Distribution of points on spherical objects and applications

By

Alessandro Selvitella

A Thesis
Submitted to the School of Graduate Studies
in Partial Fulfilment of the Requirements
for the Degree
Master of Science

McMaster University

MASTER OF SCIENCE (2013) McMaster University
(Statistics) Hamilton, Ontario

TITLE: Distribution of points on spherical objects and applications
AUTHOR: Alessandro Selvitella
SUPERVISOR: Professor Narayanaswamy Balakrishnan
NUMBER OF PAGES: 43

Abstract

In this thesis, we discuss some results on the distribution of points on the sphere, asymptotically when both the number of points and the dimension of the sphere tend to infinity. We then give some applications of these results to some statistical problems and especially to hypothesis testing.

The thesis is organized as follows. In Chapter 2, we discuss the properties of spherical and elliptical distributions. Here there are most of the new contributions of the thesis, especially the geometric characterization of elliptical distributions and the definition of *streched orthogonal* matrices are not known in the literature, as far as the author knows. Also, the generalizations from spherical to elliptical distributions of the theorems in this section are original. The subsequent chapters are mainly about spherical distributions and the results are recent but already known in the literature. In Chapter 3, we use the results of Chapter 2 to study the distribution of correlation coefficients. Chapter 4 forms the main part of this thesis and presents some results concerning the distribution of points on the sphere in random packing problems. Chapter 5 connects the discussed problems to the 7th Smale's problem, while in Chapter 6 we mention some open problems of possible future interest. Chapter 7 presents the Appendix, in which some technical tools, that are pertinent to the discussions in the preceding chapters, are described.

Acknowledgements

Foremost, I would like to thank my thesis advisor, Dr. Narayanaswamy Balakrishnan, whose enthusiasm and knowledge inspired and motivated me at every stage of this project; for his patient and expert guidance, I am truly grateful.

I would also like to thank the Department of Mathematics and Statistics at McMaster, in particular Dr. Maung Min-Oo, Dr. Viveros, Dr. Feng, Dr. Canty, Dr. Volterman whose courses and seminars provided a rich environment in which to learn and do statistics.

I thank all my friends who have made my permanence in Canada so special. In particular I say thank you to Chris Cappadocia who has always helped me in the good and bad moments which have been a lot.

Infine ringrazio la mia mamma e i mie nonni a cui voglio tantissimo bene e che mi sono sempre vicini anche se sono in Italia e non li posso vedere quanto vorrei!

Contents

Chapter 1

Introduction

This thesis deals with a topic in stochastic geometry and multivariate analysis, namely, the distribution of points on spherical objects. In particular, we discuss the distribution of points on the sphere and its asymptotic behaviour as the number of points n or the dimension p tend to $+\infty$.

Stochastic geometry is a mathematical discipline in which one studies the relations between geometry and probability theory. We refer to [11] for a detailed description of this field. Multivariate analysis deals with probability distributions of more than one dependent variable. We refer to [8] for a complete treatment of this subject.

The rest of this thesis is organized as follows. In Chapter 2, we discuss the properties of spherical and elliptical distributions. In Chapter 3, we use those properties to study the distribution of correlation coefficients. Chapter 4 forms the main part of this thesis and presents some results concerning the distribution of points on the sphere in random packing problems. Chapter 5 connects the discussed problems to the 7th Smale's problem, while Chapter 7 presents the Appendix, in which some technical tools, that are pertinent to the discussions in the preceding chapters, are described.

Chapter 2

Spherical and Elliptical Distributions

In this chapter we present a brief description of some of the main properties and main features of spherical and elliptical distributions.

2.1 Definitions

In multivariate analysis one of the most important distributions that plays a key role is the Multivariate Normal Distribution. For developing inference with more flexibility, one after considers more general models like spherical and elliptical distributions. In fact it is quite natural to consider a class of densities whose contours of constant density have the same elliptical shape as the Gaussian. The first extensions considered in the literature are *Spherical Distributions.*

Definition 2.1.1. *A* $m \times 1$ *random vector* $\mathbf{X}$ *is said to have a* spherical distribution *(see [8]) if* $\mathbf{X}$ *and* $\mathbf{HX}$ *have the same distribution for all* $m \times m$ *orthogonal matrices* $\mathbf{H}$ *(*$\mathbf{H^T} = \mathbf{H^{-1}}$*).*

Remark 2.1.2. *If* X *has a spherical distribution, then its pdf depends only on* $\mathbf{x}' \cdot \mathbf{x}$*.*

Example 2.1.3. *We now give some examples of Spherical Distributions:*

- *The* Multivariate Normal Distribution $\mathbf{N}_m(\mathbf{0}, \sigma^2 \mathbf{I}_m)$ *with pdf*

$$p(\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{\frac{m}{2}}} e^{\left(-\frac{1}{2\sigma^2}\mathbf{x}' \cdot \mathbf{x}\right)} \tag{2.1}$$

  for $\mathbf{x} \in \mathbf{R}^m$*;*

- *the* ε-Contaminated Normal distribution *(or* Two-component Mixture Normal Distribution*) with pdf*

$$p(\mathbf{x}) = (1-\varepsilon)\frac{1}{(2\pi)^{\frac{m}{2}}}e^{\left(-\frac{1}{2}\mathbf{x}'\cdot\mathbf{x}\right)} + \varepsilon\frac{1}{(2\pi\sigma^2)^{\frac{m}{2}}}e^{\left(-\frac{1}{2\sigma^2}\mathbf{x}'\cdot\mathbf{x}\right)} \tag{2.2}$$

for $\mathbf{x} \in \mathbf{R}^m$*;*

- *the* Multivariate t-distribution *with* n*-degrees of freedom with pdf*

$$p(\mathbf{x}) = \frac{\Gamma[\frac{1}{2}(n+m)]}{\Gamma[\frac{1}{2}n](n\pi)^{\frac{m}{2}}}\frac{1}{\left(1+\frac{1}{n}\mathbf{x}'\cdot\mathbf{x}\right)^{\frac{n+m}{2}}}, \tag{2.3}$$

for $\mathbf{x} \in \mathbf{R}^m$ *(called also* Multivariate Cauchy Distribution *when* $n = 1$*).*

Remark 2.1.4. *A natural way to generate spherical distributions is as follows. Let* $\mathbf{X}_1, \ldots, \mathbf{X}_m$*,* $\mathbf{Z} > 0$ *be random variables such that given* $\mathbf{Z}$*,* $\mathbf{X}_1, \ldots, \mathbf{X}_m$ *have independent* $\mathbf{N}_1(\mathbf{0}, \mathbf{Z})$ *distributions. If* $\mathbf{Z}$ *has cdf* G*, then the joint pdf of* $\mathbf{X}_1, \ldots, \mathbf{X}_m$ *is*

$$p(x_1, \ldots, x_m) = \int_0^{+\infty} \left(2\pi\mathbf{Z}\right)^{-\frac{m}{2}} e^{-\frac{1}{2\mathbf{Z}^T\cdot\mathbf{Z}}\Sigma_{i=1}^m x_i^2} dG(\mathbf{Z}),$$

which is spherical and it is indeed a scale mixture *of Gaussians. Th class of Spherical Distributions formed varying* G *is called* Compound Normal Distributions*. It follows that* $\mathbf{X} = \mathbf{Z}^{\frac{1}{2}}\mathbf{Y}$*, where* $\mathbf{Y}$ *is* $\mathbf{N}_m(\mathbf{0}, \mathbf{I}_m)$ *and* $\mathbf{Z}$ *and* $\mathbf{Y}$ *are independent, so that the values of* $\mathbf{X}$ *can be generated by generating values of independent* $N(0,1)$ *variables and multiplying them by values of an independent variable* $\mathbf{Z}$*. If* Z *takes values* 1 *with probability* $1-\varepsilon$ *and* σ^2 *with probability* ε*, then* $\mathbf{X}$ *has the* ε*-contaminated normal distribution in (2.2). Also, if* $\frac{n}{\mathbf{Z}}$ *is* χ^2_n*, then* $\mathbf{X}$ *has the* m*-variate t-distribution with* n*-degrees of freedom in (2.3).*

Now a further generalization are the so called *Elliptical Distributions.*

Definition 2.1.5. *An* $m \times 1$ *random vector* $\mathbf{X}$ *is said to have an* elliptical distribution *(see [8]) with parameters the* $m\times 1$*-vector* μ *and a symmetric positive definite* $m\times m$*-matrix* V *if its density function is of the form*

$$f(\mathbf{x}) = c_m(det(V))^{-\frac{1}{2}}h\Big((\mathbf{x}-\mu)^T V^{-1}(\mathbf{x}-\mu)\Big), \tag{2.4}$$

for some positive function h*. If* $\mathbf{X}$ *has an elliptical distribution we denote it by* $\mathbf{E}_m(\mu, \mathbf{V})$*.*

Example 2.1.6. *We now list some key properties of the family of elliptical distributions:*

- *Every spherical distribution is an elliptical distribution* $\mathbf{X} \sim \mathbf{E}_m(\mathbf{0}, \mathbf{Id}_m)$.
- *If* $\mathbf{Y}$ *has an* m*-variate spherical distribution and* $\mathbf{X} = \mathbf{C}\mathbf{Y} + \mu$, *where* $\mathbf{C}$ *is a non-singular* $m \times m$ *matrix, then* $\mathbf{X} \sim \mathbf{E}_m(\mu, \mathbf{V})$ *with* $V = \mathbf{C}\mathbf{C}^T$.
- *For some other examples, one may refer to [8].*

2.2 Parametrization of the Ellipsoid and Independence of Radius and Angle

An interesting property of a spherically distributed vector $\mathbf{X}$ is that a transformation to polar coordinates yields angles having the same distributions for all $\mathbf{X}$.

Theorem 2.2.1. *If* $\mathbf{X} \sim \mathbf{E}_m(\mathbf{0}, \mathbf{Id}_m)$ *with density function* $c_m h(x^T x)$ *and* $\mathbf{X} = R\omega$ *with* $R > 0$ *and* $\omega = f(\theta_1, \ldots, \theta_{m-1}) \in \mathbf{S}^{m-1}$ *then* $R, \theta_1, \ldots, \theta_{m-1}$ *are statistically independent with the distributions of* $\theta_1, \ldots, \theta_{m-1}$ *being the same for all* $\mathbf{X}$, $p(\theta_k) \propto \sin^{m-1-k}\theta_k$ *and* $R^2 = X^T \cdot X$ *has its density function as* $f_{R^2}(y) = \frac{c_m \pi^{\frac{m}{2}}}{\Gamma(\frac{m}{2})} y^{\frac{m}{2}-1} h(y)$ *with* $y > 0$*; see [8].*

Remark 2.2.2. *This theorem, with a suitable modification, will work for general Riemannian Manifolds (not necessarily spheres) with proper definitions of radius and angles.*

Proof. We now provide an outline on how it should work for elliptical distributions in dimension $m = 2$.
We will consider just ellipsoids $\mathbf{E}_{a,b}$ of the form $\frac{x^2}{a^2} + \frac{y^2}{b^2} = R^2$, with $R > 0$, since one can reduce to these by just rotating the coordinate axis. Now, the issue is how to properly parametrize this geometric object. Since the axis are stretched one with respect to the other, we are tempted to use a parametrization of the type $x = ab\cos(\theta),\ y = ab\sin(\theta)$. Since the axis should scale as R one can take $a = R\alpha$ and $b = R\beta$, where α and β are two constants independent of R. The problem here is that now it seems we have 3 parameters α, β and R for a 2-dimensional object. But, these three parameters are dependent, because

$$R^2 = \frac{x^2}{a^2} + \frac{y^2}{b^2} = R^2\beta^2\cos^2(\theta) + R^2\alpha^2\sin^2(\theta),$$

which implies

$$1 = \beta^2\cos^2(\theta) + \alpha^2\sin^2(\theta)$$

and so

$$\beta = \Big(\frac{1 - \alpha^2\sin^2(\theta)}{\cos^2(\theta)}\Big)^{\frac{1}{2}}.$$

So a natural parametrization of the ellipsoid for subsequent developments is:

$$\begin{cases} x = R^2\alpha\left(\frac{1-\alpha^2\sin^2(\theta)}{\cos^2(\theta)}\right)^{\frac{1}{2}}\cos(\theta) \\ y = R^2\alpha\left(\frac{1-\alpha^2\sin^2(\theta)}{\cos^2(\theta)}\right)^{\frac{1}{2}}\sin(\theta). \end{cases}$$

Here, $R > 0$ and $\theta \in [0, 2\pi]$. We can now compute the Jacobian of the transformation as

$$J_{R,\theta} := \begin{pmatrix} \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial R} \\ \frac{\partial y}{\partial \theta} & \frac{\partial y}{\partial \theta} \end{pmatrix}.$$

We then have $|det(J_{R,\theta})| = 2R^3\alpha^2\left(\frac{1-\alpha^2\sin^2(\theta)}{\cos^2(\theta)}\right)^{\frac{1}{2}}$. We can use to rewrite the expression of for Elliptical Distributions in (2.4) in Elliptical Coordinates. First $(\mathbf{x}-\mu)^T V^{-1}(\mathbf{x}-\mu) = \frac{x^2}{a^2} + \frac{y^2}{b^2} = R^2$, so that by the change of variable formula, we get

$$f(\mathbf{x}) = f(\theta, R) = \frac{c_m h(R^2)}{R^2\alpha\left(\frac{1-\alpha^2\sin^2(\theta)}{\cos^2(\theta)}\right)^{\frac{1}{2}}} 2R^3\alpha^2\left(\frac{1-\alpha^2\sin^2(\theta)}{\cos^2(\theta)}\right)^{\frac{1}{2}},$$

which simplifies to

$$f(\mathbf{x}) = f(\theta, R) = 2c_2 R\alpha h(R^2). \tag{2.5}$$

Then by Factorization Theorem, we readily obtain the independence of R and θ.

□

Remark 2.2.3. *Here, we want to make some key observations from the last theorem say. If $\mathbf{X}$ is spherically distributed, then $\mathbf{X}$ may be expressed as $\mathbf{X} = R(\mathbf{X})T(\mathbf{X})$, with $R(\mathbf{X}) = \sqrt{\mathbf{X}^T \cdot \mathbf{X}}$ and $T(\mathbf{X}) = \frac{\mathbf{X}}{R(\mathbf{X})}$ being independent random variables. From the form of the joint pdf, it is clear that $T(\mathbf{X})$ is uniformly distributed on S_m. We will show below that the uniform distribution on S_m is the unique distribution on S_m which is invariant under orthogonal transformation (see [8] and [6]). This decomposition suggests a general method of decomposing distributions with clear geometric properties.*
This can be utilized for hypothesis testing problems for example. The radius-angle decomposition suggests $t(m-1)$-type tests for the angles and a $\chi^2(1)$ test for the radius giving more importance to the radius-test. This gives m-independent tests to reject the null hypothesis. We believe this is especially useful in high dimensional analysis, because one can do a single $\chi^2(1)$ test before doing all the other tests instead of doing a $\chi^2(m)$ that can be computationally demanding when m is large. We will talk about applications in statistics later on.

2.3 Transformations which are invariant under an ellipsoid

Now, we have to find transformations that leave the ellipsoid invariant. Of course, they are not anymore *orthogonal* matrices, since they will not respect the asymmetry of the axis. It turns out that these transformations are *Stretched Orthogonal.*

Definition 2.3.1. *A $m \times m$ matrix $M_{\theta;a,b}$ is said to be $\frac{a}{b}$-*Stretched Orthogonal*, if it is of the form*

$$M_{\theta;a,b} := \begin{pmatrix} \cos(\theta) & \frac{a}{b}\sin(\theta) \\ -\frac{b}{a}\sin(\theta) & \cos(\theta) \end{pmatrix},$$

where $a, b \in \mathbf{R}$ and $\theta \in [0, 2\pi]$.

Remark 2.3.2. *These matrices are not orthogonal, but they are if $a = -b$ or $a = b$, namely, when the ellipsoid reduces to a sphere.*

Now we prove that these transformations leave the ellipsoid invariant.

Proposition 2.3.3. *Let $\mathbf{X} = (X_1, X_2)$ and $\mathbf{X}' = (X_1', X_2')$ be such that $\mathbf{X} = M_{\theta;a,b}\mathbf{X}'$. Then, $\frac{x^2}{a^2} + \frac{y^2}{b^2} = \frac{x'^2}{a^2} + \frac{y'^2}{b^2}$.*

Proof. It is straight forward to show this, but we report it here for the sake of completeness. By definition of $M_{\theta;a,b}$, we have $x = x' cos(\theta) + y' \frac{a}{b}\sin(\theta)$ and $y = -x'\frac{b}{a}\sin(\theta) + y'\cos(\theta)$. Now, substituting this in $\frac{x^2}{a^2} + \frac{y^2}{b^2}$, we obtain:

$$\begin{aligned} & \frac{x^2}{a^2} + \frac{y^2}{b^2} \\ = \quad & \frac{1}{a^2}\Big\{x' cos(\theta) + y'\frac{a}{b}\sin(\theta)\Big\}^2 + \frac{1}{b^2}\Big\{-x'\frac{b}{a}\sin(\theta) + y'\cos(\theta)\Big\}^2 \\ = \quad & \frac{1}{a^2}\Big\{x'^2 cos^2(\theta) + 2x'y'\frac{a}{b} cos(\theta)\sin(\theta) + y'\frac{a^2}{b^2}\sin^2(\theta)\Big\} \\ + \quad & \frac{1}{b^2}\Big\{x'^2\frac{b^2}{a^2}\sin(\theta) + y'^2\cos^2(\theta) - 2x'y'\frac{a}{b} cos(\theta)\sin(\theta)\Big\} \\ = \quad & \frac{1}{a^2}\Big\{x'^2 cos^2(\theta) + x'^2\sin^2(\theta)\Big\} + \frac{1}{b^2}\Big\{y'^2 cos^2(\theta) + y'^2\sin^2(\theta)\Big\} \\ = \quad & \frac{x'^2}{a^2} + \frac{y'^2}{b^2}. \end{aligned}$$

□

Now we define a norm which reflects the geometry of $E_{a,b}$ as follows

Definition 2.3.4. *The* $\frac{a}{b}$-Stretched Norm *is defined as follows:*

$$||X||_V := \mathbf{X}^T V^{-1} \mathbf{X}.$$

Theorem 2.3.5. *If* $\mathbf{X}$ *is an* Elliptical Distribution *with* $P(\mathbf{X} = \mathbf{0}) = 0$ *and*

$$r = ||\mathbf{X}||_V = (\mathbf{X}' \cdot \mathbf{X})^{\frac{1}{2}}, \quad \mathbf{T}(\mathbf{X}) = \frac{\mathbf{X}}{||\mathbf{X}||_V},$$

then $\mathbf{T}(\mathbf{X})$ *is uniformly distributed on* $E_{a,b}$*; moreover,* $\mathbf{T}(\mathbf{X})$ *and* r *are independent.*

Proof. We first prove that $\mathbf{X}$ and $\mathbf{Y} = M_{\theta;a,b}\mathbf{X}$ have the same distribution. In fact, since $\tilde{f}(y) = f(M_{\theta;a,b}y)det(M_{\theta;a,b})$, we have

$$\begin{aligned}&\tilde{f}(\mathbf{y}) = f(M_{\theta;a,b}\mathbf{y})det(M_{\theta;a,b})\\ &= c_m det(M)(det(V))^{-\frac{1}{2}} h\Big((\mathbf{y})^T M^T V^{-1} M(\mathbf{y})\Big) = c_m (det(V))^{-\frac{1}{2}} h\Big((\mathbf{y})^T V^{-1}(\mathbf{y})\Big) = f(\mathbf{y}).\end{aligned}$$

Now, for any $\frac{a}{b}$-*stretched orthogonal* matrix $M_{\theta;a,b}$, we have

$$T(M_{\theta;a,b}\mathbf{X}) = \frac{\mathbf{M}_{\theta;\mathbf{a},\mathbf{b}}\mathbf{X}}{||M_{\theta;a,b}\mathbf{X}||_V} = \frac{\mathbf{M}_{\theta;\mathbf{a},\mathbf{b}}\mathbf{X}}{||\mathbf{X}||_V} = M_{\theta;a,b}T(X),$$

so that $T(M_{\theta;a,b}\mathbf{X})$ and $M_{\theta;a,b}T(\mathbf{X})$ have the same distribution. We have established before that $\mathbf{X}$ and $T(\mathbf{X})$ have the same distribution, and so $T(M_{\theta;a,b}\mathbf{X})$ and $T(\mathbf{X})$ have the same distribution. Then, both $T(\mathbf{X})$ and $M_{\theta;a,b}T(\mathbf{X})$ have the same distribution. Since the uniform distribution is the unique distribution invariant under $\frac{a}{b}$-*stretched orthogonal* transformation on $E_{a,b}$ (see next section below), $T(\mathbf{X})$ is uniformly distributed on $E_{a,b}$. Hence, the theorem. $\square$

2.4 Uniqueness of the uniform distribution on S_m

Here we present a result by Kariya and Eaton (see [6]) which states that the uniform distribution on S_m is the unique distribution which is invariant under orthogonal transformations.

Theorem 2.4.1. *Suppose* $\mathbf{Z} \sim \mathbf{N}_m(\mathbf{0}, \mathbf{I}_m)$, $\mathbf{X} \sim \mathbf{E}_m(\mathbf{0}, \mathbf{Id}_m)$ *and* $\mathbf{U}$ *is the uniform distribution on the sphere. Then,* $\frac{\mathbf{Z}}{R(\mathbf{Z})} \sim \frac{\mathbf{X}}{R(\mathbf{X})} \sim \mathbf{U}$*, where* $R(\mathbf{X}) = \sqrt{\mathbf{X}^t \cdot \mathbf{X}}$.

Proof. Let $T(\mathbf{X}) = \frac{\mathbf{X}}{R(\mathbf{X})} \in S_m$. Then, $T(g\mathbf{X}) = gT(\mathbf{X})$ for every $g \in O(n)$ and so $T(g\mathbf{X}) \sim T(\mathbf{X}) \sim gT(\mathbf{X})$. We just need to prove the uniqueness of invariant probability measures on S_m. Suppose there exists a pdf $q(\mathbf{x})$ which is not the uniform distribution on S_m, then there exists a set A such that $P(A) > 0$ and where $q(\mathbf{x})$ pdf is not constant. But then there exists $g \in O(n)$ such that $P(gA) \neq P(A)$, which implies that $q(\mathbf{x})$ is not invariant for that particular g and this leads to a contradiction. hence, the result. $\square$

Remark 2.4.2. *A similar proof works for every manifold where a group of symmetries acts on it. For example a very similar argument leads to the uniqueness of the uniform distribution on the ellipsoid among the stretched orthogonal transformations.*

The result is used to generalize a well-known result for normal random variables.

Theorem 2.4.3. *Let $\mathbf{X} \sim \mathbf{E}_m(\mu, \mathbf{V})$, with $P(\mathbf{X} = \mathbf{0}) = 0$. If $\mathbf{W} = \frac{\alpha^T \cdot \mathbf{X}}{||\mathbf{X}||_V}$, where $\alpha \in \mathbf{R}^m$, $\alpha^T \cdot \alpha = 1$ then*

$$\mathbf{Y} = \frac{(m-1)^{\frac{1}{2}} \mathbf{W}}{(1 - \mathbf{W}^2)^{\frac{1}{2}}}$$

has the t_{m-1} distribution.

Proof. Since for every $\mathbf{X} \sim \mathbf{E}_m(\mu, \mathbf{V})$, $T(\mathbf{X})$ is uniformly distributed on $E_{a,b}$ and $\mathbf{W}$ is a function of $\mathbf{X}$ just through $T(\mathbf{X})$, we can assume without loss of generality that $\mathbf{X} \sim \mathbf{N}_m(\mathbf{0}, \mathbf{I}_m)$ and take $\alpha^T = (1, 0, \ldots, 0)$. Then $\mathbf{Y}$ is the ratio between a standard normal and a $\chi^2(m-1)$ (see for example [5]), and hence it has the t_{m-1} distribution. $\square$

Chapter 3

The Distribution of Correlation Coefficients

In this section we will discuss correlation coefficients and their distributions. We will see their relation with spherical and elliptical distribution, but we shall deal only with spherical distributions and the treatment concerning elliptical distribution is left for future work. Here, we follow closely the presentation of [3].

3.1 Definitions

If the $(m \times 1)$-random vector $\mathbf{X}$ has covariance matrix $\Sigma = (\sigma_{i,j})_{i,j}$, the correlation coefficients between two components of $\mathbf{X}$, say X_i and X_j, is defined by

$$\rho_{ij} = \frac{\sigma_{i,j}}{\sqrt{\sigma_{ii}\sigma_{jj}}} = \frac{Cov(X_i, X_j)}{\sqrt{Var(X_i)Var(X_j)}}.$$

By Cauchy-Schwartz inequality, we have $|\rho_{ij}| \leq 1$ and $\rho_{ij} = +1, -1$ if and only if X_i and X_j are linearly related. Correlation coefficients are hence a measure of linear dependence between X_i and X_j.

Now let $\mathbf{X}_1, \ldots, \mathbf{X}_N$ be N independent observations of $\mathbf{X}$, and let

$$S = \frac{1}{N-1}\Sigma_{i=1}^{N}(\mathbf{X_i} - \bar{\mathbf{X}}) \cdot (\mathbf{X_j} - \bar{\mathbf{X}})^T.$$

Then, the sample correlation coefficient between $\mathbf{X}_i$ and $\mathbf{X}_j$ is

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}}.$$

Remark 3.1.1. *If the random sample is from a multivariate normal distribution with all parameters unknown, then* r_{ij} *is the maximum likelihood estimate of* ρ_{ij}*.*

Remark 3.1.2. *It is well known that in this case* $\rho_{ij} = 0$ *if and only if* X_i *and* X_j *are independent. This, in fact, characterizes the multivariate normal distribution.*

Let $\mathbf{X} = (\mathbf{X}_1, \cdots, \mathbf{X}_p) = (x_{ij})_{n\times m}$ be an $n \times m$ random matrix. We shall assume:

Assumption (A): the columns $\mathbf{X}_1, \cdots, \mathbf{X}_p$ are independent n-dimensional random vectors with a common spherical distribution (which may depend on n) and $P(\mathbf{X}_1 = \mathbf{0}) = 0$. The condition $P(\mathbf{X}_1 = \mathbf{0}) = 0$ is to ensure that the correlation coefficients are well defined. Let r_{ij} be the Pearson correlation coefficient of $\mathbf{X}_i$ and $\mathbf{X}_j$ for $1 \le i < j \le p$. Then, $R_n := (r_{ij})_{p\times p}$ is the correlation matrix of $\mathbf{X}$, and L_n is *the coherence of the random matrix* defined as follows:

Definition 3.1.3. *The coherence of the random matrix* $\mathbf{X}$ *is defined as*

$$L_n := \max_{1\le i<j\le p} |r_{ij}|, \tag{3.1}$$

namely, it is the largest magnitude of the off-diagonal entries of the sample correlation matrix R_n*.*

3.2 Properties

In the following we discuss the exact distribution of the sample correlation coefficients under not too strong assumptions. We look now at a single sample correlation coefficient. Let us consider N pairs of variables $(X_1, Y_1), \ldots, (X_N, Y_N)$ and form the sample correlation coefficient

$$r = \frac{\Sigma_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\left[\Sigma_{i=1}^N (X_i - \bar{X})^2 \Sigma_{i=1}^N (Y_i - \bar{Y})^2\right]^{\frac{1}{2}}},$$

where $\bar{X} = \frac{1}{N}\Sigma_{i=1}^N X_i$ and $\bar{Y} = \frac{1}{N}\Sigma_{i=1}^N Y_i$. If we assume that the $\mathbf{X}$'s are independent of the $\mathbf{Y}$'s, the normality assumption is not important as long as one set of these variables has a spherical distribution. Let us set $\mathbf{1} = (1, \ldots, 1)^T$ and $\{1\} = \{k\mathbf{1}, k \in \mathbf{R}\}$.

Theorem 3.2.1 (Kariya-Eaton, 1977)**.** *Let* $\mathbf{X} = (X_1, \ldots, X_N)^T$ *and* $\mathbf{Y} = (Y_1, \ldots, Y_N)^T$*, with* $N > 2$*, be two independent random vectors, where* X *has a* m*-variate spherical* *distribution with* $P(\mathbf{X}=\mathbf{0})=0$ *and* $\mathbf{Y}$ *has any distribution with* $P(\mathbf{Y}\in\mathbf{1})=0$. *If*

$$r=\frac{\Sigma_{i=1}^{N}(X_i-\bar{X})(Y_i-\bar{Y})}{\left[\Sigma_{i=1}^{N}(X_i-\bar{X})^2\Sigma_{i=1}^{N}(Y_i-\bar{Y})^2\right]^{\frac{1}{2}}},$$

then

$$T:=(N-2)^{\frac{1}{2}}\frac{r}{(1-r^2)^{\frac{1}{2}}}$$

has the t_{N-2} *distribution*

Proof. See [8] for the details. □

Remark 3.2.2. *We believe a similar theorem works if one requires* $\mathbf{X}$ *to be just elliptical.*

Remark 3.2.3. *The correlation coefficient* r *can be seen as the cosine of the angle between the two vectors* $\mathbf{X}$ *and* $\mathbf{Y}$.

Remark 3.2.4. *The case when the vectors are normally distributed is a special case of the previous theorem.*

Corollary 3.2.5. *In the hypotheses of Theorem 3.2.1, the pdf of* r *is*

$$p(r)=\frac{\Gamma(\frac{N-1}{2})(1-r^2)^{\frac{N-4}{2}}}{\pi^{\frac{1}{2}}\Gamma(\frac{N-2}{2})},\quad -1<r<1.$$

Equivalently, r^2 *has the beta distribution with parameters* $\frac{1}{2}$ *and* $\frac{N-2}{2}$, *namely,* $r^2\sim B(\frac{1}{2},\frac{N-2}{2})$.

Proof. The required result follows by the fact that $p(r)=(N-2)^{\frac{1}{2}}\frac{r}{(1-r^2)^{\frac{1}{2}}}$ has the t_{N-2} distribution and the change of variable formula.

□

Remark 3.2.6. *Although* $\{r_{ij};\ 1\le i<j\le p\}$ *are pairwise independent, they are not mutually independent. In fact, recalling* $R=R_N=(r_{ij})_{p\times p}$, *the probability density function of* R *is given by*

$$h(R)=B_{n,p}\cdot(\det(R))^{(N-p-2)/2}\qquad(|r_{ij}|<1,\ i<j)\tag{3.2}$$

for $1\le p<N$, *where* $B_{N,p}:=\frac{\Gamma(\frac{N}{2})^p}{\Gamma_p(\frac{N}{2})}$ *is an (explicit) normalizing constant (see page 148 from [8]). Obviously,* $h(R)$ *is not a product of functions of individual* r_{ij}*'s, the entries of* R, *and so* $\{r_{ij};\ 1\le i<j\le p\}$ *are not mutually independent.*

The distribution of the correlation coefficients r_{ij} in the small sample cases are as follows.

Corollary 3.2.7. *Under Assumption (A), the following holds for all* $1 \le i < j \le p$*:*

(i) *When* $N = 2$, r_{ij} *has the symmetric Bernoulli distribution, i.e.,* $P(r_{ij} = \mathbf{1}) = 1/2$*;*

(ii) *When* $N = 3$, r_{ij} *has the density* $f(r) = \frac{1}{\pi}\frac{1}{\sqrt{1-r^2}}$ *on* $(-1, 1)$*;*
that is, r_{ij}^2 *follows the arcsine law on* $[0, 1]$*;*

(iii) *When* $N = 4$, r_{ij} *follows the uniform distribution on* $[-1, 1]$*;*

(iv) *When* $N = 5$, r_{ij} *has the density* $f(r) = \frac{2}{\pi}\sqrt{1 - r^2}$ *for* $|r| \le 1$*; that is,* r_{ij} *follows the semi-circle law.*

3.3 Limiting laws of coherence

In this section discuss limiting distributions of some random variables, in particular the Pearson correlation coefficients r_{ij}. Historically, the most used tools to prove asymptotic results in probability have been the Chen-Stein Method, the Large Deviation Principle and Strong approximations. We describe these methods in Chapter 7.
In [3], a different approach is developed to derive the limiting distributions of L_n.
Assuming the $\mathbf{X}_i$'s have the spherical distribution, the authors find an interesting and useful property of the correlation coefficients $\{r_{ij};\, 1 \le i < j \le p\}$.

Remark 3.3.1. *A first very simple asymptotic result can be given as*

$$f_N(w) = \frac{1}{\sqrt{N}} \cdot \frac{1}{\sqrt{\pi}} \frac{\Gamma(\frac{N-1}{2})}{\Gamma(\frac{N-2}{2})} \cdot \left(1 - \frac{w^2}{N}\right)^{\frac{N-4}{2}} \to \frac{1}{\sqrt{2\pi}} e^{-w^2/2} \tag{3.3}$$

as $N \to \infty$ *for all* $w \in \mathbb{R}$*. This shows that* W_N *converges to* $N(0, 1)$ *in distribution as* $N \to \infty$*. Set* $(x_{ij})_{N\times p} := (\mathbf{X}_1, \cdots, \mathbf{X}_p)$*. It is basically an application of the central limit theorem for i.i.d. random variables and the Slutsky theorem. However this does not give any information about the speed of convergence.*

Motivated by the applications in statistics and signal processing, we are especially interested in the ultra high dimensional case. More specifically, we consider three different regimes:

(i) *the sub-exponential case:* $\frac{1}{N}\log p \to 0$;

(ii) *the exponential case:* $\frac{1}{N}\log p \to \beta \in (0, \infty)$;

(iii) *the super-exponential case:* $\frac{1}{N}\log p \to \infty$.

Now, we shall state the main results of this section. The results presented below show that the limiting behaviour of L_N differs significantly in these different regimes and exhibits interesting phase transition phenomena.

Theorem 3.3.2 (Sub-Exponential Case)**.** *Suppose $p = p_N$ satisfies $\frac{\log p}{N} \to 0$ as $N \to \infty$. Then under Assumption (A), we have:*

(i) $L_N \to 0$ *in probability as* $N \to \infty$;

(ii) *Let* $T_N = \log(1 - L_N^2)$. *Then, as* $N \to \infty$,

$$NT_N + 4\log p - \log\log p \tag{3.4}$$

converges weakly to an extreme distribution with the distribution function $F(y) = 1 - e^{-Ke^{y/2}}$, $y \in \mathbb{R}$ *and* $K = 1/\sqrt{8\pi}$;

- *(law of large numbers)*

$$\sqrt{\frac{N}{\log p}}L_N \to 2 \tag{3.5}$$

in probability as $N \to \infty$.

We now consider the exponential case.

Theorem 3.3.3 (Exponential Case)**.** *Suppose $p = p_N$ satisfies $\frac{\log p}{N} \to \beta \in (0, \infty)$ as $N \to \infty$. Then under Assumption (A), we have:*

(i) $L_N \to \sqrt{1 - e^{-4\beta}}$ *in probability as* $N \to \infty$;

(ii) *Let* $T_N = \log(1 - L_N^2)$. *Then, as* $N \to \infty$,

$$NT_N + 4\log p - \log\log p \tag{3.6}$$

converges weakly to the distribution function

$$F(y) = 1 - \exp\left\{-K(\beta)e^{\frac{y+8\beta}{2}}\right\},\ y \in \mathbf{R},\ \textit{where } K(\beta) = \left(\frac{\beta}{2\pi(1 - e^{-4\beta})}\right)^{1/2}. \tag{3.7}$$

Finally, we turn to the super-exponential case where $(\log p)/N \to \infty$.

Theorem 3.3.4 (Super-Exponential Case)**.** *Suppose $p = p_N$ satisfies $\frac{\log p}{N} \to \infty$ as $N \to \infty$. Let $T_N = \log(1 - L_N^2)$. Then under Assumption (A), we have:*

(i) *$L_N \to 1$ in probability as $N \to \infty$. Further, $\frac{N}{\log p} T_N \to -4$ in probability as $N \to \infty$.*

(ii) *As $N \to \infty$,*

$$NT_N + \frac{4N}{N-2} \log p - \log N \tag{3.8}$$

converges weakly to the distribution function $F(y) = 1 - e^{-Ke^{y/2}}$, $y \in \mathbb{R}$ with $K = \frac{1}{\sqrt{2\pi}}$.

Remark 3.3.5. *Knowing the distribution of the correlation coefficient is a key part of the proof of the limiting distribution results. The starting step in the proofs of the theorems presented above is the Chen-Stein method (see Chapter 7) which requires the evaluation of two quantities: $P(r_{ij} \geq C)$ and $P(r_{ij} \geq C, r_{kl} \geq C)$. By using the explicit density of the correlation coefficients, one is able to evaluate the first probability precisely. The pairwise independence gives $P(r_{ij} \geq C, r_{kl} \geq C) = P(r_{ij} \geq C)^2$ for $\{i,j\} \neq \{k,l\}$. In other words, the evaluation of the second quantity is reduced to the study of the first one. This greatly simplifies some of the technical arguments. Moreover, with the understanding of the pairwise independence among $\{r_{ij}; 1 \leq i < j \leq p\}$ and the exact distribution of r_{ij} the authors of [3] have been able to get the limiting distribution of L_N for the full range of values of p and to fully characterize the phase transition phenomena in the limiting behaviours of the coherence (Theorems 3.3.2, 3.3.3 and 3.3.4).*

3.4 Applications of the limiting laws of coherence

As mentioned before, the limiting laws of coherence have a wide range of applications. Here, we discuss briefly two applications, one in high-dimensional statistics and another in signal processing.

3.4.1 High-dimensional statistics

A typical hypothesis testing in high dimensional statistical inference is the following. Let $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ be a random sample from a p-variate spherical distribution with covariance matrix $\Sigma_{p\times p} = (\sigma_{ij})$. We wish to test the hypotheses that Σ is diagonal, i.e.,

$$H_0 : \sigma_{i,j} = 0 \text{ for all } |i-j| \geq 1 \text{ vs. } H_a : \sigma_{i,j} \neq 0 \text{ for some } |i-j| \geq 1. \tag{3.9}$$

In the Gaussian case, this is the same as testing for independence. The asymptotic distribution of L_N can be used to construct a convenient test statistic for testing the hypotheses in (3.9). For example, in the case $\log p = o(N^{1/2})$, an approximate level α test is to reject the null hypothesis H_0 whenever

$$L_N^2 \geq N^{-1}\Big(4\log p - \log\log p - \log(8\pi) - 2\log\log(1-\alpha)^{-1}\Big).$$

It follows directly from Theorem 3.3.2 that the size of this test goes to α as $N \to \infty$.

3.4.2 Signal processing

Compressed sensing is an active and fast growing field in signal processing. An important problem in compressed sensing is the construction of measurement matrices $\mathbf{X}_{N\times p}$ which enables the precise recovery of a sparse signal β from linear measurements $\mathbf{y} = \mathbf{X}\beta$ using an efficient recovery algorithm. Such a measurement matrix $\mathbf{X}$ is typically randomly generated because it is difficult to construct deterministically. The best known example is perhaps the $N \times p$ random matrix $\mathbf{X}$ whose entries $x_{i,j}$ are iid normal variables

$$x_{i,j} \overset{iid}{\sim} N(0, N^{-1}). \tag{3.10}$$

A commonly used condition is the mutual incoherence property (MIP) which requires the pairwise correlations among the column vectors of $\mathbf{X}$ to be small. Write $\mathbf{X} = (\mathbf{X}_1, \cdots, \mathbf{X}_p) = (x_{ij})_{N\times p}$ with x_{ij} satisfying (3.10) and let $\tilde{L}_N := \max_{1\leq i<j\leq p} |\tilde{r}_{ij}|$, where

$$\tilde{r}_{ij} := \frac{(\mathbf{X_i} - \mu_i)(\mathbf{X} - \mu_j)}{||\mathbf{X_i} - \mu_i||||\mathbf{X} - \mu_j||},$$

$1 \leq i, j \leq p$. It has been shown that the condition

$$(2k-1)\tilde{L}_N < 1 \tag{3.11}$$

ensures the exact recovery of k-sparse signal β in the noiseless case where $y = X\beta$ (see [3]), and stable recovery of sparse signal in the noisy case where

$$y = X\beta + z. \tag{3.12}$$

Here z is an error vector, not necessarily random (see [3] and the references therein). The limiting laws derived in [3] can be used to show how likely a random matrix satisfies the MIP condition in (3.11).

Chapter 4

Distributions of Angles on Spheres

The distribution of distances between two random points on a unit sphere or other geometric objects has a wide range of applications including geometric probability, physics, statistics, machine learning and many others, and it has been well studied in different settings. In general, the angles, areas and volumes associated with random points, random lines and random planes appear in the studies of stochastic geometry; (see Stoyan and Kendall [11] and Kendall and Molchanov [7]).

In this section, we consider the empirical law and extreme laws of pairwise angles among a large number of random unit vectors. More specifically, let $\mathbf{X}_1, \cdots, \mathbf{X}_N$ be random points independently chosen with the uniform distribution on $\mathbf{S}^{p-1}$, the unit sphere in $\mathbf{R}^p$. The N points $\mathbf{X}_1, \cdots, \mathbf{X}_N$ on the sphere naturally generate N unit vectors $\overrightarrow{\mathbf{OX}}_i$, for $i = 1, 2 \cdots, N$, where $\mathbf{O}$ is the origin. Let $0 \leq \Theta_{ij} \leq \pi$ denote the angle between $\overrightarrow{\mathbf{OX}}_i$ and $\overrightarrow{\mathbf{OX}}_j$ for all $1 \leq i < j \leq N$.

In the case of a fixed dimension p, the global behaviour of the angles Θ_{ij} is captured by its empirical distribution

$$\mu_N = \frac{1}{\binom{N}{2}} \sum_{1 \leq i < j \leq N} \delta_{\Theta_{ij}}, \ \ N \geq 2. \tag{4.1}$$

When both the number of points N and the dimension p grow, it is more appropriate to consider the normalized empirical distribution

$$\mu_{N,p} = \frac{1}{\binom{N}{2}} \sum_{1 \leq i < j \leq N} \delta_{\sqrt{p-2}(\frac{\pi}{2} - \Theta_{ij})}, \ \ N \geq 2, \ p \geq 3. \tag{4.2}$$

In many applications, it is of interest to consider the extreme angles $\Theta_{\min}$ and $\Theta_{\max}$ defined by

$$\Theta_{\min} = \min\{\Theta_{ij};\ 1 \le i < j \le N\}, \tag{4.3}$$
$$\Theta_{\max} = \max\{\Theta_{ij};\ 1 \le i < j \le N\}. \tag{4.4}$$

We discuss below both the empirical distribution of the angles Θ_{ij}, $1 \le i < j \le n$, and the distributions of the extreme angles $\Theta_{\min}$ and $\Theta_{\max}$ as the number of points $n \mapsto \infty$, while the dimension p is either fixed or growing with n.
Here, following the work of Cai, Fan and Jiang [2] we investigate the asymptotic behaviours of the random angles $\{\Theta_{ij}; 1 \le i < j \le N\}$. It is shown that, when the dimension p is fixed, and as $n \mapsto \infty$, the empirical distribution μ_N converges to a distribution with the density function given by

$$h(\theta) = \frac{1}{\sqrt{\pi}} \frac{\Gamma(\frac{p}{2})}{\Gamma(\frac{p-1}{2})} \cdot (\sin\theta)^{p-2},\ \theta \in [0, \pi].$$

On the other hand, when the dimension p grows with N, it is shown that the limiting normalized empirical distribution $\mu_{N,p}$ of the random angles Θ_{ij}, $1 \le i < j \le N$ is Gaussian. When the dimension is high, most of the angles are concentrated around $\frac{\pi}{2}$. The results provide a precise description of this concentration and thus give a rigorous theoretical justification to the belief that "all high-dimensional random vectors are almost always nearly orthogonal to each other".
In addition to the empirical law of the angles Θ_{ij}, we also consider the extreme laws of the random angles in both the fixed and growing dimension settings. The limiting distributions of the extremal statistics $\Theta_{\max}$ and $\Theta_{\min}$ are derived.

4.1 Fixed dimension p and $N \mapsto +\infty$

We begin with the limiting empirical distribution of the random angles.

Theorem 4.1.1 (Empirical Law for Fixed p). *Let the empirical distribution μ_N of the angles Θ_{ij}, $1 \le i < j \le N$, be defined as in* (4.1.1)*. Then, as $N \to \infty$, with probability one, μ_N converges weakly to the distribution with density*

$$h(\theta) = \frac{1}{\sqrt{\pi}} \frac{\Gamma(\frac{p}{2})}{\Gamma(\frac{p-1}{2})} \cdot (\sin\theta)^{p-2},\ \theta \in [0, \pi]. \tag{4.5}$$

In fact, $h(\theta)$ is the probability density function of Θ_{ij} for any $i \ne j$ (Θ_{ij}'s are identically distributed). Due to the dependency of Θ_{ij}'s, some of them are large and some are small. More details are presented in subsequent sections.

Remark 4.1.2. *Theorem 4.1.1 says that the average of these angles asymptotically has the same density as that of* Θ_{12}.

Remark 4.1.3. *Notice that when* $p = 2$, $h(\theta)$ *is the uniform density on* $[0, \pi]$, *and when* $p > 2$, $h(\theta)$ *is unimodal with mode* $\theta = \frac{\pi}{2}$. *Theorem 4.1.1 implies that most of the angles in the total of* $\binom{N}{2}$ *angles are concentrated around* $\frac{\pi}{2}$. *This concentration becomes stronger as the dimension* p *grows since* $(\sin\theta)^{p-2}$ *converges to zero more quickly for* $\theta \neq \frac{\pi}{2}$. *In fact, in the extreme case when* $p \to \infty$, *almost all of* $\binom{N}{2}$ *angles go to* $\frac{\pi}{2}$ *at the rate* $\sqrt{p}$. *This can be seen from Theorem 4.2.1 later.*

We now consider the limiting distribution of the extreme angles $\Theta_{\min}$ and $\Theta_{\max}$.

Theorem 4.1.4 (Extreme Law for Fixed p)**.** *Let* $\Theta_{\min}$ *and* $\Theta_{\max}$ *be as defined in (4.3) and (4.4) respectively. Then, both* $N^{2/(p-1)}\Theta_{\min}$ *and* $N^{2/(p-1)}(\pi - \Theta_{\max})$ *converge weakly to a distribution given by*

$$F(x) = \begin{cases} 1 - e^{-Kx^{p-1}}, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0, \end{cases} \tag{4.6}$$

as $N \to \infty$, *where*

$$K = \frac{1}{4\sqrt{\pi}} \frac{\Gamma(\frac{p}{2})}{\Gamma(\frac{p+1}{2})}. \tag{4.7}$$

The above theorem says that the smallest angle $\Theta_{\min}$ is close to zero, and the largest angle $\Theta_{\max}$ is close to π as N grows. This makes sense from Theorem 4.1.1 since the support of the density function $h(\theta)$ is $[0, \pi]$.

Remark 4.1.5. *In the special case of* $p = 2$, *the scaling of* $\Theta_{\min}$ *and* $\pi - \Theta_{\max}$ *in Theorem 4.1.4 is* N^2. *This in fact can also be seen in a similar problem. Let* $\xi_1, \cdots, \xi_N$ *be i.i.d.* $U[0,1]$*-distributed random variables with order statistics* $\xi_{(1)} \leq \cdots \leq \xi_{(N)}$. *Set* $W_n := \min_{1 \leq i \leq N-1}(\xi_{(i+1)} - \xi_{(i)})$, *which is the smallest spacing among the observations of* ξ_i*'s. Then, by using the representation theorem of* $\xi_{(i)}$*'s through i.i.d. random variables with exponential distribution* $\mathrm{Exp}(1)$ *(see, for example, Arnold et al. [1]), it is easy to check that* $N^2 W_N$ *converges weakly to* $\mathrm{Exp}(1)$ *with probability density function* $e^{-x} I(x \geq 0)$.

4.2 Both $p \mapsto +\infty$ and $N \mapsto +\infty$

It is helpful to see how the density changes with dimension p:

$$\begin{aligned} h_p(\theta) &= \frac{1}{\sqrt{p-2}} h\Big(\frac{\pi}{2} - \frac{\theta}{\sqrt{p-2}}\Big) \\ &= \frac{1}{\sqrt{\pi}} \frac{\Gamma(\frac{p}{2})}{\Gamma(\frac{p-1}{2})\sqrt{p-2}} \cdot \Big(\cos \frac{\theta}{\sqrt{p-2}}\Big)^{p-2}, \quad \theta \in [0, \pi] \end{aligned} \tag{4.8}$$

which is the asymptotic density of the normalized empirical distribution $\mu_{n,p}$ defined in (4.2); in fact, we have the asymptotic approximation

$$h_p(\theta) \propto \exp\Big((p-2)\log\big\{\cos\big(\frac{\theta}{\sqrt{p-2}}\big)\big\}\Big) \approx e^{-\theta^2/2}. \tag{4.9}$$

The following result shows that the empirical distribution of the random angles, after suitable normalization, converges to a standard normal distribution. This is clearly different from the limiting distribution given in Theorem 4.1.1 when the dimension p is fixed.

Theorem 4.2.1 (Empirical Law for Growing p)**.** *Let $\mu_{N,p}$ be as defined in* (4.2)*. Assume $\lim_{N\to\infty} p_N = \infty$. Then, with probability one, $\mu_{N,p}$ converges weakly to $N(0,1)$ as $N \to \infty$.*

Theorem 4.2.1 holds regardless of the speed of p relative to n when both go to infinity.

Remark 4.2.2. *The theorem implies that most of the $\binom{N}{2}$ random angles go to $\frac{\pi}{2}$ very quickly. Take any $\gamma_p \mapsto 0$ such that $\sqrt{p}\gamma_p \mapsto \infty$ and denote by $n_{N,p}$ the number of the angles Θ_{ij} that are within γ_p of $\frac{\pi}{2}$, i.e., $|\frac{\pi}{2} - \Theta_{ij}| \le \gamma_p$. Then $\frac{n_{N,p}}{\binom{N}{2}} \mapsto 1$. Hence, most of the random vectors in the high-dimensional Euclidean spaces are nearly orthogonal.*

An interesting question is:
"Given two such random vectors, how fast is their angle close to $\frac{\pi}{2}$ as the dimension increases? “
The following result answers this question.

Proposition 4.2.3. *Let $\mathbf{U}$ and $\mathbf{V}$ be two random points on the unit sphere in $\mathbf{R}^p$. Let Θ be the angle between $\overrightarrow{\mathbf{OU}}$ and $\overrightarrow{\mathbf{OV}}$. Then*

$$P(|\Theta - \frac{\pi}{2}| \ge \epsilon) \le K\sqrt{p}(\cos\epsilon)^{p-2} \tag{4.10}$$

for all $p \ge 2$ and $\epsilon \in (0, \frac{\pi}{2})$, where K is a universal constant.

Under the spherical invariance, we can think of Θ as a function of the random point $\mathbf{U}$ only.

Remark 4.2.4. *One can see that, as the dimension p grows, the probability decays exponentially. In particular, take $\epsilon = \sqrt{(c\log p)/p}$ for some constant $c > 1$. Note that $\cos\epsilon \le 1 - \frac{\epsilon^2}{2} + \frac{\epsilon^4}{24}$ and so*

$$P\left(|\Theta - \frac{\pi}{2}| \ge \sqrt{\frac{c\log p}{p}}\right) \le K\sqrt{p}\left(1 - \frac{c\log p}{2p} + \frac{c^2\log^2 p}{24p^2}\right)^{p-2} \le K' p^{-\frac{c-1}{2}} \tag{4.11}$$

for all sufficiently large p, where K' is a constant depending only on c. Hence, in the high dimensional space,

$$\Theta_{ij} \in [\pi/2 - \sqrt{(c\log p)/p}, \pi/2 + \sqrt{(c\log p)/p}$$

with high probability. This provides a precise characterization of the statement mentioned earlier that "all high-dimensional random vectors are almost always nearly orthogonal to each other".

We now turn to the limiting extreme laws of the angles when both N and $p \mapsto \infty$. For the extreme laws, it is necessary to divide into three asymptotic regimes (as earlier in Chapter 3): sub-exponential case $\frac{1}{p}\log N \to 0$, exponential case $\frac{1}{p}\log N \to \beta \in (0,\infty)$, and super-exponential case $\frac{1}{p}\log N \to \infty$. The limiting extreme laws are different in these three regimes.

Theorem 4.2.5 (Extreme Law: Sub-Exponential Case)**.** *Let $p = p_N \to \infty$ satisfy $\frac{\log N}{p} \to 0$ as $N \to \infty$. Then:*

(i) $\max_{1\le i<j\le N} |\Theta_{ij} - \frac{\pi}{2}| \to 0$ *in probability as $N \to \infty$;*

(ii) *As $N \to \infty$, $2p\log\sin\Theta_{\min} + 4\log N - \log\log N$ converges weakly to the extreme value distribution with distribution function $F(y) = 1 - e^{-Ke^{y/2}}$, $y \in \mathbb{R}$ and $K = 1/(4\sqrt{2\pi}\,)$. The conclusion still holds if $\Theta_{\min}$ is replaced by $\Theta_{\max}$.*

In this case, both $\Theta_{\min}$ and $\Theta_{\max}$ converge to $\pi/2$ in probability. The above extreme value distribution differs from that in (4.6) where the dimension p is fixed. This is obviously caused by the fact that p is finite in Theorem 4.1.4 and goes to infinity in Theorem 4.2.5.

Corollary 4.2.6. *Let $p = p_N$ satisfy $\lim_{N\to\infty} \frac{\log N}{\sqrt{p}} = \alpha \in [0,\infty)$. Then, $p\cos^2\Theta_{\min} - 4\log N + \log\log N$ converges weakly to a distribution with cumulative distribution function $\exp\{-\frac{1}{4\sqrt{2\pi}}e^{-(y+8\alpha^2)/2}\}$, $y \in \mathbf{R}$. The conclusion still holds if $\Theta_{\min}$ is replaced by $\Theta_{\max}$.*

Now, we turn our attention to the exponential case.

Theorem 4.2.7 (Extreme Law: Exponential Case)**.** *Let $p = p_N$ satisfy $\frac{\log N}{p} \to \beta \in (0, \infty)$ as $N \to \infty$. Then:*

(i) $\Theta_{\min} \to \cos^{-1}\sqrt{1 - e^{-4\beta}}$ *and* $\Theta_{\max} \to \pi - \cos^{-1}\sqrt{1 - e^{-4\beta}}$ *in probability as* $N \to \infty$;

(ii) *As* $N \to \infty$, $2p \log \sin \Theta_{\min} + 4 \log N - \log\log N$ *converges weakly to a distribution with distribution function*

$$F(y) = 1 - \exp\left\{-K(\beta)e^{(y+8\beta)/2}\right\}, \ y \in \mathbb{R}, \tag{4.12}$$

where $K(\beta) = \left(\frac{\beta}{8\pi(1-e^{-4\beta})}\right)^{1/2}$. *The conclusion still holds if* $\Theta_{\min}$ *is replaced by* $\Theta_{\max}$.

Remark 4.2.8. *In contrast to Theorem 4.2.5, neither* $\Theta_{\max}$ *nor* $\Theta_{\min}$ *converges to* $\frac{\pi}{2}$ *under the case* $(\log N)/p \to \beta \in (0, \infty)$. *Instead, they converge to different constants depending on* β.

Theorem 4.2.9 (Extreme Law: Super-Exponential Case)**.** *Let $p = p_N$ satisfy $\frac{\log N}{p} \to \infty$ as $N \to \infty$. Then:*

(i) $\Theta_{\min} \to 0$ *and* $\Theta_{\max} \to \pi$ *in probability as* $N \to \infty$;

(ii) *As* $N \to \infty$, $2p \log \sin \Theta_{\min} + \frac{4p}{p-1} \log N - \log p$ *converges weakly to the extreme value distribution with the distribution function* $F(y) = 1 - e^{-Ke^{y/2}}$, $y \in \mathbf{R}$, *with* $K = 1/(2\sqrt{2\pi})$. *The conclusion still holds if* $\Theta_{\min}$ *is replaced by* $\Theta_{\max}$.

Remark 4.2.10. *It can be seen from Theorems 4.2.5, 4.2.7 and 4.2.9 that* $\Theta_{\max}$ *becomes larger when the rate* $\beta = \lim(\log N)/p$ *increases. They are* $\pi/2$, $\pi - \cos^{-1}\sqrt{1 - e^{-4\beta}} \in (\pi/2, \pi)$ *and* π *when* $\beta = 0$, $\beta \in (0, \infty)$ *and* $\beta = \infty$, *respectively.*

Remark 4.2.11. *Set* $f(\beta) = \pi - \cos^{-1}\sqrt{1 - e^{-4\beta}}$. *Then,* $f(0) = \pi/2$ *and* $f(+\infty) = \pi$, *which corresponds to* $\Theta_{\max}$ *in (i) of Theorem 4.2.5 and (i) of Theorem 4.2.9, respectively. So the conclusions in Theorems 4.2.5, 4.2.7 and 4.2.9 are all consistent.*

Remark 4.2.12. *As discussed in Chapter 4 and in [3], Cai and Jiang considered the limiting distribution of the coherence of a random matrix and the coherence is closely related to the minimum angle* $\Theta_{\min}$. *In the current setting, the coherence* $L_{N,p}$ *is defined by*

$$L_{N,p} := \max_{1 \le i < j \le N} |r_{ij}|, \tag{4.13}$$

where r_{ij} are as defined in Chapter 3. The proofs of the results in Theorems 4.2.5, 4.2.7 and 4.2.9 can be essentially reduced to the analysis of $L_{N,p} := \max_{1 \leq i < j \leq N} r_{ij}$. This maximum is analysed by modifying the proofs of the results for the limiting distribution of the coherence $L_{N,p}$ in Cai and Jiang (2012). The key step in the proofs is the study of the maximum and minimum of pairwise i.i.d. random variables $\{r_{ij};\ 1 \leq i < j \leq N\}$ by using the Chen-Stein method. It is noted that $\{r_{ij};\ 1 \leq i < j \leq N\}$ are not i.i.d. random variables (see Chapter 3), and so the standard techniques of analysing the extreme values (see the Chapter 7) of $\{r_{ij};\ 1 \leq i < j \leq N\}$ do not apply.

Chapter 5

Connection to the 7th Smale's Problem

The results on random angles presented in the preceding chapters may be used to study a number of open deterministic problems in mathematics and physics (see [2]). We will discuss here in particular about their relation to the seventh Smale's problem. This part is mainly based on a paper by Steven Smale [10].
Smale's problems are a list of eighteen unsolved problems in mathematics that was proposed by Steve Smale in 1998 republished in 1999. Smale composed this list in reply to a request from Vladimir Arnold, then president of the International Mathematical Union, who asked several mathematicians to propose a list of problems for the twentyfirst century. Arnold's inspiration came from the list of Hilbert's problems that had been published at the beginning of the twentieth century. The list of problems is the following:

1. The Riemann Hypothesis
2. The Poincare' Conjecture
3. Does $P = NP$?
4. Integer zeros of a polynomial of one variable
5. Height bounds for diophantine curves
6. Finiteness of the number of relative equilibria in celestial mechanics
7. Distribution of points on the 2-sphere
8. Introduction of dynamics into economic theory

9. The linear programming problem
10. The closing Lemma
11. Is one-dimensional dynamics generally hyperbolic?
12. Centralizers of diffeomorphisms
13. Hilbert's 16th Problem
14. Lorentz Attractor
15. Navier-Stokes Equations
16. The Jacobian Conjecture
17. Solving polynomial equations
18. Limits of intelligence

We concentrate here just on the seventh one and demonstrate that it is closely connected with the work presented in the preceding chapters.

5.1 Distribution of points on the 2-sphere

Let $V_N(x) = \Sigma_{1\leq i<j\leq N} \log \frac{1}{||x_i - x_j||}$, where $(x_1, \ldots, x_N)$, are distinct points on the 2-sphere $\mathbf{S}^2 \subset \mathbf{R}^3$, and $||x_i - x_j||$ is the distance in $\mathbf{R}^3$. Denote by $V_N := \min_x V_N(x)$. The problem is

$$\text{Can one find } \ (x_1, \ldots, x_n) \ \text{ such that } V_N(x) - V_N \leq c \log N,$$
$$\text{where } c \text{ is a universal constant?}$$

This problems comes from complexity theory and it is related to the question of finding a good starting polynomial for a homotopy algorithm for realizing the Fundamental Theorem of Algebra. A $(x_1, \ldots, x_N)$ is called an N-tuple of elliptic Fekete points. The function V_N as a function of N satisfies

$$V_N = -\frac{1}{4} \log \frac{4}{e} N^2 - \frac{N}{4} \log N + O(N).$$

It is also natural to consider potential of the form

$$V_N(x, s) = \Sigma_{1\leq i<j\leq N} \log \frac{1}{||x_i - x_j||^s},$$

with $V_N(s) := \min_x V_N(x, s)$, x as before and $0 < s < 2$. For $s = 1$, we recover the Coulomb potential and $V(1)$ corresponds to an equilibrium position of N electrons constrained to lie on the 2-sphere. One may also consider higher dimensional spheres $\mathbf{S}^{p-1}$. The difficulty of the problem is represented by the high number of symmetries and of saddle points that the function $V_N(s, x)$ may have.

Chapter 6

Open Problems and Future Directions

- What about Areas or Volumes?
  One can study the asymptotic distribution of Areas and Volumes of triangles on the sphere as the dimension and/or the number of points go to $+\infty$.

- What happens for a General Manifold, Riemannian or not?
  We can try to prove the asymptotic results in [3] and [2] for general manifolds, not necessarily spheres, but for example for ellipsoids or tori or other manifolds.

- Does the domain of attraction change with curvature?
  We saw that asymptotically the distribution of a point tends to one of the three extreme distributions. Is the type of distribution related to the sign of the curvature of the manifold?

- Does the Curvature play a role in the Chebishev Inequality?
  Is it possible that the curvature of the manifold plays a role in the rate of convergence of the asymptotic results?

Chapter 7

Appendix: Technical Tools

Sophisticated approximation methods such as the Chen-Stein method, large deviation bounds and strong approximations are the key ingredients in the proofs of the main results in the literature for the cases when the dimension p is not too high. Even if it is not clear if these same tools can be used to derive the limiting distributions of the coherence L_n for the three regimes considered in Chapters 4 and 5, we think we should briefly mention them for the sake of completeness.

7.1 The Chen-Stein Method

The Chen-Stein Method was first introduced by Stein in [9]. It is a very well developed and useful tool for proving convergence in distribution to special distributions like Gaussian or Poisson. Recent applications involve Random Matrix Theory (see for example [12]). The importance of this method is due to the fact that the Central Limit Theorem does not give any indication of the rate of convergence to the Gaussian distribution while Chen-Stein Method does.
The pdf of the standard normal $\rho(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ can be viewed as a solution to the ODE

$$\rho'(x) + x\rho(x) = 0,$$

which in weak form becomes

$$\int_{\mathbf{R}} \rho(x)\Big(f'(x) - xf(x)\Big) = 0$$

for any test function f. Equivalently, one can say that if $X \sim N(0,1)$, then

$$\mathbf{E}\Big[f'(X) - Xf(X)\Big] = 0.$$

It turns out that the converse is also true.

Theorem 7.1.1. *A random variable X is distributed as a standard normal $X \sim N(0,1)$ if and only if it satisfies $\mathbf{E}[f'(X) - Xf(X)] = 0$, for all continuous and piecewise differentiable $f : \mathbf{R} \to \mathbf{R}$ with $\mathbf{E}[|f'(Z)|] < +\infty$ for $Z \sim N(0,1)$.*

Proof. Suppose $X \sim N(0,1)$, then

$$\begin{aligned}
\mathbf{E}\Big[f'(X) - Xf(X)\Big] &= \frac{1}{\sqrt{2\pi}}\int_{\mathbf{R}} f'(x)e^{-\frac{x^2}{2}} - \frac{1}{\sqrt{2\pi}}\int_{\mathbf{R}} xf(x)e^{-\frac{x^2}{2}} = \\
&= \frac{1}{\sqrt{2\pi}}\int_{\mathbf{R}} f'(x)e^{-\frac{x^2}{2}} - \frac{1}{\sqrt{2\pi}}\int_{\mathbf{R}} f'(x)e^{-\frac{x^2}{2}} + \Big[\frac{1}{\sqrt{2\pi}}f(x)e^{-\frac{x^2}{2}}\Big]_{-\infty}^{+\infty} \\
&= 0.
\end{aligned}$$

Conversely, suppose $\mathbf{E}[f'(X) - Xf(X)] = 0$ for all continuous and piecewise differentiable $f : \mathbf{R} \to \mathbf{R}$ with $\mathbf{E}[|f'(Z)|] < +\infty$ for $Z \sim N(0,1)$. Then this equation holds in particular for

$$f_t(x) = e^{\frac{x^2}{2}} \int_{-\infty}^{x} \Big(\mathbf{I}_{\{y \le t\}} - \Phi(t)\Big) e^{\frac{-y^2}{2}} dy,$$

where $\Phi(t)$ is the cdf of a standard normal and t is fixed. Then

$$f_t'(X) = Xf_t(X) + \mathbf{I}_{\{y \le t\}} - \Phi(t)$$

and so, for all $t \in \mathbf{R}$ we get

$$\mathbf{E}\Big[f_t'(X) - Xf_t(X)\Big] = \mathbf{P}(X \le t) - \Phi(t) = 0,$$

which implies that $X \sim N(0,1)$.

□

Remark 7.1.2. *The relation between ODEs and convergence of random variables makes it possible to use in this context methods from functional analysis and differential equations such as the Lyapunov-Schmidt Decomposition.*

In fact, more is true, as seen below.

Theorem 7.1.3 (Stein's Continuity Theorem)**.** *Let X_n be a sequence of real random variables with uniformly bounded second moment and let $G \equiv N(0,1)$. Then, the following are equivalent:*

- $\mathbf{E}\Big[f'(X_N) - X_N f(X_N)\Big] \to 0$ *whenever $f : \mathbf{R} \to \mathbf{R}$ is continuous and piecewise differentiable $f : \mathbf{R} \to \mathbf{R}$ with $\mathbf{E}[|f'(Z)|] < +\infty$ for $Z \sim N(0,1)$;*

- *X_N converges in distribution to G.*

The above theorem gives only a qualitative result, but its proof is very quantitative (see [12]). In fact, using this type of ideas, one can prove a series of theorems which are basically Central Limit-type Theorems along with a rate. We state here the Berry-Esséen Theorem in the case of independent and identically distributed random variables (see [9]).

Theorem 7.1.4. *Suppose X_i, $i = 1, \ldots, N$, are iid random variables with $\mathbf{E}[X_i] = 0$, $\mathbf{E}[X_i^2] = 1$ and $\mathbf{E}[|X_i|^3] < \infty$, for $i = 1, \ldots, N$. Then, for $N \in \mathbf{N}$ and all $x \in \mathbf{R}$, we have*

$$\Big|\mathbf{P}\Big(\frac{1}{\sqrt{N}}\Sigma_{i=1}^N X_i \leq t\Big) - \Phi(t)\Big| \leq \frac{9\mathbf{E}[|X_i|^3]}{\sqrt{N}}.$$

Remark 7.1.5. *Another well known use of Chen-Stein Method is for Poisson Approximation. The general procedure is similar to the one for the normal distribution. We also want to emphasise that with this method one can relax the hypothesis of independence.*

7.2 Large deviations

The theory of large deviations deals with probabilities of rare events that become exponentially small as a function of some parameter. This theory has applications in many different fields, mainly in statistics. Here, we present a brief overview of this theory mainly through the example of the sum of independent and identically distributed Gaussian random variables (see [14] for more details).

7.2.1 Sum of iid Gaussian Random Variables

The pdf of a Gaussian random variable X is of the form

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbf{R},$$

where $\mu = \mathbf{E}[X]$ is the mean of X and $\sigma^2 = \mathbf{E}[(X-\mu)^2]$ is its variance. We are interested here in the distribution of the sum of N iid Gaussians:

$$S_N = \Sigma_{i=1}^N X_i,$$

which takes the formula

$$p_{S_N}(x) = \int_{\mathbf{R}} dx_1 \ldots \int_{\mathbf{R}} dx_N \delta\Big(\Sigma_{i=1}^N - Ns\Big) p(x_1, \ldots, x_N),$$

where $p(x_1, \ldots, x_N)$ is the joint pdf of the X_i's. One can easily get p_{S_N} from this formula using the method of moment generating functions. What we want to underline here is that the pdf will have this general form:

$$p_{S_N}(s) \approx e^{-NI(s)}$$

with

$$I(s) = \frac{(x-\mu)^2}{2\sigma^2}.$$

This form is common to several other cases such as the exponential and binomial, but each distribution will have its own form of $I(s)$.

Remark 7.2.1. *We will be more precise in the meaning of the approximation sign $\approx$, but it actually means that the dominant behaviour of $p(S_N)$ is exponentially decaying in N, with all other terms being subexponential in N. This type of integral expressions play an important role in a lot of different fields in mathematics like stationary phase (saddle point method), Fourier Integrals, PDEs and dynamical systems, just to cite a few.*

Hence from $p_{S_N}(s) \approx e^{-NI(s)}$ we deduce that $p(S_n)$ decays to 0 exponentially fast with N whenever $I(s)$ is positive. But $I(s) \geq 0$ and $I(s) = 0$ if and only if $s = \mu = \mathbf{E}[X_i]$. Therefore, $p_{S_N} \to \delta(s-\mu)$ in this limit.

7.2.2 The Large Deviation Principle

The general exponential form $p_{S_N}(s) \approx e^{-NI(s)}$, is the founding result of large deviation theory and it arises in several contexts and for many stochastic processes also and not just for iid sample. The rigorous theory needs many deep concepts of topology and measure theory that go beyond the scope of this thesis. So, we will present simpler but less formal definitions.

Definition 7.2.2. *We say that a random variable S_N or its pdf $p(S_N)$ satisfy a* large deviation principle *(LDP) if the following limit exists:*

$$\lim_{N\to+\infty} -\frac{1}{N} \ln p_{S_N}(s) = I(s).$$

$I(s)$ needs to be a function, not everywhere zero, and it is called the rate function.

Basically, a pdf $p(S_N)$ satisfies a *large deviation principle* (LDP) if it is of the form

$$p_{S_N}(s) = e^{-NI(s)+o(N)}$$

for large N. For this reason, often this theory is said to estimate probability at a logarithmic scale. The rigorous mathematical definition is due to Varadhan and we refer to [13] for a deeper explanation of this concept.
We now state two of the main theorems of this theory.

Theorem 7.2.3 (Gärtner-Ellis Theorem). *Define*

$$\lambda(k) := \lim_{N \to +\infty} \frac{1}{N} \ln \mathbf{E}[e^{NkS_N}].$$

If $\lambda(k)$ is differentiable in k, then

- *S_N satisfies an LDP, namely $\lim_{N\to+\infty} -\frac{1}{N} \ln p_{S_N}(s) = I(s)$;*
- *the rate function $I(s)$ is given by the* Legendre-Fenchel Transform *of $\lambda(k)$:*

$$I(s) = \sup_{k \in \mathbf{R}} \{ks - \lambda(k)\}.$$

Remark 7.2.4. *Typically, $\lambda(k)$ does not exist when a pdf $p(S_N)$ does not satisfy a LDP.*

Another very important result is *Varadhan's Theorem.* This theorem is concerned with the evaluation of a functional expectation of the form

$$W_N(f) = \mathbf{E}[e^{Nf(S_N)}] = \int_{\mathbf{R}} p_{S_N}(s) e^{nf(s)} ds.$$

If we assume that S_N satisfies a LDP with rate function $I(s)$, we can get

$$W_N(f) \approx e^{N \sup_N [f(s) - I(s)]},$$

which is called *Laplace approximation* and it is justified because corrections are subexponential in N.

Theorem 7.2.5 (Varadhan's Theorem). *If $\lambda[f] := \lim_{N\to+\infty} \frac{1}{N} \ln W_N(f)$, then we have*

$$\lambda[f] = \sup_{\mathbf{R}} \{f(s) - I(s)\}.$$

Varadhan proved this result for a large class of random variables not necessarily iid. We refer to [15] and [13] for a more detailed and rigorous discussion of this topic. The two theorems are indeed connected basically by noticing that we get one from the other upon choosing $f(s) = ks$.

7.3 Strong Approximations

Strong approximations in Probability and Statistics are results that describe the closeness almost surely of random processes such as partial sums and empirical processes to certain Gaussian processes. As a result, strong laws such as the law of the iterated logarithm and weak laws such as the central limit theorem (see Central Limit Theorems) follow.
Let $\mathbf{X}_1, \mathbf{X}_2, \ldots$ be a sequence of iid random variables. Let $S_N = \Sigma_{i=1}^N X_i$ be defined as before. If the mean $\mu = E[\mathbf{X}_1]$ exists (finite), then the strong law of large numbers states that $S_N/N \to \mu$, almost surely, as $N \to +\infty$. As before the question is: at what rate does this convergence take place? This question was answered, in 1941 by Hartman and Wintner, who proved the *Law of the Iterated Logarithm* (LIL). If, in addition, the variance σ^2 of $\mathbf{X}_1$ is finite, then

$$\limsup_{N\to+\infty} \frac{S_N - N\mu}{\sigma\sqrt{2N\ln\ln N}} \to_{a.s.} 1 \tag{7.1}$$

$$\liminf_{N\to+\infty} \frac{S_N - N\mu}{\sigma\sqrt{2N\ln\ln N}} \to_{a.s.} -1. \tag{7.2}$$

Questions of this type and extensions of this result can be found in the more general context of random processes. We refer to [4] for a detailed exposition of the subject.

7.4 The Semicircle Law

Here, we give a brief overview of the semicircle law in the context of random matrix theory. In particular we will discuss the Wigner semi-circle law for Wigner matrices. We will follow the presentation of [12].

Definition 7.4.1. *A Wigner Hermitian matrix ensemble is a random matrix ensemble $M_p = (\xi_{ij})_{1\le i,j\le p}$ of Hermitian matrices (thus, $\xi_{ij} = \overline{\xi_{ji}}$) in which the upper-triangular entries ξ_{ij}, $i > j$, are iid complex random variables with mean zero and unit variance, and the diagonal entries ξ_{ii} are iid real variables, independent of the upper-triangular entries, with bounded mean and variance.*

Remark 7.4.2. *An important special case of Wigner Hermitian Matrix includes real symmetric matrices, as the ones treated in the preceding chapters.*

Since the operator norm of M_p is of order $O(\sqrt{p})$ (see [12]), it is natural to work with the normalized matrix $\frac{1}{\sqrt{p}}M_p$. Given any $p \times p$ Hermitian matrix M_p, we can form the (normalized) empirical spectral distribution (or ESD for short) as

$\mu_{\frac{1}{\sqrt{p}}M_p} := \frac{1}{p}\sum_{j=1}^{p} \delta_{\lambda_j(M_p)/\sqrt{p}}$, of M_p, where $\lambda_1(M_p) \leq \ldots \leq \lambda_p(M_p)$ are the eigenvalues of M_p, including multiplicity.

Remark 7.4.3. *The ESD is a probability measure, which can be viewed as a distribution of the normalized eigenvalues of M_p. When M_p is a random matrix ensemble, then the ESD $\mu_{\frac{1}{\sqrt{p}}M_p}$ is now a random measure, i.e. a random variable taking values in the space $Pr(\mathbb{R})$ of probability measures on the real line. Basically, the distribution of $\mu_{\frac{1}{\sqrt{p}}M_p}$ is a probability measure on probability measures.*

Now, we consider the behaviour of the ESD of a sequence of Hermitian matrix ensembles M_p as $p \to \infty$.

Definition 7.4.4. *A sequence of random ESDs $\mu_{\frac{1}{\sqrt{p}}M_p}$ converge in probability (resp., converge almost surely) to a deterministic limit $\mu \in Pr(\mathbb{R})$ if, for every test function $\varphi \in C_c(\mathbb{R})$, the quantities $\int_{\mathbb{R}} \varphi \, d\mu_{\frac{1}{\sqrt{p}}M_p}$ converge in probability (resp. converge almost surely) to $\int_{\mathbb{R}} \varphi \, d\mu$.*

We can now state the Wigner semi-circular law.

Theorem 7.4.5 (Semicircular Law)**.** *Let M_p be the top left $p \times p$ minors of an infinite Wigner matrix $(\xi_{ij})_{i,j\geq 1}$. Then, the ESDs $\mu_{\frac{1}{\sqrt{p}}M_p}$ converge almost surely (and hence also in probability) to the Wigner semi-circular distribution*

$$\mu_{sc} := \frac{1}{2\pi}(4 - |x|^2)_+^{1/2} \, dx.$$

Proof. See [12]. □

Remark 7.4.6. *The semi-circular law is an analogue of the central limit theorem, with the semi-circular distribution taking on the role of the normal distribution. Of course, there is a striking difference between these two distributions, in that the former is compactly supported while the latter is merely subgaussian. One reason for this is that the concentration of measure phenomenon is more powerful in the case of ESDs of Wigner matrices than it is for averages of iid variables. We refer to [12] for a more detailed explanation.*

7.5 Extreme value distributions

In statistics, the *Fisher-Tippett-Gnedenko theorem* (also the *Fisher-Tippett theorem* or the *Extreme value theorem*) is a general result in extreme value theory regarding asymptotic distribution of extreme order statistics. The maximum of a sample of iid random variables, after proper renormalization, converges in distribution to one of three possible distributions, the Gumbel distribution, the Fréchét distribution, or the Weibull distribution. Credit for the extreme value theorem (or convergence to types theorem) was given to Gnedenko (1948), previous versions were stated by Fisher and Tippett and by Fréchét. One may refer to Chapter 8 of Arnold et al. [1] for a detailed exposition of this topic. The role of the extremal types theorem for maxima is similar to that of central limit theorem for averages.

Theorem 7.5.1. *Let $(X_1, X_2 \dots, X_N)$ be a sequence of independent and identically-distributed random variables, and let $M_N = \max\{X_1, \dots, X_N\}$. If a sequence of pairs of real numbers (a_N, b_N) exists such that each $a_N > 0$ and $\lim_{N\to\infty} P\left(\frac{M_N - b_N}{a_N} \leq x\right) = F(x)$, where F is a nondegenerate distribution function, then the limit distribution F belongs to either the Gumbel, the Fréchét or the Weibull family.*

These distributions can all be grouped into the family of *generalized extreme value distributions.*

Definition 7.5.2. *The generalized extreme value distribution has cumulative distribution function*

$$F(x; \mu, \sigma, \xi) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}$$

for $1 + \xi(x-\mu)/\sigma > 0$, where $\mu \in \mathbf{R}$ is the location parameter, $\sigma > 0$ the scale parameter, and $\xi \in \mathbf{R}$ the shape parameter. For $\xi = 0$, the expression is formally undefined and is understood as a limiting case. The density function is, consequently,

$$f(x; \mu, \sigma, \xi) = \frac{1}{\sigma}\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{(-1/\xi)-1} \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}$$

again, for $1 + \xi(x-\mu)/\sigma > 0$.

Example 7.5.3. *The shape parameter ξ governs the tail behaviour of the distribution. The sub-families defined by $\xi = 0$, $\xi > 0$ and $\xi < 0$ correspond, respectively, to the Gumbel, Fréchét and Weibull families, whose cumulative distribution functions are displayed below.*

- Gumbel or type I extreme value distribution *($\xi = 0$) with*

$$F(x; \mu, \sigma, 0) = e^{-e^{-(x-\mu)/\sigma}} \quad \text{for } x \in \mathbf{R};$$

- Fréchét or type II extreme value distribution *($\xi = \alpha^{-1} > 0$) with*

$$F(x; \mu, \sigma, \xi) = e^{-((x-\mu)/\sigma)^{-\alpha}} \qquad \text{for } x > \mu;$$

- Reversed Weibull or type III extreme value distribution *($\xi = -\alpha^{-1} < 0$) with*

$$F(x;\mu,\sigma,\xi) = e^{-(-(x-\mu)/\sigma)^{\alpha}} \quad x \le \mu,$$

where $\sigma > 0$.

Bibliography

[1] B. C. Arnold, N. Balakrishnan and H. N. Nagaraja, A First Course in Order Statistics, *John Wiley and Sons*, New York (1992).

[2] T.T. Cai, J. Fan and T. Jiang, Distributions of Angles in Random Packing on Spheres, *Journal of Machine Learning Research* (2013) arXiv:1306.0256v1.

[3] T.T. Cai and T. Jiang, Phase Transition in Limiting Distributions of Coherence of High-Dimensional Random Matrices, *Journal of Multivariate Analysis* **107** (2012) 24-39.

[4] M. Csörgő, Strong Approximations in Probability and Statistics, *Academic*, New York (1981).

[5] A. Dobson, An introduction to Generalized Linear Models, *Chapman and Hall/CRC Press* , Boca Raton - Florida (2008).

[6] T. Kariya and M. Eaton, Robust Tests for Spherical Symmetry, *The Annals of Statistics* **5** (1977) 206-215.

[7] W.S. Kendall and I. Molchanov, New Perspectives in Stochastic Geometry *Oxford University Press,* New York (2010).

[8] R.J. Muirhead, Aspects of Multivariate Statistical Theory *John Wiley and Sons*, New York (1982).

[9] C. Stein, A bound for the error in the normal approximation to the distribution of a sum of dependent random variables *Proc. of Berk. Symp. Math. Star.* **11** (1970) 503-602.

[10] S. Smale, Mathematical Problems for the Next Century *Mathematical Intelligencer* **20** (1999) 7-15.

[11] D. Stoyan and W.S. Kendall, Stochastic Geometry and its Applications *Second edition. John Wiley and Sons*, Hoboken-New Jersey (2008) .

[12] T. Tao, Topics in Random Matrix Theory *Graduate Studies in Mathematics* **132** (2012).

[13] H. Touchette, The large deviation approach to statistical mechanics, *Phys.Rep.* **478 (1-3)** (2009) 1-69.

[14] H. Touchette, A basic introduction to large deviations: Theory, applications, simulations, *arxiv:1106.4146v3.*

[15] S. Varadhan, Asymptotic probabilities and differential equations. *Comm. Pure Appl. Math.* **19** (1966) 261-286.