Gaussian Random Fields

Infinite, Improper and Intrinsic

Jakob Nebeling Hedegaard

Aalborg University

Master's thesis, spring 2019

Title: Gaussian Markov Random Field

Subtitle: Infinite, Improper and Intrinsic

Project Period: Spring Semester 2019

Project Group: 5.213c

Participants: Jakob Nebeling Hedegaard

Supervisor: Rasmus Waagepetersen

Date of Completion: June, 2019

Abstract: In this project Gaussian measures in $\mathbb{R}^n$ for $n \in \mathbb{N}$ and $\mathbb{R}^{\mathbb{Z}^n}$ are considered. Various facts about them are presented and the theory underlying the infinite dimensional case is treated carefully. Improper finite dimensional Gaussians are also presented and their use motivated, and an analogue of autoregressive models for stationary fields is described for intrinsic random fields.

Contents

Introduction

The Gaussian distribution is arguably one of the most used distributions in statistics. This is likely so due to a combination of central limit theorems and the fact that the Gaussian distribution has nice theoretical properties which makes working with it analytically tractable. The theory of the so-called non-degenerate Gaussians in $\mathbb{R}^n$ is well known. Non-degenerate Gaussians are those characterised by having positive definite covariance matrices. In this project we describe three ways of extending this theory.

Firstly, let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\mathbf{X}\colon \Omega \to \mathbb{R}^n$ be a random vector. In some cases, it may be reasonable to assume that all linear combinations of $\mathbf{X}$ are univariate normals, so that $\mathbf{X}$ is Gaussian, while at the same time $\mathbf{X}$ is constrained to some affine subspace of $\mathbb{R}^n$. Such a Gaussian does not have a covariance matrix of full rank, and the measure $P_{\mathbf{X}}$ does not have density wrt. the Lebesgue measure on $\mathbb{R}^n$. Such Gaussians are known as degenerate Gaussian. The theory about them can be unified with the theory of non-degenerate Gaussians, which we intend to do. This is considered briefly in Section 1.1. This unification depends on some preliminary facts such as e.g. the Hausdorff measures. These are considered in A.1.3 in the appendix.

Secondly, it may be theoretically reasonable to assume that we have infinitely many variables $\mathbf{X} = (\mathbf{X}_i)$ indexed by e.g. $i \in \mathbb{Z}^n$. We will then clarify what it means for $\mathbf{X}$ to have a Gaussian distribution. We will do this thoroughly which includes defining a σ-algebra on $\mathbb{R}^{\mathbb{Z}^n}$ and a topology generating that σ-algebra. We will also consider Kolmogorov's consistency theorem. In these cases, finite and infinite Gaussians, we have so-called marginal specification. In a marginal specification of a Gaussian measure we directly say what its distribution is. That is, if $\mathbf{X}$ is a Gaussian vector or process, then we can uniquely determine the distribution of $\mathbf{X}$ in terms of its mean and covariance functions. Marginal specification is considered in Chapter 1.

Thirdly, let us assume that we can argue that $\mathbf{X}$ is a degenerate Gaussian on some linear affine subspace $V + \alpha u$ for some vector $u \in \mathbb{R}^n \setminus V$ and scalar α, but that we do not know the value of α. In the complete absence of information about α it seems reasonable to assume a so-called uniform prior on $\mathbb{R}$ for α. This means that we believe that the likelihood of $\alpha \in (a, b]$ only depends on the length of the interval $b - a$. The problem, of course, is that there exists no probability distribution with this property. We therefore extend the notion of probability measures to include the so-called improper probability measures. Improper probability measures

can be compared with *intrinsic* random fields. These fields can be seen as a generalization of stationary random fields. We leave these two cases for Chapter 4 since they depend on notions from the second and third chapter.

This summarizes the first goal of the project. Our second goal will be to consider a special kind of model for Gaussians known as conditional specification. These are models where instead of specifying the mean and covariance structure of a Gaussian $\mathbf{X}$ directly, we express beliefs about $\mathbf{X}$ through the conditional distributions $\mathbf{X}_J \mid \mathbf{X}_{-J} = x_{-J}$. The question is then whether those beliefs are reasonable in that some $\mathbf{X}$ exists with those conditional distributions, and if that is the case, what the mean and covariance functions of $\mathbf{X}$ are.

We prepare for conditional specification in Chapter 2 by making precise what $P(\mathbf{X} \in A\cdot \mid \mathbf{X}_{-J} = x_{-J})$ means when $\mathbf{X}$ is a random field. We first consider the general notion of conditional distributions in Section 2.1, then regular conditional distributions in Section 2.2 and finally 'deterministic' conditional distributions in 2.3. Chapter 2 is *only* about these notions so that a reader, who is comfortable with them, can skip it.

In Chapter 3 we consider conditional specifications. We summarize well-known results about conditional distributions for finite dimensional Gaussians. After this we consider conditional specification in the finite dimensional case in Section 3.1.1. Conditional specification of Gaussian random fields is then considered in Section 3.3. This depends on the notion of *specifications* which is motivated and defined in Section 3.2. Regarding conditional specification of Gaussian fields, two questions are considered: what does a specification for a Gaussian field look like, and secondly, for a given specification, does a Gaussian field specified by it exist? These questions are answered in Sections 3.3.1 and 3.3.2 respectively. In the case that the specification has the so-called property of homogeneity, more things can be said. This is considered in 3.3.3. Preliminaries are considered in the appendix. Proofs of propositions that are important but not of primary interest are also included in the appendix.

I would like to thank my advisor Rasmus Waagepetersen for recommending literature, giving ideas for directions, and generally being very helpful and forthcoming this semester and the semester before.

Notation

Theorems, propositions and similar are highlighted in the usual LaTeX-fashion such as

Theorem 1 (This is the name of the theorem)**.** *Content goes here*

A number of the propositions in this thesis have been formulated and proved by the author without 'seeking help' in the literature. If this is the case the proposition (or similar) will be formatted with an asterisk such as

Theorem* 2 (This theorem has an original proof)**.** *Content goes here*

Random variables, vectors and fields are all denoted with bold font and upper case, Roman letters, e.g. random fields $\mathbf{X}\colon \Omega \to \mathbb{R}^{\mathbb{Z}^n}$. In some cases we will consider measurable maps from a probability space into some measurable space. Such maps are also denoted the same way, e.g. $\mathbf{Y}\colon \Omega \to \mathbb{Y}$. Random vectors (or fields) $\mathbf{X}\colon \Omega \to \mathbb{R}^J$ for some index set J are also denoted by $\mathbf{X} = (\mathbf{X}_i)_{i\in J}$. Here $\mathbf{X}_i\colon \Omega \to \mathbb{R}$ for $i \in J$ is known as a component of $\mathbf{X}$ and is a random variable and if J is countable. This will always be the case.

Realizations of random variables will be denoted with the corresponding lower case Roman letters and are not bold fonted. In general, elements of $\mathbb{R}^n$ will be denoted lower case letters which are not bold. Symbols such as $\{\mathbf{X} = x\}$, $\{\mathbf{X} \in A\}$ are preimages and contain the elements

ω such $\mathbf{X}(\omega) = x$ or $\mathbf{X}(\omega) \in A$ respectively. The clauses $\{\}$ are often omitted such as in e.g. $P(\mathbf{X} \in A)$. Sometimes sums are taking over sets that are given by the context. In such cases we often omit writing which set is being summed over. E.g. if (α_i) is a finite set of scalars then $\sum_i \alpha_i$ is the sum of all these α_i. In integrals we often have to make explicit which variable is being integrated over. We use the following notation

$$\int f(x) d\lambda(x)$$

where λ is some measure. This is contrasted with the notation $\int f(x)\lambda(dx)$ which is often used. For an index set J, Cartesian products of sets $(A_i)_{i \in J}$ is denoted by $\prod_{i \in J} A_i$.

Chapter 1

Marginal Specification of Gaussians

It is well known that the distribution of a multivariate normal in $\mathbb{R}^n$ is uniquely determined by its mean vector and covariance matrix. Thus to specify a multivariate normal in $\mathbb{R}^n$ it suffices to know these objects. Furthermore, we can also specify the distribution directly in terms of densities. This is more complicated in light of the fact that if a Gaussian distribution is degenerate, it does not have density wrt. the Lebesgue measure. We briefly consider these topics in the first part of this chapter.

In the second part we attempt to define what a Gaussian measure on $\mathbb{R}^{\mathbb{Z}^n}$ is. To do this we first need define a σ-algebra on $\mathbb{R}^{\mathbb{Z}^n}$. It turns out that there is a particularly nice choice of this. With this choice, Gaussian measures on $\mathbb{R}^{\mathbb{Z}^n}$ can be defined in terms of finite dimensional projections. For such a measure we have so-called mean and covariance functions and just like in the finite dimensional case, Gaussian measures are determined by these.

1.1 Gaussian Measures in $\mathbb{R}^n$

There are several equivalent definitions of Gaussian vectors. One can define them directly in terms of densities or by reduction to simpler cases. We take the latter approach in the following.

Definition 1 (Reduction to standard Gaussians)**.** *Let* $\mathbf{Z}$ *be a random vector with values in* $\mathbb{R}^k$*. Then we say that* $\mathbf{Z}$ *is standard Gaussian of dimension* k *if* $\mathbf{Z}$ *has density wrt. the Lebesgue measure* λ_k *given by*

$$f(z) = \frac{1}{\sqrt{(2\pi)^k}} \exp(-\frac{1}{2}\|z\|^2).$$

A stochastic vector $\mathbf{X}$ *mapping into* $\mathbb{R}^n$ *is said to be multivariate normal if there exists a standard Gaussian* $\mathbf{Z}$ *of dimension* k*, a matrix* $A \in \mathbb{R}^{n\times k}$ *and vector* $\mu \in \mathbb{R}^n$ *such that*

$$\mathbf{X} = A\mathbf{Z} + \mu.$$

The fact that a standard Gaussian has density f wrt. λ_k just means that $P(\mathbf{Z} \in U) = \int_U f d\lambda_k$ for Borel sets $U \subset \mathbb{R}^k$. We call the matrix A the transition matrix and μ the mean vector of $\mathbf{X}$. Note that in the definition we do not require $n \geq k$ nor that A is injective. It follows straightforwardly from this definition that any linear affine transformation of a multivariate Gaussian is also multivariate Gaussian. If $\mathbf{X}$ is Gaussian of dimension n with mean μ and covariance Σ then we write $\mathbf{X} \sim N_n(\mu, \Sigma)$. Using properties of covariance matrices and mean vectors we get

$$\mathbf{X} \sim N_n(\mu, \Sigma) \implies A\mathbf{X} + \nu \sim N_m(\nu + A\mu, A\Sigma A^\top) \quad \forall \nu \in \mathbb{R}^m, A \in \mathbb{R}^{m\times n} \tag{1.1}$$

We note that there is a peculiar case in the definition of Gaussians. This is the case where the transition matrix $A \in \mathbb{R}^{nk}$ is the 0 matrix. If this is the transition matrix of $\mathbf{X}$ then $\mathbf{X}$ is *concentrated* on the mean vector μ, in the sense that $\mathbf{X} = \mu$ P-a.s. This will be known as a Dirac measure in $\mathbb{R}^n$. A more generally, there are cases where A is not surjective onto $\mathbb{R}^n$. In that case, $\mathbf{X}$ lives on some linear affine space of $\mathbb{R}^n$. Such Gaussians are known as degenerate Gaussian vectors. These are of course contrasted with non-degenerate Gaussians which have support in all of $\mathbb{R}^n$. We have already noted that the transition matrix need not be injective, although this is a theoretically attractive case. It turns out that unless A is the zero matrix, we can always reduce to the injective case.

Lemma* 2. *Let $\mathbf{X}$ be multivariate normal with transition matrix $A \in \mathbb{R}^{nk}$ and mean vector $\mu \in \mathbb{R}^n$. Assume that A is not the zero matrix. Then there exists a matrix of full rank $B \in \mathbb{R}^{nh}$ and a standard Gaussian $\tilde{\mathbf{Z}}$ of dimension h such that $\mathbf{X} = B\tilde{\mathbf{Z}} + \mu$.*

See Lemma 92 in the Appendix for a proof of this. Using the lemma and the definition of Gaussians it is easy to derive several standard results, such as e.g. that the sum of independent Gaussians is a Gaussian variable.

1.1.1 Equivalent definitions

We will now give equivalent definitions of Gaussian vectors. Two such definitions are of importance. Firstly, note that in the sense of Definition 1, any Gaussian vector $\mathbf{X}$ of dimension n has the property that any linear combination $a^\top \mathbf{X}$ is univariate normal. Conversely, this property is also sufficient for being Gaussian.

Theorem 3 (Reduction to univariate case)**.** *A stochastic vector $\mathbf{X}$ with values in $\mathbb{R}^n$ is Gaussian if and only if for all $a \in \mathbb{R}^n$, $a^\top \mathbf{X}$ is a univariate normal.*

See e.g. [6]. For our second alternative definition we define the distribution of $\mathbf{X}$ directly in terms of densities. It is well known that when $\mathbf{X}$ is Gaussian of dimension n and has a covariance matrix that is invertible, then $\mathbf{X}$ has density wrt. the n-dimensional Lebesgue measure λ_n. However, when $\mathbf{X}$ is a degenerate Gaussian, the support of $\mathbf{X}$ is some linear affine proper subspace of $\mathbb{R}^n$. Such a space has Lebesgue measure 0 implying that a degenerate Gaussian can not have density wrt. λ_n. We can take care of this technicality by using Hausdorff measures.

Theorem 4 (Gaussian densities)**.** *Let $\mathbf{X}$ be a random vector where $V := \text{range}(\mathbf{X})$ is a linear affine subspace of $\mathbb{R}^n$ of dimension $k > 0$. Then $\mathbf{X}$ is Gaussian iff. there exist a symmetric and positive semi-definite matrix (SPSD) matrix Q with $\text{range}(Q) = V$ and a vector $\mu \in V$ such that the measure $P_{\mathbf{X}}$ has density wrt. the (n, k) Hausdorff measure given by*

$$f(x) = \mathbb{I}_V(x) \frac{\sqrt{\det^*(Q)}}{\sqrt{(2\pi)^k}} \exp(-\frac{1}{2}(x-\mu)^\top Q(x-\mu)) \tag{1.2}$$

where $\det^*(Q)$ is the product of the non-zero eigenvalues of Q. If $k = 0$, $P_{\mathbf{X}}$ is a Dirac measure which has density wrt. to the $(n, 0)$ Hausdorff measure given by

$$f(x) = \begin{cases} 1 & x = \mu \\ 0 & otherwise \end{cases}, \tag{1.3}$$

We may assume that Q is the Moore-Penrose inverse of the covariance, $Q = (\mathbb{V}\text{ar}[\mathbf{X}])^+$. We have a nice uniqueness property of $(\mathbb{V}\text{ar}[\mathbf{X}])^+$.

Theorem 5. *Assume that $k > 0$ and that $\mathbf{X}$ have density wrt. λ_{nk} given by*

$$f(x) = \mathbb{I}_V(x)\frac{\det^*(Q)}{\sqrt{(2\pi)^k}}\exp\left(-\frac{1}{2}(x-\mu)^\top Q(x-\mu)\right)$$

where V is an affine subspace of $\mathbb{R}^n$, $\mu \in V$ and Q is SPSD with range equal to $V - \mu$. Then $Q = (\mathbb{V}\mathrm{ar}[\mathbf{x}])^+$.

We can therefore, under some restrictions, think of the matrix $(\mathbb{V}\mathrm{ar}[\mathbf{X}])^+$ as the 'unique matrix in the density of $\mathbf{X}$'. We therefore define the precision of $\mathbf{X}$ as $\mathrm{Prec}[\mathbf{X}] = (\mathbb{V}\mathrm{ar}[\mathbf{X}])^+$. Proof of claims 4 and 5 can be found in [13].

1.2 Gaussians in Infinite Dimensional Spaces

Now that we have considered finite dimensional spaces and Gaussian vectors we are ready to begin the discussion of random vectors with range an infinite dimensional space. To distinguish such random vectors from their finite dimensional variants, we will call them random (or stochastic) *processes*.

There are many possible infinite dimensional vector spaces over $\mathbb{R}$. For our current purposes we will consider the sets $\mathbb{R}^{\mathbb{Z}^n}$ for $n \in \mathbb{N}$, $n \geq 1$, consisting of functions from $\mathbb{Z}^n$ to $\mathbb{R}$. The sets $\mathbb{Z}^n$ are lattices, and can be thought of as subspaces of $\mathbb{R}^n$. For $n = 1$ we can therefore think of an element in $\mathbb{R}^{\mathbb{Z}}$ as a countably infinite collection of real values placed at each of the integer points on the real number line. For $n = 2$ they can be represented as countably many values placed at integer points in the plane. We have similar interpretations for the remaining $n \in \mathbb{N}$.

Since elements in both of $\mathbb{R}^{\mathbb{Z}^2}$ and $\mathbb{R}^{\mathbb{Z}}$ can be represented as collections of countably many real values it is quite clear that these spaces are isomorphic as vector spaces. Indeed, pick your favourite bijection $f\colon \mathbb{Z}^2 \to \mathbb{Z}$ and define $J\colon \mathbb{R}^{\mathbb{Z}^2} \to \mathbb{R}^{\mathbb{Z}}$ by $J((x_{ij})_{ij}) = (x_{f^{-1}(k)})_k$. Then J is a linear isomorphism. From a geometric perspective, such isomorphisms are unlikely to be of much use.

Lemma* 6. *Let $f : \mathbb{Z}^2 \to \mathbb{Z}$ be a bijection. Then for every $d \in \mathbb{N}$, there exists $(i, j) \in \mathbb{Z}^2$ such that $|f(i, j) - f(i, j + 1)| \geq d$.*

This lemma says that points which are close in $\mathbb{Z}^2$ have bijective images that are arbitrarily far apart in $\mathbb{Z}$. Such bijections do not preserve locality. The proof of this lemma can be found in the Appendix. Because of this, there is not much hope that linear isomorphisms $\mathbb{R}^{\mathbb{Z}^n} \to \mathbb{R}^{\mathbb{Z}}$ will lead to nice ways of reducing probability theory on $\mathbb{R}^{\mathbb{Z}^n}$ to probability theory on $\mathbb{R}^{\mathbb{Z}}$.

1.2.1 A σ-algebra on $\mathbb{R}^{\mathbb{Z}^n}$

We restate the context as follows. We intend to endow $\mathbb{R}^{\mathbb{Z}^n}$ with a σ-algebra so that we can define what it means for a map $\mathbf{X}\colon \Omega \to \mathbb{R}^{\mathbb{Z}^n}$ to be measurable, and thus a stochastic process. Secondly, this σ-algebra should be reasonable in that it should not be too large so that only a few maps are measurable, nor should it be too small so that we can not say anything interesting about e.g. probabilities associated with $\mathbf{X}$.

Denote the σ-algebra which we intend to define by $\mathbb{B}(\mathbb{R}^{\mathbb{Z}^n})$. Let some map $\mathbf{X}\colon \Omega \to \mathbb{R}^{\mathbb{Z}}$ and projection $\pi_i\colon \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ be given. Suppose that we know that $\mathbf{X}$ is measurable. Then a natural property of $\mathbb{B}(\mathbb{R}^{\mathbb{Z}^n})$ is that composition $\pi_i \circ \mathbf{X}\colon \Omega \to \mathbb{R}$ is measurable too. Let us therefore consider the preimages $\pi^{-1}(B)$ for Borel sets $B \subset \mathbb{R}$. Elements $x \in \pi_i^{-1}(B) \in \mathbb{R}^{\mathbb{Z}^n}$ are of the form $x_i \in B$ and otherwise arbitrary. We write this preimage as

$$\pi_i^{-1}(B) = \prod_{k\in\mathbb{Z}^n} B_k \tag{1.4}$$

where $B_k = \mathbb{R}$ unless $k = i$ in which case $B_i = B$. Sets of this form will henceforth be known as *cylinders*. We denote the class of cylinders as $\mathcal{U}$. We will then define $\mathbb{B}(\mathbb{R}^{\mathbb{Z}^n})$ in terms of $\mathcal{U}$

Definition 7. *Define $\mathbb{B}(\mathbb{R}^{\mathbb{Z}^n}) := \sigma(\mathcal{U})$ where $\mathcal{U}$ is the basis containing all cylinders in $\mathbb{R}^{\mathbb{Z}^n}$.*

The notation $\mathbb{B}(\mathbb{R}^{\mathbb{Z}^n})$ usually implies that $\mathbb{B}(\mathbb{R}^{\mathbb{Z}^n})$ contains Borel sets, i.e. sets generated by the open sets in some topology. We have not yet considered a topology on $\mathbb{R}^{\mathbb{Z}^n}$ and we are therefore not justified in using this notation. However, we will use it and think of it as foreshadowing the results in Section 1.2.3 where we show that $\mathbb{B}(\mathbb{R}^{\mathbb{Z}^n})$ are Borel sets wrt. a generalization of the Euclidean topology.

We will henceforth always understand measurability of stochastic processes mapping into $\mathbb{R}^{\mathbb{Z}^n}$ wrt. the σ-algebra $\mathbb{B}(\mathbb{R}^{\mathbb{Z}^n})$.

The set $\mathbb{B}(\mathbb{R}^{\mathbb{Z}^n})$ is also known as the infinite product algebra. It turns out that measurability of a process $\mathbf{X} = (\mathbf{X}_i)_{i\in\mathbb{Z}^n}$ wrt. this product algebra can be decided by considering its component functions $\mathbf{X}_i$.

Theorem* 8. *A stochastic process $\mathbf{X}\colon \Omega \to \mathbb{R}^{\mathbb{Z}^n}$ is measurable if and only if for any projection $\pi_i\colon \mathbb{R}^{\mathbb{Z}^n} \to \mathbb{R}$, the composition $\pi_i \circ \mathbf{X}$ is measurable.*

Proof. Since $\mathcal{U}$ is a basis for $\mathbb{B}(\mathbb{R}^{\mathbb{Z}^n})$, measurability of a stochastic process $\mathbf{X}\colon \Omega \to \mathbb{R}^{\mathbb{Z}^n}$ is equivalent with $\mathbf{X}^{-1}(U)$ being a measurable set for any set $U \in \mathcal{U}$. Every such set U is a preimage of the form $\pi_i^{-1}(B)$ where B is a Borel set in $\mathbb{R}$. Thus if $\mathbf{X}$ is measurable then for any i and any Borel set B, $(\pi_i \circ \mathbf{X})^{-1}(B) = \mathbf{X}^{-1}(\pi_i^{-1}(B))$ is a measurable set. This shows that $\pi_i \circ \mathbf{X}$ is measurable.

Conversely, assume $\pi_i \circ \mathbf{X}$ is measurable for any i. For any $U \in \mathcal{U}$ write $U = \pi_i^{-1}(B)$ for some Borel set B in $\mathbb{R}$. Then by assumption we have that $\mathbf{X}^{-1}(U)$ is a measurable set. This holds for any U in the basis $\mathcal{U}$ so that $\mathbf{X}$ is measurable. □

Therefore a function $\mathbf{X} = (\mathbf{X}_i)_{i\in\mathbb{Z}^n}$ is measurable if and only if all its components functions $\mathbf{X}_i$ are measurable as functions into $\mathbb{R}$. We note also that this is the coarsest such σ-algebra.

Proposition 9. *Let $\mathcal{F}$ be a σ-algebra on $\mathbb{R}^{\mathbb{Z}^n}$ such that a function $\mathbf{X}\colon \Omega \to \mathbb{R}^{\mathbb{Z}^n}$ is measurable if and only if $p_i \circ \mathbf{X}$ is measurable for all projections p_i. Then $\mathbb{B}(\mathbb{R}^{\mathbb{Z}^n}) \subset \mathcal{F}$.*

Proof. We note that all projections π_i are measurable, since π_i composed with the identity map on $\mathbb{R}^{\mathbb{Z}^n}$ is measurable. But then $\mathcal{F}$ must contain all sets of the form $\pi_i^{-1}(B)$ for Borel sets $B \subset \mathbb{R}$. It follows that $\mathbb{B}(\mathbb{Z}^n) \subset \mathcal{F}$. □

Measurability can be verified by checking whether preimages $\mathbf{X}^{-1}(U)$ are measurable in Ω for $U \in \mathcal{U}$. If we extend to the slightly larger basis set $\tilde{\mathcal{U}}$ consisting of all sets $\prod B_i$, where B_i are measurable in $\mathbb{R}$ and $B_i \neq \mathbb{R}$ for only finitely many i, we get the follow uniqueness property for stochastic processes $\mathbf{X}\colon \Omega \to \mathbb{R}^{\mathbb{Z}^n}$.

Theorem* 10. *Let $\mathbf{X}$, $\mathbf{Y}$ be real stochastic processes on $\mathbb{Z}^n$ such that $P(\mathbf{X} \in \tilde{U}) = P(\mathbf{Y} \in \tilde{U})$ for all $\tilde{U} \in \tilde{\mathcal{U}}$. Then $P(\mathbf{X} \in B) = P(\mathbf{Y} \in B)$ for all $B \in \mathbb{B}(\mathbb{R}^{\mathbb{Z}^n})$.*

Proof. The proof is an immediate corollary to the uniqueness theorem for measures as stated in Theorem 90 if we can show that the assumptions of that theorem are true. This is indeed the case since $P_{\mathbf{X}}$ and $P_{\mathbf{Y}}$ are finite measures on $\mathbb{R}^{\mathbb{Z}^n}$ and $\tilde{\mathcal{U}}$ is stable under intersections. □

We note that the result also holds when finiteness of the measures is replaced with σ-finiteness. As a direct corollary we see that two random fields $\mathbf{X}$ and $\mathbf{Y}$ have identical distributions if their finite dimensional parts $\mathbf{X}_J = \pi_J \circ \mathbf{X}$ and $\mathbf{Y}_J = \pi_J \circ \mathbf{Y}$ have identical measures where π_J is a the projection $\pi_J(x) = (x_j)_{j\in J}$ for finite subsets $J \subset \mathbb{Z}^n$.

1.2.2 Kolmogorov consistency theorem

Suppose we have a well-defined stochastic process $\mathbf{X}\colon \Omega \to \mathbb{Z}^n$. Then for any subset $J \subset \mathbb{R}^{\mathbb{Z}^n}$, we have the following trivial facts of the finite dimensional distributions of $\mathbf{X}$

1. For any Borel sets B_i, $i \in J$, and bijection $b\colon J \to J$

$$P(\mathbf{X}_i \in B_i : i \in J) = P(\mathbf{X}_{b(i)} \in B_{b(i)} : i \in J).$$

   This should be read as saying that any reordering of the terms $\mathbf{X}_i \in B_i$ does not change probabilities.

2. For any $k \notin J$,

$$P(\mathbf{X}_k \in \mathbb{R}, \mathbf{X}_i \in B_i : i \in J) = P(\mathbf{X}_i \in B_i : i \in J).$$

The colon ": " should not be interpreted as a condition on $i \in J$, but should instead be understood as intersection of the sets. I.e. $\mathbf{X}_i \in B_i\colon i \in J$ means $\cap_{i\in J}\{\mathbf{X}_i \in B_i\}$. These properties are sometimes known as the consistency conditions. Assume that we have a collection of measures μ_J on $\mathbb{R}^J$ for finite $J \subset \mathbb{Z}^n$. Kolmogorov's consistency theorem then says, in a precise sense, that the consistency conditions are sufficient for ensuring that the measures μ_J 'come from' a measure on $\mathbb{R}^{\mathbb{Z}^n}$.

Theorem 11 (Kolmogorov consistency theorem)**.** *Assume that we have a class (μ_J) indexed by finite $J \subset \mathbb{Z}^n$ of probability measures on $\mathbb{R}^J$ with the following properties:*

1. *For a given J and bijection $b\colon J \to J$ and Borel sets $B_j \subset \mathbb{R}$ we have*

$$\mu_J(\prod_{j\in J} B_j) = \mu_J(\prod_{j\in J} B_{b(j)})$$

2. *Let $i \notin J$ and $B_j \subset \mathbb{R}$ be Borel sets for $j \in J$. Denote $B_i = \mathbb{R}$ and assume that*

$$\mu_{J\cup\{i\}}(\prod_{j\in J\cup\{i\}} B_j) = \mu_J(\prod_{j\in J} B_j)$$

Then there exists a measure μ on $\mathbb{R}^{\mathbb{Z}^n}$ such that for $J \subset \mathbb{Z}^n$ we have $\mu \circ \pi_J^{-1} = \mu_J$.

Here the notation π_J^{-1} refers to a set function and does not imply that π_J is invertible (which it is not). We note also that if the μ_J are probability measures then so is μ since e.g. $1 = \mu_i(\mathbb{R}) = \mu \circ \pi_i^{-1}(\mathbb{R}) = \mu(\mathbb{R}^{\mathbb{Z}^n})$. Kolmogorov's consistency theorem allows us to verify the existence of stochastic processes. Indeed, assume we have a collection of measures on $\mathbb{R}^J$ for finite J and we wish to know whether these measures are the push-forward measures of $\mathbf{X}_J$ for some random field $\mathbf{X}\colon \Omega \to \mathbb{R}^{\mathbb{Z}^n}$.

Corollary* 12. *Let assumptions about the collection (μ_J) of probability measures be as in Kolomogorov consistency theorem. Then there exists a random field $\mathbf{X}$ with domain a probability space such that $P_{\mathbf{X}_J} = \mu_J$ for all J.*

Proof. Let μ be the measure on $\mathbb{R}^{\mathbb{Z}^n}$ such that $\mu \circ \pi_J^{-1} = \mu_J$ which exists by Kolmogorov's theorem. Now choose $\Omega = \mathbb{R}^{\mathbb{Z}^n}$, $\mathbf{X} = I\colon \mathbb{R}^{\mathbb{Z}^n} \to \mathbb{R}^{\mathbb{Z}^n}$ the identity map and $P = \mu$. Then clearly $P_{\mathbf{X}} = \mu$ and $P_{\mathbf{X}_J} = P_{(\pi_J \circ \mathbf{X})} = P_{\mathbf{X}} \circ \pi_J^{-1} = \mu_J$. □

1.2.3 $\mathbb{B}(\mathbb{R}^{\mathbb{Z}^n})$ as induced by a topology

To construct of a topology on $\mathbb{R}^{\mathbb{Z}^n}$ we can use the same ideas as in the construction of the σ-algebra $\mathbb{B}(\mathbb{R}^{\mathbb{Z}^n})$. Indeed, we can define projections $\pi_i \colon \mathbb{R}^{\mathbb{Z}^n} \to \mathbb{R}$ for all $i \in \mathbb{Z}^n$. We then require that these projections are continuous and the topology on $\mathbb{R}^{\mathbb{Z}^n}$ should therefore contain preimages $\pi_i^{-1}(\mathcal{O})$ where $\mathcal{O}$ is an open set of $\mathbb{R}$.

Consider the set of sets $\prod_{i\in\mathbb{Z}^n} \mathcal{O}_i$ where $\mathcal{O}_i$ is open in $\mathbb{R}$ and $\mathcal{O}_i \neq \mathbb{R}$ for only *finitely* many i. Note that this set is closed under finite intersections. It is therefore a basis for some topology on $\mathbb{R}^{\mathbb{Z}^n}$. See [5] for details. This topology will be known as the Euclidean topology. This topology differs from the box-topology which is generated by the sets $\prod_{i\in\mathbb{Z}^n} \mathcal{O}_i$ where the $\mathcal{O}_i$ are only assumed to be open, and thus can be proper subsets of $\mathbb{R}$ for *infinitely* many i. [5]

Theorem 13. *The product topology on $\mathbb{R}^{\mathbb{Z}^n}$ is the unique topology on $\mathbb{R}^{\mathbb{Z}^n}$ with the property that a map $f \colon \mathbb{Y} \to \mathbb{R}^{\mathbb{Z}^n}$ from a topological space $\mathbb{Y}$ is continuous if and only if $\pi_i \circ f$ is continuous.*

This property is sometimes known as the characteristic property of the product topology. See [5] for further details. To connect this topology with $\mathbb{B}(\mathbb{R}^{\mathbb{Z}^n})$ we will need to recall some facts from topology.

1. A countable product of metric spaces is metrizable.
2. A countable product of separable spaces is separable.
3. In a metrizable separable space (Ω, τ), for any basis $\mathcal{B}$ for the topology τ, every element in τ can be written as a countable union of elements in $\mathcal{B}$.
4. A countable topological product of completely metrizable spaces is completely metrizable.

See e.g. [10]. We elaborate on these claims in Section 1.2.4 including defining the terms used. For now it suffices to note that $\mathbb{R}^{\mathbb{Z}^n}$ is countably generated by a basis in the sense of property 3. above.

Theorem* 14. *The σ-algebra defined in Definition 7 is the least σ-algebra on $\mathbb{R}^{\mathbb{Z}^n}$ containing the Euclidean topology.*

Proof. We denote the σ-algebra in Definition 7 by $\mathbb{B}(\mathbb{R}^{\mathbb{Z}^n})$ as usual. We similarly denote the least σ-algebra containing the Euclidean topology τ by $\mathbb{B}(\tau)$. We then intend to prove that $\mathbb{B}(\mathbb{R}^{\mathbb{Z}^n}) = \mathbb{B}(\tau)$.

Note that τ contains all cylinders of the form $\prod_{i\in\mathbb{Z}^n} \mathcal{O}_i$ where $\mathcal{O}_i = \mathbb{R}$ for all $i \neq j \in \mathbb{Z}^n$ and $\mathcal{O}_j$ is any open set in $\mathbb{R}$. But then $\mathbb{B}(\tau)$ contains sets $\prod_{i\in\mathbb{Z}^n} B_i$ where $B_i = \mathbb{R}$ for all $i \neq j$ and B_j is any Borel set in $\mathbb{R}$. The collection of these sets is a basis for $\mathbb{B}(\mathbb{R}^{\mathbb{Z}^n})$ so that $\mathbb{B}(\mathbb{R}^{\mathbb{Z}^n}) \subset \mathbb{B}(\tau)$.

The converse inclusion follows by almost the same argument. Clearly, $\mathbb{B}(\mathbb{R}^{\mathbb{Z}^n})$ contains finite intersections of cylinders of the form $\prod_{i\in\mathbb{Z}^n} \mathcal{O}_i$ where only finitely many $\mathcal{O}_i \neq \mathbb{R}$ and $\mathcal{O}_i$ are open. This is a basis for τ. Then using property 3. above, we see that an open set is a *countable* union of elements in this basis. This shows that $\tau \subset \mathbb{B}(\mathbb{R}^{\mathbb{Z}^n})$ from which we derive $\mathbb{B}(\tau) \subset \mathbb{B}(\mathbb{R}^{\mathbb{Z}^n})$ □

1.2.4 Further properties of $\mathbb{R}^{\mathbb{Z}^n}$

Here we verify some of the claims made in the section 1.2.3. As it turns out, $\mathbb{R}^{\mathbb{Z}^n}$ belongs to a particularly nice class of spaces.

A topological space is said to be metrizable if its topology can be induced by a metric. For a metrizable space there is not necessarily any preferred metric inducing the topology; merely the fact that it is metrizable implies several nice properties such as the Hausdorff property. In the context of countable topological products we have the following.

Theorem 15. *The product topology on $\prod \mathbb{X}_i$ of countably many metrizable spaces $\mathbb{X}_i$ is metrizable. If ρ_i generates the topology on $\mathbb{X}_i$, then*

$$\rho = \sum_{i=1}^{\infty} 2^{-i} \frac{\rho_i}{1+\rho_i}$$

generates the product topology on $\prod \mathbb{X}_i$.

A proof of this claim can be found in the Appendix. Note that $\mathbb{R}^{\mathbb{Z}^n}$ is a countable product of countably many copies of the metric space $\mathbb{R}$. As a direct corollary to Theorem 15 we therefore get that $\mathbb{R}^{\mathbb{Z}^n}$ is metrizable. More than this is true. A topological space for which the topology can be induced by a complete metric is said to be *completely metrizable*. It is well known that the Euclidean metric on $\mathbb{R}$ makes $\mathbb{R}$ into a complete metric space. A natural question is then whether this suffices to make $\mathbb{R}^{\mathbb{Z}^n}$ completely metrizable.

Theorem* 16. *A countable product of completely metrizable spaces is completely metrizable*

Proof. In this proof we will deviate from our usual notation. We will let elements in $\mathbb{X}_i$ be denoted with small non-bold fonted letters such as e.g. x. Elements in $\prod \mathbb{X}_i$ will be denoted by bold font and individual components of $\mathbf{x}$ using function notation as $\mathbf{x} = (x(i))$.

Let $(\mathbb{X}_i, \tau_i)$ be the completely metrizable spaces and assume that the topologies are induced by the complete metrics ρ_i. By Theorem 15, the product topology on $\prod \mathbb{X}_i$ is induced by the metric

$$\rho = \sum_{i=1}^{\infty} 2^{-i} \frac{\rho_i}{1+\rho_i}.$$

It thus suffices to show that this metric is complete. We start by noting that we have the following implications.

$$\rho_i < r \implies \frac{\rho_i}{1+\rho_i} < r \tag{1.5}$$

$$\frac{\rho_i}{1+\rho_i} < r \implies \rho_i < 2r \text{ if } r < \frac{1}{2} \tag{1.6}$$

Now to show that ρ is complete let $(x_n) \subset \mathbb{X}_i$ be a $\frac{\rho_i}{1+\rho_i}$ Cauchy sequence. Then using (1.6) one can show that (x_n) is also ρ_i Cauchy so that it converges to some element $x \in \mathbb{X}_i$ wrt. ρ_i. It then follows by (1.5) that (x_n) also converges to x wrt. $\frac{\rho_i}{1+\rho_i}$.

Thus, ρ is an infinite sum of (scaled) complete metrics. Assume that $(\mathbf{x}_n)$ is a ρ-Cauchy sequence in $\prod \mathbb{X}_i$. The inequality

$$2^{-i} \frac{\rho_i}{1+\rho_i} \leq \sum_{k=1}^{\infty} 2^{-i} \frac{\rho_i}{1+\rho_i}$$

implies that each component sequence $(x_n(i))$ for fixed i is Cauchy too, so that they converge to elements $x(i)$. Define $\mathbf{x} = (x(i))$. The theorem follows if we can show that $(\mathbf{x}_n)$ converges to $\mathbf{x}$. We thus let $\epsilon > 0$ be given.

Note the obvious fact that each $\frac{\rho_i}{1+\rho_i}$ is bounded by 1. It follows that there exists some K big enough such that $\sum_{i=K}^{\infty} 2^{-i} \frac{\rho_i}{1+\rho_i} < \frac{\epsilon}{2}$. Now the sum

$$\sum_{i=1}^{K-1} 2^{-i} \frac{\rho_i(x_n(i), x(i))}{\rho_i(x_n(i), x(i)) + 1}$$

is a finite sum of sequences which converge to 0 as $n \to \infty$. Thus we can find N large enough so that this sum is less than $\frac{\epsilon}{2}$ when $n \geq N$. Combining this we get that if $n \geq N$ we have $\rho(\mathbf{x}_n, \mathbf{x}) < \epsilon$. □

We have thus seen that $\mathbb{R}^{\mathbb{Z}^n}$ is a completely metrizable. A separable topological space is a topological space $\mathbb{X}$ containing a countable set Q which is dense in $\mathbb{X}$. By definition this means that the smallest closed set containing Q is equal to $\mathbb{X}$. Equivalently, there exists a countable set Q, such that any non-empty open set in $\mathbb{X}$ will have non-empty intersection with Q. As an example of this, $\mathbb{R}$ is separable with $\mathbb{Q}$ being a countable dense subset.

Theorem 17. *A countable product of separable spaces is separable*

Proof. Let the $\mathbb{N}$-many separable spaces be denoted by $\mathbb{X}_i$. By hypothesis, each space $\mathbb{X}_i$ contains a countable set $Q_i \subset \mathbb{X}_i$ which is dense in $\mathbb{X}_i$. Let $x = (x_i)$ be any element in $\prod_{i\in\mathbb{N}} \mathbb{X}_i$. Define sets $U_k = \prod_{i=1}^{k} Q_i \times \prod_{i=k+1}^{\infty} \{x_i\}$. These sets are countable since finite Cartesian products of countable sets are countable. Furthermore, countable unions of countable sets are countable too. Incidentally $U = \cup_{k\in\mathbb{N}} U_k$ is countable. We propose that U is dense in $\prod_{i\in\mathbb{N}} \mathbb{X}_i$.

Let $\mathcal{O}$ be any open set in $\prod \mathbb{X}_i$. By definition, the product topology has a basis given by the set of cylinders $C = \prod_{i\in\mathbb{N}} \mathcal{O}_i$ where the $\mathcal{O}_i$ are non-empty and open in $\mathbb{X}_i$, and for at most finitely many i we have $\mathcal{O}_i \neq \mathbb{X}_i$. Since this is a basis, there must exist a cylinder such that $C \subset \mathcal{O}$. To show $\mathcal{O} \cap U \neq \emptyset$ it suffices to show $C \cap U \neq \emptyset$.

Note that we may write any basis set C as $C_k = \prod_{i=1}^{k} \mathcal{O}_i \times \prod_{i=k+1}^{\infty} \mathbb{X}_i$ for some k. Then

$$C_k \cap U_k = \prod_{i=1}^{k} (\mathcal{O}_i \cap Q_i) \prod_{i=k+1}^{\infty} (X_i \cap \{x_i\}) \neq \emptyset$$

so that $C_k \cap U \neq \emptyset$. □

As an immediate corollary we get that $\mathbb{R}^{\mathbb{Z}^n}$ is completely metrizable and separable. The class of such spaces is well known.

Definition 18. *Let $(\mathbb{X}, \tau)$ be a topological space which is separable and completely metrizable. Then we say that $(\mathbb{X}, \tau)$ is a Polish space.*

Informally, separability gives some kind of limitation of size of a topological space, since any neighbourhood of points in $\mathbb{X}$ contain points in some fixed countable subset. Separability in itself, however, does not say much on its own since the notion of neighbourhood can mean many things in general topological spaces. However, when we restrict to metric topologies we get the following.

Lemma* 19. *A metrizable separable space $\mathbb{X}$ has cardinality at most $|\mathbb{X}| \leq |\mathbb{R}|$.*

Proof. Assume that $|\mathbb{X}| \geq |\mathbb{R}|$. By definition, $\mathbb{X}$ contains a countable subset Q which is dense in $\mathbb{X}$. Now since $\mathbb{X}$ is metrizable, limits of convergent sequences in $\mathbb{X}$ are unique. Furthermore, since Q is dense, every element in $\mathbb{X}$ is the limit of some sequence in Q. But then there is an injection from $f\colon \mathbb{X} \to Q^{\mathbb{N}}$. Indeed, for all $x \in \mathbb{X}$ we can find a sequence $(q_n) \in Q^{\mathbb{N}}$ such that x is the limit of (q_n). Define $f(x) = (q_n)$. Uniqueness of limits implies injectivity of f. Now facts about cardinal numbers give

$$|\mathbb{R}| \leq |\mathbb{X}| \leq |Q^{\mathbb{N}}| = |\mathbb{Q}^{\mathbb{N}}| = |\mathbb{R}|$$

which implies $|\mathbb{X}| = |\mathbb{R}|$. □

This gives a strong limitation of the size of a Polish space. Their cardinalities can not be greater than that of the real line. What is maybe more surprising is the fascinating fact that Polish spaces are in a sense determined (up to isomorphism) by their cardinalities as stated in the following.

Theorem 20. *Let $\mathbb{X}$ and $\mathbb{Y}$ be Polish spaces and let $\mathcal{X}$ and $\mathcal{Y}$ be the Borel sets induced by the topologies on $\mathbb{X}$ and $\mathbb{Y}$. Then if $|\mathbb{X}| = |\mathbb{Y}|$ we have that $\mathbb{X} \cong \mathbb{Y}$ as measurable spaces. I.e. there exists a bijective map $\theta\colon \mathbb{X} \to \mathbb{Y}$ such that both θ and θ^{-1} are measurable.*

This is a classical result. See e.g. [15] for a proof. In conjunction with Lemma 19, The theorem shows that if we know $(\Omega, \mathcal{F})$ is a measurable space such that $\mathcal{F}$ are the Borel sets of a Polish topology, then $(\Omega, \mathcal{F})$ is Borel equivalent with some subset of $\mathbb{R}$. This is precisely the case for $(\mathbb{R}^{\mathbb{Z}^n}, \mathbb{B}(\mathbb{R}^{\mathbb{Z}^n}))$ so that we may use the theorem to get the following corollary.

Corollary 21. *We have that $\mathbb{R}^{\mathbb{Z}^n}$ with the product topology is Borel isomorphic with $([0,1], \mathbb{B})$ where $[0,1] \subset \mathbb{R}$ is the closed interval and $\mathbb{B}$ is the restriction of the Borel algebra on $\mathbb{R}$ to $[0,1]$.*

In fact, more than this is true since all Polish spaces are isomorphic with one of the following

1. $\{1, 2, \ldots, k\}$ for some $k \in \mathbb{N}$ with the discrete algebra
2. $\mathbb{N}$ with the discrete algebra, or
3. $\mathbb{R}$ with the Borel algebra induced by the Euclidean topology.

1.3 Gaussian Processes in $\mathbb{R}^{\mathbb{Z}^n}$

We will start this section with the general notion of stochastic processes $\mathbf{Y} = (\mathbf{Y}_i)_{i \in \mathbb{Z}^n}$ in $\mathbb{R}^{\mathbb{Z}^n}$. For such a process we may define the obvious generalizations of the mean vector and covariance matrix as the functions $\mu\colon \mathbb{Z}^n \to \mathbb{R}$ and $C\colon \mathbb{Z}^n \times \mathbb{Z}^n \to \mathbb{R}$ given by

$$\mu(i) = \mathbb{E}[\mathbf{Y}_i], \quad C(i,j) = \operatorname{Cov}(\mathbf{Y}_i, \mathbf{Y}_j). \tag{1.7}$$

These functions, provided they are well-defined, are known respectively as the mean function and covariance function. The covariance function C encodes the covariance matrix of all the finite subsets of components in the sense that for $n = |J| < \infty$, the matrix $C_J \in \mathbb{R}^{n \times n}$ given by

$$[C_J]_{ij} = C(i,j) \quad i,j \in J$$

must be the covariance matrix of the components $\mathbf{Y}_J = (\mathbf{Y}_i)_{i \in J}$.. This matrix must be SPSD. In terms of the covariance function, this means that $C(i,j) = C(j,i)$ and for any $n \in \mathbb{N}$, any vector λ in $\mathbb{R}^n$ and any collection of n indices $J \subset \mathbb{Z}^n$ we have

$$\sum_{i,j \in J} \lambda_i \lambda_j C(i,j) \geq 0.$$

We say that a function $\mathbb{Z}^n \times \mathbb{Z}^n \to \mathbb{R}$ with this property is SPSD, too. Furthermore any stochastic process $\mathbf{Y}$ must also satisfy the natural permutation and marginalization laws described in Section 1.2.2.

Definition 22. *A stochastic process $\mathbf{X}\colon \Omega \mapsto \mathbb{R}^{\mathbb{Z}^n}$, $\mathbf{X} = (\mathbf{X}_i)_{i \in \mathbb{Z}^n}$ is said to be Gaussian if for all finite subsets of components $J \subset \mathbb{Z}^n$ the stochastic vector $\mathbf{X}_J = (\mathbf{X}_i)_{i \in J}$ has a multivariate normal distribution.*

1.3.1 Marginal specification of Gaussian random fields

Section 1.1.1 can be used to derive equivalent definitions for Gaussian processes. For instance by Theorem 3, the assumptions in Definition 22 hold if and only if any finite linear combination of components is univariate Gaussian. To prove existence of Gaussian processes in $\mathbb{R}^{\mathbb{Z}^n}$ we will use a combination of Kolmogorov's extension theorem and results from the finite dimensional case.

Lemma* 23. *Let $\mu\colon \mathbb{Z}^n \to \mathbb{R}$ be any function and let $C\colon \mathbb{Z}^n \times \mathbb{Z}^n \to \mathbb{R}$ be SPSD. Then there exists a Gaussian stochastic process $\mathbf{X}$ with mean function μ and covariance function C. Furthermore, the distribution of a Gaussian process $\mathbf{X}$ is uniquely determined by the mean and covariances functions.*

Remark 24. *We note that uniqueness in this regard refers to the fact that if $\mathbf{X}$ and $\tilde{\mathbf{X}}$ are Gaussian processes and have the same mean and covariance functions then their distribution measures are identical, $P_{\mathbf{X}} = P_{\tilde{\mathbf{X}}}$. This of course does not imply that $\mathbf{X} = \tilde{\mathbf{X}}$.*

Proof. We will use Kolmogorov's extension theorem. To do so we define a class of finite dimensional distributions as follows. We will denote these distributions by F_J for finite sets of indices $J \subset \mathbb{Z}^n$, $n = |J| < \infty$. For one such J define F_J as the distribution of a multivariate normal vector in $\mathbb{R}^n$ with mean vectors $\mu_J = (\mu(i))_{i\in J} \in \mathbb{R}^n$ and covariance matrices $C_J = [C(i,j)]_{i,j\in J} \in \mathbb{R}^{n\times n}$. This can be made precise by giving J a linear ordering, but for the sake of clarity we will not do this.

These distributions satisfy the consistency condition regarding permutations since they are Gaussian. An easy proof of this can be obtained using the (Hausdorff) density of $\mathbf{X}_J$ and the fact that for permutation matrices P we have $P^\top P = I$ and $|P| = \pm 1$. It thus remains to show that they satisfy the consistency condition regarding marginalization.

This means that we have to show that for $k \in J$ the distributions F_J and $F_{J\setminus\{k\}}$ agree for products of Borel sets B_i if we choose $B_k = \mathbb{R}$. The distribution F_J is the same as the distribution of an affine transformation of a standard Gaussian. That is, we can take f_J to be the distribution of $\mathbf{X}_J := A\mathbf{Z} + b$ where $\mathbf{Z} \sim N_m(0, I)$. Thus $f_J(\prod_{i\in J} B_i) = P(\mathbf{X}_J \in \prod_{i\in J} B_i)$, so that if we restrict to the case $B_k = \mathbb{R}$ this is the same as the distribution of $\mathbf{X}_J$ minus the k'th component. This stochastic vector can be written as $\mathbf{X}_{J\setminus\{k\}} = \bar{A}\mathbf{Z} + \bar{\mu}$ where $\bar{A}$ is obtained from A by removing the k'th row and similarly for $\bar{\mu}$. It follows that $\mathbf{X}_{J\setminus\{k\}}$ is multivariate Gaussian with mean vector given by μ_J except that the k'th component is removed, and with covariance matrix given by C_J where both the k'th row and column are removed. This can be seen by considering the identities

$$\mathbb{E}[\bar{A}\mathbf{Z} + \bar{\mu}] = \bar{\mu}, \quad \mathbb{V}\mathrm{ar}[\bar{A}\mathbf{Z} + \bar{\mu}] = \bar{A}\bar{A}^\top$$

In other words, $\mathbf{X}_{J\setminus\{k\}}$ has distribution given by $F_{J\setminus\{k\}}$. From this it follows that the marginalization condition holds. Existence then follows by Kolomogorov's extension theorem.

It thus remains to consider the uniqueness proposition. Assume that $\tilde{\mathbf{X}}$ is another Gaussian process with the same mean and covariance function. Then the finite dimensional distributions of $\tilde{\mathbf{X}}$ and $\mathbf{X}$ agree since they are Gaussian (and thus determined by mean vectors and covariance matrices which are determined by the mean and covariance functions). The fact that $P_{\mathbf{X}}$ and $P_{\tilde{\mathbf{X}}}$ are equal then immediately follows from Theorem 10.

□

Lemma 23 gives that the distribution of a Gaussian random field is uniquely specified by its mean and covariance functions, and furthermore, for given functions μ and C, there exists a Gaussian random field with those functions as its mean and covariance functions. In this

sense, in order to specify Gaussian random fields, it suffices to know the mean and covariance functions. However, in practical applications, it is not entirely obvious that we will know a priori which mean and covariance functions will suit a given data set. Indeed, a covariance function could in principle be any positive definite function, and it is not easy to argue a priori that one of these functions is better than another. To simplify things, a common assumption that is made is that the mean function is the constant function $\mu(i) \equiv \mu_0$ for all i and that $C(i,j)$ only depends on the difference i,j. That is, there exists a function $\tilde{C}\colon \mathbb{Z}^n \to \mathbb{R}$ such that $C(i,j) = \tilde{C}(i-j)$. In general, a random field $\mathbf{X}$ with mean and covariance function satisfying these properties is known as a weakly stationary random field. One can then estimate the covariance function empirically or model it further using parametric families of functions. A commonly used parametric model is the Matérn model. We will not consider such matters further.

We can characterize this approach by the fact that we model the mean and covariance functions directly. As discussed, it may not be the case that we have any strong opinions a priori about the covariance function. However, we may have some reasonable idea about the conditional distributions $\mathbf{X}_i \mid \mathbf{X}_{-i}$. If that is the case it is desirable to use these ideas in order to specify a model. In less formal terms, we may have indirect knowledge about the mean and covariance functions. An indirect approach to modelling a random field is the content of Section 3.3.

1.3.2 Independence

If $\mathbf{X}$ is a Gaussian process in $\mathbb{R}^{\mathbb{Z}^n}$ with mean function μ and covariance function C, we write $\mathbf{X} \sim N_{\mathbb{Z}^n}(\mu, C)$ or just $\mathbf{X} \sim N(\mu, C)$. The components of $\mathbf{X}$ are independent of all other components if C has the property that $i \neq j \implies C(i,j) = 0$.

Proposition* 25. *Let $\mathbf{X} \sim N(\mu, C)$ where $i \neq j \implies C(i,j) = 0$. Then for any collection of Borel sets $(B_i)_{i\in\mathbb{Z}^n}$ we have*

$$P(\mathbf{X} \in \prod_{i\in\mathbb{Z}^n} B_i) = \prod_{i\in\mathbb{Z}^n} P(\mathbf{X}_i \in B_i).$$

Proof. Since $\mathbb{Z}^n$ is countable there exists a bijection $f\colon \mathbb{N} \to \mathbb{Z}^n$. Thus every components of $\mathbf{X}$ are contained once in the sequence $(\mathbf{X}_{f(n)})_{n\in\mathbb{N}}$. The sets $D_N := \{\omega \in \Omega(\mathbf{x}_{f(n)}(\omega))_{n\leq N} \in \prod_{n=1}^N B_{f(n)}\}$ for $N \in \mathbb{N}$ are decreasing in the sense that $D_{N+1} \subset D_N$ for all N. Thus, continuity of measures gives

$$\begin{aligned} P((\mathbf{X}_{f(n)})_{n\in\mathbb{N}} \in \prod_{n=1}^{\infty} B_{f(n)}) &= P(\lim_N D_N) \\ &= \lim_N P(D_n) \\ &= \lim_N \prod_{n=1}^{N} P(\mathbf{X}_{f(n)} \in B_n) \\ &= \prod_{n=1}^{\infty} P(\mathbf{X}_{f(n)} \in B_n). \end{aligned}$$

The result thus follows if we can show $P(\mathbf{X} \in \prod_{i\in\mathbb{Z}^n} B_i) = P((\mathbf{X}_{f(n)})_{n\in\mathbb{N}} \in \prod_{n=1}^{\infty} B_{f(n)})$ and $\prod_{n=1}^{\infty} P(\mathbf{X}_{f(n)} \in B_{f(n)}) = \prod_{i\in\mathbb{Z}^n} P(\mathbf{X}_i \in B_i)$. The former follows since as sets in Ω,

$$\{\mathbf{X} \in \prod_{i\in J} B_i\} = \{(\mathbf{X}_{f(n)})_{n\in\mathbb{N}} \in \prod_{n=1}^{\infty} B_{f(n)}\}.$$

The latter depends on what we mean by an infinite product over the set $\mathbb{Z}^n$. This problem is considered in Appendix A.2 $\square$

Just as in the case of Gaussian vectors in $\mathbb{R}^n$ we have a special Gaussian process which we will call the standard Gaussian process on $\mathbb{Z}^n$. A Gaussian process is said to be a standard Gaussian process on $\mathbb{R}^{\mathbb{Z}^n}$ if all its finite dimensional distributions are standard Gaussians in $\mathbb{R}^k$ (for the appropiate k). Equivalently, $\mathbf{X}$ is standard Gaussian if $\mathbf{X} \sim N(0, I)$ where $I(i,j) = 1$ if and only if $i = j$ and 0 otherwise. We then get

$$P(\mathbf{X} \in \prod_{i \in \mathbb{Z}^n} B_i) = \prod_{i \in \mathbb{Z}^n} P(\mathbf{X}_i \in B_i)$$

for many choices of $(B_i)_{i \in \mathbb{Z}^n}$, we have $P(\mathbf{X} \in \prod_{i \in J} B_i) = 0$. Indeed, as a direct corollary to Lemma 86 in the appendix, if $P(\mathbf{X} \in \prod_{i \in J} B_i) > 0$, then for any bijection $b \colon \mathbb{N} \to \mathbb{Z}^n$, $P(\mathbf{X}_{b(n)} \in B_{b(n)})$ must converge to 1 as $n \to \infty$.

In practice this is unlikely to be of much importance, since it is a rare occasion where we need to put strong bounds B_i on *every* component of a stochastic process $\mathbf{X} = (\mathbf{X}_i)_{i \in \mathbb{Z}^n}$. Indeed, in practical applications, where we only have finitely many measurements, some finite dimensional distribution $\mathbf{X}_J$ will be of central importance, whereas the remaining components of $\mathbf{X}$ are merely a helpful theoretical abstraction.

Chapter 2

Conditional Distributions

In this chapter we introduce some preliminary theory about conditional probability. We start with the notion of conditional distributions for measurable maps $\mathbf{X}$ which has a domain a probability space. Conditional distributions are directly related to conditional expectations. These distributions have some undesirable properties. For this reason, one is therefore interested in defining the so-called *regular* conditional distributions which as the name implies is a sort of regular version of conditional distributions. Regular conditional distributions can be hard to interpret. Thankfully, in some cases they induce so-called 'deterministic' conditional distributions which one may find to be a little more tangible. We consider these concepts in this chapter.

2.1 Conditional Distributions

We intend to give meaning to symbols such as $P_{\mathbf{X}}(\cdot \mid \mathbf{X}_{-i} = x_{-i})$ which are supposed to be something like a probability measure for $\mathbf{X}$ where we know about $\mathbf{X}_{-i} = x_{-i}$. In finite dimensions we can use e.g. the Lebesgue measure to define such things. In the fully general, measure-theoretic case, such an approach is not possible. We thus have to approach the problem differently.

Recall the basic fact from measure theory that $P(A) = \mathbb{E}[\mathbb{I}_A]$ when expectation is taken wrt. P. We will connect conditional probabilities with conditional expectations in a similar fashion. We start with the definition of conditional expectation.

Definition 26. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $\mathbf{X}\colon \Omega \to \mathbb{R}$ be $\mathcal{F}$-measurable and $\mathcal{L}^1(\mu)$. Let $\mathcal{G} \subset \mathcal{F}$ be a (sub)-σ-algebra. Then we say that $\mathbf{Z}\colon \Omega \to \mathbb{R}$ is a conditional expectation of $\mathbf{X}$ given $\mathcal{G}$ if*

- *$\mathbf{Z}$ is $\mathcal{G}$ measurable.*
- *For any $B \in \mathcal{G}$ we have*
$$\int_B \mathbf{Z} d\mu = \int_B \mathbf{X} d\mu.$$

Note that $B_1 = \{\omega \in \Omega\colon \mathbf{Z}(\omega) \geq 0\}$ and $B_2 = \Omega \setminus B_1$ are measurable sets. It follows that the conditional expectation of $\mathbf{X}$ given $\mathcal{G}$ is determined μ almost everywhere. We are therefore somewhat justified in saying that a function is *the* conditional expectation of $\mathbf{X}$ and denoting it by $\mathbb{E}[\mathbf{X} \mid \mathcal{G}]$.

Lemma 27 (Existence of conditional expectations)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $\mathbf{X}\colon \Omega \to \mathbb{R}$ from be $\mathcal{F}$ measurable and $\mathcal{L}^1(\mu)$. Assume that μ is σ-finite on $\mathcal{G}$. Then the conditional expectation of $\mathbf{X}$ given $\mathcal{G}$ exists.*

Proof. Define the positive and negative parts of $\mathbf{X}$ as the functions $\mathbf{X}^+ = \mathbb{I}[\mathbf{X} \geq 0]\mathbf{X}$ and $\mathbf{X}^- = -\mathbb{I}[\mathbf{X} < 0]\mathbf{X}$ and note that they are measurable, non-negative, integrable and satisfy $\mathbf{X} = \mathbf{X}^+ - \mathbf{X}^-$. Define the measure ν^+ on $\mathcal{G}$ by letting

$$\nu^+(G) = \int_G \mathbf{X}^+ d\mu.$$

Then ν^+ is finite and absolutely continuous wrt. the restriction of μ to $\mathcal{G}$, $\mu_{|\mathcal{G}}$. Since $\mu_{|\mathcal{G}}$ is σ-finite on $\mathcal{G}$. The Radon-Nikodym theorem implies that a $\mathcal{G}$ measurable function $\mathbf{Z}^+$ exists such that

$$\nu^+(G) = \int_G \mathbf{Z}^+ d\mu.$$

We can similarly define a measure ν^- using the negative part of $\mathbf{X}$ to find a $\mathcal{G}$ measurable function $\mathbf{Z}^-$. Define $\mathbf{Z} = \mathbf{Z}^+ - \mathbf{Z}^-$ and note that it is $\mathcal{G}$ measurable, integrable and linearity of the Lebesgue integral gives

$$\int_G \mathbf{Z} d\mu = \int_G \mathbf{X} d\mu, \quad G \in \mathcal{G}.$$

□

Assume that $\mathbf{Y} \colon \Omega \to \mathbb{Y}$ is a $\mathcal{F} - \mathcal{Y}$ measurable function, where $\mathcal{Y}$ is a σ-algebra on $\mathbb{Y}$. We can then define the σ-algebra generated by $\mathbf{Y}$ as the

$$\sigma(\mathbf{Y}) := \sigma(\{\mathbf{Y}^{-1}(B) \mid B \in \mathcal{Y}\}).$$

We can then condition on this σ-algebra to obtain conditional expectations. Notationally we write this as $\mathbb{E}[\mathbf{X} \mid \mathbf{Y}] := \mathbb{E}[\mathbf{X} \mid \sigma(\mathbf{Y})]$.

Definition 28. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\mathbf{X} \colon \Omega \to \mathbb{X}$ be a random variable onto a space $(\mathbb{X}, \mathcal{X})$. Let $\mathcal{G} \subset \mathcal{F}$ be a $\sigma-$algebra. Then a conditional probability for $\mathbf{X}$ given $\mathcal{G}$ is a function $\gamma \colon \mathcal{X} \times \Omega \to \mathbb{R}$ such that for all $A \in \mathcal{X}$*

$$\gamma(A, \cdot) = \mathbb{E}[\mathbb{I}_{\mathbf{X} \in A} \mid \mathcal{G}].$$

If $\mathbf{X}$ is the identity map and $(\mathbb{X}, \mathcal{X}) = (\Omega, \mathcal{F})$ we say that such a function is a conditional distribution for P given $\mathcal{G}$.

Remark 29. *The equation simply means that $\gamma(A, \cdot)$ is a version of the conditional expectation $\mathbb{E}[\mathbb{I}_{\mathbf{X} \in A} \mid \mathcal{G}]$. That is, $\gamma(A, \cdot)$ is some $\mathcal{G}$ measurable map satisfying $\int_G \gamma(A, \cdot) dP = \int \mathbb{I}_{\mathbf{X} \in A} dP = P(\mathbf{X} \in A)$.*

Clearly, conditional probabilities for P given $\mathcal{G}$ are not uniquely defined since they can be changed on null-sets. One conditional expectation for $\mathbf{X}$ given $\mathcal{G}$ exists, since we can define $\gamma(A, \omega)$ using Radon-Nikodym derivatives. In the following we will investigate properties of this function.

Theorem* 30. *Let $(\Omega, \mathcal{F}, P)$, $\mathbf{X} \colon \Omega \to \mathbb{X}$ and $\mathcal{G} \subset \mathcal{F}$ be as above and define $\gamma(A, \omega) = \mathbb{E}[\mathbb{I}_{\mathbf{X} \in A} \mid \mathcal{G}]$. Then*

1. *For every $A \in \mathcal{X}$ the map $\omega \mapsto \gamma(A, \omega)$ is $\mathcal{G}$ measurable.*
2. *For all A, $0 \leq \gamma(A, \cdot) \leq 1$ P-a.e. Furthermore, for P-a.a. ω $\gamma(\emptyset, \cdot) = 0$ and $\gamma(\Omega, \cdot) = 1$.*
3. *Let (A_n) be a sequence of disjoint measurable sets. Then there exists some set B with $P(B) = 1$ such that for P-a.a. ω we have*

$$\sum \mathbb{I}_B \gamma(A_n, \cdot) = \gamma(\sqcup A_n, \cdot)$$

4. *For $A \in \mathcal{X}$ we have*

$$P(\{\mathbf{X} \in A\} \cap B) = \int_B \gamma(A, \cdot) dP$$

The proof of this claim is neither particularly interesting nor pretty and has therefore been placed in the appendix.

2.2 Regular Conditional Distributions

Theorem 30 shows that the map $A \mapsto \gamma(A \mid \mathcal{G})$ is *almost* a probability measure. This is a problem that can not be handwaved away by saying that it is a probability measure outside some null set. Indeed the set in Ω for which

$$\sum \gamma(A_n, \cdot) = \gamma(\bigcup A_n, \cdot)$$

fails to be true for disjoint sets (A_n) might depend on the particular sequence (A_n). The set $\{\omega \colon A \mapsto \gamma(A, \omega) \text{ is not } \sigma\text{-additive}\}$ is therefore the union of many null-sets, possible uncountably many and so may not be a null-set at all. It is rather dissatisfying that we can not ensure that it is a probability measure for all ω since there is then no certainty that $A \mapsto \gamma(A \mid \mathcal{G})(\omega)$ has any practical meaning in a probabilistic context. We therefore introduce the notion of *regular* conditional distributions.

Definition 31. *Let $(\Omega, \mathcal{F}, P)$ be probability space and let $\mathbf{X} \colon \Omega \to \mathbb{X}$ be a random variable onto a measurable space $(\mathbb{X}, \mathcal{X})$. Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra. Then we say that $\gamma \colon \mathcal{X} \times \Omega \to \mathbb{R}$ is a regular conditional distribution for $\mathbf{X}$ given $\mathcal{G}$ if*

1. *For any $A \in \mathcal{X}$, the map $\omega \mapsto \gamma(A, \omega)$ is $\mathcal{G}$ measurable.*
2. *For* all *$\omega \in \Omega$, the map $A \mapsto \gamma(A, \omega)$ is a probability measure on $(\mathbb{X}, \mathcal{X})$*
3. *For all $B \in \mathcal{G}$ we have*

$$P(\{\mathbf{X} \in A\} \cap B\}) = \int_B \gamma(A, \cdot) dP$$

In this case we denote γ by $P_{\mathbf{X}}(\cdot \mid \mathcal{G})$ or $P(\mathbf{X} \in \cdot \mid \mathcal{G})$. If $\mathbf{X}$ is the identity map and $(\mathbb{X}, \mathcal{X}) = (\Omega, \mathcal{F})$ then we call the function γ a regular conditional distribution for P given $\mathcal{G}$. In this case we denote γ by $P(\cdot \mid \mathcal{G})$.

As already noted (non-regular) conditional distributions exist by the existence of Radon-Nikodym theorem. For regular conditional distributions we have a non-trivial existence problem. We will now consider this problem.

Assume we have the situation that $\mathbf{X} \colon \Omega \to \mathbb{X}$ has a regular conditional probability $P_{\mathbf{X}}(\cdot \mid \mathcal{G})$. Assume that $\mathbb{X}$ is Borel isomorphic to some other measurable space $\mathbb{Y}$ with ϕ the isomorphism. It is then perhaps not much of a surprise that it possible to use the regular conditional probability for $\mathbf{X}$ to construct a regular conditional probability for $\phi \circ \mathbf{X}$. More generally, we do not need ϕ to be an isomorphism, it suffices for ϕ to be measurable.

Lemma* 32. *Assume that $\mathbf{X} \colon \Omega \to \mathbb{X}$ be a random variable which has a regular conditional distribution $P_{\mathbf{X}}(\cdot \mid \mathcal{G})$, and let $\phi \colon \mathbb{X} \to \mathbb{Y}$ be a measurable onto a measurable space $(\mathbb{Y}, \mathcal{Y})$. Then the map $\gamma \colon \mathcal{Y} \times \Omega \to \mathbb{R}$ given by $\gamma(A, \omega) = P_{\mathbf{X}}(\phi^{-1}(A) \mid \mathcal{G})(\omega)$ is a regular conditional distribution for $\phi \circ \mathbf{X}$ given $\mathcal{G}$.*

Now in the context of Polish spaces $\mathbb{X}, \mathbb{Y}$ we have that $\mathbb{X}$ and $\mathbb{Y}$ are Borel isomorphic if $|\mathbb{X}| = |\mathbb{Y}|$. This is the content of Theorem 20. Let us assume that $\mathbf{X}\colon \Omega \to \mathbb{X}$ where $\mathbb{X}$ is an uncountable Polish space which is the case for for random fields on $\mathbb{Z}^n$. We then have that $\mathbb{X}$ is Borel isomorphic to the interval $[0,1] \subset \mathbb{R}$ equipped with the (sub)-Borel algebra with ϕ the isomorphism.

To show that $\mathbf{X}$ has a regular conditional probability it suffices to first construct one for $\phi \circ \mathbf{X}\colon \Omega \to [0,1]$. Indeed, using Lemma 32 we see that for $A \subset \mathbb{X}$ a Borel set and $\omega \in \Omega$, the map $(A, \omega) \mapsto P_{\phi\circ\mathbf{X}}(\phi(A) \mid \mathcal{G})(\omega)$ defines a regular conditional distribution for $\phi^{-1} \circ \phi \circ \mathbf{X} = \mathbf{X}$. This is useful since constructing regular conditional distributions for maps into $[0,1]$ can be aided by the structure that $[0,1]$ is known to have.

This approach to constructing regular conditional probabilities for measurable maps into Polish spaces can be found in several sources. This is the case in e.g. [16]. The proof in [16], however, fails to show existence of regular conditional probabilities as we have defined it here. Indeed, in e.g. [16] a regular conditional probability for $\mathbf{X}$ given $\mathcal{G}$ is defined as maps $\tilde{\gamma}\colon \mathcal{X} \times \Omega \to \mathbb{R}$ such that for all $\omega \in \Omega$, $\tilde{\gamma}(\cdot, \omega)$ is a probability measure, and $\tilde{\gamma}(A, \omega) = \mathbb{E}[\mathbb{I}_{\mathbf{X}\in A} \mid \mathcal{G}]$ P-a.s. If we define regular conditional probabilities this way, the proof works. A detailed version of the proof can be found in the Appendix, Lemma 96.

This is a weaker definition though. Indeed, there is no certainty that a map $\tilde{\gamma}$ equal to $\mathbb{E}[\mathbb{I}_{\mathbf{X}\in A} \mid \mathcal{G}]$ P-a.s. is $\mathcal{G}$-measurable. Indeed, if we can find a set $\mathcal{N}$ which is a P null-set and a value $x \notin \text{range}(\tilde{\gamma}(A, \cdot))$ we can define the function

$$\psi(A, \omega) = \mathbb{I}_{\mathcal{N}^c}\tilde{\gamma}(A, \omega) + \mathbb{I}_{\mathcal{N}}x$$

which is equal to $\gamma(A, \cdot)$ P-a.s. and thus also to $\mathbb{E}[\mathbb{I}_{\mathbf{X}\in A} \mid \mathcal{G}]$ but it is *not* $\mathcal{G}$-measurable.

We will therefore not use this definition here, and when we use the term "regular conditional distribution" we mean it in the sense of Definition 31. As it turns out, a regular conditional distribution for a random variable $\mathbf{X}\colon \Omega \to \mathbb{X}$ mapping into Polish spaces *can* be constructed. This can be done as in e.g. [17].

Theorem 33. *Let $\mathbf{X}\colon \Omega \to \mathbb{X}$ be measurable, where $(\Omega, \mathcal{F}, P)$ is a probability space and $\mathbb{X}$ is a Polish space. Then for any σ-algebra $\mathcal{G} \subset \mathcal{F}$ a conditional probability distribution for $\mathbf{X}$ given $\mathcal{G}$ exists.*

Let $\mathbf{X}\colon \Omega \to \mathbb{R}^{\mathbb{Z}^n}$ is a Gaussian random field. Then the range of $\mathbf{X}$ is a Polish space such that for any finite set of indices $J \subset \mathbb{Z}^n$ we have that a regular conditional probability for $\mathbf{X}$ given $\sigma(\mathbf{X}_J)$ exists. We denote such a regular conditional probability by $P(\mathbf{X} \in \cdot \mid \mathbf{X}_J)$.

In elementary probability theory, expectation of a random variable $\mathbf{X}\colon \Omega \to \mathbb{R}$ is often defined in terms of integrals such as e.g. $\mathbb{E}[\mathbf{X}] = \int \mathbf{X} dP$. Now that we have defined conditional probabilities it is of interest whether we can relate them to conditional expectations in a similar way.

Proposition* 34. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\mathcal{G}$ be a σ-algebra contained in $\mathcal{F}$. Assume that a regular conditional probability for P wrt. to $\mathcal{G} \subset$ exists. Then if $\mathbf{X}\colon \Omega \to \mathbb{R}$ is integrable wrt. $P(\cdot \mid \mathcal{G})(\omega)$ for all ω then*

$$\mathbb{E}[\mathbf{X} \mid \mathcal{G}] = \int \mathbf{X} dP(\cdot \mid \mathcal{G}) \quad P\text{-a.s.} \tag{2.1}$$

Proof. The proof follows straightforwardly from the standard argument. Indeed, take $\mathbf{X} = \mathbb{I}_A$ for some set $A \in \mathcal{F}$. Then

$$\int \mathbb{I}_A dP(\cdot \mid \mathcal{G})(\omega) = P(A \mid \mathcal{G})(\omega) = \mathbb{E}[\mathbb{I}_A \mid \mathcal{G}](\omega).$$

Now if $\mathbf{X}$ is simple $\mathbf{X} = \sum_i^n \alpha_i \mathbb{I}_{A_i}$ then

$$\int \mathbf{X} dP(\cdot \mid \mathcal{G}) = \sum_i^n \alpha_i \mathbb{E}[\mathbb{I}_{A_i} \mid \mathcal{G}] = \mathbb{E}[\mathbf{X} \mid \mathcal{G}] \quad P\text{-a.s.}$$

by almost sure linearity of conditional expectation. We also have a monotone convergence theorem for conditional expectations in that if $\mathbf{X}_n \nearrow \mathbf{X}$ then $\mathbb{E}[\mathbf{X}_n \mid \mathcal{G}] \nearrow \mathbb{E}[\mathbf{X} \mid \mathcal{G}]$ P-a.s. Using this and monotone convergence for integrals we get the result for $\mathbf{X}$ measurable and positive if we write $\mathbf{X}$ as the pointwise limit of simple functions $\mathbf{X}_n$. □

2.3 Probabilities of the Form $P(\mathbf{X} \in A \mid \mathbf{Y} = y)$

In the previous section we made sense of $P(\mathbf{X} \in A \mid \mathcal{G})$. This can be interpreted as a kind of probability measure given the information in $\mathcal{G}$. This is of course a rather vague notion. In this section we will define $P(\mathbf{X} \in A \mid \mathbf{Y} = y)$. On an informal level this is the probability of $\mathbf{X}$ being in A given that we know $\mathbf{Y} = y$ and should thus be a real number.

In terms of probability theory we are often interested in the codomain of random variables; the domain Ω is merely a useful abstraction. It is thus of interest whether we can decompose $f := P(\mathbf{X} \in A \mid \mathbf{X}_J)$ into two functions as $f = \phi \circ \mathbf{X}_J$ where $\phi \colon \text{range}(\mathbf{X}_J) \to \mathbb{R}$ is some function we are interested in finding. I.e. we seek a function ϕ such that the following diagram commutes:

$$\begin{array}{ccc} \Omega & \xrightarrow{\mathbf{X}} & \mathbb{X} \\ & \searrow^{f} & \downarrow{\scriptstyle \phi} \\ & & \mathbb{R} \end{array} \tag{2.2}$$

This is possible as the following lemma shows.

Lemma* 35. *Let $(\Omega, \mathcal{F}), (\mathbb{X}, \mathcal{Y}), (\mathbb{X}, \mathcal{X})$ be a measurable spaces and $\mathbf{X} \colon \Omega \to \mathbb{X}$ and $\mathbf{Y} \colon \Omega \to \mathbb{Y}$. Assume that $\mathcal{X}$ and $\mathcal{Y}$ contain all singletons $\{x\}, \{y\}$ and that $\mathbf{X}$ is surjective. If $\mathbf{Y}$ is $\sigma(\mathbf{X})$ measurable then there exists a measurable function $\phi \colon \mathbb{X} \to \mathbb{Y}$ such that $\mathbf{Y} = \phi \circ \mathbf{X}$.*

Remark 36. *Let $f \colon \mathbb{A} \to \mathbb{B}$ be any function between sets $\mathbb{A}$ and $\mathbb{B}$. Then for subsets A of $\mathbb{A}$ we denote*

$$f(A) := \{f(a) \colon a \in A\}.$$

We will need to move sets around in this way using many different functions in the proof below. Then, if g has domain equal to $\mathbb{B}$, we denote

$$g \circ f(A) = g(f(A))$$

to make notation cleaner.

Proof. We prove that $\mathbf{Y}$ is constant on sets where $\mathbf{X}$ is constant. Then we can define ϕ as the map from $x \in \mathbb{X}$ to the unique element in $\mathbf{Y}(\mathbf{X}^{-1}(\{x\}))$. We start by making some observations about the σ-algebra $\sigma(\mathbf{X})$.

Let $A \in \sigma(\mathbf{X})$. By definition we have $A = \mathbf{X}^{-1}(B)$ for some $B \in \mathcal{X}$. Define $A_x := \mathbf{X}^{-1}(\{x\})$ then we can decompose A as

$$A = \mathbf{X}^{-1}(B) = \bigcup_{x \in B} A_x.$$

By surjectivity the A_x are non-empty and satisfy $A_x \cap A_y \neq \emptyset \implies A_x = A_y$. Combine these facts to get that for any $A \in \sigma(\mathbf{X})$ we have the implication $A \cap A_x \neq \emptyset \implies A \cap A_x = A_x$.

Let $y \in \mathbf{Y}(A_x)$. Then clearly $\mathbf{Y}^{-1}(\{y\}) \cap A_x \neq \emptyset$ so that $\mathbf{Y}^{-1}(\{y\}) \cap A_x = A_x$. Applying $\mathbf{Y}$ to this equality directly yields $\{y\} = \mathbf{Y}(A_x)$. This shows that $\mathbf{Y}$ is constant on all sets which $\mathbf{X}$ is constant on. We can then define $\phi(x)$ as the unique element in $\mathbf{Y}(A_x)$. Thus for all $\omega \in A_x$ we get $\mathbf{Y}(\omega) = \phi(x) = \phi(\mathbf{X}(\omega))$. This holds everywhere so that $\mathbf{Y} = \phi \circ \mathbf{X}$.

To prove measurability of ϕ, let $C \in \mathcal{Y}$. Then using surjectivity of $\mathbf{X}$ we get

$$\phi^{-1}(C) = \mathbf{X} \circ \mathbf{X}^{-1} \circ \phi^{-1}(C) = \mathbf{X} \circ \mathbf{Y}^{-1}(C)$$

The "$\circ$" should not be interpreted as implying that $\mathbf{X} \circ \mathbf{X}^{-1} \circ \phi^{-1}$ is a function. Since $\mathbf{Y}$ is $\sigma(\mathbf{X})$ measurable we have $\mathbf{Y}^{-1}(C) = A \in \sigma(\mathbf{X})$ so that $A = \mathbf{X}^{-1}(B)$ for some $B \in \mathcal{X}$. But then

$$\phi^{-1}(C) = \mathbf{X} \circ \mathbf{X}^{-1}(B) = B \in \mathcal{X}.$$

□

Note that we can always assume that $\mathbf{X}$ is surjective by restricting the codomain to the range of $\mathbf{X}$ and endowing it with the natural subspace σ-algebra consisting of sets $\text{range}(\mathbf{X}) \cap B$ for B measurable in $\mathbb{X}$.

With the concrete choices we have in this project $\mathbf{X}$ will be a random field mapping into $\mathbb{R}^{\mathbb{Z}^n}$. We defined the σ-algebra on $\mathbb{R}^{\mathbb{Z}^n}$ as the least σ-algebra such that all projections $\pi_i \colon \mathbb{R}^{\mathbb{Z}^n} \to \mathbb{R}$, $i \in \mathbb{Z}^n$ are measurable. Since all singletons in $\mathbb{R}$ are measurable we see $\mathbb{B}(\mathbb{R}^{\mathbb{Z}^n})$ contains sets of the form $\prod_{j \in \mathbb{Z}^n} \mathcal{O}_j$ where $\mathcal{O}_j = \mathbb{R}$ for all $j \neq i$ and $\mathcal{O}_i = \{x_i\}$ for some $x_i \in \mathbb{R}$. Since we can take countable intersections and $\mathbb{Z}^n$ is countable it follows that all singletons $\prod_{i \in \mathbb{Z}} \{x_i\}$ are measurable in $\mathbb{R}^{\mathbb{Z}^n}$. Thus the maps $\mathbf{X}_J$ and $P(\mathbf{X} \in A \mid \mathbf{X}_J)$ for any subset $J \subset \mathbb{Z}^n$ satisfy the hypotheses of Lemma 35 so that we can find a function $\phi_A \colon \mathbb{R}^{\mathbb{Z}^n} \to \mathbb{R}$ such that

$$P(\mathbf{X} \in A \mid \mathbf{X}_J) = \phi_A \circ \mathbf{X}_J. \tag{2.3}$$

Furthermore we can define $\phi \colon \mathcal{F} \times \mathbb{R}^{\mathbb{Z}^n} \to \mathbb{R}$ by $\phi(A, x_J) = \phi_A(x_J)$.

Definition 37. *Let $\mathbf{X} \colon \Omega \to \mathbb{R}^{\mathbb{Z}^n}$ be a Gaussian random field. Then we define*

$$P(\mathbf{X} \in A \mid \mathbf{X}_J = x_J) := \phi_A(x_J)$$

where $\phi_A \colon \text{range}(\mathbf{X}_J) \to \mathbb{R}$ is the measurable function solving $P(\mathbf{X} \in A \mid \mathbf{X}_J) = \phi_A \circ \mathbf{X}_J$.

We will henceforth use the $P(\mathbf{X} \in A \mid \mathbf{X}_J = x_j)$ notation. We denote the function $x_J \mapsto P(\mathbf{X} \in A \mid \mathbf{X}_J = x_j)$ by $P(\mathbf{X} \in A \mid \mathbf{X}_J = \cdot)$. This function is sometimes known as a deterministic conditional probability. It is not much of a surprise that this function 'inherits' properties of regular conditional probabilities.

Proposition 38. *Let $\mathbf{X} \colon \Omega \to \mathbb{R}^{\mathbb{Z}^n}$ be a Gaussian random field. Then for any subset $J \subset \mathbb{Z}^n$*

1. *The map $x_J \mapsto P(\mathbf{X} \in A \mid \mathbf{X}_J = x_J)$ is $\mathbb{B}(\mathbb{R}^J)$ measurable.*
2. *The map $A \mapsto P(\mathbf{X} \in A \mid \mathbf{X}_J = x_J)$ is a probability measure for all $x_J \in \text{range}(\mathbf{X}_J)$.*
3. *For any $B \in \mathbb{B}(\mathbb{R}^J)$ we have*

$$\int_B P(\mathbf{X} \in A \mid \mathbf{X}_J = \cdot) dP_{\mathbf{X}_J} = P(\mathbf{X}^{-1}(A) \cap \mathbf{X}_J^{-1}(B)).$$

Proof. The first part follows by definition of $P(\mathbf{X} \in A \mid \mathbf{X}_J = \cdot)$. The second follows since for every x_j we can find an $\omega \in \Omega$ such that $P(\mathbf{X} \in A \mid \mathbf{X}_J = x_j) = P(\mathbf{X} \in A \mid \mathbf{X}_J)(\omega)$. For the third part use ϕ_A as shorthand for $P(\mathbf{X} \in A \mid \mathbf{X}_j = \cdot)$. Then

$$\begin{aligned}\int_B P(\mathbf{X} \in A \mid \mathbf{X}_J = \cdot) dP_{\mathbf{X}_J} &= \int_B \phi_A dP_{\mathbf{X}_J} \\ &= \int_{\mathbf{X}_J^{-1}(B)} \phi_A \circ \mathbf{X}_J dP \\ &= \int_{\mathbf{X}_J^{-1}(B)} P(\mathbf{X} \in A \mid \mathbf{X}_J) dP \\ &= P(\mathbf{X}^{-1}(A) \cap \mathbf{X}_J^{-1}(B))\end{aligned}$$

where we have used Lemma 97 in the Appendix. □

It turns out that the properties 1.-3. are actually sufficient for defining the deterministic function $P(A \mid \mathbf{X} = x)$ in the sense that these three properties determine a function almost everywhere.

Proposition* 39. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\mathbf{X}\colon \Omega \to \mathbb{X}$ be measurable and surjective onto a measurable space $(\mathbb{X}, \mathcal{X})$. Assume that a regular conditional probability for P given $\mathbf{X}$ exists. Furthermore, let $\phi_A\colon \mathbb{X} \to \mathbb{R}$ be any function satisfying*

1. *$\tilde{\phi}_A$ is $\mathcal{X}$-$\mathbb{B}(\mathbb{R})$ measurable.*
2. *For any B we have*
$$P(A \cap \{\mathbf{X} \in B\}) = \int_B \tilde{\phi}_A P_{\mathbf{X}}$$

then $\phi_A = P(A \mid \mathbf{X} = \cdot)$ $P_{\mathbf{X}}$-a.e.

Proof. The proof follows easily once we notice that the sets $B = \{P(A \mid \mathbf{X} = \cdot) \geq \phi_A(\cdot)\}$ and $\{P(A \mid \mathbf{X} = \cdot) < \phi_A(\cdot)\}$ are $\mathcal{X}$ measurable. Then we have e.g.

$$\int_B P(A \mid \mathbf{X} = \cdot) - \tilde{\phi}_A(\cdot) dP_{\mathbf{X}} = P(A \cap \{\mathbf{X} \in B\}) - P(A \cap \{\mathbf{X} \in B\}) = 0.$$

Since $P(A \mid \mathbf{X} = \cdot) - \tilde{\phi}_A(\cdot)$ is a positive function when restricted to B, we get that it is equal to the zero function P-a.s. on B. A similar argument can be made on B^c so that $P(A \mid \mathbf{X} = \cdot) = \tilde{\phi}_A(\cdot)$ P-a.s. □

This uniqueness property shows that we do not need to go through the abstract construction of $P(A \mid \mathbf{X} = X)$ outlined in the proof of Lemma 35 and that it suffices to find (using educated guesswork) a function as described in Proposition 39.

Example 40. *Assume that $(\mathbf{X}_1, \mathbf{X}_2) \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right)$. Well known properties of the Gaussian distributions give that the marginal distribution of $\mathbf{X}_2$ is $N(0,1)$. By Proposition 39 a function ϕ_A is a.e. equal to $P(\mathbf{X}_1 \in A \mid \mathbf{X}_2 = \cdot)$ if it is Borel measurable and solves*

$$P(\mathbf{X}_1 \in A, \mathbf{X}_2 \in B) = \int_B \phi_A dP_{\mathbf{X}_2}.$$

Using basic probability theory as a suggestion we define

$$\phi_A(x_2) = \int_A \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp(-\frac{(x_1 - \rho x_2)^2}{2(1-\rho^2)}) dx_1.$$

Using the fact that $P_{\mathbf{X}_2}$ has density wrt. the one dimensional Lebesgue measure we calculate as

$$\begin{aligned}
\int_B \phi_A dP_{\mathbf{X}_2} &= \int_B \phi_A \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}) dx_2 \\
&= \int_B \int_A \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp(-\frac{(x_1-\rho x_2)^2}{2(1-\rho^2)}) dx_1 \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}) dx_2 \\
&= \int_B \int_A \frac{1}{\sqrt{(2\pi)^2(1-\rho^2)}} \exp\left(-\frac{1}{2(1-\rho^2)}(x_1^2 - 2\rho x_1 x_2 + \rho^2 x_2^2 + (1-\rho^2)x^2)\right) dx_1 dx_2 \\
&= \int_B \int_A \frac{1}{\sqrt{(2\pi)^2(1-\rho^2)}} \exp\left(-\frac{1}{2(1-\rho^2)} \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) dx_1 dx_2 \\
&= \int_B \int_A \frac{1}{\sqrt{(2\pi)^2(1-\rho^2)}} \exp\left(-\frac{1}{2} \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) dx_1 dx_2 \\
&= P(\mathbf{X}_1 \in A, \mathbf{X}_2 \in B)
\end{aligned}$$

Let $\mathbf{X}, \mathbf{Y} \colon \Omega \to \mathbb{R}$ be two random variables with joint density $f \colon \mathbb{R}^2 \to \mathbb{R}$ wrt. the Lebesgue measure on $\mathbb{R}^2$. That is, $P((\mathbf{X}, \mathbf{Y}) \in A) = \int_A f(x,y) d(x,y)$. In light of the example above, it may be particularly easy to find the conditional distribution $P(\mathbf{X} \in A \mid \mathbf{Y} = y)$.

Proposition 41. *Let $\mathbf{X}, \mathbf{Y}$ be random variables with joint density f wrt. the Lebesgue measure. Assume that the marginal density for $\mathbf{Y}$ is non-zero everywhere and that f is continuous. Then we have*

$$P(\mathbf{X} \in A \mid \mathbf{Y} = y) = \int_A \frac{f(x,y)}{f_{\mathbf{Y}}(y)} dx.$$

Proof. Using Proposition 39 we see firstly that we firstly have to argue that the function $y \mapsto \int_A \frac{f(x,y)}{f_{\mathbf{Y}}(y)} dx$ is Borel measurable. This follows by continuity of f. Secondly we calculate

$$\begin{aligned}
\int_B \int_A \frac{f(x,y)}{f_{\mathbf{Y}}(y)} dx dP_{\mathbf{Y}} &= \int_B \left(\int_A \frac{f(x,y)}{f_{\mathbf{Y}}(y)} dx \right) f_{\mathbf{Y}}(y) dy \\
&= P(\mathbf{X} \in A, \mathbf{Y} \in B).
\end{aligned}$$

Hence $\int_A \frac{f(x,y)}{f_{\mathbf{Y}}(y)} dx = P(\mathbf{X} \in A \mid \mathbf{Y} = y)$, $P_{\mathbf{Y}}$-a.s. □

Chapter 3

Conditional Specification of Gaussians

In Chapter 1 we considered three ways of extending the notion of non-degenerate Gaussian vector. With one notable exception in Section 4.1 we described Gaussian measures directly in terms of their covariance matrices (functions) and mean vectors (functions). As noted in the end of 1.3 it may not always be reasonable to assume that Gaussian measures can be specified in this way. In this section we thus seek to dispose of this shortcoming.

3.1 The Finite Dimensional Case

The theory of conditional Gaussian distributions in finite dimensions is fairly well known. We will therefore only consider it briefly.

In several applications it is useful to consider the conditional distribution of a Gaussian vector $\mathbf{X}$ given an observation $\mathbf{Y} = y$ of another Gaussian $\mathbf{Y}$. A simple example of this is to let $\mathbf{Y}$ be a linear transformation of $\mathbf{X}$. We then say that $\mathbf{X} \mid A\mathbf{X} = b$ is a Gaussian vector under a linear constraint.

Theorem 42. *Let* $\mathbf{X} \sim N_n(\mu, \Sigma)$. *Then* $\mathbf{X} \mid A\mathbf{X} = b \sim N_n(\bar{\mu}, \bar{\Sigma})$ *where*

$$\bar{\mu} = \mu + \Sigma A^\top (A\Sigma A^\top)^+ (b - A\mu), \quad \bar{\Sigma} = \Sigma - \Sigma A^\top (A\Sigma A^\top)^+ A\Sigma$$

A proof can be found in [2]. The case described in Theorem 42 may seem rather restrictive, but every case $\mathbf{X} \mid \mathbf{Y} = Y$ can in fact be deduced from this theorem fairly easily. For the general case, set $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$ and let P_1 be the matrix defining the linear map $\mathbf{Z} \mapsto \mathbf{X}$ and let P_2 be the matrix for the map $\mathbf{Z} \mapsto \mathbf{Y}$. Partition the covariance $\Sigma_{\mathbf{Z}}$ of $\mathbf{Z}$ as follows

$$\Sigma_{\mathbf{Z}} = \begin{pmatrix} \Sigma_{\mathbf{X}} & \Sigma_{\mathbf{X}:\mathbf{Y}} \\ \Sigma_{\mathbf{Y}:\mathbf{X}} & \Sigma_{\mathbf{Y}} \end{pmatrix}$$

where e.g. $\Sigma_{\mathbf{X}}$ is the covariance of $\mathbf{X}$ and $\Sigma_{\mathbf{X}:\mathbf{Y}} = \mathrm{Cov}[\mathbf{X}, \mathbf{Y}]$. We can similarly partition the mean of $\mathbf{Z}$ into $\mathbf{X}$ and $\mathbf{Y}$ components. Then $\mathbf{X} \mid \mathbf{Y} = Y$ is exactly the same as $P_1\mathbf{Z} \mid P_2\mathbf{Z} = y$. Apply Theorem 42 to get the conditional mean and covariances:

$$\bar{\mu} = \mu_{\mathbf{X}} + \Sigma_{\mathbf{X}:\mathbf{Y}} \Sigma_{\mathbf{Y}}^+ (y - \mu_{\mathbf{Y}}), \quad \bar{\Sigma} = \Sigma_{\mathbf{X}} - \Sigma_{\mathbf{X}:\mathbf{Y}} \Sigma_{\mathbf{Y}}^+ \Sigma_{\mathbf{Y}:\mathbf{X}} \tag{3.1}$$

3.1.1 Specification through local characteristics in $\mathbb{R}^n$

In the following we will need to specify subsets of components of random vectors $\mathbf{X} = (\mathbf{X}_i)$. The vector $\mathbf{X}_{-i}$ consists of all components except the i'th. Similarly, $\mathbf{X}_{-ij}$ are all components except the i'th and j'th. For a set of indices J we let $\mathbf{X}_J$ be a vector consisting of all components with indices in J. This convention also works in $\mathbb{R}^{\mathbb{Z}^n}$ if we agree that $\mathbf{X} \in \mathbb{R}^{\mathbb{Z}^n}$ is a function from $\mathbb{Z}^n$ to $\mathbb{R}$. Then $\mathbf{X}_J$ is just the restriction of that function to $J \subset \mathbb{Z}^n$.

Let $\mathbf{X} \sim N_n(\mu, \Sigma)$ and consider the conditional distributions of the following form $\mathbf{X}_i \mid \mathbf{X}_{-i} = x_{-i}$ where x_{-i} is some realization of $\mathbf{X}_{-i}$. These conditional distributions are known as the local characteristics. In the case that $\mathbf{X}$ is non-degenerate with $Q = \Sigma^{-1}$, one can show that

$$\begin{aligned} \mathbb{E}[\mathbf{X}_i \mid \mathbf{X}_{-i} = x_{-i}] &= \mu_i - \frac{1}{Q_{ii}} \sum_{j \neq i} Q_{ij}(x_j - \mu_j) \\ \operatorname{Prec}[\mathbf{X}_i \mid \mathbf{X}_{-i} = x_{-i}] &= Q_{ii} \end{aligned} \tag{3.2}$$

which can do done by directly comparing the conditional Lebesgue density $p(\cdot \mid \mathbf{X}_{-i} = x_{-i})$ with the Lebesgue density of a univariate Gaussian. [3].

It turns out that local characteristics also suffice for defining a Gaussian distribution given certain conditions of the conditional distributions.

Theorem 43. *Assume that* $\mathbf{X}$ *is a non-degenerate Gaussian and satisfies*

$$\begin{aligned} \mathbb{E}[\mathbf{X}_i \mid \mathbf{X}_{-i} = x_{-i}] &= \mu_i - \sum_{j \neq i} \beta_{ij}(x_j - \mu_j) \\ \mathbb{V}\mathrm{ar}[\mathbf{X}_i \mid \mathbf{X}_{-i} = x_{-i}] &= \kappa_i^{-1} \end{aligned} \tag{3.3}$$

for some vectors $\mu, \kappa \in \mathbb{R}^n$ *and a matrix* $\beta \in \mathbb{R}^{n \times n}$*. Then* μ *is the mean of* $\mathbf{X}$ *and the matrix* C *given by*

$$C_{ij} = \begin{cases} \kappa_i & i = j \\ \kappa_i \beta_{ij} & i \neq j \end{cases} \tag{3.4}$$

is the precision.

Using equations (3.2) it is quite evident that we also have a converse statement. That is, if β_{ij} and κ_i are such that C in (3.4) is symmetric and positive definite (SPD), then we there exists a non-degenerate Gaussian satisfying (3.3).

Statistical models specified through these local characteristics are known in [1] as (Gaussian) conditional autoregressions or as conditional autoregressive models (CAR models). The conditional approach can be extended to non-Gaussian models which are discussed in [1]. We will not consider such models here. CAR models are contrasted by (Gaussian) simultaneous autoregressive models (SAR) which are defined through equations of the form $B\mathbf{X} = \epsilon$ where ϵ is some Gaussian vector defined on the range of B.

3.1.2 Markov properties in $\mathbb{R}^n$

As in the previous sections let $\mathbf{X}$ be a Gaussian vector. We say that the component $\mathbf{X}_i$ is conditionally independent of $\mathbf{X}_j$ given the remaining variables $\mathbf{X}_{-ij}$ if $\mathbf{X}_i \mid \mathbf{X}_{-ij}$ is independent of $\mathbf{X}_j \mid \mathbf{X}_{-ij}$. One defines conditional independence of two subsets of components similarly. One can show that $\mathbf{X}_i$ is conditionally independent of $\mathbf{X}_j$ given the rest of the variables if and only if the (i, j)'th component of the precision matrix of $\mathbf{X}$ is 0. We can therefore say that the precision matrix of a Gaussian vector encodes the conditional independence properties of $\mathbf{X}$. [3].

To visualize this structure, define a matrix $\mathcal{G} = (V, E)$ such that the vertices V correspond to components of $\mathbf{X}$, i.e. numerate V by $1, \dots, n$, and let E contain an edge (i, j) if and only if $q_{ij} \neq 0$. I.e. components i, j are connected in the graph if and only if $\mathbf{X}_i$ is conditionally independent of $\mathbf{X}_j$ given the remaining components. This graph makes $\mathbf{X}$ into a *graphical* model. Graphical models in general can satisfy one of three increasingly strong Markov properties. For a distribution with Lebesgue density that is strictly positive, these properties are equivalent [1]. This is of course satisfied for non-degenerate Gaussians, so that we get.

Theorem 44. *Let $\mathbf{X}$ be non-degenerate Gaussian vector and $\mathcal{G} = (V, E)$ be a graph. Then the following Markov properties are equivalent:*

1. $\mathbf{X}_i \perp \mathbf{x}_j \mid \mathbf{X}_{-ij} = x_{-ij} \iff \{i, j\} \notin E$ *for all* $i, j \in V$
2. $p(x \mid \mathbf{X}_{-i} = x_{-i}) = p(x \mid \mathbf{X}_{N(i)} = x_{N(i)})$ *for all* $i \in V$ *where* $N(i)$ *is the set of indices* j *such that* $Q(i, j) \neq 0$.
3. *Let* A, B, C *be non-empty and pairwise disjoint, and such that* C *separates* A *and* B. *Then* $\mathbf{X}_A \perp \mathbf{X}_B \mid \mathbf{X}_C = x_C$.

Properties $1, 2$ and 3 are sometimes known respectively as the pairwise, local and global Markov properties. The Markov properties show that the conditional independence graph of a Gaussian vector is quite useful. We therefore define the following.

Definition 45 (Finite dimensional GMRF)**.** *Let $\mathbf{X}$ be a finite dimensional Gaussian vector with n components and let $G = (V, E)$ be an undirected graph. Then we say that $\mathbf{X}$ is a Gaussian Markov Random Field wrt. G (or just a Gaussian Markov Random Field) if $|V| = n$ and the precision Q of $\mathbf{X}$ satisfies $\{i, j\} \in E$ iff $q_{ij} \neq 0$.*

Note that the Markov part of a GMRF refers to the fact that every component $\mathbf{X}_i$ is only conditionally dependent on *some* of the other variables. Theoretically, this does not appear to say much since we are working with finitely many components. In infinite dimensional space, Markov properties mean that $\mathbf{X}_i$ is conditionally dependent on only finitely many other components, which from a theoretical stand point seems to be a fairly strong claim. Markov properties in finite dimensions are interesting for practical reason though. This is so since Markov properties can be used to make calculations more efficient. See e.g. [3] for many results and examples of this.

This concludes our introduction to conditional specification of finite dimensional Gaussian vectors. In the next section we will prepare for the theory of conditional specification of Gaussian processes.

3.2 Specifications

Regarding conditional specification of Gaussian random fields $\mathbf{X}$ we will be modelling the deterministic conditional distributions $P(\mathbf{X} \in A \mid \mathbf{X}_{-J} = \cdot)$ when J is a finite subset of $\mathbb{Z}^n$. That is, we are interested in some finite subset J where we condition on everything outside this subset. To continue we will define so-called specifications, which are objects γ that are supposed to model the conditional distributions $P(\mathbf{X} \in A \mid \mathbf{X}_{-J})$. Obviously such objects are bound to have certain properties. We will consider them in this section.

Now let some random field $\mathbf{X} \colon \Omega \to \mathbb{R}^{\mathbb{Z}^n}$ be given. We will change the formalism slightly and instead of considering $P(\mathbf{X} \in A \mid \mathbf{X}_{-J} = \cdot)$ we define things as in [14]. Recall that the σ-algebra on $\mathbb{R}^{\mathbb{Z}^n}$ was defined in terms of projections. Indeed, $\mathbb{B}(\mathbb{R}^{\mathbb{Z}^n})$ was defined as the coarsest σ-algebra on $\mathbb{R}^{\mathbb{Z}^n}$ such that all projections π_i are measurable. For $J \subset \mathbb{Z}^n$ we can define a smaller

σ-algebra $\mathcal{F}_J$ as the coarsest σ-algebra such that all projection π_i for $i \in J$ are measurable. If H is some other subset of $\mathbb{Z}^n$ such that $J \subset H$ then we have the inclusions

$$\mathcal{F}_J \subset \mathcal{F}_H \subset \mathbb{B}(\mathbb{R}^{\mathbb{Z}^n}) \tag{3.5}$$

The conditional probabilities $P(\mathbf{X} \in A \mid \mathbf{X}_{-J} = \cdot)$ and $P_{\mathbf{X}}(A \mid \mathcal{F}_{-J})$ are connected. Indeed, let B_{-J} be measurable in the subspace $\mathbb{R}^{\mathbb{Z}^n \setminus J}$. This just means that the preimage $B := \pi_{-J}^{-1}(B_{-J})$ is measurable in $\mathbb{R}^{\mathbb{Z}^n}$. By definition of $\mathcal{F}_{-J}$ this means $B \in \mathcal{F}_{-J}$. Since we can write $\mathbf{X}_{-J} = \pi_{-J} \circ \mathbf{X}$ we see that $\mathbf{X}_{-J} \in B_{-J}$ if and only if $\mathbf{X} \in \pi_{-J}^{-1}(B_{-J}) = B$. We now get

$$\begin{aligned}\int_{B_{-J}} P(\mathbf{X} \in A \mid \mathbf{X}_{-J} = \cdot) P_{\mathbf{X}_{-J}} &= \int_{\mathbf{X}_{-J}^{-1}(B_{-J})} P(\mathbf{X} \in A \mid \mathbf{X}_{-J}) dP \\ &= P(\mathbf{X} \in A, \mathbf{X}_{-J} \in B_{-J}) \\ &= P(\mathbf{X} \in A, \mathbf{X} \in B)\end{aligned}$$

while integrating $P_{\mathbf{X}}(A \mid \mathcal{F}_{-J})$ over B we get

$$\int_B P_{\mathbf{X}}(A \mid \mathcal{F}_{-J}) dP_{\mathbf{X}} = P_{\mathbf{X}}(A \cap B) = P(\mathbf{X} \in A, \mathbf{X} \in B).$$

In this sense, using one formalism or the other is equivalent. However, there is a certain advantage to the $\mathcal{F}_{-J}$ formalism. This will be clear later. Hence, we denote

$$\gamma_J = P_{\mathbf{X}}(\cdot \mid \mathcal{F}_{-J}) \colon \mathbb{B}(\mathbb{R}^{\mathbb{Z}^n}) \times \mathbb{R}^{\mathbb{Z}^n} \to \mathbb{R}$$

for finite $J \in \mathbb{Z}^n$. This is as a regular conditional probability for $P_{\mathbf{X}}$ given $\mathcal{F}_{-J}$. By Lemma 96, γ_J exists. Then $\gamma_J(\cdot, \cdot)$ is a probability measure on $\mathbb{B}(\mathbb{R}^{\mathbb{Z}^n})$ in its first argument and $\mathcal{F}_{-J}$ measurable in its second argument. These properties are important in their own right. We therefore define

Definition 46. *Let $(\mathbb{X}, \mathcal{X}), (\mathbb{Y}, \mathcal{Y})$ be two measurable spaces. Let $\psi \colon \mathcal{Y} \times \mathbb{X} \to \mathbb{R}$ be a function such that*

1. *For any $x \in \mathbb{X}$, $U \mapsto \psi(U, x)$ is a measure.*
2. *For any $U \in \mathcal{Y}$, $x \mapsto \psi(U, x)$ is $\mathcal{X}$-$\mathbb{B}(\mathbb{R})$ measurable.*

Then ψ is known as a Markov kernel from $(\mathbb{X}, \mathcal{X})$ to $(\mathbb{Y}, \mathcal{Y})$. If furthermore $U \mapsto \gamma(U, x)$ is a probability measure, then ψ is known as a probability kernel.

It is quite clear from definition that any regular conditional probability is also a probability kernel. Now assume that we have three measurable spaces $(\mathbb{X}, \mathcal{X}), (\mathbb{Y}, \mathcal{Y}), (\mathbb{U}, \mathcal{U})$, with probability kernels ψ_1 from $(\mathbb{X}, \mathcal{X})$ to $(\mathbb{Y}, \mathcal{Y})$ and ψ_2 from $(\mathbb{Y}, \mathcal{Y})$ to $(\mathbb{U}, \mathcal{U})$. Then we can define a composition of ψ_1 and ψ_2 as a function $\psi_1 \psi_2 \colon \mathcal{U} \times \mathbb{X} \to \mathbb{R}$

$$\psi_1 \psi_2(U, x) = \int \psi_2(U, \cdot) d\psi_1(\cdot, x) \tag{3.6}$$

The integral should be understood as integration wrt. the measure $Y \mapsto \psi_1(Y, x)$ on $(\mathbb{Y}, \mathcal{Y})$. This composition is well-defined, since $\psi_2(U, \cdot)$ is $\mathcal{Y}$-measurable and bounded by 1 so that it is integrable wrt. $\psi_1(\cdot, x)$. We have the following.

Lemma 47. *With assumptions as above, the composition $\psi_1 \psi_2$ is a probability kernel from $(\mathbb{X}, \mathcal{X})$ to $(\mathbb{U}, \mathcal{U})$.*

Assume as before that we have three measurable spaces and probability kernels ψ_1 and ψ_2. Let us in addition assume that $\mathbb{Y}$ has two σ-algebras $\mathcal{Y}$ and $\tilde{\mathcal{Y}}$, with $\tilde{\mathcal{Y}} \subset \mathcal{Y}$. Assume that ψ_1 is a probability kernel from $(\mathbb{X}, \mathcal{X})$ to $(\mathbb{Y}, \mathcal{Y})$ and ψ_2 from $(\mathbb{Y}, \tilde{\mathcal{Y}})$ to $(\mathbb{U}, \mathcal{U})$. Then the right-hand side of (3.6) is clearly still well-defined and we will define the composition of ψ_1 and ψ_2 the same way. This composition is also a probability kernel.

Let us exemplify this situation. Denote as before $\gamma_J = P_{\mathbf{X}}(\cdot \mid \mathcal{F}_{-J})$. Then γ_J is a probability kernel from $(\mathbb{R}^{\mathbb{Z}^n}, \mathcal{F}_{-J})$ to $(\mathbb{R}^{\mathbb{Z}^n}, \mathbb{B}(\mathbb{R}^{\mathbb{Z}^n}))$. Then if we let, $K \subset J$, then $\mathcal{F}_{-J} \subset \mathcal{F}_{-K}$ and the composition

$$\gamma_J \gamma_K(A, x) = \int \gamma_K(A, \cdot) d\gamma_J(\cdot, x)$$

has meaning since $X \mapsto \gamma_K(A, \cdot)$ is $\mathcal{F}_{-K}$ measurable and thus also $\mathbb{B}(\mathbb{R}^{\mathbb{Z}^n})$ measurable. This is a clear advantage of the $\mathcal{F}_{-J}$ formalism over the $P(\mathbf{X} \in \cdot \mid \mathbf{X}_{-J} = \cdot)$ formalism since the latter takes sets in $\mathbb{R}^{\mathbb{Z}^n}$ in its first argument and elements in $\mathbb{R}^{\mathbb{Z}\setminus J}$ in its second. Thus seen as Markov kernels, γ_J 'stays' in the same space while $P(\mathbf{X} \in \cdot \mid \mathbf{X}_{-J} = \cdot)$ does not. We can therefore take compositions of the γ_J in the sense of (3.6), which we can not with $P(\mathbf{X} \in \cdot \mid \mathbf{X}_{-J} = \cdot)$.

Apply Proposition 34 to obtain

$$\begin{aligned}
\int \gamma_K(A, \cdot) d\gamma_J(\cdot, x) &= \mathbb{E}[\gamma_K(A, \cdot) \mid \mathcal{F}_{-J}](x) \\
&= \mathbb{E}[\mathbb{E}[\mathbb{I}_A \mid \mathcal{F}_{-K}] \mid \mathcal{F}_{-J}](x) \\
&= \mathbb{E}[\mathbb{I}_A \mid \mathcal{F}_{-J}](x) \quad P_{\mathbf{X}}\text{-a.s.} \\
&= \gamma_J(A, x).
\end{aligned}$$

In total we have for all A that $\gamma_J \gamma_K(A, \cdot) = \gamma_J(A, \cdot)$ for $P_{\mathbf{X}}$-a.s. Furthermore, assume that $B \in \mathcal{F}_{-J}$. Then we have

$$\begin{aligned}
\gamma_J(A \cap B, \cdot) &= \mathbb{E}[\mathbb{I}_A \mathbb{I}_B \mid \mathcal{F}_{-J}] \\
&= \mathbb{E}[\mathbb{I}_A \mid \mathcal{F}_{-J}] \mathbb{I}_B \quad P_{\mathbf{X}}\text{-a.s.} \\
&= \gamma_J(A, \cdot) \mathbb{I}_B
\end{aligned}$$

Note that these properties are necessary conditions that a class of probability kernels (γ_J) must satisfy if $\gamma_J = P_{\mathbf{X}}(A \mid \mathcal{F}_{-J})$.

We can therefore restrict our attention to the following class of objects.

Definition 48 (Specifications)**.** *A specification is a class $\psi = (\psi_J)$ of probability kernels indexed by finite $J \subset \mathbb{Z}^n$ such that ψ_J is a probability kernel from $(\mathbb{R}^{\mathbb{Z}^n}, \mathcal{F}_{-J})$ to $(\mathbb{R}^{\mathbb{Z}^n}, \mathbb{B}(\mathbb{R}^{\mathbb{Z}^n}))$ satisfying*

1. *For $B \in \mathcal{F}_{-J}$, $A \in \mathbb{B}(\mathbb{R}^{\mathbb{Z}^n})$, $\psi_J(A \cap B, \cdot) = \psi_J(A, \cdot)\mathbb{I}_B$*
2. *If $K \subset J$ then $\psi_J \psi_K = \psi_J$.*

Assume that $\psi = (\psi_J)$ is a specification. Then we say that ψ specifies a random field $\mathbf{X} \colon \Omega \to \mathbb{R}$ if for all finite $J \subset \mathbb{Z}^n$ we have $\psi_J = P_{\mathbf{X}}(\cdot \mid \mathcal{F}_{-J})$.

Note that we make the stronger assumption of *sure* equality of e.g. $\psi_J \psi_K = \psi_J$ instead of $P_{\mathbf{X}}$-a.s. The reason for this is practical since when modelling we do not know $P_{\mathbf{X}}$.

3.3 Specifications for Gaussian Random Fields

In this section we will follow chapter 13 of [14] closely. Proofs of various claims and propositions, unless otherwise stated, can be found therein. The book [14] is a challenging read and the

notation is decidedly *not* the standard notation used in the statistical sciences. The contribution of this part of the project is thus consists of expressing the statements in [14] that are more accessible to readers with a background in the statistical and mathematical sciences.

In Section 3.2 we motivated and defined specifications, and what it means for a specification to specify a random field $\mathbf{X}\colon \Omega \to \mathbb{R}^{\mathbb{Z}^n}$. This definition works for general random fields $\mathbf{X}$ mapping into $\mathbb{R}^{\mathbb{Z}^n}$. This generality, while nice, will not be of much importance in the remaining sections as we will only be considering Gaussian random fields. We are interested in two problems.

Firstly, let us say that we have an a priori Gaussian random field $\mathbf{X}$. We can then define a specification for $\mathbf{X}$ setting $\psi_J = P_{\mathbf{X}}(\cdot \mid \mathcal{F}_{-J})$. This construction, however, is so trivial that it is hardly satisfying. We will therefore be interested in whether there are other ways of constructing a specification for $\mathbf{X}$. Such a construction is given in in Section 3.3.1.

Secondly, a specification (ψ_J) may be given a priori. It is then a question whether there exists a Gaussian random field which is specified by it. Properties, and indeed existence itself, of such a Gaussian random field quite clearly depend on properties of the specification. In Section 3.3.2 we consider the problem of existence under quite general assumptions on the specification. In Section 3.3.3 these general results are specialized to a homogenous case where the so-called 'precision' function is translationally invariant, and certain weighted sums of the mean $\mathbb{E}[\mathbf{X}]$ are proportional to the vector of all 1's.

3.3.1 Specifications for an a priori given Gaussian random field

We know that the conditional distributions $\psi_J = P_{\mathbf{X}}(\cdot \mid \mathcal{F}_{-J})$ can be collected into a class of probability kernels that specify $\mathbf{X}$. To define another specification for $\mathbf{X}$, a natural starting point is therefore functions and scalars related to ψ_J such as the conditional means $\mathbb{E}[\mathbf{X}_i \mid \mathcal{F}_{-i}]$. Therefore, we begin by introducing some notation.

1. $\bar{\mu}_i = \mathbb{E}[\mathbf{X}_i \mid \mathcal{F}_{-i}]$.

2. The function $\nu_{ij} = \mathbb{E}[(\mathbf{X}_i - \bar{\mu}_i)(\mathbf{X}_j - \bar{\mu}_j)]$. Note that it is quite natural to assume that $\nu_{ii} \neq 0$, since otherwise $\mathbf{X}_i = \bar{\mu}_i$ P-a.s.

3. If we assume $\nu_{ii} \neq 0$ for all i, we can define the 'precision' function Q by $Q(i,j) = \nu_{ij}/\nu_{ii}\nu_{jj}$. We will justify the name precision later, but for now it suffices to note that $Q(i,i) = \frac{1}{\nu_{ii}}$ where ν_{ii} is prediction error $\mathbf{X}_i - \bar{\mu}_i$.

4. The vector of weighted sums h defined by $h_i = -\sum_{j\in\mathbb{Z}^n} Q(i,j)\mu_j$ where $\mu_j = \mathbb{E}[\mathbf{X}_i]$.

The function Q and the vector h will be of primary interest in this section. Indeed, Q plays a role similar to the precision matrix of a finite dimensional Gaussian distribution and h is similar to the term $-\sum Q_{ij}\mu_i$ in (3.3).

For finite dimensional GMRF, we have the assumption that the conditional expectations can be calculated locally in the sense that $\mathbb{E}[\mathbf{X}_i \mid \mathbf{X}_{-i}]$ only depends on those variables j which are *neighbours* of i, in whichever sense we use that term. In the infinite dimensional case there is the possibility that $\mathbf{X}_i$ depends on infinitely many $\mathbf{X}_j$, infinitely far away. To avoid this, we can assume a Markovian property. Formally, we assume that $\bar{\mu}_i = \mathbb{E}[\mathbf{X}_i \mid \mathcal{F}_{-i}] = \mathbb{E}[\mathbf{X}_i \mid \mathcal{F}_J]$ P-a.s. for some *finite* subset $J \subset \mathbb{Z}^n$ with $i \notin J$. Note that we can always find a smallest such subset J by taking intersections. We can denote this set by $N(i)$. The Markovian property then just means that $N(i)$ is finite. We will always assume this is the case. Recall that in the finite dimensional case we have that $\mathrm{Prec}[\mathbf{X}_i, \mathbf{X}_j] = 0$ if and only if $\mathbf{X}_i$ is conditionally independent of $\mathbf{X}_j$ given the rest of the variables. A similar result holds in the infinite dimensional case.

Proposition 49. *Let* $\mathbf{X}$ *be a Gaussian random field on* $\mathbb{Z}^n$*. Suppose* $\nu_{ii} > 0$ *and* $N(i)$ *is finite for all* $i \in \mathbb{Z}^n$*. Then* Q *is a symmetric, positive definite function and* $Q(i,j) = 0$ *if and only if* $j \notin N(i) \cup \{i\}$ *and we have that*

$$\mathbf{X}_i = \bar{\mu}_i + Q(i,i)^{-1}(h_i + \sum_{j \in \mathbb{Z}^n} Q(i,j)\mathbf{X}_j) \quad P\text{-a.s.} \tag{3.7}$$

$$\mathbb{V}\mathrm{ar}[\mathbf{X}_i \mid \mathcal{F}_{-i}] = Q(i,i)^{-1} \tag{3.8}$$

Equation (3.7) should be compared with (3.2). Indeed

$$h_i + \sum_{j \in \mathbb{Z}^n} Q(i,j)\mathbf{X}_j = \sum_{j \in \mathbb{Z}^n} Q(i,j)(\mathbf{X}_j - \mu_j)$$

where we have used the Markovian property to get that the sums are finite. Now using this in (3.7) we directly get

$$\mathbb{E}[\mathbf{X}_i \mid \mathcal{F}_{-i}] = \mu_i - \frac{1}{Q(i,i)} \sum_{j \neq i} Q(i,j)(\mathbf{X}_j - \mu_j). \tag{3.9}$$

We have thus seen that Proposition 49 gives properties of the local characteristics. In [14], this is used to motivate the following definition of a specification. For $J \subset \mathbb{Z}^n$, denote $k = |J|$. Since J is finite, we can collect the elements $Q(i,j)$ into a matrix Q_J. Since $Q(\cdot,\cdot)$ is a positive definite function, Q_J is invertible. We denote the components of the inverse by $Q_J^{-1}(i,j)$. Furthermore, denote $c = \sqrt{\det Q_J}/\sqrt{(2\pi)^k}$, and let yx_{-J} be defined by

$$(yx_{-J})_i = \begin{cases} y_i & i \in J \\ x_i & \notin J. \end{cases}.$$

We can then define our object of interest as

$$\gamma_J(A \mid x) = c \int \mathbb{I}_A(yx_{-J}) \exp(-\frac{1}{2}\left(\sum_{i \in J} Q(i,i)y_i^2 + h_i y_i + \sum_{i \in J}\sum_{j \neq i} Q(i,j)y_i(yx_{-J})_j\right))d\lambda_J(y) \tag{3.10}$$

One can then show, as in [14], that γ_J is a Gaussian distribution on $\mathbb{R}^{\mathbb{Z}^n}$ with mean given by

$$\bar{\mu}_i = \begin{cases} -\sum_{k \in J} Q_J^{-1}(i,k)\left(h_k + \sum_{j \notin J} Q(k,j)x_j\right) & \text{if } i \in J \\ x_i & \text{if } i \notin J \end{cases} \tag{3.11}$$

and covariance by function

$$\bar{C}_J(i,j) = \begin{cases} Q_J^{-1}(i,j) & i,j \in J \\ 0 & \text{otherwise.} \end{cases} \tag{3.12}$$

We can interpret this as a degenerate Gaussian field since if $\mathbf{Z}$ has distribution given by $\gamma_J(\cdot \mid x)$ then $\mathbf{Z}$ is a.s. constrained to the space where $\mathbf{Z}_{-J} = x_{-j}$.

Proposition 50. *The set of functions* $\gamma = (\gamma_J)$ *defined in* (3.10) *is a specification which specifies* $\mathbf{X}$.

We thus have an example of a specification for $\mathbf{X}$. Our next question is whether all specifications for $\mathbf{X}$ are of this form. To this end, note that if $\tilde{Q}$ is any symmetric positive definite function and $\tilde{h}$ is some vector, we can define a function $\tilde{\gamma}_J$ as in (3.10) using $\tilde{Q}$ and $\tilde{h}$ instead of Q and h. Assume furthermore that $\tilde{Q}$ has a Markov property, i.e. for fixed i, there are only finitely many j s.t. $\tilde{Q}(i,j) \neq 0$. This function will then be a specification and we have the following [14].

Proposition 51. *Let $\tilde{Q}\colon \mathbb{Z}^n \times \mathbb{Z}^n \to \mathbb{R}$ be SPD and have the Markov property described above. Let also $\tilde{h} \in \mathbb{R}^{\mathbb{Z}^n}$ and define the specification $\tilde{\gamma} = (\tilde{\gamma}_J)$ using* (3.10)*. Then a Gaussian field $\mathbf{X}$ with mean μ and covariance C is specified by $\tilde{\gamma}$ if and only if the following equations hold*

$$\tilde{h}_i + \sum_{j \in \mathbb{Z}^n} \tilde{Q}(i,j)\mu_j = 0 \quad \forall i \in \mathbb{Z}^n \tag{3.13}$$

$$\sum_{k \in \mathbb{Z}^n} \tilde{Q}(i,k)C(k,j) = \delta_{ij} \quad \forall i,j \in \mathbb{Z}^n \tag{3.14}$$

Equation (3.14) shows that, seen as infinite matrices, $\tilde{Q}$ must be an inverse of C if it is to be included in a specification for $\mathbf{X}$. Thus $\tilde{Q}$ has the interpretation of a precision matrix for $\mathbf{X}$. Note that this proposition does not give any restraints on what a specification for $\mathbf{X}$ must look like in general, as it only pertains to specifications of the form (3.10). It does, however, give a necessary and sufficient condition which specifications of the form (3.10) must satisfy in order to specify $\mathbf{X}$. We will not attempt to describe general properties of specifications for $\mathbf{X}$ and only consider specifications of the form (3.10) about which we now understand quite a bit.

3.3.2 Gaussian random fields specified by an a priori given specification

Assume we have a specification γ of the form (3.10), where Q is some symmetric, positive definite function and $h = (h_i) \in \mathbb{R}^{\mathbb{Z}^n}$ is any vector. Assume additionally that Q has the Markovian property that for fixed i, $Q(i,j) \neq 0$ for only finitely many j. We are then interested in whether there exists a Gaussian field $\mathbf{X}$ which is specified by γ. Let us therefore define $\mathfrak{G}(\gamma)$ to be the set of random fields specified by γ, so that our working question is whether $\mathfrak{G}(\gamma) \neq \emptyset$. Restating the problem this way is not empty formalism since it turns out that $\mathfrak{G}(\gamma)$ has a special structure.

We start by noting that for two measures ρ, ν on the same measurable space any linear combination of ρ and ν defines a measure on that space too. Thus any convex combination of ρ, ν is a measure too. It turns out that for a given specification γ, $\mathfrak{G}(\gamma)$ is a convex set. We can then define extremal elements of $\mathfrak{G}(\gamma)$ as those elements (if they exist) that can *not* be written as a convex combination of other elements in $\mathfrak{G}(\gamma)$. More precisely, we say that $\rho \in \mathfrak{G}(\gamma)$ is extremal in $\mathfrak{G}(\gamma)$ if, for every pair $\nu, \nu' \in \mathfrak{G}(\gamma)$ and $t \in (0,1)$, we have $\rho \neq t\nu + (1-t)\nu'$. The set of extremal measures in $\mathfrak{G}(\gamma)$ is denoted by $\operatorname{ex}\mathfrak{G}(\gamma)$. An interesting fact about $\operatorname{ex}\mathfrak{G}(\gamma)$ is that it is non-empty if $\mathfrak{G}(\gamma)$ is non-empty and each $\rho \in \mathfrak{G}(\gamma)$ can be decomposed into extremal elements in $\mathfrak{G}(\gamma)$. For details see [14].

Assume that there exists some Gaussian field $\mathbf{X} \in \mathfrak{G}(\gamma)$. Then in view of Proposition 51, the function Q should behave like a kind of inverse for the covariance function for $\mathbf{X}$. It is therefore natural to consider the inverse matrices Q_J^{-1}.

Theorem 52. *Let γ be a specification of the form* (3.10) *where Q is some SPD function with the Markovian property and h is some vector in $\mathbb{R}^{\mathbb{Z}^n}$. Assume furthermore that $\mathfrak{G}(\gamma)$ is non-empty. Then $\mathbf{X} \in \operatorname{ex}\mathfrak{G}(\gamma)$ if and only if $\mathbf{X}$ is Gaussian with mean μ satisfying*

$$h_i + \sum_{j \in \mathbb{Z}^n} Q(i,j)\mu_j = 0 \quad \forall i \in \mathbb{Z}^n \tag{3.15}$$

and the covariance function C of $\mathbf{X}$ satisfies

$$C(i,j) = \lim_J Q_J^{-1}(i,j) \tag{3.16}$$

Remark 53. *The limit in* (3.16) *should be understood as the limit of real numbers $\alpha_k = Q_{J_k}^{-1}(i,j)$ where J_k is a sequence of finite subsets in $\mathbb{Z}^n$ such that $\mathbb{Z}^n = \cup_{k\in\mathbb{N}} J_k$. The proposition shows, that if there exists some Gaussian field in $\operatorname{ex}\mathfrak{G}(\gamma)$, then this sequence must converge and is determined by the covariance of the Gaussian field. In particular, any Gaussian field in $\operatorname{ex}\mathfrak{G}(\gamma)$ must have the same covariance structure. The same is not true for the mean structure. Indeed, assume that $\mathbf{X} \in \mathfrak{G}(\gamma)$ is Gaussian. Then if we can solve $\sum_{j\in\mathbb{Z}^n} Q(i,j)\alpha_j = 0$ for all i for some vector α then $\mathbf{X} + \alpha \in \operatorname{ex}\mathfrak{G}(\gamma)$.*

Theorem 52 gives a description of $\operatorname{ex}\mathfrak{G}(\gamma)$ provided that it is non-empty. It does not, however, give any conditions under which it is non-empty. We continue by noting the following facts. Assume that Q is SPD, Markovian and has an inverse in the sense of (3.16). Assume also that $h = 0 \in \mathbb{R}^{\mathbb{Z}^n}$ and that $\gamma_{Q,0}$ is defined by (3.10). Define C as the inverse of Q and let $\mathbf{X} \sim N_{\mathbb{Z}^n}(0, C)$. Then by Proposition 51 we get $\mathbf{X} \in \mathfrak{G}(\gamma_{Q,0})$. Now let h be arbitrary except that we assume that there exists some μ such that (3.15) holds. It then follows that the Gaussian field $\mathbf{X} + \mu$ is specified by $\gamma_{Q,h}$ where $\gamma_{Q,h}$ is again a specification defined using (3.10) using Q and the arbitrary vector h [14]. This shows that if we assume Q is SPD, Markovian and has an inverse in the sense of (3.16), then $\mathfrak{G}(\gamma_{Q,h})$ is non-empty if there exists a μ solving (3.15). The converse also holds. With these assumptions we can therefore say that existence of a Gaussian random field in $\mathfrak{G}(\gamma_{Q,h})$ is equivalent with existence of solutions to (3.15). We summarize this as follows.

Corollary 54. *Let $Q\colon \mathbb{Z}^n \times \mathbb{Z}^n \to \mathbb{R}$ be SPD and Markovian. Let $h \in \mathbb{R}^{\mathbb{Z}^n}$ and let $\gamma_{Q,h}$ be defined by* (3.10)*. Then a Gaussian field specified by $\gamma_{Q,h}$ exists if and only if a function $C\colon \mathbb{Z}^n \times \mathbb{Z}^n \to \mathbb{R}$ and a vector μ satisfying*

$$h_i + \sum_{j\in\mathbb{Z}^n} Q(i,j)\mu_j = 0 \quad \forall i \in \mathbb{Z}^n$$
$$C(i,j) = \lim_J Q_J^{-1}(i,j).$$

exist. In that case a Gaussian random field $\mathbf{X}$ with mean μ and covariance C is specified by $\gamma_{Q,h}$.

3.3.3 Homogenous conditional specification of Gaussian fields

Let a specification $\gamma_{Q,h}$ of the form (3.10) for a SPD function Q and vector h be given. Then we say that $\gamma_{Q,h}$ is *homogenous* on $\mathbb{Z}^n$ if $h_i = h_0$ and there exists a function $\tilde{Q}\colon \mathbb{Z}^n \to \mathbb{R}$ such that $Q(i,j) = \tilde{Q}(i-j)$ for all $i, j \in \mathbb{Z}^n$.

Symmetry of Q implies that $\tilde{Q}$ is an even function, i.e. $\tilde{Q}(i) = \tilde{Q}(-i)$. Positive definiteness of Q implies that for finite subsets $J \subset \mathbb{Z}^n$ and sequences of real numbers λ_i not all equal to 0 indexed by $i \in J$ we have

$$\sum_{i,j} \lambda_i \lambda_j \tilde{Q}(i-j) > 0 \tag{3.17}$$

We refer to this property (of $\tilde{Q}$) as positive definiteness. Henceforth, we will denote $\tilde{Q}$ as Q, so that with abuse of notation we can write e.g. $Q(i,j) = Q(i-j)$. Note that if Q has a Markovian property then $Q(i) \neq 0$ for only finitely many i.

Homogeneity is a strong property, and it is therefore no surprise that it allows us to improve upon the results about conditional specifications which we considered in Section 3.3.2. These

additional properties can be related to a Fourier transform of Q. To make things precise, we will be using the concept of Haar measures, which are briefly described in Appendix A.3

As in example 88 we let K^n be the n-fold product of complex unit circles with normalized Haar measure ν. For elements $i = (i_1, \ldots, i_n) \in \mathbb{Z}^n$ and $z \in K^n$ define the following operation $z^i = z_1^{i_1} \cdot \ldots \cdot z_n^{i_n}$ and note $z^i \in K^1$. With notation as in the beginning of this section let $Q\colon \mathbb{Z}^n \to \mathbb{R}$ correspond to a SPD and Markovian function. Then we can define a Fourier transform $\hat{Q}\colon K^n \to \mathbb{C}$ of Q by

$$\hat{Q}(z) = \sum_{i \in \mathbb{Z}^n} z^i Q(i).$$

Since Q is even, we have $z^i Q(i) + z^{-i} Q(-i) = (z^i + z^{-i}) Q(i)$. For $k \in \mathbb{Z}^n$ we can write $z^k = \exp(i\pi\theta)$ for some angle θ. Then $z^{-k} = \exp(i\pi(-\theta))$. Using Euler's formula it follows that $z^k + z^{-k}$ is real. Thus evenness of Q implies that the Fourier transform is real. It therefore makes sense to ask whether $\int_{K^n} \hat{Q} d\nu < \infty$. Finally, we have that $\hat{Q}$ is non-negative if and only if Q is positive definite [14]. In that case we will define the symbol $\hat{Q}(i)^{-1}$ by

$$\hat{Q}(z)^{-1} = \begin{cases} (\hat{Q}(z))^{-1} & \text{if } \hat{Q}(i) \neq 0 \\ \infty & \text{otherwise} \end{cases} \tag{3.18}$$

This should not be confused with an inverse of $\hat{Q}$.

Proposition 55. *Let $\gamma_{Q,0}$ be of the form* (3.10) *where Q is SPD and has the Markovian property and the vector $h = 0$ is just the zero vector. Assume also that $\gamma_{Q,h}$ is homogenous and that*

$$\int_{K^n} \hat{Q} d\nu < \infty.$$

Then the centered Gaussian random field $\mathbf{X} \sim N(0, C)$ where

$$C(i,j) = \int_{K^n} z^{i-j} \hat{Q}(z)^{-1} d\nu \tag{3.19}$$

is specified by $\gamma_{Q,0}$.

This proposition takes care of the centered case. In what follows we will consider two things. Firstly, to investigate the non-centered case, we will describe the space $V_{Q,h}$ of solutions μ to $h_i + \sum_{j \in \mathbb{Z}^n} Q(i-j)\mu_j = 0$ for all $i \in \mathbb{Z}^n$. Secondly, we will consider more carefully the covariance function in (3.19). Note that C is evidently translationally invariant since the integral in (3.19) only depends on (i,j) through $i - j$

Proposition 56. *Assume as usual that $Q\colon \mathbb{Z}^n \to \mathbb{R}$ is SPD and $Q(i) \neq 0$ for only finitely many i. Let $h \in \mathbb{R}^{\mathbb{Z}^n}$ be any vector. Then $V_{Q,h}$ contains at least one element. If $Q(i) \neq 0$ for some $i \neq 0$ then $V_{Q,h}$ is uncountable.*

Note that in the case where $Q(i) = 0$ for all $i \neq 0$ we get precisely one solution given by $\mu_i = h_i / Q(0)$ for all i. The proposition shows that we in general will not have uniqueness of Gaussian fields specified by $\gamma_{Q,h}$. Note that we do have uniqueness in the case of no interaction, i.e. in the case where $Q(i) \neq 0$ implies $i = 0$.

We will now consider the covariance function further. In Example 88 in the Appendix it was noted that the measure ν in (3.19) can be expressed in terms of the Lebesgue measure on $(-1,1]^n$ using the function f in Equation (A.5). This function is given by $f(x_1, \ldots, x_n) = (\exp(i\pi x_1), \ldots, \exp(i\pi x_n))$. Combining these facts we find that

$$C(i,j) = \int_K^n z^{i-j} \hat{Q}(z)^{-1} d\nu = \int_{(-1,1]^n} \exp(i\pi \langle x, i-j \rangle) \hat{Q}(f(x)) d\lambda_n(x)$$

where $\langle x, i-j\rangle$ is the dot product of $x = (x_1, \ldots, x_n)$ and $i-j \in \mathbb{Z}^n$. This gives $C(i-j) := C(i,j)$ the interpretation as the inverse Fourier transform of $\hat{Q} \circ f$. We will end this section with an example of a covariance function of the form (3.19) for a given Q.

Example 57. *Let $\{e_k\}$ denote the standard basis for $\mathbb{Z}^n$. Let $Q\colon \mathbb{Z}^n \to \mathbb{R}$ be given by*

$$Q(i) = \begin{cases} 1 & i = 0 \\ 1/4n & i \in \{e_k\} \cup \{-e_k\} \\ 0 & \textit{otherwise} \end{cases}$$

This function is clearly Markovian and symmetric. We claim that it is positive definite too. To show this let, $J \subset \mathbb{Z}^n$ be finite and let Q_J be the $|J| \times |J|$ matrix containing the components $Q(i-j)$ for $i, j \in J$. More precisely, since J is finite we can endow it with a linear ordering $\tilde{\leq}$. Then J is isomorphic with the set $\{1, \ldots, |J|\}$ seen as ordered sets. Let ϕ be such an isomorphism and define Q_J by $[Q_J]_{mk} = Q(\phi^{-1}(m), \phi^{-1}(k))$. Then Q_J is symmetric. We intend to use diagonal dominance. Thus note that we have $[Q_J]_{mm} = 1$. Let m be fixed. Then we can only have $[Q_J]_{mk} = \frac{1}{4n}$ for at most $2n$ (where $n = \dim \mathbb{Z}^n$) choices of k. For the remaining $k \in \{1, \ldots, |J|\}$ we have $[Q_J]_{mk} = 0$. We thus have diagonal dominance with positive diagonal elements and symmetry so that Q_J is positive definite. It hence, Q is positive definite too. It follows that the transform $\hat{Q}$ is well-defined. Let $z \in K^n$, so that $z = (z_1, \ldots, z_n)$ where each z_i is an element of the complex unit circle. Then $\hat{Q}$ can be explicitly expressed as

$$\begin{aligned} \hat{Q}(z) &= \sum_{i \in \mathbb{Z}^n} z^i Q(i) \\ &= \sum_{i \in \{0\} \cup \{e_k\} \cup \{-e_k\}} z_1^{i_1} \ldots z_n^{i_n} Q(i) \\ &= 1 + \sum_{i=1}^{n} \frac{1}{4n} z_i + \sum_{i=1}^{n} \frac{1}{4n} \frac{1}{z_i} \end{aligned}$$

It has been shown that $\hat{Q}(z)$ is generally real. This is exemplified by this case. Calculating as complex numbers, we get the bound $|\sum_{i=1}^n \frac{1}{4n} z_i + \sum_{i=1}^n \frac{1}{4n} \frac{1}{z_i}| \leq 1/2$ by using the triangle inequality. It follows that $\left|\hat{Q}(z)\right| \geq 1/2$. In fact, using that $\sum_{i=1}^n \frac{1}{4n} z_i + \sum_{i=1}^n \frac{1}{4n} \frac{1}{z_i}$ is real we get $\hat{Q}(z) \geq 1/2$. Also by the triangle inequality, we have $\hat{Q}(z) \leq 3/2$. Thus $\hat{Q}(z)^{-1}$ is well-defined as a real number and contained in the interval $[\frac{2}{3}, 2]$. Since $\nu(K^n) = 1$ and since $\hat{Q}$ is positive and bounded by $3/2$ on K^n we get that $\hat{Q}$ is integrable with

$$\int_{K^n} \hat{Q}(z) d\nu(z) \leq 3/2.$$

We now define C as in (3.19). Then we need to integrate functions of the form $z^{i-j}\hat{Q}(z)^{-1}$ wrt. ν on K^n. To this end we use that $\nu(B) = \lambda_n(f^{-1}(B))$, $B \subset K^n$ a Borel set, where f is the isomorphism between K^n and $]-1, 1]^n$ in (A.5). We can then calculate integrals of functions

wrt. ν *as* $\int_{K^n} g d\nu = \int_{]-1,1]^n} g \circ f d\lambda_n$. *Finally, denote* $d_k = i_k - j_k$, *then we calculate*

$$\begin{aligned}
C(i,j) &= \int_{K^n} z^{i-j} \hat{Q}(z)^{-1} d\nu \\
&= \int_{]-1,1]^n} \prod_{k=1}^{n} \exp(i\pi d_k x_k)(1 + \sum_{k=1}^{n} \frac{1}{4n}(\exp(i\pi x_k) + \exp(-i\pi x_k)))^{-1} d\lambda_n(x) \\
&= \int_{]-1,1]^n} \prod_{k=1}^{n} \exp(i\pi d_k x_k)(1 + \sum_{k=1}^{n} \frac{1}{2n}\cos(\pi x_k))^{-1} d\lambda_n(x) \\
&= \int_{]-1,1]^n} \frac{\prod_{k=1}^{n} \cos(\pi d_k x_k)}{1 + \frac{1}{2n}\sum_{k=1}^{n} \cos(\pi x_k)} d\lambda_n(x).
\end{aligned}$$

In the case that $n = 2$, *numerical integration yields the values in Table 3.1 of* $C(d)$ *for* $d = (d_1, d_2)$ *with* $-2 \leq d_1, d_2 \leq 2$. *Note* $C(d)$ *only appears to depend on* d *through* $|d_1| + |d_2|$ *so that* C *is isotropic. Furthermore,* $C(d)$ *decreases fast as* $|d_1| + |d_2|$ *grows.*

$d_2 = -2$	$d_2 = -1$	$d_2 = 0$	$d_2 = 1$	$d_2 = 2$	
0.008	-0.031	0.083	-0.031	0.008	$d_1 = 2$
-0.031	0.154	-0.585	0.154	-0.031	$d_1 = 1$
0.083	-0.585	4.293	-0.585	0.083	$d_1 = 0$
-0.031	0.154	-0.585	0.154	-0.031	$d_1 = -1$
0.008	-0.031	0.083	-0.031	0.008	$d_1 = -2$

Table 3.1: Values of $C(d)$ around the origin in Example 57 when $n = 2$

Chapter 4

Improper Probability and Intrinsic Fields

So far we have considered measurable maps $\mathbf{X}\colon \Omega \to \mathbb{X}$ which are *random* in the classical sense. That is, we assume $P(\mathbf{X} \in \mathbb{X}) = 1$. However, in some instances, natural assumptions about $\mathbf{X}$ may imply that $P(\mathbf{X} \in \mathbb{X}) = \infty$. In that case $\mathbf{X}$ is not a random vector or process in the classical sense, but many methods from probability theory still apply. We begin this chapter by considering the class of measurable maps $\mathbf{X}\colon \Omega \to \mathbb{X}$ for which $P_{\mathbf{X}}$ is σ-finite. In that case $\mathbf{X}$ is said to be improper. After this we introduce so-called intrinsic random fields which can be thought of as a generalization of stationary fields and may in some cases be improper. We conclude the chapter by considering a model for intrinsic fields known as intrinsic autoregressions which are related to the conditional specifications of Chapter 2.

4.1 Improper Gaussians

Following Kolmogorov we may formalize the notions in probability using probability spaces. These are simply measure spaces $(\Omega, \mathcal{F}, P)$ where the measure P has the property that $P(\Omega) = 1$. However, in some situations it may be tempting to define a measure which can not have a non-zero and finite total mass. We we illustrate this with the following, arguably contrived, examples.

Example 58. *Let us consider some random variable $\mathbf{X}\colon \Omega \to \mathbb{R}$. If $\mathbf{X}$ has some special meaning in a statistical model, and if we do not know anything about $\mathbf{X}$, it is not very attractive to endow it with a probability distribution which, in a certain sense, favours $\mathbf{X}$ being in some areas of $\mathbb{R}$ rather than others. More precisely, $P(\mathbf{X} \in (a, b))$ should only depend on the size of the interval $b - a$. Using this assumption we get $P(\mathbf{X} \in (a, b]) = c(a - b)$ for all intervals (a, b) and some fixed scalar c. But then $1 = P(\mathbf{X} \in \mathbb{R})$ leads to a contradiction. Indeed, if $c > 0$ then $P(\mathbf{X} \in \mathbb{R}) = \sum_{n\in\mathbb{Z}} P(\mathbf{X} \in (n, n + 1]) = \sum_{n\in\mathbb{Z}} c = \infty$, while if $c = 0$ then the same calculation yields $P(\mathbf{X} \in \mathbb{R}) = 0$.*

The axiom $P(\mathbf{X} \in \mathbb{R}) = 0$ appears entirely useless, while on the other hand $P(\mathbf{X} \in \mathbb{R}) = \infty$ seems at least capable of being useful. Indeed, if we take $c = 1$, then we say that the "likelihood" that $\mathbf{X} \in (a, b]$ is equal to $a - b$ which seems reasonable and permits interpretation.

We say that $\mathbf{X}\colon \Omega \to \mathbb{R}$ has a uniform distribution on $\mathbb{R}$ if $P_{\mathbf{X}}$ is the Lebesgue measure.

Example 59. *To build on this, assume we have $\mathbf{X}\colon \Omega \to \mathbb{R}^n$ where we know that $\mathbf{X}$ is a degenerate Gaussian living on one of the infinitely many affine subspaces orthogonal to the vector $e_n = (\delta_{in})_i \in \mathbb{R}^n$. Here $\delta_{in} = 1$ if $i = n$ and 0 otherwise. In practical applications it*

may be that we do not want to impose any restrictions on which of these subspaces $\mathbf{X}$ *actually belongs. Then any of these affine subspaces* V *should have the same probability. If* $\mathbf{X}$ *were a degenerate Gaussian we should then have that* $\mathbf{X}$ *has density* f *wrt. the* $\lambda_{n,n-1}$ *Hausdorff measure given by*

$$f(x) = \frac{\sqrt{\det{}^*(Q)}}{\sqrt{(2\pi)^{k-1}}} \exp(-\frac{1}{2}(x-\mu)^\top Q(x-\mu))$$

for some vector $\mu \in V$*. Here we have that* Q *is symmetric and* $\operatorname{range}(Q) = V$ *so that* $\ker Q = \operatorname{span}(e_n)$*. Now let us pretend* f *is also a density wrt. the* n*-dimensional Lebesgue measure. This density is invariant to whether* $\mathbf{X} \in V$ *or* $\mathbf{X} \in V + \alpha e_n$ *for any* $\alpha \in \mathbb{R}$*. For this reason* f *seems like the kind of density we are looking for. Now since* f *is a density wrt.* λ_n *we can calculate probabilities of sets such as* $\mathbf{X} \in V + (a,b]e_n$*. This clearly depends on the size of* $(a,b]$*. Indeed, it can be shown that*

$$P(\mathbf{X} \in V + (a,b]e_n) = \int_{V+(a,b]e_n} f d\lambda_n = c(b-a)$$

Using this we get a contradiction if we assume $P(\mathbf{X} \in \mathbb{R}^n) = 1$ *and* $c > 0$*.*

While these examples were designed to show that it may be natural to make assumptions that imply infinite total mass, it is not obvious that extending probability theory in this way is a strict necessity. The question of whether it is strictly necessary borders almost on the philosophical, and we will thus not attempt to answer it here. We will be satisfied with noting measures exist, which do not seem to correspond to anything in standard probability theory, nonetheless are measures which are susceptible to analysis and appear related to probability.

4.1.1 Improper priors leading to proper posteriors

Following [12] we have the following well known axioms for probability theory. A triplet $(\Omega, \mathcal{F}, P)$ is called a probability space if $\mathcal{F}$ is a σ-algebra on Ω and P is a measure defined on $\mathcal{F}$ assigning total mass 1 to Ω.

Let us make explicit some well known ideas and definitions. Consider any measurable map $\mathbf{X}$ from a measure space Ω to some other measurable space $\mathbb{X}$. Then we say that $\mathbf{X}$ is a random vector if Ω is a probability space, i.e. the measure on Ω has total mass 1. We also say that the pushforward measure $P_{\mathbf{X}}$ on $\mathbb{X}$ defined by $P_{\mathbf{X}}(A) = P \circ \mathbf{X}^{-1}(A) = P(\mathbf{X} \in A)$ is the distribution of $\mathbf{X}$. This measure will of course satisfy $P_{\mathbf{X}}(\mathbb{X}) = 1$ and thus be a probability measure on $\mathbb{X}$ if P is a probability measure on Ω.

In the Examples 58 and 59 we considered measurable maps $\mathbf{X}$ with $P_{\mathbf{X}}$ having infinite total mass. Such a map is therefore not a probability measure as usually understood. Note in addition to this, that these measures were σ-finite. Thus, if the usual axioms of probability are extended slightly to incorporate σ-finite measures, such functions can be incorporated into the theory. Such an approach is proposed in [11]. We therefore define as follows.

Definition 60. *Let* Ω *be a measurable space. Then a measure on* Ω *is said to be an improper probability measure if it is* σ*-finite and not finite.*

The more general theory including improper probability measures contains standard probability theory as a subset and all results and constructions about proper probability measures are true in the larger theory as well. Many basic results and constructions from standard probability theory can be generalized to improper measures. See e.g. [11] for a survey of this.

The commonly known approach to defining conditional probabilities using the Radon-Nikodym theorem can also be used in the larger theory involving improper distributions. In the context of probability we have the following interesting property.

Proposition 61. For an improper distribution $P_{\mathbf{X}}$ and any realization x of $\mathbf{X}$ the conditional distribution $A \mapsto P(A \mid \mathbf{X} = x)$ is a proper probability distribution. [11]

In more informal words, conditioning on improper processes as defined in Definition 60 lead to proper distributions. To define the distribution of conditional vectors or processes $\mathbf{Y} \mid \mathbf{X} = x$ we compose $P(\cdot \mid \mathbf{X} = x)$ with $\mathbf{Y}$. This makes sense if $\mathbf{Y}$ and $\mathbf{X}$ both map from the same probability space $(\Omega, \mathcal{F}, P)$. One can then derive the usual Bayes' rule involving densitites

$$p(x \mid y) = \frac{p(y \mid x)p(x)}{p(y)} \tag{4.1}$$

even when the density $p(x)$ is not proper. See [11] and the references mentioned therein for details.

We have already seen examples of improper probability distributions. Let us end this section by considering a simple example of an improper distribution being used as a prior. Assume we have a linear model $\mathbf{Y} = \Theta\mathbf{X} + \epsilon$, where $\epsilon \sim N_n(0, I)$ and $\Theta \colon \Omega \to \mathbb{R}$ is some random variable. If we have prior knowledge that Θ is close to some value Θ_0 it may be reasonable to equip Θ with prior distribution $\Theta \sim N(\Theta_0, \tau^2)$. Then, if we assume independence of $\mathbf{X}$ and Θ, we get a posterior given by

$$p(\Theta \mid x, y) \propto p(y \mid x, \Theta)p(\Theta) \propto \exp(-\frac{1}{2\sigma^2}\|y - \Theta x\|^2) \exp(-\frac{1}{2\tau^2}(\Theta - \Theta_0)^2).$$

The first factor penalizes residual error $\|y - \Theta x\|^2$ whereas the second penalizes differences $(\Theta - \Theta_0)^2$. We can therefore think of this term as skewing the maximum a posteriori estimate of Θ towards Θ_0. Indeed we get the estimate

$$\mathbb{E}[\Theta \mid \mathbf{X} = x, \mathbf{Y} = y] = \frac{x^\top y/\sigma^2 + \Theta_0/\tau^2}{\|x\|^2/\sigma^2 + 1/\tau^2} \tag{4.2}$$

If we do not know anything about Θ, we would like our prior to not skew estimate it in any direction. If we knew that $\Theta \in [a, b]$ for some interval $[a, b] \subset \mathbb{R}$, this could be accomplished by using a uniform distribution on $[a, b]$. Then the posterior reduces to

$$p(\Theta \mid x, y) \propto \mathbb{I}[\Theta \in [a, b]] \exp(-\frac{1}{2\sigma^2}\|y - \Theta x\|^2).$$

The term $\mathbb{I}[\Theta \in [a, b]]$ can be made practically unimportant by letting $[a, b]$ be very large. Using this prior, even for very large intervals $[a, b]$ is a bit unfortunate, if we really do not want to say anything at all about Θ. This motivates defining a uniform distribution on the infinite interval $(-\infty, \infty)$. We thus formally set $U(B) = \int_B 1 d\lambda_1(x) = \lambda_1(B)$ for Borel sets B. In other words, the uniform measure on $\mathbb{R}$ is exactly the one dimensional Lebesgue measure. This is clearly not a proper distribution. It is however σ-finite so its use may be justified as in [11]. Using the uniform distribution as a prior for Θ we obtain

$$p(\Theta \mid x, y) \propto \exp(-\frac{1}{2\sigma^2}\|y - \Theta x\|^2)$$

which has maximizer wrt. Θ given by $\hat{\Theta} = \frac{x^\top y}{\|x\|^2}$ which is exactly the same as the frequentist estimate for Θ. If we compare this with the estimate in (4.2), we see that these estimates agree in the limit $\tau^2 \to \infty$. This reflects the fact that the density for $N(\Theta_0, \tau^2)$ gets increasingly flat as τ^2. We can therefore think of the limit of $N(\Theta_0, \tau^2)$ as $\tau^2 \to \infty$ as something like the uniform distribution on $\mathbb{R}$. This is not a precise statement since the density for $N(\Theta_0, \tau^2)$ converges uniformly to the zero function $\Theta \mapsto 0$ as $\tau^2 \to \infty$, so that the limit measure has total mass 0.

4.1.2 A model for noisy images

We have not seen how the theory of impropriety expresses itself in the context of Gaussian vectors. Indeed, we have not defined what it means for an improper distribution to be Gaussian. As seen in Example 59, impropriety of something like a Gaussian distribution can occur if we take the Hausdorff density of a degenerate Gaussian and simply claim that this density is a density with respect to the Lebesgue measure. In this section we give a more detailed example of this theme.

Suppose that we have data $Y = \{y_{ij}\}$, $1 \leq i \leq N$, $1 \leq j \leq M$ where each y_{ij} is the value of a pixel in a noisy image. Assume for simplicity that y_{ij} are greyscale values which are transformed to take values in all of $\mathbb{R}$. Furthermore, we may convert indices $(i, j) \in \mathbb{Z}^2$ to indices $k \in \mathbb{Z}$. Let $n = NM$ and assume everything is modelled using vectors in $\mathbb{R}^n$. We will use a GMRF model in the following.

In a probabilistic interpretation we propose that there exists a random vector $\mathbf{X}$ generating images and a random vector $\mathbf{Y}$ generating noisy images. We will assume that $\mathbf{Y}$ is connected to $\mathbf{X}$ through the simple model $\mathbf{Y} = \mathbf{X} + \epsilon$ where $\epsilon = \{\epsilon_k\}$ is Gaussian white noise, i.e. $\epsilon \sim N_n(0, I)$ which is independent of $\mathbf{X}$. In the context of denoising y we are therefore interested in finding some estimate of $\mathbf{X}$ given $\mathbf{Y} = y$. We will use the estimate $\hat{\mathbf{X}} = \mathbb{E}[\mathbf{X} \mid \mathbf{Y} = y]$. To do this we need to have some sensible prior distribution for $\mathbf{X}$. Note that in a noisy image, values of pixels next to each other may fluctuate wildly, whereas in a noiseless image, there may be significant correlations between neighbouring pixels. In order to model correlations between pixels we can form a conditional autoregressive model for $\mathbf{X}$ where we let the conditional mean of $\mathbf{X}_{ij}$ depend on its four neighbours $N(i) = \{i \pm (1, 0), i \pm (0, 1)\}$

$$\mathbb{E}[\mathbf{X}_i \mid \mathbf{X}_{-i} = x_{-i}] = \alpha - \sum_{k \in N(i)} \beta_{ik}(x_k - \alpha) \tag{4.3}$$

$$\mathbb{V}\mathrm{ar}[\mathbf{X}_i \mid \mathbf{X}_{-i} = x_{-i}] = \kappa_i^2. \tag{4.4}$$

whenever i does not correspond to a pixel on an edge. It is not entirely clear what to do with vertices on the boundary. If we hold on to the idea that a pixel should depend on four neighbours, we may:

- let missing values be interpolated by an estimated mean value
- let them be equal to the theoretical mean, so that their contribution is exactly 0 to the conditional mean,
- condition on the boundary values, or
- use the so-called toroidal approximation.

For our current purposes we will use the latter approach. In this model we assume that vertices on edges opposite each other are neighbours, so that, for example, pixels $(1, j)$ and (N, j) are neighbours. This of course induces strong correlations between nodes that geometrically are far apart, but we will ignore such problems.

We motivate the use of an improper prior as follows: The arguably simplest choice of model is to let $\kappa_i = \kappa$ for some $\kappa > 0$ for all i, $\beta_{ik} = -\frac{1}{4}$ whenever $k \in N(i)$. In that case the conditional expectation of $\mathbf{X}_i$ is just the arithmetic average of the values in its neighbouring sites. Intuitively, such a prior would discourage wild fluctuations in neighbouring pixels, thus having a smoothing effect on the image, and therefore removing some of the noise. Note that

with this prior, the mean $\mathbb{E}[\mathbf{X}_i] = \alpha$ disappears from the conditional expectation (4.3) so that it takes the simple form

$$\mathbb{E}[\mathbf{X}_i \mid \mathbf{X}_{-i} = x_{-i}] = \sum_{k \in N(i)} \beta_{ik} x_k.$$

If $\mathbf{X}$ is non-degenerate and $\kappa = 1$ we can use Theorem 43 to find that $\mathbf{X}$ has precision given by Q given by

$$Q_{ij} = \begin{cases} 1 & i = j \\ -\frac{1}{4} & j \in N(i). \end{cases} \tag{4.5}$$

This matrix is *not* invertible since the vector $1_n = (1, \ldots, 1)$ is in the kernel of Q. Theorem 43 can therefore not be used to say anything about $\mathbf{X}$. We can, however, *define* a distribution for $\mathbf{X}$ such that $\mathbf{X}$ is a Gaussian with precision Q given by (4.5).

We then get a well-defined degenerate Gaussian if we state that $\mathbf{X}$ lives on the affine subspace $\text{range}(Q) + \alpha 1_n$ for a given α. However, it is rather undesirable that we have to specify the *overall level* of the greyscale values α. It is therefore tempting to say that $\mathbf{X}$ is a Gaussian vector defined on all of $\mathbb{R}^n$ with density proportional to

$$f(x) \propto \exp(-\frac{1}{2} x^\top Q x) \tag{4.6}$$

Much like in Example 59, we find that the measure with Lebesgue density f is not a probability measure since $\int_{\mathbb{R}^n} f(x) d\lambda_n(x) = \infty$. However, it can be shown that this measure is σ-finite so that $\mathbf{X}$ with this distribution is improper.

4.1.3 Interpretation of the improper prior

Let us use this improper distribution of $\mathbf{X}$ as a prior. We are then interested in the posterior $\mathbf{X} \mid \mathbf{Y} = y$. In light of [11] we should treat $\mathbf{X}$ with this prior as if it were proper and consider the posterior density $p(x \mid y) \propto p(y \mid x)p(x)$ as in (4.1).

Proposition* 62. *Using the improper prior* (4.6)*, we get the proper posterior for* $\mathbf{X} \mid \mathbf{Y} = y \sim N((I+Q)^{-1}y, (I+Q)^{-1})$.

Proof. Since $\mathbf{Y} = \epsilon + \mathbf{X}$ we get that $\mathbf{Y} \mid \mathbf{X} = x$ follows a non-degenerate Gaussian distribution with mean x and covariance matrix equal to the identity matrix. Therefore using 4.1 we get

$$\begin{aligned} p(x \mid y) &\propto \exp(-\frac{1}{2}\|y - x\|^2) \exp(-\frac{1}{2} x^\top Q x) \\ &\propto \exp(-\frac{1}{2} x^\top x + x^\top y - \frac{1}{2} x^\top Q x) \\ &= \exp(-\frac{1}{2} x^\top (I+Q) x + x^\top y) \end{aligned}$$

We think of this a density on $\mathbb{R}^n$. The density of a non-degenerate Gaussian $\mathbf{Z} \sim N(\mu, \Sigma)$ can be written as

$$p(z) \propto \exp(-\frac{1}{2} z^\top \Sigma^{-1} z + z^\top \Sigma^{-1} \mu).$$

Comparing this with the posterior, we see that precision of $\mathbf{X} \mid \mathbf{Y} = y$ is given by $I+Q$ which *is* invertible. Write the term $x^\top y$ as $x^\top (I+Q)(I+Q)^{-1} y$ to find that $(I+Q)^{-1}y$ is the conditional mean of $\mathbf{X}$ given $\mathbf{Y} = y$. □

Let us think about this improper prior from a different angle and form a model $\mathbf{X} = \mathbf{W} + \alpha 1_n$ where $\mathbf{W}$ is a degenerate Gaussian on range(Q) with precision Q and mean 0 and $\alpha \sim N(0, \tau^2)$ is independent of $\mathbf{W}$. As before let $\mathbf{Y} = \mathbf{X} + \epsilon$ where ϵ is independent of both $\mathbf{X}$ and α. As noted in the end of Section 4.1.1, the limit measure of $N(0, \tau^2)$ as $\tau^2 \to \infty$ resembles something like a uniform distribution on $\mathbb{R}$. It is therefore expected that we get similar results using the improper prior as we let τ^2 grow without bound. To investigate this we start with the following lemma.

Lemma* 63. *Using the proper prior $\mathbf{X} = \mathbf{W} + \alpha 1_n$ we get*

$$\mathbb{E}[\mathbf{X} \mid \mathbf{Y} = y] = \Sigma(\Sigma + I)^{-1} y, \quad \mathbb{V}\mathrm{ar}[\mathbf{X} \mid \mathbf{Y} = y] = \Sigma - \Sigma(\Sigma + I)^{-1}\Sigma$$

where $\Sigma := Q^+ + \tau^2 1_n 1_n^\top$ is the covariance of $(\mathbf{X}, \alpha)$. We also get $\mathbb{E}[\alpha \mid \mathbf{Y} = y] = \tau^2 1_n^\top (\Sigma + I)^{-1} y$

Proof. We start by noting that $\Sigma + I$ has full rank, which follows from the fact that it is positive definite. Furthermore

$$(\mathbf{Y}, \mathbf{X}, \alpha) \sim N\left(0, \begin{pmatrix} \Sigma + I & \Sigma & \tau^2 1_n \\ \Sigma & \Sigma & \tau^2 1_n \\ \tau^2 1_n^\top & \tau^2 1_n^\top & \tau^2 \end{pmatrix}\right)$$

Denote the covariance matrix of $(\mathbf{X}, \alpha)$ by Σ. We can then find the conditional distribution $(\mathbf{X}, \alpha) \mid \mathbf{Y} = y$ using (3.1). We then get conditional means and variances

$$\mathbb{E}\left[\begin{pmatrix} \mathbf{X} \\ \alpha \end{pmatrix} \mid \mathbf{Y} = y\right] = \begin{pmatrix} \Sigma \\ \tau^2 1_n^\top \end{pmatrix} (\Sigma + I)^{-1} y$$

$$\mathbb{V}\mathrm{ar}\left[\begin{pmatrix} \mathbf{X} \\ \alpha \end{pmatrix} \mid \mathbf{Y} = y\right] = \begin{pmatrix} \Sigma & \tau^2 1_n \\ \tau^2 1_n & \tau^2 \end{pmatrix} - \begin{pmatrix} \Sigma \\ \tau^2 1_n^\top \end{pmatrix} (\Sigma + I)^{-1} \begin{pmatrix} \Sigma & \tau 1_n \end{pmatrix}.$$

The lemma follows directly from these equations. □

We intend to compare the estimate $\Sigma(\Sigma + I)^{-1} Y$ in Lemma 63 with the estimate $\hat{\mathbf{X}} = (I + Q)^{-1} Y$ that was obtained using the improper prior. Our conjecture is that in the limit $\tau^2 \to \infty$ these estimates agree.

Lemma* 64. *We have $\Sigma(\Sigma + I)^{-1} y \to (I + Q)^{-1} y$ for any y as $\tau^2 \to \infty$.*

Proof. Recall that Q is SPSD. By the spectral theorem, Q can be diagonalized wrt. an orthonormal basis $\{q_i\}$ for $\mathbb{R}^n$ of eigenvectors of Q. Note that $(\sqrt{n})^{-1} 1_n = q_j$ for some j. Assume this is true for $j = 1$. It easily shown that $\{q_i\}$ are also eigenvectors of Q^+. If q_i is associated with eigenvalue $\lambda_i \neq 0$ then $Q^+ q_i = \lambda_i^{-1} q_i$. If $\lambda_i = 0$, then $Q^+ q_i = 0$. We now intend to prove that $\Sigma(\Sigma + I)^{-1} q_i \to (I + Q)^{-1} q_i$ as $\tau^2 \to \infty$ for all these eigenvectors. If we can show this, the lemma will follow since $\{q_i\}$ is a basis for $\mathbb{R}^n$.

By definition of inverse matrices, if K is invertible, then $Kx = y$ holds for vectors x, y if and only if $x = K^{-1} y$. Observe that

$$(\Sigma + I) 1_n = (Q^+ + \tau^2 1_n 1_n^\top + I) 1_n = (\tau^2 n + 1) 1_n.$$

From this we get $(\Sigma + I)^{-1} 1_n = \frac{1}{\tau^2 n + 1} 1_n$. Use this to obtain

$$\begin{aligned} \Sigma(\Sigma + I)^{-1} 1_n &= \frac{1}{\tau^2 n + 1} \Sigma 1_n \\ &= \frac{1}{\tau^2 n + 1} (Q^+ + \tau^2 1_n 1_n^\top) 1_n \\ &= \frac{\tau^2 n}{\tau^2 n + 1} 1_n. \end{aligned}$$

Note that this converges to 1_n as $\tau \to \infty$. The same holds for any scalar multiple of 1_n and in particular for $(\sqrt{n})^{-1}1_n$. Let q_i be any of other eigenvectors of Q so that $q_i \perp 1_n$. Then q_i is associated with one of the non-zero eigenvalues λ_i, of Q, and we get $Q^+ q_i = \lambda_i^{-1} q_i$. Since $q_i \perp 1_n$ we get $\Sigma q_i = (Q^+ + \tau^2 1_n 1_n^\top) q_i = \lambda_i^{-1} q_i$. Using this and the same technique as used above we get $(\Sigma + I)^{-1} q_i = (1 + \lambda_i^{-1})^{-1} q_i$. Combining this we get

$$\begin{aligned}\Sigma(\Sigma + I)^{-1} q_i &= \frac{1}{1 + \frac{1}{\lambda_i}} \Sigma q_i \\ &= \frac{1}{1 + \frac{1}{\lambda_i}} \frac{1}{\lambda_i} q_i \\ &= \frac{1}{1 + \lambda_i} q_i\end{aligned}$$

We compare this with the action of $(I + Q)^{-1}$ on this basis. In this case we get $(I + Q)1_n = 1_n$ and $(I + Q)q_i = (1 + \lambda_i)q_i$ whenever $q_i \perp 1_n$ is an eigenvector. Thus $(1 + Q)^{-1}1_n = 1_n$ and $(I + Q)^{-1} q_i = \frac{1}{1+\lambda_i} q_i$. The lemma follows from these identities. □

In the limit $\tau^2 \to \infty$ we thus get the same posterior estimate for $\mathbf{X}$ as what we obtained using the improper prior. By Lemma 63 we have the estimate $\hat{\alpha} = \tau^2 1_n^\top (\Sigma + I)^{-1} y$ in the proper model. If we define the overall level of $x \in \mathbb{R}^n$ as $\frac{1}{n} 1_n X$, which is just the arithmetic average of the components of x, we see as in the next lemma that α must be related to the overall level of $\mathbf{X}$.

Lemma* 65. *In the proper model, we have*

$$\hat{\alpha} = \frac{\tau^2}{\tau^2 n + 1} \left(\frac{1}{n} 1_n^\top y \right) = \frac{1}{n} 1_n^\top \hat{\mathbf{X}}.$$

Proof. By Lemma 63 we have $\hat{\alpha} = \tau^2 1_n^\top (\Sigma + I)^{-1} y$ and $\frac{1}{n} 1_n^\top \hat{\mathbf{X}} = \frac{1}{n} 1_n^\top \Sigma (\Sigma + I)^{-1} y$. We show that both equal to $\frac{\tau^2}{\tau^2 n + 1} 1_n^\top y$.

Recall from the proof of Lemma 64 that the orthonormal eigenvectors $\{q_i\}$ of Q are also eigenvectors of $(\Sigma + I)^{-1}$ and Σ. Assume as before that $q_1 = (\sqrt{n})^{-1} 1_n$ and express y in terms of this basis as $y = \sum_{i=1}^n y_i q_i$. Here we have $y_1 = \frac{1}{\sqrt{n}} 1_n^\top y$. Note that both $\hat{\alpha}$ and $\frac{1}{n} 1_n^\top \hat{\mathbf{X}}$ only depend on y through the $y_1 q_1$ term since the remaining q_i are orthogonal to 1_n. We therefore get

$$\begin{aligned}\hat{\alpha} &= \tau^2 1_n^\top (\Sigma + I)^{-1} y_1 q_1 \\ &= \frac{\tau^2 y_1}{\sqrt{n}(\tau^2 n + 1)} 1_n^\top 1_n \\ &= \frac{\tau^2 n}{\tau^2 n + 1} \left(\frac{1}{n} 1_n^\top y \right).\end{aligned}$$

Where we have used the identity $(\Sigma + I)^{-1} 1_n = (\tau^2 n + 1)^{-1} 1_n$. Calculating similarly for $\frac{1}{n} 1_n^\top \hat{\mathbf{X}}$ we get

$$\begin{aligned}\frac{1}{n} 1_n^\top \hat{\mathbf{X}} &= \frac{1}{n} 1_n^\top \Sigma (\Sigma + I)^{-1} y_1 q_1 \\ &= \frac{y_1 \tau^2 n}{n\sqrt{n}(\tau^2 n + 1)} 1_n^\top 1_n \\ &= \frac{\tau^2 n}{\tau^2 n + 1} \left(\frac{1}{n} 1_n^\top y \right).\end{aligned}$$

□

Note that taking the limit $\tau^2 \to \infty$, we see that $\hat{\alpha} \to \frac{1}{n}1_n^\top y$, which is just the overall level of y. It is of interest to compare this estimate with the overall level of the estimate $\hat{\mathbf{X}} = (I+Q)^{-1}y$.

Lemma* 66. *We have* $\frac{1}{n}1_n^\top (I+Q)^{-1}Y = \frac{1}{n}1_n^\top Y$.

Proof. As before let $\{q_i\}$ be an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of Q and write y as a linear combination $y = \sum_{i=1}^n y_i q_i$. Calculating as in the previous proofs, we get

$$\begin{aligned}\frac{1}{n}1_n^\top (I+Q)^{-1}y &= \frac{y_1}{n\sqrt{n}}1_n^\top (I+Q)^{-1}1_n \\ &= \frac{y_1}{n\sqrt{n}}1_n^\top 1_n \\ &= \frac{1}{n}1_n^\top y.\end{aligned}$$

□

This shows that the overall level in the proper estimate converges to the overall level in the improper estimate as $\tau^2 \to \infty$. From this we see that there is a strong connection between the use of the improper prior and the proper prior. Since $N(0,\tau^2)$ resembles the uniform distribution on $\mathbb{R}$ in the limit $\tau^2 \to \infty$ we can therefore informally think of using the improper prior $f(x) \propto \exp(-\frac{1}{2}x^\top Qx)$ to estimate $\mathbf{X}$ given $\mathbf{Y} = y$ as equivalent to using a uniform distribution for the overall level $\alpha 1_n$. We will now make this more precise in the following. To do this we will directly compare the improper distribution $f(x) \propto \exp(-\frac{1}{2}x^\top Qx)$ with the distribution of $\mathbf{W} + \alpha 1_n$ when $\mathbf{W}$ is degenerate on range(Q) with precision Q and α has a uniform distribution on $\mathbb{R}$.

Proposition* 67. *The distribution induced by the improper density in* (4.6) *is equal to the distribution of* $\mathbf{W} + \alpha 1_n$ *when* $\mathbf{W}$ *is a degenerate Gaussian with precision* Q *and mean* 0*, and* α *is independent of* $\mathbf{W}$ *and has a uniform distribution on* $\mathbb{R}$

Proof. We will use Theorem 90 in the appendix to prove equality. Write $Q = \mathcal{O}D\mathcal{O}^\top$, where D is a diagonal matrix containing the eigenvalues of Q. We assume that the first column of $\mathcal{O}$ is the $\frac{1}{\sqrt{n}}1_n$ vector so that $\lambda_1 := D_{11} = 0$.

It is well known that the set of Borel sets in $\mathbb{R}^n$ are generated by sets of the form $I \times B$ where $I = (a,b] \subset \mathbb{R}$ is an interval and $B \subset \mathbb{R}^{n-1}$ is of the $\prod_{i=2}^n (a_i, b_i]$. Similarly, sets of the form $\mathcal{O}(I \times B)$ also generate the set of Borel sets in $\mathbb{R}^n$ and are stable under taking intersections. Denote $\kappa(\mathcal{U}) = c\int_{\mathcal{U}} \exp(-\frac{1}{2}x^\top Qx)dx$ where c is the 'integration constant' and $\nu(\mathcal{U}) = P(\mathbf{W} + \alpha 1_n \in \mathcal{U})$. We have

$$\begin{aligned}\kappa(\mathcal{O}(I \times B)) &= \frac{\sqrt{\det{}^* Q}}{\sqrt{(2\pi)^{n-1}}} \int_{\mathcal{O}(I\times b)} \exp\left(-\frac{1}{2}x^\top Qx\right) dx \\ &= \frac{\sqrt{\det{}^* Q}}{\sqrt{(2\pi)^{n-1}}} \int_{I\times B} \exp\left(-\frac{1}{2}\sum_{i=1}^n \lambda_i x_i^2\right) dx \\ &= (b-a)\frac{\sqrt{\det{}^* Q}}{\sqrt{(2\pi)^{n-1}}} \int_B \exp\left(-\frac{1}{2}\sum_{i=2}^n \lambda_i x_i\right) dx_{-1}\end{aligned}$$

where the last equation follows by using Tonelli's theorem twice and x_{-1} is short hand for the vector $(x_2, \dots x_n)$. Note that $P(\mathbf{W} + \alpha 1_n \in \mathcal{O}(I \times B)) = P(\mathcal{O}^\top(\mathbf{W} + \alpha 1_n) \in I \times B)$. We can write $\mathcal{O}^\top(\mathbf{W} + \alpha 1_n)$ as the vector $(\alpha, \tilde{\mathbf{W}}_2, \dots, \tilde{\mathbf{W}}_n)$ where $\tilde{\mathbf{W}} \sim N(0, \tilde{D})$ with $\tilde{D}$ is a diagonal matrix consisting of the non-zero eigenvalues of Q. The latter claim follows by noting that

$\tilde{\mathbf{W}} = \tilde{\mathcal{O}}\mathbf{W}$ where the $\tilde{\mathcal{O}}$ is obtained from $\mathcal{O}$ by removing the first row and column. Then, since the map $w \mapsto \tilde{\mathcal{O}}w$ is measurable, we get that α is also independent of $\tilde{\mathbf{W}}$. Using these facts we get

$$\begin{aligned} P(\mathbf{W} + \alpha 1_n \in \mathcal{O}(I \times B)) &= P((\alpha, \tilde{\mathbf{w}}_2, \ldots, \tilde{\mathbf{w}}_n) \in I \times B) \\ &= P(\alpha \in I)P(\tilde{\mathbf{W}} \in B) \\ &= (b-a)\frac{\sqrt{\prod_{i=2}^n d_i}}{\sqrt{(2\pi)^{n-1}}} \int_B \exp(-\frac{1}{2}\sum_{i=2}^n \lambda_i \tilde{w}_i^2) d\tilde{W} \end{aligned}$$

The measures κ and ν are identical on sets of the form $\mathcal{O}(I \times B)$. Furthermore, it is quite clear that κ and ν assign finite measure to $U_k = \mathcal{O}(\prod_{i=1}^n (-k, k])$, $k \in \mathbb{N}$ and since $\mathbb{R}^n = \cup_{k=1}^\infty (U_k)$ we get that $\kappa = \nu$ by Theorem 90 in the appendix. □

Thus the distribution measure associated with the improper prior is identical to the distribution of the degenerate Gaussian $\mathbf{W}$ plus the random vector $\alpha 1_n$ where α has a uniform distribution on $\mathbb{R}$. This gives another way of interpreting the improper prior as a degenerate Gaussian plus a uniform distribution on the overall level $\alpha 1_n$.

4.1.4 Improper Gaussians

In Section 4.1.3 we saw that a random variable $\mathbf{X}$ with density $f(x) \propto \exp(-\frac{1}{2}x^\top Q x)$ wrt. the Lebesgue measure for a rank deficient Q can be directly compared with the distribution of the sum $\mathbf{W} + \alpha 1_n$ where $\mathbf{W}$ is degenerate with precision Q and α has a zero mean Gaussian distribution, and if we assume α has a uniform distribution on $\mathbb{R}$ we get equality of distributions. To define improper Gaussians we will let this inspire us. The improper density is seemingly easier to work with than things such as $\mathbf{W} + \alpha 1n$. We therefore define as follows.

Definition 68. *Let $\mathbf{X} \colon \Omega \to \mathbb{R}^n$ be a measurable map from some probability space Ω. Furthermore, let a SPSD matrix Q of rank $k < n$ and a vector μ be given. Then we say that $\mathbf{X}$ is an improper Gaussian if the map*

$$f(x) = \frac{\sqrt{\det{}^* Q}}{\sqrt{(2\pi)^k}} \exp\left(-\frac{1}{2}(x-\mu)^\top Q(x-\mu)\right)$$

is the density of $P_{\mathbf{X}}$ wrt. the n-dimensional Lebesgue measure.

We can compare improper Gaussians with degenerate Gaussians and variables with uniform distributions on $\mathbb{R}$ the same way we did in Proposition 67.

Theorem* 69. *Let $\mathbf{X}$ be an improper Gaussian with density*

$$f(x) = \frac{\sqrt{\det{}^* Q}}{\sqrt{(2\pi)^k}} \exp\left(-\frac{1}{2}(x-\mu)^\top Q(x-\mu)\right).$$

Assume that Q has rank $k < n$. Then $\mathbf{X}$ has the same distribution as the sum $\mathbf{W} + \sum_{j=k+1}^n \alpha_j u_j$ for a degenerate Gaussian $\mathbf{W} \sim N(\mu, Q)$, some vectors $u_j \in \mathbb{R}^n$ and random variables α_j independent of $\mathbf{W}$ and each other with a uniform distribution on $\mathbb{R}$.

Proof. Let $Q = \mathcal{O}D\mathcal{O}^\top$ be an eigenvalue decomposition where we denote the eigenvalues by λ_j. We let $\mathbf{W}$ have a degenerate Gaussian distribution on $\text{range}(Q) + \mu$ with precision Q and mean μ. We set $u_j = q_j$ if $\lambda_j = 0$. We can then prove this theorem just as we did Proposition 67 with only minor adjustments. □

4.2 Intrinsic Random Fields

In Section 3.3 we considered Gaussian fields for which $P(\mathbf{X}_i \neq \mathbb{E}[\mathbf{X}_i \mid \mathbf{X}_{-i}]) \neq 0$. We then derived a specification $\gamma_{Q,h}$ which $\mathbf{X}$ is specified by. We then took the converse approach, starting with a specification of the form $\gamma_{Q,h}$ and sought to derive properties (and existence) of Gaussian fields specified by $\gamma_{Q,h}$. We only considered the case where Q was positive definite. It was then expressed, as in e.g. Corollary 54, that only if Q has an inverse C in the sense of (3.16) a Gaussian random field specified by $\gamma_{Q,h}$ exists, and that if such a field $\mathbf{X}$ exists, then C is the covariance of $\mathbf{X}$. It is of interest to model C a bit more flexibly. In Proposition 55 we connected a covariance function $C\colon \mathbb{Z}^n \times \mathbb{Z}^n \to \mathbb{R}$ with an integral of the form

$$C(i,j) = \int_{(-1,1]^n} e^{i\pi\langle \omega, i-j\rangle} f(\omega) d\lambda_n(\omega).$$

for some appropriate function f. Note that by substitution we can rewrite this as an integral over $(-\pi,\pi]^n$, if we set $\tilde{f}(x) = \frac{1}{\pi^n} f(\frac{1}{\pi}x)$ we get

$$C(i,j) = \int_{(-\pi,\pi]^n} e^{i\langle \omega, i-j\rangle} \tilde{f}(\omega) d\lambda_n(\omega).$$

This relationship between C and $\tilde{f}$ has a name. Indeed, we say that $\tilde{f}$ is the spectral density C. In [18] such equations are taken as the starting point for modelling so-called intrinsic autoregressions. Before we continue with this, we will consider the spectrum of a random field.

Assume that we have a function $C\colon \mathbb{Z}^n \to \mathbb{R}$. Then C is positive semi-definite if and only if there exists a measure ν on $(-\pi,\pi]^n$ such that

$$C(k) = \int_{(-\pi,\pi]^n} e^{i\langle \omega, k\rangle} d\nu(\omega). \tag{4.7}$$

This result is sometimes known as Herglotz' theorem. Thus in the case that $\mathbf{X}$ is a random field with covariance function C which is translationally invariant we can define the spectrum of $\mathbf{X}$ as the measure ν in (4.7). We furthermore define the spectral distribution of $\mathbf{X}$ as the function $F(\omega) = \nu(\prod_{i=1}^n (-\pi, \omega_i])$ for $x \in (-\pi,\pi]^n$. Furthermore if ν has density f wrt. the Lebesgue measure on $(-\pi,\pi]^n$ we call the function f the spectral density of $\mathbf{X}$.

4.2.1 Intrinsic autoregressions in the plane

This section and the next is based on [18].

Intrinsic random fields can be thought of as generalizing stationary random fields. Indeed, we say that $\mathbf{X}$ is an intrinsic random field if certain collections of linear combinations of $\mathbf{X}_i$ are stationary. We motivate this definition with an example. Let $\epsilon \sim N_{\mathbb{Z}}(0, C)$ be Gaussian white noise, i.e. $C(i,j) = \delta_{ij}$ and assume that $\mathbf{Z} = (\mathbf{Z}_i)_{i\in\mathbb{Z}}$ satisfies $\mathbf{Z}_{i+1} = \mathbf{Z}_i + \epsilon_i$. Let us condition on $\mathbf{Z}_0 = b$ for some $b \in \mathbb{R}$. Then we get $\mathbf{Z}_i \mid \mathbf{Z}_0 = b \sim N(b, |i|)$. Since the variance depends on i, we get that $\mathbf{Z} \mid \mathbf{Z}_0 = b$ is not a stationary. However, the *difference* process defined by $\mathbf{Z}_i - \mathbf{Z}_{i-1} = \epsilon_{i-1}$ is stationary. Note that if we give some improper prior, say the Lebesgue prior on b, then $\mathbf{Z}_i - \mathbf{Z}_{i-1}$ still defines a stationary process. To define intrinsic random fields in the plane we therefore start by generalizing this idea of differencing.

Definition 70. *Let (λ_i) be a finite collection of coefficients indexed by $i \in J \subset \mathbb{Z}^2$ which satisfies $\sum_{i\in J} \lambda_i i_1^\alpha i_2^{\alpha'} = 0$ for all non-negative exponents α, α' satisfying $\alpha + \alpha' \leq d$.*

Let such a collection of coefficients be given, and let $\mathbf{X}\colon \Omega \to \mathbb{R}^{\mathbb{Z}^2}$ be any random field in the plane. Then for any $j \in \mathbb{Z}^2$ we call $\sum_{i\in J} \lambda_i \mathbf{X}_{i+j}$ an increment of order d. The process $\mathbf{Y}$ defined by $\mathbf{Y}_j = \sum_{i\in J} \lambda_i \mathbf{X}_{i+j}$ is called an increment process of order d of $\mathbf{X}$.

Assume that $d = 0$. Then the condition $\sum_{i\in J} \lambda_i i_1^{\alpha} i_2^{\alpha'} = 0$ just means $\sum_{i\in J} \lambda_i = 0$. In this case the increment $\sum_{i\in J} \lambda_i \mathbf{X}_{i+j}$ is known as a contrast. Note that the process defined by $\mathbf{Z}_i - \mathbf{Z}_{i-1} = \epsilon_{i-1}$ is an example of this.

Definition 71. *We say that a random field $\mathbf{X}\colon \Omega \to \mathbb{R}^{\mathbb{Z}^2}$ is intrinsic of order d if all increment processes of $\mathbf{X}$ have zero mean and are weakly stationary.*

Remark 72. *In light of the one dimensional example, an intrinsic process may be improper.*

Now that we have defined intrinsic random fields, we will develop the concept of autoregressions for them. For stationary Gaussian fields $\mathbf{X}$, autoregressive models are defined using a neighbourhood system N containing 0 and are of the form

$$\mathbb{E}[\mathbf{X}_i \mid \mathbf{X}_{-i}] = \mu_i - \sum_{k\in N} \beta_k(\mathbf{X}_{i+k} - \mu_k). \tag{4.8}$$

In an autoregressive model we thus express that the right hand side of (4.8) is a best linear predictor of $\mathbf{X}_i$ given $\mathbf{X}_{-i}$. As noted in [18] there are problems with generalizing such models to intrinsic fields since for an intrinsic field, $\mathbb{E}[\mathbf{X}_i \mid \mathbf{X}_{-i}]$ is no longer well-defined. We therefore have to approach the idea of autoregressions differently.

The spectral density of a stationary random field is by definition connected with the covariance structure of the field. Since $\mathbb{E}[\mathbf{X}_i \mid \mathbf{X}_{-i}]$ is also connected to the covariance structure, it is no surprise that there is a connection between $\mathbb{E}[\mathbf{X}_i \mid \mathbf{X}_{-i}]$ and the spectral density of f. In the case of autoregressions for stationary processes this expresses itself in terms of the coefficients (β_k).

Proposition* 73. *Let $\mathbf{X}$ be a stationary Gaussian field satisfying* (4.8)*. Assume that $\mathbf{X}$ has a spectral density f. Then if we denote $\sigma^2 = \mathbb{E}[(\mathbf{X}_i - \mathbb{E}[\mathbf{X}_i \mid \mathbf{X}_{-i}])^2]$ we have*

$$f(\omega) = \sigma^2 \left(1 + \sum_{k\in N} \beta_k \cos(\pi\langle k, \omega\rangle)\right)^{-1}. \tag{4.9}$$

Remark 74. *The scalar $\sigma^2 := \mathbb{E}[(\mathbf{X}_i - \mathbb{E}[\mathbf{X}_i \mid \mathbf{X}_{-i}])^2]$ is known as the prediction error.*

Proof. General spectral theory for random fields gives that the spectral density, if it exists, is given by

$$f(\omega) = \sum_{h\in\mathbb{Z}^2} C(h) e^{-i\langle h,\omega\rangle}. \tag{4.10}$$

We therefore start by calculating the covariances $C(h)$. For convenience let us denote $\hat{\mathbf{X}}_i = \mathbb{E}[\mathbf{X}_i \mid \mathbf{X}_{-i}]$. Assume that $i - j \neq 0$ then

$$\begin{aligned} C(i-j) &= \operatorname{Cov}(\mathbf{X}_i, \mathbf{X}_j) \\ &= \operatorname{Cov}(\mathbf{X}_i - \hat{\mathbf{X}}_i, \mathbf{X}_j) + \operatorname{Cov}(\hat{\mathbf{X}}_i, \mathbf{X}_j). \end{aligned}$$

We will now prove that $\operatorname{Cov}(\mathbf{X}_i - \hat{\mathbf{X}}_i, \mathbf{X}_j) = 0$. To see this, we will use the law of total covariance which says that $\operatorname{Cov}(\mathbf{X}, \mathbf{Y}) = \mathbb{E}[\operatorname{Cov}(\mathbf{X}, \mathbf{Y} \mid \mathbf{Z})] + \operatorname{Cov}(\mathbb{E}[\mathbf{X} \mid \mathbf{Z}], \mathbb{E}[\mathbf{Y} \mid \mathbf{Z}])$. Here we define conditional covariance by $\operatorname{Cov}(\mathbf{X}, \mathbf{Y} \mid \mathbf{Z}) = \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X} \mid \mathbf{Z}])(\mathbf{Y} - \mathbb{E}[\mathbf{Y} \mid \mathbf{Z}]) \mid \mathbf{Z}]$. We will thus condition on $\mathbf{X}_{-i}$ to obtain

$$\begin{aligned} \operatorname{Cov}(\mathbf{X}_i - \hat{\mathbf{X}}_i, \mathbf{X}_j) &= \mathbb{E}[\operatorname{Cov}(\mathbf{X}_i - \hat{\mathbf{X}}_i, \mathbf{X}_j \mid \mathbf{X}_{-i})] + \operatorname{Cov}[\mathbb{E}[\mathbf{X}_i - \hat{\mathbf{X}}_i \mid \mathbf{X}_{-i}], \mathbb{E}[\mathbf{X}_j, \mathbf{X}_{-i}]] \\ &= \mathbb{E}[\mathbb{E}[(\mathbf{X}_i - \hat{\mathbf{X}}_i - (\hat{\mathbf{X}}_i - \hat{\mathbf{X}}_i))(\mathbf{X}_j - \mathbf{X}_j) \mid \mathbf{X}_{-i}]] + \operatorname{Cov}[\hat{\mathbf{X}}_i - \hat{\mathbf{X}}_i, \mathbf{X}_j] \\ &= \mathbb{E}[0] + \operatorname{Cov}(0, \mathbf{X}_j) \\ &= 0 \end{aligned} \tag{4.11}$$

where we have used e.g. $\mathbb{E}[\mathbf{X}_j \mid \mathbf{X}_{-i}] = \mathbf{X}_j$ since $j \neq i$. Thus $C(i-j) = \mathrm{Cov}(\hat{\mathbf{X}}_i, \mathbf{X}_j)$. We will now need the hypothesis that $\hat{\mathbf{X}}_i = \alpha - \sum_{k \in N} \beta_k \mathbf{X}_{i+k}$ for a scalar α which does not depend on i. Note also that we can assume $N = \mathbb{Z}^d$ by letting $\beta_k = 0$ whenever $k \notin N$. Note in particular that $a_0 = 0$. Now using the expression for $\hat{\mathbf{X}}_i$, let $d = i - j$, then

$$\begin{aligned} C(d) &= \mathrm{Cov}(\hat{\mathbf{X}}_i, \mathbf{X}_j) \\ &= \sum_l \beta_l \mathrm{Cov}[\mathbf{X}_{i+l}, \mathbf{X}_j] \\ &= \sum \beta_l C(d+l) \end{aligned}$$

Using (4.11) several times we get $\mathrm{Cov}(\mathbf{X}_i - \hat{\mathbf{X}}_i, \hat{\mathbf{X}}_i) = 0$ by linearity. Now to calculate $C(0)$ chose some i to get

$$\begin{aligned} C(0) &= \mathbb{V}\mathrm{ar}[\mathbf{X}_i] \\ &= \mathbb{V}\mathrm{ar}[\mathbf{X}_i - \hat{\mathbf{X}}_i + \hat{\mathbf{X}}_i] \\ &= \mathbb{V}\mathrm{ar}[\mathbf{X}_i - \hat{\mathbf{X}}_i] + 2\mathrm{Cov}(\mathbf{X}_i - \hat{\mathbf{X}}_i, \hat{\mathbf{X}}_i) + \mathbb{V}\mathrm{ar}[\hat{\mathbf{X}}_i] \\ &= \sigma^2 - \mathbb{V}\mathrm{ar}[\hat{\mathbf{X}}_i] \\ &= \sigma^2 - \mathbb{V}\mathrm{ar}[\sum_d \beta_d \mathbf{X}_{i+d}] \\ &= \sigma^2 - \sum_d \sum_h \beta_d \beta_h C(i+d-(i+h)) \\ &= \sigma^2 - \sum_d \beta_d \sum_h \beta_h C(-d+h) \\ &= \sigma^2 - \sum_d \beta_k C(-d) \\ &= \sigma^2 - \sum_d \beta_k C(d) \end{aligned}$$

where we have used the expression for $C(d)$ and evenness of C. Expressions for $C(0)$ and $C(i-j)$, $i - j \neq 0$ can be summarized as

$$C(h) = \sigma^2 \, \mathbb{I}_{h=0} - \sum_k \beta_k C(k+h) \tag{4.12}$$

Insert this into (4.10).

$$\begin{aligned} f(\omega) &= \sum_k C(k) e^{-i\langle k, \omega \rangle} \\ &= C(0)e^0 + \sum_{k \neq 0} C(k) e^{-i\langle k, \omega \rangle} \\ &= \sigma^2 - \sum_h a_h C(h) - \sum_{k \neq 0} \left[\sum_h \beta_h C(k+h) \right] e^{-i\langle k, \omega \rangle} \\ &= \sigma^2 - \sum_k \left[\sum_h \beta_h C(k+h) \right] e^{-i\langle k, \omega \rangle} \end{aligned}$$

Since the β_h is non-zero for only finitely many h we can change the order of summation to get

$$\begin{aligned} f(\omega) &= \sigma^2 - \sum_h \beta_h \left[\sum_k C(k+h) e^{-i\langle k,\omega\rangle} \right] \\ &= \sigma^2 - \sum_h \beta_h \left[\sum_k C(k+h) e^{-i\langle k+h,\omega\rangle} e^{i\langle h,\omega\rangle} \right] \\ &= \sigma^2 - \sum_h \beta_h \left[\sum_k C(k) e^{-i\langle k,\omega\rangle} \right] e^{i\langle h,\omega\rangle} \\ &= \sigma^2 - \sum_h \beta_h f(\omega) e^{i\langle h,\omega\rangle} \end{aligned} \tag{4.13}$$

Finally solving for $f(\omega)$ yields

$$f(\omega) = \sigma^2 / (1 + \sum_h a_h e^{i\langle h,\omega\rangle})$$

provided $\sigma^2 \neq 0$. □

Remark 75. *Note as an immediate corollary to* (4.13) *that if* $\sigma^2 \neq 0$ *then* $1 + \sum_h a_h e^{i\langle h,\omega\rangle} > 0$ *and* $f(\omega) \neq 0$.

This connection shows that we could define autoregressions for stationary fields in terms of spectral densities. This is the approach we will take for intrinsic random fields. There is a problem though since we can not, in general, expect covariances $\mathbb{E}[(\mathbf{X}_i - \mu_i)(\mathbf{X}_j - \mu)]$ to be well-defined when $\mathbf{X}$ is intrinsic. We therefore have to define what it means for a function to be a spectral density for an intrinsic random field. Since certain increments are assumed stationary, it is more natural to connect 'intrinsic spectral densities' to covariances between increments as is done in [18].

Definition 76. *Let* $\mathbf{X}$ *be an intrinsic Gaussian field of order* d *and let* $f\colon (-\pi,\pi]^2 \to \mathbb{R}$ *be a function such that* f *is non-negative, even and* $\int_{(-\pi,\pi]^2} \|\omega\|^{2d+2} f(\omega) d\omega < \infty$. *If for increments* $\sum_j \lambda_j \mathbf{X}_j$ *and* $\sum_k \lambda'_k \mathbf{X}_k$ *we have*

$$\mathbb{E}[\sum_j \lambda_j \mathbf{X}_j \sum_k \lambda'_k \mathbf{X}_k] = (2\pi)^{-2} \int_{(-\pi,\pi]^2} \sum_j \lambda_j e^{i\langle j,\omega\rangle} \sum_k \lambda'_k e^{-i\langle k,\omega\rangle} f(\omega) d\omega \tag{4.14}$$

then f *is known as the (intrinsic) spectral density of* $\mathbf{X}$.

Remark 77. *The sums in* (4.14) *are finite by definition. Since we assume increments have zero mean, the left hand side of* (4.14) *is just a covariance.*

We now make some non-trivial claims. If $\mathbf{X}$ is an intrinsic Gaussian field of order d where every increment process of order $\leq d$ has an absolutely continuous spectrum, then $\mathbf{X}$ has a spectral density. Conversely, every function $f\colon (-\pi,\pi]^2 \to \mathbb{R}$ which is non-negative, even and

$$\int_{(-\pi,\pi]^2} \|x\|^{2d+2} f(\omega) d\omega < \infty$$

is the intrinsic spectral density of some intrinsic Gaussian field [18]. We do not have uniqueness, since a spectral density determines an intrinsic random field only up to addition with a polynomial of at most order d on $\mathbb{Z}^n$. [18] With this in mind, for a function f as in Definition 76, we define a collection of intrinsic field having f as their spectral density.

To define intrinsic autoregressions we use the connection between (4.9) and (4.8) as inspiration. Let $N \subset \mathbb{Z}^2$ be some a set of indices not containing 0 and closed under inversion, i.e. $N = -N$ and let β_k , $k \in N$ be scalars such that $\beta_k = \beta_{-k}$. If we assume that

$$P(\omega) := 1 + \sum_{k \in N} \beta_k \cos(\langle k, \omega \rangle) \geq \alpha \|\omega\|^{2d+2}$$

for all $x \in (-\pi, \pi]^2$ for some constant α, then the function f defined by $f(\omega) = \sigma^2 / P(\omega)$ satisfies the properties of an intrinsic spectral density of order d. Then a class of intrinsic random fields having f as their spectral density thus exist.

Definition 78 (Intrinsic autoregressions)**.** *With assumptions as above, the class of intrinsic fields corresponding to the spectral density function f is called an intrinsic autoregressive model with neighbourhood N and coefficients β_k.*

4.2.2 Best intrinsic predictors

The right hand side of (4.8) has the interpretation of being a best linear predictor. In this section we extend this notion to intrinsic random fields.

For intrinsic processes, it is not necessarily possible to express that some linear combination $\sum \lambda_j \mathbf{X}_j$ is the best linear predictor $\mathbf{X}_i$ given $\mathbf{X}_{-i}$. Indeed, variances of prediction errors are not always defined. [18]. We therefore define a smaller class of predictors for which variances of prediction errors are defined.

Definition 79 (Intrinsic predictors)**.** *Let $\mathbf{X}$ be an intrinsic random field. Then a finite linear combination $\sum \lambda_j \mathbf{X}_j$ over indexes $j \neq i$ is called an intrinsic predictor of $\mathbf{X}_i$ given $\mathbf{X}_{-i}$ if $\mathbf{X}_i - \sum \lambda_j \mathbf{X}_j$ is an increment. We say that $\sum \lambda_j \mathbf{X}_j$ is the best intrinsic predictor of $\mathbf{X}_i$ given $\mathbf{X}_{-i}$ if for any other intrinsic predictor $\sum \lambda'_j \mathbf{X}_j$ of $\mathbf{X}_i$ given $\mathbf{X}_{-i}$ we have*

$$\mathbb{E}[(\mathbf{X}_i - \sum \lambda_j \mathbf{X}_j)^2] \leq \mathbb{E}[(\mathbf{X}_i - \sum \lambda'_j \mathbf{X}_j)^2]$$

It is of interest whether best intrinsic predictors are connected to intrinsic densities similarly to how these concepts are connected in the stationary case. There is indeed such a connection as expressed by the following theorem.

Theorem 80. *Let f be of the form $f(\omega) = \sigma^2 / P(\omega)$ where $P(\omega) := 1 + \sum_{k \in N} \beta_k \cos(\langle k, \omega \rangle) \geq \alpha \|\omega\|^{2d+2}$ for $k \in N \subset \mathbb{Z}^2$, where N is closed under inversion, $0 \notin N$, and $\beta_k = \beta_{-k}$ for every $k \in N$.*

Assume that an intrinsic Gaussian field $\mathbf{X}$ has f as its spectral density. Then $\sum_{k \in N} \beta_k \mathbf{X}_{i+k}$ is the best intrinsic predictor of $\mathbf{X}_i$ given $\mathbf{X}_{-i}$ for all i.

Conversely, if $\mathbf{X}$ has some spectral density f and the finite linear combination $\sum \lambda_k \mathbf{X}_{i+k}$ for $k \in N = -N$ is the best intrinsic predictor of $\mathbf{X}_i$ given $\mathbf{X}_{-i}$ then

- *$\lambda_k = \lambda_{-k}$ and*
- *$f(x) = \mathbb{E}[\mathbf{X}_i - \sum \lambda_k \mathbf{X}_{i+k}] / (1 + \sum \lambda_k \cos(\langle k, x \rangle))$*

This theorem does two things. First off, it gives a precise sense in which an intrinsic autoregression is an autoregression; for an intrinsic autoregression we have some distinguished linear predictor $\sum \lambda_k \mathbf{X}_{i+k}$ which in a certain sense is the best. Secondly, it allows us to specify intrinsic autoregressions in terms linear combinations $\sum \lambda_k \mathbf{X}_{i+k}$ if we can argue that this linear combination is the best intrinsic predictor of $\mathbf{X}_i$ given $\mathbf{X}_{-i}$.

4.3 Final Remarks

Note that one of the ideas behind the definition of intrinsic random fields can roughly be phrased as follows; that while the random field $\mathbf{X}$ has unwieldy properties, some transformation of $\mathbf{X}$ may be well behaved. By analysing this transformation we may then be able to infer properties of $\mathbf{X}$. This is a general and useful idea. Indeed, this is reflected by the following observation.

Let $\mathbf{X}$ be a finite dimensional proper. Assume also that $\mathbf{X}$ is degenerate on some subspace V with precision matrix Q. Hausdorff measures give us complete control of calculating probabilities for $\mathbf{X}$, but let us pretend anyway, that this is an unwieldy distribution. Let an eigenvalue decomposition $Q = \mathcal{O}D\mathcal{O}^\top \in \mathbb{R}^{n\times n}$ be given and assume that the n'th diagonal component of D is zero and the remaining diagonal elements of Q are non-zero. This just means that $\mathbf{X}$ lives on some subspace orthogonal to the n'th column o_n of $\mathcal{O}^\top$. In light of this, let us transform $\mathbf{X}$ in such a way that we get rid of this extra dimension o_n. To do this let $\mathcal{O}_{1:n-1}$ be the matrix obtained by removing the n'th column of $\mathcal{O}$ and define $\mathbf{Y} = \mathcal{O}_{1:n-1}^\top \mathbf{X}$. Then $\mathbf{Y}$ is Gaussian and has precision given by $\tilde{D} = \operatorname{diag}(d_1, \dots, d_{n-1)})$ which has full rank. We have thusly transformed $\mathbf{X}$ into a non-degenerate Gaussian vector in $\mathbb{R}^{n-1}$. Furthermore, to calculate probabilities for $P(\mathbf{X} \in A)$ we can instead calculate probabilities

$$P(\mathcal{O}_{1:n-1}\mathbf{X} \in \mathcal{O}_{1:n-1}A) = P(\mathbf{Y} \in \tilde{A})$$

since the map $x \mapsto \mathcal{O}_{1:n-1}x$ is bijective when we restrict to $x \in V$.

Thus we may transform degenerate Gaussians $\mathbf{X}$ into non-degenerate Gaussians $\mathbf{Y}$ in such a way that we can calculate probabilities for $\mathbf{X}$ on the basis of $\mathbf{Y}$. Some theorems about finite dimensional GMRF depends on the precision matrix being full rank. In light of transformations of degenerate into non-degenerate Gaussians, this is not a problem. Indeed we can form simply form a model for $\mathbf{X}$ in which we claim that $A\mathbf{X}$ is non-degenerate for some known matrix A. One could then form a GMRF model for $A\mathbf{X}$. This could arguably be seen as a finite dimensional analogue of intrinsic models.

This approach completely fails if we assume that $\mathbf{X}$ is *improper* with precision Q. Indeed, in that case we can not assume that x is contained in some subspace V of dimension $n-1$. Thus the map $x \mapsto \mathcal{O}_{1:n-1}x$ is not bijective if we define it for all $x \in \mathbb{R}^n$ since its kernel is given by $\operatorname{span}(o_n)$. Thus $\mathbf{Y}^{-1}(\tilde{A}) = \mathbf{X}^{-1}(A + \operatorname{span}(o_n))$ for some set $A \subset V$. Now if $\tilde{A}$ has Lebesgue measure strictly greater than zero, the Hausdorff measure of A will be non-zero too, and we find that $\mathbf{Y}^{-1}(\tilde{A}) = \infty$. Conversely, if A has Lebesgue measure 0 then $\mathbf{Y}^{-1}(\tilde{A}) = 0$. The transformation $\mathbf{Y}$ thus contains no useful information about $\mathbf{X}$. Intuitively, this is not strange since $\mathbf{Y}$ only has information about the position of $\mathbf{X}$ in the directions orthogonal to o_n, while at the same time $\mathbf{X}$, in an arguably sense, has infinite variance in the o_n direction. This shows two things: that (a) linear transformations of improper Gaussians may yield rather pathological distributions, and thus (b) that one should be careful about treating improper distributions the same way one is treating proper distributions.

Appendix A

Preliminaries

A.1 Finite Dimensional Vector Spaces

In linear algebra we are concerned with finite dimensional vector spaces. We define these as follows: Let V be any vector space over $\mathbb{R}$. Then we say that a set B is a Hamel basis for $\mathbb{R}$ if every vector $v \in V$ can be written as a finite linear combination of vectors in B. By application of Zorn's lemma, one may show every vector space V has a Hamel basis. [7] A Hamel basis may be either finite or infinite. In both cases we have an invariance property known as the invariance of dimension which says that every Hamel basis for some vector space V has the same cardinality. E.g. if V has a finite Hamel basis, then every Hamel basis for V has the same size n. [7]. If the cardinality of a Hamel basis for V is finite, we say that V is finite dimensional and its dimension is defined as the cardinality of the basis. We otherwise say that V is infinite dimensional. In the following we will consider finite dimensional spaces.

A.1.1 Topology and Borel sets

Many things make working in finite dimensional vector spaces over $\mathbb{R}$ nice. Every vector space over $\mathbb{R}$ of dimension $n < \infty$ is linearly isomorphic with $\mathbb{R}^n$. Furthermore every inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^n$ is given by $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\top A \mathbf{y}$ for some symmetric positive definite (SPD) matrix $A \in \mathbb{R}^{n \times n}$. All norms on $\mathbb{R}^n$ are equivalent in the sense that if $\|\cdot\|_1$ and $\|\cdot\|_2$ are two norms on $\mathbb{R}^n$ then there exists scalars α_1, α_2 such that

$$\alpha_1 \|\mathbf{x}\|_1 \leq \|\mathbf{x}\|_2 \leq \alpha_2 \|\mathbf{x}\|_1, \quad \forall \mathbf{x} \in \mathbb{R}^n. \tag{A.1}$$

From this it follows that all norm induced topologies on $\mathbb{R}^n$ are the same. This topology is known as the Euclidean topology which we denote by $\tau(\mathbb{R}^n)$. The Euclidean topology is Hausdorff and second countable. Since it is second countable it is separable with the rational points being a well known example of a countable dense set. [5] We note that there exists metric topologies not equal to the Euclidean topology with one example being the discrete topology.

Note that for $n, m \in \mathbb{N}$ we can identify as vector spaces the Cartesian product $\mathbb{R}^n \times \mathbb{R}^m$ with $\mathbb{R}^{n+m}$. Something similar is true when we view this relationship topologically. Indeed, as topological spaces $\mathbb{R}^n$ and $\mathbb{R}^m$ induce a topology on $\mathbb{R}^{nm}$ known as the product topology, which is the least topology generated by the basis

$$\{U_n \times U_m \mid U_n, U_m \text{ are open in } \mathbb{R}^n \text{ and } \mathbb{R}^m \text{ respectively}\}.$$

Denote this topology by $\tau(\mathbb{R}^n) \times \tau(\mathbb{R}^m)$. A useful fact is that $\tau(\mathbb{R}^n) \times \tau(\mathbb{R}^m) = \tau(\mathbb{R}^{n+m})$ so that we can equally see the Euclidean topology as being induced by the Euclidean metric or as

products of lower dimensional Euclidean topologies [5]. This product topology can be defined for any finite number of arbitrary topological spaces and has a range of useful properties as described in [5]. We note then that we have two characterizations of the Euclidean topology in $\mathbb{R}^n$, either as induced by a norm (or metric), or as given by a product topology.

Now that we have a topology on $\mathbb{R}^n$ we can define the Borel algebra on $\mathbb{R}^n$ as the least σ-algebra containing all open sets. What 'Borel algebra' means depends on the chosen topology, but when working with $\mathbb{R}^n$ it will henceforth always be assumed that 'Borel algebra' means the Borel algebra induced by the Euclidean topology. We will similarly always understand measurability from or to these spaces in terms of this Borel algebra. We denote this by $\mathbb{B}(\mathbb{R}^n)$.

We have a construction for measurable spaces very similar to the product topology. We can equip $\mathbb{R}^{n+m}$ with Borel sets coming from the Euclidean topology or we can use the *product σ-algebra* of $\mathbb{R}^n$ and $\mathbb{R}^m$ which is defined as the least σ-algebra containing

$$\{\mathcal{O}_n \times \mathcal{O}_m \mid \mathcal{O}_n, \mathcal{O}_m \text{ are Borel sets in } \mathbb{R}^n \text{ and } \mathbb{R}^m \text{ respectively}\}$$

One can then show that this σ-algebra is equal to the Borel algebra $\mathbb{B}(\mathbb{R}^{n+m})$. [4] There is one interesting class of measures defined on the Borel algebra in $\mathbb{R}^n$ known as the Hausdorff measures. These measures contain the Lebesgue measure on $\mathbb{R}^n$ as a special case. We will consider these later.

A.1.2 Some useful linear algebra

We have two very useful results known as the Spectral Theorem and Singular Value Decomposition. These results are about linear maps and are most easily expressed when these maps are represented using matrices. The set of real valued matrices with dimension $n \times m$ will be denoted by $\mathbb{R}^{n\times m}$.

Theorem 81 (Spectral theorem for symmetric matrices)**.** *Let $A \in \mathbb{R}^{n\times n}$ be symmetric. Then there exists an orthogonal matrix $\mathcal{O} \in \mathbb{R}^{n\times n}$ and a diagonal matrix $D \in \mathbb{R}^{n\times n}$ such that*

$$A = \mathcal{O}D\mathcal{O}^\top. \tag{A.2}$$

A representation as the one in (A.2) is sometimes known as an eigenvalue decomposition. This is so because $\mathcal{O}$ contains eigenvectors of A and D contains eigenvalues. This theorem of course applies to symmetric positive semidefinite (SPSD) matrices. SPSD matrices have non-negative eigenvalues such that in that diagonal elements in the diagonal matrix in (A.2) are non-negative. In the symmetric and positive definite (SPD) case the diagonal elements are strictly positive. The singular value decomposition is similar but applies more generally.

Theorem 82 (Singular value decomposition)**.** *Let $A \in \mathbb{R}^{n\times k}$. Then there exists an orthogonal matrices $U \in \mathbb{R}^{n\times n}$ and $V \in \mathbb{R}^{k\times k}$ a diagonal matrix $D \in \mathbb{R}^{n\times k}$ such that*

$$A = UDV. \tag{A.3}$$

The term singular value decomposition is often abbreviated as SVD. Some linear maps T between finite dimensional spaces have an inverse denoted T^{-1}. It is well known that in order for T to be invertible, the domain and codomain of T must have the same dimension. But even if T is not invertible we have something similar to an inverse known as the Moore-Penrose inverse (MPI). This special 'pseudo'-inverse is denoted by T^+. If we represent T as a matrix A with SVD $A = UDV$ we may define A^+ as

$$A^+ = V^\top D^+ U^\top$$

where D^+ is obtained by transposing D and inverting all non-zero terms. One can then show that AA^+ is the orthogonal projection onto the range of A, and A^+A is the orthogonal projection onto the orthogonal complement of the kernel of A. In the case that A is invertible, the MPI of A is equal to the inverse.

Assume that V is any vector space over $\mathbb{R}$. Then a functional is a linear map $f\colon V \to \mathbb{R}$. The set of functionals are denoted by V^*. In finite dimensions, functionals take a specific form as given by the following proposition.

Proposition 83. *Every functional f in $(\mathbb{R}^n)^*$ is given by $f(\mathbf{x}) = a^\top \mathbf{x}$ for some $a \in \mathbb{R}^n$.*

Proof. The proof is a straightforward application of Riesz's representation theorem since every finite dimensional space is an inner product space when equipped with the dot product. □

A.1.3 The Lebesgue and Hausdorff Measures

Two classes of measures will be of special importance in this project, the Lebesgue and Hausdorff measures. The n dimensional Lebesgue measure, denoted λ_n, is a measure on $\mathbb{R}^n$ which formalizes and generalises our notions of length, area and volume. In particular, for any hyperrectangle $B = (a_1, b_1) \times \ldots \times (a_n, b_n) \subset \mathbb{R}^n$ we have

$$\lambda_n(B) = \prod_{i=1}^{n} b_i - a_i$$

and λ_n is in fact the only measure on $\mathbb{R}^n$ with this property. Consider the case $n = 3$ and let $S^2 = \{\mathbf{x} \in \mathbb{R}^3 \mid \|\mathbf{x}\| = 1\}$ be the unit sphere. Since S^2 is a two dimensional manifold it is no surprise that $\lambda_3(S^2) = 0$. The unit sphere, however, does have a non-zero surface area, but the Lebesgue measures are not directly capable of expressing that fact.

The Hausdorff measures solve this problem by, in part acting like a k dimensional Lebesgue measure pushed forward into $\mathbb{R}^n$. These measures are denoted by λ_{nk} and are defined for every pair (n, k) where $n \in \mathbb{N}$ and $k \geq 0$ is a real number. The Hausdorff measures are related to the Lebesgue measures through the following theorem.

Theorem 84. *Let $k \leq n$ be positive integers, $D \subset \mathbb{R}^k$ be Borel set and let $T\colon D \to \mathbb{R}^n$ be an a.e. smooth Borel measurable function with derivative T' . If $g\colon \mathbb{R}^n \to \mathbb{R}$ is non-negative and measurable, then for any measurable set $A \subset D$ we have*

$$\int_A g(T(x)) J_T(x) d\lambda_k(x) = \int_{T(A)} g(y) |A \cap T^{-1}(y)| d\lambda_{nk}(y).$$

where $J_T(x) = |\det(T'(x)T(x))|$ is known as the absolute Jacobian and $|A \cap T^{-1}(\mathbf{y})|$ is the cardinality of the set $A \cap T^{-1}(y)$. [2]

This theorem shows that calculating Hausdorff integrals in $\mathbb{R}^n$ for $k \in \mathbb{N}$ can be reduced to calculating Lebesgue integrals in $\mathbb{R}^k$ when a suitable transformation, $T\colon D \mapsto \mathbb{R}^n$, $D \subset \mathbb{R}^k$is known. For fixed n, and $k = 0$, λ_{n0} is the counting measure. For $k = 1$ and L a parametrised smooth curve , $\lambda_{n1}(L)$ is the arc length of L. For $k = n$, λ_{nn} is exactly the Lebesgue measure.[1][2] When $T(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$ where $A \in \mathbb{R}^{n \times k}$, $A^\top A = I$ Theorem 84 takes the particularly simple form

$$\int_A g(T(x)) d\lambda_k(x) = \int_{T(A)} g(y) d\lambda_{nk}(y),$$

[1]Some authors prefer to define Hausdorff measures in such a way that they are not equal but only proportional to the Lebesgue measures when $n = k$.

which shows that integrals on k-dimensional affine subspaces of $\mathbb{R}^n$ can be calculated as k dimensional Lebesgue integrals. See [2] for details. It should be noted that Hausdorff measures in general are not merely a pushed forward Lebesgue measure. Indeed, for $k \geq 0$ not an integer we can not use Theorem 84 to characterize the (n, k) Hausdorff measure.

A.2 Infinite products

Let (α_i) be an infinite sequence of real numbers. Then we say that $\prod_{i=1}^{\infty} \alpha_i$ exists if the sequence of partial products (p_n), $p_n = \prod_{i=1}^{n} \alpha_i$ converges to some real number p and in that case we define the value of $\prod_{i=1}^{\infty} \alpha_i$ to be p. Note that if at any index $\alpha_i = 0$, then the infinite product converges with value one. Note that any real $\alpha \neq 0$ has a unique representation as $\alpha = \operatorname{sign}(\alpha) \exp(\log |\alpha|)$. Assume for the time being that all $\alpha_i > 0$. Then we get

$$\log \left(\prod_{i=1}^{n} \alpha_i \right) = \sum_{i=1}^{n} \log(\alpha_i)$$

so that for a positive sequence (α_i) the convergence of the product of is equivalent to convergence of the series $\sum_{i=1}^{\infty} \log(\alpha_i)$.

It is not easy to chose a definition for infinite products over arbitrary index sets T. Luckily we will only need to consider the case where we take infinite products over values $(\alpha_i)_{i \in T}$ where $\alpha_i \in [0, 1]$ and T is countable. Such products have the following property.

Proposition* 85. *Let T be a countable and let $(\alpha_i)_{i \in T}$ be a real sequence with $\alpha_i \in [0, 1]$ for all $i \in T$. Then for any bijection $b \colon \mathbb{N} \to T$, the product $\prod_{n=1}^{\infty} \alpha_{b(n)}$ converges. If d is any other such bijection then $\prod_{n=1}^{\infty} \alpha_{b(n)} = \prod_{n=1}^{\infty} \alpha_{d(n)}$.*

Proof. Assume WLOG that $\alpha_i \neq 0$ for all $i \in T$ since otherwise all products $\prod_{n=1}^{\infty} \alpha_{b(n)}$ converge to 0 regardless of the bijection b. Note that for any bijection $b \colon \mathbb{N} \to T$, the sequence defined by $\prod_{n=1}^{N} \alpha_{b(n)}$ is monotonously decreasing and bounded from below by 0 and thus converges. It thus suffices to show that all such sequences converge to the same limit.

Note that using any other bijection d instead of b merely amounts to permuting the order in which the elements α_i are introduced into the product. More precisely, there exists a bijection $\pi \colon T \to T$ such that $d = \pi \circ b$. We thus only need to show that the infinite product is stable under arbitrary permutations.

Take logarithms to obtain

$$\lim_{N \to \infty} \log \left(\prod_{n=1}^{N} \alpha_{b(n)} \right) = \sum_{n=1}^{\infty} \log(\alpha_{b(n)}). \tag{A.4}$$

The elements $\log(\alpha_{b(n)})$ are in the interval $(-\infty, 0]$, so that the right-hand side of (A.4) either diverges monotonously to minus infinity or is absolutely convergent. In either case rearranging the summands lead to the same sum. In the case of absolute convergence, continuity of the exponential function yields that $\prod_{n=1}^{\infty} \alpha_{b(n)}$ is stable under rearranging terms. In the case of divergence to minus infinity, the fact that $\lim_{x \to -\infty} \exp(x) = 0$ implies stability under rearrangement. □

Thus if T is countable and $\alpha_i \in [0, 1]$ for all $i \in T$ we define $\prod_{i \in T} \alpha_i$ as the value of $\prod_{n=1}^{\infty} \alpha_{b(n)}$ whenever $b \colon \mathbb{N} \to T$ is any bijection.

Lemma* 86. *Let T be countable and $\alpha_i \in (0, 1]$ for all $i \in T$. If $\prod_{i \in T} \alpha_i > 0$ then $\alpha_{b(n)} \to 1$ for any bijection $b \colon \mathbb{N} \to T$.*

Proof. Let a bijection b such as in the statement be given. Then $\prod_{n=1}^{\infty} \alpha_{b(n)}$ converges to say $\Pi > 0$. Take logarithms to obtain $\sum_{n=1}^{\infty} \log(\alpha_{b(n)}) = \log(\Pi) \leq 0$, where the left-hand side is a sum of terms in the interval $(-\infty, 0]$ and $\log(\Pi) \in \mathbb{R}$. In order for this to be true, the sequence $(\log(\alpha_{b(n)}))$ can not contain infinitely values less than $-\epsilon$, for any $\epsilon > 0$ since otherwise, $\sum_{n=1}^{\infty} \log(\alpha_{b(n)})$ diverges to $-\infty$. But then $\log(\alpha_{b(n)}) \to 0$. □

A.3 Haar measures

Let $(\Omega, \tau, \circ)$ be a topological *group*, i.e. a $(\Omega, \circ)$ is a group, τ is a topology on Ω such that both the 'inversion operator' $i\colon \Omega \to \Omega$ and the group $\circ\colon \Omega \times \Omega \to \Omega$ are continuous. The inversion operator is the function mapping $\omega \in \Omega$ to its unique inverse element. Well known examples include the real line with addition or the complex unit circle with multiplication.

Furthermore we say that (Ω, τ) is locally compact, if each $\omega \in \Omega$ has a compact neighbourhood, i.e. there exists an open neighbour U of ω which is contained in a compact set $K \subset \Omega$. Assume furthermore that Ω is Hausdorff. Then it can be shown that any compact set is also closed. From this it follows that it is a Borel set wrt. to τ.

It is well known that on any locally compact Hausdorff topological group $(\Omega, \tau, \circ)$ we can define a special measure μ on the Borel sets of τ with the following properties:

1. μ is not identically zero.
2. For any Borel set A and element $\omega \in \Omega$ we have $\mu(A) = \mu(\omega \circ A)$, where $\omega \circ A = \{\omega \circ a \in \Omega\colon a \in A\}$.
3. For any compact set K we have $\mu(K) < \infty$.
4. For Borel sets $A \in \mathbb{B}(\tau)$ we have $\mu(A) = \inf\{\mu(U) \mid U \in \tau, A \subset U\}$
5. For any open set $U \in \tau$ we have $\mu(U) = \sup\{\mu(K) \mid K \text{ is compact with } K \subset U\}$

Such a measure μ is known as a Haar measure and exists and is unique upto scale, i.e. if μ and $\tilde{\mu}$ are two Haar measures on the same locally compact Hausdorff topological group then $\mu = c\tilde{\mu}$ for some real number $c > 0$. Properties number 4. and 5. are respectively known as outer and inner regularity.

Example 87. *As noted above the real line with addition is a topological group. The same holds true for $\mathbb{R}^n$ with addition. Furthermore these spaces are locally compact and metrizable and thus Hausdorff. Incidentally there exists a Haar measure on $(\mathbb{R}^n, +)$ which is unique up to scale. Now note that the Lebesgue measure λ_n is translationally invariant and is finite on bounded sets and thus also on compact sets. It can also be shown that λ_n is outer and inner regular too. It thus follows that λ_n is the unique Haar measure on $(\mathbb{R}^n, +)$ assigning unit measure to the unit cube $[0, 1]^n$.*

Example 88. *Denote the complex unit circle $K = \{z \in \mathbb{C} \mid |z| = 1\}$. With the natural metric induced by modulus this is a topological group. Denote by K^n the product space of n copies of K which we endow with the product topology and a composition defined by component wise multiplication. Then this is a topological group which is locally compact and Hausdorff. Thus it has a Haar measure. Let ν be such a Haar measure which assigns total measure 1 to K^n. Then ν is equal to the image of the Lebesgue measure on $]-1, 1]^n$ under the map*

$$(x_1, \ldots, x_n) \mapsto (\exp(i\pi x_1), \ldots, \exp(i\pi x_n)). \tag{A.5}$$

See [14]. Note that this map is a group isomorphism. If we denote the isomorphism by f this means that we can calcuate the Haar measure as $\nu(A) = \lambda_n(f^{-1}(A))$.

A.4 Some measure defined in [14]

Define the measures $\lambda_i(\cdot \mid X_{-i})$ for vectors $X \in \mathbb{R}^{\mathbb{Z}^n}$ on $\mathbb{R}^{\mathbb{Z}^n}$ by

$$\lambda^J(A \mid X) = \lambda_J(\{Y_J \in \mathbb{R} \mid Y \in \mathbb{R}^{\mathbb{Z}^n} \text{ with } Y_{-J} = X_{-i}\})$$

where λ_J is the usual Lebesgue measure on $\mathbb{R}^J$. Note that λ_i is something like a Lebesgue measure of $A \subset \mathbb{R}^{\mathbb{Z}^n}$ where we condition on $A_{-i} = \{X_{-i}\}$. This is formalized by the following lemma.

Lemma* 89. *For $\lambda_J(\cdot \mid X)$ integrable functions $f \colon \mathbb{R}^{\mathbb{Z}^n} \to \mathbb{R}$ we have*

$$\int_A f d\lambda^J(\cdot \mid x) = \int f(yx_{-J}) \mathbb{I}_A(yx_{-J}) d\lambda_J(y)$$

where yx_{-J} for $y \in \mathbb{R}^J$ is the vector defined by

$$(yx_{-J}) = \begin{cases} y_i & i \in J \\ x_i & i \notin J \end{cases}$$

Proof. We start by noting that $\lambda^J(A \mid x) = \int \mathbb{I}_A(yx_{-J}) d\lambda_J(y)$ which follows directly from the definition of λ^J. From this we get

$$\int \mathbb{I}_B(yx_{-J}) \mathbb{I}_A(yx_{-J}) d\lambda_J(y) = \lambda^J(A \cap B \mid x) = \int_A \mathbb{I}_B d\lambda^J(\cdot \mid x).$$

The result then follows from the standard proof. □

A.5 Miscellaneous

Theorem 90. *Let $(\Omega, \mathcal{F})$ be a measurable space and assume we have a class $\mathcal{G} \subset \mathcal{F}$ such that $\mathcal{G}$ generates $\mathcal{F}$ and is stable under intersections. Assume that μ, ν are measures on Ω such that $\mu(G) = \nu(G)$ for all $G \in \mathcal{G}$. Assume furthermore that there is a sequence of sets (G_n) in $\mathcal{G}$ such that $\mu(G_n) < \infty$ for all n and that $\cup G_n = \Omega$. Then $\mu = \nu$ everywhere.*

See e.g. [4].

Appendix B

Some proofs

Proposition 91. *The sum of independent Gaussians is also Gaussian.*

Proof. Let $\mathbf{V} = A_1\mathbf{Z}_1 + b_1$ and $\mathbf{W} = A_2\mathbf{Z}_2 + b_2$ where $\mathbf{V}$ and $\mathbf{W}$ are independent. Assume that A_1 and A_2 are injective, so that neither $\mathbf{V}$ nor $\mathbf{W}$ are completely degenerate. The completely degenerate case is trivial. Then

$$\mathbf{V} + \mathbf{W} = (A_1, A_2)\begin{pmatrix}\mathbf{Z}_1\\ \mathbf{Z}_2\end{pmatrix} + b_1 + b_2$$

where $(\mathbf{Z}_1, \mathbf{Z}_2)^\top$ is a standard normal. Independence of $\mathbf{Z}_1$ and $\mathbf{Z}_2$ can be shown using the fact that A_1 and A_2 are injective. □

Lemma* 92. *Let $\mathbf{X}$ be multivariate normal with transition matrix $A \in \mathbb{R}^{nk}$ and mean vector $\mu \in \mathbb{R}^n$. Assume that A is not the zero matrix. Then there exists a matrix of full rank $B \in \mathbb{R}^{nh}$ and a standard Gaussian $\tilde{\mathbf{Z}}$ of dimension h such that $\mathbf{X} = B\tilde{\mathbf{Z}} + \mu$.*

Proof. Let $A = UDV$ be the singular value decomposition of A. Let $\mathbf{Y} = DV\mathbf{Z}$. Note that $\mathbf{Y}$ has precisely h (generally) non-zero components. Let $\tilde{\mathbf{Y}}$ be those components. We can find a $(0,1)$ matrix $P : \mathbb{R}^n \to \mathbb{R}^h$ such that $\tilde{\mathbf{Y}} = PDV\mathbf{Z}$. We note that in order for this to work, the range of A needs to be strictly greater than 0 (otherwise all components are always equal to 0)

The matrix $\tilde{\Sigma} = (PDV)(PDV)^\top \in \mathbb{R}^{h\times h}$ is the covariance matrix of $\tilde{\mathbf{Y}}$. One can show that this matrix has full rank, so that it has a so-called reciprocal square root. I.e. a matrix $B^{-1/2}$ satisfying $B^{-1/2}\tilde{\Sigma}(B^{-1/2})^\top = I$. This reciprocal square root can be derived using the spectral theorem. Define $\tilde{\mathbf{Z}} = B^{-1/2}\tilde{\mathbf{Y}}$. It then follows that $\tilde{\mathbf{Z}}$ is standard normal. Thus

$$\mathbf{X} = UP^\top B^{1/2}\tilde{\mathbf{Z}} + \mu$$

where the map $UP^\top B^{1/2}$ has full rank and $B^{1/2}$ is the inverse of $B^{-1/2}$. □

Lemma* 93. *Let $f : \mathbb{Z}^2 \to \mathbb{Z}$ be a bijection. Then for every $d \in \mathbb{N}$, there exists a $(i,j) \in \mathbb{Z}^2$ such that $|f(i,j) - f(i,j+1)| \geq d$.*

Proof. Assume to the contrary that there exists some $D \in \mathbb{N}$ such that for all (i,j) we have $|f(i,j) - f(i,j+1)| \leq D$. Consider the sets $F_i = f(i,\mathbb{Z})$. Call two elements $x, y \in F_i$ neighbours if they are not equal, and there are no elements in F_i strictly between them. We claim that the distance between neighbours must be at most D.

Assume to the contrary that $x, y \in F_i$ are neighbours with $x - y > D$. Find j, k such that $x = f(i,j)$ and $y = f(i,k)$. By assumption we have $|j - k| > 2$. Assume WLOG that $j < k$. Note that $f(i,j+1) < y$ since otherwise we have $y < f(i,j+1) \leq y + d < x$ which contradicts y

and x being neighbours. Similarly, by induction, for any $0 < h < k - j$ we have $f(i, j + h) < y$. But then $f(i, k - 1) < y$ while $f(i, k) = x$ so that $|f(i, k - 1) - f(i, k)| > D$ which contradicts our assumption. The proof is similar if $y - x > D$, since in this case we can prove $f(i, k - h) < x$ for all $0 < h < k - j$.

By injectivity of f, all image sets $f(i, \mathbb{Z})$ are infinite for every $i \in \mathbb{Z}$ so that one of the sets $U_i\{z \in f(i, \mathbb{Z}) \mid z \geq 0\}$, $B_i = \{z \in f(i, \mathbb{Z}) \mid z < 0\}$ is infinite. Furthermore, again by injectivity, U_i is infinite for infinitely many i or B_i is. To see this, assume it were not the case. Then U_i would be infinite for only finitely many i and likewise for B_i. But then we can find j such that both U_i and B_i were finite which contradicts $f(i, \mathbb{Z})$ being infinite.

Assume WLOG that U_i is infinite for infinitely many i and let the indices such that this is true be written as $i_1, i_2, \ldots$ or more simply, and with some abuse of notation, as $1, 2, \ldots$. Let $z_i = \min U_i$ which exist by the well-ordering of $\mathbb{N}$. Set $w = \max_{1 \leq i \leq D+2} z_i$. We claim that all U_i for $1 \leq i \leq D + 2$ must at least one element between w and $w + D$.

All U_i for $1 \leq i \leq D + 2$ have a value less than or equal to w, and since they by construction have infinitely many values greater than 0, all U_i also contain elements greater than $w + D$. But then since neighbours have distance less than D, the U_i must also contain an element between w and $w + D$. But there is only enough space for $D + 1$ such values, but we have chosen $D + 2$ sets so by the Pigeon hole principle there are at least two U_i with non-empty intersection. But this contradicts f being injective. □

Theorem 94. *The product topology on $\prod \mathbb{X}_i$ of countably many metrizable spaces $\mathbb{X}_i$ is metrizable. If ρ_i generates the topology on $\mathbb{X}_i$, then*

$$\rho = \sum_{i=1}^{\infty} 2^{-i} \frac{\rho_i}{1 + \rho_i}$$

induces the product topology on $\prod \mathbb{X}_i$.

Proof. This proof is based on a proof found in [10]. We do not show that $\sum_{i=1}^{\infty} 2^{-i} \frac{\rho_i}{1+\rho_i}$ is a metric.

We start by showing that the metric topology $\tau(\rho)$ induced by ρ is finer than the product topology τ on $\prod \mathbb{X}_i$. To do this, we will use Theorem 13. Let an index i be given and let π_i be the projection $\pi_i(x) = x_i$. We have to show that π_i is continuous between the metric spaces $(\prod \mathbb{X}_i, \rho)$ and $(\mathbb{X}_i, \rho_i)$. Note that if $\delta < \frac{1}{2}$ we have

$$\frac{\rho_i}{1 + \rho_i} < \delta \implies \rho_i < \frac{\delta}{1 - \delta} < 2\delta$$

Since we also have that $\rho(\mathbf{X}, \mathbf{y}) < \delta$ implies $\frac{\rho_i}{1+\rho_i} < 2^i\delta$ we have $\rho_i(\pi_i(x), \pi_i(y)) < \epsilon$ if $\rho(x, y) < \delta$ where $2\delta < \max\{2^{-i}\epsilon, \frac{1}{2}\}$. Continuity of π_i follows. Since the product topology is the coarsest topology such that these projections are continuous we get that $\tau \subset \tau(\rho)$

The theorem follows if we can show that every $\tau(\rho)$ open set is also open wrt. τ. It suffices to show that for every $x \in \prod \mathbb{X}_i$ and open ball $B_\rho(x, r)$ there exists a $U \in \tau$ such that $x \in U$ and $U \subset B_\rho(x, r)$.

The open balls $B_{\rho_i}^{(i)}(x_i, r_i) = \{y_i \in X_i \mid \rho_i(x_i, y_i) < r_i\}$ are open in $\mathbb{X}_i$. Thus $U_i(r_i) := \pi_i^{-1}(B_{\rho_i}^{(i)}(x_i, r))$ is also open in $\prod \mathbb{X}_i$ wrt. to the product topology. Furthermore any finite intersection of such sets is also open. Note that $y \in U_i(r_i)$ if and only if $\rho_i(y_i, x_i) < r_i$.

Choose k large enough such that $\sum_{i=k+1}^{\infty} 2^{-i} < r/2$ and assume that $y \in \bigcap_{i=1}^{k} U_i$.

$$\begin{aligned}\rho(\mathbf{Y},\mathbf{X}) &= \sum_{i=1}^{k} 2^{-i}\frac{\rho_i(y_i,x_i)}{1+\rho_i(y_i,x_i)} + \sum_{i=k+1}^{\infty} 2^{-i}\frac{\rho_i(y_i,x_i)}{1+\rho_i(y_i,x_i)} \\ &= \sum_{i=1}^{k} 2^{-i}\rho_i(x_i,y_i) + \sum_{i=k+1}^{\infty} 2^{-i} \\ &= \sum_{i=1}^{k} 2^{-i} r_i + \frac{1}{2}r\end{aligned}$$

Where we have used that $\rho_i/(1+\rho_i) \leq 1$. Thus if we chose $r_i = \frac{1}{2}r$ for all $i \leq k$ we get $\sum_{i=1}^{k} 2^{-i} r_i < \frac{1}{2}r$. It follows that $\bigcap_{i=1}^{k} U_i \subset B_\rho(x,r)$.

□

Theorem* 95. *Let $(\Omega,\mathcal{F},P)$, $\mathbf{X}\colon \Omega \to \mathbb{X}$ and $\mathcal{G} \subset \mathcal{F}$ be as above and define $\gamma(A,\omega) = \mathbb{E}[\mathbb{I}_{\mathbf{X}\in A} \mid \mathcal{G}](\omega)$. Then*

1. *For every $A \in \mathcal{X}$ the map $\omega \mapsto \gamma(A,\omega)$ is $\mathcal{G}$ measurable.*
2. *For all A, $0 \leq \gamma(A,\cdot) \leq 1$ P-a.e. Furthermore, for P-a.a. ω $\gamma(\emptyset,\omega) = 0$ and $\gamma(\Omega,\omega) = 1$.*
3. *Let (A_n) be a sequence of disjoint measurable sets. Then for P-a.a. ω we have*
$$\sum \gamma(A_n,\omega) = \gamma(\bigcup A_n,\omega)$$
4. *For $A \in \mathcal{X}$ we have*
$$P(\{\mathbf{X}\in A\}\cap B) = \int_B \gamma(A,\cdot)dP$$

Proof. 1. We only prove the result in the case that $\mathbf{X}$ is the identity map and $(\mathbb{X},\mathcal{X}) = (\Omega,\mathcal{F})$. The proof is essentially the same in the general case. The first part is trivial, since by definition $\gamma(A,\omega) = \mathbb{E}[\mathbb{I}_A \mid \mathcal{G}](\omega)$ is $\mathcal{G}$ measurable as a map of ω.

2. Note that for any $A \in \mathcal{F}$ and $G \in \mathcal{G}$ we have

$$\int_G \gamma(A,\cdot)dP = \int_G \mathbb{I}_A dP = P(A\cap G) \geq 0.$$

By integrating over the sets $\{\gamma(A,\cdot) < 0\}$ and $\{\gamma(A,\cdot) \geq 0\}$, both of which are $\mathcal{G}$-measurable. It follows from this that $\gamma(A,\cdot)$ is positive P-almost everywhere since. One similarly shows that $\gamma(A,\cdot) \leq 1$, $\gamma(\emptyset,\cdot) = 0$ and $\gamma(\Omega,\cdot) = 1$ P- a.e.

3. Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of disjoint sets. Define $B_n = \{\gamma(A_n,\cdot) \geq 0\}$. Note that B_n is measurable and its complement is a P-null set. Define $B := \cap B_n$. This set is also measurable and its complement a P-null set. Note that the functions $\sum_{n=1}^{N} \mathbb{I}_B\gamma(A_n,\cdot)$ are $\mathcal{F}$-measurable and that $\sum_{n=1}^{N} \mathbb{I}_B\gamma(A_n,\cdot) \nearrow \sum \mathbb{I}_B\gamma(A_n,\cdot)$ so that the limit is $\mathcal{F}$-measurable too. Note that the series always exist as an element in the extended real line since $\mathbb{I}_B\gamma(A_n,\cdot) \geq 0$. Define the set $C_1 = \{\sum \mathbb{I}_B\gamma(A_n,\cdot) \geq \gamma(\bigcup A_n,\omega)\}$ and note that it is $\mathcal{F}$-measurable. Apply Lebesgue

monotone convergence to obtain

$$\begin{aligned}\int_{C_1}\left(\sum \mathbb{I}_B\gamma(A_n,\cdot)\right)-\gamma(\cup A_n,\cdot)dP &= \int_{C_1}\sum \mathbb{I}_B\gamma(A_n,\cdot)dP-\int_{B_1}\gamma(\cup A_n,\cdot)dP\\ &= \sum\int_{C_1}\mathbb{I}_B\gamma(A_n,\cdot)dP-\int_{B_1}\gamma(\cup A_n,\cdot)dP\\ &= \sum\int_{C_1}\mathbb{I}_{A_n}dP-\int\mathbb{I}_{\cup A_n}dP\\ &= \sum P(A_n\cap C_1)-P((\cup A_n)\cap C_1)\\ &= 0.\end{aligned}$$

Which shows that $\sum \mathbb{I}_B\gamma(A_n,\omega)=\gamma(\bigcup A_n,\omega)$ P-almost everywhere on C_1. A similar result holds on $C_2=\Omega\setminus C_1$ so that $\sum \mathbb{I}_B\gamma(A_n,\omega)=\gamma(\bigcup A_n,\omega)$ everywhere except on the union of complements, $C_1^c\cup C_2^c$, which is a P-null set.

4. By definition of conditional expectation, for $B\in\mathcal{G}$, we get

$$\begin{aligned}\int_B\gamma(A,\cdot)dP &= \int_B\mathbb{E}[\mathbb{I}_{\mathbf{X}\in A}\mid\mathcal{G}]dP\\ &= \int_B\mathbb{I}_{\mathbf{X}\in A}dP\\ &= P(\{\mathbf{X}\in A\}\cap B).\end{aligned}$$

□

Lemma 96. *Let $(\Omega,\mathcal{F},P)$ be a probability space and $\mathbf{X}\colon\Omega\to\mathbb{X}$ is a measurable map into a Polish space $\mathbb{X}$. Then for any σ-algebra $\mathcal{G}\subset\mathcal{F}$ there exists a function $\gamma\colon\mathcal{X}\times\Omega\to[0,1]$ s.t. for all $\omega\in\Omega$, $\gamma(\cdot,\omega)$ is a probability measure and for all A, $\gamma(A,\cdot)=\mathbb{E}[\mathbb{I}_{\mathbf{X}\in A}\mid\mathcal{G}]$ P-a.e.*

Proof. We proceed as in [16]. We only consider the case where $\mathbb{X}$ is uncountable. Then per the discussion above we may assume that $(\mathbb{X},\mathcal{X})=([0,1],\mathcal{B})$.

Step 1: Construction a function $F\colon\mathbb{R}\times\Omega\to\mathbb{R}$
Start by considering the sets $A=[0,q)$ for $q\in\mathbb{Q}$ of which there are only countably many. We thus define a function $F\colon\mathbb{Q}\times\Omega\to\mathbb{R}$ by

$$F(q,\cdot)=\mathbb{E}[\mathbb{I}_{\mathbf{X}\in(-\infty,q]}\mid\mathcal{G}].$$

We now use Theorem 30 and general properties of conditional expectations to get the following:

1. For all $q\in\mathbb{Q}$ there exists a null set $\mathcal{N}_q$ outside of which $F(q,\cdot)$ satisfies $0\le F(q,\cdot)\le 1$.
2. There is a null set $\mathcal{B}$ such that for all $q\in\mathbb{Q}$ with $q<0$ $F(q,\cdot)=0$ on $\mathcal{B}$. This follows from the fact that range$(\mathbf{X})\subset[0,1]$. Similarly there is a null set $\mathcal{U}$ such that for rational number $q>1$ we have $F(q,\cdot)=1$ on $\mathcal{U}$.
3. For all $q\le\tilde{q}$ both in $\mathbb{Q}\cap[0,1]$ we have $\mathbb{I}_{\mathbf{X}\in[0,q)}\le\mathbb{I}_{\mathbf{X}\in[0,\tilde{q})}$ so that there exists a null set $\mathcal{N}_{q\le\tilde{q}}$ out side of which $F(q,\cdot)\le F(\tilde{q},\cdot)$.

Note that there are only countably many such sets $\mathcal{N}_q$ and $\mathcal{N}_{q\le\tilde{q}}$. Thus the union of these sets and $\mathcal{B}$ and $\mathcal{U}$, denoted $\mathcal{N}$, is a null set too. Note that for $\omega\notin\mathcal{N}$, F has the nice properties that for any monotonously increasing sequence of rationals $q_n\in\mathbb{Q}$ the sequence $F(q_n,\omega)$ is

monotonously increasing and bounded. Thus the restriction of F to $\mathbb{Q}\times\Omega\setminus\mathcal{N}$ has an extension to $\mathbb{R}\times\Omega\setminus\mathcal{N}$. Indeed, for irrational x we can find a sequence q_n converging to x monotonously from above. We then define $F(x,\omega)=\lim F(q_n,\omega)$ for $\omega\notin\mathcal{N}$. With some abuse of notation we denote this extension by F too. We can also extend it in the other variable. To do this, for $\omega\in\mathcal{N}$, we set $F(x,\omega):=P(\mathbf{X}^{-1}([0,x)))$. We thus have a function $F\colon\mathbb{R}\times\Omega\to[0,1]$ which is non-decreasing and right-continuous and $\lim_{x\nearrow\infty}F(x,\omega)=1$ and $\lim_{x\searrow-\infty}F(x,\omega)=0$.

Step 2: Defining a probability kernel $\gamma\colon\mathbb{B}(\mathbb{R})\colon\mathbb{R}$
For such functions there exists a unique probability measure $\gamma\colon\mathbb{B}(\mathbb{R})\times\Omega\to[0,1]$ such that $\gamma((\infty,x],\omega)=F(x,\omega)$. This measure can be defined by e.g. the Stieltjes integral

$$\gamma(B,\omega)=\int_B dF(\cdot,\omega).$$

See e.g. [16] for details. This function γ is our candidate for a regular conditional distribution for $\mathbf{X}$ given $\mathcal{G}$. We do this directly by verifying that γ has the defining properties. By construction, $A\mapsto\gamma(A,\omega)$ is a probability measure on $\mathbb{R}$ for *all* $\omega\in\Omega$. It remains to prove that for all $A\in\mathbb{B}(\mathbb{R})$, $\gamma(A\cdot)$ is equal to $\mathbb{E}[\mathbb{I}_{\mathbf{X}\in A}\mid\mathcal{G}]$ P-a.e.

Step 3: Relating the probability kernel to the distribution of X
Define the set $\mathbb{D}=\{A\in\mathbb{B}(\mathbb{R})\colon\gamma(A,\cdot)=\mathbb{E}[\mathbb{I}_{\mathbf{X}\in A}\mid\mathcal{G}]P\text{-a.e.}\}$. We will use the π-λ theorem to show that $\mathbb{D}=\mathbb{B}(\mathbb{R})$.

- Clearly, $\mathbb{R}\in\mathbb{D}$ since $\gamma(\mathbb{R},\cdot)=1$ and $\mathbb{E}[\mathbb{I}_{\mathbf{X}\in\mathbb{R}}\mid\mathcal{G}]=1$ P-a.e.

- Let $A\in\mathcal{D}$. Then

$$\begin{aligned}\gamma(A^c,\cdot)&=1-\gamma(A,\cdot)\\&=1-\mathbb{E}[\mathbb{I}_{\mathbf{X}\in A}\mid\mathcal{G}]\\&=\mathbb{E}[\mathbb{I}_{\mathbf{X}\in\mathbb{R}}\mid\mathcal{G}]-\mathbb{E}[\mathbb{I}_{\mathbf{X}\in A}\mid\mathcal{G}]\quad P\text{-a.e.}\\&=\mathbb{E}[\mathbb{I}_{\mathbf{X}\in A^c}\mid\mathcal{G}]\quad P\text{-a.e.}\end{aligned}$$

  where we have used almost sure linearity of conditional expectation.

- Let (A_n) be a sequence of disjoint sets in $\mathbb{D}$. Since the sets are disjoint we have $\sum_{n=1}^N\mathbb{I}_{\mathbf{X}\in A_n}\nearrow\mathbb{I}_{\mathbf{X}\in\cup A_n}$ so that $\sum_{n=1}^N\mathbb{E}[\mathbb{I}_{\mathbf{X}\in A_n}\mathcal{G}]\nearrow\mathbb{E}[\mathbb{I}_{\mathbf{X}\in\cup A_n}\mid\mathcal{G}]$ P-a.e. by monotone convergence. Thus

$$\begin{aligned}\gamma(\bigcup A_n,\cdot)&=\sum\gamma(A_n,\cdot)\\&=\sum\mathbb{E}[\mathbb{I}_{\mathbf{X}\in A_n}\mid\mathcal{G}]\\&=\mathbb{E}[\mathbb{I}_{\mathbf{X}\in\cup A_n}\mid\mathcal{G}].\end{aligned}$$

This shows that $\mathbb{D}$ is a λ-system. Next we will show that $\mathbb{D}$ contains the sub-class $\mathcal{P}=\{I\subset\mathbb{R}\mid I=(a,b]\text{ is a half-open interval}\}$, where a is may be $-\infty$. By construction of γ,

$$\begin{aligned}\gamma((-\infty,x],\cdot)&=F(x,\cdot)\\&=\lim_{q\searrow x}\mathbb{E}[\mathbb{I}_{\mathbf{X}\leq q}\mid\mathcal{G}]\\&=\mathbb{E}[\mathbb{I}_{\mathbf{X}\leq x}\mid\mathcal{G}]\quad P\text{-a.e.}\end{aligned}$$

where the last equality follows by dominated convergence for conditional expectation. Now let $a < b$. Then we can write $\mathbb{I}_{a<\mathbf{X}\leq b} = \mathbb{I}_{\mathbf{X}\leq b} - \mathbb{I}_{\mathbf{X}\leq a}$ and we get

$$\gamma((a,b],\cdot) = \mathbb{E}[\mathbb{I}_{\mathbf{X}\leq b} \mid \mathcal{G}] - \mathbb{E}[\mathbb{I}_{\mathbf{X}\leq a} \mid \mathcal{G}] = \mathbb{E}[\mathbb{I}_{a<\mathbf{X}\leq b} \mid \mathcal{G}]P\text{-a.e..}$$

This shows that $\mathbb{P} \subset \mathbb{D}$. Clearly $\mathbb{P}$ is a stable under taking finite intersections, i.e. a π-system. Furthermore, it is well known that $\mathbb{P}$ generates $\mathbb{B}(\mathbb{R})$. Now the π-λ Theorem implies shows that $\sigma(\mathbb{P}) = \mathbb{B}(\mathbb{R}) \subset \mathbb{D} \subset \mathbb{B}(\mathbb{R})$. □

Lemma* 97. *Let $\mathbf{X}\colon \Omega \to \mathbb{X}$ be measurable from a measure space $(\Omega, \mathcal{F}, P)$ to a measurable space $(\mathbb{X}, \mathcal{X})$. Denote $P_{\mathbf{X}} = P \circ \mathbf{X}^{-1}$. For $B \in \mathcal{X}$ and $f\colon \mathbb{X} \to \mathbb{R}$ measurable and integrable we have*

$$\int_B f dP_{\mathbf{X}} = \int_{\mathbf{X}^{-1}(B)} f \circ \mathbf{X} dP.$$

Proof. The lemma follows from the standard proof. To see that it holds for indicator functions, let $A \in \mathcal{X}$ and calculate

$$\int_B \mathbb{I}_A dP_{\mathbf{X}} = P_{\mathbf{X}}(A \cap B) = P(\mathbf{X}^{-1}(A \cap B)).$$

Similarly, note that $\mathbb{I}_A \circ \mathbf{X} = \mathbb{I}_{\mathbf{X}^{-1}(A)}$. Therefore

$$\int_{\mathbf{X}^{-1}(B)} \mathbb{I}_A \circ \mathbf{X} dP = P(\mathbf{X}^{-1}(A) \cap \mathbf{X}^{-1}(B))$$

where the right hand sides are equal. □

Danish summary

Dette er et projekt, der omhandler generaliseringer af normalfordelingen og modeller for disse. To brede tilgange til modellering af normalfordelinger, direkte og indirekte specifikation, samt forudsætninger for teorien om disse er beskrevet. Tre ting er værd at nævne.

1. Der er fire måder man kan generalisere fordelingen for en multivariat normal fordeling på $\mathbb{R}^n$ med kovariansmatrix, der har fuld rang, og disse måder karakteriseres ved, at man

1. lader kovariansmatricen ikke have fuld rank, så fordelingen ikke har tæthed mht. Lebesgue målet på $\mathbb{R}^n$,

2. lader fordelingen havde tæthed mht. Lebesgue målet på formen

$$f(x) \propto \exp(-\frac{1}{2}(x-\mu)^\top Q(x-\mu))$$

   for en matrix Q der ikke er fuld rang og $\mu \in \mathbb{R}^n$,

3. lader indexmængden for den underliggende målelige afbildning have indexmængde $\mathbb{Z}^n$ i stedet for $\{1, \ldots, n\}$, så denne er en stokastisk process, eller

4. lader den underliggende målelige afbildning have indexmængde $\mathbb{Z}^n$ og have egenskaben at visse linear kombinationer af processens komponenter er stationære.

2. Tre stringente begreber kan defineres, der relaterer sig til vores informelle forståelse af betinget sandsynlighed, og disse er

1. "Betingede fordelinger", som er direkte relateret til det abstrakte begreb "betingede forventet værdi mht. en σ-algebra

2. "Regulære betingede fordelinger", som kan ses som en opstrammet version af betingede fordelinger med bedre egenskaber

3. "Deterministiske betingede fordelinger" som er en version af regulære betingede fordelinger som er nemmere at tolke

3. To måder at modellere Gaussiske vektorer er fremhævet, idet det er beskrevet, at man enten kan specificere dem direkte ud fra deres kovariansmatrix (funktion) eller middelværdivektor (funktion) (kapitel 1) eller indirekte gennem betingede sandsynligheder på formen $\mathbf{X} \mid \mathbf{X}_{-J} = x_{-J}$ (kapitel 2).

Disse ting, som er nævnt i dette resume, der skal fylde en side, bliver beskrevet i projektet. Se [8] for detaljer.

Bibliography

[1] Julian Besag, *Spatial Interaction and the Statistical Analysis of Lattice Systems*, Journal of the Royal Statistical Society. Series B (Methodological), 1 January 1974, Vol.36(2), pp.192-236 ISSN: 00359246

[2] J. Hoffmann-Jørgensen *Probability With a View Toward Statistics* Chapman & Hall Chapman and Hall/CRC July 1, 1994 Volume II ISBN: 0-412-05231-8

[3] Håvard Rue & Leonhard Held *Gaussian Markov Random Fields: Theory and Applications* Chapman & Hall Chapman and Hall/CRC 2005 ISBN: 1-58488-432-0

[4] Christian Berg & Tage Gutmann Madsen *Mål- og integralteori. Matematik 3MI.* 2001 ISBN 87-91180-15-5

[5] John M. Lee *Introduction to Topological Manifolds* 2010 Second Edition Springer ISBN 978-1-4419-7939-1

[6] Alan Gut *Probability: A Graduate Course* 2005 Springer ISBN 0-387-22833-0

[7] Erwin Kreyszig *Introductory Functional Analysis with Applications* 1989 Second Edition Wiley ISBN 978-0-471-50459-7

[8] Jakob Nebeling Hedegaard *Gaussian Random Fields: Infinite, improper and intrinsic*

[9] Olav Kallenberg *Foundations of Modern Probability* 1997 Springer ISBN 978-0-387-22704-7

[10] Ryszard Engelking *General Topology* 1989 Helderman Verlag Berlin ISBN 3-88538-006-4

[11] Gunnar Taraldsen & Bo Henry Lindqvist: *Improper Priors are not Improper* 2010 The American Statistician, Vol. 64, No. 2 , pp. 154-158

[12] Andrey Nikolaevich Kolmogorov: *Foundations of the Theory of Probability* Second English Edition Translated: Nathan Morrison Chelsea Publishing Company New York 1956

[13] Jakob Nebeling Hedegaard *Gaussian Markov Random Fields*

[14] Hans-Otto Georgii *Gibbs Measures and Phase Transitions* ISBN: 978-3-11-025029-9

[15] Donald L. Cohn *Measure Theory* ISBN 978-1-4614-6955-1

[16] A.N. Shiryaev *Probability (Second Edition)* Springer ISBN 0-387-94549-0

[17] Aurel Spătaru *Analysis and Probability* Elsevier ISBN 978-0-12-401665-1

[18] Hans R. Künsch *Intrinsic Autoregressions and Related Models on the Two-Dimensional Lattice* Biometrika, Vol. 74, No. 3 (Sep., 1987), pp. 517-524 Oxford University Press on behalf of Biometrika Trust