One-dependent Point Processes

Carries Process (mod $\infty$)

Harry Sapranidis Mantelos

Supervisors: Professors Roger Tribe and Oleg Zaboronski

First year PhD summer project

Mathematics Institute
University of Warwick
United Kingdom
29/08/2022
## Contents

1 Introduction 1

2 Preliminaries 4
   2.1 Point processes, DPPs and their properties, thinning of point processes 4
   2.2 Gap probabilities 8
   2.3 Stationarity, 1-dependence, the carries process and its properties 10

3 Toeplitz matrices 18
   3.1 Introduction to Toeplitz matrices 18
   3.2 Szegő-type asymptotics for Toeplitz matrices 22

4 1-dependent processes on the integers 40

5 Asymptotics for the thinned carries process 65
   5.1 The Combinatorial Approach 65
   5.2 The Determinantal Structure Approach 74

6 Conclusion 101
Abstract

Let \((X_i)_{i \in \mathbb{Z}}\) be any 1-dependent point process on \(\mathbb{Z}\). It was discovered in [3] that \((X_i)_{i \in \mathbb{Z}}\) must then have a determinantal structure, and an explicit kernel for the process was presented. By making the further assumption that \((X_i)_{i \in \mathbb{Z}}\) is translation-invariant we find explicit expressions for the associated kernel of this process, which turns out (unsurprisingly) to be Toeplitz, and non-Hermitian. By achieving this we are then able to derive the result from [3] mentioned previously with rigorous detail as well as to provide a couple of other acceptable kernels. We then study the Carries/Descents process \((\text{mod } b)\) from [3] (a 1-dependent translation-invariant point process on \(\mathbb{Z}\)) and consider the case where we take \(b \to \infty\). We obtain an asymptotic (as well as an exact expression) for the probability of a large gap of the corresponding thinned process via two different approaches: the combinatorial approach, utilizing Eulerian numbers; and the determinantal structure approach, exploiting the determinantal structure of the point process as well as the Toeplitz structure of the associated kernels for this process by making use of the aforementioned theoretical results and also by developing analogous Sz"ego-type results for non-Hermitian Toeplitz matrices.

1 Introduction

Point processes are probabilistic models for random scatterings of points in some mathematical space. More precisely they’re random integer-valued positive radon measures (i.e., Borel measures that are finite on compact sets) on this space (see Section 1.2, [11], for more details). This space is usually taken to be some subset of the real line or even \(\mathbb{R}^d\). In simple words, point processes count the number of occurrences of some event in some mathematical space. This is why they are also often called counting processes. One can thus imagine the vast amount of applications such mathematical objects have in disciplines such as telecommunications, to model the number of radio stations in a network; finance, to model the number of orders received in a stock exchange at some point in time; and physics, to model the number of particles in a Geiger counter, just to name a few specific examples.

The simplest and most popular point process is the Poisson point process, which, without getting into the technical details, is a point process on the real line, whereby the real line is usually thought of as the time variable and subsets of it can therefore be thought of as time intervals where some number of occurrences of some specified event (e.g., number of orders in the stock exchange) can take place in. It is assumed for Poisson point processes that the number of occurrences (of some specified event) in different (disjoint) time intervals are independent from each other. It is also assumed that the number of occurrences of an event in some time interval follow a Poisson distribution with parameter given by the size of the interval multiplied by some other specified
parameter \( c > 0 \) known as the intensity of the process. Thus, the probability of getting some number of occurrences in some time interval only depends on the size of the interval and not on the position of the interval on the real line.

Although the Poisson point process is quite a simple process it is usually not the most realistic model in most cases. Take for example fermions in quantum mechanics where the likelihood of observing two fermions close together is small. In this case the Poisson point process fails to capture this phenomenon - there is a clear dependency between the random points (the fermions in this case) whereby there is a tendency to repel each other - but the Poisson process assumes no dependence! This is where determinantal point processes (abbr. DPP) enter the scene, capable of modelling such negative associations in a mathematically elegant way, whereby probabilities of the process are principle minors of a known matrix called the kernel for the DPP.

This particular example with the fermions was actually the reason why this class of point processes (DPPs) was initially created and studied (see [16]). DPPs’ original name, in fact, was the fermion process. It first made its appearance in 1975 in [16]. Thereafter the properties of this class of point processes were found to be desirable in areas other than quantum mechanics such as in random matrix theory to model the positions of eigenvalues of random matrices (e.g., see [11], [12]). Actually, the very first instance of the word ”DPP” being used interchangeably with the word ”fermion process” was in the year 2000 in [4].

There are two types of DPPs: discrete and continuous. Discrete DPPs are defined on spaces with finite cardinality and continuous DPPs are defined on Polish spaces (such as the real line). In this project we will only be concerned with the discrete setting. With that said, various detailed treatments of DPPs in the discrete setting can be found in [21], [2] and [14] where there is also an emphasis on their ever-growing popularity in solving machine learning tasks that require the selection of diverse objects. As for DPPs on the continuum, a detailed treatment can be found in [11].

In this project we study a specific class of point processes in detail: 1-dependent, translation-invariant point processes on the integers (precise definitions will be given in the Preliminary sections). It was first discovered in [3] that 1-dependent point processes on the integers are determinantal and an explicit expression of an associated kernel was given (even without the translation-invariance assumption). In this project, however, we are mostly interested in the additional assumption of translation-invariance. Now, there was no explanation as to how the kernel given in [3], Theorem 7.1, for an arbitrary 1-dependent point process on the integers was derived in the first place. Only an inductive verification (with minimal details) that it works was given. However, by the kernels we derived and made sense of in the translation-invariant setting we were able to decipher their (possible) meaning in the more general setting (by relaxing the additional
translation-invariance assumption) and thus we were able to derive the formula given in Theorem 7.1, [3], from scratch with full details, as well as provide some additional kernels that work. This was done in Section 4. Only the statements of these alternative kernels, and minimal details for their derivation, however, are given in this project because, otherwise, the project loses its main focus and goes well over the page limit. Moreover it is not clear whether these alternative kernels we found are any better (for the time being, at least) than the one given in Theorem 7.1, [3], and hence are better kept for further work later.

Having found explicit expressions for kernels of an arbitrary 1-dependent, translation-invariant point process on \( \mathbb{Z} \), which all turn out to be Toeplitz, we were able to derive the symbol associated to one of these (the symbol being the Fourier series that has Fourier coefficients being the entries of the Toeplitz matrix). We then studied the carries/descents point process (mod \( b \)) from [3] and considered the case where we take the base \( b \to \infty \). The resulting point process is both 1-dependent and translation-invariant. Therefore, the previous theoretical results to this point process were applicable here and so we were able to establish its determinantal structure with Toeplitz kernel in rigorous detail (whereas only statements of these results were given in [3]). We were then interested in obtaining an asymptotic (as well as an exact) expression for the probability of a large gap of the corresponding thinned process (precise definitions in the preliminary sections) and see how this asymptotic varies as we alter the thinning parameter. This is generally, as we will discuss later on in Section 2.1, a difficult (perhaps even impossible) problem even for simple point processes. But the many nice properties that DPPs posses (the closure under complementation and thinning properties in this case - see Section 2.1 for details) as well as the fact that the kernel in our case is of a special type (Toeplitz), abolishes the impossibility of obtaining such statistics for the point process. This is the approach we take to obtain the desired asymptotic in Section 5.2.

Moreover, since we are in a discrete-probability setting there is usually some counting/combinatorics that we can use to obtain probabilities, and this provided one other way of obtaining our desired statistics (see Section 5.1). However, more interesting is the approach explained in the previous paragraph because of the many challenges it presented in our analysis due to the fact that kernels of the Carries process turned out to be non-Hermitian, which meant that all the classic Szegö theorems (e.g., see [7], [8], [23], [13]) on the asymptotic behaviours of Toeplitz determinants and eigenvalues (which only hold for Hermitian matrices) were not applicable, and so adjustments and adaptations to these results had to be made (that is, new results had to be constructed - see Sections 3.2 and 5) in addition to applying a theorem from a rather elusive and old paper (Theorem 2, [1]), as well as refining some results already existing in the non-Hermitian Toeplitz matrix literature, such as deriving a tighter uniform (in the matrix size) bound on the modulus of the eigenvalues of a sequence of non-Hermitian Toeplitz matrix (from the bound given in Lemma 4.1, [7]). The latter indeed had presented a bit of an obstacle in the project because papers on non-Hermitian Toeplitz matrices
were, besides scarce, only seeming to focus on the behaviour of their singular values rather than of their eigenvalues (e.g., see [5], [24], [9], [19], [6]); and papers that specifically discussed eigenvalues of non-Hermitian Toeplitz matrices, such as [26], would remark that such results are very sparse and of mostly theoretical interest. Actually, [26] presented a conjecture for the behaviour of eigenvalues of non-Hermitian Toeplitz matrices as well as some evidence supporting this conjecture. Despite these difficulties, however, we were eventually able to derive the uniform-bound-on-the-eigenvalues result we needed that solved the puzzle in the project (see Theorem 5.9, Section 5).

2 Preliminaries

2.1 Point processes, DPPs and their properties, thinning of point processes

We begin by formally defining DPPs in the discrete setting.

**Definition (from Section 3, [3])** Let $V$ be a set of finite cardinality. Without loss of generality we may assume that $V = \{1, \ldots, n\}$ (so $|V| = n < \infty$). We say that the random subset $Y \subseteq V$ is a **(discrete) determinantal point process** on $V$ with kernel $K := (K_{ij})_{1 \leq i, j \leq n} \in \mathbb{C}^{n \times n}$ if

$$P(A \subseteq Y) = \det(K_A), \quad A \subseteq V,$$

where $K_A$ denotes the principle submatrix of $K$ that is restricted to entries of the original matrix $K$ that are indexed by $A$, that is, $K_A := (K_{ij})_{i, j \in A} \in \mathbb{C}^{|A| \times |A|}$.

**Remark** Since probabilities are by definition non-negative, the above definition implies that every principle minor of the matrix $K$ is non-negative. This, by Sylvester’s criterion, is equivalent to $K$ being positive semidefinite (provided that $K$ is Hermitian).

**Remark** The repulsion property that DPPs exhibit that was mentioned in the introduction can easily be seen from the above definition (with the additional assumption that $K$ is Hermitian):

Let $i, j \in V$. Then,

$$P(i, j \in Y) = P(\{i, j\} \subseteq Y) = \det(K_{\{i,j\}}) = \begin{vmatrix} K_{ii} & K_{ij} \\ K_{ji} & K_{jj} \end{vmatrix}.$$
\[ K_{ii}K_{jj} - |K_{ij}|^2 = K_{ii}K_{jj} - K_{ij}K_{ji} \]  
\[ \leq K_{ii}K_{jj} \text{ (Hermitianity assumption on } K) \]  
\[ = P(i \in Y)P(j \in Y) \]

Observe how the off-diagonal entries of $K$ are the entries that determine the magnitude of the repulsion between elements of $V$. Indeed, notice how the larger $|K_{ij}|$ above is, the smaller the probability of observing the elements $i, j \in V$ together. In particular, if $K_{ij}$ is exactly equal to $\sqrt{K_{ii}K_{jj}}$, then almost surely the elements $i$ and $j$ are never together. On the other hand if $K$ is a diagonal matrix (and so $K_{ij} = 0$ for all $i, j \in V$ such that $i \neq j$), then $P(i, j \in Y) = P(i \in Y)P(j \in Y)$ and so elements of $V$ in this case occur independently (just like for Poisson point processes mentioned earlier).

It is also important to note that DPPs (with Hermitian kernels), though appropriate for modelling repulsion (negative association), they are unable to model positive association. That is, if two elements are likely to be observed together, DPPs (with Hermitian kernels) are unable to capture this - this is obvious from the previous equations.

Below we state a well-known property of DPPs which we will make extensive use of later on:

**Theorem 2.1 (Closure of DPPs under complementation, p.12, [14])** Let $Y$ be a DPP on $V := \{1, \ldots, n\}$ with kernel $K \in \mathbb{C}^{n \times n}$. Then $\overline{Y} := V \setminus Y$ is a DPP with kernel $I_n - K$, where $I_n$ is the $n \times n$ identity matrix.

We have defined DPPs as random subsets of some finite set. We can, however, also identify them in the following way (as explained in [21]):

First let $Y$ be a DPP on $V := \{1, \ldots, n\}$ with kernel $K \in \mathbb{C}^{n \times n}$. Then, for $i \in \{1, \ldots, n\}$, define the random variable $X_i$ by

\[ X_i = \begin{cases} 
1, & \text{if } i \in Y \\
0, & \text{if } i \notin Y 
\end{cases} 
\]

So by definition,

\[ P(X_i = 1) = P(\{i\} \subseteq Y) = \det((K_{ii})) = K_{ii}. \]

Then define for $A \subseteq V$, the (random) vector given by

\[ X_A = (X_i)_{i \in A} \in \{0, 1\}^{|A|}. \]
So then by definition,

\[ P(X_A = 1_A) = P(A \subseteq Y) = \det(K_A), \]

where \( 1_A := (1, \ldots, 1)^\top \in \mathbb{R}^{|A|} \) denotes the \(|A|\)-sized vector with all its entries being equal to 1.

We can therefore identify our DPP \( Y \) by the random variables \( X_1, \ldots, X_n \) defined above.

Let’s now introduce the concept of a correlation function.

**Definition (from [21])** Let \( Y \) be a DPP on \( V := \{1, \ldots, n\} \) with kernel \( K \in \mathbb{C}^{n \times n} \). Let \( X_1, \ldots, X_n \) be the binary sequence of random variables associated with \( Y \) as described previously.

We define the **correlation function** of \( Y \) as

\[ \rho(A) := P(X_A = 1_A) = P(A \subseteq Y) = \det(K_A), \quad \text{where } A \subseteq V. \]

**Remark** This identification of DPPs by a sequence of binary random variables as defined above together with Theorem 2.1 imply the following:

Let \( X_1, \ldots, X_n \) be the binary random variables that identify the DPP \( Y \subseteq V \) with kernel \( K \in \mathbb{C}^{n \times n} \) and let \( \rho \) denote the correlation function of \( Y \). Then, Theorem 2.1 implies that \( V \setminus Y \) is a DPP with kernel \( I_n - K \in \mathbb{C}^{n \times n} \). Let \( \tilde{X}_1, \ldots, \tilde{X}_n \) be the binary random variables that identify \( V \setminus Y \) and let \( \tilde{\rho} \) denote the correlation function of \( V \setminus Y \). Then, by making use of the well-known notation \( [n] := \{1, \ldots, n\} \),

\[ \tilde{\rho}([n]) := P(\tilde{X}_1 = 1, \ldots, \tilde{X}_n = 1) = P([n] \subseteq V \setminus Y) = P(X_1 = 0, \ldots, X_n = 0) \]

(1)

and

\[ \tilde{\rho}([n]) = \det(I_n - K). \]

(2)

And so

\[ P(X_1 = 0, \ldots, X_n = 0) = \det(I_n - K) \]

(3)

gives the probability of an \( n \)-gap of the point process as a determinant of a known matrix.

Let’s now introduce the concept of *thinning* a point process (see, e.g., [2]).

**Thinning** is a (random) operation applied to the points of an underlying point process \( Y \), where the points are thinned (i.e., removed/killed) or retained (i.e., kept) according to some probabilistic rule. By considering the previously explained identification of a DPP by the sequence \( X_1, \ldots, X_n \),
to thin a point \(i \in V\) means, informally, that if we had \(X_i = 1\) (i.e., \(i \in Y\)) initially, we have \(X_i = 0\) (i.e., \(i \notin Y\)) after thinning the point.

There are two types of thinning: *spatially independent thinning* and *spatially dependent thinning*. The latter type of thinning essentially has a dependence on the location of the points. In other words it favours thinning specific points more than others, i.e., the thinning happens with a (possibly) different probability at different points. Spatially independent thinning on the other hand has no favour on which points it thins - all points are equally likely to be thinned. We shall consider spatially independent thinning from now on.

**Definition** We can define the \(p\)-thinned point process of the DPP \(X_1, \ldots, X_n\) from above, where \(0 \leq p \leq 1\), in the following way:

\[
X^{(p)}_1, \ldots, X^{(p)}_n,
\]

where

\[
X^{(p)}_i = B_i X_i,
\]

where

\[
B_i = \begin{cases} 
0, & \text{with probability } 1-p \\
1, & \text{with probability } p 
\end{cases}
\]

for \(i \in V\), is a sequence of independent \(\text{Ber}(p)\) random variables that are independent of \(X_1, \ldots, X_n\) (this is the aforementioned notion of spatial independence).

In other words we kill a particle \(i \in V\) with probability \(1 - p\) and we keep it with probability \(p\).

It turns out that a thinned DPP is still determinantal:

**Proposition 2.2 (Closure of DPPs under thinning)** Let \(X_1, \ldots, X_n\) be the DPP with kernel \(K\) defined previously. Then the (spatially independently) \(p\)-thinned process \(X^{(p)}_1, \ldots, X^{(p)}_n\) defined previously is a DPP with kernel \(pK\).

**Proof** Let the correlation functions of \(X_1, \ldots, X_n\) and \(X^{(p)}_1, \ldots, X^{(p)}_n\) be \(\rho\) and \(\tilde{\rho}\) respectively. Observe how, for \(A \subseteq V\),

\[
\tilde{\rho}(A) := P(X^{(p)}_A = 1_A)
\]

\[
= P(X_A = 1_A, B_A = 1_A)
\]

\[
= P(X_A = 1_A) P(B_A = 1_A)
\]

\[
= \rho(A) \cdot \prod_{i \in A} P(B_i = 1)
\]

(we assumed the \(B_1, \ldots, B_n\) are independent)
\[ = \rho(A) \cdot p^{|A|} \]
\[ = \det(K_A) \cdot p^{|A|} \quad (X_1, \ldots, X_n \text{ is determinantal with kernel } K) \]
\[ = p^{|A|} \cdot \sum_{\tau \in S_{|A|}} (-1)^{sgn(\tau)} \prod_{i=1}^{|A|} (K_A)_{i,\tau(i)} \quad \text{(Leibniz formula for determinants)} \]
\[ = \sum_{\tau \in S_{|A|}} (-1)^{sgn(\tau)} \prod_{i=1}^{|A|} p(K_A)_{i,\tau(i)} \]
\[ = \sum_{\tau \in S_{|A|}} (-1)^{sgn(\tau)} \prod_{i=1}^{|A|} (pK_A)_{i,\tau(i)} \quad \text{(Leibniz formula for determinants)} \]
\[ = \det(pK_A) \]
\[ = \det((pK)_A). \]

This implies that the thinned process \( X_1^{(p)}, \ldots, X_n^{(p)} \) is determinantal with kernel \( pK \).  

2.2 Gap probabilities

In this section we discuss the challenges of obtaining gap probabilities for general (discrete) point processes as mentioned in the introductory section.

Given a general point process \( Y \) on \( V := [n] \) (not necessarily having a determinantal structure, i.e., \( Y \) not necessarily being a DPP), which, as we’ve explained in the previous section, can be identified by a sequence of random variables \( X_1, \ldots, X_n \) such that

\[ X_i = \begin{cases} 
1, & \text{if } i \in Y \\
0, & \text{if } i \notin Y 
\end{cases} \]

where

\[ X_i = \begin{cases} 
1, & \text{with probability } P(i \in Y) \\
0, & \text{with probability } P(i \notin Y) 
\end{cases} \]

what is the probability that we get a long consecutive sequence of particles (i.e., a long sequence of 1’s)? What about the probability of a large gap (i.e., a long sequence of 0’s)? How do these probabilities change as we increase the cardinality of our finite subset \( V := [n] \), that is, how do these probabilities decay as \( n \) gets larger (of course, it wouldn’t make sense for these probabilities to grow as \( n \) gets larger and larger since we are demanding for more and more from the sequence \( X_1, \ldots, X_n \))? The answer to the first question is simply the definition of the correlation function of
the point process. More precisely, it’s equal to \( \rho(V) \), which, in terms of the \( X_1, \ldots, X_n \) sequence, is just equal to \( P(X_V = 1_V) \equiv P(X_1 = 1, \ldots, X_n = 1) \). Therefore, since we have full knowledge of the collection of numbers

\[
\{\rho(A) : A \subseteq V\}
\]

the answer to this question is trivial.

Remark Actually, the numbers given in (4) uniquely determine the point process, since, in terms of these numbers, we can get unique expressions for any probability of the point process \( Y \) through inclusion-exclusion ([21]). More precisely, by using the vector shorthand notations that were introduced previously, for any \( \emptyset \neq A \subseteq V \),

\[
P(Y = A) = P(X_A = 1_A, X_{V \setminus A} = 0_{V \setminus A}) = \sum_{B \subseteq V : A \subseteq B} (-1)^{|B \setminus A|} \rho(B),
\]

where, as one can easily guess, \( 0_{V \setminus A} \) denotes the \(|V \setminus A|\)-sized vector having only 0 entries. We define \( \rho(\emptyset) = 1 \).

Now, what about the second question: the probability of a run of 0’s? In that case, full knowledge of the numbers given in (4) does not give us a simple and straightforward answer to this question. On the contrary, it only gives us complicated expressions through inclusion-exclusion of these \( \rho(A) \)'s similar to the one given in (5). One such expression is the following:

\[
P(X_1 = \ldots = X_n = 0) \equiv P(X_V = 0_V)
\]

\[
= 1 - P(\exists i \in V : X_i = 1 \text{ in } (X_1, \ldots, X_n))
\]

\[
= 1 - P\left( \bigcup_{k=1}^{n} \bigcup_{A \subseteq V : |A|=k} \{X_A = 1_A, X_{V \setminus A} = 0_{V \setminus A}\} \right)
\]

\[
= 1 - \sum_{k=1}^{n} \sum_{A \subseteq V : |A|=k} P(X_A = 1_A, X_{V \setminus A} = 0_{V \setminus A}) \quad \text{(by mutual-disjointness of events)}
\]

\[
= 1 - \sum_{k=1}^{n} \sum_{A \subseteq V : |A|=k} \sum_{B \subseteq V : A \subseteq B} (-1)^{|B \setminus A|} \rho(B) \quad \text{(by (5))}
\]

One could try and get a more simplified expression than the one above (but not by much), however one cannot hope to get much meaning out of it, even less so get the faintest idea as to how to answer the last question from before regarding the decay of this probability as \( n \) gets large.

To be able to say more we need the point processes to have some "special" structures. This
is where determinantal point processes enter the scene. Let’s now assume that \( Y \) is a DPP with kernel \( K \in \mathbb{C}^{n \times n} \). We saw in (3) that we can then write the probability from (6) as the determinant of a matrix:

\[
P(X_1 = \ldots = X_n = 0) \equiv P(X_V = 0_V) \\
= \det((I - K)_V) \quad \text{(definition of a DPP with kernel } I - K) \\
= \det(I - K) \quad \text{(} V = [n] \text{ is the whole set)}
\]

One can argue that even now we cannot hope to be able to answer the question regarding the decay of the above probability as \( n \) gets large since computing determinants of large matrices is usually far from straightforward if the matrix does not have a particularly nice structure. But, for some special types of matrices there are some very well-known results regarding the limiting behaviour of their determinants and eigenvalues (as the dimension of the matrix gets large) and those are, as mentioned and discussed in the introductory section, (Hermitian) Toeplitz matrices. This then means that if we have a DPP with a Hermitian Toeplitz kernel, then we can say a lot about the decay of the probability of gaps as the dimension \( n \) gets large.

### 2.3 Stationarity, 1-dependence, the carries process and its properties

In this section we introduce the notion of a stationary and 1-dependent point process and we also introduce the Carries process from [3] that we want to study.

**Definition (Page 3, [3])** We say that the point process \( X_1, \ldots, X_n \) is **stationary** if for every \( J \subseteq [n] \) and every \( i \in [n] \) such that \( \{j + i : j \in J\} =: J + i \subseteq [n] \) we have that \( \{X_j : j \in J\} \) has the same distribution as \( \{X_j : j \in J + i\} \) (i.e., \( \{X_j : j \in J\} \overset{d}{=} \{X_j : j \in J + i\} \)).

**Remark** It is immediate from the definition above that the respective correlation function, \( \rho \), of the process for every \( k \in \mathbb{Z} \) satisfies:

\[
\rho([k]) \equiv P(X_1 = \ldots = X_k = 1) \\
= P(X_{1+i} = 1, \ldots, X_{k+i} = 1) \quad \text{(12)} \\
=: \rho_k \quad \text{(13)}
\]

is independent of \( i \in \mathbb{Z} \).
If we further assume that the point process at hand is determinantal, then we have that

$$\rho_k = \det(K_{1+i, \ldots, k+i})$$

for some matrix $K$ for all $i \in \mathbb{Z}$. One would then rightly suspect that there must be a link between Toeplitz matrices and stationary DPPs - and in fact there is! It is proved in Section 4 that for 1-dependent (see the definition below) stationary point processes on (a segment of) $\mathbb{Z}$ there exists a kernel for the process (not necessarily unique) that is a Toeplitz matrix.

Let’s now introduce the notion of a 1-dependent point process.

**Definition (Page 3, [3])** Let $X_1, \ldots, X_n$ be a point process. We say that the process is 1-dependent if

$$X_1, \ldots, X_{i-1} \perp \perp X_{i+1}, \ldots, X_n$$

for every $i \in \{2, \ldots, n-1\}$.

A point process on (a segment of) the integers is just a point process with associated binary random variables $X_i$ that are indexed by (that segment of) the integers. That is, given a segment $A \subseteq \mathbb{Z}$, a point process on $A$ is the sequence of binary random variables $(X_i)_{i \in A}$ that, as we’ve discussed many times, depend on the correlation function $\rho$ of the point process.

**Remark** It is then easy to see from the above definition that, for a point process, in terms of its correlation function $\rho$, 1-dependence on (a segment of) $\mathbb{Z}$ is equivalent to the statement

$$\forall X, Y \subseteq \mathbb{Z}, \quad \min_{x \in X, y \in Y} |x - y| =: \text{dist}(X, Y) \geq 2 \implies \rho(X \cup Y) = \rho(X)\rho(Y) \quad (14)$$

The point process and its desired statistics (the gap probabilities mentioned before) that we explore in this project is the carries and descents process introduced in [3], where we further explore the case where the base $b$ is taken to be $\infty$. We now introduce this point process in detail and in a similar way that was done in [3]:

First consider the task of adding digits, say

$$3 + 7 + 4 + 5 + 9 + 2 + 1 + 4 + 8 + 5.$$
Let’s write this list of added digits as a sequence

\[(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}) = (3, 7, 4, 5, 9, 2, 1, 4, 8, 5).\]  \hspace{1cm} (15)

When we add 7 to 3 we get 10 and so there is a carry at position 1 (since we do exceed 9), and in this case there is no remainder (i.e., remainder 0 at position 2) since \(7 + 3 = 10 = 0 \mod 10\).

When we then add 4 to 7 + 3 (= 10) we get 14 and so there is no carry at position 2 (since we do not exceed 19), and in this case we get a remainder of 4 at position 3 since \((7 + 3) + 4 = 10 + 4 = 4 \mod 10\).

When we then add 5 to 14 we get 19 and so there is no carry at position 3 (since we still do not exceed 19), and in this case we get a remainder of 9 at position 4 since \(14 + 5 = 19 = 9 \mod 10\).

When we then add 9 to 19 we get 28 and so there is a carry at position 4 (since we still do not exceed 29) with remainder 8 at position 5 since \(19 + 9 = 28 = 8 \mod 10\).

When we then add 2 to 28 we get 30 and so there is a carry at position 5 (since we do exceed 29) and in this case there is no remainder (i.e., remainder 0) at position 6 since \(28 + 2 = 30 = 0 \mod 10\).

When we then add 1 to 30 we get 31 and so there is no carry at position 6, and in this case we get remainder 1 at position 7 since \(30 + 1 = 31 = 1 \mod 10\).

When we then add 4 to 31 we get 35 and so there is no carry at position 7 (since we still do not exceed 39), and in this case we get a remainder of 5 at position 8 since \(31 + 4 = 35 = 5 \mod 10\).

When we then add 8 to 35 we get 43 and so there is a carry at position 8 (since we now exceed 39) with remainder 3 at position 9 since \(35 + 8 = 43 = 3 \mod 10\).

When we then add 5 to 43 we get 48 and so there is no carry at position 9 (since we do not exceed 49), and in this case we get a remainder of 8 at position 10 since \(43 + 5 = 48 = 8 \mod 10\).

Let’s write the remainders we obtain after each addition as a sequence (with \(y_1 \equiv x_1\))

\[(y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8, y_9, y_{10}) = (3, 0, 4, 9, 8, 0, 1, 5, 3, 8).\]  \hspace{1cm} (16)

Let us also write the sequence from (15) with dots on top of the corresponding digits at which a carry occurs:

\[(\hat{x}_1, x_2, x_3, \hat{x}_4, x_5, x_6, x_7, \hat{x}_8, x_9, x_{10}) = (3, 7, 4, \hat{5}, 9, 2, 1, \hat{4}, 8, 5).\]  \hspace{1cm} (17)
The key observation to be made here by comparing the sequence from (16) with the sequence from (17) is that there is a carry in the sequence from (15) at position \(i \in \{10\}\) (in the sense explained in the previous explanations) if and only if \(y_i > y_{i+1}\) (i.e., there is a descent at \(i\) in the sequence from (16)).

Let’s now define our "alphabet" \(B := \{0, \ldots, b - 1\}\). In this case we say that we are working with "base \(b\)". Let \(B_1, \ldots, B_n\) be a sequence of i.i.d. randomly and uniformly chosen elements from \(B\). In the previous example that we went through we were working with base 10 and \((x_1, \ldots, x_n)\) was a particular realization of the sequence of random variables \((B_1, \ldots, B_n)\) chosen from \(B\) with \(b = 10\).

We say that there is a \underline{descent} at \(i \in \{n - 1\}\) if \(B_i > B_{i+1}\). Define the random variables \(X_1, \ldots, X_{n-1}\) as follows:

\[
X_i = \begin{cases} 
1, & \text{if there is a descent at } i \\
0, & \text{if there is no descent at } i 
\end{cases}
\]

In the example we went through we made the observation that there is a carry if and only if there is a descent, and so \(X_i = 1\) says that there is a carry at position \(i\) (in some random sequence of uniformly and independently chosen numbers from \([b - 1]\)). Therefore we shall call the sequence \(X_1, \ldots, X_{n-1}\) the Carries process.

We now explain in detail the probability theory for the Carries process (adding more detail to Section 2 of [3]).

**Proposition 2.3 (Single carry)** For \(i \in \{n - 1\}\), \(P(X_i = 1) = \left(\frac{b}{b^2}\right)\).

**Proof** The number of distinct pairs of distinct numbers from \(B\) is by definition \(\left(\binom{b}{2}\right)\) and hence so is the number of ways we could have \(B_i > B_{i+1}\). We then divide this quantity by the total number of (not necessarily distinct) pairs of (not necessarily distinct) numbers from \(B\) which there are precisely \(b^2\)-many of them.

**Proposition 2.4 (Runs of carries)** For \(1 \leq i < i + j \leq n\),

\[
P(X_i = X_{i+1} = \ldots = X_{i+j-1} = 1) = \frac{\binom{b}{j+1}}{b^{j+1}}
\]

**Proof** The probability of having \(j\) consecutive descents starting from \(i\) is the number of ways we could get a sequence \(B_i > B_{i+1} > \ldots > B_{i+j-1} > B_{i+j}\) from the set \(B\) (which is equal to the number of distinct \((j + 1)\)-sized pairs of distinct numbers from \(B\), which is given by \(\binom{b}{j+1}\)) divided
by the total number of \((j + 1)\)-sized pairs of (not necessarily distinct) elements from \(B\) (which is equal to \(b^{j+1}\)).

**Remark** Of course it’s impossible to have a run of \(b\) or more carries since there is no way we can get a sequence \(B_1 > \ldots > B_{b+k}\), where \(k \geq 1\), from the set \(B\) as we would need \(b + k\) distinct elements from \(B\) which we know only has \(b\).

**Proposition 2.5 (Stationarity of the Carries process)** The distribution of the Carries process is stationary.

**Proof** It is clear from the construction of the Carries process that it does not matter from which point \(i \in [n-1]\) we start the process. Proposition 2.4 clearly depicts this.

**Proposition 2.6 (1-dependence of the Carries process)** The distribution of the Carries process is one-dependent.

**Proof** Again this is clear from the definition of one-dependent processes and the construction of the Carries process since the value of \(X_i\) (either 0 or 1) is only dependent on the values of the random variables \(B_i\) and \(B_{i+1}\) (recall how \(X_i = 0\) if \(B_i \leq B_{i+1}\) and \(X_i = 1\) if \(B_i > B_{i+1}\) and therefore, since we had assumed that the \(B_k\)'s were i.i.d., \(X_i \perp \!\!\!\!\perp X_{i+2}\) (since \(X_{i+2}\) is only dependent on the random variables \(B_{i+2}\) and \(B_{i+3}\)). However, \(X_i\) and \(X_{i+1}\) are not independent since they both have a mutual random variable they depend on, namely \(B_{i+1}\).

As mentioned earlier, in this project we will explore the Carries process with the base \(b \to \infty\).

First observe how, by Proposition 2.4, for any \(1 \leq i < i + j \leq n\),

\[
\lim_{b \to \infty} P(X_i = X_{i+1} = \ldots = X_{i+j-1} = 1) = \lim_{b \to \infty} \frac{b^n}{b^{n+1}} = \frac{1}{(n + 1)!}.
\]

So it is natural to define the Carries point process with base \(b \to \infty\) as follows:

**Definition** We call the point process \((X_i)_{i \in \mathbb{N}}\) that is both translation-invariant and 1-dependent, and which has a correlation function that satisfies

\[
\rho_n = P(X_{1+i} = 1, \ldots, X_{n+i} = 1) = \frac{1}{(n + 1)!},
\]

which, by stationarity, is independent of \(i \in \mathbb{N}_0\), the Carries process \((\text{mod } \infty)\).
Observe how for any \( U_1, \ldots, U_{n+1} \) i.i.d. continuous random variables with density function \( f \) each,
\[
P(U_1 < \ldots < U_{n+1}) = \frac{1}{(n+1)!}.
\] (18)

Indeed, notice how
\[
\sum_{\sigma \in S_{n+1}} P(U_{\sigma(1)} < \ldots < U_{\sigma(n+1)}) + P(\exists i, j \in [n+1]: i \neq j, U_i = U_j) = 1
\]

But \( P(\exists i, j \in [n+1]: i \neq j, U_i = U_j) = 0 \) since the \( U_i \)'s are continuous random variables. This implies that
\[
\sum_{\sigma \in S_{n+1}} P(U_{\sigma(1)} < \ldots < U_{\sigma(n+1)}) = 1.
\] (19)

Then notice how for any \( \sigma \in S_{n+1}, \)
\[
P(U_1 < \ldots < U_{n+1}) = \int_{\{ (\tilde{u}_1, \ldots, \tilde{u}_{n+1}) \in \text{supp}(f)_{n+1}; \tilde{u}_1 < \ldots < \tilde{u}_{n+1} \}} f_{U_1, \ldots, U_{n+1}}(u_1, \ldots, u_{n+1}) du_1 \ldots du_{n+1}
\] (20)
\[
= \int_{\{ (\tilde{u}_1, \ldots, \tilde{u}_{n+1}) \in \text{supp}(f)_{n+1}; \tilde{u}_1 < \ldots < \tilde{u}_{n+1} \}} f_{U_1}(u_1) \ldots f_{U_{n+1}}(u_{n+1}) du_1 \ldots du_{n+1} \quad \text{(independence)}
\] (21)
\[
= \int_{\{ \tilde{u}_1 < \ldots < \tilde{u}_{n+1} \}} f(u_1) \ldots f(u_{n+1}) du_1 \ldots du_{n+1} \quad \text{(id. distribution)}
\] (22)
\[
= \int_{\{ \tilde{u}_1 < \ldots < \tilde{u}_{n+1} \}} f_{U_{\sigma(1)}}(u_1) \ldots f_{U_{\sigma(n+1)}}(u_{n+1}) du_1 \ldots du_{n+1}
\] (23)
\[
= \int_{\{ \tilde{u}_1 < \ldots < \tilde{u}_{n+1} \}} f_{U_{\sigma(1)}, \ldots, U_{\sigma(n+1)}}(u_1, \ldots, u_{n+1}) du_1 \ldots du_{n+1} \quad \text{(independence)}
\] (24)
\[
= P(U_{\sigma(1)} < \ldots < U_{\sigma(n+1)}).
\] (25)

Therefore, (19) becomes
\[
(n+1)! \cdot P(U_1 < \ldots < U_{n+1}) = 1
\]
as stated.

One, additionally, could prove this via integrals:

Denote the cumulative distribution function of every \( U_i \) by \( F \) (and so \( F(u) := P(U_i < u) \) and \( f(u) = \frac{dF(u)}{du} \)).
We carry out the proof for \( n = 3 \) inasmuch as it contains all the essential features of the general case. So,

\[
P(U_1 < U_2 < U_3) = \int_{\text{supp}(f)} \int_{\text{supp}(f)} P(U_1 < U_2 < U_3 | u_2 = U_2 < U_3 = u_3) f_{U_2,U_3}(u_2, u_3) du_2 du_3
\]

\[
= \int_{\text{supp}(f)} \int_{\text{supp}(f)} P(U_1 < U_2 < U_3 | u_2 = U_2 < U_3 = u_3) f(u_2)f(u_3) du_2 du_3 \quad \text{(i.i.d. property)}
\]

\[
= \int_{\text{supp}(f)} \int_{\{u_2 < u_3\}} F(u_2) dF(u_2) dF(u_3)
\]

\[
= \int_{\text{supp}(f)} \int_{\{u_2 < u_3\}} \frac{F^2(u_2)}{2} dF(u_3)
\]

\[
= \frac{1}{2} \int_{\text{supp}(f)} F^2(u_3) dF(u_3)
\]

\[
= \frac{1}{2} \int_{\text{supp}(f)} \frac{F^3(u_3)}{3} du_3
\]

\[
= \frac{1}{2} \left( \frac{1^3}{3} - \frac{0^3}{3} \right)
\]

\[
= \frac{1}{3!}.
\]

as required.

So for simplicity let’s let \( U_1,\ldots,U_{n+1} \) be \( Unif[0,1] \) i.i.d. random variables.

Then by (18) we have

\[
\rho_n \equiv P(X_1 = 1,\ldots,X_n = 1) = P(U_1 < \ldots < U_{n+1}) = \frac{1}{(n+1)!}.
\]

(26)

So similar to having \( X_i = 1 \iff B_i > B_{i+1} \) when we were working in the base \( b < \infty \) case, we now have \( X_i = 1 \iff U_i < U_{i+1} \) in the base \( b \to \infty \) case.

It is also immediate from (25) that

\[
\rho_n \equiv P(X_1 = 1,\ldots,X_n = 1) = P(U_1 < \ldots < U_{n+1})
\]

\[
= P(U_1 > \ldots > U_{n+1}) \quad \text{(by the previous symmetry argument)}
\]

\[
= P(U_1 \geq \ldots \geq U_{n+1}) \quad \text{(since the } U_i's \text{ are continuous random variables)}
\]

\[
= P(X_1 = 0,\ldots,X_n = 0).
\]

So for this Carries process the probability of getting a gap of size \( n \) is equal to \( \frac{1}{(n+1)!} \). Even though
the stationarity and 1-dependence of the Carries process with base \( b < \infty \) that were established in Propositions 2.5 and 2.6 are inherited in the limit \( b \to \infty \), and hence by results from Section 4 this Carries process is determinantal, the determinant of the associated kernel is not needed to compute the statistic we are interested in - we already obtain a very simple expression for the \( n \)-gap probabilities!

But what if we considered the corresponding \( p \)-spatially independently thinned process of this \((X_i)_{i \in \mathbb{N}}\) Carries process, denoted by \((X_i^{(p)})_{i \in \mathbb{N}}\), that was introduced in Section 2.1? Then the corresponding gap probability

\[
P(X_1^{(p)} = 0, \ldots, X_n^{(p)} = 0) =: \tilde{\rho}_n
\]

is no longer immediately straightforward, and there is a dependence on \( q := 1 - p \) which we would like to explore. One need only calculate these probabilities for some small \( n \) to be convinced of this:

\[
P(X_1^{(p)} = 0) = P\left(\{X_1 = 0\} \cup \{X_1 = 1, \text{point 1 gets thinned}\}\right)
= P(X_i = 0) + P(X_i = 1) \cdot q
= \frac{1 + q}{2}.
\]

\[
P(X_1^{(p)} = X_2^{(p)} = 0) = P\left(\{X_1 = X_2 = 0\}\right) + P\left(\{X_1 = 1, X_2 = 0, \text{point 1 gets thinned}\}\right)
+ P\left(\{X_1 = 0, X_2 = 1, \text{point 2 gets thinned}\}\right)
+ P\left(\{X_1 = 1, X_2 = 1, \text{points 1 and 2 get thinned}\}\right)
= \frac{1}{3!} + q \cdot P(X_1 = 1, X_2 = 0) + q \cdot P(X_1 = 0, X_2 = 1) + q^2 \cdot P(X_1 = 1, X_2 = 1)
= \frac{1}{3!} + q \cdot \left(P(X_1 = 1) - P(X_1 = X_2 = 1)\right) + q \cdot \left(P(X_1 = 0) - P(X_1 = X_2 = 0)\right) + q^2 \cdot \frac{1}{3!}
= \frac{1}{3!} + 2q \left(\frac{1}{2} - \frac{1}{3!}\right) + q^2 \cdot \frac{1}{3!}.
\]

We are already at \( n = 2 \) and we are beginning to rely on inclusion-exclusion as in (6) in order to get the gap probabilities.

It is in this thinned Carries process \((X_i^{(p)})_{i \in \mathbb{N}}\), therefore, where we need to exploit the fact that it has some special structure, namely that it is determinantal (recall Theorem 2.2) with associated kernel being Toeplitz (as we shall see).

We shall further develop the theory of 1-dependent point processes later on in Section 4.
3 Toeplitz matrices

In this section we give an introduction to Toeplitz matrices, their basic properties and the terminology used in the literature (which we’ll also use extensively). We also provide an exposition and discussion of some well known results regarding the asymptotic behaviour of their determinants and eigenvalues for large orders (the preeminent Szégo theorems). Furthermore, we discuss the challenges faced in this project of having to deal with non-Hermitian Toeplitz matrices (recall we had mentioned earlier that the kernel for the point process of interest in this project is non-Hermitian Toeplitz) and how we overcame these as discussed in the introductory section - the main obstacle being, recall, that the aforementioned Szégo theorems cannot hold in the non-Hermitian setting (as explained in [19] also).

3.1 Introduction to Toeplitz matrices

Let’s first recall the definition of a Toeplitz matrix.

**Definition (from [7])** A Toeplitz matrix is an $n \times n$ matrix $T_n = (t_{ij})_{1 \leq i,j \leq n} \in \mathbb{C}^{n \times n}$ with entries $t_{ij} = t_{j-i}$ only depending on the difference between the column $j$ and row $i$ of their coordinates in the matrix, i.e., it’s a matrix of the form

$$T_n = \begin{pmatrix} t_0 & t_1 & t_2 & \cdots & t_{n-1} \\ t_{-1} & t_0 & t_1 & \cdots & t_{n-2} \\ t_{-2} & t_{-1} & t_0 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_{-(n-1)} & \cdots & \cdots & \cdots & t_0 \end{pmatrix}$$  \hspace{1cm} (27)

These special kinds of matrices have many applications, besides for this project, throughout mathematics and statistics (e.g., see Section 1.2 of [7]).

We can then define a sequence $(T_n)_{n \geq 1}$ of Toeplitz matrices, where $T_n \in \mathbb{C}^{n \times n}$ is as in the previous definition, by defining an infinite sequence $(t_n)_{n \in \mathbb{Z}}$ of complex numbers.

Now, the classic Szégo theorems (see [23], [8]) only require the weaker constraint

$$\sum_{n \in \mathbb{Z}} |t_n|^2 < \infty,$$

18
but the proof is significantly more involved than if we were to assume the stronger constraint

\[ \sum_{n \in \mathbb{Z}} |t_n| < \infty. \] (28)

For simplicity and for the purposes of the project, we shall only assume the (28) constraint for the \( t_n \)'s.

It is basic Fourier analysis that this constraint ensures both the existence of the Fourier series

\[ f(\lambda) = \sum_{n \in \mathbb{Z}} t_n e^{in\lambda}, \quad \lambda \in [0, 2\pi], \]

and its uniform convergence.

Indeed,

\[ \sum_{n \in \mathbb{Z}} |t_n e^{in\lambda}| = \sum_{n \in \mathbb{Z}} |t_n| < \infty \]

exists, which proves the existence of the Fourier series (since absolute convergence implies convergence), and, since

\[ |t_n e^{in\lambda}| = |t_n| \quad \forall n \in \mathbb{Z} \quad \forall \lambda \in [0, 2\pi], \]

then (28) implies, by the Weierstrass M-test, that the Fourier series is uniformly convergent.

Therefore, since \((\sum_{k=-n}^{n} t_k e^{ik\lambda})_{n\geq1}\) is a sequence of continuous functions (it’s a sequence of polynomials), then, by a classic analysis result that states that any uniformly convergent sequence of continuous functions is again continuous, we get that the function \( f \) from above is continuous as well.

We can then recover the \( t_n \) Fourier coefficients of \( f \):

\[
\frac{1}{2\pi} \int_{0}^{2\pi} f(\lambda)e^{-in\lambda}d\lambda = \frac{1}{2\pi} \int_{0}^{2\pi} \left( \sum_{k \in \mathbb{Z}} t_k e^{ik\lambda} \right)e^{-in\lambda}d\lambda
\]

\[ = \sum_{k \in \mathbb{Z}} t_k \cdot \frac{1}{2\pi} \int_{0}^{2\pi} e^{i(k-n)\lambda}d\lambda. \quad \text{(by uniform convergence of the Fourier series)} \]

\[ = t_n \quad \text{(well known trick in Fourier analysis).} \]

This is commonly known as the Fourier inversion formula.

Having set the groundwork, we introduce some common terminology which will be used frequently:
Definition Let
\[ f(\lambda) = \sum_{n \in \mathbb{Z}} t_n e^{in\lambda}, \quad \lambda \in [0, 2\pi], \]
be a Fourier series that satisfies (28). Then define the sequence of Toeplitz matrices \((T_n(f))_{n \geq 1}\), where \(T_n(f) \in \mathbb{C}^{n \times n}\) is the matrix from (27).

As explained previously, the (28) assumption implies that for all \(n \in \mathbb{Z}\),
\[ t_n = \frac{1}{2\pi} \int_{0}^{2\pi} f(\lambda) e^{-in\lambda} d\lambda. \tag{29} \]
Therefore, \(f\) determines this sequence of Toeplitz matrices. We call this \(f\) the symbol for the sequence \((T_n(f))_{n \geq 1}\).

Remark It is easy to check (see p.40, [7]) from the previous explanations and the previous definition that for \(n \in \mathbb{N}\),
\[ T_n(f) \text{ is Hermitian } \iff \text{the symbol } f \text{ is real}. \]

The reason why we introduced the notion of a symbol is because it is often of interest to begin with a Fourier series \(f\) that satisfies (28) and then define the sequence of Toeplitz matrices \((T_n(f))_{n \geq 1}\) via the Fourier inversion formula as explained previously. One reason being that it is often a difficult task, even for matrices of a special type (like Toeplitz), to compute eigenvalues and find bounds for their magnitude - let alone figuring out their limiting behaviour - and so having some function (the symbol in this case) that determines the matrices means that there is a way to learn about the eigenvalues' behaviour by studying just this one function. This is indeed the case here with Toeplitz matrices and their symbols. To proceed we need the notion of an essential supremum/infimum of a function:

Definition (p.48, [7]) Let \(f\) be a real-valued function.

We define the essential supremum \(M_f = \text{ess sup} f\) of \(f\) as the smallest number \(a\) for which \(f(x) \leq a\) except on a set of Lebesgue measure 0.

Similarly, we define the essential infimum \(m_f = \text{ess inf} f\) of \(f\) as the largest number \(a\) for which \(f(x) \geq a\) except on a set of Lebesgue measure 0.

The necessity of the above two definitions for our analysis comes from the fact that all the entries of the Toeplitz matrix \(T_n(f)\) are determined by an integral involving \(f\) (namely, the Fourier inversion formula), and so it is naturally expected for the "behaviour" of the eigenvalues of \(T_n(f)\) to also be...
influenced by integrals involving $f$, which, in turn, are not affected by sets of measure 0. Hence, to avoid unnecessary technical difficulties arising from values of $f$ on sets that do not even contribute to integrals involving $f$, we have formulated analogous notions of the supremum and infimum of a function that circumvent these (possible) issues.

By the (28) assumption it is easy to see that

$$-\infty < m_f \leq M_f < \infty.$$ 

Let’s now state some known results for the bounds of the eigenvalues of the $(T_n(f))_{n \geq 1}$ sequence of Toeplitz matrices. Firstly, from [7] we get the following bounds.

**Lemma 3.1** Let $\lambda_0(n), \ldots, \lambda_{n-1}(n)$ be the eigenvalues of $T_n(f)$, $n \in \mathbb{N}$.

If $f$ is real (equivalently, $T_n(f)$ Hermitian), then the eigenvalues are uniformly (in $n$) bounded between $m_f$ and $M_f$, that is,

$$m_f \leq \lambda_i(n) \leq M_f \quad \forall i \in \{0, \ldots, n-1\} \quad \forall n \in \mathbb{N}. \quad (30)$$

By dropping the Hermitianity assumption on $T_n(f)$ (i.e., assuming $f$ is complex-valued),

$$|\lambda_i(n)| \leq 2M_{|f|} \quad \forall i \in \{0, \ldots, n-1\} \quad \forall n \in \mathbb{N}. \quad (31)$$

**Proof** The proof of (30) is given in the proof of Lemma 4.1 from [7].

To obtain (31) we make use of (4.12) from [7] involving the ”strong” matrix norm $\| \cdot \|$. That is,

$$\sqrt{\max_{z \in \mathbb{C}^n : z^*z = 1} z^* T_n(f)^* T_n(f) z} =: \|T_n(f)\| \leq 2M_{|f|}. \quad (32)$$

Let $x^{(i)} \in \mathbb{C}^n$ be a normalized (in the sense that $x^*x = 1$) eigenvector of $\lambda_i(n)$. Then,

$$x^* T_n(f)^* T_n(f) x = (T_n(f)x)^* (T_n(f)x)$$

$$= \lambda_i(n)^* \lambda_i(n) x^* x \quad \text{(by the definition of an eigenvalue and eigenvector)}$$

$$= |\lambda_i(n)|^2 \quad \text{(by the } x^*x = 1 \text{ assumption}).$$

And so

$$|\lambda_i(n)| = \sqrt{x^* T_n(f)^* T_n(f) x} \leq \sqrt{\max_{z \in \mathbb{C}^n : z^*z = 1} z^* T_n(f)^* T_n(f) z} \leq 2M_{|f|}. \quad (32)$$

Some of our findings in Section 5.2 on the point process of interest (the Carries process) alluded to
the existence of an even tighter bound to the bound (31) that was given in [7] for non-Hermitian matrices. Namely, our findings alluded to the removal of the 2-factor from the (31) bound. We eventually were able to prove this allusion in Theorem 5.9 and hence improve the above result from [7].

### 3.2 Szegő-type asymptotics for Toeplitz matrices

Let’s first discuss the Szegő theorems regarding the asymptotics of large Toeplitz determinants that already exist in the literature.

The first result, originally appearing in [8], p.65, presented in [7] under the (weaker) assumption that the symbol $f$ satisfies (28), which, as mentioned before, we shall always assume is the case in this section, regarding the limiting behaviour of Toeplitz determinants was:

**Theorem 3.2 (Thm. 5.4, [7])** Let $(T_n(f))_{n \geq 1}$ be a sequence of Hermitian matrices that satisfy (28) such that $\log f(\lambda)$ is Riemann-integrable and $f(\lambda) \geq m_f > 0$. Then,

$$
\lim_{n \to \infty} (\det T_n(f))^{1/n} = e^{\frac{1}{2\pi} \int_0^{2\pi} \log f(\lambda) d\lambda}.
$$

**Remark** We can take logarithms on both sides of

$$
\lim_{n \to \infty} (\det T_n(f))^{1/n} = e^{\frac{1}{2\pi} \int_0^{2\pi} \log f(\lambda) d\lambda}
$$

from the previous result because, by assumption $m_f > 0$, and so by (30) from Lemma 3.1, all eigenvalues of $T_n(f)$ are greater than 0 independently of $n$, and so since the determinant of $T_n(f)$ is the product of its eigenvalues, it must also be strictly positive and so the logarithm of it is well-defined.

So taking logarithms on both sides yields

$$
\log \left( \lim_{n \to \infty} (\det T_n(f))^{1/n} \right) = \frac{1}{2\pi} \int_0^{2\pi} \log f(\lambda) d\lambda,
$$

which, by the continuity of log, is equivalent to

$$
\lim_{n \to \infty} \left( \log (\det T_n(f))^{1/n} \right) = \frac{1}{2\pi} \int_0^{2\pi} \log f(\lambda) d\lambda.
$$
Then, by well-known properties of the logarithm, we have

\[ \lim_{n \to \infty} \frac{1}{n} \log \left( \det T_n(f) \right) = \frac{1}{2\pi} \int_0^{2\pi} \log f(\lambda) d\lambda. \]

This implies that

\[ \frac{1}{n} \log \left( \det T_n(f) \right) = \frac{1}{2\pi} \int_0^{2\pi} \log f(\lambda) d\lambda \cdot \epsilon(n), \quad n \in \mathbb{N}, \]

where \( \epsilon(n) \) is the error term as a result of having taken out the limit as \( n \to \infty \). And so

\[ \det T_n(f) = e^{\frac{n}{2\pi} \int_0^{2\pi} \log f(\lambda) d\lambda \cdot \epsilon(n)}, \quad n \in \mathbb{N}, \quad (33) \]

where \( \mathcal{E}(n) := e^{\epsilon(n)} \) is an error term.

It is then of interest the limit as \( n \to \infty \) of \( \mathcal{E}(n) \), which is exactly equal to

\[ \lim_{n \to \infty} \frac{\det T_n(f)}{e^{\frac{n}{2\pi} \int_0^{2\pi} \log f(\lambda) d\lambda}}. \quad (34) \]

By denoting the above limit by \( \alpha \), we can conclude that for large \( n \),

\[ \det T_n(f) \approx \alpha \cdot e^{\frac{n}{2\pi} \int_0^{2\pi} \log f(\lambda) d\lambda}. \quad (35) \]

We shall call

\[ e^{\frac{n}{2\pi} \int_0^{2\pi} \log f(\lambda) d\lambda} \]

the Szegő leading term, and

\[ \alpha := \lim_{n \to \infty} \frac{\det T_n(f)}{e^{\frac{n}{2\pi} \int_0^{2\pi} \log f(\lambda) d\lambda}} \]

the Szegő error term.

By making some further assumptions on the symbol \( f \) that determines the sequence of Toeplitz matrices \( (T_n(f))_{n \geq 1} \), Szegő and Grenander in [8], p.76, found a sharp and explicit expression for the Szegő error term:

**Theorem 3.3 (see [13])** If the symbol \( f \) is positive and its derivative satisfies a Lipschitz condition (with exponent \( 0 < \alpha \leq 1 \)), then

\[ \lim_{n \to \infty} \frac{\det T_n(f)}{e^{\frac{n}{2\pi} \int_0^{2\pi} \log f(\lambda) d\lambda}} = \exp \left\{ \frac{1}{4} \sum_{m=1}^{\infty} m |k_m|^2 \right\}, \]

23
where
\[
\sum_{m=1}^{\infty} k_m z^m = \frac{1}{2\pi} \int_{0}^{2\pi} \log f(\lambda) \frac{1 + ze^{-i\lambda}}{1 - ze^{-i\lambda}} d\lambda.
\]

**Remark** Of course, even without explicitly stating that we need the symbol \( f \) to satisfy the (28) assumption (which, like mentioned many times, we’ll always assume is the case in this section), the \( f' \) satisfying a Lipschitz condition implies that
\[
\sum_{n \in \mathbb{Z}} |n| t_n < \infty,
\]
where, recall, the \( t_n \)'s are the Fourier coefficients of \( f \). And from this, (28) follows.

We have just stated and explained the two main Szegő results (Theorem 3.2 and Theorem 3.3), which only hold for Hermitian Toeplitz matrices. Indeed, by following the proof explained in the 98 page book [7] (which is just a more reader-friendly version of the original 248 page long [8]), one would be convinced that all of the ingenious constructions, arguments and notions of asymptotic equivalence of matrices, the constructions of circulant matrices that are asymptotically equivalent to general Hermitian Toeplitz matrices, and many more, would break down without this Hermitianity assumption. As mentioned before, the proof is very long, but ingenious, and there are just too many constructions and technical lemmas involved for it to be summarized and explained here as to why Hermitianity is important. Very briefly, and admittedly, not very accurately, it’s due to the fact that without the Hermitianity, eigenvalues are not necessarily real, which causes problems when we want to compare them with values of the respective symbol (as in (30)), since complex numbers are incomparable in and of themselves (except for when we take their modulus), or when we take their logarithm.

This, as mentioned earlier in the project, presents a problem in our analysis of the Carries point process because as we will see, its kernel is non-Hermitian Toeplitz. Thus, none of the Szegő results just stated are applicable. Also, the literature for non-Hermitian Toeplitz matrices, as mentioned many times, is rather lacking - which presents yet another challenge.

Our first step in overcoming this difficulty is by studying all the different arguments in the previously mentioned papers on Toeplitz matrices ([13] in particular) and see if we can make any adaptations to them in order to make them work (perhaps not the most optimally) for the non-Hermitian case. We present our first “Szegő-type” theorem that works for non-Hermitian Toeplitz matrices, which treats the Szegő leading term:

**Theorem 3.4** Let \( (T_n(f))_{n \geq 1} \) be the sequence of Toeplitz matrices \( T_n(f) \in \mathbb{C}^{n \times n} \) determined by
the symbol

\[ f : [0, 2\pi] \to \mathbb{C}, \quad f(\lambda) = \sum_{k=-\infty}^{\infty} t_k e^{ik\lambda}, \]

so that \((T_n(f))_{i,j} = t_{j-i}\) for every \(n \geq 1\) and every \(i,j \in [n]\). As always, we assume

\[ \sum_{k=-\infty}^{\infty} |t_k| < \infty. \]

Denote by \(\lambda_0(n), \ldots, \lambda_{n-1}(n)\) the \(n\) eigenvalues of \(T_n(f)\). Let \(p\) be sufficiently small such that

\[ |p\lambda_i(n)| < 1 \quad \forall n \in \mathbb{N} \quad \forall i = 0, \ldots, n-1 \quad (36) \]

and

\[ |pf(\lambda)| < 1 \quad \forall \lambda \in [0, 2\pi]. \quad (37) \]

Then,

\[ \det(I_n - pT_n(f)) = \exp \left\{ \frac{n}{2\pi} \int_{0}^{2\pi} \log(1 - pf(\lambda))d\lambda \right\} \cdot \exp \left\{ - \sum_{k=1}^{\infty} \epsilon_k(n) \frac{p^k}{k} \right\}, \]

where \(I_n\) denotes the \(n \times n\) identity matrix, and

\[ \epsilon_k(n) = \sum_{i=0}^{n-1} (\lambda_i(n))^k - \frac{n}{2\pi} \int_{0}^{2\pi} (f(\lambda))^k d\lambda \quad (38) \]

is an error term which appears later in the proof.

**Proof** To ease notations we shall denote \(T_n \equiv T_n(f)\) for the purposes of this proof.

The determinant of a matrix is the product of its eigenvalues. Therefore,

\[ \det(I_n - pT_n) = \prod_{i=0}^{n-1} (1 - p\lambda_i(n)). \]

And so

\[ \log \left( \det(I_n - pT_n) \right) = \sum_{i=0}^{n-1} \log(1 - p\lambda_i(n)) \]

\[ = - \sum_{i=0}^{n-1} \sum_{k=1}^{\infty} \frac{p^k(\lambda_i(n))^k}{k} \quad ((36) \text{ ensures the existence of these Taylor expansions}) \]

\[ = - \sum_{k=1}^{\infty} \frac{p^k}{k} \cdot tr(T_k^n) \quad \left( (\lambda_i(n))^k \text{ is an eigenvalue of } T_k^n \right). \]
Now,

\[
\text{tr}(T_n^k) = \sum_{i=0}^{n-1} (T_n)_i^i
\]

\[
= \sum_{0 \leq i_0, \ldots, i_{k-1} \leq n-1} (T_n)_{i_0} (T_n)_{i_1} (T_n)_{i_2} \cdots (T_n)_{i_{k-2}} (T_n)_{i_{k-1}} i_0
\]

\[
= \sum_{0 \leq i_0, \ldots, i_{k-1} \leq n-1} t_{i_1-i_0} t_{i_2-i_1} \cdots t_{i_{k-1}-i_{k-2}} t_{i_0-i_{k-1}}.
\]

We define the following indicator functions:

\[
\psi_n(i) = \begin{cases} 
1, & \text{if } i \in \{0, \ldots, n-1\}, \\
0, & \text{otherwise}.
\end{cases}, \quad n \in \mathbb{N}
\]

These indicator functions allow us to write the expression for \(\text{tr}(T_n^k)\) we obtained above as infinite summations:

\[
\text{tr}(T_n^k) = \sum_{-\infty < i_0, \ldots, i_{k-1} < \infty} \psi_n(i_0) \psi_n(i_1) \cdots \psi_n(i_{k-1}) t_{i_1-i_0} t_{i_2-i_1} \cdots t_{i_0-i_{k-1}}.
\]

Now let’s do a change of coordinates:

\[
j_0 = i_0, j_1 = i_1 - i_0, j_2 = i_2 - i_1, \ldots, j_{k-1} = i_{k-1} - i_{k-2}.
\]

So we can now rewrite the previous equation as:

\[
\text{tr}(T_n^k) = \sum_{-\infty < j_0, \ldots, j_{k-1} < \infty} \psi_n(j_0) \psi_n(j_0 + j_1) \cdots \psi_n(j_0 + \ldots + j_{k-1}) t_{j_1} t_{j_2} \cdots t_{j_{k-1}} t_{j_1 - j_0 - \ldots - j_{k-1}}
\]

\[
= \sum_{j_0 = -\infty}^{\infty} \psi_n(j_0) \sum_{-\infty < j_1, \ldots, j_{k-1} < \infty} \psi_n(j_0 + j_1) \cdots \psi_n(j_0 + \ldots + j_{k-1}) t_{j_1} t_{j_2} \cdots t_{j_{k-1}} t_{j_1 - j_0 - \ldots - j_{k-1}},
\]

where the interchange of the sums is justified by Fubini’s theorem, since

\[
\sum_{-\infty < j_0, \ldots, j_{k-1} < \infty} |\psi_n(j_0) \psi_n(j_0 + j_1) \cdots \psi_n(j_0 + \ldots + j_{k-1}) t_{j_1} t_{j_2} \cdots t_{j_{k-1}} t_{j_1 - j_0 - \ldots - j_{k-1}}|
\]

is, by definition of the \(\psi_n\)’s, less than or equal to

\[
n \cdot \sum_{-\infty < j_1, \ldots, j_{k-1} < \infty} |t_{j_1} t_{j_2} \cdots t_{j_{k-1}} t_{j_1 - j_0 - \ldots - j_{k-1}}|,
\]
which can easily be seen to be finite by the (28) assumption.

By making use of the definition of $\psi_i(n)$ we can further evaluate the expression for $tr(T_n^k)$:

$$tr(T_n^k) = \sum_{j_0=0}^{n-1} \sum_{-\infty<j_1,\ldots,j_{k-1}<\infty} \psi_{n}(j_0 + j_1) \cdots \psi_{n}(j_0 + \ldots + j_{k-1}) t_{j_1} t_{j_2} \cdots t_{j_{k-1}} t^{\ell_{j_1} - \ldots - \ell_{j_{k-1}}}$$

$$= \sum_{j_0=0}^{n-1} \sum_{-\infty<j_1,\ldots,j_{k-1}<\infty} t_{j_1} t_{j_2} \cdots t_{j_{k-1}} t^{\ell_{j_1} - \ldots - \ell_{j_{k-1}}} + \epsilon_k(n),$$

where $\epsilon_k(n)$ is the error term due to removing the indicator functions from the above summations, and, since the latter depend on $n$, so does this error term.

We further evaluate the expression:

$$tr(T_n^k) = n \cdot \sum_{-\infty<j_0,\ldots,j_{k-1}<\infty} t_{j_1} t_{j_2} \cdots t_{j_{k-1}} t_{j_0} + \epsilon_k(n)$$

$$= n \cdot \sum_{-\infty<j_0,\ldots,j_{k-1}<\infty} t_{j_1} t_{j_2} \cdots t_{j_{k-1}} t_{j_0} + \epsilon_k(n)$$

$$= n \cdot \sum_{-\infty<j_0,\ldots,j_{k-1}<\infty} t_{j_0} t_{j_1} \cdots t_{j_{k-1}} \cdot \delta_{j_0 + j_1 + \ldots + j_{k-1} = 0} + \epsilon_k(n)$$

$$= n \cdot \sum_{-\infty<j_0,\ldots,j_{k-1}<\infty} t_{j_0} t_{j_1} \cdots t_{j_{k-1}} \cdot \left( \frac{1}{2\pi} \int_{0}^{2\pi} e^{i(j_0 + \ldots + j_{k-1}) \lambda} d\lambda \right) + \epsilon_k(n)$$

We can interchange integral and summation due to the (28) condition that is satisfied by the symbol $f$. Indeed, consider

$$\int_{0}^{2\pi} \left( \sum_{-\infty<j_0,\ldots,j_{k-1}<\infty} t_{j_0} t_{j_1} \cdots t_{j_{k-1}} \right) e^{i(j_0 + \ldots + j_{k-1}) \lambda} d\lambda.$$ 

Then notice how

$$|t_{j_0} t_{j_1} \cdots t_{j_{k-1}} e^{i(j_0 + \ldots + j_{k-1}) \lambda}| = |t_{j_0} t_{j_1} \cdots t_{j_{k-1}}| \quad \forall -\infty < j_1, \ldots, j_{k-1} < \infty \quad \forall \lambda \in [0, 2\pi],$$

and

$$\sum_{-\infty<j_0,\ldots,j_{k-1}<\infty} |t_{j_0} t_{j_1} \cdots t_{j_{k-1}}| = \left( \sum_{j_0=-\infty}^{\infty} |t_{j_0}| \right) \cdots \left( \sum_{j_{k-1}=-\infty}^{\infty} |t_{j_{k-1}}| \right) < \infty$$

due to the (28) assumption.
And so then by the Weierstrass M-test,

$$\sum_{-\infty<j_0,\ldots,j_{k-1}<\infty} t_{j_0} t_{j_1} \cdots t_{j_{k-1}} e^{i(j_0+\ldots+j_{k-1})\lambda}$$

converges uniformly in $\lambda \in [0,2\pi]$. And uniform convergence, by a fundamental Analysis result, means that we can interchange the infinite summations with the integral, that is,

$$\sum_{-\infty<j_0,\ldots,j_{k-1}<\infty} t_{j_0} t_{j_1} \cdots t_{j_{k-1}} \left( \int_0^{2\pi} e^{i(j_0+\ldots+j_{k-1})\lambda} d\lambda \right) = \int_0^{2\pi} \left( \sum_{-\infty<j_0,\ldots,j_{k-1}<\infty} t_{j_0} t_{j_1} \cdots t_{j_{k-1}} \right) e^{i(j_0+\ldots+j_{k-1})\lambda} d\lambda.$$

Therefore,

$$\text{tr}(T^k_n) = n \cdot \sum_{-\infty<j_0,\ldots,j_{k-1}<\infty} t_{j_0} t_{j_1} \cdots t_{j_{k-1}} \cdot \left( \frac{1}{2\pi} \int_0^{2\pi} e^{i(j_0+\ldots+j_{k-1})\lambda} d\lambda \right) + \epsilon_k(n)$$

$$= \frac{n}{2\pi} \cdot \int_0^{2\pi} \left( \sum_{-\infty<j_0,\ldots,j_{k-1}<\infty} t_{j_0} t_{j_1} \cdots t_{j_{k-1}} \right) e^{i(j_0+\ldots+j_{k-1})\lambda} d\lambda + \epsilon_k(n)$$

$$= \frac{n}{2\pi} \cdot \int_0^{2\pi} \left( \sum_{j_0=-\infty}^{\infty} t_{j_0} e^{ij_0\lambda} \right) \cdots \left( \sum_{j_{k-1}=-\infty}^{\infty} t_{j_{k-1}} e^{ij_{k-1}\lambda} \right) d\lambda + \epsilon_k(n)$$

$$= \frac{n}{2\pi} \int_0^{2\pi} (f(\lambda))^k d\lambda + \epsilon_k(n).$$

And so, since the trace of a matrix is the sum of its eigenvalues, and the eigenvalues of $T^k_n$ are $\lambda_0(n)^k, \ldots, \lambda_{n-1}(n)^k$, we have

$$\text{tr}(T^k_n) = \sum_{i=0}^{n-1} (\lambda_i(n))^k,$$

and so the error term $\epsilon_k(n)$ is equal to

$$\epsilon_k(n) = \sum_{i=0}^{n-1} (\lambda_i(n))^k - \frac{n}{2\pi} \int_0^{2\pi} (f(\lambda))^k d\lambda$$

as mentioned in the statement of the Theorem.

Now, recall how we had found earlier in the proof that

$$\log \left( \det(I_n - pT_n) \right) = -\sum_{k=1}^{\infty} \frac{p^k}{k} \cdot \text{tr}(T^k_n).$$
We can now plug in the expression we obtained for $tr(T_k^n)$:

$$\log \left( \det(I_n - pT_n) \right) = -\sum_{k=1}^{\infty} \frac{p^k}{k} \cdot \left( \frac{n}{2\pi} \int_0^{2\pi} (f(\lambda))^k d\lambda + \epsilon_k(n) \right)$$

$$= -\frac{n}{2\pi} \sum_{k=1}^{\infty} \int_0^{2\pi} \frac{p^k(f(\lambda))^k}{k} d\lambda - \sum_{k=1}^{\infty} \frac{p^k}{k} \cdot \epsilon_k(n).$$

We can interchange integral with infinite summation due to the (37) assumption:

Indeed, notice how

$$\left| \frac{p^k(f(\lambda))^k}{k} \right| \leq \left| \frac{p^k M^k_{f|}}{k} \right| \quad \forall k \in \mathbb{N} \ \forall \lambda \in [0, 2\pi],$$

and

$$\sum_{k=1}^{\infty} \left| \frac{p^k M^k_{f|}}{k} \right| < \sum_{k=1}^{\infty} (|p| M_{f|})^k < \infty$$

due to the (37) assumption.

Therefore, by the Weierstrass M-test,

$$\sum_{k=1}^{\infty} \frac{p^k(f(\lambda))^k}{k}$$

is uniformly convergent in $\lambda \in [0, 2\pi]$, and so, again, by a fundamental Analysis result we can interchange infinite summation and integral to get

$$\log \left( \det(I_n - pT_n) \right) = \frac{n}{2\pi} \int_0^{2\pi} \log(1 - pf(\lambda)) d\lambda - \sum_{k=1}^{\infty} \frac{p^k}{k} \cdot \epsilon_k(n),$$

which, since we’ve assumed (37) and hence the Taylor expansion around 0 of $\log(1 - pf(\lambda))$ exists for all $\lambda \in [0, 2\pi]$, is equal to

$$\log \left( \det(I_n - pT_n) \right) = \frac{n}{2\pi} \int_0^{2\pi} \log(1 - pf(\lambda)) d\lambda - \sum_{k=1}^{\infty} \frac{p^k}{k} \cdot \epsilon_k(n).$$

Then, by taking the exponential of both sides we are done. 

**Remark** This is indeed a "Szögo-type" theorem since it gives us an expression for the determinant of a Toeplitz matrix of dimension $n \in \mathbb{N}$ in terms of the Szögo leading term

$$\exp \left\{ \frac{n}{2\pi} \int_0^{2\pi} \log(1 - pf(\lambda)) d\lambda \right\}$$
of the sequence of Toeplitz matrices

\[(T_n(1 - pf))_{n \geq 1} \equiv (I_n - pT_n(f))_{n \geq 1}\]  \hspace{1cm} (39)

that are determined by the respective symbol \(1 - pf\).

To see the equivalence that was stated in (39), recall how \(T_n(1 - pf)\) denotes the \(n \times n\) Toeplitz matrix that is determined by the symbol

\[1 - pf(\lambda) = 1 - p \sum_{k=\infty}^{\infty} t_k e^{ik\lambda}, \quad \lambda \in [0, 2\pi].\]  \hspace{1cm} (40)

That is, on its \(k\)th diagonal, \(k = -(n - 1), \ldots, 0, \ldots, n - 1\), it has on all entries the \(k\)th Fourier coefficient of \(1 - pf\). Observe that on its 0th diagonal (the main diagonal) it has on all entries \(1 - pt_0\), since the 0th Fourier coefficient of \(1 - pf\) is \(1 - pt_0\) (as can be seen from (40)). And all the other diagonals are the same exact diagonals as those of the Toeplitz matrix \(T_n(pf) = pT_n(f)\), since the \(k\)th, \(k \neq 0\), Fourier coefficient of \(1 - pf\) is exactly \(pt_k\) (as can be seen from (40)).

**Remark** As for the Szegő error term, notice how, from the statement of our Theorem 3.4 and the equivalence (39) that was explained previously,

\[
\exp\left\{-\sum_{k=1}^{\infty} \epsilon_k(n) \frac{p^k}{k}\right\} = \frac{\det(T_n(1 - pf))}{\exp\left\{\frac{n}{2\pi} \int_{0}^{2\pi} \log(1 - pf(\lambda)) d\lambda\right\}}.
\]

Now, the limit of this as \(n \to \infty\) is exactly the Szegő error term we’ve discussed before and seen in the statement of Theorem 3.3 (for real-valued symbols). We now present our second "Szegő-type" theorem that works for non-Hermitian matrices, now treating the Szegő error term:

**Theorem 3.5** Let \((T_n(f))_{n \geq 1}\) be the sequence of Toeplitz matrices \(T_n(f) \in \mathbb{C}^{n \times n}\) determined by the symbol

\[f : [0, 2\pi] \to \mathbb{C}, \quad f(\lambda) = \sum_{k=\infty}^{\infty} t_k e^{ik\lambda},\]

so that \((T_n(f))_{i,j} = t_{j-i}\) for every \(n \geq 1\) and every \(i, j \in [n]\). As always, we assume

\[A := \sum_{k=\infty}^{\infty} |t_k| < \infty.\]
Denote by \( \lambda_0(n), \ldots, \lambda_{n-1}(n) \) the \( n \) eigenvalues of \( T_n(f) \). Let \( p \) be sufficiently small such that

\[
|p \lambda_i(n)| < 1 \quad \forall n \in \mathbb{N} \ \forall i = 0, \ldots, n - 1
\]

and

\[
|pf(\lambda)| < 1 \quad \forall \lambda \in [0, 2\pi].
\]

Assume further that

\[
B := \sum_{k=-\infty}^{\infty} |k||t_k| < \infty
\]

and

\[
p < \frac{1}{A}.
\]

Then, we obtain the following Szégo error term for the symbol \( 1 - pf \):

\[
\lim_{n \to \infty} \frac{\det(T_n(1 - pf))}{\exp\left\{ \frac{n}{2\pi} \int_0^{2\pi} \log(1 - pf(\lambda))d\lambda \right\}} = \exp\left\{ \sum_{k=1}^{\infty} p_k \sum_{-\infty < j_1, \ldots, j_{k-1} < \infty} t_{j_1} t_{j_2} \cdots t_{j_{k-1}} t_{j_{k-1} - 1} \cdots t_{j_{k-1}} h(j_1, \ldots, j_{k-1}) \right\},
\]

where,

\[
h(j_1, \ldots, j_{k-1}) = \max\{0, j_1, j_1 + j_2, \ldots, j_1 + \ldots + j_{k-1}\} - \min\{0, j_1, j_1 + j_2, \ldots, j_1 + \ldots + j_{k-1}\}.
\]

**Proof** As in the proof of Theorem 3.4, to ease notations we shall denote \( T_n(f) \equiv T_n \).

Recall how we had found in the proof of Theorem 3.4 that

\[
\sum_{i=0}^{n-1} (\lambda_i(n))^k = tr(T_n^k) = \sum_{-\infty < j_0, \ldots, j_{k-1} < \infty} \psi_n(j_0) \psi_n(j_0 + j_1) \cdots \psi_n(j_0 + \ldots + j_{k-1}) t_{j_1} t_{j_2} \cdots t_{j_{k-1}} t_{j_{k-1} - 1} \cdots t_{j_{k-1}}.
\]

We had seen in the proof of Theorem 3.4 that we can interchange the order of the summations:

\[
\sum_{i=0}^{n-1} (\lambda_i(n))^k = \sum_{-\infty < j_1, \ldots, j_{k-1} < \infty} t_{j_1} t_{j_2} \cdots t_{j_{k-1}} t_{j_{k-1} - 1} \cdots t_{j_{k-1}} \sum_{j_0 = -\infty}^{\infty} \psi_n(j_0) \psi_n(j_0 + j_1) \cdots \psi_n(j_0 + \ldots + j_{k-1}).
\]

By following the proof of Theorem 2.1 from [13] we can write

\[
\sum_{j_0 = -\infty}^{\infty} \psi_n(j_0) \psi_n(j_0 + j_1) \cdots \psi_n(j_0 + \ldots + j_{k-1})
\]
in a simpler way for fixed $-\infty < j_1, \ldots, j_{k-1} < \infty$. Namely, the above sum is equal to
\[ n - h(j_1, \ldots, j_{k-1}), \quad \text{if } h(j_1, \ldots, j_{k-1}) \leq n, \]
\[ 0, \quad \text{otherwise} \tag{43} \]
where $h$ is the function given at the end of the statement of Theorem 3.5.

This means that for any $-\infty < j_1, \ldots, j_{k-1} < \infty$ such that
\[ h(j_1, \ldots, j_{k-1}) \leq n, \tag{44} \]
we have
\[ \sum_{j_0 = -\infty}^{j_0 = -\infty} \psi_n(j_0) \psi_n(j_0 + j_1) \cdots \psi_n(j_0 + \ldots + j_{k-1}) = n - h(j_1, \ldots, j_{k-1}). \tag{45} \]

Therefore, for all sufficiently large $n > 0$, the inequality (44) is guaranteed and hence the equality (45) as well.

So for all sufficiently large $n > 0$,
\[ \sum_{i=0}^{n-1} (\lambda_i(n))^k = \sum_{-\infty < j_1, \ldots, j_{k-1} < \infty} t_{j_1} t_{j_2} \cdots t_{j_{k-1}} t_{-j_1 - \ldots - j_{k-1}} \sum_{j_0 = -\infty}^{j_0 = -\infty} \psi_n(j_0) \psi_n(j_0 + j_1) \cdots \psi_n(j_0 + \ldots + j_{k-1}) 
\]
\[ = n \cdot \sum_{-\infty < j_1, \ldots, j_{k-1} < \infty} t_{j_1} \cdots t_{j_{k-1}} t_{-j_1 - \ldots - j_{k-1}} - \sum_{-\infty < j_1, \ldots, j_{k-1} < \infty} t_{j_1} \cdots t_{j_{k-1}} t_{-j_1 - \ldots - j_{k-1}} \cdot h(j_1, \ldots, j_{k-1}). \]

By retracing some of the steps taken in the proof of Theorem 3.4 to evaluate $tr(T_k^n)$, one will find that
\[ \frac{n}{2\pi} \int_0^{2\pi} (f(\lambda))^k d\lambda = n \cdot \sum_{-\infty < j_1, \ldots, j_{k-1} < \infty} t_{j_1} \cdots t_{j_{k-1}} t_{-j_1 - \ldots - j_{k-1}}. \tag{46} \]

Therefore,
\[ \lim_{n \to \infty} \epsilon_k(n) = \lim_{n \to \infty} \left( \sum_{i=0}^{n-1} (\lambda_i(n))^k - \frac{n}{2\pi} \int_0^{2\pi} (f(\lambda))^k d\lambda \right) 
\]
\[ = - \sum_{-\infty < j_1, \ldots, j_{k-1} < \infty} t_{j_1} \cdots t_{j_{k-1}} t_{-j_1 - \ldots - j_{k-1}} \cdot h(j_1, \ldots, j_{k-1}). \]

We will show that the above infinite summations actually converge. It suffices to show they converge absolutely.

For notation sake we shall only consider the $k = 3$ case (the same steps, with a simple inductive argument, can be taken to prove the general $k \in \mathbb{N}$ case).
First recall
\[
\sum_{-\infty<j_1,\ldots,j_2<\infty} t_{j_1} t_{j_2} t_{-j_1-j_2} \cdot h(j_1, j_2) = \sum_{-\infty<j_1,\ldots,j_2<\infty} t_{j_1} t_{j_2} t_{-j_1-j_2} \left( \max\{0, j_1 + j_2\} - \min\{0, j_1, j_2\} \right).
\]

Then notice how
\[
\sum_{-\infty<j_1,\ldots,j_2<\infty} |t_{j_1} t_{j_2} t_{-j_1-j_2}| \cdot |h(j_1, j_2)| \leq 2 \cdot \sum_{-\infty<j_1,\ldots,j_2<\infty} |t_{j_1} t_{j_2} t_{-j_1-j_2}| \max\{|j_1|, |j_1 + j_2|\}.
\]

And so it suffices to prove that
\[
\sum_{-\infty<j_1,\ldots,j_2<\infty} |t_{j_1} t_{j_2} t_{-j_1-j_2}| \max\{|j_1|, |j_1 + j_2|\} < \infty.
\]

We will make use of the following well-known identity regarding the maximum of two numbers:
\[
\max\{a,b\} = \frac{a + b + |a − b|}{2}.
\]

So,
\[
\sum_{-\infty<j_1,\ldots,j_2<\infty} |t_{j_1} t_{j_2} t_{-j_1-j_2}| \max\{|j_1|, |j_1 + j_2|\} = \sum_{-\infty<j_1,\ldots,j_2<\infty} \frac{|j_1| + |j_1 + j_2| + |j_1 + j_2| - |j_1|}{2} |t_{j_1}||t_{j_2}||t_{-j_1-j_2}|
\]
\[
\leq \sum_{-\infty<j_1,\ldots,j_2<\infty} \left(|j_1| + |j_2| + |j_1 + j_2|\right) |t_{j_1}||t_{j_2}||t_{-j_1-j_2}|,
\]
where the above inequality is due to the reverse triangle inequality. We then further evaluate the last expression:
\[
\sum_{-\infty<j_1,j_2<\infty} \left(|j_1| + |j_2| + |j_1 + j_2|\right) |t_{j_1}||t_{j_2}||t_{-j_1-j_2}| \leq 2 \sum_{-\infty<j_1,j_2<\infty} \left(|j_1| + |j_2|\right) |t_{j_1}||t_{j_2}||t_{-j_1-j_2}| \quad (\triangle \text{-inequality})
\]
\[
\leq 2A \sum_{-\infty<j_1,j_2<\infty} \left(|j_1| + |j_2|\right) |t_{j_1}||t_{j_2}|
\]
\[
= 2A \left( \sum_{j_1=-\infty}^{\infty} |j_1||t_{j_1}| \right) \left( \sum_{j_2=-\infty}^{\infty} |t_{j_2}| \right)
\]
\[
+ 2A \left( \sum_{j_2=-\infty}^{\infty} |j_2||t_{j_2}| \right) \left( \sum_{j_1=-\infty}^{\infty} |t_{j_1}| \right)
\]
\[
= 4A^2 B < \infty
\]
This proves that
\[ \sum_{-\infty<j_1,j_2<\infty} \left( |j_1| + |j_2| + |j_1 + j_2| \right) |t_{j_1}||t_{j_2}||t_{-j_1-j_2}| < \infty, \]
which, in turn, proves
\[ \sum_{-\infty<j_1,...,j_2<\infty} |t_{j_1}t_{j_2}t_{-j_1-j_2}| \max \{|j_1|,|j_1+j_2|\} < \infty \]
as required.

Before we proceed, let’s summarize what we currently have.

We just proved that
\[ \lim_{n \to \infty} \epsilon_k(n) = -\sum_{-\infty<j_1,...,j_{k-1}<\infty} t_{j_1} \cdots t_{j_{k-1}} t_{-j_1-\ldots-j_{k-1}} \cdot h(j_1,\ldots,j_{k-1}) \] (47)
exists for all \(k \in \mathbb{N}\).

And recall
\[ \exp \left\{ -\sum_{k=1}^{\infty} \epsilon_k(n) \frac{p^k}{k} \right\} = \frac{\det(T_n(1-pf))}{\exp \left\{ \frac{n}{2\pi} \int_0^{2\pi} \log(1-pf(\lambda))d\lambda \right\}}. \] (48)

By the continuity of the exponential function we are able to pull the limit as \(n \to \infty\) inside the exponent:
\[ \lim_{n \to \infty} \exp \left\{ -\sum_{k=1}^{\infty} \epsilon_k(n) \frac{p^k}{k} \right\} = \exp \left\{ \lim_{n \to \infty} -\sum_{k=1}^{\infty} \epsilon_k(n) \frac{p^k}{k} \right\} \] (49)

But we need further justifications in order to pull the limit as \(n\) tends to infinity inside the infinite summation, and hence, be able to plug in the expression (47) we just proved, which will finally prove the Szegő error term stated in the theorem.

If we can prove uniform convergence (in \(n\)) of
\[ \sum_{k=1}^{\infty} \epsilon_k(n) \frac{p^k}{k}, \]
then, by a classic Analysis result, we will be able to pull the desired limit inside this infinite summation.

We would like to apply the Weierstrass M-test to prove the uniform convergence of the above
infinite summation. For this we need an estimate for \(|\epsilon_k(n)|\).

Recall

\[
\epsilon_k(n) = \sum_{i=0}^{n-1} (\lambda_i(n))^k - \frac{n}{2\pi} \int_0^{2\pi} (f(\lambda))^k d\lambda
\]

\[
= \sum_{-\infty < j_1, \ldots, j_{k-1} < \infty} t_{j_1} t_{j_2} \cdots t_{j_{k-1}} t_{j_1 - \cdots - j_{k-1}} \sum_{j_0 = -\infty}^{\infty} \psi_n(j_0) \psi_n(j_0 + j_1) \cdots \psi_n(j_0 + \cdots + j_{k-1})
\]

\[
- n \cdot \sum_{-\infty < j_1, \ldots, j_{k-1} < \infty} t_{j_1} \cdots t_{j_{k-1}} t_{j_1 - \cdots - j_{k-1}},
\]

where the last equality holds due to (42) and (46).

Define

\[\mathcal{A} := \{(j_1, \ldots, j_{k-1}) \in \mathbb{Z}^{k-1} : h(j_1, \ldots, j_{k-1}) \leq n\} \]

By (43), we have that for \((j_1, \ldots, j_{k-1}) \in \mathcal{A},

\[
\sum_{j_0 = -\infty}^{\infty} \psi_n(j_0) \psi_n(j_0 + j_1) \cdots \psi_n(j_0 + \cdots + j_{k-1}) = n - h(j_1, \ldots, j_{k-1}), \quad (50)
\]

and for \((j_1, \ldots, j_{k-1}) \in \mathcal{A}^C,

\[
\sum_{j_0 = -\infty}^{\infty} \psi_n(j_0) \psi_n(j_0 + j_1) \cdots \psi_n(j_0 + \cdots + j_{k-1}) = 0. \quad (51)
\]

Then, (50) and (51) imply

\[
\epsilon_k(n) = \sum_{-\infty < j_1, \ldots, j_{k-1} < \infty} t_{j_1} t_{j_2} \cdots t_{j_{k-1}} t_{j_1 - \cdots - j_{k-1}} \sum_{j_0 = -\infty}^{\infty} \psi_n(j_0) \psi_n(j_0 + j_1) \cdots \psi_n(j_0 + \cdots + j_{k-1})
\]

\[
- n \cdot \sum_{-\infty < j_1, \ldots, j_{k-1} < \infty} t_{j_1} \cdots t_{j_{k-1}} t_{j_1 - \cdots - j_{k-1}}
\]

\[
= \sum_{(j_1, \ldots, j_{k-1}) \in \mathcal{A}} t_{j_1} t_{j_2} \cdots t_{j_{k-1}} t_{j_1 - \cdots - j_{k-1}} (n - h(j_1, \ldots, j_{k-1}))
\]

\[
- n \cdot \sum_{(j_1, \ldots, j_{k-1}) \in \mathcal{A}} t_{j_1} \cdots t_{j_{k-1}} t_{j_1 - \cdots - j_{k-1}}
\]

\[
- n \cdot \sum_{(j_1, \ldots, j_{k-1}) \in \mathcal{A}^C} t_{j_1} \cdots t_{j_{k-1}} t_{j_1 - \cdots - j_{k-1}}.
\]
Finally,

\[ \epsilon_k(n) = - \sum_{(j_1, \ldots, j_{k-1}) \in A} t_{j_1} t_{j_2} \cdots t_{j_{k-1}} t_{j_1 - \ldots - j_{k-1}} h(j_1, \ldots, j_{k-1}) \]

\[ - n \cdot \sum_{(j_1, \ldots, j_{k-1}) \in A^C} t_{j_1} \cdots t_{j_{k-1}} t_{j_1 - \ldots - j_{k-1}}. \]

Having done this analysis we are now in a position to obtain an estimate for \(|\epsilon_k(n)|\).

By recalling the definition of \(h\), it is easy to see that

\[ |\epsilon_k(n)| \leq \sum_{(j_1, \ldots, j_{k-1}) \in A} |t_{j_1} t_{j_2} \cdots t_{j_{k-1}} t_{j_1 - \ldots - j_{k-1}}| \cdot \left( \max \{0, j_1, j_1 + j_2, \ldots, j_1 + \ldots + j_{k-1}\} - \min \{0, j_1, j_1 + j_2, \ldots, j_1 + \ldots + j_{k-1}\} \right) \]

\[ + n \cdot \sum_{(j_1, \ldots, j_{k-1}) \in A^C} |t_{j_1} \cdots t_{j_{k-1}} t_{j_1 - \ldots - j_{k-1}}|. \]

By the definition of \(A^C\),

\[ n \leq h(j_1, \ldots, j_{k-1}), \]

and so

\[ |\epsilon_k(n)| \leq \sum_{(j_1, \ldots, j_{k-1}) \in A} |t_{j_1} t_{j_2} \cdots t_{j_{k-1}} t_{j_1 - \ldots - j_{k-1}}| \cdot \left( \max \{0, j_1, j_1 + j_2, \ldots, j_1 + \ldots + j_{k-1}\} - \min \{0, j_1, j_1 + j_2, \ldots, j_1 + \ldots + j_{k-1}\} \right) \]

\[ + \sum_{(j_1, \ldots, j_{k-1}) \in A^C} |t_{j_1} \cdots t_{j_{k-1}} t_{j_1 - \ldots - j_{k-1}}| \cdot \left( \max \{0, j_1, j_1 + j_2, \ldots, j_1 + \ldots + j_{k-1}\} - \min \{0, j_1, j_1 + j_2, \ldots, j_1 + \ldots + j_{k-1}\} \right) \]

\[ = \sum_{-\infty < j_1, \ldots, j_{k-1} < \infty} |t_{j_1} t_{j_2} \cdots t_{j_{k-1}} t_{j_1 - \ldots - j_{k-1}}| \cdot \left( \max \{0, j_1, j_1 + j_2, \ldots, j_1 + \ldots + j_{k-1}\} - \min \{0, j_1, j_1 + j_2, \ldots, j_1 + \ldots + j_{k-1}\} \right) \]

\[ \leq 2 \cdot \sum_{-\infty < j_1, \ldots, j_{k-1} < \infty} |t_{j_1} t_{j_2} \cdots t_{j_{k-1}} t_{j_1 - \ldots - j_{k-1}}| \cdot \max \{|j_1|, |j_1 + j_2|, \ldots, |j_1 + \ldots + j_{k-1}|\}. \]

Notice how the above is a uniform estimate (in \(n\)). Let’s further evaluate it.

\[ |\epsilon_k(n)| \leq 2 \cdot \sum_{-\infty < j_1, \ldots, j_{k-1} < \infty} |t_{j_1} t_{j_2} \cdots t_{j_{k-1}} t_{j_1 - \ldots - j_{k-1}}| \cdot \max \{|j_1|, |j_1 + j_2|, \ldots, |j_1 + \ldots + j_{k-1}|\} \]
\[
\leq 2 \cdot A \sum_{-\infty < j_1, \ldots, j_{k-1} < \infty} |t_{j_1} t_{j_2} \cdots t_{j_{k-1}}| \cdot \max \{|j_1|, |j_1 + j_2|, \ldots, |j_1 + \ldots + j_{k-1}|\}
\]
\[
\leq 2 \cdot A \sum_{-\infty < j_1, \ldots, j_{k-1} < \infty} |t_{j_1} t_{j_2} \cdots t_{j_{k-1}}| \cdot (|j_1| + |j_2| + \ldots + |j_{k-1}|)
\]
\[
= 2 \cdot A \cdot \sum_{r=1}^{k-1} \left( \sum_{j_r=\infty}^{\infty} |j_r| |t_{j_r}| \right) \left( \prod_{1 \leq l \leq k-1, j_l=\infty} \sum_{l \neq r} |t_{j_l}| \right)
\]
\[
= 2 \cdot A \cdot \sum_{r=1}^{k-1} B \left( \prod_{1 \leq l \leq k-1, l \neq r} A \right) = 2 \cdot A \cdot B \sum_{r=1}^{k-1} A^{k-2}
\]
\[
= 2(k-1)A^{k-1}B.
\]

Now consider
\[
\sum_{k=1}^{\infty} \epsilon_k(n) \frac{p^k}{k}.
\]

We have
\[
\left| \frac{\epsilon_k(n) p^k}{k} \right| \leq \frac{2B}{A} \cdot \frac{k-1}{k} \cdot (pA)^k \quad \forall k \in \mathbb{N} \ \forall n \in \mathbb{N},
\]
and
\[
2 \frac{B}{A} \sum_{k=1}^{\infty} \frac{k-1}{k} \cdot (pA)^k \leq 2 \frac{B}{A} \sum_{k=1}^{\infty} (pA)^k
\]
which converges since we assume \( p < \frac{1}{A} \).

Therefore, by the Weierstrass M-test,
\[
\sum_{k=1}^{\infty} \frac{\epsilon_k(n) p^k}{k}
\]
converges uniformly (in \( n \)).

As explained previously, this allows us to pass the limit as \( n \to \infty \) inside the infinite sum from (49), which then allows us to plug in the expression for \( \lim_{n \to \infty} \epsilon_k(n) \) we had found in (47) inside this sum and finally obtain the Szegö error term given in the statement of the Theorem.

**Remark** The Szegö error term given in Theorem 3.3 was seemingly ”nicer” than the one we gave in Theorem 3.5 in the sense that there were integrals involved in the former as opposed to complicated infinite summations (as in the latter)... Yet they are the same often, that is, they are the same for sufficiently regular real symbols \( 1 - pf \).
After a deep and laborious search into the literature of non-Hermitian Toeplitz matrices we were able to find a very useful theorem, but which only holds for (non-)Hermitian Toeplitz matrices that have some very special properties (but which the kernel of our Carries process, turns out, to satisfy).

This aforementioned theorem presents the Szegő error term (for these special Toeplitz matrices) in a much nicer way, namely, in terms of integrals as opposed to complicated infinite summations. However, as stated before, to achieve this nice presentation of the Szegő error term, the special properties are vital. So, although the following theorem gives a much more neat Szegő error term, it does so at the expense of its strength (in regards to the conditions that have to be satisfied by the symbol).

We state it here:

**Theorem 3.6 (Theorem 2, [1])**

For symbols \( f(\lambda) = e^{-ik\lambda} \sum_{m=0}^{\infty} a_m e^{im\lambda} \), with \( a_0 = 1 \) and \( a_m = 0 \ \forall \ m < 0 \), for some \( k \in \mathbb{N} \), such that

\[
\sum_{m=0}^{\infty} |a_m| < \infty
\]

and

\[
\log(f(\lambda)) = \sum_{m=-\infty}^{\infty} h_m e^{im\lambda}
\]

with

\[
\sum_{m=-\infty}^{\infty} |h_m| < \infty,
\]

we have the following Szegő error term:

\[
\lim_{n \to \infty} \frac{\det(T_n(f))}{\exp\left\{ \frac{n}{2\pi J_0^2 \pi} \log(f(\lambda))d\lambda \right\}} = \exp\left\{ \sum_{m=1}^{\infty} mh_m h_{-m} \right\}.
\]

**Remark** By recalling the definition of a symbol, the conditions of Theorem 3.6 stated on the first line say that this theorem is only applicable for Toeplitz matrices that on their \( k \)th sub-diagonal have all entries equal to 1 and on all sub-diagonals lower than the \( k \)th one 0. That is, the above
Theorem assumes that the Toeplitz matrices are of the form

\[
\begin{pmatrix}
  a_0 & a_1 & a_2 & \cdots \\
  a_{-1} & a_0 & a_1 & \cdots \\
  \vdots & \ddots & \ddots & \ddots \\
  a_{-(k-1)} & \cdots & 1 & \ddots \\
  & \ddots & \ddots & \ddots \\
  1 & \ddots & \ddots & \ddots \\
  & & \ddots & \ddots \\
  & & & 1 \\
\end{pmatrix}
\]

(53)

Remark The above Theorem, besides the (28) condition it also assumes (as our Theorem 3.5 does), it further assumes that the logarithm of the symbol has a Fourier series with absolutely summable coefficients. If one retraces the proof of Theorem 2 given in [1], then one will see that the latter condition on the symbol is primarily used to actually get the Szegő error term in a nice form involving integrals, and the integrals being exactly equal to

\[
h_m = \frac{1}{2\pi} \int_0^{2\pi} \log(f(\lambda)) e^{-im\lambda}, \quad m \in \mathbb{Z},
\]

(54)

by the Fourier inversion formula since (52) holds (which ensures the uniform convergence of the Fourier series in \( \lambda \in [0, 2\pi] \)).

The way it achieves this is by getting the inverse of the symbol in a particularly "nice" form through some complex analysis, which then allows for a rather technical result (Theorem 1, [1]) to be applied having to do with Toeplitz matrices of the form (53) which states a special relation between their respective determinants under the assumption that the respective symbols are mutual inverses of each other.

Remark It is also worth noting that our Theorem 3.4 (to get the Szegő leading term) will work for symbols \( 1 - pf \) with countably many Fourier coefficients (with the right conditions on \( p \)); whereas Theorem 3.6 cannot.

We shall further explore Theorem 3.6 later on in Section 5.2 when we look to apply it on our carries process.
4 1-dependent processes on the integers

We had introduced the concept of 1-dependent point processes on the integers in Section 2.3 and had stated that they are in fact determinantal. This was proved in [3]. We’ve provide the full details of this very complicated ”proof” (it’s, rather, an inductive verification) below. Additionally, by gaining a better understanding of the determinantal structure of this class of point processes, we were able to actually derive the original result from scratch, and, as a result provide some other, more general, kernels.

To save writing, we shall introduce the following notations: denote by $\mathcal{X}$ the class of 1-dependent point processes on (a segment of) the integers, and by $\mathcal{Y}$ the subclass of $\mathcal{X}$ of translation-invariant 1-dependent point processes on (a segment of) the integers.

Although it is a known fact that the kernel of a DPP is not unique, it is usually a challenge to be able to figure out the different kernels by just making use of results from matrix theory. Interestingly, in the continuum, [22] classifies all the possible transformations of a symmetric kernel of a DPP that yield acceptable kernels for the same DPP. What we’ve done in this section is, by deriving some of the different matrices that make valid kernels for proving the determinantal structure of point processes from $\mathcal{Y}$, decipher their possible meaning as we make the transition into the more general, not necessarily translation-invariant, setting. As a result we were able to come up with some more kernels for point processes from $\mathcal{X}$ that would have otherwise been much more difficult to derive by solely making use of results from matrix theory. For the reasons mentioned in the introductory section we have not included these here, leaving them for further work in the future.

Besides for theoretical interest, such results potentially make for a practical one as well, since, there could be a result from a paper that one would like to apply to their DPP but its kernel is not in the appropriate form (initially).

We need a (fundamental) lemma from linear algebra before we begin.

**Lemma 4.1 (see, e.g., [10])** For a block-upper-triangular matrix $\Gamma = \begin{pmatrix} A & B \\ 0 & D \end{pmatrix}$, where $A$ and $D$ are square matrices, we have $\det(\Gamma) = \det(A) \det(D)$.

**Theorem 4.2** (Theorem 7.1, [3]) Any point process, with correlation function $\rho$, from $\mathcal{X}$ is determinantal.

A kernel for such point process is given by the matrix with $(x,y)$-entry being equal to $K(x,y)$ given
by:

\[ K(x, y) = \begin{cases} 
0, & \text{if } x - y \geq 2 \\
-1, & \text{if } x - y = 1 \\
\sum_{r=1}^{y-x+1} (-1)^{r-1} \sum_{x = l_0 < l_1 < \ldots < l_r = y+1} \rho([l_0, l_1]) \rho([l_1, l_2]) \cdots \rho([l_{r-1}, l_r]) & \text{if } x < y,
\end{cases} \]

where we make use of the notation \([a, b) := \{a, \ldots, b - 1\}\).

**Proof** To show that the point process is determinantal with kernel \(K\) as described above, we want to show that for any subset \(A\) of our segment,

\[ \rho(A) = \det((K(x, y))_{x, y \in A}) \quad (55) \]

In other words we want to show that for any \(k \in \mathbb{N}\) and for any subset \(A\) of our segment such that \(|A| = k\), (55) holds.

So let us break \(A\) into non-empty disjoint blocks of consecutive integers \(A_1, \ldots, A_l\), that is, \(A = \bigcup_{i=1}^{l} A_i\) with \(\text{dist}(A_i, A_j) \geq 2 \forall i \neq j, \|A_i\| = \alpha_i \forall i \in [l]\) such that \(\sum_{i=1}^{l} \alpha_i = k\). For example, if we were to break down \(A = \{1, 3, 4\}\) in this way, we would have \(A = A_1 \cup A_2\) with \(A_1 = \{1\}\) and \(A_2 = \{3, 4\}\) and \(\text{dist}(A_1, A_2) = |3 - 1| = 2\).

So by 1-dependence of the point process (recall (14) and the fact that, by construction, \(A_i\) and \(A_j\) for \(i \neq j\) are subsets of the segment such that \(\text{dist}(A_i, A_j) \geq 2\),

\[ \rho(A) = \rho(\bigcup_{i=1}^{l} A_i) = \prod_{i=1}^{l} \rho(A_i) \quad (56) \]

Assume without loss of generality that \(A_1, \ldots, A_l\) are arranged in ascending order in the sense that for every \(i \in \{1, \ldots, l - 1\}\),

\[ A_i = \{a_i^{(m)} := a_i^1 < a_i^2 < \ldots < a_i^{\alpha_i} =: a_i^{(M)}\} \]

consists of all the \(\alpha_i\) consecutive integers that make up \(A_i\), which are all at least \((-2)\) smaller than those of \(A_{i+1}\), which are also arranged from smallest to largest:

\[ A_{i+1} = \{a_{i+1}^{(m)} := a_{i+1}^1 < a_{i+1}^2 < \ldots < a_{i+1}^{\alpha_{i+1}} =: a_{i+1}^{(M)}\}. \]

Then, by this construction, we obtain a matrix \((K(x, y))_{x, y \in A}\) that has on its block diagonal the following matrices in this specific order: the \(\alpha_1 \times \alpha_1\) matrix \((K(x, y))_{x, y \in A_1}\), the \(\alpha_2 \times \alpha_2\) matrix \((K(x, y))_{x, y \in A_2}\), \ldots, the \(\alpha_l \times \alpha_l\) matrix \((K(x, y))_{x, y \in A_l}\). Namely, we get the following \(k \times k\) block
matrix:

\[
(K(x, y))_{x,y \in A} = \begin{bmatrix}
(K(x, y))_{x,y \in A_1} & ** & ** & ** \\
* & (K(x, y))_{x,y \in A_2} & ** & ** \\
* & * & (K(x, y))_{x,y \in A_3} & ** \\
* & * & * & (K(x, y))_{x,y \in A_l}
\end{bmatrix}
\]

If we can show that the matrix \((K(x, y))_{x,y \in A}\) is block-upper-triangular with the diagonal blocks being the matrices \((K(x, y))_{x,y \in A_1}, (K(x, y))_{x,y \in A_2}, \ldots, (K(x, y))_{x,y \in A_l}\), then we can apply Lemma 4.1 inductively and get that

\[
\det((K(x, y))_{x,y \in A}) = \prod_{i=1}^{l} \det((K(x, y))_{x,y \in A_i}).
\]  

(57)

In order to achieve this, we must prove that all the * blocks in the lower-triangular part above are just 0 blocks (i.e., matrices that only have the entry 0). The ** blocks in the upper-triangular part above are, of course, not of any interest to us in this pursuit.

It should be noted that this is not exactly a trivial task (even though the formula for \(K(x, y)\) given by the Theorem alludes otherwise) and we need to be careful in our justifications. More precisely, the construction of our sets \(A_1, \ldots, A_l\) play a crucial role in the proof of this, and without this specific construction the proof would break down.

So, essentially, we need to show that for any \(j \in \{2, \ldots, l\}\) and for any \(1 \leq i < j\), we have the following:

\[
K(a_j^{(m)}, a_i^{(m)}) = K(a_j^{(m)} + 1, a_i^{(m)})
\]

\[
= K(a_j^{(m)} + 2, a_i^{(m)})
\]

\[
= \ldots
\]

42
\[ K(a_j^{(M)}, a_i^{(M)}) = 0 \]

and

\[ K(a_j^{(m)}, a_i^{(m)}) = K(a_j^{(m)}, a_i^{(m)} + 1) = K(a_j^{(m)}, a_i^{(m)} + 2) = \ldots = K(a_j^{(m)}, a_i^{(M)}) = 0. \]

Now, since \( i < j \), we have, by construction of the \( A_k \)'s, that \( \text{dist}(A_i, A_j) \geq 2 \) and all elements of \( A_j \) are at least +2 bigger than all the elements of \( A_i \), i.e.,

\[ \forall p \in \{0, 1, \ldots, a_i^{(M)} - a_j^{(M)}\}, \forall q \in \{0, 1, \ldots, a_j^{(M)} - a_j^{(m)}\}, \ a_j^{(m)} + q \geq a_i^{(m)} + p + 2. \quad (58) \]

This implies that

\[ (a_j^{(m)} + q) - a_i^{(m)} \geq 2 \quad \forall q \in \{0, 1, \ldots, a_j^{(M)} - a_j^{(m)}\}. \]

So then by the definition of the kernel from the theorem,

\[ K(a_j^{(m)} + q, a_i^{(m)}) = 0 \quad \forall q \in \{0, 1, \ldots, a_j^{(M)} - a_j^{(m)}\} \]

as required.

Similarly, (58) implies that

\[ a_j^{(m)} - (a_i^{(m)} + p) \geq 2 \quad \forall p \in \{0, 1, \ldots, a_i^{(M)} - a_i^{(m)}\}. \]

So then, again, by the definition of the kernel from the theorem,

\[ K(a_j^{(m)}, a_i^{(m)} + p) = 0 \quad \forall p \in \{0, 1, \ldots, a_i^{(M)} - a_i^{(m)}\} \]

as required.

So indeed, \( (K(x, y))_{x,y \in A} \) is a block-upper-triangular matrix with the blocks \( (K(x, y))_{x,y \in A_i}, 1 \leq i \leq l \), on the block-diagonal. So we proved (57).
Having proved (57), if we are now able to show that
\[ \rho(A_i) = \det((K(x,y))_{x,y \in A_i}) \quad \forall i \in [l], \] (59)
then, by (56), we have proved (55) and we are done.

Recall how \( A_i \) is just a set containing consecutive integers. So for simplicity, it suffices to show that
\[ \rho([x,y+1)) = \det((K(x+i,x+j))_{0 \leq i,j \leq y-x}) \quad \forall x \leq y. \] (60)

We proceed by strong induction on \( y - x \):

The base case when \( y - x = 0 \) is easy:
\[
\det((K(x+i,x+j))_{0 \leq i,j \leq y-x}) = \det((K(x+i,x+j))_{0 \leq i,j \leq 0}) \\
= K(x,x) \\
= \rho(\{x\}) \quad \text{(by the definition of the kernel } K) \\
= \rho([x,x+1)) = \rho([x,y+1))
\]
as required.

To help us in the more complicated and challenging inductive step, we provide a simple sketch of the matrix \((K(x+i,x+j))_{0 \leq i,j \leq y-x}\):

\[
\begin{pmatrix}
K(x,x) & K(x,x+1) & K(x,x+2) & K(x,x+3) & \cdots & K(x,y-1) & K(x,y) \\
-1 & K(x+1,x+1) & K(x+1,x+2) & K(x+1,x+3) & \cdots & K(x+1,y-1) & K(x+1,y) \\
-1 & K(x+2,x+2) & K(x+2,x+3) & \cdots & K(x+2,y-1) & K(x+2,y) \\
-1 & K(x+3,x+3) & \cdots & K(x+3,y-1) & K(x+3,y) \\
\ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
-1 & K(y-1,y-1) & K(y-1,y) & \cdots & K(y-1,y) \\
-1 & & & & & & K(y,y)
\end{pmatrix},
\]

where, in the blank space below the sub-diagonal we have only 0 entries.

We will compute the determinant of the above matrix by consecutive cofactor expansion via the first column:

\[
\det((K(x+i,x+j))_{0 \leq i,j \leq y-x}) = K(x,x) \cdot \det((K(x+1+i,x+1+j))_{0 \leq i,j \leq y-(x+1)})
\]
\[
\begin{vmatrix}
K(x, x+1) & K(x, x+2) & K(x, x+3) & \cdots & K(x, y) \\
-1 & K(x+2, x+2) & K(x+2, x+3) & \cdots & K(x+2, y) \\
-1 & K(x+3, x+3) & \cdots & K(x+3, y) \\
\vdots & \vdots & \ddots & \vdots \\
& & & & K(y, y)
\end{vmatrix}
\]

\[
= K(x, x) \cdot \det((K(x+1+i, x+1+j))_{0 \leq i,j \leq y-(x+1)})
\]

\[
+ K(x, x+1) \cdot \det((K(x+2+i, x+2+j))_{0 \leq i,j \leq y-(x+2)})
\]

\[
\begin{vmatrix}
K(x, x+2) & K(x, x+3) & \cdots & K(x, y) \\
-1 & K(x+3, x+3) & \cdots & K(x+3, y) \\
-1 & \cdots & \vdots \\
& & & K(y, y)
\end{vmatrix}
\]

\[
= \sum_{k=0}^{(y-2)-x} K(x, x+k) \cdot \det((K(x+k+1+i, x+k+1+j))_{0 \leq i,j \leq y-(x+k+1)})
\]

\[
+ K(x, y-1) \cdot K(x, y)
\]

\[
\begin{vmatrix}
K(x, y-1) & K(x, y) \\
-1 & K(y, y)
\end{vmatrix}
\]

\[
= \sum_{k=0}^{(y-1)-x} K(x, x+k) \cdot \det((K(x+k+1+i, x+k+1+j))_{0 \leq i,j \leq y-(x+k+1)})
\]

\[
+ K(x, y)
\]

\[
= \sum_{k=0}^{(y-1)-x} K(x, x+k) \cdot \rho([x+k+1, y+1]) + K(x, y),
\]

where the last equality holds due to the inductive hypothesis, since

\[
y - (x + k + 1) = (y - x) - (k + 1) < y - x
\]

for every \( k = 0, \ldots, (y - 1) - x \).

Plugging in the formula for \( K(x, x+k) \) given by the theorem yields:

\[
\det((K(x+i, x+j))_{0 \leq i,j \leq y-x}) = \sum_{k=0}^{(y-1)-x} \left( \sum_{r=1}^{k+1} (-1)^{r-1} \sum_{x=l_0 < l_1 < \ldots < l_r = x+k+1} \rho([l_0, l_1]) \cdots \rho([l_{r-1}, l_r]) \right)
\]

\[
\cdot \rho([x+k+1, y+1]) + K(x, y)
\]  \( (61) \)

\[
\sum_{k=0}^{(y-1)-x} \left( \sum_{r=1}^{k+1} (-1)^{r-1} \sum_{x=l_0 < l_1 < \ldots < l_r = x+k+1} \rho([l_0, l_1]) \cdots \rho([l_{r-1}, l_r]) \right)
\]

\[
\cdot \rho([x+k+1, y+1]) + K(x, y)
\]  \( (62) \)
Observe how by the definition of the kernel $K$ given in the Theorem,

$$K(x, y) = \rho([x, y + 1]) + \sum_{r=2}^{y-x+1} (-1)^{r-1} \sum_{x=l_0 < l_1 < \ldots < l_r = y+1} \rho([l_0, l_1]) \cdots \rho([l_{r-1}, l_r]).$$

Thus, in order to show that

$$\det((K(x + i, x + j))_{0 \leq i, j \leq y-x}) = \rho([x, y + 1]),$$

it will suffice to show that we have the following identity

$$\sum_{r=2}^{y-x+1} (-1)^{r-1} \sum_{x=l_0 < l_1 < \ldots < l_r = y+1} \rho([l_0, l_1]) \cdots \rho([l_{r-1}, l_r]) = \rho([x + k + 1, y + 1]).$$

Observe how

$$- \sum_{k=0}^{(y-1)x} \left( \sum_{r=1}^{k+1} (-1)^{r-1} \sum_{x=l_0 < l_1 < \ldots < l_r = x+k+1} \rho([l_0, l_1]) \cdots \rho([l_{r-1}, l_r]) \right) \cdot \rho([x + k + 1, y + 1])$$

is equal to

$$\sum_{k=0}^{y-x-1} \sum_{r=1}^{k+1} (-1)^{r+2} \sum_{x=l_0 < l_1 < \ldots < l_r = x+k+1} \rho([l_0, l_1]) \cdots \rho([l_{r-1}, l_r]) \rho([x + k + 1, y + 1]),$$

which is equal to

$$\sum_{k=0}^{y-x-1} \sum_{r=2}^{k+2} (-1)^{r+1} \sum_{x=l_0 < l_1 < \ldots < l_{r-1} = x+k+1} \rho([l_0, l_1]) \cdots \rho([l_{r-2}, l_{r-1}]) \rho([x + k + 1, y + 1]),$$

which, after a simple change of variables, is finally equal to

$$\sum_{r=2}^{y-x+1} (-1)^{r-1} \sum_{k=r-2}^{y-x-1} \sum_{x=l_0 < l_1 < \ldots < l_{r-1} = x+k+1} \rho([l_0, l_1]) \cdots \rho([l_{r-2}, l_{r-1}]) \rho([x + k + 1, y + 1]).$$
It is then easy to see that

\[
\sum_{k=r-2}^{y-x-1} \sum_{x=l_0<l_1<\ldots<l_{r-1}=x+k+1} \rho([l_0, l_1]) \cdots \rho([l_{r-2}, l_{r-1}]) \rho([x+k+1, y+1])
= \sum_{x=l_0<l_1<\ldots<l_r=y+1} \rho([l_0, l_1]) \cdots \rho([l_{r-1}, l_r]).
\]

And so the equality from (63) holds. Hence we are done by strong induction.

Before we move on we should take a moment to really appreciate this result and what it states. What this result states is that any 1-dependent point process on the integers has a “special” structure as discussed in the previous sections: any probability can be obtained by computing principle minors of a known matrix. In particular, the above theorem is applicable to the point process introduced in Section 2.3 (the Carries process with base \(\infty\)) which we are interested in studying its gap probabilities.

The proof of Theorem 4.2 was very technical and it merely verified that the formula for the kernel given by the theorem "works". It begs the question, however: how was this formula discovered and how could anyone possibly conceive that such an enormous class of point processes would be determinantal? On top of that, is the formula given by the theorem unique for all such processes? In other words, is this the only formula that works in order for 1-dependent point processes on the integers to be determinantal?

None of the above questions were answered in [3], so we had to investigate and find a way to somehow derive this mysterious formula from scratch. To do this, a helpful assumption would be that our 1-dependent point process on the integers is stationary - this would greatly simplify computations in our search and would also make it easier for us to spot potential patterns. Once we (hopefully) spotted a pattern and got an expression for the translation-invariant/stationary case we would then try to translate it to the more general case and (hopefully) obtain Theorem 4.2, and possibly, some other valid kernels as well and hence establish non-uniqueness of the kernel at hand.

To this end, we consider a point process from \(\mathcal{Y}\) with correlation function \(\rho\), which, by recalling remark (13) from Section 2.3 regarding the stationarity assumption, satisfies

\[\rho_k := P(X_{1+i} = 1, \ldots, X_{k+i} = 1)\]

for every \(i \in \mathbb{N}_0\) and \(k \in \mathbb{N}\), and, by recalling remark (14) regarding the 1-dependence assumption,
also satisfies

\[
\rho(\{x, x+1, \ldots, x+l_1, x+l_1+k, x+l_1+k+1, \ldots, x+l_1+k+l_2\}) = (64)
\]

\[
\rho(\{x, x+1, \ldots, x+l_1\})\rho(\{x+l_1+k, x+l_1+k+1, \ldots, x+l_1+k+l_2\}) = \rho_{l_1+1} \cdot \rho_{l_2+1},
\]

for any \(x \in \mathbb{Z}, l_1, l_2 \in \mathbb{N}\) and \(k \geq 2\), where the last equality holds due to the translation-invariance/stationarity property (13) of the process.

In addition, we assume a priori that this process is in fact determinantal with some, for the time being, unknown kernel \(K\), and in this way we can hopefully work out what this kernel should be - or perhaps realize that there can be no kernel such that the process is determinantal (even though we already know the answer to this is nay from Theorem 4.2). We should note that the equalities given in (13) and (64) would, reasonably, give someone hope for a determinantal structure of the process.

Let’s begin. Under all of the above assumptions we would first require that for every \(x \in \mathbb{Z}\),

\[
\rho(\{x\}) = \rho_1 = \det((K(z,w))_{x \leq z, w \leq x}) = K(x, x),
\]

(65)

where the first equality is due to the stationarity assumption (13) of the process and the second equality due to the determinantal structure assumption of the process.

By the same reasoning we would also require that for every \(x \in \mathbb{Z}\),

\[
\rho(\{x, x+1\}) = \rho_2 = \det((K(z,w))_{x \leq z, w \leq x+1}) = \begin{vmatrix} K(x, x) & K(x, x+1) \\ K(x+1, x) & K(x+1, x+1) \end{vmatrix} = \begin{vmatrix} \rho_1 & K(x, x+1) \\ K(x+1, x) & \rho_1 \end{vmatrix} \quad \text{(by (65))}
\]

\[
= \rho_1^2 - K(x, x+1)K(x+1, x),
\]

which is equivalent to requiring that

\[
K(x, x+1)K(x+1, x) = \rho_1^2 - \rho_2.
\]

(66)

The choices for the values of \(K(x, x+1)\) and \(K(x+1, x)\) that would yield the simplest kernel \(K\)
would be to choose values that are consistent in the sense that

\[ K(x, x + 1) = K(y, y + 1) \quad \forall x, y \in \mathbb{Z} \]

and

\[ K(x + 1, x) = K(y + 1, y) \quad \forall x, y \in \mathbb{Z}. \]

In this way we would be on track in constructing \( K \) as a Toeplitz matrix. We should note however that not abiding to the above two equations would not necessarily disqualify the resulting matrix \( K \); but given the fact that the resulting matrix, in that case, would definitely not be Toeplitz - and wanting/expecting a Toeplitz matrix is only natural when dealing with translation-invariant processes - we are not interested in it. Not to mention the many nice properties and theorems (which we discussed/developed in Section 3 and which we will see a bit more in Section 5.2) for Toeplitz matrices that will come in handy. For these reasons we will abide to the above two equalities.

Again in the same way, under all our assumptions on our process, for every \( x, y \in \mathbb{Z} \) such that \( y - x \geq 2 \),

\[
\rho(\{x, y\}) = \rho(\{x\})\rho(\{y\}) = \rho_1^2 = \det(K_{\{x,y\}}) \\
= |K(x, x) K(x, y)| \\
= |K(y, x) K(yy)| \\
= K(x, x)K(y, y) - K(x, y)K(y, x) \\
= \rho_1^2 - K(x, y)K(y, x),
\]

which is equivalent to requiring that

\[ K(x, y)K(y, x) = 0. \] (67)

Since we endeavour in constructing \( K \) as a Toeplitz matrix, we again have to be consistent as to which of the two entries of the matrix \( K \), \( K(x, y) \) and \( K(y, x) \), is equal to 0. We cannot have both of them equal to 0 and the reason why will become clear soon. If we chose \( K(x, y) = 0 \) we would get a matrix of the form

\[
\begin{pmatrix}
\rho_1 & * & \cdots \\
* & \rho_1 & * & \cdots \\
* & * & \rho_1 & * & \cdots \\
\vdots & \vdots & \vdots & \ddots & \ddots & \cdots
\end{pmatrix},
\]

where all the entries above the upper-sub-diagonal are 0.
And if we chose \( K(y, x) = 0 \) we would get a matrix of the form

\[
\begin{pmatrix}
\rho_1 & * & * & \cdots \\
* & \rho_1 & * & \cdots \\
* & * & \rho_1 & \cdots \\
* & * & * & \rho_1 \\
& \ddots & \ddots & \ddots \\
\end{pmatrix},
\]

where all the entries below the lower-sub-diagonal are 0.

So up to now we have established that \( K(x, x) = \rho_1 \) for every \( x \in \mathbb{Z} \) (by (65)), which implies that our matrix \( K \) will need to have a constant diagonal with the value \( \rho_1 \). Furthermore we need to make consistent choices (in the sense described previously) for the entries \( K(x, x+1) \) and \( K(x+1, x) \) so that (66) is satisfied, as well as a consistent choice for which \( K(x, y) \) or \( K(y, x) \) equals 0 for \( x, y \in \mathbb{Z} \) such that \( y - x \geq 2 \).

To summarize, we have two options:

\[
K(x, x+1) = \alpha \quad \text{and} \quad K(x+1, x) = \frac{\rho_1^2 - \rho_2}{\alpha} \quad \text{for some } \alpha \in \mathbb{R}, \; \forall x \in \mathbb{Z}
\]  

(68)

or

\[
K(x, x+1) = \frac{\rho_1^2 - \rho_2}{\alpha} \quad \text{and} \quad K(x+1, x) = \alpha \quad \text{for some } \alpha \in \mathbb{R}, \; \forall x \in \mathbb{Z}.
\]  

(69)

And after we make this choice, we need to choose one of the following two options:

\[
K(y, x) = 0 \quad \forall x, y \in \mathbb{Z} \quad \text{such that} \quad y - x \geq 2
\]  

(70)

or

\[
K(x, y) = 0 \quad \forall x, y \in \mathbb{Z} \quad \text{such that} \quad y - x \geq 2
\]  

(71)

For now let’s choose (69) and (70) and see where this gets us. Having made these two choices, our Toeplitz matrix \( K \) in the size \( 2 \times 2 \) case looks like

\[
\begin{pmatrix}
\rho_1 & \frac{\rho_1^2 - \rho_2}{\alpha} \\
\alpha & \rho_1
\end{pmatrix}.
\]
and in the dimension 3 case looks like

\[
\begin{pmatrix}
\rho_1 & \frac{\rho_1^2 - \rho_2}{\alpha} & K(x, x + 2) \\
\alpha & \rho_1 & \frac{\rho_1^2 - \rho_2}{\alpha} \\
0 & \alpha & \rho_1
\end{pmatrix},
\]

where \( K(x, x + 2) \) is to be determined. Actually, under all the assumptions of our process, the following equation needs to be satisfied:

\[
\rho(\{x, x + 1, x + 2\}) = \rho_3 = \det((K(z, w))_{x \leq z, w \leq x+2})
\]

\[
= \begin{vmatrix}
K(x, x) & K(x, x + 1) & K(x, x + 2) \\
K(x + 1, x) & K(x + 1, x + 1) & K(x + 1, x + 2) \\
K(x + 2, x) & K(x + 2, x + 1) & K(x + 2, x + 2)
\end{vmatrix}
\]

\[
= \begin{vmatrix}
\rho_1 & \frac{\rho_1^2 - \rho_2}{\alpha} & K(x, x + 2) \\
\alpha & \rho_1 & \frac{\rho_1^2 - \rho_2}{\alpha} \\
0 & \alpha & \rho_1
\end{vmatrix}
\]

\[
= \rho_1 \rho_2 - \rho_1^3 + \rho_1 \rho_2 + \alpha^2 K(x, x + 2).
\]

Therefore, we must have for every \( x \in \mathbb{Z} \),

\[
K(x, x + 2) = \frac{\rho_3 - 2\rho_1 \rho_2 + \rho_1^3}{\alpha^2}.
\] (72)

And so this is the reason why when we were explaining (67) we stated that we could not have both \( K(y, x) = 0 \) and \( K(x, y) = 0 \), where \( x, y \in \mathbb{Z} \) such that \( y - x \geq 2 \) - because there are further requirements that need to be adhered to in order for the process to be determinantal (such as the one above).

Now we have for every \( x \in \mathbb{Z} \),

\[
K(x + 1, x) = \alpha,
\]

\[
K(x, x) = \rho_1,
\]

\[
K(x, x + 1) = \frac{\rho_1^2 - \rho_2}{\alpha},
\]

\[
K(x, x + 2) = \frac{\rho_3 - 2\rho_1 \rho_2 + \rho_1^3}{\alpha^2},
\]

and

\[
K(y, x) = 0 \ \forall x, y \in \mathbb{Z} \text{ such that } y - x \geq 2.
\]
If one looks closely, there is already a pattern emerging for the $K(x, x + k)$, $k \in \mathbb{N}_0$, entries. Because we aim to construct the matrix $K$ to be Toeplitz, let’s simplify our notations and set $K(x, y) \equiv k(y - x)$. Hence, we have

\[ k(-1) = \alpha, \]
\[ k(0) = \rho_1, \]
\[ k(1) = \frac{\rho_1^2 - \rho_2}{\alpha}, \]
\[ k(2) = \frac{\rho_3 - 2\rho_1\rho_2 + \rho_1^3}{\alpha^2}, \]
\[ k(n) = 0 \ \forall n \leq -2. \]

The pattern for $\left( K(x, x + n) \equiv k(n), n \in \mathbb{N}_0, \right.$ in words, seems to be the following: all the terms appearing on the numerator - ignoring signs - are, each of them, products of the $\rho$’s with subscripts coming from the set $\{1, 2, \ldots, n + 1\}$ such that the sum of the subscripts of each of the $\rho$’s in the product that makes up the particular term equals $n + 1$; and the only term that appears in the denominator is simply just $\alpha$ to the power $n$. This description may be a little confusing, so we shall go through the entries $k(0)$, $k(1)$ and $k(2)$ in detail to illustrate the previous explanation.

We start with $k(2)$: Observe how in the numerator the terms $\rho_3$, $2\rho_1\rho_2$ and $\rho_1^3$ all appear (ignoring signs for now). More explicitly, the terms $\rho_3$, $\rho_1\rho_2$, $\rho_2\rho_1$ and $\rho_1\rho_1\rho_1$ all appear in the numerator. Notice how the subscripts of each of the $\rho$’s in the product that makes up the particular term in the numerator all add up to 3. Indeed, for the first term $\rho_3$, this is a product involving only one $\rho$ and its subscript exactly equals 3. As for the second term $\rho_1\rho_2$, this is a product involving two $\rho$’s (the $\rho_1$ and the $\rho_2$) and the sum of the subscripts of both of these $(1 + 2)$ exactly equals 3. Likewise for the $\rho_2\rho_1$ term, which is just a rearranging of the term described previously. As for the last term $\rho_1^3 = \rho_1\rho_1\rho_1$, this is a product involving two $\rho$’s (the $\rho_1$ three times) and the sum of the subscripts of these three $\rho$’s $(1+1+1)$ is exactly equal to 3. Therefore, we see that all the terms just described correspond to all the possible ways we can sum elements from the set $\{1, 2, 3\}$ to get 3. More precisely, the term $\rho_3$ corresponds to the sum with one summand, namely 3 itself; the term $\rho_1\rho_2$ corresponds to a sum with two summands from $\{1, 2, 3\}$, namely $1 + 2(= 3)$; the term $\rho_2\rho_1$ corresponds to the sum with two summands from $\{1, 2, 3\}$, namely $2 + 1(= 3)$; and lastly, the term $\rho_1^3 = \rho_1\rho_1\rho_1$ corresponds to the sum with three summands from $\{1, 2, 3\}$, namely $1 + 1 + 1(= 3)$.

We move next to $k(1)$: Observe how the terms that are appear in the numerator (ignoring signs for now) are just $\rho_2$ and $\rho_1^2 = \rho_1\rho_1$. Notice how the subscripts of each of the $\rho$’s in the product that makes up the particular term in the numerator all add up to 2. Indeed, for the first term
\( \rho_2 \), this is a product involving only one \( \rho \) and its subscript exactly equals 2. As for the last term \( \rho_1^2 = \rho_1 \rho_1 \), this is a product involving two \( \rho \)'s (the \( \rho_1 \) two times) and the sum of the subscripts of these two \( \rho \)'s (1 + 1) is exactly equal to 2. Therefore, again we see that these two terms coming from the numerator just described correspond to all the possible ways we can sum elements from the set \{1, 2\} to get 2. More precisely, the term \( \rho_1^2 \) corresponds to the sum with one summand, namely 2 itself; and the term \( \rho_1 \rho_1 \) corresponds to the sum with two summands from \{1, 2\}, namely 1 + 1( = 2).

And finally, for \( k(0) \), since the only way we can sum up elements from the singleton set \{1\} to get 1 is by taking the sum with one summand 1 itself, the only term appearing in the numerator is simply just \( \rho_1 \).

Hopefully these descriptions and illustrations were clear and have made the reader understand and recognize the structure of the \( k(n) \) entries. Furthermore, we are able to express these descriptions in very precise ways mathematically - taking the signs of the terms from the numerator into account as well via the following equations: for \( n \in \mathbb{N}_0 \),

\[
k(n) = (-1)^n \frac{1}{\alpha^n} \sum_{r=1}^{n+1} (-1)^{r-1} \sum_{1 \leq l_1, \ldots, l_r \leq n+1 \atop \sum l_i = n+1} \rho_1 \cdots \rho_r \]  \quad (73)

\[
= (-1)^n \frac{1}{\alpha^n} \sum_{r=1}^{n+1} (-1)^{r-1} \sum_{0 \leq m(1), \ldots, m(n+1) \leq r \atop \sum m(l) = n+1} \prod_{j=1}^{n+1} \left( r - \sum_{l=1}^{j-1} m(l) \right) \rho_j^{m(j)}. \]  \quad (74)

It should be emphasized that the Toeplitz matrix \( (K(x,y))_{x,y \in \mathbb{Z}} \equiv (k(x-y))_{x,y \in \mathbb{Z}} \), where \( k(n) \) is as described above, is only just a guess of a matrix that could possibly be a kernel for our process. For \( n = -1, 0, 1, 2 \) we’ve seen that \( k(n) \) works in this regard, but this does not necessarily mean that the expression for \( k(n) \) given in (73) works for our cause for greater values of \( n \) - we need to formally prove that it does, and strong induction seems like the most appropriate way of doing so. In other words, the Toeplitz matrix constructed by the coefficients \{\( k(n) : n \in \mathbb{Z} \)\} from above is merely a conjecture of a kernel for our process at the moment.

Recall how we are assuming that our process is stationary and hence are using the \( \rho \) subscript notation from (13). If we now relax this assumption and consider a more general (not necessarily translation-invariant) 1-dependent point process on (a segment of) \( \mathbb{Z} \) (i.e., a point process from the \( \mathcal{X} \) class), we can attempt to translate the above derived expressions for \( k(n) \) from (73) to this new, more general, point process in the following way: For the matrix entry \( K(x,y) \), where \( y - x = n \), we
adjust the inner sum from (73) from
\[
\sum_{1 \leq l_1, \ldots, l_r \leq n+1, \sum_i l_i = n+1} \rho_{l_1} \cdots \rho_{l_r},
\]
(75)
to
\[
\sum_{x = l_0 < \cdots < l_r = y+1} \rho([l_0, l_1]) \cdots \rho([l_{r-1}, l_r]).
\]
(76)
The reasoning behind this is that, since we are no longer in the translation-invariant setting, we no longer are able to use (13), that is, it is no longer the case that
\[
P(X_{l_i} = 1, \ldots, X_{l_i+1} = 1) =: \rho([l_i, l_{i+1}]) = \rho([l_i + j, l_{i+1} + j]) := P(X_{l_{i+j}} = 1, \ldots, X_{l_{i+j+1}} = 1),
\]
and so besides for the number of consecutive 1’s, the starting point of this sequence of consecutive 1’s equally matters. More precisely, in the translation-invariant setting, the sum from (75) is used for all entries $K(x, y)$ such that $y - x = n$, and hence, we have not used the variables $x$ or $y$ in the sum, but rather, we’ve only used their difference $n$. When we move to the more general setting where we don’t necessarily have translation-invariance, both variables $x$ and $y$ matter - not just their difference $y - x = n$, and so we need to somehow translate what (75) is saying but also take into account the associated variables $x$ and $y$. Now, what (75) is saying is precisely what was explained and illustrated earlier about the lengths of the runs of 1’s of the point process, $l_1, \ldots, l_r$, and them having to add up to $n + 1$. To incorporate the positions $x$ and $y$ into this sum (so as to translate (75) to the general setting) we think about all the possible lengths of runs of 1’s that start from position $x$ and finish at position $y + 1$, so that, again, as in the translation-invariant case, all the lengths add up to $y + 1 - x = n + 1$. This is precisely what (76) is describing.

Therefore, we conjecture the following theorem and hope to be able to prove it rigorously.

**Theorem 4.3** Any point process from $X$, with correlation function $\rho$, is determinantal with kernel $K$ with $(x, y)$-entry given by $K(x, y)$, where

\[
K(x, y) = \begin{cases} 
0, & \text{if } x - y \geq 2 \\
\alpha, & \text{if } x - y = 1 \\
(-1)^{y-x} \frac{1}{\alpha^{y-x}} \frac{y-x+1}{x+1} \sum_{r=1}^{y-x} (-1)^{r-1} \sum_{x = y < l_0 < \cdots < l_r = y+1} \rho([l_0, l_1]) \rho([l_1, l_2]) \cdots \rho([l_{r-1}, l_r]) & \text{if } x < y,
\end{cases}
\]

where we make use of the notation $[a, b) := \{a, \ldots, b-1\}$.

**Proof** To show that the point process is determinantal with kernel $K$ as described above, we want
to show that for any subset $A$ of our segment,

$$\rho(A) = \det((K(x, y))_{x, y \in A})$$

We then break $A$ into non-empty disjoint blocks of consecutive integers $A_1, \ldots, A_l$ in the same we had done in the proof of Theorem 4.2 and obtain the same identity (56) we had obtained in that proof.

Then, by the same analysis as that at the start of the proof of Theorem 4.2, we get that $(K(x, y))_{x, y \in A}$ is a block-upper-triangular matrix with the diagonal blocks being the matrices $(K(x, y))_{x, y \in A_1}, (K(x, y))_{x, y \in A_2}, \ldots, (K(x, y))_{x, y \in A_l}$ in this order. And so we can then apply Lemma 4.1 inductively and get the same identity (57) we had obtained in the proof of Theorem 4.2.

So then, as in the proof of Theorem 4.2, it suffices to show that

$$\rho([x, y + 1)) = \det((K(x + i, x + j))_{0 \leq i, j \leq y - x}) \quad \forall x \leq y.$$ 

We proceed by strong induction on $y - x$:

The base case when $y - x = 0$ is easy:

$$\det((K(x + i, x + j))_{0 \leq i, j \leq 0}) = \det((K(x + i, x + j))_{0 \leq i, j \leq 0})$$

$$= K(x, x)$$

$$= (-1)^{y-x} \cdot \frac{1}{\alpha^{y-x}} \rho(\{x\})$$

$$= (-1)^0 \cdot \frac{1}{\alpha^0} \rho(\{x\})$$

$$= \rho(\{x\})$$

$$= \rho([x, x + 1)) = \rho([x, y + 1))$$

as required.

To help us in the more complicated and challenging inductive step, we provide a simples sketch of
the matrix \((K(x+i,x+j))_{0 \leq i,j \leq y-x}\):

\[
\begin{pmatrix}
  K(x,x) & K(x,x+1) & K(x,x+2) & K(x,x+3) & \cdots & K(x,x+y-1) & K(x,y) \\
  \alpha & K(x+1,x+1) & K(x+1,x+2) & K(x+1,x+3) & \cdots & K(x+1,x+y-1) & K(x+1,y) \\
  \alpha & K(x+2,x+2) & K(x+2,x+3) & \cdots & K(x+2,y-1) & K(x+2,y) \\
  \alpha & K(x+3,x+3) & \cdots & K(x+3,y-1) & K(x+3,y) \\
  \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
  \alpha & K(y-1,y-1) & K(y-1,y) & \cdots & K(y,y) \\
\end{pmatrix},
\]

where, in the black space below the sub-diagonal we have only 0 entries.

As in the proof of Theorem 4.2, we will compute the determinant of the above matrix by consecutive cofactor expansion via the first column:

\[
\begin{align*}
\det((K(x+i,x+j))_{0 \leq i,j \leq y-x}) &= K(x,x) \cdot \det((K(x+1+i,x+1+j))_{0 \leq i,j \leq y-(x+1)}) \\
&= \sum_{k=0}^{(y-2)-x} (-1)^k \alpha^k K(x,x+k) \det((K(x+k+1+i,x+k+1+j))_{0 \leq i,j \leq y-(x+k+1)}) \\
&\quad + (-1)^{(y-1)-x} \cdot \alpha^{(y-1)-x} \cdot K(x,y-1) \\
&\quad \cdot \det((K(x,x+y-1))_{0 \leq i,j \leq y-(x+y-1)}) \\
&\quad + (-1)^{(y-1)-x} \cdot \alpha^{(y-1)-x} \cdot K(x,x+y-1) \\
&\quad \cdot \det((K(x,y))_{0 \leq i,j \leq y-(x+y-1)}) \\
&\quad + (-1)^{(y-1)-x} \cdot \alpha^{y-x} K(x,y) \\
&\quad \cdot \sum_{k=0}^{(y-1)-x} \sum_{r=1}^{(y-1)-x-k} (-1)^{r-1} \cdot \rho([l_0,l_1]) \rho([l_1,l_2]) \cdots \rho([l_{r-1},l_r]) \\
&\quad \cdot \rho([x+k+1,y+1]) \\
&\quad + \sum_{r=1}^{y-x+1} (-1)^{r-1} \cdot \rho([l_0,l_1]) \rho([l_1,l_2]) \cdots \rho([l_{r-1},l_r]),
\end{align*}
\]
where the last equality holds by plugging in the expressions for $K(x, x + k)$ and $K(x, y)$ given by the Theorem/Conjecture and applying the inductive hypothesis since

$$y - (x + k + 1) = (y - x) - (k + 1) < y - x \quad \forall k = 0, \ldots, (y - 1) - x.$$ 

Note that we skipped writing most of the initial determinant computations in the above series of equations because they are very similar to those we had done in the proof of Theorem 4.2.

In the above equations we have managed to manipulate the expressions in such a way so that the final expression (in the last equality from above) is exactly equal to the RHS of (61) from the proof of Theorem 4.2, which means that we can apply the identity (63), which we had showed at the end of the proof of Theorem 4.2, to conclude that

$$\det((K(x + i, x + j))_{0 \leq i, j \leq y - x}) = \rho([x, y + 1]).$$

And so we are done by strong induction.  

So we were correct in our deductive reasoning. Actually, Theorem 4.3 is the more general Theorem 4.2 since we can obtain the latter from the former by setting $\alpha = -1$!

**Remark** We derived the above kernel by attempting to derive the fact that point processes from $X$ are determinantal from scratch. There is a way, however, that one could derive the above kernel *directly from* the kernel given in Theorem 4.2 via the following results from matrix theory:

$$\begin{vmatrix}
    a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,n} \\
    a_{2,1} & a_{2,2} & a_{2,3} & \cdots & a_{2,n} \\
    a_{3,1} & a_{3,2} & a_{3,3} & \cdots & a_{3,n} \\
    \vdots & \vdots & \vdots & \ddots & \vdots \\
    a_{n,1} & a_{n,2} & a_{n,3} & \cdots & a_{n,n}
\end{vmatrix} = \begin{vmatrix}
    a_{1,1} & ra_{1,2} & r^2a_{1,3} & \cdots & r^{n-1}a_{1,n} \\
    r^{-1}a_{2,1} & a_{2,2} & ra_{2,3} & \cdots & r^{n-2}a_{2,n} \\
    r^{-2}a_{3,1} & r^{-1}a_{3,2} & a_{3,3} & \cdots & r^{n-3}a_{3,n} \\
    \vdots & \vdots & \vdots & \ddots & \vdots \\
    r^{-(n-1)}a_{n,1} & r^{-(n-2)}a_{n,2} & r^{-(n-3)}a_{n,3} & \cdots & a_{n,n}
\end{vmatrix},$$

and all the principal minors of the above two matrices agree as well.

Indeed, the equality from above can be derived via elementary row/column operations. More precisely, if we multiply row 1 of the matrix on the RHS from above by $r^{-1}$ (which is a type 2 row operation), it’s well-known that this just multiplies the determinant of the RHS matrix by $r^{-1}$. If we then multiply column 1 of the resulting matrix by $r$ (which is a type 2 column operation), it’s well-known that this just multiplies the determinant by $r$. Now, having performed these two row/column operations on the RHS matrix, we have multiplied its determinant by $r^{-1}$ and then by
r (and so we haven’t changed its determinant), and we have also gotten its $2 \times 2$ leading principle submatrix equal to that of the matrix on the LHS (that is, we’ve gotten rid of the $r$ and $r^{-1}$ factors from it). Continuing to perform these types of elementary type 2 row/column operations on the remaining rows/columns inductively in this way on the RHS matrix yields the above equality. The same exact arguments can be used to prove the assertion below the above equality regarding the principal minors of the above two matrices. Finally, having proved these two assertions, the fact that probabilities of a DPP are given by principal minors, we are done.

By recalling our arguments from (73) to (76) regarding the translation of the expressions of the entries of the kernels from the translation-invariant setting to the general one and vice-versa, we get the following corollary regarding translation-invariant 1-dependent processes on the integers:

**Corollary 4.4** Any point process from $\mathcal{Y}$ is determinantal with kernel $K$ being a Toeplitz matrix with entries $k(n)$ on the $n^{th}$ diagonal given by:

$$
k(n) = \begin{cases} 
0, & \text{if } n \leq -2 \\
\alpha, & \text{if } n = -1 \\
(-1)^n \frac{1}{\alpha^n} \sum_{r=1}^{n+1} (-1)^{r-1} \sum_{1 \leq l_1, \ldots, l_r \leq n+1} \rho_{l_1} \cdots \rho_{l_r}, & \text{if } n \geq 0
\end{cases}
$$

where $K(x, y) \equiv k(y - x)$ for any $x, y \in \mathbb{Z}$.

Since $\alpha = -1$ simplifies the above expression a bit, we state the following corollary as well.

**Corollary 4.5** Any point process from $\mathcal{Y}$ is determinantal with kernel $K$ being a Toeplitz matrix with entries $k(n)$ on the $n^{th}$ diagonal given by:

$$
k(n) = \begin{cases} 
0, & \text{if } n \leq -2 \\
-1, & \text{if } n = -1 \\
\sum_{r=1}^{n+1} (-1)^{r-1} \sum_{1 \leq l_1, \ldots, l_r \leq n+1} \rho_{l_1} \cdots \rho_{l_r}, & \text{if } n \geq 0
\end{cases}
$$

where $K(x, y) \equiv k(y - x)$ for any $x, y \in \mathbb{Z}$.

It would be useful (for later on in the dissertation) to be able to find the symbol of the above
Toeplitz matrix, which, recall from Section 3.1, is defined as

\[ f : [0, 2\pi] \to \mathbb{C}, \quad f(\lambda) = \sum_{n \in \mathbb{Z}} k(n) e^{in\lambda}. \]  \hspace{1cm} (77)

To this end, the following result from [3] will be useful, which we’ve filled in all the missing details for.

**Theorem 4.6 (Corollary 7.3, [3])** \( k(n), \ n \in \mathbb{Z}, \) from Corollary 4.5 has generating function given by:

\[ \sum_{n \in \mathbb{Z}} k(n)z^n = \frac{1}{1 - R(z)}, \quad \text{for all } z \in \mathbb{C} \text{ with } |z| \text{ sufficiently small}, \]

where \( R(z) = 1 + z + \sum_{k=1}^{\infty} \rho_k z^{k+1} \).

**Proof** First make the following observation:

\[
\sum_{r=1}^{\infty} (-1)^{r-1} \left( \sum_{m=1}^{\infty} \rho_m z^m \right)^r = \sum_{m_1=1}^{\infty} \rho_{m_1} z^{m_1} - \sum_{m_1,m_2=1}^{\infty} \rho_{m_1} \rho_{m_2} z^{m_1+m_2} + \sum_{m_1,m_2,m_3=1}^{\infty} \rho_{m_1} \rho_{m_2} \rho_{m_3} z^{m_1+m_2+m_3} - \ldots
\]

\[
= \left( \rho_1 z + \rho_2 z^2 + \rho_3 z^3 + \sum_{m_1=4}^{\infty} \rho_{m_1} z^{m_1} \right)
\]

\[
- \left( \rho_1\rho_1 z^1 z^1 + \rho_1 \rho_2 z^1 z^2 + \rho_2 \rho_1 z^2 z^1 + \sum_{m_1,m_2=2}^{\infty} \rho_{m_1} \rho_{m_2} z^{m_1+m_2} \right)
\]

\[
+ \left( \rho_1\rho_1\rho_1 z^1 z^1 z^1 + 3 \cdot \rho_1^2 \rho_2 z^4 + \sum_{m_1,m_2,m_3=2}^{\infty} \rho_{m_1} \rho_{m_2} \rho_{m_3} z^{m_1+m_2+m_3} \right)
\]

\[- \ldots
\]

\[
= \left( \rho_1 z + \rho_2 z^2 + \rho_3 z^3 + \sum_{m_1=4}^{\infty} \rho_{m_1} z^{m_1} \right)
\]

\[
- \left( \rho_1^2 z^2 + 2 \cdot \rho_1 \rho_2 z^3 + \sum_{m_1,m_2=2}^{\infty} \rho_{m_1} \rho_{m_2} z^{m_1+m_2} \right)
\]

\[
+ \left( \rho_1^3 z^3 + 3 \cdot \rho_1^2 \rho_2 z^4 + \sum_{m_1,m_2,m_3=2}^{\infty} \rho_{m_1} \rho_{m_2} \rho_{m_3} z^{m_1+m_2+m_3} \right)
\]

\[- \ldots
\]
By then comparing the coefficients of $z$, $z^2$, $z^3$, ... and recalling the expression for $k(n)$ we had found in (73) with $\alpha = -1$, we get

$$\sum_{r=1}^{\infty} (-1)^{r-1} \left( \sum_{m=1}^{\infty} \rho_m z^m \right)^r = k(z) = k(0) z + k(1) z^2 + k(2) z^3 + \ldots$$

This implies that

$$\sum_{n \in \mathbb{Z}} k(n) z^n = \sum_{n=0}^{\infty} k(n) z^n + k(-1) z^{-1} + \frac{1}{z} \cdot \sum_{n=0}^{\infty} k(n) z^{n+1}$$

(78)

as required.

Recall from our earlier discussion that these last few results regarding the kernel of a 1-dependent point process on the integers were due to following the options (69) and (70), which we did not know for certain a priori that they would lead to a legitimate kernel that would work for proving that these specific class of processes are determinantal, but we were then able to prove that it actually is a legitimate kernel by strong induction. It is then of independent interest whether or not following a different combination of options from the range (68)-(71) other than (69) and (70), which we already explored, would also lead to legitimate kernels, and moreover, how more/less complicated
(and, perhaps, different) would they be in contrast to the ones we proved in Theorem 4.2, Corollary 4.5, Theorem 4.3 and Corollary 4.4 (which we got by following the options (69) and (70)). For the reasons explained in the introductory section of the project (page limitations mostly), we have decided to omit this exploration.

Also recall how the options (68)-(71) are but a few of the (possibly) many more options that we could explore that yield legitimate kernels for the class of processes. Moreover, these options yield non-Hermitian Toeplitz matrices for point processes in $\mathcal{Y}$. This is a bit of an inconvenience because the literature for non-Hermitian Toeplitz matrices is rather scarce (as already mentioned many times). Indeed, the renowned Szegö theorems for Toeplitz matrices require Hermitianity. Thus, it would be worth finding out whether we could construct a matrix that is Toeplitz and Hermitian and that is a kernel for our class of point processes:

**Theorem 4.7** If a point process from $\mathcal{Y}$ is determinantal with (complex) kernel $K$ being a Hermitian Toeplitz matrix, then it must be tridiagonal of the following form:

$$
\begin{pmatrix}
\rho_1 & e^{i\theta} \sqrt{\rho_1^2 - \rho_2} & 0 & 0 & \cdots \\
e^{-i\theta} \sqrt{\rho_1^2 - \rho_2} & \rho_1 & 0 & 0 & \cdots \\
e^{-i\theta} \sqrt{\rho_1^2 - \rho_2} & e^{i\theta} \sqrt{\rho_1^2 - \rho_2} & \rho_1 & 0 & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots
\end{pmatrix}
$$

**Proof** The diagonal entries must all be $\rho_1$:

Indeed, for any $x \in \mathbb{Z}$,

$$
\rho(\{x\}) = \rho_1 = \det\left(K(x,x)\right) = K(x,x),  \quad (81)
$$

where the first equality is due to the translation-invariance assumption and the second equality is due to the determinantal structure assumption.

Next, notice how for any $x \in \mathbb{Z}$,

$$
\rho(\{x, x+1\}) = \rho_2
$$

$$
= \begin{vmatrix}
K(x, x) & K(x, x+1) \\
K(x+1, x) & K(x+1, x+1)
\end{vmatrix}
= \begin{vmatrix}
\rho_1 & K(x, x+1) \\
K(x+1, x) & \rho_1
\end{vmatrix}
= \rho_1^2 - K(x, x+1)K(x+1, x)  \quad (by \ (81))
$$

61
This implies that
\[ K(x, x + 1)K(x + 1, x) = \rho_1^2 - \rho_2. \]

And so by the Hermitianity assumption \( K(x, x + 1) = K(x + 1, x) \), we must have
\[ K(x, x + 1) = e^{i\theta} \sqrt{\rho_1^2 - \rho_2} \]
and
\[ K(x + 1, x) = e^{-i\theta} \sqrt{\rho_1^2 - \rho_2}, \]
for \( \theta \in [0, 2\pi] \).

Now let \( x, y \in \mathbb{Z} \) such that \( \text{dist}(x, y) \geq 2 \). Then, by the 1-dependence assumption,
\[
\rho(\{x, y\}) = \rho(\{x\})\rho(\{y\}) \\
= \rho_1^2 \quad \text{(by translation-invariance)}
\]

Also, by the determinantal structure assumption,
\[
\rho(\{x, y\}) = \begin{vmatrix} K(x, x) & K(x, y) \\ K(y, x) & K(y, y) \end{vmatrix} \\
= \begin{vmatrix} \rho_1 & K(x, y) \\ K(y, x) & \rho_1 \end{vmatrix} \quad \text{(by (81))} \\
= \rho_1^2 - K(x, y)K(y, x).
\]

And so we must have
\[ K(x, y)K(y, x) = 0 \]
for every \( x, y \in \mathbb{Z} \) such that \( \text{dist}(x, y) \geq 2 \). By the Hermitianity assumption, this can only happen if
\[ K(x, y) = 0 = K(y, x). \]

This then means that the resulting matrix is tridiagonal. \( \square \)

**Remark** One should be careful and note that the above theorem does not provide any sufficient conditions for a point process from \( \mathcal{Y} \) to be determinantal with a Hermitian Toeplitz kernel; it merely gives a necessary condition, namely, that it cannot be anything other than tridiagonal if it does happen to be Hermitian Toeplitz. A lot is known about tridiagonal Toeplitz matrices, both in the behaviour of their eigenvalues and of their determinants. Indeed, one can readily obtain precise expressions for the determinants of such matrices through recurrence relations. With that said, the
previous theorem can be used to rule out point processes from \( \mathcal{Y} \) that cannot have a Hermitian Toeplitz kernel by comparing the growth of the determinant of a tridiagonal matrix with \( \rho_n \) of the point process (since by definition of a point process \( X_1, \ldots, X_n \) being determinantal with kernel \( K \),
\[ \rho_n = \rho([n]) = \det(K). \]

**Remark** If one were to endeavour to actually provide sufficient conditions for a point process from \( \mathcal{Y} \) to have a kernel that is (real) Hermitian and Toeplitz, then one would find that there would have to be an abundance of necessary conditions on its correlation function that would have to be obeyed. We shall provide a few of these conditions here so as to provide yet another way of quickly ruling out point processes from \( \mathcal{Y} \) that cannot have their kernel being (real) Hermitian and Toeplitz.

Firstly, by the definition of a point process from \( \mathcal{Y} \), the associated (real) Hermitian kernel would have to satisfy, by the arguments from the proof of Theorem 4.7, the following equations for all \( x \in \mathbb{Z} \):
\[
\rho_3 = \begin{vmatrix} K(x, x) & K(x, x+1) & K(x, x+2) \\ K(x+1, x) & K(x+1, x+1) & K(x+1, x+2) \\ K(x+2, x) & K(x+2, x+1) & K(x+2, x+2) \end{vmatrix}
\]
\[
= \rho_1 \begin{vmatrix} \pm \sqrt{\rho_1^2 - \rho_2} & K(x, x+2) \\ \pm \sqrt{\rho_1^2 - \rho_2} & \rho_1 \end{vmatrix} \left( \begin{array}{c} \pm \sqrt{\rho_1^2 - \rho_2} \\ \rho_1 \end{array} \right) + \begin{vmatrix} K(x, x+2) \\ \rho_1 \end{vmatrix} \left( \begin{array}{c} \sqrt{\rho_1^2 - \rho_2} \\ \rho_1 \end{array} \right)
\]
\[
= \pm \sqrt{\rho_1^2 - \rho_2} K(x, x+2) \pm \sqrt{\rho_1^2 - \rho_2} \rho_1 \left( \begin{array}{c} \pm \sqrt{\rho_1^2 - \rho_2} \\ \rho_1 \end{array} \right) + \begin{vmatrix} K(x, x+2) \\ \rho_1 \end{vmatrix} \left( \begin{array}{c} \sqrt{\rho_1^2 - \rho_2} \\ \rho_1 \end{array} \right)
\]
\[
= \rho_1 \rho_2 - \sqrt{\rho_1^2 - \rho_2} \rho_1 \begin{vmatrix} \rho_1 \sqrt{\rho_1^2 - \rho_2} - K(x+2, x) \rho_1 \begin{vmatrix} \rho_1 \sqrt{\rho_1^2 - \rho_2} - \rho_2 - K(x+2, x) \rho_1 \end{vmatrix}
\]
\[
+ K(x, x+2) (\rho_1^2 - \rho_2 - K(x+2, x) \rho_1) = 0.
\]

By the Hermitianity assumption \( K(x, x+2) = K(x+2, x) \) we get the following quadratic equation
\[
\rho_1 K(x, x+2)^2 - 2(\rho_1^2 - \rho_2) K(x, x+2) + (\rho_3 - 2\rho_1 \rho_2 + \rho_1^3) = 0.
\]

Solving this equation yields
\[
K(x, x+2) = \frac{(\rho_1^2 - \rho_2) \pm \sqrt{\rho_1^2 - \rho_1 \rho_3}}{\rho_1}.
\]

By Theorem 4.7 a necessary condition for our process to have a Hermitian Toeplitz kernel is to have the above equal to 0. This can only happen, however, if the correlation function of the point
process satisfies the following equation

\[ \rho_1^2 - \rho_2 = \pm \sqrt{\rho_2^2 - \rho_1 \rho_3}. \]  \hspace{1cm} (82)

In the same way we can get another necessary condition, namely, the following equations need to be satisfied also:

\[
\rho_4 = \begin{vmatrix}
\rho_1 & \sqrt{\rho_1^2 - \rho_2} & 0 & K(x, x + 3) \\
\sqrt{\rho_1^2 - \rho_2} & \rho_1 & \sqrt{\rho_1^2 - \rho_2} & 0 \\
0 & \sqrt{\rho_1^2 - \rho_2} & \rho_1 & \sqrt{\rho_1^2 - \rho_2} \\
K(x + 3, x) & 0 & \sqrt{\rho_1^2 - \rho_2} & \rho_1
\end{vmatrix} \hspace{1cm} \text{(wlog)}
\]

\[
= \rho_1 \rho_3 - \rho_2 (\rho_1^2 - \rho_2) - K(x + 3, x) \sqrt{\rho_1^2 - \rho_2 (\rho_1^2 - \rho_2)} - K(x, x + 3) \sqrt{\rho_1^2 - \rho_2 (\rho_1^2 - \rho_2)} \\
- K(x, x + 3) K(x + 3, x) \rho_2.
\]

Again, by the Hermitianity assumption \( K(x, x + 3) = K(x + 3, x) \) we get the following quadratic equation

\[
\rho_2 K(x, x + 3)^2 + 2(\rho_1^2 - \rho_2)^{3/2} K(x, x + 3) + (\rho_4 - \rho_1 \rho_3 + \rho_1^2 \rho_2 - \rho_2^3) = 0.
\]

Solving this equation yields

\[
K(x, x + 3) = \frac{-(\rho_1^2 - \rho_2)^{3/2} \pm \sqrt{\rho_1^2 - 3 \rho_1^2 \rho_2 + 2 \rho_1^2 \rho_2^2 - \rho_2^4 + \rho_1 \rho_2 \rho_3}}{\rho_2}.
\]

Again, by Theorem 4.7 a necessary condition for our process to have Hermitian Toeplitz kernel is to have the above equal to 0. This can only happen, however, if the correlation function of the point process satisfies the following equation

\[
(\rho_1^2 - \rho_2)^{3/2} = \pm \sqrt{\rho_1^2 - 3 \rho_1^2 \rho_2 + 2 \rho_1^2 \rho_2^2 - \rho_2^4 + \rho_1 \rho_2 \rho_3}. \hspace{1cm} (83)
\]

So we now have two additional necessary conditions (83) and (82) which we can use to rule out point process from \( \mathcal{Y} \) that cannot have a (real) Hermitian Toeplitz kernel.

**Remark** At first glance there does not seem to be an obvious pattern in the two necessary conditions (82) and (83) so as for us to provide the (countably) many more necessary conditions on the correlation function of the point process from \( \mathcal{Y} \) that guarantee a Hermitian Toeplitz kernel for them.
Moreover, there are just so many of them to the point where even if we were to completely populate the list of necessary conditions, the results would only be of theoretical interest rather than of a practical one since it’s highly unlikely for the correlation function of an arbitrary point process from our class to satisfy all these, rather demanding, conditions. But this, however, does not rule out the possibility of the existence of some special point process from \( \mathcal{Y} \) that does indeed satisfy all the necessary conditions, and hence possesses a Hermitian Toeplitz kernel. Actually, one can easily check that the Bernoulli process is an example (it has a kernel being a diagonal matrix with \( \rho_1 = p \) as each diagonal entry). But so as not to deviate too much from the aim of the project, we stop our analysis on the possibility of Hermitian Toeplitz kernels for point processes from \( \mathcal{Y} \) here and return to it in Section 5.2 when we look at the possibility of kernels of our point process of interest (the Carries process with base \( \infty \)) being Hermitian Toeplitz.

5 Asymptotics for the thinned carries process

5.1 The Combinatorial Approach

We recall some basic facts and notations that were introduced in Section 2 regarding the \( p \)-spatially independently thinned carries process with \( b = \infty \), denoted by \( (X_i^{(p)})_{i \in \mathbb{N}} \), which we are interested in.

We saw that \( (X_i^{(p)})_{i \in \mathbb{N}} \) is both stationary and 1-dependent and we had denoted its \( n \)-gap probability by

\[
\hat{\rho}_n := P(X_1^{(p)} = 0, \ldots, X_n^{(p)} = 0).
\]

We had also computed in Section 2.3 some of the first few \( n \)-gap probabilities to illustrate the tediousness in computing these probabilities manually by relying on inclusion-exclusion without making use of appropriate mathematical tools which we’ll explore in this section. Just as a reminder, we had computed and found

\[
\hat{\rho}_1 = \frac{q + 1}{2!},
\]

\[
\hat{\rho}_2 = \frac{q^2 + 4q + 1}{3!},
\]

\[
\hat{\rho}_3 = \frac{q^3 + 11q^2 + 11q + 1}{4!},
\]

where \( q := 1 - p \).

One could argue that there is a pattern that can be observed here. The pattern in the denominator
is obvious. But what is the pattern in the numerator and how could we express it mathematically? This is where we have to turn to the definition/construction of our process.

Recall that for the corresponding un-thinned process \((X_i)_{i \in \mathbb{N}}\) we had that

\[
P(X_1 = 1, \ldots, X_n = 1) = P(U_1 < \ldots < U_{n+1})
\]

\[
= \frac{1}{(n + 1)!}
\]

\[
= P(U_1 \geq \ldots \geq U_{n+1})
\]

\[
= P(X_1 = 0, \ldots, X_n = 0),
\]

where \((U_i)_{i \in \mathbb{N}}\) are \(Unif[0,1]\) i.i.d. random variables independent of \((X_i)_{i \in \mathbb{N}}\).

From this we get

\[
\tilde{\rho}_n = P(\bigcup_{m=0}^{n} \{\text{there are exactly } m \text{ 1’s in } (X_1, \ldots, X_n) \text{ and we kill them}\})
\]

\[
= \sum_{m=0}^{n} P(\text{there are exactly } m \text{ 1’s in } (X_1, \ldots, X_n)) \cdot P(\text{we kill these } m \text{ 1’s})
\]

\[
= \sum_{m=0}^{n} P(\text{there are exactly } m \text{ ascents in } (U_1, \ldots, U_{n+1})) \cdot q^m \quad \text{(spatial-independent thinning)},
\]

where the second equality is due to the independence of the underlying Bernoulli random variables of the thinned process (defined in Section 2.1) and the \(X_i\)’s.

Let’s analyse the \(P(\text{there are exactly } m \text{ ascents in } (U_1, \ldots, U_{n+1}))\) term.

If we denote \(A(n+1, m)\) as the number of permutations in \(S_{n+1}\) that have exactly \(m\) ascents, then, since the \(U_i\)’s are independent and identically distributed we have

\[
P(\exists \text{ exactly } m \text{ ascents in } (U_1, \ldots, U_{n+1})) = \frac{A(n+1, m)}{(n + 1)!},
\]

where this equality is due to the fact that the \(U_i\)’s are continuous random variables (and therefore the events where any of them are exactly equal to one another are improbable) and the fact that there is the symmetry we had explained in (25) and (19).

And so

\[
\tilde{\rho}_n = \frac{1}{(n + 1)!} \cdot \sum_{m=0}^{n} A(n+1, m) \cdot q^m.
\]
Thankfully, it turns out, the numbers \( A(n+1, m) \) are well studied. They are called the Eulerian numbers, and moreover,

\[
\sum_{m=0}^{n} A(n+1, m) \cdot q^m =: A_{n+1}(q)
\]

is called the \((n+1)^{th}\) Eulerian polynomial. Therefore,

\[
\tilde{\rho}_n = \frac{A_{n+1}(q)}{(n+1)!},
\]

and so we’ve found the exact pattern on the numerator of the \( \tilde{\rho}_n \)’s we were looking for as well.

To further evaluate these gap probabilities and also obtain asymptotics for them let’s formally introduce Eulerian numbers and polynomials.

**Definition (e.g., see [20])** \( A(n, m) \), the **Eulerian number**, is the number of permutations of \([n]\) in which exactly \( m \) elements are greater than the previous element (i.e., the total number of permutations with \( m \) "ascents"). It is clear that \( A(n, 0) = 1 \) since the only permutation of \([n]\) that has 0 "ascents" is precisely \((n, n−1, n−2, \ldots, 1) \in S_n\). \( A(n, n−1) \), similarly, equals 1 since the only permutation of \([n]\) that has \( n−1 \) "ascents" is precisely \((1, 2, \ldots, n−1, n) \in S_n\). We also define \( A(n, n) \equiv 0 \) since it’s impossible to have \( n \) "ascents" in a permutation that only has \( n \) elements.

We then define

\[
A_n(t) := \sum_{m=0}^{n} A(n, m)t^m
\]

to be the \( n^{th} \) **Eulerian polynomial**. We also define \( A_0(t) \equiv 1 \).

As mentioned previously, a fair amount is known about Eulerian numbers and polynomials. In particular, we have full knowledge of the exponential generating function for the latter:

**Lemma 5.1 (Theorem 1.6, [20])** The Eulerian polynomials are defined by the exponential generating function

\[
\sum_{n=0}^{\infty} A_n(t) \cdot \frac{x^n}{n!} = \frac{t - 1}{t - e^{(t-1)x}}, \quad t \neq 1, \text{ and } x \in \mathbb{C} \text{ with } |x| \text{ sufficiently small.}
\]

**Remark** Eulerian polynomials can be computed by the recurrence

\[
A_0(t) = 1,
\]
\[ A_n(t) = t(1 - t) \cdot A'_{n-1}(t) + A_{n-1}(t) \cdot (1 + (n-1)t), \quad n \geq 1. \]

This implies in particular that \( A_1(t) = 1. \)

Lemma 5.1 will allow us to obtain the asymptotics for the \( n \)-gap probability, \( \tilde{\rho}_n \), that we are interested in:

First recall that

\[
\sum_{n=1}^{\infty} \tilde{\rho}_n z^n = \sum_{n=1}^{\infty} A_{n+1}(q) \cdot \frac{z^n}{(n+1)!} \\
= \frac{1}{z} \sum_{n=2}^{\infty} A_n(q) \cdot \frac{z^n}{n!} \\
= \frac{1}{z} \left( \sum_{n=0}^{\infty} A_n(q) \cdot \frac{z^n}{n!} - A_0(q) \cdot \frac{z^0}{0!} - A_1(q) \cdot \frac{z^1}{1!} \right) \\
= \frac{1}{z} \left( \sum_{n=0}^{\infty} A_n(q) \cdot \frac{z^n}{n!} - 1 - z \right)
\]

And so by Lemma 5.1,

\[
\sum_{n=1}^{\infty} \tilde{\rho}_n z^n = \frac{1}{z} \left( \frac{q-1}{q-e(q-1)z} - 1 - z \right). 
\tag{84}
\]

So then, for \( n \in \mathbb{N} \), and \( r < 1 \) sufficiently small so that the exponential generating function for the Eulerian polynomials is well-defined,

\[
\tilde{\rho}_n = \frac{1}{2\pi} \int_0^{2\pi} \left( \sum_{k=1}^{\infty} \tilde{\rho}_k e^{ikx} \right) r^{-n} e^{-inx} dx \\
= \frac{1}{2\pi} \int_0^{2\pi} \left( \sum_{k=1}^{\infty} \tilde{\rho}_k (re^{ix})^k \right) (re^{ix})^{-n} dx \\
= \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{re^{ix}} \left( \frac{q-1}{q-e(q-1)re^{ix}} - 1 - re^{ix} \right) (re^{ix})^{-n} dx \
\quad \text{(by setting } z = re^{ix} \text{ in (84))} \\
= \frac{1}{2\pi i} \oint_{\partial B_r(0)^+} \frac{1}{z^{n+2}} \left( \frac{q-1}{q-e(q-1)z} - 1 - z \right) dz \
\quad \text{(converting into a Contour integral)} \\
= \frac{1}{2\pi i} \oint_{\partial B_r(0)^+} \frac{1}{z^{n+2}} \cdot \frac{q-1}{q-e(q-1)z} dz - \frac{1}{2\pi i} \oint_{\partial B_r(0)^+} \frac{1}{z^{n+2}} \cdot (1 + z) dz \\
= \frac{1}{2\pi i} \oint_{\partial B_r(0)^+} \frac{1}{z^{n+2}} \cdot \frac{q-1}{q-e(q-1)z} dz \
\quad \text{(since } n \in \mathbb{N})
\]
We would like to apply the Residue Theorem. For that we need to find all the poles of the above integrand. Clearly \( z = 0 \) is a pole (a high order one). To determine the other poles of the integrand, we solve for \( z \)

\[
q - e^{(q-1)z} = 0.
\]

So we find that

\[
z_k = \frac{\log(q) + 2\pi ik}{q - 1}, \quad k \in \mathbb{Z},
\]

are all the other (simple) poles.

But how many of our poles are in the interior of \( B_r(0) \)?

First, it’s easy to see that

\[
\frac{\log(q)}{q - 1} > 1 \quad \forall \ 0 < q < 1.
\]

Therefore, the real contribution of \( z_k \) is never inside our curve \( \partial B_r(0)^+ \) (for all values \( k \in \mathbb{Z} \)). It’s also clear that the imaginary contribution of \( z_k, \frac{2\pi k}{q-1} \), does not lie in the interior of \( \partial B_r(0)^+ \) since \( |q - 1| < 1 \) for \( 0 < q < 1 \). As a result the (simple) poles of the integrand \( z_k, k \in \mathbb{Z} \), are not in the interior of our curve and we only have \( z = 0 \) (a high order pole) inside \( B_r(0) \).

Now, by computing the residue of the high order pole \( z = 0 \) of the integrand, which involves computations of high order derivatives, we are able to get a very precise (but complicated) expression for \( \tilde{\rho}_n \). Additionally, by doing so, we also obtain an exact expression for the \( n \)-th Eulerian polynomial \( A_n(q) \), which is of independent interest, and, to the best of our knowledge having read published work on Eulerian polynomials and numbers where authors seem to only give a recursive, and not exact, expression for Eulerian polynomials, is a novel result. We shall do this at the end of this subchapter. For now, since we are interested in the asymptotic behaviour of \( \tilde{\rho}_n \) for large \( n \), we would like to only have to deal with the simple poles \( z_k \). The way we do this is by utilizing a neat trick from complex analysis that was used in the famous calculation of the Fourier transform of \( \frac{1}{\cosh(x)} \).

Namely, we reverse the orientation of our contour of integration from anti-clockwise to clockwise and apply Cauchy’s Deformation Theorem, that is,

\[
\tilde{\rho}_n = \frac{1}{2\pi i} \oint_{\partial B_r(0)^+} \frac{z^{n+2}}{z^{n+2} \cdot \frac{q - 1}{q - e^{(q-1)z}}} \, dz
\]

\[
= -\frac{1}{2\pi i} \oint_{\partial B_r(0)^-} \frac{z^{n+2}}{z^{n+2} \cdot \frac{q - 1}{q - e^{(q-1)z}}} \, dz, \quad \text{(by Cauchy’s Deformation Theorem)},
\]

69
and as a result, in the interior of \( \partial B_r(0) \) there are only the simple poles \( z_k, \ k \in \mathbb{Z} \), without the high order pole \( z = 0 \). Let’s now apply the Residue Theorem:

\[
\hat{\rho}_n = \frac{1 - q}{2\pi i} \cdot 2\pi i \sum_{k \in \mathbb{Z}} \text{Res}(\frac{1}{z^{n+2}} \cdot \frac{1}{q - e(q-1)z}, z_k)
\]

\[
= (1 - q) \sum_{k \in \mathbb{Z}} \text{Res}(\frac{1}{z^{n+2}} \cdot \frac{1}{q - e(q-1)z}, z_k).
\]

All that’s left now is to compute the residues of the integrand. We do so by considering its Laurent series: Observe how for \( k \in \mathbb{Z} \),

\[
\text{Res}(\frac{1}{z^{n+2}} \cdot \frac{1}{q - e(q-1)z}, z_k) = \lim_{z \to z_k} (z - z_k) \cdot \frac{1}{z^{n+2}(q - e(q-1)z)}
\]

\[
= \lim_{z \to z_k} \frac{1}{(n + 2)qz^{n+1} - (n + 2)z^{n+1}e(q-1)z - (q - 1)z^{n+2}e(q-1)z} \quad \text{(by L'Hospital's rule)}
\]

\[
= \lim_{z \to z_k} \frac{1}{(n + 2)z^{n+1}(q - e(q-1)z) - (q - 1)z^{n+2}e(q-1)z}
\]

\[
= -\frac{1}{q(q - 1)z_k^{n+2}}.
\]

Therefore,

\[
\hat{\rho}_n = (1 - q) \sum_{k \in \mathbb{Z}} \text{Res}(\frac{1}{z^{n+2}} \cdot \frac{1}{q - e(q-1)z}, z_k)
\]

\[
= (1 - q) \sum_{k \in \mathbb{Z}} -\frac{1}{q(q - 1)z_k^{n+2}}
\]

\[
= \frac{1}{q} \sum_{k \in \mathbb{Z}} \frac{1}{z_k^{n+2}}.
\]

It is clear that

\[
z_k = z_0 + \frac{2\pi ik}{q - 1}, \quad k \in \mathbb{Z}.
\]

Therefore,

\[
\hat{\rho}_n = \frac{1}{q} \left( z_0^{-n+2} + \sum_{k \in \mathbb{Z} \setminus \{0\}} (z_0 + \frac{2\pi ik}{q - 1})^{-n-2} \right)
\]

\[
= \frac{1}{q} z_0^{-n+2} \left( 1 + \frac{1}{z_0^{-n-2}} \sum_{k \in \mathbb{Z} \setminus \{0\}} (z_0 + \frac{2\pi ik}{q - 1})^{-n-2} \right).
\]

Now, observe how

\[
\frac{1}{z_0^{-n+2}} \sum_{k \in \mathbb{Z} \setminus \{0\}} (z_0 + \frac{2\pi ik}{q - 1})^{-n-2} = \sum_{k \in \mathbb{Z} \setminus \{0\}} \frac{z_0}{z_k^{n+2}}
\]
tends to 0 as \( n \to \infty \):

Indeed, since \( |z_k| = \sqrt{z_0^2 + \left( \frac{2\pi k}{q-1} \right)^2} > 1 \) for all \( k \in \mathbb{Z} \setminus \{0\} \) and \( |z_0| = \frac{\log(q)}{q-1} \), we have

\[
\left| \frac{z_0}{z_k} \right| < 1, \quad (85)
\]

and so

\[
\left| \frac{z_0}{z_k} \right|^{n+2} < \left| \frac{z_0}{z_k} \right|^2 = \left( \frac{\log(q)}{q-1} \right)^2 \cdot \frac{1}{z_0^2 + \left( \frac{2\pi k}{q-1} \right)^2} < \left( \frac{\log(q)}{2\pi} \right)^2 \cdot \frac{1}{k^2}, \quad \forall k \in \mathbb{Z} \setminus \{0\} \; \forall n \in \mathbb{N},
\]

and we know that

\[
\left( \frac{\log(q)}{2\pi} \right)^2 \sum_{k \in \mathbb{Z} \setminus \{0\}} \frac{1}{k^2} < \infty.
\]

Therefore, by the Weierstrass M-test,

\[
\sum_{k \in \mathbb{Z} \setminus \{0\}} \left( \frac{z_0}{z_k} \right)^{n+2}
\]

converges uniformly in \( n \in \mathbb{N} \). This allows us to take the limit as \( n \to \infty \) inside the above infinite summation and get, by (85), that

\[
E(n) := \sum_{k \in \mathbb{Z} \setminus \{0\}} \left( \frac{z_0}{z_k} \right)^{n+2}
\]

converges to 0 as \( n \to \infty \), as previously stated.

Finally,

\[
\hat{\rho}_n = \frac{1}{q} z_0^{-n-2} \left( 1 + \frac{1}{z_0^{-n-2}} \sum_{k \in \mathbb{Z} \setminus \{0\}} \left( z_0 + \frac{2\pi ik}{q-1} \right)^{-n-2} \right) \quad (86)
\]

\[
= \frac{1}{q} z_0^{-n-2} (1 + E(n)) \quad (87)
\]

\[
= \frac{1}{q} \left( \frac{\log(q)}{q-1} \right)^{-n-2} \cdot (1 + E(n)) \quad (88)
\]

\[
= \frac{(1 - q)^n}{(\log(\frac{1}{q}))^n} \cdot \frac{(1 - q)^2}{q \cdot (\log(q))^2} \cdot (1 + E(n)) \quad (89)
\]

\[
= \frac{p^n}{(\log(\frac{1}{1-p}))^n} \cdot \frac{p^2}{(1 - p) \cdot (\log(1-p))^2} \cdot (1 + E(n)) \quad (90)
\]

And so we obtain the following asymptotic for the \( n \)-gap probabilities of our process:
**Theorem 5.2** The $n$-gap probability of the $p$-thinned Carries process $(X_i^{(p)})_{i \in \mathbb{N}}$ with base $b \to \infty$ for large $n > 0$ decays at the following rate

\[ \tilde{\rho}_n \approx \alpha \cdot \frac{p^n}{(\log(\frac{1}{1-p}))^n}, \]

where

\[ \alpha = \frac{p^2}{(1-p) \cdot (\log(1-p))^2} \]

(for thinning parameters $0 < p < 1$).

**Remark** Observe how, if we take the limit as $p \to 0$ on the above expression, then the expression tends to 1. This makes sense because $p = 0$ implies that we kill particles with probability 1, and so, of course, $\tilde{\rho}_n = 1$ would have to be the case.

Now that we’ve obtained the asymptotic we were after, let’s turn our attention to the high order pole of the integrand we had mentioned a while ago at $z = 0$. We had mentioned that if we successfully computed the residue of the integrand at that pole, then we would get a very precise (non-asymptotic) expression for $\tilde{\rho}_n$ and we would also obtain an exact (non-recursive) expression for the $n$-th Eulerian polynomials as well. We first need a lemma from complex analysis:

**Lemma 5.3** (formula for calculating high order poles, [17], p.435) If $a \in \mathbb{C}$ is an $m$th order pole of a complex-valued function $f$, then

\[ \text{Res}(f(z), a) = \frac{1}{(m-1)!} \cdot \frac{d^{m-1}}{dz^{m-1}}((z-a)^m f(z)) \bigg|_{z=a}. \] \hspace{1cm} (91)

Recall we had found that

\[ \tilde{\rho}_n = \frac{q-1}{2\pi i} \int_{\partial B_r(0)^+} \frac{1}{z^{n+2}} \cdot \frac{1}{q - e^{(q-1)z}} dz, \]

where the integrand has only $z = 0$ as a pole (of order $n + 2$) that lies in the interior of our curve $\partial B_r(0)^+$.

We shall find the corresponding residue by applying Lemma 5.3. So the first step is to multiply the integrand by $z^{n+2}$ and the second step is to find an expression for the $m$th derivative of the resulting term.
After some computations we find the general \( m \)th derivative
\[
\frac{d^m}{dz^m} \left( \frac{1}{q - e(q-1)z} \right) = \sum_{k=1}^{m} A_k \cdot \frac{(q - 1)^m e^{k(q-1)z}}{(q - e(q-1)z)^{k+1}},
\]
where,
\[
A_1 = 1,
\]
\[
A_k = \sum_{j_1=k-1}^{m-1} k^{n-j_1} \sum_{l=1}^{k-2} \sum_{j_{i+1}=1}^{j_l} (k - l)^{j_{i+1}}, \quad k = 2, \ldots, m.
\]
Therefore, by Lemma 5.3,
\[
\text{Res}(\frac{1}{z^{n+2}} \frac{1}{q - e(q-1)z}, 0) = \frac{1}{(n+1)!} \left( (q - 1)^{n-1} + \sum_{k=2}^{n+1} (q - 1)^{n-k} \sum_{j_1=k-1}^{n} k^{n+1-j_1} \sum_{l=1}^{k-2} \sum_{j_{i+1}=1}^{j_l} (k - l)^{j_{i+1}} \right).
\]
And so by the Residue Theorem,
\[
\tilde{\rho}_n = \frac{1}{(n+1)!} \left( (q - 1)^{n} + \sum_{k=2}^{n+1} (q - 1)^{n+1-k} \sum_{j_1=k-1}^{n} k^{n+1-j_1} \sum_{l=1}^{k-2} \sum_{j_{i+1}=1}^{j_l} (k - l)^{j_{i+1}} \right).
\]
Now recall that
\[
\tilde{\rho}_n = \frac{A_{n+1}(q)}{(n+1)!}.
\]
And so the explicit expression we found for \( \tilde{\rho}_n \) in (92) implies the following (to the best of our knowledge) novel result regarding Eulerian polynomials:

**Theorem 5.4** An exact (non-recursive) expression for the \( n \)-th Eulerian polynomial, \( A_n(q) \), is as follows
\[
A_n(q) = (q - 1)^{n-1} + \sum_{k=2}^{n} (q - 1)^{n-k} \sum_{j_1=k-1}^{n-1} k^{n-j_1} \sum_{l=1}^{k-2} \sum_{j_{i+1}=1}^{j_l} (k - l)^{j_{i+1}}.
\]

**Proof** This theorem can be proved by induction.  

To summarize this subsection, we have found, via Eulerian polynomials, an exact expression for the \( n \)-gap probabilities of our process (92) as well as an asymptotic for them (90).
In the next subsection we will try and obtain this same asymptotic from Theorem 5.2 by applying (and developing) results from Section 3.

5.2 The Determinantal Structure Approach

We recall the basic facts and results we had derived from earlier sections regarding the carries processes with base $b = \infty$, $(X_i)_{i \in \mathbb{N}}$, and its $p$-thinned version, $(X_i^{(p)})_{i \in \mathbb{N}}$, which we are interested in studying its gap probabilities.

**Fact 1**: $(X_i)_{i \in \mathbb{N}}$ is in the $\mathcal{Y}$ class of point processes (i.e., it is both 1-dependent and translation-invariant).

**Fact 2**: $(X_i)_{i \in \mathbb{N}}$ has
\[
\rho_n := P(X_1 = 1, \ldots, X_n = 1) = \frac{1}{(n+1)!}.
\]

**Fact 3**: We denote the $n$-gap probabilities of $(X_i^{(p)})_{i \in \mathbb{N}}$ by
\[
\tilde{\rho}_n := P(X_1^{(p)} = 0, \ldots, X_n^{(p)} = 0).
\]

We had established in Section 4 that point processes from the $\mathcal{Y}$ class are determinantal and we had given a range of valid kernels (Corollaries 4.4 and 4.5) that were expressed in terms of the $\rho_n$ function of the respective point process. So by Fact 1 we have that $(X_i)_{i \in \mathbb{N}}$ is determinantal with some kernel $T$ expressed in terms of the $\rho_n$ given in Fact 2. Therefore, by Theorem 2.2, $(X_i^{(p)})_{i \in \mathbb{N}}$ is also determinantal but with kernel $pT$. So then, by Theorem 2.1 and (1)-(3),
\[
\tilde{\rho}_n = \det(I_n - pT_n),
\]
where $T_n$ denotes the $n \times n$ leading principal submatrix of $T$.

Now, all the kernels mentioned on the previous paragraph are non-Hermitian Toeplitz, but we had also given, towards the end of Section 4, some necessary conditions for the existence of Hermitian kernels for the DPP. As discussed previously, it is worth checking whether there is any hope for a Hermitian kernel for our Carries process $(X_i)_{i \in \mathbb{N}}$ (so that we can apply the actual Szegő Theorems 3.2, 3.3 to the kernel to get an asymptotic for $\tilde{\rho}_n$ given in (93) instead of having to use the weaker non-Hermitian analogues (Theorems 3.4, 3.5 and 3.6) we had developed and reported).

So let’s find out if the Carries process $(X_i)_{i \in \mathbb{N}}$ satisfies conditions (82) and (83).
By Fact 2,

\[ \rho_1^2 - \rho_2 = \frac{1}{12} = \sqrt{\rho_2^2 - \rho_1 \rho_3}, \]

and so the necessary condition (82) is satisfied, which gives us a bit of hope for a Hermitian kernel - this is, after all, quite a demanding property on the function \( \rho_n \) that is satisfied.

Let’s see if (83) is satisfied. Again by Fact 2,

\[ (\rho_1^2 - \rho_2)^{3/2} = (1/12)^{3/2} \neq 0.01863389981 = \sqrt{\rho_1^4 - 3\rho_1^2 \rho_2 + 2\rho_1 \rho_2^2 - \rho_2^4 + \rho_1 \rho_2^3}. \]

Thus, the necessary condition for a Hermitian kernel, (83), is not satisfied.

This rules out any possibility for a kernel of \((X_i)_{i \in \mathbb{N}}\) to be Hermitian.

There is actually a neater and more elegant way of checking the above fact by making use of Theorem 4.7, Fact 2, and the known formulae for the determinants of tridiagonal Toeplitz matrices (e.g., see [18]). Namely, if it were the case that there exists a Hermitian Toeplitz kernel \( T \) for \((X_i)_{i \in \mathbb{N}}, \) then by Theorem 4.7 it would have to be tridiagonal, and so the determinant of \( T_n, \)

\[ \det(T_n) = \frac{1}{\sqrt{4 \rho_2 - 3 \rho_1^2}} \left( \left( \frac{\rho_1 + \sqrt{4 \rho_2 - 3 \rho_1^2}}{2} \right)^{n+1} - \left( \frac{\rho_1 - \sqrt{4 \rho_2 - 3 \rho_1^2}}{2} \right)^{n+1} \right). \]

But we also know that

\[ \rho_n = \det(T_n) \]

since \((X_i)_{i \in \mathbb{N}}\) is determinantal. And so by Fact 2, we also have

\[ \det(T_n) = \frac{1}{(n + 1)!}. \]

This is a contradiction since factorials grow faster than exponential functions!

Now that we have established the fact that a kernel of \((X_i)_{i \in \mathbb{N}}\) must be non-Hermitian Toeplitz, let’s choose for it the simplest kernel out of all those we constructed in Section 4. Recall how we had found the generating function for the entries of the matrix given in Corollary 4.5 in Theorem 4.6. Applying these two results on \((X_i)_{i \in \mathbb{N}}\) by utilising Fact 2 yield the following results:

**Theorem 5.5** The carries process \((X_i)_{i \in \mathbb{N}}\) has a non-Hermitian Toeplitz kernel \( T := (T_{ij})_{i,j \in \mathbb{N}} \equiv \)
\[(t_{j-i})_{i,j \in \mathbb{N}}\] given by

\[
t_n = \begin{cases} 
0, & \text{if } n \leq -2 \\
-1, & \text{if } n = -1 \\
\sum_{r=1}^{n+1} (-1)^{r-1} \frac{1}{(l_1+1)! \cdots (l_r+1)!}, & \text{if } n \geq 0 
\end{cases}
\]

\textbf{Theorem 5.6} \( t_n, n \in \mathbb{Z} \), from Theorem 5.5 has generating function given by:

\[
\sum_{n \in \mathbb{Z}} t_n z^n = \frac{1}{1 - ez}, \quad \text{for all } z \in \mathbb{C} \text{ with sufficiently small } |z|.
\]

\textbf{Proof} By Fact 2, the function \( R(z) \) given in Theorem 4.6 applied to the Carries process is precisely

\[
R(z) = 1 + z + \sum_{n=1}^{\infty} \frac{1}{(n+1)!} z^{n+1} = e^z.
\]

\textbf{Remark} We were not very precise as to how small we need \(|z|\) (for \( z \in \mathbb{C} \)) to be for the statement of Theorem 5.6 to hold. In order to use the statement of Theorem 5.6 to derive the symbol of the Toeplitz matrix given in Theorem 5.5, we would need to prove that the former theorem holds for any \( z \in \mathbb{C} \) with \(|z| = 1\). In this way we would be justified in plugging in \( z = e^{i\lambda}, \lambda \in [0, 2\pi] \), into Theorem 5.6 and get the desired symbol.

Essentially, we need to prove that for \( z \in \mathbb{C} \) with \(|z| = 1\), the step (79) is justified (for \( \rho_n = \frac{1}{(n+1)!} \), by Fact 2). That is, we need to prove that

\[
\left| \sum_{m=1}^{\infty} -\rho_m z^m \right| = \left| \sum_{m=1}^{\infty} -\frac{1}{(m+1)!} z^m \right| < 1.
\]

But

\[
\left| \sum_{m=1}^{\infty} -\frac{1}{(m+1)!} z^m \right| \leq \sum_{m=1}^{\infty} \frac{1}{(m+1)!} \quad \text{(by the triangle inequality and } |z| = 1) \\
= \sum_{m=0}^{\infty} \frac{1}{m!} - 2 \\
= e - 2 < 1,
\]
as required.

So the symbol $f : [0, 2\pi] \to \mathbb{C}$ for the Toeplitz matrix given in Theorem 5.5 is given by

$$f(\lambda) = \sum_{n=-\infty}^{\infty} t_n e^{in\lambda} = \frac{1}{1 - e^{i\lambda}}. \quad (94)$$

We would like to find more precise expressions (from those we gave in Theorem 5.5) for the entries $t_n$ for $n \in \mathbb{N}_0$ by utilizing the generating function from above.

**Corollary 5.7 (p.12, remark 2, [3])** For $n \in \mathbb{N}_0$, $t_n$ from Theorem 5.5 is exactly equal to

$$t_n = -\frac{B_{n+1}}{(n+1)!},$$

where $B_n$ denotes the $n^{th}$ Bernoulli number.

**Proof** By Theorem 5.6,

$$\sum_{n\in\mathbb{Z}} t_n z^{n+1} = \frac{z}{1 - e^z} = -\frac{z}{e^z - 1} = -\sum_{n=0}^{\infty} \frac{B_n z^n}{n!},$$

where the last equality holds from the definition of the exponential generating function for Bernoulli numbers (e.g., see [15]).

By comparing coefficients of the far LHS with the far RHS of the above equation, one gets the result.

**Remark** Further to this, since $B_0 = 1$ and $B_n = 0$ for every $n < 0$, it’s easy to see that

$$t_n = -\frac{B_{n+1}}{(n+1)!}$$

actually holds for every $n \in \mathbb{Z}$.

We would now like to make use of the non-Hermitian Szegő theorem analogues we have at our disposal from Section 3.2 to evaluate (93).

Let’s start off by working out the Szegő leading term of $(I_n - pT_n)_{n \in \mathbb{N}}$ by making use of our Theorem 3.4.

77
We first check whether our symbol \( f \) satisfies condition (28) of Theorem 3.4. For this we need a lemma from the theory of Bernoulli numbers.

**Lemma 5.8** (p.18, [15]) For \( n \in \mathbb{N} \),

\[
|B_{2n}| = \frac{2\zeta(2n)(2n)!}{(2\pi)^{2n}},
\]

where \( \zeta(n) := \sum_{k=1}^{\infty} \frac{1}{k^n} \) is the Riemann zeta function.

We want to see if

\[
\sum_{n=-\infty}^{\infty} |t_n|
\]

converges.

Notice how, by Corollary 5.7 and Theorem 5.5,

\[
\sum_{n=-\infty}^{\infty} |t_n| = 1 + 1/2 + \sum_{n=2}^{\infty} \frac{|B_n|}{n!}
= 3/2 + \sum_{n=1}^{\infty} \frac{|B_{2n}|}{(2n)!},
\]

where the last equality holds due to the known fact that

\[
B_{2n+1} = 0 \quad \forall n \geq 1.
\]

Let’s see if

\[
\sum_{n=1}^{\infty} \frac{|B_{2n}|}{(2n)!} = \sum_{n=1}^{\infty} a_n
\]

converges by using the ratio test.

Observe how

\[
\frac{|a_{n+1}|}{a_n} = \frac{(2n)!}{(2n + 2)!} \cdot \frac{|B_{2n+2}|}{|B_{2n}|}
= \frac{1}{(2n + 2)(2n + 1)} \cdot \frac{|B_{2n+2}|}{|B_{2n}|}
= \frac{1}{(2n + 2)(2n + 1)} \cdot \frac{\zeta(2n + 2)}{\zeta(2n)} \cdot \frac{(2n + 2)!2n}{(2n)!2n+2}.
\]
It is easy to see that
\[ \zeta(2n + 2) = \sum_{k=1}^{\infty} \frac{1}{k^{2n}} \cdot \frac{1}{k^2} \leq \sum_{k=1}^{\infty} \frac{1}{k^{2n}} = \zeta(2n). \]

This implies that
\[
\lim_{n \to \infty} \left| \frac{a_{n+1}}{a_n} \right| \leq \lim_{n \to \infty} \frac{1}{(2n + 2)(2n + 1)} \cdot \frac{(2n + 2)! (2\pi)^{2n}}{(2n)! (2\pi)^{2n+2}} \\
= \lim_{n \to \infty} \frac{1}{(2n + 2)(2n + 1)} \cdot \frac{(2n + 2)(2n + 1)}{(2\pi)^2} \\
= \frac{1}{(2\pi)^2} \\
< 1.
\]

Therefore
\[ \sum_{n=1}^{\infty} \frac{|B_{2n}|}{(2n)!}, \]

and hence,
\[ \sum_{n=-\infty}^{\infty} |t_n| \]

converges by the ratio test.

So the (28) condition is in fact satisfied by our symbol \( f \).

Next we need to check for which thinning parameters \( 0 < p < 1 \) of our thinned Carries process Theorem 3.4 holds for. That is, for what range of values of \( 0 < p < 1 \) do the conditions (36) and (37) hold for our symbol \( f \)?

To satisfy condition (37), we need \( 0 < p < 1 \) sufficiently small such that
\[ pM_{|f|} < 1. \]

It is easy to check that \( M_{|f|} = \frac{1}{1-e^{-1}} \). Therefore, we need
\[ p < 1 - e^{-1} \]  
(95)
for (37) to be satisfied.

As for the condition (36), Lemma 3.1 established a uniform bound (in \( n \) - the size of the kernel) on the modulus of the eigenvalues of \( (T_n)_{n \geq 1} \), namely, \( 2M_{|f|} \). This bound means that for (36) to be
satisfied, we would definitely need
\[ p < \frac{1 - e^{-1}}{2}. \]  \hspace{1cm} (96)

The factor of a $\frac{1}{2}$ above is slightly inconvenient as it means that we only know that Theorem 3.4 holds for just half of the thinning parameters $p$ given in (95).

Thus, for $p < \frac{1 - e^{-1}}{2}$, by applying Theorem 3.4, we get the Szegő leading term for $\det(I_n - pT_n(f))$:

\[ \exp \left\{ \frac{n}{2\pi} \int_0^{2\pi} \log(1 - pf(\lambda))d\lambda \right\}. \]

We use a trick to obtain
\[ I(p) := \frac{1}{2\pi} \int_0^{2\pi} \log(1 - pf(\lambda))d\lambda = \frac{1}{2\pi} \int_0^{2\pi} \log(1 - \frac{p}{1 - e^{i\lambda}})d\lambda. \]

Namely, we differentiate the above with respect to $p$ and by the Leibniz rule we are able to take the derivative inside the integral and obtain:

\[ I'(p) = -\frac{1}{2\pi} \int_0^{2\pi} \frac{1}{1 - e^{i\lambda}} \cdot \frac{1}{1 - \frac{p}{1 - e^{i\lambda}}} d\lambda = -\frac{1}{2\pi} \int_0^{2\pi} \frac{1}{1 - e^{i\lambda} - p} d\lambda \]
\[ = -\frac{1}{2\pi i} \int_0^{2\pi} \frac{1}{e^{i\lambda} \cdot 1 - e^{i\lambda} - p} \cdot i \cdot e^{i\lambda} d\lambda \]
\[ = -\frac{1}{2\pi i} \oint_{\partial B_1(0)^+} \frac{1}{z} \cdot \frac{1}{1 - e^{z} - p} dz. \]

We would like to apply the Residue theorem. For that we need to find the poles of the above integrand that lie in the unit circle centred at the origin.

The above integrand has poles at $z = 0$ and $z_n = \log(1 - p) + 2\pi in$ for $n \in \mathbb{Z}$.

Notice how $-1 < \log(1 - p) < 1$ for $0 < p < \frac{1 - e^{-1}}{2}$ that we’ve assumed. Hence, the integrand has only poles $z = 0$ and $z = z_0 = \log(1 - p)$ that are inside $B_1(0)$.

Actually, the previous statement holds true for exactly $0 < p < 1 - e^{-1}$ thinning parameters. Interestingly, this inequality was exactly the requirement we had found in (95) in order for our symbol to satisfy the condition (37) of Theorem 3.4. So we are getting the correct answer (that is, we are getting the correct residues of the integrand) for a bigger range of values of $p$ than the
one that the bound we currently have, (31), on the eigenvalues of our kernel elicits (the $p < \frac{1-e^{-1}}{2}$ restriction). This alludes to the possibility of even tighter bounds for eigenvalues of non-Hermitian Toeplitz matrices. Namely, since $M[f] = \frac{1}{1-e^{-1}}$, it alludes to the possibility that we can somehow get rid of the 2 from the (31) bound. If this turns out to indeed be the case, then we can further strengthen the result on the thinned carries process we obtain by having it work for a bigger range of thinning parameters $p$, namely, for $0 < p < 1 - e^{-1}$. We will discuss this further after we do the computations to obtain the Szégo leading term.

Now, by the Residue theorem, we have

$$I'(p) = -\text{Res}\left(\frac{1}{z} \cdot \frac{1}{1 - e^z - p}, 0\right) - \text{Res}\left(\frac{1}{z} \cdot \frac{1}{1 - e^z - p}, \log(1 - p)\right).$$

We find these two residues by considering the Laurent series of $\frac{1}{z} \cdot \frac{1}{1 - e^z - p}$.

First,

$$\text{Res}\left(\frac{1}{z} \cdot \frac{1}{1 - e^z - p}, 0\right) = \lim_{z \to 0} z \cdot \frac{1}{1 - e^z - p} = -\frac{1}{p}.$$

Second,

$$\text{Res}\left(\frac{1}{z} \cdot \frac{1}{1 - e^z - p}, \log(1 - p)\right) = \lim_{z \to z_0} (z - z_0) \cdot \frac{1}{1 - e^z - p - ze^z} = -\frac{1}{(1 - p) \log(1 - p)}.$$ (by L'Hospital’s Rule)

Therefore,

$$I'(p) = \frac{1}{p} + \frac{1}{(1 - p) \log(1 - p)}.$$

Now we take the anti-derivative of both sides:

$$I(p) = \log(p) - \log(-\log(1 - p)) + C,$$

where $C$ is the constant of integration.

To find this constant of integration we note that by the definition of $I(p)$, it satisfies the

$$I(0) = 0$$
boundary condition.

Therefore,

\[
0 = \lim_{p \to 0} I(p) = \lim_{p \to 0} \left( \log(p) - \log(-\log(1 - p)) \right) + C \\
= \lim_{p \to 0} \log\left(\frac{p}{-\log(1 - p)}\right) + C \\
= \log\left(\lim_{p \to 0} \left(\frac{p}{-\log(1 - p)}\right)\right) + C \quad \text{(by continuity of log)} \\
= \log\left(\lim_{p \to 0} (1 - p)\right) + C \quad \text{(by L’Hospital’s Rule)} \\
= C.
\]

Therefore,

\[
\exp\left\{ \frac{n}{2\pi} \int_0^{2\pi} \log(1 - pf(\lambda))d\lambda \right\} = \exp\left\{ n \cdot I(p) \right\}
= \exp\left\{ n \cdot \log(p) - n \cdot \log(-\log(1 - p)) \right\}
= \exp\left\{ \log\left(\frac{p^n}{(\log(1 - p))^n}\right) \right\}
= \frac{p^n}{(\log(1 - p))^n}
\]

is the Szegő leading term for \( \det(I_n - pT_n(f)) \) for thinning parameters \( p < \frac{1-e^{-1}}{2} \). But as we had discussed earlier, we get this exact same Szegő leading term for an even bigger range of thinning parameters, namely, for \( p < 1 - e^{-1} \), which alludes to the existence of an even tighter bound on the eigenvalues of a non-Hermitian Toeplitz matrix from those in the literature (31); or it could just be by chance that we’re getting the correct Szegő leading term for some thinning parameters bigger than \( \frac{1-e^{-1}}{2} \).

It is worth investigating, as this would further strengthen the result we just obtained by having it work for a larger range of thinning parameters \( p \).

Eventually we were able to prove this allusion:

**Theorem 5.9** Let \( (T_n(f))_{n \geq 1} \equiv (t_{j-i})_{1 \leq i, j \leq n} \) be the sequence of Toeplitz matrices that is generated by the symbol (either real or complex-valued)

\[
f(\theta) = \sum_{m=-\infty}^{\infty} t_me^{im\theta}, \quad \theta \in [0, 2\pi].
\]
Let $\lambda_0(n), \ldots, \lambda_{n-1}(n)$ be the eigenvalues of $T_n(f) \equiv T_n$, for some $n \in \mathbb{N}$. Then,

$$|\lambda_i(n)| \leq M_{|f|} \quad \forall n \in \mathbb{N} \forall i \in \{0, \ldots, n - 1\}.$$  

**Proof** Let $\lambda \in \mathbb{C}$ be an eigenvalue of $T_n$ with eigenvector $x = (x_1, \ldots, x_n)^T \in \mathbb{C}^n$. Then,

$$T_n x = \lambda x \iff \sum_{j=1}^{n} t_{j-k} x_j = \lambda x_k, \quad k = 1, \ldots, n.$$  

Let’s now multiply both sides by $\bar{x}_k$, $k \in \{1, \ldots, n\}$:

$$\bar{x}_k \sum_{j=1}^{n} t_{j-k} x_j = \lambda |x_k|^2, \quad k = 1, \ldots, n.$$  

Now take the sum over all $k = 1, \ldots, n$:

$$\sum_{k=1}^{n} \sum_{j=1}^{n} t_{j-k} \bar{x}_k x_j = \lambda \sum_{k=1}^{n} |x_k|^2.$$  

Now apply the Fourier inversion formula:

$$\sum_{k=1}^{n} \sum_{j=1}^{n} t_{j-k} \bar{x}_k x_j = \sum_{k=1}^{n} \sum_{j=1}^{n} \left( \frac{1}{2\pi} \int_{0}^{2\pi} f(\theta) e^{-(j-k)\theta} d\theta \right) \bar{x}_k x_j$$

$$= \frac{1}{2\pi} \int_{0}^{2\pi} f(\theta) \left( \sum_{k=1}^{n} \sum_{j=1}^{n} \bar{x}_k x_j e^{-i(j-k)\theta} \right) d\theta$$

$$= \frac{1}{2\pi} \int_{0}^{2\pi} f(\theta) \left( \sum_{k=1}^{n} \bar{x}_k e^{ik\theta} \sum_{j=1}^{n} x_j e^{-ij\theta} \right) d\theta$$

$$= \frac{1}{2\pi} \int_{0}^{2\pi} f(\theta) \left( \sum_{k=1}^{n} x_k e^{-ik\theta} \sum_{j=1}^{n} x_j e^{ij\theta} \right) d\theta$$

$$= \frac{1}{2\pi} \int_{0}^{2\pi} f(\theta) \left( \sum_{k=1}^{n} x_k e^{-ik\theta} \right)^2 d\theta.$$  

Therefore,

$$|\lambda| \sum_{k=1}^{n} |x_k|^2 \leq M_{|f|} \cdot \frac{1}{2\pi} \int_{0}^{2\pi} \left| \sum_{k=1}^{n} x_k e^{-ik\theta} \right|^2 d\theta$$

$$= M_{|f|} \cdot \frac{1}{2\pi} \int_{0}^{2\pi} \left( \sum_{k=1}^{n} |x_k|^2 + \sum_{k,j=1, \ldots, n, k \neq j} x_k e^{-ik\theta} x_j e^{-ij\theta} \right) d\theta \quad (|a+b|^2 = |a|^2 + |b|^2 + \bar{a}b + ab).$$
\[ M[f] \sum_{k=1}^{n} |x_k|^2 + M[f] \cdot \frac{1}{2\pi} \int_{0}^{2\pi} \sum_{k,j=1,\ldots,n} \bar{x}_k x_j e^{i(k-j)\theta} d\theta \]

\[ = M[f] \sum_{k=1}^{n} |x_k|^2 + M[f] \cdot \sum_{k,j=1,\ldots,n} \bar{x}_k x_j \cdot \frac{1}{2\pi} \int_{0}^{2\pi} e^{i(k-j)\theta} d\theta \]

\[ = M[f] \sum_{k=1}^{n} |x_k|^2 \quad \text{(for} \ k \neq j \text{ the integral is 0).} \]

By the definition of \( x \) being an eigenvector,

\[ \sum_{k=1}^{n} |x_k|^2 \neq 0. \]

Thus, we can divide both LHS and RHS of the above inequality by \( \sum_{k=1}^{n} |x_k|^2 \) and get the result.

We can now apply Theorem 5.9 to the symbol \( f \) of our Carries process to get that both conditions (36) and (37) of Theorem 3.4 hold for thinning parameters \( 0 < p < 1 - e^{-1} \). And so from our earlier computations, we get:

**Theorem 5.10** The Szőgo leading term of \( \tilde{\rho}_n = \det(I_n - pT_n) \) is

\[ \frac{p^n}{(\log(\frac{1}{1-p}))^n}, \]

for thinning parameters \( p < 1 - e^{-1} \).

By comparing Theorem 5.10 with the expression for \( \tilde{\rho}_n \) we had obtained in Section 5.1 (Theorem 5.2), we get that the Szőgo error term of

\[ \tilde{\rho}_n = \det(I_n - pT_n) \]

is exactly equal to

\[ \frac{p^2}{(1-p) \cdot (\log(1-p))^2}. \]

More precisely, by Theorem 3.5,
Theorem 5.11 The Szego error term of $\tilde{\rho}_n = \det(I_n - pT_n)$ is

$$\exp\left\{ \sum_{k=1}^{\infty} \frac{(-p)^k}{k} \sum_{-\infty < j_1, \ldots, j_{k-1} < \infty} \frac{B_{j_1+1}}{(j_1+1)!} \cdots \frac{B_{j_{k-1}+1}}{(j_{k-1}+1)!} t'_{j_1-\ldots-j_{k-1}} \cdot h(j_1, \ldots, j_{k-1}) \right\} = \frac{p^2}{(1-p) \cdot (\log(1-p))^2},$$

where

$$t'_{j_1-\ldots-j_{k-1}} = \frac{B_{-j_1-\ldots-j_{k-1}+1}}{(-j_1 - \ldots - j_{k-1} + 1)!},$$

for thinning parameters $0 < p < \frac{1}{A}$, where $h$ is the function from Theorem 3.5.

Proof This is a direct application of Theorem 3.5 from earlier in the project and the fact that $t_k = \frac{B_{k+1}}{(k+1)!}$, $k \in \mathbb{Z}$, are the Fourier coefficients of the symbol $f$ of the Carries process.

To prove this applicability of Theorem 3.5 on the symbol $f$, we need to prove that $f$ satisfies the condition (41) of the theorem.

Let's begin:

$$\sum_{n=-\infty}^{\infty} n |t_n| = 1 + \sum_{n=1}^{\infty} \frac{|B_{n+1}|}{(n+1)!} = 1 + \sum_{n=2}^{\infty} (n-1) \cdot \frac{|B_n|}{n!} = 1 + \sum_{n=1}^{\infty} (2n-1) \cdot \frac{|B_{2n}|}{(2n)!},$$

where the last equality holds due to the known fact that $B_{2n+1} = 0$ for all $n \in \mathbb{N}$.

Let’s use the ratio test to determine whether the above infinite sum converges.

Set $\sum_{n=1}^{\infty} (2n-1) \cdot \frac{|B_{2n}|}{(2n)!} = \sum_{n=1}^{\infty} a_n$.

Then,

$$\frac{a_{n+1}}{a_n} = \frac{\frac{2n+1}{(2n+2)!} \cdot |B_{2n+2}|}{\frac{2n-1}{(2n)!} \cdot |B_{2n}|} = \frac{(2n)!(2n+1)}{(2n+2)!(2n-1)} \cdot \frac{|B_{2n+2}|}{|B_{2n}|} = \frac{1}{(2n+2)(2n-1)} \cdot \frac{|B_{2n+2}|}{|B_{2n}|}$$
\[
\frac{1}{(2n+2)(2n-1)} \cdot \frac{\zeta(2n+2)}{\zeta(2n)} \cdot \frac{(2n+2)!}{(2n)!(2\pi)^{2n+2}} \quad \text{(by Lemma 5.8)}
\]

\[
\leq \frac{1}{(2\pi)^2} \cdot \frac{2n+1}{2n-1}
\]

(saw earlier in this section that \(\zeta(2n+2) \leq \zeta(2n)\)).

Hence,

\[
\lim_{n \to \infty} \left| \frac{a_{n+1}}{a_n} \right| \leq \frac{1}{(2\pi)^2} < 1,
\]

and so by the ratio test, the above infinite series converges.

Therefore, (41) is indeed satisfied by the symbol \(f\) of the Carries process.

As promised at the end of Section 3.2, we would like to see if we can apply Theorem 3.6 on our \(p\)-thinned carries process to get the respective Szegő error term in terms of integrals as opposed to complicated summations as we just did in Theorem 5.11.

Clearly the main assumption of Theorem 3.6 is satisfied by \(1 - pf\), that is, \(1 - pf\) is of the form

\[
1 - pf(\lambda) = e^{-ik\lambda} \sum_{m=0}^{\infty} t'_m e^{im\lambda}, \quad \lambda \in [0, 2\pi],
\]

and \(t'_m = 0\) for all \(m < 0\), where \(k = 1\).

However, \(1 - pf\) fails the \(t'_0 = 1\) assumption of the Theorem (since \(t'_0 = p\), recall). Thankfully, Theorem 3.6 can hold for all \(a_0 \neq 0\):

**Corollary 5.12** Theorem 3.6 holds even for general \(a_0 \neq 0\).

**Proof** Suppose we have the symbol

\[
f(\lambda) = e^{-ik\lambda} \sum_{m=0}^{\infty} a_m e^{im\lambda}, \quad \lambda \in [0, 2\pi],
\]

with \(a_0 \neq 0\) and \(a_m = 0\) for all \(m < 0\). Suppose \(f\) also satisfies the conditions (28) and (52) of Theorem 3.6. That is,

\[
\sum_{m=0}^{\infty} |a_m| < \infty
\]

and \(\log(f(\lambda)) = \sum_{m=-\infty}^{\infty} h_m e^{im\lambda}\) with

\[
\sum_{m=-\infty}^{\infty} |h_m| < \infty.
\]
Now define
\[ g(\lambda) := \frac{1}{a_0} f(\lambda) = e^{-ik\lambda} \sum_{m=0}^{\infty} a_m' e^{im\lambda}, \quad \lambda \in [0, 2\pi]. \]

Then \( g \) satisfies the \( a_m' = \frac{a_m}{a_0} = 1 \) assumption. It also satisfies the conditions (28) and (52) of Theorem 3.6. Indeed,
\[ \sum_{m=0}^{\infty} |a_m'| = \frac{1}{|a_0|} \sum_{m=0}^{\infty} |a_m| < \infty \]
and \( \log(g(\lambda)) = \sum_{m=-\infty}^{\infty} h_m' e^{im\lambda} \), where, since the (52) assumption is satisfied by \( f \) (and hence the Fourier inversion formula for \( \log(f(\lambda)) \) is invokable), for all \( m \neq 0 \),
\[
\begin{align*}
 h_m &= \frac{1}{2\pi} \int_{0}^{2\pi} \log(f(\lambda)) e^{-im\lambda} d\lambda \\
 &= \frac{1}{2\pi} \int_{0}^{2\pi} \log(f(\lambda)) e^{-im\lambda} d\lambda - \frac{1}{2\pi} \log(a_0) \int_{0}^{2\pi} e^{-im\lambda} d\lambda \quad \text{(since the latter integral is 0 for } m \neq 0) \\
 &= \frac{1}{2\pi} \int_{0}^{2\pi} \log\left(\frac{f(\lambda)}{a_0}\right) e^{-im\lambda} d\lambda \\
 &= \frac{1}{2\pi} \int_{0}^{2\pi} \log(g(e^{i\lambda})) e^{-im\lambda} d\lambda \\
 &= h_m'.
\end{align*}
\]
As for the \( m = 0 \) case, it is easy to see from the above equalities that
\[ h_0' = h_0 - \log(a_0). \]
So since the \( h_m \)'s are absolutely summable, so are the \( h_m' \)'s and hence (52) is also satisfied by \( g \).

Moreover,
\[ \det(T_n(g)) = \det\left(\frac{1}{a_0} T_n(f)\right) = \frac{1}{a_0^n} \det(T_n(f)). \]

Then observe how
\[
\begin{align*}
\frac{\det(T_n(f))}{\exp \left\{ n \frac{2\pi}{2\pi} \int_{0}^{2\pi} \log(f(\lambda)) d\lambda \right\}} &= \frac{\det(T_n(f))}{e^{nh_0}} \\
&= \frac{\det(T_n(f))}{a_0^ne^{n(h_0 - \log(a_0))}} \\
&= \frac{\det(T_n(g))}{e^{nh_0'}} \\
&= \frac{\det(T_n(g))}{\exp \left\{ n \frac{2\pi}{2\pi} \int_{0}^{2\pi} \log(g(\lambda)) d\lambda \right\}} \quad \text{(by Fourier inversion formula).}
\end{align*}
\]
So since $g$ satisfies all conditions of Theorem 3.6 and $h_m = h_m'$ for every $m \neq 0$, we get

$$\lim_{n \to \infty} \det(T_n(f)) \exp\left\{ \frac{n}{2\pi} \int_0^{2\pi} \log(f(\lambda))d\lambda \right\} = \lim_{n \to \infty} \det(T_n(g)) \exp\left\{ \frac{n}{2\pi} \int_0^{2\pi} \log(g(\lambda))d\lambda \right\}
= \exp\left\{ \sum_{m=1}^\infty m h_m h_{-m} \right\}$$
(by Theorem 3.6 applied on $g$),

which is exactly the statement of Theorem 3.6.

Therefore, by Corollary 5.12, the fact that $t_0' \neq 1$ (but $t_0' = p$ instead) does not pose a problem in applying Theorem 3.6 to the symbol $1 - pf$.

As for condition (28), it’s clear that $1 - pf$ satisfies it (that is, that $\sum_{m=0}^\infty |t_m'| < \infty$) since $t_m' = -p \cdot t_{m-1}$ for $m \neq 1$, and $t_1' = 1 - p \cdot t_1$, and we’ve proved long ago that the $t_m$’s are absolutely summable.

We now need to check if $1 - pf$ satisfies condition (52) of Theorem 3.6. It does, and to prove this, we will make use of the Wiener-Levy theorem from complex analysis (e.g., see Chapter VI, Section 5, Theorem 5.2 (i) from [27]):

Since the symbol $1 - pf$ satisfies the (28) condition (that is, it has an absolutely summable Fourier series), then, by the Wiener-Levy theorem, since the principal complex logarithm (that is defined on the principal branch), $\log(z)$, is an analytic function (except along the negative real axis), and $1 - pf(\lambda) = 1 - \frac{p}{1 - e^{i\lambda}}$ for $0 < p < 1 - e^{-1}, \lambda \in [0, 2\pi]$, thought of as a path in the complex plane, has winding number 0 around the origin (and so does not cross the negative real axis), $\log(1 - pf(\lambda)), \lambda \in [0, 2\pi]$, also has an absolutely summable Fourier series, and so condition (52) is indeed satisfied by $1 - pf$ for $0 < p < 1 - e^{-1}$.

We stated above that $1 - pf$ for $0 < p < 1 - e^{-1}$ has winding number 0 around the origin. Let’s prove this.

We would like to use the Argument Principle from complex analysis to prove this. So we first check if the Argument Principle’s conditions are satisfied.

Define

$$F(z) := \sum_{n=-\infty}^{\infty} t_n z^n = \frac{1}{1 - e^z}, \quad z \in \mathbb{C},$$
i.e.,

$$f(\lambda) = F(e^{i\lambda}), \quad \lambda \in [0, 2\pi].$$
Let’s check if $1 - pF$, for $p < 1 - e^{-1}$, satisfies the requirements of the Argument principle.

Firstly, $p < 1 - e^{-1}$, we’ve seen, ensures that $|pF(e^{i\lambda})| < 1$ for every $\lambda \in [0, 2\pi]$, which implies that $1 - pF(e^{i\lambda}) \neq 0$ for every $\lambda \in [0, 2\pi]$, i.e., $1 - pF$ is non-zero on $\partial B_1(0)$.

Secondly, $1 - pF$ has only the poles $w_k = 2\pi ki$, $k \in \mathbb{Z}$, and all of these poles have modulus not equal to 1. Therefore, $1 - pF$ has no poles on $\partial B_1(0)$.

Thirdly, $F$, and therefore $1 - pF$, has only 1 pole inside $B_1(0)$ (namely, $z = 0$), and so $1 - pF$ is meromorphic inside $B_1(0)$.

So $1 - pF$, for $0 < p < 1 - e^{-1}$, does indeed satisfy all three conditions of the Argument Principle, and so we can apply it.

By the Argument Principle, for $0 < p < 1 - e^{-1}$, the winding number of $1 - pf$ around the origin is exactly equal to

$$\frac{1}{2\pi i} \oint_{\partial B_1(0)^+} \frac{(1 - pF(z))'}{1 - pF(z)} \, dz = \frac{1}{2\pi i} \oint_{\partial B_1(0)^+} \frac{-pe^z}{(1 - e^z)(1 - e^z - p)} \, dz$$

$$= \text{Res}\left(\frac{-pe^z}{(1 - e^z)(1 - e^z - p)}, 0\right) + \text{Res}\left(\frac{-pe^z}{(1 - e^z)(1 - e^z - p)}, \log(1 - p)\right),$$

where the last equality holds by the Residue theorem, since the pole of the integrand $\log(1 - p)$ lies inside $B_1(0)$ under the $0 < p < 1 - e^{-1}$ assumption.

Now,

$$\text{Res}\left(\frac{-pe^z}{(1 - e^z)(1 - e^z - p)}, 0\right) = \lim_{z \to 0} z \cdot \frac{-pe^z}{(1 - e^z)(1 - e^z - p)}$$

$$= \lim_{z \to 0} \frac{-pe^z}{-pe^z - pze^z} \quad \text{(by L'Hospital's rule)}$$

$$= \frac{-p}{p} = -1,$$

and

$$\text{Res}\left(\frac{-pe^z}{(1 - e^z)(1 - e^z - p)}, \log(1 - p)\right) = \lim_{z \to \log(1 - p)} (z - \log(1 - p)) \cdot \frac{-pe^z}{(1 - e^z)(1 - e^z - p)}$$

$$= \lim_{z \to \log(1 - p)} \frac{-pe^z(z - \log(1 - p) + 1)}{-e^z(1 - e^z - p) - e^z(1 - e^z)} \quad \text{(by L'Hospital's rule)}$$

$$= \frac{-p(1 - p)}{-p(1 - p)} = 1.$$
And so
\[ \frac{1}{2\pi i} \oint_{\partial B_1(0)^+} \frac{(1 - pF(z)')}{1 - pF(z)} \, dz = -1 + 1 = 0 \]
is indeed the winding number of \( 1 - pf \) for \( 0 < p < 1 - e^{-1} \) around the origin.

The fact that the winding number of \( 1 - pf \) around the origin is 0, together with the fact that \( 1 - pf(\pi) = 1 - \frac{p}{1 - e^{i\pi}} > 0 \) (since \( p < 1 - e^{-1} \)), which means that \( 1 - pf \) does cross the positive real axis, implies that \( 1 - pf \) never crosses the negative real axis, as was stated.

So now that we’ve established that \( 1 - pf \) satisfies all the conditions of Theorem 3.6 (for \( 0 < p < 1 - e^{-1} \)), we obtain the following expression of the Szegő error term for \( \det(I_n - pT_n(f)) \) by Theorem 3.6:

**Theorem 5.13** The Szegő error term of \( \tilde{\rho}_n = \det(I_n - pT_n(f)) \) for \( 0 < p < 1 - e^{-1} \) is

\[ \exp \left\{ \sum_{m=1}^{\infty} mh_m h_{-m} \right\} = \frac{p^2}{(1 - p) \cdot (\log(1 - p))^2}, \]

where

\[ h_m = \frac{1}{2\pi i} \oint_{\partial B_1(0)^+} \frac{1}{z^{m+1}} \log(1 - \frac{p}{1 - e^{i\lambda}}) \, dz \]

Observe how, if we choose \( |\log(1 - p)| < \delta < 1 \), then our integrand \( \frac{1}{z^{m+1}} \log(1 - \frac{p}{1 - e^{i\lambda}}) \) is analytic in the curve \( \partial B_\delta(0)^+ \) since no singularities of the integrand are crossed by this curve (the singularities being \( z = 0 \) (when \( m \geq 1 \)) and \( z = \log(1 - p) + 2\pi ki, k \in \mathbb{Z} \)).

Moreover, the curve \( \partial B_\delta(0)^+ \) contains (in its interior) the same poles of the integrand as \( \partial B_1(0)^+ \).
does (the poles \( z = 0 \) (for \( m \geq 1 \)) and \( z = \log(1 - p) \)).

Hence, we can deform the contour of integration and obtain, for \( |\log(1 - p)| < \delta < 1 \) and \( m \in \mathbb{Z} \setminus 0 \),

\[
    h_m = \frac{1}{2\pi i} \oint_{\partial B_1(0)^+} \frac{1}{z^{m+1}} \log(1 - \frac{p}{1 - e^z}) \, dz
    = \frac{1}{2\pi i} \oint_{\partial B_\delta(0)^+} \frac{1}{z^{m+1}} \log(1 - \frac{p}{1 - e^z}) \, dz.
\]

This then allows us to write

\[
    \sum_{m=1}^{\infty} mh_m h_{-m} = \sum_{m=1}^{\infty} m \oint_{\partial B_1(0)^+} \frac{1}{2\pi i} \frac{1}{z^{m+1}} \log(1 - \frac{p}{1 - e^z}) \, dz \oint_{\partial B_\delta(0)^+} \frac{w^{m-1}}{2\pi i} \log(1 - \frac{p}{1 - e^w}) \, dw
    = \sum_{m=1}^{\infty} m \oint_{\partial B_1(0)^+} \oint_{\partial B_\delta(0)^+} \frac{1}{(2\pi i)^2 z w} \log(1 - \frac{p}{1 - e^z}) \log(1 - \frac{p}{1 - e^w}) (\frac{w}{z})^m \, dw \, dz.
\]

To pull the infinite summation inside the two integrals it suffices to prove uniform convergence (in \((z, w) \in (\partial B_1(0), \partial B_\delta(0)))\) of

\[
    \sum_{m=1}^{\infty} \frac{1}{(2\pi i)^2 z w} \log(1 - \frac{p}{1 - e^z}) \log(1 - \frac{p}{1 - e^w}) (\frac{w}{z})^m.
\]

Let \((z, w) \in (\partial B_1(0), \partial B_\delta(0)), i.e., |z| = 1 and |w| = \delta, \) and let \( m \in \mathbb{N} \).

We claim that there exists a constant \( C_1 > 0 \) (independent of the \( z \) variable) such that

\[
    |\log(1 - \frac{p}{1 - e^z})| < C_1.
\]

Indeed, if this weren’t the case, then we would necessarily have that \( \log(|1 - \frac{p}{1 - e^x}|) = \infty \) for some \( x \in \partial B_1(0) \). This follows from the definition of the complex modulus \(|\cdot|\) and the definition of the logarithm of a complex number. This could only happen, however, if \( |1 - \frac{p}{1 - e^x}| \) were equal to 0 or infinity. Now, if we had that \(|1 - \frac{p}{1 - e^x}| = 0\), then, necessarily we would have \( \frac{p}{1 - e^x} = 1 \); but this is impossible since \( p < 1 - e^{-1} \) and \( x = e^{i\theta} \) (for some \( \theta \in [0, 2\pi] \)) implies that \( |\frac{p}{1 - e^x}| < \frac{1 - e^{-1}}{1 - e^{-1}} = 1 \) - clearly a contradiction. Also, \(|1 - \frac{p}{1 - e^x}|\) being infinity is impossible since this could only hold if \( x = 2\pi ki \), for some \( k \in \mathbb{Z} \), contradicting the fact that \(|x| = 1\). This proves the existence of such a constant \( C_1 \).
Next we claim that there exists a constant $C_2 > 0$ (independent of the $w$ variable) such that

$$| \log(1 - \frac{p}{1 - e^w}) | < C_2.$$ 

Indeed, if this weren’t the case, then, as before, we would necessarily have that $\log(|1 - \frac{p}{1 - e^x}|) = \infty$ for some $x \in \partial B_\delta(0)$. Let $x = \delta e^{i\phi}$ for some $\phi \in [0, 2\pi]$. This, again, could only happen, however, if $|1 - \frac{p}{1 - e^x}|$ were equal to 0 or infinity. Now, $|1 - \frac{p}{1 - e^x}|$ being infinity is impossible since that could only hold if $x = 2\pi ki$, for some $k \in \mathbb{Z}$, contradicting the fact that $0 < |x| = \delta < 1$. Also, $|1 - \frac{p}{1 - e^x}| = 0$ could only happen if

$$\frac{p}{1 - e^x} = 1 \iff e^x = 1 - p \iff e^{\delta e^{i\phi}} = 1 - p \iff \delta e^{i\phi} = \log(1 - p) + 2\pi li$$

for some $l \in \mathbb{Z}$, which could only happen if

$$\delta \cos \phi = \log(1 - p) \text{ and } \delta \sin \phi = 2\pi l.$$

Now, $l = 0$ has to be the case. Indeed, if we instead had $l \neq 0$, then $|\frac{2\pi l}{\delta}| > 1$ (since $0 < \delta < 1$); but, of course, there does not exist such a $\phi \in [0, 2\pi]$ such that $\sin \phi = \frac{2\pi l}{\delta}$ (the sin function is bounded between $-1$ and 1). So in fact it must be that $l = 0$, which implies that $\delta e^{i\phi} = \log(1 - p)$, which then implies that $\delta = |\log(1 - p)|$; contradicting the fact that $\delta > |\log(1 - p)|$. And so this proves that $|1 - \frac{p}{1 - e^x}| = 0$ is also impossible. Thus, there does exist a constant $C_2$ as stated previously.

Therefore,

$$| \frac{1}{(2\pi i)^2 zw} \log(1 - \frac{p}{1 - e^z}) \log(1 - \frac{p}{1 - e^w})| < \frac{1}{4\pi^2 \delta} \cdot C_1 C_2 \cdot m \delta^m \cdot m^2 \cdot \delta < \infty$$

by the ratio test since, if we set $a_m := m \delta^m$, then

$$\lim_{m \to \infty} \frac{a_{m+1}}{a_m} = \delta \cdot \lim_{m \to \infty} \frac{m + 1}{m} = \delta < 1.$$ 

And so by the Weierstrass $M$-test we get the uniform convergence (in $(z, w) \in (\partial B_1(0), \partial B_\delta(0))$) of

$$\sum_{m=1}^{\infty} \frac{1}{(2\pi i)^2 zw} \log(1 - \frac{p}{1 - e^z}) \log(1 - \frac{p}{1 - e^w}) \cdot \delta^m,$$

as desired.
This, by a classic analysis result, justifies the following interchange of infinite summation and integrals:

\[
\sum_{m=1}^{\infty} m h_m h_{-m} = \sum_{m=1}^{\infty} m \int_{\partial B_1(0)^+ \cup \partial B_{\delta}(0)^+} \frac{1}{2\pi i} \frac{1}{z} \log(1 - \frac{p}{1 - e^z}) \log(1 - \frac{p}{1 - e^w}) \frac{w}{z}^m dw dz
\]

Now, notice how for \( x \in \mathbb{C} \) with \(|x| < 1\),

\[
\sum_{m=1}^{\infty} m x^m = x \cdot \frac{d}{dx} \left( \sum_{m=1}^{\infty} x^m \right) \quad \text{(by uniform convergence)}
\]

\[
= x \cdot \frac{d}{dx} \left( \frac{x}{1-x} \right) \quad \text{(since } |x| < 1\text{)}
\]

\[
= x \cdot \frac{(1-x) + x}{(1-x)^2} = \frac{x}{(1-x)^2}.
\]

So, since \(|\frac{w}{z}| < 1\) for \( w \in \partial B_{\delta}(0) \) and \( z \in \partial B_1(0) \) (since \( 0 < \delta < 1\)),

\[
\sum_{m=1}^{\infty} m h_m h_{-m} = \int_{\partial B_1(0)^+ \cup \partial B_{\delta}(0)^+} \frac{1}{2\pi i} \log(1 - \frac{p}{1 - e^z}) \log(1 - \frac{p}{1 - e^w}) \cdot \frac{w}{z}^m dw dz
\]

Let’s now compute the inner integral

\[
\frac{1}{2\pi i} \int_{\partial B_{\delta}(0)^+} \log(1 - \frac{p}{1 - e^w}) \cdot \frac{1}{(w-z)^2} dw.
\]

By our choice of \( \delta \), \( \log(1 - \frac{p}{1 - e^w}) \) is analytic in the contour \( \partial B_{\delta}(0) \), and so we can integrate by parts:

\[
\frac{1}{2\pi i} \int_{\partial B_{\delta}(0)^+} \log(1 - \frac{p}{1 - e^w}) \cdot \frac{1}{(w-z)^2} dw = -\frac{1}{2\pi i} \int_{\partial B_{\delta}(0)^+} \left( -\frac{1}{w-z} \right) \left( -\frac{1}{(1 - e^w)^2} \right) \left( \frac{p}{1 - e^w} \right) dw
\]

\[
= -p \cdot \frac{1}{2\pi i} \int_{\partial B_{\delta}(0)^+} \frac{1}{w-z} \cdot \frac{e^w}{1-e^w} \cdot \frac{1}{1-e^w-p} dw.
\]

To apply the Residue theorem, we first need to find the poles of the integrand. It’s easy to see that these are precisely \( w = z, w = 2\pi ki \ (k \in \mathbb{Z}) \), \( w = \log(1 - p) + 2\pi ki \ (k \in \mathbb{Z}) \). Next we need to find
which of these lie in the interior of $B_δ(0)$. Since $δ < 1$ and $|z| = 1$, the pole $w = z$ does not lie inside $B_δ(0)$. Then, the fact that $|\log(1 - p)| < δ < 1$ implies that $w = 0$ and $w = \log(1 - p)$ are the only poles of the integrand that lie in the interior of $B_δ(0)$. Let’s now compute the residues of these two poles:

\[
\text{Res}_{w=0} \left( \frac{1}{w - z} \cdot \frac{e^w}{1 - e^w} \cdot \frac{1}{1 - e^w - p}, 0 \right) = \lim_{w \to 0} \frac{we^w}{(w - z)(1 - e^w)(1 - e^w - p)} = \lim_{w \to 0} \frac{w(e^w - 1)}{(w - z)(1 - e^w)(1 - e^w - p)} = -\frac{1}{p^z},
\]

where the second-last equality holds by L’Hospital’s rule; and, for $w_0 := \log(1 - p)$,

\[
\text{Res}_{w=w_0} \left( \frac{1}{w - z} \cdot \frac{e^w}{1 - e^w} \cdot \frac{1}{1 - e^w - p}, w_0 \right) = \lim_{w \to w_0} \frac{(w - \log(1 - p))e^w}{(w - z)(1 - e^w)(1 - e^w - p)} = \lim_{w \to w_0} \frac{(w - \log(1 - p))e^w}{(w - z)(1 - e^w)(1 - e^w - p)} = \frac{1}{(1 - p)(\log(1 - p) - z)p}.
\]

So, by the Residue theorem,

\[
\frac{1}{2\pi i} \oint_{\partial B_δ(0)^+} \log(1 - \frac{p}{1 - e^z}) \frac{1}{(w - z)^2} \, dw = -p\text{Res}_{w=0} \left( \frac{1}{w - z} \cdot \frac{e^w}{1 - e^w} \cdot \frac{1}{1 - e^w - p}, 0 \right) - p\text{Res}_{w=w_0} \left( \frac{1}{w - z} \cdot \frac{e^w}{1 - e^w} \cdot \frac{1}{1 - e^w - p}, w_0 \right)
\]

\[
= \frac{1}{z} + \frac{1}{\log(1 - p) - z}.
\]

Therefore,

\[
\sum_{m=1}^{\infty} m h_m h_{-m} = \frac{1}{2\pi i} \oint_{\partial B_1(0)^+} \log(1 - \frac{p}{1 - e^z}) \frac{1}{z} \, dz + \frac{1}{2\pi} \oint_{\partial B_1(0)^+} \log(1 - \frac{p}{1 - e^z}) \frac{1}{\log(1 - p) - z} \, dz.
\]

Let’s compute $I_2(p)$ by first rewriting it as a Riemann integral:

\[
I_2(p) = \frac{1}{2\pi i} \int_{0}^{2\pi} \log(1 - \frac{p}{1 - e^{i\theta}}) \cdot \frac{1}{\log(1 - p) - e^{i\theta}} \cdot ie^{i\theta} \, d\theta
\]

\[
= \frac{1}{2\pi i} \int_{0}^{2\pi} \log(1 - \frac{p}{1 - e^{-\log(1 - p) + e^{i\theta}} + \log(1 - p)}) \cdot \frac{1}{-(\log(1 - p) + e^{i\theta})} \cdot ie^{i\theta} \, d\theta.
\]

94
By setting $z = -\log(1 - p) + e^{i\theta}$ we can rewrite this integral as a contour integral with contour being $\partial B_1(-\log(1 - p))^+$ (the positively-oriented circle of radius 1 centred at $-\log(1 - p)$):

$$I_2(p) = -\frac{1}{2\pi i} \oint_{\partial B_1(-\log(1 - p))^+} \log(1 - \frac{p}{1 - (1 - p)e^z}) \cdot \frac{1}{z} dz.$$ 

Now observe how all the poles of the above integrand that lie in the interior of $B_1(-\log(1 - p))$ are the same as those that lie inside $B_1(0)$. These poles are precisely $z = 0$ and $z = -\log(1 - p)$. Indeed, since $0 < p < 1 - e^{-1}$, the latter pole is strictly between 0 and 1 (and so is in the interior of $B_1(0)$), and it is clear that the pole $z = 0$ is in the interior of $B_1(0)$ as well. Moreover, none of the poles of the integrand are crossed by the curve $\partial B_1(0)$ (as explained before, due to the choice of $p$, $-\log(1 - p)$ is never 0 or 1), and so the integrand is analytic in $\partial B_1(0)$. Therefore, we can deform the contour back to $\partial B_1(0)^+$:

$$I_2(p) = -\frac{1}{2\pi i} \oint_{\partial B_1(0)^+} \log(1 - \frac{p}{1 - (1 - p)e^z}) \cdot \frac{1}{z} dz.$$ 

So we now have

$$\sum_{m=1}^{\infty} m h_m h_{-m} = \frac{1}{2\pi i} \oint_{\partial B_1(0)^+} \frac{1}{z} \left( \log(\frac{1 - p - e^z}{1 - e^z}) - \log(1 - \frac{p}{1 - (1 - p)e^z}) \right)$$

$$= -\log(1 - p) \cdot \frac{1}{2\pi i} \oint_{\partial B_1(0)^+} \frac{1}{z} dz + 2 \cdot \frac{1}{2\pi i} \oint_{\partial B_1(0)^+} \frac{1}{z} \cdot \log(\frac{1 - p - e^z}{1 - e^z}) dz$$

$$= -\log(1 - p) + 2 \cdot \frac{1}{2\pi i} \oint_{\partial B_1(0)^+} \frac{1}{z} \cdot \log(\frac{1 - p - e^z}{1 - e^z}) dz$$

(by Cauchy’s formula).

$$= I(p)$$

(99)

By the Leibniz integral rule,

$$I'(p) = -\frac{1}{2\pi i} \oint_{\partial B_1(0)^+} \frac{1}{z} \cdot \frac{1}{1 - p - e^z} dz.$$ 

It is easy to see that the only poles of the above integrand that are in the interior of $B_1(0)$ are exactly $z = 0$ and $z = \log(1 - p)$. Now,

$$\text{Res}(\frac{1}{z} \cdot \frac{1}{1 - p - e^z}, 0) = \lim_{z \to 0} \frac{1}{1 - p - e^z} = -\frac{1}{p}$$
\[
\text{Res}_z \left( \frac{1}{z} \cdot \frac{1}{1 - p - e^z}, \log(1 - p) \right) = \lim_{z \to \log(1-p)} \frac{z - \log(1 - p)}{z(1 - p - e^z)} \\
= \lim_{z \to \log(1-p)} \frac{1}{1 - p - e^z - ze^z} \quad \text{(by L'Hospital's Rule)} \\
= \frac{1}{(1 - p) \log(1 - p)}.
\]

So by the Residue theorem,
\[
I'(p) = \frac{1}{p} + \frac{1}{(1 - p) \log(1 - p)}.
\]

By integrating both sides with respect to \( p \) and making use of the \( I(0) = 0 \) boundary condition, we get
\[
I(p) = \log(p) + \log(- \log(1 - p)).
\]

Finally,
\[
\sum_{m=1}^\infty mh_mh_{-m} = -\log(1 - p) + 2 \cdot (\log(p) + \log(- \log(1 - p))) \\
= \log \left( \frac{p^2}{(1 - p) \cdot (\log(1 - p)^2)} \right).
\]

Taking the exponential of both sides yields the result. 

There is yet another crafty way of deriving the above error term, but the technical justifications are seemingly trickier. Therefore, we caution the reader that the following "proof" is incomplete and merely a sketch (at least for the time being). However, it is interesting as it makes use of some slick algebraic manipulations and identities/infinite product representations of functions, and possibly provides a different insight into the inner workings of Theorem 3.6. The common "key idea" between the proof we have just given and the following sketch is the fact that there is a "duplication" present in the sense that was shown in (99), which is, essentially, responsible for the squaring of some parts of the \( O(1) \) error term we derived.

**Alternative "proof" of Theorem 5.13** In the beginning of the previous proof we had established, by the Fourier inversion formula, that for \( m \in \mathbb{Z} \),
\[
h_m = \frac{1}{2\pi} \int_0^{2\pi} \log(1 - \frac{p}{1 - e^{i\lambda}}) e^{-im\lambda} d\lambda.
\]
Let’s first compute the above quantity for \( m \leq -1 \):
This is exactly the same as computing, for \( m \geq 1 \), the quantity
\[
 \frac{1}{2\pi} \int_0^{2\pi} \log(1 - \frac{p}{1-e^{i\lambda}})e^{im\lambda}d\lambda =: I(p).
\]
We do the usual trick of taking the derivative with respect to \( p \) inside the integral by the Leibniz rule and working out \( I'(p) \):
\[
 I'(p) = \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{p + e^{i\lambda} - 1} e^{im\lambda}d\lambda = \frac{1}{2\pi i} \oint_{\partial B_1(0)^+} \frac{\frac{z^{m+1}}{p + e^z - 1}}{z^m - p} dz \tag{100}
\]
(converting into a Contour integral).

Since \( m \geq 1 \), the only pole of the integrand that lies in the interior of \( \partial B_1(0)^+ \) is exactly \( z_0 = \log(1 - p) \). Now,
\[
 Res(\frac{\frac{z^{m+1}}{p + e^z - 1}}{z_0}, z_0) = \lim_{z \to z_0} \frac{z^{m+1} - \frac{z^{m-1}(z - z_0)}{p + e^z - 1}}{z^m - p}
 = \lim_{z \to z_0} \frac{mz^{m-1} - (m - 1)z^{m-2}z_0}{e^z}
 = \frac{(\log(1 - p))^{m-1}}{1 - p}.
\]
So, by the Residue theorem,
\[
 I'(p) = \frac{(\log(1 - p))^{m-1}}{1 - p}.
\]
Then, by integrating both sides with respect to \( p \) and making use of the \( I(0) = 0 \) boundary condition, we get for \( m \geq 1 \),
\[
 h_{-m} = -\frac{(\log(1 - p))^{m}}{m}.
\]
Let’s now compute the quantity \( h_m \) for \( m \geq 1 \):

Let’s set
\[
 h_m = \frac{1}{2\pi} \int_0^{2\pi} \log(1 - \frac{p}{1-e^{i\lambda}})e^{-im\lambda}d\lambda =: I(p).
\]
Next do the usual trick of differentiating with respect to \( p \) inside the integral and then converting to a contour integral to obtain
\[
 I'(p) = \frac{1}{2\pi i} \oint_{\partial B_1(0)^+} \frac{\frac{z^{m+1}}{p + e^z - 1}}{z^m - p} dz.
\]
Now, the poles of the above integrand are precisely \( z = 0 \) (a high order pole) and \( z_k = \log(1 - p) + \)
$2\pi ki, k \in \mathbb{Z}$ (simple poles). Only the poles $z = 0$ and $z = z_0$ are inside the interior of the curve $\partial B_1(0)^+$. This means that only the poles $z = z_k$ for $k \in \mathbb{Z} \setminus \{0\}$ (which we know are simple) are in the interior of the curve $\partial B_1(0)^-$. So, to avoid the high order pole and only deal with simple poles when applying the Residue theorem, we make use of the switching-of-the-orientation-of-the-contour trick we performed in Section 5.1 to obtain

$$I'(p) = \frac{-1}{2\pi i} \int_{\partial B_1(0)^-} \frac{1}{z^{m+1}} \cdot \frac{1}{p + e^z - 1} \, dz \quad \text{(by Cauchy's deformation theorem)}$$

$$= - \sum_{k \in \mathbb{Z} \setminus \{0\}} \text{Res}(\frac{1}{z^{m+1}} \cdot \frac{1}{p + e^z - 1}, z_k) \quad \text{(by the Residue theorem).}$$

Now,

$$\text{Res}(\frac{1}{z^{m+1}} \cdot \frac{1}{p + e^z - 1}, z_k) = \lim_{z \to z_k} \frac{z - z_k}{z^{m+1} (p + e^z - 1)}$$

$$= \lim_{z \to z_k} \frac{1}{(m + 1)z^m(p + e^z - 1) + e^z z^{m+1}} \quad \text{(by L'Hospital's rule)}$$

$$= \frac{1}{(1 - p)z_k^{m+1}}.$$

Therefore,

$$I'(p) = \frac{1}{p - 1} \sum_{k \in \mathbb{Z} \setminus \{0\}} \frac{1}{z_k^{m+1}}.$$

We would now like to integrate both sides with respect to $p$. To be able to pull the integral inside the infinite summation it suffices to prove uniform convergence (in $p$) of the above infinite sum.

Notice how, for $k \neq 0$,

$$|z_k| = \sqrt{(\log(1 - p))^2 + (2\pi k)^2} > 2\pi |k| > 1.$$

This then means that, $\forall k \in \mathbb{Z} \setminus \{0\} \ \forall p \in (0, 1 - e^{-1})$,

$$|\frac{1}{p - 1} \frac{1}{z_k^{m+1}}| \leq \frac{1}{1 - p} \frac{1}{|z_k|^2} \quad \text{(since } p < 1, \ |z_k| > 1 \text{ and } m \geq 1)$$

$$< e \cdot \frac{1}{(\log(1 - p))^2 + (2\pi k)^2} \quad \text{(since } p < 1 - e^{-1})$$

$$< \frac{e}{4\pi^2} \cdot \frac{1}{k^2},$$

and we know $\frac{e}{4\pi^2} \cdot \sum_{k \in \mathbb{Z} \setminus \{0\}} \frac{1}{k^2} < \infty$.

And so the desired infinite summation is in fact uniformly convergent in $p$ by the Weierstrass M-test.
This, by a classic analysis result, allows to write

$$I(p) = \int \left( \frac{1}{p-1} \sum_{k \in \mathbb{Z} \setminus \{0\}} \frac{1}{m+1} \right) dp + C$$

$$= \sum_{k \in \mathbb{Z} \setminus \{0\}} \int \frac{1}{(p-1)(\log(1-p) + 2\pi ki)^{m+1}} dp + C,$$

where $C$ is the constant of integration.

Now,

$$\int \frac{1}{(p-1)(\log(1-p) + 2\pi ki)^{m+1}} dp = \int \frac{1}{(\log(1-p) + 2\pi ki)^{m+1}} d(\log(1-p))$$

$$= -\frac{1}{m} \cdot \frac{1}{(\log(1-p) + 2\pi ki)^m}.$$

To find the constant of integration we make use of the $I(0) = 0$ boundary condition, which yields:

$$C = \frac{1}{m} \sum_{k \in \mathbb{Z} \setminus \{0\}} \left( \frac{1}{2\pi ki} \right)^m.$$

Therefore, for $m \geq 1$,

$$h_m = -\frac{1}{m} \cdot \frac{1}{(\log(1-p) + 2\pi ki)^m} + \frac{1}{m} \sum_{k \in \mathbb{Z} \setminus \{0\}} \left( \frac{1}{2\pi ki} \right)^m. \quad (101)$$

We now move onto the more sketchy/incomplete arguments used in deriving the $O(1)$ error term.

Plugging in the expressions obtained in (100) and (101) into $\sum_{m=1}^{\infty} mh_m h_{-m}$, we get

$$\sum_{m=1}^{\infty} mh_m h_{-m} = \sum_{m=1}^{\infty} \frac{1}{m} (\log(1-p))^m \sum_{k \in \mathbb{Z} \setminus \{0\}} \frac{1}{(\log(1-p) + 2\pi ki)^m} - \sum_{m=1}^{\infty} \frac{1}{m} (\log(1-p))^m \sum_{k \in \mathbb{Z} \setminus \{0\}} \frac{1}{(2\pi ki)^m}.$$

Recall how we had mentioned that there was a "duplication" appearing in (99) in the previous proof. The analogous "duplication" step in this "proof" (quotations marks used to emphasise that it is incomplete/sketchy) is in arguing that the infinite summations $I_1(p)$ and $I_2(p)$ from above are equal.

If one were to argue, by justifying the interchange of the $m$-summation with the $k$-summation and using the Taylor expansions around 0 of $\log(1 - \frac{\log(1-p)}{\log(1-p) + 2\pi ki})$ and $\log(1 - \frac{\log(1-p)}{2\pi ki})$, for $k \neq 0$ (since
\[
\left| \frac{\log(1-p)}{\log(1-p) + 2\pi k} \right| < 1 \text{ and } \left| \frac{\log(1-p)}{2\pi k} \right| < 1 \text{ for } 0 < p < 1 - e^{-1}, \]
that

\[
I_1(p) = \log \left( \prod_{k=1}^{\infty} \left( 1 + \left( \frac{\log(1-p)}{2\pi k} \right)^2 \right) \right)
= I_2(p),
\] (102)

then, by making use of the infinite product representation/identity given on line (41) of [25], one can write

\[
\prod_{k=1}^{\infty} \left( 1 + \left( \frac{\log(1-p)}{2\pi k} \right)^2 \right) = \frac{1}{\left( \frac{\log(1-p)}{2\pi} \right)^2} \left| \Gamma \left( \frac{\log(1-p)}{2\pi} e^{\pi i (2(1-1)-1)} \right) \right|^2
= \frac{4\pi^2}{(\log(1-p))^2} \left| \Gamma \left( -\frac{\log(1-p)}{2\pi} i \right) \right|^2
= \frac{4\pi^2}{(\log(1-p))^2} \cdot \frac{-\log(1-p) \cdot \sinh(\pi \cdot -\frac{\log(1-p)}{2\pi})}{\pi}
= -\frac{2}{\log(1-p)} \cdot \sinh(-\frac{\log(1-p)}{2})
= -\frac{2}{\log(1-p)} \cdot \left( e^{-\frac{\log(1-p)}{2}} - e^{\frac{\log(1-p)}{2}} \right)
= -\frac{1}{\log(1-p)} \cdot \left( (1-p)^{-1/2} - (1-p)^{1/2} \right)
= -\frac{p}{\sqrt{1-p} \cdot \log(1-p)}.
\] (application of Euler’s reflection formula)

And so, by the "duplication" from (102),

\[
\exp \left\{ \sum_{m=1}^{\infty} mh_m h_{-m} \right\} = \left( -\frac{p}{\sqrt{1-p} \cdot \log(1-p)} \right)^2
= \frac{p^2}{(1-p) \cdot (\log(1-p)^2)},
\]
as required. “Q.E.D.”

**Remark** We caution the reader once again that this is an incomplete/sketchy proof with many technical details unjustified, hence the use of quotation marks around Q.E.D. in the end.

And so Theorems 5.10, 5.11, and 5.13, together with (35), provide another way of obtaining the result of Theorem 5.2 (but for thinning parameters $0 < p < 1 - e^{-1}$).
6 Conclusion

In conclusion, we have provided all the missing details for the derivation of the determinantal structure of 1-dependent point processes on the integers (cf. [3]) as well as explicit expressions for kernels in the more specific translation-invariance setting. Having done this we were able to decipher the meaning of these kernels as we transition into the more general (having the translation-invariance assumption relaxed) setting and hence obtain various other valid kernels for general 1-dependent point processes other than the one that was given in Theorem 7.1, [3]. We thus gained a deep understanding of the determinantal structure of this class of point processes. All of this was done in Section 4.

By applying these expressions for the kernels we obtained for translation-invariant 1-dependent point processes onto the carries process from [3] with base $b = \infty$, we were able to derive the symbol for one of its kernels (being Toeplitz) from scratch. We then studied this particular point process extensively and found a way to identify it (by a sequence of iid uniform random variables), which provided one method (with the aid of some combinatorics) for finding an asymptotic for the rate of decay of the $n$-gap probability of its $p$-thinned version as $n$ gets very large (done in Section 5.1). This asymptotic was valid for all (non-trivial) thinning parameters $0 < p < 1$.

We then wanted to exploit the determinantal and Toeplitz structure of the aforementioned carries process and its kernel to obtain the same asymptotic. This was a lot more challenging due to the non-Hermitianity of the kernel (as explained many times). Thus we needed to construct quite a few novel results that work for non-Hermitian Toeplitz matrices so as to obtain the Szegő-type asymptotics for large Toeplitz determinants, and hence the asymptotic for large $n$-gaps of the thinned Carries process we studied (done in Sections 3.2 and 5.2). However, due to the weaknesses of Szegő-type theorems when adapted to work for non-Hermitian Toeplitz matrices, our asymptotic worked for thinning parameters $0 < p < 1 - e^{-1}$ only.

With that said, there are many ways this project could be further developed. For one, getting the asymptotic mentioned on the previous paragraph (via the determinantal and Toeplitz structure approach) to work for every thinning parameter $p$ would be quite remarkable due to the fact that one would first need to derive a result stronger than Szegő’s theorem on the asymptotic of large Toeplitz determinants altogether with completely different arguments and constructions, and perhaps, making use of more sophisticated tools from complex analysis (so that the result works for non-real-valued symbols as well). This would be quite innovative and ambitious given the scarce literature on non-Hermitian Toeplitz matrices.

Adding to the previous paragraph, in this project we have reported three different Szegő error term
results: Theorem 3.3 (the original and renowned theorem due to Szégo that requires Hermitianity), Theorem 3.5 (our extension of the aforementioned result to the non-Hermitian case), and Theorem 3.6 (which works for some very particular, not necessarily real-valued, symbols). The proof of the latter is seemingly quite different to the other two and makes use of a technical lemma. All these theorems have slightly different hypotheses and give slightly different formulae for the error term. But they should somehow be trying to have the same hypotheses and trying to have the same error term also (and indeed will simultaneously apply to a great range of examples - by considering examples that are in the intersection of the three theorems’ hypotheses). Coming up with some easy such examples and carefully studying them might be a way in to how to manipulate these types of formulae, and thus, a way into perhaps obtaining a stronger-than-Szégo result on the asymptotics of large Toeplitz determinants for any symbol.

We had briefly touched upon the idea of equivalent kernels for DPPs in the introductory paragraphs of Section 4 where we mentioned [22] which provided all the equivalent symmetric kernels for DPPs. It would be interesting to get analogue results for the non-symmetric case.

Another possible further development could be to further explore the possibility of a Hermitian Toeplitz kernel for point processes from the $Y$ class (this would allow for the renowned Szégo theorems to be applicable). As we saw in Theorem 4.7 the kernel would then necessarily have to be tridiagonal. Moreover, searching for point processes whose correlation function satisfies the many necessary (and rather demanding) conditions for Hermitian Toeplitz kernels (besides the simple Bernoulli process), which we had given at the end of Section 4, seems like an interesting endeavour.

**Acknowledgements**

I would like to thank my supervisors Roger Tribe and Oleg Zaboronski. This project would not have been possible without them. I am particularly grateful for all our insightful discussions - my aim in all our meetings is to absorb as much as possible from their many years of scientific thinking such as what the right questions to ask are and how to go about finding answers. I am also grateful to my supervisors for assigning me this particular project in my first year, as it gave me the opportunity to gain experience and learn a lot of new material, tools and tricks from a wide range of areas (complex analysis, Fourier analysis, linear algebra, combinatorics, to name a few) as well as the opportunity to study such a great class of matrices (Toeplitz matrices, that is) in depth.
References


