

Estimation Problems Related to Random Matrix Ensembles

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch–Naturwissenschaftlichen Fakultäten
der Georg–August–Universität zu Göttingen

vorgelegt von
Rada Matić
aus
Loznica, Serbien und Montenegro

Göttingen 2006

D7

Referent: Prof. Dr. Manfred Denker

Koreferent: Prof. Mikhail Gordin

Tag der Mündlichen Prüfung:

Contents

1	Introduction	5
2	Exponential Families	11
2.1	Introduction	11
2.2	Definitions and Basic Properties	12
2.3	Parametrization and Maximum Likelihood Estimation	14
3	β-Ensembles	17
3.1	Introduction	17
3.2	Historical Development	19
3.3	Definitions and Summary of Known Results	20
3.4	Scaled β -Hermite Ensemble	28
3.5	Scaled β -Laguerre Ensemble	31
3.5.1	Maximum Likelihood Estimation	35
3.6	β -Jacobi Ensemble	37
3.6.1	Maximum Likelihood Estimation	40
3.7	Cauchy Ensemble	40
3.7.1	Maximum Likelihood Estimation	44
3.8	Discrete Orthogonal Ensembles	44
3.8.1	Krawtchouk Polynomials Ensemble	44
3.8.2	Charlier Polynomials Ensemble	47
4	Asymptotical analysis	49
4.1	Introduction	49

4.2	Historical Overview and Motivation	50
4.3	Central Limit Theorem for Sufficient Statistics	54
4.3.1	β -Hermite Ensemble	54
4.3.2	β -Laguerre Ensemble	55
4.3.3	β -Jacobi Ensemble	57
4.3.4	Cauchy Ensemble	59
4.4	Maximum Likelihood Estimation	60
4.4.1	β -Hermite Ensemble	60
4.4.2	β -Laguerre Ensemble	61
4.4.3	β -Jacobi Ensemble	67
4.4.4	Cauchy Ensemble	69
5	Circular Ensembles	71
5.1	Introduction	71
5.2	Circular Ensembles and Generalizations	73
5.3	Weyl's Integration Formula	75
5.4	Toeplitz Determinants and Strong Szegö Theorem	78
5.5	Asymptotic Normality	82
5.6	Asymptotic Maximum Likelihood Estimation	85
5.7	The Third-Order Phase Transition Model	90
6	Summary and Open Questions	95
	References	96
	Curriculum Vitae	106

Chapter 1

Introduction

Random matrix theory (RMT) has been attracting attention of researchers in different branches of science since it was introduced by Wishart in the late 1920s. Nowadays, it has been successfully applied to an extraordinarily large variety of problems in fields as diverse as multivariate statistics [4], [88], harmonic analysis on groups [30], combinatorics [9], nuclear physics [87], quantum gravity [41], wireless communications [113], etc.

Although random matrices were first encountered by Weyl [114, 115] in connection with the integration over the unitary group, the explicit study of their properties began in 1928 with Wishart [121], who obtained the joint distribution of sample variances and covariances from multivariate normal population. After a relatively slow start, investigation of random matrix ensembles intensified in the 1950s, when Wigner [118, 119] proposed the use of RMT to characterize certain properties of complex many-body systems such as heavy nuclei, complex atoms and molecules. Predictions of RMT were successfully applied in describing statistical properties of excited states of atomic nuclei. Seemingly, that was the first time when matrix models were used to simulate physical reality. Later the scope of RMT has broadened enormously, and connections to other branches of physics and mathematics were established. A number of interesting models were found to lead to random matrix ensembles, among which are the random growth model, the length of the longest increasing subsequence of a random permutation, non-colliding random processes and tandem queues, directed last percolation, and others. Random matrices became widely used in experimental studies in economics, physics, electrical engineering, etc. In these studies a stochastic phenomenon (e.g. a channel in wireless communications) can be simulated by a matrix with random elements. Such an extension of the range of possible applications of RMT led to the rigorous mathematical treatment of certain properties of random matrices and resulted in many important discoveries.

The purpose of this thesis is to develop an exact (that is, non-asymptotic) and asymptotic estimation theory for a class of probability distributions arising in RMT. The models that we consider have been studied extensively, but the problem of *statistical estimation* of their parameters seems to be unaddressed in the literature, although estimation issues are of interest for both theoretical and practical applications.

In Chapter 3 of the thesis we consider a statistical experiment generated by an observation $\mathbf{X} = (X_1, \dots, X_n) \in \mathbb{R}^n(\mathbb{Z}^n)$. Here X_1, \dots, X_n are (strongly dependent) eigenvalues of a random matrix H . We assume that H belongs to a certain class of parametric random matrix ensembles. One of the main questions in Chapter 3 is whether one is able to obtain explicit expressions for the maximum likelihood estimators of model parameters for arbitrary fixed $n \geq 2$. This leads to certain non-trivial problems such as evaluation of Selberg-type integrals.

Chapters 4 and 5 are devoted to an *asymptotical* estimation in the framework of spectral random matrix theory. For a class of models under consideration we derive central limit theorems for sufficient statistics. Moreover, we obtain that the information about the true parameter value, contained in a sample of size n , is of order $O(n^2)$ under certain natural conditions (e.g. convergence of empirical measures to a proper distribution as $n \rightarrow \infty$).

Formally, we consider the sequence of statistical experiments $\{\mathcal{E}_n\}_{n \geq 2}$, with

$$\mathcal{E}_n = (\mathbb{R}^n, \mathcal{B}^n, P_{\boldsymbol{\theta}}^n, \boldsymbol{\theta} \in \Theta), \quad \text{or} \quad \mathcal{E}_n = (\mathbb{Z}^n, \mathcal{A}^n, P_{\boldsymbol{\theta}}^n, \boldsymbol{\theta} \in \Theta), \quad (1.1)$$

where $\mathcal{B}^n(\mathcal{A}^n)$ is the σ -algebra of Borel (all) subsets of $\mathbb{R}^n(\mathbb{Z}^n)$, and $P_{\boldsymbol{\theta}}^n$ is the probability measure with the density

$$p_{n; w_{\boldsymbol{\theta}; n}}(\mathbf{x}) = C_{\beta, n}^{-1}(w_{\boldsymbol{\theta}; n}) \prod_{1 \leq i < j \leq n} |x_i - x_j|^\beta \prod_{i=1}^n w_{\boldsymbol{\theta}; n}(x_i), \quad \mathbf{x} \in \mathbb{R}^n(\mathbb{Z}^n), \quad (1.2)$$

with respect to the Lebesgue (counting) measure on $\mathbb{R}^n(\mathbb{Z}^n)$. Here, $\beta > 0$ is the fixed *repulsiveness* parameter, $C_{\beta, n}(w_{\boldsymbol{\theta}; n})$ is a normalizing constant, and $w_{\boldsymbol{\theta}; n}$ is a weight function depending on the dimension parameter n , and on an r -dimensional parameter $\boldsymbol{\theta} = (\theta_1, \dots, \theta_r) \in \Theta \subset \mathbb{R}^r$ to be estimated. We assume that the weight function is of exponential form

$$w_{\boldsymbol{\theta}; n}(x) = \exp \left\{ n \sum_{i=1}^r \theta_i V_i(x) \right\}, \quad x \in \mathbb{R}(\mathbb{Z}), \quad (1.3)$$

where V_1, \dots, V_r are suitable known real-valued functions on $\mathbb{R}(\mathbb{Z})$. Under this as-

sumption the probability densities (1.2), known as β -ensembles in RMT, belong to the class of exponential (geometric) families of probability density functions which are briefly reviewed in Chapter 2. The exponential form of the weight function $w_{\theta;n}(x)$ in (1.3) is justified from several arguments (see Section 3.1), while the presence of the dimension parameter n in $w_{\theta;n}(x)$ is discussed in Chapter 4.

It is known that β -ensembles arise in many different settings in physics: the spectral theory of complex quantum systems and of single particle systems with chaotic dynamics, in the theory of log-potential gases, etc. In random matrix theory β -ensembles represent the joint probability distribution of eigenvalues of a certain class of $n \times n$ random matrices, while in log-potential theory they coincide with the Boltzmann factor of the one-dimensional Coulomb gas consisting of n particles free to move on the real line \mathbb{R} (lattice \mathbb{Z}) in an external field with the potential $-n \sum_{i=1}^r \theta_i V_i(x)/\beta$. In the Coulomb gas interpretation, the parameter $\beta > 0$ is the inverse temperature of the system.

Several properties of the probability densities (1.2) should be emphasized here.

- (a) β -ensembles with $\beta = 1, 2, 4$, known as *orthogonal*, *unitary* and *symplectic* ensembles respectively, have been seen as special primarily due to two reasons: existence of explicit matrix models from which they can be constructed, and availability of analytical methods for the derivation of their exact or asymptotical properties. Recently, ensembles of sparse random matrices which lead to the probability densities (1.2) with general β (but specific choice of the weight function $w_{\theta;n}$) were identified in Dumitriu, Edelman [42], Killip, Nenciu [82]. In the following, we will allow β to be arbitrary positive number, unless stated otherwise.
- (b) We interpret X_1, \dots, X_n as coordinates of n particles on the real line \mathbb{R} (lattice \mathbb{Z}). Then presence of the absolute value of the Vandermonde determinant

$$\Delta(\mathbf{x}) = \prod_{1 \leq i < j \leq n} (x_i - x_j) = (-1)^{\frac{n(n-1)}{2}} \begin{vmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \\ \dots & \dots & \dots & \dots \\ x_1^{n-1} & x_2^{n-1} & \dots & x_n^{n-1} \end{vmatrix}, \quad (1.4)$$

in formula (1.2) models a strong repulsive interaction between the particles. The larger β is, the stronger are the repelling forces between the observations with joint probability density (1.2). Due to repulsiveness structure, the asymptotical behavior of the maximum likelihood estimator $\hat{\theta}_n$ is in deep contrast to the classical theory of i.i.d. samples under regularity conditions (see Lehmann, Casella [85]). In Chapter 4 we show that the maximum likelihood estimator $\hat{\theta}_n$, based on n dependent observations with joint probability density (1.2),

converges to the true parameter value $\boldsymbol{\theta}_0$ significantly faster than in the case of n independent observations, precisely

$$n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \Sigma(\boldsymbol{\theta}_0)), \quad n \rightarrow \infty. \quad (1.5)$$

(c) Multiple integrals of the functions

$$f_n(\mathbf{x}) = \prod_{1 \leq i < j \leq n} |x_i - x_j|^\beta \prod_{i=1}^n w_{\boldsymbol{\theta}; n}(x_i), \quad \mathbf{x} \in \mathbb{R}^n \quad (1.6)$$

from which the normalizing constant $C_{\beta, n}(w_{\boldsymbol{\theta}; n})$ is evaluated, have been the subject of investigation of many mathematicians, including Selberg [97], Aomoto [6], Anderson [3], Andrews, Askey, Roy [5], amongst others. The explicit expressions for $C_{\beta, n}(w_{\boldsymbol{\theta}; n})$ with general β are known for $w_{\boldsymbol{\theta}; n}$ being one of the classical weight functions: Hermite, Laguerre, Jacobi or Cauchy. They are derived from Selberg's integral

$$\begin{aligned} I(a, b, \beta, n) &= \int_{[0,1]^n} |\Delta(\mathbf{x})|^\beta \prod_{i=1}^n x_i^{a-1} (1-x_i)^{b-1} d\mathbf{x} \\ &= \prod_{i=0}^{n-1} \frac{\Gamma(1 + (i+1)\beta/2) \Gamma(a + i\beta/2) \Gamma(b + i\beta/2)}{\Gamma(1 + \beta/2) \Gamma(a + b + (n-1+i)\beta/2)}, \quad \operatorname{Re} a, \operatorname{Re} b, \beta > 0, \end{aligned} \quad (1.7)$$

by changing variables of integration, choosing appropriate values for a and b , or taking limits. The computation of $C_{\beta, n}(w_{\boldsymbol{\theta}; n})$ for general β and other choices of the weight function $w_{\boldsymbol{\theta}; n}$ of the form (1.3) seems to be an open problem. However, if $\beta = 2$ and $w_{\boldsymbol{\theta}; n}$ is a weight function on the unit circle \mathbb{T} , it follows from the determinant identity due to Heine and Szegö (see e.g. [105]) that

$$C_{2, n}(w_{\boldsymbol{\theta}; n}) = (2\pi)^n n! D_n[w_{\boldsymbol{\theta}; n}],$$

where $D_n[w_{\boldsymbol{\theta}; n}]$ is the n^{th} Toeplitz determinant with respect to $w_{\boldsymbol{\theta}; n}$. For detailed discussion of this remarkable connection, the reader is referred to Chapter 5.

The main results of this thesis are the following. We obtain the exact distribution of sufficient statistics for the four classical β -ensembles: β -Hermite, β -Laguerre, β -Jacobi and Cauchy unitary ensemble. Exact distributions of certain linear statistics for β -Hermite and β -Laguerre ensemble have been computed by Baker and Forrester [18]. We reestablish and extend some of their results. Furthermore, we prove central limit theorems for sufficient statistics of the four classical β -ensembles and

the generalized circular unitary ensemble. Central limit theorems were obtained (in a broader context) for β -Hermite and Dyson's circular ensembles in Johansson [69], [71], Dumitriu, Edelman [43], Diaconis, Evans [40]. Corresponding theorems for sufficient statistics of β -Laguerre, β -Jacobi, the Cauchy unitary and the generalized circular unitary ensemble seem to be new results. In the author's opinion, the main contribution of the thesis are results regarding the asymptotical behavior of the maximum likelihood estimators (MLE) of model parameters. We analyze the four classical β -ensembles, two discrete orthogonal polynomial ensembles and the generalized circular unitary ensemble. Existence, uniqueness, consistency and asymptotical normality of MLE for random matrix ensembles seem to be unexamined previously in the literature. We hope that these and related results may be of interest to specialists on asymptotic estimation theory.

The organization of this thesis is as follows.

In Chapter 2 we review some basic properties of exponential families of distributions. In particular, we formulate a theorem concerning existence and uniqueness of the maximum likelihood estimator for standard exponential families.

In Chapter 3 we focus our attention on the *exact* analysis of four classical β -ensembles: β -Hermite, β -Laguerre, β -Jacobi and β -Cauchy. We review the historical development of RMT, and discuss its connection to the theory of orthogonal polynomials on the real line. Results regarding the *exact* distribution of the sufficient (vector) statistics for these four models are established. Moreover, we obtain existence and uniqueness of the maximum likelihood estimators for the (vector-valued) parameters of these ensembles. Additionally, we consider two discrete orthogonal polynomial models (the Krawtchouk and Charlier ensemble), and derive the maximum likelihood estimators of their parameters. Throughout Chapter 3, the dimension parameter n is fixed.

In Chapter 4 we develop an *asymptotical* theory for four classical random matrix ensembles. We derive a central limit theorem for the sufficient (vector) statistics in β -Hermite, β -Laguerre, β -Jacobi and Cauchy unitary ensemble. Furthermore, we show that the maximum likelihood estimator of the (vector-valued) parameter is consistent and asymptotically normal in the sense of convergence in (1.5).

Chapter 5 is devoted to statistical analysis of *circular unitary ensembles*. Observations $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)$ are the phases of the eigenvalues of a random $n \times n$ unitary matrix. We introduce the formal definition of the parametric generalized Dyson's circular unitary ensemble by setting the probability density of n eigenphases to be

of the form

$$p_{\boldsymbol{\theta};n}(\boldsymbol{\zeta}) = C_n^{-1}(\boldsymbol{\theta}) \prod_{j=1}^n w_{\boldsymbol{\theta}}(e^{i\zeta_j}) \prod_{1 \leq j < k \leq n} |e^{i\zeta_j} - e^{i\zeta_k}|^2, \quad \boldsymbol{\zeta} \in [0, 2\pi]^n, \quad (1.8)$$

where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_r)$ is an r -dimensional parameter, $C_n(\boldsymbol{\theta})$ is a normalizing constant, and $w_{\boldsymbol{\theta}}$ is a suitable weight function on the unit circle \mathbb{T} . We show that consistent estimation of parameter $\boldsymbol{\theta}$ is not possible unless the dimension parameter n is included into the weight function $w_{\boldsymbol{\theta}}$. Finally, we consider the third-order phase transition model of Gross and Witten [57], and show that its concentration parameter $\gamma_0 \in (0, 1 - \epsilon]$, $0 < \epsilon < 1$, can be estimated consistently. The asymptotical maximum likelihood estimator $\hat{\gamma}_n$ is given in closed form. The variance of $\hat{\gamma}_n$ is shown to asymptotically attain the Cramér-Rao lower bound.

Chapter 6 contains a short summary and discussion of open problems to be considered in further investigations.

Acknowledgements

I would like to express my deep gratitude to my mentors, Prof. Dr. Manfred Denker and Prof. Mikhail Gordin, for proposing the topic of my dissertation, their excellent supervision and many helpful suggestions during this research work. I also thank Prof. Dr. Walter Zucchini for his encouraging support and comments.

I am grateful to all the people from the Institute for Mathematical Stochastics and the Center for Statistics for providing me with an excellent working environment. Thanks go to Dr. Alexei Min, Dr. Manuel Stadlbauer, Andreas Cordes, Thomas Hotz, Sachar Kablutschko, Anna Levina, Mikhail Langovoy, Nora Serdyukova, Janis Valeinis and all my other colleagues and friends.

My sincerest thanks goes to my family and friends for supporting me with patience, love and understanding.

Financial support from the Center for Statistics is gratefully acknowledged.

Chapter 2

Exponential Families

2.1 Introduction

This chapter is an overview of the general theory of statistical exponential families. We recall some well known facts, which are needed in the sequel, and we base the terminology on Brown [34].

The fundamental role played by exponential families in the theory and applications of statistics is generally known. Specific examples of exponential families of distributions include the normal, exponential, gamma and beta, binomial and multinomial, geometric, negative binomial and Poisson family. These families and their properties have been studied extensively, individually and from a general perspective of belonging to the exponential family of distributions. For a more detailed discussion, we refer the reader to Barndorff-Nielsen [20, 21], Brown [34] and Chentsov [35].

The chapter is divided into three sections. In section 2 we introduce exponential families and standard exponential families of probability distributions on \mathbb{R}^k and recall some of their basic properties, while section 3 is devoted to initiation of the mean-value parametrization and discussion of the maximum likelihood estimation problem. The proofs of results formulated here can be found in Brown [34].

2.2 Definitions and Basic Properties

Let μ be a σ -finite measure on $(\mathbb{R}^k, \mathcal{B})$, where \mathcal{B} is the σ -algebra of Borel subsets of \mathbb{R}^k . Define

$$\lambda(\boldsymbol{\theta}) = \int \exp\langle \boldsymbol{\theta}, \mathbf{x} \rangle d\mu(\mathbf{x}) \quad (2.1)$$

on the set

$$N = \{\boldsymbol{\theta} \in \mathbb{R}^k : \int \exp\langle \boldsymbol{\theta}, \mathbf{x} \rangle d\mu(\mathbf{x}) < \infty\}, \quad (2.2)$$

and let

$$\psi(\boldsymbol{\theta}) = \log \lambda(\boldsymbol{\theta}), \quad \boldsymbol{\theta} \in N, \quad (2.3)$$

$$p_{\boldsymbol{\theta}}(\mathbf{x}) = \exp(\langle \boldsymbol{\theta}, \mathbf{x} \rangle - \psi(\boldsymbol{\theta})), \quad \boldsymbol{\theta} \in N, \quad (2.4)$$

where $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^k x_i y_i$ is the scalar product in \mathbb{R}^k .

Let $\Theta \subset N$. The family of probability densities $\{p_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ is called a *k-dimensional standard exponential family*. The convex set N is called the *natural parameter space*, $\boldsymbol{\theta} \in \Theta$ is referred to as a *canonical parameter* and $\psi(\boldsymbol{\theta})$ is the *cumulant-generating function*.

The family is called *full* if $\Theta = N$, and *regular* if N is an open set in \mathbb{R}^k .

We define a convex support of μ as

$$K = \text{cl conv supp}(\mu), \quad (2.5)$$

where $\text{supp}(\mu)$ is the support of μ , $\text{cl } S$ is the closure, and $\text{conv } S$ is the convex hull of $S \subset \mathbb{R}^k$. The dimension of S , $\dim S$, is the dimension of the linear space spanned by the set of vectors $\{(x_1 - x_2) : x_1, x_2 \in S\}$. A *k*-dimensional standard exponential family is *minimal* if

$$\dim K = \dim N = k. \quad (2.6)$$

A full minimal standard exponential family is called a *canonical* exponential family.

In the following definition we introduce a class of exponential families of distributions.

Definition 1. Let $\{P_{\boldsymbol{\omega}}, \boldsymbol{\omega} \in \Omega\}$ be a family of probability distributions on the measurable space (Y, \mathcal{A}) , such that $P_{\boldsymbol{\omega}} \ll \nu, \boldsymbol{\omega} \in \Omega$, for some σ -finite measure ν on

(Y, \mathcal{A}) . Suppose there exist functions

$$\begin{aligned} C : \Omega &\rightarrow (0, \infty), \\ R : \Omega &\rightarrow \mathbb{R}^k, \\ T : Y &\rightarrow \mathbb{R}^k, && \text{(Borel measurable),} \\ h : Y &\rightarrow [0, \infty), && \text{(Borel measurable),} \end{aligned}$$

such that

$$p_{\omega}(y) = \frac{dP_{\omega}}{d\nu}(y) = C(\omega)h(y) \exp\langle R(\omega), T(y) \rangle. \quad (2.7)$$

Then $\{p_{\omega}, \omega \in \Omega\}$ is called a *k-dimensional exponential family* of probability densities.

Proposition 2.1. *Any k-dimensional exponential family (2.7) can be reduced by sufficiency, reparametrization, and proper choice of measure μ to a k-dimensional standard exponential family (2.4). The distributions of the sufficient statistics $\mathbf{X} = T(\mathbf{Y})$ form an exponential family with canonical parameter $\boldsymbol{\theta} = R(\omega)$.*

The moments of the sufficient statistics \mathbf{X} can be calculated by differentiating under the integral sign. This result is formulated as theorem.

Theorem 2.2. *Let $\{p_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ be a standard exponential family and $\boldsymbol{\theta}_0 \in \text{int } N$. Then all derivatives of λ and ψ exist at $\boldsymbol{\theta}_0$ and are obtained by differentiation under the integral sign. In particular, the following relations are valid*

$$\frac{\partial^l}{\partial \theta_1^{l_1} \dots \partial \theta_k^{l_k}} \lambda(\boldsymbol{\theta}_0) = \int \prod_{i=1}^k x_i^{l_i} \exp\langle \boldsymbol{\theta}_0, \mathbf{x} \rangle d\mu(\mathbf{x}), \quad \forall \mathbf{l} = (l_1, \dots, l_k) \in \mathbb{Z}_+^k, \sum_{i=1}^k l_i = l, \quad (2.8)$$

$$E_{\boldsymbol{\theta}_0}[\mathbf{X}] = \nabla \psi(\boldsymbol{\theta}_0). \quad (2.9)$$

One of the properties of the normalizing factor $\lambda(\boldsymbol{\theta})$ that we will need in the sequel is its analyticity. We extend the definition of λ to the complex domain by setting

$$\lambda(\boldsymbol{\theta}) = \int \exp\langle \boldsymbol{\theta}, \mathbf{x} \rangle d\mu(\mathbf{x}) \quad (2.10)$$

on $L := \{\boldsymbol{\theta} \in \mathbb{C}^k : \text{Re}(\boldsymbol{\theta}) = (\text{Re}(\theta_1), \dots, \text{Re}(\theta_k)) \in N\}$. Then

Theorem 2.3. *Let $\{p_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ be a regular canonical exponential family. Then $\lambda(\boldsymbol{\theta})$ is the analytic function on the set L .*

A family $\{P_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ of probability distributions with corresponding densities $\{p_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ relative to a σ -finite measure μ on \mathbb{R}^k is called *complete*, if for every

$T : \mathbb{R}^k \rightarrow \mathbb{R}$ such that

$$\int T(\mathbf{x})p_{\boldsymbol{\theta}}(\mathbf{x})d\mu(\mathbf{x}) = 0, \quad \forall \boldsymbol{\theta} \in \Theta,$$

we have that

$$T(\mathbf{x}) = 0 \quad P_{\boldsymbol{\theta}}\text{- a.e.} \quad \forall \boldsymbol{\theta} \in \Theta.$$

The following theorem provides a sufficient condition of the completeness of an exponential family.

Theorem 2.4. ?? *Let $\{p_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ be a standard exponential family with $\text{int}(\Theta) \neq \emptyset$. Then $\{p_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ is complete.*

2.3 Parametrization and Maximum Likelihood Estimation

The standard parametrization of the canonical family is convenient for the analytical treatment of the probability distributions under consideration. Nevertheless, it does not allow a direct probabilistic interpretation. In order to introduce the parametrization with probabilistic meaning, we define the map $\xi : N \rightarrow \text{int } K$ by

$$\xi(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}(\mathbf{X}), \quad \boldsymbol{\theta} \in N. \quad (2.11)$$

Theorem 2.5. *Let $\{p_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in N\}$ be a canonical regular exponential family. Then the map*

$$\xi(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}(\mathbf{X}) \quad (2.12)$$

defines a homeomorphism between N and $\text{int } K$.

Under the conditions of Theorem 2.5, the family $\{p_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ can be reparametrized by $\boldsymbol{\xi} = \xi(\boldsymbol{\theta})$, and such a parametrization is called the *mean-value parametrization*. The map $\xi(\cdot)$ is invertible on $\text{int } K$, and we denote its inverse $\xi^{-1}(\cdot)$ by $\boldsymbol{\theta}(\cdot)$. In regular exponential families the mean-value parametrization is closely related to the maximum likelihood estimation as described below.

Let $\{p_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ be an exponential family. Define

$$l(\boldsymbol{\theta}, \mathbf{x}) = \log p_{\boldsymbol{\theta}}(\mathbf{x}), \quad \boldsymbol{\theta} \in N. \quad (2.13)$$

For $S \subset N$ let

$$l(S, \mathbf{x}) = \sup_{\boldsymbol{\theta} \in S} l(\boldsymbol{\theta}, \mathbf{x}) \quad (2.14)$$

and

$$\hat{\theta}_S(\mathbf{x}) = \{\boldsymbol{\theta} \in S : l(\boldsymbol{\theta}, \mathbf{x}) = l(S, \mathbf{x})\}. \quad (2.15)$$

Then $\hat{\boldsymbol{\theta}} \in \hat{\theta}_S(\mathbf{x})$ is called the *maximum likelihood estimate* of $\boldsymbol{\theta}$ at \mathbf{x} relative to $S \subset N$. A function $\delta : \text{int } K \rightarrow \Theta$ such that $\delta(\mathbf{x}) \in \hat{\theta}_\Theta(\mathbf{x})$ μ -a.e. is called the *maximum likelihood estimator*.

The following theorem establishes the conditions for the existence and the uniqueness of the maximum likelihood estimate for exponential families of distributions.

Theorem 2.6. *Let $\{p_\theta : \theta \in \Theta\}$ be a regular canonical exponential family. If $\mathbf{x} \in \text{int } K$, then*

$$\hat{\theta}_\Theta(\mathbf{x}) = \{\theta(\mathbf{x})\} \in \Theta. \quad (2.16)$$

Otherwise, $\hat{\theta}_\Theta(\mathbf{x}) = \emptyset$.

Remark 2.7. Since $\xi(\boldsymbol{\theta}) = \mathbb{E}_\theta(\mathbf{X})$ is the homeomorphism of N and $\text{int } K$, we have that \mathbf{x} is the maximum likelihood estimate of the mean-value parameter $\boldsymbol{\xi}$, i.e.

$$\hat{\boldsymbol{\xi}}(\mathbf{x}) = \mathbf{x}, \quad \mathbf{x} \in \text{int } K.$$

It is important to notice that from the definition of the mean-value parameter $\boldsymbol{\xi}$ it immediately follows that its method-of-moments estimator and the maximum likelihood estimator $\hat{\boldsymbol{\xi}}$ agree.

Chapter 3

β -Ensembles

3.1 Introduction

The aim of this chapter is to develop an exact (non-asymptotical) estimation theory for a class of distribution families arising in RMT. For every integer $n \geq 2$, the class of probabilities under consideration consists of collections of joint n -dimensional probability distributions of exchangeable random variables. The random variables take values in the real line \mathbb{R} , the positive ray \mathbb{R}_+ , the unit interval $(0, 1)$, or their discrete subsets. The estimation theory for observations taking values in the unit circle will be developed in Chapter 5.

A paradigmatic example of the measure under consideration is given by the joint probability density in \mathbb{R}^n

$$C_{\beta,n}^{-1}(w) \prod_{1 \leq i < j \leq n} |x_i - x_j|^\beta \prod_{i=1}^n w(x_i), \quad (3.1)$$

where w is a suitable weight function on the real line \mathbb{R} (or its subset), $\beta > 0$ is a fixed parameter and $C_{\beta,n}(w)$ is the normalization constant. Function w can be viewed as an infinite-dimensional functional parameter and its estimation can be dealt with within a non-parametric framework. However, we focus our attention on the examples where

$$w(x) = \exp \left[\sum_{i=1}^r \theta_i V_i(x) \right], \quad x \in \mathbb{R}. \quad (3.2)$$

Here, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_r) \in \Theta \subset \mathbb{R}^r$ is an r -dimensional parameter to be estimated,

and V_i are suitable (known) functions, called confining potentials in the theory of one-dimensional Coulomb gases on the real line. An alternative problem, when β in (3.1) is regarded as a parameter to be estimated is also possible, but not considered here. The generalities for all exponential families do not depend on β .

Such a choice of the weight function w is motivated by several reasons. First of all, a considerable number of models discussed in the literature allow such a description. Secondly, the argument of Shannon's information theory, applied in the situation common for RMT, leads to distributions that have an exponential form (see Balian [19], Mehta [87] for details). These distributions are distinguished among others by the property of maximizing the entropy of the system under certain prescribed conditions. As the third source of justification one can see the applicability of the theory of exponential families which is well-developed in the framework of mathematical statistics. However, our models are in deep contrast to the classical theory of i.i.d. samples of size n : the order of random fluctuations of certain quantities is different. This can be explained by a strong (*repulsive*) interaction between the observations, due to the presence of the Vandermonde determinant in (3.1).

Throughout this section, we regard n and $\beta > 0$ as fixed. Presence of the size parameter n , and applications of the models in physics suggest consideration of our estimation problem in an asymptotical setting. This situation will be discussed in Chapter 4, while here we obtain rigorous analytical results for every fixed n . Asymptotic behavior of certain quantities and estimators will be considered just for the sake of comparison with the respective situation in the i.i.d. case.

The chapter is organized as follows. In section 2 we give a short review of historical development of RMT and describe ensembles that will be treated statistically in the subsequent sections. Section 3 is devoted to the analytical initiation of the Gaussian (Hermite), Wishart (Laguerre), MANOVA (Jacobi) and Cauchy ensemble, and to an overview of the theoretical results related to our further analysis. We focus our attention on the concepts and contributions that are of significant relevance for our later investigation, providing the reader with the list of references where more detailed account of the analytical tools and wide-ranging applications is given. In sections 4, 5, 6, 7 we derive exact probability distributions for certain linear statistics arising in the study of random matrix ensembles. In particular, we reestablish and extend some of the results of Baker, Forrester [18]. These results are further incorporated into the maximum likelihood estimation problem. Existence and uniqueness of the maximum likelihood estimators is derived. Finally, two examples of exponential (geometric) discrete matrix ensembles are considered in section 8. The exact distributions of the sufficient statistics arising in their study are obtained. Additionally, the variance of the maximum likelihood estimators is computed.

3.2 Historical Development

In many applications of mathematical statistics, a random phenomenon can be modeled by a matrix whose elements are random variables. The problems arising in related studies are very involved theoretically, as well as computationally. These problems produced the need to obtain rigorous methods that will give efficient analytical tools under such circumstances. An enormously large class of issues in RMT is related to the *spectral* properties of random matrices, rather than to the matrix structure itself. We follow the pattern of *spectral* RMT and investigate the probability distributions on spectra, providing the reader with the underlying matrix models when possible.

Random matrices were introduced by Wishart [121] in 1928, as sample covariance matrices for n independent identically distributed p -variate normal vectors. In fact, in 1915, Fisher [46] derived this distribution for the bivariate case ($p = 2$) in his study of sample correlation coefficient from normal population. Wishart's model, which came to be known as the real *Wishart ensemble*, was further extended and modified. One may see Anderson [4], or Johnson, Kotz [75] for details. One of the extensions led to the definition of zonal polynomials in James [67], and to the study of eigenvalues of sample covariance matrices in terms of special functions. A thorough account of the results related to the spectral properties of the Wishart ensembles can be found in Muirhead [88].

Another matrix model originating from mathematical statistics, which will be treated in this chapter is connected to the multivariate analysis of variance (and thus called the *MANOVA ensemble*). The references include Hsu [63], Olkin, Rubin [92], etc. We refer the reader to Muirhead [88] for the detailed description of this model.

Independently of the developments in multivariate statistics, another type of matrix ensembles was initiated in the framework of nuclear physics by Wigner. The model is commonly known as the *Gaussian ensemble*, since the entries of matrices follow Gaussian distribution. This model is dominant in physics literature and it has found broad extensions of its domain of applications connecting fields as diverse as log-potential theory, quantum chromodynamics, transport properties of disordered systems, etc. We refer the reader to Mehta [87] and Guhr, Müller-Groeling, Weidenmüller [58] for recent applications of RMT concepts in physics. Although formally introduced in nuclear physics, spectral theory of Gaussian ensembles appeared to be closely connected to many classical problems in mathematics such as zeros of Riemann ζ -function, distribution of primes on the real line, length of longest increasing subsequence in a random permutation, etc. (see Mehta [87], Forrester [47], Katz, Sarnak [80], Keating, Snaith [81] for details). Since the analytical tools are

the most developed for Gaussian ensembles, particular attention will be paid to this specific case in our further investigation.

Recently, a new spectral model, called the *Cauchy ensemble*, was introduced by Forrester [47]. A particular case of this model is independently discussed in physics under the name *Laurentzian ensemble* (see Brouwer [33], Tierz [106]). The Cauchy ensemble will be described in detail in section 7.

It should be mentioned here that there are two independent sources that encouraged us to consider four ensembles described above: the Gaussian, Wishart, MANOVA and Cauchy. Apart from the fact that they are the most discussed in the literature, one can obtain connections to the

- theory of classical orthogonal polynomials on the real line
- theory of exponential families in mathematical statistics.

We will pay particular attention to these connections after the analytical initiation of ensembles. Following this pattern of thinking, we have decided to include two discrete ensembles (the *Krawtchouk* and the *Charlier* polynomial ensemble), since they are the natural analogue of the four ensembles on the real line from the point of view of connections to the theory of orthogonal polynomials and theory of exponential families.

3.3 Definitions and Summary of Known Results

This section is devoted to the formulation of basic concepts and analytical results in RMT, definitions of orthogonal, unitary and symplectic random matrix ensembles and their β -generalizations. In particular cases, asymptotical results are given.

For almost 80 years, researchers in different branches of science are using matrix models to simulate random phenomena. Special attention is focused on the study of the *Gaussian ensembles* for a number of reasons. The Gaussian ensembles were introduced as a statistical model of an (infinite-dimensional) Hamiltonian of the physical system in order to describe its spectral properties. In such a model, physical system (e.g. heavy nucleus or complex atom) is represented by its Hamiltonian, which in turn can be modeled by a random Hermitian matrix H of large size. Physical consideration of the system imposes constraints on possible choices of the probability measure $P(H)$ on the set of all Hermitian matrices, such as invariance under orthogonal, unitary or symplectic transformations. The invariance properties of $P(H)$ depend on certain symmetries of the system like time-reversal

or rotational symmetry. This reasoning led Wigner and Dyson to the definitions of three Gaussian ensembles.

Definition 2. The *Gaussian orthogonal ensemble* (GOE) is a probability measure $P(H)$ defined on the space T_{1G} of all $n \times n$ real symmetric matrices by two requirements

1. The measure is invariant under every transformation

$$H \rightarrow WHW^T$$

of T_{1G} into itself, where W is an arbitrary orthogonal matrix.

2. Various elements $H_{j,k}, j \leq k$ are statistically independent.

In physics literature the Gaussian orthogonal ensemble is used to model systems with even spin in the presence of rotational symmetry and invariance under time reversal. Requirement of statistical independence of various elements $H_{j,k}, j \leq k$ is added for convenience and lacks physical motivated interpretation. The definition of the *Gaussian symplectic ensemble* (GSE) is similar (see Mehta [87]).

Much simpler mathematically is the Gaussian unitary ensemble defined as follows.

Definition 3. The *Gaussian unitary ensemble* (GUE) is a probability measure $P(H)$ defined on the space T_{2G} of all $n \times n$ Hermitian matrices by the following properties

1. The measure is invariant under every automorphism

$$H \rightarrow UHU^{-1}$$

of T_{2G} into itself, where U is an arbitrary unitary matrix.

2. The (real) diagonal entries and the (complex) upper-diagonal entries of H are independent random variables.

GUE is used to model systems without invariance under time-reversal.

It appeared that the two postulates of statistical independence and invariance under orthogonal, unitary and symplectic transformations, lead to the measure of the form

$$P(H) = \exp[-a \operatorname{tr} H^2 + b \operatorname{tr} H + c], \quad (3.3)$$

on the spaces of real symmetric, Hermitian and self-dual Hermitian matrices respectively. Here a is real and positive, b and c are real. For the details of the derivation we refer the reader to Porter, Rosenzweig [96] or Mehta [87].

Further developments in nuclear physics were primarily connected to the spectral properties of three ensembles and their comparison with experimental results. The joint probability density of the eigenvalues x_1, \dots, x_n was derived and shown to be equal to

$$p_{\beta,n}^*(\mathbf{x}) = \prod_{1 \leq i < j \leq n} |x_i - x_j|^\beta \prod_{i=1}^n \exp[-ax_i^2 + bx_i + c], \quad \mathbf{x} \in \mathbb{R}^n, \quad (3.4)$$

where β is the symmetry parameter equal to 1, 2, 4 for orthogonal, unitary, symplectic ensemble respectively. Many authors prefer the scaled version of the density (3.4)

$$p_{\beta,n}(\mathbf{x}) = C_{\beta,n}^{-1} \prod_{1 \leq i < j \leq n} |x_i - x_j|^\beta \prod_{i=1}^n \exp[-x_i^2/2], \quad \mathbf{x} \in \mathbb{R}^n, \quad (3.5)$$

which can be obtained from $p_{\beta,n}^*$ by the change of coordinates. $C_{\beta,n}$ is the normalizing constant dependent on the symmetry parameter β . This form of the density is more convenient in examining the statistical properties of the system, such as 1- and 2-point correlation functions, extremal eigenvalues, e.t.c. Computationally, a sample following the density (3.5), can be obtained by the diagonalization of a random real symmetric, Hermitian or self-dual Hermitian matrix with diagonal elements having standard normal distribution and off-diagonal elements following normal distribution with variance $\sigma^2 = 1/2$. All the elements are supposed to be independent up to the symmetry constraints.

For many years, the attention of mathematicians and physicists has been attached to the three values of β equal to 1, 2 or 4. The analytical methods employed in the investigation of various statistical properties of orthogonal, unitary and symplectic ensembles were highly parameter-specific. The case $\beta = 2$ is considered as the easiest one, since for its evaluation one needs to know properties of orthogonal polynomials (the Hermite polynomials for the Gaussian ensemble). The analysis in orthogonal and symplectic case involves Pfaffians, anti-symmetric scalar products and the associated skew-orthogonal polynomials. Recently, the three isolated cases ($\beta = 1, 2, 4$) have been unified and studied in the general β framework. The initial step in this direction was made by Baker and Forrester [17], who considered Calogero-Sutherland-type models for quantum systems. Further development of general β concept was made in Dumitriu, Edelman [42] through the construction of the model of random sparse matrices whose eigenvalues follow $p_{\beta,n}(x)$ for general

$\beta > 0$. To formulate their result, we introduce the symmetric random matrix

$$H_\beta = \begin{pmatrix} Y_n & Z_{n-1} & \dots & 0 \\ Z_{n-1} & Y_{n-1} & \dots & 0 \\ 0 & \dots & \dots & Z_1 \\ 0 & \dots & Z_1 & Y_1 \end{pmatrix}, \quad (3.6)$$

where Y_i , $1 \leq i \leq n$ have the standard normal distribution. The sub-diagonal elements Z_i are distributed as square roots of non-homogeneous random variables following the gamma distribution with shape parameter $i\beta$ and scale parameter 1 for arbitrary $i \in \{1, \dots, n-1\}$. All $2n-1$ distinct elements of H_β are supposed to be independent. We can formulate the theorem.

Theorem 3.1 (Dumitriu, Edelman [42]). *Let $H_\beta = Q_\beta \Lambda_\beta Q_\beta^T$ be the eigenvalue decomposition of H_β , $\beta > 0$. Fix the signs of the first row \mathbf{q} of the matrix Q_β to be non-negative, and order the eigenvalues in non-increasing order. Then \mathbf{q} and $\boldsymbol{\lambda} = \text{diag } \Lambda_\beta$ are independent. Furthermore, the joint density of eigenvalues $\boldsymbol{\lambda}$ is*

$$p_{\beta,n}(\boldsymbol{\lambda}) = C_{n,\beta}^{-1} \prod_{1 \leq i < j \leq n} |\lambda_i - \lambda_j|^\beta \prod_{i=1}^n \exp[-\lambda_i^2/2], \quad \boldsymbol{\lambda} \in \mathbb{R}^n. \quad (3.7)$$

Proof. See [42]. □

The ensemble of sparse random matrices H_β is called β -Hermite ensemble in Dumitriu, Edelman [42]. We adopt their terminology, since such a description may be considered as more accurate because of its connection to the type of weight function in (3.7).

The model under our consideration is closely connected to β -Hermite ensemble, but it contains a scaling parameter which will be estimated based on the observation of respective eigenvalues. To be more precise, we consider the density of the *scaled β -Hermite ensemble*

$$p_\beta(\mathbf{x}; \sigma) = C_{n,\beta}^{-1}(\sigma) \prod_{1 \leq i < j \leq n} |x_i - x_j|^\beta \prod_{i=1}^n e^{-\frac{x_i^2}{2\sigma^2}}, \quad \mathbf{x} \in \mathbb{R}^n, \quad (3.8)$$

and estimate the parameter σ^2 from a sample (X_1, \dots, X_n) of exchangeable random variables with the joint probability density as specified in (3.8). Our estimation procedure does not rely on any inference about correlation functions of the system such as level density, distribution of level spacings etc. Instead, we apply the theory of exponential families and use the fact that the information about the scaling

parameter σ^2 can be extracted from a sample by means of sufficient statistic

$$H(X_1, \dots, X_n) = \sum_{i=1}^n X_i^2.$$

Remark 3.2. We notice that numerical generation procedure of the scaled β -Hermite ensemble is rather simple. Namely, the density $p_\beta(\mathbf{x}; \sigma)$ is easily obtained from Theorem 3.1 by diagonalizing the matrix σH_β .

In section 5 of this chapter, we apply the maximum likelihood procedure to the eigenvalue density of *scaled β -Laguerre (Wishart) ensemble* given by

$$p_\beta(\mathbf{x}; \alpha, \theta) = C_{n,\beta}^{-1}(\alpha, \theta) \prod_{1 \leq i < j \leq n} |x_i - x_j|^\beta \prod_{i=1}^n x_i^{\alpha-1} e^{-x_i/\theta}, \quad \mathbf{x} \in (0, \infty)^n. \quad (3.9)$$

As in the Gaussian case, the density (3.9) was first obtained for "classical" values of $\beta = 1, 2, 4$ from the probability density

$$p_{\beta,n,p}(H) = C_{\beta;n,p}^{-1} \det H^{\beta(n-p+1)/2-1} \exp[-\text{tr } H/2], \quad H > 0, n \geq p-1, \quad (3.10)$$

on the set of $p \times p$ positive-definite real symmetric, Hermitian and self-dual Hermitian matrices, respectively. The induced measure on the set of eigenvalues is absolutely continuous with respect to the Lebesgue measure on \mathbb{R}_+^p and its density coincides with $p_\beta(\mathbf{x}; \alpha, \theta)$, $\mathbf{x} \in \mathbb{R}_+^p$, where $\alpha = \beta(n-p+1)/2$, $\theta = 2$, and $\beta \in \{1, 2, 4\}$.

The joint density of matrix elements (3.10) was derived by Wishart [121], who considered the probability distribution of the sample covariance matrix $X^T X$ from multivariate normal population. Here, X is $n \times p$ matrix with standard normal real entries, and X^T is its transpose. Further, the complex ($\beta = 2$) and quaternion ($\beta = 4$) cases were considered. Many other derivations of the density (3.10) were given subsequently (see James [67]). Recently, the unifying β theory has been established by construction of sparse random matrix models with eigenvalues as in (3.9). To formulate the result, we introduce the random matrix $L_{\beta;\alpha,\theta} = B_{\beta;\alpha,\theta} B_{\beta;\alpha,\theta}^T$, where

$$B_{\beta;\alpha,\theta} = \begin{pmatrix} Y_n & 0 & \dots & 0 \\ Z_{n-1} & Y_{n-1} & \dots & 0 \\ 0 & \dots & \dots & 0 \\ 0 & \dots & Z_1 & Y_1 \end{pmatrix}. \quad (3.11)$$

Here Y_i^2 follows the gamma distribution with shape parameter $\alpha + (i-1)\beta/2$ and

scale parameter θ , while Z_i^2 has the probability density function

$$p(x) = \frac{\theta^{-\beta i/2}}{\Gamma(\beta i/2)} x^{\beta i/2-1} e^{-x/\theta}, \quad x > 0,$$

and all $2n - 1$ random variables are independent.

Theorem 3.3. *Let $L_{\beta;\alpha,\theta} = Q_{\beta;\alpha,\theta} \Lambda_{\beta;\alpha,\theta} Q_{\beta;\alpha,\theta}^T$ be the spectral decomposition of the tridiagonal random matrix $L_{\beta;\alpha,\theta} = B_{\beta;\alpha,\theta} B_{\beta;\alpha,\theta}^T$, where $B_{\beta;\alpha,\theta}$ is given in (3.11). Fix the signs of the first row \mathbf{q} of the matrix $Q_{\beta;\alpha,\theta}$ to be non-negative, and order the eigenvalues in non-increasing order. Then \mathbf{q} and $\boldsymbol{\lambda} = \text{diag } \Lambda_{\beta;\alpha,\theta}$ are independent. Furthermore, the joint density of eigenvalues $\boldsymbol{\lambda}$ is*

$$p_{\beta,n}(\boldsymbol{\lambda}) = C_{n,\beta}^{-1} \prod_{1 \leq i < j \leq n} |\lambda_i - \lambda_j|^\beta \prod_{i=1}^n \lambda_i^{\alpha-1} e^{-\lambda_i/\theta}, \quad \boldsymbol{\lambda} \in \mathbb{R}_+^n. \quad (3.12)$$

Again, as in the case of β -Hermite ensemble, we consider the *scaled* version of β -Laguerre ensemble (3.9), and obtain the MLE of parameters α and θ . The information about their properties is extracted from the the corresponding sufficient statistics

$$L_1(X_1, \dots, X_n) = \sum_{i=1}^n \log X_i,$$

$$L_2(X_1, \dots, X_n) = \sum_{i=1}^n X_i,$$

and the results on their exact distributions.

The third classical ensemble in RMT is MANOVA (Jacobi) ensemble. It was introduced in multivariate statistics in testing the independence of two Gaussian populations. Random matrix from MANOVA ensemble is numerically generated by setting $C = (A + B)^{-1/2} A (A + B)^{-1/2}$, where A and B are independent Wishart matrices with the densities $p_{\beta,n_1,p}(A)$ and $p_{\beta,n_2,p}(B)$ as in (3.10), and $X^{1/2}$ denotes a unique positive definite square root of a matrix X (notice that a Wishart matrix is always positive definite by its construction, and consequently, its square root can be defined uniquely). The probability density of the eigenvalues of matrix C is

$$p_\beta(\mathbf{x}; a, b) = C_{p,\beta}^{-1}(a, b) \prod_{1 \leq i < j \leq p} |x_i - x_j|^\beta \prod_{i=1}^p x_i^{a-1} (1 - x_i)^{b-1}, \quad \mathbf{x} \in [0, 1]^p, \quad (3.13)$$

where $a = \beta(n_1 - p + 1)/2$, $b = \beta(n_2 - p + 1)/2$ and $\beta = 1, 2, 4$ for real symmetric, Hermitian and self-dual Hermitian matrices respectively. In (3.13) we recognize the celebrated Jacobi weight $w(x) = x^{a-1}(1-x)^{b-1}$, $x \in (0, 1)$. Generalizations over β were established shortly after β -generalizations of Hermite and Laguerre ensemble in Killip, Nenciu [82]. Respective densities are called the β -Jacobi ensembles. In section 6, we obtain the exact distribution of the statistics

$$J_1(X_1, \dots, X_p) = \sum_{i=1}^p \log X_i,$$

$$J_2(X_1, \dots, X_p) = \sum_{i=1}^p \log(1 - X_i)$$

based on a sample (X_1, \dots, X_p) with the joint eigenvalue density $p_{\beta,p}(\mathbf{x}; a, b)$ as in (3.13), and incorporate them into the maximum likelihood estimation procedure.

The last ensemble on the real line that is considered in this chapter is β -Cauchy ensemble. The ensemble is defined in terms of its eigenvalues

$$p_{\beta}(\mathbf{x}; \alpha) = C_{n,\beta}^{-1}(\alpha) \prod_{1 \leq i < j \leq n} |x_i - x_j|^{\beta} \prod_{i=1}^n (1 + x_i^2)^{-\alpha}, \quad \mathbf{x} \in \mathbb{R}^n, \quad (3.14)$$

where $\alpha > \beta(n-1)/2 + 1/2$, $\beta > 0$, and $C_{n,\beta}(\alpha)$ is the normalizing constant. Further details are given in Section 7.

To summarize this section, we want to point out the connections between the four ensembles described above. First of all, the density of their eigenvalues can be written in the form

$$p_{\beta,w}(\mathbf{x}) = C_{\beta,n}^{-1}(w) \prod_{1 \leq i < j \leq n} |x_i - x_j|^{\beta} \prod_{i=1}^n w(x_i), \quad \mathbf{x} \in \mathbb{R}^n \quad (3.15)$$

where

$$w(x) = \begin{cases} e^{-x^2/2\sigma^2}, & x \in \mathbb{R}, & \text{Hermite,} \\ x^{a-1}e^{-x/\theta}, & x \in \mathbb{R}_+, & \text{Laguerre,} \\ x^{a-1}(1-x)^{b-1}, & x \in (0, 1), & \text{Jacobi,} \\ (1+x^2)^{-\alpha}, & x \in \mathbb{R}, & \text{Cauchy.} \end{cases} \quad (3.16)$$

The weight functions in (3.16) have the common property that their logarithmic

derivative is a rational function. Namely, if we consider

$$\frac{d}{dx} \log w(x) = \frac{w'(x)}{w(x)} =: -\frac{g(x)}{f(x)}, \quad (3.17)$$

where polynomials f and g have no common factors and $f > 0$ on the support of the corresponding weight, then

$$(f, g) = \begin{cases} (\sigma^2, x), & \text{Hermite,} \\ (\theta x, x - (a - 1)\theta), & \text{Laguerre,} \\ (x(1 - x), -a(1 - x) - bx + 1), & \text{Jacobi,} \\ (1 + x^2, 2\alpha x), & \text{Cauchy.} \end{cases} \quad (3.18)$$

Note that

$$\text{degree } f \leq 2, \quad \text{degree } g \leq 1.$$

The weight functions (3.16) are called *classical* in the theory of orthogonal polynomials on the real line. For more details regarding the connection of random matrix theory and the theory of orthogonal polynomials, the reader is referred to Deift [37] and König [83].

The second connection is established through the theory of exponential families. Namely, every density in (3.15) with the weight function as specified in (3.16), can be written in a general form

$$C_{\beta, n}^{-1}(\boldsymbol{\theta}) \prod_{1 \leq i < j \leq n} |x_i - x_j|^\beta \prod_{i=1}^n \exp\left[\sum_{j=1}^r t_j(\theta_j) V_j(x_i)\right], \quad (3.19)$$

where r is the number of parameters, $\boldsymbol{\theta} \in \Omega \subset \mathbb{R}^r$ is the r -dimensional parameter and t_i, V_i are suitable functions. The density (3.15) with classical weight functions w from (3.16), can be written in the form (3.19) with $r = 1, 2, 2, 1$ for scaled β -Hermite, scaled β -Laguerre, β -Jacobi and β -Cauchy ensembles, respectively.

Notice also that the weights in (3.16) are proportional to the probability densities of well-known distributions from exponential families on the real line: the Gaussian, Gamma, Beta and Pearson type VII distribution. From this perspective, one can think of (3.15) as a spectral random matrix modification of

$$p_{0, w}(\mathbf{x}) = C_{0, n}^{-1}(w) \prod_{i=1}^n w(x_i), \quad \mathbf{x} \in \mathbb{R}^n,$$

which corresponds to a sample of n i.i.d. random variables with the density propor-

tional to $w(x)$.

This connection to the standard situation of i.i.d. samples encouraged us to include into our consideration the following two discrete ensembles: the Krawtchouk and Charlier unitary polynomial ensembles, which can be considered as a spectral random matrix modifications of the binomial and Poisson distribution, respectively. Formally, we will consider the densities

$$p_w(\mathbf{x}) = C_n^{-1}(w) \prod_{1 \leq i < j \leq n} |x_i - x_j|^2 \prod_{i=1}^n w(x_i), \quad \mathbf{x} \in \mathbb{Z}^n, \quad (3.20)$$

where

$$w(x) = \begin{cases} \binom{N}{x} p^x (1-p)^{N-x}, & x \in \{0, \dots, N\}, & \text{Krawtchouk,} \\ a^x / x!, & x \in \mathbb{Z}_+, & \text{Charlier.} \end{cases} \quad (3.21)$$

We confine our attention to the unitary case ($\beta = 2$) since the normalizing constant $C_n(w)$ is explicitly given for this case only. The references where these two ensembles were studied are given in Section 8.

3.4 Scaled β -Hermite Ensemble

In this section we derive the exact distribution of the sufficient statistics for the scaled β -Hermite ensemble, and obtain the maximum likelihood estimator $\hat{\sigma}_n^2$ of the scaling parameter σ^2 in the explicit form.

Consider the probability density of the scaled β -Hermite ensemble

$$p_\beta(\mathbf{x}; \sigma) = C_{n,\beta}^{-1}(\sigma) \prod_{1 \leq i < j \leq n} |x_i - x_j|^\beta \prod_{i=1}^n e^{-\frac{x_i^2}{2\sigma^2}}, \quad \mathbf{x} \in \mathbb{R}^n, \quad (3.22)$$

where

$$C_{n,\beta}(\sigma) = \sigma^{n[\beta(n-1)/2+1]} (2\pi)^{n/2} \prod_{i=1}^n \frac{\Gamma(1 + i\beta/2)}{\Gamma(1 + \beta/2)}$$

is the normalizing constant, obtained from the change of coordinates in Selberg's integral (1.7), and $n \geq 2, \sigma^2 > 0, \beta > 0$. The logarithmic derivative of the partition

function $C_{n,\beta}(\sigma)$ is given by

$$\frac{n[\beta(n-1)/2+1]}{\sigma} = \frac{1}{\sigma^3} \mathbb{E}_n \left[\sum_{i=1}^n X_i^2 \right],$$

where \mathbb{E}_n denotes the expectation with respect to the measure (3.22), and $\mathbf{X} = (X_1, \dots, X_n)$ is a sample of exchangeable random variables with the joint probability density (3.22). Thus, the unbiased estimator $\hat{\sigma}_n^2$ of the scaling parameter σ^2 , obtained from a sample (X_1, \dots, X_n) , is given by

$$\hat{\sigma}_n^2 = \frac{\sum_{i=1}^n X_i^2}{n[\beta(n-1)/2+1]}. \quad (3.23)$$

The probability distribution of the statistic

$$H_\beta(X_1, \dots, X_n) = \sum_{i=1}^n X_i^2 \quad (3.24)$$

is derived from its Laplace transform in the following proposition.

Proposition 3.4. *Let (X_1, \dots, X_n) be a sample with the joint probability density*

$$p_\beta(\mathbf{x}; \sigma) = C_{n,\beta}^{-1}(\sigma) \prod_{1 \leq i < j \leq n} |x_i - x_j|^\beta \prod_{i=1}^n e^{-\frac{x_i^2}{2\sigma^2}}, \quad \mathbf{x} \in \mathbb{R}^n,$$

where $C_{n,\beta}(\sigma)$ is the normalizing constant. The probability density of the statistic

$$H_\beta(X_1, \dots, X_n) = \sum_{i=1}^n X_i^2,$$

is given by

$$h_\beta(x) = \frac{(2\sigma^2)^{-n[\beta(n-1)/2+1]/2}}{\Gamma(n[\beta(n-1)/2+1]/2)} x^{n[\beta(n-1)/2+1]/2-1} e^{-\frac{x}{2\sigma^2}}, \quad x > 0. \quad (3.25)$$

Proof. Consider the Laplace transform of $H_\beta(X_1, \dots, X_n)$

$$\begin{aligned} \mathcal{L}[H_\beta](s) &= \mathbb{E}_n[e^{-sH_\beta(X_1, \dots, X_n)}] \\ &= (1 + 2s\sigma^2)^{-n[\beta(n-1)/2+1]/2}, \end{aligned}$$

for $s > 0$. Since the Laplace transform of the random variable Y that follows the

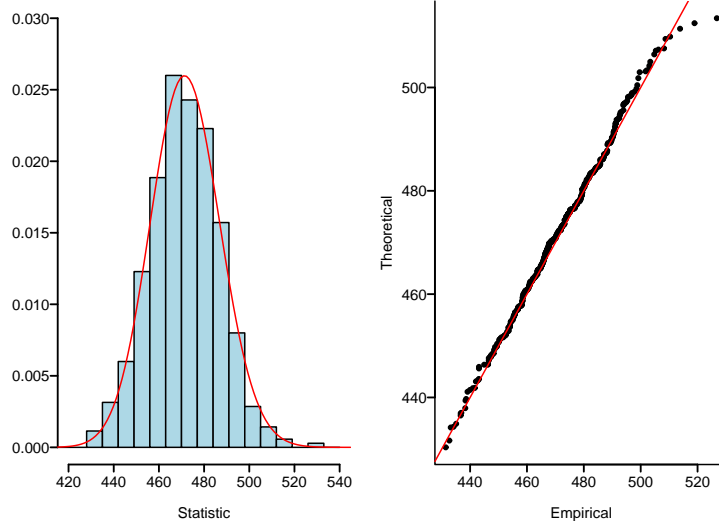


Figure 3.1: The probability density and qq-plot of $H_\beta(\mathbf{X})$ for $n = 50, \beta = 3/2, \sigma^2 = 1/4$.

gamma distribution with the scale parameter a and the shape parameter ρ is

$$\mathcal{L}[Y](s) = (1 + sa)^{-\rho}, \quad s > 0,$$

we arrive at the conclusion that the statistics $H_\beta(X_1, \dots, X_n)$ follows the gamma distribution with the scale parameter $2\sigma^2$ and shape parameter $n[\beta(n-1)/2 + 1]/2$. Its density is exactly as specified in (3.25). \square

In Figure 3.1 we illustrate the result of Proposition 3.4 by empirically calculating 500 values of $H_\beta(\mathbf{X})$ based on the spectra of random matrices from the scaled β -Hermite ensemble with $\beta = 3/2, \sigma^2 = 1/4, n = 50$, and comparing the obtained histogram with the theoretical density of the corresponding gamma distribution. The comparison of respective quantiles is provided in the qq-plot.

The proposition that we obtained has a consequence related to the maximum likelihood estimation procedure. More precisely, $\hat{\sigma}_n^2$ given in (3.23) is the maximum likelihood estimator of σ^2 based on a sample (X_1, \dots, X_n) , and its mean and variance can be deduced from the corresponding moments of the gamma distribution, namely

$$\mathbb{E}_n[\hat{\sigma}_n^2] = \sigma^2, \quad (3.26)$$

$$\text{Var}_n[\hat{\sigma}_n^2] = \frac{2\sigma^4}{n[\beta(n-1)/2 + 1]}. \quad (3.27)$$

Remark 3.5. Notice that the variance of $\hat{\sigma}_n^2$ is of order $O(1/n^2)$.

Remark 3.6. In order to point out the difference between the random matrix setting and the common situation in mathematical statistics, where independent observations are usually considered, we recall some basic results from the theory of point estimation. The standard problem encountered in mathematical statistics is the estimation of the mean and variance of a population following normal distribution. If we consider a sample (X_1, \dots, X_n) of n independent observations with the Gaussian distribution with mean 0 and variance σ^2 , the joint probability density is given by

$$p_0(\mathbf{x}; \sigma) = C_{n,0}^{-1}(\sigma) \prod_{i=1}^n e^{-\frac{x_i^2}{2\sigma^2}}, \quad \mathbf{x} \in \mathbb{R}^n,$$

and the maximum likelihood estimator $\tilde{\sigma}_n^2$ of the population variance σ^2 is

$$\tilde{\sigma}_n^2 = \frac{\sum_{i=1}^n X_i^2}{n}.$$

Notice that $\tilde{\sigma}_n^2$ coincides with $\hat{\sigma}_n^2$ if $\beta = 0$ in (3.23). Observe that when dependence between the observations is introduced into the population in a way specific for random matrix theory, where the joint eigenvalue density is as in (3.22), we arrive at the maximum likelihood estimator given by the sufficient statistic $H_\beta(X_1, \dots, X_n)$ rescaled *quadratically* in n . The variance of the estimator $\tilde{\sigma}_n^2$, obtained from the sample of n independent observations is of order $O(1/n)$. The rescaling procedure for dependent observations reduces the variance of the respective estimator to the order $O(1/n^2)$.

3.5 Scaled β -Laguerre Ensemble

In this section we derive the exact distribution of the sufficient vector-statistics for the scaled β -Laguerre ensemble. Additionally, we prove that the maximum likelihood estimators of the shape parameter α and scale parameter θ exist and are unique for every $n \geq 2$.

The probability density function of the scaled β -Laguerre ensemble is

$$p_\beta(\mathbf{x}; \alpha, \theta) = C_{n,\beta}^{-1}(\alpha, \theta) \prod_{1 \leq i < j \leq n} |x_i - x_j|^\beta \prod_{i=1}^n x_i^{\alpha-1} e^{-x_i/\theta}, \quad \mathbf{x} \in (0, \infty)^n, \quad (3.28)$$

where

$$C_{n,\beta}(\alpha, \theta) = \theta^{n[\beta(n-1)/2+\alpha]} \prod_{i=0}^{n-1} \frac{\Gamma(1 + (i+1)\beta/2)\Gamma(\alpha + i\beta/2)}{\Gamma(1 + \beta/2)} \quad (3.29)$$

is the normalizing constant and $n \geq 2, \alpha > 0, \theta > 0$. Applying Theorem 2.2, we get the equalities

$$\begin{aligned} \frac{\partial}{\partial \theta} \log C_{n,\beta}(\alpha, \theta) &= \frac{1}{\theta^2} \mathbb{E}_n \left[\sum_{i=1}^n X_i \right], \\ \frac{\partial}{\partial \alpha} \log C_{n,\beta}(\alpha, \theta) &= \mathbb{E}_n \left[\sum_{i=1}^n \log X_i \right], \end{aligned}$$

that provide us with unbiased estimators

$$L_{1,\beta}(\mathbf{X}) = \sum_{i=1}^n X_i, \quad (3.30)$$

$$L_{2,\beta}(\mathbf{X}) = \sum_{i=1}^n \log X_i, \quad (3.31)$$

of the respective quantities

$$\theta n[\beta(n-1)/2 + \alpha]$$

and

$$\sum_{i=1}^n \psi(\alpha + i\beta/2) + n \log \theta,$$

where $\psi(x)$ is the logarithmic derivative of the Gamma function, $\mathbf{X} = (X_1, \dots, X_n)$ is a sample of exchangeable random variables with the joint probability density (3.28). The exact distribution of the vector statistic $\mathbf{L}_\beta(\mathbf{X}) = (L_{1,\beta}(\mathbf{X}), L_{2,\beta}(\mathbf{X}))$ is derived in the following proposition.

Proposition 3.7. *Let (X_1, \dots, X_n) be the sample with joint probability density (3.28). Assume that $Y_i^2, Z_i^2, 1 \leq i \leq n$, are as specified in Theorem 3.3. Then*

$$L_{1,\beta}(\mathbf{X}) = \sum_{i=1}^n X_i \stackrel{\mathcal{D}}{=} \sum_{i=1}^n Y_i^2 + \sum_{i=1}^{n-1} Z_i^2,$$

and its probability density function is

$$l_\beta(x) = \frac{\theta^{-n[\beta(n-1)/2+\alpha]}}{\Gamma(n[\beta(n-1)/2 + \alpha])} x^{n[\beta(n-1)/2+\alpha]-1} e^{-x/\theta}, \quad x > 0,$$

while

$$L_{2,\beta}(\mathbf{X}) = \sum_{i=1}^n \log X_i \stackrel{\mathcal{D}}{=} \sum_{i=1}^n \log Y_i^2.$$

Proof. Although the distributions of statistics $L_{1,\beta}(\mathbf{X}), L_{2,\beta}(\mathbf{X})$ can be obtained from their Laplace transforms, we present a different (hopefully more elegant) approach to their derivation. Recall that the density (3.28) is obtained from diagonalization of $n \times n$ random matrix $L_{\beta;\alpha,\theta} = B_{\beta;\alpha,\theta} B_{\beta;\alpha,\theta}^T$, where

$$B_{\beta;\alpha,\theta} = \begin{pmatrix} Y_n & 0 & \dots & 0 \\ Z_{n-1} & Y_{n-1} & \dots & 0 \\ 0 & \dots & \dots & 0 \\ 0 & \dots & Z_1 & Y_1 \end{pmatrix}.$$

Here Y_i^2 follows the gamma distribution with the shape parameter $\alpha + (i-1)\beta/2$ and scale parameter θ , while Z_i^2 has the probability density function

$$p(x) = \frac{\theta^{-\beta i/2}}{\Gamma(\beta i/2)} x^{\beta i/2-1} e^{-x/\theta}, \quad x > 0,$$

and all $2n-1$ random variables are independent.

Let $L_{\beta;\alpha,\theta} = Q_{\beta;\alpha,\theta} X_{\beta;\alpha,\theta} Q_{\beta;\alpha,\theta}^T$ be the spectral decomposition of the tridiagonal matrix $L_{\beta;\alpha,\theta}$. Since the determinant and the trace of a matrix are invariant under orthogonal transformations, we have

$$\det L_{\beta;\alpha,\theta} = \prod_{i=1}^n Y_i^2, \quad \text{tr } L_{\beta;\alpha,\theta} = \sum_{i=1}^n Y_i^2 + \sum_{i=1}^{n-1} Z_i^2.$$

On the other hand, we have the following relations

$$\det X_{\beta;\alpha,\theta} = \prod_{i=1}^n X_i, \quad \text{tr } X_{\beta;\alpha,\theta} = \sum_{i=1}^n X_i.$$

As a consequence, we obtain that

$$\sum_{i=1}^n \log Y_i^2 \stackrel{\mathcal{D}}{=} \sum_{i=1}^n \log X_i, \quad \sum_{i=1}^n Y_i^2 + \sum_{i=1}^{n-1} Z_i^2 \stackrel{\mathcal{D}}{=} \sum_{i=1}^n X_i$$

providing us with the distribution of the statistics $L_{1,\beta}(\mathbf{X}), L_{2,\beta}(\mathbf{X})$. \square

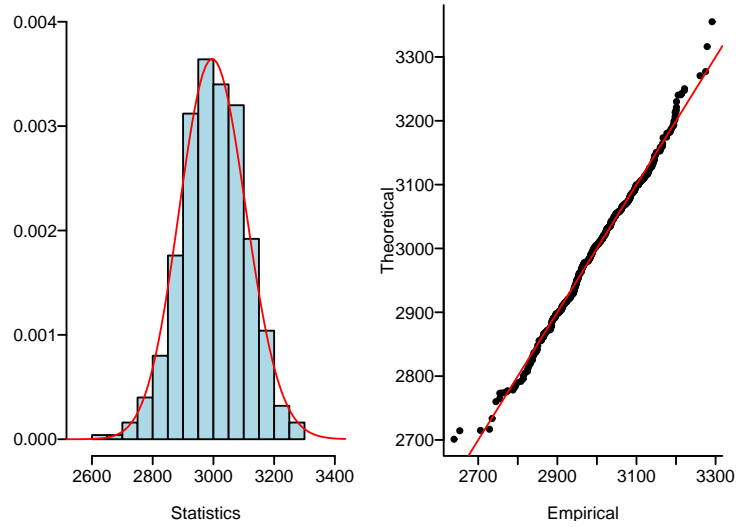


Figure 3.2: The probability density function and qq-plot of 500 statistics $L_{1,\beta}(\mathbf{X})$ for $n = 30, \beta = 3/2, \theta = 4, \alpha = 5/2$.

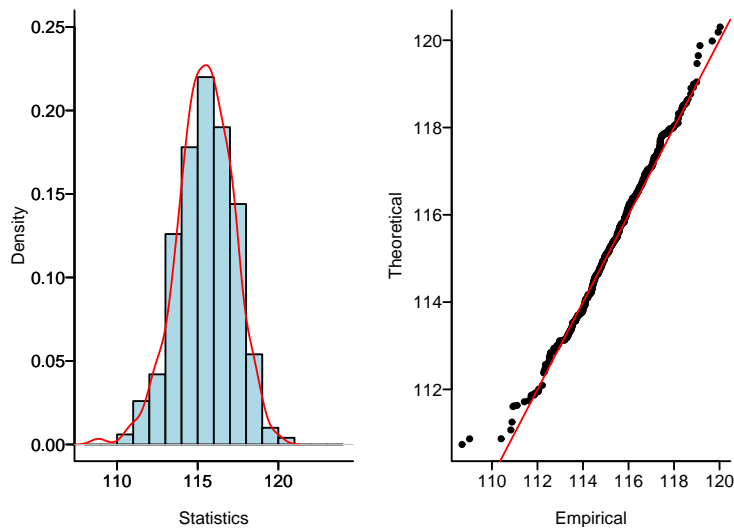


Figure 3.3: The probability density function and qq-plot of 500 statistics $L_{2,\beta}(\mathbf{X})$ for $n = 30, \beta = 3/2, \theta = 4, \alpha = 5/2$.

In Figures 3.2 and 3.3 we illustrate the result of Proposition 3.7 by empirically calculating 500 values of $L_{1,\beta}(\mathbf{X})$, $L_{2,\beta}(\mathbf{X})$ based on the spectra of random matrices from the scaled β -Laguerre ensemble with $n = 30$, $\beta = 3/2$, $\theta = 4$, $\alpha = 5/2$, and comparing the obtained histogram with the density of corresponding theoretical distribution. The comparison of respective quantiles is provided in the qq-plot.

Remark 3.8. The analysis of asymptotical behavior of correlation between $L_{1,\beta}(\mathbf{X})$ and $L_{2,\beta}(\mathbf{X})$ appropriately rescaled will be discussed in Chapter 4.

Remark 3.9. The statistics $L_{1,\beta}(\mathbf{X})$, $L_{2,\beta}(\mathbf{X})$ have a natural Coulomb gas interpretation. Namely, as indicated in Baker, Forrester [17], the linear statistic $L_{1,\beta}(\mathbf{X})$ corresponds to the dipole moment, while $L_{2,\beta}(\mathbf{X})$ represents the potential at the origin of the one-dimensional Coulomb gas of n particles free to move on the half-line \mathbb{R}_+ in the external field with the potential

$$V_{\alpha,\theta}(x) = \theta^{-1}x - (\alpha - 1) \log x, \quad x \in (0, \infty), \quad (3.32)$$

when the inverse temperature is equal to β .

3.5.1 Maximum Likelihood Estimation

Estimation of parameter θ , when α is known. First we consider the maximum likelihood estimation of the scaling parameter θ in the situation when the shape parameter α is known. The maximum likelihood estimator $\hat{\theta}_n$ of θ is given in closed form

$$\hat{\theta}_n = \frac{\sum_{i=1}^n X_i}{n[\beta(n-1)/2 + \alpha]}. \quad (3.33)$$

Again we observe that the asymptotic variance of the scaling parameter is of order $O(1/n^2)$ (compared to $O(1/n)$ in the case of independent observations).

Estimation of the shape parameter α , when θ is known. The maximum likelihood estimator $\hat{\alpha}_n$ is obtained from the equation

$$\sum_{i=0}^{n-1} \psi(\hat{\alpha}_n + i\beta/2) + n \log \theta = \sum_{i=1}^n \log X_i. \quad (3.34)$$

Since the explicit theoretical solution of (3.34) is not available, one can use the iterative procedure and incorporate asymptotical expansions for function $\psi(x)$ in order to solve the equation numerically. As a starting value in the Newton–Raphson procedure, one could use an unbiased estimator $\tilde{\alpha}_n$ defined as

$$\tilde{\alpha}_n = \frac{\sum_{i=1}^n X_i}{n\theta} - \frac{\beta(n-1)}{2}. \quad (3.35)$$

Notice that from Proposition 3.7 it immediately follows that

$$\begin{aligned} \mathbb{E}_n[\tilde{\alpha}_n] &= \alpha, \\ \text{Var}_n[\tilde{\alpha}_n] &= \frac{\beta(n-1)/2 + \alpha}{n}. \end{aligned}$$

We observe that the estimator $\tilde{\alpha}_n$ is unbiased, but not consistent since its asymptotical variance is constant. Thus we suggest to use it just as a starting value when solving the maximum likelihood equation (3.34) iteratively. The derivation of the asymptotical variance of α_n is presented in Chapter 4.

Existence and uniqueness of the maximum likelihood estimator $\hat{\alpha}_n$ follow from the properties of the function $f^{(n)}(\alpha) = \sum_{i=1}^n \psi(\alpha + j\beta/2)$. Namely, $f^{(n)}(\alpha)$ is increasing and

$$\begin{aligned} \lim_{\alpha \rightarrow 0^+} f^{(n)}(\alpha) &= 0, \\ \lim_{\alpha \rightarrow +\infty} f^{(n)}(\alpha) &= +\infty, \end{aligned}$$

for every fixed $n \in \mathbb{N}$. Therefore the equation (3.34), whose right hand side is positive with probability 1, has a unique root in the interval $(0, +\infty)$ for arbitrary fixed $n \in \mathbb{N}$.

Estimation of the shape parameter α and scale parameter θ . The maximum likelihood estimator $(\hat{\theta}_n, \hat{\alpha}_n)$ satisfies

$$\hat{\theta}_n n [\beta(n-1)/2 + \hat{\alpha}_n] = \sum_{i=1}^n X_i, \quad (3.36)$$

$$\sum_{j=0}^{n-1} \psi(\hat{\alpha}_n + j\beta/2) + n \log \hat{\theta}_n = \sum_{i=1}^n \log X_i. \quad (3.37)$$

Consequently, we have

$$\sum_{j=0}^{n-1} \psi(\hat{\alpha}_n + j\beta/2) - n \log(\beta(n-1)/2 + \hat{\alpha}_n) = n \log \frac{\prod_{i=1}^n X_i^{1/n}}{\sum_{i=1}^n X_i/n}. \quad (3.38)$$

The right hand side of the equation (3.38) is negative with probability 1 since

$$\mathbb{P}(X_i > 0) = 1, \quad 1 \leq i \leq n, \quad \mathbb{P}(X_i = X_j) = 0, \quad 1 \leq i < j \leq n,$$

and therefore

$$\mathbb{P}\left(\sum_{i=1}^n X_i/n > \prod_{i=1}^n X_i^{1/n}\right) = 1, \quad n \geq 2.$$

The function

$$f^{(n)}(\alpha) = \sum_{j=0}^{n-1} \psi(\alpha + j\beta/2) - n \log(\beta(n-1)/2 + \alpha)$$

is increasing, $f^{(n)}(\alpha) < 0$ in the interval $(0, +\infty)$ and

$$\lim_{\alpha \rightarrow 0^+} f^{(n)}(\alpha) = -\infty, \quad (3.39)$$

$$\lim_{\alpha \rightarrow +\infty} f^{(n)}(\alpha) = 0. \quad (3.40)$$

While monotonicity and the limit (3.39) follow immediately from the properties of the digamma function $\psi(x)$, equality (3.40) needs more detailed discussion. The asymptotical expansion of digamma function (see Abramowitz, Stegun [1])

$$\psi(x) = \log x - \frac{1}{2x} + O\left(\frac{1}{x^2}\right), \quad x \rightarrow \infty, \quad (3.41)$$

induces the following asymptotical identity

$$\begin{aligned} f^{(n)}(\alpha) &= \sum_{j=0}^{n-1} \log\left(\alpha + \frac{j\beta}{2}\right) - n \log\left(\alpha + \frac{(n-1)\beta}{2}\right) + o(1) \\ &= \sum_{j=0}^{n-2} \log\left(1 - \frac{(n-1-j)\beta/2}{\alpha + (n-1)\beta/2}\right) + o(1), \quad \alpha \rightarrow \infty, \end{aligned}$$

for arbitrary fixed $n \in \mathbb{N}$. Therefore, the limit (3.40) is established.

3.6 β -Jacobi Ensemble

In this section we focus our attention on the derivation of the exact distribution of sufficient statistics for β -Jacobi ensemble. Further, we show that the maximum likelihood estimators of parameters a and b exist and are unique for $n \geq 2$.

Consider the probability density of the β -Jacobi ensemble

$$p_\beta(\mathbf{x}; a, b) = C_{n,\beta}^{-1}(a, b) \prod_{1 \leq i < j \leq n} |x_i - x_j|^\beta \prod_{i=1}^n x_i^{a-1} (1 - x_i)^{b-1}, \quad \mathbf{x} \in [0, 1]^n, \quad (3.42)$$

where

$$C_{n,\beta}(a, b) = \prod_{i=0}^{n-1} \frac{\Gamma(1 + (i+1)\beta/2) \Gamma(a + i\beta/2) \Gamma(b + i\beta/2)}{\Gamma(1 + \beta/2) \Gamma(a + b + (n-1+i)\beta/2)}$$

is the normalizing constant and $n \geq 2, a > 0, b > 0$.

From Theorem 2.2, we have equalities

$$\begin{aligned} \frac{\partial}{\partial a} \log C_{n,\beta}(a, b) &= \mathbb{E}_n \left[\sum_{i=1}^n \log X_i \right], \\ \frac{\partial}{\partial b} \log C_{n,\beta}(a, b) &= \mathbb{E}_n \left[\sum_{i=1}^n \log(1 - X_i) \right], \end{aligned}$$

and the unbiased estimators

$$J_{1,\beta}(X_1, \dots, X_n) = \sum_{i=1}^n \log X_i, \quad (3.43)$$

$$J_{2,\beta}(X_1, \dots, X_n) = \sum_{i=1}^n \log(1 - X_i) \quad (3.44)$$

of the respective quantities

$$\sum_{i=0}^{n-1} \psi(a + i\beta/2) - \sum_{i=0}^{n-1} \psi(a + b + (n-1+i)\beta/2),$$

and

$$\sum_{i=0}^{n-1} \psi(b + i\beta/2) - \sum_{i=0}^{n-1} \psi(a + b + (n-1+i)\beta/2).$$

Here, $\mathbf{X} = (X_1, \dots, X_n)$ is the sample of exchangeable random variables with joint probability density (3.42). In the following proposition we derive the exact distribution of the sufficient statistics $J_{1,\beta}(\mathbf{X})$ and $J_{2,\beta}(\mathbf{X})$.

Proposition 3.10. *Let (X_1, \dots, X_n) be a sample with the joint probability density*

function (3.42). The distribution of the statistic

$$J_{1,\beta}(X_1, \dots, X_n) = \sum_{i=1}^n \log X_i,$$

coincides with the distribution of the sum of n independent random variables Z_i , $1 \leq i \leq n$, where $Z_i = \log Y_i$, and Y_i follows the beta distribution with parameters $a+i\beta/2$ and $b+(n-1)\beta/2$ for $i \in \{0, \dots, n-1\}$. Similarly, the distribution of the statistic

$$J_{2,\beta}(X_1, \dots, X_n) = \sum_{i=1}^n \log(1 - X_i),$$

coincides with the distribution of the sum of n independent random variables with components $\tilde{Z}_i = \log \tilde{Y}_i$, where \tilde{Y}_i is beta distributed with parameters $b+i\beta/2$ and $a+(n-1)\beta/2$ for $i \in \{0, \dots, n-1\}$.

Proof. The distribution of the statistics $J_{1,\beta}(\mathbf{X})$ and $J_{2,\beta}(\mathbf{X})$ can be obtained from their moment-generating functions, namely

$$\begin{aligned} M[J_{1,\beta}](s) &= \mathbf{E}[e^{sJ_{1,\beta}(X_1, \dots, X_n)}] \\ &= \prod_{i=0}^{n-1} \frac{\Gamma(a+i\beta/2+s)}{\Gamma(a+i\beta/2)} \frac{\Gamma(a+b+(n-1+i)\beta/2)}{\Gamma(a+b+(n-1+i)\beta/2+s)}, \quad s > -a. \end{aligned}$$

The moment-generating function of the random variable $Z = \log Y$, where Y follows the beta distribution with parameters p and q is

$$\begin{aligned} M[Z](s) &= \mathbf{E}[e^{s \log Y}] \\ &= \frac{\Gamma(p+s)}{\Gamma(p)} \frac{\Gamma(p+q)}{\Gamma(p+q+s)}, \quad s > -p, \end{aligned}$$

Moreover, the moment-generating function of the sum of independent random variables is the product of moment-generating functions of summands. Since the distribution of a random variable X is determined by its moment-generating function $M[X](s)$ if $M[X](s)$ exists in a neighborhood of 0 (see e.g. Bilingsley [27]), we obtain the statement of the proposition for $J_{1,\beta}(\mathbf{X})$. The proof of the statement for $J_{2,\beta}(\mathbf{X})$ is analogous. \square

3.6.1 Maximum Likelihood Estimation

The maximum likelihood estimators \hat{a}_n and \hat{b}_n satisfy the following equations

$$\begin{aligned} \sum_{j=0}^{n-1} \psi(\hat{a}_n + j\beta/2) - \sum_{j=0}^{n-1} \psi(\hat{a}_n + \hat{b}_n + (n-1+j)\beta/2) &= \sum_{i=1}^n \log X_i, \\ \sum_{j=0}^{n-1} \psi(\hat{b}_n + j\beta/2) - \sum_{j=0}^{n-1} \psi(\hat{a}_n + \hat{b}_n + (n-1+j)\beta/2) &= \sum_{i=1}^n \log(1 - X_i). \end{aligned} \quad (3.45)$$

Existence and uniqueness of the maximum likelihood estimators for $n \geq 2$ follow from Theorem 2.6 since the support of the vector sufficient statistic

$$\mathbf{J}_\beta(\mathbf{X}) = (J_{1,\beta}(\mathbf{X}), J_{2,\beta}(\mathbf{X}))$$

is the set $C = \{(t_1, t_2) \in \mathbb{R}^2 : t_1 < 0, t_2 < \log(1 - e^{t_1})\}$. The condition $\mathbf{J}_\beta(\mathbf{X}) \in \text{int } C$ of Theorem 2.6 is satisfied if and only if $n \geq 2$ and $\sum_{i=1}^n (X_i - \bar{X})^2 > 0$. The event $\sum_{i=1}^n (X_i - \bar{X})^2 = 0$ occurs with probability 0 when $n \geq 2$, and therefore the maximum likelihood estimators \hat{a}_n and \hat{b}_n exist with probability one if $n \geq 2$.

Remark 3.11. Closed algebraic forms for \hat{a}_n and \hat{b}_n seem to be unobtainable. Therefore, iterative procedures should be used for solving the system of equations (3.45), or direct numerical maximization of the log-likelihood function should be performed.

3.7 Cauchy Ensemble

This section is devoted to the derivation of the probability distribution of a sufficient statistics for Cauchy unitary ensemble. The β -Cauchy ensemble was analytically introduced in Forrester [47] as a counterpart to β -circular Jacobi ensemble, where the probability density of the phases of n eigenvalues is

$$p_\beta(\boldsymbol{\zeta}) = C_\beta^{-1}(b) \prod_{1 \leq j < k \leq n} |e^{i\zeta_j} - e^{i\zeta_k}|^\beta \prod_{j=1}^n |1 - e^{i\zeta_k}|^{2b}, \quad \boldsymbol{\zeta} \in [0, 2\pi]^n. \quad (3.46)$$

The correspondence is obtained via the Cayley transform

$$e^{i\zeta} = \frac{1 + ix}{1 - ix}, \quad \zeta \in [0, 2\pi], \quad (3.47)$$

which maps a point $e^{i\theta}$ on the unit circle into a unique point x on the real line, and transforms the density (3.46) into

$$p_\beta(\mathbf{x}; \alpha) = C_{n,\beta}^{-1}(\alpha) \prod_{1 \leq i < j \leq n} |x_i - x_j|^\beta \prod_{i=1}^n \frac{1}{(1 + x_i^2)^\alpha}, \quad \mathbf{x} \in \mathbb{R}^n, \quad (3.48)$$

with $\alpha = b + 1 + \beta(n - 1)/2$ and $C_{n,b}(\alpha)$ as the normalizing constant. In the physics literature the ensemble with the joint eigenvalue density (3.48) and $\alpha = 1 + \beta(n - 1)/2$ is known as the Lorentzian ensemble (see Brouwer [33], Tierz [106], and references therein). It arises in the study of scattering matrices in a chaotic cavity with non-ideal leads.

The analysis that follows is performed for the density of the Cauchy unitary ensemble ($\beta = 2$) with arbitrary $\alpha > n - 1/2$. We notice that our investigation can be generalized and similar results can be obtained for arbitrary $\beta > 0, \alpha > \beta(n - 1)/2 + 1/2$.

Consider the probability density

$$p(\mathbf{x}; \alpha) = C_n^{-1}(\alpha) \prod_{1 \leq i < j \leq n} |x_i - x_j|^2 \prod_{i=1}^n (1 + x_i^2)^{-\alpha}, \quad \mathbf{x} \in \mathbb{R}^n, \quad (3.49)$$

where

$$C_n(\alpha) = \pi^n 2^{-n^2 - 2\alpha n + n} \prod_{i=0}^{n-1} \frac{\Gamma(2\alpha - n - i)}{[\Gamma(\alpha - i)]^2}$$

is the normalizing constant and $\alpha > n - 1/2$. From Theorem 2.2 we know that

$$\frac{d}{d\alpha} \log C_n(\alpha) = E_n \left[- \sum_{i=1}^n \log(1 + X_i^2) \right],$$

where $\mathbf{X} = (X_1, \dots, X_n)$ is a sample of exchangeable random variables with the joint probability density (3.49). Thus,

$$2 \left\{ \sum_{i=0}^{n-1} \psi(2\alpha - n - i) - \sum_{i=0}^{n-1} \psi(\alpha - i) - n \log 2 \right\} = E_n \left[- \sum_{i=1}^n \log(1 + X_i^2) \right],$$

where $\psi(x)$ is the logarithmic derivative of the Gamma function.

In order to find the distribution of the statistic

$$C(X_1, \dots, X_n) = - \sum_{i=1}^n \log(1 + X_i^2), \quad (3.50)$$

we will consider its moment-generating function, namely

$$M[C](s) = 2^{-2sn} \prod_{i=0}^{n-1} \frac{\Gamma(2(\alpha + s) - n - i)}{\Gamma(2\alpha - n - i)} \frac{[\Gamma(\alpha - i)]^2}{[\Gamma(\alpha + s - i)]^2}, \quad s > n - \alpha - 1/2.$$

Note that

$$C_\alpha := \int_{\mathbb{R}} (1 + x^2)^{-\alpha} = \frac{4\pi}{2^{2\alpha}} \frac{\Gamma(2\alpha - 1)}{[\Gamma(\alpha)]^2} \quad (3.51)$$

and thus the moment-generating function of the random variable $Y = -\log[1 + Z^2]$, where Z is the Pearson type VII random variable with probability density function

$$p_\alpha(x) = C_\alpha^{-1} (1 + x^2)^{-\alpha}, \quad x \in \mathbb{R}, \alpha > -1/2, \quad (3.52)$$

is equal to

$$M[Y](s) = \frac{1}{2^{2s}} \frac{\Gamma(2(\alpha + s) - 1)}{\Gamma(2\alpha - 1)} \frac{[\Gamma(\alpha)]^2}{[\Gamma(\alpha + s)]^2}, \quad s > -\alpha - 1/2.$$

For more detailed account of Pearson type VII distribution see Nadarajah, Kotz [90].

Using the property of the Gamma function $\Gamma(x + 1) = x\Gamma(x)$, one can for $i \in \{0, \dots, n - 2\}$ rewrite the term

$$\frac{\Gamma(2(\alpha + s) - n - i)}{\Gamma(2\alpha - n - i)} \frac{[\Gamma(\alpha - i)]^2}{[\Gamma(\alpha + s - i)]^2}$$

as

$$\frac{\Gamma(2(\alpha + s) - 2i - 1)}{\Gamma(2\alpha - 2i - 1)} \frac{[\Gamma(\alpha - i)]^2}{[\Gamma(\alpha + s - i)]^2} \prod_{j=i+2}^n \frac{2\alpha - j - i}{2(\alpha + s) - j - i}.$$

The function

$$\phi_1(s) = \frac{\Gamma(2(\alpha + s) - 2i - 1)}{\Gamma(2\alpha - 2i - 1)} \frac{[\Gamma(\alpha - i)]^2}{[\Gamma(\alpha + s - i)]^2}$$

is the moment-generating function of the random variable $Y_i = -\log[1 + Z_i^2]$, where

Z_i has the probability density function $p_{\alpha-i}(x)$ defined in (3.52), and

$$\phi_2(s) = \prod_{j=i+2}^n \frac{\alpha - (j+i)/2}{\alpha + s - (j+i)/2}$$

is the Laplace transform of $-W_i$, where W_i

$$W_i = \sum_{j=i+2}^n E_j,$$

is distributed as the sum of $n - i - 1$ independent exponentially distributed random variables

$$E_j \stackrel{\mathcal{D}}{=} \text{Exp} \left(\alpha - \frac{i+j}{2} \right).$$

Since the sum of independent identically exponentially distributed random variables follows the gamma distribution, we obtain the proposition

Proposition 3.12. *The statistic*

$$C(X_1, \dots, X_n) = \sum_{i=1}^n \log(1 + X_i^2),$$

obtained from a sample (X_1, \dots, X_n) with the joint probability density function (3.49), is distributed as the sum of $3n - 3$ independent random variables

$$C(X_1, \dots, X_n) \stackrel{\mathcal{D}}{=} \sum_{i=0}^{n-1} Y_i + \sum_{i=2}^{2n-2} U_i.$$

Here, $Y_i = \log[1 + Z_i^2]$ with Z_i having the Pearson type VII probability density function $p_{\alpha-i}(x)$ introduced in (3.52), and U_i follows the gamma distribution with the scale parameter $\alpha - i/2$ and shape parameter $g_n(i)$ defined by the function

$$g_n(i) = \begin{cases} \lfloor \frac{i}{2} \rfloor & \text{if } i = 2, \dots, n \\ \lfloor \frac{2n-i}{2} \rfloor & \text{if } i = n+1, \dots, 2n-2. \end{cases}$$

3.7.1 Maximum Likelihood Estimation

The maximum likelihood estimator $\hat{\alpha}_n$ based on a sample (X_1, \dots, X_n) of dependent random variables with joint probability density (3.49) satisfies

$$2 \left\{ \sum_{i=0}^{n-1} \psi(2\alpha - n - i) - \sum_{i=0}^{n-1} \psi(\alpha - i) - n \log 2 \right\} = - \sum_{i=1}^n \log(1 + X_i^2). \quad (3.53)$$

From Theorem 2.6 we have that $\hat{\alpha}_n$ exists and is unique with probability one for $n \geq 1$ because $P_n(C(X_1, \dots, X_n) \in (-\infty, 0)) = 1$ if $n \geq 1$.

3.8 Discrete Orthogonal Ensembles

This section is devoted to the analysis of examples of discrete distributions from exponential families arising in random matrix theory: the Krawtchouk and Charlier orthogonal polynomial ensembles. Discrete orthogonal polynomials have been investigated in detail by Johansson [72, 73, 74] in connection with the random growth models and the Plancherel measure on partitions. For a survey of models in physics, statistical mechanics, probability theory and combinatorics which are described in terms of orthogonal polynomial ensembles, the reader is referred to König [83]. Since the analytical results and practical applications in the discrete case are discussed exclusively for $\beta = 2$, we restrict our attention to this particular case. For this specific instance, the theory of discrete Riemann-Hilbert problems was developed in Borodin, Boyarchenko [28]. Moreover, the distribution of the first particle was obtained through a certain recurrence procedure for a class of weight functions. We show that the results regarding the distribution function of sufficient statistics are similar to those obtained for ensembles on the real line: sufficient statistics from n -dimensional orthogonal polynomial ensemble appears to be distributed as a random variable with probability density specified by the discrete weight function under consideration, rescaled and shifted appropriately.

3.8.1 Krawtchouk Polynomials Ensemble

In this section we consider a point process with fixed configuration size n on the set $X = \{0, 1, \dots, N\}$, where $2 \leq n \leq N + 1$. The probability distribution is given on

all n -point subsets of X by the probabilities

$$P(x_1, \dots, x_n) = C_{n,N}^{-1}(p) \prod_{1 \leq i < j \leq n} (x_i - x_j)^2 \prod_{i=1}^n \binom{N}{x_i} p^{x_i} (1-p)^{N-x_i}, \quad (3.54)$$

where the normalizing constant is obtained from

$$C_{n,N}(p) = \prod_{i=0}^{n-1} (P_i, P_i)_{w_p},$$

and $P_i(x)$ is the i^{th} normalized Krawtchouk polynomial

$$P_i(x) = (-N)_i p^i {}_2F_1 \left(\begin{matrix} -i, -x \\ -N \end{matrix} \middle| \frac{1}{p} \right).$$

Here, $(N)_i$ is Pochhammer symbol defined as

$$(a)_i = \begin{cases} 1 & \text{if } i = 0, \\ a(a+1) \dots (a+i-1) & \text{if } i \geq 1, \end{cases}$$

for $a \in \mathbb{R}, i \in \mathbb{Z}_{\geq 0}$, while P_i satisfy

$$\begin{aligned} (P_i, P_i)_{w_p} &= \sum_{x=0}^N \binom{N}{x} p^x (1-p)^{N-x} P_i(x)^2 \\ &= (-1)^i i! (-N)_i p^i (1-p)^i, \quad 0 \leq i \leq N. \end{aligned}$$

For detailed information about the system of Krawtchouk orthogonal polynomials, we refer the reader to Borodin, Boyarchenko [28] and references therein. We observe that

$$C_{n,N}(p) = p^{\frac{n(n-1)}{2}} (1-p)^{\frac{n(n-1)}{2}} \prod_{i=1}^{n-1} (-1)^i i! (-N)_i.$$

Rewriting the probabilities in (3.54) as

$$P(x_1, \dots, x_n) = \frac{(1-p)^{Nn}}{C_{n,N}(p)} \prod_{1 \leq i < j \leq n} (x_i - x_j)^2 \prod_{i=1}^n \binom{N}{x_i} \exp \left[x_i \log \frac{p}{1-p} \right],$$

we obtain the equality

$$\frac{C_{n,N}(p)}{(1-p)^{Nn}} = \sum_{\{x_1, \dots, x_n\} \subset \{0, \dots, N\}} \prod_{1 \leq i < j \leq n} (x_i - x_j)^2 \prod_{i=1}^n \binom{N}{x_i} \exp \left[x_i \log \frac{p}{1-p} \right].$$

Taking the logarithms on both sides and differentiating them with respect to p , we get

$$\begin{aligned} \mathbb{E}_n \left[\sum_{i=1}^n X_i \right] &= \frac{n(n-1)}{2} (1-p) - \frac{n(n-1)}{2} p + Nnp \\ &= \frac{n(n-1)}{2} + n(N-n+1)p, \end{aligned}$$

where (X_1, \dots, X_n) is a sample of random variables with the joint probability (3.54). The distribution of the statistic

$$K(X_1, \dots, X_n) = \sum_{i=1}^n X_i$$

can be easily obtained from its moment-generating function

$$\begin{aligned} \mathbb{E}_n [e^{sK(X_1, \dots, X_n)}] &= \sum_{\{x_1, \dots, x_n\} \subset \{0, \dots, N\}} P(x_1, \dots, x_n) e^{s \sum_{i=1}^n x_i} \\ &= e^{sn(n-1)/2} (1-p + pe^s)^{Nn-n(n-1)}, \quad s > 0, \end{aligned}$$

that coincides with the moment-generating function of the random variable following the binomial distribution with the probability of success equal to p and $n(N-n+1)$ trials, shifted by the factor $n(n-1)/2$.

Proposition 3.13. *Consider a statistic $K(\mathbf{X}) = \sum_{i=1}^n X_i$, obtained from a sample $\mathbf{X} = (X_1, \dots, X_n)$ with the joint distribution (3.54). The probability distribution of $K(\mathbf{X})$ is given by*

$$P \left(K(X_1, \dots, X_n) = i + \frac{n(n-1)}{2} \right) = \binom{n(N-n+1)}{i} p^i (1-p)^{n(N-n+1)-i}, \quad (3.55)$$

for $0 \leq i \leq n(N-n+1)$.

The maximum likelihood estimator \hat{p}_n of parameter p for the Krawtchouk unitary ensemble is equal to

$$\hat{p}_n = \frac{\sum_{i=1}^n X_i - n(n-1)/2}{n(N-n-1)}. \quad (3.56)$$

Notice that

$$\text{Var}_n[\hat{p}_n] = \frac{p(1-p)}{n(N-n-1)},$$

and if $N = \kappa n, \kappa > 1$, the order of the variance is $O(n^{-2})$.

3.8.2 Charlier Polynomials Ensemble

Fix a positive integer $n \geq 2$ (the number of particles in the configuration) and consider the probability measure on all n -point subsets of $\mathbb{Z}_{\geq 0}$ given by

$$P(x_1, \dots, x_n) = C_n^{-1}(a) \prod_{1 \leq i < j \leq n} (x_i - x_j)^2 \prod_{i=1}^n \frac{a^{x_i}}{x_i!}, \quad (3.57)$$

where $C_n(a)$ is the normalizing constant and a is a parameter.

Constant $C_n(a)$ can be estimated from the formula

$$C_n(a) = \prod_{i=0}^{n-1} (P_i, P_i)_{w_a},$$

where P_n is the n^{th} normalized Charlier polynomial (the n^{th} monic orthogonal polynomial with respect to the weight $w(x) = a^x/x!$ defined on $\mathbb{Z}_{\geq 0}$),

$$P_n(x) = (-a)^n {}_2F_1\left(-n, -x \mid -\frac{1}{a}\right).$$

Since $(P_n, P_n)_{w_a} = a^n e^a n!$ (see, e.g. Borodin, Boyarchenko [28]), we conclude that

$$C_n(a) = a^{\frac{n(n-1)}{2}} e^{na} \prod_{i=0}^{n-1} i!.$$

From the definition of the joint probability distribution (3.57) it follows that

$$\mathbb{E}_n\left[\sum_{i=1}^n X_i\right] = \frac{n(n-1)}{2} + na,$$

where (X_1, \dots, X_n) is a sample of random variables with the joint probability distribution (3.57).

In order to find the distribution of the sufficient statistic

$$\text{Ch}(X_1, \dots, X_n) = \sum_{i=1}^n X_i, \quad (3.58)$$

we consider its moment generating function, namely

$$\begin{aligned} \mathbb{E}_n[e^{s \text{Ch}(X_1, \dots, X_n)}] &= \sum_{\{x_1, \dots, x_n\} \subset \mathbb{Z}_{\geq 0}} \prod_{1 \leq i < j \leq n} (x_i - x_j)^2 \prod_{i=1}^n \frac{a^{x_i} e^{sx_i}}{x_i!} \\ &= e^{na(e^s - 1)} e^{\frac{sn(n-1)}{2}}, \quad s > 0. \end{aligned}$$

The first term in the product above is the moment generating function of a random variable following the Poisson distribution with parameter an , and the second term is the moment generating function of the point mass at $n(n-1)/2$. Thus, the following proposition is obtained.

Proposition 3.14. *Let $\text{Ch}(X_1, \dots, X_n)$ be a statistic (3.58) obtained from a sample (X_1, \dots, X_n) with the joint distribution (3.57). The probability distribution of $\text{Ch}(X_1, \dots, X_n)$ is*

$$\mathbb{P}(\text{Ch}(X_1, \dots, X_n) = \frac{n(n-1)}{2} + y) = e^{-an} \frac{(an)^y}{y!}, \quad y \in \mathbb{Z}_{\geq 0}. \quad (3.59)$$

The maximum likelihood estimator \hat{a}_n of parameter a is thus given in closed form

$$\hat{a}_n = \sum_{i=1}^n X_i/n - (n-1)/2. \quad (3.60)$$

Its variance is therefore

$$\text{Var}_n[\hat{a}_n] = a/n.$$

If parameter a is proportional to the sample size n , i.e. $a = hn$ for some $h > 0$, the mean spectral measure $\mathbb{E}_n[n^{-1} \sum_{i=1}^n \delta_{X_i/n}]$ converges in distribution to a random variable specified in Lemma 4.4, and the maximum likelihood estimator \hat{h}_n of parameter h is

$$\hat{h}_n = \sum_{i=1}^n X_i/n^2 - (n-1)/(2n). \quad (3.61)$$

Notice that under such a parametrization, the variance of \hat{h}_n is of order $O(n^{-2})$.

Chapter 4

Asymptotical analysis

4.1 Introduction

The asymptotical theory plays the central role in the analysis of random matrix models since the practical applications involve matrices of growing dimension. Proofs of the asymptotical results employ various combinatorial techniques and properties of orthogonal polynomials related to a certain class of weight functions, combined with such elaborated tools of asymptotical analysis as the mean-field method or the steepest descent method of the Riemann-Hilbert problem. Among distinguished features that are analyzed asymptotically, one can identify two groups

- **global properties**, such as the limiting level density (the distribution of randomly chosen eigenvalue) and the fluctuations around the limiting distribution,
- **local properties**, such as the distribution of level spacings, distribution of appropriately rescaled extremal eigenvalue, etc.

The term local refers to the properties of an individual or a constant number of eigenvalues, while global refers to characteristics of the whole spectrum or a significant portion of the eigenvalues. In most cases the analysis is performed for classical values of $\beta = 1, 2, 4$, while results of Johansson [71], Dumitriu [44] and Dumitriu, Edelman [43] cover general β ensembles. We allow β to be arbitrary positive number, and organize this chapter as follows. Section 2 is a review of the main asymptotical results for β -ensembles introduced in Chapter 3. In section 3 we prove the central limit theorem for the sufficient statistics in "classical" β -ensembles, while section 4 is devoted to the derivation of the asymptotical properties of the maximum likelihood estimators $\hat{\theta}_n$. We establish their consistency and asymptotic

normality in the sense that $n(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \Sigma_{\boldsymbol{\theta}_0})$ as $n \rightarrow \infty$, where $\boldsymbol{\theta}_0$ denotes the true parameter value and $\Sigma_{\boldsymbol{\theta}_0}$ is explicitly specified matrix.

4.2 Historical Overview and Motivation

In this section we overview the asymptotical results in RMT. Our attention is restricted to *global* properties since they are of higher relevance for our further analysis. The reader interested in local properties of random matrix spectra is referred to Mehta [87], Tracy, Widom [107, 108, 109], where the properties of level spacings are derived, and to Tracy, Widom [108, 110, 112], Soshnikov [101, 104], Johnstone [76], for the results related to the distribution of extremal eigenvalue from Hermite and Laguerre ensembles.

The nontrivial question in global asymptotic theory of random matrix ensembles is: "What is the right way to plug the dimension parameter into the parameter set in order to ensure convergence of empirical spectral measures to some proper distribution?". For a certain class of models the answer is known (see Deift [37]): the weight function of a model should depend on n as in (1.2). This choice of the model can be explained by the argument of log-potential theory: the external field of a system of n repulsive charges should be proportional to n to assure a stable behavior of the system. In Chapter 5 we show that the naive idea to follow the i.i.d. pattern where n is not plugged into the weight function, fails completely: dependence on parameters disappears in the large n limit, and consistent estimation of parameters is impossible.

The first and the most celebrated global asymptotical result in RMT is due to Wigner [119, 120], who established the weak convergence of spectral distribution for Wigner (in particular, Gaussian hermitian and symmetric) random matrices to the semi-circle law. The result was improved to convergence in probability by Grenander [56], and to almost sure convergence by Arnold [7, 8]. The rate of convergence to the limiting distribution has been considered by several authors including Girko [49, 50], Bai and coauthors [15, 11, 12, 16], Götze, Tikhomirov [51, 52]. Although the asymptotical results mentioned above apply to symmetric or Hermitian Wigner matrices, they are constrained to particular values of $\beta = 1$ or 2 . The general β result for Hermite ensembles was obtained in Dumitriu [44], and formulated as follows.

Theorem 4.1. *Let $\beta > 0$ and $H_{\beta,n}$ be the matrix from β -Hermite (3.6) ensemble scaled by $1/\sqrt{2\beta n}$. Let $F_{\beta,n}(x)$ be the empirical distribution function for the eigenvalues of $H_{\beta,n}$. Then, $F_{\beta,n}(x) \rightarrow S(x)$ a.s. as $n \rightarrow \infty$, where $S(x)$ is the cumulative*

distribution function with the density

$$s(x) = \begin{cases} \frac{2}{\pi} \sqrt{1-x^2}, & x \in [-1, 1], \\ 0, & x \notin [-1, 1]. \end{cases} \quad (4.1)$$

Remark 4.2. The function $s(x)$ is the probability density of the celebrated *Wigner's semi-circle law*. Notice that if we consider β -Hermite ensemble with the varying weight $w_n(x) = \exp(-nx^2/2)$, then the empirical distribution of the eigenvalues rescaled by a constant $1/\sqrt{2\beta}$ converges to the semi-circle law.

The asymptotical analysis of spectral density for Wishart matrices has a long history. The weak convergence of spectral distribution to the *Marčenko-Pastur law* is due to [86]. The result was strengthened to convergence in probability by Jonsson [77], while the almost sure convergence is established in Silverstein, Bai [98]. The rate of convergence is investigated by Bai [14], Goetze, Tikhomirov [53] (see references therein for more detailed historical account). The general β result (for β -Laguerre ensemble) may be found in Dumitriu [44].

The spectral density for a number of discrete orthogonal polynomial ensembles is derived in Ledoux [84]. For the Charlier and Krawtchouk ensembles the following results are valid.

Lemma 4.3. *Let (X_1, \dots, X_n) be a sample of random variables with the joint probability distribution (3.54) of the Krawtchouk orthogonal polynomial ensemble with $0 < p < 1$ and $N = N(n) \sim \kappa n, n \rightarrow \infty, \kappa > 1$. Then the mean spectral measure $E_n [n^{-1} \sum_{i=1}^n \delta_{X_i/n}]$ converges in distribution to*

$$Z = 2\sqrt{p(1-p)U(\kappa-U)}Y + p(\kappa-U) + (1-p)U,$$

where Y has the arcsine law on $(-1, 1)$, and U is uniformly distributed on $[0, 1]$ and independent from Y .

Lemma 4.4. *Let (X_1, \dots, X_n) be a sample of random variables with the joint probability distribution (3.57) of the Charlier polynomial ensemble with $a = a(n) \sim hn, n \rightarrow \infty, h > 0$. Then*

$$E_n [n^{-1} \sum_{i=1}^n \delta_{X_i/n}] \xrightarrow{\mathcal{D}} 2\sqrt{hUY} + h + U$$

as $n \rightarrow \infty$, where Y has the arcsine law on $(-1, 1)$, and U is uniformly distributed on $[0, 1]$ and independent from Y .

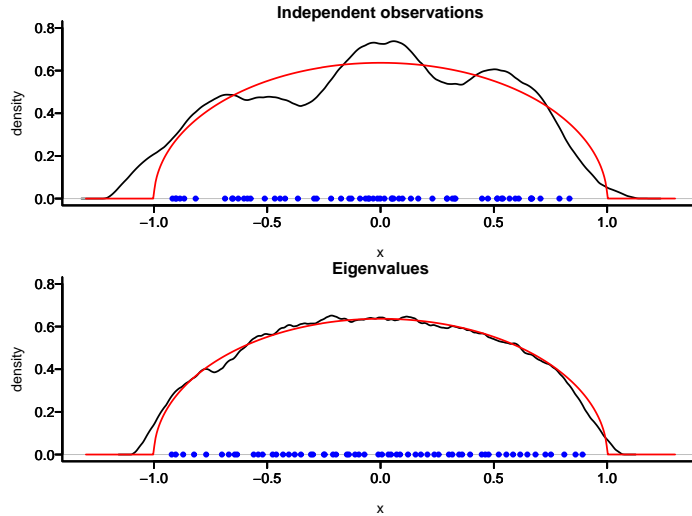


Figure 4.1: The probability density of Wigner's semi-circle law (red line). The independent observations (upper plot), (rescaled) spectrum of a random matrix from GUE (lower plot), and the kernel estimators of their probability densities (black line).

One of the key issues in global asymptotics of random matrix models is the *central limit theorem* for linear statistics of eigenvalues $F(\mathbf{X}) = \sum_{i=1}^n f(X_i)$. The random variable $F(\mathbf{X})$ is called a *linear statistics* since products of different eigenvalues X_i do not appear, although the function $f(\cdot)$ may depend nonlinearly on its argument. The distribution of linear statistics $F(\mathbf{X})$ reflects the *fluctuations* of random spectra around the limiting distribution. To illustrate this fact, we consider β -Hermite ensemble and the limiting density $s(x)$ of Wigner's semi-circle law. Convergence in Theorem 4.1 implies that for arbitrary bounded $f \in C(\mathbb{R})$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f(X_i) = \int_{-1}^1 f(t) s(t) dt, \quad a.s. \quad (4.2)$$

but does not provide us with the information about fluctuations of random variable $\sum_{i=1}^n f(X_i)$. Remarkably, it appears that the variance of statistics $\sum_{i=1}^n f(X_i)$ does not grow to infinity if function $f(\cdot)$ satisfies certain regularity conditions (see Johansson [71] for general β , Soshnikov [103] for $\beta = 2$). The fact that the variance of $F(\mathbf{X})$ remains bounded when n tends to infinity can be explained by the strong repulsiveness of eigenvalues, which results in effective cancelations in the sum $\sum_{i=1}^n f(X_i)$. To stress the difference from the situation when a sample of i.i.d. random variables is considered, we propose the following setting: let $\mathbf{Y} = (Y_1, \dots, Y_n)$ be the sample

of independent observations from Wigner's semi-circle law and $\mathbf{X} = (X_1, \dots, X_n)$ be the eigenvalues of Gaussian unitary (2-Hermitic) ensemble scaled by $1/2\sqrt{n}$ (notice that from Theorem 4.1 we have that limiting densities of \mathbf{X} and \mathbf{Y} coincide as $n \rightarrow \infty$). Then, from the central limit theorem for i.i.d random variables we have that as $n \rightarrow \infty$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1/4), \quad n \rightarrow \infty \quad (4.3)$$

while

$$\sum_{i=1}^n X_i \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1/4), \quad n \rightarrow \infty. \quad (4.4)$$

Notice that $\sum_{i=1}^n X_i$ is *not normalized* by $1/\sqrt{n}$. The difference between the samples \mathbf{X} and \mathbf{Y} for $n = 50$ is illustrated in Figure 4.1. Although the limiting density for both samples is Wigner's semi-circle, X_i 's are more regularly spaced on the support $[-1, 1]$, and this fact results in different scalings in central limit theorems (4.3) and (4.4).

The fluctuation properties of the random matrix spectra influence asymptotical behavior of the maximum likelihood estimators discussed in Chapter 3. Namely, it will be shown that the asymptotical variance of MLE is of order $O(1/n^2)$, compared to $O(1/n)$ when the sample of i.i.d. random variables is considered. In Section 3 we prove the CLT for the complete sufficient statistics in classical random matrix ensembles, while Section 4 is devoted to discussion of the asymptotical properties of the maximum likelihood estimators. Throughout this chapter we consider the models with *varying weight*, i.e. we will assume that a weight function depends on the dimension parameter n .

We conclude this section with the list of references where CLT for linear statistics in various random matrix models and global fluctuations of eigenvalues of large random matrices are discussed. Heuristical derivations of Gaussian limit for linear statistics may be found in Politzer [95], Beenakker [24, 25], Costin, Lebowitz [36] while some of the references where rigorous analysis is performed include Johansson [71], Basor [22], Sinai, Soshnikov [100], Soshnikov [103], Bai, Silverstein [13] and Dumitriu, Edelman [43]. It should be noticed that similar results are obtained for spectrum fluctuations of matrices from classical compact groups, but they will be discussed in detail in Chapter 5.

4.3 Central Limit Theorem for Sufficient Statistics

4.3.1 β -Hermite Ensemble

Consider the joint probability density of eigenvalues of β -Hermite ensemble

$$p_{\beta,n}(\mathbf{x}) = C_{\beta,n}(\sigma)^{-1} \prod_{1 \leq i < j \leq n} |x_i - x_j|^\beta \prod_{i=1}^n \exp \left\{ -\frac{nx_i^2}{2\sigma^2} \right\}, \quad \mathbf{x} \in \mathbb{R}^n, \quad (4.5)$$

with scaling parameter dependent on n and $n \geq 2$. The result that we state here has been established in Johansson [71] and Dumitriu, Edelman [43] in a broader context.

Proposition 4.5. *Let $\mathbf{X} = (X_1, \dots, X_n)$ be the sample with joint probability density (4.5). The random variable*

$$H_\beta(\mathbf{X}) = \sum_{i=1}^n X_i^2 - \sigma^2 \left[\frac{\beta}{2}(n-1) + 1 \right] \quad (4.6)$$

converges weakly to the normal distribution with 0 mean and variance $\beta\sigma^4$ as $n \rightarrow \infty$.

Proof. From the expression for the normalizing constant $C_{\beta,n}(\sigma)$, we have that the characteristic function of the random variable $H_\beta(\mathbf{X})$ is

$$\begin{aligned} \phi_n(t) &= \exp \left\{ it\sigma^2 \left[\frac{\beta}{2}(n-1) + 1 \right] \right\} \left(\frac{n - 2it\sigma^2}{n} \right)^{-n(\beta(n-1)/2 + 1)/2} \\ &= \exp \left\{ it\sigma^2 \left[\frac{\beta}{2}(n-1) + 1 \right] - \frac{n}{2} \left[\frac{\beta}{2}(n-1) + 1 \right] \log \left(1 - \frac{2it\sigma^2}{n} \right) \right\} \\ &= \exp \left\{ it\sigma^2 \left[\frac{\beta}{2}(n-1) + 1 \right] - \frac{n}{2} \left[\frac{\beta}{2}(n-1) + 1 \right] \left(-\frac{2it\sigma^2}{n} - \frac{(it)^2 4\sigma^4}{2n^2} + O\left(\frac{t^3}{n^3}\right) \right) \right\} \\ &= \exp \left\{ -\frac{t^2}{2} \frac{2\sigma^4 n [\beta(n-1)/2 + 1]}{n^2} + O\left(\frac{t^3}{n}\right) \right\}, \end{aligned}$$

and consequently

$$\lim_{n \rightarrow \infty} \phi_n(t) = \exp \left\{ -\frac{t^2}{2} \sigma^4 \beta \right\}.$$

Thus, the asymptotic distribution of the statistic (4.6) is normal with 0 mean and

variance $\beta\sigma^4$. □

4.3.2 β -Laguerre Ensemble

The proofs of asymptotical normality for minimal sufficient statistics of β -Laguerre and β -Jacobi ensemble involve asymptotical result for sums of trigamma function. The result is formulated as proposition.

Proposition 4.6. *Let $\alpha, \beta \geq 0$, $n \in \mathbb{N}$ and define*

$$f_n(\alpha, \beta) = \sum_{j=0}^{n-1} \psi'(\alpha n + j\beta/2), \quad (4.7)$$

where $\psi'(x)$ is the second derivative of $\log \Gamma(x)$. Then

$$\lim_{n \rightarrow \infty} f_n(\alpha, \beta) = \frac{2}{\beta} \log \left(1 + \frac{\beta}{2\alpha} \right). \quad (4.8)$$

Proof. Using the asymptotic expansion for $\psi'(x)$ (see Gradshteyn, Ryzhik [54])

$$\psi'(x) = \frac{1}{x} + \frac{1}{2x^2} + o\left(\frac{1}{x^2}\right),$$

we obtain that

$$\begin{aligned} f_n(\alpha, \beta) &= \sum_{j=0}^{n-1} \frac{1}{\alpha n + \beta j/2} + o(1) \\ &= \frac{2}{\beta} \log \left(1 + \frac{\beta}{2\alpha} \right) + o(1), \end{aligned}$$

and consequently

$$\lim_{n \rightarrow \infty} f_n(\alpha, \beta) = \frac{2}{\beta} \log \left(1 + \frac{\beta}{2\alpha} \right). \quad \square$$

Consider the probability density of scaled β -Laguerre ensemble

$$p_\beta(\mathbf{x}; \alpha, \theta) = C_{n,\beta}^{-1}(\alpha, \theta) \prod_{1 \leq i < j \leq n} |x_i - x_j|^\beta \prod_{i=1}^n x_i^{\alpha n - 1} e^{-n x_i / \theta}, \quad \mathbf{x} \in (0, \infty)^n, \quad (4.9)$$

where $\alpha, \theta > 0, n \geq 2$. Notice that both shape and scale parameter depend on dimension parameter n . We examine the asymptotical distribution of the sufficient vector statistic and formulate the result as lemma.

Lemma 4.7. *Let $\mathbf{X} = (X_1, \dots, X_n)$ be a sample with the joint probability density (4.9) and define the vector statistics $\mathbf{L}(\mathbf{X}) = (L_1(\mathbf{X}), L_2(\mathbf{X}))$, where*

$$L_1(\mathbf{X}) = \sum_{j=1}^n X_j - \theta[\alpha n + (n-1)\beta/2], \quad (4.10)$$

$$L_2(\mathbf{X}) = \sum_{j=1}^n \log X_j - \sum_{j=0}^{n-1} \psi(\alpha n + j\beta/2) - n \log \frac{\theta}{n}. \quad (4.11)$$

The distribution of the random vector $\mathbf{L}(\mathbf{X})$ converges weakly to the centered normal distribution with the covariance matrix

$$\Sigma_{\theta, \alpha} = \begin{pmatrix} \theta^2(\alpha + \beta/2) & \theta \\ \theta & \frac{2}{\beta} \log \left(1 + \frac{\beta}{2\alpha}\right) \end{pmatrix}. \quad (4.12)$$

Proof. Consider the characteristic function of $\mathbf{L}(\mathbf{X})$

$$\begin{aligned} \phi_n(t_1, t_2) &= E_n \exp\{it_1 L_1(\mathbf{X}) + it_2 L_2(\mathbf{X})\} \\ &= I_1^{(n)}(t_1, t_2) I_2^{(n)}(t_1, t_2) I_3^{(n)}(t_1, t_2), \end{aligned}$$

with

$$\begin{aligned} I_1^{(n)}(t_1, t_2) &:= \frac{\left(\frac{\theta}{n-it_1\theta}\right)^{n[\alpha n + \beta(n-1)/2 + it_2]}}{\left(\frac{\theta}{n}\right)^{n[\alpha n + \beta(n-1)/2]}}, \\ I_2^{(n)}(t_1, t_2) &:= \prod_{j=0}^{n-1} \frac{\Gamma(\alpha n + j\beta/2 + it_2)}{\Gamma(\alpha n + j\beta/2)}, \\ I_3^{(n)}(t_1, t_2) &:= \exp\{-it_1\theta[\beta(n-1)/2 + \alpha n]\} \\ &\quad \times \exp\left\{-it_2 \sum_{j=0}^{n-1} \psi(\alpha n + j\beta/2) - it_2 n \log(\theta/n)\right\}. \end{aligned}$$

Notice that

$$I_1^{(n)}(t_1, t_2) = \left(\frac{\theta}{n}\right)^{it_2 n} \left(\frac{n-it_1\theta}{n}\right)^{-n[\alpha n + \beta(n-1)/2 + it_2]}$$

$$\begin{aligned}
&= \exp \left\{ it_2 n \log \left(\frac{\theta}{n} \right) - n \left[\alpha n + \frac{\beta}{2}(n-1) + it_2 \right] \log \left(1 - \frac{it_1 \theta}{n} \right) \right\} \\
&= \exp \left\{ it_2 n \log \left(\frac{\theta}{n} \right) - n \left[\alpha n + \frac{\beta}{2}(n-1) + it_2 \right] \left(-\frac{it_1 \theta}{n} + \frac{t_1^2 \theta^2}{2n^2} + O \left(\frac{t_1^3}{n^3} \right) \right) \right\},
\end{aligned}$$

and that the Taylor expansion for $\log \Gamma(x)$ gives us

$$\begin{aligned}
I_2^{(n)}(t_1, t_2) &= \exp \left\{ \sum_{j=0}^{n-1} \log \Gamma(\alpha n + j\beta/2 + it_2) - \sum_{j=0}^{n-1} \log \Gamma(\alpha n + j\beta/2) \right\} \\
&= \exp \left\{ it_2 \sum_{j=0}^{n-1} \psi(\alpha n + j\beta/2) - \frac{t_2^2}{2} \sum_{j=0}^{n-1} \psi'(\alpha n + j\beta/2) + o(t_2^2) \right\}.
\end{aligned}$$

The last two expressions and Proposition 4.6 result in

$$\lim_{n \rightarrow \infty} \phi_n(t_1, t_2) = \exp \left\{ -\frac{t_1^2}{2} \theta^2 \left(\alpha + \frac{\beta}{2} \right) - \frac{t_2^2}{2} \frac{2}{\beta} \log \left(1 + \frac{\beta}{2\alpha} \right) - t_1 t_2 \theta \right\}.$$

Since the convergence of characteristic functions is equivalent to the weak convergence of distributions, we obtain that the asymptotic distribution of $\mathbf{L}(\mathbf{X})$ is normal with the covariance matrix as specified in (4.12). \square

4.3.3 β -Jacobi Ensemble

Consider the probability density of β -Jacobi ensemble with varying weight

$$p_\beta(\mathbf{x}; a, b) = C_{n,\beta}^{-1}(a, b) \prod_{1 \leq i < j \leq n} |x_i - x_j|^\beta \prod_{i=1}^n x_i^{a-1} (1 - x_i)^{b-1}, \quad \mathbf{x} \in [0, 1]^n, \quad (4.13)$$

where $C_{n,\beta}(a, b)$ is the normalizing constant, $a, b > 0, n \geq 2$. We prove the following lemma.

Lemma 4.8. *Let $\mathbf{X} = (X_1, \dots, X_n)$ be a sample with the joint probability distribution and define the vector statistic $\mathbf{J}(\mathbf{X}) = (J_1(\mathbf{X}), J_2(\mathbf{X}))$ by*

$$J_1(\mathbf{X}) = \sum_{j=1}^n \log X_j - \mathbb{E}_n \left[\sum_{j=1}^n \log X_j \right], \quad (4.14)$$

$$J_2(\mathbf{X}) = \sum_{j=1}^n \log(1 - X_j) - \mathbb{E}_n \left[\sum_{j=1}^n \log(1 - X_j) \right]. \quad (4.15)$$

The distribution of the random vector $\mathbf{J}(\mathbf{X})$ converges in distribution to the centered normal distribution with the covariance matrix

$$\Sigma_{(a,b)} = \begin{pmatrix} \frac{2}{\beta} \log \frac{(2a+\beta)(2a+2b+\beta)}{2a(2a+2b+2\beta)} & -\frac{2}{\beta} \log \frac{2a+2b+2\beta}{2a+2b+\beta} \\ -\frac{2}{\beta} \log \frac{2a+2b+2\beta}{2a+2b+\beta} & \frac{2}{\beta} \log \frac{(2b+\beta)(2a+2b+\beta)}{2b(2a+2b+2\beta)} \end{pmatrix}. \quad (4.16)$$

Proof. The characteristic function of the random vector $\mathbf{J}(\mathbf{X})$ is

$$\begin{aligned} \phi_n(t_1, t_2) &= \mathbb{E}_n \exp \{it_1 J_1(\mathbf{X}) + it_2 J_2(\mathbf{X})\} \\ &= I_1^{(n)}(t_1, t_2) I_2^{(n)}(t_1, t_2) I_3^{(n)}(t_1, t_2) I_4^{(n)}(t_1, t_2), \end{aligned}$$

where

$$\begin{aligned} I_1^{(n)}(t_1, t_2) &:= \prod_{j=0}^{n-1} \frac{\Gamma(an + it_1 + j\beta/2)}{\Gamma(an + j\beta/2)}, \\ I_2^{(n)}(t_1, t_2) &:= \prod_{j=0}^{n-1} \frac{\Gamma(bn + it_2 + j\beta/2)}{\Gamma(bn + j\beta/2)}, \\ I_3^{(n)}(t_1, t_2) &:= \prod_{j=0}^{n-1} \frac{\Gamma(an + bn + (n-1+j)\beta/2)}{\Gamma(an + bn + i(t_1 + t_2) + (n-1+j)\beta/2)}, \\ I_4^{(n)}(t_1, t_2) &:= \exp \left\{ -it_1 \mathbb{E}_n \left[\sum_{j=1}^n \log X_j \right] - it_2 \mathbb{E}_n \left[\sum_{j=1}^n \log(1 - X_j) \right] \right\} \end{aligned}$$

From the Taylor expansion for $\log \Gamma(x)$, we have the following asymptotic expressions

$$\begin{aligned} I_1^{(n)}(t_1, t_2) &= \exp \left\{ it_1 \sum_{j=0}^{n-1} \psi(an + j\beta/2) - \frac{t_1^2}{2} \sum_{j=0}^{n-1} \psi'(an + j\beta/2) + o(t_1^2) \right\}, \\ I_2^{(n)}(t_1, t_2) &= \exp \left\{ it_2 \sum_{j=0}^{n-1} \psi(bn + j\beta/2) - \frac{t_2^2}{2} \sum_{j=0}^{n-1} \psi'(bn + j\beta/2) + o(t_2^2) \right\}, \\ I_3^{(n)}(t_1, t_2) &= \exp \left\{ -i(t_1 + t_2) \sum_{j=0}^{n-1} \psi(an + bn + (n-1+j)\beta/2) \right\} \\ &\quad \times \exp \left\{ \frac{(t_1 + t_2)^2}{2} \sum_{j=0}^{n-1} \psi'(an + bn + (n-1+j)\beta/2) + o((t_1 + t_2)^2) \right\}. \end{aligned}$$

Consequently, from Proposition 4.6 and asymptotic expansions above, it follows that

$$\begin{aligned} \lim_{n \rightarrow \infty} \phi_n(t_1, t_2) &= \exp \left\{ -\frac{t_1^2}{2} \frac{2}{\beta} \left[\log \left(1 + \frac{\beta}{2a} \right) - \log \left(1 + \frac{\beta}{2a + 2b + \beta} \right) \right] \right\} \\ &\times \exp \left\{ -\frac{t_2^2}{2} \frac{2}{\beta} \left[\log \left(1 + \frac{\beta}{2b} \right) - \log \left(1 + \frac{\beta}{2a + 2b + \beta} \right) \right] \right\} \\ &\times \exp \left\{ t_1 t_2 \frac{2}{\beta} \log \left(1 + \frac{\beta}{2a + 2b + \beta} \right) \right\}, \end{aligned}$$

and the convergence of $\phi_n(t_1, t_2)$ to characteristic function of bivariate normal distribution with 0 mean and covariance matrix (4.16) is established. \square

4.3.4 Cauchy Ensemble

Consider the Cauchy unitary ensemble with the probability density function

$$p(\mathbf{x}; \alpha) = C_n^{-1}(\alpha) \prod_{1 \leq i < j \leq n} |x_i - x_j|^2 \prod_{i=1}^n (1 + x_i^2)^{-\alpha n}, \quad \mathbf{x} \in \mathbb{R}^n, \quad (4.17)$$

with a varying parameter $\alpha n > n - 1/2$. The reasoning similar to the derivation of CLT for sufficient statistics in Laguerre and Jacobi ensembles leads to the following proposition.

Proposition 4.9. *Let (X_1, \dots, X_n) be a sample of exchangeable random variables with the joint probability density (4.17). Then*

$$C(\mathbf{X}) = \sum_{i=1}^n \log(1 + X_i^2) - E_n \left[\sum_{i=1}^n \log(1 + X_i^2) \right] \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_\alpha^2), \quad n \rightarrow \infty, \quad (4.18)$$

where

$$\sigma_\alpha^2 = 4 \log \left(1 + \frac{1}{2(\alpha - 1)} \right) - 2 \log \left(1 + \frac{1}{\alpha - 1} \right), \quad \alpha > 1. \quad (4.19)$$

Proof. The characteristic function of the statistic $\sum_{i=1}^n \log(1 + X_i^2)$ is equal to

$$\tilde{\phi}_n(t) = 2^{itn} \prod_{j=0}^{n-1} \frac{\Gamma(2(\alpha n - it) - n - j)}{\Gamma(2\alpha n - n - j)} \frac{[\Gamma(\alpha n - j)]^2}{[\Gamma(\alpha n - it - j)]^2}.$$

The Taylor expansion of the right hand side and Proposition 4.6 result in

$$\lim_{n \rightarrow \infty} \phi_n(t) = \exp \left\{ -\frac{t^2}{2} \left(4 \log \left(1 + \frac{1}{2(\alpha - 1)} \right) - 2 \log \left(1 + \frac{1}{\alpha - 1} \right) \right) \right\}, \quad t \in \mathbb{R},$$

where $\phi_n(t)$ denotes the characteristic function of the statistic $C(\mathbf{X})$. \square

4.4 Maximum Likelihood Estimation

4.4.1 β -Hermite Ensemble

Since for β -Hermite ensemble the maximum likelihood estimator of σ^2 is available in closed form, and its asymptotical properties can be derived from the properties of the sufficient statistics $\sum_{i=1}^n X_i^2$, we formulate the following proposition.

Proposition 4.10. *Let (X_1, \dots, X_n) be a sample of random variables with the joint probability density of β -Hermite ensemble (4.5) with the true parameter value σ_0^2 . The maximum likelihood estimator $\hat{\sigma}_n^2$ is consistent and asymptotically normal, i.e.*

$$n(\hat{\sigma}_n^2 - \sigma_0^2) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 4\sigma_0^4/\beta). \quad n \rightarrow \infty. \quad (4.20)$$

Proof. The consistency of $\hat{\sigma}_n^2$ follows immediately from the expressions (3.26, 3.27) for the mean and the variance of the maximum likelihood estimator. From Proposition 4.5 we have that

$$\sum_{i=1}^n X_i^2 - \sigma_0^2 \left[\frac{\beta}{2}(n-1) + 1 \right] \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_0^4\beta), \quad n \rightarrow \infty.$$

Since

$$\hat{\sigma}_n^2 = \frac{\sum_{i=1}^n X_i^2}{\beta(n-1)/2 + 1},$$

we obtain the convergence (4.20). \square

In Figure 4.2 we illustrate the result of Proposition 4.10 in the qq-plot. The empirical quantiles of $n(\hat{\sigma}_n^2 - \sigma_0^2)$ are compared with the theoretical quantiles of the limiting normal distribution for β -Hermite ensemble with $\beta = 1.2$, $\sigma_0^2 = 4$, $n = 50$.

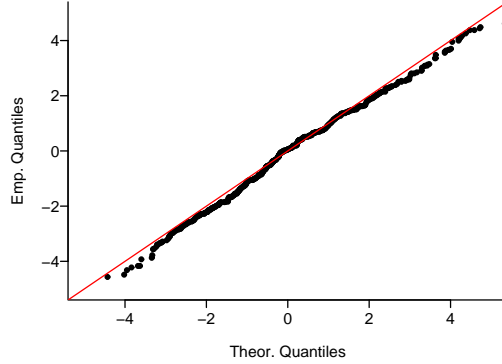


Figure 4.2: The asymptotic normality of the maximum likelihood estimator $\hat{\sigma}_n^2$ in the scaled β -Hermite ensemble with $\beta = 1.2$, $\sigma_0^2 = 4$, $n = 50$.

4.4.2 β -Laguerre Ensemble

In order to prove consistency and asymptotical efficiency of the MLE for β -Laguerre and β -Jacobi ensemble, we will use the unified approach described below. Namely, consider the probability density function

$$p_\beta(\mathbf{x}; \boldsymbol{\theta}) = C_{\beta, n}^{-1}(\boldsymbol{\theta}) \prod_{1 \leq i < j \leq n} |x_i - x_j|^\beta \prod_{i=1}^n w_{n, \boldsymbol{\theta}}(x_i), \quad (4.21)$$

with $\boldsymbol{\theta} = (\theta_1, \theta_2) \in \mathbb{R}_+^2$ and

$$w_{n, \boldsymbol{\theta}}(x) = \begin{cases} e^{-nx/\theta_1} x^{n\theta_2-1}, & x \in \mathbb{R}_+, & \text{Laguerre,} \\ x^{n\theta_1-1} (1-x)^{n\theta_2-1}, & x \in (0, 1), & \text{Jacobi.} \end{cases} \quad (4.22)$$

The true parameter value will be denoted by $\boldsymbol{\theta}^0$. We will consider the behavior of the log-likelihood function $l(\boldsymbol{\theta}, \mathbf{X})$ on the sphere Q_r of radius r with center at the true parameter $\boldsymbol{\theta}^0$, and show that for sufficiently small r and $\boldsymbol{\theta} \in Q_r$, the probability

$$P_{\boldsymbol{\theta}^0}(l(\boldsymbol{\theta}^0, \mathbf{X}) > l(\boldsymbol{\theta}, \mathbf{X})) \rightarrow 1, \quad n \rightarrow \infty.$$

Consequently, $l(\boldsymbol{\theta})$ has a local maximum in the interior of Q_r . Since the maximum likelihood equations must be satisfied at the local maximum and it was shown that the solution of the maximum likelihood equation is unique, it will follow that with probability tending to 1 as $n \rightarrow \infty$ for any $r > 0$ sufficiently small, the maximum

likelihood equations have the unique solution $\hat{\boldsymbol{\theta}}_n$ within Q_r , and hence

$$P_{\boldsymbol{\theta}^0}(\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0\| < r) \rightarrow 1, \quad n \rightarrow \infty.$$

In the proof of the asymptotic efficiency we utilize the following lemma.

Lemma 4.11 (Lehmann, Casella [85]). *Let (T_{1n}, \dots, T_{sn}) be a sequence of random vectors converging in distribution to (T_1, \dots, T_s) , and suppose that for arbitrary fixed j and k , A_{jkn} is a sequence of random variables tending in probability to constants a_{jk} for which the matrix $A = \|a_{jk}\|$ is nonsingular. Let $B = \|b_{jk}\| = A^{-1}$. Then, if the distribution of (T_1, \dots, T_s) has a density with respect to Lebesgue measure over E_s , the solutions (Y_{1n}, \dots, Y_{sn}) of the system of random linear equations in s unknowns*

$$\sum_{k=1}^s A_{jkn} Y_{kn} = T_{jn}, \quad j \in \{1, \dots, s\}, \quad (4.23)$$

tend in probability to the solutions (Y_1, \dots, Y_s) of

$$\sum_{k=1}^s a_{jk} Y_k = T_j, \quad j \in \{1, \dots, s\}, \quad (4.24)$$

given by

$$Y_j = \sum_{k=1}^s b_{jk} T_k, \quad j \in \{1, \dots, s\}. \quad (4.25)$$

In the following lemma we show that the maximum likelihood estimators of the scale and shape parameter for β -Laguerre ensemble with varying coefficients are consistent and asymptotically efficient.

Lemma 4.12. *Let $\beta > 0$ and $\mathbf{X} = (X_1, \dots, X_n)$ be a sample of exchangeable random variables with the joint probability density of β -Laguerre ensemble with varying coefficients (4.9) and the true parameter value (α_0, θ_0) . The maximum likelihood estimator $(\hat{\alpha}_n, \hat{\theta}_n)$ is consistent and asymptotically efficient in the sense that*

$$n(\hat{\alpha}_n - \alpha_0, \hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_{(\alpha_0, \theta_0)}), \quad (4.26)$$

with

$$\Sigma_{(\alpha_0, \theta_0)} = \begin{pmatrix} \frac{2}{\beta} \log \left(1 + \frac{\beta}{2\alpha_0} \right) & 1/\theta_0 \\ 1/\theta_0 & (\alpha_0 + \beta/2)/\theta_0^2 \end{pmatrix}^{-1}. \quad (4.27)$$

Proof. First we prove consistency. To obtain the needed facts concerning the behavior of the log-likelihood function $l_n(\alpha, \theta)$ on the sphere Q_r of radius r with center at

the true point (α_0, θ_0) , we will use the Taylor expansion about (α_0, θ_0) and divide it by n^2

$$\frac{1}{n^2} l_n(\alpha, \theta) - \frac{1}{n^2} l_n(\alpha_0, \theta_0) = I_1^{(n)}(\alpha, \theta) + I_2^{(n)}(\alpha, \theta) + I_3^{(n)}(\alpha, \theta),$$

with

$$I_1^{(n)}(\alpha, \theta) := (\theta - \theta_0) \left\{ \frac{1}{n} L_1(\mathbf{X}; \alpha_0, \theta_0) \right\} \\ + (\alpha - \alpha_0) \left\{ \frac{1}{n} L_2(\mathbf{X}; \alpha_0, \theta_0) \right\},$$

$$I_2^{(n)}(\alpha, \theta) := \frac{1}{2}(\theta - \theta_0)^2 \left\{ \frac{1}{n} \frac{1}{\theta_0^2} [\beta(n-1)/2 + \alpha_0 n] - \frac{1}{n} \frac{2}{\theta_0^3} \sum_{i=1}^n X_i \right\} \\ + \frac{1}{2}(\alpha - \alpha_0)^2 \left\{ - \sum_{j=0}^{n-1} \psi'(\alpha_0 n + j\beta/2) \right\} + (\alpha - \alpha_0)(\theta - \theta_0) \left\{ -\frac{1}{\theta_0} \right\},$$

$$I_3^{(n)}(\alpha, \theta) := \frac{1}{6}(\alpha - \alpha_0)^3 \left\{ -n \sum_{j=0}^{n-1} \psi''(\alpha_* n + j\beta/2) \right\} \\ + \frac{1}{6}(\theta - \theta_0)^3 \left\{ \frac{6}{n} \frac{1}{\theta_*^4} \sum_{i=1}^n X_i - \frac{2}{n} \frac{1}{\theta_*^3} [\beta(n-1)/2 + \alpha_* n] \right\} \\ + \frac{1}{2}(\theta - \theta_0)^2 (\alpha - \alpha_0) \left\{ \frac{1}{\theta_*^2} \right\},$$

where (α_*, θ_*) is a point on the line segment connecting (α, θ) and (α_0, θ_0) . We notice that from Lemma 4.7 it follows

$$\frac{1}{n} \mathbf{L}(\mathbf{X}; \alpha_0, \theta_0) \rightarrow 0, \quad n \rightarrow \infty$$

in probability, and consequently

$$\frac{1}{n} |L_1(\mathbf{X}; \alpha_0, \theta_0)| + \frac{1}{n} |L_2(\mathbf{X}; \alpha_0, \theta_0)| < r^2$$

with probability tending to 1. Since the norms $\|\mathbf{x}\|_1 = |x_1| + |x_2|$ and $\|\mathbf{x}\|_2 = (|x_1|^2 + |x_2|^2)^{1/2}$ are equivalent in \mathbb{R}^2 , we have that there exists $s > 0$ such that with

probability tending to 1 and arbitrary r

$$|I_1^{(n)}(\alpha, \theta)| \leq sr^3.$$

Further, from Lemma 4.7 we have that the term

$$\frac{1}{n} \frac{1}{\theta_0^2} [\beta(n-1)/2 + \alpha_0 n] - \frac{1}{n} \frac{2}{\theta_0^3} \sum_{i=1}^n X_i \rightarrow -\frac{\alpha_0 + \beta/2}{\theta_0^2}, \quad n \rightarrow \infty$$

in probability, and hence

$$I_2^{(n)}(\alpha, \theta) = I_{21}^{(n)}(\alpha, \theta) + I_{22}^{(n)}(\alpha, \theta),$$

where

$$I_{21}^{(n)}(\alpha, \theta) := -\frac{1}{2}(\theta - \theta_0)^2 \frac{\alpha_0 + \beta/2}{\theta_0^2} - (\theta - \theta_0)(\alpha - \alpha_0) \frac{1}{\theta_0} - \frac{1}{2}(\alpha - \alpha_0)^2 \sum_{j=1}^{n-1} \psi(\alpha_0 n + j\beta/2)$$

is negative (nonrandom) quadratic form, while

$$I_{22}^{(n)}(\alpha, \theta) := -\frac{2}{n\theta_0^2}(\theta - \theta_0)^2 L_1(\mathbf{X}; \alpha_0, \theta_0)$$

converges to 0 in probability. Consequently, there exist constants $C > 0$ and $r_0 > 0$ such that for every $r < r_0$

$$I_2^{(n)}(\alpha, \theta) \leq -Cr^2.$$

Finally, with probability tending to 1, there exists a constant $C_1 > 0$ such that

$$|I_3^{(n)}(\alpha, \theta)| \leq C_1 s^3 r^3 / 6$$

since $n \sum_{j=0}^{n-1} \psi''(\alpha_* n + j\beta/2)$ is bounded for α_* between α and α_0 , and

$$\begin{aligned} & \frac{6}{n} \frac{1}{\theta_*^4} \sum_{i=1}^n X_i - \frac{2}{n} \frac{1}{\theta_*^3} [\beta(n-1)/2 + \alpha_* n] \\ & \leq \frac{6}{\theta_*^4} \left| \frac{1}{n} \sum_{i=1}^n X_i \right| + \frac{2}{n} \frac{1}{\theta_*^3} [\alpha_* n + \beta(n-1)/2] \\ & \leq C_2 \end{aligned}$$

for n large enough and (α_*, θ_*) lying in the segment connecting (α, θ) and (α_0, θ_0) .

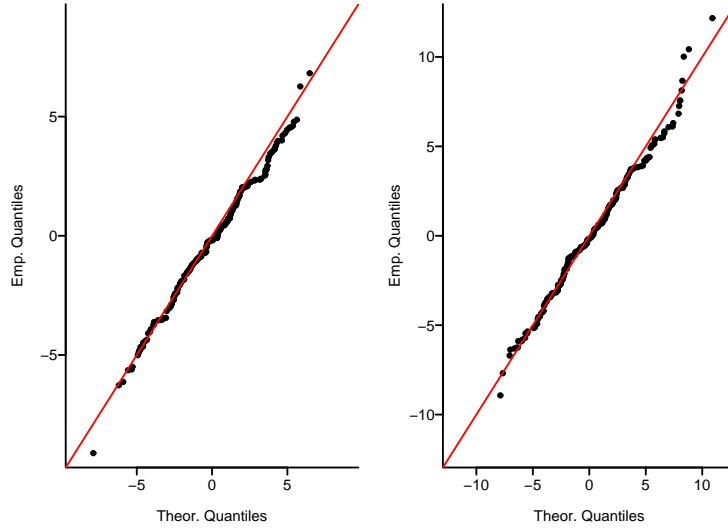


Figure 4.3: The asymptotic efficiency of the maximum likelihood estimators $\hat{\alpha}_n$ (left plot), $\hat{\theta}_n$ (right plot) in β -Laguerre ensemble with $\beta = 1.5, \alpha = 1.25, \theta = 2.4, n = 1000$.

Since

$$\max_{(\alpha, \theta) \in Q_r} \left\{ I_1^{(n)}(\alpha, \theta) + I_2^{(n)}(\alpha, \theta) + I_3^{(n)}(\alpha, \theta) \right\} \leq (s + C_1 s^3/6)r^3 - Cr^2 < 0$$

with probability tending to 1, if $r < C(s + C_1 s^3/6)^{-1}$, we obtain that

$$\lim_{n \rightarrow \infty} P_{(\alpha_0, \theta_0)}(l_n(\alpha_0, \theta_0; \mathbf{X}) > l_n(\alpha, \theta; \mathbf{X})) = 1, \quad \forall (\alpha, \theta) \in Q_r$$

and therefore $l_n(\alpha, \theta; \mathbf{X})$ has a local maximum $(\hat{\alpha}_n, \hat{\theta}_n)$ inside Q_r for every r small enough. Since the maximum likelihood equations have to be satisfied at the point $(\hat{\alpha}_n, \hat{\theta}_n)$ and it was shown that the solution of the maximum likelihood equations is unique for β -Laguerre ensemble, we have that

$$\lim_{n \rightarrow \infty} P_{(\alpha_0, \theta_0)}(|\hat{\theta}_n - \theta_0|^2 + |\hat{\alpha}_n - \alpha_0|^2 < r^2) = 1,$$

and the consistency of the maximum likelihood estimator $(\hat{\alpha}_n, \hat{\theta}_n)$ follows.

To prove asymptotic efficiency we consider the partial derivatives of the log-

likelihood function and their Taylor expansions

$$\begin{aligned} \frac{\partial}{\partial \theta} l_n(\hat{\alpha}_n, \hat{\theta}_n; \mathbf{X}) &= \frac{n}{\theta_0^2} L_1(\mathbf{X}; \alpha_0, \theta_0) + (\hat{\theta}_n - \theta_0) \left\{ -\frac{2n}{\theta_0^3} \sum_{i=1}^n X_i + \frac{n}{\theta_0^2} [\beta(n-1)/2 + \alpha_0 n] \right\} \\ &\quad - \frac{n}{\theta_0} (\hat{\alpha}_n - \alpha_0) + \frac{1}{2} (\hat{\theta}_n - \theta_0)^2 \left\{ \frac{6n}{\theta_*^4} \sum_{i=1}^n X_i - \frac{2n}{\theta_*^3} [\beta(n-1)/2 + \alpha_* n] \right\} \\ &\quad + \frac{n^2}{\theta_*^2} (\hat{\theta}_n - \theta_0) (\hat{\alpha}_n - \alpha_0), \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial \alpha} l_n(\hat{\alpha}_n, \hat{\theta}_n; \mathbf{X}) &= n L_2(\mathbf{X}; \alpha_0, \theta_0) + (\hat{\alpha}_n - \alpha_0) \left\{ -n^2 \sum_{j=0}^{n-1} \psi'(\alpha_0 n + j\beta/2) \right\} \\ &\quad - \frac{n^2}{\theta_0} (\hat{\theta}_n - \theta_0) + \frac{1}{2} (\hat{\alpha}_n - \alpha_0)^2 \left\{ -n^3 \sum_{j=0}^{n-1} \psi''(\alpha_* n + j\beta/2) \right\} \\ &\quad + \frac{n^2}{2\theta_*^2} (\hat{\theta}_n - \theta_0)^2. \end{aligned}$$

Since

$$\frac{\partial}{\partial \theta} l_n(\hat{\alpha}_n, \hat{\theta}_n; \mathbf{X}) = 0, \quad \frac{\partial}{\partial \alpha} l_n(\hat{\alpha}_n, \hat{\theta}_n; \mathbf{X}) = 0,$$

we have that

$$\begin{aligned} Y_{1n} &= n(\hat{\alpha}_n - \alpha_0), \\ Y_{2n} &= n(\hat{\theta}_n - \theta_0), \end{aligned}$$

satisfy the system of (random) linear equations

$$\sum_{k=1}^2 A_{jkn} Y_{kn} = T_{jn}, \quad j = 1, 2,$$

with

$$\begin{aligned} T_{1n} &= L_2(\mathbf{X}; \alpha_0, \theta_0), \\ T_{2n} &= L_1(\mathbf{X}; \alpha_0, \theta_0) / \theta_0^2, \end{aligned}$$

and

$$A_{11n} \rightarrow \frac{2}{\beta} \log \left(1 + \frac{\beta}{2\alpha_0} \right), \quad A_{12n} = A_{21n} \rightarrow \frac{1}{\theta_0}, \quad A_{22n} \rightarrow \frac{\alpha_0 + \beta/2}{\theta_0^2}, \quad n \rightarrow \infty,$$

in probability. Consequently, from Lemma 4.7 and Lemma 4.11, we have that the random vector $(Y_{1n}, Y_{2n}) = n(\hat{\alpha}_n - \alpha_0, \hat{\theta}_n - \theta_0)$ converges weakly to the bivariate normal distribution with 0 mean and covariance matrix $\Sigma_{(\alpha_0, \theta_0)}$ as specified in (4.27). \square

In Figure 4.3 we illustrate the result of Lemma 4.12. The quantiles of the random variables $n(\hat{\alpha}_n - \alpha_0)$ and $n(\hat{\theta}_n - \theta_0)$ are compared with the quantiles of the limiting normal distribution in (4.26).

4.4.3 β -Jacobi Ensemble

In the following lemma we obtain consistency and asymptotic efficiency for the maximum likelihood estimators of parameters in β -Jacobi ensemble.

Lemma 4.13. *Let $\beta > 0$ and $\mathbf{X} = (X_1, \dots, X_n)$ be a sample of exchangeable random variables with the joint probability density of β -Jacobi ensemble with varying coefficients (4.13), and the true parameter value (a_0, b_0) . The maximum likelihood estimator (\hat{a}_n, \hat{b}_n) is consistent and asymptotically efficient in the sense that*

$$n(\hat{a}_n - a_0, \hat{b}_n - b_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_{(a_0, b_0)}^{-1}). \quad (4.28)$$

with

$$\Sigma_{(a_0, b_0)} = \begin{pmatrix} \frac{2}{\beta} \log \frac{(2a_0 + \beta)(2a_0 + 2b_0 + \beta)}{2a_0(2a_0 + 2b_0 + 2\beta)} & -\frac{2}{\beta} \log \frac{2a_0 + 2b_0 + 2\beta}{2a_0 + 2b_0 + \beta} \\ -\frac{2}{\beta} \log \frac{2a_0 + 2b_0 + 2\beta}{2a_0 + 2b_0 + \beta} & \frac{2}{\beta} \log \frac{(2b_0 + \beta)(2a_0 + 2b_0 + \beta)}{2b_0(2a_0 + 2b_0 + 2\beta)} \end{pmatrix}. \quad (4.29)$$

Proof. The proof of consistency and asymptotic efficiency is analogous to the proof of Lemma 4.12. Namely, the Taylor expansion of the log-likelihood function divided by n^2 is

$$\frac{1}{n^2} l_n(a, b) - \frac{1}{n^2} l_n(a_0, b_0) = I_1^{(n)}(a, b) + I_2^{(n)}(a, b) + I_3^{(n)}(a, b),$$

where

$$I_1^{(n)}(a, b) = \frac{1}{n}(a - a_0)J_1(\mathbf{X}; a_0, b_0) - \frac{1}{n}(b - b_0)J_2(\mathbf{X}; a_0, b_0),$$

$$\begin{aligned}
I_2^{(n)}(a, b) &= \frac{1}{2}(a - a_0)^2 \left\{ -\sum_{j=0}^{n-1} \psi(a_0n + j\beta/2) + \sum_{j=0}^{n-1} \psi(a_0n + b_0n + (n - j + 1)\beta/2) \right\} \\
&+ \frac{1}{2}(b - b_0)^2 \left\{ -\sum_{j=0}^{n-1} \psi(b_0n + j\beta/2) + \sum_{j=0}^{n-1} \psi(a_0n + b_0n + (n - j + 1)\beta/2) \right\} \\
&+ (a - a_0)(b - b_0) \left\{ \sum_{j=0}^{n-1} \psi(a_0n + b_0n + (n - j + 1)\beta/2) \right\},
\end{aligned}$$

and

$$|I_3^{(n)}(a, b)| \leq C_1 s^3 r^3 / 6$$

for n sufficiently large, $(a, b) \in Q_r$. From Lemma 4.8 we have that

$$\frac{1}{n} \mathbf{J}(\mathbf{X}; a_0, b_0) \rightarrow 0, \quad n \rightarrow \infty$$

in probability, hence

$$|I_1^{(n)}(a, b)| \leq sr^3$$

with probability tending to 1, while $I_2^{(n)}(a, b)$ is negative (nonrandom) quadratic form such that

$$|I_2^{(n)}(a, b)| \leq -Cr^2$$

for some constant $C > 0$. Consequently,

$$\max_{(a, b) \in Q_r} \left\{ I_1^{(n)}(a, b) + I_2^{(n)}(a, b) + I_3^{(n)}(a, b) \right\} \leq (s + C_1 s^3 / 6) r^3 - Cr^2 < 0$$

for $r < C(s + s^3 C_1 / 6)^{-1}$ with probability tending to 1, and

$$\lim_{n \rightarrow \infty} \mathbb{P}_{(a_0, b_0)}(l_n(a_0, b_0; \mathbf{X}) > l_n(a, b; \mathbf{X})) = 1, \quad \forall (a, b) \in Q_r.$$

Therefore, the maximum likelihood function has a maximum (\hat{a}_n, \hat{b}_n) at the interior of Q_r with probability tending to 1. Since the solution of the maximum likelihood equations for β -Jacobi ensemble is unique with probability 1 for every $n \geq 2$, we obtain that

$$\lim_{n \rightarrow \infty} \mathbb{P}_{(a_0, b_0)}(|\hat{b}_n - b_0|^2 + |\hat{a}_n - a_0|^2 < r^2) = 1,$$

and the consistency of (\hat{a}_n, \hat{b}_n) follows.

Asymptotic efficiency is obtained like in Lemma 4.12. The Taylor expansions of the partial derivatives of the log-likelihood function lead to the system of random

linear equations

$$\sum_{k=1}^2 A_{jkn} Y_{kn} = T_{jn}, \quad j = 1, 2,$$

with

$$\begin{aligned} Y_{1n} &= n(\hat{a}_n - a_0), & Y_{2n} &= n(\hat{b}_n - b_0), \\ T_{1n} &= J_1(\mathbf{X}; a_0, b_0), & T_{2n} &= J_2(\mathbf{X}; a_0, b_0), \end{aligned}$$

and

$$A_{jkn} \rightarrow a_{jk}, \quad n \rightarrow \infty,$$

in probability, where $A = \|a_{jk}\| = \Sigma_{(a_0, b_0)}$ is the covariance matrix in (4.16). From Lemma 4.8 and Lemma 4.11, we obtain that

$$n(\hat{a}_n - a_0, \hat{b}_n - b_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_{(a_0, b_0)}^{-1}).$$

□

4.4.4 Cauchy Ensemble

Proposition 4.14. *Let $\mathbf{X} = (X_1, \dots, X_n)$ be a sample of exchangeable random variables with the joint probability density of Cauchy unitary ensemble with varying coefficients (4.17) and the true parameter value $\alpha_0 > 1$. The maximum likelihood estimator $\hat{\alpha}_n$ is consistent and asymptotically efficient in the sense that*

$$n(\hat{\alpha}_n - \alpha_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_{\alpha_0}^{-2}), \quad (4.30)$$

where $\sigma_{\alpha_0}^2$ is given in (4.19).

Proof. The Taylor expansion of the log-likelihood function divided by n^2 is

$$\frac{1}{n^2} l_n(\alpha) - \frac{1}{n^2} l_n(\alpha_0) = I_1^{(n)}(\alpha) + I_2^{(n)}(\alpha) + I_3^{(n)}(\alpha),$$

where

$$\begin{aligned} I_1^{(n)}(\alpha) &:= \frac{1}{n}(\alpha - \alpha_0)C(\mathbf{X}; \alpha_0), \\ I_2^{(n)}(\alpha) &:= \frac{1}{2}(\alpha - \alpha_0)^2 \left\{ -4 \sum_{i=0}^{n-1} \psi'(2\alpha n - n - i) + 2 \sum_{i=0}^{n-1} \psi'(\alpha n - i) \right\}, \end{aligned}$$

$$I_3^{(n)}(\alpha) := \frac{1}{6}(\alpha - \alpha_0)^3 \left\{ -8n \sum_{i=0}^{n-1} \psi''(2\alpha_* n - n - i) + 2n \sum_{i=0}^{n-1} \psi''(\alpha_* n - i) \right\},$$

The arguments analogous to those in the proof of Lemma 4.12 lead to the limit

$$\lim_{n \rightarrow \infty} P_{\alpha_0}(|\hat{\alpha}_n - \alpha_0| < r) = 1, \quad \forall r > 0,$$

and consistency of $\hat{\alpha}_n$ is obtained. The Taylor expansion for the derivative of $l_n(\hat{\alpha}_n, \mathbf{X})$ results in the linear random equation $A_n Y_n = T_n$, where

$$Y_n = n(\hat{\alpha}_n - \alpha_0), \quad T_n = C(\mathbf{X}; \alpha_0),$$

$A_n \rightarrow \sigma_{\alpha_0}^2$ in probability, as $n \rightarrow \infty$, and $C(\mathbf{X}; \alpha_0)$ is the sufficient statistics defined in Proposition 4.9. From Lemma 4.11 and Proposition 4.9 asymptotic efficiency (4.30) follows. \square

Chapter 5

Circular Ensembles

5.1 Introduction

The theory of random matrices has been developed in order to investigate properties of complex spectra. Although the Gaussian ensembles, introduced in Wigner [118], allow wide applications in different branches of mathematics and physics, they have a drawback of being defined on the space of matrices which is not compact. As a consequence, there is no way to assign the same weight to every matrix and hence matrices representing different quantum systems can not be treated in the same manner. To avoid this deficiency, Dyson [45] modified Wigner's treatment of a nucleus and defined three ensembles, similar to Gaussian, but mathematically simpler to deal with. In Dyson's paper, the system is characterized by a random unitary $n \times n$ matrix U , which is supposed to be a function of the Hamiltonian H of the system, but no detailed specification of this functional relationship can be given globally. As mentioned in Mehta [87], one can think of a connection such as

$$U = \exp[i\tau H] \quad \text{or} \quad U = \frac{1 - i\tau H}{1 + i\tau H},$$

although these relations can be valid only locally. In such a setting, the eigenvalues $\exp[i\zeta_j]$, $1 \leq j \leq n$ of U are supposed to be functions of the eigenvalues E_j of H , and the following basic assumption is formulated as "the behavior of n_0 consecutive levels of an actual system, whenever n_0 is small compared to the total number of levels, is the same as that of n_0 consecutive angles $\zeta_1, \dots, \zeta_{n_0}$, where n_0 is small compared with n ".

The purpose of this chapter is to generalize Dyson's circular ensembles and develop their rigorous statistical treatment. Attention of physicists and mathematicians is

centered around the asymptotical behavior of the spectra of random matrices rather than around the structure of random matrices themselves. Therefore, our asymptotical analysis and estimation procedures are developed for the densities of eigenvalues of matrices taken at random from the *generalized circular unitary ensemble*. Although the estimation problem can be dealt with within a general non-parametric framework, we restrict our attention to the densities belonging to exponential families of probability distributions, since numerous examples discussed in the physics literature allow this interpretation. Additionally, remarkable analytical tools from the theory of exponential families described in Chapter 2 can be applied to this setting.

The outline of this chapter is as follows. In section 2, we state the basic facts concerning Dyson circular ensembles and introduce their generalizations by setting the probability density of the phases of n eigenvalues to be of the form

$$p_{\boldsymbol{\theta};n}(\boldsymbol{\zeta}) = C_n^{-1}(\boldsymbol{\theta}) \prod_{j=1}^n w_{\boldsymbol{\theta}}(e^{i\zeta_j}) \prod_{1 \leq j < k \leq n} |e^{i\zeta_j} - e^{i\zeta_k}|^2, \quad \boldsymbol{\zeta} \in [0, 2\pi]^n, \quad (5.1)$$

where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_r)$ is an r -dimensional parameter, $C_n(\boldsymbol{\theta})$ is the normalizing constant, and $w_{\boldsymbol{\theta}}$ is a suitable weight function on the unit circle \mathbb{T} . Our investigation is motivated by their numerous applications in physics in the areas such as chaotic scattering (Jalabert, Pichard [66]), conductance in mesoscopic systems (Beenaker [26]), periodically driven systems (Haake [59]), and mathematics, where they arise in the theory of orthogonal polynomials on the unit circle (Ismail, Witte [65], Simon [99]), the theory of Toeplitz determinants (Adler, van Moerbeke [2], Borodin [31]), etc. Section 3 is devoted to Weyl's integration formula for unitary group $U(n)$, while section 4 provides a brief overview of the results related to the theory of Toeplitz determinants. In particular, the strong Szegő theorem is formulated. In section 5 the asymptotic normality of the sufficient statistics for the parametric generalized Dyson circular unitary ensemble (GCUE) is established. Finally, the asymptotic maximum likelihood procedure is applied to estimate parameters of the joint eigenvalue distribution of GCUE, and the properties of the estimators are derived in section 6. From asymptotical behavior of the Fisher information matrix and explicit expressions for the asymptotic maximum likelihood estimators, we conclude that consistent estimation of the parameter $\boldsymbol{\theta}$ is not possible, unless the dimension parameter n is included into the weight function $w_{\boldsymbol{\theta}}$ in (5.1). For this reason in section 8, we analyze the model with *varying weight* from Gross, Witten [57], where the probability density of n eigenphases is equal to

$$p_{\gamma;n}(\boldsymbol{\zeta}) = C_n^{-1}(\gamma) |\Delta(\boldsymbol{\zeta})|^2 \prod_{i=1}^n \exp \left\{ \gamma n \sum_{i=1}^n \cos \zeta_i \right\}, \quad \boldsymbol{\zeta} \in [0, 2\pi]^n, \quad (5.2)$$

with $C_n(\gamma)$ as the normalizing constant and $\gamma > 0$. We show that the true parameter value γ_0 can be estimated consistently, provided that $\gamma_0 \in (0, 1 - \epsilon)$, $\epsilon > 0$. Additionally, we obtain that the maximum likelihood estimator $\hat{\gamma}_n$ of γ_0 is asymptotically normal in the sense that

$$n(\hat{\gamma}_n - \gamma_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 2), \quad n \rightarrow \infty, \quad (5.3)$$

if $\gamma_0 \in (0, 1 - \epsilon)$, $\epsilon > 0$.

5.2 Circular Ensembles and Generalizations

For motivational purposes in this section we review the construction of the circular unitary ensemble of Dyson, introduce its parametric as well as nonparametric generalizations and, for the sake of completeness, state basic properties of circular orthogonal and symplectic ensembles. Our attention is especially restricted to the unitary case since the analytical tools are the most developed for this specific instance. In particular, the eigenvalue density and its asymptotic behavior are easily derived by methods described in sections 3 and 4.

Circular ensembles were defined by Dyson [45] as subsets of a set $U(n)$ of all $n \times n$ unitary matrices. Uniqueness of the ensembles is imposed by introducing measures invariant under appropriate groups of transformations. As mentioned above, unitary matrices have applications in quantum physics, particularly in the theory of periodic quantum systems (kicked tops) and in the scattering of plane waves within an irregular shaped domain (cavity), or a cavity of arbitrary shape containing random impurities (see Forrester [47] and references therein).

Recall that $U(n)$ consists of all $n \times n$ invertible complex matrices U which satisfy the relation $UU^* = I$, where U^* is the complex conjugate transpose of the matrix U and I denotes the identity matrix. In order to define the circular unitary ensemble, we endow $U(n)$ with the normalized Haar measure dU . The normalized Haar measure is the unique probability measure on $U(n)$ invariant under left (also right) multiplications by an arbitrary unitary matrix. As a consequence, we have that all the matrices in the group receive equal weights in group averages. Notice that this homogeneity requirement makes the measure invariant under unitary changes of basis. This remarkable invariance property implies the explicit definition of an ensemble of random unitary matrices.

Definition 4. A *circular unitary ensemble* (CUE) is a group of unitary matrices endowed with the normalized Haar measure.

Since circular orthogonal and symplectic ensembles, defined on the spaces of symmetric and self-dual unitary quaternion matrices respectively, are not of primary interest for our analysis, and their introduction involves more detailed discussions about the definition of a homogeneous measure on relevant spaces (on the space of symmetric, respectively self-dual unitary quaternion matrices) which do not form groups, we will escape their formal definitions and refer the reader to Forrester [47] and Mehta [87].

As mentioned above, motivation for the initiation of circular ensembles primarily came from the need to assign the same probability to each matrix in the ensemble. Enormously wide applications of this concept were discovered subsequently. Nevertheless, it appeared that some physical systems which could be described by random unitary matrices show deviations from this homogeneity property. As mentioned in Muttalib, Ismail [89], numerical studies related to disordered conductors [66] and periodically driven systems [78], exhibit statistical behavior different from that of Dyson's models. To remedy this deficiency, a new analytical model, containing CUE, is considered. Namely, the probability of a matrix U to be taken from the unitary group is set to be proportional to $L(U)dU$, where $L : U(n) \rightarrow \mathbb{R}$ is an arbitrary integrable function invariant under conjugacy transformations $U \rightarrow WUW^{-1}$, for any $W \in U(n)$. Of our particular interest are the exponential parametric generalizations, where the function L is assumed to be of the form

$$L(U) = \exp \left\{ \sum_{j=1}^r \theta_j \operatorname{tr} V_j(U) \right\}. \quad (5.4)$$

Here θ_j are the parameters and V_j are the real-valued functions. For the moment, we are being informal about conditions on V_j under which such a generalization is possible. Notice that all powers and their linear combinations (provided that they are real-valued) give an example of such a parametrization, e.g. functions $V_j(U) = U^j + U^{-j}$ can be used in (5.4). The initiation of the weight function on the unitary group may come from any system-dependent physical constraint as the measure under consideration may depend on various physical parameters (see Muttalib, Ismail [89]). To summarize the discussion, we introduce the formal definition of a generalized circular unitary ensemble.

Definition 5. A *generalized circular unitary ensemble* (CGUE) is a unitary group $U(n)$ endowed with the probability measure

$$P(dU) = L(U)dU, \quad (5.5)$$

where $L : U(n) \rightarrow \mathbb{R}$ is a function invariant under conjugacy transforms $U \rightarrow WUW^{-1}$ for every $W \in U(n)$, and dU denotes the Haar measure on $U(n)$.

Remark 5.1. We should mention a particular class of probability densities (5.5) with

$$L(U) = \exp(\operatorname{tr} V(U)),$$

where $V : \mathbb{T} \rightarrow \mathbb{R}$ is such that $\exp(V(z))$ is integrable on \mathbb{T} . Ensembles of this form can be seen as dominant in RMT and the theory of log-potential systems (see e.g. Deift [37]).

5.3 Weyl's Integration Formula

The aim of this section is to obtain the joint density of eigenvalues of a matrix taken at random from GCUE. For this purpose, some well-known results from representation theory will be given, and their applications to our problem will be discussed.

In the middle 1920s Hermann Weyl obtained an explicit formula to integrate functions on compact simple Lie groups. The formula, commonly known as *Weyl's integration formula*, is an analogue of the classical formula for the transformation of integrals over \mathbb{R}^3 under the change of Cartesian to spherical coordinates. We will restrict our attention to its specific case—integration over the unitary group with respect to the normalized Haar measure, since Weyl's formula in its full generality involves a number of analytical details, which are not of main interest for our investigation. For the full description, we refer the reader to Weyl [116], noticing that a nice review over the unitary case may be found in Pasquale [93].

Consider the group of unitary matrices $U(n)$ endowed with the normalized Haar measure and its subgroup $T(n)$ consisting of diagonal matrices, i.e.

$$T(n) := \{D(\zeta) = \operatorname{diag}\{e^{i\zeta_1}, \dots, e^{i\zeta_n}\} : \zeta_j \in [0, 2\pi), 1 \leq j \leq n\}.$$

Recall that every U in $U(n)$ is conjugate to a diagonal unitary matrix, namely for arbitrary $U \in U(n)$ there exist $V \in U(n)$ and $D \in T(n)$, such that $U = VDV^{-1}$, where the diagonal elements of D are the eigenvalues of U . In order to identify all the matrices with the same spectrum, we introduce the conjugacy classes as subsets of $U(n)$ consisting of all matrices with the same set of eigenvalues. It is easy to show that unitary matrices U and V belong to the same conjugacy class if and only if there exists a unitary W such that $U = WVW^{-1}$. To formulate Weyl's formula we need a notion of a class function.

Definition 6. A *class function* is a function $f : U(n) \rightarrow \mathbb{C}$ which is invariant on conjugacy classes, i.e. $f(VUV^{-1}) = f(U)$ for arbitrary $U, V \in U(n)$.

The celebrated integration formula due to Weyl can be stated now.

Lemma 5.2 (Weyl). *Let $f \in L^1(U(n))$ be a class function. Then the following integration formula holds*

$$\int_{U(n)} f(U) dU = \frac{1}{(2\pi)^n n!} \int_{[0, 2\pi]^n} f(D(\zeta)) |\Delta(\zeta)|^2 d^n \zeta, \quad (5.6)$$

where

$$\Delta(\zeta) = \prod_{1 \leq j < k \leq n} (e^{i\zeta_j} - e^{i\zeta_k}) = \begin{vmatrix} 1 & 1 & \dots & 1 \\ e^{i\zeta_1} & e^{i\zeta_2} & \dots & e^{i\zeta_n} \\ \dots & \dots & \dots & \dots \\ e^{i(n-1)\zeta_1} & e^{i(n-1)\zeta_2} & \dots & e^{i(n-1)\zeta_n} \end{vmatrix} \quad (5.7)$$

is the Vandermonde determinant.

From Weyl's formula the probability density of the spectrum for GCUE can be easily obtained. Since the eigenvalues λ_j of a unitary matrix lie on the unit circle, they can be written in the form $\lambda_j = \exp[i\zeta_j]$, where the eigenphases ζ_j are assumed to be from the interval $[0, 2\pi]$ for every j . The probability measure of the eigenvalues of a matrix chosen at random from GCUE is given in the following proposition.

Proposition 5.3. *The joint distribution of the eigenphases of $n \times n$ matrix taken randomly from the generalized circular unitary ensemble (5.5) is absolutely continuous with respect to the Lebesgue measure $\prod_{j=1}^n d\zeta_j$ on $[0, 2\pi]^n$ and its density is*

$$p_n(\zeta) = C_n^{-1} V(D(\zeta)) |\Delta(\zeta)|^2, \quad \zeta \in [0, 2\pi]^n, \quad (5.8)$$

where C_n is the normalizing constant.

Proof. The statement immediately follows from Weyl's integration formula applied to the function $V(U)$ in Definition 5. \square

It should be mentioned that a certain inconsistency related to the terminology can be observed in the literature on random matrix theory. Namely, some authors use the term "ensemble" exclusively for the space of matrices endowed with the probability measure, while in other sources the same notion refers to the probability distribution on the spectra of random matrices. The latter case is commonly used in the issue of generalizations of Dyson's model, namely, both Nagao, Wadati [91] and Muttalib, Ismail [89] initiate generalized circular ensembles through the joint probability density of the eigenvalues without specifying the measure on the set of matrices. The brief outline of their reasoning is given below.

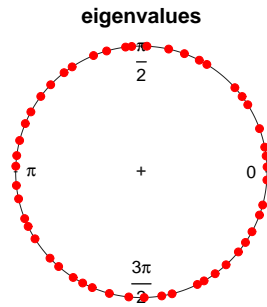


Figure 5.1: The eigenvalues of Dyson's circular unitary ensemble, $n = 50$.

Dyson's circular ensembles can be defined by the probability density function of n eigenphases

$$p_{\beta;n}(\zeta) = C_{\beta;n}^{-1} \prod_{1 \leq j < k \leq n} |e^{i\zeta_j} - e^{i\zeta_k}|^{\beta}, \quad \zeta \in [0, 2\pi]^n. \quad (5.9)$$

Here β is the symmetry parameter equal to 1, 2 and 4 for Dyson orthogonal, unitary and symplectic ensemble, respectively. The eigenvalues of circular unitary ensemble ($n = 50$) are plotted in Figure 5.1. In two papers mentioned above [89], [91], Dyson's model is generalized by introducing an additional weight w on the unit circle and setting the density of n eigenvalues to be of the form

$$p_{w,\beta;n}(\zeta) = C_{w,\beta;n}^{-1} \prod_{j=1}^n w(e^{i\zeta_j}) \prod_{1 \leq j < k \leq n} |e^{i\zeta_j} - e^{i\zeta_k}|^{\beta}, \quad \zeta \in [0, 2\pi]^n. \quad (5.10)$$

The weight w could take into account some physical-dependent constraints of the system. For instance, in the Coulomb gas model, this generalization is equivalent to applying an external field to the system of n particles free to move on the circular wire. One of the main problems regarding such a definition is the evaluation of the normalizing constant $C_{w,\beta;n}$, which is known as the partition function in the theory of log-potential systems. However, for the unitary case ($\beta = 2$), the explicit expression for $C_{w,\beta;n}$ can be given in terms of *Toeplitz determinants* with respect to the weight function w . For this purpose, Toeplitz determinants together with their asymptotical behavior are reviewed in the next section.

Remark 5.4. Notice that for the parametric model introduced in (5.4), the density of eigenphases is exactly of the form (5.10), where the weight w depends on parameters

$\theta_1, \dots, \theta_r$, and can be written as

$$w(e^{i\zeta}) = \exp \left\{ \sum_{j=1}^r \theta_j V_j(e^{i\zeta}) \right\}, \quad \zeta \in [0, 2\pi].$$

Remark 5.5. The problem of specifying CGUE from the numerical generation sense remains open as a part of our further development of this subject.

Remark 5.6. It should be noted here that circular ensembles may be generalized in a way different from the one described above. The recent paper by Killip, Nenciu [82] discusses an ensemble of random sparse matrices whose eigenvalues follow the distribution given in (5.9) for general $\beta > 0$. Our approach allows further investigation from this perspective as well. Namely, not only the unitary case can be considered, but also generalizations over β may be established. We suppose that the numerical generation procedure of such a model would bring new insights into the subject and give rise to an additional analysis of this matter.

5.4 Toeplitz Determinants and Strong Szegő Theorem

In this section we briefly review the theory of Toeplitz determinants and give the exact expression for the normalization constant $C_{w,\beta;n}$ in (5.10). As a consequence, the rigorous definition of the parametric generalized circular unitary ensemble will be obtained.

Various remarkable issues in the modern mathematics are closely connected to the analysis of certain hermitian forms whose matrices, finite or infinite, have the elements of the structure a_{i-j} . A very attractive field in this respect is the theory of Toeplitz operators, in particular—asymptotics of Toeplitz determinants, which is associated to a variety of problems in physics, probability theory and mathematical statistics, theory of orthogonal polynomials, information and control theory and several other branches of mathematics. In this section we present some classical results on the asymptotics of Toeplitz determinants which have relevance to our analysis of the generalized circular ensemble. The selection of the material is determined by the applications of the strong Szegő theorem to the ensembles on the unitary group, while the reader with particular interest in this subject may consider the celebrated book by Grenander, Szegő [55]. For the recent insight into the theory of Toeplitz operators one is referred to Böttcher, Silbermann [32]. The classical theory of orthogonal polynomials on the unit circle together with historical remarks related to its development is thoroughly described in Simon [99].

Recall that for a function $f \in L^1(\mathbb{T})$, the n^{th} -order *Toeplitz determinant* is defined as

$$D_n[f] = \det[\hat{f}(j-k)]_{j,k=0}^{n-1} = \begin{vmatrix} \hat{f}(0) & \hat{f}(-1) & \dots & \hat{f}(-n+1) \\ \hat{f}(1) & \hat{f}(0) & \dots & \hat{f}(-n+2) \\ & \dots & \dots & \\ \hat{f}(n-1) & \hat{f}(n-2) & \dots & \hat{f}(0) \end{vmatrix}, \quad (5.11)$$

where $\{\hat{f}(k)\}_{k \in \mathbb{Z}}$ is the sequence of Fourier coefficients of f . In this context, the function f is frequently referred to as the symbol of the corresponding determinant. The analysis of the asymptotical behavior of $D_n[f]$ has had an illustrious history. The initial step was made by Szegö in 1915, when the theorem regarding the first-order asymptotics of $D_n[f]$ was established. This result was subsequently extended in different directions. Particularly, the problem of analyzing the second-order asymptotics was raised by Onsager, who encountered it in his study of the Ising model. Motivated by Onsager's question, Szegö sharpened his 1915 result. For this reason, the theorem and its further generalizations, are nowadays commonly known as the *strong Szegö theorem*. The result was formulated as follows

Theorem 5.7 (Szegö). *Let $f \in C^1(\mathbb{T})$ be a positive function on the unit circle with the derivative f' satisfying the Lipschitz condition of order α ($0 < \alpha < 1$) and $g = \log f$. Then*

$$\lim_{n \rightarrow \infty} \frac{D_n[f]}{G[f]^n} = E[f], \quad (5.12)$$

where

$$G[f] = \exp \{ \hat{g}(0) \}, \quad (5.13)$$

$$E[f] = \exp \left\{ \sum_{k \in \mathbb{N}} k \hat{g}(k) \hat{g}(-k) \right\}, \quad (5.14)$$

and $\{\hat{g}(k)\}_{k \in \mathbb{Z}}$ is the sequence of Fourier coefficients of g .

The smoothness condition needed by Szegö was subsequently relaxed in the series of articles including Kac [79], Baxter [23], Hirschman [60, 61, 62], Geronimus [48], Devinatz [38, 39], etc. Finally, the conclusive contribution was made by Ibragimov [64], who replaced the Lipschitz condition on f' by the requirement $\log f \in L^1(\mathbb{T})$. The ultimate result can be rephrased as follows (see Johansson [69]).

Theorem 5.8 (Ibragimov). *Let $g \in L^1(\mathbb{T})$ be a complex-valued function on \mathbb{T} with*

Fourier coefficients $\{\hat{g}(k)\}_{k \in \mathbb{Z}}$. Assume that

$$\sum_{k \in \mathbb{Z}} |k| |\hat{g}(k)|^2 < \infty. \quad (5.15)$$

Then

$$D_n[\exp g] = \exp \left\{ n\hat{g}(0) + \sum_{k \in \mathbb{N}} k\hat{g}(k)\hat{g}(-k) + o(1) \right\} \quad (5.16)$$

as $n \rightarrow \infty$.

The condition (5.15) is of crucial importance for our definition of the parametric generalized circular unitary ensemble. Its precise formulation is as follows.

Definition 7. The *parametric generalized circular unitary ensemble* is the ensemble of random unitary matrices whose joint distribution of n eigenphases is given by the density

$$p_{\boldsymbol{\theta}; n}(\boldsymbol{\zeta}) = C_n^{-1}(\boldsymbol{\theta}) \prod_{j=1}^n w_{\boldsymbol{\theta}}(e^{i\zeta_j}) \prod_{1 \leq j < k \leq n} |e^{i\zeta_j} - e^{i\zeta_k}|^2, \quad \boldsymbol{\zeta} \in [0, 2\pi]^n. \quad (5.17)$$

with $\boldsymbol{\theta} = (\theta_1, \dots, \theta_r)$ as the r -dimensional parameter, $C_n(\boldsymbol{\theta})$ as the normalizing constant and

$$w_{\boldsymbol{\theta}}(e^{i\zeta}) = \exp \left\{ \sum_{i=1}^r \theta_i V_i(e^{i\zeta}) \right\}, \quad \zeta \in [0, 2\pi], \quad (5.18)$$

where real-valued functions V_j with Fourier coefficients $\{\hat{V}_j(k)\}_{k \in \mathbb{Z}}$ are assumed to satisfy the condition

$$E_j = \sum_{k \in \mathbb{N}} k \left| \hat{V}_j(k) \right|^2 < \infty, \quad 1 \leq j \leq r. \quad (5.19)$$

Notice that the function $w_{\boldsymbol{\theta}}$ is integrable on the unit circle if the condition (5.19) is satisfied. This fact is obtained from the proposition

Proposition 5.9. Let $V \in L^1(\mathbb{T})$ be a real-valued function on the unit circle with Fourier coefficients $\{\hat{V}(k)\}_{k \in \mathbb{Z}}$, such that

$$E = \sum_{k \in \mathbb{N}} k \left| \hat{V}(k) \right|^2 < +\infty,$$

and define $w(e^{i\zeta}) = \exp V(e^{i\zeta})$, $\zeta \in [0, 2\pi]$. Then $w^p \in L^1(\mathbb{T})$ for every $p \in \mathbb{R}$.

Proof. See Simon [99]. □

From this it follows that $\exp(\theta_j V_j) \in L^1(\mathbb{T})$ for every $\theta_j \in \mathbb{R}, 1 \leq j \leq r$, and consequently $w_\theta \in L^1(\mathbb{T})$ for every $\theta \in \mathbb{R}^r$, if the condition (5.19) is satisfied.

We proceed with giving a list of arguments for the initiation of the model described in (5.17). First of all, most models from practical applications can be written in this form and although a non-parametrical set up (5.10) can be considered, our class of models is sufficient for most practical objectives. Secondly, statistical analysis of our model can be partially carried out by means of the theory of exponential families: the model allows sufficient reduction of data, the normalizing constant $C_{\theta;n}$ is an analytical function of parameter $\theta \in \mathbb{R}^r$ for arbitrary fixed $n \in \mathbb{N}$, etc. Another justification of our choice of the model is supplied from an argument of the information theory: densities with exponential form maximize the entropy of the system, provided that extra constraints on the system performance are imposed (see Balian[19]). In addition to these properties, we must mention that parameters $\theta_1, \dots, \theta_r$ may have a clear physical interpretation, and their estimation may be of high importance for our understanding of the system behavior.

The main examples of the weight function w_θ as described in (5.18), include

$$\begin{aligned} w_\theta^{(1)}(e^{i\zeta}) &= \exp(\theta \cos \zeta), & \zeta \in [0, 2\pi], \\ w_\theta^{(2)}(e^{i\zeta}) &= \exp(\theta_1 \cos \zeta + \theta_2 \cos 2\zeta), & \zeta \in [0, 2\pi], \\ w_\theta^{(3)}(e^{i\zeta}) &= (1 + \rho^2 - 2\rho \cos \zeta)^\theta, & \zeta \in [0, 2\pi]. \end{aligned}$$

The weight function $w_\theta^{(1)}$ and the corresponding system of orthogonal polynomials have arisen from the studies on the length of the longest increasing subsequence of a random permutation in Baik, Deift, Johansson [9], and random matrix models in Gross, Witten [57], Periwal, Shewitz [94]. Recursion relations for Toeplitz determinants of this symbol can be found in Borodin [31], or Adler, van Moerbeke [2], while the properties of orthogonal polynomials with respect to this weight (called the Bessel weight) appear in detail in Ismail, Witte [65]. It should be noted here that in the literature on circular statistics, $w_\theta^{(1)}(e^{i\zeta})$ is known as the density of von Mises-Fisher distribution on 2-dimensional sphere (see e.g. Jammalamadaka, SenGupta [68]). The weight $w_\theta^{(2)}$ arises in the studies of the longest increasing subsequence in a random odd permutation (see Tracy, Widom [111]). Rational recursion relations for the respective Toeplitz determinants are described in detail in Adler, van Moerbeke [2]. For the discussion on the weight $w_\theta^{(3)}$ and the Toeplitz determinants with respect to this symbol, one is referred to Borodin [31], Borodin, Okounkov [29], Adler, van Moerbeke [2].

The key connection of the parametric GCUE with the theory of Toeplitz determinants lies in the following determinant identity due to Heine and Szegö, from which the explicit expression for the normalizing constant $C_{\boldsymbol{\theta};n}$ can be deduced.

Lemma 5.10 (Szegö [105]). *Let $f \in L^1(\mathbb{T})$ be the function on the unit circle with Fourier coefficients $\{\hat{f}(k)\}_{k \in \mathbb{Z}}$. Then the following identity holds*

$$\frac{1}{(2\pi)^n n!} \int_{[0,2\pi]^n} \prod_{k=1}^n f(e^{i\zeta_k}) |\Delta(\boldsymbol{\zeta})|^2 d^n(\boldsymbol{\zeta}) = D_n[f]. \quad (5.20)$$

From this we can see that $C_n(\boldsymbol{\theta})$ in (5.17) is a multiple of Toeplitz determinant with respect to the symbol $w_{\boldsymbol{\theta}}$, namely

$$C_n(\boldsymbol{\theta}) = (2\pi)^n n! D_n[w_{\boldsymbol{\theta}}]. \quad (5.21)$$

We conclude this section with an important remark regarding the applications of identity (5.20) and strong Szegö theorem in asymptotical analysis of the sufficient statistics of random unitary matrices. The connection of the identity (5.20) with the limiting behavior of linear statistics for Dyson's circular unitary ensemble was first observed by Johansson [69], who in particular established the asymptotic normality of $\text{tr } U^k$, $k \geq 1$ for matrix U taken from circular unitary ensemble. Our approach in the next section is similar to Johansson's—it incorporates results on the asymptotical behavior of Toeplitz determinants with the identity (5.21), resulting in the central limit theorem for sufficient statistics of the parametric GCUE.

5.5 Asymptotic Normality

The aim of this section is to obtain a central limit theorem (CLT) for the vector statistic

$$\mathbf{V}^{(n)}(\boldsymbol{\xi}) = \left(\sum_{k=1}^n V_1(e^{i\xi_k}), \dots, \sum_{k=1}^n V_r(e^{i\xi_k}) \right), \quad (5.22)$$

where $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)$ is a sample of exchangeable random variables with the probability density function

$$p_{\boldsymbol{\theta};n}(\boldsymbol{\zeta}) = C_n^{-1}(\boldsymbol{\theta}) \prod_{j=1}^n w_{\boldsymbol{\theta}}(e^{i\zeta_j}) \prod_{1 \leq j < k \leq n} |e^{i\zeta_j} - e^{i\zeta_k}|^2, \quad \boldsymbol{\zeta} \in [0, 2\pi]^n. \quad (5.23)$$

introduced in Definition 7. This issue is of special interest for comparison with experiments, since physicists are mainly interested in those features of the statistical model that tend to definite limits as $n \rightarrow \infty$. Our approach is based on the strong Szegő theorem and identity (5.21), which will be used to prove point-wise convergence of the characteristic functions to the corresponding limit.

In the case of Dyson's circular unitary ensemble

$$p_{\mathbf{0};n}(\boldsymbol{\zeta}) = C_n^{-1}(\mathbf{0}) \prod_{1 \leq j < k \leq n} |e^{i\zeta_j} - e^{i\zeta_k}|^2, \quad \boldsymbol{\zeta} \in [0, 2\pi]^n,$$

the central limit theorem was obtained for the statistics of the form

$$\sum_{j=1}^n f(e^{i\zeta_j}),$$

where $f \in L^1(\mathbb{T})$ is assumed to be real-valued function such that

$$\sum_{k \in \mathbb{N}} k |\hat{f}(k)|^2 < \infty. \quad (5.24)$$

This result together with the rate of convergence for f equal to a trigonometric polynomial is due to Johansson [69]. For further developments of this subject see Soshnikov [102] and Diaconis, Evans [40].

In order to state the CLT for our model, we need to define a non-negative bilinear form on the space $H_2^{1/2}$ of real-valued functions $f \in L^1(\mathbb{T})$ satisfying the condition (5.24). The bilinear form is defined by

$$\langle f, g \rangle_{1/2} = \sum_{k \in \mathbb{Z}} \hat{f}(k) \hat{g}(-k) |k|, \quad f, g \in H_2^{1/2}. \quad (5.25)$$

The space $H_2^{1/2}$ is the example of Besov potential space and alternative definitions of the bilinear form (5.25) are possible. For further details we refer the reader to [40] and references therein. Once the bilinear form is defined, we can state the theorem

Theorem 5.11. *Consider the sequence of probability density functions $\{p_{\mathbf{0};n}(\boldsymbol{\zeta})\}_{n \geq 2}$ defined in (5.17), where the real-valued functions V_j belong to $H_2^{1/2}$, $1 \leq j \leq r$. Assume in addition that*

$$\hat{V}_j(0) = 0, \quad 1 \leq j \leq r. \quad (5.26)$$

Then the vector statistic $\mathbf{V}^{(n)}(\boldsymbol{\xi})$ introduced in (5.22), has asymptotically normal

distribution with mean

$$\boldsymbol{\mu} = \left(\sum_{j=1}^r \theta_j \langle V_1, V_j \rangle_{1/2}, \dots, \sum_{j=1}^r \theta_j \langle V_r, V_j \rangle_{1/2} \right), \quad (5.27)$$

and covariance matrix

$$\Sigma = \begin{pmatrix} \langle V_1, V_1 \rangle_{1/2} & \langle V_1, V_2 \rangle_{1/2} & \cdots & \langle V_1, V_r \rangle_{1/2} \\ \langle V_2, V_1 \rangle_{1/2} & \langle V_2, V_2 \rangle_{1/2} & \cdots & \langle V_2, V_r \rangle_{1/2} \\ \cdots & \cdots & \cdots & \cdots \\ \langle V_r, V_1 \rangle_{1/2} & \langle V_r, V_2 \rangle_{1/2} & \cdots & \langle V_r, V_r \rangle_{1/2} \end{pmatrix}. \quad (5.28)$$

Proof. The asymptotical distribution of the statistics $\mathbf{V}^{(n)}(\boldsymbol{\xi})$ is obtained from the convergence of its characteristic function

$$\psi_n(\mathbf{t}) = \mathbb{E}_n \exp \left\{ \sum_{j=1}^r \sum_{k=1}^n it_j V_j(e^{i\xi_k}) \right\}, \quad \mathbf{t} \in \mathbb{R}^r,$$

where \mathbb{E}_n denotes the expectation with respect to the measure (5.17). Substitution of the expression for the density $p_{\boldsymbol{\theta}, n}(\boldsymbol{\zeta})$ into the mathematical expectation above gives

$$\begin{aligned} \psi_n(\mathbf{t}) &= C_n^{-1}(\boldsymbol{\theta}) \int_{[0, 2\pi]^n} |\Delta(\boldsymbol{\zeta})|^2 \prod_{k=1}^n \exp \left\{ \sum_{j=1}^r (\theta_j + it_j) V_j(e^{i\xi_k}) \right\} d^n \boldsymbol{\zeta} \\ &= \frac{C_n(\boldsymbol{\theta} + i\mathbf{t})}{C_n(\boldsymbol{\theta})} \end{aligned}$$

Taking into account identity (5.21) and Ibragimov's version of the strong Szegő theorem (Theorem 5.8), we arrive at the expression that has the following form

$$\begin{aligned} \lim_{n \rightarrow \infty} \psi_n(\mathbf{t}) &= \exp \left\{ - \sum_{k \in \mathbb{N}} k \left(\sum_{j=1}^r \theta_j \hat{V}_j(k) \right) \left(\sum_{j=1}^r \theta_j \hat{V}_j(-k) \right) \right\} \\ &\times \exp \left\{ \sum_{k \in \mathbb{N}} k \left(\sum_{j=1}^r (\theta_j + it_j) \hat{V}_j(k) \right) \left(\sum_{j=1}^r (\theta_j + it_j) \hat{V}_j(-k) \right) \right\}. \end{aligned}$$

Rewriting the expression on the right hand side, we observe that

$$\lim_{n \rightarrow \infty} \psi_n(\mathbf{t}) = \exp \left[\sum_{j=1}^r it_j \left(\sum_{k=1}^r \theta_k \langle V_j, V_k \rangle_{1/2} \right) \right]$$

$$\times \exp \left[-\frac{1}{2} \sum_{j,k=1}^r t_j t_k \langle V_j, V_k \rangle_{1/2} \right],$$

where the limiting function coincides with the characteristic function of the multivariate normal distribution with mean and covariance matrix as specified in (5.27), (5.28). Since asymptotic normality is equivalent to the pointwise convergence of characteristic functions to the corresponding limit, the proof of our theorem is completed. \square

Remark 5.12. Notice that

$$\mathbf{V}^{(n)}(\boldsymbol{\xi}) \xrightarrow{\mathcal{D}} \mathcal{N}(\boldsymbol{\mu}, \Sigma), \quad n \rightarrow \infty,$$

without the normalization by \sqrt{n} , and that the components of the vector statistics $\mathbf{V}^{(n)}(\boldsymbol{\xi})$ are asymptotically independent if and only if the functions $V_j, 1 \leq j \leq r$, are orthogonal in the Sobolev space $H_2^{1/2}$, i.e. $\langle V_j, V_k \rangle_{1/2} = 0$ for $j \neq k$. In that case, the covariance matrix (5.28) obtains the diagonal structure

$$\Sigma = \text{diag} \{ \langle V_1, V_1 \rangle_{1/2}, \dots, \langle V_r, V_r \rangle_{1/2} \}.$$

5.6 Asymptotic Maximum Likelihood Estimation

The aim of this section is to construct the asymptotic maximum likelihood estimator $\hat{\boldsymbol{\theta}}^{(n)}$ for parameter $\boldsymbol{\theta}$ of the generalized circular unitary ensemble. The maximum likelihood estimator $\tilde{\boldsymbol{\theta}}^{(n)}$ is defined as the solution of the system of equations

$$\frac{\partial}{\partial \theta_j} \log C_n(\tilde{\boldsymbol{\theta}}^{(n)}) = \sum_{l=1}^n V_j(e^{i\zeta_l}), \quad 1 \leq j \leq r, \quad (5.29)$$

and is not available in the closed form. We show that

$$\frac{\partial}{\partial \theta_j} \log C_n(\boldsymbol{\theta}) \xrightarrow{n \rightarrow \infty} \sum_{k=1}^r \theta_k \langle V_j, V_k \rangle_{1/2}, \quad 1 \leq j \leq r,$$

uniformly over $\boldsymbol{\theta} \in K, K \subset \mathbb{R}^r, K$ compact, and replace (5.29) by the system of linear equations

$$\sum_{k=1}^r \hat{\theta}_k^{(n)} \langle V_j, V_k \rangle_{1/2} = \sum_{l=1}^n V_j(e^{i\zeta_l}), \quad 1 \leq j \leq r, \quad (5.30)$$

which defines the asymptotic MLE $\hat{\boldsymbol{\theta}}^{(n)}$ in closed form. The estimator $\hat{\boldsymbol{\theta}}^{(n)}$ is asymptotically unbiased, but not consistent. We conclude that consistent estimation of parameter $\boldsymbol{\theta}$ is not possible, unless the dimension parameter n is included into the weight function $w_{\boldsymbol{\theta}}$.

It is known from Szegő's theory that the normalizing constant $C_n(\boldsymbol{\theta})$ is the multiple of the n^{th} -order Toeplitz determinant with respect to the symbol $w_{\boldsymbol{\theta}}$, namely

$$C_n(\boldsymbol{\theta}) = (2\pi)^n n! D_n[w_{\boldsymbol{\theta}}].$$

Since $p_{\boldsymbol{\theta},n}(\boldsymbol{\zeta})$ is the density of the distribution from an exponential family for any $\boldsymbol{\theta} \in \mathbb{R}^r, n \geq 2$, it follows from Theorem 2.2 that the derivatives of $\lambda_n(\boldsymbol{\theta}) = C_n(\boldsymbol{\theta})$ may be obtained by differentiation under the integral sign. Additionally, $\lambda_n(\boldsymbol{\theta})$ admits an analytical continuation to \mathbb{C}^r by the identity

$$\tilde{\lambda}_n(\boldsymbol{\theta}) = \int_{[0,2\pi]^n} |\Delta(\boldsymbol{\zeta})|^2 \prod_{k=1}^n w_{\boldsymbol{\theta}}(e^{i\zeta_k}) d^n \boldsymbol{\zeta}, \quad \boldsymbol{\theta} \in \mathbb{C}^r.$$

Assume that the conditions of Theorem 5.11 are satisfied and consider the sequence of Toeplitz determinants $\{D_n[w_{\boldsymbol{\theta}}]\}_{n \geq 2}$ with respect to the symbol $w_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \mathbb{C}^r$. Since the functions $V_j, 1 \leq j \leq r$ are fixed, the Toeplitz determinants $D_n[w_{\boldsymbol{\theta}}]$ become functions of parameter $\boldsymbol{\theta} \in \mathbb{C}^r$. In order to avoid misinterpretation, we define the new sequence of functions $\{d_n(\boldsymbol{\theta})\}_{n \geq 2}$ by setting

$$d_n(\boldsymbol{\theta}) = D_n[w_{\boldsymbol{\theta}}], \quad \boldsymbol{\theta} \in \mathbb{C}^r, n \geq 2. \quad (5.31)$$

From the theory of exponential families and the determinant identity (5.20), we obtain that $d_n(\boldsymbol{\theta}) = \tilde{\lambda}_n(\boldsymbol{\theta})(2\pi)^{-n}/n!$ is an entire function for every $n \geq 2$. In addition to that, as shown in the following lemma, the sequence $\{d_n(\boldsymbol{\theta})\}_{n \geq 2}$ is locally uniformly bounded.

Lemma 5.13. *The sequence $\{d_n(\boldsymbol{\theta})\}_{n \geq 2}$ of entire functions on \mathbb{C}^r is locally uniformly bounded, i.e. for every compact set $K \subset \mathbb{C}^r$ there exists a constant $C_K > 0$ such that*

$$|d_n(\boldsymbol{\theta})| \leq C_K, \quad \forall \boldsymbol{\theta} \in K.$$

Proof. From the determinant identity (5.20) we have

$$d_n(\boldsymbol{\theta}) = \frac{1}{(2\pi)^n n!} \int_{[0,2\pi]^n} |\Delta(\boldsymbol{\zeta})|^2 \prod_{k=1}^n w_{\boldsymbol{\theta}}(e^{i\zeta_k}) d^n \boldsymbol{\zeta}, \quad \boldsymbol{\theta} \in \mathbb{C}^r,$$

and consequently

$$|d_n(\boldsymbol{\theta})| \leq \frac{1}{(2\pi)^n n!} \int_{[0, 2\pi]^n} |\Delta(\boldsymbol{\zeta})|^2 \prod_{k=1}^n |w_{\boldsymbol{\theta}}(e^{i\zeta_k})| d^n \boldsymbol{\zeta}, \quad \boldsymbol{\theta} \in \mathbb{C}^r.$$

Since the functions $V_j, 1 \leq j \leq r$, are real-valued, we observe that

$$|w_{\boldsymbol{\theta}}(e^{i\zeta_k})| = \exp \left\{ \sum_{l=1}^r \operatorname{Re}(\theta_l) V_l(e^{i\zeta_k}) \right\}.$$

From the last equality it follows that it is enough to prove the locally uniform boundness for the sequence $\{d_n(\boldsymbol{\theta})\}_{n \geq 2}$ with respect to the real-valued symbols $w_{\boldsymbol{\theta}}$ with parameter $\boldsymbol{\theta} \in \mathbb{R}^r$. For the real-valued symbol $w_{\boldsymbol{\theta}}$, we have from Szegő's theory that

$$\frac{D_n[w_{\boldsymbol{\theta}}]}{D_{n-1}[w_{\boldsymbol{\theta}}]} = \min_{p \in \mathcal{P}_{n-1}} \int_0^{2\pi} |p(e^{i\zeta})|^2 w_{\boldsymbol{\theta}}(e^{i\zeta}) d\zeta, \quad n \geq 2, \quad (5.32)$$

where \mathcal{P}_n is the set of all polynomials of degree non-exceeding n , with leading coefficient equal to 1. Observe that the left-hand side in (5.32) is non-increasing in n and it tends to 1 as $n \rightarrow \infty$ because $\lim_{n \rightarrow \infty} D_n[w_{\boldsymbol{\theta}}]$ exists. Thus, for every $\boldsymbol{\theta}$, the sequence $\{D_n[w_{\boldsymbol{\theta}}]\}_{n \in \mathbb{N}}$ increases to its limit

$$\exp \left\{ \sum_{k \in \mathbb{N}} k \left(\sum_{i=1}^r \theta_i \hat{V}_i(k) \right) \left(\sum_{i=1}^r \theta_i \hat{V}_i(-k) \right) \right\}$$

as n increases. As an immediate consequence of this fact, we obtain that

$$|d_n(\boldsymbol{\theta})| \leq \exp \left\{ \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r \theta_i \theta_j \langle V_i, V_j \rangle_{1/2} \right\}, \quad \forall n \in \mathbb{N}, \boldsymbol{\theta} \in \mathbb{R}^r, \quad (5.33)$$

where $\langle V_i, V_j \rangle_{1/2}$ is the bilinear form in the Sobolev space $H_2^{1/2}$. The function on the right-hand side is bounded on compacts in \mathbb{R}^r , and the lemma is proved. \square

The strong Szegő theorem applied to the symbol $w_{\boldsymbol{\theta}} \in L^1(\mathbb{T})$, states that

$$\lim_{n \rightarrow \infty} d_n(\boldsymbol{\theta}) = \exp \left\{ \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r \theta_i \theta_j \langle V_i, V_j \rangle_{1/2} \right\}, \quad (5.34)$$

under the conditions of Theorem 5.11. Thus, we have the following lemma

Lemma 5.14. *Suppose the conditions of Theorem 5.11 are satisfied. The sequence $\{d_n(\boldsymbol{\theta})\}_{n \geq 2}$ of Toeplitz determinants with respect to the (complex-valued) generating*

function $w_{\boldsymbol{\theta}}$, converges locally uniformly in \mathbb{C}^r to its limit

$$d(\boldsymbol{\theta}) = \exp \left\{ \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^r \theta_i \theta_j \langle V_i, V_j \rangle_{1/2} \right\}. \quad (5.35)$$

Moreover, all partial derivatives of $d_n(\boldsymbol{\theta})$ converge locally uniformly to the corresponding derivatives of the limiting function (5.35).

Proof. Since the sequence of analytic functions $\{d_n(\boldsymbol{\theta})\}_{n \geq 2}$ is locally bounded and it converges to the entire function on \mathbb{C}^r , from Vitali's theorem for several complex variables it follows that it converges locally uniformly on \mathbb{C}^r . If we apply Weierstrass' theorem for several complex variables to the sequence of Toeplitz determinants $\{d_n(\boldsymbol{\theta})\}_{n \geq 2}$, we get that for arbitrary multi-index $\mathbf{l} = (l_1, \dots, l_r)$, with $\sum_{i=1}^r l_i = l, l_i \geq 0, 1 \leq i \leq r$, derivatives

$$\frac{\partial^l}{\partial \theta_1^{l_1} \dots \partial \theta_r^{l_r}} d_n(\boldsymbol{\theta})$$

exist and converge uniformly on compact sets in \mathbb{C}^r to the corresponding derivatives of the limiting function (5.35). \square

In order to obtain the probabilistic interpretation of the uniform convergence of derivatives in Lemma 5.14, we will consider the cumulant-generating function $\psi_n(\boldsymbol{\theta}) = \log \lambda_n(\boldsymbol{\theta})$ corresponding to the density $p_{\boldsymbol{\theta};n}(\boldsymbol{\zeta})$ and formulate the following corollary.

Corollary 5.15. *Let $\{p_{\boldsymbol{\theta};n}(\boldsymbol{\zeta})\}_{n \geq 2}$ be the sequence of probability density functions defined in (5.17) and suppose that the conditions of Theorem 5.11 are satisfied. Denote the cumulant-generating function of $p_{\boldsymbol{\theta};n}(\boldsymbol{\zeta})$ by $\psi_n(\boldsymbol{\theta})$. Then the sequence of gradients $\{\nabla \psi_n(\boldsymbol{\theta})\}_{n \geq 2}$ converges locally uniformly in $\boldsymbol{\theta} \in \mathbb{R}^r$ to the vector*

$$\left(\sum_{j=1}^r \theta_j \langle V_1, V_j \rangle_{1/2}, \dots, \sum_{j=1}^r \theta_j \langle V_r, V_j \rangle_{1/2} \right).$$

Proof. The statement follows from Lemma 5.14 applied to compact sets $K \subset \mathbb{R}^r$ and functions

$$\frac{\partial}{\partial \theta_j} \psi_n(\boldsymbol{\theta}) = \frac{\partial}{\partial \theta_j} d_n(\boldsymbol{\theta}) / d_n(\boldsymbol{\theta}).$$

\square

As discussed in the beginning of the section, our aim is to obtain the asymptotic maximum likelihood estimator for the vector parameter $\boldsymbol{\theta} \in \mathbb{R}^r$. Since all preliminary results are stated, we may summarize the discussion in the following theorem. We denote the true parameter value by $\boldsymbol{\theta}_0$.

Theorem 5.16. *Let $\{p_{\boldsymbol{\theta};n}(\boldsymbol{\zeta})\}_{n \geq 2}$ be the sequence of probability density functions given in (5.17), and assume that the conditions of Theorem 5.11 are satisfied. Then the asymptotic maximum likelihood estimator $\hat{\boldsymbol{\theta}}^{(n)}$ of $\boldsymbol{\theta}_0$, defined as the unique solution of the system of linear equations*

$$\sum_{k=1}^r \hat{\theta}_k^{(n)} \langle V_j, V_k \rangle_{1/2} = \sum_{l=1}^n V_j(e^{i\xi_l}), \quad 1 \leq j \leq r, \quad (5.36)$$

converges weakly to $\mathcal{N}(\boldsymbol{\theta}_0, \Sigma^{-1})$, where Σ is as specified in (5.28).

Proof. The maximum likelihood estimator $\tilde{\boldsymbol{\theta}}^{(n)}$ is defined as the solution of the system of equations

$$\frac{\partial}{\partial \theta_j} \psi_n(\tilde{\boldsymbol{\theta}}^{(n)}) = \sum_{l=1}^n V_j(e^{i\xi_l}), \quad 1 \leq j \leq r. \quad (5.37)$$

The convergence in Corollary 5.15 implies that the left-hand side in (5.37) can be approximated by the expression

$$\sum_{k=1}^r \tilde{\theta}_k^{(n)} \langle V_j, V_k \rangle_{1/2}$$

and we arrive at the system of linear equations (5.36), which defines the asymptotic maximum likelihood estimator $\hat{\boldsymbol{\theta}}^{(n)}$. Consequently, we have that

$$\hat{\boldsymbol{\theta}}^{(n)} = \mathbf{V}^{(n)}(\boldsymbol{\xi})\Sigma^{-1}. \quad (5.38)$$

From Theorem 5.11 the convergence

$$\hat{\boldsymbol{\theta}}^{(n)} \xrightarrow{\mathcal{D}} \mathcal{N}(\boldsymbol{\theta}_0, \Sigma^{-1}), \quad n \rightarrow \infty, \quad (5.39)$$

follows. □

Remark 5.17. We notice that although $\hat{\boldsymbol{\theta}}^{(n)}$ is asymptotically unbiased and follows asymptotically normal distribution, it is not consistent. This fact can be explained by the behavior of the Fisher information matrix $I_n(\boldsymbol{\theta})$ corresponding to the density

$p_{\boldsymbol{\theta};n}(\boldsymbol{\zeta})$. Namely,

$$I_n(\boldsymbol{\theta}) = \left\| \frac{\partial^2}{\partial \theta_j \partial \theta_k} \psi(\boldsymbol{\theta}) \right\|_{j,k=1,\dots,r} \quad (5.40)$$

converges to a constant matrix Σ defined in (5.28), and therefore, the information contained in the sample (ξ_1, \dots, ξ_n) of exchangeable random variables with probability density $p_{\boldsymbol{\theta};n}(\boldsymbol{\zeta})$ remains bounded as n increases. We conclude that the consistent estimation of parameter $\boldsymbol{\theta}$ is not possible, unless the dimension parameter n is introduced into the weight function $w_{\boldsymbol{\theta}}$ defined in (5.18). To remedy this deficiency, we consider the model (5.2) with varying weight and the third-order phase transition, and show that under such a model, the concentration parameter γ can be estimated consistently.

5.7 The Third-Order Phase Transition Model

In this section we consider the model from two-dimensional lattice gauge theories which was analyzed heuristically by the steepest descent method in Gross, Witten [57]. The same model arose in the studies of the length of the longest increasing subsequence in a random permutation. Its properties were rigorously analyzed in a series of papers including Johansson [70], Baik, Deift, Johansson [9], Baik, Deift, Rains [10] and Widom [117]. It was shown that the ensemble exhibits the third-order phase transition at $\gamma = 1$. Formally, the model is defined by the joint probability density of n eigenphases equal to

$$p_{\gamma;n}(\boldsymbol{\zeta}) = C_n^{-1}(\gamma) |\Delta(\boldsymbol{\zeta})|^2 \exp \left\{ \gamma n \sum_{i=1}^n \cos \zeta_i \right\}, \quad \boldsymbol{\zeta} \in [0, 2\pi]^n, \quad (5.41)$$

where $C_n(\gamma) = (2\pi)^n n! D_n[w_{\gamma n}]$ is the normalizing constant and $\gamma > 0$.

The following lemma, due to Gross, Witten [57] and Johansson [70], describes the asymptotical behavior of the free energy $f_n(\gamma) = n^{-2} \log C_n(\gamma)$.

Lemma 5.18. *If $f_n(\gamma) = n^{-2} \log C_n(\gamma)$, then*

$$\lim_{n \rightarrow \infty} f_n(\gamma) = f(\gamma) = \begin{cases} \frac{\gamma^2}{4}, & \text{if } 0 \leq \gamma \leq 1, \\ \gamma - \frac{1}{2} \log \gamma - \frac{3}{4}, & \text{if } 1 < \gamma. \end{cases} \quad (5.42)$$

Remark 5.19. The limiting function $f(\gamma)$ in Lemma 5.18 is not analytic. Its derivative $d^3 f/d\gamma^3$ is discontinuous at $\gamma = 1$, thus the third-order phase transition occurs at this point. The asymptotic eigenvalue distribution is supported on the whole unit

circle for $\gamma < 1$, whereas for $\gamma > 1$ its support is the subset of \mathbb{T} . For more details regarding the limiting distribution one is referred to Johansson [70].

In the lemma that follows we show that the limit (5.42) can be used to prove the asymptotic normality of the sufficient statistics $T_n(\xi_1, \dots, \xi_n)$ defined below.

Lemma 5.20. *Let $\epsilon > 0, \gamma \in (0, 1 - \epsilon]$, and $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)$ be a sample with the joint probability density function (5.41). The sufficient statistics*

$$T_n(\xi_1, \dots, \xi_n) = \sum_{i=1}^n \cos \xi_i - n\gamma/2 \quad (5.43)$$

converges in distribution to the normal distribution with 0 mean and variance 1/2.

Proof. The moment-generating function of the statistics $T_n(\boldsymbol{\xi})$ is

$$\begin{aligned} M_n(s) &= C_n^{-1}(\gamma) e^{-s\gamma n/2} \int_{[0, 2\pi]^n} |\Delta(\boldsymbol{\zeta})|^2 \prod_{i=1}^n \exp \left\{ (\gamma n + s) \sum_{i=1}^n \cos \zeta_i \right\} d^n \boldsymbol{\zeta} \\ &= e^{-s\gamma n/2} \frac{D_n[w_{\gamma n + s}]}{D_n[w_{\gamma n}]}, \end{aligned}$$

for $s \in \mathbb{R}$. Consider $\gamma \in (0, 1 - \epsilon]$ and the sequence of functions $\{M_n(s)\}_{n \in \mathbb{N}}$ in the interval $s \in I_\gamma = [-\min(\gamma, 1 - \gamma), \min(\gamma, 1 - \gamma)]$. From the construction of I_γ , we have that $0 \leq \gamma + s/n \leq 1$ for every $s \in I_\gamma, n \geq 1$. It was shown in Johansson [72] that for every $\epsilon > 0$ there exists a constant $C > 0$ such that

$$\left| \log C_n(\gamma) - \frac{n^2 \gamma^2}{4} \right| \leq \frac{C}{n}, \quad (5.44)$$

for arbitrary $n \geq 1, \gamma \in [0, 1 - \epsilon]$. Consequently,

$$e^{-C/n} \leq C_n(\gamma) e^{-\gamma^2 n^2/4} \leq e^{C/n}, \quad n \geq 1 \quad (5.45)$$

uniformly for $\gamma \in [0, 1 - \epsilon]$. Thus, we have the following asymptotics

$$\begin{aligned} M_n(s) &= \exp \left\{ -\frac{sn\gamma}{2} \right\} \exp \left\{ \frac{n^2}{4} \left(\gamma + \frac{s}{n} \right)^2 - \frac{n^2}{4} \gamma^2 + O \left(\frac{1}{n} \right) \right\} \\ &= \exp \left\{ \frac{s^2}{4} + O \left(\frac{1}{n} \right) \right\}, \quad \forall s \in I_\gamma. \end{aligned}$$

Therefore,

$$\lim_{n \rightarrow \infty} M_n(s) = \exp(s^2/4), \quad s \in I_\gamma,$$

where the function on the right hand side is the moment generating function of normal distribution. Since the convergence of the moment-generating functions in the neighborhood of the origin implies weak convergence of distributions (see Billingsley [27]), we have obtained the asymptotical normality of the statistics $T_n(\boldsymbol{\xi})$. \square

Corollary 5.21. *If conditions of Lemma 5.20 are satisfied, the asymptotical maximum likelihood estimator*

$$\tilde{\gamma}_n = \frac{2}{n} \sum_{i=1}^n \cos \xi_i \quad (5.46)$$

is asymptotically unbiased and

$$n(\tilde{\gamma}_n - \gamma) \xrightarrow{\mathcal{D}} N(0, 2) \quad (5.47)$$

as $n \rightarrow \infty$.

Proof. The maximum likelihood estimator $\hat{\gamma}_n$ of parameter γ is obtained from the equation

$$\frac{d}{d\gamma} \log C_n(\hat{\gamma}_n) = n \sum_{i=1}^n \cos \xi_i. \quad (5.48)$$

Since it was proved in Johansson [70] that there exists a constant $C > 0$, such that

$$\left| \frac{d}{d\gamma} \log C_n(\gamma) - \frac{n^2 \gamma}{2} \right| \leq \frac{C}{n},$$

for every $n \geq 1, \gamma \in (0, 1 - \epsilon]$, we replace the equation (5.48) by

$$\frac{n^2 \tilde{\gamma}_n}{2} = n \sum_{i=1}^n \cos \xi_i, \quad (5.49)$$

which leads to the estimator

$$\tilde{\gamma}_n = \frac{2 \sum_{i=1}^n \cos \xi_i}{n}. \quad (5.50)$$

We notice that

$$\mathbb{E}_n[\tilde{\gamma}_n - \gamma] = \mathbb{E}_n[2T_n(\boldsymbol{\xi})/n] \rightarrow 0, \quad n \rightarrow \infty,$$

and

$$n(\tilde{\gamma}_n - \gamma) \xrightarrow{\mathcal{D}} N(0, 2), \quad n \rightarrow \infty.$$

The statement of the corollary is obtained. \square

Remark 5.22. The case $\gamma > 1$ is not considered here. In this instance, the maximum likelihood equation (5.48) can be replaced by the asymptotical expression

$$1 - \frac{1}{2\tilde{\gamma}_n} = \frac{\sum_{i=1}^n \cos \xi_i}{n} \quad (5.51)$$

that follows from Lemma 5.42 and leads us to the equation which defines the asymptotic maximum likelihood estimator $\tilde{\gamma}_n$. Asymptotical properties of $\tilde{\gamma}_n$ are the subject of our further investigation. Heuristical derivation of the asymptotical distribution of sufficient statistics

$$T_n(\xi_1, \dots, \xi_n) = \sum_{i=1}^n \cos \xi_i - n \left(1 - \frac{1}{2\gamma}\right)$$

leads us to the following convergence of moment-generating functions

$$\lim_{n \rightarrow \infty} M_n(s) = \exp \left[\frac{s^2}{2} \frac{1}{2\gamma^2} \right], \quad s \in [1 - \gamma, \gamma - 1], \quad (5.52)$$

and consequently, the asymptotical distribution of $T_n(\boldsymbol{\xi})$ is normal with variance $(2\gamma^2)^{-1}$. However, to prove the convergence (5.52) rigorously, one needs to establish that for $\gamma > 1$, the rate of convergence to the limiting function in Lemma 5.42 is of order $o(n^{-2})$.

Chapter 6

Summary and Open Questions

This thesis presented a brief historical overview of random matrix theory and some existing theoretical results. We developed statistical estimation theory for certain classes of random matrix ensembles. We derived exact and asymptotical distributions of the sufficient statistics for β -Hermite, β -Laguerre, β -Jacobi and Cauchy unitary ensemble. Furthermore, we examined properties of the maximum likelihood estimators of their parameters. Two discrete orthogonal polynomial ensembles (the Krawtchouk and Charlier ensemble) have been considered, and the maximum likelihood estimators of their parameters were obtained in closed form. We generalize Dyson's circular unitary ensemble, and analyze the properties of the sufficient statistics and asymptotic maximum likelihood estimator for the generalized circular unitary ensemble. This was followed by the discussion of the third-order phase transition model of Gross and Witten [57].

Throughout the thesis, we have formulated numerous proposals for future research. Here we highlight the major open questions to be addressed from pure mathematics point of view, as well as from the perspective of applications.

- (1) In Section 3.1 we have noticed that the estimation problem can be dealt with within the framework of non-parametric mathematical statistics. One of the possible approaches in non-parametrical setting would be the implementation of the results of Deift [37] into the estimation procedure. Numerical simulations strongly support further investigation in this direction.
- (2) Among the major non-parametrical questions related to the further development of the statistical estimation theory for random matrix ensembles we mention the *kernel density estimation* of the level density. The motivation for the theoretical analysis comes from numerical simulations that we per-

formed. Though the numerical results are not presented here, we observed that the kernel density estimator $\hat{\rho}_n(x)$, based on n eigenvalues of random matrix ensemble, converges to the theoretical limiting density $\rho(x)$ considerably faster than the corresponding estimator based on n i.i.d. observations with the probability density $\rho(x)$. The greater rate of convergence can be explained by repulsiveness property. The main problem in this issue is the evaluation of multiple integrals from which the properties of the bias, variance and mean-squared error of $\hat{\rho}_n(x)$ should be derived.

- (3) Possible extensions of the results of Chapter 4 for the weight function with general exponential form (1.2) should be carried out. Additionally, as mentioned in Section 3.1, the repulsiveness parameter β can be regarded as a parameter to be estimated.
- (4) The results of Chapter 5 can be extended in several directions. The problem of construction of the explicit random matrix models for the generalized circular unitary ensembles remains open. The estimation of $\gamma > 1$ in the Gross–Witten third order phase transition model (5.41) is one of the main questions to be addressed in our future research.

Bibliography

- [1] M. Abramowitz and I. Stegun, editors. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Dover Publications Inc., New York, 1992. Reprint of the 1972 edition.
- [2] M. Adler and P. van Moerbeke. Recursion relations for unitary integrals, combinatorics and the Toeplitz lattice. *Comm. Math. Phys.*, 237(3):397–440, 2003.
- [3] G. W. Anderson. A short proof of Selberg’s generalized beta formula. *Forum Math.*, 3(4):415–417, 1991.
- [4] T. W. Anderson. *An introduction to multivariate statistical analysis*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, third edition, 2003.
- [5] G. E. Andrews, R. Askey, and R. Roy. *Special functions*, volume 71 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1999.
- [6] K. Aomoto. Jacobi polynomials associated with Selberg integrals. *SIAM J. Math. Anal.*, 18(2):545–549, 1987.
- [7] L. Arnold. On the asymptotic distribution of the eigenvalues of random matrices. *J. Math. Anal. Appl.*, 20:262–268, 1967.
- [8] L. Arnold. On Wigner’s semicircle law for the eigenvalues of random matrices. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 19:191–198, 1971.
- [9] J. Baik, P. Deift, and K. Johansson. On the distribution of the length of the longest increasing subsequence of random permutations. *J. Amer. Math. Soc.*, 12(4):1119–1178, 1999.

-
- [10] J. Baik, P. Deift, and E. Rains. A Fredholm determinant identity and the convergence of moments for random Young tableaux. *Comm. Math. Phys.*, 223(3):627–672, 2001.
- [11] Z. D. Bai, B. Miao, and J. Tsay. Remarks on the convergence rate of the spectral distributions of Wigner matrices. *J. Theoret. Probab.*, 12(2):301–311, 1999.
- [12] Z. D. Bai, B. Miao, and J. Tsay. Convergence rates of the spectral distributions of large Wigner matrices. *Int. Math. J.*, 1(1):65–90, 2002.
- [13] Z. D. Bai and J. W. Silverstein. CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.*, 32(1A):553–605, 2004.
- [14] Z. D. Bai. Convergence rate of expected spectral distributions of large random matrices. II. Sample covariance matrices. *Ann. Probab.*, 21(2):649–672, 1993.
- [15] Z. D. Bai. Convergence rate of expected spectral distributions of large random matrices. I. Wigner matrices. *Ann. Probab.*, 21(2):625–648, 1993.
- [16] Z. D. Bai. Methodologies in spectral analysis of large-dimensional random matrices, a review. *Statist. Sinica*, 9(3):611–677, 1999. With comments by G. J. Rodgers and Jack W. Silverstein; and a rejoinder by the author.
- [17] T. H. Baker and P. J. Forrester. The Calogero-Sutherland model and generalized classical polynomials. *Comm. Math. Phys.*, 188(1):175–216, 1997.
- [18] T. H. Baker and P. J. Forrester. Finite- N fluctuation formulas for random matrices. *J. Statist. Phys.*, 88(5-6):1371–1386, 1997.
- [19] R. Balian. Random matrices and information theory. *Nuovo Cimento B*, (57):183–193, 1968.
- [20] O. Barndorff-Nielsen. *Exponential families: Exact theory*. Matematisk Institut, Aarhus Universitet, Aarhus, 1970. Various Publication Series, No. 19.
- [21] O. Barndorff-Nielsen. *Information and exponential families in statistical theory*. John Wiley & Sons Ltd., Chichester, 1978. Wiley Series in Probability and Mathematical Statistics.
- [22] E. L. Basor. Distribution functions for random variables for ensembles of positive Hermitian matrices. *Comm. Math. Phys.*, 188(2):327–350, 1997.
- [23] G. Baxter. A convergence equivalence related to polynomials orthogonal on the unit circle. *Trans. Amer. Math. Soc.*, 99:471–487, 1961.

-
- [24] C. W. J. Beenakker. Random-matrix theory of mesoscopic fluctuations in conductors and superconductors. *Phys. Rev. B*, 47:15763–15775, 1993.
- [25] C. W. J. Beenakker. Universality in the random-matrix theory of quantum transport. *Phys. Rev. Lett.*, 70:1155–1158, 1993.
- [26] C. W. J. Beenakker. Random-matrix theory of quantum transport. *Rev. Mod. Phys.*, 69(3):731–808, Jul 1997.
- [27] P. Billingsley. *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, third edition, 1995. A Wiley-Interscience Publication.
- [28] A. Borodin and D. Boyarchenko. Distribution of the first particle in discrete orthogonal polynomial ensembles. *Comm. Math. Phys.*, 234(2):287–338, 2003.
- [29] A. Borodin and A. Okounkov. A Fredholm determinant formula for Toeplitz determinants. *Integral Equations Operator Theory*, 37(4):386–396, 2000.
- [30] A. Borodin and G. Olshanski. Harmonic analysis on the infinite-dimensional unitary group and determinantal point processes. *Ann. of Math. (2)*, 161(3):1319–1422, 2005.
- [31] A. Borodin. Discrete gap probabilities and discrete Painlevé equations. *Duke Math. J.*, 117(3):489–542, 2003.
- [32] A. Böttcher and B. Silbermann. *Analysis of Toeplitz operators*. Springer-Verlag, Berlin, 1990.
- [33] P. W. Brouwer. Generalized circular ensemble of scattering matrices for a chaotic cavity with non-ideal leads. *Phys. Rev. B*, 51:16878, 1995.
- [34] L. D. Brown. *Fundamentals of statistical exponential families with applications in statistical decision theory*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 9. Institute of Mathematical Statistics, Hayward, CA, 1986.
- [35] N. N. Chentsov. *Statisticheskie reshayushchie pravila i optimalnye vyvody*. 1972.
- [36] O. Costin and J. Lebowitz. Gaussian fluctuations in random matrices. *Phys. Rev. Lett.*, 75:69–72, 1995.
- [37] P. A. Deift. *Orthogonal polynomials and random matrices: a Riemann-Hilbert approach*, volume 3 of *Courant Lecture Notes in Mathematics*. New York University Courant Institute of Mathematical Sciences, New York, 1999.

- [38] A. Devinatz. An extension of a limit theorem of G. Szegő. *J. Math. Anal. Appl.*, 14:499–510, 1966.
- [39] A. Devinatz. The strong Szegő limit theorem. *Illinois J. Math.*, 11:160–175, 1967.
- [40] P. Diaconis and S. N. Evans. Linear functionals of eigenvalues of random matrices. *Trans. Amer. Math. Soc.*, 353(7):2615–2633 (electronic), 2001.
- [41] P. Di Francesco, P. Ginsparg, and J. Zinn-Justin. 2D gravity and random matrices. *Phys. Rep.*, 254(1-2):133, 1995.
- [42] I. Dumitriu and A. Edelman. Matrix models for beta ensembles. *J. Math. Phys.*, 43(11):5830–5847, 2002.
- [43] I. Dumitriu and A. Edelman. Global spectrum fluctuations for the β -Hemite and β -Lagurre ensembles via matrix models. Preprint lanl.arxiv.org/abs/math-ph/0510043. 2005.
- [44] I. Dumitriu. Eigenvalue statistics for β -ensembles. PhD thesis. Massachusetts Institute of Technology. 2003.
- [45] F. J. Dyson. Statistical theory of the energy levels of complex systems. I. *J. Mathematical Phys.*, 3:140–156, 1962.
- [46] R. A. Fisher. Frequency distribution of the values of the correlation coefficient in samples from an infinitely large population. *Biometrika*, 10:507–521, 1915.
- [47] P. Forrester. *Log-gases and Random Matrices*. Book Manuscript www.ms.unimelb.edu.au/~matpjf/matpjf.html.
- [48] J. L. Geronimus. On a problem of G. Szegő, M. Kac, G. Baxter and I. Hirschman. *Izv. Akad. Nauk SSSR Ser. Mat.*, 31:289–304, 1967.
- [49] V. L. Girko. Asymptotics of the distribution of the spectrum of random matrices. *Uspekhi Mat. Nauk*, 44(4(268)):7–34, 256, 1989.
- [50] V. L. Girko. Convergence rate of the expected spectral functions of symmetric random matrices is equal to $O(n^{-1/2})$. *Random Oper. Stochastic Equations*, 6(4):359–408, 1998.
- [51] F. Götze and A. Tikhomirov. Rate of convergence to the semi-circular law for the Gaussian unitary ensemble. *Teor. Veroyatnost. i Primenen.*, 47(2):381–387, 2002.

- [52] F. Götze and A. Tikhomirov. Rate of convergence to the semi-circular law. *Probab. Theory Related Fields*, 127(2):228–276, 2003.
- [53] F. Götze and A. Tikhomirov. Rate of convergence in probability to the Marchenko-Pastur law. *Bernoulli*, 10(3):503–548, 2004.
- [54] I. S. Gradshteyn and I. M. Ryzhik. *Table of integrals, series, and products*. Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1980.
- [55] U. Grenander and G. Szegő. *Toeplitz forms and their applications*. Chelsea Publishing Co., New York, second edition, 1984.
- [56] U. Grenander. *Probabilities on algebraic structures*. John Wiley & Sons Inc., New York, 1963.
- [57] D. J. Gross and E. Witten. Possible third-order phase transition in the large- n lattice gauge theory. *Phys. Rev. D*, 21(2):446–453, Jan 1980.
- [58] T. Guhr, A. Müller-Groeling, and H. A. Weidenmüller. Random-matrix theories in quantum physics: common concepts. *Phys. Rep.*, 299(4-6):189–425, 1998.
- [59] F. Haake. *Quantum signatures of chaos*. Springer Series in Synergetics. Springer-Verlag, Berlin, second edition, 2001. With a foreword by H. Haken.
- [60] I. I. Hirschman, Jr. Finite sections of Wiener-Hopf equations and Szegő polynomials. *J. Math. Anal. appl.*, 11:290–320, 1965.
- [61] I. I. Hirschman, Jr. On a theorem of Szegő, Kac, and Baxter. *J. Analyse Math.*, 14:225–234, 1965.
- [62] I. I. Hirschman, Jr. The strong Szegő limit theorem for Toeplitz determinants. *Amer. J. Math.*, 88:577–614, 1966.
- [63] P. L. Hsu. On the distribution of roots of certain determinantal equations. *Ann. Eugenics*, 9:250–258, 1939.
- [64] I. A. Ibragimov. A theorem of Gabor Szegő. *Mat. Zametki*, 3:693–702, 1968.
- [65] M. E. H. Ismail and N. S. Witte. Discriminants and functional equations for polynomials orthogonal on the unit circle. *J. Approx. Theory*, 110(2):200–228, 2001.
- [66] R.A. Jalabert and Pichard J. Quantum mesoscopic scattering: Disordered systems and dyson circular ensembles. *J. Phys. I France*, 5:287–324, 1995.

- [67] A. T. James. Distributions of matrix variates and latent roots derived from normal samples. *Ann. Math. Statist.*, 35:475–501, 1964.
- [68] S. R. Jammalamadaka and A. SenGupta. *Topics in circular statistics*, volume 5 of *Series on Multivariate Analysis*. World Scientific Publishing Co. Inc., River Edge, NJ, 2001. With 1 IBM-PC floppy disk (3.5 inch; HD).
- [69] K. Johansson. On random matrices from the compact classical groups. *Ann. of Math. (2)*, 145(3):519–545, 1997.
- [70] K. Johansson. The longest increasing subsequence in a random permutation and a unitary random matrix model. *Math. Res. Lett.*, 5(1-2):63–82, 1998.
- [71] K. Johansson. On fluctuations of eigenvalues of random Hermitian matrices. *Duke Math. J.*, 91(1):151–204, 1998.
- [72] K. Johansson. Shape fluctuations and random matrices. *Comm. Math. Phys.*, 209(2):437–476, 2000.
- [73] K. Johansson. Discrete orthogonal polynomial ensembles and the Plancherel measure. *Ann. of Math. (2)*, 153(1):259–296, 2001.
- [74] K. Johansson. Non-intersecting paths, random tilings and random matrices. *Probab. Theory Related Fields*, 123(2):225–280, 2002.
- [75] N. L. Johnson and S. Kotz. *Distributions in statistics: continuous multivariate distributions*. John Wiley & Sons Inc., New York, 1972. Wiley Series in Probability and Mathematical Statistics.
- [76] I. M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.*, 29(2):295–327, 2001.
- [77] D. Jonsson. Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.*, 12(1):1–38, 1982.
- [78] J. V. José and R. Cordery. Study of a quantum fermi-acceleration model. *Phys. Rev. Lett.*, 56(4):290–293, Jan 1986.
- [79] M. Kac. Toeplitz matrices, translation kernels and a related problem in probability theory. *Duke Math. J.*, 21:501–509, 1954.
- [80] N. M. Katz and P. Sarnak. *Random matrices, Frobenius eigenvalues, and monodromy*, volume 45 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 1999.

- [81] J. P. Keating and N. C. Snaith. Random matrices and L -functions. *J. Phys. A*, 36(12):2859–2881, 2003.
- [82] R. Killip and I. Nenciu. Matrix models for circular ensembles. *Int. Math. Res. Not.*, (50):2665–2701, 2004.
- [83] W. König. Orthogonal polynomial ensembles in probability theory. *Probab. Surv.*, 2:385–447 (electronic), 2005.
- [84] M. Ledoux. Differential operators and spectral distributions of invariant ensembles from the classical orthogonal polynomials: the discrete case. *Electron. J. Probab.*, 10:no. 34, 1116–1146 (electronic), 2005.
- [85] E. L. Lehmann and G. Casella. *Theory of point estimation*. Springer Texts in Statistics. Springer-Verlag, New York, second edition, 1998.
- [86] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)*, 72 (114):507–536, 1967.
- [87] M. L. Mehta. *Random matrices*, volume 142 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, third edition, 2004.
- [88] R. J. Muirhead. *Aspects of multivariate statistical theory*. John Wiley & Sons Inc., New York, 1982. Wiley Series in Probability and Mathematical Statistics.
- [89] K.E. Muttalib and M.E.H. Ismail. Impact of localization on dyson’s circular ensemble. *J. Phys. A: Math. Gen.*, (28):L541–L548, 1995.
- [90] S. Nadarajah and S. Kotz. On the ratio of Pearson type VII and Bessel random variables. *J. Appl. Mat. Dec. Sci.*, (4):191–199, 2005.
- [91] T. Nagao and M. Wadati. An integration method on generalized circular ensembles. *J. Phy. Soc. Jpn.*, 61(6):1903–1909, 1992.
- [92] I. Olkin and H. Rubin. Multivariate beta distributions and independence properties of the Wishart distribution. *Ann. Math. Statist*, 35:261–269, 1964.
- [93] A. Pasquale. Weyl’s integration formula for $U(N)$. Lecture given at the DMV Seminar “The Riemann Zeta Function and Random Matrix Theory”, Oberwolfach, www.math.tau.ac.il/~rudnick/dmv.html.
- [94] V. Periwal and D. Shevitz. Unitary-matrix models as exactly solvable string theories. *Phys. Rev. Lett.*, 64(12):1326–1329, Mar 1990.
- [95] D. Politzer. Random-matrix description of the distribution of mesoscopic conductance. *Phys, Rev. B*, 40:11917–11919, 1989.

- [96] C. E. Porter and N. Rosenzweig. Statistical properties of atomic and nuclear spectra. *Ann. Acad. Sci. Fenn. Ser. A VI No.*, 44:66, 1960.
- [97] A. Selberg. Remarks on a multiple integral. *Norsk Mat. Tidsskr.*, 26:71–78, 1944.
- [98] J. W. Silverstein and Z. D. Bai. On the empirical distribution of eigenvalues of a class of large-dimensional random matrices. *J. Multivariate Anal.*, 54(2):175–192, 1995.
- [99] B. Simon. *Orthogonal polynomials on the unit circle. Part 1*, volume 54 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2005. Classical theory.
- [100] Y. Sinai and A. Soshnikov. Central limit theorem for traces of large random symmetric matrices with independent matrix elements. *Bol. Soc. Brasil. Mat. (N.S.)*, 29(1):1–24, 1998.
- [101] A. Soshnikov. Universality at the edge of the spectrum in Wigner random matrices. *Comm. Math. Phys.*, 207(3):697–733, 1999.
- [102] A. Soshnikov. The central limit theorem for local linear statistics in classical compact groups and related combinatorial identities. *Ann. Probab.*, 28(3):1353–1370, 2000.
- [103] A. Soshnikov. Gaussian limit for determinantal random point fields. *Ann. Probab.*, 30(1):171–187, 2002.
- [104] A. Soshnikov. A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *J. Statist. Phys.*, 108(5-6):1033–1056, 2002. Dedicated to David Ruelle and Yasha Sinai on the occasion of their 65th birthdays.
- [105] G. Szegő. *Orthogonal polynomials*. American Mathematical Society, Providence, R.I., fourth edition, 1975. American Mathematical Society, Colloquium Publications, Vol. XXIII.
- [106] M. Tierz. The stable random matrix models. Preprint [lanl.arxiv.org/abs/cond-mat/0106485](https://arxiv.org/abs/cond-mat/0106485). 2003.
- [107] C. A. Tracy and H. Widom. Fredholm determinants, differential equations and matrix models. *Comm. Math. Phys.*, 163, 1994.
- [108] C. A. Tracy and H. Widom. Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.*, 159(1):151–174, 1994.

-
- [109] C. A. Tracy and H. Widom. Level spacing distributions and the Bessel kernel. *Comm. Math. Phys.*, 161(2):289–309, 1994.
- [110] C. A. Tracy and H. Widom. On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.*, 177(3):727–754, 1996.
- [111] C. A. Tracy and H. Widom. Random unitary matrices, permutations and Painlevé. *Comm. Math. Phys.*, 207(3):665–685, 1999.
- [112] C. A. Tracy and H. Widom. Distribution functions for largest eigenvalues and their applications. In *Proceedings of the International Congress of Mathematicians, Vol. I (Beijing, 2002)*, pages 587–596, Beijing, 2002. Higher Ed. Press.
- [113] A. M. Tulino and S. Verdú. Random matrix theory and wireless communications. *Foundation and Trends in Communications and Information Theory*, 1(1):1–182, 2004.
- [114] H. Weyl. Theorie der darstellungen kontinuierlicher halbeinfacher gruppen durch linear transformationen. i. *Math. Z.*, (23):271 – 309, 1924.
- [115] H. Weyl. Theorie der darstellungen kontinuierlicher halbeinfacher gruppen durch linear transformationen. ii. *Math. Z.*, (24):328 – 376, 1925.
- [116] H. Weyl. *The classical groups. Their invariants and representations*. Princeton University Press, Princeton, NJ, 1997.
- [117] H. Widom. Asymptotic behavior of block Toeplitz matrices and determinants. II. *Advances in Math.*, 21(1):1–29, 1976.
- [118] E. P. Wigner. On the statistical distribution of the widths and spacings of nuclear resonance levels. *Proc. Cam. Phil. Soc.*, 47:790–798, 1951.
- [119] E. P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math. (2)*, 62:548–564, 1955.
- [120] E. P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. of Math. (2)*, 67:325–327, 1958.
- [121] J. Wishart. The generalized product moment distribution in samples from a normal multivariate population. *Biometrika*, 20A:32–43, 1928.

Curriculum Vitae

Geburtsdatum:	24.04.1979
Geburtsort :	Loznica, Serbien
1986–1994 :	Grundschule Dositej Obradović, Klupci, Loznica
1994–1998 :	Gymnasium Vuk Karadžić, Loznica
Sept. 1998 bis Juni 2003:	Mathematikstudium, Studienrichtung Theoretische Mathematik, Abteilung Stochastik, Staatliche Universität Sankt-Petersburg, Russland
11.06.2003	Abschluss: Diplom
seit September 2003:	Doktorandin am Institut für Mathematische Stochastik der Universität Göttingen bei Herrn Prof. Dr. Manfred Denker mit dem Ziel der Promotion
seit September 2003:	Teilnahme am interdisziplinären Promotionsstudiengang "Angewandte Statistik und Empirische Methoden" der Universität Göttingen
Mai 2005 bis Okt. 2005:	wissenschaftliche Mitarbeiterin am Institut für Mathematische Stochastik der Universität Göttingen bei Herrn Prof. Dr. Manfred Denker und Herrn Prof. Dr. Edgar Brunner

