theory and applications.*

[48] Neal, Radford M. (2003) *Slice sampling.* Ann. Statist. 31 , no. 3, 705–767.

[49] Nechita, I. (2007) *Asymptotics of random density matrices.* Ann. Henri Poincar 8 , no. 8, 15211538.

[50] Nica, A; Speicher, R; Voiculescu, D. (2004) *Free Probability Theory.*

[51] Nica, A; Speicher, R. (2006) *Lectures on the combinatorics of free probability.* London Mathematical Society Lecture Note Series, 335. Cambridge University Press, Cambridge.

[52] Nummelin, E. (1984) *General irreducible Markov chains and non-negative operators.* Cambridge University press

[53] O'Brien, G. L. (1974) *Limit theorems for sums of chain-dependent processes.* J. Appl. Probability 11 , 582587.

[54] Richard, Ch. (2007) *On q-functional equations and excursion moments* . Fakultät fr Mathematik, Universität Bielefeld.

[55] Robert, Christian P.; Casella, G. (2004) *Monte Carlo Statistical Methods.* Series: Springer Texts in Statistics 2nd ed.

[56] Roberts, G. O.; Smith, A. F. M. (1994) *Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms.* (English summary) Stochastic Process. Appl. 49, no. 2, 207216.

[57] Roberts, G. O.; Rosenthal, Jeffrey S. (1997) *Geometric ergodicity and hybrid Markov chains.* Electron. Comm. Probab. 2, no. 2, 1325 (electronic).

[58] Roberts, G. O.; Rosenthal, Jeffrey S. (1999) *Convergence of slice sampler Markov chains.* J. R. Stat. Soc. Ser. B Stat. Methodol. 61, no. 3, 643660.

[59] Roberts, G. O. ; Rosenthal, Jeffrey S. (2006) *Harris recurrence of Metropolis-within-Gibbs and trans-dimensional Markov chains.* Ann. Appl. Probab. 16 , no. 4, 21232139.

[60] Swendsen, R. H.; Wang, J. (1986) *Monte Carlo simulation of spin-glasses.* Phys. Rev. Lett. 57, no. 21, 26072609, 82A57 (82-08)

[61] Tierney, L. (1994) *Markov chains for exploring posterior distributions.* Ann. Statist. 22, no. 4, 17011762.

[62] Trotter, H.; Tukey, J. (1956) *Conditional Monte Carlo for normal samples. Symposium on Monte Carlo methods.* University of Florida, 1954, pp. 6479. John Wiley and Sons, Inc., New York; Chapman and Hall, Limited, London.

[63] Tulino, A. M; Verdu. S. (2004) *Random matrix theory and wireless communications*. Foundations and Trends in Communications and Information Theory, 1(1).

[64] Vervaat, W. (1979) *A Relation between Brownian Bridge and Brownian Excursion*. Source: Ann. Probab. Volume 7, Number 1, 143-149.

[65] Voiculescu, D. (1998) *strengthened asymptotic freeness result for random matrices with applications to free entropy.* Internat. Math. Res. Notices , no. 1, 4163, 46L54 (81S25 82B10)

[66] Voiculescu, D. V.; Dykema, K. J.; Nica, A. (1992) *Free random variables. A noncommutative probability approach to free products with applications to random matrices, operator algebras and harmonic analysis on free groups.* CRM Monograph Series, 1. American Mathematical Society, Providence, RI.

[67] von Neumann, J.; Goldstine, H.H. (1947) *Numerical inverting of matrices of high order.* Bull. Amer. Math. Soc. 53 (11): 10211099.

[68] Wigner, E. (1955) *Characteristic vectors of bordered matrices with infinite dimensions.* Ann. Of Math. 62 (3): 548564.

[69] Wishart, J. (1928) *Generalized product moment distribution in samples.* Biometrika 20A (12): 3252.