ESTIMATES FOR CORRELATION FUNCTIONS OF REAL ROOTS FOR RANDOM POLYNOMIALS AND APPLICATIONS

Nhan Du Vi Nguyen
Binh Dinh, Vietnam

BS in Mathematics, Quy Nhon University, 2007
MS in Mathematics, University of West Georgia, 2017

A Dissertation presented to the Graduate Faculty
of the University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

Department of Mathematics

University of Virginia
May 2023

Yen Do, Advisor
Tai Melcher
Leonid Petrov
Jianhui Zhou

Abstract

A random polynomial is a polynomial whose coefficients are random variables. A major task in the theory of random polynomials is to examine how the real roots are distributed and correlated in situations where the degree of the polynomial is large. In this dissertation, we investigate two classes of random polynomials that have piqued the interest of researchers in probability theory and mathematical physics: elliptic polynomials and generalized Kac polynomials.

Regarding elliptic polynomials, we obtain asymptotic expansions for the variances of the number of real roots on intervals whose endpoints may vary based on the polynomial's degree. Additionally, we provide sharp estimates for the cumulants of these quantities. As applications, we can determine intervals on which the number of real roots satisfies a central limit theorem and a strong law of large numbers.

Our next objective is to compute the precise leading asymptotics of the variance of the number of real roots for generalized Kac polynomials whose coefficients have polynomial asymptotics. Examples of this class of random polynomials include Kac polynomials, hyperbolic polynomials, and any linear combinations of their derivatives. Before our work, such variance asymptotics had only been established for the Kac polynomials in the 1970s, thanks to Maslova's influential contribution. Our proof relies on novel asymptotic estimates for the real roots' two-point correlation function, which exposes geometric features in the distribution of real roots for these random polynomials. As a corollary, we establish asymptotic normality for the number of real roots of these random polynomials, extending and enhancing a related result of O. Nguyen and V. Vu.

Acknowledgments

First and foremost, I am deeply indebted to Yen Do, my esteemed advisor, for his invaluable advice, support, and assistance. I could not have undertaken this journey without the other members of my dissertation committee, Tai Melcher, Leonid Petrov, and Jianhui Zhou, who kindly contributed their knowledge and experience.

Furthermore, my studies would not have been feasible without the kind funding of the University of Virginia and the Jefferson Scholars Foundation.

I am grateful to the professors, colleagues, and friends at the University of Virginia for their generous aid and support, which helped to make my education and life in the United States so delightful.

I would also like to extend my deepest gratitude to Professors Dinh Thanh Duc and Vu Kim Tuan for their help throughout my life. Special thanks go to everyone in the Department of Mathematics and Statistics at Quy Nhon University for their encouragement.

Finally, I would be remiss in not mentioning my family, especially my parents, spouse, and children. Their belief in me has kept my spirits and motivation high during this process.

Contents

Chapter 1

Introduction

1.1 Real roots of random polynomials

Random polynomials, so simple and innocent at first sight but difficult to understand, have attracted generations of mathematicians. Apart from being of interest from a probabilistic viewpoint, random polynomials arise naturally and have applications in various fields of physics, engineering, and economics. Random polynomials serve as a basic model for the eigenfunctions of chaotic quantum systems [15, 16]. We can also find their use in filtering theory, statistical communication theory, and the analysis of capital and investment in mathematical economics [10].

Let n be a positive integer, $p_0, ..., p_n$ be deterministic polynomials, and let $\xi_0, ..., \xi_n$ be independent random variables. The linear combination

$$P_n(x) := \sum_{j=0}^{n} \xi_j p_j(x)$$

is an example of a random polynomial. If we normalize ξ_j so that $\textbf{Var}[\xi_j] = 1$, then different definitions of $p_j(x)$ give rise to various classes of random polynomials. Some classes of random polynomials that are objects of much interest in probability theory and have attracted research attention in mathematical physics include

1. elliptic polynomials (or binomial polynomials) in which $p_j(x) = \sqrt{\binom{n}{j}} x^j$;
2. Kac polynomials in which $p_j(x) = x^j$, and more generally, generalized Kac polynomials in which $p_j(x) = c_j x^j$ with c_j having power growth;
3. Weyl polynomials (or flat polynomials) in which $p_j(x) = \frac{1}{\sqrt{j!}} x^j$;
4. orthogonal polynomials in which $p_j(x)$ form a system of orthonormal polynomials with respect to a fixed compactly supported measure; and
5. trigonometric polynomials in which $p_j(x)$ are trigonometric polynomials.

For any subset I of $\mathbb{R}$, where I may depend on n, we denote by $N_n(I)$ the number of real roots of $P_n(x)$ in I (counting multiplicity). In particular, $N_n(\mathbb{R})$ is the total number of real roots. These are random variables taking values in $\{0, 1, ..., n\}$, and a central research direction in the theory of random polynomials is to characterize their statistics when n is large [69].

Most earlier studies about real roots of random polynomials focused on computing the average value for $N_n(\mathbb{R})$ for the Kac polynomials, starting from Bloch and Pólya [14], with seminal contributions of Littlewood and Offord [49, 50, 51], Kac [43, 44], Ibragimov and Maslova [38, 39, 40, 41]. Many classical results with numerous references related to the subject are given in the books by Bharucha-Reid and Sambandham [10] and by Farahmand [32]. We refer the reader to the articles [69, 27] for more comprehensive literature reviews.

When ξ_j are independent standard normal random variables, Kac [43] derived an exact formula for $\mathbf{E}[N_n(\mathbb{R})]$, nowadays known as the Kac-Rice formula, and showed that

$$\mathbf{E}[N_n(\mathbb{R})] = \frac{2}{\pi}\log n + o(\log n).$$

An elementary geometric derivation of the Kac-Rice formula was provided by Edelman and Kostlan [30], who showed that $\mathbf{E}[N_n(\mathbb{R})]$ is simply the length of the moment curve $x \mapsto (p_0(x), ..., p_n(x))$ projected onto the surface of the unit sphere, divided by π. If ξ_j are not Gaussian, the key ingredient is the universality method, whose general idea is to reduce the problem of calculating the distribution of the roots and the interaction between them to the case where ξ_j are Gaussian (see Nguyen and Vu [59], Tao and Vu [69]). By now, we can determine $\mathbf{E}[N_n(\mathbb{R})]$ for many classes of random polynomials, with various choices for $p_j(x)$ and under very general assumptions for ξ_j (see Do [22], Do, H. Nguyen and Vu [25], Do, O. Nguyen and Vu [27, 28], Nguyen and Vu [59], and the references given there).

Estimating the variance $\mathbf{Var}[N_n(\mathbb{R})]$, however, has proved to be a much more difficult task, and it is apparent that this problem still awaits rigorous treatment. Despite a large number of prior studies, only a few are about $\mathbf{Var}[N_n(\mathbb{R})]$. For Kac polynomials, Maslova [54] proved that if $\mathbf{P}(\{\xi_j = 0\}) = 0$, $\mathbf{E}[\xi_j] = 0$, and $\mathbf{E}[|\xi_j|^{2+\varepsilon}] < \infty$ for some $\varepsilon > 0$, then

$$\mathbf{Var}[N_n(\mathbb{R})] = \left[\frac{4}{\pi}\left(1 - \frac{2}{\pi}\right) + o(1)\right]\log n.$$

While the condition $\mathbf{P}(\{\xi_j = 0\}) = 0$ has been removed by O. Nguyen and V. Vu [60], there has been no other result of this type for other generalized Kac polynomials (even for the Gaussian setting when ξ_j are all Gaussian). Beyond Kac polynomials, examining the asymptotics of $\mathbf{Var}[N_n(\mathbb{R})]$ for other models of random polynomials has been extensively considered since the 1990s and has become an active direction of research in recent years. Thanks to the Kac-Rice formula and the universality method, the leading asymptotics of the variance of the real roots were established for elliptic polynomials (Bleher and Di [11], Dalmao [18]), Weyl polynomials (Do and Vu [29], Schehr and Majumdar [66]), orthogonal polynomials (Lubinsky and Pritsker [52, 53]), and for trigonometric polynomials (Bally, Caramellino, and Poly [9], Do, H. Nguyen and O. Nguyen [23], Granville and Wigman [37]). The important point to note here is that most works focus on the case where ξ_j are all Gaussian and that the second terms in the variance asymptotics for these random models are still unknown. This dissertation was intended as an attempt to find the true nature of the error term in the asymptotic estimates for the variance of the number of real roots.

Establishing the limiting law of $N_n(\mathbb{R})$ is a more challenging problem. We say that $N_n(\mathbb{R})$ satisfies the central limit theorem (CLT for short) if we have the following convergence in distribution:

$$\frac{N_n(\mathbb{R}) - \mathbf{E}[N_n(\mathbb{R})]}{\sqrt{\mathbf{Var}[N_n(\mathbb{R})]}} \xrightarrow{d} \mathcal{N}(0,1) \quad \text{as} \quad n \to \infty,$$

where $\mathcal{N}(0,1)$ denotes the standard normal distribution. In 1974, Maslova [55] proved that the number of real roots of Kac polynomials satisfies the CLT. Almost forty years after the publication of Maslova's result, CLTs have been established for other classes of random polynomials. The CLT for Gaussian Qualls' trigonometric polynomials was first proved by Granville and Wigman [37], and subsequently by Azaïs and León [6] via different methods. Azaïs, Dalmao, and León [5] extended this result to classical trigonometric polynomials. Dalmao [18] did the same for elliptic polynomials, and his result was recently generalized by Ancona and Letendre [1] by the method of moments. The main tool used in [5], [6], and [18] is an L^2 expansion of the number of real roots. CLTs for Weyl polynomials and Weyl series were obtained by Do and Vu [29], using the cumulant convergence theorem. In 2022, Nguyen and Vu [60] proved the CLT for generalized Kac polynomials. The proof of Nguyen and Vu has adapted the universality method and the argument in Maslova [55], which is to approximate the number of real roots by a sum of independent random variables. Recently, Do et al. [24] used the method of Wiener chaos and proved the CLT for random orthogonal polynomials. We emphasize that most of these works deal with the CLT for the number of real roots in a fixed interval. It is of interest to know whether the CLT holds for the number of real roots in an interval whose endpoints depend on the degree of the polynomial.

In this dissertation, we focus on two specific classes of random polynomials that have been studied for a long time: elliptic polynomials and generalized Kac polynomials. More precisely, we study the correlations between the real roots of such polynomials. As applications, we obtain various statistical properties for the number of real roots. Namely, for elliptic polynomials, we aim to find

- a complete asymptotic expansion for the variance of the number $N_n(a,b)$ of real roots in an arbitrary interval (a,b), where a and b may depend on n;
- sharp estimates for the cumulants $s_k[N_n(a,b)]$ and central moments $\mu_k[N_n(a,b)]$; and
- sufficient conditions on the interval (a,b) under which $N_n(a,b)$ satisfies a central limit theorem and a strong law of large numbers.

For generalized Kac polynomials, we establish the leading asymptotics of $\mathbf{Var}[N_n(\mathbb{R})]$. As a consequence, we extend the asymptotic normality result for $N_n(\mathbb{R})$ of O. Nguyen and V. Vu in [60, Theorem 1.2 and Lemma 1.3] to new random polynomials in the generalized Kac regime.

1.2 Correlation functions

In this section, we recall the notion of correlation functions, truncated correlation functions, and their applications (see, e.g., Do and Vu [29], Nazarov and Sodin [56]).

Let Z denote a random point process on $\mathbb{R}$. For example, we can take Z to be the set of real roots of a random polynomial. For $k \geq 1$, the function $\rho_k : \mathbb{R}^k \to \mathbb{R}$ is called the k-point correlation function of Z if, for any compactly supported C^∞ function $f : \mathbb{R}^k \to \mathbb{R}$, it holds that

$$\mathbf{E} \sum_{(z_1,...,z_k)} f(z_1, ..., z_k) = \int_{\mathbb{R}^k} f(x_1, ..., x_k)\rho_k(x_1, ..., x_k)dx_1 \cdots dx_k,$$

where the sum is taken over all possible ordered k-tuples of different elements in Z. So if $(z_\alpha)_{\alpha\in I}$ is a labeling of elements of Z, then we are summing over all $(z_{\alpha_1}, \dots, z_{\alpha_k})$ where $(\alpha_1, \dots, \alpha_k) \in I^k$ such that $\alpha_i \neq \alpha_j$ if $i \neq j$. The correlation function ρ_k is symmetric, and the definition does not depend on the choice of labeling. This implies that ρ_k is locally integrable on $\mathbb{R}^k$. If there is $\varepsilon > 0$ such that ρ_k is locally $L^{1+\varepsilon}$ integrable, then by a simple approximation argument, it follows that the above equality holds when f is only bounded and compactly supported. In particular, for every interval $I \subset \mathbb{R}$ and $N(I) = |Z \cap I|$, it holds that

$$\mathbf{E}[N(I)(N(I)-1)\dots(N(I)-k+1)] = \int_{I^k} \rho_k(x_1,\dots,x_k)dx_1\dots dx_k.$$

We denote by $\Pi(k)$ the collection of all unordered partitions of the set $\{1,2,...,k\}$ into disjoint nonempty blocks and by $\Pi(k,j)$ the collection of all unordered partitions of the set $\{1,2,...,k\}$ into j disjoint nonempty blocks. For $\gamma \in \Pi(k,j)$, we denote the blocks by $\{\gamma_1,\gamma_2,...,\gamma_j\}$ with an arbitrarily chosen enumeration, and denote the lengths of the blocks by $l_i = |\gamma_i|$, for $1 \le i \le j$. If $\boldsymbol{x} = (x_1,...,x_k)$ and $\gamma_j \subset \{1,...,k\}$, let $\boldsymbol{x}_{\gamma_j}$ denote $(x_i)_{i\in\gamma_j}$. The function ρ_k^T defined by the formula

$$\rho_k^T(\boldsymbol{x}) = \sum_{j=1}^{k}(-1)^{j-1}(j-1)! \sum_{\gamma\in\Pi(k,j)} \rho_{l_1}(\boldsymbol{x}_{\gamma_1})\cdots\rho_{l_j}(\boldsymbol{x}_{\gamma_j})$$

is said to be the k-point truncated correlation function of Z (see, e.g., Nazarov and Sodin [56]). We see at once that

$$\rho_1^T(x_1) = \rho_1(x_1), \quad \rho_2^T(x_1,x_2) = \rho_2(x_1,x_2) - \rho_1(x_1)\rho_1(x_2),$$

and so on.

The cumulants $s_k[N(I)]$ of the random variable $N(I)$ is defined by the formal equation

$$\log \mathbf{E}[e^{\lambda N(I)}] = \sum_{k\ge1} \frac{s_k[N(I)]}{k!}\lambda^k.$$

In particular,

$$s_1[N(I)] = \frac{d}{d\lambda}\log \mathbf{E}[e^{\lambda N(I)}]\Big|_{\lambda=0} = \mathbf{E}[N(I)]$$

and

$$s_2[N(I)] = \frac{d^2}{d\lambda^2}\log \mathbf{E}[e^{\lambda N(I)}]\Big|_{\lambda=0} = \mathbf{E}[N(I)^2] - (\mathbf{E}[N(I)])^2 = \mathbf{Var}[N(I)].$$

Lemma 1.1. *For each $k \ge 1$, let ρ_k^T denote the k-point truncated correlation function of Z. For every interval $I \subset \mathbb{R}$ and $N(I) = |Z \cap I|$, it holds that*

$$s_k[N(I)] = \sum_{\gamma\in\Pi(k)} \int_{I^{|\gamma|}} \rho_{|\gamma|}^T(\boldsymbol{x}_\gamma)d\boldsymbol{x}_\gamma, \tag{1.1}$$

where $|\gamma|$ is the number of blocks in the partition γ and $d\boldsymbol{x}_\gamma$ is the Lebesgue measure on $I^{|\gamma|}$.

For the proofs, we refer the reader to Do and Vu [29, Appendix B], in which a general version of Lemma 1.1 was proved. Moreover, a complex variant of Lemma 1.1 was earlier considered by Nazarov and Sodin [56, Claim 4.3].

One should point out that the correlation functions of a general point process Z do not always exist. When Z is the set of real roots of a smooth Gaussian process, the existence of the correlation functions is a consequence of the Kac-Rice formula, which will be recalled below.

1.3 The Kac-Rice formula

Let $\mathcal{G} = \{G(x), x \in I\}$, I an interval on the real line, be a non-degenerate, centered Gaussian process having C^1 paths. Let ρ_k be the k-point correlation function of the real roots of G and let $N(I)$ denote the number of real roots in I. The Kac-Rice formula asserts that (see, e.g., [8, Chapter 3]) for pairwise different $x_1, ..., x_k$,

$$\begin{aligned}\rho_k(x_1, ..., x_k) = \mathbf{E}[|G'(x_1)\cdots G'(x_k)| \mid G(x_1) = 0, ..., G(x_k) = 0] \\ \times p_{G(x_1),...,G(x_k)}(0, ..., 0),\end{aligned} \tag{1.2}$$

where $p_{G(x_1),...,G(x_k)}$ is a joint distribution density of the vector $(G(x_1), ..., G(x_k))$.

Let $r(x, y) = \mathbf{E}[G(x)G(y)]$. We normalize G so that $r(x, x) = 1$. Then

$$\begin{aligned}\mathbf{E}[G(x)G'(x)] &= \mathbf{E}[G(y)G'(y)] = 0, \\ \mathbf{E}[G'(x)G(y)] &= \frac{\partial r}{\partial x}(x, y) =: r_{10}(x, y), \\ \mathbf{E}[G(x)G'(y)] &= \frac{\partial r}{\partial y}(x, y) =: r_{01}(x, y), \\ \mathbf{E}[G'(x)G'(y)] &= \frac{\partial^2 r}{\partial x \partial y}(x, y) =: r_{11}(x, y).\end{aligned}$$

Then the 1-point correlation function ρ_1 is the density of $N(I)$, so that

$$\mathbf{E}[N(I)] = \int_I \rho_1(x)dx.$$

By (1.2),

$$\rho_1(x) = \mathbf{E}[|G'(x)| \mid G(x) = 0] p_{G(x)}(0) = \int_{-\infty}^{\infty} |s| D(0, s; x) ds,$$

where $D(t, s; x)$ is a joint distribution density of $G(x)$ and $G'(x)$,

$$\mathbf{P}(\{a < G(x) \le b; c < G'(x) \le d\}) = \int_a^b dt \int_c^d D(t, s; x) ds.$$

Since $G(x)$ and $G'(x)$ are Gaussian, $D(t, s; x)$ is a Gaussian distribution density with the covariance matrix

$$\begin{pmatrix} \mathbf{E}[G(x)G(x)] & \mathbf{E}[G(x)G'(x)] \\ \mathbf{E}[G(x)G'(x)] & \mathbf{E}[G'(x)G'(x)] \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & r_{11}(x, x) \end{pmatrix}.$$

Therefore,

$$\rho_1(x) = \int_{-\infty}^{\infty} \frac{|s|}{2\pi\sqrt{r_{11}(x, x)}} \exp\left(- \frac{s^2}{2r_{11}(x, x)} \right) ds = \frac{1}{\pi}\sqrt{r_{11}(x, x)}.$$

Next, we show that the 2-point correlation function $\rho_2(x, y)$ is given by

$$\rho_2(x, y) = \frac{1}{\pi^2}\left(\sqrt{1 - \delta^2(x, y)} + \delta(x, y) \arcsin \delta(x, y)\right) \frac{\sigma(x, y)}{\sqrt{1 - r^2(x, y)}}, \tag{1.3}$$

where

$$\sigma(x, y) := \sqrt{\left(r_{11}(x, x) - \frac{r_{10}^2(x, y)}{1 - r^2(x, y)}\right)\left(r_{11}(y, y) - \frac{r_{01}^2(x, y)}{1 - r^2(x, y)}\right)}$$

and

$$\delta(x, y) := \frac{1}{\sigma(x, y)}\left[r_{11}(x, y) + \frac{r(x, y) r_{10}(x, y) r_{01}(x, y)}{1 - r^2(x, y)}\right].$$

Indeed, using (1.2) we get

$$\rho_2(x, y) = \mathbf{E}[|G'(x)G'(y)| \mid G(x) = 0, G(y) = 0] p_{G(x),G(y)}(0, 0),$$

where $p_{G(x),G(y)}$ is the joint density of $(G(x), G(y))$, so that

$$p_{G(x),G(y)}(0, 0) = \frac{1}{2\pi\sqrt{1 - r^2(x, y)}}.$$

Observe that conditionally on $\mathcal{C} := \{G(x) = 0, G(y) = 0\}$, $G'(x)$ and $G'(y)$ have a joint Gaussian distribution with expectations, variances, and covariances given by the following formulas, which can be obtained using regression formulas,

$$\begin{aligned}
\mathbf{E}[G'(x) \mid \mathcal{C}] &= \mathbf{E}[G'(y) \mid \mathcal{C}] = 0, \\
\mathbf{Var}[G'(x) \mid \mathcal{C}] &= r_{11}(x,x) - \frac{r_{10}^2(x,y)}{1 - r^2(x,y)}, \\
\mathbf{Var}[G'(y) \mid \mathcal{C}] &= r_{11}(y,y) - \frac{r_{01}^2(x,y)}{1 - r^2(x,y)}, \\
\mathbf{E}[G'(x)G'(y) \mid \mathcal{C}] &= r_{11}(x,y) + \frac{r(x,y)r_{10}(x,y)r_{01}(x,y)}{1 - r^2(x,y)}.
\end{aligned}$$

But then, computation shows that

$$\sqrt{\mathbf{Var}[G'(x) \mid \mathcal{C}]\, \mathbf{Var}[G'(y) \mid \mathcal{C}]} = \sigma(x,y) \quad \text{and} \quad \mathbf{E}[G'(x)G'(y) \mid \mathcal{C}] = \delta(x,y)\sigma(x,y).$$

Therefore, using [48, Corollary 3.1] yields

$$\mathbf{E}[|G'(x)G'(y)| \mid \mathcal{C}] = \frac{2}{\pi} \left(\sqrt{1 - \delta^2(x,y)} + \delta(x,y) \arcsin \delta(x,y) \right) \sigma(x,y),$$

which implies the assertion.

For other proofs of (1.3), we refer the reader to [11, §3] and [52, Lemma 2.2].

According to Lemma 1.1, we get the following lemma.

Lemma 1.2. *It holds that*

$$\mathbf{Var}[N(I)] = \int_{I^2} [\rho_2(x,y) - \rho_1(x)\rho_1(y)]\, dy dx + \int_I \rho_1(x) dx. \tag{1.4}$$

Remark 1.3. It should be suggested that with Lemma 1.2 and thorough treatment, we can deduce the leading asymptotics of $\mathbf{Var}[N_n(\mathbb{R})]$ for all Gaussian polynomials listed in §1.1.

Generally, the k-point correlation function $\rho_k(x_1, ..., x_k)$ for pairwise different $x_1, ..., x_k$ is given by

$$\rho_k(x_1, ..., x_k) = \int_{\mathbb{R}^k} |s_1 \cdots s_k| D_k(0, s_1, ..., 0, s_k; x_1, ..., x_k) ds_1 \cdots ds_k,$$

where $D_k(0, s_1, ..., 0, s_k; x_1, ..., x_k)$ is a joint distribution density of the vector

$$(G(x_1), G'(x_1), ..., G(x_k), G'(x_k)).$$

Chapter 2

Correlations between the real roots of elliptic polynomials

This chapter aims to further explore the number $N_n(a,b)$ of real roots of elliptic polynomials of degree n in an arbitrary interval (a,b), where a and b may depend on n (see [58]). We first develop an exact and accessible formula for the variance of $N_n(a,b)$ (Theorem 2.1) and exploit it to derive all terms in the large n asymptotic expansion (Theorem 2.2). We then provide sharp estimates for the cumulants (Theorem 2.3) and central moments (Corollary 2.6) of this quantity. These estimates play an important role in determining sufficient conditions on the interval (a,b) under which $N_n(a,b)$ satisfies a central limit theorem (Theorem 2.8) and a strong law of large numbers (Theorem 2.9).

2.1 Introduction and main results

This chapter deals with the number of real roots of elliptic polynomials,

$$P_n(x) = \sum_{j=0}^{n} \xi_j \sqrt{\binom{n}{j}} x^j,$$

where ξ_j are i.i.d. standard normal random variables. This kind of polynomial arises when considering the quantum mechanics of a spin S system whose modulus S is conserved (see, e.g., Bogomolny, Bohias, and Lebœuf [15]). Therefore, the geometric structure of elliptic polynomials is of significant interest for applications to quantum chaos. In addition, such a random polynomial has been intensively studied because of its mathematical properties (Ancona and Letendre [1], Bleher and Di [11, 12], Dalmao [18], Flasche and Kabluchko [33], Edelman and Kostlan [30], Nguyen and Vu [59], Schehr and Majumdar [66], Tao and Vu [69]). To illustrate, let us quote here a sentence from [30] by Edelman and Kostlan: “This particular random polynomial is probably the more natural definition of a random polynomial”. In the literature, elliptic polynomials are sometimes called $SO(2)$ random polynomials because their k-point joint probability distribution of real roots is $SO(2)$-invariant for all $k \geq 1$ (see Bleher and Di [11] for more details).

In 1995, Edelman and Kostlan [30] showed that the expected number $\mathbf{E}[N_n(a,b)]$ is given by

$$\mathbf{E}[N_n(a,b)] = \frac{1}{\pi}\int_a^b \frac{\sqrt{n}}{1+x^2}dx = \frac{1}{\pi}\sqrt{n}(\arctan b - \arctan a). \tag{2.1}$$

In 1997, Bleher and Di [11] obtained the leading term in the large n expansion of the variance $\mathbf{Var}[N_n(a,b)]$, where a and b are fixed. Namely, let

$$\delta_0(s) = \frac{e^{-s^2/2}(1-s^2-e^{-s^2})}{1-e^{-s^2}-s^2e^{-s^2}}, \quad \gamma_0(s) = \frac{1-e^{-s^2}-s^2e^{-s^2}}{(1-e^{-s^2})^{3/2}},$$

and

$$f_0(s) = \left(\sqrt{1-\delta_0^2(s)} + \delta_0(s)\arcsin\delta_0(s)\right)\gamma_0(s) - 1.$$

It was shown in [11, §6] that

$$\mathbf{Var}[N_n(a,b)] = (1+\kappa_0+o(1))\mathbf{E}[N_n(a,b)] \quad \text{as} \quad n \to \infty, \tag{2.2}$$

where

$$\kappa_0 := \frac{2}{\pi}\int_0^\infty f_0(s)ds \tag{2.3}$$

and $(1+\kappa_0) \approx 0.5717310486$. The asymptotics for $\mathbf{Var}[N_n(\mathbb{R})]$ was also considered by Dalmao in [18].

It is worth pointing out that for fixed a and b, (2.1) provides an exact formula for $\mathbf{E}[N_n(a,b)]$, whereas (2.2) gives an asymptotic bound with a worse error term $o(n^{1/2})$. To our knowledge, the true nature of the error term in $\mathbf{Var}[N_n(a,b)]$ has not been known. Finding the precise error in the variance asymptotics is not a trivial task, and it is quite technical. It is apparent that this problem still awaits rigorous treatment because it seems very difficult to improve the order of magnitude of the error term by using all of the existing approaches. As a matter of fact, the second asymptotic term in the variance expansion of the number of real roots for all classes of random polynomials listed in §1.1 has never been obtained yet, which leaves plenty of room for further improvement.

Our purpose is to solve this problem for elliptic polynomials. More generally, we are interested in finding a complete asymptotic expansion for the variance of real roots in the large degree limit. To this end, we first obtain an exact and accessible formula for $\mathbf{Var}[N_n(a,b)]$ which can be used to derive the precise error in the variance asymptotics. To formulate our results, we introduce the functions

$$\begin{aligned}\Delta_n(s) &:= (1+s^2)^{-n/2}\frac{(1+s^2)[1-(1+s^2)^{-n}]-ns^2}{1-(1+s^2)^{-n}-ns^2(1+s^2)^{-n}},\\ \Gamma_n(s) &:= \frac{1-(1+s^2)^{-n}-ns^2(1+s^2)^{-n}}{[1-(1+s^2)^{-n}]^{3/2}},\\ F_n(s) &:= \left(\sqrt{1-\delta^2(s)}+\delta(s)\arcsin\delta(s)\right)\Gamma_n(s)-1,\end{aligned}$$

and the integrals

$$\begin{aligned}K_n(a,b) &:= \frac{2}{\pi}\int_0^{\sqrt{n}|\alpha(a,b)|}\frac{F_n(s/\sqrt{n})}{1+s^2/n}ds,\\ L_n(a,b) &:= \frac{2}{\pi^2}\int_0^{\sqrt{n}|\alpha(a,b)|}\frac{F_n(s/\sqrt{n})}{1+s^2/n}\sqrt{n}\arctan(s/\sqrt{n})ds,\end{aligned}$$

where $\alpha(a,b) := (b-a)/(1+ab)$, for $-\infty \le a, b \le \infty$. We will write K_n and L_n instead of $K_n(a,b)$ and $L_n(a,b)$, respectively, if $\sqrt{n}|\alpha(a,b)|$ is replaced by ∞.

Theorem 2.1 (Exact formulas for the variances)**.**

1. *If* $\alpha(a,b) > 0$*, then*

$$\mathbf{Var}[N_n(a,b)] = (1+K_n(a,b))\mathbf{E}[N_n(a,b)] - L_n(a,b). \tag{2.4}$$

2. *For* $\alpha(a,b) = 0$*, one has* $(a,b) = \mathbb{R}$ *and*

$$\mathbf{Var}[N_n(\mathbb{R})] = (1+K_n)\mathbf{E}[N_n(\mathbb{R})]. \tag{2.5}$$

3. *Otherwise, when* $\alpha(a,b) < 0$*, it holds that*

$$\begin{aligned}\mathbf{Var}[N_n(a,b)] &= (1+K_n)\mathbf{E}[N_n(a,b)]\\ &\quad + (K_n - K_n(a,b))\left(\mathbf{E}[N_n(a,b)] - \sqrt{n}\right) - L_n(a,b).\end{aligned} \tag{2.6}$$

With Theorem 2.1 and delicate analytical tools, we can establish precise asymptotics for the variance $\mathbf{Var}[N_n(a,b)]$. The advantage of using Theorem 2.1 lies in the fact that we can handle the case where the interval (a,b) depends on n. In particular, we show that $\mathbf{Var}[N_n(a,b)]$ admits a complete asymptotic expansion provided that (a,b) does not shrink too rapidly as $n \to \infty$.

Theorem 2.2 (Asymptotic expansions for the variances)**.** *Write* $\alpha_n = \sqrt{n}\alpha(a,b)$.

1. *Assume first that* $|\alpha_n| \to \infty$ *as* $n \to \infty$. *Let*

$$d_n = \left\lfloor \frac{\alpha_n^2 + 3\log|\alpha_n|}{\log n} \right\rfloor, \tag{2.7}$$

where $\lfloor\cdot\rfloor$ *denotes the integer part. Then*

$$\mathbf{Var}[N_n(a,b)] = \left(1 + \sum_{k=0}^{d_n} \frac{\kappa_k}{n^k}\right)\mathbf{E}[N_n(a,b)] - \sum_{k=0}^{d_n} \frac{\ell_k}{n^k} + O(\alpha_n^4 e^{-\alpha_n^2}), \tag{2.8}$$

in which κ_k *and* ℓ_k *are real constants independent of* n, a, *and* b. *In particular,* κ_0 *is defined as in* (2.3) *and*

$$\ell_0 = \frac{2}{\pi^2}\int_0^\infty s f_0(s)ds.$$

Consequently, if $\alpha_n^2/\log n \to \infty$ *as* $n \to \infty$, *then* $\mathbf{Var}[N_n(a,b)]$ *admits a full asymptotic expansion of the form*

$$\mathbf{Var}[N_n(a,b)] \sim \left(1 + \sum_{k=0}^{\infty} \frac{\kappa_k}{n^k}\right)\mathbf{E}[N_n(a,b)] - \sum_{k=0}^{\infty} \frac{\ell_k}{n^k}. \tag{2.9}$$

2. *Assume now that* $|\alpha_n| = O(1)$ *as* $n \to \infty$. *If* $\alpha_n = c > 0$, *then*

$$\mathbf{Var}[N_n(a,b)] \sim \left(1 + \sum_{k=0}^{\infty} \frac{\kappa_{c,k}}{n^k}\right)\mathbf{E}[N_n(a,b)] - \sum_{k=0}^{\infty} \frac{\ell_{c,k}}{n^k}, \tag{2.10}$$

in which $\kappa_{c,k}$ *and* $\ell_{c,k}$ *are real constants. In particular,*

$$\kappa_{c,0} = \frac{2}{\pi}\int_0^c f_0(s)ds \quad \textit{and} \quad \ell_{c,0} = \frac{2}{\pi^2}\int_0^c s f_0(s)ds.$$

For $\alpha_n = -c < 0$, *we have*

$$\begin{aligned}\mathbf{Var}[N_n(a,b)] \sim{}& \left(1 + \sum_{k=0}^{\infty} \frac{\kappa_k}{n^k}\right)\mathbf{E}[N_n(a,b)] \\ &- \left(\sum_{k=0}^{\infty} \frac{\kappa_k - \kappa_{c,k}}{n^k}\right)\frac{\sqrt{n}}{\pi}\arctan\frac{c}{\sqrt{n}} - \sum_{k=0}^{\infty} \frac{\ell_{c,k}}{n^k}.\end{aligned} \tag{2.11}$$

3. Finally, assume that $\alpha_n = o(1)$ as $n \to \infty$. If $\alpha_n > 0$, then

$$\mathbf{Var}[N_n(a,b)] = \frac{1}{\pi}\alpha_n - \frac{1}{\pi^2}\alpha_n^2 + \frac{1}{12\pi}\alpha_n^3 - \frac{5}{12\pi}\frac{\alpha_n^3}{n} + \frac{2}{3\pi^2}\frac{\alpha_n^4}{n} + O(\alpha_n^5). \quad (2.12)$$

If $\alpha_n < 0$, then

$$\begin{aligned}\mathbf{Var}[N_n(a,b)] = &\left(1 + \sum_{k=0}^{q_n} \frac{\kappa_k}{n^k}\right)\sqrt{n} + \frac{2}{\pi}\left(\alpha_n - \frac{\alpha_n^3}{3n}\right)\sum_{k=0}^{r_n}\frac{\kappa_k}{n^k} \\ &+ \frac{1}{\pi}\alpha_n - \frac{1}{\pi^2}\alpha_n^2 - \frac{1}{12\pi}\alpha_n^3 - \frac{1}{4\pi}\frac{\alpha_n^3}{n} + \frac{2}{3\pi^2}\frac{\alpha_n^4}{n} + O(|\alpha_n|^5),\end{aligned} \quad (2.13)$$

where

$$q_n := \left\lfloor \frac{1}{2} - \frac{5\log|\alpha_n|}{\log n} \right\rfloor \quad \text{and} \quad r_n := \left\lfloor -\frac{4\log|\alpha_n|}{\log n} \right\rfloor.$$

If, in addition, $r_n \to \infty$ as $n \to \infty$, then

$$\mathbf{Var}[N_n(a,b)] \sim \left(1 + \sum_{k=0}^{\infty} \frac{\kappa_k}{n^k}\right)\sqrt{n}. \quad (2.14)$$

In particular, $\mathbf{Var}[N_n(\mathbb{R})]$ has a full asymptotic expansion of the form

$$\mathbf{Var}[N_n(\mathbb{R})] \sim \left(1 + \sum_{k=0}^{\infty} \frac{\kappa_k}{n^k}\right)\sqrt{n}. \quad (2.15)$$

The proof of Theorem 2.1 is given in §2.2. In §2.3, we derive asymptotic expansions of $K_n(a,b)$ and $L_n(a,b)$ and use them to prove Theorem 2.2. The definitions of κ_k, $\kappa_{c,k}$, ℓ_k, and $\ell_{c,k}$ are also given in §2.3, along with some detailed numerical computations (see Table 2.1).

Our next goal is to find the asymptotics of the cumulants $s_k[N_n(a,b)]$.

Theorem 2.3 (Asymptotics of the cumulants)**.** *For each positive integer k, there exists a real constant β_k, independent of n, a, and b, such that*

$$s_k[N_n(a,b)] = \beta_k \mathbf{E}[N_n(a,b)] + O(1) \quad \text{as} \quad n \to \infty. \quad (2.16)$$

Remark 2.4. Since $s_1[N_n(a,b)] = \mathbf{E}[N_n(a,b)]$, (2.16) is trivial when $k = 1$, with $\beta_1 = 1$. Using Theorem 2.2 and $s_2[N_n(a,b)] = \mathbf{Var}[N_n(a,b)]$, we see that (2.16) holds for $k = 2$, with $\beta_2 = 1 + \kappa_0$. It should be mentioned that when studying the gap probabilities for elliptic polynomials, Schehr and Majumdar [66, Appendix E] proved that

$$s_3[N_n(a,b)] \sim \beta_3 \mathbf{E}[N_n(a,b)] \quad \text{as} \quad n \to \infty,$$

under the assumptions that β_3 is well-defined and that $\mathbf{E}[N_n(a,b)] \sim \sqrt{n}$ in the large n limit. They expected a similar mechanism to hold for higher values of k (see [66, Equation 93]). Accordingly, our theorem provides a fuller treatment.

Remark 2.5. The error term in (2.16) is only $O(1)$, which is best possible, as shown in Theorem 2.2 for the case $k = 2$. Recently, Gass [34] also computed the cumulant asymptotics for random models with rapidly decreasing covariance functions. Gass's method refines the recent approach by Ancona and Letendre [1, 2], where it has been proved that the k-th central moment, when properly rescaled, converges towards the k-th moment of a Gaussian random variable. As a matter of fact, the methods in these works cannot lead to the true nature of the error term in $s_k[N_n(a,b)]$. Furthermore, it should be noted that the combinatorics of cumulants is a bit more accurate than the method of moments since the asymptotics of cumulants allow us to recover the asymptotics of moments. Indeed, we can exploit Theorem 2.3 to deduce the asymptotics of the central moments $\mu_k[N_n(a,b)]$ as follows.

Corollary 2.6 (Asymptotics of the central moments). *Fix $k \geq 1$. As $n \to \infty$, it holds that*

$$\mu_{2k}[N_n(a,b)] = \frac{(2k)!\beta_2^k}{k!2^k}(\mathbf{E}[N_n(a,b)])^k + O((\mathbf{E}[N_n(a,b)])^{k-1}) \tag{2.17}$$

and

$$\mu_{2k+1}[N_n(a,b)] = \frac{(2k+1)!\beta_2^{k-1}\beta_3}{(k-1)!2^{k-1}3!}(\mathbf{E}[N_n(a,b)])^k + O((\mathbf{E}[N_n(a,b)])^{k-1}). \tag{2.18}$$

Remark 2.7. The asymptotics of the central moments of $N_n(\mathbb{R})$ in the large n limit were earlier studied by Ancona and Letendre [1] (see [2] for more general settings). They showed that, as $n \to \infty$,

$$\mu_k[N_n(\mathbb{R})] = \mu_k[\mathcal{N}(0,1)]\beta_2^{k/2} n^{k/4} + O(n^{(k-1)/4}\log^k(n)), \tag{2.19}$$

where $\mu_k[\mathcal{N}(0,1)]$, for $k \geq 1$, are the moments of the standard normal distribution. Since $\mu_{2k+1}[\mathcal{N}(0,1)] = 0$, formula (2.19) does not imply the leading asymptotics of $\mu_{2k+1}[N_n(\mathbb{R})]$. Thus, our results in Corollary 2.6 not only fill this gap but also offer an improvement of (2.19), as our error terms are only $O(n^{(\lfloor k/2 \rfloor - 1)/2})$.

The proofs of Theorem 2.3 and Corollary 2.6 are given in §2.5 and §2.6, respectively.

To derive our results, in §2.4, we provide a detailed exposition of the correlation and truncated correlation functions of the real roots for elliptic polynomials.

Finally, we consider the limiting law of $N_n(a,b)$. In 2015, Dalmao [18] proved that $N_n(\mathbb{R})$ satisfies the CLT. The proof of Dalmao used the Wiener-Itô expansion of the standardized number of real roots and the fourth-moment theorem. In 2021, Ancona and Letendre [1] recovered Dalmao's CLT by the method of moments. They also proved a strong law of large numbers for $N_n(\mathbb{R})$. In this dissertation, via a different method, we examine a more general framework in which we propose sufficient conditions on (a,b) under which $N_n(a,b)$ satisfies the CLT and a strong law of large numbers. Roughly speaking, it is required that the interval (a,b) does not shrink too rapidly as $n \to \infty$.

Theorem 2.8 (Central limit theorem)**.** *Let α_n be defined as in Theorem 2.2. If either $\alpha_n \le 0$ or $\alpha_n \to \infty$ as $n \to \infty$, then $N_n(a,b)$ satisfies the CLT.*

Theorem 2.9 (Strong law of large numbers)**.** *If either $\alpha_n \le 0$ or $\sum_{n=1}^{\infty} 1/\alpha_n^k < \infty$ for some $k > 0$, then*

$$\frac{N_n(a,b)}{\mathbf{E}[N_n(a,b)]} \xrightarrow{a.s.} 1 \quad \text{as } n \to \infty.$$

Theorem 2.8 follows from applying Theorem 2.3 and the cumulant convergence theorem given by Janson [42, Theorem 1] (see Proposition 2.26), while Theorem 2.9 follows from Corollary 2.6 by a Borel-Cantelli type argument.

Remark 2.10. Similar considerations may apply to the linear statistics $N_n(\phi)$ defined by

$$N_n(\phi) = \sum_{x \in Z_n} \phi(x),$$

where Z_n is the real zero set of an elliptic polynomial of degree n and ϕ satisfies appropriate assumptions. Note that $N_n(\phi)$ reduces to $N_n(a,b)$ if we take $\phi(x) = \mathbf{1}_{(a,b)}(x)$, the indicator function of the interval (a,b).

Remark 2.11. We conclude this section with some suggestions for further work.

First, it follows from (2.1) that, for the Gaussian case, $\mathbf{E}[N_n(\mathbb{R})] = \sqrt{n}$ exactly for all n. In [12], among other results, Bleher and Di extended this result to the non-Gaussian setting.

Proposition 2.12 ([12])**.** *Let ξ_j be i.i.d. random variables with mean zero and variance one. Assume furthermore that for some $c, C > 0$, the characteristic function*

$\varphi(s)$ of ξ_j satisfies

$$|\varphi(s)| \leq \frac{1}{(1+c|s|)^6}, \quad \left|\frac{d^j\varphi(s)}{ds^j}\right| \leq \frac{C}{(1+c|s|)^6}, \quad j=1,2,3, \quad s\in\mathbb{R}.$$

Then, as $n\to\infty$,

$$\mathbf{E}[N_n(\mathbb{R})] = \sqrt{n} + o(n^{1/2}). \tag{2.20}$$

The same result with the assumption on $\varphi(s)$ being removed was obtained in a recent work of Flasche and Kabluchko [33]. In [69, Theorem 5.6], Tao and Vu showed that the same result holds when the random variables ξ_j are only required to be independent with mean zero, variance one, and finite $(2+\varepsilon)$-moments. A more quantitative version of (2.20) was recently given by Nguyen and Vu [59, Corollary 6.4]:

$$\mathbf{E}[N_n(\mathbb{R})] = \sqrt{n} + O(n^{1/2-d}), \quad d>0.$$

By (2.2), we see that for the Gaussian case,

$$\mathbf{Var}[N_n(\mathbb{R})] = (1+\kappa_0)\sqrt{n} + o(n^{1/2}) \quad \text{as} \quad n\to\infty. \tag{2.21}$$

One may ask whether (2.21) is still true if ξ_j are only required to be independent with mean zero, variance one, and finite $(2+\varepsilon)$-moments. More generally, it would be desirable to extend the results of this dissertation to the non-Gaussian setting.

Second, it might be interesting to extend the above results to the number of real zeros of a square system $\boldsymbol{P}=(P_1,\dots,P_m)$ of m polynomial equations in m variables with common degree $n>1$,

$$P_\ell(\boldsymbol{x}) = \sum_{|\boldsymbol{j}|\leq n} \boldsymbol{\omega}_{\boldsymbol{j}}^{(\ell)} \boldsymbol{x}^{\boldsymbol{j}},$$

where

- $\boldsymbol{j}=(j_1,\dots,j_m)\in\mathbb{N}^m$ and $|\boldsymbol{j}|=\sum_{k=1}^m j_k$;
- $\boldsymbol{\omega}_{\boldsymbol{j}}^{(\ell)} = \omega_{j_1\cdots j_m}^{(\ell)}\in\mathbb{R}$, $\ell=1,\dots,m$, $|\boldsymbol{j}|\leq n$, and the coefficients $\boldsymbol{\omega}_{\boldsymbol{j}}^{(\ell)}$ are independent centered normally distributed random variables with variances

$$\mathbf{Var}[\boldsymbol{\omega}_{\boldsymbol{j}}^{(\ell)}] = \binom{n}{\boldsymbol{j}} = \frac{n!}{j_1!\dots j_m!(n-|\boldsymbol{j}|)!};$$

- $\boldsymbol{x}=(x_1,\dots,x_m)$ and $\boldsymbol{x}^{\boldsymbol{j}}=\prod_{k=1}^m x_k^{j_k}$.

Such a system, also known as a Kostlan-Shub-Smale system, was first introduced and studied by Kostlan [45], and subsequently developed by Armentano et al.[3, 4], Azaïs and Wschebor [7], Bleher and Di [12], Edelman and Kostlan [30], Shub and Smale [67], Wschebor [72]. Accordingly, we hope that the concepts and techniques of this dissertation may stimulate further research in this fascinating area.

2.2 Exact formulas for the variances

This section deals with the proof of Theorem 2.1. Let us now apply Lemma 1.2 to $\mathcal{G} = \{P_n(x) : x \in (a,b)\}$, where $P_n(x)$ are elliptic polynomials. Using the binomial theorem we see that the normalized correlator of elliptic polynomials is given by

$$r(x,y) = \frac{\mathbf{E}[P_n(x)P_n(y)]}{\sqrt{\mathbf{Var}[P_n(x)]\,\mathbf{Var}[P_n(y)]}} = \frac{(1+xy)^n}{\sqrt{(1+x^2)^n(1+y^2)^n}}.$$

A direct computation now shows that

$$r_{10}(x,y) = nr(x,y)\frac{(y-x)}{(1+xy)(1+x^2)}, \quad r_{01}(x,y) = nr(x,y)\frac{(x-y)}{(1+xy)(1+y^2)},$$
$$r_{11}(x,y) = nr(x,y)\left(\frac{1}{(1+xy)^2} - \frac{n(x-y)^2}{(1+xy)^2(1+x^2)(1+y^2)}\right).$$

Using $\alpha(x,y) = (y-x)/(1+xy)$ and $(1+x^2)(1+y^2) = (1+xy)^2 + (x-y)^2$, we can write (1.3) in the form

$$\rho_2(x,y) = \frac{1}{\pi^2}\frac{n}{(1+x^2)(1+y^2)}(F_n(\alpha(x,y)) - 1).$$

Together with

$$\rho_1(x)\rho_1(y) = \frac{1}{\pi^2}\frac{n}{(1+x^2)(1+y^2)},$$

we deduce from (1.4) that

$$\mathbf{Var}[N_n(a,b)] = I_{n,2}(a,b) + \mathbf{E}[N_n(a,b)], \tag{2.22}$$

where

$$I_{n,2}(a,b) := \frac{1}{\pi^2}\int_a^b\int_a^b \frac{n}{(1+x^2)(1+y^2)}F_n(\alpha(x,y))dydx. \tag{2.23}$$

The proof of Theorem 2.1 now falls naturally into three following lemmas.

Lemma 2.13. *If $\alpha(a,b) > 0$, then*

$$I_{n,2}(a,b) = K_n(a,b)\mathbf{E}[N_n(a,b)] - L_n(a,b). \tag{2.24}$$

This gives (2.4) when substituted in (2.22).

Proof. The condition $\alpha(a,b) > 0$ implies $ab > -1$. Thus, $1 + xy = 0$ has no solutions in $(a,b) \times (a,b)$. Fix $x \in (a,b)$ and make the change of variables $s = \sqrt{n}\alpha(x,y)$ for the integral $\int_a^b \frac{\sqrt{n}F_n(\alpha(x,y))}{1+y^2}dy$, we see that

$$I_{n,2}(a,b) = \frac{1}{\pi^2}\int_a^b \frac{\sqrt{n}dx}{1+x^2}\int_{\sqrt{n}\alpha(x,a)}^{\sqrt{n}\alpha(x,b)} \frac{F_n(s/\sqrt{n})}{1+s^2/n}ds.$$

By Fubini's Theorem, $\arctan\alpha(s/\sqrt{n},a) = \arctan a - \arctan(s/\sqrt{n})$, and (2.1),

$$\begin{aligned}\int_a^b \frac{\sqrt{n}dx}{1+x^2}\int_{\sqrt{n}\alpha(x,a)}^{0} \frac{F_n(s/\sqrt{n})}{1+s^2/n}ds &= \int_{\sqrt{n}\alpha(b,a)}^{0} \frac{F_n(s/\sqrt{n})}{1+s^2/n}ds\int_{\alpha(s/\sqrt{n},a)}^{b} \frac{\sqrt{n}dx}{1+x^2} \\ &= \frac{\pi^2}{2}K_n(a,b)\mathbf{E}[N_n(a,b)] - \frac{\pi^2}{2}L_n(a,b).\end{aligned}$$

Similarly,

$$\int_a^b \frac{\sqrt{n}dx}{1+x^2}\int_0^{\sqrt{n}\alpha(x,b)} \frac{F_n(s/\sqrt{n})}{1+s^2/n}ds = \frac{\pi^2}{2}K_n(a,b)\mathbf{E}[N_n(a,b)] - \frac{\pi^2}{2}L_n(a,b).$$

Combining these we obtain (2.24) as required. □

Lemma 2.14. *Equation (2.5) follows from the fact that*

$$I_{n,2}(\mathbb{R}) = K_n\sqrt{n}. \tag{2.25}$$

Proof. It follows from (2.23) that

$$I_{n,2}(\mathbb{R}) = \frac{1}{\pi^2}\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} \frac{nF_n(\alpha(x,y))}{(1+x^2)(1+y^2)}dy. \tag{2.26}$$

Fix $x \in (-\infty, 0)$ and substitute $s = \sqrt{n}\alpha(x,y)$, we see that

$$\frac{1}{\pi^2}\int_{-\infty}^{0} dx \int_{-\infty}^{\infty} \frac{nF_n(\alpha(x,y))}{(1+x^2)(1+y^2)}dy = \frac{1}{\pi^2}\int_{-\infty}^{0} \frac{\sqrt{n}}{1+x^2}dx\int_{-\infty}^{\infty} \frac{F_n(s/\sqrt{n})}{1+s^2/n}ds = \frac{\sqrt{n}}{2}K_n.$$

Similarly,
$$\frac{1}{\pi^2}\int_0^\infty dx \int_{-\infty}^\infty \frac{nF_n(\alpha(x,y))}{(1+x^2)(1+y^2)}dy = \frac{\sqrt{n}}{2}K_n.$$
Substituting the results just obtained into (2.26) yields (2.25) as claimed. □

Lemma 2.15. *If* $\alpha(a,b) < 0$, *then*

$$I_{n,2}(a,b) = K_n\mathbf{E}[N_n(a,b)] + (K_n - K_n(a,b))\left(\mathbf{E}[N_n(a,b)] - \sqrt{n}\right) - L_n(a,b), \quad (2.27)$$

which implies (2.6) *when combined with* (2.22).

Proof. Write

$$\begin{aligned} I_{n,2}(a,b) = &\frac{1}{\pi^2}\int_a^{-1/b} \frac{\sqrt{n}dx}{1+x^2}\int_a^{-1/x} \frac{\sqrt{n}F_n(\alpha(x,y))}{1+y^2}dy \\ &+ \frac{1}{\pi^2}\int_a^{-1/b} \frac{\sqrt{n}dx}{1+x^2}\int_{-1/x}^{b} \frac{\sqrt{n}F_n(\alpha(x,y))}{1+y^2}dy \\ &+ \frac{1}{\pi^2}\int_{-1/b}^{-1/a} \frac{\sqrt{n}dx}{1+x^2}\int_a^{b} \frac{\sqrt{n}F_n(\alpha(x,y))}{1+y^2}dy \\ &+ \frac{1}{\pi^2}\int_{-1/a}^{b} \frac{\sqrt{n}dx}{1+x^2}\int_a^{-1/x} \frac{\sqrt{n}F_n(\alpha(x,y))}{1+y^2}dy \\ &+ \frac{1}{\pi^2}\int_{-1/a}^{b} \frac{\sqrt{n}dx}{1+x^2}\int_{-1/x}^{b} \frac{\sqrt{n}F_n(\alpha(x,y))}{1+y^2}dy. \end{aligned}$$

We can now proceed analogously to the proof of Lemma 2.13. Using the substitution $s = \sqrt{n}\alpha(x,y)$, Fubini's Theorem, and the facts that

$$\arctan\alpha(x,y) = \begin{cases} \arctan y - \arctan x & \text{if} \quad 1+xy>0, \\ \arctan y - \arctan x - \pi & \text{if} \quad 1+xy<0 \text{ and } y>0, \\ \arctan y - \arctan x + \pi & \text{if} \quad 1+xy<0 \text{ and } y<0, \end{cases}$$

and
$$\arctan x + \arctan(1/x) = \begin{cases} +\pi/2 & \text{if} \quad x>0, \\ -\pi/2 & \text{if} \quad x<0, \end{cases}$$

we conclude that

$$\begin{aligned}
I_{n,2}(a,b) &= \frac{1}{2}(K_n + K_n(-1/a,b))(\mathbf{E}[N_n(a,b)] - \sqrt{n}/2) - \frac{1}{2}L_n(-1/a,b) \\
&\quad + \frac{1}{2}(K_n - K_n(a,b))(\mathbf{E}[N_n(a,b)] - \sqrt{n}) + \frac{1}{2}(L_n - L_n(a,b)) \\
&\quad + \frac{\sqrt{n}}{2}K_n - K_n(-1/a,b)(\mathbf{E}[N_n(a,b)] - \sqrt{n}/2) - L_n + L_n(-1/a,b) \\
&\quad + \frac{1}{2}(K_n - K_n(a,b))(\mathbf{E}[N_n(a,b)] - \sqrt{n}) + \frac{1}{2}(L_n - L_n(a,b)) \\
&\quad + \frac{1}{2}(K_n + K_n(-1/a,b))(\mathbf{E}[N_n(a,b)] - \sqrt{n}/2) - \frac{1}{2}L_n(-1/a,b) \\
&= K_n\mathbf{E}[N_n(a,b)] + (K_n - K_n(a,b))(\mathbf{E}[N_n(a,b)] - \sqrt{n}) - L_n(a,b),
\end{aligned}$$

which gives (2.27). □

2.3 Asymptotic expansions for the variances

This section provides a detailed proof of Theorem 2.2. Let us first mention some lemmas that will be imperative to the proof. Note that Theorem 2.1 allows us to derive the large n expansion of $\mathbf{Var}[N_n(a,b)]$ using that of $K_n(a,b)$ and $L_n(a,b)$. To expand $K_n(a,b)$ and $L_n(a,b)$, we first show that $\frac{F_n(s/\sqrt{n})}{1+s^2/n}$ and $\frac{F_n(s/\sqrt{n})}{1+s^2/n}\sqrt{n}\arctan(s/\sqrt{n})$ can be transformed into series of terms which are powers of $1/n$.

Lemma 2.16. *Given $0 < c_n < \sqrt{n}$, one has*

$$\frac{F_n(s/\sqrt{n})}{1+s^2/n} = \sum_{k=0}^{\infty} \frac{f_k(s)}{n^k} \quad \text{uniformly for } s \in [0, c_n], \tag{2.28}$$

in which $f_k(s)$ have continuous extensions to $[0,\infty)$ such that, as $s \to 0$,

$$f_k(s) = \begin{cases} -1 + \frac{\pi}{4}s + O(s^3) & \text{if} \quad k = 0, \\ -\frac{\pi}{4}s + s^2 + O(s^3) & \text{if} \quad k = 1, \\ O(s^3) & \text{if} \quad k \geq 2, \end{cases} \tag{2.29}$$

and, as $s \to \infty$,

$$f_k(s) = \frac{1}{2^{k+1}k!}s^{4k+4}e^{-s^2} + O(s^{4k+2}e^{-s^2}), \quad k \geq 0. \tag{2.30}$$

Proof. The proof will be divided into four steps.

Step 1. Expand $\Delta_n(s/\sqrt{n})$.

Observe that

$$\Delta_n(s/\sqrt{n}) = \left(1+\frac{s^2}{n}\right)^{-n/2} \frac{(1+s^2/n)[1-(1+s^2/n)^{-n}]-s^2}{1-(1+s^2)(1+s^2/n)^{-n}}.$$

If $s \in [0, c_n]$, then $s^2/n \le c_n^2/n < 1$. Hence, for any $c > 0$,

$$-cn\log(1+s^2/n) = -cs^2 + c\sum_{k=1}^{\infty} \frac{q_k(s)}{k!}\frac{1}{n^k} \quad \text{uniformly for } s \in [0, c_n],$$

in which $q_k(s) = (-s^2)^{k+1}k!/(k+1)$. But then

$$\left(1+\frac{s^2}{n}\right)^{-cn} = e^{-cs^2}\left(1+\sum_{k=1}^{\infty}\frac{e_{c,k}(s)}{n^k}\right), \tag{2.31}$$

where

$$e_{c,k}(s) = \frac{1}{k!}\sum_{j=1}^{k} c^j B_{k,j}(q_1(s), ..., q_{k-j+1}(s)) \tag{2.32}$$

with $B_{k,j}$ denoting the exponential partial Bell polynomials (see [17, §3.3]). Explicit formulas for these polynomials are as follows

$$B_{k,j}(q_1(s), ..., q_{k-j+1}(s)) = \sum \frac{k!}{m_1!\cdots m_{k-j+1}!}\prod_{r=1}^{k-j+1}\left(\frac{(-s^2)^{r+1}}{r+1}\right)^{m_r} \tag{2.33}$$

where the sum is over all solutions in non-negative integers of the equations

$$\begin{aligned} m_1 + 2m_2 + \cdots + (k-j+1)m_{k-j+1} &= k, \\ m_1 + m_2 + \cdots + m_{k-j+1} &= j. \end{aligned}$$

Combining (2.32) with (2.33) yields

$$e_{c,k}(s) = \begin{cases} \frac{c(-1)^{k+1}}{k+1}s^{2k+2} + O(s^{2k+4}) & \text{as} \quad s \to 0, \\ \frac{c^k}{k!2^k}s^{4k} + O(s^{4k-2}) & \text{as} \quad s \to \infty. \end{cases} \tag{2.34}$$

With (2.31) and a bit of work, we can write

$$\left(1+\frac{s^2}{n}\right)^{-n/2}\left[(1+s^2/n)[1-(1+s^2/n)^{-n}]-s^2\right]=\sum_{k=0}^{\infty}\frac{u_k(s)}{n^k},$$

in which

$$u_0(s)=e^{-s^2/2}\left(1-s^2-e^{-s^2}\right),\quad u_1(s)=e^{-s^2/2}\left[s^2+\frac{s^4}{4}-\frac{s^6}{4}-e^{-s^2}\left(s^2+\frac{3s^4}{4}\right)\right],$$

and, for $k\geq 2$,

$$u_k(s)=e^{-s^2/2}\left[s^2e_{1/2,k-1}(s)+(1-s^2)e_{1/2,k}(s)\right]-e^{-3s^2/2}\left[e_{3/2,k}(s)+s^2e_{3/2,k-1}(s)\right].$$

Notice that

$$u_0(s)=\begin{cases}-\frac{1}{2}s^4+O(s^6) & \text{as}\quad s\to 0,\\ -s^2e^{-s^2/2}+O(e^{-s^2/2}) & \text{as}\quad s\to\infty,\end{cases}\tag{2.35}$$

$$u_1(s)=\begin{cases}\frac{1}{2}s^4+O(s^6) & \text{as}\quad s\to 0,\\ -\frac{1}{4}s^6e^{-s^2/2}+O(s^4e^{-s^2/2}) & \text{as}\quad s\to\infty,\end{cases}\tag{2.36}$$

and, by (2.34), for $k\geq 2$,

$$u_k(s)=\begin{cases}O(s^{2k+2}) & \text{as}\quad s\to 0,\\ -\frac{1}{4^kk!}s^{4k+2}e^{-s^2/2}+O(s^{4k}e^{-s^2/2}) & \text{as}\quad s\to\infty.\end{cases}\tag{2.37}$$

For $s\in(0,c_n]$, one has $0<(1+s^2)(1+s^2/n)^{-n}<1$, and so

$$\begin{aligned}\frac{1}{1-(1+s^2)(1+s^2/n)^{-n}}&=1+\sum_{m=1}^{\infty}(1+s^2)^m\left(1+\frac{s^2}{n}\right)^{-mn}\\&=1+\sum_{m=1}^{\infty}(1+s^2)^me^{-ms^2}\left(1+\sum_{k=1}^{\infty}\frac{e_{m,k}(s)}{n^k}\right)\\&=:v_0(s)+\sum_{k=1}^{\infty}\frac{v_k(s)}{n^k}.\end{aligned}$$

Clearly,

$$v_0(s)=1+\sum_{m=1}^{\infty}(1+s^2)^me^{-ms^2}=\frac{1}{1-(1+s^2)e^{-s^2}},$$

which gives

$$v_0(s) = \begin{cases} 2s^{-4} + O(s^{-2}) & \text{as} \quad s \to 0, \\ 1 + O(s^2 e^{-s^2}) & \text{as} \quad s \to \infty. \end{cases} \tag{2.38}$$

For $k \geq 1$, $v_k(s)$ can be expressed in terms of the polylogarithm functions (see [47]) defined by

$$\mathrm{Li}_j(z) := \sum_{m=1}^{\infty} \frac{z^m}{m^j}.$$

Indeed, by definition of $v_k(s)$ and (2.32),

$$\begin{aligned} v_k(s) &= \sum_{m=1}^{\infty} (1+s^2)^m e^{-ms^2} e_{m,k}(s) \\ &= \sum_{m=1}^{\infty} (1+s^2)^m e^{-ms^2} \frac{1}{k!} \sum_{j=1}^{k} m^j B_{k,j}(q_1(s), ..., q_{k-j+1}(s)) \\ &= \frac{1}{k!} \sum_{j=1}^{k} B_{k,j}(q_1(s), ..., q_{k-j+1}(s)) \, \mathrm{Li}_{-j}((1+s^2)e^{-s^2}). \end{aligned}$$

In particular,

$$v_1(s) = B_{1,1}(q_1(s)) \, \mathrm{Li}_{-1}((1+s^2)e^{-s^2}) = \frac{s^4}{2} \frac{(1+s^2)e^{-s^2}}{(1-(1+s^2)e^{-s^2})^2}.$$

Since, for $1 \leq j \leq k$,

$$\mathrm{Li}_{-j}((1+s^2)e^{-s^2}) \sim \begin{cases} j! 2^{j+1} s^{-4(j+1)} & \text{as} \quad s \to 0, \\ (1+s^2)e^{-s^2} & \text{as} \quad s \to \infty, \end{cases}$$

it follows that

$$v_k(s) = \begin{cases} 2s^{-4} + O(s^{-2}) & \text{as} \quad s \to 0, \\ \frac{1}{2^k k!} s^{4k+2} e^{-s^2} + O(s^{4k} e^{-s^2}) & \text{as} \quad s \to \infty. \end{cases} \tag{2.39}$$

Next, by the Cauchy product, for $s \in (0, c_n]$,

$$\left(\sum_{k=0}^{\infty} \frac{u_k(s)}{n^k} \right) \left(\sum_{k=0}^{\infty} \frac{v_k(s)}{n^k} \right) = \sum_{k=0}^{\infty} \frac{\delta_k(s)}{n^k},$$

where

$$\delta_k(s) := \sum_{j=0}^{k} u_j(s) v_{k-j}(s), \quad k \geq 0.$$

In particular,

$$\begin{aligned}\delta_0(s) &= \frac{e^{-s^2/2}(1 - s^2 - e^{-s^2})}{1 - (1+s^2)e^{-s^2}},\\ \delta_1(s) &= \frac{s^4 e^{-s^2/2}}{2} \frac{(1 - s^2 - e^{-s^2})(1+s^2)e^{-s^2}}{(1-(1+s^2)e^{-s^2})^2}\\ &\quad + \frac{e^{-s^2/2}}{1-(1+s^2)e^{-s^2}} \left[s^2 + \frac{s^4}{4} - \frac{s^6}{4} - e^{-s^2}\left(s^2 + \frac{3s^4}{4} \right) \right].\end{aligned}$$

We check at once that

$$\delta_0(s) = \begin{cases} -1 + O(s^2) & \text{as} \quad s \to 0, \\ -s^2 e^{-s^2/2} + O(e^{-s^2/2}) & \text{as} \quad s \to \infty. \end{cases} \tag{2.40}$$

Using (2.37), (2.38), and (2.39), we see that, for $k \geq 1$,

$$\delta_k(s) = \begin{cases} u_0(s)v_k(s) + u_1(s)v_{k-1}(s) + O(s^2) & \text{as} \quad s \to 0, \\ u_k(s)v_0(s) + O(s^{4k+4}e^{-3s^2/2}) & \text{as} \quad s \to \infty. \end{cases}$$

Together with (2.35) and (2.36), we arrive at

$$\delta_k(s) = \begin{cases} O(s^2) & \text{as} \quad s \to 0, \\ -\frac{1}{4^k k!} s^{4k+2} e^{-s^2/2} + O(s^{4k+4}e^{-3s^2/2}) & \text{as} \quad s \to \infty. \end{cases} \tag{2.41}$$

This implies that the functions $\delta_k(s)$ extend by continuity at $s = 0$. Hence,

$$\Delta_n(s/\sqrt{n}) = \sum_{k=0}^{\infty} \frac{\delta_k(s)}{n^k} \quad \text{uniformly for } s \in [0, c_n]. \tag{2.42}$$

Step 2. Expand $h(\Delta_n(s/\sqrt{n}))$, where $h(x) := \sqrt{1-x^2} + x \arcsin x$.

For $s > 0$, we see that $-1 < \Delta_n(s/\sqrt{n}) < 1$. Thus, by (2.42) and Faà di Bruno's formula (see [17, §3.4]),

$$h(\Delta_n(s/\sqrt{n})) = h(0) + \sum_{m=1}^{\infty} \frac{h^{(m)}(0)}{m!}(\Delta_n(s/\sqrt{n}))^m$$
$$= 1 + \sum_{m=1}^{\infty} \frac{h^{(m)}(0)}{m!} \left(\sum_{k=0}^{\infty} \frac{\delta_k(s)}{n^k} \right)^m =: z_0(s) + \sum_{k=1}^{\infty} \frac{z_k(s)}{n^k},$$

where

$$z_0(s) = 1 + \sum_{m=1}^{\infty} \frac{h^{(m)}(0)}{m!} \delta_0^m(s) = h(\delta_0(s)),$$

and, for $k \geq 1$,

$$z_k(s) = \frac{1}{k!} \sum_{j=1}^{k} h^{(j)}(\delta_0(s)) B_{k,j}(1!\delta_1(s), ..., (k-j+1)!\delta_{k-j+1}(s)).$$

In particular,

$$z_1(s) = B_{1,1}(\delta_1(s)) h'(\delta_0(s)) = \delta_1(s) \arcsin(\delta_0(s)).$$

By (2.40) and the asymptotic behaviors of $h(x)$ as $x \to 0$ and as $x \to -1^+$,

$$z_0(s) = \begin{cases} \pi/2 + O(s^2) & \text{as} \quad s \to 0, \\ 1 + \frac{1}{2} s^4 e^{-s^2} + O(s^2 e^{-s^2}) & \text{as} \quad s \to \infty. \end{cases} \tag{2.43}$$

Note that $h'(x) = \arcsin x$, so

$$h'(x) = \begin{cases} O(1) & \text{as} \quad x \to -1^+, \\ x + O(x^3) & \text{as} \quad x \to 0, \end{cases}$$

and, for $j \geq 2$,

$$h^{(j)}(x) = \begin{cases} O((1-x^2)^{(3-2j)/2}) & \text{as} \quad x \to -1^+, \\ \frac{1+(-1)^j}{2} + O(x^2) & \text{as} \quad x \to 0. \end{cases}$$

Together with (2.40), we see that

$$h'(\delta_0(s)) = \begin{cases} O(1) & \text{as} \quad s \to 0, \\ -s^2 e^{-s^2/2} + O(s^6 e^{-3s^2/2}) & \text{as} \quad s \to \infty, \end{cases}$$

and, for $j \geq 2$,

$$h^{(j)}(\delta_0(s)) = \begin{cases} O(s^{3-2j}) & \text{as} \quad s \to 0, \\ \frac{1+(-1)^j}{2} + O(s^4 e^{-s^2}) & \text{as} \quad s \to \infty. \end{cases}$$

Thus, using (2.41) and the fact that $1 + B_{k,2}(1, ..., 1) = 2^{k-1}$, we get

$$z_k(s) = \begin{cases} O(s^2) & \text{as} \quad s \to 0, \\ \frac{1}{2^{k+1}k!} s^{4k+4} e^{-s^2} + O(s^{4k+2} e^{-s^2}) & \text{as} \quad s \to \infty. \end{cases} \tag{2.44}$$

Summarizing, we have

$$h(\Delta_n(s/\sqrt{n})) = \sum_{k=0}^{\infty} \frac{z_k(s)}{n^k} \quad \text{uniformly for } s \in [0, c_n]. \tag{2.45}$$

Step 3. Expand $\Gamma_n(s/\sqrt{n})$.

For this purpose, let us consider the function $x \mapsto g_s(x)$ given by

$$g_s(x) = \frac{1 - (1 + s^2)x}{(1 - x)^{3/2}}, \quad x \in (-1, 1).$$

For $s > 0$, we have $0 < (1 + s^2/n)^{-n} < 1$ and

$$\Gamma_n(s/\sqrt{n}) = \frac{1 - (1 + s^2)(1 + s^2/n)^{-n}}{[1 - (1 + s^2/n)^{-n}]^{3/2}} = g_s\left((1 + s^2/n)^{-n}\right).$$

Therefore,

$$\begin{aligned} \Gamma_n(s/\sqrt{n}) &= g_s(0) + \sum_{m=1}^{\infty} \frac{g_s^{(m)}(0)}{m!} \left(1 + \frac{s^2}{n}\right)^{-mn} \\ &= 1 + \sum_{m=1}^{\infty} \frac{g_s^{(m)}(0)}{m!} e^{-ms^2} \sum_{k=0}^{\infty} \frac{e_{m,k}(s)}{n^k} \\ &= 1 + \sum_{m=1}^{\infty} \frac{g_s^{(m)}(0)}{m!} e^{-ms^2} + \sum_{k=1}^{\infty} \left(\sum_{m=1}^{\infty} \frac{g_s^{(m)}(0)}{m!} e^{-ms^2} e_{m,k}(s) \right) \frac{1}{n^k} \\ &=: \gamma_0(s) + \sum_{k=1}^{\infty} \frac{\gamma_k(s)}{n^k}. \end{aligned}$$

In particular,

$$\gamma_0(s) = 1 + \sum_{m=1}^{\infty} \frac{g_s^{(m)}(0)}{m!} e^{-ms^2} = g_s(e^{-s^2}) = \frac{1-(1+s^2)e^{-s^2}}{(1-e^{-s^2})^{3/2}}.$$

To determine $\gamma_k(s)$, for $k \geq 1$, we utilize the following identity (see [17, §5.1]), for $m \geq 1$ and $1 \leq j \leq k$,

$$m^j = \sum_{r=1}^{j} S(j,r)(m)_r,$$

where $S(j,r)$ are the Stirling numbers of the second kind, and $(m)_r$ are the falling factorials defined by $(m)_r = m(m-1)\cdots(m-r+1)$. This implies

$$\begin{aligned}\gamma_k(s) &= \sum_{m=1}^{\infty} \frac{g_s^{(m)}(0)}{m!} e^{-ms^2} e_{m,k}(s)\\ &= \sum_{m=1}^{\infty} \frac{g_s^{(m)}(0)}{m!} e^{-ms^2} \frac{1}{k!} \sum_{j=1}^{k} m^j B_{k,j}(q_1(s), ..., q_{k-j+1}(s))\\ &= \frac{1}{k!} \sum_{j=1}^{k} B_{k,j}(q_1(s), ..., q_{k-j+1}(s)) \sum_{r=1}^{j} S(j,r) e^{-rs^2} g_s^{(r)}(e^{-s^2}).\end{aligned}$$

In particular,

$$\gamma_1(s) = B_{1,1}(q_1(s))S(1,1)e^{-s^2} g_s'(e^{-s^2}) = \frac{s^4 e^{-s^2}(1-2s^2-(1+s^2)e^{-s^2})}{4(1-e^{-s^2})^{5/2}}.$$

A trivial verification shows that

$$\gamma_0(s) = \begin{cases} \frac{1}{2}s + O(s^3) & \text{as} \quad s \to 0, \\ 1 - s^2 e^{-s^2} + O(e^{-s^2}) & \text{as} \quad s \to \infty, \end{cases} \tag{2.46}$$

$$\gamma_1(s) = \begin{cases} -\frac{1}{2}s + O(s^3) & \text{as} \quad s \to 0, \\ -\frac{1}{2}s^6 e^{-s^2} + O(s^4 e^{-s^2}) & \text{as} \quad s \to \infty. \end{cases} \tag{2.47}$$

Notice that

$$g_s(e^{-s^2}) = \frac{1-(1+s^2)e^{-s^2}}{(1-e^{-s^2})^{3/2}} = \begin{cases} \frac{1}{2}s + O(s^3) & \text{as} \quad s \to 0, \\ 1 + O(s^2 e^{-s^2}) & \text{as} \quad s \to \infty \end{cases}$$

and

$$g_s^{(r)}(e^{-s^2}) = \frac{2^r}{(2r+1)!!}\frac{1-(1+s^2)e^{-s^2}}{(1-e^{-s^2})^{(2r+3)/2}} - \frac{r2^{r-1}}{(2r-1)!!}\frac{1+s^2}{(1-e^{-s^2})^{(2r+1)/2}}$$
$$= \begin{cases} -\frac{r2^{r-1}}{(2r-1)!!}s^{-(2r+1)} + O(s^{-(2r-1)}) & \text{as} \quad s \to 0, \\ -\frac{r2^{r-1}}{(2r-1)!!}s^2 + O(1) & \text{as} \quad s \to \infty. \end{cases}$$

Combining with (2.33) yields, for $k \geq 2$,

$$\gamma_k(s) = \begin{cases} O(s^{2k-1}) & \text{as} \quad s \to 0, \\ -\frac{1}{2^k k!}s^{4k+2}e^{-s^2} + O(s^{4k}e^{-s^2}) & \text{as} \quad s \to \infty. \end{cases} \tag{2.48}$$

Therefore,

$$\Gamma_n(s/\sqrt{n}) = \sum_{k=0}^{\infty} \frac{\gamma_k(s)}{n^k} \quad \text{uniformly for } s \in [0, c_n]. \tag{2.49}$$

Step 4. Expand $\frac{F_n(s/\sqrt{n})}{1+s^2/n}$.

Combining (2.45) with (2.49), we can write

$$F_n(s/\sqrt{n}) = \left(\sum_{k=0}^{\infty} \frac{z_k(s)}{n^k}\right)\left(\sum_{k=0}^{\infty} \frac{\gamma_k(s)}{n^k}\right) - 1 = \sum_{k=0}^{\infty} \frac{a_k(s)}{n^k} \tag{2.50}$$

uniformly for $s \in [0, c_n]$, in which

$$a_0(s) = z_0(s)\gamma_0(s) - 1 \quad \text{and} \quad a_k(s) = \sum_{j=0}^{k} z_j(s)\gamma_{k-j}(s), \quad k \geq 1.$$

On account of (2.43), (2.44), (2.46), (2.47), and (2.48), we have

$$a_0(s) = \begin{cases} -1 + \frac{\pi}{4}s + O(s^3) & \text{as} \quad s \to 0, \\ \frac{1}{2}s^4 e^{-s^2} + O(s^2 e^{-s^2}) & \text{as} \quad s \to \infty, \end{cases}$$
$$a_1(s) = \begin{cases} -\frac{\pi}{4}s + O(s^3) & \text{as} \quad s \to 0, \\ \frac{1}{4}s^8 e^{-s^2} + O(s^6 e^{-s^2}) & \text{as} \quad s \to \infty, \end{cases}$$

and, for $k \geq 2$,

$$a_k(s) = \begin{cases} O(s^3) & \text{as} \quad s \to 0, \\ \frac{1}{2^{k+1}k!}s^{4k+4}e^{-s^2} + O(s^{4k+2}e^{-s^2}) & \text{as} \quad s \to \infty. \end{cases}$$

Since

$$\frac{1}{1+s^2/n} = \sum_{k=0}^{\infty} \frac{(-s^2)^k}{n^k},$$

it follows from (2.50) that

$$\frac{F_n(s/\sqrt{n})}{1+s^2/n} = \left(\sum_{k=0}^{\infty} \frac{a_k(s)}{n^k}\right)\left(\sum_{k=0}^{\infty} \frac{(-s^2)^k}{n^k}\right) = \sum_{k=0}^{\infty} \frac{f_k(s)}{n^k}$$

uniformly for $s \in [0, c_n]$, in which

$$f_k(s) := \sum_{j=0}^{k} (-1)^j s^{2j} a_{k-j}(s), \quad k \geq 0.$$

By the above, $f_k(s)$ have continuous extensions to $[0, \infty)$ such that

$$f_0(s) = \begin{cases} -1 + \frac{\pi}{4}s + O(s^3) & \text{as} \quad s \to 0, \\ \frac{1}{2}s^4 e^{-s^2} + O(s^2 e^{-s^2}) & \text{as} \quad s \to \infty, \end{cases}$$
$$f_1(s) = \begin{cases} -\frac{\pi}{4}s + s^2 + O(s^3) & \text{as} \quad s \to 0, \\ \frac{1}{4}s^8 e^{-s^2} + O(s^6 e^{-s^2}) & \text{as} \quad s \to \infty, \end{cases}$$

and, for $k \geq 2$,

$$f_k(s) = \begin{cases} O(s^3) & \text{as} \quad s \to 0, \\ \frac{1}{2^{k+1}k!}s^{4k+4} e^{-s^2} + O(s^{4k+2} e^{-s^2}) & \text{as} \quad s \to \infty. \end{cases}$$

Thus, Lemma 2.16 is verified. □

Recall that

$$f_0(s) = h(\delta_0(s))\gamma_0(s) - 1$$

and

$$f_1(s) = h(\delta_0(s))\gamma_1(s) + \delta_1(s)\arcsin(\delta_0(s))\gamma_0(s) - s^2 f_0(s),$$

where explicit formulas for $h(x)$, $\delta_0(s)$, $\delta_1(s)$, $\gamma_0(s)$, and $\gamma_1(s)$ are provided. This means that one can also obtain explicit formulas for both $f_0(s)$ and $f_1(s)$. In Figure 2.1, we show plots of $f_0(s)$ and $f_1(s)$ for $s \in [0, 5]$.

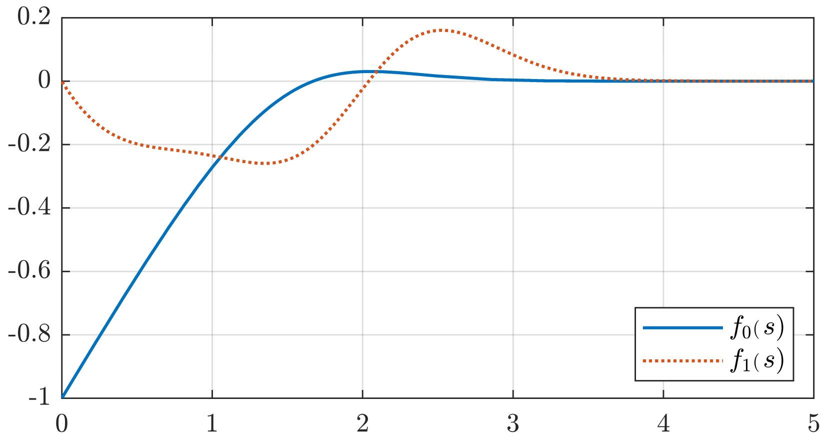

Figure 2.1: Plots of $f_0(s)$ and $f_1(s)$.

Note that, for $-1 \le s/\sqrt{n} \le 1$,

$$\sqrt{n}\arctan(s/\sqrt{n}) = \sum_{k=0}^{\infty} \frac{(-1)^k s^{2k+1}}{2k+1} \frac{1}{n^k}.$$

Together with Lemma 2.16, we obtain the following lemma.

Lemma 2.17. *Given $0 < c_n < \sqrt{n}$, one has*

$$\frac{F_n(s/\sqrt{n})}{1+s^2/n}\sqrt{n}\arctan(s/\sqrt{n}) = \sum_{k=0}^{\infty} \frac{g_k(s)}{n^k} \quad \textit{uniformly for } s \in [0, c_n], \tag{2.51}$$

in which

$$g_k(s) = \sum_{j=0}^{k} \frac{(-1)^j s^{2j+1}}{2j+1} f_{k-j}(s), \quad k \ge 0.$$

Moreover, $g_k(s)$ have continuous extensions to $[0, \infty)$ such that, as $s \to 0$,

$$g_k(s) = \begin{cases} -s + \frac{\pi}{4}s^2 + O(s^4) & \textit{if} \quad k = 0, \\ -\frac{\pi}{4}s^2 + \frac{4}{3}s^3 + O(s^4) & \textit{if} \quad k = 1, \\ O(s^4) & \textit{if} \quad k \ge 2, \end{cases} \tag{2.52}$$

and, as $s \to \infty$,

$$g_k(s) = \frac{1}{2^{k+1}k!} s^{4k+5} e^{-s^2} + O(s^{4k+3}e^{-s^2}), \quad k \ge 0. \tag{2.53}$$

Note that explicit formulas for $g_0(s)$ and $g_1(s)$ can be obtained from $g_0(s) = sf_0(s)$

and $g_1(s) = sf_1(s) - \frac{s^3}{3}f_0(s)$. Plots of $g_0(s)$ and $g_1(s)$ for $s \in [0,5]$ are included in Figure 2.2.

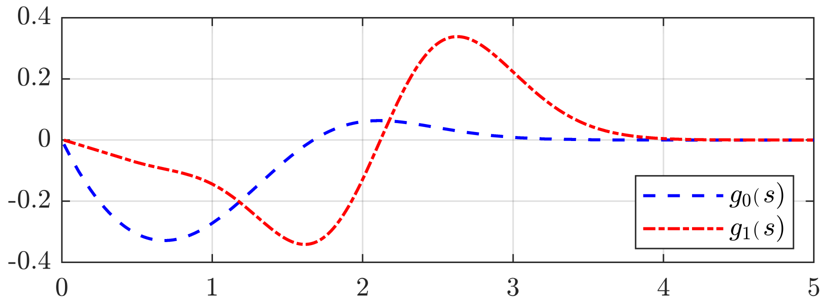

Figure 2.2: Plots of $g_0(s)$ and $g_1(s)$.

We can now define, for $k \geq 0$ and $c > 0$,

$$\kappa_k := \frac{2}{\pi}\int_0^\infty f_k(s)ds, \qquad \kappa_{c,k} := \frac{2}{\pi}\int_0^c f_k(s)ds,$$
$$\ell_k := \frac{2}{\pi^2}\int_0^\infty g_k(s)ds, \qquad \ell_{c,k} := \frac{2}{\pi^2}\int_0^c g_k(s)ds.$$

The continuity of $f_k(s)$ and $g_k(s)$, and the asymptotic behaviors given in (2.29), (2.30), (2.52), and (2.53) make the definitions of κ_k, ℓ_k, $\kappa_{c,k}$, and $\ell_{c,k}$ allowable. Using explicit formulas for $f_k(s)$ and $g_k(s)$, for $k = 0, 1$, we can numerically compute κ_k, ℓ_k, $\kappa_{c,k}$, and $\ell_{c,k}$, for $k = 0, 1$, and $c > 0$. Some such numerical values are listed in Table 2.1, in which the integrals involved were evaluated numerically using MATLAB.

Table 2.1: Numerical values of κ_k, ℓ_k, $\kappa_{1,k}$, and $\ell_{1,k}$ for $k = 0, 1$.

k	κ_k	ℓ_k	$\kappa_{1,k}$	$\ell_{1,k}$
0	-0.4282689510	-0.0580365252	-0.3955313789	-0.0505415303
1	-0.1522064957	-0.0082122652	-0.1093878905	-0.0138350833

We emphasize that the expansions given in (2.28) and (2.51) allow us to expand $K_n(a,b)$ and $L_n(a,b)$ into series of terms which are powers of $1/n$, provided that $|\alpha_n| < \sqrt{n}$. Our next task is thus to estimate $K_n - K_n(a,b)$ and $L_n - L_n(a,b)$ when $|\alpha_n|$ is arbitrarily large.

Lemma 2.18. *If $|\alpha_n| \to \infty$ as $n \to \infty$, then for all n sufficiently large,*

$$K_n - K_n(a,b) = O\left(|\alpha_n|^3\left(1+\alpha_n^2/n\right)^{-n}\right), \tag{2.54}$$
$$L_n - L_n(a,b) = O\left(\sqrt{n}|\alpha_n|^3\left(1+\alpha_n^2/n\right)^{-n}\right). \tag{2.55}$$

Proof. Observe that

$$K_n - K_n(a,b) = \frac{2}{\pi}\int_{|\alpha_n|}^{\infty} \frac{F_n(s/\sqrt{n})}{1+s^2/n} ds,$$
$$L_n - L_n(a,b) = \frac{2}{\pi^2}\int_{|\alpha_n|}^{\infty} \frac{F_n(s/\sqrt{n})}{1+s^2/n} \sqrt{n}\arctan(s/\sqrt{n}) ds.$$

Since $\arctan(s/\sqrt{n})$ is bounded, (2.55) is a consequence of (2.54). To prove (2.54), we first show that, for any fixed $n \geq 1$, as $s \to \infty$,

$$\frac{F_n(s/\sqrt{n})}{1+s^2/n} = \left(1-\frac{1}{n}\right)^2 \frac{s^4}{2}\left(1+\frac{s^2}{n}\right)^{-n-1} + O\left(\left(1+\frac{s^2}{n}\right)^{-n}\right). \tag{2.56}$$

Indeed, as $s \to \infty$,

$$|\Delta_n(s/\sqrt{n})| = \left(1-\frac{1}{n}\right) s^2 \left(1+\frac{s^2}{n}\right)^{-n/2} + O\left(\left(1+\frac{s^2}{n}\right)^{-n/2}\right),$$
$$\Gamma_n(s/\sqrt{n}) = 1 + O\left(\left(1+\frac{s^2}{n}\right)^{-n}\right),$$

and therefore,

$$F_n(s/\sqrt{n}) = \left(1-\frac{1}{n}\right)^2 \frac{s^4}{2}\left(1+\frac{s^2}{n}\right)^{-n} + O\left(s^2\left(1+\frac{s^2}{n}\right)^{-n}\right),$$

which implies (2.56).

We now prove (2.54). Since $|\alpha_n| \to \infty$ as $n \to \infty$, it follows that for all n sufficiently large, the function $s \mapsto s^3\left(1+s^2/n\right)^{-n/2}$ achieves its maximum value on $[|\alpha_n|, \infty)$ at $s = |\alpha_n|$. So

$$\int_{|\alpha_n|}^{\infty} \frac{s^4}{2}\left(1+\frac{s^2}{n}\right)^{-n-1} ds \leq \frac{|\alpha_n|^3}{2}\left(1+\frac{\alpha_n^2}{n}\right)^{-n/2}\int_{|\alpha_n|}^{\infty} s\left(1+\frac{s^2}{n}\right)^{-n/2-1} ds$$
$$= \frac{|\alpha_n|^3}{4}\left(1+\frac{\alpha_n^2}{n}\right)^{-n}.$$

Combining with (2.56) yields (2.54) as required. □

We can now formulate the asymptotic expansions of $K_n(a,b)$ and $L_n(a,b)$.

Lemma 2.19. *As $n \to \infty$, if $\alpha_n^2/\log n \to \infty$, then*

$$K_n(a,b) \sim \sum_{k=0}^{\infty} \frac{\kappa_k}{n^k} \quad \textit{and} \quad L_n(a,b) \sim \sum_{k=0}^{\infty} \frac{\ell_k}{n^k}. \tag{2.57}$$

Consequently, as $n \to \infty$,

$$K_n \sim \sum_{k=0}^{\infty} \frac{\kappa_k}{n^k} \quad \textit{and} \quad L_n \sim \sum_{k=0}^{\infty} \frac{\ell_k}{n^k}. \tag{2.58}$$

Proof. We first prove (2.57). If $\alpha_n^2 \geq n$, one has

$$|\alpha_n|^3 \left(1 + \alpha_n^2/n\right)^{-n} \leq n^{3/2} 2^{-n}.$$

But then (2.54) implies that the integral $\frac{2}{\pi}\int_{\sqrt{n}}^{|\alpha_n|} \frac{F_n(s/\sqrt{n})}{1+s^2/n} ds$ is negligible. Thus, it suffices to assume that $\alpha_n^2 < n$. By (2.28),

$$K_n(a,b) = \sum_{k=0}^{\infty} \left(\frac{2}{\pi} \int_0^{|\alpha_n|} f_k(s) ds \right) \frac{1}{n^k}. \tag{2.59}$$

We now show that

$$\sum_{k=0}^{\infty} \left(\frac{2}{\pi} \int_{|\alpha_n|}^{\infty} f_k(s) ds \right) \frac{1}{n^k} = O\left(|\alpha_n|^3 e^{-\alpha_n^2 + \alpha_n^4/2n} \right). \tag{2.60}$$

In fact, since $|\alpha_n| \to \infty$ as $n \to \infty$, we see that, for any fixed $k \geq 0$ and all n sufficiently large, the function $s \mapsto s^{4k+3} e^{-s^2/2}$ achieves its maximum value on $[|\alpha_n|, \infty)$ at $s = |\alpha_n|$. This implies

$$0 \leq \int_{|\alpha_n|}^{\infty} \frac{s^{4k+4} e^{-s^2}}{2^{k+1} k!} ds \leq \frac{|\alpha_n|^{4k+3} e^{-\alpha_n^2/2}}{2^{k+1} k!} \int_{|\alpha_n|}^{\infty} s e^{-s^2/2} ds = \frac{|\alpha_n|^{4k+3} e^{-\alpha_n^2}}{2^{k+1} k!}.$$

Therefore,

$$\sum_{k=0}^{\infty} \left(\frac{2}{\pi} \int_{|\alpha_n|}^{\infty} \frac{s^{4k+4} e^{-s^2}}{2^{k+1} k!} ds \right) \frac{1}{n^k} \leq \frac{|\alpha_n|^3 e^{-\alpha_n^2 + \alpha_n^4/2n}}{\pi},$$

which gives (2.60) when combined with (2.30). Next, in view of (2.59) and (2.60), the series on the right-hand side of (2.57) converges and

$$K_n(a,b) = \sum_{k=0}^{\infty} \frac{\kappa_k}{n^k} + O\left(|\alpha_n|^3 e^{-\alpha_n^2+\alpha_n^4/2n}\right).$$

Since $O(|\alpha_n|^3 e^{-\alpha_n^2+\alpha_n^4/2n})$ is negligible when $\alpha_n^2/\log n \to \infty$, we get (2.57). The term $L_n(a,b)$ can be handled in much the same way.

Finally, by Lemma 2.18, (2.58) follows from (2.57). □

Note that if α_n^2 does not grow faster than $\log n$, then $|\alpha_n|^3 e^{-\alpha_n^2+\alpha_n^4/2n} \geq n^{-d}$ for some constant $d > 0$. We are thus looking for finite expansions of $K_n(a,b)$ and $L_n(a,b)$.

Lemma 2.20. *If $|\alpha_n| \to \infty$ as $n \to \infty$ and $\alpha_n^2 = O(\log n)$, then for d_n given by (2.7), we have*

$$K_n(a,b) = \sum_{k=0}^{d_n} \frac{\kappa_k}{n^k} + O(|\alpha_n|^3 e^{-\alpha_n^2}) \quad \text{and} \quad L_n(a,b) = \sum_{k=0}^{d_n} \frac{\ell_k}{n^k} + O(\alpha_n^4 e^{-\alpha_n^2}).$$

Proof. Since $\alpha_n^2 = O(\log n)$, d_n is bounded. A slight change in the proof of Lemma 2.16 actually shows that, for $s \in [0, |\alpha_n|]$ and any integer $d \geq 0$,

$$\frac{F_n(s/\sqrt{n})}{1+s^2/n} = \sum_{k=0}^{d} \frac{f_k(s)}{n^k} + O\left(\frac{\alpha_n^{4d}}{n^{d+1}}\right).$$

It follows that

$$\begin{aligned} K_n(a,b) &= \sum_{k=0}^{d_n} \left(\frac{2}{\pi}\int_0^{|\alpha_n|} f_k(s)ds\right)\frac{1}{n^k} + O\left(\frac{|\alpha_n|^{4d_n+1}}{n^{d_n+1}}\right) \\ &= \sum_{k=0}^{d_n} \frac{\kappa_k}{n^k} - \sum_{k=0}^{d_n} \left(\frac{2}{\pi}\int_{|\alpha_n|}^{\infty} f_k(s)ds\right)\frac{1}{n^k} + O\left(\frac{|\alpha_n|^{4d_n+1}}{n^{d_n+1}}\right). \end{aligned}$$

Analysis similar to that in the proof of (2.60) shows

$$\sum_{k=0}^{d_n} \left(\frac{2}{\pi}\int_{|\alpha_n|}^{\infty} f_k(s)ds\right)\frac{1}{n^k} \leq \frac{1+d_n}{\pi}|\alpha_n|^3 e^{-\alpha_n^2} = O(|\alpha_n|^3 e^{-\alpha_n^2}).$$

This clearly forces

$$K_n(a,b) = \sum_{k=0}^{d_n} \frac{\kappa_k}{n^k} + O(|\alpha_n|^3 e^{-\alpha_n^2}).$$

The term $L_n(a,b)$ can be handled in much the same way. □

Our next goal is to establish the asymptotics of $K_n(a,b)$ and $L_n(a,b)$ when $\alpha_n = o(1)$.

Lemma 2.21. *If $\alpha_n = o(1)$ as $n \to \infty$, then*

$$K_n(a,b) = -\frac{2}{\pi}|\alpha_n| + \frac{1}{4}\alpha_n^2 - \frac{1}{4}\frac{\alpha_n^2}{n} + \frac{2}{3\pi}\frac{|\alpha_n|^3}{n} + O(\alpha_n^4), \tag{2.61}$$

$$L_n(a,b) = -\frac{1}{\pi^2}\alpha_n^2 + \frac{1}{6\pi}|\alpha_n|^3 - \frac{1}{6\pi}\frac{|\alpha_n|^3}{n} + \frac{2}{3\pi^2}\frac{\alpha_n^4}{n} + O(|\alpha_n|^5). \tag{2.62}$$

Proof. By (2.28) and (2.29),

$$\begin{aligned} K_n(a,b) &= \frac{2}{\pi}\int_0^{|\alpha_n|}\left[-1 + \frac{\pi}{4}s + \frac{1}{n}\left(-\frac{\pi}{4}s + s^2\right)\right] ds + O(\alpha_n^4) \\ &= -\frac{2}{\pi}|\alpha_n| + \frac{1}{4}\alpha_n^2 - \frac{1}{4}\frac{\alpha_n^2}{n} + \frac{2}{3\pi}\frac{|\alpha_n|^3}{n} + O(\alpha_n^4), \end{aligned}$$

and (2.61) is proved. Similarly, (2.62) follows from applying Lemma 2.17. □

We are now in a position to prove Theorem 2.2.

Proof of Theorem 2.2. The proof is based on the large n behavior of α_n.

1. Assume first that $|\alpha_n| \to \infty$ as $n \to \infty$. Then (2.9) is a consequence of Theorem 2.1 and Lemma 2.19, while (2.8) follows from Theorem 2.1, the relation (2.58), and Lemma 2.20.

2. If $\alpha_n = c > 0$, then

$$\mathbf{E}[N_n(a,b)] = \frac{\sqrt{n}}{\pi}\arctan\frac{c}{\sqrt{n}}.$$

   Thus, using (2.4), Lemmas 2.16 and 2.17, we see that (2.10) holds true. When $\alpha_n = -c$, (2.11) is deduced from (2.6), (2.58), Lemmas 2.16 and 2.17, and

$$\mathbf{E}[N_n(a,b)] = \sqrt{n} + \frac{\sqrt{n}}{\pi}\arctan\frac{c}{\sqrt{n}}.$$

3. Suppose that $\alpha_n = o(1)$ as $n \to \infty$. If $\alpha_n > 0$, then

$$\mathbf{E}[N_n(a,b)] = \frac{\sqrt{n}}{\pi}\arctan\alpha(a,b) = \frac{1}{\pi}\left(\alpha_n - \frac{\alpha_n^3}{3n}\right) + O(\alpha_n^5).$$

Hence, (2.12) follows from (2.4) and Lemma 2.21. Next, for $\alpha_n < 0$,

$$\mathbf{E}[N_n(a,b)] = \sqrt{n} + \frac{1}{\pi}\left(\alpha_n - \frac{\alpha_n^3}{3n}\right) + O(|\alpha_n|^5).$$

Together with (2.6), Lemma 2.21, and the facts that $\sqrt{n}/n^{q_n} = O(|\alpha_n|^5)$ and $\alpha_n/n^{r_n} = O(|\alpha_n|^5)$, we deduce (2.13). If, in addition, $r_n \to \infty$, we have $q_n \to \infty$ and $\alpha_n = o(n^{-r_n/4})$ which is negligible, so (2.14) is indeed a consequence of (2.13). Finally, substituting (2.58) into (2.5) yields (2.15).

□

2.4 Correlations between the real roots

Let $\rho_{n,k}$ and $\rho^T_{n,k}$, respectively, be the k-point correlation and truncated correlation functions of the real roots of the elliptic polynomials. As shown in [11] that if $x_1, ..., x_k$ are k distinct fixed points, then

$$\rho_{n,k}(x_1,...,x_k) = \prod_{j=1}^{k}\left(\frac{\sqrt{n}}{1+x_j^2}\right)\int_{\mathbb{R}^k}|y_1\cdots y_k|D_{n,k}(0,y_1,...,0,y_k;x_1,...,x_k)dy_1\cdots dy_k, \tag{2.63}$$

where $D_{n,k}(s_1,y_1,...,s_k,y_k;x_1,...,x_k)$ is a $(2k)\times(2k)$ Gaussian density with the covariance matrix

$$\Sigma_n = \left(\Sigma_{ij}^{(n)}\right)_{i,j=1}^{k},$$

in which

$$\Sigma_{ij}^{(n)} = (1+\alpha^2(x_i,x_j))^{-n/2}\begin{pmatrix} 1 & -\sqrt{n}\alpha(x_i,x_j) \\ \sqrt{n}\alpha(x_i,x_j) & 1+(1-n)\alpha^2(x_i,x_j)\end{pmatrix}. \tag{2.64}$$

In particular,

$$\rho^T_{n,1}(x_1) = \rho_{n,1}(x_1) = \frac{\sqrt{n}}{\pi(1+x_1^2)}.$$

For $k \geq 2$, to find a scaling limit of $\rho^T_{n,k}$, let us make the change of variables

$$t_j = \sqrt{n}\alpha(x_1,x_j), \quad j = 2,...,k. \tag{2.65}$$

But then

$$\alpha(x_i,x_j) = \alpha(t_i/\sqrt{n}, t_j/\sqrt{n}), \quad i,j = 2,...,k. \tag{2.66}$$

The integral $\int_{\mathbb{R}^k} |y_1 \cdots y_k| D_{n,k}(0, y_1, ..., 0, y_k; x_1, ..., x_k) dy_1 \cdots dy_k$ appeared in (2.63) can be interpreted as a function of $(k-1)$ variables $t_2, ..., t_k$. More precisely, by letting $t_1 = 0$, we deduce from (2.64) and (2.66) that

$$\int_{\mathbb{R}^k} |y_1 \cdots y_k| D_{n,k}(0, y_1, ..., 0, y_k; x_1, x_2, ..., x_k) dy_1 \cdots dy_k \\ = \int_{\mathbb{R}^k} |y_1 \cdots y_k| d_{n,k}(0, y_1, ..., 0, y_k; 0, t_2, ..., t_k) dy_1 \cdots dy_k,$$

in which $d_{n,k}(s_1, y_1, ..., s_k, y_k; t_1, t_2, ..., t_k)$ is a Gaussian density with the covariance matrix

$$\Omega_n = \left(\Omega_{ij}^{(n)}\right)_{i,j=1}^{k},$$

in which

$$\Omega_{ij}^{(n)} = \left(1 + \alpha^2 \left(\frac{t_i}{\sqrt{n}}, \frac{t_j}{\sqrt{n}}\right)\right)^{-\frac{n}{2}} \begin{pmatrix} 1 & -\sqrt{n}\alpha(\frac{t_i}{\sqrt{n}}, \frac{t_j}{\sqrt{n}}) \\ \sqrt{n}\alpha(\frac{t_i}{\sqrt{n}}, \frac{t_j}{\sqrt{n}}) & 1 + (1-n)\alpha^2(\frac{t_i}{\sqrt{n}}, \frac{t_j}{\sqrt{n}}) \end{pmatrix}. \tag{2.67}$$

For $\gamma_j \subset \{1, ..., k\}$ with $l_j = |\gamma_j| \geq 1$, let us introduce the l_j-point functions

$$\Theta_{n,l_j}(\boldsymbol{t}_{\gamma_j}) = \int_{\mathbb{R}^{l_j}} |y_1 \cdots y_{l_j}| d_{n,l_j}(0, y_1, ..., 0, y_{l_j}; \boldsymbol{t}_{\gamma_j}) dy_1 \cdots dy_{l_j},$$

where $\boldsymbol{t}_{\gamma_j} = (t_i)_{i \in \gamma_j}$. We also consider the corresponding truncated functions

$$\Theta_{n,k}^T(t_1, ..., t_k) = \sum_{j=1}^{k} (-1)^{j-1}(j-1)! \sum_{\gamma \in \Pi(k,j)} \Theta_{n,l_1}(\boldsymbol{t}_{\gamma_1}) \cdots \Theta_{n,l_j}(\boldsymbol{t}_{\gamma_j}). \tag{2.68}$$

Put this way, one has

$$\rho_{n,k}^T(\boldsymbol{x}) = \left(\prod_{j=1}^{k} \frac{\sqrt{n}}{1 + x_j^2}\right) \Theta_{n,k}^T(0, t_2, ..., t_k). \tag{2.69}$$

Note that, for fixed $t_i, t_j \in \mathbb{R}$, we have

$$\lim_{n\to\infty} \sqrt{n}\alpha\left(\frac{t_i}{\sqrt{n}}, \frac{t_j}{\sqrt{n}}\right) = t_j - t_i,$$

and by (2.67),

$$\lim_{n\to\infty} \Omega_n = \Omega = (\Omega_{ij})_{i,j=1}^{k},$$

where

$$\Omega_{ij} = e^{-(t_j-t_i)^2/2}\begin{pmatrix} 1 & -(t_j-t_i) \\ t_j-t_i & 1-(t_j-t_i)^2 \end{pmatrix}. \tag{2.70}$$

Therefore,

$$\lim_{n\to\infty}\Theta_{n,k}(t_1,...,t_k) = \int_{\mathbb{R}^k} |y_1\cdots y_k| d_k(0,y_1,...,0,y_k;t_1,...,t_k)dy_1\cdots dy_k,$$

where $d_k(s_1,y_1,...,s_k,y_k;t_1,...,t_k)$ is a Gaussian density with the covariance matrix Ω. Generally, for $\gamma_j \subset \{1,...,k\}$ with $l_j = |\gamma_j| \geq 1$, we can define

$$\Theta_{l_j}(\boldsymbol{t}_{\gamma_j}) := \lim_{n\to\infty}\Theta_{n,l_j}(\boldsymbol{t}_{\gamma_j}). \tag{2.71}$$

Thus, it follows from (2.68) and (2.71) that

$$\lim_{n\to\infty}\Theta^T_{n,k}(t_1,...,t_k) = \sum_{j=1}^{k}(-1)^{j-1}(j-1)!\sum_{\gamma\in\Pi(k,j)}\Theta_{l_1}(\boldsymbol{t}_{\gamma_1})\cdots\Theta_{l_j}(\boldsymbol{t}_{\gamma_j}).$$

Next, we restrict out attention to this scaling limit. Namely, let

$$\Theta^T_{k}(t_1,...,t_k) = \sum_{j=1}^{k}(-1)^{j-1}(j-1)!\sum_{\gamma\in\Pi(k,j)}\Theta_{l_1}(\boldsymbol{t}_{\gamma_1})\cdots\Theta_{l_j}(\boldsymbol{t}_{\gamma_j}). \tag{2.72}$$

Using $\Theta_1 \equiv 1/\pi$, we get

$$\Theta^T_1(t_1) = \frac{1}{\pi}, \quad \Theta^T_2(t_1,t_2) = \Theta_2(t_1,t_2) - \frac{1}{\pi^2},$$
$$\Theta^T_3(t_1,t_2,t_3) = \Theta_3(t_1,t_2,t_3) - \frac{1}{\pi}[\Theta_2(t_1,t_2)+\Theta_2(t_1,t_3)+\Theta_2(t_2,t_3)] + \frac{2}{\pi^3},$$

and so on. The inversions to (2.72) have the form

$$\Theta_k(t_1,...,t_k) = \sum_{j=1}^{k}\sum_{\gamma\in\Pi(k,j)}\Theta^T_{l_1}(\boldsymbol{t}_{\gamma_1})\cdots\Theta^T_{l_j}(\boldsymbol{t}_{\gamma_j}). \tag{2.73}$$

Lemma 2.22. *For $k \geq 2$, we have $\Theta^T_k(0,t_2,...,t_k) \in L^1(\mathbb{R}^{k-1}, dt_2\cdots dt_k)$ and*

$$\int_{\mathbb{R}^{k-1}}\Theta^T_k(0,t_2,...,t_k)dt_2\cdots dt_k = \int_{A_n^{k-1}}\Theta^T_k(0,t_2,...,t_k)dt_2\cdots dt_k + O(e^{-\alpha_n^2/k^2}), \tag{2.74}$$

where $A_n := (-|\alpha_n|, |\alpha_n|)$.

Proof. We first rewrite $\Theta_k(t_1, t_2, ..., t_k)$ in a more explicit form

$$\Theta_k(t_1, t_2, ..., t_k) = \frac{1}{(2\pi)^k \sqrt{\det \Omega}} \int_{\mathbb{R}^k} |y_1 \cdots y_k| e^{-\frac{1}{2}\langle \Omega^{-1}\boldsymbol{y}, \boldsymbol{y}\rangle} dy_1 \cdots dy_k,$$

where $\boldsymbol{y} = (0, y_1, ..., 0, y_k)$. It was shown in [11, Appendix C] that $\Omega > 0$ at distinct points t_j, so $\Theta_k(t_1, t_2, ..., t_k)$ is well-defined when the point t_j are distinct. Let $\operatorname{adj}(\Omega)$ denote the adjugate of Ω, so $\Omega^{-1} = (\det \Omega)^{-1} \operatorname{adj}(\Omega)$. Making the change of variables $y_j = \eta_j \sqrt{\det \Omega}$, $j = 1, ..., k$, we obtain

$$\Theta_k(t_1, t_2, ..., t_k) = \frac{(\det \Omega)^{(k-1)/2}}{(2\pi)^k} \int_{\mathbb{R}^k} |\eta_1 \cdots \eta_k| e^{-\frac{1}{2}\langle \operatorname{adj}(\Omega)\boldsymbol{\eta}, \boldsymbol{\eta}\rangle} d\eta_1 \cdots d\eta_k,$$

where $\boldsymbol{\eta} = (0, \eta_1, ..., 0, \eta_k)$. This formula implies that $\Theta_k(t_1, t_2, ..., t_k)$ has a continuous extension to the entire space such that $\Theta_k(t_1, t_2, ..., t_k) = 0$ whenever $t_i = t_j$ for some $i \neq j$. By (2.72), $\Theta_k^T(t_1, t_2, ..., t_k)$ also has a continuous extension to the entire space. Thus, $\int_{A_n^{k-1}} \Theta_k^T(0, t_2, ..., t_k) dt_2 \cdots dt_k$ is well-defined. Let $\mathbb{R}^{k-1}(\alpha_n) = \mathbb{R}^{k-1} \backslash A_n^{k-1}$. It remains to show that

$$\int_{\mathbb{R}^{k-1}(\alpha_n)} \Theta_k^T(0, t_2, ..., t_k) dt_2 \cdots dt_k = O(e^{-\alpha_n^2/k^2}). \tag{2.75}$$

By (2.73),

$$\Theta_k^T(t_1, t_2, ..., t_k) = \Theta_k(t_1, t_2, ..., t_k) - \sum_{j=2}^{k} \sum_{\gamma \in \Pi(k,j)} \Theta_{l_1}^T(\boldsymbol{t}_{\gamma_1}) \cdots \Theta_{l_j}^T(\boldsymbol{t}_{\gamma_j}).$$

Thus, if $t_i = t_j$ for some $i \neq j$, then $\Theta_k^T(t_1, t_2, ..., t_k)$ is completely expressed in terms of the j-point truncated functions for $j < k$. Moreover, we have $\Theta_2^T(0, t_2) = f_0(t_2)/\pi^2$ which satisfies the conclusion of the lemma. Hence, by induction on k, it suffices to treat the case where

$$\min_{i \neq j} |t_i - t_j| \geq r > 0.$$

Under this condition, we conclude from (2.70) that, as $r \to \infty$, the matrix Ω approaches the unit matrix with the rate of convergence being $O(r^2 e^{-r^2/2})$. This gives

$$\Theta_k(t_1, t_2, ..., t_k) = \frac{1}{\pi^k} + O(r^4 e^{-r^2}) \quad \text{as} \quad r \to \infty.$$

Consequently, in exactly the same way as in [13, Corollary 5.8], we infer that

$$\Theta_k^T(t_1, t_2, ..., t_k) = o(e^{-R^2/k}) \quad \text{as} \quad R \to \infty,$$

where $R = \max_{i \neq j} |t_i - t_j|$. Since $t_1 = 0$, it follows that $R \geq |t_j|$, for $j = 2, ..., k$. Hence, as $t_j \to \infty$,

$$\Theta_k^T(0, t_2, ..., t_k) = O\left(\prod_{j=2}^{k} e^{-t_j^2/k^2}\right),$$

which gives (2.75). The lemma is proved. □

Remark 2.23. The proof above gives more, namely for $k \geq 2$ and a polynomial $P(t_2, ..., t_k)$, we have $P(t_2, ..., t_k)\Theta_k^T(0, t_2, ..., t_k) \in L^1(\mathbb{R}^{k-1}, dt_2 \cdots dt_k)$.

Lemma 2.24. *For $k \geq 2$ and $A_n = (-|\alpha_n|, |\alpha_n|)$, we have*

$$\int_{\mathbb{R}^{k-1}} \Theta_{n,k}^T(0, t_2, ..., t_k) dt_2 \cdots dt_k = \int_{A_n^{k-1}} \Theta_{n,k}^T(0, t_2, ..., t_k) dt_2 \cdots dt_k + O\left((1 + \alpha_n^2/n)^{-n/k^2}\right). \tag{2.76}$$

If, in addition, $|\alpha_n| < \sqrt{n}$, one has

$$\int_{A_n^{k-1}} \Theta_{n,k}^T(0, t_2, ..., t_k) dt_2 \cdots dt_k = \int_{A_n^{k-1}} \Theta_k^T(0, t_2, ..., t_k) dt_2 \cdots dt_k + O\left(\alpha_n^2/n\right). \tag{2.77}$$

Proof. The estimate (2.76) follows the same method as in the proof of Lemma 2.22. Next, if $t_i, t_j \in A_n$ and $|\alpha_n| < \sqrt{n}$,

$$\sqrt{n}\alpha\left(\frac{t_i}{\sqrt{n}}, \frac{t_j}{\sqrt{n}}\right) = (t_j - t_i)\left(1 + O\left(\alpha_n^2/n\right)\right)$$

and by (2.67),

$$\Omega_n = \Omega\left(1 + O\left(\alpha_n^2/n\right)\right).$$

This gives

$$\Theta_{n,k}^T(t_1, t_2, ..., t_k) = \Theta_k^T(t_1, t_2, ..., t_k)\left(1 + O\left(\alpha_n^2/n\right)\right),$$

which yields (2.77). □

2.5 Asymptotics of the cumulants

In this section, we give the proof of Theorem 2.3. We begin by recalling the relation between the cumulant $s_k[N_n(a, b)]$ and the truncated correlation functions $\rho_{n,j}^T$, for

$1 \leq j \leq k$. By Lemma 1.1,

$$s_k[N_n(a,b)] = \sum_{\gamma \in \Pi(k)} \int_{(a,b)^{|\gamma|}} \rho^T_{n,|\gamma|}(\boldsymbol{x}_\gamma) d\boldsymbol{x}_\gamma, \tag{2.78}$$

where $|\gamma|$ is the number of blocks in the partition γ and $d\boldsymbol{x}_\gamma$ is the Lebesgue measure on $(a,b)^{|\gamma|}$.

Thus, the task is now to estimate the integrals $\int_{(a,b)^{|\gamma|}} \rho^T_{n,|\gamma|}(\boldsymbol{x}) d\boldsymbol{x}$.

Lemma 2.25. *For $k \geq 1$, we have, as $n \to \infty$,*

$$\int_{(a,b)^k} \rho^T_{n,k}(x_1, ..., x_k) dx_1 \cdots dx_k = C_k \mathbf{E}[N_n(a,b)] + O(1) \tag{2.79}$$

in which $C_1 = 1$ and

$$C_k = \pi \int_{\mathbb{R}^{k-1}} \Theta^T_k(0, t_2, ..., t_k) dt_2 \cdots dt_k, \quad k \geq 2.$$

Notice that Theorem 2.3 immediately follows from applying (2.78) and Lemma 2.25. Indeed, substituting (2.79) into (1.1), we obtain (2.16), in which

$$\beta_k := \sum_{\gamma \in \Pi(k)} C_{|\gamma|}.$$

For $k = 1$, Lemma 2.25 is trivial. Assume now that $k \geq 2$. By Lemma 2.22, the constants C_k, for $k \geq 2$, are well-defined. Making the change of variables (2.65), we see that

$$\frac{\sqrt{n}}{1 + x_j^2} dx_j = \frac{1}{1 + t_j^2/n} dt_j, \quad j = 2, ..., k,$$

which together with (2.69) yields

$$\rho^T_{n,k}(x_1, ..., x_k) dx_1 \cdots dx_k = \frac{\sqrt{n}}{1 + x_1^2} \frac{\Theta^T_{n,k}(0, t_2, ..., t_k)}{\prod_{j=2}^k (1 + t_j^2/n)} dx_1 dt_2 \cdots dt_k, \tag{2.80}$$

where $\Theta^T_{n,k}(t_1, t_2, ..., t_k)$ is given by (2.68).

To prove Lemma 2.25, we consider three cases of α_n. To shorten the notation, let $I_{n,k}(a,b)$ stand for the integral on the left-hand side of (2.79).

Claim 1. *If $\alpha_n > 0$, then (2.79) holds.*

Proof. Using (2.80), we obtain

$$I_{n,k}(a,b) = \int_a^b \frac{\sqrt{n}}{1+x_1^2} dx_1 \int_{R_n(x_1)} \frac{\Theta^T_{n,k}(0,t_2,...,t_k)}{\prod_{j=2}^k (1+t_j^2/n)} dt_2 \cdots dt_k,$$

where

$$R_n(x_1) := \left\{(t_2,...,t_k) \in (a,b)^{k-1} : \sqrt{n}\alpha(x_1,a) < t_2,...,t_k < \sqrt{n}\alpha(x_1,b)\right\}.$$

By Fubini's Theorem,

$$I_{n,k}(a,b) = \int_{(-\alpha_n,\alpha_n)^{k-1}} \frac{\Theta^T_{n,k}(0,t_2,...,t_k)}{\prod_{j=2}^k (1+t_j^2/n)} dt_2 \cdots dt_k \int_a^b \sqrt{n}\frac{G(x_1,t_2,...,t_k)}{1+x_1^2} dx_1,$$

where

$$G(x_1,t_2,...,t_k) := \prod_{j=2}^k \left(\mathbf{1}_{(-\alpha_n,0)}(t_j)\mathbf{1}_{(\alpha(t_j/\sqrt{n},a),b)}(x_1) + \mathbf{1}_{(0,\alpha_n)}(t_j)\mathbf{1}_{(a,\alpha(t_j/\sqrt{n},b))}(x_1)\right).$$

For $k \geq 2$, let $\Lambda(k)$ be the set of all ordered pair (λ_1,λ_2) of disjoint subsets of $\{2,...,k\}$ such that $\lambda_1 \cup \lambda_2 = \{2,...,k\}$. For each $\lambda = (\lambda_1,\lambda_2) \in \Lambda(k)$, we introduce the function

$$G_\lambda(x_1,t_2,...,t_k) := \prod_{j\in\lambda_1} \mathbf{1}_{(-\alpha_n,0)}(t_j)\mathbf{1}_{(\alpha(t_j/\sqrt{n},a),b)}(x_1) \prod_{i\in\lambda_2} \mathbf{1}_{(0,\alpha_n)}(t_i)\mathbf{1}_{(a,\alpha(t_i/\sqrt{n},b))}(x_1)$$

so that

$$I_{n,k}(a,b) = \int_{(-\alpha_n,\alpha_n)^{k-1}} \frac{\Theta^T_{n,k}(0,t_2,...,t_k)}{\prod_{j=2}^k (1+t_j^2/n)} dt_2 \cdots dt_k \sum_{\lambda\in\Lambda(k)} \int_a^b \sqrt{n}\frac{G_\lambda(x_1,t_2,...,t_k)}{1+x_1^2} dx_1.$$

For each $\lambda = (\lambda_1,\lambda_2) \in \Lambda(k)$ and $(t_2,...,t_k) \in \mathbb{R}^{k-1}$, let

$$t^{\min}_{\lambda_1} = \begin{cases} 0 & \text{if} \quad \lambda_1 = \varnothing, \\ \min_{j\in\lambda_1} t_j & \text{if} \quad \lambda_1 \neq \varnothing, \end{cases} \quad \text{and} \quad t^{\max}_{\lambda_2} = \begin{cases} 0 & \text{if} \quad \lambda_2 = \varnothing, \\ \max_{i\in\lambda_2} t_i & \text{if} \quad \lambda_2 \neq \varnothing. \end{cases}$$

By a direct computation, we obtain

$$\int_a^b \sqrt{n}\frac{G_\lambda(x_1,t_2,...,t_k)}{1+x_1^2}dx_1 = \pi \prod_{j\in\lambda_1}\mathbf{1}_{(-\alpha_n,0)}(t_j)\prod_{i\in\lambda_2}\mathbf{1}_{(0,\alpha_n)}(t_i)\mathbf{E}[N_n(a,b)]$$
$$+\prod_{j\in\lambda_1}\mathbf{1}_{(-\alpha_n,0)}(t_j)\prod_{i\in\lambda_2}\mathbf{1}_{(0,\alpha_n)}(t_i)\sqrt{n}\left(\arctan\frac{t_{\lambda_1}^{\min}}{\sqrt{n}}-\arctan\frac{t_{\lambda_2}^{\max}}{\sqrt{n}}\right).$$

For any fixed $(t_2,...,t_k)\in\mathbb{R}^{k-1}$, we have

$$\lim_{n\to\infty}\sqrt{n}\left(\arctan\frac{t_{\lambda_1}^{\min}}{\sqrt{n}}-\arctan\frac{t_{\lambda_2}^{\max}}{\sqrt{n}}\right)=t_{\lambda_1}^{\min}-t_{\lambda_2}^{\max}.$$

Together with Remark 2.23, we can assert that, as $n\to\infty$,

$$I_{n,k}(a,b)=\left(\pi\int_{(-\alpha_n,\alpha_n)^{k-1}}\frac{\Theta_{n,k}^T(0,t_2,...,t_k)}{\prod_{j=2}^k(1+t_j^2/n)}dt_2\cdots dt_k\right)\mathbf{E}[N_n(a,b)]+O(1).$$

Note that if $\alpha_n>\log n$, then by (2.76),

$$\int_{(-\alpha_n,\alpha_n)^{k-1}}\frac{\Theta_{n,k}^T(0,t_2,...,t_k)}{\prod_{j=2}^k(1+t_j^2/n)}dt_2\cdots dt_k$$
$$=\int_{(-\log n,\log n)^{k-1}}\frac{\Theta_{n,k}^T(0,t_2,...,t_k)}{\prod_{j=2}^k(1+t_j^2/n)}dt_2\cdots dt_k+O((1+\log^2 n/n)^{-n/k^2}),$$

in which the term $O((1+\log^2 n/n)^{-n/k^2})$ is negligible because

$$O((1+\log^2 n/n)^{-n/k^2})\mathbf{E}[N_n(a,b)]=o(1).$$

Thus, it suffices to assume that $\alpha_n\le\log n$. Using (2.74), (2.77), and the fact that

$$\frac{\Theta_{n,k}^T(0,t_2,...,t_k)}{\prod_{j=2}^k(1+t_j^2/n)}=\Theta_{n,k}^T(0,t_2,...,t_k)(1+O(\alpha_n^2/n)),$$

we get

$$\pi\int_{(-\alpha_n,\alpha_n)^{k-1}}\frac{\Theta_{n,k}^T(0,t_2,...,t_k)}{\prod_{j=2}^k(1+t_j^2/n)}dt_2\cdots dt_k=C_k+O(e^{-\alpha_n^2/k^2}+\alpha_n^2/n).$$

Since $O(e^{-\alpha_n^2/k^2}+\alpha_n^2/n)\mathbf{E}[N_n(a,b)]=O(1)$, it follows that the asymptotic formula (2.79) holds true. □

Claim 2. *The asymptotic formula* (2.79) *holds for* $\alpha_n = 0$.

Proof. When $\alpha_n = 0$, we have $(a, b) = \mathbb{R}$. By (2.80), Lemmas 2.22 and 2.24,

$$\begin{aligned} I_{n,k}(\mathbb{R}) &= \int_{-\infty}^{\infty} \frac{\sqrt{n}}{1+x_1^2} dx_1 \int_{\mathbb{R}^{k-1}} \frac{\Theta_{n,k}^T(0, t_2, ..., t_k)}{\prod_{j=2}^k (1+t_j^2/n)} dt_2 \cdots dt_k \\ &= \left(\pi \int_{\mathbb{R}^{k-1}} \frac{\Theta_{n,k}^T(0, t_2, ..., t_k)}{\prod_{j=2}^k (1+t_j^2/n)} dt_2 \cdots dt_k \right) \mathbf{E}[N_n(\mathbb{R})] \\ &= C_k \mathbf{E}[N_n(\mathbb{R})] + o(1), \end{aligned}$$

which yields (2.79). □

Claim 3. *If* $\alpha_n < 0$, *then* (2.79) *takes place.*

Proof. We first write

$$I_{n,k}(a, b) = \left(\int_a^{-1/b} + \int_{-1/b}^{-1/a} + \int_{-1/a}^{b} \right) dx_1 \int_{(a,b)^{k-1}} \rho_{n,k}^T(\boldsymbol{x}) dx_2 \cdots dx_k.$$

By (2.80) and Fubini's Theorem,

$$I_{n,k}(a, b) = \int_{\mathbb{R}^{k-1}} \frac{\Theta_{n,k}^T(0, t_2, ..., t_k)}{\prod_{j=2}^k (1+t_j^2/n)} dt_2 \cdots dt_k \int_a^b \sqrt{n} \frac{H(x_1, t_2, ..., t_k)}{1+x_1^2} dx_1,$$

where

$$\begin{aligned} H(x_1, t_2, ..., t_k) = \mathbf{1}_{(a,-1/b)}(x_1) \prod_{j=2}^k \Big(&\mathbf{1}_{(-\infty,\alpha_n)}(t_j) \mathbf{1}_{(a,\alpha(t_j/\sqrt{n},b))}(x_1) \\ &+ \mathbf{1}_{(n/\alpha_n,0)}(t_j) \mathbf{1}_{(\alpha(t_j/\sqrt{n},a),-1/b)}(x_1) + \mathbf{1}_{(0,\infty)}(t_j) \Big) \\ + \mathbf{1}_{(-1/b,-1/a)}(x_1) \prod_{j=2}^k \Big(&\mathbf{1}_{(-\infty,n/\alpha_n)}(t_j) \mathbf{1}_{(\alpha(t_j/\sqrt{n},a),-1/a)}(x_1) \\ &+ \mathbf{1}_{(n/\alpha_n,-n/\alpha_n)}(t_j) + \mathbf{1}_{(-n/\alpha_n,\infty)}(t_j) \mathbf{1}_{(-1/b,\alpha(t_j/\sqrt{n},b))}(x_1) \Big) \\ + \mathbf{1}_{(-1/a,b)}(x_1) \prod_{j=2}^k \Big(&\mathbf{1}_{(-\infty,0)}(t_j) + \mathbf{1}_{(0,-n/\alpha_n)}(t_j) \mathbf{1}_{(-1/a,\alpha(t_j/\sqrt{n},b))}(x_1) \\ &+ \mathbf{1}_{(-\alpha_n,\infty)}(t_j) \mathbf{1}_{(\alpha(t_j/\sqrt{n},a),b)}(x_1) \Big). \end{aligned}$$

Analysis similar to that in the proof of Claim 1 shows

$$I_{n,k}(a,b) = \left(\pi \int_{\mathbb{R}^{k-1}} \frac{\Theta_{n,k}^{T}(0, t_2, ..., t_k)}{\prod_{j=2}^{k}(1 + t_j^2/n)} dt_2 \cdots dt_k \right) \mathbf{E}[N_n(a,b)] + O(1),$$

which establishes the asymptotic formula (2.79). □

2.6 Asymptotics of the central moments

This section is devoted to the proof of Corollary 2.6. Observe first that the explicit expression for the k-th central moment in terms of the first k cumulants can be obtained by using Faà di Bruno's formula for higher derivatives of composite functions. More precisely, let

$$K(t) = \log \mathbf{E}\left[e^{tN_n(a,b)}\right] \quad \text{and} \quad C(t) = \mathbf{E}\left[e^{t(N_n(a,b) - \mathbf{E}[N_n(a,b)])}\right].$$

Then

$$s_k[N_n(a,b)] = \frac{d^k}{dt^k} K(t)\bigg|_{t=0} \quad \text{and} \quad \mu_k[N_n(a,b)] = \frac{d^k}{dt^k} C(t)\bigg|_{t=0}.$$

Since

$$C(t) = e^{K(t) - t\mathbf{E}[N_n(a,b)]},$$

it follows from Faà di Bruno's formula that

$$\mu_k[N_n(a,b)] = \frac{d^k}{dt^k} e^{K(t) - t\mathbf{E}[N_n(a,b)]}\bigg|_{t=0} = \sum_{j=1}^{k} B_{k,j}(0, s_2[N_n(a,b)], ..., s_{k-j+1}[N_n(a,b)]).$$

Recall that, for $1 \leq j \leq k$,

$$B_{k,j}(x_1, ..., x_{k-j+1}) = \sum \frac{k!}{m_1! \cdots m_{k-j+1}!} \prod_{r=1}^{k-j+1} \left(\frac{x_r}{r!}\right)^{m_r},$$

where the sum is over all solutions in non-negative integers of the equations

$$\begin{aligned} m_1 + 2m_2 + \cdots + (k-j+1)m_{k-j+1} &= k, \\ m_1 + m_2 + \cdots + m_{k-j+1} &= j. \end{aligned}$$

Note that $m_1 \geq 1$ whenever $j > k/2$, so $B_{k,j}(0, x_2, ..., x_{k-j+1}) = 0$ for all $j > k/2$. Therefore, for $k \geq 2$,

$$\mu_k[N_n(a,b)] = \sum_{j=1}^{\lfloor k/2 \rfloor} B_{k,j}(0, s_2[N_n(a,b)], ..., s_{k-j+1}[N_n(a,b)]).$$

Together with (2.16), we obtain

$$\begin{aligned}\mu_{2k}[N_n(a,b)] &= B_{2k,k}(0, s_2[N_n(a,b)], ..., s_{k+1}[N_n(a,b)]) + O((\mathbf{E}[N_n(a,b)])^{k-1}) \\ &= \frac{(2k)!}{k!}\left(\frac{s_2[N_n(a,b)]}{2!}\right)^k + O((\mathbf{E}[N_n(a,b)])^{k-1}),\end{aligned}$$

which yields (2.17). Similar arguments apply to (2.18).

2.7 Asymptotic normality for the real roots

We now prove Theorem 2.8. The main idea is the cumulant convergence theorem of Janson [42, Theorem 1].

Proposition 2.26 ([42]). *Let $\{X_n\}$ be a sequence of random variables such that, as $n \to \infty$,*

- $s_1[X_n] \to 0$,
- $s_2[X_n] \to 1$, *and*
- $s_k[X_n] \to 0$ *for every* $k \geq m$,

where $m \geq 3$. Then $X_n \xrightarrow{d} \mathcal{N}(0,1)$ as $n \to \infty$. Furthermore, all moments of X_n converge to the corresponding moments of $\mathcal{N}(0,1)$.

To deduce Theorem 2.8, let us consider

$$X_n := \frac{N_n(a,b) - \mathbf{E}[N_n(a,b)]}{\sqrt{\mathbf{Var}[N_n(a,b)]}}, \quad n \geq 1.$$

We must show that as $n \to \infty$, $X_n \xrightarrow{d} \mathcal{N}(0,1)$, provided that either $\alpha_n \leq 0$ or $\alpha_n \to \infty$ as $n \to \infty$.

Evidently, $s_1[X_n] = \mathbf{E}[X_n] = 0$ and $s_2[X_n] = \mathbf{Var}[X_n] = 1$. By Proposition 2.26, it remains to show that the higher cumulants of X_n converge to 0 as $n \to \infty$. In fact, for $k \geq 3$, we have

$$s_k[X_n] = \frac{s_k[N_n(a,b)]}{(\mathbf{Var}[N_n(a,b)])^{k/2}}.$$

If either $\alpha_n \leq 0$ or $\alpha_n \to \infty$ as $n \to \infty$, then $\mathbf{E}[N_n(a,b)] \to \infty$. By Theorem 2.3,

$$s_k[X_n] = \frac{\beta_k \mathbf{E}[N_n(a,b)] + O(1)}{(\beta_2 \mathbf{E}[N_n(a,b)] + O(1))^{k/2}} \to 0 \quad \text{as} \quad n \to \infty.$$

2.8 Strong law of large numbers

This section deals with the proof of Theorem 2.9. The given assumption implies that

$$\sum_{n=1}^{\infty} \frac{1}{(\mathbf{E}[N_n(a,b)])^k} < \infty.$$

But then, due to Corollary 2.6,

$$\mathbf{E}\left[\sum_{n=1}^{\infty} \left(\frac{N_n(a,b)}{\mathbf{E}[N_n(a,b)]} - 1\right)^{2k}\right] = \sum_{n=1}^{\infty} \frac{\mu_{2k}[N_n(a,b)]}{(\mathbf{E}[N_n(a,b)])^{2k}} = O\left(\sum_{n=1}^{\infty} \frac{1}{(\mathbf{E}[N_n(a,b)])^k}\right) < \infty.$$

Then, almost surely, we have

$$\sum_{n=1}^{\infty} \left(\frac{N_n(a,b)}{\mathbf{E}[N_n(a,b)]} - 1\right)^{2k} < \infty,$$

which implies

$$\frac{N_n(a,b)}{\mathbf{E}[N_n(a,b)]} - 1 \xrightarrow{a.s.} 0 \quad \text{as } n \to \infty,$$

and the theorem is proved.

Chapter 3

Variance of the real roots of generalized Kac polynomials

This chapter deals with the real roots of random polynomials whose deterministic coefficients have polynomial asymptotics [26]. We first compute the precise leading asymptotics of the variance of the number of real roots (Theorem 3.1). The main ingredients of the proof are new asymptotic estimates for the two-point correlation function of the real roots, revealing geometric structures in the distribution of the real roots of these random polynomials. As a corollary, we obtain asymptotic normality for the number of real roots for these random polynomials (Corollary 3.5).

3.1 Introduction and main results

We consider random polynomials of the form

$$P_n(x) = \sum_{j=0}^{n} \xi_j c_j x^j, \tag{3.1}$$

where ξ_j are independent real-valued random variables with zero mean $\mathbf{E}[\xi_j] = 0$ and unit variance $\mathbf{Var}[\xi_j] = 1$, and c_j are deterministic real-valued coefficients with polynomial asymptotics. The precise technical conditions for c_j will be formulated shortly. The class of random polynomials in our main results, Theorem 3.1, contains several interesting ensembles of random polynomials, including (but not limited to) the following examples:

- the Kac polynomials, where $c_j = 1$ for all j;
- hyperbolic polynomials, where $c_j = \sqrt{\frac{L(L+1)\cdots(L+j-1)}{j!}}$ (for some fixed $L > 0$); and
- linear combinations of the derivatives of the above examples.

Recall that for the Gaussian Kac polynomials, Kac [43] showed that

$$\mathbf{E}[N_n(\mathbb{R})] = \frac{2}{\pi}\log n + o(\log n). \tag{3.2}$$

Edelman and Kostlan [30] improved (3.2) by finding two more asymptotic terms. The complete asymptotic expansion of $\mathbf{E}[N_n(\mathbb{R})]$ was earlier given by Wilkins [71].

The asymptotic formula (3.2) has been extended to Kac polynomials with general random coefficients. Namely, Kac [44] proved (3.2) when ξ_j are independent and uniformly distributed on $[-1, 1]$, and Stevens [68] extended this result further to cover a large class of smooth distributions with certain regularity properties. In 1956, Erdős and Offord [31] extended the result to the Bernoulli distribution case. In the late 1960s and early 1970s, Ibragimov and Maslova [38, 40] successfully refined Erdős-Offord's method to extend the result to all mean-zero distributions in the domain of attraction of the normal law, with the extra assumption that $\mathbf{P}(\{\xi_j = 0\}) = 0$. In 2016, H. Nguyen, O. Nguyen, and V. Vu [57] removed this extra condition and furthermore showed that the error term in (3.2) is bounded. For several interesting classes of Kac polynomials, the nature of the error term in (3.2) was shown to be of the form $C + o(1)$ in a joint work of the first author with H. Nguyen and V. Vu [25].

The formula (3.2) has recently been extended to generalized Kac polynomials, which are random polynomials of the form (3.1) where the deterministic coefficients c_j have polynomial growth. In Do, Nguyen, and Vu [27], the polynomial growth condition is formulated as follows. For some $C_1, C_2, C_3, N_0 > 0$ fixed (independent of n) assume that

$$\begin{cases} C_1 j^\tau \le |c_j| \le C_2 j^\tau & \text{if} \quad N_0 \le j \le n, \\ c_j^2 \le C_3 & \text{if} \quad 0 \le j < N_0. \end{cases} \tag{3.3}$$

The order of growth τ is also assumed to be independent of n. Note that c_j are allowed to depend on n; otherwise, the second condition in (3.3) is superfluous. In [27], it was proved that if ξ_j have uniformly bounded $2 + \varepsilon$ moments and $\tau > -1/2$, then $\mathbf{E}[N_n(\mathbb{R})]$ grows logarithmically with respect to n. In [27], it was furthermore shown that if $|c_j|$ have polynomial asymptotics, namely, there is a fixed constant $C_1 > 0$ such that[1]

$$|c_j| = C_1 j^\tau (1 + o_j(1)), \tag{3.4}$$

then

$$\mathbf{E}[N_n(\mathbb{R})] = \frac{1 + \sqrt{2\tau + 1}}{\pi}\log n + o(\log n). \tag{3.5}$$

[1]The $o_j(1)$ notation in (3.4) means that this term can be bounded by some o_j independent of n such that $\lim_{j\to\infty} o_j = 0$.

See also Do [22] for an extension of [27] to settings with non-centered random coefficients. The asymptotic formula (3.5) in the special cases when ξ_j are Gaussian and $c_j = j^\tau$ for some $\tau > 0$ was previously formulated by Das [19, 20], Sambandham [62], Sambandham, Gore, and Farahmand [63], Schehr and Majumdar [65, 66]. For non-Gaussian cases, (3.5) recovers Maslova's result in [54, Theorem 2] for the first derivatives of the Kac polynomials.

Generalized Kac polynomials appear naturally when considering derivatives of the Kac polynomials and hyperbolic polynomials. Recently, they have also attracted research attention in the mathematical physics community. In particular, Schehr and Majumdar [65, 66] made a connection between the persistence exponent of the diffusion equation with random initial conditions and the probability that a certain generalized Kac polynomial has no real root in a given interval. A more complete treatment was given later by Dembo and Mukherjee [21], who derived general criteria for continuity of persistence exponents for centered Gaussian processes. The authors of [21] then used these criteria to study the gap probabilities for both real roots of random polynomials and zero-crossings of solutions to the heat equation initiated by Gaussian white noise.

Evaluating the variance of $N_n(\mathbb{R})$ for generalized Kac polynomials has proved to be a much more difficult task. As far as we know, despite a large number of prior studies, the only result that establishes the leading asymptotics for $\mathbf{Var}[N_n(\mathbb{R})]$ is for the Kac polynomials, a celebrated result of Maslova [54] from the 1970s, who proved that if ξ_j are independent identically distributed random variables such that $\mathbf{P}(\{\xi_j = 0\}) = 0$, $\mathbf{E}[\xi_j] = 0$, and $\mathbf{E}[|\xi_j|^{2+\varepsilon}] = O(1)$, then

$$\mathbf{Var}[N_n(\mathbb{R})] = \left[\frac{4}{\pi}\left(1 - \frac{2}{\pi}\right) + o(1)\right] \log n. \tag{3.6}$$

While the condition $\mathbf{P}(\{\xi_j = 0\}) = 0$ has been removed by O. Nguyen and V. Vu [60], there has been no other result of this type for other generalized Kac polynomials (even for the Gaussian setting when ξ_j are all Gaussian). We mention here another result due to Sambandham, Thangaraj, and Bharucha-Reid [64], who proved an estimate of $\mathbf{Var}[N_n(\mathbb{R})]$ for random Kac polynomials with dependent Gaussian coefficients.

For generalized Kac polynomials, O. Nguyen and V. Vu [60] have recently proved the following lower bound:

$$\log n \ll \mathbf{Var}[N_n(\mathbb{R})], \tag{3.7}$$

provided that

$$\frac{|c_j|}{|c_n|} - 1 = O\left(e^{-(\log\log n)^{1+\varepsilon}}\right), \quad n - ne^{-\log^{1/5} n} \le j \le n - e^{\log^{1/5} n}. \tag{3.8}$$

Here we use the usual asymptotic notation $X \ll Y$ or $X = O(Y)$ to denote the bound $|X| \le cY$ where c is independent of Y. The other assumptions for (3.7) needed in [60] include $\sup_{1\le j\le n} \mathbf{E}[|\xi_j|^{2+\varepsilon}] = O(1)$ and the polynomial growth condition (3.3).

In this dissertation, we are interested in establishing the leading asymptotics of $\mathbf{Var}[N_n(\mathbb{R})]$ for generalized Kac polynomials whose deterministic coefficients have polynomial asymptotics, as described in condition (A2). We recall that the leading asymptotics for $\mathbf{E}[N_n(\mathbb{R})]$ in [27] was also established under the same assumption; therefore, this setting seems reasonable for us to consider the leading asymptotics for the variance.

We assume that there are fixed positive constants C_0, C_1, C_2, N_0, ε, and a fixed constant $\tau > -1/2$, where

(A1) $\xi_0, \dots, \xi_n$ are independent real-valued random variables, with $\mathbf{E}[\xi_j] = 0$ for $j \ge N_0$, $\mathbf{Var}[\xi_j] = 1$ for $j \ge 0$, and $\sup_{0\le j\le n} \mathbf{E}[|\xi_j|^{2+\varepsilon}] < C_0$,

(A2) each c_j is real and may depend on both j and n, such that

$$\begin{cases} |c_j| = C_1 j^\tau (1 + o_j(1)), & \text{for} \quad N_0 \le j \le n, \\ |c_j| \le C_2, & \text{for} \quad 0 \le j < N_0. \end{cases}$$

To formulate our results, we first fix some notations. Let

$$f_\tau(u) = \left(\sqrt{1 - \Delta_\tau^2(u)} + \Delta_\tau(u) \arcsin \Delta_\tau(u)\right) \Sigma_\tau(u) - 1, \tag{3.9}$$

where

$$\begin{aligned} \Delta_\tau(u) &:= u^{\tau+1/2} \frac{u(1-u^{2\tau+1}) - (2\tau+1)(1-u)}{1 - u^{2\tau+1} - (2\tau+1)u^{2\tau+1}(1-u)}, \\ \Sigma_\tau(u) &:= \frac{1 - u^{2\tau+1} - (2\tau+1)(1-u)u^{2\tau+1}}{(1-u^{2\tau+1})^{3/2}}, \end{aligned} \tag{3.10}$$

and let

$$\kappa_\tau = \left(\frac{2\tau+1}{\pi} \int_0^\infty f_\tau(\operatorname{sech}^2 v) dv + \frac{\sqrt{2\tau+1}}{2}\right) \frac{1}{\pi}. \tag{3.11}$$

Some basic properties of f_τ (including integrability) are collected in Lemma 3.22.

Theorem 3.1 (Asymptotics of variances)**.** *Assume that the polynomial P_n defined by* (3.1) *satisfies conditions* (A1) *and* (A2)*. Then*

$$\mathbf{Var}[N_n(\mathbb{R})] = \left[2\kappa_\tau + \frac{2}{\pi}\left(1 - \frac{2}{\pi}\right) + o(1)\right] \log n,$$

where the implicit constants in the $o(1)$ term depend only on N_0, C_0, C_1, C_2, ε, τ, and the rate of decay of $o_j(1)$ in condition (A2)*.*

Remark 3.2. When $\tau = 0$ we have

$$\int_0^\infty f_\tau(\operatorname{sech}^2 v)dv = \int_0^\infty \big(\tanh^2 v + \tanh v \operatorname{sech} v \arcsin(\operatorname{sech} v) - 1\big)\, dv = \frac{\pi}{2} - 2,$$

thus it follows from (3.11) that

$$\kappa_\tau = \frac{1}{\pi}\left(1 - \frac{2}{\pi}\right),$$

recovering Maslova's result given in (3.6) for the Kac polynomials.

Let us mention an important consequence of Theorem 3.1.

Corollary 3.3. *Let $\xi_0, \dots, \xi_n$ be real-valued independent random variables with zero mean, unit variance, and uniform bounded $(2+\varepsilon)$ moments, for some $\varepsilon > 0$. Let $L > 0$ and consider the random hyperbolic polynomial*

$$P_{n,L}(x) = \xi_0 + \sqrt{L}\xi_1 x + \cdots + \sqrt{\frac{L(L+1)\cdots(L+n-1)}{n!}}\xi_n x^n.$$

For any $k \geq 0$, let $N_{n,k}(\mathbb{R})$ be the number of real roots of the kth derivative of $P_{n,L}$ (so $k = 0$ means $P_{n,L}$ itself). Then for $\tau = k + \frac{L-1}{2}$, we have

$$\mathbf{Var}[N_{n,k}(\mathbb{R})] = \left[2\kappa_\tau + \frac{2}{\pi}\left(1 - \frac{2}{\pi}\right) + o(1)\right] \log n.$$

Recall that when $L = 1$ the random hyperbolic polynomial $P_{n,L}$ becomes a random Kac polynomial; thus, this corollary also applies to derivatives of the Kac polynomials.

In [60], the lower bound (3.7) enabled O. Nguyen and V. Vu to show asymptotic normality for the number of real roots for such random polynomials, thanks to their following result.

Proposition 3.4 ([60]). *Assume that the polynomial P_n defined by (3.1) satisfies conditions (A1) and (3.3). Assume further that $\log n \ll \mathbf{Var}[N_n(\mathbb{R})]$. Then $N_n(\mathbb{R})$ satisfies the CLT.*

From Theorem 3.1 and this result, we obtain the following corollary.

Corollary 3.5 (Central limit theorem). *Suppose that the polynomial P_n satisfies conditions (A1) and (A2). Then $N_n(\mathbb{R})$ satisfies the CLT.*

Remark 3.6. Corollary 3.5 extends the asymptotic normality result for $N_n(\mathbb{R})$ of O. Nguyen and V. Vu in [60, Theorem 1.2 and Lemma 1.3] to new random polynomials in the generalized Kac regime. For the convenience of the reader, we include here an example to demonstrate that the asymptotic condition (3.4) (which is part of condition (A2)) is not equivalent to Nguyen-Vu's condition (3.8).

Consider the sequence $c_j = j^\tau\big(1 + \frac{(-1)^j}{\log j}\big)$, which satisfies (3.4), we will show that it does not satisfy (3.8). Observe that for $n - e^{(\log n)^{1/5}} \geq j \geq n - ne^{-(\log n)^{1/5}}$ we have

$$\frac{j^\tau}{n^\tau} = \left(1 + o\left(\frac{1}{\log n}\right)\right)^\tau = 1 + o\left(\frac{1}{\log n}\right).$$

Now if $n - j$ is odd then we claim that

$$\left|\frac{1 + \frac{(-1)^j}{\log j}}{1 + \frac{(-1)^n}{\log n}} - 1\right| \geq \frac{1}{\log n}.$$

To see this, consider two cases. First, if n is even then

$$\frac{1 - \frac{1}{\log j}}{1 + \frac{1}{\log n}} \leq 1 - \frac{1}{\log j} \leq 1 - \frac{1}{\log n},$$

and if n is odd then

$$\frac{1 + \frac{1}{\log j}}{1 - \frac{1}{\log n}} \geq 1 + \frac{1}{\log j} \geq 1 + \frac{1}{\log n},$$

so the claim is proved. It follows that, for n large,

$$\left|\frac{|c_j|}{|c_n|} - 1\right| = \left|\frac{j^\tau}{n^\tau}\frac{1 + \frac{(-1)^j}{\log j}}{1 + \frac{(-1)^n}{\log n}} - 1\right| \geq \frac{1}{2}\frac{1}{\log n} = \frac{1}{2}e^{-\log\log n},$$

so condition (3.8) is not satisfied.

We remark that this example can be modified to show that the asymptotic condition (3.4) does not imply conditions similar to (3.8) where one requires a decay estimate for $\left|\frac{|c_j|}{|c_n|} - 1\right|$ (as $n \to \infty$) that is stronger than the uniform decay rate for $|\frac{j^\tau}{n^\tau} - 1|$ (over the range of j under consideration).

Note that Proposition 3.4 strengthens Maslova's result in [55, Theorem] for the Kac polynomials. For the derivatives of the Kac polynomials and random hyperbolic polynomials for which the CLT in [60] applies, our asymptotic estimates for the variances also strengthen the CLT in [60], since they provide the details about the denominator of $\frac{N_n(\mathbb{R}) - \mathbf{E}[N_n(\mathbb{R})]}{\sqrt{\mathbf{Var}[N_n(\mathbb{R})]}}$.

Remark 3.7. The assumption $\tau > -1/2$ is used in many places in our proof. It might be interesting to consider generalized Kac polynomials in the setting when $\tau \leq -1/2$. It is curious to see if the method given here can be extended to estimate $\mathbf{Var}[N_n(\mathbb{R})]$, and this will be left for further investigation. For a recent account of these polynomials, we refer the reader to the work of Krishnapur, Lundberg, and Nguyen [46], who provided asymptotics for the expected number of real roots and answered the question on bifurcating limit cycles.

Remark 3.8. As a numerical illustration, let us consider the first derivatives of the Kac polynomials. By Theorem 3.1, we have

$$\mathbf{Var}[N_n(\mathbb{R})] = C \log n + o(\log n),$$

where

$$C = 2\kappa_1 + \frac{2}{\pi}\left(1 - \frac{2}{\pi}\right) \approx 0.575737841.$$

Figures 3.1 and 3.2 provide some numerical simulations of this result. The numerical evidence given in Figure 3.3 seems to support the conjecture that $\mathbf{Var}[N_n(\mathbb{R})] - C \log n$ converges to a limit as $n \to \infty$, and the limit may depend on the distribution of ξ_j.

We now discuss some of the main ideas of our proof. To prove Theorem 3.1, our starting point is the universality argument of O. Nguyen and V. Vu in [60], reducing the proof to the Gaussian case. However, our consideration of the Gaussian setting differs from O. Nguyen and V. Vu's argument. In [60], the authors used a novel swapping argument to compare the Gaussian version of P_n with a classical Kac polynomial (using the assumption (3.8) and via the reciprocal formulation of P_n) and deduced the lower bound (3.7) from Maslova's variance estimate for real roots inside $[-1, 1]$. This elegant approach, however, only involves the real roots outside of $[-1, 1]$ of P_n and the consequential lower bound in [60] for the variance is unfortunately not sharp. We

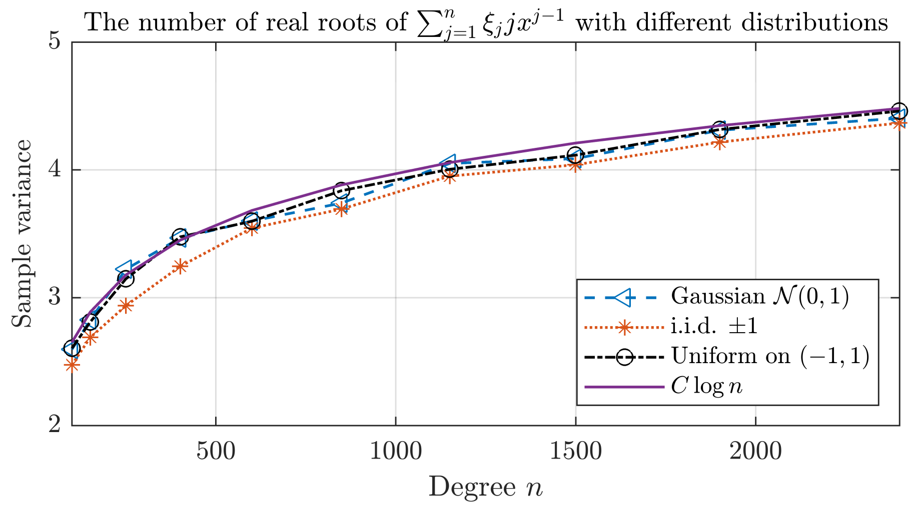

Figure 3.1: Plot of sample variances versus the degree n.

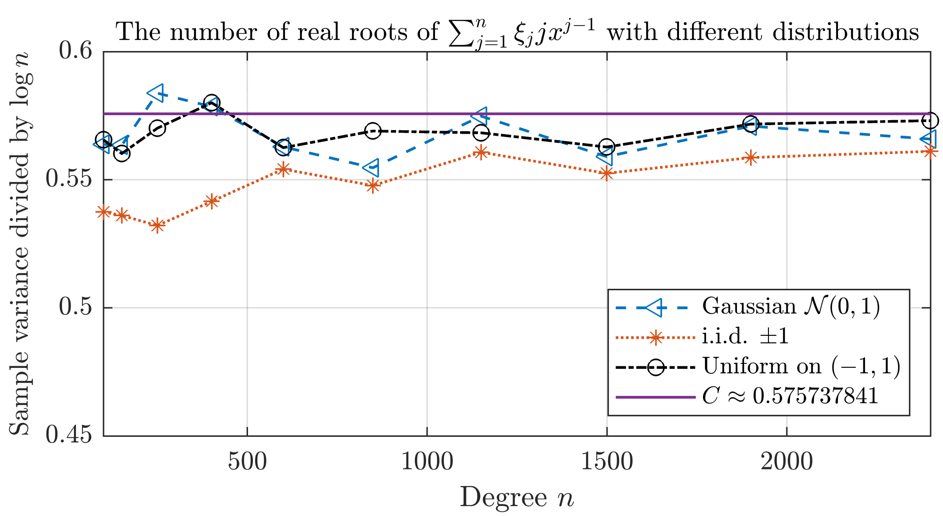

Figure 3.2: Sample variances divided by $\log(n)$ are approaching C.

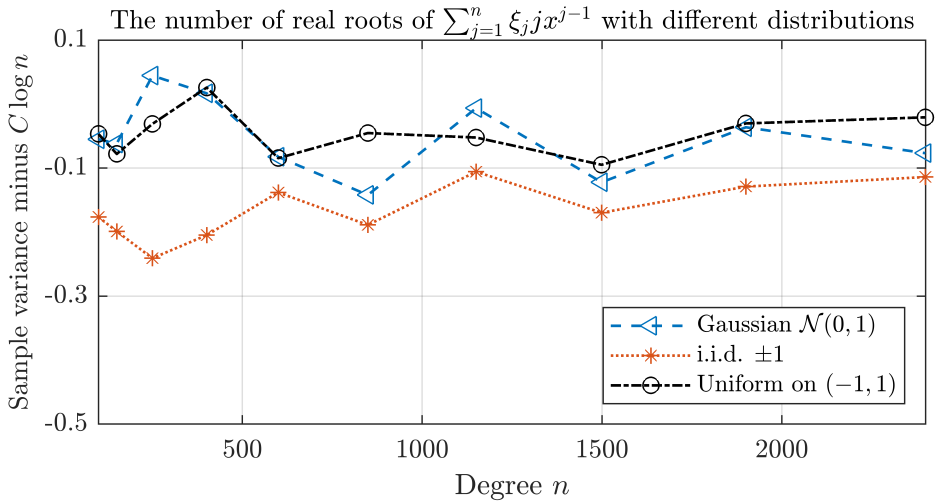

Figure 3.3: We subtract $C \log(n)$ from the sample variances, and the curves seem to converge to different values.

do not use Maslova's estimates as black boxes in our approach. The main ingredients in our evaluation of the variance for the Gaussian case are new asymptotic estimates for the two-point correlation function of the real roots of $\widetilde{P}_n$ (the Gaussian analog of P_n), see Section 3.3. These estimates reveal some underlying hyperbolic geometry inside the distribution of the real roots of generalized Kac polynomials (especially under the hypothesis (3.4)): the asymptotics depend on a certain notion of pseudo-hyperbolic distance between the real roots. One of the main difficulties in the proof is the fact that there are various instants when one has to find the leading asymptotics for an algebraic sum where the asymptotics of the summands may negate each other, especially when the mentioned pseudo-hyperbolic distance is very small. A separate consideration is often required in such situations, where a geometric property of this distance (Lemma 3.13) becomes useful in the proof.

We would also like to mention that Maslova's proof in [54] for the Kac polynomials is based on very delicate computations for the variance of real roots inside carefully chosen local intervals. It seems very difficult to extend such explicit computations to the setting of the current paper, where there are no closed-form formulas for c_j. The estimates for the correlation functions in the current paper can certainly be used to derive asymptotic estimates for the variances of the number of real roots inside local intervals, and we include some local estimates in Theorem 3.10.

Our results also demonstrate universality for $\mathbf{Var}[N_n(\mathbb{R})]$, and this is part of an active area of research to understand the universality of the distribution of real roots for random polynomials. While there have been many studies of universality for the expectation $\mathbf{E}[N_n(\mathbb{R})]$, results about universality for the variance $\mathbf{Var}[N_n(\mathbb{R})]$ are harder to come by. Besides the Kac polynomials, the class of random trigonometric polynomials is another model where universality for the variance of real roots is also well understood; see the work of Bally, Caramellino, and Poly [9] and Do, H. Nguyen, and O. Nguyen [23] for more details. Unlike (generalized) Kac polynomials, the variance for the trigonometric model is not universal and depends on the distribution of the underlying coefficients through their kurtosis [9, 23].

For Gaussian random polynomials, leading asymptotics for the variances of real roots have also been established for several other models. We refer the reader to the works of Lubinsky and Pritsker [52, 53] for random orthogonal polynomials associated with varying weights and exponential weights (essentially generalizing Azaïs, Dalmao, and Leon [5]), Gass [35] for random trigonometric polynomials with dependent coefficients (generalizing Granville and Wigman [37]), Do and Vu [29] for the Weyl polynomials, Bleher and Di [11] and Dalmao [18] for the elliptic polynomials (see also Chapter 2 for more precise asymptotics).

The rest of this chapter is organized as follows. In §3.2, we recall the universality method of [60] to reduce to the Gaussian case. Estimates for correlation functions are presented in §3.3, and the proof of the Gaussian case is presented in §3.4.

3.2 Reduction to the Gaussian case

We begin by recalling the universality arguments in [60] to reduce Theorem 3.1 to the Gaussian case, and also to localize N_n to the core region $\mathcal{I}_n$, defined as follows.

Here and subsequently, fix $d \in (0, 1/2)$ and let $d_n := e^{\log^{\frac{d}{4}} n}$, $a_n := 1/d_n$, $b_n := d_n/n$, and $I_n := [1 - a_n, 1 - b_n]$. We define

$$\mathcal{I}_n = I_n \cup -I_n \cup I_n^{-1} \cup -I_n^{-1},$$

where for any given set S, we define $-S := \{-x : x \in S\}$ and $S^{-1} := \{x^{-1} : x \in S\}$.

Let $\widetilde{P}_n(x)$ stand for the Gaussian analog of $P_n(x)$; that is,

$$\widetilde{P}_n(x) := \sum_{k=0}^{n} \widetilde{\xi}_j c_j x^j,$$

where $\widetilde{\xi}_j$'s are i.i.d. standard Gaussian random variables and c_j's satisfy assumption (A2). For $S \subset \mathbb{R}$, we denote by $\widetilde{N}_n(S)$ the number of real roots of $\widetilde{P}_n(x)$ inside S.

The following results were proved in [60, Corollary 2.2 and Proposition 2.3].

Lemma 3.9 ([60]). *There exist positive constants c and λ such that for sufficiently large n,*

$$|\,\mathbf{Var}[N_n(\mathcal{I}_n)] - \mathbf{Var}[\widetilde{N}_n(\mathcal{I}_n)]| \le c a_n^{\lambda} + c n^{-\lambda}$$

and

$$\mathbf{E}[N_n^2(\mathbb{R}\backslash\mathcal{I}_n)] \le \begin{cases} c((\log a_n)^4 + \log^2(nb_n)) & \text{if} \quad b_n \ge 1/n, \\ c(\log a_n)^4 & \text{if} \quad b_n < 1/n. \end{cases}$$

With the aid of Lemma 3.9, Theorem 3.1 will be proved once we proved the following theorem for the Gaussian case.

Theorem 3.10 (Gaussian case). *Fix $S_n \in \{-I_n, I_n\}$. As $n \to \infty$, it holds that*

$$\mathbf{Var}[\widetilde{N}_n(S_n)] = (\kappa_\tau + o(1)) \log n,$$
$$\mathbf{Var}[\widetilde{N}_n(S_n^{-1})] = \left[\frac{1}{\pi}\left(1 - \frac{2}{\pi}\right) + o(1)\right] \log n,$$

and

$$\mathbf{Var}[\widetilde{N}_n(\mathcal{I}_n)] = \left[2\kappa_\tau + \frac{2}{\pi}\left(1 - \frac{2}{\pi}\right) + o(1)\right] \log n,$$

where the implicit constants in the $o(1)$ terms depend only on the constants N_0, C_1, C_2, τ, and the rate of decay of $o_j(1)$ in condition (A2).

3.3 Estimates for the correlation functions

The proof of Theorem 3.10 relies on Lemma 1.2. Let $r(x, y)$ denote the normalized correlator of $\widetilde{P}_n$ defined as

$$r(x, y) := \frac{\mathbf{E}[\widetilde{P}_n(x)\widetilde{P}_n(y)]}{\sqrt{\mathbf{Var}[\widetilde{P}_n(x)]\,\mathbf{Var}[\widetilde{P}_n(y)]}}.$$

Set $k(x) := \sum_{j=0}^{n} c_j^2 x^j$, we see that

$$r(x, y) = \frac{k(xy)}{\sqrt{k(x^2)k(y^2)}}. \tag{3.12}$$

The estimates for the variances rely on the asymptotic estimates for the correlation functions ρ_1 and ρ_2, which will be established shortly. To this end, we first investigate the behavior of $r(x, y)$ for $x, y \in I_n \cup (-I_n)$, and thanks to (3.12) this will be done via estimates for $k(x)$ for $|x| \in I_n^2 := \{uv : u, v \in I_n\}$. In what follows, we will assume that n is sufficiently large and $S_n \in \{I_n, -I_n\}$.

By assumption (A2), we can write $c_j^2 = C_1^2 j^{2\tau}(1 + o_{j,n})$ for $N_0 \le j \le n$, where $o_{j,n} = o_j(1)$ as $j \to \infty$.

Lemma 3.11. *Let*

$$\tau_n^0 := \max_{\log(\log n) \le j \le n} \left\{ |o_{j,n}| + a_n^{\tau+1/2} + a_n \right\}.$$

Then it holds uniformly for $x \in I_n^2$ that

$$k(x) = \frac{C_1^2 \Gamma(2\tau + 1)}{(1-x)^{2\tau+1}} (1 + O(\tau_n^0))$$

and

$$k(-x) = O(\tau_n^0) k(x).$$

Here the implicit constants have the same possible dependence mentioned in Theorem 3.10.

Proof. Clearly $\tau_n^0 = o(1)$. By scaling invariant we may assume $C_1 = 1$. For $x \in I_n^2$ we have

$$\begin{aligned} k(x) &= \sum_{j=0}^{N_0-1} c_j^2 x^j + \sum_{j=N_0}^{n} j^{2\tau}(1 + o_{j,n}) x^j \\ &= \sum_{j=0}^{N_0-1} c_j^2 x^j + (1 + O(\tau_n^0)) \sum_{j=1}^{n} j^{2\tau} x^j + O\left(\sum_{j=1}^{\lfloor \log(\log n) \rfloor} j^{2\tau} x^j \right) \\ &=: \varphi(x) + (1 + O(\tau_n^0)) v_n(x) + O(t_n(x)). \end{aligned}$$

It is clear that $\varphi(x)$ is bounded uniformly on any compact subset of $\mathbb{R}$, and the bounds are independent of n. For $O(t_n(x))$, we note that

$$|t_n(x)| = O\left([\log(\log n)]^{2\tau+1} \right) = \frac{O(a_n^{\tau+1/2})}{(1-x)^{2\tau+1}} = \frac{O(\tau_n^0)}{(1-x)^{2\tau+1}}, \quad x \in I_n^2.$$

The estimate for the middle term is based on the asymptotics of $v_n(x)$. For $|x| < 1$, $v_n(x)$ converges to $v_\infty(x) = \mathrm{Li}_{-2\tau}(x)$ as $n \to \infty$, where $\mathrm{Li}_s(z)$ is the polylogarithm function defined by

$$\mathrm{Li}_s(z) = \sum_{j=1}^{\infty} \frac{z^j}{j^s}, \quad |z| < 1.$$

It is well-known that (see [70, p. 149])

$$\mathrm{Li}_s(z) = \Gamma(1-s)(-\log z)^{s-1} + \sum_{m=0}^{\infty} \zeta(s-m)\frac{(\log z)^m}{m!},$$

for $|\log z| < 2\pi$ and $s \notin \{1, 2, 3, ...\}$, where $\zeta(s)$ is the Riemann zeta function. Thus, uniformly for $x \in I_n^2$ (and one could also let $x \in I_n^4$),

$$\mathrm{Li}_{-2\tau}(x) = \Gamma(2\tau+1)(-\log x)^{-2\tau-1} + O(1) = \frac{\Gamma(2\tau+1)}{(1-x)^{2\tau+1}}(1 + O(\tau_n^0)),$$

here in the second estimate we implicitly used the fact that $1 - x = O(a_n) = O(\tau_n^0)$. Now, to estimate $\mathrm{Li}_{-2\tau}(-x)$ for $x \in I_n^2$, we use the first estimate of the last display and the duplication formula (see [47, §7.12]),

$$\mathrm{Li}_{-2\tau}(-x) + \mathrm{Li}_{-2\tau}(x) = 2^{2\tau+1}\,\mathrm{Li}_{-2\tau}(x^2),$$

and find that uniformly for $x \in I_n^2$,

$$\mathrm{Li}_{-2\tau}(-x) = O(1).$$

For $|x| \in I_n^2$, we have $1 - |x| \geq b_n = d_n/n$ and $|x|^{n+1} \leq (1-b_n)^{2(n+1)} = O(e^{-2d_n})$, so

$$\begin{aligned}
\left|\sum_{j=n+1}^{\infty} j^{2\tau} x^j\right| &= |x|^{n+1}\left|\sum_{j=0}^{\infty}(j+n+1)^{2\tau}x^j\right| \\
&= |x|^{n+1} O\left(\sum_{j=0}^{\infty} j^{2\tau}|x|^j + (n+1)^{2\tau}\sum_{j=0}^{\infty}|x|^j\right) \\
&= |x|^{n+1} O\left(\mathrm{Li}_{-2\tau}(|x|) + \frac{(n+1)^{2\tau}}{1-|x|}\right) \\
&= o(e^{-d_n}).
\end{aligned}$$

Thus, uniformly for $x \in I_n^2$,

$$v_n(x) = \mathrm{Li}_{-2\tau}(x) - \sum_{j=n+1}^{\infty} j^{2\tau}x^j = \frac{\Gamma(2\tau+1)}{(1-x)^{2\tau+1}}(1 + O(\tau_n^0))$$

and

$$v_n(-x) = O(1).$$

Therefore, uniformly for $x \in I_n^2$,

$$k(x) = \frac{C_1^2 \Gamma(2\tau+1)}{(1-x)^{2\tau+1}}(1 + O(\tau_n^0)).$$

Since

$$k(-x) = \varphi(-x) + C_1^2 \sum_{j=N_0}^{n} j^{2\tau}(-x)^j + C_1^2 \sum_{j=N_0}^{n} j^{2\tau} o_{j,n}(-x)^j,$$

it follows that

$$|k(-x)| \le O(1) + O(\tau_n^0 v_n(x)) = O(\tau_n^0)k(x), \quad x \in I_n^2,$$

here in the last estimate we used $a_n^{2\tau+1} = O(\tau_n^0)$. This completes the proof of the lemma. □

The proof can also be applied to $k_n^{(i)}$ (where τ is replaced by $\tau + \frac{1}{2}i$). Note that the $o_{j,n}$ term may change, but it is not hard to see that, for $i \ge 1$,

$$|c_j^2 j(j-1)\dots(j-i+1)| = C_1^2 j^{2\tau+i}\left(1 + O(o_{n,j}) + O\left(1/j\right)\right).$$

Since $(a_n)^c = O(\frac{1}{\log\log n})$ for any positive constant c, we then let

$$\tau_n := \max_{\log(\log n) \le j \le n} \left|\frac{|c_j|}{C_1 j^\tau} - 1\right| + \frac{1}{\log\log n} = o(1),$$

and obtain the following corollary.

Corollary 3.12. *For any $0 \le i \le 4$ it holds uniformly for $x \in I_n^2$ that*

$$k^{(i)}(x) = \frac{C_1^2 \Gamma(2\tau+i+1)}{(1-x)^{2\tau+i+1}}(1 + O(\tau_n))$$

and

$$k^{(i)}(-x) = O(\tau_n)k^{(i)}(x).$$

Here the implicit constants have the same possible dependence mentioned in Theorem 3.10.

For $(x, y) \in \mathbb{R} \times \mathbb{R}$ with $1 - xy \neq 0$, let us introduce the function

$$\alpha := \alpha(x, y) := 1 - \left(\frac{y - x}{1 - xy}\right)^2 = \frac{(1 - x^2)(1 - y^2)}{(1 - xy)^2}.$$

Clearly $0 \leq \alpha \leq 1$. It is well-known in complex analysis that

$$\varrho(z, w) = \frac{|z - w|}{|1 - \overline{w}z|},$$

defines a metric on the hyperbolic disk $\mathbb{D} := \{z \in \mathbb{C} : |z| < 1\}$ and is known as the pseudo-hyperbolic distance on $\mathbb{D}$ (see, e.g. [36]). A related notion is

$$\frac{|x - y|^2}{(1 - |x|^2)(1 - |y|^2)} \equiv \frac{1}{\alpha} - 1,$$

which can be naturally extended to $\mathbb{R}^n$ where it is an isometric invariant for the conformal ball model $\mathbb{B}^n := \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\| < 1\}$, and the classical Poincaré metric on $\mathbb{B}^n$ can also be computed from this invariant (see, e.g. [61, §4.5]).

We first prove a property of the pseudo-hyperbolic distance that will be convenient later.

Lemma 3.13. *Let $0 \leq c < \frac{1}{\sqrt{5}}$ be a fixed constant. Suppose that for some $x, y \in (-1, 1)$ that have the same sign we have $\varrho(x, y) \leq c$. Then for every z_1, z_2, z_3, z_4 between x and y it holds that*

$$\frac{1}{1 - z_1 z_2} = \frac{1 + O(\varrho(x, y))}{1 - z_3 z_4}$$

and the implicit constant may depend on c. Consequently,

$$\varrho(z_1, z_2) \leq \varrho(x, y)[1 + O(\varrho(x, y))].$$

Proof. It is clear that in the conclusion of Lemma 3.13 the second desired estimate follows immediately from the first desired estimate and the inequality $|z_1 - z_2| \leq |x - y|$. Below we prove the first estimate.

Without loss of generality, we may assume $|x| \leq |y|$. Since z_1, z_2, z_3, z_4 will be of the same sign, we have $x^2 \leq z_1 z_2, z_3 z_4 \leq y^2$, thus it suffices to show that

$$\frac{1}{1 - y^2} = \frac{1 + O(\varrho(x, y))}{1 - x^2}.$$

It follows from the given hypothesis that $\alpha = 1 - \varrho^2(x,y) \geq 1 - c^2$. Consequently,

$$\frac{|x-y|}{\sqrt{(1-x^2)(1-y^2)}} = \sqrt{\frac{1}{\alpha} - 1} \leq \frac{\varrho(x,y)}{\sqrt{1-c^2}}.$$

Therefore

$$0 \leq \frac{1}{1-y^2} - \frac{1}{1-x^2} = \frac{y^2-x^2}{(1-x^2)(1-y^2)} \leq \frac{2|x-y|}{(1-x^2)(1-y^2)} \\ \leq \frac{2\varrho(x,y)/\sqrt{1-c^2}}{\sqrt{(1-x^2)(1-y^2)}} \leq \frac{2\varrho(x,y)/\sqrt{1-c^2}}{1-y^2}.$$

Since $\frac{2\varrho(x,y)}{\sqrt{1-c^2}} \leq \frac{2c}{\sqrt{1-c^2}} < 1$, we obtain

$$\frac{1}{1-y^2} \leq \left(1 - \frac{2\varrho(x,y)}{\sqrt{1-c^2}}\right)^{-1} \frac{1}{1-x^2} = (1 + O(\varrho(x,y)))\frac{1}{1-x^2},$$

and the lemma follows. □

In the following, we will prove asymptotic estimates for r_n and its partial derivatives. Under hypothesis (A2), $\widetilde{P}_n$ is very similar to a hyperbolic random polynomial, and it is well-known that the root distributions of (complex) Gaussian hyperbolic polynomials are asymptotically invariant with respect to isometries of the hyperbolic disk $\mathbb{D}$. Thus, it seems natural to expect that the asymptotic estimates for the correlation functions of the real roots of $\widetilde{P}_n$ will involve isometric invariants (such as the pseudo-hyperbolic distance). The next few lemmas will demonstrate this heuristic.

Lemma 3.14. *It holds uniformly for* $(x,y) \in S_n \times S_n$ *that*

$$r(x,y) = \alpha^{\tau+1/2}(1 + O(\tau_n)) \tag{3.13}$$

and

$$1 - r^2(x,y) = (1 - \alpha^{2\tau+1})(1 + O(\sqrt[4]{\tau_n})). \tag{3.14}$$

Proof. Inequality (3.13) follows immediately from (3.12) and Corollary 3.12.

To prove (3.14), we consider two cases depending on whether x and y are close in the pseudo-hyperbolic distance. Let $D := \{(x,y) \in S_n \times S_n : |\frac{y-x}{1-xy}| > \sqrt[4]{\tau_n}\}$ and $D' := (S_n \times S_n) \backslash D$ (see Figure 3.4).

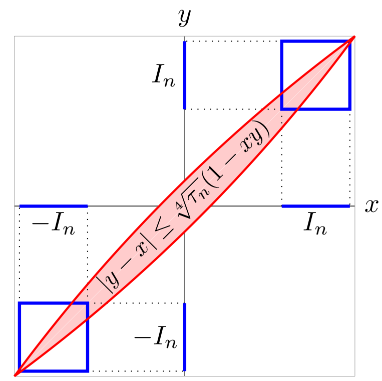

Figure 3.4: D' is the intersection of the red region with one of the blue squares.

First, if $(x, y) \in D$ then $1 - \alpha > \sqrt{\tau_n}$ and so by the mean value theorem we have

$$\begin{aligned} 1 - \alpha^{2\tau+1} &\geq (2\tau + 1)(1 - \alpha) \min(\alpha^{2\tau}, 1) \\ &\geq (2\tau + 1)\sqrt{\tau_n}\alpha^{2\tau+1}. \end{aligned}$$

Therefore we can use (3.13) to obtain

$$1 - r^2(x, y) = 1 - \alpha^{2\tau+1}(1 + O(\tau_n)) = (1 - \alpha^{2\tau+1})(1 + O(\sqrt{\tau_n})),$$

which implies (3.14).

Now, if $(x, y) \in D'$ we have $\alpha = 1 + O(\sqrt{\tau_n})$, so using the mean value theorem we obtain

$$1 - \alpha^{2\tau+1} = (2\tau + 1)(1 - \alpha)(1 + O(\sqrt{\tau_n})). \tag{3.15}$$

Using Lemma 3.13, for any z_1, z_2 between x and y we gave

$$\frac{1}{1 - z_1 z_2} = \frac{1 + O(\sqrt[4]{\tau_n})}{1 - y^2} = \frac{1 + O(\sqrt[4]{\tau_n})}{1 - x^2}, \quad \text{and} \quad \frac{|z_1 - z_2|}{|1 - z_1 z_2|} = O(\sqrt[4]{\tau_n}). \tag{3.16}$$

Fix x. We now have

$$1 - r^2(x, y) = \frac{A(y)}{k(x^2)k(y^2)}$$

where $A(y) := k(x^2)k(y^2) - k^2(xy)$ for $y \in S_n$. Using (3.16) we have

$$k(x^2)k(y^2) = \left(\frac{C_1^2 \Gamma(2\tau + 1)}{(1 - x^2)^{2\tau+1}} \right)^2 (1 + O(\sqrt[4]{\tau_n}))\,.$$

Now, $A(x) = 0$, $A'(x) = 0$, and for any z between x and y we have

$$A''(z) = 2k(x^2)k'(z^2) + 4z^2k(x^2)k''(z^2) - 2x^2k''(xz)k(xz) - 2x^2(k'(xz))^2.$$

Using Corollary 3.12 and (3.16), we obtain

$$\begin{aligned} A''(z) &= \frac{C_1^4\Gamma(2\tau+1)\Gamma(2\tau+2)}{(1-x^2)^{4\tau+4}}\Big[2 + 4(2\tau+2)(z^2-x^2) + O(\sqrt[4]{\tau_n})\Big] \\ &= \frac{2C_1^4\Gamma(2\tau+1)\Gamma(2\tau+2)}{(1-x^2)^{4\tau+4}}(1+O(\sqrt[4]{\tau_n})). \end{aligned}$$

Since $A(x) = A'(x) = 0$, by the mean value theorem there exists some z between x and y such that $A(y) = \frac{1}{2}A''(z)(y-x)^2$. Consequently,

$$\begin{aligned} 1 - r^2(x,y) = \frac{A(y)}{k(x^2)k(y^2)} &= \frac{(y-x)^2A''(z)}{2k(x^2)k(y^2)} \\ &= (2\tau+1)\left(\frac{y-x}{1-x^2}\right)^2(1+O(\sqrt[4]{\tau_n})) \\ &= (2\tau+1)\left(\frac{y-x}{1-xy}\right)^2(1+O(\sqrt[4]{\tau_n})), \end{aligned}$$

which gives (3.14) when combined with (3.15). □

Our next task is to estimate the partial derivatives of $r(x,y)$. To avoid the messy algebra in the computations, let $\ell(x,y) := \log|r(x,y)|$. Note that

$$r_{10}(x,y) = r(x,y)\ell_{10}(x,y), \quad r_{01}(x,y) = r(x,y)\ell_{01}(x,y), \tag{3.17}$$

and

$$r_{11}(x,y) = r(x,y)\left(\ell_{11}(x,y) + \ell_{10}(x,y)\ell_{01}(x,y)\right). \tag{3.18}$$

The following lemma indicates that one can take the natural log of (3.13) and then differentiate and the estimates remain essentially valid.

Lemma 3.15. *It holds uniformly for $(x,y) \in S_n \times S_n$ that*

$$\ell_{10}(x,y) = \frac{2\tau+1}{1-x^2}\left(\frac{y-x}{1-xy}\right)(1+O(\sqrt{\tau_n})) \tag{3.19}$$

and

$$\ell_{11}(x,y) = \frac{2\tau+1}{(1-xy)^2}\left(1+O(\tau_n)\right). \tag{3.20}$$

Proof. We start with the proof for (3.19). From (3.12), we have

$$\ell(x,y) = \log|k(xy)| - \frac{1}{2}\log k(x^2) - \frac{1}{2}\log k(y^2),$$

whence

$$\ell_{10}(x,y) = y\frac{k'(xy)}{k(xy)} - x\frac{k'(x^2)}{k(x^2)} = \frac{yk'(xy)k(x^2) - xk'(x^2)k(xy)}{k(xy)k(x^2)}. \tag{3.21}$$

Using Corollary 3.12, it holds uniformly for $(x,y) \in S_n \times S_n$ that

$$\frac{k'(xy)}{k(xy)} = \frac{2\tau+1}{1-xy}(1+O(\tau_n)) \quad \text{and} \quad \frac{k'(x^2)}{k(x^2)} = \frac{2\tau+1}{1-x^2}(1+O(\tau_n)). \tag{3.22}$$

We now divide the proof into two cases, similar to the proof of Lemma 3.14.

Let $R := \{(x,y) \in S_n \times S_n : |\frac{y-x}{1-xy}| \geq \sqrt{\tau_n}\}$ and $R' := (S_n \times S_n)\backslash R$.

For $(x,y) \in R$, one has

$$\frac{1-x^2}{|y-x|} = \frac{1-xy}{|y-x|} + x\frac{y-x}{|y-x|} < \frac{2}{\sqrt{\tau_n}},$$

therefore

$$\max\left\{\frac{\tau_n}{1-x^2}, \frac{\tau_n}{1-xy}\right\} < 2\sqrt{\tau_n}\frac{|y-x|}{(1-x^2)(1-xy)},$$

so (3.22) implies

$$\ell_{10}(x,y) = y\frac{k'(xy)}{k(xy)} - x\frac{k'(x^2)}{k(x^2)} = \frac{2\tau+1}{1-x^2}\left(\frac{y-x}{1-xy}\right)\left(1+O(\sqrt{\tau_n})\right).$$

We now suppose that $(x,y) \in R'$, then $\alpha = 1 + O(\tau_n)$. Using Lemma 3.13, for all z_1, z_2 between x and y we have

$$\frac{1}{1-z_1z_2} = \frac{1+O(\sqrt{\tau_n})}{1-y^2} = \frac{1+O(\sqrt{\tau_n})}{1-x^2}, \quad \text{and} \quad \frac{|z_1-z_2|}{|1-z_1z_2|} = O(\sqrt{\tau_n}). \tag{3.23}$$

Fix $x \in S_n$ and write

$$\ell_{10}(x,y) = \frac{B(y)}{k(xy)k(x^2)},$$

where $B(y) := yk'(xy)k(x^2) - xk'(x^2)k(xy)$ viewed as a function of $y \in S_n$. Then

$B(x) = 0$ and

$$B'(y) = k(x^2)[k'(xy) + xyk''(xy)] - x^2k'(x^2)k'(xy).$$

Using Corollary 3.12 and (3.23), for any z between x and y we have

$$\begin{aligned} \frac{B'(z)}{k(x^2)k(xz)} &= \frac{2\tau+1}{(1-x^2)^2}\Big[(1-x^2) + xz(2\tau+2) - x^2(2\tau+1) + O(\sqrt{\tau_n})\Big] \\ &= \frac{2\tau+1}{(1-x^2)^2}\left[1 + (2\tau+2)x(z-x) + O(\sqrt{\tau_n})\right] \\ &= \frac{2\tau+1}{(1-x^2)^2}\left[1 + O(\sqrt{\tau_n})\right]. \end{aligned}$$

Now, using the mean value theorem and (3.23), we see that, for some z between x and y,

$$B(y) = B(x) + B'(z)(y-x) = B'(z)(y-x),$$

therefore

$$\begin{aligned} \ell_{10}(x,y) = \frac{(y-x)B'(z)}{k(xy)k(x^2)} &= \frac{(2\tau+1)(y-x)}{(1-x^2)^2}\left(1 + O(\sqrt{\tau_n})\right), \\ &= \frac{2\tau+1}{1-x^2}\left(\frac{y-x}{1-xy}\right)\left(1 + O(\sqrt{\tau_n})\right), \end{aligned}$$

and (3.19) is proved.

To show (3.20), we will use Corollary 3.12 and obtain

$$\begin{aligned} \ell_{11}(x,y) &= \frac{k'(xy)}{k(xy)} + xy\frac{k''(xy)}{k(xy)} - xy\left(\frac{k'(xy)}{k(xy)}\right)^2 \\ &= \frac{2\tau+1}{1-xy} + xy\frac{(2\tau+1)(2\tau+2)}{(1-xy)^2} - xy\left(\frac{2\tau+1}{1-xy}\right)^2 + O\left(\frac{\tau_n}{(1-xy)^2}\right) \\ &= \frac{2\tau+1}{(1-xy)^2}\left(1 + O(\tau_n)\right), \end{aligned} \tag{3.24}$$

and the proof is complete. □

We obtain, as a corollary of the above estimates, an asymptotic estimate for ρ_1.

Corollary 3.16. *Uniformly for $x \in S_n \cup (-S_n)$, it holds that*

$$\rho_1(x) = \frac{1}{\pi}\frac{\sqrt{2\tau+1}}{1-x^2}\left(1 + O(\tau_n)\right). \tag{3.25}$$

We remark that a variant of (3.25) is also implicit in [27], with a stronger bound for the error term (but more stringent assumptions on c_j).

Proof. By symmetry, it suffices to consider $x \in S_n$. Using $r(x,x) = 1$ and $\ell_{10}(x,x) = \ell_{01}(y,y) = 0$, we see that

$$\rho_1(x) = \frac{1}{\pi}\sqrt{r_{11}(x,x)} = \frac{1}{\pi}\sqrt{\ell_{11}(x,x)}.$$

Thus, (3.25) follows immediately from (3.20). ☐

We now prove asymptotic estimates for ρ_2. We recall that

$$\rho_2(x,y) = \frac{1}{\pi^2}\left(\sqrt{1-\delta^2(x,y)} + \delta(x,y)\arcsin\delta(x,y)\right)\frac{\sigma(x,y)}{\sqrt{1-r^2(x,y)}}, \tag{3.26}$$

where using (3.17) and (3.18) we may rewrite σ and δ as

$$\begin{aligned}\sigma(x,y) &= \pi^2\rho_1(x)\rho_1(y)\\ &\times\sqrt{\left(1-\frac{r^2(x,y)\ell_{10}^2(x,y)}{(1-r^2(x,y))\ell_{11}(x,x)}\right)\left(1-\frac{r^2(x,y)\ell_{01}^2(x,y)}{(1-r^2(x,y))\ell_{11}(y,y)}\right)}\end{aligned} \tag{3.27}$$

and

$$\delta(x,y) = \frac{r(x,y)}{\sigma(x,y)}\left(\ell_{11}(x,y) + \frac{\ell_{10}(x,y)\ell_{01}(x,y)}{1-r^2(x,y)}\right). \tag{3.28}$$

To keep the proof from being too long, we separate the estimates into several lemmas.

Lemma 3.17. *Uniformly for $(x,y) \in S_n \times S_n$, it holds that*

$$\rho_2(x,y) = \frac{2\tau+1}{\pi^2}\frac{1+f_\tau(\alpha)}{(1-x^2)(1-y^2)}\left(1+O(\sqrt[16]{\tau_n})\right), \tag{3.29}$$

where f_τ is defined as in (3.9).

Furthermore, there is a positive constant $\alpha_0 > 0$ (independent of n but may depend on the implicit constants and rate of convergence in conditions (A1) and (A2)) such that when $\alpha \le \alpha_0$ the following holds

$$\rho_2(x,y) - \rho_1(x)\rho_1(y) = \frac{O(\alpha^{2\tau+1})}{(1-x^2)(1-y^2)}. \tag{3.30}$$

Proof. (i) We start with (3.29). For this, we first derive asymptotic estimates for σ and δ.

For $\sigma(x,y)$, we first show that

$$\begin{aligned}&(1-r^2(x,y))\ell_{11}(x,x)-(r(x,y)\ell_{10}(x,y))^2\\&=\frac{2\tau+1}{(1-x^2)^2}\left(1-\alpha^{2\tau+1}-(2\tau+1)(1-\alpha)\alpha^{2\tau+1}\right)\left(1+O(\sqrt[16]{\tau_n})\right).\end{aligned}\tag{3.31}$$

Let $g(u)=1-u^{2\tau+1}-(2\tau+1)(1-u)u^{2\tau+1}$, defined for $u\in[0,1]$. Then it can be seen that g is non-increasing on $[0,1]$ and $g(1)=0$. It follows that $g(\alpha)\geq 0$, i.e.

$$1-\alpha^{2\tau+1}\geq(2\tau+1)(1-\alpha)\alpha^{2\tau+1}.$$

We consider two cases. First, if $\sqrt[8]{\tau_n}(1-\alpha^{2\tau+1})\leq g(\alpha)$ then (3.31) follows immediately from Lemmas 3.14 and 3.15. Now, if $\sqrt[8]{\tau_n}(1-\alpha^{2\tau+1})>g(\alpha)$ we will show that x and y are close in the pseudo-hyperbolic distance, namely

$$\frac{|x-y|}{|1-xy|}=O(\sqrt[16]{\tau_n}).$$

To see this, note that on $[0,1)$ the inequality $g(u)-\sqrt[8]{\tau_n}(1-u^{2\tau+1})\leq 0$ implies

$$u^{2\tau+1}\geq(1-\sqrt[8]{\tau_n})\frac{1-u^{2\tau+1}}{(2\tau+1)(1-u)}\geq(1-\sqrt[8]{\tau_n})\min(1,u^{2\tau}),$$

thanks to the mean value theorem, which then implies

$$u\geq\min\left(1-\sqrt[8]{\tau_n},(1-\sqrt[8]{\tau_n})^{1/(2\tau+1)}\right)=1-O(\sqrt[8]{\tau_n}),$$

therefore $\alpha=1+O(\sqrt[8]{\tau_n})$ as desired.

Now, since $g(1)=g'(1)=0$ and $g''(t)=(2\tau+2)(2\tau+1)(1+O(\sqrt[8]{\tau_n}))$ for every $t\in[\alpha,1]$, using the mean value theorem we may rewrite the right-hand side (RHS) of (3.31) as

$$\text{RHS}=\frac{(2\tau+1)^2(\tau+1)}{(1-x^2)^2}(1-\alpha)^2(1+O(\sqrt[8]{\tau_n})).\tag{3.32}$$

Fix x. Rewrite the left-hand side of (3.31) as

$$(1-r^2(x,y))\ell_{11}(x,x)-(r(x,y)\ell_{10}(x,y))^2=\frac{\frac{1}{2}A(y)A''(x)-B^2(y)}{k^3(x^2)k(y^2)},$$

recalling that $A(y) = k(x^2)k(y^2) - k^2(xy)$ and $B(y) = yk'(xy)k(x^2) - xk'(x^2)k(xy)$. Let $C(y) := \frac{1}{2}A(y)A''(x) - B^2(y)$. We check at once that $C(x) = 0$, $C'(x) = 0$,

$$\begin{aligned} C''(x) &= \frac{1}{2}[A''(x)]^2 - 2[B'(x)]^2 = 0, \\ C'''(x) &= \frac{1}{2}A''(x)[A'''(x) - 6B''(x)] = 0, \end{aligned}$$

and for all z between x and y,

$$C^{(4)}(z) = \frac{1}{2}A^{(4)}(z)A''(x) - 2B(z)B^{(4)}(z) - 8B'(z)B'''(z) - 6[B''(z)]^2.$$

Now, using Lemma 3.13 if follows that for all z_1, z_2 between x and y we have

$$\frac{1}{1 - z_1 z_2} = \frac{1 + O(\sqrt[16]{\tau_n})}{1 - y^2} = \frac{1 + O(\sqrt[16]{\tau_n})}{1 - x^2}, \quad \text{and} \quad \frac{|z_1 - z_2|}{|1 - z_1 z_2|} = O(\sqrt[16]{\tau_n}). \tag{3.33}$$

Using Corollary 3.12 and (3.33) we obtain

$$k^3(x^2)k(y^2) = \left(\frac{C_1^2\Gamma(2\tau + 1)}{(1 - x^2)^{2\tau+1}}\right)^4 (1 + O(\sqrt[16]{\tau_n})),$$

and for any z between x and y, arguing as in the proof of Lemma 3.14 and Lemma 3.15 we have

$$\begin{aligned} B(z) &= \frac{C_1^4\Gamma(2\tau + 1)\Gamma(2\tau + 2)}{(1 - x^2)^{4\tau+4}}(z - x)\Big(1 + O(\sqrt[16]{\tau_n})\Big) \\ &= \frac{C_1^4\Gamma(2\tau + 1)\Gamma(2\tau + 2)}{(1 - x^2)^{4\tau+3}}O(\sqrt[16]{\tau_n}), \\ B^{(i)}(z) &= k(x^2)\Big[zk^{(i+1)}(xz)x^i + ik^{(i)}(xz)x^{i-1}\Big] - k'(x^2)k^{(i)}(xz)x^{i+1} \\ &= \frac{C_1^4\Gamma(2\tau + 1)\Gamma(2\tau + i + 1)}{(1 - x^2)^{4\tau+i+3}} \\ &\quad \times \Big((2\tau + i + 1)zx^i + ix^{i-1}(1 - x^2) - (2\tau + 1)x^{i+1} + O(\sqrt[16]{\tau_n})\Big) \\ &= \frac{C_1^4\Gamma(2\tau + 1)\Gamma(2\tau + i + 1)}{(1 - x^2)^{4\tau+i+3}}\Big(ix^{i-1} + O(\sqrt[16]{\tau_n})\Big), \quad (i \geq 1). \end{aligned}$$

Since $x = 1 + O(\tau_n)$ for $x \in S_n$, we obtain

$$B^{(i)}(z) = \frac{C_1^4\Gamma(2\tau + 1)\Gamma(2\tau + i + 1)}{(1 - x^2)^{4\tau+i+3}}\Big(i + O(\sqrt[16]{\tau_n})\Big).$$

Similarly,

$$\begin{aligned}
A''(x) &= \frac{2C_1^4\Gamma(2\tau+1)\Gamma(2\tau+2)}{(1-x^2)^{4\tau+4}}(1+O(\sqrt[16]{\tau_n})),\\
A^{(4)}(z) &= 16k^{(4)}(z^2)k(x^2)(1+O(\tau_n))\\
&\quad - 2k^{(4)}(xz)k(xz) - 8k'''(xz)k'(xz) - 6(k''(xz))^2\\
&= \frac{C_1^4\Gamma(2\tau+5)\Gamma(2\tau+1)}{(1-x^2)^{4\tau+6}}\\
&\quad \times \left[14 - \frac{8(2\tau+1)}{2\tau+4} - \frac{6(2\tau+2)(2\tau+1)}{(2\tau+4)(2\tau+3)} + O(\sqrt[16]{\tau_n})\right].
\end{aligned}$$

Consequently,

$$C^{(4)}(z) = \frac{24C_1^8(\tau+1)\Gamma^2(2\tau+2)\Gamma^2(2\tau+1)}{(1-x^2)^{8\tau+10}}(1+O(\sqrt[16]{\tau_n})).$$

Remark 3.18. We may arrive at this estimate by formally differentiating the leading asymptotics of $C(y)$ (obtained using Corollary 3.12) with respect to y, then letting $y = x$. In general, when $\varrho(x,y)$ is $o(1)$ small, it is possible that there is cancellation inside the differentiated asymptotics, in which case the expression obtained from the formal differentiation may no longer be the leading asymptotics for the underlying derivative of $C(y)$. Lemma 3.13 is useful in the examination of the differentiated asymptotics, effectively allowing us to let $y = x$ at the cost of error terms of (theoretically) smaller orders.

Now, applying the mean value theorem and (3.33), we find that

$$\frac{C(y)}{k^3(x^2)k(y^2)} = \frac{C^{(4)}(z)\frac{(y-x)^4}{4!}}{k^3(x^2)k(y^2)} = \frac{(2\tau+1)^2(\tau+1)}{(1-x^2)^2}(1-\alpha)^2(1+O(\sqrt[16]{\tau_n})),$$

which gives (3.31) when combined with (3.32).

From (3.31), it holds uniformly for $(x,y) \in S_n \times S_n$ that

$$1 - \frac{r^2(x,y)\ell_{10}^2(x,y)}{(1-r^2(x,y))\ell_{11}(x,x)} = \left(1 - \frac{(2\tau+1)(1-\alpha)\alpha^{2\tau+1}}{1-\alpha^{2\tau+1}}\right)(1+O(\sqrt[16]{\tau_n}))\,.$$

Likewise,

$$1 - \frac{r^2(x,y)\ell_{01}^2(x,y)}{(1-r^2(x,y))\ell_{11}(y,y)} = \left(1 - \frac{(2\tau+1)(1-\alpha)\alpha^{2\tau+1}}{1-\alpha^{2\tau+1}}\right)(1+O(\sqrt[16]{\tau_n}))\,.$$

Combining the above estimates with (3.27) we get

$$\frac{\sigma(x,y)}{\pi^2\rho_1(x)\rho_1(y)} = \left(1 - \frac{(2\tau+1)(1-\alpha)\alpha^{2\tau+1}}{1-\alpha^{2\tau+1}}\right)(1+O(\sqrt[16]{\tau_n}))\,. \tag{3.34}$$

For the asymptotics for δ, we now show that

$$\begin{aligned} &\ell_{11}(x,y) + \frac{\ell_{10}(x,y)\ell_{01}(x,y)}{1-r^2(x,y)} \\ &\quad = \frac{2\tau+1}{(1-x^2)(1-y^2)}\left(\alpha - \frac{(2\tau+1)(1-\alpha)}{1-\alpha^{2\tau+1}}\right)(1+O(\sqrt[16]{\tau_n})). \end{aligned} \tag{3.35}$$

The argument is similar to the proof of (3.31), so we will only mention the key steps. We may assume $\alpha = 1 + O(\sqrt[8]{\tau_n})$, otherwise (3.35) will follow from Lemma 3.14 and Lemma 3.15. With this constraint on α we have

$$\alpha - \frac{(2\tau+1)(1-\alpha)}{1-\alpha^{2\tau+1}} = -\frac{1}{2}\left(\frac{(\alpha-1)^2(2\tau+1)(2\tau+2)}{1-\alpha^{2\tau+1}}\right)(1+O(\sqrt[8]{\tau_n})).$$

Now, arguing as before and taking advantage of Lemma 3.14, Lemma 3.13, and Corollary 3.12, it suffices to show

$$\ell_{11}(x,y)(1-r^2(x,y)) + \ell_{10}(x,y)\ell_{01}(x,y) = -\frac{(2\tau+1)^2(\tau+1)(y-x)^4}{(1-x^2)^4}(1+O(\sqrt[8]{\tau_n})).$$

Fix x. We then write the left-hand side as $\frac{E(y)}{k_n^2(xy)k(x^2)k(y^2)}$, where

$$\begin{aligned} E(y) &= A(y)a(y) + B(y)b(y), \\ a(y) &:= [k'(xy) + k''(xy)xy]k(xy) - xy[k'(xy)]^2, \\ b(y) &:= xk'(xy)k(y^2) - yk'(y^2)k(xy). \end{aligned}$$

One can check that, as a function of y, $E(x) = E'(x) = E''(x) = E'''(x) = 0$. Indeed, by direct computation,

$$\begin{aligned} A(x) = A'(x) = b(x) = B(x) = 0, \\ A''(x) = -2b'(x) = 2B'(x) = 2a(x), \\ A'''(x) = -2b''(x) = 6B''(x) = 6a'(x) \end{aligned}$$

from there one can see that E vanishes up to the third derivative at $y = x$. Furthermore, using Lemma 3.13, and Corollary 3.12, we can show that, for all z between x

and y,

$$E^{(4)}(z) = -\frac{24(2\tau+1)^2(\tau+1)C_1^4(\Gamma(2\tau+1))^4}{(1-x^2)^{8\tau+8}}(1+O(\sqrt[16]{\tau_n})),$$

which implies the desired estimate, thanks to an application of the mean value theorem.

On account of (3.28), (3.34), and Lemma 3.14, we conclude that

$$\begin{aligned}\delta(x,y) &= \alpha^{\tau+1/2}\frac{\alpha(1-\alpha^{2\tau+1})-(2\tau+1)(1-\alpha)}{1-\alpha^{2\tau+1}-(2\tau+1)\alpha^{2\tau+1}(1-\alpha)}\,(1+O(\sqrt[16]{\tau_n}))\\ &= \Delta_\tau(\alpha)\,(1+O(\sqrt[16]{\tau_n}))\,,\end{aligned}$$

where Δ_τ is defined by (3.10). Let

$$\Lambda(\delta) := \sqrt{1-\delta^2}+\delta\arcsin\delta, \quad \delta\in[-1,1].$$

Since $\Lambda(\delta)\geq 1$ and $|\Lambda'(\delta)| = |\arcsin\delta| \leq \pi/2$ for all $\delta\in(-1,1)$, and $|\Delta_\tau(\alpha)|\leq 1$ for all $(x,y)\in S_n\times S_n$, it follows from the mean value theorem that

$$\begin{aligned}\Lambda(\delta(x,y)) &= \Lambda(\Delta_\tau(\alpha)) + O(|\Delta_\tau(\alpha)O(\sqrt[16]{\tau_n})|)\\ &= \Lambda(\Delta_\tau(\alpha))(1+O(\sqrt[16]{\tau_n})).\end{aligned} \tag{3.36}$$

Substituting (3.14), (3.34), and (3.36) into (3.26), we deduce (3.29) as claimed.

(ii) We now discuss the proof of (3.30). In the computation below we will assume $\alpha\leq\alpha_0$, a sufficiently small positive constant. Using Lemma 3.14 and Lemma 3.15 we have

$$\frac{r^2(x,y)\ell_{10}^2(x,y)}{(1-r^2(x,y))\ell_{11}(x,x)} = \frac{O(\alpha^{2\tau+1})O\left(\frac{1}{(1-x^2)^2}\right)}{(1+O(\alpha^{2\tau+1}))\frac{1}{(1-x^2)^2}} = O(\alpha^{2\tau+1}).$$

Similarly,

$$\frac{r^2(x,y)\ell_{01}^2(x,y)}{(1-r^2(x,y))\ell_{11}(y,y)} = O(\alpha^{2\tau+1}).$$

Using (3.27), it follows that

$$\sigma(x,y) = \pi^2\rho_1(x)\rho_1(y)(1+O(\alpha^{2\tau+1})).$$

From Lemma 3.15 we also have

$$\begin{aligned}&\ell_{11}(x,y)+\frac{\ell_{10}(x,y)\ell_{01}(x,y)}{1-r^2(x,y)}\\&\quad=O\left(\frac{1}{(1-xy)^2}\right)+O\left(\frac{(x-y)^2}{(1-xy)^2(1-x^2)(1-y^2)}\frac{1}{1+O(\alpha^{2\tau+1})}\right)\\&\quad=O\left(\frac{1}{(1-x^2)(1-y^2)}\right).\end{aligned}$$

Thus, it follows from (3.28) that

$$\delta(x,y)=O(\alpha^{\tau+\frac{1}{2}}).$$

Recall $\Lambda(\delta)=\sqrt{1-\delta^2}+\delta\arcsin\delta$ satisfies $\Lambda'(0)=0$ and $\Lambda''(\delta)=O(1)$ for δ near 0, thus $\Lambda(\delta)=1+O(\delta^2)$ near 0. Consequently, using Lemma 3.14 again, we obtain

$$\begin{aligned}\frac{\rho_2(x,y)}{\rho_1(x)\rho_1(y)}&=\Lambda(\delta(x,y))\frac{\sigma(x,y)}{\pi^2\rho_1(x)\rho_1(y)\sqrt{1-r^2(x,y)}}\\&=(1+O(\delta^2(x,y)))\frac{1+O(\alpha^{2\tau+1})}{\sqrt{1+O(\alpha^{2\tau+1})}}\\&=1+O(\alpha^{2\tau+1}).\end{aligned}$$

This completes the proof of Lemma 3.17. □

Lemma 3.19. *Uniformly for* $(x,y)\in S_n\times S_n$ *it holds that*

$$\rho_2(-x,y)-\rho_1(-x)\rho_1(y)=\frac{2\tau+1}{\pi^2}\frac{\alpha^{2\tau+1}}{(1-x^2)(1-y^2)}o(1). \tag{3.37}$$

Proof. Using (3.12) and Corollary 3.12, we have

$$r(-x,y)=\frac{k(-xy)}{\sqrt{k(x^2)k(y^2)}}=\frac{O(\tau_n)k(xy)}{\sqrt{k(x^2)k(y^2)}}=O(\tau_n)r(x,y)=o(1).$$

Now, by explicit computation (see also (3.21) and (3.24)),

$$\ell_{10}(-x,y)=y\frac{k'(-xy)}{k(-xy)}+x\frac{k'(x^2)}{k(x^2)}.$$

Therefore, using (3.20) and Corollary 3.12, we obtain

$$\begin{aligned}\left|\frac{r^2(-x,y)\ell_{10}^2(-x,y)}{\ell_{11}(-x,-x)}\right| &\le \frac{2}{|\ell_{11}(-x,-x)|}\frac{(k'(-xy))^2}{k(x^2)k(y^2)} + 2r^2(-x,y)\frac{\left(\frac{k'(x^2)}{k(x^2)}\right)^2}{|\ell_{11}(-x,-x)|}\\ &= \frac{(1-x^2)^2}{(1-xy)^2}O(\tau_n^2)r^2(x,y) + O(r^2(-x,y))\\ &= O(\tau_n^2)r^2(x,y) = O(\tau_n^2)\alpha^{2\tau+1}.\end{aligned}$$

Similarly,

$$\frac{(r(-x,y)\ell_{01}(-x,y))^2}{(1-r^2(-x,y))\ell_{11}(y,y)} = O(\tau_n^2)\alpha^{2\tau+1}.$$

Substituting these estimates into (3.27) yields

$$\sigma(-x,y) = \pi^2\rho_1(-x)\rho_1(y)\left(1 + O(\tau_n^2)\alpha^{2\tau+1}\right), \quad (x,y)\in S_n\times S_n.$$

Similarly, it holds uniformly for $(x,y)\in S_n\times S_n$ that

$$\delta(-x,y) = O(\tau_n)|r(x,y)| = O(\tau_n)\alpha^{\tau+1/2}. \tag{3.38}$$

While the proof is fairly similar, we include the details since in the proof there is an artificial singular term that appears because we use ℓ (instead of r) to compute δ via (3.28). To start, by explicit computation (see also (3.24)) we have

$$\ell_{11}(-x,y) = \frac{k'(-xy)}{k(-xy)} - xy\frac{k_n''(-xy)}{k(-xy)} + xy\left(\frac{k'(-xy)}{k(-xy)}\right)^2,$$

therefore

$$\begin{aligned}r(-x,y)\ell_{11}(-x,y) &= \frac{O\Big(|k'(-xy)| + |k_n''(-xy)|\Big)}{\sqrt{k(x^2)k(y^2)}} + r(-x,y)xy\left(\frac{k'(-xy)}{k(-xy)}\right)^2\\ &= O(\tau_n)\frac{r(x,y)}{(1-xy)^2} + r(-x,y)xy\left(\frac{k'(-xy)}{k(-xy)}\right)^2\\ &= \frac{O(\tau_n r(x,y))}{(1-x^2)(1-y^2)} + r(-x,y)xy\left(\frac{k'(-xy)}{k(-xy)}\right)^2.\end{aligned}$$

Similarly,

$$\begin{aligned}
&r(-x,y)\ell_{10}(-x,y)\ell_{01}(-x,y)\\
&\quad= r(-x,y)\left(y\frac{k'(-xy)}{k(-xy)}+x\frac{k'(x^2)}{k(x^2)}\right)\left(-x\frac{k'(-xy)}{k(-xy)}-y\frac{k'(y^2)}{k(y^2)}\right)\\
&\quad= -r(-x,y)xy\left(\frac{k'(-xy)}{k(-xy)}\right)^2\\
&\qquad+O(\tau_n r(x,y))\left(\frac{1}{1-xy}\frac{1}{1-x^2}+\frac{1}{1-xy}\frac{1}{1-y^2}+\frac{1}{1-x^2}\frac{1}{1-y^2}\right)\\
&\quad= -r(-x,y)xy\left(\frac{k'(-xy)}{k(-xy)}\right)^2+\frac{O(\tau_n r(x,y))}{(1-x^2)(1-y^2)}.
\end{aligned}$$

Thus, from (3.28) and the above estimates we have

$$\begin{aligned}
\delta(-x,y) &= \frac{r(-x,y)}{\sigma(-x,y)}\left(\ell_{11}(-x,y)+\frac{\ell_{10}(-x,y)\ell_{01}(-x,y)}{1-r^2(-x,y)}\right)\\
&= \frac{O(\tau_n r(x,y))}{\sigma(-x,y)(1-x^2)(1-y^2)}\\
&\quad+\frac{r(-x,y)}{\sigma(-x,y)}xy\left(\frac{k'(-xy)}{k(-xy)}\right)^2\left(1-\frac{1}{1-r^2(-x,y)}\right)\\
&= O(\tau_n r(x,y))+O\left(\frac{|r^3(-x,y)|}{|\sigma(-x,y)|}\left|\frac{k'(-xy)}{k(-xy)}\right|^2\right)\\
&= O(\tau_n r(x,y))+O(\tau_n^3 r_n^3(x,y))=O(\tau_n r(x,y)).
\end{aligned}$$

This completes the proof of (3.38).

Now, note that $\Lambda(\delta)=1+O(\delta^2)$ near 0, so it follows that

$$\Lambda(\delta(-x,y))=1+O(\tau_n^2)\alpha^{2\tau+1}.$$

But then

$$\rho_2(-x,y)=\frac{1}{\pi^2}\Lambda(\delta(-x,y))\frac{\sigma(-x,y)}{\sqrt{1-r^2(-x,y)}}=\rho_1(-x)\rho_1(y)\left(1+O(\tau_n^2)\alpha^{2\tau+1}\right),$$

which yields (3.37) when combined with (3.25). □

Note that the above analysis is only directly applicable to estimating the variances of the numbers of real roots inside subsets of $(-1,1)$. For $(-\infty,-1)\cup(1,\infty)$, we will pass to the reciprocal polynomial

$$Q_n(x) := \frac{x^n}{c_n}\widetilde{P}_n(1/x)$$

that converts the roots of $\widetilde{P}_n$ in $(-\infty,-1)\cup(1,\infty)$ to the roots of Q_n in $(-1,1)$. Note that

$$Q_n(x) = \sum_{j=0}^{n} \widetilde{\xi}_{n-j}\frac{c_{n-j}}{c_n}x^j$$

is also a Gaussian random polynomial. Let $k_{Q_n}(x)$ denote the corresponding variance function,

$$k_{Q_n}(x) = \sum_{j=0}^{n} \frac{c_{n-j}^2}{c_n^2}x^j = \frac{x^n k(1/x)}{c_n^2}.$$

Recall that $I_n = [1-a_n, 1-b_n]$ where $a_n = d_n^{-1} = \exp(-\log^{d/4} n)$ and $b_n = d_n/n$. As we will see, for $x \in I_n$, $k_{Q_n}(x)$ converges to $\frac{1}{1-x}$ as $n \to \infty$, which suggests that Q_n would behave like a classical Kac polynomial (this heuristics is well-known, see e.g. [27]). Let

$$e_n := \max_{0\le j\le n\sqrt{a_n}} \left|\frac{|c_{n-j}|}{|c_n|} - 1\right| + \frac{1}{\log\log n}, \tag{3.39}$$

we will show the following.

Lemma 3.20. *Let $\rho^{(1)}_{Q_n}$ and $\rho^{(2)}_{Q_n}$, respectively, denote the one-point and two-point correlation functions for the real roots of Q_n. Fix $S_n \in \{-I_n, I_n\}$. It holds uniformly for $x \in S_n$ that*

$$\rho^{(1)}_{Q_n}(x) = \frac{1}{\pi}\frac{1}{1-x^2}(1+O(e_n)).$$

Uniformly for $(x,y) \in S_n \times S_n$,

$$\rho^{(2)}_{Q_n}(x,y) = \frac{1}{\pi^2}\frac{1+f_0(\alpha)}{(1-x^2)(1-y^2)}\left(1+O(\sqrt[16]{e_n})\right),$$

where f_0 is defined as in (3.9) with $\tau = 0$.

Furthermore, there is a positive constant $\alpha_1 > 0$ (independent of n but may depend on the implicit constants and rate of convergence in conditions (A1) and (A2)) such that when $\alpha \le \alpha_1$ the following holds

$$\rho^{(2)}_{Q_n}(x,y) - \rho^{(1)}_{Q_n}(x)\rho^{(1)}_{Q_n}(y) = \frac{O(\alpha)}{(1-x^2)(1-y^2)}.$$

Also, uniformly for $(x,y) \in S_n \times S_n$,

$$\rho^{(2)}_{Q_n}(-x,y) - \rho^{(1)}_{Q_n}(-x)\rho^{(1)}_{Q_n}(y) = \frac{1}{\pi^2}\frac{\alpha}{(1-x^2)(1-y^2)}o(1).$$

Note that explicitly we can write

$$f_0(u) = (\sqrt{1-u} + \sqrt{u}\arcsin\sqrt{u})\sqrt{1-u} - 1.$$

Proof. Write

$$k_{Q_n}(x) = \sum_{0\le j\le n\sqrt{a_n}} \frac{c^2_{n-j}}{c^2_n}x^j + \sum_{n\sqrt{a_n}<j\le n} \frac{c^2_{n-j}}{c^2_n}x^j.$$

By the assumption (A2),

$$\frac{c^2_{n-j}}{c^2_n} = \begin{cases} 1 + o_n(1) & \text{if} \quad 0 \le j \le n\sqrt{a_n}, \\ O\left(n^{|2\tau|}\right) & \text{if} \quad n\sqrt{a_n} < j \le n, \end{cases}$$

where $o_n(1) \to 0$ as $n \to \infty$. For $x \in I^2_n$, it holds that

$$n^{|2\tau|+1}x^{n\sqrt{a_n}} \le n^{|2\tau|+1}(1-b_n)^{2n\sqrt{a_n}} = O(e^{-\sqrt{d_n}}) = O(a_n).$$

Consequently, by putting

$$e^0_n := \max_{0\le j\le n\sqrt{a_n}} \left|\frac{|c_{n-j}|}{|c_n|} - 1\right| + a_n,$$

we see that uniformly for $x \in I^2_n$,

$$k_{Q_n}(x) = \frac{1}{1-x}(1 + O(e^0_n))$$

and

$$k_{Q_n}(-x) = O(1) + O(e^0_n)k_{Q_n}(x) = O(e^0_n)k_{Q_n}(x).$$

For the derivatives of k_{Q_n}, we will similarly compare them with the corresponding derivatives of the analog of k_{Q_n} for the classical Kac polynomials, obtaining analogs of Lemma 3.11 (more precisely Corollary 3.12) for k_{Q_n}. Note that to account for the derivatives of k_{Q_n} one has to add an $O(1/j)$ term to e^0_n (where $j \ge \log\log n$ as in the proof of Lemma 3.11). Thus with e_n defined by (3.39) (which dominates e^0_n), we will have

$$k_{Q_n}^{(j)}(x) = \frac{j!}{(1-x)^{1+j}}(1+O(e_n)) \quad \text{and} \quad k_{Q_n}^{(j)}(-x) = O(e_n)k_{Q_n}^{(j)}(x) \tag{3.40}$$

uniformly for $x \in I_2^2$, for $j = 0, 1, ..., 4$. In other words, e_n plays the same role as τ_n in the prior treatment of $\widetilde{P}_n$ inside $(-1, 1)$.

Let $r_{Q_n}(x, y)$ denote the normalized correlator of Q_n,

$$r_{Q_n}(x,y) = \frac{\mathbf{E}[Q_n(x)Q_n(y)]}{\sqrt{\mathbf{Var}[Q_n(x)]\,\mathbf{Var}[Q_n(y)]}} = \frac{k_{Q_n}(xy)}{\sqrt{k_{Q_n}(x^2)k_{Q_n}(y^2)}}.$$

It follows from (3.40) that, uniformly for $(x, y) \in S_n \times S_n$,

$$r_{Q_n}(x,y) = \left[\frac{(1-x^2)(1-y^2)}{(1-xy)^2}\right]^{1/2}(1+O(e_n)) = \alpha^{1/2}(1+O(e_n))$$

and

$$r_{Q_n}(-x,y) = O(e_n)r_{Q_n}(x,y).$$

The rest of the proof is entirely similar to the prior treatment for $\widetilde{P}_n$, with $\tau = 0$ and $C_1 = 1$. □

Our next task is to estimate the two-point real correlation function for $\widetilde{P}_n$ between the roots inside $(-1, 1)$ and the roots outside $(-1, 1)$.

Lemma 3.21. *Fix $S_n \in \{-I_n, I_n\}$ and $T_n \in \{-I_n^{-1}, I_n^{-1}\}$. It holds uniformly for $(x, y) \in S_n \times T_n$ that*

$$r(x,y) = o(e^{-d_n/2}) \tag{3.41}$$

and

$$\rho_2(x,y) - \rho_1(x)\rho_1(y) = \rho_1(x)\rho_1(y)o(e^{-d_n/2}), \tag{3.42}$$

where $\rho_1(x)$ satisfies (3.25) and

$$\rho_1(y) = \frac{1}{\pi}\frac{1}{y^2-1}(1+O(e_n)), \quad y \in T_n. \tag{3.43}$$

Proof. It is well-known that $\rho_1(y) = y^{-2}\rho_{Q_n}^{(1)}(1/y)$, this can be seen via a change of variables (see e.g. [27]) or via explicit computations. Recall the definition of e_n from

(3.39). Applying (3.40) and proceed as in the proof of Lemma 3.15, we see that

$$\rho_1(y) = \frac{1}{\pi}\frac{1}{y^2-1}(1+O(e_n)), \quad y \in T_n,$$

which gives (3.43), and it also follows that $\ell_{11}(y,y) = \frac{1}{\pi^2(y^2-1)^2}(1+O(e_n))$.

Since $k(x^2) \gg 1$ for $x \in S_n$, to show (3.41) it suffices to show that

$$|k^{(m)}(xy)| = o\left(e^{-2d_n/3}\sqrt{k(y^2)}\right), \tag{3.44}$$

for any $m \geq 0$ bounded integer. To see this, note first that from the polynomial growth of c_j, we obtain

$$|k^{(m)}(xy)| = O(n^{2\tau+1+m}(|xy|^{n+1}+1)).$$

Using $k(y) = c_n^2 y^n k_{Q_n}(1/y)$ and the asymptotic behavior of k_{Q_n} given in (3.40), we see that

$$k(y^2) = c_n^2\frac{y^{2n+2}}{y^2-1}(1+o(1)) \gg n^{2\tau}|y|^{2(n+1)}, \quad y \in T_n.$$

For $(x,y) \in S_n \times T_n$ and any bounded constant c we have

$$n^c|x|^n = O(n^c(1-b_n)^n) = o(e^{-2d_n/3}), \quad \text{and similarly} \quad \frac{n^c}{|y|^n} = o(e^{-2d_n/3}).$$

Consequently,

$$\begin{aligned}\left|\frac{k_n^{(m)}(xy)}{\sqrt{k(y^2)}}\right| &= O\left(\frac{n^{O(1)}}{\sqrt{k(y^2)}} + \frac{n^{O(1)}|x|^{n+1}|y|^{n+1}}{\sqrt{k(y^2)}}\right)\\ &\leq O\left(\frac{n^{O(1)}}{|y|^{n+1}}\right) + O\left(n^{O(1)}|x|^{n+1}\right) = o(e^{-2d_n/3}),\end{aligned}$$

completing the proof of (3.44), and thus (3.41) is verified.

To prove (3.42), we first use the explicit computation of ℓ_{10} (see (3.21)) to estimate

$$\left|\frac{(r(x,y)\ell_{10}(x,y))^2}{\ell_{11}(x,x)}\right| \leq \left|\frac{2y^2}{\ell_{11}(x,x)}\frac{(k'(xy))^2}{k(x^2)k(y^2)}\right| + \left|2r^2(x,y)\frac{\left(\frac{k'(x^2)}{k(x^2)}\right)^2}{\ell_{11}(x,x)}\right|.$$

Using (3.44), we have

$$\frac{|k'(xy)|^2}{k(x^2)k(y^2)} = o(e^{-d_n}), \quad (x, y) \in S_n \times T_n.$$

It follows from Lemma 3.15 and Corollary 3.12 that

$$\frac{2y^2}{\ell_{11}(x, x)} = o(1) \quad \text{and} \quad \frac{\left(\frac{k'(x^2)}{k(x^2)}\right)^2}{\ell_{11}(x, x)} = O(1).$$

Together with (3.41), we obtain

$$1 - \frac{(r(x, y)\ell_{10}(x, y))^2}{(1 - r^2(x, y))\ell_{11}(x, x)} = 1 + o(e^{-d_n}), \quad (x, y) \in S_n \times T_n. \tag{3.45}$$

Similarly, we also have

$$1 - \frac{(r(x, y)\ell_{01}(x, y))^2}{(1 - r^2(x, y))\ell_{11}(y, y)} = 1 + o(e^{-d_n}), \quad (x, y) \in S_n \times T_n. \tag{3.46}$$

Substituting (3.45) and (3.46) into (3.27) yields

$$\sigma(x, y) = \pi^2 \rho_1(x)\rho_1(y)(1 + o(e^{-d_n})), \quad (x, y) \in S_n \times T_n.$$

Similarly, we can show that

$$\delta(x, y) = o(e^{-d_n/2}), \quad (x, y) \in S_n \times T_n.$$

To prove this, we proceed as in the proof of (3.38) in Lemma 3.19. Recall from (3.24) that

$$\ell_{11}(x, y) = \frac{k'(xy)}{k(xy)} + xy\frac{k''(xy)}{k(xy)} - xy\left(\frac{k'(xy)}{k(xy)}\right)^2,$$

therefore

$$\begin{aligned} r(x, y)\ell_{11}(x, y) &= \frac{O\Big(|k'(xy)| + |k''(xy)|\Big)}{\sqrt{k(x^2)k(y^2)}} - r(x, y)xy\left(\frac{k'(xy)}{k(xy)}\right)^2 \\ &= o(e^{-d_n/2}) - r(x, y)xy\left(\frac{k'(xy)}{k(xy)}\right)^2. \end{aligned}$$

On the other hand, using (3.21)

$$r(x,y)\ell_{10}(x,y)\ell_{01}(x,y) = r(x,y)\Big(y\frac{k'(xy)}{k(xy)} - x\frac{k'(x^2)}{k(x^2)}\Big)\Big(x\frac{k'(xy)}{k(xy)} - y\frac{k'(y^2)}{k(y^2)}\Big).$$

As we will see, the main term on the right-hand side is $r(x,y)xy\big(\frac{k'(xy)}{k(xy)}\big)^2$. In the estimate below, we will use the crude estimate $|k'(t^2)/k(t^2)| \le n$ for all $t \in \mathbb{R}$. Combining with (3.44), it follows that

$$r(x,y)\ell_{10}(x,y)\ell_{01}(x,y) = r(x,y)xy\left(\frac{k'(xy)}{k(xy)}\right)^2 + o(e^{-d_n/2}).$$

Since $r(x,y) = o(e^{-d_n/2})$ and $\sigma(x,y) \gg 1$ (as proved above), using (3.28) we obtain

$$\begin{aligned}\delta(x,y) &= \frac{r(x,y)}{\sigma(x,y)}xy\left(\frac{k'(xy)}{k(xy)}\right)^2\left(\frac{1}{1-r^2(x,y)} - 1\right) + \frac{o(e^{-d_n/2})}{\sigma(x,y)}\\ &= \frac{O(|r(x,y)|^3)}{\sigma(x,y)}\left(\frac{k'(xy)}{k(xy)}\right)^2 + \frac{o(e^{-d_n/2})}{\sigma(x,y)}\\ &= o(e^{-d_n/2}),\end{aligned}$$

which proves the desired claim.

Together with (3.26), we deduce (3.42) as desired. □

3.4 Proof of the Gaussian case

In this section, we give the proof of Theorem 3.10. It follows from Lemma 3.21 that the numbers of real roots of $\widetilde{P}_n$ in $S_n \in \{-I_n, I_n\}$ and in $T_n \in \{-I_n^{-1}, I_n^{-1}\}$ are asymptotically independent. Indeed, on account of (3.42), (3.43), and (3.25), we have

$$\begin{aligned}\mathbf{Cov}[\widetilde{N}_n(S_n), \widetilde{N}_n(T_n)] &= \int_{S_n} dx \int_{T_n} (\rho_2(x,y) - \rho_1(x)\rho_1(y))\, dy\\ &= o(e^{-d_n/2}) \int_{S_n} \frac{1}{1-x^2}dx \int_{T_n} \frac{1}{y^2-1}dy\\ &= o(e^{-d_n/2})O(\log^2 n) = o(1).\end{aligned}$$

This gives

$$\mathbf{Var}[\widetilde{N}_n(\mathcal{I}_n)] = \mathbf{Var}[\widetilde{N}_n(-I_n \cup I_n)] + \mathbf{Var}[\widetilde{N}_n(-I_n^{-1} \cup I_n^{-1})] + o(1).$$

Thus, the proof of Theorem 3.10 now falls naturally into Lemma 3.23 and Lemma 3.24. We first collect some basic facts about f_τ that will be convenient for the proof.

Lemma 3.22. *For $\tau > -1/2$ it holds that $\sup_{u\in[0,1]} |f_\tau(u)| < \infty$, and*

$$f_\tau(u) = \begin{cases} O(u^{2\tau+1}) & \text{as} \quad u \to 0^+, \\ -1 + O(\sqrt{1-u}) & \text{as} \quad u \to 1^-. \end{cases}$$

Furthermore, given any $\varepsilon' \in (0, 1/2)$, f_τ is real analytic on $(\varepsilon', 1-\varepsilon')$ and in particular the equation $f_\tau(u) = 0$ has at most finitely many real roots in $(0, 1)$, each of them has a finite vanishing order.

Proof. The estimates near 0 and 1 for f_τ are immediate consequences of the definition of f_τ and Taylor expansions, in fact for u near 0 one has

$$f_\tau(u) = \begin{cases} 2\tau^2 u^{2\tau+1}(1 + o(1)) & \text{if} \quad \tau \neq 0, \\ -\frac{1}{3}u^2(1 + o(1)) & \text{if} \quad \tau = 0. \end{cases}$$

In particular, by continuity, we know $f_0(u) = 0$ has no real root in $(0, 1)$, while the equation $f_\tau(u) = 0$, with $\tau \neq 0$, has at least one root inside $(0, 1)$ (see Figure 3.5), and the above endpoint estimates show that the roots do not accumulate to 0 or 1.

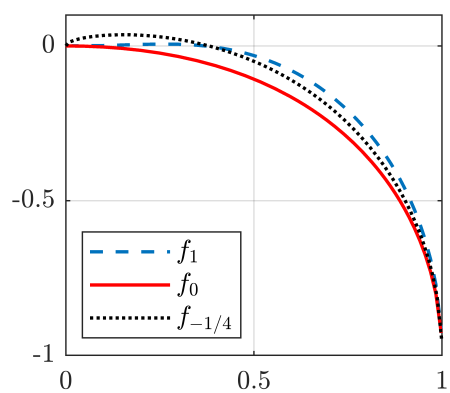

Figure 3.5: Plots of f_τ on $[0, 1]$ when $\tau = 1, 0, -\frac{1}{4}$.

Now, note that $0 \geq \Delta_\tau(u) \geq -1$ and the inequalities are strict if $u \in (0, 1)$ (see Figure 3.6). Indeed, writing $\Delta_\tau = u^{\tau+1/2}\frac{\text{Num}}{\text{Denom}}$, we will show that Num < 0 while $u^{\tau+\frac{1}{2}}\text{Num} + \text{Denom} > 0$ for $u \in (0, 1)$. First, by examination, Num is strictly increasing on $(0, 1)$, so Num $<$ Num$(1) = 0$.

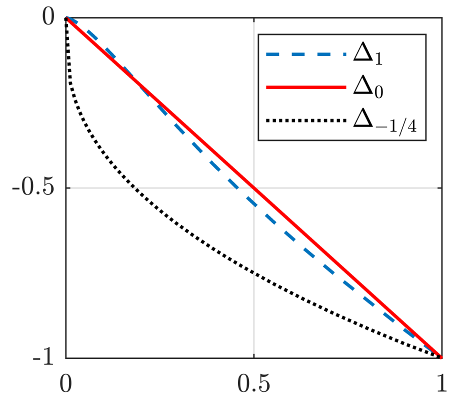

Figure 3.6: Plots of Δ_τ on $[0,1]$ when $\tau = 1, 0, -\frac{1}{4}$.

Now, let

$$\begin{aligned} h(u) := u^{\tau+\frac{1}{2}}\mathrm{Num} + \mathrm{Denom} &= (1-u^{2\tau+1})(1+u^{\tau+\frac{3}{2}}) \\ &\quad - (2\tau+1)(1-u)(u^{\tau+\frac{1}{2}} + u^{2\tau+1}). \end{aligned}$$

If $\tau \geq 0$, then one can check that $(1-u^{2\tau+1}) - (2\tau+1)(1-u)u^\tau$ is decreasing for $u \in (0,1)$, and therefore $(1-u^{2\tau+1}) \geq (2\tau+1)(1-u)u^\tau$. It follows that

$$\begin{aligned} h(u) &\geq (2\tau+1)(1-u)(u^\tau + u^{2\tau+\frac{3}{2}} - u^{\tau+\frac{1}{2}} - u^{2\tau+1}) \\ &= (2\tau+1)u^\tau(1-u)(1-\sqrt{u})(1-u^{\tau+1}) > 0. \end{aligned}$$

If $-\frac{1}{2} < \tau < 0$, then we will show that h is strictly decreasing for $u \in (0,1)$. One has

$$\begin{aligned} h'(u) = (2\tau+2)(\tau+\frac{3}{2})u^{\tau+\frac{1}{2}} + (2\tau+2)(2\tau+1)u^{2\tau+1} - (3\tau+\frac{5}{2})u^{3\tau+\frac{3}{2}} \\ - (2\tau+2)(2\tau+1)u^{2\tau} - (2\tau+1)(\tau+\frac{1}{2})u^{\tau-\frac{1}{2}} \end{aligned}$$

and

$$\begin{aligned} \frac{d}{du}(u^{-2\tau}h'(u)) &= (2\tau+2)(\tau+\frac{3}{2})(\frac{1}{2}-\tau)u^{-\tau-\frac{1}{2}} + (2\tau+2)(2\tau+1) \\ &\quad - (3\tau+\frac{5}{2})(\tau+\frac{3}{2})u^{\tau+\frac{1}{2}} + (2\tau+1)(\tau+\frac{1}{2})(\tau+\frac{1}{2})u^{-\tau-\frac{3}{2}} \\ &> u^{\tau+\frac{1}{2}}\Big[(2\tau+2)(\tau+\frac{3}{2})(\frac{1}{2}-\tau) + (2\tau+2)(2\tau+1) \\ &\quad + (2\tau+1)(\tau+\frac{1}{2})(\tau+\frac{1}{2}) - (3\tau+\frac{5}{2})(\tau+\frac{3}{2})\Big] \\ &= 0. \end{aligned}$$

It follows that $u^{-2\tau}h'(u)$ is strictly increasing on $(0,1)$ and so $u^{-2\tau}h'(u) < h'(1) = 0$ for $u \in (0,1)$. Consequently, $h' < 0$ and h is strictly decreasing on $(0,1)$, and so $h(u) > h(1) = 0$. This completes the proof of the claimed estimates for Δ_τ.

By continuity, the above consideration implies $\max_{u\in[\varepsilon',1-\varepsilon']} |\Delta_\tau(u)| < 1$. Thus, using the principal branch of log it is clear from the definition that f_τ has analytic continuation to a neighborhood of $[\varepsilon', 1-\varepsilon']$ in $\mathbb{C}$, thus the claimed properties about real zeros of f_τ follow. □

Lemma 3.23. *It holds that*

$$\mathbf{Var}[\widetilde{N}_n(S_n)] = (\kappa_\tau + o(1))\log n, \quad S_n \in \{-I_n, I_n\}, \tag{3.47}$$

and

$$\mathbf{Var}[\widetilde{N}_n(-I_n \cup I_n)] = (2\kappa_\tau + o(1))\log n. \tag{3.48}$$

Proof. (i) We start with (3.47). Let $\varepsilon' > 0$ be arbitrary, it suffices to show that for n large enough (depending on ε') the following holds

$$|\,\mathbf{Var}[\widetilde{N}_n(S_n)] - \kappa_\tau \log n| = O(\varepsilon' \log n).$$

By (3.25),

$$\mathbf{E}[\widetilde{N}_n(S_n)] = \int_{S_n} \frac{\sqrt{2\tau+1}}{\pi}\frac{1}{1-x^2}(1+o(1))\,dx = \left(\frac{\sqrt{2\tau+1}}{2\pi} + o(1)\right)\log n.$$

Now, using the change of variables $x = \tanh t$ and $y = \tanh s$, we see that

$$\frac{dxdy}{(1-x^2)(1-y^2)} = dtds \quad \text{and} \quad \alpha = \operatorname{sech}^2(s-t), \quad (t,s) \in J_n \times J_n,$$

where

$$J_n := \begin{cases} (\frac{1}{2}\log\frac{2-a_n}{a_n}, \frac{1}{2}\log\frac{2-b_n}{b_n}) & \text{if} \quad S_n = I_n, \\ (\frac{1}{2}\log\frac{b_n}{2-b_n}, \frac{1}{2}\log\frac{a_n}{2-a_n}) & \text{if} \quad S_n = -I_n, \end{cases}$$

and it is clear that

$$|J_n| = \frac{1}{2}\log n - \log^{\frac{d}{4}} n + o(1).$$

Recall the constant $\alpha_0 > 0$ from Lemma 3.17. Then there is a constant $M_0 > 0$ such that $\operatorname{sech}^2(t) \le \alpha_0$ is equivalent to $|t| \ge M_0$. It follows that

$$\begin{aligned}
&\iint_{S_n \times S_n : \alpha \le \alpha_0} |\rho_2(x,y) - \rho_1(x)\rho_1(y)|\, dxdy \\
&\quad = O\left(\iint_{J_n \times J_n : |s-t| \ge M_0} \operatorname{sech}^{4\tau+2}(s-t) dsdt\right) \\
&\quad = O\left(\int_{M_0}^{|J_n|} (|J_n| - v) \operatorname{sech}^{4\tau+2}(v) dv\right) \\
&\quad = |J_n| O\left(\int_{M_0}^{\infty} \operatorname{sech}^{4\tau+2}(v) dv\right) + O(1).
\end{aligned}$$

We now can refine α_0 (making it smaller) so that

$$\int_{M_0}^{\infty} \operatorname{sech}^{4\tau+2}(v) dv < \varepsilon',$$

and it follows that

$$\iint_{S_n \times S_n : \alpha < \alpha_0} |\rho_2(x,y) - \rho_1(x)\rho_1(y)| dxdy = O(\varepsilon' \log n). \tag{3.49}$$

We now consider $\iint_{S_n \times S_n : \alpha \ge \alpha_0}$. We separate the integration region into

$$\begin{aligned}
(I) &:= \{(x,y) \in S_n \times S_n : \alpha \ge \alpha_0, |f_\tau(\alpha)| < \sqrt[32]{\tau_n}\}, \\
(II) &:= \{(x,y) \in S_n \times S_n : \alpha \ge \alpha_0, |f_\tau(\alpha)| \ge \sqrt[32]{\tau_n}\}.
\end{aligned}$$

We will use the same change of variable $x = \tanh t$ and $y = \tanh s$, so that $\alpha = \operatorname{sech}^2(s-t)$.

For (I), using Lemma 3.22 it is clear that the set $E = \{u \in [\alpha_0, 1] : |f(u)| < \sqrt[32]{\tau_n}\}$ can be covered by a union of finitely many subintervals of $[\alpha_0, 1]$, each having length $o(1)$. Let $F = \{v : \operatorname{sech}^2(v) \in E\}$. Then it is clear that F may be covered by a union of finitely many intervals, each having length $o(1)$. (The implicit constant may depend on α_0). Thus, using boundedness of f_τ and Lemma 3.17 we have

$$\begin{aligned}
\iint_{(I)} |\rho_2(x,y) - \rho_1(x)\rho_1(y)|\, dxdy &\le O\left(\iint_{(I)} \rho_1(x)\rho_1(y) dxdy\right) \\
&\le \iint_{J_n \times J_n : |s-t| \in F} O(1) dsdt \\
&= o(|J_n|) = o(\log n).
\end{aligned}$$

For (II), we note that

$$(1+f_\tau(\alpha))(1+O(\sqrt[16]{\tau_n})) = 1+f_\tau(\alpha)+O(|f_\tau(\alpha)|\sqrt[32]{\tau_n}).$$

Thus, using Lemma 3.17 and the above estimates (for the region (I)) and boundedness of f_τ, we obtain

$$\begin{aligned}
&\iint_{(II)} [\rho_2(x,y)-\rho_1(x)\rho_1(y)]\, dxdy \\
&= \iint_{(II)} \Big(f_\tau(\alpha)+O(\sqrt[32]{\tau_n})|f_\tau(\alpha)|\Big)\rho_1(x)\rho_1(y)dxdy \\
&= o(\log n) + \iint_{S_n\times S_n:\alpha\ge\alpha_0} \Big(f_\tau(\alpha)+O(\sqrt[32]{\tau_n})|f_\tau(\alpha)|\Big)\rho_1(x)\rho_1(y)dxdy.
\end{aligned}$$

Making the change of variables $x=\tanh t$ and $y=\tanh s$ again. From Lemma 3.22 we know $\int_0^\infty |f_\tau(\operatorname{sech}^2 v)|dv$ and $\int_0^\infty v|f_\tau(\operatorname{sech}^2 v)|dv$ both converge. It follows that

$$\begin{aligned}
&\iint_{S_n\times S_n:\alpha\ge\alpha_0} f_\tau(\alpha)\rho_1(x)\rho_1(y)dxdy \\
&= \frac{2\tau+1}{\pi^2}\iint_{J_n\times J_n:|t-s|\le M_0} f_\tau(\operatorname{sech}^2(t-s))dsdt \\
&= \frac{2(2\tau+1)}{\pi^2}\int_0^{M_0}(|J_n|-v)f_\tau(\operatorname{sech}^2 v)dv \\
&= \frac{2\tau+1}{\pi^2}\left(\int_0^\infty f_\tau(\operatorname{sech}^2 v)dv + O(\varepsilon')\right)\log n + o(\log n).
\end{aligned}$$

Similar computation works for $|f_\tau(\alpha)|$, giving a contribution of order $O(\sqrt[32]{\tau_n}\log n) = o(\log n)$, and we obtain

$$\begin{aligned}
&\iint_{S_n\times S_n:\alpha\ge\alpha_0} [\rho_2(x,y)-\rho_1(x)\rho_1(y)]\, dxdy \\
&= \left(\frac{2\tau+1}{\pi^2}\int_0^\infty f_\tau(\operatorname{sech}^2 v)dv + O(\varepsilon')\right)\log n.
\end{aligned}$$

Combining these with (3.49) and (1.4), we get

$$\mathbf{Var}[\widetilde{N}_n(S_n)] = \left(\frac{2\tau+1}{\pi}\int_0^\infty f_\tau(\operatorname{sech}^2 v)dv + \frac{\sqrt{2\tau+1}}{2} + O(\varepsilon')\right)\frac{1}{\pi}\log n,$$

for any $\varepsilon'>0$, which gives (3.47), recalling the definition of κ_τ in (3.11).

(ii) We now prove (3.48). Using (3.37), we get

$$\begin{aligned}
\mathbf{Cov}[\widetilde{N}_n(-I_n), \widetilde{N}_n(I_n)] &= \iint_{I_n \times I_n} (\rho_2(-x, y) - \rho_1(-x)\rho_1(y))\, dxdy \\
&= \iint_{I_n \times I_n} \frac{2\tau+1}{\pi^2} \frac{\alpha^{2\tau+1}(x,y)}{(1-x^2)(1-y^2)} o(1) dxdy \\
&= o(1) \int_0^{|J_n|} (|J_n| - v)(\operatorname{sech}^2 v)^{2\tau+1} dv \\
&= o(1) \log n.
\end{aligned}$$

Combining with (3.47) and

$$\mathbf{Var}[\widetilde{N}_n(-I_n \cup I_n)] = \mathbf{Var}[\widetilde{N}_n(-I_n)] + \mathbf{Var}[\widetilde{N}_n(I_n)] + 2\,\mathbf{Cov}[\widetilde{N}_n(-I_n), \widetilde{N}_n(I_n)],$$

we deduce (3.48) as desired. □

Lemma 3.24. *It holds that*

$$\mathbf{Var}[\widetilde{N}_n(T_n)] = \left[\frac{1}{\pi}\left(1 - \frac{2}{\pi}\right) + o(1)\right] \log n, \quad T_n \in \{-I_n^{-1}, I_n^{-1}\},$$

and

$$\mathbf{Var}[\widetilde{N}_n(-I_n^{-1} \cup I_n^{-1})] = \left[\frac{2}{\pi}\left(1 - \frac{2}{\pi}\right) + o(1)\right] \log n.$$

Proof. The proof is entirely similar to the proof of Lemma 3.23 presented above (specialized to the case $\tau = 0$), making use of Lemma 3.20. □

Bibliography

[1] Ancona, M. and Letendre, T.: Roots of Kostlan polynomials: moments, strong law of large numbers and central limit theorem. *Ann. H. Lebesgue* **4** (2021), 1659–1703. MR4353974

[2] Ancona, M. and Letendre, T.: Zeros of smooth stationary Gaussian processes. *Electron. J. Probab.* **26** (2021), Paper No. 68, 81 pp. MR4262341

[3] Armentano, D., Azaïs, J.-M., Dalmao, F. and León, J. R.: Asymptotic variance of the number of real roots of random polynomial systems. *Proc. Amer. Math. Soc.* **146** (2018), 5437–5449. MR3866880

[4] Armentano, D., Azaïs, J.-M., Dalmao, F. and León, J. R.: Central limit theorem for the number of real roots of Kostlan Shub Smale random polynomial systems. *Amer. J. Math.* **143** (2021), 1011–1042. MR4291248

[5] Azaïs, J.-M., Dalmao, F. and León, J. R.: CLT for the zeros of classical random trigonometric polynomials. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** (2016), 804–820. MR3498010

[6] Azaïs, J.-M. and León, J. R.: CLT for crossings of random trigonometric polynomials. *Electron. J. Probab.* **18** (2013), 17 pp. MR3084654

[7] Azaïs, J.-M. and Wschebor, M.: On the roots of a random system of equations. The theorem on Shub and Smale and some extensions. *Found. Comput. Math.* **5** (2005), 125–144. MR2149413

[8] Azaïs, J.-M. and Wschebor, M.: Level Sets and Extrema of Random Processes and Fields. *John Wiley & Sons, Inc.*, Hoboken, NJ, 2009. MR2478201

[9] Bally, V., Caramellino, L. and Poly, G.: Non universality for the variance of the number of real roots of random trigonometric polynomials. *Probab. Theory Relat. Fields* **174** (2019), 887–927. MR3980307

[10] Bharucha-Reid, A. T. and Sambandham, M.: Random polynomials. Probability and Mathematical Statistics. *Academic Press, Inc.*, Orlando, FL, 1986. MR0856019

[11] Bleher, P. and Di, X.: Correlations between zeros of a random polynomial. *J. Stat. Phys.* **88** (1997), 269–305. MR1468385

[12] Bleher, P. and Di, X.: Correlations between zeros of non-Gaussian random polynomials. *Int. Math. Res. Not.* **2004** (2004), 2443–2484. MR2078308

[13] Bleher, P., Shiffman, B. and Zelditch, S.: Universality and scaling of zeros on symplectic manifolds. Random matrix models and their applications, 31–69, Math. Sci. Res. Inst. Publ. **40**, *Cambridge Univ. Press*, Cambridge, 2001. MR1842782

[14] Bloch, A. and Pólya, G.: On the roots of certain algebraic equations. *Proc. London Math. Soc. (2)* **33** (1932), 102–114. MR1576817

[15] Bogomolny, E., Bohias, O. and Lebœuf, P.: Distribution of roots of random polynomials. *Phys. Rev. Lett.* **68** (1992), 2726–2729. MR1160289

[16] Bogomolny, E., Bohias, O. and Lebœuf, P.: Quantum chaotic dynamics and random polynomials. *J. Statist. Phys.* **85** (1996), 639–679. MR1418808

[17] Comtet, L.: Advanced Combinatorics. The Art of Finite and Infinite Expansions. Revised and enlarged edition. *D. Reidel Publishing Co.*, Dordrecht, 1974. MR0460128

[18] Dalmao, F.: Asymptotic variance and CLT for the number of zeros of Kostlan Shub Smale random polynomials. *C. R. Math. Acad. Sci. Paris* **353** (2015), 1141–1145. MR3427922

[19] Das, M.: The average number of maxima of a random algebraic curve, *Proc. Cambridge Philos. Soc.* **65** (1969), 741–753. MR0239669

[20] Das, M.: Real zeros of a class of random algebraic polynomials, *J. Indian Math. Soc. (N.S.)* **36** (1972), 53–63. MR0322960

[21] Dembo, A. and Mukherjee, S.: No zero-crossings for random polynomials and the heat equation, *Ann. Probab.* **43** (2015), no. 1, 85–118. MR3298469

[22] Do, Y.: Real roots of random polynomials with coefficients of polynomial growth: a comparison principle and applications. *Electron. J. Probab.* **26** (2021), Paper No. 144, 45 pp. MR4346676

[23] Do, Y., Nguyen, H. and Nguyen, O.: Random trigonometric polynomials: universality and non-universality of the variance for the number of real roots. *Ann. Inst. Henri Poincaré Probab. Stat.* **58** (2022), 1460–1504. MR4452640

[24] Do, Y., Nguyen, H., Nguyen, O., and Pritsker, I.E.: Central limit theorem for the number of real roots of random orthogonal polynomials. *Ann. Inst. Henri Poincaré Probab. Stat.* (to appear). arXiv:2111.09015

[25] Do, Y., Nguyen, H. and Vu, V.: Real roots of random polynomials: expectation and repulsion. *Proc. Lond. Math. Soc. (3)* **111** (2015), 1231–1260. MR3447793

[26] Do, Y. and Nguyen, N. D. V.: Real roots of random polynomials with coefficients of polynomial growth: asymptotics of the variance, submitted, arXiv:2303.05478.

[27] Do, Y., Nguyen, O. and Vu, V.: Roots of random polynomials with coefficients of polynomial growth. *Ann. Probab.* **46** (2018), 2407–2494. MR3846831

[28] Do, Y., Nguyen, O. and Vu, V.: Random orthonormal polynomials: local universality and expected number of real roots. *Trans. Am. Math. Soc.* (to appear). DOI: https://doi.org/10.1090/tran/8901.

[29] Do, Y. and Vu, V.: Central limit theorems for the real zeros of Weyl polynomials. *Amer. J. Math.* **142** (2020), 1327–1369. MR4150647

[30] Edelman, A. and Kostlan, E.: How many zeros of a random polynomial are real? *Bull. Amer. Math. Soc. (N.S.)* **32** (1995), 1–37. MR1290398

[31] Erdős, P. and Offord, A. C.: On the number of real roots of a random algebraic equation. *Proc. London Math. Soc. (3)* **6** (1956), 139–160. MR0073870

[32] Farahmand, K.: Topics in Random Polynomials. Pitman Research Notes in Mathematics Series, **393**. *Longman*, Harlow, 1998. MR1679392

[33] Flasche, H. and Kabluchko, Z.: Real zeroes of random analytic functions associated with geometries of constant curvature. *J. Theor. Probab.* **33** (2020), 103–133. MR4064295

[34] Gass, L.: Cumulants asymptotics for the zeros counting measure of real Gaussian processes, preprint, arXiv:2112.08247v2.

[35] Gass, L.: Variance of the number of zeros of dependent Gaussian trigonometric polynomials, *Proc. Amer. Math. Soc.* **151** (2023), no. 5, 2225–2239. MR4556213

[36] Garnett, J. B.: *Bounded analytic functions*, Revised first edition, Graduate Texts in Mathematics, 236, Springer, New York, 2007. MR2261424

[37] Granville, A. and Wigman, I.: The distribution of the zeros of random trigonometric polynomials. *Amer. J. Math.* **133** (2011), 295–357. MR2797349

[38] Ibragimov, I. A. and Maslova, N. B.: The average number of zeros of random polynomials, *Vestnik Leningrad. Univ.* **23** (1968), no. 19, 171–172. MR0238376

[39] Ibragimov, I. A. and Maslova, N. B.: The average number of real roots of random polynomials, *Dokl. Akad. Nauk SSSR* **199** (1971), 13–16. MR0292134

[40] Ibragimov, I. A. and Maslova, N. B.: The mean number of real zeros of random polynomials. I. Coefficients with zero means, *Theory Probab. Appl.* **16** (1971), no. 2, 228–248. MR0286157

[41] Ibragimov, I. A. and Maslova, N. B.: The mean number of real zeros of random polynomials. II. Coefficients with a nonzero mean, *Theory Probab. Appl.* **16** (1971), no. 3, 485–493. MR0288824

[42] Janson, S.: Normal convergence by higher semi-invariants with applications to sums of dependent random variables and random graphs. *Ann. Probab.* **16** (1988), 305–312. MR0920273

[43] Kac, M.: On the average number of real roots of a random algebraic equation. *Bull. Amer. Math. Soc.* **49** (1943), 314–320. MR0007812

[44] Kac, M.: On the average number of real roots of a random algebraic equation. II. *Proc. London Math. Soc. (2)* **50** (1948), 390–408. MR0030713

[45] Kostlan, E.: On the distribution of roots of random polynomials. From topology to computation: Proceedings of the Smalefest (Berkeley, CA, 1990), 419–431, *Springer*, New York, 1993. MR1246137

[46] Krishnapur, M., Lundberg, E. and Nguyen, O.: The number of limit cycles bifurcating from a randomly perturbed center, preprint, arXiv:2112.05672v2.

[47] Lewin, L.: Polylogarithms and associated functions. With a foreword by A. J. Van der Poorten. *North-Holland Publishing Co.*, New York-Amsterdam, 1981. MR0618278

[48] Li, W. V. and Wei, A.: Gaussian Integrals Involving Absolute Value Functions. High dimensional probability V: the Luminy volume, 43–59, Inst. Math. Stat. (IMS) Collect. **5**, *Inst. Math. Statist.*, Beachwood, OH, 2009. MR2797939

[49] Littlewood, J. E. and Offord, A. C.: On the number of real roots of a random algebraic equation. III. *Rec. Math. [Mat. Sbornik] N.S.* **12(54)** (1943), 277–286. MR0009656

[50] Littlewood, J. E. and Offord, A. C.: On the distribution of the zeros and a-values of a random integral function. I. *J. London Math. Soc.* **20** (1945), 130–136. MR0019123

[51] Littlewood, J. E. and Offord, A. C.: On the distribution of zeros and a-values of a random integral function. II. *Ann. of Math. (2)* **49** (1948), 885–952; errata **50**, (1949), 990–991. MR0029981

[52] Lubinsky, D. S. and Pritsker, I. E.: Variance of real zeros of random orthogonal polynomials. *J. Math. Anal. Appl.* **498** (2021), Paper No. 124954, 32 pp. MR4202193

[53] Lubinsky, D. S. and Pritsker, I. E.: Variance of real zeros of random orthogonal polynomials for varying and exponential weights, *Electron. J. Probab.* **27** (2022), Paper No. 83, 32 pp. MR4444378

[54] Maslova, N. B.: The variance of the number of real roots of random polynomials, *Theory Probab. Appl.* **19** (1974), no. 1, 35–52. MR0334327

[55] Maslova, N. B.: The distribution of the number of real roots of random polynomials. *Theory Probab. Appl.* **19** (1974), no. 3, 461–473. MR0368136

[56] Nazarov, F. and Sodin, M.: Correlation functions for random complex zeroes: strong clustering and local universality. *Comm. Math. Phys.* **310** (2012), 75–98. MR2885614

[57] Nguyen, H. H., Nguyen, O. and Vu, V.: On the number of real roots of random polynomials, *Commun. Contemp. Math.* **18** (2016), no. 4, 1550052, 17 pp. MR3493213

[58] Nguyen, N. D. V.: The number of real zeros of elliptic polynomials, submitted, arXiv:2111.10875v2.

[59] Nguyen, O. and Vu, V.: Roots of random functions: a framework for local universality. *Amer. J. Math.* **144** (2022), 1–74. MR4367414

[60] Nguyen, O. and Vu, V.: Random polynomials: central limit theorems for the real roots. *Duke Math. J.* **170** (2022), 3745–3813. MR4340724

[61] Ratcliffe, J. G.: Foundations of Hyperbolic Manifolds. Second edition. Graduate Texts in Mathematics, **149**. *Springer*, New York, 2006. MR2249478

[62] Sambandham, M.: On the average number of real zeros of a class of random algebraic curves, *Pacific J. Math.* **81** (1979), no. 1, 207–215. MR0543744

[63] Sambandham, M., Gore, H. and Farahmand, K.: The average number of point [points] of inflection of random algebraic polynomials, *Stochastic Anal. Appl.* **16** (1998), no. 4, 721–731. MR1632566

[64] Sambandham, M., Thangaraj, V. and Bharucha-Reid, A. T.: On the variance of the number of real roots of random algebraic polynomials, *Stochastic Anal. Appl.* **1** (1983), no. 2, 215–238. MR0699265

[65] Schehr, G. and Majumdar, S. N.: Statistics of the number of zero crossings: from random polynomials to the diffusion equation, *Phys. Rev. Lett.* **99** (2007), 060603. DOI: 10.1103/PhysRevLett.99.060603

[66] Schehr, G. and Majumdar, S. N.: Real roots of random polynomials and zero crossing properties of diffusion equation. *J. Stat. Phys.* **132** (2008), 235–273. MR2415102

[67] Shub, M. and Smale, S.: Complexity of Bezout's theorem. II. Volumes and probabilities. *Computational algebraic geometry* (Nice, 1992), 267–285, Progr. Math. **109**, *Birkhäuser Boston*, Boston, MA, 1993. MR1230872

[68] Stevens, D. C.: The average number of real zeros of a random polynomial, *Comm. Pure Appl. Math.* **22** (1969), 457–477. MR0251003

[69] Tao, T. and Van, V.: Local universality of zeroes of random polynomials. *Int. Math. Res. Not. IMRN* **2015** (2015), 5053–5139. MR3439098

[70] Truesdell, C.: On a function which occurs in the theory of the structure of polymers, *Ann. of Math. (2)* **46** (1945), 144–157. MR0011344

[71] Wilkins, J. E., Jr.: An asymptotic expansion for the expected number of real zeros of a random polynomial, *Proc. Amer. Math. Soc.* **103** (1988), no. 4, 1249–1258. MR0955018

[72] Wschebor, M.: On the Kostlan-Shub-Smale model for random polynomial systems. Variance of the number of roots. *J. Complexity* **21** (2005), 773–789. MR2182444