

CHAPTER 1

Introduction

1.1. Random polynomials and their zeros

The primary objects of study in this book are *point processes*, which are random variables taking values in the space of discrete subsets of a metric space, where, by a discrete set we mean a countable set with no accumulation points. Precise definitions of relevant notions will be given later. Many physical phenomena can be modeled by random discrete sets. For example, the arrival times of people in a queue, the arrangement of stars in a galaxy, energy levels of heavy nuclei of atoms etc. This calls upon probabilists to find point processes that can be mathematically analysed in some detail, as well as capture various qualitative properties of naturally occurring random point sets.

The single most important such process, known as the **Poisson process** has been widely studied and applied. The Poisson process is characterized by independence of the process when restricted to disjoint subsets of the underlying space. More precisely, for any collection of mutually disjoint measurable subsets of the underlying space, the numbers of points of a Poisson process that fall in these subsets are stochastically independent. The number of points that fall in A has Poisson distribution with a certain mean $\mu(A)$ depending on A . Then, it is easy to see then that μ must be a measure, and it is called the intensity measure of the Poisson process. This assumption of independence is acceptable in some examples, but naturally, not all. For instance if one looks at outbreaks of a rare disease in a province, then knowing that there is a case in a particular location makes it more likely that there are more such cases in a neighbourhood of that location. On the other hand, if one looks at the distribution of like-charged particles confined by an external field (physicists call it a “one component plasma”), then the opposite is true. Knowing that a particular location holds a particle makes it unlikely for there to be any others close to it. These two examples indicate two ways of breaking the independence assumption, inducing more clumping (“positively correlated”) as in the first example or less clumping (“negatively correlated”) as in the second.

A natural question is whether there are probabilistic mechanisms to generate such clumping or anti-clumping behaviour? A simple recipe that gives rise to positively correlated point processes is well-known to statisticians: First sample $X(\cdot)$, a continuous random function on the underlying space that takes values in \mathbb{R}_+ , and then, sample a Poisson process whose intensity measure has density $X(\cdot)$ with respect to a fixed reference measure ν on the underlying space. These kinds of processes are now called Cox processes, and it is clear why they exhibit clumping - more points fall where X is large, and if X is large at one location in space, it is large in a neighbourhood. We shall encounter a particular subclass of Cox processes, known

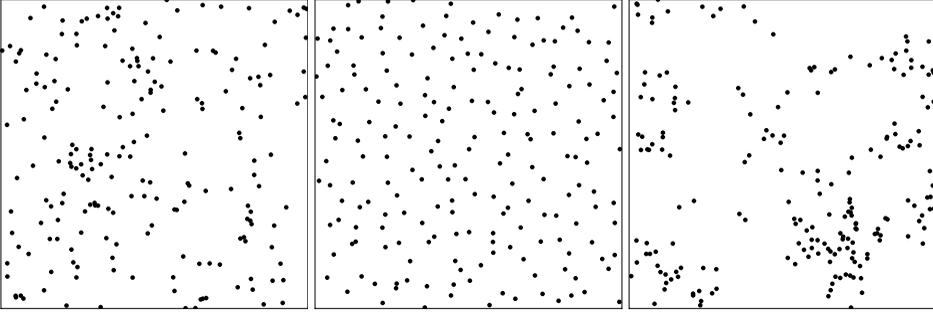


FIGURE 1. Samples of translation invariant point processes in the plane: Poisson (left), determinantal (center) and permanental for $\mathbb{K}(z, w) = \frac{1}{\pi} e^{z\bar{w} - \frac{1}{2}(|z|^2 + |w|^2)}$. Determinantal processes exhibit repulsion, while permanental processes exhibit clumping.

as permanental processes, in Chapter 4, only to compare their properties with determinantal processes, one of two important classes of point processes having negative correlations that we study in this book.

This brings us to the next natural question and that is of central importance to this book. Are there interesting point processes that have *less* clumping than Poisson processes? As we shall see, one natural way of getting such a process without putting in the anti-clumping property by hand, is to extract zero sets of random polynomials or analytic functions, for instance, zeros of random polynomials with stochastically independent coefficients. On the other hand it is also possible to build anti-clumping into the very definition. A particularly nice class of such processes, known as determinantal point processes, is another important object of study in this book.

We study these point processes only in the plane and give some examples on the line, that is, we restrict ourselves to random analytic functions in one variable. One can get point processes in \mathbb{R}^{2n} by considering the joint zeros of n random analytic functions on \mathbb{C}^n , but we do not consider them in this book. Determinantal processes have no dimensional barrier, but it should be admitted that most of the determinantal processes studied have been in one and two dimensions. In contrast to Cox processes that we described earlier, determinantal point processes seem mathematically more interesting to study because, for one, they are apparently not just built out of Poisson processes¹.

Next we turn to the reason why these processes (zeros of random polynomials and determinantal processes) have less clustering of points than Poisson processes. Determinantal processes have this anti-clustering or repulsion built into their definition (chapter 4, definition 4.2.1), and below we give an explanation as to why zeros of random polynomials tend to repel in general. Before going into this, we invite the reader to look at Figure 1. All the three samples shown are portions of certain translation invariant point processes in the plane, with the same average number of points per unit area. Nevertheless, they visibly differ from each other qualitatively, in terms of the clustering they exhibit.

¹“Do not listen to the prophets of doom who preach that every point process will eventually be found out to be a Poisson process in disguise!” - Gian-Carlo Rota.

Now we “explain” the repulsion of points in point processes arising from zeros of random analytic functions (Of course, any point process in the plane is the zero set of a random analytic function, and hence one may wonder if we are making an empty or false claim. However, when we use the term random analytic function, we tacitly mean that we have somehow specified the distribution of coefficients, and that there is a certain amount of independence therein). Consider a polynomial

$$(1.1.1) \quad p(z) = z^n + a_{n-1}z^{n-1} + \dots + a_1z + a_0.$$

We let the coefficients be random variables and see how the (now random) roots of the polynomial are distributed. This is just a matter of change of variables, from coefficients to the roots, and the Jacobian determinant of this transformation is given by the following well known fact (see the book (2) p. 411-412, for instance).

LEMMA 1.1.1. *Let $p(z) = \prod_{k=1}^n (z - z_k)$ have coefficients a_k , $0 \leq k \leq n-1$ as in (1.1.1). Then the transformation $T : \mathbb{C}^n \rightarrow \mathbb{C}^n$ defined by*

$$T(z_1, \dots, z_n) = (a_{n-1}, \dots, a_0),$$

has Jacobian determinant $\prod_{i < j} |z_i - z_j|^2$.

PROOF. Note that we are looking for the real Jacobian determinant, which is equal to

$$\left| \det \left(\frac{\partial T(z_1, \dots, z_n)}{\partial (z_1, \dots, z_n)} \right) \right|^2.$$

To see this in the simplest case of one complex variable, observe that if $f = u + iv : \mathbb{C} \rightarrow \mathbb{C}$, its Jacobian determinant is

$$\det \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix},$$

which is equal to $|f'|^2$, provided f is complex analytic. See Exercise 1.1.2 for the relationship between real and complex Jacobian determinants in general.

Let us write

$$T_n(k) = a_{n-k} = (-1)^k \sum_{1 \leq i_1 < \dots < i_k \leq n} z_{i_1} \dots z_{i_k}.$$

$T_n(k)$ and all its partial derivatives are polynomials in z_j s. Moreover, by the symmetry of $T_n(k)$ in the z_j s, it follows that if $z_i = z_j$ for some $i \neq j$, then the i^{th} and j^{th} columns of $\frac{\partial T(z_1, \dots, z_n)}{\partial (z_1, \dots, z_n)}$ are equal, and hence the determinant vanishes. Therefore, the polynomial $\det \left(\frac{\partial T_n(k)}{\partial z_j} \right)_{1 \leq j, k \leq n}$ is divisible by $\prod_{i < j} (z_i - z_j)$. As the degree of

$\det \left(\frac{\partial T_n(k)}{\partial z_j} \right)_{1 \leq j, k \leq n}$ is equal to $\sum_{k=1}^n (k-1) = \frac{1}{2}n(n-1)$, it must be that

$$\det \left(\frac{\partial T(z_1, \dots, z_n)}{\partial (z_1, \dots, z_n)} \right) = C_n \prod_{i < j} (z_i - z_j).$$

To find the constant C_n , we compute the coefficient of the monomial $\prod z_j^{j-1}$ on both sides. On the right hand side the coefficient is easily seen to be $D_n := (-1)^{n(n-1)/2} C_n$. On the left, we begin by observing that $T_n(k) = -z_n T_{n-1}(k-1) + T_{n-1}(k)$, whence

$$(1.1.2) \quad \frac{\partial T_n(k)}{\partial z_j} = -z_n \frac{\partial T_{n-1}(k-1)}{\partial z_j} + \frac{\partial T_{n-1}(k)}{\partial z_j} - \delta_{jn} T_{n-1}(k-1).$$

The first row in the Jacobian matrix of T has all entries equal to -1 . Further, the entries in the last column (when $j = n$) are just $-T_{n-1}(k-1)$, in particular, independent of z_n . Thus when we expand $\det\left(\frac{\partial T_n(k)}{\partial z_j}\right)$ by the first row, to get z_n^{n-1} we must take the $(1, n)$ entry in the first row and in every other row we must use the first summand in (1.1.2) to get a factor of z_n . Therefore

$$\begin{aligned} D_n &= \text{coefficient of } \prod_{j=1}^n z_j^{j-1} \text{ in } \det\left(\frac{\partial T_n(k)}{\partial z_j}\right)_{1 \leq k, j \leq n} \\ &= (-1)^n \text{coefficient of } \prod_{j=1}^{n-1} z_j^{j-1} \text{ in } \det\left(-\frac{\partial T_{n-1}(k-1)}{\partial z_j}\right)_{\substack{2 \leq k \leq n \\ 1 \leq j \leq n-1}} \\ &= -D_{n-1}. \end{aligned}$$

Thus $C_n = (-1)^n C_{n-1} = (-1)^{n(n+1)/2}$ because $C_1 = -1$. Therefore the real Jacobian determinant of T is $\prod_{i < j} |z_i - z_j|^2$. \square

The following relationship between complex and real Jacobians was used in the proof of the lemma.

EXERCISE 1.1.2. Let $(T_1, \dots, T_n) : \mathbb{C}^n \rightarrow \mathbb{C}^n$ be complex analytic in each argument. Let $A_{ij} = \frac{\partial \operatorname{Re} T_i(\mathbf{z})}{\partial x_j}$ and $B_{ij} = \frac{\partial \operatorname{Re} T_i(\mathbf{z})}{\partial y_j}$ where $z_j = x_j + iy_j$. Then the real Jacobian determinant of $(\operatorname{Re} T_1, \dots, \operatorname{Re} T_n, \operatorname{Im} T_1, \dots, \operatorname{Im} T_n)$ at $(x_1, \dots, x_n, y_1, \dots, y_n)$, is

$$\det \begin{bmatrix} A & B \\ -B & A \end{bmatrix}$$

which is equal to $|\det(A - iB)|^2$, the absolute square of the complex Jacobian determinant.

We may state Lemma 1.1.1 in the reverse direction. But first a remark that will be relevant throughout the book.

REMARK 1.1.3. Let z_k , $1 \leq k \leq n$ be the zeros of a polynomial. Then z_i s do not come with any natural order, and usually we do not care to order them. In that case we identify the set $\{z_k\}$ with the measure $\sum \delta_{z_k}$. However sometimes we might also arrange the zeros as a vector $(z_{\pi_1}, \dots, z_{\pi_n})$ where π is any permutation. If we randomly pick π with equal probability to be one of the $n!$ permutations, we say that the zeros are in **exchangeable random order** or uniform random order. We do this when we want to present joint probability densities of zeros of a random polynomial. Needless to say, the same applies to eigenvalues of matrices or any other (finite) collection of unlabeled points.

Endow the coefficients of a monic polynomial with product Lebesgue measure. The induced measure on the vector of zeros of the polynomial (taken in exchangeable random order) is

$$\left(\prod_{i < j} |z_i - z_j|^2 \right) \prod_{k=1}^n dm(z_k).$$

Here dm denotes the Lebesgue measure on the complex plane.

One can get a probabilistic version of this by choosing the coefficients from Lebesgue measure on a domain in \mathbb{C}^n . Then the roots will be distributed with density proportional to $\prod_{i < j} |z_i - z_j|^2$ for (z_1, \dots, z_n) in a certain symmetric domain of \mathbb{C}^n .

A similar phenomenon occurs in random matrix theory. We just informally state the result here and refer the reader to (6.3.5) in chapter 6 for a precise statement and proof.

FACT 1.1.4. Let $(a_{i,j})_{i,j \leq n}$ be a matrix with complex entries and let z_1, \dots, z_n be the eigenvalues of the matrix. Then it is possible to choose a set of auxiliary variables which we just denote \mathbf{u} (so that \mathbf{u} has $2n(n-1)$ real parameters) so that the transformation $T(z, \mathbf{u}) = (a_{i,j})$ is essentially one-to-one and onto and has Jacobian determinant

$$f(\mathbf{u}) \prod_{i < j} |z_i - z_j|^2$$

for some function f .

REMARK 1.1.5. Unlike in Lemma 1.1.1, to make a change of variables from the entries of the matrix, we needed auxiliary variables in addition to eigenvalues. If we impose product Lebesgue measure on $a_{i,j}$ s, the measure induced on $(z_1, \dots, z_n, \mathbf{u})$ is a product of a measure on the eigenvalues and a measure on \mathbf{u} . However, the measures are infinite and hence it does not quite make sense to talk of "integrating out the auxiliary variables" to obtain

$$(1.1.3) \quad \prod_{i < j} |z_i - z_j|^2 \prod_{k=1}^n dm(z_k)$$

as the "induced measure on the eigenvalues". We can however make sense of similar statements as explained below.

Lemma 1.1.1 and Fact 1.1.4 give a technical intuition as to why zeros of random analytic functions as well as eigenvalues of random matrices often exhibit repulsion. To make genuine probability statements however, we would have to endow the coefficients (or entries) with a probability distribution and use the Jacobian determinant to compute the distribution of zeros (or eigenvalues). In very special cases, one can get an explicit and useful answer, often of the kind

$$(1.1.4) \quad \prod_{i < j} |z_i - z_j|^2 \prod_k e^{-V(z_k)} \prod_{k=1}^n dm(z_k) = \exp \left\{ - \left[\sum_{k=1}^n V(z_k) - \sum_{i \neq j} \log |z_i - z_j| \right] \right\} \prod_{k=1}^n dm(z_k).$$

This density may be regarded as a one component plasma with external potential V and at a particular temperature (see Remark 1.1.6 below). Alternately one may regard it as a "determinantal point process". However it should be pointed out that in most cases, the distribution of zeros (or eigenvalues) is not exactly of this form, and then it is not to be hoped that one can get any explicit and tractable expression of the density. Nevertheless the property of repulsion is generally valid at short distances. Figure 2 shows a determinantal process and a process of zeros of a random analytic function both having the same intensity (the average number of points per unit area).

REMARK 1.1.6. Let us make precise the notion of a one component plasma of n particles with unit charge in the plane with potential V and temperature β^{-1} . This is just the probability density (with respect to Lebesgue measure on \mathbb{C}^n) proportional to

$$\exp \left\{ - \frac{\beta}{2} \left[\sum_{k=1}^n V(z_k) - \sum_{j \neq k} \log |z_j - z_k| \right] \right\} \prod_{k=1}^n dm(z_k).$$

This expression fits the statistical mechanical paradigm, namely it is of the form $\exp\{-\beta H(x)\}$, where H has the interpretation of the energy of a configuration and $1/\beta$ has the physical interpretation of temperature. In our setting we have

$$(1.1.5) \quad H(z_1, \dots, z_n) = \sum_{k=1}^n V(z_k) - \sum_{j \neq k} \log |z_j - z_k|.$$

If we consider n unit negative charges placed in an external potential V at locations z_1, \dots, z_n , then the first term gives the total potential energy due the external field and the second term the energy due to repulsion between the charges. According to Coulomb's law, in three dimensional space the electrical potential due to a point charge is proportional to the inverse distance from the charge. Since we are in two dimensions, the appropriate potential is $\log |z - w|$, which is the Green's function for the Laplacian on \mathbb{R}^2 . However in the density (1.1.4) that (sometimes) comes from random matrices, the temperature parameter is set equal to the particular value $\beta = 2$, which correspond to determinantal processes. Surprisingly, this particular case is much easier to analyse as compared to other values of β !

We study here two kinds of processes (determinantal and zero sets), focusing particularly on specific examples that are invariant under a large group of transformations of the underlying space (translation-invariance in the plane, for instance). Moreover there are certain very special cases of random analytic functions, whose zero sets turn out to be determinantal and we study them in some detail. Finally, apart from these questions of exact distributional calculations, we also present results on large deviations, central limit theorems and also (in a specific case) the stochastic geometry of the zeros. In the rest of the chapter we define some basic notions needed throughout, and give a more detailed overview of the contents of the book.

1.2. Basic notions and definitions

Now we give precise definitions of the basic concepts that will be used throughout the book. Let Λ be a locally compact Polish space (i.e., a topological space that can be topologized by a complete and separable metric). Let μ be a Radon measure on Λ (recall that a Radon measure is a Borel measure which is finite on compact sets). For all examples of interest it suffices to keep the following two cases in mind.

- Λ is an open subset of R^d and μ is the d -dimensional Lebesgue measure restricted to Λ .
- Λ is a finite or countable set and μ assigns unit mass to each element of Λ (the **counting measure** on Λ).

Our point processes (to be defined) will have points in Λ and μ will be a reference measure with respect to which we shall express the probability densities and other similar quantities. So far we informally defined a point process to be a random discrete subset of Λ . However the standard setting in probability theory is to have a sample space that is a complete separable metric space and the set of all discrete subsets of Λ is not such a space, in general. However, a discrete subset of Λ may be identified with the counting measure on the subset (the Borel measure on Λ that assigns unit mass to each element of the subset), and therefore we may define a point process as a random variable taking values in the space $\mathcal{M}(\Lambda)$ of sigma-finite Borel measures on Λ . This latter space is well-known to be a complete separable metric space (see (69), for example).

A **point process** \mathcal{X} on Λ is a random integer-valued positive Radon measure on Λ . If \mathcal{X} almost surely assigns at most measure 1 to singletons, it is a **simple** point process; in this case \mathcal{X} can be identified with a random discrete subset of Λ , and $\mathcal{X}(D)$ represents the number of points of this set that fall in D .

How does one describe the distribution of a point process? Given any $m \geq 1$, any Borel sets D_1, \dots, D_m of Λ , and open intervals $I_1, \dots, I_m \subset [0, \infty)$, we define a subset of $\mathcal{M}(\Lambda)$ consisting of all measures θ such that $\theta(D_k) \in I_k$, for each $k \leq m$. These are called cylinder sets and they generate the sigma field on $\mathcal{M}(\Lambda)$. Therefore, the distribution of a point process \mathcal{X} is determined by the probabilities of cylinder sets, i.e., by the numbers $\mathbf{P}[\mathcal{X}(D_k) = n_k, 1 \leq k \leq m]$ for Borel subsets D_1, \dots, D_m of Λ .

Conversely, one may *define* a point process by consistently assigning probabilities to cylinder sets. Consistency means that

$$\sum_{0 \leq n_m \leq \infty} \mathbf{P}[\mathcal{X}(D_k) = n_k, 1 \leq k \leq m]$$

should be the same as $\mathbf{P}[\mathcal{X}(D_k) = n_k, 1 \leq k \leq m-1]$. (Of course, the usual properties of finite additivity should hold as should the fact that these numbers are between zero and one!). For example the Poisson process may be defined in this manner.

EXAMPLE 1.2.1. For $m \geq 1$ and mutually disjoint Borel subsets $D_k, 1 \leq k \leq m$, of Λ , let

$$p((D_1, n_1), \dots, (D_m, n_m)) = \prod_{k=1}^m e^{-\mu(D_k)} \frac{\mu(D_k)^{n_k}}{n_k!}.$$

The right hand side is to be interpreted as zero if at least one of the D_k s has infinite μ -measure. Then Kolmogorov's existence theorem asserts that there exists a point process \mathcal{X} such that

$$\mathbf{P}[\mathcal{X}(D_k) = n_k, 1 \leq k \leq m] = p((D_1, n_1), \dots, (D_m, n_m)).$$

This is exactly what we informally defined as the Poisson process with intensity measure μ .

Nevertheless, specifying the joint distributions of the counts $\mathcal{X}(D)$, $D \subset \Lambda$ may not be the simplest or the most useful way to define or to think about the distribution of a point process. Alternately, the distribution of a point process can be described by its **joint intensities** (also known as **correlation functions**). We give the definition for simple point processes only, but see remark 1.2.3 for trick to extend the same to general point processes.

DEFINITION 1.2.2. Let \mathcal{X} be a simple point process. The joint intensities of a point process \mathcal{X} w.r.t. μ are functions (if any exist) $\rho_k : \Lambda^k \rightarrow [0, \infty)$ for $k \geq 1$, such that for any family of mutually disjoint subsets D_1, \dots, D_k of Λ ,

$$(1.2.1) \quad \mathbf{E} \left[\prod_{i=1}^k \mathcal{X}(D_i) \right] = \int_{\prod_i D_i} \rho_k(x_1, \dots, x_k) d\mu(x_1) \dots d\mu(x_k).$$

In addition, we shall require that $\rho_k(x_1, \dots, x_k)$ vanish if $x_i = x_j$ for some $i \neq j$.

As joint intensities are used extensively throughout the book, we spend the rest of the section clarifying various points about their definition.

The first intensity is the easiest to understand - we just define the measure $\mu_1(D) := \mathbf{E}[\mathcal{X}(D)]$, we call it the **first intensity measure** of \mathcal{X} . If it happens to be absolutely continuous to the given measure μ , then the Radon Nikodym derivative

ρ_1 is called the first intensity function. From definition 1.2.2 it may appear that the k -point intensity measure μ_k is the first intensity measure of $\mathcal{X}^{\otimes k}$ (the k -fold product measure on Λ^k) and that the k -point intensity function is the Radon Nikodym derivative of μ_k with respect to $\mu^{\otimes k}$, in cases when μ_k is absolutely continuous to $\mu^{\otimes k}$. However, this is incorrect, because (1.2.1) is valid only for pairwise disjoint D_i s. For general subsets of Λ^k , for example, $D_1 \times \dots \times D_k$ with overlapping D_i s, the situation is more complicated as we explain now.

REMARK 1.2.3. Restricting attention to simple point processes, ρ_k is not the intensity measure of \mathcal{X}^k , but that of $\mathcal{X}^{\wedge k}$, the set of ordered k -tuples of distinct points of \mathcal{X} . First note that (1.2.1) by itself does not say anything about ρ_k on the diagonals, that is, for (x_1, \dots, x_k) with $x_i = x_j$ for some $i \neq j$. That is why we added to the definition, the requirement that ρ_k vanish on the diagonal. Then, as we shall explain, equation (1.2.1) implies that for any Borel set $B \subset \Lambda^k$ we have

$$(1.2.2) \quad \mathbf{E}\#(B \cap \mathcal{X}^{\wedge k}) = \int_B \rho_k(x_1, \dots, x_k) d\mu(x_1) \dots d\mu(x_k).$$

When $B = \prod_{i=1}^r D_i^{\otimes k_i}$ for a mutually disjoint family of subsets D_1, \dots, D_r of Λ , and $k = \sum_{i=1}^r k_i$, the left hand side becomes

$$(1.2.3) \quad \mathbf{E} \left[\prod_{i=1}^r \binom{\mathcal{X}(D_i)}{k_i} k_i! \right].$$

For a general point process \mathcal{X} , observe that it can be identified with a simple point process \mathcal{X}^* on $\Lambda \times \{1, 2, 3, \dots\}$ such that $\mathcal{X}^*(D \times \{1, 2, 3, \dots\}) = \mathcal{X}(D)$ for Borel $D \subset \Lambda$. This way, one can deduce many facts about non-simple point processes from those for simple ones.

But why are (1.2.2) and (1.2.3) valid for a simple point process? It suffices to prove the latter. To make the idea transparent, we shall assume that Λ is a countable set and that μ is the counting measure and leave the general case to the reader (consult (55; 56; 70) for details). For simplicity, we restrict to $r = 1$ and $k_1 = 2$ in (1.2.3)) and again leave the general case to the reader. We begin by computing $\mathbf{E}[\mathcal{X}(D)^2]$.

$$\begin{aligned} \mathbf{E}[\mathcal{X}(D)^2] &= \mathbf{E} \left[\left(\sum_{x \in D} \mathcal{X}(\{x\}) \right)^2 \right] \\ &= \mathbf{E} \left[\sum_{x \in D} \mathcal{X}(\{x\}) \right] + \sum_{x \neq y} \mathbf{E}[\mathcal{X}(\{x\})\mathcal{X}(\{y\})] \\ &= \mathbf{E}[\mathcal{X}(D)] + \int_{D \times D} \rho_2(x, y) d\mu(x) d\mu(y). \end{aligned}$$

Here we used two facts. Firstly, $\mathcal{X}(\{x\})$ is 0 or 1 (and 0 for all but finitely many $x \in D$) and secondly, from (1.2.1), for $x \neq y$ we get $\mathbf{E}[\mathcal{X}(\{x\})\mathcal{X}(\{y\})] = \rho_2(x, y)$ while $\rho_2(x, x) = 0$ for all x . Thus

$$(1.2.4) \quad \mathbf{E}[\mathcal{X}(D)(\mathcal{X}(D) - 1)] = \int_{D \times D} \rho_2(x, y) d\mu(x) d\mu(y)$$

as claimed.

Do joint intensities determine the distribution of a point process? The following remark says yes, under certain restrictions.

REMARK 1.2.4. Suppose that $\mathcal{X}(D)$ has exponential tails for all compact $D \subset \Lambda$. In other words, for every compact D , there is a constant $c > 0$ such that $\mathbf{P}[\mathcal{X}(D) > k] \leq e^{-ck}$ for all $k \geq 1$. We claim that under this assumption, the joint intensities (provided they exist) determine the law of \mathcal{X} .

This is because exponential tails for $\mathcal{X}(D)$ for any compact D ensures that for any compact D_1, \dots, D_k , the random vector $(\mathcal{X}(D_1), \dots, \mathcal{X}(D_k))$ has a convergent Laplace transform in a neighbourhood of 0. That is, for some $c > 0$ and any $s_1, \dots, s_k \in (-c, c)$, we have

$$(1.2.5) \quad \mathbf{E}[\exp\{s_1 \mathcal{X}(D_1) + \dots + s_k \mathcal{X}(D_k)\}] < \infty.$$

The Laplace transform determines the law of a random variable and is in turn determined by the moments, whence the conclusion. For these basic facts about moments and Laplace transform consult Billingsley's book (6).

Joint intensities are akin to densities: Assume that \mathcal{X} is simple. Then, the joint intensity functions may be interpreted as follows.

- If Λ is finite and μ is the *counting measure* on Λ , i.e., the measure that assigns unit mass to each element of Λ , then for distinct x_1, \dots, x_k , the quantity $\rho_k(x_1, \dots, x_k)$ is just the probability that $x_1, \dots, x_k \in \mathcal{X}$.
- If Λ is open in \mathbb{R}^d and μ is the Lebesgue measure, then for distinct x_1, \dots, x_k , and $\epsilon > 0$ small enough so that the balls $B_\epsilon(x_j)$ are mutually disjoint, by definition 1.2.2, we get

$$(1.2.6) \quad \int_{\prod_{j=1}^k B_\epsilon(x_j)} \rho_k(y_1, \dots, y_k) \prod_{j=1}^k dm(y_j) = \mathbf{E} \left[\prod_{j=1}^k \mathcal{X}(B_\epsilon(x_j)) \right] \\ = \sum_{\substack{(n_j) \\ n_j \geq 1}} \mathbf{P}(\mathcal{X}(B_\epsilon(x_j)) = n_j, j \leq k) \prod_{j=1}^k n_j.$$

In many examples the last sum is dominated by the term $n_1 = \dots = n_k = 1$. For instance, if we assume that for any compact K , the power series

$$(1.2.7) \quad \sum_{(n_j): j \leq k} \max\{\rho_{n_1+\dots+n_k}(t_1, \dots, t_{n_1+\dots+n_k}) : t_i \in K\} \frac{z_1^{n_1} \dots z_k^{n_k}}{n_1! \dots n_k!}$$

converges for z_i in a neighbourhood of 0, then it follows that for $n_j \geq 1$, by (1.2.2) and (1.2.3) that if $B_\epsilon(x_j) \subset K$ for $j \leq k$, then

$$\mathbf{P}(\mathcal{X}(B_\epsilon(x_j)) = n_j, j \leq k) \leq \mathbf{E} \left[\prod_{j=1}^k \binom{\mathcal{X}(B_\epsilon(x_j))}{n_j} \right] \\ = \frac{1}{n_1! \dots n_k!} \int_{B_\epsilon(x_1)^{n_1} \times \dots \times B_\epsilon(x_k)^{n_k}} \rho_{n_1+\dots+n_k}(y_1, \dots, y_{n_1+\dots+n_k}) \prod_j dm(y_j) \\ \leq \frac{\max\{\rho_{n_1+\dots+n_k}(t_1, \dots, t_{n_1+\dots+n_k}) : t_i \in K\}}{n_1! \dots n_k!} \prod_{j=1}^k m(B_\epsilon)^{n_j}.$$

Under our assumption 1.2.7, it follows that the term $\mathbf{P}(\mathcal{X}(B_\epsilon(x_j)) = 1, j \leq k)$ dominates the sum in (1.2.6). Further, as ρ_k is locally integrable, a.e.

(x_1, \dots, x_k) is a Lebesgue point and for such points we get

$$(1.2.8) \quad \rho_k(x_1, \dots, x_k) = \lim_{\epsilon \rightarrow 0} \frac{\mathbf{P}(\mathcal{X} \text{ has a point in } B_\epsilon(x_j) \text{ for each } j \leq k)}{m(B_\epsilon)^k}.$$

If a continuous version of ρ_k exists, then (1.2.8) holds for every $x_1, \dots, x_k \in \Lambda$. The following exercise demonstrates that for simple point processes with a deterministic finite total number of points, the joint intensities are determined by the top correlation (meaning k -point intensity for the largest k for which it is not identically zero). This fails if the number of points is random or infinite.

EXERCISE 1.2.5. (1) Let X_1, \dots, X_n be exchangeable real valued random variables with joint density $p(x_1, \dots, x_n)$ with respect to Lebesgue measure on \mathbb{R}^n . Let $\mathcal{X} = \sum \delta_{X_i}$ be the point process on \mathbb{R} that assigns unit mass to each X_i . Then show that the joint intensities of \mathcal{X} are given by

$$(1.2.9) \quad \rho_k(x_1, \dots, x_k) = \frac{n!}{(n-k)!} \int_{\mathbb{R}^{n-k}} p(x_1, \dots, x_n) dx_{k+1} \dots dx_n.$$

(2) Construct two simple point process on $\Lambda = \{1, 2, 3\}$ that have the same two-point intensities but not the same one-point intensities.

Moments of linear statistics: Joint intensities will be used extensively throughout the book. Therefore we give yet another way to understand them, this time in terms of **linear statistics**. If \mathcal{X} is a point process on Λ , and $\varphi : \Lambda \rightarrow \mathbb{R}$ is a measurable function, then the random variable

$$(1.2.10) \quad \mathcal{X}(\varphi) := \int_{\Lambda} \varphi d\mathcal{X} = \sum_{\alpha \in \Lambda} \varphi(\alpha) \mathcal{X}(\{\alpha\})$$

is called a linear statistic. If $\varphi = \mathbf{1}_D$ for some $D \subset \Lambda$, then $\mathcal{X}(\varphi)$ is just $\mathcal{X}(D)$.

Knowing the joint distributions of $\mathcal{X}(\varphi)$ for a sufficiently rich class of test functions φ , one can recover the distribution of the point process. For instance, the class of all indicator functions of compact subsets of Λ is rich enough, as explained earlier. Another example is the class of compactly supported continuous functions on Λ . Joint intensities determine the moments of linear statistics corresponding to indicator functions, as made clear in definition 1.2.2 and remark 1.2.4. Now we show how moments of any linear statistics can be expressed in terms of joint intensities. This is done below, but we state it so as to make it into an alternative definition of joint intensities. This is really a more detailed explanation of remark 1.2.3.

Let \mathcal{X} be a point process on Λ and let $C_c(\Lambda)$ be the space of compactly supported continuous functions on Λ . As always, we have a Radon measure μ on Λ .

(1) Define $T_1(\varphi) = \mathbf{E}[\mathcal{X}(\varphi)]$. Then, T_1 is a positive linear functional on $C_c(\Lambda)$. By Riesz's representation theorem, there exists a unique positive regular Borel measure μ_1 such that

$$(1.2.11) \quad T_1(\varphi) = \int \varphi d\mu_1.$$

The measure μ_1 is called the **first intensity measure** of \mathcal{X} .

If it happens that μ_1 is absolutely continuous to μ , then we write $d\mu_1 = \rho_1 d\mu$ and call ρ_1 the first intensity function of \mathcal{X} (with respect to the measure μ). We leave it to the reader to check that this coincides with ρ_1 in definition 1.2.2.

- (2) Define a positive bilinear functional on
- $C_c(\Lambda) \times C_c(\Lambda)$
- by

$$T_2(\varphi, \psi) = \mathbf{E}[\mathcal{X}(\varphi)\mathcal{X}(\psi)]$$

which induces a positive linear functional on $C_c(\Lambda^2)$. Hence, there is a unique positive regular Borel measure $\tilde{\mu}_2$ on Λ^2 such that

$$T_2(\varphi, \psi) = \int_{\Lambda^2} \varphi(x)\psi(y)d\tilde{\mu}_2(x, y).$$

However, in general $\tilde{\mu}_2$ should not be expected to be absolutely continuous to $\mu \otimes \mu$. This is because the random measure $\mathcal{X} \otimes \mathcal{X}$ has atoms on the diagonal $\{(x, x) : x \in \Lambda\}$. In fact,

$$(1.2.12) \quad \mathbf{E}[\mathcal{X}(\varphi)\mathcal{X}(\psi)] = \mathbf{E}[\mathcal{X}(\varphi\psi)] + \mathbf{E}\left[\sum_{(x,y) \in \Lambda^2} \varphi(x)\psi(y)\mathbf{1}_{x \neq y}\mathcal{X}(\{x\})\mathcal{X}(\{y\})\right].$$

Both terms define positive bilinear functionals on $C_c(\Lambda) \times C_c(\Lambda)$ and are represented by two measures $\hat{\mu}_2$ and μ_2 that are supported on the diagonal $D := \{(x, x) : x \in \Lambda\}$ and $\Lambda^2 \setminus D$, respectively. Naturally, $\tilde{\mu}_2 = \mu_2 + \hat{\mu}_2$.

The measure $\hat{\mu}_2$ is singular with respect to $\mu \otimes \mu$ and is in fact the same as the first intensity measure μ_1 , under the natural identification of D with Λ . The second measure μ_2 is called the **two point intensity measure** of \mathcal{X} and if it so happens that μ_2 is absolutely continuous to $\mu \otimes \mu$, then its Radon-Nikodym derivative $\rho_2(x, y)$ is called the two point intensity function. The reader may check that this coincides with the earlier definition. For an example where the second intensity measure is not absolutely continuous to $\mu \otimes \mu$, look at the point process $\mathcal{X} = \delta_a + \delta_{a+1}$ on \mathbb{R} , where a has $N(0, 1)$ distribution.

- (3) Continuing, for any
- $k \geq 1$
- we define a positive multilinear functional on
- $C_c(\Lambda)^k$
- by

$$(1.2.13) \quad T_k(\psi_1, \dots, \psi_k) = \mathbf{E}\left[\prod_{i=1}^k \mathcal{X}(\psi_i)\right]$$

which induces a linear functional on $C_c(\Lambda)^{\otimes k}$ and hence, is represented by a unique positive regular Borel measure $\tilde{\mu}_k$ on Λ^k . We write $\tilde{\mu}_k$ as $\hat{\mu}_k + \mu_k$, where $\hat{\mu}_k$ is supported on the diagonal $D_k = \{(x_1, \dots, x_k) : x_i = x_j \text{ for some } i \neq j\}$ and μ_k is supported on the complement of the diagonal in Λ^k . We call μ_k the **k point intensity measure** and if it happens to be absolutely continuous to $\mu^{\otimes k}$, then we refer to its Radon Nikodym derivative as the k -point intensity function. This agrees with our earlier definition.

1.3. Hints and solutions

Exercise 1.1.2 Consider $\begin{bmatrix} A & B \\ -B & A \end{bmatrix}$. Multiply the second row by i and add to the first to get $\begin{bmatrix} A - iB & B + iA \\ -B & A \end{bmatrix}$. Then multiply the first column by $-i$ and add to the second to get $\begin{bmatrix} A - iB & 0 \\ -B & A + iB \end{bmatrix}$. Since both these operations do not change the determinant, we see that the original matrix has determinant equal to $\det(A - iB)\det(A + iB) = |\det(A - iB)|^2$.

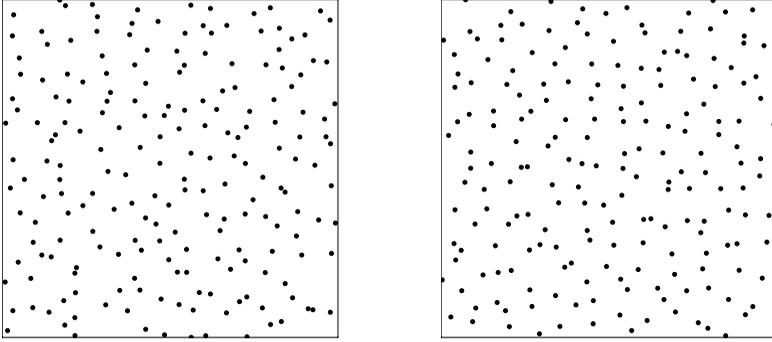


FIGURE 2. Samples of a translation invariant determinantal process (left) and zeros of a Gaussian analytic function. Determinantal processes exhibit repulsion at all distances, and the zeros repel at short distances only. However, the distinction is not evident in the pictures.

Gaussian Analytic Functions

2.1. Complex Gaussian distribution

Throughout this book, we shall encounter complex Gaussian random variables. As conventions vary, we begin by establishing our terminology. By $N(\mu, \sigma^2)$, we mean the distribution of the real-valued random variable with probability density $\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. Here $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ are the mean and variance respectively.

A **standard complex Gaussian** is a complex-valued random variable with probability density $\frac{1}{\pi} e^{-|z|^2}$ w.r.t the Lebesgue measure on the complex plane. Equivalently, one may define it as $X + iY$, where X and Y are i.i.d. $N(0, \frac{1}{2})$ random variables.

Let a_k , $1 \leq k \leq n$ be i.i.d. standard complex Gaussians. Then we say that $\mathbf{a} := (a_1, \dots, a_n)^t$ is a standard complex Gaussian vector. Then if B is a (complex) $m \times n$ matrix, $B\mathbf{a} + \mu$ is said to be an m -dimensional complex Gaussian vector with mean μ (an $m \times 1$ vector) and covariance $\Sigma = BB^*$ (an $m \times m$ matrix). We denote its distribution by $N_C^m(\mu, \Sigma)$.

- EXERCISE 2.1.1. i. Let U be an $n \times n$ unitary matrix, i.e. $UU^* = I_n$, (here U^* is the conjugate transpose of U), and \mathbf{a} an n -dimensional standard complex Gaussian vector. Show that $U\mathbf{a}$ is also an n -dimensional standard complex Gaussian vector.
- ii. Show that the mean and covariance of a complex Gaussian random vector determines its distribution.

REMARK 2.1.2. Although a complex Gaussian can be defined as one having i.i.d. $N(0, \frac{1}{2})$ real and imaginary parts, we advocate thinking of it as a single entity, if not to think of a real Gaussian as merely the real part of a complex Gaussian! Indeed, one encounters the complex Gaussian variable in basic probability courses, for instance in computing the normalizing constant for the density $e^{-x^2/2}$ on the line (by computing the normalizing constant for a complex Gaussian and then taking square roots); and also in generating a random normal on the computer (by generating a complex Gaussian and taking its real part). The complex Gaussian is sometimes easier to work with because it can be represented as a pair of independent random variables in two co-ordinate systems, Cartesian as well as polar (as explained below in more detail). At a higher level, in the theory of random analytic functions and random matrix theory, it is again true that many more exact computations are possible when we use complex Gaussian coefficients (or entries) than when real Gaussians are used.

Here are some other basic properties of complex Gaussian random variables.

- If \mathbf{a} has $N_{\mathbb{C}}^m(\mu, \Sigma)$ distribution, then for every $j, k \leq n$ (not necessarily distinct), we have

$$\mathbf{E}[(a_k - \mu_k)(a_j - \mu_j)] = 0 \text{ and } \mathbf{E}[(a_j - \mu_j)\overline{(a_k - \mu_k)}] = \Sigma_{j,k}.$$

- If a is a standard complex Gaussian, then $|a|^2$ and $\frac{a}{|a|}$ are independent, and have exponential distribution with mean 1 and uniform distribution on the circle $\{z : |z| = 1\}$, respectively.
- Suppose \mathbf{a} and \mathbf{b} are m and n -dimensional random vectors such that

$$\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \sim N_{\mathbb{C}}^{m+n} \left(\begin{bmatrix} \mu \\ \nu \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right),$$

where the mean vector and covariance matrices are partitioned in the obvious way. Then Σ_{11} and Σ_{22} are Hermitian, while $\Sigma_{12}^* = \Sigma_{21}$. Assume that Σ_{11} is non-singular. Then the distribution of \mathbf{a} is $N_{\mathbb{C}}^m(\mu, \Sigma_{11})$ and the conditional distribution of \mathbf{b} given \mathbf{a} is

$$N_{\mathbb{C}}^n(\nu + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{a} - \mu), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}).$$

EXERCISE 2.1.3. Prove this.

- Weak limits of complex Gaussians are complex Gaussians. More precisely,

EXERCISE 2.1.4. If \mathbf{a}_n has $N_{\mathbb{C}}(\mu_n, \Sigma_n)$ distribution and $\mathbf{a}_n \xrightarrow{d} \mathbf{a}$, then $\{\mu_n\}$ and $\{\Sigma_n\}$ must converge, say to μ and Σ , and \mathbf{a} must have $N_{\mathbb{C}}(\mu, \Sigma)$ distribution.

Conversely, if $\{\mu_n\}$ and $\{\Sigma_n\}$ converge to μ and Σ , then \mathbf{a}_n converges weakly to $N_{\mathbb{C}}(\mu, \Sigma)$ distribution.

- The moments of products of complex Gaussians can be computed in terms of the covariance matrix by the Wick or the Feynman diagram formula. First we recall the notion of “permanent” of a matrix, well-known to combinatorists but less ubiquitous in mathematics than its more famous sibling, the determinant.

DEFINITION 2.1.5. For an $n \times n$ matrix M , its permanent, denoted $\text{per}(M)$ is defined by

$$\text{per}(M) = \sum_{\pi \in \mathbb{S}_n} \prod_{k=1}^n M_{k\pi_k}.$$

The sum is over all permutations of $\{1, 2, \dots, n\}$.

REMARK 2.1.6. The analogy with the determinant is clear - the signs of the permutations have been omitted in the definition. But note that this makes a huge difference in that $\text{per}(A^{-1}MA)$ is not in general equal to $\text{per}(M)$. This means that the permanent is a basis-dependent notion and thus has no geometric meaning unlike the determinant. As such, it can be expected to occur only in those contexts where the entries of the matrices themselves are important, as often happens in combinatorics and also in probability.

Now we return to computing moments of products of complex Gaussians. The books of Janson (40) or Simon (79) have such formulas, also in the real Gaussian case.

LEMMA 2.1.7 (Wick formula). Let $(\mathbf{a}, \mathbf{b}) = (a_1, \dots, a_n, b_1, \dots, b_n)^t$ have $N_{\mathbb{C}}(\mathbf{0}, \Sigma)$ distribution, where

$$(2.1.1) \quad \Sigma = \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{bmatrix}.$$

Then,

$$(2.1.2) \quad \mathbf{E} \left[a_1 \cdots a_n \bar{b}_1 \cdots \bar{b}_n \right] = \text{per}(\Sigma_{1,2}).$$

In particular

$$\mathbf{E} \left[|a_1 \cdots a_n|^2 \right] = \text{per}(\Sigma_{1,1}).$$

PROOF. First we prove that

$$\mathbf{E}[a_1 \cdots a_n \bar{b}_1 \cdots \bar{b}_n] = \sum_{\pi} \prod_{j=1}^k \mathbf{E} a_j \bar{b}_{\pi(j)} = \text{per} \left(\mathbf{E} a_j \bar{b}_k \right)_{jk},$$

where the sum is over all permutations $\pi \in S_n$. Both sides are linear in each a_j and \bar{b}_j , and we may assume that the a_j, b_j are complex linear combinations of some finite i.i.d. standard complex Gaussian sequence $\{V_j\}$. The formula is proved by induction on the total number of nonzero coefficients that appear in the expression of the a_j and b_j in terms of the V_j . If the number of nonzero coefficients is more than one for one of a_j or b_j , then we may write that variable as a sum and use induction and linearity. If it is 1 or 0 for all a_j, b_j , then the formula is straightforward to verify; in fact, using independence it suffices to check that $V = V_j$ has $\mathbf{E} V^n \bar{V}^m = n! \mathbf{1}_{\{m=n\}}$. For $n \neq m$ this follows from the fact that V has a rotationally symmetric distribution. Otherwise, $|V|^{2n}$ has the distribution of the n th power of a rate 1 exponential random variable, so its expectation equals $n!$.

The second statement follows immediately from the first, applied to the vector (\mathbf{a}, \mathbf{a}) . \square

- If $a_n, n \geq 1$ are i.i.d. $N_{\mathbb{C}}(0, 1)$, then

$$(2.1.3) \quad \limsup_{n \rightarrow \infty} |a_n|^{\frac{1}{n}} = 1, \quad \text{almost surely.}$$

In fact, equation (2.1.3) is valid for any i.i.d. sequence of complex valued random variables a_n , such that

$$(2.1.4) \quad \mathbf{E}[\max\{\log|a_1|, 0\}] < \infty, \quad \text{provided } \mathbf{P}[a_1 = 0] < 1.$$

We leave the proof as a simple exercise for the reader not already familiar with it. We shall need this fact later, to compute the radii of convergence of random power series with independent coefficients.

2.2. Gaussian analytic functions

Endow the space of analytic functions on a region $\Lambda \subset \mathbb{C}$ with the topology of uniform convergence on compact sets. This makes it a complete separable metric space which is the standard setting for doing probability theory (To see completeness, if $\{f_n\}$ is a Cauchy sequence, then f_n converges uniformly on compact sets to some continuous function f . Then Morera's theorem assures that that f must be analytic because its contour integral vanishes on any closed contour in Λ , since $\int_{\gamma} f = \lim_{n \rightarrow \infty} \int_{\gamma} f_n$ and the latter vanishes for every n by analyticity of f_n).

DEFINITION 2.2.1. Let \mathbf{f} be a random variable on a probability space taking values in the space of analytic functions on a region $\Lambda \subset \mathbb{C}$. We say \mathbf{f} is a Gaussian analytic function (GAF) on Λ if $(\mathbf{f}(z_1), \dots, \mathbf{f}(z_n))$ has a mean zero complex Gaussian distribution for every $n \geq 1$ and every $z_1, \dots, z_n \in \Lambda$.

It is easy to see the following properties of GAFs

- $\{\mathbf{f}^{(k)}\}$ are jointly Gaussian, i.e., the joint distribution of \mathbf{f} and finitely many derivatives of \mathbf{f} at finitely many points,

$$\left\{ \mathbf{f}^{(k)}(z_j) : 0 \leq k \leq n, 1 \leq j \leq m \right\},$$

has a (mean zero) complex Gaussian distribution. (Hint: Weak limits of Gaussians are Gaussians and derivatives are limits of difference coefficients).

- For any $n \geq 1$ and any $z_1, \dots, z_n \in \Lambda$, the random vector $(\mathbf{f}(z_1), \dots, \mathbf{f}(z_n))$ has a complex Gaussian distribution with mean zero and covariance matrix $(K(z_i, z_j))_{i, j \leq n}$. By Exercise 2.1.1 it follows that the covariance kernel K determines all the finite dimensional marginals of \mathbf{f} . Since \mathbf{f} is almost surely continuous, it follows that the distribution of \mathbf{f} is determined by K .
- Analytic extensions of GAFs are GAFs.

EXERCISE 2.2.2. In other words, if \mathbf{f} is a random analytic function on Λ and is Gaussian when restricted to a domain $D \subset \Lambda$, then \mathbf{f} is a GAF on the whole of Λ .

The following lemma gives a general recipe to construct Gaussian analytic functions.

LEMMA 2.2.3. Let ψ_n be holomorphic functions on Λ . Assume that $\sum_n |\psi_n(z)|^2$ converges uniformly on compact sets in Λ . Let a_n be i.i.d. random variables with zero mean and unit variance. Then, almost surely, $\sum_n a_n \psi_n(z)$ converges uniformly on compact subsets of Λ and hence defines a random analytic function.

In particular, if a_n has standard complex Gaussian distribution, then $\mathbf{f}(z) := \sum_n a_n \psi_n(z)$ is a GAF with covariance kernel $K(z, w) = \sum_n \psi_n(z) \overline{\psi_n(w)}$.

If (c_n) is any square summable sequence of complex numbers, and a_n s are i.i.d. with zero mean and unit variance, then $\sum c_n a_n$ converges almost surely, because by Kolmogorov's inequality

$$\mathbf{P} \left[\sup_{k \geq N} \left| \sum_{j=N}^k c_j a_j \right| \geq t \right] \leq \frac{1}{t^2} \sum_{j=N}^{\infty} |c_j|^2$$

$$\rightarrow 0 \quad \text{as } N \rightarrow \infty.$$

Thus, for fixed z , the series of partial sums for $\mathbf{f}(z)$ converge almost surely. However, it is not clear that the series converges for all z simultaneously, even for a single sample point. The idea of the proof is to regard $\sum a_n \psi_n$ as a Hilbert space valued series and prove a version of Kolmogorov's inequality for such series. This part is taken from chapter 3 of Kahane's book (44). That gives convergence in the Hilbert space, and by Cauchy's formulas we may deduce uniform convergence on compacta.

PROOF. Let K be any compact subset of Λ . Regard the sequence $X_n = \sum_{k=1}^n a_k \psi_k$ as taking values in $L^2(K)$ (with respect to Lebesgue measure). Let $\|\cdot\|^2$ denote the

norm in $L^2(K)$. It is easy to check that for any $k < n$ we have

$$(2.2.1) \quad \mathbf{E}[\|X_n\|^2 \mid a_j, j \leq k] = \|X_k\|^2 + \sum_{j=k+1}^n \|\psi_j\|^2.$$

Define the stopping time $\tau = \inf\{n : \|X_n\| > \epsilon\}$. Then,

$$\begin{aligned} \mathbf{E}[\|X_n\|^2] &\geq \sum_{k=1}^n \mathbf{E}[\|X_n\|^2 \mathbf{1}_{\tau=k}] \\ &= \sum_{k=1}^n \mathbf{E}[\mathbf{1}_{\tau=k} \mathbf{E}[\|X_n\|^2 \mid a_j, j \leq k]] \\ &\geq \sum_{k=1}^n \mathbf{E}[\mathbf{1}_{\tau=k} \|X_k\|^2] \quad \text{by (2.2.1)} \\ &\geq \epsilon^2 \mathbf{P}[\tau \leq n]. \end{aligned}$$

Thus

$$(2.2.2) \quad \mathbf{P}\left[\sup_{j \leq n} \|X_j\| \geq \epsilon\right] \leq \frac{1}{\epsilon^2} \sum_{j=1}^n \|\psi_j\|^2.$$

We have just proved Kolmogorov's inequality for Hilbert space valued random variables. Apply this to the sequence $\{X_{N+n} - X_N\}_n$ to get

$$(2.2.3) \quad \mathbf{P}\left[\sup_{m, n \geq N} \|X_m - X_n\| \geq 2\epsilon\right] \leq \mathbf{P}\left[\sup_{n \geq 1} \|X_{N+n} - X_N\| \geq \epsilon\right] \leq \frac{1}{\epsilon^2} \sum_{j=N+1}^{\infty} \|\psi_j\|^2$$

which converges to zero as $N \rightarrow \infty$. Thus

$$\mathbf{P}[\exists N \text{ such that } \forall n, \|X_{N+n} - X_N\| \leq \epsilon] = 1.$$

In other words, almost surely X_n is a Cauchy sequence in $L^2(K)$.

To show uniform convergence on compact subsets, consider any disk $D(z_0, 4R)$ contained in Λ . Since X_n is an analytic function on Λ for each n , Cauchy's formula says

$$(2.2.4) \quad X_n(z) = \frac{1}{2\pi i} \int_{C_r} \frac{X_n(\zeta)}{\zeta - z} d\zeta$$

where $C_r(t) = z_0 + re^{it}$, $0 \leq t \leq 2\pi$ and $|z - z_0| < r$. For any $z \in D(z_0, R)$, average equation (2.2.4) over $r \in (2R, 3R)$ to deduce that

$$\begin{aligned} X_n(z) &= \frac{1}{2\pi i R} \int_{2R}^{3R} \int_0^{2\pi} \frac{X_n(z_0 + re^{i\theta})}{z_0 + re^{i\theta} - z} ie^{i\theta} d\theta r dr \\ &= \frac{1}{2\pi} \int_A X_n(\zeta) \varphi_z(\zeta) dm(\zeta) \end{aligned}$$

where A denotes the annulus around z_0 of radii $2R$ and $3R$ and $\varphi_z(\cdot)$ is defined by the equality. The observation that we shall need is that the collection $\{\varphi_z\}_{z \in D(z_0, R)}$ is uniformly bounded in $L^2(A)$.

We proved that almost surely $\{X_n\}$ is a Cauchy sequence in $L^2(K)$ where $K := D(z_0, 4R)$. Therefore there exists $X \in L^2(K)$ such that $X_n \rightarrow X$ in $L^2(K)$. Therefore the integral above converges to $\frac{1}{2\pi} \int_A X(\zeta) \varphi_z(\zeta) dm(\zeta)$ uniformly over $z \in D(z_0, R)$.

Thus we conclude that $X_n \rightarrow X$ uniformly on compact sets in Λ and that X is an analytic function on Λ .

If a_n s are complex Gaussian, it is clear that X_n is a GAF for each n . Since limits of Gaussians are Gaussians, we see that X is also a GAF. The formula for the covariance $\mathbf{E}[\mathbf{f}(z)\overline{\mathbf{f}(w)}]$ is obvious. \square

2.3. Isometry-invariant zero sets

As explained in Chapter 1, our interest is in the zero set of a random analytic function. Unless one's intention is to model a particular physical phenomenon by a point process, there is one criterion that makes some point processes more interesting than others, namely, *invariance* under a large group of transformations (invariance of a measure means that its distribution does not change under the action of a group, i.e., symmetry). There are three particular two dimensional domains (up to conformal equivalence) on which the group of conformal automorphisms act transitively (There are two others that we do not consider here, the cylinder or the punctured plane, and the two dimensional torus). We introduce these domains now.

- **The Complex Plane \mathbb{C} :** The group of transformations

$$(2.3.1) \quad \varphi_{\lambda, \beta}(z) = \lambda z + \beta, \quad z \in \mathbb{C}$$

where $|\lambda| = 1$ and $\beta \in \mathbb{C}$, is nothing but the Euclidean motion group. These transformations preserve the Euclidean metric $ds^2 = dx^2 + dy^2$ and the Lebesgue measure $dm(z) = dxdy$ on the plane.

- **The Sphere \mathbb{S}^2 :** The group of rotations act transitively on the two dimensional sphere. Moreover the sphere inherits a complex structure from the complex plane by stereographic projection which identifies the sphere with the extended complex plane. In this book we shall always refer to $\mathbb{C} \cup \{\infty\}$ as the sphere. The rotations of the sphere become linear fractional transformations mapping $\mathbb{C} \cup \{\infty\}$ to itself bijectively. That is, they are given by

$$(2.3.2) \quad \varphi_{\alpha, \beta}(z) = \frac{\alpha z + \beta}{-\beta z + \bar{\alpha}}, \quad z \in \mathbb{C} \cup \{\infty\}$$

where $\alpha, \beta \in \mathbb{C}$ and $|\alpha|^2 + |\beta|^2 = 1$. These transformations preserve the spherical metric $ds^2 = \frac{dx^2 + dy^2}{(1 + |z|^2)^2}$ and the spherical measure $\frac{dm(z)}{(1 + |z|^2)^2}$. It is called the spherical metric because it is the push forward of the usual metric (inherited from \mathbb{R}^3) on the sphere onto $\mathbb{C} \cup \{\infty\}$ under the stereographic projection, and the measure is the push forward of the spherical area measure.

EXERCISE 2.3.1. (i) Show that the transformations $\varphi_{\alpha, \beta}$ defined by (2.3.2) preserve the spherical metric and the spherical measure.

(ii) Show that the radius and area of the disk $D(0, r)$ in the spherical metric and spherical measure are $\arctan(r)$ and $\frac{\pi r^2}{1+r^2}$, respectively.

- **The Hyperbolic Plane \mathbb{D} :** The group of transformations

$$(2.3.3) \quad \varphi_{\alpha, \beta}(z) = \frac{\alpha z + \beta}{\beta z + \bar{\alpha}}, \quad z \in \mathbb{D}$$

where $\alpha, \beta \in \mathbb{C}$ and $|\alpha|^2 - |\beta|^2 = 1$, is the group of linear fractional transformations mapping the unit disk $\mathbb{D} = \{z : |z| < 1\}$ to itself bijectively. These

transformations preserve the hyperbolic metric $ds^2 = \frac{dx^2+dy^2}{(1-|z|^2)^2}$ and the hyperbolic area measure $\frac{dm(z)}{(1-|z|^2)^2}$ (this normalization differs from the usual one, with curvature -1 , by a factor of 4, but it makes the analogy with the other two cases more formally similar). This is one of the many models for the hyperbolic geometry of Bolyai, Gauss and Lobachevsky (see (13) or (36) for an introduction to hyperbolic geometry).

- EXERCISE 2.3.2. (i) Show that $\varphi_{\alpha,\beta}$ defined in (2.3.3) preserves the hyperbolic metric and the hyperbolic measure.
- (ii) Show that the radius and area of the disk $D(0,r)$, $r < 1$ in the hyperbolic metric and hyperbolic measure are $\operatorname{arctanh}(r)$ and $\frac{\pi r^2}{1-r^2}$, respectively.

Note that in each case, the group of transformations acts transitively on the corresponding space, i.e., for every z,w in the domain, there is a transformation φ such that $\varphi(z) = w$. This means that in these spaces every point is just like every other point. Now we introduce three families of GAFs whose relation to these symmetric spaces will be made clear in Proposition 2.3.4.

In each case, the domain of the random analytic function can be found using Lemma 2.2.3 or directly from equation (2.1.3).

- **The Complex Plane** \mathbb{C} : Define for $L > 0$,

$$(2.3.4) \quad \mathbf{f}(z) = \sum_{n=0}^{\infty} a_n \frac{\sqrt{L^n}}{\sqrt{n!}} z^n.$$

For every $L > 0$, this is a random analytic function in the entire plane with covariance kernel $\exp\{Lz\bar{w}\}$.

- **The Sphere** \mathbb{S}^2 : Define for $L \in \mathbb{N} = \{1, 2, 3, \dots\}$,

$$(2.3.5) \quad \mathbf{f}(z) = \sum_{n=0}^L a_n \frac{\sqrt{L(L-1)\dots(L-n+1)}}{\sqrt{n!}} z^n.$$

For every $L \in \mathbb{N}$, this is a random analytic function on the complex plane with covariance kernel $(1+z\bar{w})^L$. Since it is a polynomial, we may also think of it as an analytic function on $\mathbb{S}^2 = \mathbb{C} \cup \{\infty\}$ with a pole at ∞ .

- **The Hyperbolic Plane** \mathbb{D} : Define for $L > 0$,

$$(2.3.6) \quad \mathbf{f}(z) = \sum_{n=0}^{\infty} a_n \frac{\sqrt{L(L+1)\dots(L+n-1)}}{\sqrt{n!}} z^n.$$

For every $L > 0$, this is a random analytic function in the unit disk $\mathbb{D} = \{z : |z| < 1\}$ with covariance kernel $(1-z\bar{w})^{-L}$. When L is not an integer, the question of what branch of the fractional power to take, is resolved by the requirement that $K(z,z)$ be positive.

It is natural to ask whether the unit disk is the natural domain for the hyperbolic GAF or if it has an analytic continuation to a larger region. To see that almost surely it does not extend to any larger open set, consider an open disk D intersecting \mathbb{D} but not contained in \mathbb{D} , and let C_D be the event that there exists an analytic continuation of \mathbf{f} to $\mathbb{D} \cup D$. Note that C_D is a tail event, and therefore by Kolmogorov's zero-one law, if it has positive probability then it occurs almost surely. If $\mathbf{P}(C_D) = 1$ for some D , then by the rotational symmetry of complex Gaussian distribution, we see

that $\mathbf{P}(C_{e^{i\theta}D}) = 1$ for any $\theta \in [0, 2\pi]$. Choose finitely many rotations of D so that their union contains the unit circle. With probability 1, \mathbf{f} extends to all of these rotates of D , whence we get an extension of \mathbf{f} to a disk of radius strictly greater than 1. But the radius of convergence is 1 a.s. Therefore $\mathbf{P}(C_D) = 0$ for any D , which establishes our claim.

Another argument is pointed out in the notes. However, these arguments used the rotational invariance of complex Gaussian distribution very strongly. One may adapt an argument given in Billingsley (6), p. 292 to give a more robust proof that works for any symmetric distribution of the coefficients (that is, $-a \stackrel{d}{=} a$).

LEMMA 2.3.3. *Let a_n be i.i.d. random variables with a symmetric distribution in the complex plane. Assume that conditions (2.1.4) hold. Then $\sum_{n=0}^{\infty} a_n \frac{\sqrt{L(L+1)\dots(L+n-1)}}{\sqrt{n!}} z^n$ does not extend analytically to any domain larger than the unit disk.*

PROOF. Assuming (2.1.4), Borel-Cantelli lemmas show that the radius of convergence is at most 1. We need to consider only the case when it is equal to 1. As before, suppose that $\mathbf{P}(C_D) = 1$ for some disk D intersecting the unit disk but not contained in it. Fix k large enough so that an arc of the unit circle of length $\frac{2\pi}{k}$ is contained in D and set

$$(2.3.7) \quad \tilde{a}_n = \begin{cases} a_n & \text{if } n \not\equiv 0 \pmod{k} \\ -a_n & \text{if } n \equiv 0 \pmod{k} \end{cases}.$$

Let

$$(2.3.8) \quad \tilde{\mathbf{f}}(z) = \sum_{n=0}^{\infty} \tilde{a}_n \frac{\sqrt{L(L+1)\dots(L+n-1)}}{\sqrt{n!}} z^n$$

and define \tilde{C}_D in the obvious way. Since $\tilde{\mathbf{f}} \stackrel{d}{=} \mathbf{f}$ it follows that $\mathbf{P}(C_D) = \mathbf{P}(\tilde{C}_D)$. Now suppose both these events have probability one so that the function

$$(2.3.9) \quad \mathbf{g}(z) \stackrel{\text{def}}{=} \mathbf{f}(z) - \tilde{\mathbf{f}}(z) = 2 \sum_{n=0}^{\infty} a_{kn} \frac{\sqrt{L(L+1)\dots(L+kn-1)}}{\sqrt{(kn)!}} z^{kn}$$

may be analytically extended to $\mathbb{D} \cup D$ almost surely. Replacing z by $ze^{2\pi i/k}$ leaves $\mathbf{g}(z)$ unchanged, hence \mathbf{g} can be extended to $\mathbb{D} \cup (\cup_{\ell} D_{\ell})$ where $D_{\ell} = e^{2\pi i \ell/k} D$. In particular, \mathbf{g} can be analytically extended to $(1+\epsilon)\mathbb{D}$ for some $\epsilon > 0$ which is impossible since \mathbf{g} has radius of convergence equal to one. We conclude that C_D has probability zero. \square

Next we prove that the zero sets of the above analytic functions are isometry-invariant.

PROPOSITION 2.3.4. *The zero sets of the GAF \mathbf{f} in equations (2.3.4), (2.3.5) and (2.3.6) are invariant (in distribution) under the transformations defined in equations (2.3.1), (2.3.2) and (2.3.3) respectively. This holds for every allowed value of the parameter L , namely $L > 0$ for the plane and the disk and $L \in \mathbb{N}$ for the sphere.*

PROOF. For definiteness, let us consider the case of the plane. Fix $L > 0$. Then

$$\mathbf{f}(z) = \sum_{n=0}^{\infty} a_n \frac{\sqrt{L^n}}{\sqrt{n!}} z^n,$$

is a centered (mean zero) complex Gaussian process, and as such, its distribution is characterized by its covariance kernel $\exp\{Lz\bar{w}\}$. Now consider the function obtained by translating \mathbf{f} by an isometry in (2.3.1), i.e., fix $|\lambda| = 1$ and $\beta \in \mathbb{C}$, and set

$$\mathbf{g}(z) = \mathbf{f}(\lambda z + \beta).$$

\mathbf{g} is also a centered complex Gaussian process with covariance kernel

$$\begin{aligned} K_{\mathbf{g}}(z, w) &= K_{\mathbf{f}}(\lambda z + \beta, \lambda w + \beta) \\ &= e^{Lz\bar{w} + Lz\lambda\bar{\beta} + L\bar{w}\beta\bar{\lambda} + L|\beta|^2}. \end{aligned}$$

If we set

$$\mathbf{h}(z) = \mathbf{f}(z)e^{Lz\lambda\bar{\beta} + \frac{1}{2}L|\beta|^2},$$

then it is again a centered complex Gaussian process. Its covariance kernel $K_{\mathbf{h}}(z, w)$ is easily checked to be equal to $K_{\mathbf{g}}(z, w)$. This implies that

$$(2.3.10) \quad \mathbf{f}(\lambda z + \beta) \stackrel{d}{=} \mathbf{f}(z)e^{Lz\lambda\bar{\beta} + \frac{1}{2}L|\beta|^2},$$

where the equality in distribution is for the whole processes (functions), not just for a fixed z . Since the exponential function on the right hand side has no zeros, it follows that the zeros of $\mathbf{f}(\lambda z + \beta)$ and the zeros of $\mathbf{f}(z)$ have the same distribution. This proves that the zero set is translationally invariant in distribution.

The proof in the other two cases is exactly the same. If \mathbf{f} is one of the GAFs under consideration, and φ is an isometry of the corresponding domain, then by computing the covariance kernels one can easily prove that

$$(2.3.11) \quad \mathbf{f}(\varphi(\cdot)) \stackrel{d}{=} \mathbf{f}(\cdot)\Delta(\varphi, \cdot),$$

where, $\Delta(\varphi, z)$ is a deterministic nowhere vanishing analytic function of z . That immediately implies the desired invariance of the zero set of \mathbf{f} .

The function $\Delta(\varphi, z)$ is given explicitly by (we are using the expression for φ from equations (2.3.1), (2.3.2) and (2.3.3) respectively).

$$\Delta(\varphi, z) = \begin{cases} e^{Lz\lambda\bar{\beta} + \frac{1}{2}L|\beta|^2} & \text{domain} = \mathbb{C}. \\ \varphi'(z)^{\frac{L}{2}} & \text{domain} = \mathbb{S}^2. \\ \varphi'(z)^{-\frac{L}{2}} & \text{domain} = \mathbb{D}. \end{cases}$$

It is important to notice the following two facts or else the above statements do not make sense.

- (1) In the case of the sphere, by explicit computation we can see that $\varphi'(z)$ is $(-\bar{\beta}z + \bar{\alpha})^{-2}$. Therefore one may raise φ' to half-integer powers and get (single-valued) analytic functions.
- (2) In the case of the disk, again by explicit computation we can see that $\varphi'(z)$ is $(\bar{\beta}z + \bar{\alpha})^{-2}$, but since L is any positive number, to raise φ' to the power $L/2$ we should notice that $\varphi'(z)$ does not vanish for z in the unit disk (because $|\alpha|^2 - |\beta|^2 = 1$). And hence, a holomorphic branch of $\log \varphi'$ may be chosen and thus we may define φ' to the power $L/2$.

□

We shall see later (remark 2.4.5) that the first intensity of zero sets for these canonical GAFs is not zero. Translation invariance implies that the expected number of zeros of the planar and hyperbolic GAFs is almost surely infinite. However, mere translation invariance leaves open the possibility that with positive probability there

are no zeros at all! We rule out this ridiculous possibility by showing that the zero set is in fact ergodic. We briefly recall the definition of ergodicity.

DEFINITION 2.3.5. Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space and let G be a group of measure preserving transformations of Ω to itself, that is, $\mathbf{P} \circ \tau^{-1} = \mathbf{P}$ for every $\tau \in G$. An invariant event is a set $A \in \mathcal{F}$ such that $\tau(A) = A$ for every $\tau \in G$. The action of G is said to be **ergodic** if every invariant set has probability equal to zero or one. In this case we may also say that \mathbf{P} is ergodic under the transformations G .

EXAMPLE 2.3.6. Let \mathbf{P} be the distribution of the zero set of the planar GAF \mathbf{f} . Then by Proposition 2.3.4 we know that the Euclidean motion group acts in a measure preserving manner. The event that \mathbf{f} has infinitely many zeros is an invariant set. Another example is the event that

$$(2.3.12) \quad \lim_{a \rightarrow \infty} \frac{1}{4a^2} \{\text{Number of zeros of } \mathbf{f} \text{ in } [-a, a]^2\} = c$$

where c is a fixed constant. In Proposition 2.3.7 below, we shall see that the action of the translation group (and hence the whole motion group) is ergodic and hence all these invariant events have probability zero or one. We shall see later that the expected number of zeros is positive, which shows that the number of zeros is almost surely infinite. Similarly, the event in (2.3.12) has probability 1 for $c = 1/\pi$ and zero for any other c .

PROPOSITION 2.3.7. *The zero sets of the GAF \mathbf{f} in equations (2.3.4), and (2.3.6) are ergodic under the action of the corresponding isometry groups.*

PROOF. We show the details in the planar case ($\Lambda = \mathbb{C}$) with $L = 1$. The proof is virtually identical in the hyperbolic case. For $\beta \in \mathbb{C}$, let $\mathbf{f}_\beta(z) = \mathbf{f}(z + \beta)e^{-z\bar{\beta} - \frac{1}{2}|\beta|^2}$. We saw in the proof of Proposition 2.3.4 that $\mathbf{f}_\beta \stackrel{d}{=} \mathbf{f}$. We compute

$$\mathbf{E} \left[\mathbf{f}_\beta(z) \bar{\mathbf{f}}(w) \right] = e^{-z\bar{\beta} - \frac{1}{2}|\beta|^2 + z\bar{w} + \beta\bar{w}}.$$

As $\beta \rightarrow \infty$ this goes to 0 uniformly for z, w in any compact set. By Cauchy's formula, the coefficients of the power series expansion of \mathbf{f}_β around 0 are given by

$$\frac{1}{2\pi i} \int_C \frac{\mathbf{f}_\beta(\zeta)}{\zeta^{n+1}} d\zeta,$$

where $C(t) = e^{it}$, $0 \leq t \leq 2\pi$. Therefore, for any n , the first n coefficients in the power series of \mathbf{f} and the first n coefficients in the power series of \mathbf{f}_β become uncorrelated and hence (by joint Gaussianity) independent, as $\beta \rightarrow \infty$.

Now let A be any invariant event. Then we can find an event A_n that depends only on the first n power series coefficients and satisfies $\mathbf{P}[A \Delta A_n] \leq \epsilon$. Then,

$$\left| \mathbf{E} [\mathbf{1}_A(\mathbf{f}) \mathbf{1}_A(\mathbf{f}_\beta)] - \mathbf{E} [\mathbf{1}_{A_n}(\mathbf{f}) \mathbf{1}_{A_n}(\mathbf{f}_\beta)] \right| \leq 2\epsilon.$$

Further, by the asymptotic independence of the coefficients of \mathbf{f} and \mathbf{f}_β , as $\beta \rightarrow \infty$,

$$\mathbf{E} [\mathbf{1}_{A_n}(\mathbf{f}) \mathbf{1}_{A_n}(\mathbf{f}_\beta)] \rightarrow \mathbf{E} [\mathbf{1}_{A_n}(\mathbf{f})] \mathbf{E} [\mathbf{1}_{A_n}(\mathbf{f}_\beta)] = (\mathbf{E} [\mathbf{1}_{A_n}(\mathbf{f})])^2.$$

Thus we get

$$(2.3.13) \quad \limsup_{\beta \rightarrow \infty} \left| \mathbf{E} [\mathbf{1}_A(\mathbf{f}) \mathbf{1}_A(\mathbf{f}_\beta)] - (\mathbf{E} [\mathbf{1}_A(\mathbf{f})])^2 \right| \leq 4\epsilon.$$

This is true for any $\epsilon > 0$ and further, by the invariance of A , we have $\mathbf{1}_A(\mathbf{f})\mathbf{1}_A(\mathbf{f}_\beta) = \mathbf{1}_A(\mathbf{f})$. Therefore

$$(2.3.14) \quad \mathbf{E}[\mathbf{1}_A(\mathbf{f})] = (\mathbf{E}[\mathbf{1}_A(\mathbf{f})])^2$$

showing that the probability of A is zero or one. Since the zeros of \mathbf{f}_β are just translates of the zeros \mathbf{f} , any invariant event that is a function of the zero set must have probability zero or one. In other words, the zero set is ergodic under translations. \square

REMARK 2.3.8. It is natural to ask whether these are the only GAFs on these domains with isometry-invariant zero sets. The answer is essentially yes, but we need to know a little more in general about zeros of GAFs before we can justify that claim.

2.4. Distribution of zeros - The first intensity

In this section, we show how to compute the first intensity or the one-point correlation function (see definition 1.2.2). The setting is that we have a GAF \mathbf{f} and the point process under consideration is the counting measure on $\mathbf{f}^{-1}\{0\}$ with multiplicities where \mathbf{f} is a GAF. The following lemma from (70) shows that in great generality almost surely each zero has multiplicity equal to 1.

LEMMA 2.4.1. *Let \mathbf{f} be a nonzero GAF in a domain Λ . Then \mathbf{f} has no nondeterministic zeros of multiplicity greater than 1. Furthermore, for any fixed complex number $w \neq 0$, $\mathbf{f} - w$ has no zeros of multiplicity greater than 1 (there can be no deterministic zeros for $w \neq 0$ since \mathbf{f} has zero mean).*

PROOF. To prove the first statement in the theorem, we must show that almost surely, there is no z such that $\mathbf{f}(z) = \mathbf{f}'(z) = 0$. Fix $z_0 \in \Lambda$ such that $K(z_0, z_0) \neq 0$. Then $\mathbf{h}(z) := \mathbf{f}(z) - \frac{K(z, z_0)}{K(z_0, z_0)}\mathbf{f}(z_0)$ is a GAF that is independent of $\mathbf{f}(z_0)$. For z such that $K(z, z_0) \neq 0$, we can also write

$$(2.4.1) \quad \frac{\mathbf{f}(z)}{K(z, z_0)} = \frac{\mathbf{h}(z)}{K(z, z_0)} + \frac{\mathbf{f}(z_0)}{K(z_0, z_0)}.$$

Thus if z is a multiple zero of \mathbf{f} , then either $K(z, z_0) = 0$ or z is also a multiple zero of the right hand side of (2.4.1). Since $K(\cdot, z_0)$ is an analytic function, its zeros constitute a deterministic countable set. Therefore, \mathbf{f} has no multiple zeros in that set unless it has a deterministic one. Thus we only need to consider the complement of this set.

Now restrict to the reduced domain Λ' got by removing from Λ all z for which $K(z, z_0) = 0$. Condition on \mathbf{h} . The double zeros of \mathbf{f} in Λ' are those z for which the right hand side of (2.4.1) as well as its derivative vanish. In other words, we must have

$$(2.4.2) \quad \left(\frac{\mathbf{h}(z)}{K(z, z_0)} \right)' = 0 \text{ and } \frac{\mathbf{f}(z_0)}{K(z_0, z_0)} = - \frac{\mathbf{h}(z)}{K(z, z_0)}.$$

Let S be the set of z such that $\left(\frac{\mathbf{h}(z)}{K(z, z_0)} \right)' = 0$. Almost surely, S is a countable set. Then the second event in (2.4.2) occurs if and only if

$$\frac{\mathbf{f}(z_0)}{K(z_0, z_0)} \in \left\{ - \frac{\mathbf{h}(z)}{K(z, z_0)} : z \in S \right\}.$$

The probability of this event is zero because the set on the right is countable and the conditional distribution of $\mathbf{f}(z_0)$ given $\mathbf{h}(\cdot)$ is not degenerate.

The same proof works with \mathbf{f} replaced by $\mathbf{f} - w$ because the mean 0 nature of \mathbf{f} did not really play a role. \square

We give three different ways to find a formula for the first intensity of $n_{\mathbf{f}}$, the counting measure (with multiplicities) on $\mathbf{f}^{-1}\{0\}$, when \mathbf{f} is a Gaussian analytic function. Part of the outcome will be that the first intensity does exist, except at the deterministic zeros (if any) of \mathbf{f} . The expressions that we obtain in the end can be easily seen to be equivalent.

2.4.1. First intensity by Green's formula. The first step is to note that for any analytic function f (not random), we have

$$(2.4.3) \quad dn_f(z) = \frac{1}{2\pi} \Delta \log |f(z)|.$$

Here the Laplacian Δ on the right hand side should be interpreted in the distributional sense. In other words, the meaning of (2.4.3) is just that for any smooth function φ compactly supported in Λ ,

$$(2.4.4) \quad \int_{\Lambda} \varphi(z) dn_f(z) = \int_{\Lambda} \Delta \varphi(z) \frac{1}{2\pi} \log |f(z)| dm(z).$$

To see this, write $f(z) = g(z) \prod_k (z - \alpha_k)^{m_k}$, where α_k are zeros of f (with multiplicities m_k) that are in the support of φ and g is an analytic function with no zeros in the support of φ . Since φ is compactly supported, there are only finitely many α_k . Thus

$$\log |f(z)| = \log |g(z)| + \sum_k m_k \log |z - \alpha_k|.$$

Now, $\Delta \log |g|$ is identically zero on the support of φ because $\log |g|$ is, locally, the real part of an analytic function (of any continuous branch of $\log(g)$). Moreover, $\frac{1}{2\pi} \log |z - \alpha_k| = G(\alpha_k, z)$, the Green's function for the Laplacian in the plane implying that

$$\int_{\Lambda} \Delta \varphi(z) \frac{1}{2\pi} \log |z - \alpha_k| = \varphi(\alpha_k).$$

Therefore (2.4.4) follows.

Now for a random analytic function \mathbf{f} , we get

$$(2.4.5) \quad \mathbf{E} \left[\int_{\Lambda} \varphi(z) dn_{\mathbf{f}}(z) \right] = \mathbf{E} \left[\int_{\Lambda} \Delta \varphi(z) \frac{1}{2\pi} \log |\mathbf{f}(z)| dm(z) \right]$$

$$(2.4.6) \quad = \int_{\Lambda} \Delta \varphi(z) \frac{1}{2\pi} \mathbf{E} [|\log |\mathbf{f}(z)||] dm(z)$$

by Fubini's theorem. To justify applying Fubini's theorem, note that

$$\mathbf{E} \left[\int_{\Lambda} |\Delta \varphi(z)| \frac{1}{2\pi} |\log |\mathbf{f}(z)|| dm(z) \right] = \int_{\Lambda} |\Delta \varphi(z)| \frac{1}{2\pi} \mathbf{E} [|\log |\mathbf{f}(z)||] dm(z).$$

Now for a fixed $z \in \Lambda$, $\mathbf{f}(z)$ is a complex Gaussian with mean zero and variance $K(z, z)$. Therefore, if a denotes a standard complex Gaussian, then

$$\begin{aligned} \mathbf{E} [|\log |\mathbf{f}(z)||] &\leq \mathbf{E} [|\log |a|| + |\log \sqrt{K(z, z)}|] \\ &= \frac{1}{2} \int_0^{\infty} |\log(r)| e^{-r} dr + \frac{1}{2} |\log K(z, z)| \\ &= C + \frac{1}{2} |\log K(z, z)| \end{aligned}$$

for a finite constant C . Observe that $|\log K(z, z)|$ is locally integrable everywhere in z . The only potential problem is at points z_0 for which $K(z_0, z_0) = 0$. But then, in a neighbourhood of z_0 we may write $K(z, z) = |z - z_0|^{2p} L(z, z)$ where $L(z_0, z_0)$ is not zero. Thus $\log K(z, z)$ grows as $\log|z - z_0|$ as $z \rightarrow z_0$, whence it is integrable in a neighbourhood of z_0 . Thus

$$\mathbf{E} \left[\int_{\Lambda} |\Delta \varphi(z)| |\log |\mathbf{f}(z)|| \frac{dm(z)}{2\pi} \right] < \infty.$$

This justifies the use of Fubini's theorem in (2.4.6) and we get

$$(2.4.7) \quad \mathbf{E} \left[\int_{\Lambda} \varphi(z) dn_{\mathbf{f}}(z) \right] = \int_{\Lambda} \varphi(z) \frac{1}{2\pi} \Delta \mathbf{E}[\log |\mathbf{f}(z)|] dm(z).$$

Again using the fact that $\frac{\mathbf{f}(z)}{\sqrt{K(z, z)}}$ is a standard complex Gaussian, we deduce that

$$\begin{aligned} \mathbf{E}[\log |\mathbf{f}(z)|] &= \mathbf{E}[\log |a|] + \frac{1}{2} \log K(z, z) \\ &= -\frac{\gamma}{2} + \log \sqrt{K(z, z)} \end{aligned}$$

where

$$\gamma = -\int_0^{\infty} \log(r) e^{-r} dr.$$

is in fact the negative of Euler's constant, but for our purpose we need only observe that it does not depend on z . Thus by comparing (2.4.7) which is valid for all C_c^2 functions, with (1.2.11) we deduce that the first intensity of $\mathbf{f}^{-1}\{0\}$ with respect to Lebesgue measure is given by

$$(2.4.8) \quad \rho_1(z) = \frac{1}{4\pi} \Delta \log K(z, z).$$

This is sometimes known as the **Edelman-Kostlan** formula. There is no problem with differentiating $\log K(z, z)$ which is real analytic. Exceptions are points where $K(z, z)$ vanish, and at such points the first intensity function does not exist and the first intensity measure has an atom (\mathbf{f} has a deterministic zero).

2.4.2. First intensity by linearization. This is a more probabilistic approach. Let $z \in \Lambda$. We want to estimate the probability that $\mathbf{f}(w) = 0$ for some $w \in D(z, \epsilon)$, up to order ϵ^2 . Expand \mathbf{f} as a power series around z :

$$\mathbf{f}(w) = \mathbf{f}(z) + \mathbf{f}'(z)(w - z) + \mathbf{f}''(z) \frac{(w - z)^2}{2!} + \dots$$

The idea is that up to an event of probability $o(\epsilon^2)$, \mathbf{f} and its linear approximant,

$$\mathbf{g}(w) := \mathbf{f}(z) + (w - z)\mathbf{f}'(z),$$

have the same number of zeros in $D(z, \epsilon)$. Assuming this, it follows from (1.2.8) that

$$\begin{aligned} \rho_1(z) &= \lim_{\epsilon \rightarrow 0} \frac{\mathbf{P}[\mathbf{f} \text{ has a zero in } D(z, \epsilon)]}{\pi \epsilon^2} \\ &= \lim_{\epsilon \rightarrow 0} \frac{\mathbf{P}[\mathbf{g} \text{ has a zero in } D(z, \epsilon)]}{\pi \epsilon^2} \\ &= \lim_{\epsilon \rightarrow 0} \frac{\mathbf{P}\left[\frac{-\mathbf{f}(z)}{\mathbf{f}'(z)} \in D(0, \epsilon)\right]}{\pi \epsilon^2} \\ &= \text{Probability density of } \frac{-\mathbf{f}(z)}{\mathbf{f}'(z)} \text{ at } 0. \end{aligned}$$

If a, b are complex-valued random variables then, by an elementary change of variables, we see that the density of a/b at 0 is equal to $\chi_a(0)\mathbf{E}[|b|^2 \mid a=0]$, where χ_a is the density of a at 0 (assuming the density a and the second moment of b given $a=0$ do exist).

When \mathbf{f} is Gaussian, $(\mathbf{f}(z), \mathbf{f}'(z))$ is jointly complex Gaussian with mean zero and covariance

$$\begin{bmatrix} K(z, z) & \frac{\partial}{\partial \bar{z}} K(z, z) \\ \frac{\partial}{\partial z} K(z, z) & \frac{\partial}{\partial z} \frac{\partial}{\partial \bar{z}} K(z, z) \end{bmatrix}.$$

Here we use the standard notation

$$\frac{\partial}{\partial z} = \frac{1}{2} \left(\frac{\partial}{\partial x} - i \frac{\partial}{\partial y} \right) \text{ and } \frac{\partial}{\partial \bar{z}} = \frac{1}{2} \left(\frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right).$$

The density of $\mathbf{f}(z)$ at 0 is $\frac{1}{\pi K(z, z)}$. Moreover, $\mathbf{f}'(z) \big|_{\mathbf{f}(z)=0}$ has

$$N_{\mathbb{C}} \left(0, \frac{\partial}{\partial z} \frac{\partial}{\partial \bar{z}} K(z, z) - \frac{1}{K(z, z)} \left(\frac{\partial}{\partial \bar{z}} K(z, z) \right) \left(\frac{\partial}{\partial z} K(z, z) \right) \right)$$

distribution. Thus we can write the first intensity as

$$\rho_1(z) = \frac{\frac{\partial}{\partial z} \frac{\partial}{\partial \bar{z}} K(z, z) - \frac{1}{K(z, z)} \frac{\partial}{\partial \bar{z}} K(z, z) \frac{\partial}{\partial z} K(z, z)}{\pi K(z, z)}.$$

This is equivalent to the Edelman-Kostlan formula (2.4.8) as can be seen by differentiating $\log K(z, z)$ (since $\Delta = 4 \frac{\partial}{\partial z} \frac{\partial}{\partial \bar{z}}$).

Now we justify replacing \mathbf{f} by its linearization \mathbf{g} . Without loss of generality, we can assume that $z=0$ and expand \mathbf{f} as a power series. The following lemma is from Peres and Virág (70).

LEMMA 2.4.2. *Let $\mathbf{f}(z) = a_0 + a_1 z + \dots$ be a GAF. Assume that a_0 is not constant. Let A_ϵ denote the event that the number of zeros of \mathbf{f} in the disk $D(0, \epsilon)$ differs from the number of zeros of $\mathbf{g}(z) := a_0 + a_1 z$ in the same disk. Then for any $\delta > 0$, there exists $c > 0$ so that for all $\epsilon > 0$ we have*

$$\mathbf{P}[A_\epsilon] \leq c \epsilon^{3-2\delta}.$$

PROOF. By Rouché's theorem, if $|\mathbf{g}| > |\mathbf{f} - \mathbf{g}|$ on $\partial D(0, \epsilon)$, then \mathbf{f} and \mathbf{g} have the same number of zeros in $D(0, \epsilon)$.

We bound the maximum of $|\mathbf{f} - \mathbf{g}|$ by Lemma 2.4.4. For this we observe that for small enough ϵ ,

$$(2.4.9) \quad \max_{|z| < 2\epsilon} \mathbf{E}[|\mathbf{f}(z) - \mathbf{g}(z)|^2] \leq C \epsilon^4$$

since $\mathbf{f} - \mathbf{g}$ has a double root at 0. Thus Lemma 2.4.4 gives a constant γ such that

$$(2.4.10) \quad \mathbf{P} \left[\max\{|\mathbf{f}(z) - \mathbf{g}(z)| : z \in D(0, \epsilon)\} > \epsilon^{2-\delta} \right] < c_0 e^{-\gamma \epsilon^{-2\delta}}.$$

Now let Θ be the annulus $\partial D(0, |a_1| \epsilon) + D(0, \epsilon^{2-\delta})$ (the Minkowski sum of the two sets), and consider the following events:

$$D_0 = \{|a_0| < 2\epsilon^{1-\delta}\},$$

$$E = \{|a_1| < \epsilon^{-\delta}\},$$

$$F = \{\min\{|\mathbf{g}(z)| : z \in \partial D(0, \epsilon)\} < \epsilon^{2-\delta}\} = \{-a_0 \in \Theta\}.$$

Note that $\mathbf{P}[E^c] \leq c_2 \epsilon^3$ and that $E \cap F \subset D_0$. Given D_0 , the distribution of a_0 (recall our assumption that a_0 is not a constant) is approximately uniform on $D(0, 2\epsilon^{1-\delta})$ (in particular, its conditional density is $O(\epsilon^{2\delta-2})$). Since $\mathbf{P}[E]$ tends to one as $\epsilon \rightarrow 0$, this implies that

$$\mathbf{P}[F] \leq \mathbf{P}[F \cap E | D_0] \mathbf{P}[D_0] + \mathbf{P}[E^c] \leq c_4 \epsilon c_5 \epsilon^{2-2\delta} + c_2 \epsilon^3 \leq c_6 \epsilon^{3-2\delta}.$$

In the first term, the factor of ϵ comes from the area of Θ (as a fraction of the area of D_0) and the factor of $\epsilon^{2-2\delta}$ from the probability of D_0 . Together with (2.4.10), this gives the desired result. \square

REMARK 2.4.3. In the proof we used Lemma 2.4.4 to bound the maximum modulus of a Gaussian analytic function on a disk. In the literature there are deep and powerful theorems about the maximum of a general Gaussian process which we could have used instead. For instance, Borell's isoperimetric inequality (see Pollard (71); the inequality was also shown independently by Tsirelson-Ibragimov-Sudakov (88)) implies that for any collection of mean-zero (real) Gaussian variables with maximal standard deviation σ , the maximum M of the collection satisfies

$$(2.4.11) \quad \mathbf{P}[M > \text{median}(M) + b\sigma] \leq \mathbf{P}[\chi > b],$$

where χ is standard normal. We could have arrived at (2.4.10) by an application of (2.4.11) separately to the real and imaginary parts of $\frac{\mathbf{f}(z) - \mathbf{g}(z)}{z^2}$ (note that the median is just a finite quantity). However we preferred to use Lemma 2.4.4 as it is elementary and also exhibits some new tools for working with Gaussian analytic functions. One idea in the proof below comes from the paper of Nazarov, Sodin and Volberg (61), see Lemma 2.1 therein.

LEMMA 2.4.4. *Let \mathbf{f} be a Gaussian analytic function in a neighbourhood of the unit disk with covariance kernel K . Then for $r < \frac{1}{2}$, we have*

$$(2.4.12) \quad \mathbf{P} \left[\max_{|z| < r} |\mathbf{f}(z)| > t \right] \leq 2e^{-t^2/8\sigma_{2r}^2}$$

where $\sigma_{2r}^2 = \max\{K(z, z) : |z| \leq 2r\}$.

PROOF. Let $\gamma(t) = 2re^{it}$, $0 \leq t \leq 2\pi$. By Cauchy's integral formula, for $|z| < r$,

$$\begin{aligned} |\mathbf{f}(z)| &\leq \int_0^{2\pi} \frac{|\mathbf{f}(\gamma(t))|}{|z - \gamma(t)|} |\gamma'(t)| \frac{dt}{2\pi} \\ &\leq 2\sigma \int_0^{2\pi} |\hat{\mathbf{f}}(2re^{it})| \frac{dt}{2\pi} \end{aligned}$$

where $\hat{\mathbf{f}}(z) = \mathbf{f}(z)/\sqrt{K(z,z)}$ and we have written just σ for σ_{2r} .

$$\begin{aligned} \mathbf{P} \left[\max_{|z|<r} |\mathbf{f}(z)| > t \right] &\leq \mathbf{P} \left[\int_0^{2\pi} |\hat{\mathbf{f}}(2re^{it})| \frac{dt}{2\pi} > \frac{t}{2\sigma} \right] \\ &\leq e^{-t^2/8\sigma^2} \mathbf{E} \left[\exp \left\{ \frac{1}{2} \left(\int_0^{2\pi} |\hat{\mathbf{f}}(2re^{it})| \frac{dt}{2\pi} \right)^2 \right\} \right] \\ &\leq e^{-t^2/8\sigma^2} \mathbf{E} \left[\exp \left\{ \frac{1}{2} \int_0^{2\pi} |\hat{\mathbf{f}}(2re^{it})|^2 \frac{dt}{2\pi} \right\} \right] \end{aligned}$$

by Cauchy-Schwarz inequality. Now use the convexity of the exponential function to get

$$\mathbf{P} \left[\max_{|z|<r} |\mathbf{f}(z)| > t \right] \leq e^{-t^2/8\sigma^2} \mathbf{E} \left[\int_0^{2\pi} \exp \left\{ \frac{1}{2} |\hat{\mathbf{f}}(2re^{it})|^2 \right\} \frac{dt}{2\pi} \right].$$

Since $|\hat{\mathbf{f}}(w)|^2$ has exponential distribution with mean 1 for any w , the expectation of $\exp\{\frac{1}{2}|\hat{\mathbf{f}}(2re^{it})|^2\}$ is 2. Thus we arrive at

$$\mathbf{P} \left[\max_{|z|<r} |\mathbf{f}(z)| > t \right] \leq 2e^{-t^2/8\sigma^2}.$$

□

2.4.3. First intensity by integral geometry. This is a geometric approach to get the first intensity. We shall sketch the idea briefly. Interested readers are recommended to read the beautiful paper (23) for more along these lines.

Let \mathbf{f} be a GAF with covariance kernel K . Since K is Hermitian and positive definite, we can write $K(z,w) = \sum \psi_n(z)\bar{\psi}_n(w)$, where ψ_n are analytic functions on some domain in the plane. Then we see that $\mathbf{f}(z) = \sum a_n \psi_n(z)$, where a_n are i.i.d. standard complex Gaussians. (What we just said may be seen as a converse to Lemma 2.2.3).

First suppose that $\mathbf{f}(z) = \sum_{n=1}^N a_n \psi_n(z)$, where $N < \infty$. In the end let $N \rightarrow \infty$ to get the general case. This is possible by Rouché's theorem, for if the series $\mathbf{f}_N(z) = \sum_{n=1}^N a_n \psi_n(z)$ converges uniformly on compact sets to $\mathbf{f}(z) = \sum_{n=1}^{\infty} a_n \psi_n(z)$, then for any compact set, the number of zeros of \mathbf{f} and \mathbf{f}_N are equal, with high probability, for large N .

When N is finite, setting $\psi(z) = (\psi_1(z), \dots, \psi_N(z))$, we may write

$$\mathbf{f}(z) = \langle \psi(z), (\bar{a}_1, \dots, \bar{a}_N) \rangle$$

where $\langle \cdot, \cdot \rangle$ is the standard inner product in \mathbb{C}^N . As z varies over Λ , $\psi(z)$ defines a complex curve in \mathbb{C}^N . Also $(\bar{a}_1, \dots, \bar{a}_N)$ has a spherically invariant distribution. Thus asking for the number of zeros of \mathbf{f} is equivalent to the following.

Choose a point uniformly at random on the unit sphere $\{(z_1, \dots, z_N) : \sum |z_k|^2 = 1\}$ in \mathbb{C}^N and ask for the number of times (counted with multiplicities) the hyper plane orthogonal to the chosen point intersects the fixed curve ψ .

Turning the problem around, fix z and let w vary over $D(z, \epsilon)$. Then the hyperplane orthogonal to $\psi(w)$ sweeps out a certain portion of the unit sphere. The expected number of zeroes of \mathbf{f} in $D(z, \epsilon)$ is precisely the area of the region swept out (again counting multiplicities).

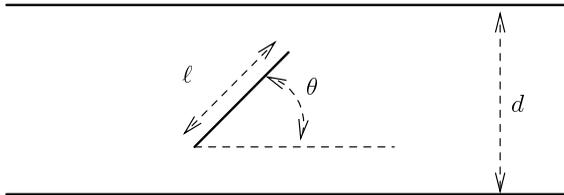


FIGURE 1. The Buffon needle problem.

Now as w varies over $D(z, \epsilon)$, $\psi(w)$ varies over a disk of radius approximately $\|\psi'(z)\|\epsilon$ on the image of the curve ψ . However what matters to us is the projection of this disk orthogonal to the radial vector $\psi(z)$, and this projection has area

$$\left(\|\psi'(z)\|^2 - \frac{|\psi'(z) \cdot \psi(z)|^2}{\|\psi(z)\|^2} \right) \pi \epsilon^2.$$

However this disk is located at a distance $\|\psi(z)\|$ from the origin.

When a particle P moves a distance δ on a geodesic of the sphere of radius r , the hyper-plane P^\perp orthogonal to P , rotates by an angle of $\frac{\delta}{r}$. When $\delta = \pi$, the entire sphere is swept out by P^\perp exactly once. Putting these together, we find that the probability of having a zero in $D(z, \epsilon)$ is

$$\frac{\left(\|\psi'(z)\|^2 - \frac{|\psi'(z) \cdot \psi(z)|^2}{\|\psi(z)\|^2} \right)}{\pi \|\psi(z)\|^2} \epsilon^2,$$

and this gives $p(z)$. Since $K(z, w) = \psi(z) \cdot \psi(w)$, this is the same as what we got earlier.

REMARK 2.4.5. As a simple application of (2.4.8), one can check that the zero sets of the GAFs described in equations (2.3.4), (2.3.5) and (2.3.6) have first intensities equal to $\frac{L}{\pi}$, w.r.t, the Lebesgue measure $dm(z)$ on the plane, the Spherical measure $\frac{dm(z)}{(1+|z|^2)^2}$ on $\mathbb{S}^2 = \mathbb{C} \cup \{\infty\}$ and the Hyperbolic measure $\frac{dm(z)}{(1-|z|^2)^2}$ on the unit disk \mathbb{D} , respectively.

EXERCISE 2.4.6. Follow the steps outlined below to give a geometric solution to the classical Buffon needle problem: Consider a family of parallel lines in the plane with adjacent lines separated by a distance d . Drop a needle of length ℓ “at random” on the plane. What is the probability that the needle crosses one of the lines? See figure 1.

- i. Show that the probability of a crossing is $c\ell$ for some constant c , provided that $\ell < d$.
- ii. If a circle of circumference ℓ is dropped on the plane, deduce that the expected number of intersections of the circle with the family of parallel lines is again $c\ell$. Use this to compute c .

2.5. Intensity of zeros determines the GAF

In this section we present the result of Sodin (80) that two GAFs on Λ having the same intensity $\rho_1(z)dm(z)$ are essentially equal. In particular we get the remarkable conclusion that the distribution of the zero set $\mathbf{f}^{-1}\{0\}$ is completely determined by its first intensity! We first prove a standard fact from complex analysis that will be used in the proof of Theorem 2.5.2.

LEMMA 2.5.1. *Let $K(z, w)$ be analytic in z and anti-analytic in w (i.e., analytic in \bar{w}) for $(z, w) \in \Lambda \times \Lambda$. If $K(z, z) = 0 \forall z \in \Lambda$, then $K(z, w) = 0 \forall z, w \in \Lambda$.*

PROOF. It is enough to prove that K vanishes in a neighbourhood of (z, z) for every $z \in \Lambda$. Without loss of generality take $z = 0$. Then around $(0, 0)$ we can expand K as $K(z, w) = \sum_{m, n \geq 1} a_{m, n} z^m \bar{w}^n$. Then $K(z, z) = \sum_{m, n \geq 1} a_{m, n} z^m \bar{z}^n$. Let $z = x + iy$. Note that

$$\frac{\partial^{m+n}}{\partial z^m \partial \bar{z}^n} z^k \bar{z}^\ell \Big|_{z=0} = \delta_{(m, n), (k, \ell)} m! n!.$$

Returning to $K(z, z) = \sum_{k, \ell \geq 1} a_{k, \ell} z^k \bar{z}^\ell$, this gives (since we have assumed that $K(z, z)$ is identically zero)

$$\begin{aligned} 0 &= \frac{\partial^{m+n}}{\partial z^m \partial \bar{z}^n} K(z, z) \Big|_{z=0} \\ &= m! n! a_{m, n}. \end{aligned}$$

Thus $K(z, w)$ vanishes identically in $\Lambda \times \Lambda$. \square

Sodin (80) discovered the following result and related it to Calabi's rigidity theorem in complex geometry.

THEOREM 2.5.2 (Calabi's rigidity). *Suppose \mathbf{f} and \mathbf{g} are two GAFs in a region Λ such that the first intensity measures of $\mathbf{f}^{-1}\{0\}$ and $\mathbf{g}^{-1}\{0\}$ are equal. Then there exists a nonrandom analytic function φ on Λ that does not vanish anywhere, such that $\mathbf{f} \stackrel{d}{=} \varphi \mathbf{g}$. In particular $\mathbf{f}^{-1}\{0\} \stackrel{d}{=} \mathbf{g}^{-1}\{0\}$.*

PROOF. For a $z \in \Omega$, we have $K_{\mathbf{f}}(z, z) = 0$ if and only if z is almost surely a zero of \mathbf{f} (and the corresponding orders of vanishing of $K_{\mathbf{f}}$ and \mathbf{f} at z match). Since \mathbf{f} and \mathbf{g} are assumed to have the same first intensity of zeros, the set of deterministic zeros of \mathbf{f} must coincide and have the same order of vanishing for \mathbf{f} and \mathbf{g} . By omitting all such zeros from Λ , we assume that $K_{\mathbf{f}}(z, z)$ and $K_{\mathbf{g}}(z, z)$ do not vanish anywhere in Λ . It suffices to prove the theorem for this reduced domain, for suppose that $\mathbf{f} = \varphi \mathbf{g}$ on $\Lambda - D$ where D is the discrete set that we have omitted, where φ is a non-vanishing analytic function on $\Lambda - D$. Since at each point z of D , the functions \mathbf{f} and \mathbf{g} vanish to the same order, we see that φ is bounded in a neighbourhood of z and thus φ extends as an analytic function to all of Λ . Again because \mathbf{f} and \mathbf{g} have the same order of vanishing at points of D , it is clear that φ cannot vanish anywhere.

Hence we assume that $K_{\mathbf{f}}(z, z)$ and $K_{\mathbf{g}}(z, z)$ are non-vanishing on Λ . By (2.4.8), the hypotheses imply that $\log K_{\mathbf{f}}(z, z) - \log K_{\mathbf{g}}(z, z)$ is harmonic in Λ . Therefore we can write

$$(2.5.1) \quad K_{\mathbf{f}}(z, z) = e^{u(z)} K_{\mathbf{g}}(z, z)$$

where u is a harmonic function in Λ .

If Λ is simply connected, we can find an analytic function ψ on Λ with $2\operatorname{Re}(\psi) = u$. Set $\varphi = e^\psi$. Then the above equation says that the two functions $K_{\mathbf{f}}(z, w)$ and $\varphi(z)\overline{\varphi(w)}K_{\mathbf{g}}(z, w)$ are equal on the diagonal. As both of these are analytic in z and anti-analytic in w , Lemma 2.5.1 shows that they are identically equal. Hence $\mathbf{f} \stackrel{d}{=} \varphi \mathbf{g}$. As φ does not vanish this shows that $\mathbf{f}^{-1}\{0\}$ and $\mathbf{g}^{-1}\{0\}$ have the same distribution.

If Λ is not simply connected, fix a $z_0 \in \Lambda$ and an $r > 0$ such that $D(z_0, r) \subset \Lambda$. Then there exists a non-vanishing analytic function φ on $D(z_0, r)$ such that

$$(2.5.2) \quad K_{\mathbf{f}}(z, w) = \varphi(z)\overline{\varphi(w)}K_{\mathbf{g}}(z, w)$$

for every $z, w \in D(z_0, r)$. Then fix $w \in D(z_0, r)$ such that $\varphi(w) \neq 0$, and note that $\frac{K_{\mathbf{f}}(z, w)}{\varphi(w)K_{\mathbf{g}}(z, w)}$ is an analytic function on $\Lambda - \{z : K_{\mathbf{g}}(z, w) = 0\}$ and is equal to φ on $D(z_0, r)$. Taking the union over $w \in D(z_0, r)$ of all these analytic functions we get an analytic extension of φ to the whole of

$$(2.5.3) \quad \Lambda \setminus \{z : K_{\mathbf{g}}(z, w) = 0 \quad \forall w \in D(z_0, r) \text{ s.t. } \varphi(w) \neq 0\}.$$

But if $K_{\mathbf{g}}(z, w) = 0$ for all w in an open set, then $K_{\mathbf{g}}(z, z) = 0$. By assumption this does not happen. Thus φ extends to the whole of Λ and the relationship

$$K_{\mathbf{f}}(z, w) = \varphi(z)\overline{\varphi(w)}K_{\mathbf{g}}(z, w)$$

persists. Thus $\varphi\mathbf{g}$ and \mathbf{f} have the same covariance kernel and by Gaussianity we get $\mathbf{f} \stackrel{d}{=} \varphi\mathbf{g}$ and φ is analytic on Λ . By inverting the roles of \mathbf{f} and \mathbf{g} , we see that $1/\varphi$ must also be analytic on Λ , which means that φ cannot vanish anywhere. \square

REMARK 2.5.3. Alternately, for the non-simply connected case, one could use the uniformization theorem to argue as follows. If Λ is a region of \mathbb{C} , let π be the covering map from \mathbb{D} or \mathbb{C} to Λ . Recall the definition of u from (2.5.1). Let $K_{\mathbf{f}}^*, K_{\mathbf{g}}^*, u^*$ be pull backs of $K_{\mathbf{f}}$ and $K_{\mathbf{g}}$ and u to \mathbb{D} . Then as before we can write $K_{\mathbf{f}}(z, w) = \varphi^*(z)\overline{\varphi^*(w)}K_{\mathbf{g}}^*(z, w)$ for a non-vanishing analytic function φ^* on \mathbb{D} . If $\pi(z_1) = \pi(z_2)$, then $\varphi^*(z_1) = \varphi^*(z_2)$ (Fix w and note that $K_{\mathbf{f}}^*(z_1, w) = K_{\mathbf{f}}^*(z_2, w)$ and $K_{\mathbf{g}}^*(z_1, w) = K_{\mathbf{g}}^*(z_2, w)$). Thus $\varphi = \varphi^*\pi^{-1}$ is well defined, does not vanish on Λ and satisfies, $K_{\mathbf{f}}(z, w) = \varphi(z)\overline{\varphi(w)}K_{\mathbf{g}}(z, w)$.

An immediate consequence is

COROLLARY 2.5.4. *The random power series described in (2.3.4), (2.3.5) and (2.3.6) are the only GAFs, up to multiplication by deterministic nowhere vanishing analytic functions, whose zeros are isometry-invariant under the respective group of isometries.*

Unfortunately, Theorem 2.5.2 is not constructive in that it does not tell us how to determine the k -point intensities of the zero set of a GAF if we know the first intensity. However, in the next chapter we shall see that it is possible to write general, although often intractable, formulas for the joint intensities of a GAF.

2.6. Notes

- The study of zeros of random polynomials goes back to Mark Kac (43) (but see also Paley and Wiener (68) which preceded Kac). He obtained the density of real zeros of various models of random polynomials, for example $a_0 + a_1x + \dots + a_nx^n$, where a_k are i.i.d. standard (real) Gaussians. These results can be obtained by the geometric proof presented here due to Edelman and Kostlan. See (23) for details. Following his papers, there was a significant amount of work done on this subject. Apart from zeros, there are many other interesting questions about random powers series as can be seen in the book of Kahane (44).
- The recent resurgence of interest in complex zeros is at least partly due to the work of many physicists such as Bogomolny, Bohigas and Leboeuf (7),(8), Hanay (32) and others. Apart from the Probabilistic perspective of these notes, there are other frameworks in which these objects are studied. For instance see Shiffman, Zelditch (76) (and references therein) who study random sections of line bundles.

- The planar GAF models were introduced (in parts) by Bogomolny, Bohigas and Leboeuf (7) and (8), Kostlan (50), Shub and Smale (78). Some of them are natural generalizations to complex coefficients of random polynomials studied by Mark Kac. A special case in the unit disk ($L = 2$) was found by Diaconis and Evans (19) as the limit of the logarithmic derivative of characteristic polynomials of random unitary matrices.
- Subhroshekhhar Ghosh pointed to us another proof that the hyperbolic GAF does not extend to any domain larger than the unit disk. If it did, the covariance kernel would be an extension of $(1 - z\bar{w})^{-L}$ and would remain analytic in z and anti-analytic in w . This is clearly impossible, as $(1 - |z|^2)^{-L}$ does not extend continuously to any point on the boundary of the unit disk.
- Theorem 2.5.2 is from Sodin (80), who found the result and related it to Calabi's rigidity theorem from differential geometry. A constructive way of recovering higher intensities from the first one is not known, and would be very desirable to have.

2.7. Hints and solutions

Exercise 2.1.1

- i. Note that X has density

$$(2.7.1) \quad f(z_1, \dots, z_d) = \frac{1}{(2\pi)^{d/2}} e^{-|z|^2/2}$$

where $|\cdot|$ denotes the Euclidean norm. By the transformation formula, the density of AX is $f(A^{-1}x)|\det(A^{-1})|^2$ (note that we use the real Jacobian here). The determinant is 1 and unitary matrices preserve the Euclidean norm, hence the density of X is invariant under A .

- ii. It suffices to consider the case $\mathbf{E}X = \mathbf{E}Y = 0$. By definition there are standard Gaussian random vectors \tilde{X} and \tilde{Y} and matrices A and B with $X = A\tilde{X}$ and $Y = B\tilde{Y}$. By adding columns of zeros to A or B , if necessary, we can assume that \tilde{X} and \tilde{Y} are both k -vectors, for some k , and A, B are both $d \times k$ matrices. Let \mathcal{A} and \mathcal{B} be the vector subspaces of \mathbb{C}^k generated by the row vectors of A and B , respectively. Suppose, WLOG, that the first $\ell \leq d$ row vectors of A form a basis of \mathcal{A} . Define the linear map $L: \mathcal{A} \rightarrow \mathcal{B}$ by

$$(2.7.2) \quad L(A_i) = B_i \text{ for } i = 1, \dots, \ell.$$

Here A_i is the i^{th} row vector of A , and B_i is the i^{th} row vector of B . Our aim is to show that L is an orthogonal isomorphism and then use the previous proposition. Let us first show that L is an isomorphism. The covariance assumption implies $AA^* = BB^*$. Suppose there is a vector $v_1A_1 + \dots + v_\ell A_\ell$ which maps to 0 under L . Then the vector

$$(2.7.3) \quad v = (v_1, \dots, v_\ell, 0, \dots, 0)$$

satisfies $vB = 0$. Hence

$$(2.7.4) \quad |vA|^2 = vAA^*v^* = vBB^*v^* = 0,$$

so $vA = 0$. Thus L is injective and $\dim \mathcal{A} \leq \dim \mathcal{B}$. Interchanging the roles of A and B shows that L is an isomorphism. The entry (i, j) of $AA^* = BB^*$ is the inner product of A_i and A_j as well as B_i and B_j , so the mapping L preserves inner products. Thus it can be extended on the orthocomplement of \mathcal{A} to give a unitary map $L: \mathbb{C}^k \rightarrow \mathbb{C}^k$ (or a unitary $k \times k$ matrix). Then $X = A\tilde{X}$ and $Y = B\tilde{Y} = AL^*\tilde{Y}$. From part i. we know that $L^*\tilde{Y}$ is standard complex normal, hence X and Y have the same distribution.

Exercise 2.2.2 If $z \in D$ and $D(z, r) \subset \Lambda$, then \mathbf{f} has a power series expansion in $D(z, r)$. By virtue of it being Gaussian in D , the coefficients of the power series have a jointly complex Gaussian distribution and hence \mathbf{f} is Gaussian on the whole of $D(z, r)$. In general, for any $w \in \Lambda \setminus D$, we can find a sequence of disks $D(z_1, r_1), \dots, D(z_n, r_n)$ contained in Λ such that $D(z_1, r_1) \subset D$, $D(z_n, r_n) \ni w$ and such that $z_j \in D(z_{j-1}, r_{j-1})$. Inductively, we apply our earlier observation about concentric disks to conclude that \mathbf{f} is Gaussian near z_2, \dots, z_n and hence near w .

Exercise 2.4.6

- i. Consider the needle as a union of shorter needles, use linearity of expectations. Each of the shorter needles will give the same expected number of intersections, provided that it is not too far from the center of mass of the original needle. And lastly, for $\ell < d$, the number of intersections is at most one, hence we get the probability of intersection.
- ii. The same argument used in i. shows that if a polygonal path of length ℓ is dropped uniformly between the two lines, the expected number of intersections is $c\ell$. Circles can be approximated arbitrarily well by polygonal paths, so the same is true for circles. A circle of diameter d has exactly two intersection, which yields $c = \frac{2}{\pi d}$.

Exercise 2.3.1

- i. Direct calculation shows that for φ as in (2.3.2), we have

$$\frac{|\varphi'(z)|}{1 + |\varphi(z)|^2} = \frac{1}{1 + |z|^2}.$$

This shows that the metric and area are preserved by φ .

- ii. The radius of $D(0, r)$ is given by $\int_0^r \frac{1}{1+t^2} dt$ and the area is given by $\int_{D(0,r)} \frac{dm(z)}{(1+|z|^2)^2}$. Straightforward calculations show that these integrals are equal to $\arctan(r)$ and $\frac{\pi r^2}{1+r^2}$, respectively.

Exercise 2.3.2

- i. This time we check easily that for φ as in (2.3.3), we have

$$\frac{|\varphi'(z)|}{1 - |\varphi(z)|^2} = \frac{1}{1 - |z|^2}.$$

This shows that the hyperbolic metric and area are preserved by φ .

- ii. The radius of $D(0, r)$ is given by $\int_0^r \frac{1}{1-t^2} dt$ and the area is given by $\int_{D(0,r)} \frac{dm(z)}{(1-|z|^2)^2}$. Straightforward calculations show that these integrals are equal to $\operatorname{arctanh}(r)$ and $\frac{\pi r^2}{1-r^2}$, respectively.