Spacing Statistics for Sequences modulo 1

A Thesis

submitted to the
Indian Institute of Science Education and Research Pune
in partial fulfilment of the requirements for the
BS-MS Dual Degree Programme

by

Ameya Abhijit Tilgul

Indian Institute of Science Education and Research Pune
Dr. Homi Bhabha Road,
Pashan, Pune 411008, INDIA.

April, 2025

Supervisor: Prof. Kaneenika Sinha

Certificate

This is to certify that this dissertation entitled '*Spacing Statistics for Sequences modulo* 1' towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research Pune represents study/work carried out by Ameya Abhijit Tilgul at the Indian Institute of Science Education and Research under the supervision of Prof. Kaneenika Sinha, Associate Professor, Department of Mathematics, during the academic year 2024-2025.

K Sinha

Prof. Kaneenika Sinha

B. Baskar

Prof. Baskar Balasubramanyam

Committee:

Prof. Kaneenika Sinha

Prof. Baskar Balasubramanyam

This thesis is dedicated to my parents, on whose shoulders I stand tall.

Declaration

I hereby declare that the matter embodied in the report entitled '*Spacing Statistics for Sequences modulo* 1' are the results of the work carried out by me at the Department of Mathematics, Indian Institute of Science Education and Research Pune under the supervision of Prof. Kaneenika Sinha, and the same has not been submitted elsewhere for any other degree.

Ameya

Ameya Abhijit Tilgul

Acknowledgements

This year has been especially fruitful and smooth for me, with several things that were not in my control going my way. Hence, I would like to begin by thanking the Almighty God for always watching over me. I would also like to thank my parents for their continual support during the year-long thesis, and for always standing by me.

I would like to express my most sincere gratitude to my thesis supervisor, Prof. Kaneenika Sinha. Her invaluable guidance, support and encouragement throughout the year made this project an extremely enriching experience. She introduced me to the topic of spacing statistics in an engaging manner, sparking my curiosity and interest in the field. Her expertise and deep insights at every step have played a key role in shaping my understanding of the subject, which has finally culminated into this thesis. I deeply appreciate her patience, dedication and thoughtful advice, which have helped me overcome the challenges I faced. I am truly fortunate to have had the opportunity to work under her supervision.

I am grateful to my expert, Prof. Baskar Balasubramanyam, for his valuable advice and helpful discussions that helped me shape the direction of my approach to the subject.

I would also like to express my gratitude to Prof. Chandrasheel Bhagwat for addressing and solving all the issues I faced while completing the formalities of the thesis, and for always being available for a chat, academic or otherwise.

I consider myself extremely blessed to be a part of the lovely mathematical community at IISER Pune. I thank all the professors for nurturing my interest in Mathematics throughout my undergraduate years. I also thank my seniors and peers for reaching out whenever I required help. I would, in particular, like to thank my close friend, Prasanna Bhat, for having several intriguing mathematical discussions with me. These discussions added a new perspective to various concepts and strengthened my fundamentals in Mathematics.

I would like to thank my dearest friend and roommate, Vaishakh Kargudri, for always being there for me, for constantly uplifting my spirits with his musical taste, and for introducing me to a whole new world — both, academic and non-academic.

I am grateful to all my friends for the light-hearted dinner-table jokes and discussions, which reinvigorated me and helped start the next day with a fresh and happy mind.

Finally, I would like to thank all my local guardians and relatives, whose presence made it feel like I was never too far from home.

Abstract

This thesis aims at exploring the spacing statistics of sequences modulo one. We focus particularly on uniform distribution, correlation statistics and level-spacing (gap) distribution of sequences modulo 1.

Uniform distribution has a rich history of its own, that is said to have begun with Hermann Weyl in 1916. With the development of correlation statistics, particularly, pair correlations, people began to explore the relation between these two notions. We study this relation between uniform distribution and correlation statistics, especially when the latter is Poissonian. We aim to gain further insight by collecting several examples spread through the literature, and observing the spacing statistics they possess.

There has also been an emergence of smooth analogues of these statistics, and we explore the interplay between the classical and smooth analogues. In particular, we show that the smooth analogues imply the classical definitions in the case of Poissonian statistics. We further try to derive a criterion for the existence of Poissonian pair correlation of a sequence modulo 1.

It was shown by Pär Kurlberg and Zeév Rudnick in [KR99, Appendix A] that if a sequence has Poissonian m-level correlation function for all integers $m \geq 2$, then we can recover the Poissonian level-spacing distribution function of the sequence. We keenly study this argument, and fill in some minor details to improve its readability.

Contents

Introduction

A sequence $(x_n)_{n\in\mathbb{N}}$ is said to be *uniformly distributed* or *equidistributed* modulo one (*abbreviated* u.d. mod 1) if the relative number of the fractional parts of the elements of the sequence lying in any subinterval of $[0,1)$ is proportional to the length of that interval. The theory of uniform distribution of sequences is said to have originated from Hermann Weyl's paper "Über die Gleichverteilung von Zahlen mod. Eins" ([Wey16]) in 1916, after which, it received the attention of the mathematical community. Weyl's work built on the theory of Kronecker's approximation theorem. Leopold Kronecker had shown that the fractional parts of the integer multiples of any irrational number are dense in $[0,1)$. Weyl refined this result by proving that they are not only dense, but also uniformly distributed in the unit interval. Weyl went on to give an equivalent criterion for uniform distribution, aptly named the 'Weyl criterion'. The theory of uniform distribution modulo one is important in the sense of random number generation. The sequences that are u.d. mod 1 are essentially deterministic sequences that exhibit *pseudo*-randomness properties.

Another measure of the randomness of a sequence is its level-spacing or gap distribution function. To compute this statistic, we consider the first N elements of the sequence, order them, and check the distribution of gaps between consecutive pairs, scaled to have an average of 1. If this distribution approaches the exponential (Poisson) distribution, then we say that the sequence exhibits the Poissonian level-spacing distribution function. This is a measure of randomness because independent and uniform random variables in the unit interval exhibit the Poisson distribution. However, as will be explained later, the level-spacing distribution function of a sequence is extremely challenging to compute, leading to multiple conjectures and few concrete results.

Thus, researchers focused on weaker measures of randomness, namely, the correlation functions. In particular, the pair correlation function received a lot of attention. Pair correlation is the study of distances between elements of a sequence. We do this by considering the first N elements, but now, we consider the distance between *all* pairs, and the statistic is to count the number of pairs that are at a ‘small’ distance from each other, and to average it over N. We quantify this ‘small’ distance using symmetric intervals about 0. The sequence is said to possess Poissonian pair correlation function if this statistic approaches the length of the interval. Later, mathematicians generalized the pair correlation to higher dimensions, and called it the m-level correlation, for $m \geq 2$. In this case, instead of looking at the distance between one pair, we simultaneously observe the distance between $(m-1)$ distinct pairs as an $(m-1)$-tuple in $\mathbb{R}^{m-1}$, count the number of such tuples that lie in a ‘small’ bounded set, and average it over N. The sequence is said to possess Poissonian m-level correlation function if this average approaches the volume of the set.

At this point, we would like to observe that while uniform distribution can be viewed as a ‘global’ statistic concerning the randomness of a sequence, level-spacings and correlations are more ‘local’ in their nature. A major result in the field of spacing statistics is that if a sequence possesses Poissonian m-level correlation function for all natural numbers $m \geq 2$, then the level-spacing distribution function of the sequence can be recovered from these correlations, and it turns out to be Poissonian as well. This was done by Pär Kurlberg and Zeév Rudnick in Appendix A of the paper [KR99]. However, note that, as remarked in [LT24], the existence of all the Poissonian correlations is much stronger than the existence of Poissonian level-spacings. Another key result in this field that connects the global and local statistics is that Poissonian pair correlation is a stronger statistic than uniform distribution. This was proved independently in [ALP18; GL17; Ste17]. Thus, Poissonian pair correlation is a refinement of the uniform distribution mod 1 of a sequence.

It is instructive to note that, historically, the Poissonian pair correlations originated in the realm of quantum physics. This concept is related to the discrete energy spectra of a Hamiltonian operator of a quantum system. In particular, the Berry-Tabor Conjecture ([BT77]) states that this spectrum has the Poissonian pair correlation. Since the spectrum for several quantum systems has the form

$$(\lambda_n)_{n\in\mathbb{N}} = (x_n\,\alpha)_{n\in\mathbb{N}},$$

where α is a constant and $(x_n)_{n\in\mathbb{N}}$ is a sequence of real numbers, the mathematical community in the 1990s, notably, Rudnick, Sarnak, Zaharescu and others, began investigating the correlation properties of these sequences, which led to several interesting results in a purely mathematical set-up. We point the reader to [LS20] for more details.

The goal of this thesis is to bring together these spacing statistics, and investigate the relation between them. Our aim is to provide the reader with a concise introduction to this theory, and summarize the key results that have been derived over the past century. We also include several examples to familiarize the reader with the application of these results.

Taking inspiration from the smooth versions in the theory of uniform distribution, there has been a shift from the classical versions involving characteristic functions to the respective smooth analogues in the case of the correlations as well. This is particularly prominent in the case of Poissonian correlations. However, the justification for this shift has been vaguely written in the literature, assuming the reader to know the appropriate approximation theorems to be used. Thus, to help the reader, we try to give rigorous explanations for these transitions. This has been done using the theory of the Beurling-Selberg trigonometric polynomials.

Motivated by the recent deterministic results for Poissonian correlations of all orders ([LT21]), and for the Poissonian level-spacing distribution function ([LT24]), we go back to [KR99, Appendix A], and study the method of the recovery of the level-spacings from the correlations of all orders. We provide the rigorous argument, and fill in some details left out by the authors.

We provide a brief plan of the thesis here.
Chapter 1 deals with the notations used in the thesis, and gives a quick summary of the prerequisites, including Fourier series and trigonometric polynomials. The theory of the Beurling-Selberg trigonometric polynomials is briefly discussed in this chapter too.
Chapter 2 gives a concise introduction to the theory of uniform distribution of sequences modulo one, including the Weyl criterion and Fejér's theorem.
Chapter 3 introduces the reader to the basic definitions of pair correlation, higher-level correlation, and level-spacing statistics of sequences modulo one. We also state some landmark results relating Poissonian correlations to uniform distribution.
Chapter 4 provides a transition of the correlation statistics from the classical version to the periodized and smooth versions. We also discuss the possibility of a Weyl-type criterion for the Poissonian pair correlation function.
Chapter 5 is devoted to the discussion of recovering the (Poissonian) level-spacing distribution function from the Poissonian correlations of all orders.

This concludes the introduction of the thesis. The material has been compiled and written with the hope that the reader gets an overview of the "big-picture" of the spacing statistics, and is pointed to the required reference wherever necessary for further details. We hope that our efforts of including the omitted details in the literature would help future readers grasp the concepts with ease.

Original Contribution: This thesis is primarily expository in nature, and is meant to introduce new learners to the theory of spacing statistics. The main results in Subsections 4.1.1 and 4.1.2, that is, Theorem 4.1.2, Lemma 4.1.4, and Remark 4.1.3, including Equations 4.5 and 4.6 are original, and as far as we know, have not been mentioned in the literature before.

Chapter 1

Preliminaries

The first chapter is devoted to familiarizing the reader with the required pre-requisite concepts and recalling key definitions. We begin by providing a list of frequently used notations. We then go on to the important definitions and ideas from Fourier analysis, and an introduction to the trigonometric polynomials. We then move on to a special class of trigonometric polynomials called the Beurling-Selberg polynomials, and conclude the chapter with some properties of the same.

1.1 Notations Used

1. $\mathbb{R}$ denotes the set of real numbers, and $\mathbb{Z}$ denotes the set of integers.

2. For any $x \in \mathbb{R}$, $[x]$ denotes the greatest integer less than or equal to x.

3. For any $x \in \mathbb{R}$, $\{x\} := x - [x]$ denotes the fractional part of x. Note that $\{x\} \in [0, 1)$.

4. For any $x \in \mathbb{R}$,
$$||x|| = \begin{cases} \{x\} & \text{if } 0 \leq \{x\} < \frac{1}{2} \\ 1 - \{x\} & \text{if } \frac{1}{2} \leq \{x\} < 1 \end{cases}$$
denotes the distance from x to its closest integer. Note that $||x|| \in \left[0, \dfrac{1}{2}\right]$.

5. For any $x \in \mathbb{R}$,

$$((x)) = \begin{cases} \{x\} & \text{if } 0 \leq \{x\} < \frac{1}{2} \\ \{x\} - 1 & \text{if } \frac{1}{2} \leq \{x\} < 1 \end{cases}$$

denotes the *signed* distance from x to its closest integer. Note that $((x)) \in \left[-\frac{1}{2}, \frac{1}{2}\right)$.

6. $\chi_B(\cdot)$ denotes the characteristic function of any set B. It is defined as

$$\chi_B(x) := \begin{cases} 1 & \text{if } x \in B \\ 0 & \text{otherwise.} \end{cases}$$

7. By $\mathscr{V}(\Omega)$, we denote the set of all complex-valued functions defined on the set Ω.

8. By $C(\Omega)$, we denote the subset of $\mathscr{V}(\Omega)$ containing all continuous functions.

9. By $C^m(\Omega)$, where $m \in \mathbb{N} \cup \{\infty\}$, we denote the subset of $\mathscr{V}(\Omega)$ containing all the m-times continuously differentiable functions.

10. By $C_c(\Omega)$, we denote the subset of $\mathscr{V}(\Omega)$ containing all the continuous functions having compact support.

11. By $C_c^m(\Omega)$, where $m \in \mathbb{N} \cup \{\infty\}$, we denote the subset of $\mathscr{V}(\Omega)$ containing all the m-times continuously differentiable functions having compact support.

12. By $L^p(\Omega)$, where $p \in [1, \infty)$, we denote the subset of $\mathscr{V}(\Omega)$, containing all the functions f satisfying

$$(||f||_p)^p := \int_\Omega |f(x)|^p dx < \infty.$$

13. By $L^\infty(\Omega)$, we denote the subset of $\mathscr{V}(\Omega)$, containing all the functions f satisfying

$$||f||_\infty := \sup_{x \in \Omega} |f(x)| < \infty.$$

14. The symbol $\mathbb{T}$ denotes the set $\mathbb{R}/\mathbb{Z}$.

15. Given any $x \in \mathbb{C}$, $e(x) := e^{2\pi i x}$.

16. For $N \in \mathbb{N}$, $[N] := \{1, \dots, N\}$.

1.2 Fourier Series

This section aims at providing a brisk recollection of the elementary concepts in Fourier series. We would also like to introduce the reader to trigonometric polynomials and their Fourier series.

Definition 1.2.1 (Fourier Coefficients and Series). *Let $f \in L^1(\mathbb{T})$ be a periodic function such that*

$$\int_{\mathbb{T}} |f(x)| dx < \infty.$$

*Then, given any integer n, we define the n-th **Fourier coefficient** of f as:*

$$\hat{f}(n) := \int_{\mathbb{T}} f(x) e(-nx) dx.$$

*We define the **Fourier series** of f as*

$$\sum_{n \in \mathbb{Z}} \hat{f}(n) e(nx).$$

In what follows, we shall refer to [Mon14, Chapters 3 and 4]. We refer the reader to this reference for several properties of Fourier series and detailed proofs of the results stated below. Note that, for the rest of the section, we will assume $x \in \mathbb{T}$.

We now define an operation on functions — the convolution.

Definition 1.2.2 (Convolution of functions). *Let $f, g \in L^1(\mathbb{T})$. The convolution of f and g is defined by the function:*

$$(f * g)(x) := \int_{\mathbb{T}} f(t) g(x - t) dt.$$

Remark 1.2.1. *Changing variables: $u = x - t$, we get*

$$(f * g)(x) = \int_0^1 f(t) g(x - t) dt = \int_{x-1}^{x} g(u) f(x - u) du = \int_0^1 g(u) f(x - u) du = (g * f)(x).$$

Thus, the convolution is commutative.

Before moving ahead, let us define the 'trigonometric polynomials', and have a look at their Fourier series.

Definition 1.2.3 (Trigonometric Polynomial). *Given any non-negative integer N, we define a trigonometric polynomial,*

$$T(x) = \sum_{n=-N}^{N} t_n e(nx),$$

where $t_n \in \mathbb{C}$ are the coefficients, $-N \leq n \leq N$.
The degree of a non-zero trigonometric polynomial T is defined to be

$$\deg(T) := \max\{|n| : -N \leq n \leq N, t_n \neq 0\}.$$

Remark 1.2.2. *Let $f \in L^1(\mathbb{T})$ and T be a trigonometric polynomial of degree N. Then,*

$$(T * f)(x) = \sum_{n=-N}^{N} t_n \hat{f}(n) e(nx).$$

Remark 1.2.3 (Fourier Series of a Trigonometric Polynomial). *Note that since $e(nx)$ is continuous of period 1 for all $-N \leq n \leq N$, therefore $T(x)$ is continuous and of period 1. Hence, $T \in L^1(\mathbb{T}) \cap C(\mathbb{T})$. Hence, we can define its Fourier coefficients. In fact, for any integer k,*

$$\hat{T}(k) = \int_0^1 T(x)e(-kx)dx = \sum_{n=-N}^{N} t_n \int_0^1 e((n-k)x)\,dx = \begin{cases} t_k & \text{if } |k| \leq N, \\ 0 & \text{if } |k| > N. \end{cases}$$

Thus, the Fourier series of T is

$$\sum_{n=-\infty}^{\infty} \hat{T}(n)e(nx) = \sum_{n=-N}^{N} t_n e(nx) = T(x).$$

This is an instance where the Fourier series of a function converges to the function itself.

The above remark prompts us to ask whether the Fourier series of any function converges to the function. This is not true in general. We state two instances where the (infinite) Fourier series converges to the function itself.

Theorem 1.2.1. *Let $f \in L^1(\mathbb{T})$. Suppose that the Fourier series of f is absolutely convergent, that is,*

$$\sum_{n \in \mathbb{Z}} |\hat{f}(n)| < \infty.$$

Then, given any x at which f is continuous,

$$f(x) = \sum_{n\in\mathbb{Z}} \hat{f}(n)\, e(nx).$$

If f is continuous with period 1, then clearly, it is bounded in $\mathbb{T}$ and thus, the integral over $\mathbb{T}$ will be bounded. Thus, $f \in L^1(\mathbb{T})$. Therefore, we have the following corollary.

Corollary 1.2.2. *Let $f \in C(\mathbb{T})$. Suppose that the Fourier series of f is absolutely convergent, that is,*

$$\sum_{n\in\mathbb{Z}} \left|\hat{f}(n)\right| < \infty.$$

Then, given any $x \in \mathbb{T}$,

$$f(x) = \sum_{n\in\mathbb{Z}} \hat{f}(n)\, e(nx).$$

We now define an important partial sum related to the Fourier series of a function.

Definition 1.2.4 (Cesàro Sums of Fourier Series). *Let $f \in L^1(\mathbb{T})$, and $\sum_{n=-\infty}^{\infty} \hat{f}(n)e(nx)$ be the corresponding Fourier series. Define the **partial sums** of the Fourier series for any non-negative integer k to be*

$$s_k(x) = \sum_{n=-k}^{k} \hat{f}(n)e(nx).$$

*Define the **Cesàro partial sums** of the Fourier series for any natural number N to be*

$$\sigma_N(x) := \frac{1}{N}\sum_{k=0}^{N-1} s_k(x) = \frac{1}{N}\sum_{k=0}^{N-1}\sum_{n=-k}^{k} \hat{f}(n)e(nx).$$

Remark 1.2.4. *We can write the Cesàro partial sum for some natural number N as*

$$\sigma_N(x) = \frac{1}{N}\sum_{k=0}^{N-1}\sum_{n=-k}^{k} \hat{f}(n)e(nx) = \frac{1}{N}\sum_{n=-(N-1)}^{N-1} \hat{f}(n)e(nx)\sum_{k=|n|}^{N-1} 1 = \sum_{n=-(N-1)}^{N-1}\left(1-\frac{|n|}{N}\right)\hat{f}(n)e(nx),$$

and since the contribution from $n = N$ and $n = -N$ is zero, we can add those terms to get

$$\sigma_N(x) = \sum_{n=-N}^{N}\left(1-\frac{|n|}{N}\right)\hat{f}(n)e(nx) \tag{1.1}$$

Thus, σ_N is a trigonometric polynomial of degree at most $(N-1)$.

We would like to conclude by stating an important result relating the Cesàro partial sums to the function.

Theorem 1.2.3. *Let $f \in C(\mathbb{T})$. Then,*

$$\sigma_N(x) \to f(x) \text{ uniformly in } \mathbb{T} \text{ as } N \to \infty.$$

The above theorem essentially states that we can approximate any continuous function on $\mathbb{T}$ by trigonometric polynomials (which are the Cesáro partial sums).

This concludes the section.

1.3 The Beurling-Selberg Polynomials

This section introduces the reader to a special class of trigonometric polynomials, called the Beurling-Selberg polynomials. These polynomials were first studied by Arne Beurling in the 1930s, wherein he observed that these can be used to approximate the sawtooth function (defined below) from above and below. The importance of these functions is that these are smooth, and have some nice extremal properties. These polynomials later caught the attention of Atle Selberg, who, in the 1970s, modified them to approximate characteristic functions of intervals. The modifications are also smooth, and have similar extremal properties. Thus, Selberg found a sequence of majorants and minorants of the characteristic function. A detailed introduction and history of the origin of these functions can be found in [Vaa85].

Our goal is to look at some useful and important properties of these functions, and study the appropriate modification needed for approximating the characteristic functions of subintervals in $\mathbb{T}$.

We begin by looking at a seemingly simple but extremely important and recurring function in Fourier analysis — the 'sawtooth' function.

1.3.1 The Sawtooth Function

The importance of the sawtooth function, other than it recurring multiple times in Fourier analysis, is that it is an example of a function whose Fourier series converges to the function itself. Let us begin by defining this function.

Definition 1.3.1 (The Sawtooth Function). *The sawtooth function s, defined on all real numbers, is given by*

$$s(x) = \begin{cases} \frac{1}{2} - \{x\} & \text{if } x \notin \mathbb{Z}, \\ 0 & \text{if } x \in \mathbb{Z}. \end{cases}$$

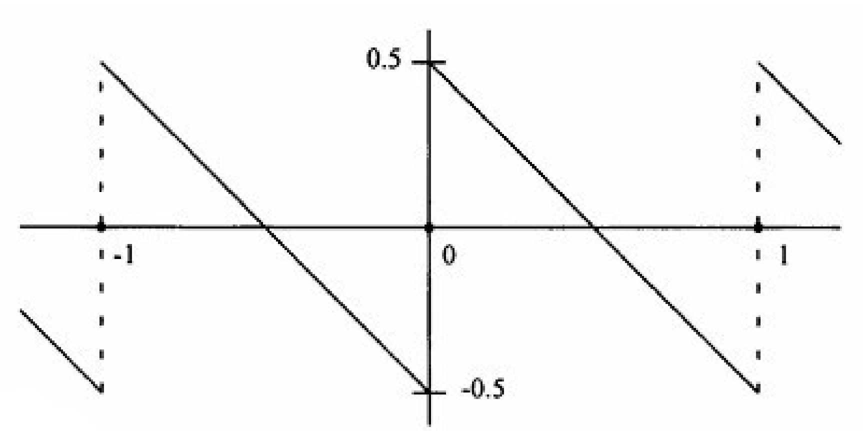

Figure 1.1: [Mon14, Chapter 3] The Sawtooth Function

Let us now derive the Fourier coefficients and the Fourier series of $s(x)$. We will also quantitatively see that the Fourier series of the sawtooth function converges to the sawtooth function, as shown in [Mon14, Theorem 3.16]. We state the theorem here. A proof of the same, with some remarks about its qualitative and quantitative aspects can be found in the reference.

Theorem 1.3.1. *Consider the sawtooth function $s(x)$ as defined above. Then, the Fourier series of $s(x)$ is given by*

$$\sum_{\substack{n \in Z \\ n \neq 0}} \frac{e(nx)}{2\pi in} = \sum_{n=1}^{\infty} \frac{\sin(2n\pi x)}{n\pi}.$$

For any natural number N, define the partial sum

$$s_N(x) = \sum_{n=1}^{N} \frac{\sin(2n\pi x)}{n\pi},$$

and now, define

$$E_N(x) := s(x) - s_N(x).$$

Then, given any x, and any natural number N,

$$|E_N(x)| \leq \min\left\{\frac{1}{2}, \frac{1}{2N\pi||x||}\right\}.$$

Corollary 1.3.2. *$s_N(x) \to s(x)$ point-wise for all real x. Thus,*

$$s(x) = \lim_{N\to\infty} s_N(x) = \sum_{n=1}^{\infty} \frac{\sin(2n\pi x)}{n\pi}.$$

Moreover, since a sequence of continuous functions (that is, the s_N's) converges to a discontinuous function (that is, s), the convergence is not uniform.

1.3.2 Definition and Properties of the Beurling-Selberg Polynomials

We begin by defining the following.

Definition 1.3.2 (The Fejér Kernel). *Given any natural number N, we define the N-th Fejér kernel as*

$$\Delta_N(x) := \sum_{n=-N}^{N} \left(1 - \frac{|n|}{N}\right) e(nx).$$

Remark 1.3.1. *$\Delta_N(x)$ is a trigonometric polynomial of degree $(N-1)$ with the coefficients $t_n = 1 - \frac{|n|}{N}$, where $-N \leq n \leq N$. Also, we have, by Remark 1.2.2,*

$$\sigma_N(x) = (\Delta_N * f)(x).$$

Definition 1.3.3 (The Beurling Trigonometric Polynomials). *Given any non-negative integer N, we define the N-th Beurling trigonometric polynomial by*

$$B_N(x) = \frac{1}{N+1} \sum_{n=0}^{N} \left(\frac{1}{2} - \frac{n}{N+1}\right) \Delta_{N+1}\left(x - \frac{n}{N+1}\right)$$
$$- \frac{1}{2(N+1)\pi} \sin(2(N+1)\pi x) + \frac{1}{2\pi} \Delta_{N+1}(x) \sin(2\pi x),$$

where $\Delta_{N+1}(x)$ is the $(N+1)$-th Fejér kernel.

We list some properties of B_N's in the following theorem, including the fact that these are trigonometric polynomials of degree N. A well-written detailed proof of the same can be found in [Mon14, Theorem 6.5].

Theorem 1.3.3 (Beurling). *Given any non-negative integer N, consider $B_N(x)$. Then,*

i. B_N is a real-valued, trigonometric polynomial of degree N.

ii. $B_N(x) \geq s(x)$ for all x, and we have

$$\widehat{B_N}(0) = \int_{\mathbb{T}} B_N(x)\, dx = \frac{1}{2(N+1)}.$$

iii. If T is a trigonometric polynomial of degree N such that

$$T(x) \geq s(x) \text{ for all } x,$$

then,

$$\int_{\mathbb{T}} T(x) dx \geq \frac{1}{2(N+1)},$$

and the equality holds if and only if

$$T(x) = B_N(x).$$

iv. Further, using (i), $B_N(-x) \geq s(-x) = -s(x)$, and hence, for any x,

$$-B_N(-x) \leq s(x) \leq B_N(x)$$

with

$$\int_{\mathbb{T}} -B_N(-x)\, dx = -\frac{1}{2(N+1)}.$$

We now look at the modification made by Selberg to approximate characteristic functions for intervals in $\mathbb{T}$. In particular, we are interested in approximating characteristic functions of the form $\chi_{[\alpha,\beta]}(\cdot)$, for some real numbers $\alpha < \beta$ such that $\beta - \alpha < 1$. Assume that $[\alpha,\beta] \subsetneq I$, where I is any unit interval.

Definition 1.3.4 (The Beurling-Selberg Trigonometric Polynomials). *Given any non-negative integer N and any real numbers $\alpha < \beta$ such that $\beta - \alpha < 1$, we define the N-th Beurling-Selberg trigonometric polynomials by*

$$S_N^+(x) = (\beta - \alpha) + B_N(\beta - x) + B_N(x - \alpha),$$

and

$$S_N^-(x) = (\beta - \alpha) - B_N(x - \beta) - B_N(\alpha - x).$$

We shall show that these trigonometric polynomials approximate the characteristic function $\chi_{[\alpha,\beta]}(\cdot)$ from above and below.

To do that, we first define the following function.

$$C_{[\alpha,\beta]}(x) := \begin{cases} 1 & \text{if } x \in (\alpha, \beta), \\ \frac{1}{2} & \text{if } x \in \{\alpha, \beta\}, \\ 0 & \text{otherwise.} \end{cases}$$

Observe that this function is almost identical to $\chi_{[\alpha,\beta]}$ in the sense that

$$C_{[\alpha,\beta]}(x) = \chi_{[\alpha,\beta]}(x) \text{ for all } x \in I \setminus \{\alpha, \beta\}$$

We can also easily see that, for any $x \in I$,

$$C_{[\alpha,\beta]}(x) = (\beta - \alpha) + s(\beta - x) + s(x - \alpha).$$

Thus, using Theorem 1.3.3(iv), we get that for any non-negative integer N and some real numbers $\alpha < \beta$ such that $\beta - \alpha < 1$,

$$S_N^-(x) \leq C_{[\alpha,\beta]}(x) \leq S_N^+(x) \text{ for all } x \in I.$$

This in fact means that

$$S_N^-(x) \leq \chi_{[\alpha,\beta]}(x) \leq S_N^+(x) \text{ for all } x \in I \setminus \{\alpha, \beta\}.$$

Now, we have

$$\lim_{x\to\beta^-} S_N^-(x) \le \lim_{x\to\beta^-} \chi_{[\alpha,\beta]}(x) \le \lim_{x\to\beta^-} S_N^+(x).$$

But note that, by Theorem 1.3.3(i), both $S_N^\pm$ are trigonometric polynomials of degree N, and hence, are continuous. Also, the characteristic function is continuous on $[\alpha,\beta]$. Thus,

$$S_N^-(\beta) \le \chi_{[\alpha,\beta]}(\beta) \le S_N^+(\beta).$$

Similarly, taking the limit $x \to \alpha^+$, we get

$$S_N^-(\alpha) \le \chi_{[\alpha,\beta]}(\alpha) \le S_N^+(\alpha),$$

and hence,

$$S_N^-(x) \le \chi_{[\alpha,\beta]}(x) \le S_N^+(x) \text{ for all } x \in I. \tag{1.2}$$

Now, consider the difference

$$S_N^+(x) - S_N^-(x) = B_N(\beta - x) + B_N(x-\alpha) + B_N(x-\beta) + B_N(\alpha - x),$$

and hence,

$$\int_{\mathbb{T}} \left(S_N^+(x) - S_N^-(x)\right) dx = \frac{4}{2(N+1)} = \frac{2}{N+1}.$$

Thus, given any $\varepsilon > 0$, there exists some large enough non-negative integer N_0 such that $\dfrac{2}{N_0+1} < \varepsilon$, and therefore, for any $N \ge N_0$,

$$\int_0^1 \left(S_N^+(x) - S_N^-(x)\right) dx < \varepsilon. \tag{1.3}$$

Finally, by Equations 1.2 and 1.3, we see that the Beurling-Selberg trigonometric polynomials approximate the characteristic functions of symmetric intervals. Since this is an important result in the context of this report, let us conclude this section by stating it in the form of a theorem.

Theorem 1.3.4 (Approximating Characteristic Functions by Beurling-Selberg Polynomials). *Given any real numbers $\alpha < \beta$ such that $\beta - \alpha < 1$, and any non-negative integer N, we have*

$$S_N^-(x) \le \chi_{[\alpha,\beta]}(x) \le S_N^+(x) \textit{ for all } x \in I.$$

Moreover, given any $\varepsilon > 0$, there exists a large enough integer N_0 such that for any $N \geq N_0$,

$$\int_{\mathbb{T}} \left(S_N^+(x) - S_N^-(x)\right) \, dx < \varepsilon.$$

This concludes the section, and the chapter.

Chapter 2

Uniform Distribution of Sequences

The theory of uniform distribution of sequences modulo one deals primarily with the distribution of the fractional parts of any real sequence in the interval $[0,1)$. The field is said to have originated with Hermann Weyl's seminal paper [Wey16] in 1916, that addressed the distribution of the fractional parts of integer multiples of an irrational number.
This chapter aims at introducing the reader to the concept of uniform distribution via the classical definition. We then transition to a functional definition that ultimately leads us to exponential sums and finally culminates in the Weyl criterion.

2.1 Uniform Distribution

Let us begin with the classical definition of uniform distribution.

Definition 2.1.1 (Uniform Distribution modulo 1). *Let $X = (x_n)_{n\in\mathbb{N}}$ be a sequence of real numbers. Define, for some subset E of $[0,1)$ and some natural number N,*

$$A(E,X,N) := \#\{1 \leq n \leq N : \{x_n\} \in E\}.$$

The sequence X is said to be ***uniformly distributed*** *(or* ***equidistributed****) modulo 1 (abbreviated u.d. mod 1) if, given any interval $[a,b) \subset [0,1)$,*

$$\lim_{N\to\infty} \frac{1}{N} A([a,b),X,N) = b-a.$$

Remark 2.1.1. *Observe that $A(E,X,N)$ can be written as*

$$A(E,X,N) := \#\{1 \le n \le N : \{x_n\} \in E\} = \sum_{n=1}^{N} \chi_E(\{x_n\}).$$

Thus, the sequence X is u.d. mod 1 if, given any $[a,b) \subset [0,1)$, we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \chi_{[a,b)}(\{x_n\}) = b - a = \int_0^1 \chi_{[a,b)}(x)dx.$$

This observation hints at a smooth version of uniform distribution. This is indeed true, as can be seen from the following theorem. We will follow the treatment from [KN74, Chapter 1].

Theorem 2.1.1. *Let $X = (x_n)_{n\in\mathbb{N}}$ be a sequence of real numbers. X is u.d. mod 1 if and only if, given any continuous function $f : \mathbb{T} \to \mathbb{C}$, we have*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(x_n) = \int_{\mathbb{T}} f(x)dx.$$

Proof.
Let X be u.d. mod 1. We will first consider the case for real-valued functions. Hence, let $f : \mathbb{T} \to \mathbb{R}$ be a continuous function. Consider $\mathbb{T}$ as the unit interval $[0,1)$, and partition it into k equal parts: $\left\{0, \frac{1}{k}, \frac{2}{k}, \ldots, \frac{k-1}{k}, 1\right\}$. Now define, for $1 \le i \le k$,

$$M_{k,i} = \max\left\{f(x) : x \in \left[\frac{i-1}{k}, \frac{i}{k}\right)\right\} \quad \text{and} \quad m_{k,i} = \min\left\{f(x) : x \in \left[\frac{i-1}{k}, \frac{i}{k}\right)\right\}$$

Define the functions

$$L_k(x) = \sum_{i=1}^{k} M_{k,i}\, \chi_{\left[\frac{i-1}{k}, \frac{i}{k}\right)}(x) \quad \text{and} \quad l_k(x) = \sum_{i=1}^{k} m_{k,i}\, \chi_{\left[\frac{i-1}{k}, \frac{i}{k}\right)}(x)$$

Clearly, for any natural number k,

$$l_k(x) \le f(x) \le L_k(x) \ \forall\, x \in [0,1).$$

Since f is continuous, for each i, the difference $(M_{k,i} - m_{k,i}) \to 0$ as $k \to \infty$. Thus, given any $\varepsilon > 0$, there exists a natural number N_0 such that

$$\int_0^1 (L_k(x) - l_k(x))\, dx < \varepsilon \text{ for any } k \geq N_0.$$

Now, for any natural numbers N and k,

$$\frac{1}{N}\sum_{n=1}^{N} l_k(x_n) \leq \frac{1}{N}\sum_{n=1}^{N} f(x_n) \leq \frac{1}{N}\sum_{n=1}^{N} L_k(x_n),$$

and hence, taking the limit as $N \to \infty$,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} l_k(x_n) \leq \liminf_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} f(x_n) \leq \limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} f(x_n) \leq \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} L_k(x_n).$$

Note that L_k and l_k are finite linear combinations of characteristic functions, and hence, by our assumption that X is u.d. mod 1, we get, for any natural number k,

$$\int_0^1 l_k(x)dx \leq \liminf_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} f(x_n) \leq \limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} f(x_n) \leq \int_0^1 L_k(x)dx$$

Thus,

$$\begin{aligned}\int_0^1 l_k(x)dx - \int_0^1 f(x)dx &\leq \liminf_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} f(x_n) - \int_0^1 f(x)dx \\ &\leq \limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} f(x_n) - \int_0^1 f(x)dx \leq \int_0^1 L_k(x)dx - \int_0^1 f(x)dx,\end{aligned}$$

which means, for any $k \geq N_0$

$$-\varepsilon \leq \liminf_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} f(x_n) - \int_0^1 f(x)dx \leq \limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} f(x_n) - \int_0^1 f(x)dx \leq \varepsilon.$$

Since this is true for any $\varepsilon > 0$, we finally have

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} f(x_n) = \int_0^1 f(x)dx.$$

We now come to the case of complex-valued, continuous functions. We know that for a continuous function $f : \mathbb{T} \to \mathbb{C}$, there exist continuous, *real*-valued functions u and v on $\mathbb{T}$ such that

$$f(x) = u(x) + i\,v(x) \text{ for all } x \in \mathbb{T}.$$

Thus, whenever X is u.d. mod 1, we can apply the above argument to u and v to get the forward implication for complex-valued, continuous functions f.

For the **converse**, assume that for any continuous function $f : \mathbb{T} \to \mathbb{C}$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(x_n) = \int_{\mathbb{T}} f(x)dx.$$

Let $[a,b)$ be any interval in $[0,1)$, and consider its characteristic function $\chi_{[a,b)}$. Let $\varepsilon > 0$ be given. Consider the functions $f_\varepsilon^{\pm} : \mathbb{T} \to \mathbb{C}$ such that

$$f_\varepsilon^{-}(x) = \begin{cases} 0 & \text{if } x \in [0,a] \cup [b,1), \\ \frac{1}{\varepsilon}(x-a) & \text{if } x \in (a, a+\varepsilon), \\ 1 & \text{if } x \in [a+\varepsilon, b-\varepsilon], \\ -\frac{1}{\varepsilon}(x-b) & \text{if } x \in (b-\varepsilon, b), \end{cases}$$

and

$$f_\varepsilon^{+}(x) = \begin{cases} 0 & \text{if } x \in [0,a-\varepsilon] \cup [b+\varepsilon,1), \\ \frac{1}{\varepsilon}(x-(a-\varepsilon)) & \text{if } x \in (a-\varepsilon, a), \\ 1 & \text{if } x \in [a, b], \\ -\frac{1}{\varepsilon}(x-(b+\varepsilon)) & \text{if } x \in (b, b+\varepsilon), \end{cases}$$

Thus, for any $\varepsilon > 0$, we have

$$f_\varepsilon^{-}(x) \leq \chi_{[a,b)}(x) \leq f_\varepsilon^{+}(x) \ \forall\ x \in \mathbb{T},$$

with

$$\int_0^1 \left(f_\varepsilon^{+}(x) - \chi_{[a,b)}(x)\right) dx = \varepsilon = \int_0^1 \left(\chi_{[a,b)}(x) - f_\varepsilon^{-}(x)\right) dx.$$

Hence, for any natural number N,

$$\frac{1}{N}\sum_{n=1}^{N} f_\varepsilon^-(x_n) \leq \frac{1}{N}\sum_{n=1}^{N} \chi_{[a,b)}(x_n) \leq \frac{1}{N}\sum_{n=1}^{N} f_\varepsilon^+(x_n).$$

Thus, taking limit as $N \to \infty$,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} f_\varepsilon^-(x_n) \leq \liminf_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} \chi_{[a,b)}(x_n) \leq \limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} \chi_{[a,b)}(x_n) \leq \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} f_\varepsilon^+(x_n),$$

which is equivalent to saying

$$\int_{\mathbb{T}} f_\varepsilon^-(x)dx \leq \liminf_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} \chi_{[a,b)}(x_n) \leq \limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} \chi_{[a,b)}(x_n) \leq \int_{\mathbb{T}} f_\varepsilon^+(x)dx.$$

Now,

$$\begin{aligned}\int_{\mathbb{T}} f_\varepsilon^-(x)dx - \int_{\mathbb{T}}\chi_{[a,b)}(x)dx &\leq \liminf_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} \chi_{[a,b)}(x_n) - \int_{\mathbb{T}}\chi_{[a,b)}(x)dx \\ &\leq \limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} \chi_{[a,b)}(x_n) - \int_{\mathbb{T}}\chi_{[a,b)}(x)dx \leq \int_{\mathbb{T}} f_\varepsilon^+(x)dx - \int_{\mathbb{T}}\chi_{[a,b)}(x)dx,\end{aligned}$$

which means that,

$$-\varepsilon \leq \liminf_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} \chi_{[a,b)}(x_n) - \int_{\mathbb{T}}\chi_{[a,b)}(x)dx \leq \limsup_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} \chi_{[a,b)}(x_n) - \int_{\mathbb{T}}\chi_{[a,b)}(x)dx \leq \varepsilon.$$

Thus, as $\varepsilon > 0$,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} \chi_{[a,b]}(x_n) = \int_{\mathbb{T}}\chi_{[a,b]}(x)dx,$$

which proves that X is u.d. mod 1. $\square$

This leads us to the celebrated **Weyl Criterion**.

2.2 The Weyl Criterion

This section introduces the reader to the Weyl criterion, and an important consequence of the same, namely, Fejér's theorem. These theorems provide a lot of examples of sequences that are u.d. mod

1. These results have been taken from [KN74, Chapter 1, Section 2]. We begin with the Weyl criterion.

Theorem 2.2.1 (The Weyl Criterion). *The sequence of real numbers $X = (x_n)_{n\in\mathbb{N}}$ is u.d. mod 1 if and only if, for any non-zero integer m, we have*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} e(mx_n) = 0.$$

Proof.
Suppose that the sequence X is u.d. mod 1. Then, by Theorem 2.1.1, given any continuous function $f : \mathbb{T}\to\mathbb{C}$,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} f(x_n) = \int_0^1 f(x)dx.$$

In particular, for $f_m(x) = e^{2\pi imx}$, where $m\in\mathbb{Z}$ (which is continuous and has period 1), we have

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} e(mx_n) = \int_0^1 e^{2\pi imx} = \begin{cases} 1 & \text{if } m = 0,\\ 0 & \text{if } m\in\mathbb{Z}\setminus\{0\}.\end{cases}$$

Conversely, suppose that given any non-zero integer m,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} e(mx_n) = 0.$$

Consider a continuous function $f : \mathbb{T}\to\mathbb{C}$. We know by Theorem 1.2.3 that f can be approximated by trigonometric polynomials. That is, given any $\varepsilon > 0$, there exists some trigonometric polynomial $T(x)$ such that

$$\sup_{x\in\mathbb{T}}|f(x)-T(x)| < \varepsilon. \tag{2.1}$$

Now, consider the following.

$$\begin{aligned}&\left|\int_0^1 f(x)dx - \frac{1}{N}\sum_{n=1}^{N} f(x_n)\right|\\ \leq&\left|\int_0^1 f(x)dx - \int_0^1 T(x)dx\right| + \left|\int_0^1 T(x)dx - \frac{1}{N}\sum_{n=1}^{N} T(x_n)\right| + \left|\frac{1}{N}\sum_{n=1}^{N} T(x_n) - \frac{1}{N}\sum_{n=1}^{N} f(x_n)\right|\end{aligned}$$

Now, as $N\to\infty$, the first and third term are less than ε due to Equation 2.1, and the second term goes to zero by our assumption. This proves the theorem. $\square$

Remark 2.2.1. *This criterion tells us that functions of the form $f_m(x) = e^{2\pi imx}$, $m \in \mathbb{Z}$, are sufficient to check whether a given sequence is u.d. mod 1. This converts the study of uniform distributions into a study of the behaviour of appropriate exponential sums.*

We now state a couple of direct consequences of the Weyl criterion 2.2.1.

Theorem 2.2.2. *Let $X = (x_n)_{n\in\mathbb{N}}$ be a sequence of real numbers such that the difference*

$$\Delta x_n := x_{n+1} - x_n$$

is monotonic with respect to n. Further assume that

$$\lim_{n\to\infty} \Delta x_n = 0 \quad \text{and} \quad \lim_{n\to\infty} n|\Delta x_n| = \infty.$$

Then, the sequence is u.d. mod 1.

Corollary 2.2.3 (Fejér's Theorem). *Let $f : [1,\infty) \to \mathbb{R}$ be a function such that there exists some $x_0 \geq 1$ such that f is differentiable for all $x \geq x_0$. If $f'(x) \to 0$ monotonically as $x \to \infty$, and if $\lim\limits_{x\to\infty} x|f'(x)| = \infty$, then, the sequence $(f(n))_{n\in\mathbb{N}}$ is u.d. mod 1.*

We conclude the section with a list of equivalent criteria for uniform distribution.

Theorem 2.2.4. *Let $X = (x_n)_{n\in\mathbb{N}}$ be a sequence of real numbers. Then, the following are equivalent.*

i. The sequence X is uniformly distributed modulo 1.

ii. Given any $[a,b] \subset [0,1)$, the limit

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \chi_{[a,b]}(\{x_n\}) = \int_0^1 \chi_{[a,b]}(x)\,dx = b-a.$$

iii. Given a continuous function $f : \mathbb{T} \to \mathbb{C}$, the limit

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(x_n) = \int_0^1 f(x)\,dx.$$

iv. *Given a Riemann-integrable function $f : \mathbb{T} \to \mathbb{C}$, the limit*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(x_n) = \int_{\mathbb{T}} f(x)\, dx.$$

v. *Let $\mathscr{H}$ be a countable subset of $C(\mathbb{T})$ such that its span is dense in $C(\mathbb{T})$. Then, given any $f \in \mathscr{H}$, the limit*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(x_n) = \int_{\mathbb{T}} f(x)\, dx.$$

vi. *(**The Weyl Criterion**) Given any non-zero integer m, the limit*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e(mx_n) = 0.$$

vii. *Given any natural number m, the limit*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e(mx_n) = 0.$$

2.3 Examples of Uniformly Distributed Sequences

In the final section of the chapter, let us compile some examples of sequences that are u.d. mod 1. These concrete examples are aimed at familiarizing the reader with the abstract concepts, and displaying the application of the Weyl criterion 2.2.1 and Fejér's theorem 2.2.3.

Example 2.1. *Let α be an irrational number. Then, the sequence of all integer multiples of α, $X_\alpha = \{\alpha n\}_{n\in\mathbb{N}}$, is uniformly distributed modulo* 1.
This can be easily seen using the Weyl criterion. Consider any non-zero integer m. We have, for a natural number N,

$$\frac{1}{N} \sum_{n=1}^{N} e(mn\alpha) = \frac{1}{N} e(m\alpha) \frac{1 - e(mN\alpha)}{1 - e(m\alpha)}$$

Thus,

$$\left| \frac{1}{N} \sum_{n=1}^{N} e(mn\alpha) \right| = \left| \frac{1}{N} e(m\alpha) \frac{1 - e(mN\alpha)}{1 - e(m\alpha)} \right| \leq \frac{1}{N} \frac{|1 - e(mN\alpha)|}{|1 - e(m\alpha)|} \leq \frac{1}{N} \frac{2}{|1 - e(m\alpha)|}.$$

Now, note that since α is irrational, $e(m\alpha)$ can never equal 1 for any non-zero integer m. Hence, $|1-e(m\alpha)|$ is a bounded quantity away from zero, and is fixed for each m. Therefore, we have

$$\lim_{N\to\infty}\left|\frac{1}{N}\sum_{n=1}^{N}e(mn\alpha)\right| \le \lim_{N\to\infty}\frac{1}{N}\frac{2}{|1-e(m\alpha)|}=0.$$

Finally,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}e(mn\alpha)=0,$$

which satisfies the Weyl criterion, and shows that the sequence is u.d. mod 1.

Example 2.2. *Let α be an irrational number and let d be any natural number. Then, the sequence $\{\alpha n^d\}_{n\in\mathbb{N}}$ is u.d. mod 1. This is shown in [Wey16].*

Example 2.3. *Consider the polynomial $p(x)=a_m x^m + a_{m-1}x^{m-1}+\cdots+a_1x+a_0$ such that at least one of the coefficients a_j's is irrational. Then, the sequence $\{p(n)\}_{n\in\mathbb{N}}$ is u.d. mod 1. This example is also due to [Wey16]. See [KN74, Chapter 1, Theorem 3.2] for further details.*

Example 2.4. *Let α and θ be fixed positive real numbers such that θ is a non-integer. Then the sequence $\left\{\alpha n^{\theta}\right\}_{n\in\mathbb{N}}$ is u.d. mod 1. Corollary 2.2.3 can be used to show uniform distribution for $0<\theta<1$, and [Csi30] shows the same for $\theta>1$.*

Example 2.5. *Let $\alpha>0$ and $\theta>1$ be fixed real numbers. Then, the sequence $\left\{\alpha\left(\log(n)\right)^{\theta}\right\}_{n\in\mathbb{N}}$ is u.d. mod 1, as can be seen from Corollary 2.2.3.*

Example 2.6. *(**Metric Result**) Let $(x_n)_{n\in\mathbb{N}}$ be a sequence of distinct integers. Then, for almost all real α, [Wey16] showed that the sequence $\{\alpha x_n\}_{n\in\mathbb{N}}$ is u.d. mod 1.*

This concludes the chapter.

Chapter 3

Spacing Statistics of Sequences in $[0,1)$

The previous chapter shed light on the basic theory of uniform distributions. This chapter goes beyond u.d. mod 1 to investigate some finer local statistical properties that a sequence may possess, namely, correlations and level-spacings.

Correlations (or, m-level correlations) measure the relation between elements of a sequence to reveal any underlying clustering properties or randomness in structure. Level-spacings deal with the spacing or distance between consecutive terms when plotted on the unit interval (or the circle $\mathbb{T}$), revealing how evenly the sequence fills the interval (or $\mathbb{T}$).

For the purpose of this chapter, and the rest of the report, we will assume without loss of generality that our sequence $X = (x_n)_{n\in\mathbb{N}}$ lies in the unit interval $[0,1)$. If not, we take the fractional part of each term and bring it into the unit interval.

3.1 Level-Spacings

Let $X = (x_n)_{n\in\mathbb{N}}$ be a sequence in $[0,1)$. We define the level-spacing distribution function of the sequence as follows.

Definition 3.1.1 (Level-Spacing Distribution Function). *Consider the sequence X and take its first N terms. Arrange these terms in ascending order and re-label them as:*

$$0 \leq x_{1,N} \leq x_{2,N} \leq \cdots \leq x_{N,N} < 1.$$

*X is said to have a **level-spacing distribution function** $P : [0,\infty) \to \mathbb{R}$ if, given any interval $[a,b] \subset [0,\infty)$, we have*

$$\lim_{N\to\infty} \frac{1}{N} \# \left\{ 1 \leq n < N : x_{n+1,N} - x_{n,N} \in \left[\frac{a}{N}, \frac{b}{N} \right] \right\} = \int_a^b P(s)ds.$$

Definition 3.1.2 (Poissonian Level-Spacing Distribution Function). *The level-spacing distribution function of the sequence X is said to be **Poissonian** if*

$$P(s) = e^{-s}.$$

We would now like to see some examples of sequences that have a level-spacing distribution function.

Example 3.1. *Let $\alpha > 0$ and $\theta > 1$ be fixed real numbers. Then, the sequence $\left\{ \alpha (\log(n))^{\theta} \right\}_{n\in\mathbb{N}}$ has a Poissonian level-spacing distribution function. This is proved in [LT24].*
Note that the sequence is u.d. mod 1 too (refer Example 2.5).

Example 3.2. *Consider the sequence $\{\sqrt{n}\}_{n\in\mathbb{N}}$. It is shown in [EM04], that the sequence has a level-spacing distribution function, given by*

$$P(s) = \begin{cases} \dfrac{6}{\pi^2} & \text{if } s \in \left[0, \dfrac{1}{2}\right], \\ F_2(s) & \text{if } x \in \left(\dfrac{1}{2}, 2\right), \\ F_3(s) & \text{if } x \in [2,\infty). \end{cases}$$

Here, $F_2(s)$ is given by

$$F_2(s) = \frac{1}{\zeta(2)} \left(\frac{2}{3}(4r-1)^{\frac{3}{2}} \psi(r) + (1-6r)\log(r) + (2r-1) \right)$$

where, $r = \dfrac{1}{2s}$ and

$$\psi(r) = \arctan\left(\frac{2r-1}{\sqrt{4r-1}} \right) - \arctan\left(\frac{1}{\sqrt{4r-1}} \right),$$

and $F_3(s)$ is given by

$$F_3(s) = \frac{1}{\zeta(2)}\left(4(1-4\alpha)(1-\alpha)^2\log(1-\alpha) - 2(1-2\alpha)^3\log(1-2\alpha) - 2\alpha^2\right),$$

where

$$\alpha = \frac{1}{2}\left(1-\sqrt{1-4r}\right).$$

*Clearly, the level-spacing distribution function is **not** Poissonian.*
Note that the sequence is u.d. mod 1 *as well (this is a special case of Example 2.4).*

Example 3.3. *Let $b > 1$ be a transcendental number. Consider the sequence $\{\log_b n\}_{n\in\mathbb{N}}$. Then, [MS13] states that the given sequence has a level-spacing distribution function, given by*

$$P(s) = \begin{cases} \frac{1}{s^2\log b}\left(F(s\log b) - F\left(\frac{s}{b}\log b\right)\right) & \text{if } 0 < s < \frac{1}{\log b}, \\ -\frac{1}{s^2\log b}F\left(\frac{s}{b}\log b\right) & \text{if } \frac{1}{\log b} \le s \le \frac{b}{\log b}, \\ 0 & \text{if } s > \frac{b}{\log b}, \end{cases}$$

where

$$F(x) = x\frac{\partial}{\partial x}\left(x;\frac{1}{b}\right) - \left(x;\frac{1}{b}\right),$$

and the symbol $(\cdot;\cdot)$ is the Pochhammer symbol, and is given by

$$(a;q) := \prod_{j=0}^{\infty}\left(1-aq^j\right), \quad \textit{and hence,} \quad \frac{\partial}{\partial x}(a;q) = -(a;q)\sum_{j=0}^{\infty}\frac{1}{q^{-j}-a}.$$

*Here too, the the level-spacing distribution function is **not** Poissonian.*
*Note that in this case, the sequence is **not** u.d. mod* 1.

Example 3.4. *Let $(x_n)_{n\in\mathbb{N}}$ be a lacunary sequence of integers. By **lacunary**, we mean:*

$$\liminf_{n\to\infty}\frac{x_{n+1}}{x_n} > 1.$$

Then, for almost all real α, the sequence $\{\alpha x_n\}_{n\in\mathbb{N}}$ has the Poissonian level-spacing distribution function.
Note that this is a metric result, and is shown in [RZ02].
Also, the sequence is u.d. mod 1*, by Example 2.6.*

Example 3.5. *The authors in [AHZ23] explicitly construct a sequence that has a Poissonian level-spacing distribution function, but is not u.d. mod 1. We do not give the construction here, and point the reader to the reference for all the details.*
This shows that a sequence having Poissonian level-spacing distribution function need not necessarily be u.d. mod 1.

Remark 3.1.1. *Even though the level-spacing distribution function is a significant statistic, it requires the elements of the given sequence to be ordered, which is a non-trivial task. Thus, we would like to study some statistics that take the difference between unordered pairs of elements from the sequence, and then try to recover the level-spacing distribution from these statistics. In this context, we introduce the correlation statistics.*

3.2 Pair Correlation Statistics

This section introduces the pair correlation sum and the pair correlation function of a sequence. The definition, as can be seen below, makes use of unordered spacings, and hence, is easier to compute.

Definition 3.2.1 (Pair Correlation Sum). *Let $X = (x_n)_{n\in\mathbb{N}}$ be a sequence in $[0,1)$. Given a real number $s > 0$ and a natural number N, define the **pair correlation sum** of X to be*

$$R_{2,X,N}(s) := \frac{1}{N}\#\left\{(j,k) : 1 \le j \ne k \le N : ||x_j - x_k|| \le \frac{s}{N}\right\}.$$

Remark 3.2.1. *As in Remark 2.1.1, we can write the pair correlation sum $R_{2,X,N}(s)$ as*

$$R_{2,X,N}(s) = \frac{1}{N}\sum_{\substack{(j,k)\in[N]^2\\ j\ne k}} \chi_{[0,s]}\left(N||x_j - x_k||\right)$$

Also, it can easily be seen that

$$||x_j - x_k|| \le \frac{s}{N} \iff ((x_j - x_k)) \in \left[-\frac{s}{N}, \frac{s}{N}\right].$$

Thus,

$$R_{2,X,N}(s) = \frac{1}{N}\sum_{\substack{(j,k)\in[N]^2\\ j\ne k}} \chi_{[0,s]}\left(N||x_j - x_k||\right) = \frac{1}{N}\sum_{\substack{(j,k)\in[N]^2\\ j\ne k}} \chi_{[-s,s]}\left(N((x_j - x_k))\right)$$

Definition 3.2.2 (Pair Correlation Function). *Let $X = (x_n)_{n\in\mathbb{N}}$ be a sequence in $[0,1)$, and define the pair correlation sum as in Definition 3.2.1. X is said to have a **pair correlation function** if the limit $\lim\limits_{N\to\infty} R_{2,X,N}(s)$ exists for all positive real s. In that case, we define the pair correlation function as*

$$R_{2,X}(s) := \lim_{N\to\infty} R_{2,X,N}(s).$$

Definition 3.2.3 (Poissonian Pair Correlation Function). *The pair correlation function of the sequence X is said to be **Poissonian** if*

$$R_{2,X}(s) := \lim_{N\to\infty} R_{2,X,N}(s) = 2s.$$

There is a beautiful relation between Poissonian pair correlations and uniform distribution. In particular, [ALP18], [GL17], and [Ste17] independently showed that Poissonian pair correlation is a stronger property than uniform distribution. Precisely stating this, we have the following theorem.

Theorem 3.2.1. *Let X be a sequence in $[0,1)$. If X has Poissonian pair correlation, then X is uniformly distributed modulo 1.*

Let us now look at some examples of sequences having a pair correlation function.

Example 3.6. *Consider the sequence $\{\sqrt{n}\}_{n\in\mathbb{N}\setminus\square}$, where $\square = \{n^2 : n \in \mathbb{N}\}$ is the set of all positive perfect squares. The authors of [EMV15] have shown that the sequence has the Poissonian pair correlation function.*

Example 3.7. *Let α and θ be positive real numbers such that $0 < \theta < \dfrac{14}{41}$. Consider the sequence $\{\alpha n^{\theta}\}_{n\in\mathbb{N}}$. Then, it is shown in[LST23] that the sequence has the Poissonian pair correlation function for all $\alpha > 0$.*

Example 3.8. *Consider the sequence $X = \{\log_2(2n-1)\}_{n\in\mathbb{N}}$. An explicit computation in [Wei23] shows that the sequence has a pair correlation function that is **not** Poissonian. The function is* *given by*

$$R_{2,X}(s) = \begin{cases} 3k - \dfrac{3k^2+2k}{4s\log 2} & \text{if } \dfrac{k}{\log 4} \le s \le \dfrac{k+1}{\log 4}, \text{ where } k \in \mathbb{N}\cup\{0\} \text{ is even, and } s \neq 0, \\ 3k+1 - \dfrac{3k^2+4k+1}{4s\log 2} & \text{if } \dfrac{k}{\log 4} \le s \le \dfrac{k+1}{\log 4}, \text{ where } k \in \mathbb{N}\cup\{0\} \text{ is odd.} \end{cases}$$

There are very few examples of sequences with an explicit description having the Poissonian pair correlation function. However, we do have several **metric results**. Some of them are listed below.

Example 3.9. *Let α and θ be real numbers such that $\theta > 1$ (contrast this with Example 3.7). Then, [AEM21] proved that the sequence $\{\alpha n^{\theta}\}_{n\in\mathbb{N}}$ has the Poissonian pair correlation function for almost all real α.*

Example 3.10. *Consider a polynomial $p(x) \in \mathbb{Z}[x]$ of degree greater than or equal to 2. Then, [RS98] showed that the sequence $\{\alpha\, p(n)\}_{n\in\mathbb{N}}$ has Poissonian pair correlation function for almost all real α.*

Example 3.11. *Let $(x_n)_{n\in\mathbb{N}}$ be a positive, real-valued, lacunary sequence (recall the definition of lacunary from Example 3.4). Then we have, from [RT20], that the sequence $\{\alpha x_n\}_{n\in\mathbb{N}}$ has Poissonian pair correlation function for almost all real α.*

Remark 3.2.2. *Theorem 3.2.1 tells us that the sequences in Examples 3.6 and 3.7 are u.d. mod 1. Similarly, we get metric results for uniform distribution from Examples 3.9, 3.10 and 3.11. However, note that the converse of the theorem is not true. Recall from Example 2.1 that the sequence $\{\alpha n\}_{n\in\mathbb{N}}$ is u.d. mod 1. But the sequence does not have Poissonian pair correlation function.*

3.3 Higher-Level Correlation Statistics

We can generalize the definition of the pair correlation sum/function to higher dimensions by simultaneously considering the difference between two or more unordered pairs of elements of a sequence.

Definition 3.3.1 (*m*-**Level Correlation Sum).** *Let $X = (x_n)_{n\in\mathbb{N}}$ be a sequence in $[0,1)$. Let $m \geq 2$*

*be a fixed natural number. Given a bounded set $B \subset \mathbb{R}^{m-1}$, and a natural number N, define the m-**level correlation sum** of X to be*

$$R_{m,X,N}(B)$$
$$:= \frac{1}{N}\#\left\{(i_1,i_2,\ldots,i_m) \in [N]^m : i_1,i_2,\ldots,i_m \text{ are all distinct, } (((x_{i_1}-x_{i_2})),\ldots,((x_{i_1}-x_{i_m}))) \in \frac{1}{N}B\right\}.$$

Remark 3.3.1. *As in Remarks 2.1.1 and 3.2.1, we can write the m-level correlation sum in the following way.*

$$R_{m,X,N}(B) = \frac{1}{N}\sum_{\vec{n}\in[N]^m}^{*} \chi_B\left(N(((x_{n_1}-x_{n_2})),\ldots,((x_{n_1}-x_{n_m})))\right),$$

where $\vec{n} = (n_1,\ldots,n_m)$ denotes an m-tuple, and the '' on the summation denotes m-tuples in which $n_1,n_2,\ldots,n_m$ are all distinct.*

Remark 3.3.2. *For $m \geq 3$, we cannot define the m-level correlations in terms of only symmetric sets, as we did for pair correlations (as in Remark 3.2.1). A detailed explanation for the same is given in [HZ23, Appendix A].*
Thus, we cannot directly define the higher-level correlations using the 'distance to the nearest integer' definition, as it does not take the symmetric nature of the differences into account. Hence, we take inspiration from Remark 3.2.1, and define it using the 'signed distance to the nearest integer'.

Definition 3.3.2 (m-Level Correlation Function). *Let $X = (x_n)_{n\in\mathbb{N}}$ be a sequence in $[0,1)$, and define the m-level correlation sum as in Definition 3.3.1. X is said to have an m-**level correlation function** if the limit $\lim_{N\to\infty} R_{m,X,N}(B)$ exists for all bounded sets $B \subset \mathbb{R}^{m-1}$. In that case, we define the m-level correlation function as*

$$R_{m,X}(B) := \lim_{N\to\infty} R_{m,X,N}(B).$$

Definition 3.3.3 (Poissonian m-Level Correlation Function). *The sequence X is said to have a **Poissonian** m-level correlation function if*

$$R_{m,X}(B) := \lim_{N\to\infty} R_{m,X,N}(B) = vol(B),$$

for any bounded set $B \subset \mathbb{R}^{m-1}$.

Analogous to Theorem 3.2.1, we have a relation between Poissonian higher-level correlations and uniform distribution, proved in [HZ23, Theorem 1.1]. We state the same here as a theorem.

Theorem 3.3.1. *Let X be a sequence in $[0,1)$. If X has Poissonian m-level correlation for some $m \geq 2$, then X is u.d. mod 1.*

As before, let us look at some examples of sequences having higher-level correlations. However, computing the higher-level correlations is much harder than computing the pair correlation. This is because the analytical and combinatorial arguments become much more involved for the higher-level correlations. Thus, most examples are in the form of metric results. We list some here.

Example 3.12. *Let $m \geq 2$ be a fixed integer, and let θ be a positive real number. Consider the sequence $\left\{n^{\theta}\right\}_{n\in\mathbb{N}}$. Then, [TY23, Theorem 1.2] states the following: The m-level correlation function of the sequence is Poissonian for almost all real $\theta > 4m^2 - 4m - 1$.*
This means that the pair correlation function of the sequence is Poissonian for almost all real $\theta > 7$.

Example 3.13. *Let $X = (x_n)_{n\in\mathbb{N}}$ be a lacunary sequence of real numbers (recall the definition of lacunary from Example 3.4). Consider the sequence $\{\alpha x_n\}_{n\in\mathbb{N}}$.*
When X is a sequence of integers, [RZ02] shows that, for almost all real α, the sequence has Poissonian m-level correlation function for each integer $m \geq 2$.
This has been refined to real-valued, positive, lacunary sequences by [CY22].

Example 3.14. *Let $\alpha > 1$ be a real number. Then, [Ais+23] shows that, for almost all $\alpha > 1$, the m-level correlation function of the sequence $\{\alpha^n\}_{n\in\mathbb{N}}$ is Poissonian for every $m \geq 2$.*

Example 3.15. *Let $\alpha > 0$ and $\theta > 1$ be fixed real numbers. Then, the sequence $\left\{\alpha\left(\log(n)\right)^{\theta}\right\}_{n\in\mathbb{N}}$ has Poissonian m-level correlations for all $m \geq 2$. This was shown in [LT24, Theorem 2].*

We would like to end the list by giving an example of a deterministic result. This remarkable result is taken from [LT21].

Example 3.16. *Let α and θ be positive real numbers. Consider the sequence $\left\{\alpha n^{\theta}\right\}_{n\in\mathbb{N}}$*
We saw in Example 3.7 that, for $0 < \theta < \dfrac{14}{41}$, the sequence has the Poissonian pair correlation function for all $\alpha > 0$.

[LT21, Theorem 1.1] states that for any $m \geq 3$, the sequence has the Poissonian m-level correlation function for any $0 < \theta < \dfrac{1}{m^2+m-1}$ and any $\alpha > 0$.

As remarked in Remark 3.1.1, it is quite challenging to compute the level-spacing distribution function of any sequence. We try to get around this by computing all the m-level correlations of the sequence, and then use these to recover the level-spacing statistics of the sequence. Chapter 5 of the thesis will focus on this recovery.

The next chapter will introduce us to the smooth analogues of the correlation functions.

Chapter 4

Smooth Version of m-Level Correlations

The previous chapter introduced the classical notion of m-level correlations in terms of the characteristic function of some bounded set. However, the characteristic function, being discontinuous, does not allow us to use the analytic machinery developed for continuous functions, thus, limiting its functionality.

Hence, we define "smooth" versions of these correlations in terms of smooth functions. These smooth versions are either in terms of continuous functions having compact support, or continuous functions whose Fourier transforms have compact support. This brings along with it the theory of Fourier analysis, which directly leads us to exponential sums. Thus, all our questions about correlations between pairs (or tuples of pairs) of a sequence transform into questions about the behaviour of exponential sums involving the first N terms of the sequence.

As before, we fix a sequence $X = (x_n)_{n\in\mathbb{N}}$ in $[0,1)$.

4.1 Pair Correlations

We first discuss the treatment for pair correlations and later generalize it to higher-level correlations in the next section.

Let us begin by periodizing the pair correlation function. This has been adapted from [Mah24, Chapter 2]. The interested reader can refer to the same for some related auxiliary results.

Lemma 4.1.1. *Let $s > 0$ be fixed. Then, there exists a positive integer N_0 such that for all $N \geq N_0$,*

$$R_{2,X,N}(s) = \frac{1}{N} \sum_{\substack{(j,k)\in[N]^2 \\ j\neq k}} \sum_{m\in\mathbb{Z}} \chi_{[-s,s]}(N(x_j - x_k + m)).$$

Thus,

$$R_{2,X}(s) = \lim_{N\to\infty} \frac{1}{N} \sum_{\substack{(j,k)\in[N]^2 \\ j\neq k}} \sum_{m\in\mathbb{Z}} \chi_{[-s,s]}(N(x_j - x_k + m)),$$

whenever the limit exists for all positive real numbers s.

Proof.

We are given some positive real number s. We know that there exists some positive integer N_0 such that $\frac{s}{N_0} < \frac{1}{2}$. Hence, for any positive integer $N \geq N_0$, we have $\frac{s}{N} \leq \frac{s}{N_0} < \frac{1}{2}$.

Let $(j,k) \in [N]^2$ be a pair such that $j \neq k$ and $0 \leq \{x_j - x_k\} < \frac{1}{2}$. We first claim that for any non-zero integer m, $\chi_{[-s,s]}(N(\{x_j - x_k\} - m)) = 0$ for all $N \geq N_0$.

Let us prove the claim by contradiction, that is, suppose there exists some $N_1 \geq N_0$ and some pair $(j_1,k_1) \in [N]^2$, with $0 \leq \{x_{j_1} - x_{k_1}\} < \frac{1}{2}$ such that $\chi_{[-s,s]}(N_1(\{x_{j_1} - x_{k_1}\} - m_1)) = 1$ for some $m_1 \neq 0$. This means that

$$N_1\left(\{x_{j_1} - x_{k_1}\} - m_1\right) \in [-s,s],$$

and thus,

$$\{x_{j_1} - x_{k_1}\} \in \left[m_1 - \frac{s}{N_1}, m_1 + \frac{s}{N_1}\right] \subsetneq \left[m_1 - \frac{1}{2}, m_1 + \frac{1}{2}\right],$$

where we have used the fact that $\frac{s}{N_1} \leq \frac{s}{N_0} < \frac{1}{2}$.

But, note that $\{x_{j_1} - x_{k_1}\} \in \left[0, \frac{1}{2}\right)$ by our assumption. Thus,

$$\{x_{j_1} - x_{k_1}\} \in \left[m_1 - \frac{1}{2}, m_1 + \frac{1}{2}\right] \cap \left[0, \frac{1}{2}\right).$$

However, for any non-zero integer m,

$$\left[m - \frac{s}{N_1}, m + \frac{s}{N_1}\right] \cap \left[0, \frac{1}{2}\right) = \emptyset.$$

This is a contradiction to our assumption. Thus, our claim is proven.

Now, for any $N \geq N_0$,

$$
\begin{aligned}
&\sum_{\substack{1\leq j\neq k\leq N\\ 0\leq\{x_j-x_k\}<\frac{1}{2}}} \chi_{[-s,s]}(N\{x_j-x_k\}) \\
&= \sum_{\substack{1\leq j\neq k\leq N\\ 0\leq\{x_j-x_k\}<\frac{1}{2}}} \chi_{[-s,s]}(N\{x_j-x_k\}) + \sum_{\substack{m\in\mathbb{Z}\\ m\neq 0}} \sum_{\substack{1\leq j\neq k\leq N\\ 0\leq\{x_j-x_k\}<\frac{1}{2}}} \chi_{[-s,s]}(N(\{x_j-x_k\}+m)) \\
&= \sum_{\substack{1\leq j\neq k\leq N\\ 0\leq\{x_j-x_k\}<\frac{1}{2}}} \sum_{m\in\mathbb{Z}} \chi_{[-s,s]}(N(\{x_j-x_k\}+m)) \qquad (4.1)\\
&= \sum_{\substack{1\leq j\neq k\leq N\\ 0\leq\{x_j-x_k\}<\frac{1}{2}}} \sum_{m'\in\mathbb{Z}} \chi_{[-s,s]}(N(\{x_j-x_k\}+[x_j-x_k]+m')) \\
&= \sum_{\substack{1\leq j\neq k\leq N\\ 0\leq\{x_j-x_k\}<\frac{1}{2}}} \sum_{m\in\mathbb{Z}} \chi_{[-s,s]}(N(x_j-x_k+m)).
\end{aligned}
$$

Let us prove an analogous claim for the case when $\{x_j-x_k\}\in\left[\frac{1}{2},1\right)$.
Here, we claim that for any integer $m\neq 1$, $\chi_{[-s,s]}\big(N(m-\{x_j-x_k\}\big)=0$ for all $N\geq N_0$ and for all pairs $(j,k)\in[N]^2$ such that $j\neq k$ and $\frac{1}{2}\leq\{x_j-x_k\}<1$.
Let us prove the claim by contradiction, that is, suppose there exists some $N_2\geq N_0$ and some pair $(j_2,k_2)\in[N]^2$, with $\frac{1}{2}\leq\{x_{j_2}-x_{k_2}\}<1$ such that $\chi_{[-s,s]}(N_2(m_2-\{x_{j_2}-x_{k_2}\}))=1$ for some $m_2\neq 1$. This means that

$$N_2(m_2-\{x_{j_2}-x_{k_2}\})\in[-s,s],$$

which is equivalent to saying that

$$N_2(\{x_{j_2}-x_{k_2}\}-m_2)\in[-s,s],$$

and hence,

$$\{x_{j_2}-x_{k_2}\}\in\left[m_2-\frac{s}{N_2},m_2+\frac{s}{N_2}\right]\subsetneq\left[m_2-\frac{1}{2},m_2+\frac{1}{2}\right],$$

where we have used the fact that $\dfrac{s}{N_2}\leq\dfrac{s}{N_0}<\dfrac{1}{2}$.
But, note that $\{x_{j_2}-x_{k_2}\}\in\left[\dfrac{1}{2},1\right)$ by our assumption. Thus,

$$\{x_{j_2}-x_{k_2}\}\in\left[m_2-\frac{1}{2},m_2+\frac{1}{2}\right]\cap\left[\frac{1}{2},1\right).$$

However, for any integer $m \neq 1$,

$$\left[m - \frac{s}{N_2}, m + \frac{s}{N_2}\right] \cap \left[\frac{1}{2}, 1\right) = \emptyset.$$

This is a contradiction to our assumption. Thus, our claim is proven.
Note that by the symmetry of $[-s,s]$, we have that for any $m \in \mathbb{Z}$,

$$m - \{x_j - x_k\} \in [-s,s] \iff \{x_j - x_k\} - m \in [-s,s].$$

Now, for any $N \geq N_0$,

$$\begin{aligned}
&\sum_{\substack{1 \leq j \neq k \leq N \\ \frac{1}{2} \leq \{x_j - x_k\} < 1}} \chi_{[-s,s]}(N(1 - \{x_j - x_k\})) \\
&= \sum_{\substack{1 \leq j \neq k \leq N \\ \frac{1}{2} \leq \{x_j - x_k\} < 1}} \chi_{[-s,s]}(N(\{x_j - x_k\} - 1)) + \sum_{\mathbb{Z} \ni m \neq 1} \sum_{\substack{1 \leq j \neq k \leq N \\ \frac{1}{2} \leq \{x_j - x_k\} < 1}} \chi_{[-s,s]}(N(\{x_j - x_k\} - m)) \\
&= \sum_{\substack{1 \leq j \neq k \leq N \\ \frac{1}{2} \leq \{x_j - x_k\} < 1}} \sum_{m \in \mathbb{Z}} \chi_{[-s,s]}(N(\{x_j - x_k\} + m)) \\
&= \sum_{\substack{1 \leq j \neq k \leq N \\ \frac{1}{2} \leq \{x_j - x_k\} < 1}} \sum_{m' \in \mathbb{Z}} \chi_{[-s,s]}(N(\{x_j - x_k\} + [x_j - x_k] + m')) \\
&= \sum_{\substack{1 \leq j \neq k \leq N \\ \frac{1}{2} \leq \{x_j - x_k\} < 1}} \sum_{m \in \mathbb{Z}} \chi_{[-s,s]}(N(x_j - x_k + m)).
\end{aligned} \tag{4.2}$$

Now, with the prerequisites in hand, let us consider $R_{2,X,N}(s)$, where $N \geq N_0$.

$$\begin{aligned}
R_{2,X,N}(s) &= \frac{1}{N} \# \left\{ (j,k) : 1 \leq j \neq k \leq N : ||x_j - x_k|| \leq \frac{s}{N} \right\} \\
&= \frac{1}{N} \sum_{1 \leq j \neq k \leq N} \chi_{[0,s]}(N||x_j - x_k||) \\
&= \frac{1}{N} \sum_{1 \leq j \neq k \leq N} \chi_{[-s,s]}(N||x_j - x_k||) \\
&= \frac{1}{N} \sum_{\substack{1 \leq j \neq k \leq N \\ 0 \leq \{x_j - x_k\} < \frac{1}{2}}} \chi_{[-s,s]}(N||x_j - x_k||) + \frac{1}{N} \sum_{\substack{1 \leq j \neq k \leq N \\ \frac{1}{2} \leq \{x_j - x_k\} < 1}} \chi_{[-s,s]}(N||x_j - x_k||)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{N} \sum_{\substack{1 \le j \ne k \le N \\ 0 \le \{x_j - x_k\} < \frac{1}{2}}} \chi_{[-s,s]}(N\{x_j - x_k\}) + \frac{1}{N} \sum_{\substack{1 \le j \ne k \le N \\ \frac{1}{2} \le \{x_j - x_k\} < 1}} \chi_{[-s,s]}(N(1 - \{x_j - x_k\})) \\
&= \frac{1}{N} \sum_{\substack{1 \le j \ne k \le N \\ 0 \le \{x_j - x_k\} < \frac{1}{2}}} \sum_{m \in \mathbb{Z}} \chi_{[-s,s]}(N(x_j - x_k + m)) + \frac{1}{N} \sum_{\substack{1 \le j \ne k \le N \\ \frac{1}{2} \le \{x_j - x_k\} < 1}} \sum_{m \in \mathbb{Z}} \chi_{[-s,s]}(N(x_j - x_k + m)) \\
&= \frac{1}{N} \sum_{\substack{(j,k) \in [N]^2 \\ j \ne k}} \sum_{m \in \mathbb{Z}} \chi_{[-s,s]}(N(x_j - x_k + m)),
\end{aligned}
$$

where we have used Equations 4.1 and 4.2 in the second-last step to periodize the sum. This completes the proof of the lemma. □

Remark 4.1.1 (Equivalent Forms for the Pair Correlation Sum). *By Lemma 4.1.1 and Remark 3.2.1, we have that, for a large enough natural number N,*

$$
\begin{aligned}
R_{2,X,N}(s) &= \frac{1}{N} \# \left\{ (j,k) : 1 \le j \ne k \le N : ||x_j - x_k|| \le \frac{s}{N} \right\} \\
&= \frac{1}{N} \sum_{\substack{(j,k) \in [N]^2 \\ j \ne k}} \chi_{[0,s]} \left(N ||x_j - x_k|| \right) \\
&= \frac{1}{N} \sum_{\substack{(j,k) \in [N]^2 \\ j \ne k}} \chi_{[-s,s]} \left(N((x_j - x_k)) \right) \\
&= \frac{1}{N} \sum_{\substack{(j,k) \in [N]^2 \\ j \ne k}} \sum_{m \in \mathbb{Z}} \chi_{[-s,s]}(N(x_j - x_k + m)).
\end{aligned}
$$

4.1.1 A Smooth Version

Taking inspiration from Lemma 4.1.1, let us now proceed to the smooth version of the pair correlation function. Before that, let us define the Fourier transform of a function.

Definition 4.1.1 (The Fourier Transform). *Given any function $f \in L^1(\mathbb{R})$, we define the Fourier transform of f by*

$$
\hat{f}(t) := \int_{-\infty}^{\infty} f(x) e(-tx) dx, \quad \text{for all real numbers } t.
$$

Definition 4.1.2 (Smooth Pair Correlation Sum). *Let $\mathscr{P}$ denote the set of real-valued, continuous, Lebesgue-integrable functions g on $\mathbb{R}$ satisfying the following conditions.*

i. For every $N \in \mathbb{N}$, the series

$$G_N(x) := \sum_{m\in\mathbb{Z}} g\left(N(x+m)\right) \tag{4.3}$$

converges absolutely and uniformly for all $x \in \mathbb{R}$.

ii. The Fourier transform of g is compactly supported. Without loss of generality, suppose that $supp(\hat{g}) = [-1,1]$.

For $g \in \mathscr{P}$, define the ***smoothened pair correlation sum****,*

$$R_{2,X,N}(g) := \frac{1}{N}\sum_{\substack{(j,k)\in[N]^2\\ j\neq k}} G_N(x_j - x_k) = \frac{1}{N}\sum_{\substack{(j,k)\in[N]^2\\ j\neq k}}\sum_{m\in\mathbb{Z}} g\left(N(x_j - x_k + m)\right).$$

Remark 4.1.2. *Observe that, by Lemma 4.1.1, we have, for large enough natural number N,*

$$R_{2,X,N}(s) = \frac{1}{N}\sum_{\substack{(j,k)\in[N]^2\\ j\neq k}}\sum_{m\in\mathbb{Z}} \chi_{[-s,s]}(N(x_j - x_k + m)) = R_{2,X,N}(\chi_{[s,s]})$$

Theorem 4.1.2 (Smooth Pair Correlation Function). *Suppose X is a sequence in $[0,1)$ such that for every $g \in \mathscr{P}$,*

$$R_{2,X}(g) := \lim_{N\to\infty} R_{2,X,N}(g) = \widehat{g}(0) = \int_{-\infty}^{\infty} g(x)dx.$$

Then, given any $s > 0$,

$$R_{2,X}(s) := \lim_{N\to\infty} R_{2,X,N}(s) = \lim_{N\to\infty} R_{2,X,N}(\chi_{[s,s]}) = \int_{-\infty}^{\infty} \chi_{[-s,s]}(x)dx = 2s$$

In other words, X admits the Poissonian pair correlation function.

Proof.

Suppose, for now, that for every $g \in \mathscr{P}$,

$$R_{2,X}(g) = \int_{-\infty}^{\infty} g(x)dx.$$

For some $s > 0$, and any large enough natural number N such that $\frac{s}{N} < \frac{1}{2}$, consider the pair correlation sum

$$R_{2,X,N}(s)=\frac{1}{N}\sum_{\substack{(j,k)\in[N]^2\\ j\neq k}}\sum_{m\in\mathbb{Z}}\chi_{[-s,s]}(N(x_j-x_k+m))=R_{2,X,N}(s)=\frac{1}{N}\sum_{\substack{(j,k)\in[N]^2\\ j\neq k}}\sum_{m\in\mathbb{Z}}\chi_{[-\frac{s}{N},\frac{s}{N}]}(x_j-x_k+m)$$
$$=\frac{1}{N}\sum_{\substack{(j,k)\in[N]^2\\ j\neq k}}\chi_{[-\frac{s}{N},\frac{s}{N}]}\left(((x_j-x_k))\right).$$

From Theorem 1.3.4, we know that this characteristic function can be approximated from above and below by the Beurling-Selberg polynomials. Thus, given any non-negative integer M, we have, for any $x\in\left[-\frac{1}{2},\frac{1}{2}\right)$,

$$S_M^-(x)\le\chi_{[-\alpha,\alpha]}(x)\le S_M^+(x),$$

where, $\alpha=\frac{s}{N}$. In particular, since $((x))\in\left[-\frac{1}{2},\frac{1}{2}\right)$ for any real x, therefore

$$S_M^-\left(((x_j-x_k))\right)\le\chi_{[-\alpha,\alpha]}\left(((x_j-x_k))\right)\le S_M^+\left(((x_j-x_k))\right).$$

This means that

$$\frac{1}{N}\sum_{\substack{(j,k)\in[N]^2\\ j\neq k}}S_M^-\left(((x_j-x_k))\right)\le\frac{1}{N}\sum_{\substack{(j,k)\in[N]^2\\ j\neq k}}\chi_{[-\alpha,\alpha]}\left(((x_j-x_k))\right)\le\frac{1}{N}\sum_{\substack{(j,k)\in[N]^2\\ j\neq k}}S_M^+\left(((x_j-x_k))\right).$$

Taking limit as $N\to\infty$, we have

$$\lim_{N\to\infty}\frac{1}{N}\sum_{\substack{(j,k)\in[N]^2\\ j\neq k}}S_M^-\left(((x_j-x_k))\right)\le\liminf_{N\to\infty}R_{2,X,N}(s)\le\limsup_{N\to\infty}R_{2,X,N}(s)\le\lim_{N\to\infty}\frac{1}{N}\sum_{\substack{(j,k)\in[N]^2\\ j\neq k}}S_M^+\left(((x_j-x_k))\right).$$

Thus,

$$\int_{\mathbb{T}}S_M^-(x)dx-\int_{\mathbb{T}}\chi_{[-\alpha,\alpha]}(x)dx\le\liminf_{N\to\infty}R_{2,X,N}(s)-\int_{\mathbb{T}}\chi_{[-\alpha,\alpha]}(x)dx$$
$$\le\limsup_{N\to\infty}R_{2,X,N}(s)-\int_{\mathbb{T}}\chi_{[-\alpha,\alpha]}(x)dx\le\int_{\mathbb{T}}S_M^+(x)dx-\int_{\mathbb{T}}\chi_{[-\alpha,\alpha]}(x)dx.$$

Using Theorem 1.3.4, given any $\varepsilon>0$, there exists a large enough integer M such that

$$-\varepsilon\le\liminf_{N\to\infty}R_{2,X,N}(s)-\int_{\mathbb{T}}\chi_{[-\alpha,\alpha]}(x)dx\le\limsup_{N\to\infty}R_{2,X,N}(s)-\int_{\mathbb{T}}\chi_{[-\alpha,\alpha]}(x)dx\le\varepsilon.$$

Taking $\varepsilon \to 0$, we finally get

$$\lim_{N\to\infty} R_{2,X,N}(s) = 2s.$$

Hence, the sequence admits the Poissonian pair correlation function. □

4.1.2 A Weyl-type Criterion for Poissonian Pair Correlations

In the previous sub-section, have defined a smooth version of the pair correlation function, and proved its equivalence to the characteristic version in the case of Poissonian pair correlation. Let us now re-frame the argument in terms of exponential sums.
Recall the smoothened pair correlation sum (Definition 4.1.2),

$$R_{2,X,N}(g) := \frac{1}{N} \sum_{\substack{(j,k)\in[N]^2 \\ j\neq k}} G_N(x_j - x_k) \text{ for some } g \in \mathscr{P}.$$

An obvious way to bring in exponential sums is to write $G_N(x)$ in terms of its Fourier coefficients, if possible. This is, in fact, possible, as we will see below.

Lemma 4.1.3. *Let $g \in \mathscr{P}$, and define $G_N(x)$ for any natural number N as in Equation 4.3. Then, the Fourier coefficients of G_N are given by*

$$\widehat{G_N}(n) = \frac{1}{N}\, \hat{g}\left(\frac{n}{N}\right) \textit{ for any integer } n,$$

where $\hat{g}(t)$ is the Fourier transform of g. Thus, the Fourier series of G_N converges to itself, and is given by

$$G_N(x) = \frac{1}{N} \sum_{|m|\leq N} \hat{g}\left(\frac{m}{N}\right) e(mx).$$

Proof.
Fix some $g \in \mathscr{P}$. Given any integer n, consider

$$\widehat{G_N}(n) = \int_0^1 G_N(x) e(-nx)\, dx = \int_0^1 \sum_{m\in\mathbb{Z}} g\left(N(x+m)\right) e(-nx)\, dx.$$

Here, since the summation is absolutely and uniformly convergent for any x, therefore, we can interchange the sum and the integral. Also, we can change the variable: $t = N(x+m)$ to get,

$$
\begin{aligned}
\widehat{G_N}(n) &= \sum_{m\in\mathbb{Z}} \int_0^1 g\left(N(x+m)\right) e(-nx)\, dx \\
&= \frac{1}{N} \sum_{m\in\mathbb{Z}} \int_{Nm}^{N(m+1)} g(t) e\left(-n\left(\frac{t}{N} - m\right)\right)\, dt \\
&= \frac{1}{N} \int_{-\infty}^{\infty} g(t) e\left(-\frac{n}{N} t\right)\, dt \\
&= \frac{1}{N} \hat{g}\left(\frac{n}{N}\right).
\end{aligned}
$$

Now, we know that $\mathrm{supp}(\hat{g}) = [-1,1]$. Thus, $\hat{g}(x) = 0$ for any x such that $|x| > 1$. Hence,

$$
\sum_{n=-\infty}^{\infty} |\widehat{G_N}(n)| = \frac{1}{N} \sum_{n=-\infty}^{\infty} \left|\hat{g}\left(\frac{n}{N}\right)\right| = \frac{1}{N} \sum_{n=-N}^{N} \left|\hat{g}\left(\frac{n}{N}\right)\right| < \infty.
$$

This means that G_N is continuous, has period 1, and $\sum_{n=-\infty}^{\infty} |\widehat{G_N}(n)| < \infty$. Therefore, we can use Corollary 1.2.2 to get

$$
G_N(x) = \sum_{n=-\infty}^{\infty} \widehat{G_N}(n) e(nx) = \frac{1}{N} \sum_{n=-N}^{N} \hat{g}\left(\frac{n}{N}\right) e(nx).
$$

This proves the lemma. $\square$

Being equipped with this, let now substitute the same in the formula of the pair correlation sum. We get, for some $g \in \mathscr{P}$,

$$
\begin{aligned}
R_{2,X,N}(g) &= \frac{1}{N} \sum_{\substack{(j,k)\in[N]^2 \\ j\neq k}} G_N(x_j - x_k) \\
&= \frac{1}{N} \sum_{\substack{(j,k)\in[N]^2 \\ j\neq k}} \frac{1}{N} \sum_{|m|\le N} \hat{g}\left(\frac{m}{N}\right) e\left(m(x_j - x_k)\right) \\
&= \frac{1}{N^2} \sum_{|m|\le N} \hat{g}\left(\frac{m}{N}\right) \sum_{\substack{(j,k)\in[N]^2 \\ j\neq k}} e\left(m(x_j - x_k)\right) \\
&= \frac{1}{N^2} \sum_{|m|\le N} \hat{g}\left(\frac{m}{N}\right) \left(\sum_{(j,k)\in[N]^2} e\left(m(x_j - x_k)\right) - N \right).
\end{aligned}
$$

Now, we can write

$$\begin{aligned}\sum_{(j,k)\in[N]^2} e\left(m(x_j - x_k)\right) &= \sum_{j=1}^{N} e(mx_j) \sum_{k=1}^{N} e(-mx_k) \\ &= \sum_{j=1}^{N} e(mx_j) \overline{\sum_{j=1}^{N} e(mx_j)} \\ &= \left| \sum_{j=1}^{N} e(mx_j) \right|^2 .\end{aligned}$$

Thus, we get

$$R_{2,X,N}(g) = \frac{1}{N^2} \sum_{|m|\leq N} \hat{g}\left(\frac{m}{N}\right) \left(\left| \sum_{j=1}^{N} e(mx_j) \right|^2 - N \right) \tag{4.4}$$

Now,

$$R_{2,X,N}(g) = \left(1 - \frac{1}{N}\right) \hat{g}(0) + \frac{1}{N^2} \sum_{\substack{|m|\leq N \\ m\neq 0}} \hat{g}\left(\frac{m}{N}\right) \left(\left| \sum_{j=1}^{N} e(mx_j) \right|^2 - N \right)$$

Therefore, for X to have Poissonian pair correlations, we need the sum

$$\frac{1}{N^2} \sum_{\substack{|m|\leq N \\ m\neq 0}} \hat{g}\left(\frac{m}{N}\right) \left(\left| \sum_{j=1}^{N} e(mx_j) \right|^2 - N \right) \to 0 \text{ as } N \to \infty.$$

We will write this result as a lemma.

Lemma 4.1.4. *Let X be a sequence in $[0,1)$. Given any function $g \in \mathscr{P}$, and any natural number N, define the sum*

$$E_N(g) := \frac{1}{N^2} \sum_{\substack{|m|\leq N \\ m\neq 0}} \hat{g}\left(\frac{m}{N}\right) \left(\left| \sum_{j=1}^{N} e(mx_j) \right|^2 - N \right).$$

If this sum

$$E_N(g) \to 0 \text{ as } N \to \infty \ \forall\ g \in \mathscr{P}, \tag{4.5}$$

then the sequence admits the Poissonian pair correlation function.

Remark 4.1.3. *Consider, for any $g \in \mathscr{P}$ and any $N \in \mathbb{N}$,*

$$\begin{aligned}
|E_N(g)| &= \frac{1}{N^2}\left|\sum_{\substack{|m|\le N\\ m\ne 0}} \hat{g}\left(\frac{m}{N}\right)\left(\left|\sum_{j=1}^{N} e(mx_j)\right|^2 - N\right)\right| \\
&\le \frac{1}{N^2}\sum_{\substack{|m|\le N\\ m\ne 0}} \left|\hat{g}\left(\frac{m}{N}\right)\right| \cdot \left|\left|\sum_{j=1}^{N} e(mx_j)\right|^2 - N\right| \\
&= C_0 \sum_{\substack{|m|\le N\\ m\ne 0}} \left|\left|\frac{1}{N}\sum_{j=1}^{N} e(mx_j)\right|^2 - \frac{1}{N}\right|,
\end{aligned}$$

where $C_0 = \max\left\{\left|\hat{g}\left(\frac{m}{N}\right)\right| : -N \le m \le N\right\}$.
Suppose that this sum goes to zero as $N \to \infty$. Then, since all summands are non-negative, each individual term has to go to zero. Thus, we get, for every non-zero integer m,

$$\left|\sum_{j=1}^{N} e(mx_j)\right| \to \sqrt{N} \ \text{ as } N \to \infty. \tag{4.6}$$

If the sequence X satisfies the above equation 4.6, then X directly satisfies Equation 4.5, and also satisfies the Weyl criterion. This not only proves that the pair correlation function of the sequence is Poissonian, but also verifies Theorem 3.2.1.

However, we do not yet know whether this condition is strictly stronger than Poissonian pair correlation, or whether they are equivalent.

With this, we end the section on pair correlations, and move to higher-level correlations.

4.2 Higher-Level Correlations

As we did for the pair correlation sum, let us begin the section by periodizing the higher-level correlations.

Lemma 4.2.1. *Let $s > 0$ be a fixed positive real number. Then, there exists a natural number N_0*

such that for all $N \geq N_0$,

$$R_{m,X,N}(B) = \frac{1}{N} \sum_{\vec{n} \in [N]^m}^{*} \sum_{\vec{k} \in \mathbb{Z}^{m-1}} \chi_B \left(N \left((x_{n_1} - x_{n_2} + k_1), (x_{n_1} - x_{n_3} + k_2), \ldots, (x_{n_1} - x_{n_m} + k_{m-1}) \right) \right).$$

Thus,

$$R_{m,X}(B) = \lim_{N \to \infty} \frac{1}{N} \sum_{\vec{n} \in [N]^m}^{*} \sum_{\vec{k} \in \mathbb{Z}^{m-1}} \chi_B \left(N \left((x_{n_1} - x_{n_2} + k_1), (x_{n_1} - x_{n_3} + k_2), \ldots, (x_{n_1} - x_{n_m} + k_{m-1}) \right) \right),$$

whenever the limit exists for all bounded sets $B \subset \mathbb{R}^{m-1}$.

4.2.1 A Smooth Version

Let us write out a smooth version of the general m-level correlation sum, taking inspiration from the version in Remark 3.3.1.

Definition 4.2.1 (Smooth m-Level Correlation Sum). *Let $m \geq 2$ be a fixed positive integer. Let g be a real-valued function on $\mathbb{R}^{m-1}$. Then, for any natural number N, the m-level correlation sum of X at N with respect to g is given by*

$$R_{m,X,N}(g) =:= \frac{1}{N} \sum_{\vec{n} \in [N]^m}^{*} g \left(N \left(((x_{n_1} - x_{n_2})), \ldots, ((x_{n_{m-1}} - x_{n_m})) \right) \right).$$

We wish to prove that this version of the m-level correlations coincides with the characteristic function version when the correlation is Poissonian. Thus, we prove the following theorem:

Theorem 4.2.2 (Smooth Version of Poissonian m-Level Correlation Function). *Fix a natural number $m \geq 2$. Then, the following are equivalent:*

i. Given any $g \in C_c(\mathbb{R}^{m-1})$,

$$\lim_{N \to \infty} \frac{1}{N} \sum_{\vec{n} \in [N]^m}^{*} g(N(((x_{n_1} - x_{n_2})), \ldots, ((x_{n_1} - x_{n_m})))) = \int_{\mathbb{R}^{m-1}} g(\vec{x})\, d\vec{x}.$$

ii. Given any bounded set $B \subset \mathbb{R}^{m-1}$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{\vec{n}\in[N]^m}^{*} \chi_B(N(((x_{n_1}-x_{n_2})),\ldots,((x_{n_1}-x_{n_m})))) = vol(B) = \int_{\mathbb{R}^{m-1}} \chi_B(\vec{x})\,d\vec{x}.$$

Proof.
Consider the forward implication. We first claim that given any bounded measurable set $B \subset \mathbb{R}^{m-1}$, and any $\varepsilon > 0$, there exist continuous functions $f, g : \mathbb{R}^{m-1} \to \mathbb{R}$ with compact support such that:

a. Given any $\vec{x} \in \mathbb{R}^{m-1}$,
$$f(\vec{x}) \leq \chi_B(\vec{x}) \leq g(\vec{x}),$$

b.
$$\int_{\mathbb{R}^{m-1}} (g(\vec{x}) - f(\vec{x}))\,d\vec{x} < \varepsilon.$$

To prove the same, we define two sets

$$B_\delta := \{\vec{x} \in B : \operatorname{dist}(\vec{x}, \partial B) > \delta\} \quad \text{and} \quad B^\delta := B \cup \left\{\vec{x} \in \mathbb{R}^{m-1} : \operatorname{dist}(\vec{x}, \partial B) < \delta\right\},$$

where, $\delta > 0$. Now, we choose a $\delta > 0$ such that

$$\operatorname{vol}(B \setminus B_\delta) < \frac{\varepsilon}{2} \quad \text{and} \quad \operatorname{vol}\left(B^\delta \setminus B\right) < \frac{\varepsilon}{2}.$$

Next, we define the functions f and g.

$$f(\vec{x}) := \begin{cases} 1 & \text{if } \vec{x} \in B_\delta, \\ 0 & \text{if } \vec{x} \in \mathbb{R}^{m-1} \setminus B_{\frac{\delta}{2}}, \end{cases}$$

such that f is continuous in $\mathbb{R}^{m-1}$, and $0 \leq f(\vec{x}) \leq 1$ for all $\vec{x} \in \mathbb{R}^{m-1}$. Also,

$$g(\vec{x}) := \begin{cases} 1 & \text{if } \vec{x} \in B, \\ 0 & \text{if } \vec{x} \in \mathbb{R}^{m-1} \setminus B^\delta, \end{cases}$$

such that g is continuous in $\mathbb{R}^{m-1}$, and $0 \leq g(\vec{x}) \leq 1$ for all $\vec{x} \in \mathbb{R}^{m-1}$.
Clearly, we see that
$$f(\vec{x}) \leq \chi_B(\vec{x}) \leq g(\vec{x}) \;\forall\; \vec{x} \in \mathbb{R}^{m-1}.$$

Also,

$$\int_{\mathbb{R}^{m-1}} (g(\vec{x}) - f(\vec{x}))\, d\vec{x} = \int_{B^\delta \setminus B_\delta} (g(\vec{x}) - f(\vec{x}))\, d\vec{x} \le \int_{B^\delta \setminus B_\delta} d\vec{x} = \mathrm{vol}\left(B^\delta \setminus B_\delta\right) < \varepsilon.$$

This proves our claim.

Now, given any natural number N, we have

$$\begin{aligned}\frac{1}{N}\sum_{\vec{n}\in[N]^m}^{*} f(N(((x_{n_1}-x_{n_2})),\dots,((x_{n_1}-x_{n_m})))) &\le \frac{1}{N}\sum_{\vec{n}\in[N]^m}^{*} \chi_B(N(((x_{n_1}-x_{n_2})),\dots,((x_{n_1}-x_{n_m}))))\\ &\le \frac{1}{N}\sum_{\vec{n}\in[N]^m}^{*} g(N(((x_{n_1}-x_{n_2})),\dots,((x_{n_1}-x_{n_m}))))\end{aligned}$$

Since this is true for all natural numbers N, we can take the limit as $N \to \infty$ to get

$$\begin{aligned}\int_{\mathbb{R}^{m-1}} f(\vec{x})d\vec{x} &\le \liminf_{N\to\infty}\frac{1}{N}\sum_{\vec{n}\in[N]^m}^{*} \chi_B(N(((x_{n_1}-x_{n_2})),\dots,((x_{n_1}-x_{n_m}))))\\ &\le \limsup_{N\to\infty}\frac{1}{N}\sum_{\vec{n}\in[N]^m}^{*} \chi_B(N(((x_{n_1}-x_{n_2})),\dots,((x_{n_1}-x_{n_m})))) \le \int_{\mathbb{R}^{m-1}} g(\vec{x})d\vec{x}.\end{aligned}$$

Thus,

$$\begin{aligned}&\int_{\mathbb{R}^{m-1}} f(\vec{x})d\vec{x} - \int_{\mathbb{R}^{m-1}} \chi_B(\vec{x})d\vec{x}\\ &\le \liminf_{N\to\infty}\frac{1}{N}\sum_{\vec{n}\in[N]^m}^{*} \chi_B(N(((x_{n_1}-x_{n_2})),\dots,((x_{n_1}-x_{n_m})))) - \int_{\mathbb{R}^{m-1}} \chi_B(\vec{x})d\vec{x}\\ &\le \limsup_{N\to\infty}\frac{1}{N}\sum_{\vec{n}\in[N]^m}^{*} \chi_B(N(((x_{n_1}-x_{n_2})),\dots,((x_{n_1}-x_{n_m})))) - \int_{\mathbb{R}^{m-1}} \chi_B(\vec{x})d\vec{x}\\ &\le \int_{\mathbb{R}^{m-1}} g(\vec{x})d\vec{x} - \int_{\mathbb{R}^{m-1}} \chi_B(\vec{x})d\vec{x},\end{aligned}$$

which means,

$$\begin{aligned}-\varepsilon &\le \liminf_{N\to\infty}\frac{1}{N}\sum_{\vec{n}\in[N]^m}^{*} \chi_B(N(((x_{n_1}-x_{n_2})),\dots,((x_{n_1}-x_{n_m})))) - \int_{\mathbb{R}^{m-1}} \chi_B(\vec{x})d\vec{x}\\ &\le \limsup_{N\to\infty}\frac{1}{N}\sum_{\vec{n}\in[N]^m}^{*} \chi_B(N(((x_{n_1}-x_{n_2})),\dots,((x_{n_1}-x_{n_m})))) - \int_{\mathbb{R}^{m-1}} \chi_B(\vec{x})d\vec{x} \le \varepsilon.\end{aligned}$$

This is true for all $\varepsilon > 0$. Therefore, taking $\varepsilon \to 0$, we get

$$\lim_{N\to\infty} \frac{1}{N} \sum_{\vec{n}\in[N]^m}^{*} \chi_B(N(((x_{n_1} - x_{n_2})), \ldots, ((x_{n_1} - x_{n_m})))) = \int_{\mathbb{R}^{m-1}} \chi_B(\vec{x}) d\vec{x}$$

For the converse, consider any continuous function $g : \mathbb{R}^{m-1} \to \mathbb{R}$ with compact support. This means that the image of g is bounded in $\mathbb{R}$. Thus, there exists some positive integer A such that $\text{Image}(g) \subset [-A, A]$. Divide the interval into n equal parts. This can be done using the partition

$$\left\{ -A, -A + \frac{2A}{n}, -A + \frac{4A}{n}, \ldots, -A + \frac{2(n-1)A}{n}, A \right\}.$$

Now, for $1 \le k \le n$, define the set

$$B_k := \left\{ \vec{x} \in \mathbb{R}^{m-1} : -A + \frac{2(k-1)A}{n} < g(\vec{x}) < -A + \frac{2kA}{n} \right\} \cap B.$$

Now, define the simple functions

$$g^-(\vec{x}) = -A\chi_{\{\vec{x}:g(\vec{x})=-A\}}(\vec{x}) + \sum_{k=1}^{n} \left(-A + \frac{2(k-1)A}{n} \right) \chi_{B_k}(\vec{x}) + A\chi_{\{\vec{x}:g(\vec{x})=A\}}(\vec{x}),$$

and

$$g^+(\vec{x}) = -A\chi_{\{\vec{x}:g(\vec{x})=-A\}}(\vec{x}) + \sum_{k=1}^{n} \left(-A + \frac{2kA}{n} \right) \chi_{B_k}(\vec{x}) + A\chi_{\{\vec{x}:g(\vec{x})=A\}}(\vec{x}).$$

Clearly,

$$g^-(\vec{x}) \le g(\vec{x}) \le g^+(\vec{x}) \ \forall \ \vec{x} \in \mathbb{R}^{m-1},$$

and

$$\int_{\mathbb{R}^{m-1}} \left(g^+(\vec{x}) - g^-(\vec{x}) \right) d\vec{x} < \frac{2A}{n} \text{vol}(B).$$

Since A is a finite number and B is a bounded set, therefore, given any $\varepsilon > 0$, there exists some $n_0 \in \mathbb{N}$ such that

$$\int_{\mathbb{R}^{m-1}} \left(g^+(\vec{x}) - g^-(\vec{x}) \right) d\vec{x} < \frac{2A}{n} \text{vol}(B) < \varepsilon \ \forall \ n \ge n_0.$$

Using the same limit argument as in the first part, we prove the converse. This completes the proof. $\square$

The following theorem tells us that the order in which we take the differences does not matter.

Before going to the theorem, let us look at a required useful lemma, which will be useful in other parts of the report as well.

Lemma 4.2.3. *Let x and y be real numbers in some unit interval $\mathbb{T}$ such that*

$$||x|| + ||y|| < \frac{1}{2}.$$

Then, we have that

$$((x)) + ((y)) = ((x+y)).$$

Extending the same to finitely many points, we have that for some fixed natural number k, if $x_1, \ldots, x_k \in \mathbb{T}$ are such that $\sum_{i=1}^{k} ||x_i|| < \frac{1}{2}$, then

$$\sum_{i=1}^{k} ((x_i)) = \left(\left(\sum_{i=1}^{k} x_i\right)\right).$$

Proof.
Since $||x|| + ||y|| < \frac{1}{2}$, we cannot have both $||x||$ and $||y||$ to be simultaneously greater than $\frac{1}{4}$. Let $x, y \in \left[-\frac{1}{2}, \frac{1}{2}\right)$. We first consider the case of both $||x||, ||y|| < \frac{1}{4}$. Then, $x, y \in \left(-\frac{1}{4}, \frac{1}{4}\right)$. Thus, $((x)) = x$ and $((y)) = y$, and $x + y \in \left(-\frac{1}{2}, \frac{1}{2}\right)$. Thus, $((x+y)) = x + y$. Hence,

$$((x)) + ((y)) = ((x+y)).$$

Now, consider the case when, without loss of generality, $||x|| \geq \frac{1}{4}$. Let $||x|| = \frac{1}{4} + \varepsilon$, for some $0 \leq \varepsilon < \frac{1}{4}$. Clearly, $||y|| < \frac{1}{4} - \varepsilon$. Again, $((x)) = x$ and $((y)) = y$, and $x + y \in \left(-\frac{1}{2}, \frac{1}{2}\right)$, and hence, $((x+y)) = x + y$. Therefore,

$$((x)) + ((y)) = ((x+y)).$$

This completes the proof of the lemma. □

Theorem 4.2.4. *Fix a natural number $m \geq 2$. Then, the following are equivalent:*

i. Given any $g \in C_c(\mathbb{R}^{m-1})$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{\vec{n}\in[N]^m}^{*} g(N(((x_{n_1}-x_{n_2})),\ldots,((x_{n_1}-x_{n_m})))) = \int_{\mathbb{R}^{m-1}} g(\vec{x})\,d\vec{x}.$$

ii. Given any $g \in C_c(\mathbb{R}^{m-1})$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{\vec{n}\in[N]^m}^{*} g(N(((x_{n_1}-x_{n_2})),\ldots,((x_{n_{m-1}}-x_{n_m})))) = \int_{\mathbb{R}^{m-1}} g(\vec{x})\,d\vec{x}.$$

Proof.
Let us try to prove (i) implies (ii). Consider a function $g \in C_c(\mathbb{R}^{m-1})$. Set

$$f(x_1,x_2,\ldots,x_{m-1}) = g(x_1,x_2-x_1,x_3-x_2,\ldots,x_{m-1}-x_{m-2}).$$

Clearly, $f \in C_c(\mathbb{R}^{m-1})$ and

$$\int_{\mathbb{R}^{m-1}} f(\vec{x})\,d\vec{x} = \int_{\mathbb{R}^{m-1}} g(\vec{x})\,d\vec{x}.$$

Also note that

$$g(x_1,x_2,\ldots,x_{m-1}) = f(x_1,x_1+x_2,x_1+x_2+x_3,\ldots,x_1+\cdots+x_{m-1}).$$

Since g has a compact support, therefore, there exists some large enough natural number N such that $\mathrm{supp}(g) \subseteq \left[-\frac{N}{2m},\frac{N}{2m}\right]^{m-1}$. Now consider, for some tuple $\vec{n} \in [N]^m$,

$$\begin{aligned}
&g\left(N\left(((x_{n_1}-x_{n_2})),\ldots,((x_{n_{m-1}}-x_{n_m}))\right)\right)\\
=&f\left(N\left(((x_{n_1}-x_{n_2})),((x_{n_1}-x_{n_2}))+((x_{n_2}-x_{n_3})),\ldots,((x_{n_1}-x_{n_2}))+\cdots+((x_{n_{m-1}}-x_{n_m}))\right)\right)\\
\neq&0.
\end{aligned}$$

Thus,

$$N\left(((x_{n_1}-x_{n_2})),\ldots,((x_{n_{m-1}}-x_{n_m}))\right) \in \mathrm{supp}(g) \subseteq \left[-\frac{N}{2m},\frac{N}{2m}\right]^{m-1},$$

which means that for every $1 \le i < m$,

$$||x_{n_i}-x_{n_{i+1}}|| < \frac{1}{2m},$$

and hence,

$$\sum_{i=1}^{k-1} ||x_{n_i} - x_{n_{i+1}}|| < \frac{k}{2m} < \frac{1}{2} \ \forall \ 2 \leq k \leq m.$$

Thus, using Lemma 4.2.3, we get, for any $2 \leq k \leq m$,

$$\sum_{i=1}^{k-1} ((x_{n_i} - x_{n_{i+1}})) = ((x_{n_1} - x_{n_k})),$$

and finally,

$$g\left(N\left(((x_{n_1} - x_{n_2})), \ldots, ((x_{n_{m-1}} - x_{n_m}))\right)\right) = f\left(N\left(((x_{n_1} - x_{n_2})), \ldots, ((x_{n_1} - x_{n_m}))\right)\right).$$

Combining everything, we have, for a large enough natural number N,

$$\sum_{\vec{n}\in[N]^m}^{*} g\left(N\left(((x_{n_1} - x_{n_2})), \ldots, ((x_{n_{m-1}} - x_{n_m}))\right)\right) = \sum_{\vec{n}\in[N]^m}^{*} f\left(N\left(((x_{n_1} - x_{n_2})), \ldots, ((x_{n_1} - x_{n_m}))\right)\right).$$

Therefore,

$$\begin{aligned}
&\lim_{N\to\infty} \frac{1}{N} \sum_{\vec{n}\in[N]^m}^{*} g\left(N\left(((x_{n_1} - x_{n_2})), \ldots, ((x_{n_{m-1}} - x_{n_m}))\right)\right) \\
&= \lim_{N\to\infty} \frac{1}{N} \sum_{\vec{n}\in[N]^m}^{*} f\left(N\left(((x_{n_1} - x_{n_2})), \ldots, ((x_{n_1} - x_{n_m}))\right)\right) \\
&= \int_{\mathbb{R}^{m-1}} f(\vec{x})\, d\vec{x} \\
&= \int_{\mathbb{R}^{m-1}} g(\vec{x})\, d\vec{x}.
\end{aligned}$$

This proves the forward implication.

For the converse, we take any function $g \in C_c(\mathbb{R}^{m-1})$, and set

$$f(x_1, x_2, \ldots, x_{m-1}) = g(x_1, x_1 + x_2, x_1 + x_2 + x_3, \ldots, x_1 + \cdots + x_{m-1}).$$

This means that

$$g(x_1, x_2, \ldots, x_{m-1}) = f(x_1, x_2 - x_1, x_3 - x_2, \ldots, x_{m-1} - x_{m-2}).$$

Now, proceeding the same way as above, we get the converse. $\square$

Remark 4.2.1. *Using Theorems 4.2.2 and 4.2.4, we can conclude that for any natural number $m \geq 2$ and given any bounded set $B \subset \mathbb{R}^{m-1}$, the following are equivalent:*

i.

$$\lim_{N\to\infty} \frac{1}{N} \sum_{\vec{n}\in[N]^m}^{*} \chi_B(N(((x_{n_1} - x_{n_2})), \ldots, ((x_{n_1} - x_{n_m})))) = vol(B).$$

ii.

$$\lim_{N\to\infty} \frac{1}{N} \sum_{\vec{n}\in[N]^m}^{*} \chi_B(N(((x_{n_1} - x_{n_2})), \ldots, ((x_{n_{m-1}} - x_{n_m})))) = vol(B).$$

Chapter 5

Recovering the Level-Spacing Distribution Function

The previous chapters have introduced the theory of spacing statistics, and what it means for a statistic to be 'Poissonian'. We have also discussed the difficulty of computing the level-spacing distribution function, and how recovering it from known values of the correlation functions would be easier.
If all the correlation functions of a sequence are known, and are known to be *Poissonian*, then, this chapter focuses on actually recovering the level-spacing distribution function of the sequence. In particular, we would like to prove the following theorem.

Theorem 5.0.1. *Let $X = (x_n)_{n\in\mathbb{N}}$ be a sequence in $[0,1)$. If X has Poissonian correlations of all orders $m \geq 2$, then the sequence has the Poissonian level-spacing distribution function.*

We will follow the general argument used in [KR99, Appendix A], and slightly tweak it for our case to prove the theorem.

As before, $X = (x_n)_{n\in\mathbb{N}}$ will denote a sequence in $[0,1)$. Also assume that the sequence has Poissonian m-level correlations for all intergers $m \geq 2$. That is, for any integer $m \geq 2$, and given any bounded set $B \subset \mathbb{R}^{m-1}$ and a large enough natural number N, we have by Remark 4.2.1,

$$
\begin{aligned}
&R_{m,X,N}(B) \\
&= \frac{1}{N}\#\left\{\vec{n} \in [N]^m : n_1, n_2, \ldots, n_m \text{ are all distinct, } \big(((x_{n_1} - x_{n_2})), \ldots, ((x_{n_{m-1}} - x_{n_m}))\big) \in \frac{1}{N}B\right\},
\end{aligned}
$$

and the limit as $N \to \infty$ exists and is given by

$$\lim_{N\to\infty} R_{m,X,N}(B) = \text{vol}(B).$$

5.1 Some Basic Terminology and Results

For the purpose of this section, we will fix a natural number N, and consider the first N terms of the sequence X. (We will let $N \to \infty$ after building up sufficient theory.) Denote the set of these terms by

$$S_N := \{x_1, \ldots, x_N : x_i \in X, 1 \le i \le N\}.$$

We can order these elements as:

$$S_N = \left\{x_{1,N} \le x_{2,N} \le \cdots \le x_{N,N}\right\}$$

Now, we will look at these terms as plotted on the 2-dimensional circle centered at the origin with *circumference* 1, denoted by $\mathbb{T}$. This can be done using any isomorphism. In particular, we shall use the map

$$x \mapsto \tilde{x} = \frac{1}{2\pi}e^{2\pi i x}.$$

Since we are going from the unit interval to the circle, we need to define a new ordering on $\mathbb{T}$:

Definition 5.1.1 (A New Ordering). *Let $x, y \in [0,1)$ and let $\tilde{x}, \tilde{y} \in \mathbb{T}$ be their respective images on the circle. We say that*

*i. $\tilde{x}$ **succeeds** $\tilde{y}$ on $\mathbb{T}$, denoted by $\tilde{x} \succ \tilde{y}$ if and only if $((x-y)) > 0$ in $[0,1)$.*

*ii. $\tilde{x}$ **precedes** $\tilde{y}$ on $\mathbb{T}$, denoted by $\tilde{x} \prec \tilde{y}$ if and only if $((x-y)) < 0$ in $[0,1)$.*

iii. $\tilde{x} = \tilde{y}$ on $\mathbb{T}$ if and only if $((x-y)) = 0$, that is, $x = y$ in $[0,1)$.

Remark 5.1.1. *A question that arises here is, why do we need a new ordering, and why go to the* *circle in the first place? This is done with the sole purpose of reducing the number of cases we need to deal with, thereby, reducing the computational hassle.*

In the unit interval $[0,1)$*, when we say* $((a-b)) > 0$*, we have to deal with two cases, one where* $a > b$ *but not so big that the parity flips, the other when* $a < b$*, where* a *is much smaller than* b*. Let us illustrate this with an example. Consider* $a = 0.7$ *and* $b = 0.1$*.* $a > b$*, but it is so big that* $((0.7-0.1)) = ((0.6)) = -0.1 < 0$*. On the other hand, for* $a = 0.2$ *and* $b = 0.4$*, even though* $b > a$*, they are so close that* $((0.2-0.4)) = -0.2 < 0$*. Thus, multiple cases need to be dealt with.*

However, on the circle (with unit circumference), $a \succ b$ *essentially means that* a *and* b *lie on an arc of length less than* $\frac{1}{2}$ *such that the arc begins at* b *and is oriented in the counter-clockwise direction. This reduces everything to a single case. Thus, we can easily see that* $\tilde{0.1} \succ \tilde{0.9}$ *but* $\tilde{0.7} \not\succ \tilde{0.1}$ *and* $\tilde{0.2} \not\succ \tilde{0.4}$*.*

Next, we denote, by T, any subset of S_N such that all the points in T lie in an arc of length less than $\frac{1}{2}$. Clearly, we can order these elements, so that T now has a (not necessarily unique) initial and final element. Thus, T can be written as

$$T = \left\{ y_{\text{init}} = y_1 \preceq y_2 \preceq \cdots \preceq y_{\text{fin}} : y_i \in S_N \text{ for all such } i,\ ((y_{\text{fin}} - y_{\text{init}})) < \frac{1}{2} \right\}.$$

We denote the **number of elements** in T by $|T|$. We define the **diameter** of T as

$$\operatorname{diam}(T) := ((y_{\text{fin}} - y_{\text{init}})).$$

T is said to be a **consecutive pair** if it contains only two points, the initial and the final, and these points are consecutive points in S_N, that is, there is no point in S_N between the two points in T.

Similarly, T is said to be a **consecutive k-tuple** of S_N if

$$T = \{y_{\text{init}} = y_1 \preceq y_2 \preceq \cdots \preceq y_k = y_{\text{fin}}\}$$

such that there are no point of S_N between y_i and y_{i+1} for any $1 \le i \le k-1$.

Now, let us define two important terms. Given any $l < \dfrac{1}{2}$,

i. define $\mathcal{N}_m(l)$ to be the number of m-tuples T of diameter less than or equal to l. Note that

$\mathscr{N}_m(l)$ counts the number of *all* the m-tuples of diameter less than or equal to l, and not just consecutive tuples.

ii. define $g(l)$ to be the number of consecutive pairs T of diameter less than or equal to l, that is, the number of spacings between consecutive elements of S_N of length less than or equal to l.

Remark 5.1.2. *Note that*

$$g(l) = \#\left\{1 \leq n \leq N : ((\tilde{x}_{n+1,N} - \tilde{x}_{n,N})) \leq l\right\},$$

where we set $\tilde{x}_{N+1,N} = \tilde{x}_{1,N}$. *Thus,*

$$\frac{1}{N} g\left(\frac{l}{N}\right) = \#\left\{1 \leq n \leq N : ((\tilde{x}_{n+1,N} - \tilde{x}_{n,N})) \leq \frac{l}{N}\right\}$$

is exactly the form of the level spacing distribution function defined in Definition 3.1.1, considering l *to be the length of the interval* $[a,b]$. *Hence, if the limit*

$$\lim_{N\to\infty} \frac{1}{N} g\left(\frac{l}{N}\right) = \#\left\{1 \leq n \leq N : ((\tilde{x}_{n+1,N} - \tilde{x}_{n,N})) \leq \frac{l}{N}\right\}$$

exists for any positive l, *then we can conclude that the sequence has level-spacing distribution function. Moreover, if the limit equals* $1 - e^{-l}$, *then we say that the distribution function is Poissonian.*

We are interested in finding the relation between these two terms. This will be done in Section 5.3. For now, we will look at a combinatorial result necessary to derive the said relation.

Lemma 5.1.1. *Let* $m \geq 0$ *be an integer. Then,*

$$\sum_{i=0}^{m} (-1)^i \binom{m}{i} = \begin{cases} 0, & \text{if } m \neq 0 \\ 1, & \text{if } m = 0. \end{cases}$$

Moreover,

$$\sum_{i=0}^{2n+1} (-1)^i \binom{m}{i} \leq \sum_{i=0}^{m} (-1)^i \binom{m}{i} \leq \sum_{i=0}^{2n} (-1)^i \binom{m}{i} \ \forall\ n \in \mathbb{N}.$$

Proof.

In the first part, for $m = 0$, it can be clearly seen that the sum becomes 1. For $m \in \mathbb{N}$, the summation

is nothing but the binomial expansion of $(1-1)^m$, and hence, is 0.
For the second part, we make use of the standard identity

$$\binom{m}{i} = \binom{m-1}{i} + \binom{m-1}{i-1}.$$

Thus, we have, for some $k \in \mathbb{N}$,

$$\begin{aligned}
\sum_{i=0}^{k}(-1)^i\binom{m}{i} &= \sum_{i=0}^{k}(-1)^i\binom{m-1}{i} + \sum_{i=0}^{k}(-1)^i\binom{m-1}{i-1} \\
&= \sum_{i=0}^{k}(-1)^i\binom{m-1}{i} + \sum_{j=0}^{k-1}(-1)^{j+1}\binom{m-1}{j} \\
&= (-1)^k\binom{m-1}{k}.
\end{aligned}$$

Here, we have used the assumption that $\binom{m-1}{-1} = 0$.

Thus, depending on whether k is even or odd, $(-1)^k\binom{m-1}{k}$ is either positive or negative respectively. This proves the inequality. $\square$

Remark 5.1.3. *Assuming that* $\binom{m}{i} = 0 \ \forall\ i > m$*, we can convert the finite sum* $\sum_{i=0}^{m}(-1)^i\binom{m}{i}$ *into an infinite sum, since all terms from* $i = m+1$ *will be zero. Thus,*

$$\sum_{i=0}^{\infty}(-1)^i\binom{m}{i} = \begin{cases} 0, & \text{if } m \neq 0 \\ 1, & \text{if } m = 0. \end{cases}$$

Similarly,

$$\sum_{i=0}^{2n+1}(-1)^i\binom{m}{i} \leq \sum_{i=0}^{\infty}(-1)^i\binom{m}{i} \leq \sum_{i=0}^{2n}(-1)^i\binom{m}{i} \ \forall\ n \in \mathbb{N}.$$

5.2 The Standard Open Simplex

This section introduces us to a very special bounded convex-open set — the 'standard open simplex'. This set is defined as follows:

Definition 5.2.1 (Standard Open Simplex). *Fix any natural number* $m \geq 2$*. The* $(m-1)$*-th di-*

mensional standard open simplex $\Delta^{m-1} \subset \mathbb{R}^{m-1}$ *is defined as*

$$\Delta^{m-1} := \left\{ (x_1, \ldots, x_{m-1}) : x_n > 0 \, \forall \, 1 \leq n \leq m-1 \textit{ and } \sum_{n=1}^{m-1} x_n < 1 \right\}.$$

Now, for any $l > 0$, set $B = l\Delta^{m-1}$ as the scaled standard open simplex. Thus,

$$B = \left\{ (x_1, \ldots, x_{m-1}) : x_n > 0 \, \forall \, 1 \leq n \leq m-1 \text{ and } \sum_{n=1}^{m-1} x_i < l \right\}.$$

Before going further, let us compute the ($(m-1)$-th dimensional) volume of a scaled standard open simplex. This would be useful later in the section.

Lemma 5.2.1. *For any positive real number l, and any natural number $m \geq 2$, the $(m-1)$-th dimensional volume of the scaled standard open simplex is given by:*

$$vol\left(l\Delta^{m-1}\right) = \frac{l^{m-1}}{(m-1)!}.$$

Proof.
Recall the definition of the simplex:

$$l\Delta^{m-1} = \left\{ (x_1, \ldots, x_{m-1}) : x_n > 0 \, \forall \, 1 \leq n \leq m-1 \text{ and } \sum_{n=1}^{m-1} x_i < l \right\}.$$

We need to integrate over the $(m-1)$ variables with the given constraints in mind to find the volume. Thus,

$$\begin{aligned}
&\int \cdots \int_{l\Delta^{m-1}} dx_1 \cdots dx_{m-1} \\
&= \int_{x_1=0}^{l} \int_{x_2=0}^{l-x_1} \int_{x_3=0}^{l-x_1-x_2} \cdots \int_{x_{m-1}=0}^{l-x_1-x_2-\cdots-x_{m-2}} dx_{m-1} \cdots dx_3 dx_2 dx_1 \\
&= \int_{x_1=0}^{l} \int_{x_2=0}^{l-x_1} \int_{x_3=0}^{l-x_1-x_2} \cdots \int_{x_{m-2}=0}^{l-x_1-x_2-\cdots-x_{m-3}} (l - x_1 - x_2 - \cdots - x_{m-3} - x_{m-2})\, dx_{m-2} \cdots dx_3 dx_2 dx_1
\end{aligned}$$

$$=\int_{x_1=0}^{l}\int_{x_2=0}^{l-x_1}\int_{x_3=0}^{l-x_1-x_2}\cdots$$
$$\left(\int_{x_{m-2}=0}^{l-x_1-x_2-\cdots-x_{m-3}}(l-x_1-x_2-\cdots-x_{m-3})\,dx_{m-2}-\int_{x_{m-2}=0}^{l-x_1-x_2-\cdots-x_{m-3}}x_{m-2}\,dx_{m-2}\right)dx_{m-2}\cdots dx_3dx_2dx_1$$
$$=\int_{x_1=0}^{l}\int_{x_2=0}^{l-x_1}\int_{x_3=0}^{l-x_1-x_2}\cdots\int_{x_{m-3}=0}^{l-x_1-x_2-\cdots-x_{m-4}}\frac{1}{2}(l-x_1-x_2-\cdots-x_{m-3})^2\,dx_{m-3}\cdots dx_3dx_2dx_1$$

We can split

$$(l-x_1-x_2-\cdots-x_{m-3})^2=(l-x_1-x_2-\cdots-x_{m-4})^2-2(l-x_1-x_2-\cdots-x_{m-4})x_{m-3}+x_{m-3}^2,$$

and integrate over x_{m-3} to get

$$\frac{1}{3}(l-x_1-x_2-\cdots-x_{m-4})^3.$$

Continuing the process, we are finally left with

$$\frac{1}{(m-2)!}\int_{x_1=0}^{l}(l-x_1)^{m-2}dx_1=\frac{l^{m-1}}{(m-1)!}.$$

Therefore,

$$\operatorname{vol}\left(l\Delta^{m-1}\right)=\frac{l^{m-1}}{(m-1)!},$$

and our proof is complete. □

Having defined the simplex (which is a bounded set), we can compute the m-level correlation sum of this set.

Fix any natural number $N>2l$. This means that $\frac{l}{N}<\frac{1}{2}$ and $\frac{1}{N}B=\frac{l}{N}\Delta^{m-1}\subsetneq\frac{1}{2}\Delta^{m-1}$.

The m-level correlation sum would be

$$R_{m,X,N}(B)$$
$$=\frac{1}{N}\#\left\{\vec{n}\in[N]^m:n_1,n_2,\ldots,n_m\text{ are all distinct, }\left(((x_{n_1}-x_{n_2})),\ldots,((x_{n_{m-1}}-x_{n_m}))\right)\in\frac{1}{N}B\right\}$$
$$=\frac{1}{N}\#\left\{\vec{n}\in[N]^m:n_1,n_2,\ldots,n_m\text{ are all distinct, }\left(((x_{n_1}-x_{n_2})),\ldots,((x_{n_{m-1}}-x_{n_m}))\right)\in\frac{l}{N}\Delta^{m-1}\right\}.$$

Let us now see what it means for a tuple to be in the scaled simplex. The tuple

$$\left(((x_{n_1}-x_{n_2})),\ldots,((x_{n_{m-1}}-x_{n_m}))\right)\in\frac{l}{N}\Delta^{m-1}$$

means that

i. $((x_{n_i}-x_{n_{i+1}}))>0$ for every $1\leq i\leq m-1$, which means that $x_{n_1}\succ\cdots\succ x_{n_m}$.

ii. $\sum_{i=1}^{m-1}((x_{n_i}-x_{n_{i+1}}))<\frac{l}{N}<\frac{1}{2}$, which means $\sum_{i=1}^{m-1}((x_{n_i}-x_{n_{i+1}}))=((x_{n_1}-x_{n_m}))<\frac{l}{N}$, where we have used Lemma 4.2.3.

Thus, combining everything, we see that, for any $l>0$ and any large enough natural number N such that $N>2l$,

$$\begin{aligned}&R_{m,X,N}\left(l\Delta^{m-1}\right)\\ =&\frac{1}{N}\#\left\{\vec{n}\in[N]^m : n_1,n_2,\ldots,n_m \text{ are all distinct, } x_{n_1}\succ\cdots\succ x_{n_m}, \text{ and they lie in an arc of length } <\frac{l}{N}\right\}\\ =&\frac{1}{N}\mathcal{N}_m\left(\frac{l}{N}\right).\end{aligned}$$

But note that we have also assumed the correlations to be Poissonian (recall Definition 3.3.3). Thus, by the definition, and by Lemma 5.2.1,

$$\lim_{N\to\infty}R_{m,X,N}\left(l\Delta^{m-1}\right)=\operatorname{vol}\left(l\Delta^{m-1}\right)=\frac{l^{m-1}}{(m-1)!}.$$

Finally, we have

$$\lim_{N\to\infty}\frac{1}{N}\mathcal{N}_m\left(\frac{l}{N}\right)=\frac{l^{m-1}}{(m-1)!}. \tag{5.1}$$

5.3 The Recovery

This brings us to the main section, where we piece together the results derived in the previous sections to finally prove Theorem 5.0.1.

As promised in Section 5.1, we shall begin the section by deriving a relation between $\mathcal{N}_m(l)$ and $g(l)$ for any $l>0$. It turns out that $g(l)$ can be written as an alternating series sum over $\mathcal{N}_m(l)$'s.

Theorem 5.3.1. *Given any $l < \frac{1}{2}$, we have*

$$g(l) = \sum_{m\geq 2} (-1)^m \mathcal{N}_m(l).$$

Moreover, for all $n \in \mathbb{N}$, we have

$$\sum_{m=2}^{2n+1} (-1)^m \mathcal{N}_m(l) \leq g(l) \leq \sum_{m=2}^{2n} (-1)^m \mathcal{N}_m(l). \tag{5.2}$$

Proof.
Consider any pair $T = \{a \prec b\}$ in S_N with diameter $l < \frac{1}{2}$. Note that this need not be a consecutive pair.
Now, let X_T be the set of all the tuples $a = y_1 \prec y_2 \prec \cdots \prec b$ in S_N whose initial element is a and final element is b.
Let $N_{m,T}$ be the number of m-tuples in X_T, and define $N_m = \sum_T N_{m,T}$, which is the number of all m-tuples in S_N with diameter l. Thus, $N_m \equiv \mathcal{N}_m(l)$.
By $\tilde{T}$, denote the consecutive tuple, $a = y_1 \prec \cdots \prec b$, that has initial element a and final element b.
Now, any m-tuple in X_T will begin at a, end at b, and have $m-2$ other elements, all from $\tilde{T}$. Thus, we can calculate the number of such m-tuples by choosing $m-2$ elements from $|\tilde{T}|-2$ elements, which means:

$$N_{m,T} = \binom{|\tilde{T}|-2}{k-2}.$$

Thus, we can apply Lemma 5.1.1 along with Remark 5.1.3 to get

$$\sum_{m\geq 2} (-1)^m N_{m,T} = \sum_{m\geq 2} (-1)^m \binom{|\tilde{T}|-2}{m-2} = \begin{cases} 1 & \text{if } |\tilde{T}|-2 = 0 \text{ i.e } |\tilde{T}| = 2 \text{ i.e. } T \text{ is a consecutive pair,} \\ 0 & \text{otherwise.} \end{cases}$$

Now, summing $\sum_{m\geq 2} (-1)^m N_{m,T}$ over all pairs $T = \{a \prec b\}$ with diameter $l < \frac{1}{2}$, we get

$$\sum_T \sum_{m\geq 2} (-1)^m N_{m,T} = \sum_{m\geq 2} (-1)^m \sum_T N_{m,T} = \sum_{m\geq 2} (-1)^m N_m = \sum_{m\geq 2} (-1)^m \mathcal{N}_m(l).$$

Evaluating the sum in another way, we have:

$$\sum_T \sum_{m\geq 2} (-1)^m N_{m,T} = \sum_{T \text{ is a consecutive pair}} 1 = g(l).$$

Finally, equating these, we get:

$$g(l) = \sum_{m\geq 2} (-1)^m \mathcal{N}_m(l).$$

Now, using the second part of Lemma 5.1.1, we have

$$\begin{aligned}
&\sum_{m=2}^{2n+1} (-1)^m \binom{|\tilde{T}|-2}{m-2} \leq \sum_{m\geq 2} (-1)^m \binom{|\tilde{T}|-2}{m-2} \leq \sum_{m=2}^{2n} (-1)^m \binom{|\tilde{T}|-2}{m-2} \ \forall\ n \in \mathbb{N} \\
\Longrightarrow &\sum_T \sum_{m=2}^{2n+1} (-1)^m N_{m,T} \leq \sum_T \sum_{m\geq 2} (-1)^m N_{m,T} \leq \sum_T \sum_{m=2}^{2n} (-1)^m N_{m,T} \ \forall\ n \in \mathbb{N} \\
\Longrightarrow &\sum_{m=2}^{2n+1} (-1)^m \mathcal{N}_m(l) \leq g(l) \leq \sum_{m=2}^{2n} (-1)^m \mathcal{N}_m(l) \ \forall\ n \in \mathbb{N}
\end{aligned}$$

This completes the proof of the theorem. □

By Equation 5.2, we have, for any $n \in \mathbb{N}$,

$$\sum_{m=2}^{2n+1} (-1)^m \mathcal{N}_m(l) \leq g(l) \leq \sum_{m=2}^{2n} (-1)^m \mathcal{N}_m(l),$$

which means that for any natural numbers N and n,

$$\frac{1}{N} \sum_{m=2}^{2n+1} (-1)^m \mathcal{N}_m(l) \leq \frac{1}{N} g(l) \leq \frac{1}{N} \sum_{m=2}^{2n} (-1)^m \mathcal{N}_m(l),$$

and pushing the factor of N inside the summation, we get, for any natural numbers N and n,

$$\sum_{m=2}^{2n+1} (-1)^m \frac{1}{N} \mathcal{N}_m(l) \leq \frac{1}{N} g(l) \leq \sum_{m=2}^{2n} (-1)^m \frac{1}{N} \mathcal{N}_m(l).$$

Now, note that for any $n \in \mathbb{N}$,

$$\liminf_{N\to\infty} \sum_{m=2}^{2n+1} (-1)^m \frac{1}{N} \mathcal{N}_m(l) \leq \liminf_{N\to\infty} \frac{1}{N} g(l)$$

and

$$\limsup_{N\to\infty} \frac{1}{N} g(l) \leq \limsup_{N\to\infty} \sum_{m=2}^{2n} (-1)^m \frac{1}{N} \mathscr{N}_m(l).$$

Using Equation 5.1, we have that for any $n \in \mathbb{N}$,

$$\sum_{m=2}^{2n+1} (-1)^m \frac{l^{m-1}}{(m-1)!} \leq \liminf_{N\to\infty} \frac{1}{N} g(l) \leq \limsup_{N\to\infty} \frac{1}{N} g(l) \leq \sum_{m=2}^{2n} (-1)^m \frac{l^{m-1}}{(m-1)!}.$$

Since this is true for every $n \in \mathbb{N}$, we can take the limit as $n \to \infty$. Also note that

$$\lim_{N\to\infty} \sum_{m=2}^{2n+1} (-1)^m \frac{l^{m-1}}{(m-1)!} = 1 - e^{-l} = \lim_{N\to\infty} \sum_{m=2}^{2n} (-1)^m \frac{l^{m-1}}{(m-1)!}.$$

Therefore,

$$\lim_{N\to\infty} \frac{1}{N} g(l) = 1 - e^{-l} = \int_0^l e^{-s} ds.$$

This shows that the level-spacing distribution function exists and is indeed Poissonian. Thus, we have proven Theorem 5.0.1.

5.4 An Example

As an example, we consider the family of sequences of squares modulo *highly composite*, square-free numbers. Let us begin by defining a highly-composite number.

Definition 5.4.1. *Let q be a natural number. Let $\omega(q)$ denote the number of distinct prime factors of q. Then, q is said to be in a sequence $\{q_i\}_{i\geq 1}$ of highly composite numbers if $\omega(q_i) \to \infty$ as $i \to \infty$. We say that q is a highly composite number if it belongs to a sequence of highly composite numbers.*

Given a highly-composite, square-free number q, let X_q denote the set of all the quadratic residues modulo q. Let N_q denote the number of quadratic residues modulo q. Thus,

$$N_q = |X_q|.$$

Pär Kurlberg and Zeév Rudnick proved the following in [KR99, Theorem 1].

Theorem 5.4.1. *Let $(q)_{q\in\mathbb{N}}$ be an increasing sequence of highly composite, square-free numbers* *such that $q \to \infty$. Let $m \geq 2$ be a fixed integer, and let $B \subset \mathbb{R}^{m-1}$ be a fixed bounded-convex set. Then, given any $\varepsilon > 0$, the m-level correlation sum satisfies the asymptotic*

$$R_{m,q}(B) \equiv R_{m,X_q,N_q}(B) = vol(B) + O\left(s^{-\frac{1}{2}-\varepsilon}\right) \quad as\ q \to \infty, \tag{5.3}$$

where s is the mean spacing, $s = \dfrac{q}{N_q}$.

We borrow the following result from elementary number theory regarding the number of quadratic residues modulo a square-free number.

Lemma 5.4.2. *For square-free q, the number of squares modulo q equals,*

$$N_q = \prod_{p|q} \frac{p+1}{2},$$

where the product is over all prime divisors of q.

Using this, we can show that

$$\frac{q}{N_q} = \frac{\prod_{p|q} p}{\prod_{p|q} \frac{p+1}{2}} = \frac{2^{\omega(q)}}{\prod_{p|q} \left(1 + \frac{1}{p}\right)} \to \infty \quad \text{as } q \to \infty.$$

Thus,

$$s = \frac{q}{N_q} \to \infty \ \text{ as } q \to \infty,$$

and hence, the error term in Equation 5.3 goes to zero as q goes to infinity. Therefore, this family of sequences has an m-level correlation function, and since the m-level correlation sum is asymptotic to the volume of the set B, the m-level correlation is Poissonian. Since this is true for all integers $m \geq 2$, the family has Poissonian m-level correlations of all orders.

Finally, using Theorem 5.0.1, we conclude that the family of sequences of quadratic residues modulo highly composite, square-free numbers has the Poissonian level-spacing distribution function.

Remark 5.4.1. *Observe that the sequence in Example 3.13 has Poissonian m-level correlations of all orders, and hence, it also exhibits Poissonian level-spacing distribution function, as in Example 3.4.*
Similarly, the correlations of the sequence in Example 3.15 are all Poissonian, and thus, it has Poissonian level-spacing distribution function, as given by Example 3.1.

Remark 5.4.2. *Note that the existence of Poissonian correlations of all orders is a strictly stronger condition than the existence of the Poissonian level-spacing distribution function. As an example, consider Example 3.5. The sequence in this example admits the Poissonian level-spacing distribution function, but is not even uniformly distributed mod* 1*. Thus, it cannot have Poissonian correlations of any order.*

This concludes the chapter.

Chapter 6

Conclusion

In this thesis, we explored the spacing statistics of sequences in the unit interval, focusing on the uniform distribution, correlation statistics, and the level-spacing distribution. In particular, we analysed the case of Poissonian level-spacings and Poissonian correlations. We also discussed the role of Poissonian correlations of all orders in recovering the Poissonian level-spacing distribution function of a sequence, filling in details to enhance the clarity and readability of the result.

We have provided rigorous explanations for the smooth analogues of the Poissonian correlations. We have shown that, in addition to retaining the essence of the classical definitions, the smooth analogues come with additional analytical advantages. In particular, we were able to use Fourier analysis on these to move towards a Weyl-type criterion for the pair correlation.

We realized that Poissonian correlation of any order is a stronger notion than uniform distribution. We also saw, through an example, that the Poissonian level-spacing distribution is not stronger than uniform distribution.

In conclusion, this thesis provides a comprehensive treatment of the statistical properties of sequences modulo one, integrating classical results with modern developments in correlation and gap distribution theory. The insights gained through this work contribute to the ongoing study of pseudo-randomness in deterministic sequences, offering a solid foundation for future research in this field.

We conclude the thesis by outlining some potential directions for future research in the theory of spacing statistics.

1. Refine the conditions stated in Lemma 4.1.4 and Remark 4.1.3 to establish a rigorous Weyl-type equivalence for the Poissonian pair correlations.

2. Using the higher dimensional analogues of the Beurling-Selberg trigonometric polynomials coupled with Lemma 4.2.1, and higher dimensional Fourier analysis, derive Weyl-type criteria for Poissonian higer-level correlations.

3. Discuss the quantitative aspects of Poissonian correlations, beginning with the pair correlation, to get feasible analogues of 'discrepancy' (like in the theory of uniform distribution), and see if they lead to Erdös-Turán-type inequalities.

4. Enhance the exponential sum treatment for Poissonian correlations of all orders, and recover the Poissonian level-spacing distribution function using this argument.

Bibliography

[AEM21] Christoph Aistleitner, Daniel El-Baz, and Marc Munsch. *A pair correlation problem, and counting lattice points with the zeta function*. 2021. arXiv: 2009.08184 [math.NT]. URL: https://arxiv.org/abs/2009.08184.

[AHZ23] Christoph Aistleitner, Manuel Hauke, and Agamemnon Zafeiropoulos. *On Sequences with Exponentially Distributed Gaps*. 2023. arXiv: 2312.08289 [math.NT]. URL: https://arxiv.org/abs/2312.08289.

[Ais+23] Christoph Aistleitner et al. "Gap statistics and higher correlations for geometric progressions modulo one". In: *Math. Ann.* 385.1-2 (2023), pp. 845–861. ISSN: 0025-5831,1432-1807. DOI: 10.1007/s00208-022-02362-3. URL: https://doi.org/10.1007/s00208-022-02362-3.

[ALP18] Christoph Aistleitner, Thomas Lachmann, and Florian Pausinger. "Pair correlations and equidistribution". In: *J. Number Theory* 182 (2018), pp. 206–220. ISSN: 0022-314X,1096-1658. DOI: 10.1016/j.jnt.2017.06.009. URL: https://doi.org/10.1016/j.jnt.2017.06.009.

[BT77] Michael Victor Berry and Michael Tabor. "Level clustering in the regular spectrum". In: *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* 356.1686 (1977), pp. 375–394.

[Csi30] P. Csillag. "Über die gleichmäßige Verteilung nichtganzer positiver Potenzen mod. 1." German. In: *Acta Litt. Sci. Szeged* 5 (1930), pp. 13–18.

[CY22] Sneha Chaubey and Nadav Yesha. "The distribution of spacings of real-valued lacunary sequences modulo one". In: *Mathematika* 68.2 (2022), pp. 416–428. ISSN: 0025-5793,2041-7942. DOI: 10.1112/mtk.12129. URL: https://doi.org/10.1112/mtk.12129.

[EM04] Noam D. Elkies and Curtis T. McMullen. "Gaps in $\sqrt{n}$ mod 1 and ergodic theory". In: *Duke Math. J.* 123.1 (2004), pp. 95–139. ISSN: 0012-7094,1547-7398. DOI: 10.1215/

S0012-7094-04-12314-0. URL: https://doi.org/10.1215/S0012-7094-04-12314-0.

[EMV15] Daniel El-Baz, Jens Marklof, and Ilya Vinogradov. "The two-point correlation function of the fractional parts of $\sqrt{n}$ is Poisson". In: *Proc. Amer. Math. Soc.* 143.7 (2015), pp. 2815–2828. ISSN: 0002-9939,1088-6826. DOI: 10.1090/S0002-9939-2015-12489-6. URL: https://doi.org/10.1090/S0002-9939-2015-12489-6.

[GL17] Sigrid Grepstad and Gerhard Larcher. *On pair correlation and discrepancy*. 2017. arXiv: 1612.08008 [math.NT]. URL: https://arxiv.org/abs/1612.08008.

[HZ23] Manuel Hauke and Agamemnon Zafeiropoulos. "Poissonian correlations of higher orders". In: *J. Number Theory* 243 (2023), pp. 202–240. ISSN: 0022-314X. DOI: 10.1016/j.jnt.2022.08.007. URL: https://doi.org/10.1016/j.jnt.2022.08.007.

[KN74] L. Kuipers and H. Niederreiter. *Uniform distribution of sequences*. Pure and Applied Mathematics. Wiley-Interscience, John Wiley and Sons, New York-London-Sydney, 1974.

[KR99] Pär Kurlberg and Zeév Rudnick. "The distribution of spacings between quadratic residues". In: *Duke Math. J.* 100.2 (1999), pp. 211–242. ISSN: 0012-7094. DOI: 10.1215/S0012-7094-99-10008-1. URL: https://doi.org/10.1215/S0012-7094-99-10008-1.

[LS20] Gerhard Larcher and Wolfgang Stockinger. "On pair correlation of sequences". In: *Discrepancy theory*. Vol. 26. Radon Ser. Comput. Appl. Math. De Gruyter, Berlin, [2020] ©2020, pp. 133–145. ISBN: 978-3-11-065115-7; 978-3-11-065258-1; 978-3-11-065120-1. DOI: 10.1515/9783110652581-007. URL: https://doi.org/10.1515/9783110652581-007.

[LST23] Christopher Lutsko, Athanasios Sourmelidis, and Niclas Technau. *Pair Correlation of the Fractional Parts of αn^{θ}*. 2023. arXiv: 2106.09800 [math.NT]. URL: https://arxiv.org/abs/2106.09800.

[LT21] Christopher Lutsko and Niclas Technau. *Correlations of the Fractional Parts of αn^{θ}*. 2021. arXiv: 2112.11524 [math.NT]. URL: https://arxiv.org/abs/2112.11524.

[LT24] Christopher Lutsko and Niclas Technau. *Full Poissonian Local Statistics of Slowly Growing Sequences*. 2024. arXiv: 2206.07809 [math.NT]. URL: https://arxiv.org/abs/2206.07809.

[Mah24] Jewel Mahajan. *The second moment of a certain pair correlation function for Sato-Tate sequences*. Indian Institute of Science Education and Research (IISER). 2024.

[Mon14] Hugh L Montgomery. *Early Fourier Analysis*. Vol. 22. American Mathematical Soc., 2014.

[MS13] Jens Marklof and Andreas Strömbergsson. "Gaps between logs". In: *Bulletin of the London Mathematical Society* 45.6 (Aug. 2013), pp. 1267–1280. ISSN: 0024-6093. DOI: 10.1112/blms/bdt060. URL: http://dx.doi.org/10.1112/blms/bdt060.

[RS98] Zeév Rudnick and Peter Sarnak. "The pair correlation function of fractional parts of polynomials". In: *Comm. Math. Phys.* 194.1 (1998), pp. 61–70. ISSN: 0010-3616,1432-0916. DOI: 10.1007/s002200050348. URL: https://doi.org/10.1007/s002200050348.

[RT20] Zeév Rudnick and Niclas Technau. "The metric theory of the pair correlation function of real-valued lacunary sequences". In: *Illinois J. Math.* 64.4 (2020), pp. 583–594. ISSN: 0019-2082,1945-6581. DOI: 10.1215/00192082-8720506. URL: https://doi.org/10.1215/00192082-8720506.

[RZ02] Zeév Rudnick and Alexandru Zaharescu. "The distribution of spacings between fractional parts of lacunary sequences". In: *Forum Math.* 14.5 (2002), pp. 691–712. ISSN: 0933-7741,1435-5337. DOI: 10.1515/form.2002.030. URL: https://doi.org/10.1515/form.2002.030.

[Ste17] Stefan Steinerberger. *Poissonian Pair Correlation and Discrepancy*. 2017. arXiv: 1711.02497 [math.NT]. URL: https://arxiv.org/abs/1711.02497.

[TY23] Niclas Technau and Nadav Yesha. "On the correlations of n^α mod 1". In: *J. Eur. Math. Soc. (JEMS)* 25.10 (2023), pp. 4123–4154. ISSN: 1435-9855. DOI: 10.4171/jems/1281. URL: https://doi.org/10.4171/jems/1281.

[Vaa85] J. D. Vaaler. "Some extremal functions in Fourier analysis". In: *Bull. Amer. Math. Soc.* 12.2 (Apr. 1985), pp. 183–216.

[Wei23] Christian Weiß. *An Explicit non-Poissonian Pair Correlation Function*. 2023. arXiv: 2304.14202 [math.NT]. URL: https://arxiv.org/abs/2304.14202.

[Wey16] H. Weyl. "Über die Gleichverteilung von Zahlen mod. Eins". In: *Math. Ann.* 77.3 (1916), pp. 313–352. ISSN: 0025-5831. URL: https://doi.org/10.1007/BF01475864.