ulm university universität
uulm

Random Fields

Lecture Notes

Prof. Dr. Evgeny Spodarev

Ulm
2022

Contents

1 Basic notions of the theory of random functions

1.1 Random functions

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, $\Omega \neq \emptyset$, and $(\mathsf{S}, \mathcal{B})$ be a measurable space constructed upon an abstract set $\mathsf{S} \neq \emptyset$.

Definition 1.1.1
A *random element* $X : \Omega \to \mathsf{S}$ is an $\mathcal{F}|\mathcal{B}$-measurable mapping of $(\Omega, \mathcal{F})$ into $(\mathsf{S}, \mathcal{B})$, i.e.

$$X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F} \quad \text{for all } B \in \mathcal{B}.$$

We write $X \in \mathcal{F}|\mathcal{B}$.

Later on, S will often be a topological space or even a metric space. Then, $\mathcal{B}$ will be chosen to be the Borel σ-algebra $\mathcal{B}(\mathsf{S})$, i.e. the σ-algebra generated by all open subsets of S: $\mathcal{B}(\mathsf{S}) = \sigma(\mathcal{M})$, where $\mathcal{M}$ is the class of all open subsets of S. "generated by $\mathcal{M}$" means: $\mathcal{B}(\mathsf{S})$ is the smallest σ-Algebra on S containing $\mathcal{M}$:

$$\mathcal{B}(\mathsf{S}) = \bigcap_{\substack{\mathcal{A}\text{: } \sigma-\text{Algebra} \\ \text{containing } \mathcal{M}}} \mathcal{A}.$$

Example 1.1.1

1. $\mathsf{S} = \mathbb{R} \Longrightarrow X$ is a *random variable*.
2. $\mathsf{S} = \mathbb{R}^k$, $k \geq 1 \Longrightarrow X$ is a *random vector*.
3. $\mathsf{S} = \{\text{functions } f : T \to E\}$, where T and E are two abstract spaces $\Longrightarrow X$ is a random (E-valued) function with index space T. If $E = \mathbb{R}$, then the notation $\mathsf{S} = \mathbb{R}^T$ is used. The choice of $\mathcal{B}$ will be specified later, see Definition 1.1.5 and Proposition 1.1.1.
4. Let N be the set of all locally finite simple point patterns $\varphi = \{x_i\}_{i=1}^{\infty}$ in $\mathbb{R}^d$. This means
   - $\varphi(B) := |\varphi \cap B| < \infty$ for all bounded subsets $B \in \mathcal{B}(\mathbb{R}^d)$ (write $B \in \mathcal{B}_o(\mathbb{R}^d)$), where $|A|$ denotes the cardinality of A
   - $x_i \neq x_j$, $i \neq j$ for all $\varphi \in \mathsf{N}$

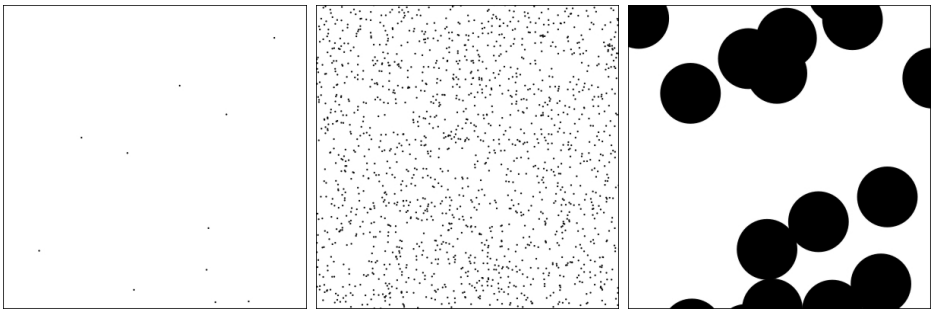
   Let $\mathcal{N}$ be the minimal σ-algebra generated by all sets of the form $\{\varphi \in \mathsf{N} : \varphi(B) = k\}$, where $k \in \mathbb{N}_0$ and B an open relatively compact[1] subset of $\mathbb{R}^d$. Then $(\mathsf{N}, \mathcal{N})$ is a measurable space. The *point process* Φ is a random element $(\Omega, \mathcal{F}) \to (\mathsf{N}, \mathcal{N})$.

   Another possibility to define Φ is to see it as a random counting measure: $\Phi(B) = \sum_{i=1}^{\infty} \delta_{x_i}(B)$, or shortly $\Phi = \mathsf{supp}\ \Phi$. The point process Φ is called *homogeneous Poisson process* on $\mathbb{R}^d$ with *intensity* λ, if $\varphi(B) \sim \mathrm{Poi}(\lambda \nu_d(B))$ for all $B \in \mathcal{B}_o(\mathbb{R}^d)$, where $\nu_d(\cdot)$ is

[1] A set $B \subset \mathbb{R}^d$ is *relatively compact* if its closure $\bar{B}$ is a compact set in $\mathbb{R}^d$.

the d-dimensional Lebesgue measure. One can show that $\Phi(B_1), \dots, \Phi(B_n)$ are independent random variables, if $B_1, \dots, B_n \in \mathcal{B}_0(\mathbb{R}^d)$ are disjoint sets. Moreover, $\lambda\nu_d(B)$ is the mean number of points of Φ in the window B.

5. Let G be the set of all closed sets in $\mathbb{R}^d$. Introduce the minimal σ-Algebra $\mathcal{G}$, which is generated by the class of sets $\{A \in \mathsf{G} : A \cap B \neq \emptyset\}$, where $B \subset \mathbb{R}^d$ is compact. The *random closed set* (RACS) is a random element $\Xi : (\Omega, \mathcal{F}) \to (\mathsf{G}, \mathcal{G})$. As an example take $\Xi = \bigcup_{i=1}^n B_r(x_i)$, $r > 0$, for any point process $\Phi = \{x_i\}_{i=1}^\infty$, which is a special case of the so called *germ-grain models*.



(a) $\lambda = 0.0001$ (b) $\lambda = 0.01$ (c) $\lambda = 0.0001, r = 40$

Fig. 1.1: A realization of a homogeneous Poisson point process with intensity $\lambda = 0.0001$ (left), $\lambda = 0.01$ (middle) and a germ-grain model of equal discs based on a homogeneous Poisson point process with intensity $\lambda = 0.0001$

If $T = \mathbb{R}$, $\mathbb{R}_+ = [0; \infty)$ or $[a; b] \subset \mathbb{R}$, then X is called a *random process with continuous time*. If $T = \mathbb{N}$ or $\mathbb{Z}$, it is called a *random process with discrete time*. For $T = \mathbb{R}^d$ ($\mathbb{Z}^d$), $d > 1$ or a (finite) subset of these spaces, we call X a *random field*. To stress that S depends on T, we will sometimes use the notation S_T.

Definition 1.1.2
The *distribution* of a random element $X : \Omega \to \mathsf{S}$ is a probability measure $\mathbf{P}_X$ defined on the measurable space $(\mathsf{S}, \mathcal{B})$ by

$$\mathbf{P}_X(B) = \mathbf{P}(X^{-1}(B)), \quad B \in \mathcal{B}.$$

Lemma 1.1.1
Any probability measure μ on $(\mathsf{S}, \mathcal{B})$ can be considered as a distribution of some random element X.

Proof Take $\Omega = \mathsf{S}$, $\mathcal{F} = \mathcal{B}$, $\mathbf{P} = \mu$ and $X(\omega) = \omega$. □

Let us give a more intuitive definition of random functions:

Definition 1.1.3
Let T be an abstract index space[2] and $(E, \mathcal{E})$ a measurable space. A family $X = \{X(t),\ t \in T\}$ of random elements $X(t) : \Omega \to E$ defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ is called a *random function*. Evidently, $X = X(\omega, t)$ is a mapping of $\Omega \times T$ onto E, which is $\mathcal{F}|\mathcal{E}$-measurable

[2]The notation T comes from "time", since for random processes $t \in T$ is often interpreted as the time parameter.

for each $t \in T$. For any fixed $\omega \in \Omega$, the function $\{X(t,\omega),\ t \in T\}$ is called a *realization* (trajectory) of X.

Let us ask the question of existence of random functions with some predefined properties. First of all, we mention the existence result for random functions $X = \{X(t), t \in T\}$, where all $X(t)$ have given distributions and are independent.

Theorem 1.1.1 (Z. Lomnicki, S. Ulam (1934)):
Let $(E_t, \mathcal{E}_t)_{t\in T}$ be a sequence of arbitrary measurable spaces equipped with a probability measure μ_t for every $t \in T$ in an index space T. Then there exists a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and a random function $X = \{X(t) : \Omega \to E_t,\ t \in T\}$ such that

1. $X(t)$ is $\mathcal{F}|\mathcal{E}_t$-measurable for all $t \in T$.
2. the random elements $X(t)$ are independent for all $t \in T$.
3. $\mathbf{P}_{X(t)} = \mu_t$ on $\mathcal{E}_t$ for all $t \in T$.

Many meaningful classes of random functions are constructed on the basis of independent ones, see examples in Section 1.2.

Definition 1.1.4
For $n \in \mathbb{N}$ and $t_1, \dots, t_n \in T$ we call the distribution of the random vector $(X(t_1), \dots, X(t_n))^\top$ a *finite-dimensional distribution* of random function $X = \{X(t),\ t \in T\}$. For $n = 1$, the distribution of $X(t)$ is called *marginal*. We write

$$\mathbf{P}_{t_1,\dots,t_n}(B_{t_1}, \dots, B_{t_n}) = \mathbf{P}(X(t_1) \in B_{t_1}, \dots, X(t_n) \in B_{t_n}),$$

where $B_{t_k} \in \mathcal{B}_{t_k}$, $k = 1, \dots, n$.

Let us analyze the question of equivalence of Definition 1.1.1, Example 1.1.1 3) and Definition 1.1.3. For that, one has to solve the problem of measurability of a mapping $\omega \longmapsto X(\omega, \cdot)$, $\omega \in \Omega$. Let $\mathsf{S} = \{\text{functions } f : T \to E\} := E^T$. Let us consider a more general space S by admitting $\mathsf{S} = \{\text{functions } f \text{ on } T : f(t) \in E_t\}$ for a family of measurable spaces $(E_t, \mathcal{E}_t)_{t\in T}$.

Introduce the class $\mathcal{M}$ of elementary cylinder sets $C(t, B_t)$ of S, i.e. $\mathcal{M}$ contains sets of the form $\{f \in \mathsf{S} : f(t) \in B_t\}$, where $t \in T$ and $B_t \in \mathcal{E}_t$. These sets contain all functions f with trajectories $f(\cdot)$ that go through the "gate" B_t.

Definition 1.1.5
The σ-algebra $\mathcal{B}_T$ on S generated by $\mathcal{M}$ is called a *cylindric σ-algebra* $\mathcal{B}_T = \sigma(\mathcal{M})$. Notation: $\mathcal{B}_T = \bigotimes_{t\in T} \mathcal{E}_t$. For $E_t = E$ for all $t \in T$ one can use $\mathcal{B}_T = \mathcal{E}^T$.

Let us prove the following result:

Proposition 1.1.1
A family $X = \{X(\omega, t) : t \in T\}$, i.e. the family of $\mathcal{F}|\mathcal{E}_t$-measurable random elements is a random function if and only if the mapping $\omega \longmapsto X(\omega, \cdot)$, $\omega \in \Omega$ is $\mathcal{F}|\mathcal{B}_T$-measurable.

Proof "$\Rightarrow$" If $X^{-1}(B_t) \in \mathcal{F}$ for all $B_t \in \mathcal{E}_t$, $t \in T$, then $\{\omega \in \Omega : X(\omega, \cdot) \in C(t, B_t)\} \in \mathcal{F}$, thus $X^{-1}(\mathcal{M}) \subseteq \mathcal{F}$ and X is $\mathcal{F}|\mathcal{B}_T$-measurable, because $\mathcal{B}_T$ is generated by $\mathcal{M}$.
"$\Leftarrow$" Let X be $\mathcal{F}|\mathcal{B}_T$-measurable. Introduce the coordinate projection $\pi_{T,t} : \mathsf{S}_T \to E_t$ by $\pi_{T,t} f = f(t)$ for $f \in \mathsf{S}_T$, $t \in T$. $\pi_{T,t}$ is $\mathcal{B}_T|\mathcal{E}_t$-measurable, since for all $B_t \in \mathcal{E}_t$ we have

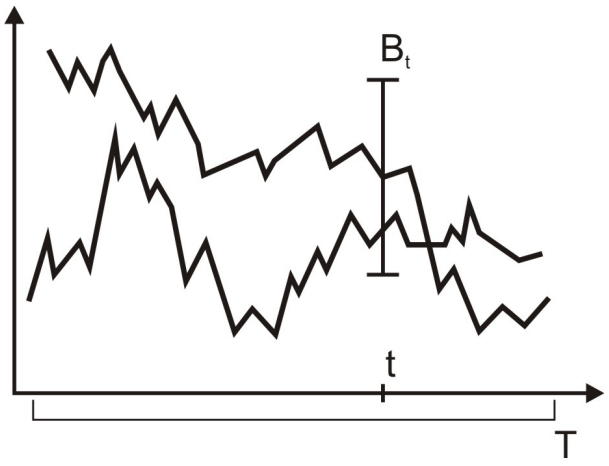


Fig. 1.2: Trajectories going through a "gate" B_t

$\pi_{T,t}^{-1}(B_t) = C(t, B_t) \in \mathcal{B}_T$. It is clear that $X(\omega, t) = \pi_{T,t} X(\omega, \cdot)$. Since $\pi_{T,t}$ is $\mathcal{B}_T|\mathcal{E}_t$-measurable and X is $\mathcal{F}|\mathcal{B}_T$-measurable, then $X(\omega, t)$ is $\mathcal{F}|\mathcal{E}_t$-measurable for all t as a composition of two measurable mappings. □

Now let us consider the question whether a random function X is defined by all its finite-dimensional distributions. To be more concrete, for all $t_1, \ldots, t_n \in T$ introduce the space $E_{t_1,\ldots,t_n} := E_{t_1} \times \ldots \times E_{t_n}$ as a phase space of the random vector $X_{t_1,\ldots,t_n} := (X(t_1), \ldots, X(t_n))^\top$. Let $\mathcal{E}_{t_1,\ldots,t_n} := \mathcal{E}_{t_1} \otimes \ldots \otimes \mathcal{E}_{t_n}$ be the σ-Algebra on $E_{t_1,\ldots,t_n}$ generated by all "parallelepipeds" $B_{t_1} \times \ldots \times B_{t_n}$, $B_{t_k} \in \mathcal{E}_{t_k}$ for all $k = 1, \ldots, n$.
Then the vector $X_{t_1,\ldots,t_n}$ is $\mathcal{F}|\mathcal{E}_{t_1,\ldots,t_n}$-measurable, since

$$X_{t_1,\ldots,t_n}^{-1}(B_{t_1} \times \ldots \times B_{t_n}) = \bigcap_{k=1}^{n} \{X(t_k) \in B_{t_k}\} \in \mathcal{F}.$$

Then, the finite-dimensional distribution $\mathbf{P}_{t_1,\ldots,t_n}$ of X at "time" points $t_1, \ldots, t_n$ is a probability measure on $(E_{t_1,\ldots,t_n}, \mathcal{E}_{t_1,\ldots,t_n})$. For parallelepipeds $B = B_{t_1} \times \ldots \times B_{t_n}$ it holds

$$\mathbf{P}_{t_1,\ldots,t_n}(B) = \mathbf{P}(X(t_1) \in B_{t_1}, \ldots, X(t_n) \in B_{t_n}). \tag{1.1.1}$$

The finite-dimensional distributions of a random function X have the following important properties:
For all $n \geq 2$, $t_1, \ldots, t_n \in T$, $B_{t_k} \in E_{t_k}$, $k = 1, \ldots, n$ and an arbitrary permutation $(i_1, \ldots, i_n)$ of $(1, \ldots, n)$ the following two conditions are satisfied:

1. **Symmetry:**
$$\mathbf{P}_{t_1,\ldots,t_n}(B_{t_1} \times \ldots \times B_{t_n}) = \mathbf{P}_{t_{i_1},\ldots,t_{i_n}}(B_{t_{i_1}}, \ldots, B_{t_{i_n}})$$

2. **Consistency:**
$$\mathbf{P}_{t_1,\ldots,t_n}(B_{t_1} \times \ldots \times B_{t_{n-1}} \times E_{t_n}) = \mathbf{P}_{t_1,\ldots,t_{n-1}}(B_{t_1}, \ldots, B_{t_{n-1}})$$

They are evident considering (1.1.1) and $\{X(t_n) \in E_{t_n}\} = \Omega$. The consistency condition can be equivalently written as

$$\mathbf{P}_{t_1,\ldots,t_m,\ldots,t_n}(B_{t_1} \times \ldots \times E_{t_m} \times \ldots \times B_{t_n}) = \mathbf{P}_{t_1,\ldots,t_{m-1},t_{m+1},\ldots,t_n}(B_{t_1} \times \ldots B_{t_{m-1}} \times B_{t_{m+1}} \times \ldots \times B_{t_n})$$

for any $m \in \{1,\ldots,n\}$. In the above conditions, only n-tuples of $t_1,\ldots,t_n$ are meaningful with $t_i \neq t_j$, $i \neq j$. Otherwise if say $t_1 = t_2 = t$, one could write $\mathbf{P}_{t_1,t_2}(B_1 \times B_2) = \mathbf{P}(X(t) \in B_1, X(t) \in B_2) = \mathbf{P}(X(t) \in B_1 \cap B_2) = \mathbf{P}_t(B_1 \cap B_2)$ for all $B_1, B_2 \in \mathcal{E}_t$.

To formulate the basic theorem of Kolmogorov on the existence of a random function with given finite-dimensional distributions, we need the following

Definition 1.1.6
Two measurable spaces $(\mathsf{C},\mathcal{C})$ and $(\mathsf{D},\mathcal{D})$ are called *isomorphic* ($(\mathsf{C},\mathcal{C}) \sim (\mathsf{D},\mathcal{D})$) if there exists a one-to-one mapping $h : \mathsf{C} \to \mathsf{D}$ such that $h \in \mathcal{C}|\mathcal{D}$ and $h^{-1} \in \mathcal{D}|\mathcal{C}$. If D is a Borel subset of $[0;1]$ and $\mathcal{D}$ is the σ-Algebra of Borel subsets of D (i.e. $\mathcal{D} = \mathsf{D} \cap \mathcal{B}([0;1])$) we call a measurable space $(\mathsf{C},\mathcal{C})$, which is isomorphic to $(\mathsf{D},\mathcal{D})$, a *Borel space*.

For instance, if C is a Borel subset of a Polish space[3] $\mathcal{P}$ with σ-Algebra $\mathcal{C} = \mathsf{C} \cap \mathcal{B}(\mathcal{P})$, then it can be shown that $(\mathsf{C},\mathcal{C})$ is a Borel space. A natural choice of $\mathcal{P}$ can be e.g. $\mathbb{R}^k$, $k \geq 1$.

Theorem 1.1.2 (A. N. Kolmogorov, 1933):
Let $(E_t,\mathcal{E}_t)_{t\in T}$ be a family of Borel spaces. For any $n \in \mathbb{N}$ and $t_1,\ldots,t_n \in T$, $t_i \neq t_j$, $i \neq j$ let measures $\mathbf{P}_{t_1,\ldots,t_n}$ be given on spaces $(E_{t_1,\ldots,t_n},\mathcal{E}_{t_1,\ldots,t_n})$ such that they satisfy the conditions of symmetry and consistency. Then there exists a probability space $(\Omega,\mathcal{F},\mathbf{P})$ and a random function $X = \{X(t) : t \in T\}$ defined on it such that its finite-dimensional distributions coincide with the measures $\mathbf{P}_{t_1,\ldots,t_n}$.

Remark 1.1.1
In Definition 1.1.6 (and consequently the formulation of Kolmogorov's Theorem) it is required that E_t is isomorphic to a Borel subset of $[0;1]$ and not to the whole $[0;1]$. It was done to admit finite and countable spaces E_t as the following result shows:

Theorem 1.1.3
The Borel space $(\mathsf{D},\mathcal{D})$ is isomorphic to

$$(\mathsf{D},\mathcal{D}) \sim \begin{cases} (\mathsf{K},\mathcal{B}(\mathsf{K})) & \text{if } \mathsf{D} \text{ is uncountable,} \\ (\mathbb{N},2^{\mathbb{N}}) & \text{if } \mathsf{D} \text{ is countable,} \\ ((1,\ldots,|\mathsf{D}|),\mathcal{A}(1,\ldots,|\mathsf{D}|)) & \text{if } \mathsf{D} \text{ is finite,} \end{cases}$$

where $|\mathsf{D}|$ is the number of elements of D, $\mathcal{A}(1,\ldots,|\mathsf{D}|)$ is the σ-algebra of all subsets of $\{1,\ldots,|\mathsf{D}|\}$, and 2^{D} is the σ-algebra of all subsets of D. The space $(\mathsf{K},\mathcal{B}(\mathsf{K}))$ is defined as follows: take the Polish space $\{0,1\}$ with metric $d(x,y) = \mathbf{1}\{x \neq y\}, x, y \in \{0,1\}$, set $\mathsf{K} = \{0,1\}^{\mathbb{N}}$ and introduce the metric d_{K} on that space by

$$d_{\mathsf{K}}(x,y) = \sum_{n=1}^{\infty} 2^{-n} \frac{d(x_n,y_n)}{1+d(x_n,y_n)} \quad \text{for } x = \{x_n\}_{n=1}^{\infty}, y = \{y_n\}_{n=1}^{\infty} \in \mathsf{K}. \tag{1.1.2}$$

It is not difficult to see that K is the space of all binary representations of real numbers from $[0,1]$. Then K is a measurable space by taking its σ-algebra $\mathcal{B}(\mathsf{K})$ of Borel sets in this metric.

[3] A Polish space is a separable complete metric space.

The proof of Kolmogorov's theorem will be given in Section 1.6. Now let us formulate an equivalent statement in terms of characteristic functions of random measures for the special case $E_t = \mathbb{R}$, $\mathcal{E}_t = \mathcal{B}(\mathbb{R})$, for all $t \in T$.

Definition 1.1.7
Let μ be a finite measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Its characteristic function $\varphi_\mu(s)$ is defined by

$$\varphi_\mu(s) = \int_{\mathbb{R}^d} \exp(i \langle s, t \rangle) \mu(dt), \quad s \in \mathbb{R}^d,$$

where $\langle s, t \rangle = \sum_{k=1}^d s_k t_k$ is the scalar product of vectors s and t in $\mathbb{R}^d$.

If μ is a probability measure, it obviously coincides with the characteristic function of a d-dimensional random vector Y with distribution μ. It is known that $\varphi_\mu(\cdot)$ determines the measure μ uniquely.

Proposition 1.1.2
The family of measures $\mathbf{P}_{t_1,\dots,t_d}$ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $(t_1, \dots, t_d) \in T^d$, $d \geq 1$, satisfies the conditions of symmetry and consistency if and only if for all $d \geq 2$, $s = (s_1, \dots, s_d) \in \mathbb{R}^d$ and $t = (t_1, \dots, t_d) \in T^d$ it holds

1.
$$\varphi_{\mathbf{P}_{t_1,\dots,t_d}}((s_1, \dots, s_d)) = \varphi_{\mathbf{P}_{t_{i_1},\dots,t_{i_d}}}((s_{i_1}, \dots, s_{i_d}))$$
for any permutation $(1, \dots, d) \longmapsto (i_1, \dots, i_d)$

2.
$$\varphi_{\mathbf{P}_{t_1,\dots,t_{d-1}}}((s_1, \dots, s_{d-1})) = \varphi_{\mathbf{P}_{t_1,\dots,t_d}}((s_1, \dots, s_{d-1}, 0))$$

Exercise 1.1.1
Prove Proposition 1.1.2.

As all stochastic objects are defined up to a set of zero probability it can happen that for a random function $X(t)$ these "exception sets" are quite different. In such case, it would be useless to speak of different realizations of $X(t)$. That is why a convention is accepted that all $X(t)$, $t \in T$ are defined simultaneously on one single set $\Omega_0 \subseteq \Omega$, $\Omega_0 \in \mathcal{F}$ of probability one. In this case, the random function $\tilde{X} : \Omega_0 \times T \to \mathbb{R}$ is called a *modification* of X, where $X : \Omega \times T \to \mathbb{R}$. The difference between $X(t)$ and $\tilde{X}(t)$ is that they may take different values on a set of zero probability. But such a difference can make the realizations of X possess some "nice" properties such as continuity or differentiability. Later on, if we say that "random function X has property A" we mean that a modification of X is considered which has this property A.

Definition 1.1.8
Random functions $X = \{X(t),\ t \in T\}$ and $Y = \{Y(t),\ t \in T\}$ defined on the same probability space $(\Omega, \mathcal{F}, \mathbf{P})$ with values in $(E_t, \mathcal{E}_t)$, $t \in T$ have *equivalent trajectories*

$$\{\omega \in \Omega : X(\omega, t) \neq Y(\omega, t) \text{ for some } t \in T\} \in \mathcal{F} \quad \text{and}$$

$$\mathbf{P}\left(\{\omega \in \Omega : X(\omega, t) \neq Y(\omega, t) \text{ for some } t \in T\}\right) = 0.$$

It means that the realizations of X and Y coincide with probability one. It is clear that for all $s \in T$ we have $\{\omega \in \Omega : X(\omega, s) \neq Y(\omega, s)\} \subset \{\omega \in \Omega : X(\omega, s) \neq Y(\omega, s) \text{ for some } s \in T\}$.

Definition 1.1.9
Random functions $X = \{X(t),\ t \in T\}$ and $Y = \{Y(t),\ t \in T\}$, defined on a common probability space $(\Omega, \mathcal{F}.\mathbb{P})$ with values in $(E_t, \mathcal{E}_t)$ are *(stochastically) equivalent*, if

$$\{\omega \in \Omega : X(\omega, t) \neq Y(\omega, t)\} \in \mathcal{F}, \quad t \in T \text{ and}$$
$$\mathbf{P}\left(\{\omega \in \Omega : X(\omega, t) \neq Y(\omega, t)\}\right) = 0, \quad t \in T.$$

We also call Y a version of X.

On a complete[4] probability space $(\Omega, \mathcal{F}, \mathbf{P})$ processes with equivalent trajectories are stochastically equivalent.

Example 1.1.2
Let $X(\omega, t) \equiv 0$, $t \in [0; 1]$ be a random function defined on $([0; 1], \mathcal{B}([0; 1]), \lambda_1)$, where λ_1 is the Lebesgue measure on $[0; 1]$. Define $Y(\omega, t) = I_{\{t\}}(\omega)$ on the same probability space. Then, X and Y are stochastically equivalent, i.e. Y is a modification of X. Note that all trajectories of X are continuous, but all trajectories of Y are discontinuous.

Definition 1.1.10
Random functions $X = \{X(t),\ t \in T\}$ and $Y = \{Y(t),\ t \in T\}$ (not necessarily defined on the same probability space) with values in $(E_t, \mathcal{E}_t)$, $t \in T$ are *equivalent in distribution* if their probability laws are identical: $\mathbf{P}_X = \mathbf{P}_Y$ on $(\mathsf{S}_T, \mathcal{B}_T)$.

By Kolmogorov's theorem, such processes must have the same finite-dimensional distributions. It is clear that equivalence in the stochastic sense implies equivalence in distribution, but not vice versa. Equivalence in distribution is often used in coupling arguments which play an important role in the theory of Markov processes.

Let us return to the question of measurability of a random function X. Let $(T, \mathcal{A})$ be a measurable space as well. It is known from the definition of a random function that for all $t \in T$ $X(t)$ is $\mathcal{F}|\mathcal{E}_t$-measurable. But nobody guarantees the measurability of $X(\omega, t)$ as a function of (ω, t).

Definition 1.1.11
Let $X = \{X(t),\ t \in T\}$ be a random function defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ with values in $(E, \mathcal{E})$ for all $t \in T$. X is said to be *measurable* if the mapping

$$X : (\omega, t) \longmapsto X(\omega, t) \in E, \quad (\omega, t) \in \Omega \times T$$

is $\mathcal{F} \otimes \mathcal{A}|\mathcal{E}$-measurable as a mapping of $\Omega \times T$ onto E.

1.2 Elementary examples

§1. White noise

Definition 1.2.1
A random function $X = \{X(t), t \in T\}$ defined on $(\Omega, \mathcal{F}, \mathbf{P})$ is called *white noise*, if $X(t), t \in T$ are independent and identically distributed.

[4] The probability space $(\Omega, \mathcal{F}, \mathbf{P})$ is complete if for all $A \in \mathcal{F} : \mathbf{P}(A) = 0$ it follows that for all $B \subset A$ we have $B \in \mathcal{F}$ (and hence $\mathbf{P}(B) = 0$ as well).

As a special case, one has $T = \mathbb{Z}^d$ or $\mathbb{R}^d$, $d \geq 1$. White noise is used to model noise in images, such as salt-and-pepper noise $X(t) \sim \mathrm{Ber}(p), t \in T$, for binary images or Gaussian white noise $X(t) \sim \mathcal{N}(0, \sigma^2), \sigma^2 > 0$, for greyscale images.
Because of mutual independence of $X(t)$ for every $t \in T$ all finite-dimensional distributions are product distributions that obviously satisfy the conditions of Theorem 1.1.2. Hence, the white noise exists, compare also Theorem 1.1.1.

§2. Gaussian random functions

Another famous simple example of a random function given by all its finite-dimensional distributions explicitly is that of a Gaussian random function.

Definition 1.2.2
The random function $X = \{X(t), t \in T\}$ is called *Gaussian* if all its finite-dimensional distributions $\mathbf{P}_{t_1,\dots,t_n}$ are Gaussian, i.e. for all $n \geq 1$ and all $t_1, \dots, t_n \in T$ the distribution of the random vector $X_{t_1,\dots,t_n} = (X(t_1), \dots, X(t_n))^\top$ is an n-dimensional normal distribution with expectation $\mu_{t_1,\dots,t_n} = (\mu(t_1), \dots, \mu(t_n))^\top$ and a covariance matrix $\Sigma_{t_1,\dots,t_n} = (\mathbf{cov}(X(t_i), X(t_j)))_{i,j=1}^n$: $X_{t_1,\dots,t_n} \sim \mathcal{N}(\mu_{t_1,\dots,t_n}, \Sigma_{t_1,\dots,t_n})$. This means that for any vector $c \in \mathbb{R}^n$ the random variable $\langle c, X_{t_1,\dots,t_n} \rangle$ is normally distributed. If $\Sigma_{t_1,\dots,t_n}$ is non-degenerate, then $\mathbf{P}_{t_1,\dots,t_n}$ has the density

$$f_{t_1,\dots,t_n}(y) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \Sigma_{t_1,\dots,t_n}}} \exp\left\{-\frac{1}{2}\left(y - \mu_{t_1,\dots,t_n}\right)^\top \Sigma^{-1}_{t_1,\dots,t_n}\left(y - \mu_{t_1,\dots,t_n}\right)\right\}, \quad y \in \mathbb{R}^n.$$

Exercise 1.2.1
Show that any Gaussian function X is uniquely defined by its expectation (or mean value) function $\mu(t) = \mathbf{E}X(t)$, $t \in T$ and its (non-centered) covariance function $C(s,t) = \mathbf{E}[X(s)X(t)]$, $s, t \in T$.

Gaussian random functions are widely used in applications starting with modelling the microstructure of surfaces in material science (e.g. metal surfaces or paper, see Figure 1.4) up to models of fluctuations of microwave cosmic background radiation (see Figure 1.3).

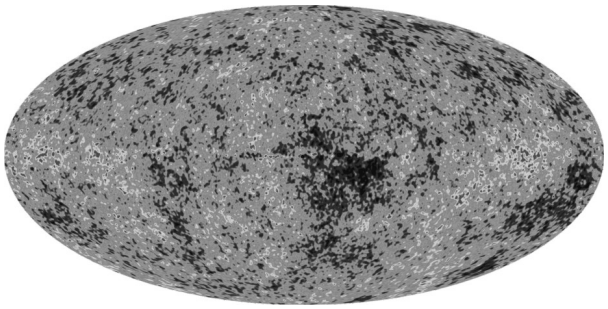


Fig. 1.3: WMAP (Wilkinson Microwave Anisotropy Probe) cosmic microwave fluctuations over the full sky with five years of data.

For the construction of many other random functions, Theorem 1.1.2 cannot be used directly as their finite-dimensional distribution cannot be written down explicitly in a convenient way. In such cases, explicit constructions of the form $X(t) = g(t, Y_1, \dots, Y_n)$, $t \in T$ are used, where g is measurable and $Y_1, \dots, Y_n$, $n \in \mathbb{N}$ are some random elements whose existence has been already proved. Let us give corresponding examples:

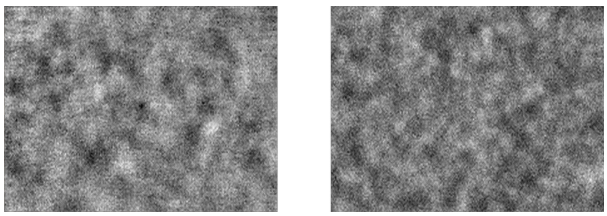


Fig. 1.4: Paper surface (left) and a simulated Gaussian random field based on the estimated data $\mathbf{E}X(t) = 126$, $\mathbf{cov}(X(0), X(t)) = 491 \exp\left(-\frac{|t|}{56}\right)$ (right)

§3. Lognormal fields

A random field $X = \{X(t), t \in T\}$ is called *lognormal* if $X(t) = e^{Y(t)}$, where $Y = \{Y(t), t \in T\}$ is a Gaussian random field.

§4. χ^2-fields

A random function $X = \{X(t), t \in \mathbb{R}^d\}$ is called χ^2-*field* if $X(t) = |Y(t)|^2$, $t \in \mathbb{R}^d$, where $Y = \{Y(t), t \in \mathbb{R}^d\}$ is an n-dimensional vector-valued random field such that $Y(t) \sim \mathcal{N}(0, I)$, $t \in \mathbb{R}^d$. I denotes the identity matrix. It is clear that $X(t)$ is χ^2_n-distributed for all $t \in \mathbb{R}^d$.

§5. Cosine fields

Let Y_1 be a random variable, Y_2 a random vector on $(\Omega, \mathcal{F}, \mathbf{P})$, $\dim Y_2 = \dim T$, where Y_1 and Y_2 are independent. Consider a random field $X = \{X(t), t \in \mathbb{R}^d\}$, $d \geq 1$ defined by $X(t) = \sqrt{2}\cos(2\pi Y_1 + \langle t, Y_2 \rangle)$. For instance, we can put Y_1 to be uniformly distributed on $[0; 1]$ (notation: $Y_1 \sim U[0; 1]$). Each realization of X is a cosine wave surface. If $X_1, \ldots, X_n$ are independent cosine waves then by the central limit theorem we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i(t) \xrightarrow{d} Z(t)$$

as $n \to \infty$, where $\xrightarrow{d}$ denotes convergence in distribution and Z is a Gaussian random field with covariance function $\mathbf{cov}$. In this case, the distribution of Y_2 is chosen to be the *spectral measure* μ of $\mathbf{cov}$, confer Definition 2.1.2. This is the so called *spectral method* for the simulation of Z. For more details on it, see e.g. [Lantuejoul].

§6. Empirical random measures

Let $\{Y_i\}_{i \in \mathbb{N}}$ be a sequence of independent identically distributed (iid) d-dimensional random vectors defined on $(\Omega, \mathcal{F}, \mathbf{P})$. Fix some $n \in \mathbb{N}$.

Definition 1.2.3
For any Borel set $B \in \mathcal{B}(\mathbb{R}^d)$ define the *empirical random measure* $\mu_n(B, \omega)$ as

$$\mu_n(B, \omega) := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{Y_i(\omega) \in B\}, \tag{1.2.1}$$

or shortly, $\mu_n(B) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}\{Y_i \in B\}$. This is obviously a random function indexed by Borel sets B. It is used in statistics to estimate the (unknown) distribution $\mathbf{P}_Y(B) = \mathbf{P}(Y_i \in B)$, $B \in \mathcal{B}(\mathbb{R}^d)$ of Y_i. For instance, in the special case of $d = 1$ and $B = (-\infty; x]$, $x \in \mathbb{R}$ one gets the well known *empirical distribution function* $\hat{F}_n(x) = \mu_n((-\infty; x])$. It is clear that by the strong law of large numbers it holds $\mu_n(B) \to \mathbf{P}_Y(B)$ in the almost-sure sense as $n \to \infty$, $B \in \mathcal{B}(\mathbb{R}^d)$. That means $\mu_n(B)$ is a reasonable approximation of $\mathbf{P}_Y(\cdot)$ for sufficiently large n.

§7. Partial sums of a random field Y

Let $Y = \{Y(t), t \in \mathbb{Z}^d\}$ be a random field on the lattice $\mathbb{Z}^d$, $d \geq 1$. Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence of σ-finite measures on $\mathcal{B}(\mathbb{R}^d)$.

Definition 1.2.4
The random function $S = \left\{S_n(B), B \in \mathcal{B}((0;1]^d), n \in \mathbb{N}\right\}$ defined by

$$S_n(B) = \sum_{t\in\mathbb{Z}^d} Y(t)\mu_n((t-1;t] \cap nB)$$

for all $B \in \mathcal{B}((0;1]^d)$ and $n \in \mathbb{N}$ is called the *process of partial (weighted) sums* of Y. Here $(t-1;t] = \bigotimes_{j=1}^d (t_j - 1; t_j)$, $t = (t_1, \ldots, t_d)^\top$ is a unit cube with left lower vertex $t - 1$ and $nB = \{nb : \ b \in B\}$.

The measures μ_n give particular weights to any cell $(t-1;t]$. Thus, for $B = (0;1]^d$ and

$$\mu_n((t;t+1] \cap (0;n]^d) = \begin{cases} \frac{1}{n^{d/2}} & \text{if } (t;t+1] \cap [0;n]^d \neq \emptyset, \\ 0 & \text{otherwise,} \end{cases}$$

we would get

$$S_n([0;1]^d) = \frac{1}{n^{d/2}} \sum_{t\in\mathbb{Z}^d\cap[0;n]^d} Y(t),$$

which corresponds to partial sums used in the central limit theorem. Under some additional assumptions we have $S_n([0;1]^d) \xrightarrow{d} \mathcal{N}(\cdot,\cdot)$.

§8. Shot noise random fields and moving averages

Let $\Phi = \{x_i, i \in \mathbb{N}\}$ be a homogeneous Poisson point process with intensity $\lambda > 0$ (see Example 1.1.1, 4.). Let $g : \mathbb{R}^d \to \mathbb{R}$ be a deterministic function, for which $\int_{\mathbb{R}^d} g(x)dx < \infty$ and $\int_{\mathbb{R}^d} g^2(x)dx < \infty$ hold. Introduce a random field $X = \{X(t), t \in \mathbb{R}^d\}$ by $X(t) = \sum_{x\in\Phi} g(t-x)$, $t \in \mathbb{R}^d$. Due to local finiteness of the point process and the integrability conditions of g it can be shown that X is a well-defined random field with finite first two moments.

Exercise 1.2.2
Please, show this!

Definition 1.2.5
The random field X introduced above is called a *shot-noise field*. The function g is called *response function*.

Which response functions $g : \mathbb{R}^d \to \mathbb{R}$ are used in applications to model real phenomena by Poisson shot-noise random fields? A large class of response functions can be constructed as

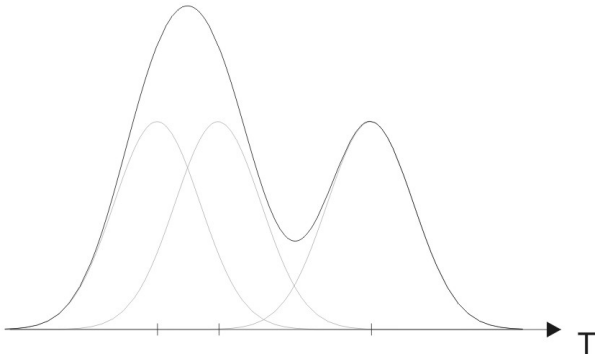


Fig. 1.5: Construction of a shot-noise process in $\mathbb{R}$

follows: take $g(x) = K(\frac{|x|}{a})$, where $|x|$ is the Euclidean norm of $x \in \mathbb{R}^d$ and $K : \mathbb{R} \to \mathbb{R}$, the so-called *kernel*, is a probability density function with compact support $\mathsf{supp}K = \{x \in \mathbb{R} : K(x) > 0\}$. For instance, K can be chosen to be the *Epanechnikov kernel* $K(x) = \frac{3}{4}(1-x^2)\mathbf{1}\{x \in [-1;1]\}$ or the *bisquare kernel* $K(x) = \frac{15}{16}(1-x^2)^2\mathbf{1}\{x \in [-1;1]\}$. Alternatively, kernels with unbounded support such as the *Gaussian kernel* $K(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ can be used as well.

Strictly speaking, we have thus defined a special case of shot-noise fields. This definition can be generalized if we use random response functions (see [40], p. 31) or allow for other (non-) homogeneous point processes for Φ. Formally, a shot-noise field can be written as a stochastic integral $X(t) = \int_{\mathbb{R}^d} g(t-x)\Phi(dx)$ if $\Phi(\cdot)$ is interpreted as a random Poisson counting measure (see [52], ch. 7). This means that it is a special case of a very general class of moving averages: $X(t) = \int_E g(t,x)\mu(dx)$, where $\mu(\cdot)$ is a random independently scattered measure on some measurable space $(E, \mathcal{E})$, i.e.

$$B_1, \ldots, B_n \in \mathcal{E},\ B_i \cap B_j = \emptyset,\ i \neq j \Longrightarrow \mu(B_1), \ldots, \mu(B_n) \text{ are independent,}$$

and $g : \mathbb{R}^d \times E \to \mathbb{R}$ is a deterministic function.

Example 1.2.1 1. μ is the Poisson counting measure Φ.

2. μ is the *Gaussian white noise measure*: $\mu(B) \sim \mathcal{N}(0, \lambda_d(B))$, if $(E, \mathcal{E}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, where $\lambda_d(B)$ denotes the volume of B.

Of course, one has to make it more precise, in which sense this stochastic integral is understood. Depending on the context, it can be defined in L^p-sense ($p \in (0;2]$), in probability or almost surely. More details will be given in Chapter 3.

Example 1.2.2 1. μ - Gaussian: Existence of the integral in the L^2-sense:

$$\mathbf{E}\left|\sum_{i=1}^{n} g(t, \xi_i^n)\mu(B_i^n) - \int_E g(t,x)\mu(dx)\right|^2 \xrightarrow[n\to\infty]{} 0.$$

2. μ - stable: Existence of the integral in probability:

$$\mathbf{P}\left(\left|\sum_{i=1}^{n} g(t, \xi_i^n)\mu(B_i^n) - \int_E g(t,x)\mu(dx)\right| > \varepsilon\right) \xrightarrow[n\to\infty]{} 0, \varepsilon > 0.$$

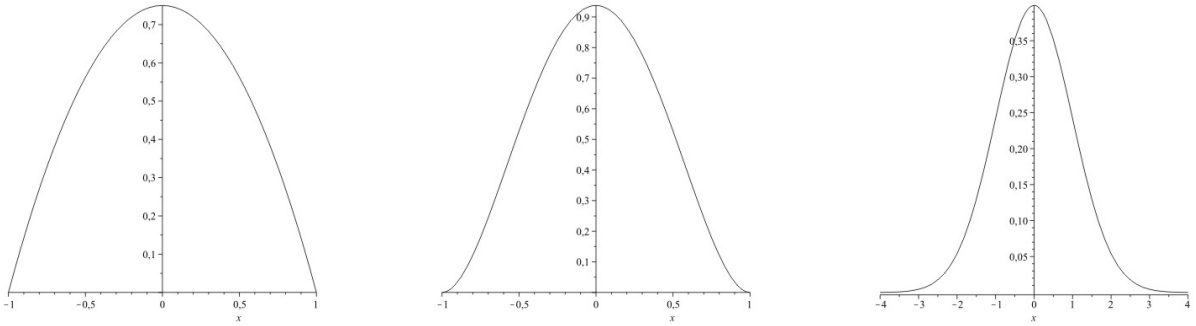


Fig. 1.6: The Epanechnikov kernel (left), Bisquare kernel (middle) and Gaussian kernel (right)

§9. Finite Markov random fields

Let T be the vertex set of a finite non-oriented graph $(T, \mathcal{T})$, where $\mathcal{T}$ is the edge set of the graph constructed by some neighborhood relation $\sim$:

$$\forall\, s,t \in T : (s,t) \in \mathcal{T} \text{ iff } s \sim t.$$

Then, the neighborhood ∂s of $s \in T$ is given by $\partial s = \{t \in T : t \neq s, s \sim t\}$. Let $E = \{e_1, \ldots, e_m\}$ be a finite space, $\mathcal{E} = 2^E$.

Definition 1.2.6
The random function $X = \{X(t), t \in T\}$ such that $X(t) : \Omega \to E$ for all $t \in T$ and $T, (E, \mathcal{E})$ introduced as above is called a (finite) *Markov random field* with respect to the neighborhood relation $\sim$ if for all $x = (x_t)_{t\in T} \in S = E^T$:

$$\mathbf{P}(X(s) = x_s | X(t) = x_t, t \neq s) = \mathbf{P}(X(s) = x_s | X(t) = x_t, t \in \partial s). \tag{1.2.2}$$

Relation (1.2.2) is called *Markov property*.

Obviously, any finite random field X on T, E is a Markov random field with respect to some neighborhood relation. In applications however, Markov random fields X with respect to small neighborhoods are sought.

The distribution of a finite random field $X = \{X(t), t \in T\}$ is given by $\mathbf{P}(X = x)$, $x \in S$. As a distribution of a finite vector of dimension $|T|$ it always exists. To give examples of T and E let $T = T_N = \mathbb{Z}^d \cap [-N; N]^d$. In image analysis we have $d = 2$ (3) and $t \in T_N$ are pixels (voxels). We distinguish between binary images ($E = \{0, 1\}$, $0\hat{=}$ white pixel, $1\hat{=}$ black pixel) and greyscale images ($E = \{0, \ldots, 255\}$, $0\hat{=}$ white pixel, ..., $255\hat{=}$ black pixel). In physics, if $d = 3$, a finite random field with index space T_N and $E = \{-1, 1\}$ describes the distribution of dipole directions (spins) at lattice points of a crystal lattice.

How can the neighborhood relation $\sim$ be defined on T_N? Two canonical ways are the so-called *+- and $*$-neigborhoods* (write $\sim_+$, $\sim_*$):

- $s \sim_+ t$ iff $\|s - t\|_1 = 1$ ($2d$ neighbors),
- $s \sim_* t$ iff $\|s - t\|_\infty = 1$ ($3^d - 1$ neighbors),

where $\|\cdot\|_1$ is the L^1-norm and $\|\cdot\|_\infty$ is the maximum norm. Accordingly, for $d = 2$ a lattice point has 4 (resp. 8) neighbors.

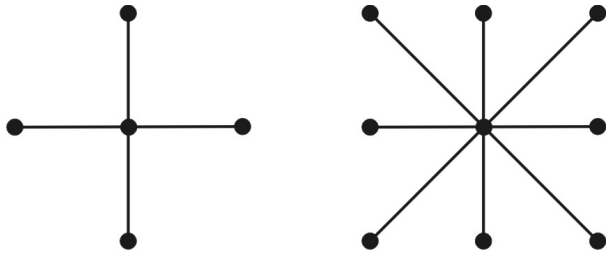


Fig. 1.7: 4-neighborhood and 8-neighborhood in $\mathbb{Z}^2$

More generally, if T is any finite vertex set with a given metric ρ, then the nearest-neighbor relation $\sim$ on T is defined as $s \sim t$, $s, t \in T$ iff $\rho(s,t)$ is the smallest distance within the class of all pairs $(s,u), u \in T, u \neq s$ or $(v,t), v \in T, v \neq t$. Then $\sim$ generates the nearest-neighbor graph out of the vertex set T.

A prominent example of finite Markov random fields yields the *Ising model* proposed by the German physicists W. Lenz and E. Ising in the 1920s. It describes ferromagnetic properties of a crystal lattice T_N.

Example 1.2.3 (Ising model):
The Ising model is given by a finite random field $X = \{X(t), t \in T_N\}$ on the phase space $S = E^{T_N}$, $E = \{-1, 1\}$ with

$$\mathbf{P}_X(x) = \mathbf{P}(X = x) = \mathbf{P}(X(s) = x_s, s \in T_N) = \frac{1}{c}\exp(-H(x)), \quad x = (x_t)_{t \in T_N},$$

where $H(x) \in (-\infty; +\infty]$ is the *energy functional*

$$H(x) = -\frac{1}{kt_o}\left(-mB\sum_{s \in T_N} x_s + J\sum_{\substack{s,t \in T_N \\ s \sim t}} x_s x_t\right).$$

The parameters have the following meaning:

- $t_0 > 0$ - absolute temperature
- $k = 1,38064852(79) \cdot 10^{-23}$ J/K - Boltzmann constant
- $B \in \mathbb{R}$ - intensity of the external ferromagnetic field
- $m > 0$ - material constant
- $J \in \mathbb{R}$ - interaction parameter ($J > 0$: ferromagnetic case, $J < 0$: antiferromagnetic case, $J = 0$: no interaction)

$c = \sum_{z \in S}\exp(-H(z))$ is a normalizing constant. Normally, the sum $\sum_{z \in S}$ runs $|S| = 2^{|T_N|} = 2^{(2N+1)^d}$ components and (for large N and d) is practically not computable. For $\sim$ one can take a 4-neighborhood relation.

§10. Gibbs random fields

Let $(T, \mathcal{T})$ be a finite non directed graph with neighborhood relation $\sim$ and E a finite state space. For any random field $X = \{X(t), t \in T\}$ with state space E, i.e. $X(t) : \Omega \to E$, $t \in T$, and distribution $\mathbf{P}_X(x) = \mathbf{P}(X = x)$, $x \in S = E^T$, $\mathbf{P}_X(x)$ can be written in the so-called *Gibbs form*:

$$\mathbf{P}_X(x) = \frac{1}{c} \exp(-H(x)),$$

where $H(x) = -\log \mathbf{P}_X(x) - \log c$ is the *energy function*, $H(x) \in (-\infty, +\infty]$, and $c = \sum_{z \in S} \exp(-H(z))$ is the *partition function*. The definition of $H(x)$ is unique up to an additive constant. One can achieve uniqueness by setting e.g. $H(x_0) = 0$ for some $x_o \in S$.

As all Gibbs fields are Markov with respect to some neighborhood relation $\sim$ on T, their consideration in this form adds nothing new to the subject, apart from the problem of the existence (and uniqueness) of infinite Gibbs random fields, i.e. fields defined as a limit of X as T expands, for instance, $T = T_N$, $T_N \nearrow \mathbb{Z}^d$ as $N \to \infty$. Does there exist a random field on $\mathbb{Z}^d$ such that it is a weak limit of finite Gibbs random fields on T_N as introduced in Examples 6 and 7 of §1.7? In general, this problem is quite complicated. Its solutions can be found in the book [39]. In particular, for the case of the Ising model it can be shown that this thermodynamic limit (as $N \to \infty$) exists for all parameters $J > 0$ (see [39], pp. 3-4).

1.3 Moments and covariance

Let $X = \{X(t), t \in T\}$ be a random function, $X(t) : \Omega \to \mathbb{R}$, $t \in T$ and T be an arbitrary index space.

Definition 1.3.1
The *(mixed) moment* $\mu^{(j_1,\ldots,j_n)}(t_1, \ldots, t_n)$ *of* X *of orders* $j_1, \ldots, j_n \in \mathbb{N}$ at index values $t_1, \ldots, t_n \in T$ is defined by

$$\mu^{(j_1,\ldots,j_n)}(t_1, \ldots, t_n) = \mathbf{E}\left[X^{j_1}(t_1) \cdot \ldots \cdot X^{j_n}(t_n)\right],$$

provided that this expectation exists and is finite. For that, it is sufficient to require $\mathbf{E}|X(t)|^j < \infty$ for all $t \in T$, where $j = j_1 + \ldots + j_n$. Special cases are

1. $\mu(t) = \mu^{(1)}(t) = \mathbf{E}X(t)$, $t \in T$ - *mean value function*
2. $\mu^{(1,1)}(s,t) = \mathbf{E}[X(s)X(t)]$, $s, t \in T$ - *(non-centered) covariance function*
3. $C(s,t) = \mathbf{cov}(X(s), X(t)) = \mu^{(1,1)}(s,t) - \mu^{(1)}(s)\mu^{(1)}(t)$, $s, t \in T$ - *(centered) covariance function*

Exercise 1.3.1
Prove that the covariance function $C(s,t)$ of a random field X

1. is symmetric, i.e. $C(s,t) = C(t,s)$, $s, t \in T$.
2. fulfills $C(t,t) = \mathbf{var}X(t)$, $t \in T$.
3. is *positive semi-definite*, i.e. for all $n \in \mathbb{N}$, $t_1, \ldots, t_n \in T$, $z_1, \ldots, z_n \in \mathbb{R}$ it holds that

$$\sum_{i=1}^{n} \sum_{j=1}^{n} C(t_i, t_j) z_i z_j \geq 0.$$

As opposed to the restrictions of positive semi-definiteness in Exercise 1.3.1, the mean value function can be arbitrary. It shows the (deterministic) *trend* of the random function X.
The correlation coefficient $R(s,t) = \frac{C(s,t)}{\sqrt{C(s,s)C(t,t)}}$ is sometimes called the *correlation function* of X. It holds $|R(s,t)| \le 1$, $s,t \in T$ by the inequality of Cauchy-Schwarz.

Exercise 1.3.2
Prove that both non-centered covariance function and correlation function are positive semi-definite.

Exercise 1.3.3
1. Prove that for any positive semi-definite function $K : T \times T \to \mathbb{R}$ there exists a centered random function X such that $C(s,t) = K(s,t)$, $s,t \in T$.
2. Let $K : T \times T \to \mathbb{R}$ be a positive semi-definite function and $m : T \to \mathbb{R}$ be any function with the property: $K(s,t) - m(s)m(t)$ is positive semi-definite. Show that there exists a random function $X : \Omega \times T \to \mathbb{R}$ such that $\mu(t) = m(t)$ and $C(s,t) = K(s,t) - m(s)m(t)$.

Exercise 1.3.4
Give examples of two different random functions X and Y with $\mathbf{E}X(t) = \mathbf{E}Y(t)$, $t \in T$ and $\mathbf{E}[X(s)X(t)] = \mathbf{E}[Y(s)Y(t)]$, $s,t \in T$.

From the solution of the above exercise it is clear that μ and C (as well as any first k moments of X) do not specify the distribution of X uniquely.

Definition 1.3.2
For any $t_1, \dots, t_n \in T$, $n \in \mathbb{N}$, let

$$\varphi(\lambda) = \varphi_{X_{t_1,\dots,t_n}}(\lambda) = \mathbf{E}\exp\{i\langle \lambda, X_{t_1,\dots,t_n}\rangle\}, \quad \lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$$

be the characteristic function of the random vector $X_{t_1,\dots,t_n} = (X(t_1), \dots, X(t_n))^\top$ (or, equivalently, of the finite-dimensional distribution $\mathbf{P}_{t_1,\dots,t_n}$ of X). For any $r = (r_1, \dots, r_n) \in \mathbb{N}^n$ let $|r| = \sum_{j=1}^n r_j$. The *semi-invariant* (cumulant, truncated correlation function) is defined by

$$s_X(t_1, \dots, t_n, r) = \frac{1}{i^{|r|}} \cdot \frac{\partial^{|r|} \log \varphi(\lambda)}{\partial \lambda_1^{r_1} \dots \partial \lambda_n^{r_n}}\bigg|_{\lambda=0}.$$

Sometimes, the notation $\langle X(t_1)^{r_1}, \dots, X(t_n)^{r_n}\rangle$ for $s_X(t_1, \dots, t_n, r)$ will be used. In particular, $\langle X(t_1), \dots, X(t_n)\rangle = s_X(t_1, \dots, t_n, \mathsf{e})$ for $\mathsf{e} = (1, \dots, 1)^\top$.

Exercise 1.3.5
Show that $\log \varphi(\lambda) \in C^\infty(\mathbb{R}^n)$, and thus $s_X(t_1, \dots, t_n, r)$ is well-defined for all $t_1, \dots, t_n \in T$, $r \in \mathbb{N}$, if

$$\mathbf{E}|X(t)|^k < \infty, \quad t \in T, \ k \in \mathbb{N}. \tag{1.3.1}$$

For later considerations, suppose (1.3.1) to hold true. The properties of s_X follow directly from the corresponding properties of characteristic functions:

1. **Symmetry:** for any permutation $\sigma : (1, \dots, n) \longmapsto (\sigma(1), \dots, \sigma(n))$ it holds

$$s_X(t_1, \dots, t_n, r) = s_X(t_{\sigma(1)}, \dots, t_{\sigma(n)}, r_{\sigma(1),\dots,\sigma(n)}), \quad n \in \mathbb{N}, \ t_1, \dots, t_n \in T.$$

2. **Multilinearity:** for any $a, b \in \mathbb{R}$, $n \in \mathbb{N}$, $t_0, \dots, t_n \in T$ it holds

$$\begin{aligned}&\langle aX(t_0) + bX(t_1), X(t_2), \dots, X(t_n)\rangle \\ &\qquad = a\langle X(t_0), X(t_2), \dots, X(t_n)\rangle + b\langle X(t_1), X(t_2), \dots, X(t_n)\rangle.\end{aligned}$$

3. Let $A \cup B = \{t_1, \dots, t_n\}$ be a non-empty partition of $\{t_1, \dots, t_n\}$ into two disjoint subsets. If $\{X(t), t \in A\}$ and $\{X(t), t \in B\}$ are independent, then $s_X(t_1, \dots, t_n, r) = 0$ for all $r \in \mathbb{N}^n$. Hence, semi-invariants characterize the degree of dependence of $X(t_1), \dots, X(t_n)$.

4. It holds $C(s,t) = s_X(s,t,(1,1)) = \langle X(s), X(t) \rangle$, $s, t \in T$.

Exercise 1.3.6
Prove the properties 1), 2), 3), 4) of the semi-invariant s_X.

Let us formulate a generalization of this result:

Theorem 1.3.1
It holds

1.
$$\mu^{(e)}(t_1, \dots, t_n) = \sum_{P_n} \prod_{\{t_{i_1}, \dots, t_{i_k}\} \in P_n} \langle X(t_{i_1}), \dots, X(t_{i_k}) \rangle ,$$
where P_n is any partition of the set $\{t_1, \dots, t_n\}$ into mutually disjoint non-empty blocks, and $\{t_{i_1}, \dots, t_{i_k}\}$ is one of such blocks.

2.
$$s_X(t_1, \dots, t_n, e) = \sum_{P_n} (-1)^{|P_n|-1} (|P_n| - 1)! \prod_{\{t_{i_1}, \dots, t_{i_k}\} \in P_n} \mu^{(e)}(t_{i_1}, \dots, t_{i_k}).$$

Proof See [39], p. 30. The idea is to compare the coefficients of the Taylor series of $\varphi(\lambda)$ and $\log \varphi(\lambda)$. □

Cumulants are used not only for the description of the dependence structure of the random vector $X_{t_1,\dots,t_n}$, but also for the characterization of its distribution.

Corollary 1.3.1
A random function X is Gaussian iff $s_X(t_1, \dots, t_n, \mathsf{e}) = 0$ for all $n > 2$, $t_1, \dots, t_n \in T$. This holds since the multivariate Gaussian distribution is the only one with $\log \varphi(\lambda) =$ polynomial of second order.

Example 1.3.1 ([42]): 1. $X \equiv a$ a.s. $\iff \langle X^r \rangle = \begin{cases} a, & r = 1, \\ 0, & r > 1 \end{cases}$

2. $X \sim \mathrm{Poi}(\lambda) \iff \langle X^r \rangle = \lambda$, $r \in \mathbb{N}$

3. $X \sim \mathrm{Exp}(\lambda) \iff \langle X^r \rangle = (r-1)! \lambda^{-r}$, $r \in \mathbb{N}$

Definition 1.3.3
For any $s, t \in T$ consider the increment $X(t) - X(s)$ and suppose that $\mathbf{E}|X(t)|^k < \infty$, $t \in T$, $k \in \mathbb{N}$.

1. $\gamma_k(s,t) = \mathbf{E}\left[X(t) - X(s)\right]^k$ is called the *mean increment of order* $k \in \mathbb{N}$.

2. In particular, $\gamma(s,t) = \frac{1}{2}\,\gamma_2(s,t) = \frac{1}{2}\,\mathbf{E}\left[X(t) - X(s)\right]^2$ is called the *variogram* of X.

Variograms are frequently used in geostatistics. One can easily see that the following relationship holds:

$$\gamma(s,t) = \frac{C(s,s) + C(t,t)}{2} - C(s,t) + \frac{1}{2}(\mu(s) - \mu(t))^2, \quad s, t \in T.$$

1.4 Stationarity and isotropy

In this section, we introduce the notions of spatial homogeneity of the distribution of random functions. Let the index space T be a linear vector space with operations $+$ and $\cdot$.

Definition 1.4.1
The random function $X = \{X(t), t \in T\}$ is called *(strictly) stationary* if for any $n \in \mathbb{N}$, $\tau, t_1, \ldots, t_n \in T$ it holds $\mathbf{P}_{t_1+\tau,\ldots,t_n+\tau} = \mathbf{P}_{t_1,\ldots,t_n}$, i.e. all finite-dimensional distributions of X are invariant with respect to shifts in T.

Definition 1.4.2
Let $X = \{X(t), t \in T\}$ be a random function with $\mathbf{E}|X(t)|^2 < \infty$, $t \in T$. X is called *stationary (in the wide sense)* if $\mu(t) \equiv \mu$, $t \in T$ and $C(s,t) = C(s+\tau, t+\tau)$ $(:= C(s-t))$, $\tau, s, t \in T$.

Notice that both definitions of stationarity do not imply each other. However, it is clear that if X is strictly stationary with a finite second moment, then it is stationary in the wide sense.

Definition 1.4.3
The random function $X = \{X(t), t \in T\}$ is *intrinsically stationary of order two* if its mean increments $\gamma_i(s,t)$, $s, t \in T$ exist up to the order 2, and

1. $\gamma_1(s,t) = 0$, $s, t \in T$.
2. $\gamma_2(s,t) = \gamma_2(s+\tau, t+\tau)$, $s, t, \tau \in T$.

It is clear that intrinsic stationarity of order two (which is widely used in practice) is a little bit more general than stationarity in the wide sense, because we require the existence of moments of increments of X and not of $X(t)$ itself. However, this distinction is of superficial nature, as most random functions which are of practical interest are stationary in the wide sense (and hence intrinsically stationary of order two).

Exactly as in the case of stationarity, the notion of isotropy can be introduced in the strict or wide sense. In order to define it, we need to assume that $T = \mathbb{R}^d$, $d \geq 2$. Isotropy can be easily explained as independence of the probabilistic properties of a random field X of a chosen observation direction. It is often required that isotropic processes are also stationary. However, we shall not do it in the following definition, as there are examples of isotropic non-stationary random fields, see Figure 1.8.

Fig. 1.8: Two realizations of isotropic non-stationary random fields

Definition 1.4.4
The random field $X = \{X(t), t \in \mathbb{R}^d\}$ is said to be *isotropic*

1. *in the strict sense*, if for any $n \in \mathbb{N}$, $t_1, \ldots, t_n \in \mathbb{R}^d$, $A \in \mathrm{SO}(d)$ [5]:

$$(X(At_1), \ldots, X(At_n))^\top \stackrel{d}{=} (X(t_1), \ldots, X(t_n))^\top.$$

2. *in the wide sense*, if for any $s, t \in \mathbb{R}^d$, $A \in \mathrm{SO}(d)$:

$$\mu(At) = \mu(t), \qquad C(As, At) = C(s, t).$$

Definition 1.4.5
The random field X is *motion invariant* if it is stationary and isotropic (both either in the strict or wide sense).

We shall mostly consider stationarity, isotropy and invariance with respect to rigid motions in the wide sense. This means that $\mu(t) = \mu$, $t \in \mathbb{R}^d$ and $C(s, t) = C_0(|t - s|)$ for such random fields, where $C_0 : \mathbb{R}_+ \to \mathbb{R}$ is a function of the distance between the points s and t.

Definition 1.4.6
The random function $X = \{X(t), t \in T\}$ on an index space T (which is a linear vector space) has *stationary increments*, if for all $n \in \mathbb{N}$ and $t, t_1, \ldots, t_n \in T$

$$(X(t_2) - X(t_1), \ldots, X(t_n) - X(t_1))^\top \stackrel{d}{=} (X(t_2 + t) - X(t_1 + t), \ldots, X(t_n + t) - X(t_1 + t))^\top.$$

In what follows (here and in §1.5) we suppose that $X = \{X(t), t \in T\}$ is a *centered* (i.e. $\mu(t) = 0, t \in T$) *random function* with values in $\mathbb{R}$ and index space T which is a Banach space as well. If X is not centered, it can be made so by considering the new random field $Y = \{Y(t) = X(t) - \mu(t), t \in T\}$. Sometimes we shall assume that X is wide sense stationary. In this case, its covariance function $C(h) = \mathbf{E}[X(t)X(t+h)]$, $h \in T$ satisfies obvious properties:

1. $C(0) = \mathbf{var} X(t) \geq 0$
2. Symmetry: $C(h) = C(-h)$, $h \in T$
3. Boundedness: $|C(h)| \leq C(0)$, $h \in T$

Further properties of stationary covariance functions will be considered in Chapter 2.

1.5 Continuity and differentiability

We consider different notions of continuity and differentiability of random functions which correspond to the respective convergence types in probability theory. As above, let X be a centered random function with index space T and phase space E (both being Banach spaces with norms $|\cdot|_T$ and $|\cdot|_E$).

Definition 1.5.1
The random function $X = \{X(t), t \in T\}$ is

[5] denotes the special orthogonal group in R^d

1. *stochastically continuous at* $t \in T$, if $X(s) \xrightarrow{p} X(t)$ as $s \to t$, where $\xrightarrow{p}$ means the convergence in probability, i.e.

$$\mathbf{P}\left(|X(s) - X(t)|_E > \varepsilon\right) \to 0, \quad \varepsilon > 0 \quad \text{as } s \to t.$$

2. *continuous in p-mean* $(p \geq 1)$ *at* $t \in T$, if $X(s) \xrightarrow{L^p} X(t)$ as $s \to t$, i.e.

$$\mathbf{E}\,|X(s) - X(t)|_E^p \to 0, \quad \text{as } s \to t.$$

   In particular, the case $p = 2$ has a special name: *continuity in quadratic mean*, which holds if $\gamma(s,t) \to 0$ as $s \to t$.

3. *almost surely continuous at* $t \in T$, if $X(s) \xrightarrow{a.s.} X(t)$ as $s \to t$, i.e.

$$\mathbf{P}\left(\lim_{s \to t} X(s) = X(t)\right) = 1.$$

In applications, the most interesting case is certainly the a.s. continuity, and an even stronger property of the continuity of almost all realizations of X. The weakest (and actually less meaningful) notion is the stochastic continuity:

cont. in p-mean $\Longrightarrow$ stochastic cont. $\Longleftarrow$ almost sure cont.

As it is shown on the above diagram, a random function is stochastically continuous if it is continuous a.s. or in p-mean, $p \geq 1$.

Let us investigate the properties of these three types of continuity in more detail. As it will follow from the next lemma, stochastic continuity is defined by the properties of the two-dimensional distribution $\mathbf{P}_{s,t}$ of X.

Lemma 1.5.1
The following statements are equivalent:

1. $$X(s) \xrightarrow{p} Y \quad \text{as } s \to t_0$$

2. $$\mathbf{P}_{s,t} \xrightarrow{\omega} \mathbf{P}_{(Y,Y)} \quad \text{as } s,t \to t_0$$

for $t_0 \in T$ and a random variable Y. Here $\xrightarrow{\omega}$ denotes the weak convergence of probability measures. For X to be stochastically continuous at t_0, set $Y = X(t_0)$.

Proof $1 \Rightarrow 2$: $X(s) \xrightarrow{p} Y$ as $s \to t_0$ implies $(X(s), X(t))^\top \xrightarrow{p} (Y,Y)^\top$ as $s,t \to t_0$. Thus $\mathbf{P}_{s,t} \xrightarrow{\omega} \mathbf{P}_{(Y,Y)}$ as $s,t \to t_0$, since convergence in probability implies weak convergence.
$2 \Rightarrow 1$: For any $\varepsilon > 0$, take a function $g_\varepsilon : \mathbb{R} \to [0;1]$, which is continuous on $\mathbb{R}$, $g_\varepsilon(0) = 0$ and $g_\varepsilon(x) = 1$ if $x \notin B_\varepsilon(0)$. For any $s,t \in T$ it holds

$$\begin{aligned}&\mathbf{E}\, g_\varepsilon(|X(s) - X(t)|_E) \\ &\quad = \mathbf{P}(|X(s) - X(t)|_E > \varepsilon) + \mathbf{E}\left[\, g_\varepsilon(|X(s) - X(t)|_E) \cdot \mathbf{1}\{|X(0) - X(t)|_E \leq \varepsilon\}\right],\end{aligned}$$

hence

$$\begin{aligned}\mathbf{P}(|X(s)-X(t)|_E>\varepsilon) &\le \mathbf{E}\, g_\varepsilon(|X(s)-X(t)|_E)\\ &= \int_{E\times E} g_\varepsilon(|x-y|_E)\mathbf{P}_{s,t}\,(d(x,y)) \to \int_{E\times E} g_\varepsilon(|x-y|_E)\mathbf{P}_{(Y,Y)}(d(x,y)) = 0\end{aligned}$$

as $s,t \to t_0$, since $\mathbf{P}_{(Y,Y)}$ is concentrated on $\{(x,y)\in E\times E : x=y\}$ and $g_\varepsilon(0)=0$. Hence $\{X(s)\}_{s\to t_0}$ is a fundamental sequence in probability, and then $X(s)\xrightarrow{p} Y$ as $s\to t_0$. □

It can happen that X is stochastically continuous whereas all its realizations are discontinuous (in the sense that X does not have an a.s. continuous modification).

Exercise 1.5.1
Show this property for the homogeneous Poisson process on the real line.

The explanation of this fact is easy: the probability to have a discontinuity at a concrete $t\in T$ is zero; hence discontinuities happen at different $t\in T$ for all realizations of X.

Exercise 1.5.2
Let T be compact. Prove that if X is stochastically continuous on T (i.e. for all $t\in T$), then it is uniformly stochastically continuous on T, which means that for all $\varepsilon,\eta>0$ there exists a $\delta>0$ such that for all $s,t\in T$ with $|s-t|_T<\delta$, it holds that

$$\mathbf{P}(|X(s)-X(t)|_E>\varepsilon)<\eta.$$

Let us turn to the continuity in quadratic mean. In order to state further results, suppose that $E=\mathbb{R}$ and $\mathbf{E}X^2(t)<\infty$, $t\in T$. Since X is centered, it holds $C(s,t)=\mathbf{E}[X(s)X(t)]$.

Lemma 1.5.2
For any $t_0\in T$ and some random variable Y with $\mathbf{E}Y^2<\infty$, the following statements are equivalent:

1. $$X(s)\xrightarrow{L^2} Y \quad \text{as } s\to t_0$$

2. $$C(s,t)\to \mathbf{E}Y^2 \quad \text{as } s,t\to t_0$$

Remark 1.5.1 1. For the mean quadratic continuity of X at t_0 just set $Y=X(t_0)$. Then, $\mathbf{E}Y^2=C(t_0,t_0)$, i.e. C must be continuous at (t_0,t_0).

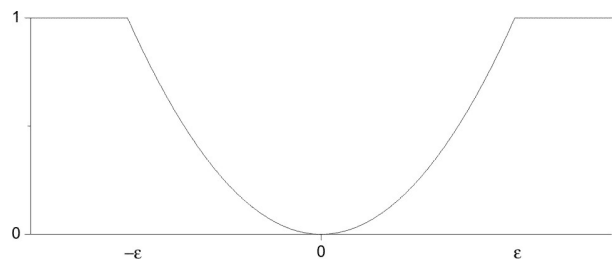


Fig. 1.9: The function $g_\varepsilon(x)$

2. If X is not centered, it is enough to require $\mu(s) \to \mu(t_0)$ as $s \to t_0$ alongside with $C(s,t) \to C(t_0,t_0)$ as $s,t \to t_0$ for the L^2-continuity of X at $t_0 \in T$.

Proof of Lemma 1.5.2. $1 \Rightarrow 2$: First prove the following general fact: If $\lim_{n\to\infty} X_n \overset{L^2}{=} X$ and $\lim_{n\to\infty} Y_n \overset{L^2}{=} Y$ for some sequences $\{X_n\}_{n\in\mathbb{N}}$ and $\{Y_n\}_{n\in\mathbb{N}}$ of random variables X and Y, then $\mathbf{E}[X_m Y_n] \to \mathbf{E}[XY]$ as $n,m \to \infty$. Indeed, it holds

$$\begin{aligned}
&|\mathbf{E}[X_m Y_n] - \mathbf{E}[XY]| \\
&\qquad = |\mathbf{E}[(X_m - X + X)(Y_n - Y + Y)] - \mathbf{E}[XY]| \\
&\qquad \le \mathbf{E}\,|(X_m - X)(Y_n - Y)| + \mathbf{E}\,|(X_m - X)Y| + \mathbf{E}\,|(Y_n - Y)X| \\
&\qquad \le \sqrt{\mathbf{E}(X_m - X)^2\,\mathbf{E}(Y_n - Y)^2} + \sqrt{\mathbf{E}Y^2\,\mathbf{E}(X_m - X)^2} + \sqrt{\mathbf{E}X^2\,\mathbf{E}(Y_n - Y)^2} \to 0
\end{aligned}$$

as $m,n \to \infty$ by the inequality of Cauchy-Schwarz. Hence, from $X(s) \xrightarrow{L^2} Y$ as $s \to t_0$ and $X(t) \xrightarrow{L^2} Y$ as $t \to t_0$ it follows

$$C(s,t) = \mathbf{E}[X(s)X(t)] \to \mathbf{E}Y^2$$

as $s,t \to t_0$.
$2 \Rightarrow 1$: Use the relation in Definition 1.3.3 for $\gamma(s,t)$:

$$\begin{aligned}
\mathbf{E}[X(s) - X(t)]^2 &= 2\gamma(s,t) = C(s,s) + C(t,t) - 2C(s,t) \\
&\to 2\mathbf{E}Y^2 - 2\mathbf{E}Y^2 = 0
\end{aligned} \tag{1.5.1}$$

as $s,t \to t_0$. Hence $\{X(s)\}_{s\to t_0}$ is a fundamental sequence in the mean quadratic sense, and it follows $X(s) \xrightarrow{L^2} Y$ as $s \to t_0$. □

A mean-square continuous random function X can still have discontinuous realizations. Still in all cases of practical relevance (i.e., except for some "pathologic" random functions X) mean-square continuous random functions have an a.s. continuous modification. Later on, this statement will be made more precise.

Exercise 1.5.3
Construct a stationary shot-noise process with a.s. discontinuous realizations which is continuous in quadratic mean.

Let us consider the special case of a stationary centered random function X. In this case, the property in Definition 1.3.3 reads

$$\gamma(s,t) = \gamma(s-t) = C(0) - C(s-t), \quad s,t \in T$$

or

$$\gamma(h) = C(0) - C(h), \quad h \in T, \tag{1.5.2}$$

where $C(0) = \mathbf{var}X(t)$. This allows us to state the following

Corollary 1.5.1 1. A stationary random function X is mean-square continuous iff its covariance function is continuous at the origin, i.e. $\lim_{h\to 0} C(h) = C(0)$.

2. If $C(h)$ is continuous at the origin, then it is continuous everywhere on T.

Proof 1. This follows directly from Remark 1.5.1, 1) and Lemma 1.5.2 with $C(h) = C(s,t)$, $s - t = h$.

2. Prove that $\lim_{h\to 0} C(h + h_0) = C(h_0)$ for all $h_0 \in T$ if it holds for $h_0 = 0$. Indeed, for any $h \in T$ it holds

$$\begin{aligned} |C(h+h_0) - C(h_0)| &= |\mathbf{E}[X(t+h)X(t-h_0)] - \mathbf{E}[X(t)X(t-h_0)]| \\ &\le \mathbf{E}\,|(X(t+h) - X(t))X(t-h_0)| \le \sqrt{2\gamma(h)}\sqrt{\mathbf{E}X^2(t-h_0)} \\ &= \sqrt{2(C(0) - C(h))C(0)} \to 0 \end{aligned}$$

as $h \to 0$ by the inequality of Cauchy-Schwarz and relation (1.5.2).

□

Now let us turn to the property of a.s. continuity. If we are looking for a condition imposed on the finite-dimensional distributions of X assuring the a.s. continuity of its realizations, it is clear that any such condition can guarantee only the existence of a modification of the random function X with a.s. continuous paths. There are many such sufficient conditions (see e.g [13], [19], v. 1., pp. 192-193, for Gaussian random functions see [1], pp. 11-19 and [11], pp. 170-171). Let us give only one example of those (attributed to N.A. Kolmogorov) in the case of X being a stochastic process defined on $T = [a;b] \subset \mathbb{R}$.

Theorem 1.5.1
The stochastic process $X = \{X(t), t \in [a;b]\}$ is a.s. continuous if for some $\alpha, c, \delta > 0$ and sufficiently small h it holds that

$$\mathbf{E}|X(t+h) - X(t)|^\alpha < c|h|^{1+\delta}, \quad t \in (a;b).$$

Proof See [11], pp. 63-65. □

The above condition of X is a condition on its two-dimensional distributions $\mathbf{P}_{t,t+h}$.

Definition 1.5.2
The random function $X = \{X(t), t \in T\}$ is *differentiable at $t \in T$ in direction h stochastically, in p-mean* ($p \ge 1$) or *almost surely*, if there exists

$$\lim_{l\to 0} \frac{X(t+hl) - X(t)}{l} := X'_h(t)$$

in the corresponding sense, i.e. in probability, in L^p or with probability one.

One says that X is *differentiable on T* if it is differentiable at any $t \in T$. The differentiability of order k is defined accordingly in the iterative way.

By Lemmas 1.5.1 and 1.5.2 it is easy to see that the stochastic differentiability is defined by the properties of the three-dimensional distributions of X (as the joint distribution of $\frac{X(t+hl)-X(t)}{l}$ and $\frac{X(s+hl')-X(s)}{l'}$ should weakly converge), whereas the differentiability in quadratic mean is regulated by the properties of the covariance function $C(s,t)$ (hence, of the two-dimensional distributions of X).

Exercise 1.5.4
Show that

1. the Wiener process is not even stochastically differentiable at any $t \in [0; \infty)$.
2. the homogeneous Poisson process on $\mathbb{R}$ is stochastically differentiable on $\mathbb{R}$, but not in p-mean, $p \geq 1$.

Let us dwell on the mean-quadratic differentiability of centered random functions $X = \{X(t), t \in T\}$.

Lemma 1.5.3
The random function $X = \{X(t), t \in T\}$ is differentiable in quadratic mean at $t \in T$ in direction h if its covariance function $C(s,t)$ is twice differentiable at (t,t) in direction h, i.e. if there exists

$$C''_{hh}(t,t) = \left.\frac{\partial^2 C(s,t)}{\partial s_h \partial t_h}\right|_{s=t}.$$

Moreover, $X'_h(t)$ is mean-square continuous at $t \in T$, if

$$C''_{hh}(s,t) = \frac{\partial^2 C(s,t)}{\partial s_h \partial t_h}$$

is continuous at $s = t$. Here $C''_{hh}(s,t)$ is the covariance function of $X'_h = \{X'_h(t), t \in T\}$.

Proof Due to Lemma 1.5.2, it suffices to show the existence of the limit

$$\lim_{l,l' \to 0} \mathbf{E}\left[\frac{X(t+lh) - X(t)}{l} \cdot \frac{X(s+l'h) - X(s)}{l'}\right]$$

for $s = t$. Indeed, we have

$$\begin{aligned}
&\mathbf{E}\left[\frac{X(t+lh) - X(t)}{l} \cdot \frac{X(s+l'h) - X(s)}{l'}\right] \\
&= \frac{1}{ll'}\left[C(t+lh, s+l'h) - C(t+lh, s) - C(t, s+l'h) + C(t,s)\right] \\
&= \frac{1}{l}\left[\frac{C(t+lh, s+l'h) - C(t+lh, s)}{l'} - \frac{C(t, s+l'h) - C(t,s)}{l'}\right] \to C''_{hh}(s,t)
\end{aligned}$$

as $l, l' \to 0$. All statements of the Lemma immediately follow from this relation. □

Corollary 1.5.2
If $X = \{X(t), t \in T\}$ is a wide sense stationary centered random function with covariance function C, then

1. X is differentiable in quadratic mean in direction $h \in T$ iff $C''_{hh}(0)$ exists and is finite.
2. If $C(\tau)$ is twice differentiable in direction $h \in T$ at $\tau = 0$ then it is twice differentiable in this direction for all $\tau \in T$. Then, it holds $C'_h(0) = 0$.
3. If $C''_{hh}(\tau)$ is continuous at $\tau = 0$, then it is continuous for any $\tau \in T$.
4. $-C''_{hh}(\tau)$ is the covariance function of X'_h, i.e. it holds $-C''_{hh}(\tau) = \mathbf{E}[X'_h(t)X'_h(t+\tau)]$.
5. If X is differentiable in quadratic mean in direction h, then X and X'_h are uncorrelated, i.e. $\mathbf{E}[X(t)X'_h(t)] = 0$, $t \in T$.

Proof Following the guidelines of the proof of Lemma 1.5.3, we get for $s, t \in T$

$$\begin{aligned}\lim_{l,l'\to 0} \mathbf{E}\left[\frac{X(t+lh)-X(t)}{l}\cdot\frac{X(s+l'h)-X(s)}{l'}\right] \\ = \lim_{l,l'\to 0}\frac{1}{l}\left[\frac{C(t-s+(l-l')h)-C(t-s+lh)}{l'} - \frac{C(t-s-l'h)-C(t-s)}{l'}\right] \\ = -C''_{hh}(t-s).\end{aligned} \tag{1.5.3}$$

This proves 1) for $s = t$ and 2), 4) for $\tau = t - s$. The property $C'_h(0) = 0$ follows from the symmetry property $C(-\tau) = C(\tau)$, $\tau \in T$.
3) Since $-C''_{hh}$ is the covariance function of X'_h due to 4), the continuity of $C''_{hh}(\tau)$, $\tau \in T$ is guaranteed by the continuity of $C''_{hh}(0)$ (see Corollary 1.5.1).
5) Similarly as in (1.5.3), it can be shown that

$$\mathbf{E}[X(t)X'_h(t+\tau)] = C'_h(\tau),$$

for $t, \tau \in T$. The rest of the proof follows from 2) $(C'_h(0) = 0)$. □

Remark 1.5.2
The properties of L^2-differentiability and a.s. differentiability are disjoint in the following sense: One can construct examples of stochastic processes that are mean-squared differentiable, but have discontinuous realizations, and vice versa, processes with a.s. differentiable realizations might have discontinuous $C'_h(\tau)$ at $\tau = 0$ and hence be not mean-squared differentiable.

Exercise 1.5.5
Give the corresponding examples!

Remark 1.5.3
If $C''_{hh}(0)$ exists for a stationary stochastic process, then the condition of Theorem 1.5.1 is fulfilled with $\alpha = 2$ and $\delta = 1$. Hence, X will have an a.s. continuous modification. Under further regularity conditions, it can be shown that $X'_h(t)$ exists in a.s. sense for almost all $t \in [a; b]$ and is a.s. integrable on $[a; b]$. See also [11], pp. 125, 171.

1.6 Proof of the Theorem of Kolmogorov

Let us prove Theorem 1.1.2. Here we give only an idea of the proof following [9], Appendix 1.

Proof According to page 4 let $(E_t, \mathcal{E}_t)_{t\in T}$ be a family of Borel spaces, and

$$E_{t_1,\ldots,t_n} = E_{t_1} \times \ldots \times E_{t_n},$$

$$\mathcal{E}_{t_1,\ldots,t_n} = \mathcal{E}_{t_1} \otimes \ldots \otimes \mathcal{E}_{t_n},$$

for all $n \in \mathbb{N}$, $t_1, \ldots, t_n \in T$, $t_i \neq t_j$, $i \neq j$. Let $\{\mathbf{P}_{t_1\ldots t_n}\}$ be a family of measures on $(E_{t_1,\ldots,t_n}, \mathcal{E}_{t_1,\ldots,t_n})$ satisfying the properties of symmetry and consistency as on page 4. We show that there exists a unique probability measure $\mu(\cdot)$ on $(\mathsf{S}, \mathcal{B}_T)$ such that the $\mathbf{P}_{t_1,\ldots,t_n}$ are "projections" of μ onto $(E_{t_1,\ldots,t_n}, \mathcal{E}_{t_1,\ldots,t_n})$. These "projections" are understood in the following sense: Let $\mathcal{F}(T)$ be the family of all finite subsets of T, consisting of index sets $J \in \mathcal{F}(T)$, $J = \{t_1, \ldots, t_n\}$. For $(E_{t_1,\ldots,t_n}, \mathcal{E}_{t_1,\ldots,t_n})$ we use the notation $(E_J, \mathcal{E}_J)$. Analogously, $\mathbf{P}_J$ denotes $\mathbf{P}_{t_1,\ldots,t_n}$. This notation is correct due to the symmetry property of the measures $\mathbf{P}_{t_1,\ldots,t_n}$.

Let $\pi_{T,J} : \mathsf{S} \to E_J$ be the coordinate projection of a function $\{y(t), t \in T\} \in \mathsf{S}$ onto coordinates with indexes in J, i.e. $\pi_{T,J}\, y = (y(t_1), \dots, y(t_n))$ or $\pi_{T,J}\, \{y(t), t \in T\} = \{y(t), t \in J\}$, if $J = \{t_1, \dots, t_n\}$. Then, "projection" of μ means that

$$\mu \circ \pi_{T,J}^{-1}(B_{t_1} \times \dots \times B_{t_n}) = \mathbf{P}_J(B_{t_1} \times \dots \times B_{t_n})$$

for any $n \in \mathbb{N}$, $J = \{t_1, \dots, t_n\} \in \mathcal{F}(T)$, $B_{t_i} \in \mathcal{E}_{t_i}$, $i = 1, \dots, n$.

If such a probability measure μ on $(\mathsf{S}, \mathcal{B}_T)$ is found, then the random function X with this distribution μ is given by the coordinate projection on the canonical space: We take $\Omega = \mathsf{S}$, $\mathcal{F} = \mathcal{B}_T$, $\mathbf{P} = \mu$ and set $X(t, \omega) = \omega(t)$ for $\omega \in \mathsf{S}$ and $t \in T$. Let us transfer measures $\mathbf{P}_J$ from $(E_J, \mathcal{E}_J)$ onto the space $(\mathsf{S}, \mathcal{B}_{T,J})$ with σ-Algebra $\mathcal{B}_{T,J} = \pi_{T,J}^{-1}\, \mathcal{E}_J \subset \mathcal{B}_T$ in the following way: A new measure μ_J is defined on $(\mathsf{S}, \mathcal{B}_{T,J})$ by

$$\mu_J(\widetilde{B}) = \mathbf{P}_J(B),$$

where $B = B_{t_1} \times \dots \times B_{t_n} \in \mathcal{E}_J$, $J = \{t_1, \dots, t_n\}$ and $\widetilde{B} = B \times \prod_{t \in T \setminus J} E_t \in \mathcal{B}_{T,J}$. It is evident that $B \in \mathcal{E}_J$ iff $\widetilde{B}$ defined as above belongs to $\mathcal{B}_{T,J}$.

So, our goal is to find a measure μ on $(\mathsf{S}, \mathcal{B}_T)$ such that $\mu = \mu_J$ on any $\mathcal{B}_{T,J}$, $J \in \mathcal{F}(T)$. Equivalently, let us define μ on the "cylindrical" algebra $\mathcal{C}_T = \bigcup_{J \in \mathcal{F}(T)} \mathcal{B}_{T,J}$ by letting $\mu(B) = \mu_J(B)$, if there exists a $J \in \mathcal{F}(T)$ such that $B \in \mathcal{B}_{T,J}$. This definition yields a finitely additive function on $\mathcal{C}_T$. By Caratheodory's Theorem, μ can be uniquely extended to a measure on $\mathcal{B}_T = \sigma(\mathcal{C}_T)$ if μ is σ-additive on $\mathcal{C}_T$. So, let us show the countable additivity, which is equivalent to showing finite additivity and continuity in $\emptyset$, i.e. for any sequence $\{C_n\} \subset \mathcal{C}_T$, $C_n \searrow \emptyset$ ($C_{n+1} \subset C_n$, $\bigcap_{n=1}^{\infty} C_n = \emptyset$) it must hold $\mu(C_n) \to 0$ as $n \to \infty$. Thus, prove that if $\mu(C_n) \geq \varepsilon_0$ for some $\varepsilon_0 > 0$ and all large enough n then

$$C_n \nrightarrow \emptyset \quad \text{as} \quad n \to \infty. \tag{1.6.1}$$

Since $C_n \in \mathcal{C}_T$ it means that there exists a $J_n \in \mathcal{F}(T)$ such that $C_n \in \mathcal{B}_{T,J_n}$.

Without loss of generality we now assume that $J_n \subset J_{n+1}$. First prove the statement (1.6.1) for spaces $E_t = [0; 1]$, $\mathcal{E}_t = \mathcal{B}([0; 1])$ equipped with the Euclidean metric $d(\cdot, \cdot)$. After that, generalize this proof to arbitrary Borel spaces. Reduce the consideration to sets $C_n = \pi_{T,J_n}^{-1} B_n$, where the B_n are compact in E_{J_n}, $J_n \in \mathcal{F}(T)$.

First formulate the following results (without proof), which are helpful in the sequel:

Lemma 1.6.1
Let E_t, $t \in T$ be separable metric spaces with metrics $d_t(\cdot, \cdot)$ and Borel σ-algebras $\mathcal{B}_t$. Then, the space $\mathsf{S}_J = \{\text{functions } y \text{ on } J : y(t) \in E_t\}$, $J \in \mathcal{F}(T)$ equipped with the metric $d_J(x, y) = \max_{t \in J} d_t(x(t), y(t))$, $x, y \in \mathsf{S}_J$ is a separable metric space with "cylindrical" σ-algebra $\mathcal{B}_J$ equal to the Borel σ-algebra $\mathcal{B}(\mathsf{S}_J)$. If E_t, $t \in T$ are complete (Polish), then S_J is complete (Polish) as well.

Lemma 1.6.2
Let E be a metric space with measure μ on the Borel σ-algebra $\mathcal{B}(E)$. Then it holds

$$\mu(B) = \sup_{\substack{A \subset B \\ A \text{ closed}}} \mu(A) = \inf_{\substack{A \supset B \\ A \text{ open}}} \mu(A), \quad B \in \mathcal{B}(E).$$

By Lemma 1.6.1, it holds $\mathcal{B}_{J_n} = \mathcal{B}(\mathsf{S}_{J_n})$. By Lemma 1.6.2, for any $B_n \in \mathcal{B}_{J_n}$ there exists a closed set $K_n \subset B_n$ in $(\mathsf{S}_{J_n}, d_{J_n})$ such that $\mu_{J_n}(B_n \setminus K_n) < 2^{-(n+1)}\varepsilon_0$, $n \in \mathbb{N}$. Since S_{J_n} is compact, so is K_n. Then, for $C_n = \pi^{-1}_{T,J_n} B_n$ and $\widetilde{K_n} = \pi^{-1}_{T,J_n} K_n$, it holds

$$\mu(C_n \setminus \widetilde{K_n}) = \mu_{J_n}(B_n \setminus K_n) < 2^{-(n+1)}\varepsilon_0, \quad n \in \mathbb{N}.$$

Introduce the sets $L_n = \bigcap_{i=1}^n \widetilde{K_i}$. It holds $L_n \searrow \emptyset$ since $L_{n+1} \subset L_n$ and $L_n \subset \widetilde{K_i} \subset C_n$. Furthermore,

$$\begin{aligned}\mu(C_n \setminus L_n) &= \mu(C_n \cap (\bigcup_{i=1}^n \widetilde{K_i})) = \mu(\bigcup_{i=1}^n (C_n \setminus \widetilde{K_i})) \\ &\le \sum_{i=1}^n \mu(C_i \setminus \widetilde{K_i}) < \sum_{i=1}^n 2^{-(i+1)}\varepsilon_0 < \varepsilon_0/2\end{aligned}$$

and thus

$$\varepsilon_0 \le \mu(C_n) = \mu(L_n) + \mu(C_n \setminus L_n) \le \mu(L_n) + \varepsilon_0/2.$$

Hence, it follows $\mu(L_n) \ge \varepsilon_0/2$ for all $n \in \mathbb{N}$. Since the basis $D_n = \bigcap_{i=1}^n \pi^{-1}_{J_n,J_i} K_i$ of the cylinder $L_n = \pi^{-1}_{T,J_n} D_n$ is compact, we can further consider only $C_n = \pi^{-1}_{T,J_n} B_n$ for compact sets B_n in S_{J_n}. Define $J = \bigcup_{n=1}^\infty J_n$. On S_J, introduce the metric

$$d_J(x,y) = \sum_{n=1}^\infty 2^{-n} \frac{d_{J_n}(x|_{J_n}, y|_{J_n})}{1 + d_{J_n}(x|_{J_n}, y|_{J_n})}.$$

Similarly to the result of Lemma 1.6.1, one can show that the metric space (S_J, d_J) is a Polish space with "cylindrical" σ-algebra $\mathcal{B}_J$ being equal to $\mathcal{B}(\mathsf{S}_J)$, which is the usual Borel σ-algebra. S_J is compact. Then, the sets $\widehat{C_n} = \pi^{-1}_{J,J_n} B_n$, $n \in \mathbb{N}$ are compact in S_J as closed subsets of a compact set (since π_{J,J_n} is continuous). It holds

$$C_n = \pi^{-1}_{T,J_n} B_n = \pi^{-1}_{T,J} \pi^{-1}_{J,J_n} B_n = \pi^{-1}_{T,J} \widehat{C_n}.$$

Since $C_n \subset C_m$, $n \ge m$ it follows $\widehat{C_n} \subset \widehat{C_m}$, $n \ge m$ and $\bigcap_{n=1}^\infty C_n = \pi^{-1}_{T,J}(\bigcap_{n=1}^\infty \widehat{C_n})$.

A sequence of contracting compacts has at least one point in its intersection, and is hence non-empty, i.e. $\bigcap_{n=1}^\infty \widehat{C_n} \ne \emptyset$ and $\bigcap_{n=1}^\infty C_n \ne \emptyset$. This is a contradiction to the assumption $\bigcap_{n=1}^\infty C_n = \emptyset$.

Now let $E_t \in \mathcal{B}([0;1])$, $\mathcal{B}_t = E_t \cap \mathcal{B}([0;1])$ and $E_t \ne [0;1]$ for at least one $t \in T$. Set $L_t = [0;1]$, $t \in T$, $L_J = \prod_{t \in J} L_t$, $J \in \mathcal{F}(T)$. Introduce the measures

$$\widetilde{\mu}_J(B) = \mu_J(B \cap \mathsf{S}_J)$$

on $\mathcal{B}(L_J)$. These measures satisfy the conditions of symmetry and consistency for any $J \in \mathcal{F}(T)$. As above, there exists a measure $\widetilde{\mu}$ on $(L_T, \mathcal{L}_T)$, $\mathcal{L}_T = \bigotimes_{t \in T} \mathcal{B}([0;1])$ with projection $\widetilde{\mu}_J$ onto $(L_J, \mathcal{B}(L_J))$, $J \in \mathcal{F}(T)$ and a random function $\widetilde{X}$ on $(\Omega, \mathcal{F}, \mathbf{P})$, where $\Omega = L_T$, $\mathcal{F} = \mathcal{L}_T$, $\mathbf{P} = \widetilde{\mu}$, with the following distribution: $\widetilde{X}(t,\omega) = \omega(t)$, $\omega \in L_T$.

Now let $E_t \ne [0;1]$ for at least one $t \in T$. We define the random function $X : T \times \Omega \to E_T$ as

$$X(t,\omega) = \begin{cases} \widetilde{X}(t,\omega) & \text{for } \widetilde{X}(t,\omega) \in E_t, \\ a_t & \text{for } \widetilde{X}(t,\omega) \notin E_t, \end{cases}$$

where $a_t \in E_t$ is some point of $E_t \neq \emptyset$. For all $t \in T$ it holds that

$$\{\omega : X(t,\omega) \neq \widetilde{X}(t,\omega)\} = \{\omega : \omega(t) \in [0;1] \setminus E_t\} = C_t.$$

C_t is an elementary cylinder and hence

$$\mathbf{P}(C_t) = \widetilde{\mu}(C_t) = \mu_{\{t\}}(\{[0;1] \setminus E_t\} \cap E_t) = 0.$$

Moreover,

$$\begin{aligned}\mathbf{P}_{t_1,\ldots,t_n}(B_{t_1} \times \ldots \times B_{t_n}) &= \mathbf{P}(X(t_i) \in B_{t_i}, i = 1,\ldots,n) \\ &= \mathbf{P}(\widetilde{X}(t_i) \in B_{t_i}, i = 1,\ldots,n) = \widetilde{\mu}_J(B_J) = \mu_J(B_J),\end{aligned}$$

where $B_J = B_{t_1} \times \ldots \times B_{t_n}$. Hence, $\mu = \mathbf{P}_X$ is the measure on $\mathcal{B}_T$ we looked for.

Now, the extension of the above result to the case of arbitrary spaces $(E_t, \mathcal{E}_t)$, which are Borel spaces, follows from the next lemma.

Lemma 1.6.3
Let $(E_t, \mathcal{E}_t)_{t\in T}$ and $(\mathsf{G}_t, \mathcal{A}_t)_{t\in T}$ be families of isomorphic measurable spaces: $(E_t, \mathcal{E}_t) \sim (\mathsf{G}_t, \mathcal{A}_t)$ for all $t \in T$. Then, for all $J \subset T$, it holds $(E_J, \mathcal{B}_J) \sim (\mathsf{G}_J, \mathcal{A}_J)$, where $\mathcal{B}_J$ and $\mathcal{A}_J$ are cylindrical σ-algebras on $E_J = \prod_{t\in J} E_t$, $\mathsf{G}_J = \prod_{t\in J} \mathsf{G}_t$.

□

1.7 Additional exercises

1. Prove the existence of a random field with the following finite dimensional distributions and specify the measurable spaces $(E_{t_1\ldots t_n}, \mathcal{E}_{t_1\ldots,t_n})$:
   a) The finite dimensional distributions are multivariate Gaussian.
   b) The marginals are Poisson and X_t, X_s are independent for $t \neq s$.
2. Give an example for a family of probability measures $\{\mathbf{P}_{t_1\ldots t_n}\}$ which do not fulfill the conditions in the theorem of Kolmogorov.
3. Give additional examples of a non-continuous random function which has a continuous modification.
4. A random process W defined on $\mathcal{B}(\mathbb{R}^d)$ such that, for all sets $A, B \in \mathcal{B}(\mathbb{R}^d)$ with $\lambda_d(A)$ and $\lambda_d(B)$ finite,
   - $W(B) \sim \mathcal{N}(0, |B|)$,
   - $A \cap B = \emptyset \Rightarrow W(A \cup B) = W(A) + W(B)$ almost surely,
   - $A \cap B = \emptyset \Rightarrow W(A)$ and $W(B)$ are independent,

   is called *Gaussian white noise indexed by Borel sets*. Define the field $X = \{X_t, t \in [0;\infty)^d\}$ by $X_t = W([0;t])$ where $[0;t]$ is the paraxial rectangle with lower left corner 0 and upper right corner t. Show (for $d = 2$) that X is a Gaussian random field with $\mathbf{E}X_t = 0$ and $\mathbf{cov}(X_s, X_t) = \prod_{j=1}^d \min\{s_j, t_j\}$.

   Note: The process X is called Brownian sheet or multiparameter Brownian motion. For $d = 1$ it is the Brownian motion.

5. Let $\eta_1, \eta_2, \ldots$ be non-negative iid random variables and $X_1, X_2, \ldots$ non-negative iid random variables, where η_j and X_j are independent for all j. Define the process $N = \{N_t, t \in [0;\infty)\}$ by

$$N_t(\omega) = \sup\left\{ n \in \mathbb{N} : \sum_{j=1}^{n} \eta_j(\omega) \le t \right\}.$$

The process $Y = \{Y_t, t \in [0,\infty)\}$ is given by

$$Y_t(\omega) = y_0 + ct - \sum_{j=1}^{N_t(\omega)} X_j(\omega),\ t \ge 0$$

with positive constants y_0 and c. Draw a sketch of a trajectory of Y.

Note: The process Y_t is the classical Cramér-Lundberg model in insurance mathematics. It describes the surplus at time t of an insurance portfolio, where y_0 is the initial capital, c is the premium intensity and $\{X_j\}_{j\ge1}$ is the sequence of claim sizes. N_t is called the claim number process, whereas $\{\eta_j\}_{j\ge1}$ are claim inter-arrival times.

6. Let Φ be a homogeneous Poisson point process in $\mathbb{R}^d$ with intensity $\lambda > 0$.
   a) Write down the finite-dimensional distributions of Φ for disjoint bounded Borel sets $B_1, \ldots, B_n$.
   b) Compute the expectation $\mathbf{E}\Phi(B)$ and the variance $\mathbf{var}\,\Phi(B)$ for bounded Borel sets B and interpret why λ is called the intensity of Φ.

7. Let Φ be a homogeneous Poisson process in $\mathbb{R}^d$ of intensity $\lambda > 0$.
   a) *Independent thinning:* Each point of Φ is retained with probability $p \in (0;1)$ and deleted with probability $1-p$, independently of other points. Prove that the thinned process $\widetilde{\Phi}$ is a homogeneous Poisson process of intensity λp.
   b) *Superposition:* The superposition of two independent Poisson processes Φ_1 with intensity λ_1 and Φ_2 with intensity λ_2 is the union $\widetilde{\Phi} = \Phi_1 \cup \Phi_2$. Prove that $\widetilde{\Phi}$ is a homogeneous Poisson process with intensity $\lambda_1 + \lambda_2$.

Hint: The void probabilities of a point process are defined by $v_B = P(\Phi(B) = 0)$ for all Borel sets B. A simple point process is characterized by the void probabilities v_B as B ranges through the Borel sets.

8. Consider the shot-noise field $X_t = \sum_{v\in\Phi} g(t-v)$ where Φ is a homogeneous Poisson process of intensity λ and $g : \mathbb{R}^d \to \mathbb{R}$ is a deterministic function fulfilling the integrability conditions on p. 31. Prove that
   a) $\mathbf{E}X_t = \lambda \int_{\mathbb{R}^d} g(t-z)\,dz$
   b) $\mathbf{cov}(X_t, X_s) = \lambda \int_{\mathbb{R}^d} g(t-z)\,g(s-z)\,dz$.

Hint: Campbell's theorem can be useful: Let Φ be a homogeneous Poisson process in $\mathbb{R}^d$ with intensity λ and $f : \mathbb{R}^d \to \mathbb{R}$ a non-negative measurable function. Then it holds that

$$\mathbf{E}\sum_{v\in\Phi} f(v) = \lambda \int_{\mathbb{R}^d} f(z)\,dz.$$

Further, it holds that

$$\mathbf{E} \sum_{\substack{v,w\in\Phi \\ v\neq w}} f(v)\, g(w) = \lambda^2 \int_{\mathbb{R}^d} f(z)\, dz \int_{\mathbb{R}^d} g(z)\, dz$$

9. Consider the Ising model given in Example 1.2.3.

   a) Prove that the Ising model has the Markov property with respect to the 4-neighborhood relation $\sim_4$. Does the Ising model possess the Markov property with respect to the 8-neighborhood relation $\sim_8$?

   b) For any $\Gamma \subset T_N$ the joint distribution of the system of random variables $\{X_t, t \in \Gamma\}$ is denoted by $\mathbf{P}^{\Gamma}_{T_N}$, i.e. for $\Gamma = \{t_1, \ldots, t_n\}$ and $x_{t_1}, \ldots, x_{t_n} = \pm 1$ it holds that $P^{\Gamma}_{T_N}(x_{t_1}, \ldots, x_{t_n}) = \mathbf{P}(X_{t_1} = x_{t_1}, \ldots, X_{t_n} = x_{t_n})$. For any such Γ the expectation $\sigma_\Gamma = \mathbf{E} \prod_{t\in\Gamma} X_t$ is called the *mixed moment*.

   Prove that the probabilities $P^{\Gamma}_{T_N}$ can be expressed in terms of the mixed moments σ_Γ as follows:

$$\mathbf{P}^{\Gamma}_{T_N}(x_{t_1}, \ldots, x_{t_n}) = \frac{(-1)^k}{2^n} \sum_{\Gamma'\subset\Gamma} C_{\Gamma'}\sigma_{\Gamma'},$$

   where k is the number of values x_{t_i} equal to -1 and $C_{\Gamma'} = \prod_{t\in\Gamma\setminus\Gamma'} x_t$.

   c) Assume that $J = 0$, i.e. there is no interaction between particles. Prove that for any $\Gamma \subset T_N$ the mixed moment σ_Γ is given by

$$\sigma_\Gamma = \left(\frac{\exp\left(\frac{mB}{kt_0}\right) - \exp\left(-\frac{mB}{kt_0}\right)}{\exp\left(\frac{mB}{kt_0}\right) + \exp\left(-\frac{mB}{kt_0}\right)} \right)^{|\Gamma|}.$$

10. The process $X = \{X_t, t \in [0;\infty)\}$ is defined as follows: X_0 takes values $+1$ and -1 with probability $1/2$ each and $X_t = X_0 \cdot (-1)^{Y([0;t])}$ for $t > 0$ where Y is a homogeneous Poisson process on $[0,\infty)$ with intensity $\lambda > 0$, independent of X_0. See Figure 1.10 (left) for a trajectory of X.

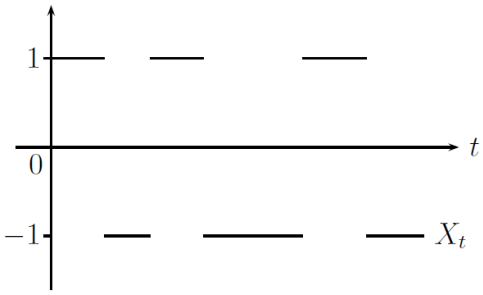
    a) Compute the expectation function of X and its covariance function.

    b) Assume that the process $\tilde{X}$, if the kth point of Y occurs, attains a random value V_k and retains that value till the next point of Y occurs, i.e. $X_0 = V_0$ and $X_t = V_{Y([0;t])}$ for $t > 0$ where $V_0, V_1, \ldots$ are iid random variables with $V_0 \sim \mathcal{N}(0, \sigma^2)$. Figure 1.10 (right) shows a trajectory of $\widetilde{X}$. Specify its marginal distribution. Is it a Gaussian process?

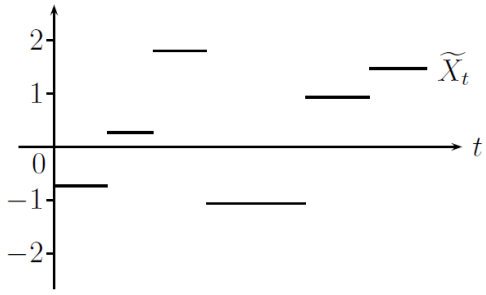
11. *One-dimensional random walk:*

    a) Let Y denote a homogeneous Poisson process on $[0;\infty)$ with intensity λ. Let $X_1, X_2, \ldots$ be i.i.d. random variables, independent of Y with $\mathbf{E}X_1 = 0$ and $\mathbf{var}X_1 = \sigma^2$. Define the random process Z on $[0;\infty)$ by $Z_t = \sum_{i=1}^{Y_t} X_i$. Compute the expectation function and the covariance function of Z.

    b) Consider the process $\widetilde{Z}_t = Y_t - \lambda t$. Compare the expectation, the covariance and the trajectories of Z and $\widetilde{Z}$.



(a) Trajectory of the process X in Exercise 10a.



(b) Trajectory of the process $\widetilde{X}$ in Exercise 10b.

Fig. 1.10: Illustration for Exercise 10a and 10b.

Hint: Wald's identity could be useful: Let N be a random variable with $\mathbf{P}(N \in \mathbb{N}) = 1$ and let $(X_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. random variables independent of N. Assume that $\mathbf{E}N$ and $\mathbf{E}X_0$ exist. Then $\mathbf{E}\sum_{n=1}^{N} X_n$ exists and is given by $\mathbf{E}\sum_{n=1}^{N} X_n = \mathbf{E}N\,\mathbf{E}X_0$. Further, if $\mathbf{var}X_1$ and $\mathbf{var}N$ exist, it holds that $\mathbf{var}\left(\sum_{n=1}^{N} X_n\right) = \mathbf{var}X_1\mathbf{E}N + (\mathbf{E}X_1)^2\,\mathbf{var}(N)$.

12. a) Compute all cumulants of a $\mathcal{N}(\mu, \sigma^2)$-distributed random variable.
    b) Prove that the cumulants $\langle X^n\rangle$ and the moments μ_n of a random variable X with $\mathbf{E}|X|^n < \infty$ are related by the following recursive formula:

$$\langle X^n\rangle = \mu_n - \sum_{i=1}^{n-1} \binom{n-1}{i-1} \langle X^i\rangle \mu_{n-i}$$

13. Give an example for a random field which is stationary in the wide sense but not strictly stationary.

14. Consider the Shot-Noise process $X_t = \sum_{v\in\Phi} g(t-v)$.
    a) Prove that X_t is stationary in the wide sense.
    b) Find conditions on the function g such that the process X_t is weakly isotropic.

15. The *cosine field* X on $\mathbb{R}^d$ can be defined by

$$X(t) = \frac{1}{\sqrt{d}} \sum_{k=1}^{d} (Y_k \cos(a_k t_k) + Z_k \sin(a_k t_k))$$

with i.i.d. centered random variables $Y_1, \ldots, Y_d, Z_1, \ldots, Z_d$ and positive constants $a_1, \ldots, a_d$. Prove that the cosine field X is weakly stationary.

16. Let $X = \{X_t, t \in \mathbb{R}^+\}$ be a process with the following properties:
    - $X_0 = 0$ almost surely
    - X has independent increments
    - there is $\sigma^2 > 0$ such that $X_t - X_s \sim \mathcal{N}(0, \sigma^2|t-s|)$, $s, t > 0$.

This process is called *Wiener process*. For $\sigma^2 = 1$ it is called *standard Wiener process*. Prove that there exists an almost surely continuous version of the Wiener process.

17. Construct a stationary shot-noise process with almost surely discontinuous realizations which is continuous in quadratic mean.

18. Give an example of a random process which is not stochastically continuous.

19. Let $T = \mathbb{N}_0$ and E be a countable phase space. The process $X = \{X_t, t \in T\}$ is called a *(discrete-time) Markov chain* if it satisfies the *Markov property*: for any $n \geq 1$, any $t, t_1, \ldots, t_n \in T$ with $t_1 < \ldots < t_n < t$, and any $i_1, \ldots, i_n, j \in E$,

$$\mathbf{P}\left(X_t = j \,|X_{t_1} = i_1, \ldots, X_{t_n} = i_n\right) = \mathbf{P}\left(X_t = j \,|X_{t_n} = i_n\right).$$

The *initial distribution* $\alpha = (\alpha_j, j \in E)$ is given by $\alpha_j = \mathbf{P}\left(X_0 = j\right)$ and the *transition probabilities* $p_{ij}\,(s,t)$ are given by $p_{ij}\,(s,t) = \mathbf{P}\left(X_t = j \,|X_s = i\right)$ for $t \geq s$, and $i, j \in E$. The chain X is called *homogeneous*, if $p_{ij}\,(s,t)$ depends on s and t only through $t - s$. Then it is enough to know α and the 1-step *transition matrix* $P = (p_{ij})$ with $p_{ij} = \mathbf{P}\left(X_{n+1} = j \,|X_n = i\right)$, $n \geq 0$. The *n-step transition matrix* $P^{(n)}$ is given by $P^{(n)} = \left(p_{ij}^{(n)}\right)$ with the *n-step transition probabilities* $p_{ij}^{(n)} = \mathbf{P}\left(X_n = j \,|X_0 = i\right)$. The *Chapman-Kolmogorov equation* states that $P^{(n+m)} = P^{(n)}P^{(m)}$ for any $n, m \in \mathbb{N}$.

a) Prove the Chapman-Kolmogorov equation.

b) Give an example for a stochastic process with $T = \mathbb{N}_0$ and countable phase space E which satisfies the Chapman-Kolmogorov equation but does not fulfill the Markov property.

2 Correlation theory of stationary random fields

In this chapter, properties of the first moments of (mostly stationary) random functions and especially random fields are considered. First, let us dwell on the (characteristic) properties and construction principles of covariance functions of random functions as positive semi-definite functions.

2.1 Positive semi-definite functions

Here we give a slightly more general definition of positive semi-definiteness than that of Exercise 1.3.1, which allows for complex-valued functions.

Definition 2.1.1 1. A function $C : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$ is called positive semi-definite, iff

- $C(x, y) = \overline{C(y, x)}$, $x, y \in \mathbb{R}^d$
- for any $n = 2, 3, \ldots$ and $x_1, \ldots, x_n \in \mathbb{R}^d$, $z_1, \ldots, z_n \in \mathbb{C}$ it holds $\sum_{i,j=1}^{n} C(x_i, x_j) z_i \overline{z}_j \geq 0$, i.e. this expression is real and non-negative. Here $\overline{z_j}$ means the complex conjugate of $z_j \in \mathbb{C}$.

2. A function $f : \mathbb{R}^d \to \mathbb{C}$ is called *positive semi-definite* if the function $C : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$ given by $C(x, y) = f(x - y)$ is positive semi-definite.

3. A function $f_o : \mathbb{R}_+ \to \mathbb{C}$ is called positive semi-definite if the function $C : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$ given by $C(x, y) = f_o(|x - y|)$, $x, y \in \mathbb{R}^d$ is positive semi-definite.

As it is shown in Exercise 1.3.1, the class of all positive semi-definite functions on $\mathbb{R}^d$ and the class of covariance functions of (complex-valued) random fields on $\mathbb{R}^d$ are identical.

The following characterization result was proved for $d = 1$ in $1932 - 1934$ independently by Bochner and Khinchin.

Theorem 2.1.1 (Bochner-Khinchin):
A continuous at the origin $o \in \mathbb{R}^d$ function $f : \mathbb{R}^d \to \mathbb{C}$ is positive semi-definite if and only if it can be represented as a characteristic function of a finite measure μ_f on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ (see Def. 1.1.7):

$$f(x) = \varphi_{\mu_f}(x) = \int_{\mathbb{R}^d} \exp(i \langle x, t \rangle) \mu_f(dt), \quad x \in \mathbb{R}^d. \tag{2.1.1}$$

For simplicity, let us prove this result in the special case ($d = 1$). For the general case, see [5]. First of all, let us mention that a positive semi-definite function is continuous iff it is continuous at the origin.

For the proof of Theorem 2.1.1, we need to introduce the so-called *Fourier transform* F: let $L^p(\mathbb{R})$ be the space of all functions $h : \mathbb{R} \to \mathbb{C}$ with $\int_{\mathbb{R}} |h(x)|^p \, dx < \infty$, $p \geq 1$. Define

$$F(h)(s) = \int_{\mathbb{R}} e^{ixs} h(x) dx, \quad h \in L^1(\mathbb{R}), \ s \in \mathbb{R}.$$

(see, e.g. [32]). If $g = F(h)$, then the *inverse Fourier transform* is defined by

$$F^{-1}(g)(t) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} g(x) dx, \quad t \in \mathbb{R}. \tag{2.1.2}$$

If additionally $h \in L^2(\mathbb{R})$, then

$$F^{-1}(F(h))(x) = h(x) \tag{2.1.3}$$

for almost all x with respect to the Lebesgue measure. Thus, $F : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ is a one-to-one mapping satisfying the following *Parseval's identity*:

$$\langle F(h), F(l) \rangle_{L^2(\mathbb{R})} = 2\pi \langle h, l \rangle_{L^2(\mathbb{R})}, \quad h, l \in L^2(\mathbb{R}), \tag{2.1.4}$$

where $\langle h, l \rangle_{L^2(\mathbb{R})} = \int_{\mathbb{R}} h(x) \bar{l}(x) dx$ is the scalar product in $L^2(\mathbb{R})$. If $h \in L^1(\mathbb{R})$, then $F(h) \in C(\mathbb{R})$ (which follows easily from Lebesgue's Theorem on dominated convergence).

Proof of Theorem 2.1.1. Necessity is trivial, see Exercise 2.6 in Section 2.6.
Sufficiency: First, we shall assume that a positive semi-definite continuous function $f : \mathbb{R} \to \mathbb{C}$ belongs to $L^1(\mathbb{R})$. Introduce $h(x) = F^{-1}(f)(x)$ and prove that $\mu(dx) = h(x)dx$ is a positive finite measure on $\mathbb{R}$ from (2.1.1). After that, we shall relax the condition $f \in L^1(\mathbb{R})$ and prove (2.1.1) for any positive semi-definite continuous f by using a smoothing argument.
If $f \in L^1(\mathbb{R})$, then $f \in L^2(\mathbb{R})$, since f is a covariance function of some centered complex-valued stationary random process $X : \Omega \times \mathbb{R} \to \mathbb{C}$, i.e. $f(t) = C_X(t) = \mathbf{E}\left[X(0)\overline{X(t)}\right]$, and then $|f(t)| \leq f(0)$, $t \in \mathbb{R}$, which follows from $|C_X(t)| \leq C_X(0) = \mathbf{var}X(0)$. Hence, $\int_{\mathbb{R}} |f(x)|^2 dx \leq f(0) \cdot \int_{\mathbb{R}} |f(x)| dx < \infty$ for $f \in L^1(\mathbb{R})$.
Take $h = F^{-1}(f)$ and prove that
a). $h(x) \geq 0$, $x \in \mathbb{R}$
b). Formula (2.1.3), which reads here $f(x) = F(h)(x) = \int_{\mathbb{R}} e^{itx} h(t) dt$ holds for *all* $x \in \mathbb{R}$, and not for almost all $x \in \mathbb{R}$.

Proof of a). Since f is a covariance function of the above random process X, it is easy to see that $e^{-itx} f(t)$ is a covariance function of $e^{-itx} X(t)$, which is again centered and stationary:

$$\begin{aligned} \mathbf{cov}\left(e^{-itx} X(t), e^{-isx} X(s)\right) &= \mathbf{E}\left[e^{-itx} X(t) \overline{e^{-isx} X(s)}\right] \\ &= e^{-i(t-s)x} f(t-s), \quad s, t \in \mathbb{R}. \end{aligned}$$

Hence, it suffices to show that for *any* continuous covariance function $C : \mathbb{R} \to \mathbb{C}$ (which is positive semi-definite), $C \in L^1(\mathbb{R})$, it holds that $\int_{\mathbb{R}} C(t) \, dt \geq 0$. If this holds true, then

$$h(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} f(t) \, dt \geq 0, \quad x \in \mathbb{R}$$

for $C(t) = e^{-itx} f(t)$, $x \in \mathbb{R}$. Since C is continuous and positive semi-definite, it holds

$$\int_{-N}^{N} \int_{-N}^{N} C(t-s) dt ds = \lim_{n \to \infty} \sum_{k=-n}^{n-1} \sum_{l=-n}^{n-1} C(t_k(n) - t_l(n)) \left(\frac{N}{n}\right)^2 \geq 0, \quad N \in \mathbb{N} \tag{2.1.5}$$

as a limit of integral sums on $[-N;N]^2$, where $\frac{N}{n}$ is the side length of the squares $[k\frac{N}{n};(k+1)\frac{N}{n}]\times[l\frac{N}{n};(l+1)\frac{N}{n}]$ building the decomposition of $[-N;N]^2$ and $t_j(n)\in[j\frac{N}{n};(j+1)\frac{N}{n}]$ are arbitrary fixed points.

Transforming the integral on the left-hand side of (2.1.5), we get by substitution $u=t-s$, $v=t+s$ that

$$\begin{aligned}0\le\frac{1}{2N}\int_{-N}^{N}\int_{-N}^{N}C(t-s)dtds&=\frac{1}{4N}\left(\int_0^{2N}C(u)\int_{-2N+u}^{2N-u}dvdu+\int_{-2N}^{0}C(u)\int_{-u-2N}^{u+2N}dvdu\right)\\&=\int_{-2N}^{2N}\left(1-\frac{|u|}{2N}\right)C(u)du=\int_{\mathbb{R}}\left(1-\frac{|u|}{2N}\right)\mathbf{1}\{u\in[-2N;2N]\}C(u)du\\&\xrightarrow[N\to\infty]{}\int_{\mathbb{R}}C(u)du\end{aligned}$$

by Lebesgue's Theorem on dominated convergence, since

$$\left(1-\frac{|u|}{2N}\right)\mathbf{1}\{u\in[-2N;2N]\}\xrightarrow[N\to\infty]{}1,$$

$$\left|\left(1-\frac{|u|}{2N}\right)\mathbf{1}\{u\in[-2N;2N]\}C(u)\right|\le|C(u)|\in L^1(\mathbb{R})\text{ for all }N\in\mathbb{N}$$

and the Jacobian $\frac{\partial(t,s)}{d(u,v)}=1/2$.

Proof of b). Show that

$$f(x)=\int_{\mathbb{R}}e^{itx}h(t)dt \tag{2.1.6}$$

holds for all $h\in L^1(\mathbb{R})$, $x\in\mathbb{R}$. It is known that $\varphi_Y(s)=e^{-\varepsilon^2s^2/2}$ for a random variable $Y\sim\mathcal{N}(0,\varepsilon^2)$, $\varepsilon>0$. Since $\varphi_Y(s)=F(f_Y)(s)$ for the density $f_Y(x)=\frac{1}{\varepsilon\sqrt{2\pi}}e^{-x^2/2\varepsilon^2}$, $x\in\mathbb{R}$, use Parseval's identity (2.1.4) for f and f_Y:

$$\begin{aligned}f(0)&\xleftarrow[\varepsilon\to0]{}\int_{\mathbb{R}}\frac{1}{\sqrt{2\pi}\varepsilon}e^{-x^2/2\varepsilon^2}f(x)dx=\langle f_Y,f\rangle_{L^2(\mathbb{R})}=2\pi\left\langle F^{-1}(f_Y),F^{-1}(f)\right\rangle_{L^2(\mathbb{R})}\\&=\int_{\mathbb{R}}e^{-(\varepsilon x)^2/2}h(x)dx\xrightarrow[\varepsilon\to0]{}\int_{\mathbb{R}}h(x)dx,\end{aligned}$$

with $h=F^{-1}(f)$, since f is continuous. By the dominated convergence theorem we get $\int_{\mathbb{R}}h(x)dx=f(0)$, i.e. $h\in L^1(\mathbb{R})$ and formula (2.1.6) holds for $x=0$. Then $F(h)\in C(\mathbb{R})$. Since both left- and right-hand sides of (2.1.6) are continuous functions of x on $\mathbb{R}$, formula (2.1.6) holds for *any* $x\in\mathbb{R}$.

Now suppose that $f\notin L^1(\mathbb{R})$. Take $f_\varepsilon(t)=f(t)e^{-(\varepsilon t)^2/2}$, $\varepsilon>0$. Then, f_ε is continuous positive semi-definite, since $e^{-(\varepsilon s)^2/2}=F(f_Y)$ for $Y\sim\mathcal{N}(0,\varepsilon^2)$, and then the product of f and $F(f_Y)$ should inherit the same properties (see Theorem 2.1.5). Moreover, $f_\varepsilon\in L^1(\mathbb{R})$. Then, $f_\varepsilon(t)/f_\varepsilon(0)=f_\varepsilon(t)/f(0)=\varphi_Z(t)=F(h_\varepsilon)$ is a characteristic function of some random vector Z with density $h_\varepsilon\ge0$, i.e. $f_\varepsilon(t)/f(0)=\int_{\mathbb{R}}e^{itx}h_\varepsilon(x)dx$, by the first part of our proof. It holds $f_\varepsilon(t)/f(0)\to f(t)/f(0)$, $t\in\mathbb{R}$, as $\varepsilon\to0$, where $\{f_\varepsilon(t)/f(0)\}_{\varepsilon>0}$ is a sequence of characteristic

functions and $f(t)/f(0)$ is continuous at zero. Then, by a very well known fact $f(t)/f(0)$ is a characteristic function of some probability distribution Ψ:

$$f(t)/f(0) = \int_{\mathbb{R}} e^{itx}\Psi(dx),$$

hence, $\mu_f(dx) = f(0)\cdot\Psi(dx)$ is the needed finite measure in (2.1.1). □

Definition 2.1.2
The measure μ_f from representation (2.1.1) is called a *spectral measure* of f. If μ_f has a density h with respect to the Lebesgue measure on $\mathbb{R}^d$, i.e. $\mu_f(dx) = h(x)dx$, then h is called *spectral density* of f.

Remark 2.1.1
If μ_f is a symmetric measure, i.e. $\mu_f(A) = \mu_f(-A)$, $A \in \mathcal{B}(\mathbb{R}^d)$, where $-A = \{-x : x \in A\}$, then $f(x)$ is real, since

$$\begin{aligned} f(x) &= \int_{\mathbb{R}^d} e^{i\langle t,x\rangle}\mu_f(dx) = \int_{\mathbb{R}^{d-1}\times\mathbb{R}_+} e^{i\langle t,x\rangle}\mu_f(dx) + \int_{\mathbb{R}^{d-1}\times\mathbb{R}_-} e^{i\langle t,x\rangle}\mu_f(dx) \\ &= \int_{\mathbb{R}^{d-1}\times\mathbb{R}_+} \cos\left(\langle t,x\rangle\right)\mu_f(dx) + i\int_{\mathbb{R}^{d-1}\times\mathbb{R}_+} \sin\left(\langle t,x\rangle\right)\mu_f(dx) \\ &\quad + \int_{\mathbb{R}^{d-1}\times\mathbb{R}_+} \cos\left(\langle t,x\rangle\right)\mu_f(dx) - i\int_{\mathbb{R}^{d-1}\times\mathbb{R}_+} \sin\left(\langle t,x\rangle\right)\mu_f(dx) \\ &= 2\int_{\mathbb{R}^{d-1}\times\mathbb{R}_+} \cos\left(\langle t,x\rangle\right)\mu_f(dx) \end{aligned}$$

by substitution $x \mapsto -x$ in the integral $\int_{\mathbb{R}^{d-1}\times\mathbb{R}_-}$

The next theorem due to Riesz (1932-1934) will characterize all *measurable* positive semi-definite functions.

Theorem 2.1.2
If f is positive semi-definite and measurable, then $f = f_c + f_0$, where f_c is positive semi-definite and continuous (and hence Bochner's theorem can be applied) and f_0 is positive semi-definite and equals zero almost everywhere on $\mathbb{R}^d$ with respect to the Lebesgue measure.

Proof See [47]. □

It follows from Theorem 2.1.2 that any function which cannot be turned into a continuous one by changing it in a set of points of Lebesgue measure zero cannot be positive semi-definite. The most popular function f_0 in applications is the so-called *nugget effect*:

$$f_0(s) = \begin{cases} a, & s = 0, \\ 0, & s \neq 0 \end{cases}$$

for $a > 0$.

Exercise 2.1.1
Show that the function $f_0(x) = a\mathbf{1}_{\mathbb{Z}^d}(x)$ is positive semi-definite. Construct the corresponding stationary random field.

2.1.1 Isotropic case

In the course of the next sections, we follow [49] and [48]. Consider covariance functions of stationary isotropic random fields in $\mathbb{R}^d$: they have the form $C(x,y) = C_0(|x-y|)$, where $|\cdot|$ is the Euclidean norm in $\mathbb{R}^d$.
It is worth noting that the function $C_0 : \mathbb{R}_+ \to \mathbb{R}$ has the following property:
$|C_0(h)| \leq C_0(0) > 0$, $h \in \mathbb{R}_+$, which follows directly from the definition of the positive semi-definiteness of C_0.
Let

$$J_\nu(r) = \sum_{j=0}^{\infty} \frac{(-1)^j}{j!\Gamma(\nu+j+1)} \left(\frac{r}{2}\right)^{\nu+2j}, \quad r \in \mathbb{R}$$

be the Bessel function of the first kind of order $\nu \geq -1/2$.

Example 2.1.1
Since $J_{-1/2}(r) = \sqrt{\frac{2}{\pi r}}\cos(r)$, $J_0(r) = \frac{2}{\pi}\int_0^\infty \sin(r\cosh(t))\,dt$, $J_{1/2}(r) = \sqrt{\frac{2}{\pi r}}\sin(r)$, $r \in \mathbb{R}$, we get e.g. in three dimensions ($d = 3$):

$$C(x,y) = C_0(|x-y|) = \int_0^\infty \frac{\sin(sr)}{sr}\mu(dr)\Big|_{s=|x-y|}$$

from (2.1.7), because for $\nu = \frac{d-2}{2} = \frac{1}{2}$

$$\begin{aligned} C_0(s) &= \int_0^\infty \sqrt{\frac{1}{sr}}\sqrt{\frac{2}{\pi sr}}\sin(sr)\mu(dr) \\ &= \frac{\sqrt{2}}{\sqrt{\pi}}\int_0^\infty \frac{\sin(sr)}{sr}\mu(dr), \end{aligned}$$

where the constant factor $\frac{\sqrt{2}}{\sqrt{\pi}}$ can be included into the measure μ.

The following characterization result can be obtained by passing to polar coordinates in Bochner's theorem:

Theorem 2.1.3 (Schoenberg, 1938):
A continuous rotation-invariant function $f : \mathbb{R}^d \to \mathbb{R}$ is positive semi-definite iff $f(x) = f_0(|x|)$, $x \in \mathbb{R}^d$ with $f_0 : \mathbb{R}_+ \to \mathbb{R}$ given by:

$$f_0(s) = \int_0^\infty \left(\frac{1}{rs}\right)^\nu J_\nu(rs)\mu(dr), \tag{2.1.7}$$

where μ is a finite measure and $\nu = \frac{(d-2)}{2}$.

Proof See [50] or [1], p. 116-117. □

For isotropic measurable positive semi-definite functions, a stronger version of Riesz theorem is available:

Theorem 2.1.4
If f is a rotation-invariant, measurable positive semi-definite function $f : \mathbb{R}^d \to \mathbb{C}$, $d \geq 2$, then

$$f(x) = f_c(x) + a \cdot \mathbf{1}\{x = 0\},$$

where $a \geq 0$ and f_c is positive semi-definite and continuous.

Proof See [21]. □

It follows from Theorem 2.1.4 that measurable isotropic covariance functions are discontinuous at most in zero.

2.1.2 Construction principles of positive semi-definite functions

Theorem 2.1.5
Let $C_i : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$, $i = 0, 1, 2, \ldots$, $C_\nu : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$, $\nu \in \mathbf{E}$ be positive semi-definite functions. A function $C : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$ is positive semi-definite iff

1. *Scaling*: $C(x, y) = C_0(ax, ay)$, $x, y \in \mathbb{R}^d$, $a \in \mathbb{R}$
2. *Linear combination*: $C(x, y) = \sum_{i=1}^n a_i C_i(x, y)$, $x, y \in \mathbb{R}^d$
3. *Multiplication*: $C(x, y) = C_1(x, y) \cdot C_2(x, y)$, $x, y \in \mathbb{R}^d$
4. *Substitution*: $C(x, y) = C_0(g(x), g(y))$, where $C_0 : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$ is positive semi-definite and $g : \mathbb{R}^d \to \mathbb{R}^d$ is an arbitrary mapping.
5. *Kernel approach*: $C(x, y) = \langle g(x), g(y) \rangle_{\mathbb{L}}$, where $\mathbb{L}$ is a Hilbert space over $\mathbb{C}$ with scalar product $\langle \cdot, \cdot \rangle_{\mathbb{L}}$ and $g : \mathbb{R}^d \to \mathbb{L}$ is an arbitrary mapping
6. *Pointwise limit*: $C(x, y) = \lim_{n \to \infty} C_n(x, y)$, $x, y \in \mathbb{R}^d$
7. *Integration*: $C(x, y) = \int_E C_\nu(x, y) \mu(d\nu)$, where μ is a finite measure over some space E and $C_\nu(x, y)$ are positive semi-definite functions. In particular,
   Series expansions: $C(x, y) = \sum_{i=0}^\infty a_i C_i(x, y)$ for a counting measure μ
8. *Scale mixture*: $C(x, y) = \int_0^\infty C_0(sx, sy) \mu(ds)$ for some finite measure μ on $[0; \infty)$ and a positive semi-definite function C_0
9. *Convolution*: $C(x, y) = f(x - y)$, $x, y \in \mathbb{R}^d$, where $f(h) = \int_{\mathbb{R}^d} f_1(h - s) \overline{f}_2(s) ds$, $h \in \mathbb{R}^d$ and $f_1, f_2 : \mathbb{R}^d \to \mathbb{C}$ are positive semi-definite and continuous.

Proof 1. This follows directly from Definition 2.1.1. See also Exercise 2.6.
2., 3. To see this, stationary random fields have to be constructed with covariances $\sum_{i=1}^n a_i C_i(x, y)$, $C_1(x, y) \cdot C_2(x, y)$.
4. Use Definition 2.1.1 directly: for any $n \in \mathbb{N}$, $z_i \in \mathbb{C}$, $i = 1, \ldots, n$ and $x_i \in \mathbb{R}^d$, $i = 1, \ldots, n$ one has

$$\sum_{i,j=1}^n C(x_i, x_j) z_i \overline{z_j} = \sum_{i,j=1}^n C_0(g(x_i), g(x_j)) z_i \overline{z_j} \geq 0,$$

since C_0 is positive definite.
5. Use Definition 2.1.1: for any $n \in \mathbb{N}$, $z_i \in \mathbb{C}$, $x_i \in \mathbb{R}^d$, $i = 1, \ldots, n$ one has

$$\begin{aligned} \sum_{i,j=1}^n C(x_i, x_j) z_i \overline{z_j} &= \sum_{i,j=1}^n \langle g(x_i), g(x_j) \rangle_{\mathbb{L}} \, z_i \overline{z_j} \\ &= \left\langle \sum_{i=1}^n g(x_i) z_i, \sum_{j=1}^n g(x_j) z_j \right\rangle_{\mathbb{L}} = \left\| \sum_{i=1}^n g(x_i) z_i \right\|_{\mathbb{L}}^2 \geq 0. \end{aligned}$$

6. This follows directly from Definition 2.1.1.
7. This follows from 6), since $\int_E C_\nu(x,y)\mu(d\nu)$ is a limit of its integral sums, each of which is a linear combination of positive semi-definite functions $C_{\nu_i}(x,y)$, $i=1,\ldots,n$, which is positive semi-definite by 2). The same reasoning works for series expansions.
8. For any $s \in [0;\infty)$, $C_0(sx,sy)$ is positive semi-definite by 1). Then use 7).
9. See [47]. □

Examples and Remarks

1. In Theorem 2.1.5, 5) take $\mathbb{L} = L^2_\mu\left(\mathbb{R}^d\right)$, $g : x \in \mathbb{R}^d \longmapsto g(x) = e^{i\langle t,x\rangle}$, $t \in \mathbb{R}^d$, where μ is a finite measure on $\mathbb{R}^d$. Then

$$C(x,y) = \left\langle e^{i\langle t,x\rangle}, e^{i\langle t,y\rangle}\right\rangle_{\mathbb{L}} = \int_{\mathbb{R}^d} e^{i\langle t,x\rangle} e^{-i\langle t,y\rangle}\mu(dt)$$
$$= \int_{\mathbb{R}^d} e^{i\langle t,x-y\rangle}\mu(dt) = \varphi_\mu(x-y),$$

   which is positive semi-definite by Bochner's theorem.

2. By means of property 2) and 3), covariance functions for *space-time random fields* $X = \{X(x,t), x \in \mathbb{R}^d, t \in \mathbb{R}\}$ can be constructed by $C((x,s),(y,t)) = C_x(x,y) + C_t(s,t)$, $x,y \in \mathbb{R}^d$, $s,t \in \mathbb{R}$, where C_x is a space covariance and C_t is a time covariance component. The same holds for $C((x,s),(y,t)) = C_x(x,y) \cdot C_t(s,t)$, $x,y \in \mathbb{R}^d$, $s,t \in \mathbb{R}$ (the so-called *separable space-time models*).

3. By means of property 7) functions like $e^{C(x,y)}$, $\cosh C(x,y)$ etc. of a positive semi-definite function C are also positive semi-definite as they can be represented as Taylor series with positive coefficients, e.g. $e^{C(x,y)} = \sum_{n=0}^{\infty} \frac{(C(x,y))^n}{n!}$, $x,y \in \mathbb{R}^d$.

2.1.3 Sufficient conditions for positive semi-definiteness

In Bochner's theorem, necessary and sufficient conditions are not so easy to verify. Here we give a series of sufficient conditions for positive semi-definiteness that are relatively easy to check in many concrete cases.

Theorem 2.1.6 (Pólya-Askey):
If a function $f : \mathbb{R}_+ \to \mathbb{R}$ is continuous with $\lim_{t\to\infty} f(t) = 0$, $f(0) = 1$, $(-1)^k f^{(k)}(t)$ convex for $k = [d/2]$ ([a] is the integral part of a), then $f(\|x\|)$ is positive semi-definite in $\mathbb{R}^d$.

Proof See [2]. □

A sharper criterion gives the following result:

Theorem 2.1.7 (Gneiting, 1999):
Let $f : \mathbb{R}_+ \to \mathbb{R}$ be continuous with $f(0) = 1$, $\lim_{t\to\infty} f(t) = 0$, $(-1)^{k+1}\frac{d^k}{dt^k}\left(f'\left(\sqrt{t}\right)\right)$ convex for $k = \left[\frac{d-2}{2}\right]$, then $f(\|x\|)$ is positive semi-definite in $\mathbb{R}^d$, $d \geq 2$.

Proof See [20]. □

2.1.4 Examples

In this section some important classes of positive semi-definite functions are introduced. They are all rotation invariant and hence can serve as covariance functions of stationary and isotropic random fields in $\mathbb{R}^d$, where d is indicated in the examples. As the lemma below shows, these models are valid for dimensions $\leq d$, but the converse statement is not always true (see example 1.5).

Lemma 2.1.1
If $\varphi : \mathbb{R}_+ \to \mathbb{R}$ is such that $g_d(t) = \varphi(|t|_d)$, $t \in \mathbb{R}^d$, $d \geq 2$ is positive semi-definite, then $g_{d-1}(t) = \varphi(|t|_{d-1})$, $t \in \mathbb{R}^{d-1}$ is positive semi-definite as well; here $|.|_d$ is the Euclidean norm in $\mathbb{R}^d$.

Proof Let $X = \{X(t) : t \in \mathbb{R}^d\}$ be a random field with covariance function g_d. Then for any hyperplane $\xi \subset \mathbb{R}^d$ the random field $X = \{X(t) : t \in \xi\}$ is a random field over ξ with the same covariance function g_d restricted on ξ. Taking $\xi = \{(x_1, \ldots, x_{d-1}, 0) : x_i \in \mathbb{R}, i = 1, \ldots, d-1)\}$ finishes the proof. □

The next result of Schoenberg shows when the converse statement holds:

Theorem 2.1.8 (Schoenberg):
Let $\varphi : \mathbb{R}_+ \to \mathbb{R}$ be a continuous function. Then $\varphi(|.|_d)$ is positive semi-definite for all d iff $\varphi(h) = \int_0^\infty e^{-rh^2}\mu(dr)$ for some finite measure μ on $[0, \infty)$, i.e., φ is scale mixture of the normal density.

Proof See [50] □

Remark 2.1.2

1. Theorem 2.1.8 implies, in particular, that $\varphi(h) \geq 0$, $h \in \mathbb{R}_+$. It follows from this property that for any function φ which is not non-negative $\varphi(||x||)$ can not be positive semi-definite for *all dimensions* $d \in \mathbb{N}$.

2. It is known that Laplace-Stieltjes transforms of finite measures $\varphi(h) = \int_0^\infty e^{-hr}\mu(dr)$, $h \in \mathbb{R}_+$ are completely monotone, i.e., $(-1)^n\varphi^{(n)}(h) \geq 0$ for all $h > 0$ and $n \in \mathbb{N}_0 = \{0, 1, 2, 3 \ldots\}$ (see e.g. [56]). Hence, under assumptions of Theorem 2.1.8, $\varphi(h)$ is a Laplace-Stieltjes transform of a measure μ evaluated at h^2, which means that $\varphi(\sqrt{h})$ should be completely monotone.

In the following examples only the function $\varphi : \mathbb{R}_+ \to \mathbb{R}$ of $\varphi(|.|_d)$ is given.

1. **Bessel family:**

$$\varphi(h) = b(ah)^{-\nu}J_\nu(ah),\ \nu \geq \tfrac{d-2}{2},\ a, b > 0$$

here J_ν is the Bessel function of the 1st kind of order ν. For $\nu = \frac{d-2}{2}$, this function (up to scaling factors, cf. Theorem 2.1.4, 1), 2)) is clearly positive semi-definite as $\varphi(h) = 2^\nu\Gamma(\nu+1)h^{-\nu}J_\nu(h)$ is a special case of Theorem 2.1.3, where $\mu(.) = \delta_{\{1\}}(.)$. For $\nu > \frac{d-2}{2}$ this statement is less trivial (cf. [58], p. 367). The spectral density of $\varphi(h)$ is given by

$$f(h) = \frac{b(a^2 - h^2)^{\nu - \frac{d}{2}}}{2^\nu \pi^{\frac{d}{2}} a^{2\nu}\Gamma(\nu + 1 - \frac{d}{2})} I(h \in [0, a])$$

A special case (for $\nu = \frac{1}{2}$, i.e. $d = 3$) yields the so-called *hole effect model* $\varphi(h) = b\frac{\sin(ah)}{ah}$. This model is valid only in $\mathbb{R}^d$, $d \leq 3$; whereas models of Bessel family are valid in $\mathbb{R}^d$ (iff $\nu = \frac{d-2}{2}$).

2. **Cauchy family:**

$$\varphi(h) = \frac{b}{(1+(ah)^2)^\nu}, \quad a, b, \nu > 0.$$

Up to scaling, this function is positive semi-definite by Theorem 2.1.8 with $\mu(dr) = cr^{\nu-1}e^{-r}dr$ for some constant c.

3. **Stable family:**

$$\varphi(h) = be^{-(ah)^\nu}, \quad \nu \in (0, 2].$$

This function is positive semi-definite since it is equal to the characteristic function of a symmetric ν-stable random variable cf. e. g. [46], p. 20. It is valid in $\mathbb{R}^d$ for all $d \geq 1$. Moreover it can be shown that for $\nu > 2$ φ is not positive semi-definite in any dimension d (cf. [49], p.19).
A special case ($\nu = 2$) of the stable family is a Gaussian model: $\varphi(h) = be^{-ah^2}$. Its spectral density is equal to $f(h) = \frac{b\sqrt{a}}{2}e^{-\frac{ah^2}{4}}$.

4. **Whittle-Matérn family:**

$$\varphi(h) = b(ah)^\nu K_\nu(ah), \quad \nu, a, b > 0$$

It is valid in any dimension d. Here K_ν is a *modified Bessel function of third kind*:

$$K_\nu(z) = \frac{\pi}{2\sin(\pi\nu)}(e^{i\frac{\pi}{2}\nu}J_{-\nu}(ze^{i\frac{\pi}{2}}) - e^{-i\frac{\pi}{2}\nu}J_\nu(ze^{-i\frac{\pi}{2}})), \quad z \in \mathbb{R}, \nu \notin \mathbb{N}.$$

For $\nu = n \in \mathbb{N}$ it should be understood in the sense of a limit as $\nu \to n$ (see [38], p. 69).

The spectral density of φ is given by

$$f(h) = b\frac{2\Gamma(\nu+\frac{d}{2})}{\Gamma(\frac{d}{2})\Gamma(\nu)}\frac{(ah)^{d-1}}{(1+(ah)^2)^{\nu+\frac{d}{2}}}I(h > 0).$$

If $\nu = \frac{2d+1}{2}$ then $\varphi \in C^{2d}(\mathbb{R}_+)$ and hence a random field with covariance function φ is d times differentiable in mean-square sense.
If $\nu = \frac{1}{2}$ then the *exponential model* $\varphi(h) = be^{-ah}$ is an important special case. The same exponential covariance belongs to the stable family for $\nu = 1$.

5. **Spherical model** is valid in $\mathbb{R}^d$, $d \leq 3$.
It is given by $\varphi(h) = b(1 - \frac{3}{2}\frac{h}{a} + \frac{1}{2}\frac{h^3}{a^3})\ I(h \in [0, a])$, $a, b > 0$, which is (for $d = 3$) proportional to the volume of $B_{\frac{a}{2}}(0) \cap B_{\frac{a}{2}}(x_0)$, where $x_o \in \mathbb{R}^3$, $|x_0| = h$. This is exactly the way how it can be generalized to higher dimensions:

$$\varphi(x) = V_d(B_{\frac{a}{2}}(0) \cap B_{\frac{a}{2}}(x)), \quad x \in \mathbb{R}^d.$$

The advantage of spherical models is that they have a compact support.
Some more covariance models are given in exercises: see Section 2.3.

2.2 Variograms

In this section we consider variograms of (intrinsically) stationary real-valued random fields of order two: they can be written as $\gamma(h) = \frac{1}{2}\mathbf{E}(X(t) - X(t+h))^2$ for any $t \in \mathbb{R}^d$.

Definition 2.2.1
A function $f : \mathbb{R}^d \to \mathbb{R}$ is *conditionally negative definite*, if it holds $\sum_{i,j=1}^n z_i \overline{z_j} f(x_i - x_j) \le 0$ for any $n \in \mathbb{N}$, points $\{x_i\}_{i=1}^n \subset \mathbb{R}^d$ and weights $\{z_i\}_{i=1}^n \subset \mathbb{C}$ such that $\sum_{i=1}^n z_i = 0$.

Proposition 2.2.1 (Properties of variograms):
Let γ be a variogram of an (intrinsically) stationary random field X. Then

1. $\gamma(0) = 0, \quad \gamma(h) \ge 0, \quad h \in \mathbb{R}^d$
2. $\gamma(h) = \gamma(-h), \quad h \in \mathbb{R}^d$ (γ is an even function)
3. If X is wide sense stationary then

$$\gamma(h) = C(0) - C(h), \quad h \in \mathbb{R}^d \tag{2.2.1}$$

   where $C(.)$ is the covariance function of X and thus $\gamma(h)$ is bounded: $|\gamma(h)| < 2C(0)$, $h \in \mathbb{R}^d$.
4. γ is conditionally negative definite.
5. If X is wide sense stationary and $C(\infty) := \lim_{|h|\to\infty} C(h) = 0$, then there exists $\gamma(\infty) = \lim_{|h|\to\infty} \gamma(h) = C(0)$ and $C(h) = \gamma(\infty) - \gamma(h)$, $h \in \mathbb{R}^d$.
6. A continuous function $\gamma : \mathbb{R}^d \to \mathbb{R}$ with $\gamma(0) = 0$ is a variogram if $e^{-\lambda\gamma(.)}$ is a positive semi-definite function for all $\lambda > 0$.
7. $\gamma(h) \le c|h|^2$, if X is mean square continuous, $\lim_{|h|\to\infty} \frac{\gamma(h)}{|h|^2} = 0$, if X is mean square differentiable.

Proof Apart from obvious properties 1), 2), 3) (3) follows from 1.3.3 on page 16), let us show: 4) Take any $n \in \mathbb{N}, \{t_i\}_{i=1}^n \in \mathbb{R}^d, \{z_i\}_{i=1}^n \subset \mathbb{C} : \sum_{i=1}^n z_i = 0$. Let us show that

$$0 \le Var\left(\sum_{i=1}^n z_i X(t_i)\right) = -\sum_{i,j=1}^n z_i \overline{z_j} \gamma(t_i - t_j)$$

Then γ will be conditionally negative definite. First $Var\left(\sum_{i=1}^n z_i X(t_i)\right)$ exists since

$$\sum_{i=1}^n z_i X(t_i) = \sum_{i=1}^n z_i (X(t_i) - X(t_0))$$

as $\sum_{i=1}^n z_i = 0$ for any $t_0 \in \mathbb{R}^d$ and the variance of increments exists. Then

$$Var\left(\sum_{i=1}^n z_i X(t_i)\right) = Var\left(\sum_{i=1}^n (X(t_i) - X(t_0))\right) \stackrel{1)}{=} \mathbf{E}\left|\sum_{i=1}^n z_i (X(t_i) - X(t_0))\right|^2$$
$$\stackrel{2)}{=} \sum_{i,j=1}^n z_i \overline{z_j}(\gamma(t_i - t_0) + \gamma(t_j - t_0) - \gamma(t_i - t_j)) = \sum_{i=1}^n z_i \gamma(t_i - t_0) \underbrace{\overline{\left(\sum_{j=1}^n z_j\right)}}_{0}$$
$$+ \underbrace{\sum_{i=1}^n z_i}_{0} \sum_{j=1}^n \overline{z_j}\gamma(t_j - t_0) - \sum_{i,j=1}^n z_i \overline{z_j}\gamma(t_i - t_j) = -\sum_{i,j=1}^n z_i \overline{z_j}\gamma(t_i - t_j).$$

[1)] since $\mathbf{E}[X(t_i) - X(t_0)] = 0, \quad \forall i$
[2)]since $\gamma(t_i - t_j) = \frac{1}{2}\mathbf{E}(X(t_i) - X(t_0) - X(t_j) + X(t_0))^2 = \frac{1}{2}(2\gamma(t_i - t_0) + 2\gamma(t_j - t_0) - 2Cov(X(t_i) - X(t_0), X(t_j) - X(t_0))) = \gamma(t_i - t_0) + \gamma(t_j - t_0) - Cov(X(t_i) - X(t_0), X(t_j) - X(t_0))$ and hence $Cov(X(t_i) - X(t_0), X(t_j) - X(t_0)) = \gamma(t_i - t_0) + \gamma(t_j - t_0) - \gamma(t_i - t_j)$.

5) Follows easily from 3). Exercise!
6) This characterization result is due to Schoenberg (1938): See the proof there [50].
7) See the proof in [58], p. 397-400.
Let us show here only the relation $\gamma(h) \le C \cdot |h|^2$.
If X is mean square continuous, then $\gamma(h)$ is a continuous function on $\mathbb{R}^d$. By Minkowski inequality, it holds

$$\sqrt{\gamma(h_1 + h_2)} \le \sqrt{\gamma(h_1)} + \sqrt{\gamma(h_2)}, \quad h_1, h_2 \in \mathbb{R}^d,$$

since

$$\sqrt{2\gamma(h_1 + h_2)}$$
$$= \sqrt{\mathbf{E}[X(t + h_1 + h_2) - X(t + h_1) + X(t + h_1) - X(t)]^2}$$
$$\le \sqrt{\mathbf{E}[X(t + h_1 + h_2) - X(t + h_1)]^2}$$
$$+ \sqrt{\mathbf{E}[X(t + h_1) - X(t)]^2} = \sqrt{2\gamma(h_1)} + \sqrt{2\gamma(h_2)}.$$

Then we can easily show by induction that $\gamma(nh) \le n^2\gamma(h)$ for any $n \in \mathbb{N}$ and $h \in \mathbb{R}^d$, or, in other words,

$$\frac{\gamma(h)}{|h|^2} \le \frac{\gamma(\frac{h}{n})}{|\frac{h}{n}|^2} \tag{2.2.2}$$

(replacing h by $\frac{h}{n}$).

Fix any $0 \ne h_0 \in \mathbb{R}^d$. Let $c = \max_{h:|h_0| \le |h| \le 2|h_0|} \left(\frac{\gamma(h)}{|h|^2}\right)$, which is finite, since $\frac{\gamma(h)}{|h^2|}$ is continuous. For any $h \in \mathbb{R}^d : |h| > |h_0|$ there exists such $n \in \mathbb{N} : \quad n|h_0| \le |h| \le 2n|h_0|$, or $|h_0| \le \left|\frac{h}{n}\right| \le 2|h_0|$. Hence, $\frac{\gamma(h)}{|h|^2} \le c$ from (2.2.2), i.e., $\gamma(h) \le c|h|^2$ for $h : |h| > |h_0|$. □

Remark 2.2.1 1. It follows from Proposition 2.2.1, 3), 5) that if the variogram γ of an intrinsically stationary random field X is unbounded, $\gamma(h) \to \infty$ as $|h| \to \infty$, then X can not be wide sense stationary.
See the corresponding examples which follow.

2. Since X is real-valued, Bochner's theorem together with formula (2.2.1) yields

$$\gamma(h) = C(0) - C(h) = \int_{\mathbb{R}^d} (1 - e^{i\langle h,x\rangle})\mu(dx) = Re\left(\int_{\mathbb{R}^d} (1 - e^{i\langle h,x\rangle})\mu(dx)\right)$$
$$= \int_{\mathbb{R}^d} (1 - \cos\langle h,x\rangle)\mu(dx), \quad h \in \mathbb{R}^d.$$

Example 2.2.1 1. Consider the case $d \equiv 1$, $X(t) = X_0 + tX_1$, where $t \in \mathbb{R}$, X_0 and X_1 are random variables, $\mathbf{E}X_1 = 0$, $\mathbf{E}X_1^2 = \sigma_1^2 \in (0,\infty)$. The process $X = \{X(t), t \in \mathbb{R}\}$ is intrinsic stationary of order two since $X(t) - X(s) = (t-s)X_1$, $s,t \in \mathbb{R}$ and hence $\mathbf{E}(X(t) - X(s)) = 0$, $\mathbf{E}(X(t) - X(s))^2 = (t-s)^2\sigma_1^2$, $t,s \in \mathbb{R}$.
Then $\gamma(h) = \frac{1}{2}\mathbf{E}(X(t+h) - X(t))^2 = \frac{\sigma_1^2}{2}h^2 \to \infty$, as $h \to \infty$. X is obviously not wide sense stationary, since X_0 may not be integrable. But even for X_0 such that $\mathbf{E}X_0 = 0$, $\mathbf{E}X_0^2 = \sigma_0^2 \in (0,\infty)$ and X_0, X_1 are independent, it holds $Cov(X(s), X(t)) = \mathbf{E}[X(s)X(t)] = \mathbf{E}[(X_0 + sX_1)(X_0 + tX_1)] = \mathbf{E}X_0^2 + (s+t)\mathbf{E}[X_0X_1] + st\mathbf{E}X_1^2 = \sigma_0^2 + st\sigma_1^2$, which contradicts the stationarity in wide sense.

2. Let $X = \{X(t), t \in \mathbb{R}^d\}$ with $X(t) = \langle \varphi, t\rangle$, $t \in \mathbb{R}^d$, where $\varphi \sim N(0, I_d)$ is a multivariate standard distributed random vector. Then X is intrinsic stationary of order two, but not wide sense stationary, since $X(t) \sim N(0, |t|^2)$, $X(t) - X(s) = X(t-s)$, $t,s \in \mathbb{R}^d$ and hence

$$C(s,t) = Cov(X(s), X(t)) = \mathbf{E}\left[\sum_{i,j=1}^{d} \varphi_i\varphi_j s_i t_j\right] = \sum_{i,j=1}^{d} \mathbf{E}[\varphi_i\varphi_j]s_it_j$$
$$= \sum_{i=1}^{d} s_it_i = \langle s,t\rangle, \quad s,t \in \mathbb{R}^d$$

with φ_i-iid$\sim N(0,1)$; $\gamma(h) = \frac{1}{2}\mathbf{E}[X(t+h) - X(t)]^2 = \frac{1}{2}\mathbf{E}[X(h)]^2 = \frac{|h|^2}{2}$, $h \in \mathbb{R}^d$.
Here we have $\gamma(h) \to \infty$ as $|h| \to \infty$ as well.

This example can be extended to the case of a field $X = \{X(t), t \in \mathbf{E}\}$ over a Hilbert space $\mathbf{E}$ with a scalar product $\langle .,. \rangle$ and orthonormal basis $\{e_i\}_{i=1}^{\infty}$:

$$X(t) = \langle \sum_{i=1}^{\infty} N_i e_i, t\rangle = \sum_{i=1}^{\infty} N_i t_i,$$

$t \in \mathbf{E}$ where $\{N_i\}_{i=1}^{\infty}$ is an iid sequence of $N(0,1)$-random variables, and
$\sum_{i=1}^{\infty} N_i t_i$ converges in the mean square sense $\mathbf{E}|\sum_{i=n}^{m} N_i t_i|^2 = \sum_{i=n}^{m} t_i^2 \to 0$ as $n,m \to \infty$, since $\sum_{i=1}^{\infty} t_i^2 < \infty$ as $t \in \mathbf{E}$.

3. *Fractional Brownian field*
Let $X = \{X(t), t \in \mathbb{R}^d\}$ be a centered Gaussian Random field with covariance function

$$C(s,t) = \frac{1}{2}(|s|^{2H} + |t|^{2H} - |s-t|^{2H}), \quad s,t \in \mathbb{R}^d, \quad H \in (0,1].$$

It is called a fractional Brownian field with Hurst index H. For $H = \frac{1}{2}$, we get the usual Brownian field. It is easy to see that X is intrinsic stationary of order two, since in this case $\gamma(h) = \frac{1}{2}|h|^{2H}$, $h \in \mathbb{R}^d$. Hurst index H is also a parameter related to the smoothness of the paths of X: the smaller H, the more rough are its realisations. In particular, for $d = 1$ and $H = 1$ X is a random line:

$$X(t) = X_1 t, \quad X_1 \sim N(0,1),$$

cf. Examples 1), 2).

Exercise 1 Show that the increments of the fractional Brownian field X

a) are stationary

b) dependent (except for the case $d = 1$, $H = \frac{1}{2}$ of the Wiener process.): for $1 > H > \frac{1}{2}$ they are positively correlated, for $O < H < \frac{1}{2}$ they are negatively correlated.

Exercise 2 Show that X is H-selfsimilar, i.e., $X(\lambda t) \stackrel{d}{=} |\lambda|^H X(t)$ for all $\lambda \in \mathbb{R}$ and $t \in \mathbb{R}$, hence, $X(s) = 0$ a.s.

Exercise 3 Show that X is long-range dependent for $\frac{1}{2} < H < 1$:

$$\sum_{n=1}^{\infty} \underbrace{\mathbf{E}\,(X(1)[X(n+1) - X(n)])}_{\text{the covariance on increments at point } n:} = +\infty$$

$Cov((X(1) - X(0), X(n+1) - X(n))$, since they are stationary.

Exercise 4 Show that there exists a version of X with a.s. Hölder-continuous paths of any order $\gamma < H$, $H \in (0,1)$.

Exercise 5 Show that paths of X are nowhere differentiable for $0 < H < 1$.

Let us prove 5): by stationarity and self-similarity,

$$\frac{X(t) - X(s)}{t-s} \stackrel{d}{=} \frac{X(t-s)}{t-s} \stackrel{d}{=} \frac{|t-s|^H X(1)}{t-s} \stackrel{d}{=}_{t>s} (t-s)^{H-1} X(1) \xrightarrow[t \downarrow s]{\text{a.s.}} \infty$$

since $H \in (0,1)$.

2.3 Stochastic integration

In this section we introduce stochastic integrals of non-random functions with respect to an independently scattered random measure on an abstract space $\mathbf{E}$. We shall need these integrals to prove spectral representations of (wide sense) stationary random fields in Section 2.4 (Theorems of Karhunen and Cramér).

2.3.1 Independently scattered random measures

Let us introduce the notion of a random measure on a semiring $K(\mathbf{E})$ of subsets of a space $\mathbf{E}$.

Definition 2.3.1
A family $K(\mathbf{E})$ of subsets of a space $\mathbf{E}$ is called a *semiring* if for all $A, B \in K(\mathbf{E})$

a) $\emptyset \in K(\mathbf{E})$

b) $A \cap B \in K(\mathbf{E})$

c) If additionally $A \subset B$ then $B \setminus A = \cup_{i=1}^{n} B_i$, where $n \in \mathbb{N}$ and $B_j \in K(\mathbf{E})$, $j = 1, \ldots, n$ are pairwise disjoint.

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space.

Definition 2.3.2 1. Let ν be a σ-finite measure on $K(\mathbf{E})$. $W = \{W(A) : A \in K(\mathbf{E})\}$ is called an *(independently scattered signed) random measure* if

a) $W(A)$, $A \in K(\mathbf{E})$ is a (complex-valued) random variable on $(\Omega, \mathcal{F}, \mathcal{P})$, which is square integrable: $\mathbf{E}|W(A)|^2 < \infty$. Later on we write $W(A) \in L^2(\Omega, \mathcal{F}, \mathcal{P})$.

b) $\mathbf{E}\left[W(A)\overline{W(B)}\right] = \nu(A \cap B)$, $A, B \in K(\mathbf{E})$.

2. ν is called the *structure measure (deriving, control measure)* of W. It is given by $\nu(A) = \mathbf{E}|W(A)|^2$, $A \in K(\mathbf{E})$.

3. If $\mathbf{E}W(A) = 0$, $A \in K(\mathbf{E})$, then W is called *centered*.

Later on we consider *only centered random measures* (for simplicity).

Proposition 2.3.1 (Properties of random measures):
A family of complex-valued random variables $W = \{W(A),\ A \in K(\mathbf{E})\}$ is a random measure iff

a) $W(A) \in L^2(\Omega, \mathcal{F}, \mathcal{P}), \quad A \in K(\mathbf{E})$

b) $\mathbf{E}(W(A)\overline{W(B)}) = 0, \quad A, B \in K(\mathbf{E}), \quad A \cap B = \emptyset$
orthogonality (uncorrelatedness) of increments, independent scattering property).

c) $W(A) \stackrel{a.s}{=} \sum_{i=1}^{\infty} W(A_i)$, for $A, \{A_i\} \subset K(\mathbf{E})$, $A = \cup_{i=1}^{\infty} A_i$, $A_i \bigcap A_j = \emptyset$, $i \neq j$. The series $\sum_{i=1}^{\infty} W(A_i)$ converges in $L^2(\Omega, \mathcal{F}, \mathcal{P})$ -sense (mean square convergence, property of σ-additivity).

Proof: Necessity: Property b) of Proposition 2.3.1 is a special case of property b), Definition 2.3.2 for disjoint sets A, B. Let us show that c) holds. First let us show finite additivity of W : $W(A \cup B) \stackrel{a.s}{=} W(A) + W(B)$, $A, B \in K(\mathbf{E})$, $A \cap B = \emptyset$. For this, it suffices to prove $\mathbf{E}|W(A \cup B) - W(A) - W(B)|^2 = 0$.

We have $\mathbf{E}|W(A\cup B)-W(A)-W(B)|^2$

$$
\begin{aligned}
&= \mathbf{E}\left[(W(A\cup B)-W(A)-W(B))(\overline{W(A\cup B)}-\overline{W(A)}-\overline{W(B)})\right]\\
&= \underbrace{\mathbf{E}|W(A\cup B)|^2}_{\nu(A\cup B)}+\underbrace{\mathbf{E}|W(A)|^2}_{\nu(A)}+\underbrace{\mathbf{E}|W(B)|^2}_{\nu(B)}\\
&-\underbrace{\mathbf{E}\left[W(A\cup B)\overline{W}(A)\right]}_{\nu(A)}-\underbrace{\mathbf{E}\left[W(A\cup B)\overline{W}(B)\right]}_{\nu(B)}-\underbrace{\mathbf{E}(W(A)\overline{W(A\cup B)})}_{\nu(A)}-\underbrace{\mathbf{E}(W(A)\overline{W(B)})}_{\nu(\emptyset)=0}\\
&-\underbrace{\mathbf{E}(W(B)\overline{W(A\cup B)})}_{\nu(B)}-\underbrace{\mathbf{E}(W(B)\overline{W(A)})}_{\nu(\emptyset)=0}=\nu(A\cup B)-\nu(A)-\nu(B)=0\\
&\Rightarrow W(A\cup B)\overset{a.s}{=}W(A)+W(B).
\end{aligned}
$$

By induction we can show the finite additivity:

$$\mathbf{E}|W(A)-\sum_{i=1}^{n}W(A_i)|^2=\nu(A)-\sum_{i=1}^{n}\nu(A_i) \tag{2.3.1}$$

for $A_1,\dots,A_n\in K(\mathbf{E})$, $A_i\cap A_j=\emptyset$, $i\neq j$. Hence for $A=\bigcup_{i=1}^{n}A_i$, it holds $W(A)\overset{a.s.}{=}\sum_{i=1}^{n}W(A_i)$. Now if

$$A=\bigcup_{i=1}^{\infty}A_i,\quad A_i\cap A_j=\emptyset,\quad i\neq j,$$

then the convergence of

$$\sum_{i=1}^{\infty}W(A_i)=W(A)$$

in the $L^2(\Omega,\mathcal{F},\mathcal{P})$-sense follows from (2.3.1) because of

$$\nu(A)=\sum_{i=1}^{\infty}\nu(A_i).$$

Sufficiency: Take any $A,B\in K(\mathbf{E})$ and prove b) of Definition 2.3.2. First we notice that $A=\underbrace{(A\setminus B)}_{C_1}\cup\underbrace{(A\cap B)}_{C_2}$, $B=\underbrace{(B\setminus A)}_{C_3}\cup(A\cap B)$, where

$$C_2\in K(\mathbf{E}),\quad C_1=\cup_{i=1}^{m}C_{1i},\quad C_3=\cup_{j=1}^{n}C_{3j},\quad C_{1i},C_{3j}\in K(\mathbf{E}),$$

and are disjoint. Then

$$
\begin{aligned}
\mathbf{E}(W(A)\overline{W(B)})=\mathbf{E}(W(C_1\cup C_2)\overline{W(C_3\cup C_2)})=\mathbf{E}\left[(W(C_2)+\sum_{i=1}^{m}W(C_{1i}))(\overline{W(C_2)}+\sum_{j=1}^{n}\overline{W(C_{3j})})\right]\\
\overset{c)of\,Prop.2.3.1}{=}\mathbf{E}|W(C_2)|^2\overset{Def.2.3.2,2)}{=}\nu(C_2)=\nu(A\cap B),
\end{aligned}
$$

since all C_{1i},C_{3j} are disjoint. □

Let us give an example of random measures.

Example 2.3.1
(Gaussian random measure)
Here we set $W(A) \sim N(0, \nu(A)), A \in K(\mathbf{E})$(cf. Example 1.2.1,2 for a special case).
In Definition 2.3.2, we postulated the existence of W with stated properties. Now we have to prove its existence by giving an explicit construction of a random measure W on some probability space $(\Omega, \mathcal{F}, \mathcal{P})$ with properties a), b) of Definition 2.3.2.

Theorem 2.3.1
Let ν be a σ-finite measure on $K(\mathbf{E})$. Then there exists a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ and a centered (independently scattered) random measure $\{W = W(A), A \in K(\mathbf{E})\}$ defined on $(\Omega, \mathcal{F}, \mathcal{P})$ and $\{A \in \sigma(K(\mathbf{E})) : \nu(A) < \infty\}$ such that ν is the control measure of W.

Proof Let $\mathcal{A} = \sigma(K(\mathbf{E}))$. By Carathéodory's theorem, we can extend the measure ν to a σ-finite measure on $\mathcal{A}$. Let us construct the required probability space $(\Omega, \mathcal{F}, \mathcal{P})$ by setting $\Omega = \mathbf{E}$, $\mathcal{F} = \mathcal{A}$. For the construction of P, we have to distinguish two non-trivial cases:

1. $\nu(\mathbf{E}) \in (0, \infty)$: set $P(B) = \frac{\nu(B)}{\nu(\mathbf{E})}$, $B \in \mathcal{F}$. And let

$$W(A, \omega) = \sqrt{\nu(\mathbf{E})} I_A(\omega), \quad \omega \in \Omega, \quad A \in \mathcal{A}.$$

We have

$$W(A, .) \in L^2(\Omega, \mathcal{F}, \mathcal{P}),$$

$$\mathbf{E}(W(A)\overline{W(B)}) = (\sqrt{\nu(\mathbf{E})})^2 \int_\Omega I_A(\omega) I_B(\omega) P(d\omega) = \nu(\mathbf{E}) P(A \cap B) = \nu(A \cap B), \quad A, B \in \mathcal{A}.$$

Hence W is a (non-centered) independently scattered random measure.

2. $\nu(\mathbf{E}) = \infty$ (the case $\nu(\mathbf{E}) = 0$ is trivial). Since ν is σ-finite, there exists a tiling

$$\mathbf{E} = \bigcup_{i=1}^{\infty} \mathbf{E}_i, \quad \mathbf{E}_i \cap \mathbf{E}_j = \emptyset, \quad i \neq j, \quad E_i \in \mathcal{A},$$

with $\nu(\mathbf{E}_i) < \infty$. Let

$$P(A) = \sum_{n=1}^{\infty} \frac{\nu(A \cap \mathbf{E}_n)}{\nu(\mathbf{E}_n) 2^n}, \quad A \in \mathcal{F}.$$

It is clear that $P(.)$ is a probability measure on $\mathcal{F}$, since

$$P(\mathbf{E}) = \sum_{n=1}^{\infty} \frac{1}{2^n} = 1.$$

For any $B \in \mathcal{A} : \nu(B) < \infty$ define

$$W(B, \omega) = \sum_{n=1}^{\infty} \sqrt{2^n \nu(E_n)} I_{B_n}(\omega),$$

where $B_n = B \cap \mathbf{E}_n$, $n \in N$. This series converges in $L^2(\Omega, \mathcal{F}, \mathcal{P})$: since $\{B_n\}$ are disjoint, it holds

$$\mathbf{E}|\sum_{n=m}^{\infty} \sqrt{2^n \nu(\mathbf{E}_n)} I_{B_n}(\omega)|^2 = \sum_{n=m}^{\infty} 2^n \nu(\mathbf{E}_n)\ \mathbf{E} {I_{B_n}}^2(\omega)$$

$$= \sum_{n=m}^{\infty} 2^n \nu(\mathbf{E}_n) P(B_n) = \sum_{n=m}^{\infty} 2^n \nu(E_n) \frac{\nu(B_n)}{\nu(\mathbf{E}_n) 2^n} = \sum_{n=m}^{\infty} \nu(B_n) \xrightarrow[m \to 0]{} 0,$$

because $\nu(B) = \sum_{n=1}^{\infty} \nu(B_n) < \infty$. Since the scalar product $< ., . >_{L^2(\Omega,\mathcal{F},\mathcal{P})}$ is continuous in its arguments, it holds

$$< W(A), W(B) >_{L^2(\Omega,\mathcal{F},\mathcal{P})} = \mathbf{E}\left[W(A) \cdot \overline{W(B)}\right]$$

$$= \sum_{n=1}^{\infty} 2^n \nu(\mathbf{E}_n) \mathbf{E}(I_{A_n} I_{B_n}) = \sum_{n=1}^{\infty} 2^n \nu(\mathbf{E}_n) P(A_n \cap B_n)$$

$$= \sum_{n=1}^{\infty} \nu(A_n \cap B_n) = \nu(A \cap B),$$

where $A_n = A \cap E_n$, $B_n = B \cap \mathbf{E}_n$, $n \in \mathbb{N}$. Hence W is a non-centered independently scattered random measure.

In order to get a centered random measure W, it does not suffice to subtract $\mathbf{E}W(A)$ from $W(A)$, since this will destroy the property b) of Definition 2.3.2. The right way is to enlarge the probability space $(\Omega, \mathcal{F}, \mathcal{P})$ as follows:

Let Z be a random variable on a new probability space $(\Omega', \mathcal{F}', \mathcal{P}')$ such that $\mathbf{E}_{P'} Z = 0$, $\mathbf{var}_{P'} Z = 1$, where $\mathbf{E}_{P'}$ and $\mathbf{var}_{P'}$ are calculated w.r.t. measure P'. Let W be the random measure constructed as above on $(\Omega, \mathcal{F}, \mathcal{P})$. Take $(\widetilde{\Omega}, \widetilde{F}, \widetilde{\mathcal{P}}) = (\Omega, \mathcal{F}, \mathcal{P}) \otimes (\Omega', \mathcal{F}', \mathcal{P}')$ and define $\widetilde{W}$ on it by $\widetilde{W}(A, \widetilde{\omega}) = W(A, \omega) Z(\omega')$ for $A \in \{\mathcal{A} : \nu(A) < \infty\}$, $\widetilde{\omega} = (\omega, \omega') \in \Omega \times \Omega'$. It can be easily seen that W is a centered independently scattered random measure on $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P})$.

□

2.3.2 Stochastic integral

Let us introduce the stochastic integral

$$\int_{\mathbf{E}} f(t) W(dt),$$

where W is a centered independently scattered random measure on $K(\mathbf{E})$ with control measure ν and $f : \mathbf{E} \to \mathbb{C}$ is a deterministic function. For this, W should be continued to the algebra generated by $K(\mathbf{E})$ and then to $\mathcal{A} = \sigma(K(\mathbf{E}))$.

Let ν be finite, i.e. $\nu(\mathbf{E}) < \infty$. Consider the algebra $\widetilde{\mathcal{A}} = \{\bigcup_{i=1}^{m} B_i$ for some $m \in \mathbb{N}$, $B_i \in K(\mathbf{E})$, $i = 1, \ldots, m\}$.

W can be continued from $K(\mathbf{E})$ to $\widetilde{\mathcal{A}}$ by setting $W(A) \stackrel{a.s.}{=} \sum_{i=1}^{m} W(A_i)$, if $A = \cup_{i=1}^{m} A_i \in \widetilde{\mathcal{A}}$, $A_i \in K(\mathbf{E})$, $A_i \cap A_j = \emptyset$, $i \neq j$. This definition is correct, since for any other representation of A by $A = \bigcup_{j=1}^{n} B_j$, $B_i \cap B_j = \emptyset$, $i \neq j$, $B_j \in K(\mathbf{E})$, it holds

$$\sum_{i=1}^{m} W(A_i) = \sum_{i=1}^{m} \sum_{j=1}^{n} W(A_i \cap B_j) = \sum_{j=1}^{n} \sum_{i=1}^{m} W(B_j \cap A_i) = \sum_{j=1}^{n} W(B_j), \quad \text{a.s.}$$

Definition 2.3.3
A function $f : \mathbf{E} \to \mathbb{C}$ is *simple* if

$$f(x) = \sum_{i=1}^{n} a_i I_{A_i}(x) \tag{2.3.2}$$

for any $n \in \mathbb{N}$, $a_i \in \mathbb{C}$, $A_i \in \mathcal{A}$, $A_i \cap A_j = \emptyset$, $i \neq j$, $\bigcup_{i=1}^{n} A_j = \mathbf{E}$, $i = 1, \ldots, n$.

Definition 2.3.4
The integral of a simple function f from (2.3.2) w.r.t. the (orthogonal) measure W is given by

$$I(f) := \int_{\mathbf{E}} f(x) W(dx) = \sum_{i=1}^{n} a_i W(A_i).$$

Lemma 2.3.1
The above definition of $I(f)$ is correct, i.e. does not depend on the representation (2.3.2) of f.

Proof Let

$$f(x) = \sum_{i=1}^{n} a_i I_{A_i}(x) = \sum_{j=1}^{m} b_j I_{B_j}(x) \quad a_i, b_j \in \mathbb{C},$$

$$A_i \cap A_k = \emptyset, \quad i \neq k, \quad B_j \cap B_l = \emptyset, \quad j \neq l, \bigcup_{i=1}^{n} A_i = \mathbf{E}, \quad \bigcup_{j=1}^{m} B_j = \mathbf{E}.$$

If $A_i \cap B_j \neq \emptyset$, then it follows that $a_i = b_j$. Hence,

$$\sum_{i=1}^{n} a_i W(A_i) = \sum_{i=1}^{n} a_i \sum_{j=1}^{m} W(A_i \cap B_j) = \sum_{j=1}^{m} \sum_{i: A_i \cap B_j \neq \emptyset} a_i W(A_i \cap B_j)$$

$$= \sum_{j=1}^{m} \sum_{i: A_i \cap B_j \neq \emptyset} b_j W(B_j \cap A_i) = \sum_{j=1}^{m} \sum_{i=1}^{n} b_j W(B_j \cap A_i) = \sum_{j=1}^{m} b_j W(B_j), \quad \text{since} \quad W(\emptyset) = 0.$$

Hence, $I(f)$ does not depend on the representation of f as a simple function. □

Lemma 2.3.2 (Properties of I):
Let f and g be two simple functions. Then,

1. Isometry:

$$< I(f), I(g) >_{L^2(\Omega, \mathcal{F}, \mathcal{P})} = \mathbf{E}(I(f)\overline{I(g)}) =< f, g >_{L^2(\mathbf{E}, \nu)} = \int_{\mathbf{E}} f(x)\overline{g(x)}\nu(dx)$$

2. Linearity: $I(af + bg) \overset{a.s.}{=} aI(f) + bI(g), \quad a, b \in \mathbb{C}$

Proof 1. Let

$$f=\sum_{i=1}^{n} a_i I(A_i), \quad g=\sum_{j=1}^{m} b_j I(B_j), \quad a_i, b_j \in \mathbb{C},$$

$A_i \cap A_k = \emptyset, \quad i \neq k \quad B_j \cap B_l = \emptyset, \quad j \neq l, \quad \bigcup_{i=1}^{n} A_i = \bigcup_{j=1}^{m} B_j = \mathbf{E}$. Consider

$$< I(f), I(g) >_{L^2(\Omega,\mathcal{F},\mathcal{P})} = \mathbf{E}(I(f)\overline{I(g)})) = \mathbf{E}(\sum_{i=1}^{n}\sum_{j=1}^{m} a_i\overline{b_j} W(A_i)\overline{W(B_j)})$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i\overline{b_j}\nu(A_i\cap B_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i\overline{b_j}\int_{\mathbf{E}} \underbrace{I_{A_i\cap B_j}(x)}_{I_{A_i}(x)I_{B_j}(x)}\nu(dx) = \int_{\mathbf{E}} \sum_{i=1}^{n} a_i I_{A_i}(x)\overline{\sum_{j=1}^{m} b_j I_{B_j}(x)}\nu(dx)$$

$$= \int_{\mathbf{E}} f(x)\overline{g(x)}\nu(dx) =< f, g >_{L^2(\mathbf{E},\nu)} .$$

2. By 1) we get

$$\begin{aligned}
&\|I(af+bg)-aI(f)-bI(g)\|^2_{L^2(\Omega,\mathcal{F},\mathcal{P})}\\
&=< I(af+bg)-aI(f)-bI(g), I(af+bg)-aI(f)-bI(g) >_{L^2(\Omega,\mathcal{F},\mathcal{P})}\\
&=< I(af+bg), I(af+bg) >_{L^2(\Omega,\mathcal{F},\mathcal{P})} -a < I(f), I(af+bg) >_{L^2(\Omega,\mathcal{F},\mathcal{P})}\\
&- b < I(g), I(af+bg) >_{L^2(\Omega,\mathcal{F},\mathcal{P})} -\overline{a} < I(af+bg), I(f) >_{L^2(\Omega,\mathcal{F},\mathcal{P})} -\overline{b} < I(af+bg), I(g) >_{L^2(\Omega,\mathcal{F},\mathcal{P})}\\
&+ a\overline{a} < I(f), I(f) >_{L^2(\Omega,\mathcal{F},\mathcal{P})} +b\overline{b} < I(g), I(g) >_{L^2(\Omega,\mathcal{F},\mathcal{P})} -\overline{b} < I(af+bg), g >_{L^2(\Omega,\mathcal{F},\mathcal{P})}\\
&+ a\overline{b} < I(f), I(g) >_{L^2(\Omega,\mathcal{F},\mathcal{P})} +b\overline{b} < I(g), I(g) >_{L^2(\Omega,\mathcal{F},\mathcal{P})}\\
&\overset{1)}{=}< af+bg, af+bg >_{L^2(\mathbf{E},\nu)} -\overline{a} < af+bg, f >_{L^2(\mathbf{E},\nu)} -\overline{b} < af+bg, g >_{L^2(\mathbf{E},\nu)}\\
&- a < f, af+bg >_{L^2(\mathbf{E},\nu)} +a\overline{a} < f, f >_{L^2(\mathbf{E},\nu)} +a\overline{b} < f, g >_{L^2(\mathbf{E},\nu)} -b < g, af+bg >_{L^2(\mathbf{E},\nu)}\\
&+ b\overline{a} < g, f >_{L^2(\mathbf{E},\nu)} +b\overline{b} < g, g >_{L^2(\mathbf{E},\nu)} = 0
\end{aligned}$$

by linearity of the scalar product $< ., . >_{L^2(\mathbf{E},\nu)}$. Hence, $I(af+bg)-aI(f)-bI(g) \overset{a.s.}{=} 0$ □

Definition 2.3.5
The *integral* of a function $f \in L^2(\mathbf{E},\nu)$ with respect to the random measure W is given by $I(f) = \lim_{n\to\infty} I(f_n)$, where $\{f_n\}_{n\in N}$ is a sequence of simple functions s.t.

$$f_n \xrightarrow[n\to\infty]{L^2(\mathbf{E},\nu)} f,$$

and the limit $\lim I(f_n)$ is understood in the mean quadratic sense, i.e., in $L^2(\Omega,\mathcal{F},\mathcal{P})$:

$$\mathbf{E}|I(f_n)-I(f)|^2 = \|I(f_n)-I(f)\|^2_{L^2(\Omega,\mathcal{F},\mathcal{P})} \to 0 \quad \text{as} \quad n\to\infty.$$

Since the set of simple functions is dense in $L^2(\mathbf{E},\nu)$, such a sequence $\{f_n\}_{n\in N}$ always exists.

Lemma 2.3.3 1. Definition 2.3.5 is correct, i.e., $I(f)$ exists and does not depend on the choice of the approximating sequence $\{f_n\}_{n\in N}$ of f.

2. For any $f, g \in L^2(\mathbf{E},\nu)$ it holds

a) $< I(f), I(g) >_{L^2(\Omega,\mathcal{F},\mathcal{P})} = < f, g >_{L^2(\mathbf{E},\nu)}$

b) $I(af + bg) = aI(f) + bI(g), \quad a, b \in \mathbb{C}$

c) If $g_n \overset{L^2(\mathbf{E},\nu)}{\longrightarrow} g$ as $n \to \infty$, then $I(g_n) \overset{L^2(\Omega,\mathcal{F},\mathcal{P})}{\longrightarrow} I(g)$ as $n \to \infty$, where $\{g_n\}_{n\in\mathbb{N}}$, $g \in L^2(\mathbf{E}, \nu)$ are not necessarily simple.

Proof 1. Let $\{f_n\}_{n\in N}$, $\{g_n\}_{n\in N}$ be simple functions s.t. $f_n \overset{L^2(\mathbf{E},\nu)}{\longrightarrow} f$ as $n \to \infty$, $g_n \overset{L^2(\mathbf{E},\nu)}{\longrightarrow} f$ as $n \to \infty$. First let us prove that $I(f)$ exists. By Lemma 2.3.2, it holds

$$\|I(f_n) - I(f_m)\|_{L^2(\Omega,\mathcal{F},\mathcal{P})} = \|I(f_n - f_m)\|_{L^2(\Omega,\mathcal{F},\mathcal{P})} = \|f_n - f_m\|_{L^2(\mathbf{E},\nu)} \xrightarrow[n,m\to\infty]{} 0,$$

because $\{f_n\}$ is a fundamental sequence in $L^2(\mathbf{E}, \nu)$ due to

$$f_n \xrightarrow[n\to\infty]{L^2(\mathbf{E},\nu)} f.$$

Then $\{I(f_n)\}_{n\in\mathbb{N}}$ is a fundamental sequence in $L^2(\Omega, \mathcal{F}, \mathcal{P})$, and since $L^2(\Omega, \mathcal{F}, \mathcal{P})$ is a complete space, $\exists$ a limit $\lim_{n\to\infty} I(f_n) \overset{def}{=} I(f)$.
Now let us show that

$$\lim_{n\to\infty} I(f_n) = \lim_{n\to\infty} I(g_n) = I(f).$$

Suppose that

$$\lim_{n\to\infty} I(f_n) = \xi, \quad \lim_{n\to\infty} I(g_n) = \eta,$$

and show that $\xi \overset{a.s.}{=} \eta$. It holds

$$\|I(f_n) - I(g_n)\|_{L^2(\Omega,\mathcal{F},\mathcal{P})} = \|I(f_n - g_n)\|_{L^2(\Omega,\mathcal{F},\mathcal{P})}$$
$$\overset{Lemma\ 2.3.2}{=} \|f_n - g_n\|_{L^2(\mathbf{E},\nu)} \le \|f_n - f\|_{L^2(\mathbf{E},\nu)} + \|g_n - f\|_{L^2(\mathbf{E},\nu)} \xrightarrow[n\to\infty]{} 0,$$

hence $I(f_n) - I(g_n) \overset{L^2(\Omega,\mathcal{F},\mathcal{P})}{\longrightarrow} 0$ as $n \to \infty$, otherwise $I(f_n) - I(g_n) \overset{L^2(\Omega,\mathcal{F},\mathcal{P})}{\longrightarrow} \xi - \eta$ as $n \to \infty$, hence $\xi \overset{a.s.}{=} \eta$.

2. Let $f_n \overset{L^2(\mathbf{E},\nu)}{\longrightarrow} f$ as $n \to \infty$, $g_n \overset{L^2(\mathbf{E},\nu)}{\longrightarrow} g$ as $n \to \infty$.

a) By Lemma 2.3.2, 1) and since the scalar product is continuous we have $< I(f), I(g) >_{L^2(\Omega,\mathcal{F},\mathcal{P})} =$

$$\lim_{n,m\to\infty} < I(f_n), I(g_m) >_{L^2(\Omega,\mathcal{F},\mathcal{P})} = \lim_{n,m\to\infty} < f_n, g_m >_{L^2(\mathbf{E},\nu)} = < f, g >_{L^2(\mathbf{E},\nu)} .$$

b) $I(af + bg) = \lim_{n\to\infty} I(af_n + bg_n)) \overset{\text{Lemma } 2.3.2,2}{=} \lim_{n\to\infty} \left(aI(f_n) + bI(g_n)\right)^2$

$$= a \lim_{n\to\infty} I(f_n) + b \lim_{n\to\infty} I(g_n) = aI(f) + bI(g).$$

c) It holds $\|I(g)-I(g_n)\|_{L^2(\Omega,\mathcal{F},\mathcal{P})} = \|I(g-g_n)\|_{L^2(\Omega,\mathcal{F},\mathcal{P})} \overset{2a)}{=} \|g-g_n\|_{L^2(\mathbf{E},\nu)} \to 0$ as $n\to\infty$, since $g_n \overset{L^2(\Omega,\mathcal{F},\mathcal{P})}{\longrightarrow} g$ as $n\to\infty$. Then $I(g_n) \overset{L^2(\Omega,\mathcal{F},\mathcal{P})}{\longrightarrow} I(g)$ as $n\to\infty$.

□

Now let us consider the case of a σ-finite ν, i.e., there exists a family of pairwise disjoint sets

$$\{\mathbf{E}_n\} \subset \mathbf{E} : \bigcup_{n=1}^{\infty} \mathbf{E}_n = \mathbf{E}, \quad \mathbf{E}_n \in K(\mathbf{E}), \nu(\mathbf{E}_n) < \infty, n \in \mathbb{N}.$$

Then, it follows from general measure theory that $f \in L^2(\mathbf{E},\mathcal{A},\nu)$ iff $f_n = f\mid_{\mathbf{E}_n} \in L^2(\mathbf{E}_n,\mathcal{A}_n,\nu_n)$ (where $\mathcal{A}_n = \sigma(K(\mathbf{E}_n))$ and ν_n is a (Lebesgue) continuation of $\nu\mid_{K(\mathbf{E}_n)}$ from $K(\mathbf{E}_n) = K(E)\cap E_n$ onto $\mathcal{A}_n$) and

$$\int_{\mathbf{E}} |f(x)|^2 \nu(dx) = \sum_{n=1}^{\infty} \int_{\mathbf{E}_n} |f_n(x)|^2 \nu_n(dx) < \infty, \tag{2.3.2}$$

where the value of $\int_{\mathbf{E}} |f(x)|^2\nu(dx)$ does not depend on the choice of $\{\mathbf{E}_n\}_{n\in\mathbb{N}}$.

Definition 2.3.6
For $f \in L^2(\mathbf{E},\nu)$, set the integral $I(f)$ w.r.t. the random measure W to be

$$I(f) = \sum_{n=1}^{\infty} I_n(f_n), \text{ where } f_n = f\mid_{\mathbf{E}_n} \text{ and } I_n(f_n) = \int_{\mathbf{E}_n} f_n(x) W(dx)$$

is the integral introduced in Definition 2.3.5.

Lemma 2.3.4 1. Definition 2.3.6 is correct i.e. $I(f)$ exists and does not depend on the choice of the sets $\{\mathbf{E}_n\}_{n\in\mathbb{N}}$.

2. Stochastic integral I preserves the scalar product (isometry):
for any $f,g \in L^2(\mathbf{E},\nu)$ it holds $\langle I(f), I(g)\rangle_{L^2(\Omega,\mathcal{F},\mathcal{P})} = \langle f,g\rangle_{L^2(\mathbf{E},\nu)}$

3. The random measure W can be continued from $\widetilde{\mathcal{A}}$ to $\{A\in\mathcal{A} : \nu(A) < \infty\}$ by the formula $W(A) = I(I_A)$.

4. If W is a centered random measure, then $\mathbf{E}(I(f)) = 0$ for all $f \in L^2(\mathbf{E},\nu)$.

Proof 1. Let us show that the series $\sum_{n=1}^{\infty} J_n(f_n)$ converges. First, it holds

$$< J_n(f_n), J_m(f_m) >_{L^2(\Omega,\mathcal{F},\mathcal{P})} = 0, \quad m \neq n.$$

Indeed, this statement is trivial for simple f_n and f_m, because

$$< W(A), W(B) >_{L^2(\Omega,\mathcal{F},\mathcal{P})} = 0 \text{ for } A \in \sigma(K(\mathbf{E}_n)), \quad B \in \sigma(K(\mathbf{E}_m)).$$

For general $f_n \in L^2(\mathbf{E}_n,\nu_n)$ and $f_m \in L^2(\mathbf{E}_m,\nu_m)$ it suffices to approximate them by a sequence of simple functions and use the continuity of $<.,.>$. Hence,

$$\|\sum_{i=1}^{n} I_i(f_i) - \sum_{i=1}^{m} I_i(f_i)\|^2_{L^2(\Omega,\mathcal{F},\mathcal{P})} = \sum_{i=\min\{m,n\}+1}^{\max\{m,n\}} \|I_i(f_i)\|^2_{L^2(\Omega,\mathcal{F},\mathcal{P})}$$
$$= \sum_{i=\min\{m,n\}+1}^{\max\{m,n\}} \|f_i\|^2_{L^2(\mathbf{E}_i,\nu_i)} = \sum_{i=\min\{m,n\}+1}^{\max\{m,n\}} \int_{\mathbf{E}_i} |f_i(x)|^2 \nu_i(dx) \to 0 \quad \text{as} \quad n,m \to \infty,$$

since the series (2.3.2) converges. Then, $\sum_{n=1}^{\infty} I_n(f_n)$ converges as well.
Now let us show that $I(f)$ does not depend on $\{\mathbf{E}_i\}_{i\in\mathbb{N}}$. Let

$$\mathbf{E} = \bigcup_{i=1}^{\infty} \mathbf{E}'_i, \quad \mathbf{E}'_i \subset K(\mathbf{E}), \quad \mathbf{E}'_i \cap \mathbf{E}'_j = \emptyset, \quad i \neq j \quad \nu(\mathbf{E}'_i) < \infty, \quad i \in \mathbb{N}$$

Denote $f_{i,j} = f\,|_{\mathbf{E}_i \cap \mathbf{E}'_j}$, $i,j \in \mathbb{N}$. Then $f_i = f\,|_{\mathbf{E}_i} = \sum_{j=1}^{\infty} f_{i,j}$, $i \in \mathbb{N}$.
Since $\sum_{j=1}^{n} f_{i,j} \overset{L^2(\mathbf{E}_i)}{\longrightarrow} f_i$ as $n \to \infty$, it holds

$$I_i(f_i) = \sum_{j=1}^{\infty} I_i(f_{i,j})$$

by Lemma 2.3.3, 2c). Consider the integral I_{ij} on

$$L^2(\mathbf{E}_i \cap \mathbf{E}'_j, \nu) : I_{ij}(g) = \int_{\mathbf{E}_i \cap \mathbf{E}'_j} g(x)\nu(x), \quad g \in L^2(\mathbf{E}_i \cap \mathbf{E}'_j, \nu)$$

Evidently, it holds $I_{ij}(g) = I_i(g) = I'_j(g)$ for $g \in L^2(\mathbf{E}_i \cap \mathbf{E}'_j, \nu)$, where I'_j is the integral on $L^2(\mathbf{E}'_j, \nu)$. Then, we get

$$\sum_{i=1}^{\infty} I_i(f_i) = \sum_{i,j=1}^{\infty} I_{ij}(f_{ij}), \quad \sum_{j=1}^{\infty} I'_j(f_j) = \sum_{j,i=1}^{\infty} I'_{ij}(f_{ij}),$$

where $I_{ij} = I'_{ij}$ on $L^2(\mathbf{E}_i \cap \mathbf{E}'_j, \nu)$. Since $< I_{ij}(t_{ij}), I_{kl}(f_{kl}) >_{L^2(\Omega,\mathcal{F},\mathcal{P})} = 0$ for $(i,j) \neq (k,l)$, we get

$$I(f) = \sum_{i=1}^{\infty} I_i(f_i) = \sum_{i,j=1}^{\infty} I_{ij}(f_{ij}) = \sum_{j,i=1}^{\infty} I_{ij}(f_{ij}) = \sum_{j,i=1}^{\infty} I'_{ij}(f_{ij})$$
$$= \sum_{j=1}^{\infty} I'_j(f_j) = I'(f), \quad \text{because}$$
$$\sum_{i,j=1}^{\infty} \|I_{ij}(f_{ij})\|^2_{L^2(\Omega,\mathcal{F},\mathcal{P})} = \sum_{i,j=1}^{\infty} \|f_{ij}\|^2_{L^2(\mathbf{E}_i \cap \mathbf{E}'_j, \nu)} = \int_{\mathbf{E}} |f(x)|^2 \nu(dx) < \infty,$$

here $I'(f)$ is the stochastic integral of f constructed with the help of tiling $\{\mathbf{E}'_j\}_{j\in\mathbb{N}}$ of $\mathbf{E}$.

2. Since $\mathbf{E} = \bigcup_{i=1}^{\infty} \mathbf{E}_i$ with $\nu(E_i) < \infty$, $i \in \mathbb{N}$, for any $f, g \in L^2(\mathbf{E}, \nu)$ it holds

$$f = \sum_{i=1}^{\infty} f_i, \quad g = \sum_{i=1}^{\infty} g_i, \quad \text{where} \quad f_i = f\,|_{\mathbf{E}_i}, \quad g_i = g\,|_{\mathbf{E}_i}, \quad i \in \mathbb{N}.$$

The above series converge in $L^2(\mathbf{E},\nu)$-sense. By the continuity of the scalar product and Lemma 2.3.3, 2) we have

$$< I(f), I(g) >_{L^2(\Omega,\mathcal{F},\mathcal{P})} = < \sum_{i=1}^{\infty} I_i(f_i), \sum_{j=1}^{\infty} I_j(g_j) >_{L^2(\Omega,\mathcal{F},\mathcal{P})} = \lim_{n,m\to\infty} < \sum_{i=1}^{n} I_i(f_i), \sum_{j=1}^{m} I_j(g_j) >_{L^2(\Omega,\mathcal{F},\mathcal{P})}$$

$$= \lim_{n,m\to\infty} \sum_{i=1}^{n}\sum_{j=1}^{m} < I_i(f_i), I_j(g_j) >_{L^2(\Omega,\mathcal{F},\mathcal{P})} \overset{Lemma\ 2.3.3}{=} \lim_{n,m\to\infty} \sum_{i=1}^{n}\sum_{j=1}^{m} < f_i, g_j >_{L^2(\mathbf{E},\nu)}$$

$$= \lim_{n,m\to\infty} < \sum_{i=1}^{n} f_i, \sum_{j=1}^{m} g_j >_{L^2(\mathbf{E},\nu)} = < f, g >_{L^2(\mathbf{E},\nu)},$$

where $< I_i(f_i), I_j(g_j) >_{L^2(\Omega,\mathcal{F},\mathcal{P})} = < f_i, g_j >_{L^2(\mathbf{E},\nu)} = \delta_{ij}$, since $\mathbf{E}_i \cap \mathbf{E}_j = \emptyset$, $i \neq j$.

3. Take a random measure $\widetilde{W}$ on $\mathcal{A}$ defined by $\widetilde{W}(A) = I(I_A)$, $A \in \mathcal{A} : \nu(A) < \infty$. It is an independently scattered random measure by linearity of the integral and isometry property, since for any $B, C \in \{A \in \mathcal{A} : \nu(A) < \infty\}$ it holds

$$< W(B), W(C) >_{L^2(\Omega,\mathcal{F},\mathcal{P})} = < I(I_B), I(I_C) >_{L^2(\Omega,\mathcal{F},\mathcal{P})} = < I_B, I_C >_{L^2(\mathbf{E},\nu)} = \nu(B \cap C).$$

Show that $\widetilde{W}$ is continuation of W from $K(\mathbf{E})$ onto $\{A \in \mathcal{A} : \nu(A) < \infty\}$. For any $B \in K(\mathbf{E})$ we have $\nu(B) < \infty$, $B = \bigcup_{i=1}^{\infty} B_i$, where $B_i = B \cap \mathbf{E}_i$, $i \in \mathbb{N}$, $\nu(B) = \sum_{i=1}^{\infty} \nu(B_i)$. Furthermore, we have

$$I_B = \sum_{i=1}^{\infty} I_{B_i} \in L^2(\mathbf{E},\nu) \text{ and thus } I(I_B) = \sum_{i=1}^{\infty} I_i(I_{B_i}).$$

Since $I_i(I_{B_i}) = W(B_i)$, $i \in \mathbb{N}$, (because $B_i \in K(\mathbf{E}_i)$), we get

$$\widetilde{W}(B) = I(I_B) = \sum_{i=1}^{\infty} W(B_i) = W(B), \quad B \in K(\mathbf{E}),$$

(all series converge in L^2-sense).

4. Follows for $\nu(\mathbf{E}) < \infty$ from the fact, that for centered W it holds $\mathbf{E}I(f_n) = 0$ obviously for any simple f_n, $n \in \mathbb{N}$. Suppose that $f_n \overset{L^2(\mathbf{E},\nu)}{\longrightarrow} f$ as $n \to \infty$, hence, by Lemma 2.3.3, we get $I(f_n) \overset{L^2(\Omega,\mathcal{F},\mathcal{P})}{\longrightarrow} I(f)$ as $n \to \infty$, and hence

$$|\mathbf{E}I(f)| = |\mathbf{E}I(f) - \mathbf{E}I(f_n)| \leq \mathbf{E}|I(f) - I(f_n)| \leq \sqrt{\mathbf{E}|I(f) - I(f_n)|^2}$$
$$= \|I(f) - I(f_n)\|_{L^2(\Omega,\mathcal{F},\mathcal{P})} \longrightarrow 0 \quad \text{as} \quad n \to \infty$$

by the Lyapunow inequality. Then, $\mathbf{E}I(f) = 0$ in case $\nu(\mathbf{E}) < \infty$.
If ν is σ-finite, the usual reasoning with

$$f = \sum_{i=1}^{\infty} f_i, \quad f_i = f\,|_{\mathbf{E}_i}, \quad i \in \mathbb{N}, \quad \mathbf{E}I_i(f_i) = 0, \quad i \in \mathbb{N}$$

proves the general cases since

$$\mathbf{E}I(f) = \sum_{i=1}^{\infty} \mathbf{E}I_i(f_i) = 0.$$

□

2.4 Spectral representation for stationary random functions

Let $X = \{X(t), t \in T\}$ be a (centered) complex-valued random function defined on $(\Omega, \mathcal{F}, \mathcal{P})$ with $\mathbf{E}|X(t)|^2 < \infty$, $t \in T$ and covariance function $C(s,t) = \mathbf{E}[X(s)\overline{X(t)}]$, $s,t \in T$. Let us discuss necessary and sufficient conditions, under which X has the following *spectral representation*:

Definition 2.4.1

$$X(t) = \int_{\mathbf{E}} f(t,x)W(dx), \quad t \in T \tag{2.4.1}$$

for some $f(t,\cdot) \in L^2(\mathbf{E}, \mathcal{A}, \nu)$, $t \in T$, where W is a independently scattered centered random measure on $(\Omega, \mathcal{F}, \mathcal{P})$ with control measure ν (ν is σ-finite on $(\mathbf{E}, \mathcal{A})$) on $\{A \in \mathcal{A} : \nu(A) < \infty\}$. Here spaces T and $\mathbf{E}$ can be any abstract spaces, and equality (2.4.1) is understood in L^2-sense, i.e. a.s..

Theorem 2.4.1 (Karhunen):
The random function X introduced above has spectral representation (2.4.1) iff its covariance function $C(s,t)$ admits the factorization

$$C(s,t) = \int_{\mathbf{E}} f(s,x)\overline{f(t,x)}\nu(dx), \quad s,t \in T \tag{2.4.2}$$

Here, the spectral measure W in (2.4.1) *is given* on an extension $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathcal{P}})$ of the original probability space $(\Omega, \mathcal{F}, \mathcal{P})$. It can be put $(\Omega, \mathcal{F}, \mathcal{P})$ if the system of functions $\{f(\cdot,t) : t \in T\}$ is complete in $L^2(\mathbf{E}, \mathcal{A}, \nu)$, i.e., for $g \in L^2(\mathbf{E}, \mathcal{A}, \nu)$ it follows from $< f(t,\cdot), g >_{L^2(\mathbf{E},\mathcal{A},\nu)} = 0$, $t \in T$ that $g = 0$ ν-a.e..

Proof

1. Necessity: If X has representation (2.4.1) then by isometry property (cf. Lemma 2.3.4, 2)) we get

$$C(s,t) = \mathbf{E}[X(s)\overline{X(t)}] =< X(s), X(t) >_{L^2(\Omega,\mathcal{F},\mathcal{P})} =< f(s,\cdot), f(t,\cdot) >_{L^2(\mathbf{E},\nu)}$$
$$= \int_{\mathbf{E}} f(s,x)\overline{f(t,x)}\nu(dx), \quad s,t \in T.$$

2. Sufficiency: First let us consider the easier case of $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathcal{P}}) = (\Omega, \mathcal{F}, \mathcal{P})$, i.e. the system $\{f(t,\cdot) : t \in T\}$ is complete. Consider the mapping ψ from the space of functions onto $(\mathbf{E}, \nu)$ into $L^2(\Omega, \mathcal{F}, \mathcal{P})$ given by $\psi : f(t,\cdot) \mapsto X(t,\cdot)$. Let us show that we can extend ψ onto $L^2(\mathbf{E}, \mathcal{A}, \nu)$. First, set

$$\psi\left(\sum_{i=1}^{n} a_i f(t_i,\cdot)\right) = \sum_{i=1}^{n} a_i X(t_i,\cdot)$$

by linearity for any $n \in \mathbb{N}$, $a_i \in \mathbb{C}$, $t_i \in T$, $i = 1, \dots, n$.
Let us show that ψ does not depend on the concrete representation of

$$\sum_{i=1}^{n} a_i f(t_i, \cdot).$$

Namely, prove that

$$\sum_{i=1}^{n} a_i X(t_i, \cdot) \overset{a.s}{=} \sum_{j=1}^{m} b_j X(s_j, \cdot) \tag{2.4.3}$$

for any $m \in \mathbb{N}$, $b_j \in \mathbb{C}$, $s_j \in T$, $j = 1, \dots, m$ such that

$$\sum_{i=1}^{n} a_i f(t_i, \cdot) = \sum_{j=1}^{m} b_j f(s_j, \cdot).$$

ψ is an isometry between the space of all linear combinations

$$L(f) = \left\{ \sum_{i=1}^{k} c_i f(y_i, \cdot) : k \in \mathbb{N}, \quad c_i \in \mathbb{C}, \quad y_i \in T, \quad i = 1, \dots, k \right\}$$

and

$$L(X) = \left\{ \sum_{i=1}^{k} c_i X(y_i, \cdot) : k \in \mathbb{N}, \quad c_i \in \mathbb{C}, \quad y_i \in T, \quad i = 1, \dots, k \right\}.$$

because

$$\left\langle \sum_{i=1}^{n} a_i f(t_i, \cdot), \sum_{j=1}^{m} b_j f(s_j, \cdot) \right\rangle_{L^2(\mathbf{E}, \nu)} = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i \overline{b_j} \int_{\mathbf{E}} f(t_i, x) \overline{f(s_j, x)} \nu(dx)$$

$$\overset{by (2.4.2)}{=} \sum_{i=1}^{n} \sum_{j=1}^{m} a_i \overline{b_j} C(t_i, s_j) = \mathbf{E} \left[\sum_{i=1}^{n} \sum_{j=1}^{m} a_i \overline{b_j} X(t_i, \cdot) \overline{X(s_j, \cdot)} \right]$$

$$= \left\langle \sum_{i=1}^{n} a_i X(t_i, \cdot), \sum_{j=1}^{m} b_j X(s_j, \cdot) \right\rangle_{L^2(\Omega, \mathcal{F}, \mathcal{P})}.$$

Hence,

$$\| \sum_{i=1}^{n} a_i X(t_i, \cdot) - \sum_{j=1}^{m} b_j X(t_j, \cdot) \|_{L^2(\Omega, \mathcal{F}, \mathcal{P})} = \| \sum_{i=1}^{n} a_i f(t_i, \cdot) - \sum_{j=1}^{m} b_j f(s_j, \cdot) \|_{L^2(\mathbf{E}, \nu)} = 0$$

and (2.4.3) is proved. Since $\{f(t, \cdot) : t \in T\}$ is complete, we have that closure $\overline{L(f)}$ of $L(f)$ in $L^2(\mathbf{E}, \nu)$ is equal to the whole $L^2(\mathbf{E}, \nu)$. As an isometry ψ can be extended to the mapping

$$\psi : \underbrace{\overline{L(f)}}_{= L^2(\mathbf{E}, \nu)} \longrightarrow \overline{L(X)}, \quad \text{where } \overline{L(X)} = \text{ closure of } L(X) \text{ in } L^2(\Omega, \mathcal{F}, \mathcal{P}).$$

Now define $W(B) = \psi(I_B)$ for any $B \in \mathcal{A} : \nu(B) < \infty$. W is an independently scattered random measure on $(\Omega, \mathcal{F}, \mathcal{P})$ with control measure ν, since

$$< W(A), W(B) >_{L^2(\Omega,\mathcal{F},\mathcal{P})} = < \psi(I_A), \psi(I_B) >_{L^2(\Omega,\mathcal{F},\mathcal{P})} = < I_A, I_B >_{L^2(\mathbf{E},\nu)} = \nu(A \cap B),$$
$$A, B \in \mathcal{A} : \nu(A) < \infty, \quad \nu(B) < \infty.$$

The measure W is centered, since $W(A) \in \overline{L(X)}$ for any $A \in \mathcal{A} : \nu(A) < \infty$, and it holds $\mathbf{E}\xi = 0$ for any $\xi \in \overline{L(X)}$ due to $\mathbf{E}X(t) = 0$, $t \in T$.
Take the stochastic integral $I(\cdot)$ on $L^2(\mathbf{E}, \nu)$ w.r.t.W. By Lemma 2.3.4, 2) I is an isometry from $L^2(\mathbf{E}, \nu)$ onto $L^2_W \subset L^2(\Omega, \mathcal{F}, \mathcal{P})$. Show that $\psi \equiv I$ (and thus $Im_\psi\ \overline{L(f)} = L^2_W$). Since $\psi(I_B) = I(I_B)$ by their definition for $B \in \mathcal{A} : \nu(B) < \infty$, and linear combinations of indicators are dense in $L^2(\mathbf{E}, \mathcal{A}, \nu)$, we get $\psi \equiv I$ as mappings $L^2(\mathbf{E}, \nu) \longrightarrow \overline{L(f)} = L^2_W$. Then, $X(t) = I(f(t, \cdot))$, $t \in T$ is the spectral representation we sought for.
Now let us return to the special case of $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathcal{P}}) \neq (\Omega, \mathcal{F}, \mathcal{P})$, i.e. $\{f(t, \cdot) : t \in T\}$ is not complete. In this case, $\overline{L(f)} \subset L^2(\mathbf{E}, \nu)$, $\overline{L(f)} \neq L^2(\mathbf{E}, \nu)$. Take the orthogonal complement of $\overline{L(f)}$ in $L^2(\mathbf{E}, \nu)$ (denote it by $\overline{L(f)}^\perp$) and consider a basis

$$\{g(s, \cdot) : s \in T'\} \subset L^2(\mathbf{E}, \nu), \quad \text{where} \quad T' \cap T = \emptyset.$$

Introduce the function

$$C_1(s, t) = \int_{\mathbf{E}} g(s, x)\overline{g(t, x)}\nu(dx), \quad s, t \in T'.$$

Since

$$\sum_{i,j=1}^{n} a_i \overline{a_j} C_1(t_i, t_j) = \int_{\mathbf{E}} \left| \sum_{i=1}^{n} a_i g(t_i, x) \right|^2 \nu(dx) \geq 0,$$

C_1 is positive semi-definite. Hence, there exists a probability space $(\Omega', \mathcal{F}', \mathcal{P}')$ and a centered (complex-valued) Gaussian random function $X' = \{X'(t) : t \in T'\}$ on $(\Omega', \mathcal{F}', \mathcal{P}')$ such that X' has the covariance function

$$C_1 : C_1(s, t) = \mathbf{E}\left[X'(s)\overline{X'(t)}\right], \quad s, t \in T'.$$

Introduce the extension of $(\Omega, \mathcal{F}, \mathcal{P})$ by $(\widetilde{\Omega}, \widetilde{F}, \widetilde{P}) = (\Omega, \mathcal{F}, \mathcal{P}) \times (\Omega', \mathcal{F}', \mathcal{P}')$.
For any $t \in T$, $s \in T'$, $X(t)$ and $X'(s)$ are independent on $(\widetilde{\Omega}, \widetilde{F}, \widetilde{P})$, where for $\widetilde{\omega} = (\omega, \omega') \in \widetilde{\Omega}$ we set $X(t, \widetilde{\omega}) = X(t, \omega)$, $X'(s, \widetilde{\omega}) = X'(s, \omega')$. Introduce the (centered) random function
$\widetilde{X} = \{\widetilde{X}(t) : t \in T \cup T'\}$ by

$$\widetilde{X}(t, \widetilde{\omega}) = \begin{cases} X(t, \widetilde{\omega}), & t \in T, \\ X'(t, \widetilde{\omega'}), & t \in T'. \end{cases}$$

It holds

$$\widetilde{C}(s,t) = \mathbf{E}\left[\widetilde{X}(s)\overline{\widetilde{X}(t)}\right] = \begin{cases} C(s,t), & s,t \in T, \\ C_1(s,t), & s,t \in T', \\ 0, & s \in T,\ t \in T' \text{ or } s \in T',\ t \in T. \end{cases}$$

Hence,

$$\widetilde{C}(s,t) = \int_{\mathbf{E}} h(s,x)\overline{h(t,x)}\nu(dx), \quad \text{with}$$

$$h(t,x) = \begin{cases} f(t,x), & t \in T, \\ g(t,x), & t \in T', \end{cases}$$

$x \in \mathbf{E}$, since $< f(t,\cdot), g(s,\cdot) >_{L^2(\mathbf{E},\nu)} = 0$ for $t \in T$, $s \in T'$.
Furthermore, $\overline{L(h)} = L^2(\mathbf{E},\nu)$, because $\{h(t,\cdot) : t \in T \cup T'\}$ is complete in $L^2(\mathbf{E},\nu)$. Then we can use the first part of the proof in 2) and write

$$\widetilde{X}(t) = \int_{\mathbf{E}} h(t,x)W(dx), \quad t \in T \cup T',$$

where W is an independently scattered random measure defined on $(\widetilde{\Omega}, \widetilde{F}, \widetilde{\mathcal{P}})$. For $t \in T$, we get

$$X(t,\omega) = \widetilde{X}(t,\widetilde{\omega}) = \int_{\mathbf{E}} h(t,x)W(dx) = \int_{\mathbf{E}} f(t,x)W(dx).$$

□

Corollary 2.4.1 (Cramér):
Let $\{X(t) : t \in \mathbb{R}^d\}$ be a centered complex-valued (wide sense) stationary random field on $(\Omega, \mathcal{F}, \mathcal{P})$ which is continuous in the mean square sense. Let ν be its spectral measure from Bochner's theorem. Then, there exists an independently scattered centered random measure $W = \{W(A) : A \in \mathcal{B}(\mathbb{R}^d)\}$ on $(\Omega, \mathcal{F}, \mathcal{P})$ with control measure ν such that

$$X(t) = \int_{\mathbb{R}^d} e^{i<t,x>}W(dx), \quad t \in \mathbb{R}^d, \quad a.s. \tag{2.4.4}$$

Proof Since X is mean square continuous, its covariance function

$$C_0(h) = \mathbf{E}\left[X(s)\overline{X(s+h)}\right]$$

is continuous at the origin. Hence, by Bochner's theorem, we get

$$C_0(h) = \int_{\mathbb{R}^d} e^{i<h,x>}\nu(dx)$$

for a finite (spectral) measure ν on $\mathcal{B}_{\mathbb{R}^d}$, in other words

$$C(s,t) = \mathbf{E}\left[X(s)\overline{X(t)}\right] = C_0(s-t) = \int_{\mathbb{R}^d} e^{i<s-t,x>}\nu(dx) = \int_{\mathbb{R}^d} e^{i<s,x>} \cdot \overline{e^{i<t,x>}}\nu(dx).$$

Put $f(t,x) = e^{i<t,x>}$, $t, x \in \mathbb{R}^d$, $\mathbf{E} = \mathbb{R}^d$. Since the set of functions $\{e^{i<t,\cdot>} : t \in \mathbb{R}^d\}$ is complete in $L^2(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$, we can apply Theorem 2.4.1 and get (2.4.4). □

Remark 2.4.1 1. If W is Gaussian, then so is X, and vice versa.

2. If X is real-valued, the spectral representation (2.4.1) can be written in its real form:

$$X(t) = \int_{\mathbb{R}^{d-1}\times\mathbb{R}_+} \cos < t,x > W_1(dx) + \int_{\mathbb{R}^{d-1}\times\mathbb{R}_+} sin < t,x > W_2(dx), \quad t \in \mathbb{R}^d,$$

where W_1 and W_2 are independent real-valued independently scattered random measures.

3. The random measure W in (2.4.1) is called a *spectral process* corresponding to X.

The proof of 1) is obvious due to the convolution stability of the Gaussian distribution. The proof of 2) is more involved and will not be given here.
How spectral representations (2.4.1) can be used in direct applications?

Example 2.4.1 (Simulation of Gaussian random fields):
In order to simulate a centered second order stationary Gaussian random field X with spectral measure ν within an observation window $A \subset \mathbb{R}^d$, A bounded Borel set, it is enough to consider sums

$$\widetilde{X}_n(t) = \sum_{i=1}^{m_n} e^{i<t,x_i>} W(A_i^n), \quad t \in A, \tag{2.4.5}$$

where

$$A = \bigcup_{i=1}^{m_n} A_i^n \quad \subset \mathbb{R}^d$$

is sufficiently large to approximate

$$X(t) = \int_{\mathbb{R}^d} e^{i<t,x>} W(dx) \quad \text{by} \quad X_A(t) = \int_A e^{i<t,x>} W(dx)$$

in the mean square sense with a desired accuracy $\|X(t) - X_A(t)\|_{L^2(\Omega,\mathcal{F},\mathcal{P})} < \varepsilon$, $A_i^n \cap A_j^n = \emptyset$, $i \neq j$ are bounded Borel sets in $\mathbb{R}^d$, $x_i \in A_i$, $W(A_i^n)$ are independent random variables such that $W(A_i^n) \sim N(0, \nu(A_i^n))$, $i = 1, \ldots, m_n$. Then, $\widetilde{X}_n(t)$ is always a Gaussian random field s.t.

$$\widetilde{X}_n(t) \overset{L^2(\Omega,\mathcal{F},\mathcal{P})}{\longrightarrow} \int_A e^{i<t,x>} W(dx), \quad \text{as} \quad n \to \infty,$$

i.e., the covariance function C_n of $\widetilde{X}_n$ tends to the covariance function C_A of X_A as $n \to \infty$. If A is a cube $[-k;k]^d$, then A_i^n can be chosen to be shifts of $[0, \frac{1}{n}]^d$ and x_i can be taken to be the diagonal crossing point of A_i. If the choice of x_i and t is appropriate, a fast Fourier transformation for the fast computation of the sums in (2.4.5) can be used.

2.5 Orthogonal expansions for random functions

In this section, we would like to prove the following representation of a centered random function $X = \{X(t), t \in T\}$ on a Hausdorff (compact) index space T satisfying some conditions:

$$X(t) = \sum_{i=1}^{\infty} \xi_i \psi_i(t), \tag{2.5.1}$$

where $\{\xi_i\}_{i\in\mathbb{N}}$ are centered uncorrelated random variables (i.e., $\mathbf{E}\xi_i^2 < \infty$, $\mathbf{E}\xi_i = 0$, $\mathbf{E}(\xi_i\xi_j) = \lambda_i\delta_{ij}$, $\lambda_i \geq 0$, $i, j \in \mathbb{N}$) and $\{\psi_i\}_{i\in\mathbb{N}}$ is an orthonormal system of $L^2(T, \nu)$ with ν a finite measure on the Borel-σ-Algebra $\mathcal{B}_T$.
This convergence is understood in $L^2(\Omega, \mathcal{F}, \mathcal{P})$, and it is uniform in t on T.
The representation (2.5.1) is known under the name of Karhunen-Loéve (named after Kari Karhunen and Michel Loéve). In order to prove it, some technical devices are at hand.

2.5.1 Mercer's Theorem

Let T be a compact Hausdorff space equipped with a finite measure ν defined on $\mathcal{B}_T$.
Let a function $K : T \times T \to \mathbb{R}$ be symmetric (i.e., $K(s,t) = K(t,s)$, $s, t \in T$) and positive semi-definite.
Then it is called a *kernel*. Let K be continuous in both arguments. Introduce the operator A_K on $L^2(T, \mathcal{B}_T, \nu)$ by

$$[A_K g](t) = \int_T K(t,s)g(s)\nu(ds), \quad t \in T, \quad g \in L^2(T, \mathcal{B}_T, \nu)$$

The following Theorem goes back to James Mercer (1905), who proved it in the case of $T = [a; b] \subset \mathbb{R}$.

Theorem 2.5.1
If the kernel K is continuous, then A_K has an orthonormal basis of eigenfunctions $\{\psi_i\}_{i\in\mathbb{N}}$ corresponding to eigenvalues $\{\lambda_i\}_{i\in\mathbb{N}}$, $\lambda_i \geq 0$ $\forall i \in \mathbb{N}$ in $L^2(T, \mathcal{B}_T, \nu)$, and

$$K(s,t) = \sum_{j=1}^{\infty} \lambda_j \psi_j(s)\psi_j(t), \quad s, t \in T, \tag{2.5.2}$$

where this convergence is absolute and uniform on T.

The proof of this result can be found in [33]. It is based on the following idea:
it is enough to show that A_K is a compact operator in $L^2(T, \mathcal{B}_T, \nu)$. Then, the spectral theorem by Hilbert-Schmidt for compact operators on Hilbert Spaces is applied.

Corollary 2.5.1
Under the assumptions of Theorem 2.5.1, it holds

$$\int_T K(t,t)\nu(dt) = \sum_{i=1}^{\infty} \lambda_i.$$

Proof The statement follows from (2.5.2) and orthonormality of $\{\psi_i\}_{i\in\mathbb{N}}$. □

Suppose that the eigenvalues $\{\lambda_i\}_{i\in\mathbb{N}}$ of T_K are given in decreasing order:
$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq 0$.
By definition of ψ_j and λ_j, it holds

$$\int_T K(s,t)\psi_j(s)\nu(ds) = \lambda_j \cdot \psi_j(t), \quad t \in T$$

as well as

$$\int_T \psi_i(t)\psi_j(t)\nu(dt) = \delta_{ij}, \quad i,j \in \mathbb{N}$$

Corollary 2.5.2
If $K : T\times T \to \mathbb{R}$ is a covariance function of a (real-valued) random function $X = \{X(t) : t \in T\}$, and K is continuous on T, then it has expansion (2.5.2), because it is a kernel.
It is not easy to get the expansion (2.5.2) analytically, because the problem of computing the eigenfunctions of T_K is not trivial. There are, however, some special cases, where it can be done explicitly.

Example 2.5.1 1. Let $T = [0;1]$, ν be the Lebesgue measure on $[0;1]$, $K(s,t) = \min\{s,t\}$ be the covariance function of the Wiener process $X = \{X(t) : t \in [0;1]\}$.
Let us show that

$$\psi_j(t) = \sqrt{2}\sin(\pi(j+\frac{1}{2})t), \quad t \in [0;1]$$

with

$$\lambda_j = \frac{1}{(j+\frac{1}{2})^2\pi^2}, \quad j \in \mathbb{N}.$$

Since

$$K(s,t) = sI\{s \leq t\} + tI\{s > t\} \tag{2.5.3}$$

is continuous, Theorem 2.5.1 can be applied to it. Let us solve the equation

$$\int_0^1 K(s,t)\psi(s)ds = \lambda \cdot \psi(t).$$

Indeed we have by (2.5.3)

$$\lambda\psi(t) = \int_0^t s\psi(s)ds + t\int_t^1 \psi(s)ds. \tag{2.5.4}$$

Assume that $\psi \in C^2[0;1]$, $\lambda > 0$ and differentiate both sides of the above equation twice. We get the boundary value problem

$$\left\{ \begin{array}{l} \psi''(t) = -\frac{1}{\lambda}\psi(t) \\ \psi(0) = 0 \\ \psi'(1) = 0, \end{array} \right\}$$

because

$$\lambda\psi'(t) = \int_t^1 \psi(s)ds, \quad t \in [0;1] \quad \text{and} \quad \psi'(1) = 0 \quad \text{follows from it.}$$

$\psi(0) = 0$ follows from (2.5.2) with $s = t = 0$. Solving the equation $\psi''(t) = -\frac{1}{\lambda}\psi(t)$, we get $\psi(t) = c \cdot \cos\left(-\frac{t}{\sqrt{\lambda}} + a\right)$, where $\psi(0) = 0 \quad \Rightarrow \quad a = \frac{\pi}{2}$, i.e.

$$\psi(t) = c\sin\left(\frac{t}{\sqrt{\lambda}}\right), \quad \psi'(t) = \frac{c}{\sqrt{\lambda}}\cos\left(\frac{t}{\sqrt{\lambda}}\right),$$

$$\psi'(1) = \frac{c}{\sqrt{\lambda}}\cos\left(\frac{1}{\sqrt{\lambda}}\right) = 0 \quad \Rightarrow \quad \frac{1}{\sqrt{\lambda}} = \pi\left(j + \frac{1}{2}\right), \quad j \in \mathbb{N}_0,$$

thus

$$\psi_j(t) = c_j \cdot \sin\left(\pi\left(j + \frac{1}{2}\right)t\right), \quad j \in \mathbb{N}_0.$$

The constant c_j can be found from

$$\int_0^1 \psi_j^2(t)dt = 1, \quad j \in \mathbb{N}_0.$$

The eigenvalue λ_j can be obtained from (2.5.4).

2. Let $T = [0;1]^d$, ν is the Lebesgue measure on $[0;1]^d$, and

$$K(s,t) = \prod_{i=1}^d \min\{s_i, t_i\}, \quad s = (s_1 \dots s_d)^T, \quad t = (t_1 \dots t_d)^T \in [0;1]^d \tag{2.5.5}$$

be the covariance function of the Brownian sheet (which is a centered Gaussian process with this covariance function). The Brownian sheet can be seen as a multidimensional generalization of the univariate Wiener process. Introduce the multiindex $j = (j_1 \cdots j_d)^T$, $j_i \in \mathbb{N}$.

Exercise 2.5.1
Show that the eigenfunctions of A_K with given K in (2.5.5) are

$$\psi_j(t) = 2^{\frac{d}{2}} \prod_{i=1}^d \sin\left(\pi\left(j_i + \frac{1}{2}\right)t_i\right), \quad j = (j_1 \dots j_d) \in \mathbb{N}^d, \quad t = (t_1 \dots t_d)^T \in [0;1]^d$$

with eigenvalues

$$\lambda_j = \prod_{i=1}^d \frac{1}{(j_i + \frac{1}{2})^2\pi^2}.$$

2.5.2 Reproducing Kernel Hilbert Spaces

Let $X = \{X(t) : t \in T\}$ be a (real-valued) random function defined on $(\Omega, \mathcal{F}, \mathcal{P})$. T being a compact separable Hausdorff space. It is often useful to assume that T is compact w.r.t. the topology induced by the following (pseudo) metric on T:

$$\rho(s,t) = \sqrt{\mathbf{E}|X(s) - X(t)|^2}.$$

It satisfies all properties of a metric except for

$$\rho(s,t) = 0 \quad \Rightarrow \quad s = t.$$

Obviously, there can exist random functions X with $X(s) \stackrel{a.s.}{=} X(t)$, but $s \neq t$ (e. g. if X is periodic). This pseudometric ρ is called a *canonical metric* for X. Let X be centered:
$\mathbf{E}X(t) \equiv 0$, $t \in T$, and $\mathbf{E}X^2(t) < \infty$, $t \in T$.
Let

$$C(s,t) = \mathbf{E}[X(s)X(t)], \quad s,t \in T$$

be the covariance function of X which is positive definite and continuous on $T \times T$. Let us construct the *reproducing kernel Hilbert space* H of C (or X) using the following step-by-step procedure:

1. Introduce the space

$$\mathbf{E} = \{g : T \to \mathbb{R} : g(\cdot) = \sum_{i=1}^{n} a_i C(s_i, \cdot), \quad a_i \in \mathbb{R}, \quad s_i \in T, \quad i = 1, \ldots, n, \quad n \in \mathbb{N}\}$$

with the scalar product

$$< f, g >_H = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j C(s_i, t_j) \tag{2.5.6}$$

on $\mathbf{E}$, where $f, g \in \mathbf{E}$:

$$f(\cdot) = \sum_{i=1}^{n} a_i C(s_i, \cdot), \quad g(\cdot) = \sum_{j=1}^{m} b_j C(t_j, \cdot).$$

Since C is positive definite, it holds

$$< g, g >_H \geq 0, \quad < g, g >_H = 0 \quad \Leftrightarrow \quad g \equiv 0, \quad g \in \mathbf{E},$$

i.e. $< \cdot, \cdot >_H$ is indeed a scalar product.

Lemma 2.5.1
The scalar product $< \cdot, \cdot >_H$ has the so-called *reproducing kernel property* on $\mathbf{E}$, i.e.,

$$< g, C(t, \cdot) >_H = g(t), \quad t \in T, \quad g \in \mathbf{E}.$$

Proof We write

$$< g, C(t,\cdot) >_H = \left\langle \sum_{i=1}^{n} a_i C(s_i,\cdot), C(t,\cdot) \right\rangle_H = \sum_{i=1}^{n} a_i C(s_i,t) = g(t), \quad t \in T$$

for

$$g(\cdot) = \sum_{i=1}^{n} a_i C(s_i,\cdot) \in \mathbf{E}.$$

□

2. Introduce the norm $\|g\|_H = \sqrt{< g, g >_H}$ on $\mathbf{E}$.

Definition 2.5.1
The closure of $\mathbf{E}$ w.r.t. the norm $\|\cdot\|_H$ is called the *reproducing kernel (Hilbert) space*:

$$H = \overline{\mathbf{E}}.$$

Lemma 2.5.2 1. The convergence $g_n \xrightarrow{\|\cdot\|} g$, as $n \to \infty$, $g_n \in \mathbf{E}$, in the norm $\|\cdot\|_H$ implies the pointwise convergence of functional sequences on T:

$$g_n(t) \longrightarrow g(t), \quad \text{as} \quad n \to \infty, \quad t \in T.$$

2. $H = \overline{\mathbf{E}}$ is indeed a separable Hilbert space.

Proof 1. Let $\{g_n\}$ converge to g in $\|\cdot\|_H$, $g_n \in \mathbf{E}$, $n \in \mathbb{N}$, then it is a Cauchy sequence:

$$\forall \varepsilon > 0 \quad \exists N: \quad \forall n, m > N \quad \|g_n - g_m\|_H < \varepsilon.$$

By the *reproducing kernel property* of $< \cdot, \cdot >_H$, we have

$$|g_n(t) - g_m(t)| = | < g_n - g_m, C(t,\cdot) >_H | \le \|g_n - g_m\|_H \cdot \|C(t,\cdot)\|_H = \|g_n - g_m\|_H \cdot \sqrt{C(t,t)}, \quad t \in T$$

using Cauchy-Schwarz inequality and the definition (2.5.6) for $\|C(t,\cdot)\|_H$. Then, $\{g_n(t)\}$, $t \in T$ is a Cauchy sequence converging to $g(t)$.

2. The only property we have to prove here is that H is separable. This property follows from the separability of T and the continuity of C.

□

Exercise 2.5.2
Please prove it!

Remark 2.5.1
Since $< \cdot, \cdot >_H$ is continuous on H and by Lemma 2.3.1, 1) reproducing kernel property extends to the whole space H :

$$< g, C(t, \cdot) >_H = g(t), \quad t \in T, \quad g \in H.$$

Example 2.5.2 (Cameron-Martin space):
Let $T = [0; 1]$, and X be the Wiener process on T with covariance function $C(s, t) = \min\{s, t\}$. Show that the reproducing kernel Hilbert space H of X, is given by

$$H = \left\{ g : [0, 1] \to \mathbb{R} : \exists\, g'(t) \text{ a.e. on } [0; 1],\ g(t) = \int_0^t g'(s)ds \text{ and } \int_0^1 (g'(t))^2 dt < \infty \right\} \quad (2.5.7)$$

with the scalar product

$$< f, g >_H = \int_0^1 f'(t)g'(t)dt.$$

First, if

$$f(t) = \sum_{i=1}^n a_i C(s_i, t) = \sum_{i=1}^n a_i \min\{s_i, t\} \quad \in \mathbf{E},$$

$$g(t) = \sum_{j=1}^m b_j C(t_j, t) = \sum_{j=1}^m b_j \min\{t_j.t\} \quad \in \mathbf{E}$$

then

$$f'(t) = \sum_{i=1}^n a_i (\min\{s_i, t\})' = \sum_{i=1}^n a_i I_{[0;s_i]}(t), \quad t \in [0; 1]$$

$$g'(t) = \sum_{j=1}^m b_j I_{[0,t_j]}(t)$$

and

$$< f, g >_H = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \min\{s_i, t_j\} = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \int_0^1 I_{[0,s_i]}(t) I_{[0;t_j]}(t)dt$$

$$= \int_0^1 \sum_{i=1}^n a_i I_{[0;s_i]}(t) \sum_{j=1}^m b_j I_{[0;t_j]}(t)dt = \int_0^1 f'(t)g'(t)dt.$$

Since $g(t) = \int_0^t g'(s)ds$ for any $g \in H$ as above, we get the reproducing kernel property:

$$< g, C(t, \cdot) >_H = \int_0^1 g'(s) I_{[0,t]}(s)ds = \int_0^t g'(s)ds = g(t), \quad t \in [0; 1].$$

Hence, H defined above is indeed the reproducing kernel Hilbert space of X, cf. the following

Exercise 2.5.3
Show that H given in (2.5.7) is indeed a Hilbert space and show that $H = \overline{E}$ in the norm $\|\cdot\|_H$.

Lemma 2.5.3
Let the covariance function C be continuous and let $\{\varphi_n\}_{n\in\mathbb{N}}$ be an orthonormal basis in the reproducing kernel Hilbert space H for C. Then, each φ_n is a continuous function on T, and

$$C(t,t) = \sum_{n=1}^{\infty} \varphi_n^2(t), \quad t \in T, \tag{2.5.8}$$

whereas the convergence of this series is uniform in $t \in T$.

Proof By the kernel reproducing property and Cauchy-Schwarz-inequality, we get $\forall n \in \mathbb{N}$ and $s, t \in T$

$$\begin{aligned}
&|\varphi_n(s) - \varphi_n(t)| = | < \varphi_n(\cdot), C(s,\cdot) >_H - < \varphi_n(\cdot), C(t,\cdot) >_H | \\
&= | < \varphi_n(\cdot), C(s,\cdot) - C(t,\cdot) >_H | \le \underbrace{\|\varphi_n\|_H}_{=1} \cdot \|C(s,\cdot) - C(t,\cdot)\|_H = \\
&= \sqrt{< C(s,\cdot), C(s,\cdot) >_H - 2 < C(s,\cdot), C(t,\cdot) >_H + < C(t,\cdot), C(t,\cdot) >_H} \\
&= \sqrt{C(s,s) - 2C(s,t) + C(t,t)} \\
&= \sqrt{\mathbf{E}[X(s) - X(t)]^2} = \rho(s,t)
\end{aligned}$$

by relation (1.5.1) on p. 21 and the definition of the canonical metric. Then $\varphi_n(\cdot)$ is continuous on T, if so is C. Expanding $C(t,\cdot)$ in a Fourier series w.r.t.$\{\varphi_n\}_{n\in\mathbb{N}}$ and using the reproducing kernel property, we get

$$C(t,\cdot) = \sum_{n=1}^{\infty} < C(t,\cdot), \varphi_n(\cdot) >_H \cdot \varphi_n(\cdot) = \sum_{n=1}^{\infty} \varphi_n(\cdot) \cdot \varphi_n(t), \quad t \in T \tag{2.5.9}$$

where the series converges in the $\|\cdot\|_H$-norm. For $s = t$, we get the pointwise convergence in the formula (2.5.8) by the reproducing kernel property. Now we have to show that this convergence is uniform in $t \in T$. The convergence of $\sum_{n=1}^{N} \varphi_n^2(t)$ to $C(t,t)$ as $N \to \infty$ is obviously monotone, by Dini's theorem it is uniform in $t \in T$.

□

Remark 2.5.2
Compare formula (2.5.9) with representation (2.5.2) in Mercer's theorem!

2.5.3 Canonical isomorphism

Let $\{X(t) : t \in T\}$ be a centered random real-valued function with a continuous covariance function C introduced in Section 2.5.2. Consider the subspace $\overline{X} \subset L^2(\Omega, \mathcal{F}, \mathcal{P})$, where $\overline{X} = \overline{\text{span}\{X(t), t \in T\}}$ is the closure of the linear hull of X in $L^2(\Omega, \mathcal{F}, \mathcal{P})$ equipped with the scalar product $< \xi, \eta >_{L^2(\Omega,\mathcal{F},\mathcal{P})} = \mathbf{E}(\xi\eta)$. Let H be the reproducing kernel Hilbert space of X. Let us show that H is isomorph to $\overline{X}$. For any

$$g(\cdot) = \sum_{i=1}^{n} a_i C(t_i, \cdot) \quad \in \mathbf{E}$$

define the mapping $\Delta : \mathbf{E} \mapsto \Delta(\mathbf{E}) \subset \overline{X}$, by

$$\Delta(g) = \sum_{i=1}^{n} a_i X(t_i).$$

Evidently, Δ is an isometry:

$$\|g\|_H^2 = \sum_{i,j=1}^{n} a_i a_j C(t_i, t_j) = \|\sum_{i=1}^{n} a_i X(t_i)\|^2_{L^2(\Omega,\mathcal{F},\mathcal{P})} = \|\Delta(g)\|^2_{L^2(\Omega,\mathcal{F},\mathcal{P})}, \quad g \in \mathbf{E}.$$

Consequently, Δ extends by limiting procedure to $H = \overline{\mathbf{E}}$ and the whole of $\overline{X}$.
So constructed, Δ is called the *canonical isomorphism*. If $\{\varphi_n\}_{n\in\mathbb{N}}$ is an orthonormal basis in H, then $\xi_n = \Delta(\varphi_n) \in L^2(\Omega, \mathcal{F}, \mathcal{P})$ yields a sequence $\{\xi_n\}_{n\in\mathbb{N}}$ of uncorrelated random variables with $\mathbf{E}\xi_n = 0$, $\mathbf{E}\xi_n^2 = 1$, which forms an orthonormal basis of $\overline{X}$. In particular, we have $\forall t \in T$

$$X(t) = \sum_{n=1}^{\infty} < X(t), \xi_n >_{L^2(\Omega,\mathcal{F},\mathcal{P})} \xi_n = \sum_{n=1}^{\infty} < C(t,\cdot), \varphi_n >_H \cdot \xi_n = \sum_{n=1}^{\infty} \varphi_n(t) \cdot \xi_n$$

by the isometry property of Δ and the reproducing kernel property of $< \cdot, \cdot >_H$, where the series converges to $X(t)$ in $L^2(\Omega, \mathcal{F}, \mathcal{P})$. We proved the following result:

Proposition 2.5.1
If $\{\varphi_n\}_{n\in\mathbb{N}}$ is an orthonormal basis in H, then there exists a sequence $\{\xi_n\}_{n\in\mathbb{N}}$ of pairwise uncorrelated random variables with $\mathbf{E}\xi_n = 0$, $\mathbf{E}\xi_n^2 = 1$ such that

$$X(t) = \sum_{n=1}^{\infty} \varphi_n(t)\xi_n \tag{2.5.10}$$

in the L^2-sense. Hence, $\xi_n = \Delta(\varphi_n)$, $n \in \mathbb{N}$.

Remark 2.5.3
If X is Gaussian then all elements of $\overline{X}$ are Gaussian random variables as well, and so are $\{\xi_n\}_{n\in\mathbb{N}}$. Hence, we get $\xi_n \sim N(0,1)$, ξ_n independent, $n \in \mathbb{N}$.

Evidently, the choice of random variables ξ_n heavily depends on the basis $\{\varphi_n\}_{n\in\mathbb{N}}$. This is a major difficulty in this theory to find an appropriate basis $\{\varphi_n\}_{n\in\mathbb{N}}$ in H.
Mercer's theorem can help us to find such a basis $\{\varphi_n\}$, which leads to the so-called Karhunen-Loéve expansion of X.

2.5.4 Karhunen-Loéve expansion

By Corollary 2.5.2, let $\{\psi_n\}_{n\in\mathbb{N}}$ be an orthonormal basis of eigenfunctions of the Fredholm operator A_C (given by $A_C g(t) = \int_T C(s,t)g(s)\nu(ds)$, $g \in L^2(T, \mathcal{B}_T, \nu)$) with non-increasing eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3 \cdots > 0$ in $L^2(T, \mathcal{B}_T, \nu)$.

Lemma 2.5.4
The reproducing kernel Hilbert space of C is given by

$$H = \left\{ g : g(t) = \sum_{n=1}^{\infty} a_n \varphi_n(t), \quad t \in T, \quad \sum_{n=1}^{\infty} a_n^2 < \infty, \quad \varphi_n = \sqrt{\lambda_n} \cdot \psi_n, \quad n \in \mathbb{N} \right\}$$

with the orthonormal basis $\{\varphi_n\}_{n\in\mathbb{N}}$ and scalar product

$$< f, g >_H = \sum_{n=1}^{\infty} a_n b_n, \quad f = \sum_{n=1}^{\infty} a_n \varphi_n, \quad g = \sum_{n=1}^{\infty} b_n \varphi_n, \quad f, g \in H.$$

Proof Let us show that $< \cdot, \cdot >_H$ has the reproducing kernel property: by Mercer's theorem, we get

$$< g(\cdot), C(t, \cdot) >_H = \left\langle \sum_{n=1}^{\infty} a_n \varphi_n, \sum_{n=1}^{\infty} \varphi_n(t) \cdot \varphi_n \right\rangle_H = \sum_{n=1}^{\infty} a_n \varphi_n(t) = g(t), \quad t \in T.$$

By Mercer's theorem, $\{\varphi_n\}_{n=1}^{\infty}$ forms a basis in H. By the definition of $< \cdot, \cdot >_H$, it is orthonormal. It remains to show that H is a Hilbert space (trivial exercise!). □

Remark 2.5.4
For

$$g = \underbrace{\sum_{n=1}^{\infty} a_n \varphi_n}_{\sum_{n=1}^{\infty} a_n \sqrt{\lambda_n} \psi_n} \in H,$$

we have $a_n =< g, \varphi_n >_H$. At the same time, we have

$$\sqrt{\lambda_n} \cdot a_n =< g, \psi_n >_{L^2(T, \mathcal{B}_T, \nu)} = \int_T g(t) \psi_n(t) \nu(dt),$$

hence

$$a_n = \sqrt{\lambda_n} < g, \psi_n >_H = \frac{1}{\sqrt{\lambda_n}} \cdot < g, \psi_n >_{L^2(T, \mathcal{B}_T, \nu)} \quad \text{for} \quad \lambda_n > 0.$$

Hence, H consists of $(L^2, \mathcal{B}_T, \nu)$-integrable functions for g with the property

$$\|g\|_H^2 = \sum_{n=1}^{\infty} a_n^2 = \sum_{n=1}^{\infty} \frac{1}{\lambda_n} \left(\int_T g(t) \psi_n(t) \nu(dt) \right)^2 < \infty$$

and is equipped with the inner product

$$< f, g >_H = \sum_{n=1}^{\infty} \frac{1}{\lambda_n} \int_T f(t) \psi_n(t) \nu(dt) \cdot \int_T g(t) \psi_n(t) \nu(dt).$$

Now take $\varphi_n = \sqrt{\lambda_n} \psi_n$ in Proposition 2.5.1 to get the following.

Corollary 2.5.3 (Karhunen-Loéve Expansion):
For any centered $X = \{X(t),\ t \in T\}$ on a compact T with continuous covariance function C, there exists a family $\{\xi_n\}_{n=1}^{\infty}$ of uncorrelated random variables with $\mathbf{E}\xi_n = 0$, $\mathbf{E}\xi_n^2 = 1$ such that

$$X(t) = \sum_{n=1}^{\infty} \sqrt{\lambda_n} \xi_n \psi_n(t), \tag{2.5.11}$$

where $\{\psi_n\}_{n=1}^{\infty}$ is an orthonormal basis of eigenfunctions of the corresponding Fredholm operator with kernel C in $L^2(T, \mathcal{B}_T, \nu)$, λ_n are the eigenvalues corresponding to ψ_n, and the above expansion holds in $L^2(\Omega, \mathcal{F}, \mathcal{P})$.

Remark 2.5.5 1. If $X(.,\omega) \in L_2(T, \mathcal{B}_T, \nu)$, it holds

$$\xi_n = \frac{1}{\sqrt{\lambda_n}} \int_T X(t) \cdot \psi_n(t) \nu(dt), \quad n \in \mathbb{N}$$

Indeed, this formula follows by applying $< \cdot, \psi_n >_{L^2(T,\mathcal{B}_T,\nu)}$ on both sides of (2.5.11) and using the orthonormality of $\{\psi_n\}_{n\in\mathbb{N}}$.

2. Karhunen-Loéve expansions hold also for complex-valued random functions with evident (minor) changes.

Example 2.5.3 (Wiener process):
By Example 2.5.1, 1), the Wiener process $X = \{X(t) : t \in [0,1]\}$ has the Karhunen-Loéve expansion

$$X(t) = \frac{\sqrt{2}}{\pi} \sum_{n=0}^{\infty} \frac{1}{n+\frac{1}{2}} \sin\left(\left[n+\frac{1}{2}\right]\pi t\right) \xi_n, \quad t \in [0,1] \tag{2.5.12}$$

where $\{\xi_n\}_{n=1}^{\infty}$ are i.i.d. $N(0,1)$-random variables. Later on, it will be shown that this convergence holds even a.s. and uniformly on $[0,1]$.

Now, let us prove that:

Theorem 2.5.2
For Gaussian a.s. continuous random functions, the series (2.5.10) converges a.s. uniformly on T, whereas $\{\xi_n\}$ are i.i.d. $N(0,1)$-distributed random variables.

For the proof of this theorem, we need the following:

Lemma 2.5.5 (Itô-Nisio, 1968):
Let $\{Y_n\}_{n\in\mathbb{N}}$ be a sequence of independent symmetric random elements with values in a separable real Banach space B, equipped with the mass topology. For

$$S_n = \sum_{i=1}^{n} Y_i$$

$\{S_n\}_{n\in\mathbb{N}}$ converges a.s. iff there exists a random element S such that

$$F(S_n) \xrightarrow[n\to\infty]{P} F(S)$$

for any $F \in B^*$, where B^* is the dual space to B.

(Without proof).

Proof of Theorem 2.5.2
We know that

$$\sum_{n=1}^{m} \varphi_n(t)\xi_n$$

converges to Gaussian $X(t)$ in L^2-sense, whereas $X(\cdot)$ is a.s. continuous. Take $Y_i = \varphi_i(t)\xi_i$. Since ξ_i are i.i.d., Y_i are independent. For any functional $F \in (L^2(T, \mathcal{B}_T, \nu))^*$, we have to show that

$$F(S_n) \xrightarrow[n\to\infty]{P} F(X(t)).$$

Show that this holds in L^2-sense:

$$\|F(S_n) - F(X(t))\|_{L^2(\Omega,\mathcal{F},\mathcal{P})} \le \|F\| \cdot \|S_n - X(t)\|_{L^2(T,\mathcal{B}_T,\nu)} \overset{n\to\infty}{\longrightarrow} 0.$$

Then applying Lemma 2.5.5, we get that S_n a.s. convergences to some random variables S, and $S = X(t)$. Now let us show that this convergence is uniform. Introduce the Banach space $C(T)$ of all continuous functions on T with sup-norm. Clearly, $X(\cdot)$, $\xi_n\varphi_n(\cdot) \in C(T)$ by Lemma 2.5.3.

Define

$$S_n(\cdot) = \sum_{k=1}^{n} \xi_k \varphi_k(\cdot)$$

and show by Lemma 2.5.5 that

$$S_n(\cdot) \xrightarrow[n\to\infty]{a.s} X(\cdot) \quad \text{in} \quad C(T).$$

It suffices to show that

$$F(S_n(\cdot)) \xrightarrow[n\to\infty]{L^1} F(X(\cdot)) \quad \text{for any} \quad F \in C^*(T).$$

It is known that

$$F(g) = \int_T g d\mu,$$

where μ is a finite signed Borel measure on $(T, \mathcal{B}_T)$. Hence, we get

$$\mathbf{E}|F(S_n(\cdot)) - F(X(\cdot))| = \mathbf{E}\left|\int_T (S_n(t) - X(t))\mu(dt)\right| \le \int_T \mathbf{E}|S_n(t) - X(t)||\mu|(dt)$$

$$\le \int_T \|S_n(t) - X(t)\|_{L^2(\Omega,\mathcal{F},\mathcal{P})}|\mu|(dt) = \int_T \sqrt{\sum_{j=n+1}^{\infty} \varphi_j^2(t)}|\mu|(dt)$$

by Lyapunov's inequality, where $|\mu|$ is the total variation measure of μ. By Lemma 2.5.3,

$$\sum_{j=n+1}^{\infty} \varphi_j^2(t) \xrightarrow[n\to\infty]{} 0$$

uniformly in $t \in T$, hence

$$F(S_n(\cdot)) \xrightarrow[n\to\infty]{P} F(X(\cdot)),$$

and we are done. □

Remark 2.5.6

We stress that the above decompositions hold only for compact T. If T is not compact (say, for stationary complex-valued X) then the formally written "Karhunen-Loéve expansion" may coincide within the spectral representation of X, as the following example shows:
Consider a second-order stationary complex-valued random field $X = \{X(t) : t \in \mathbb{R}^d\}$ with

continuous covariance function $C(s,t) = C_0(s-t)$, $s,t \in \mathbb{R}^d$ such that it has a discrete spectral measure μ in Bochner's theorem:

$$\mu(.) = \sum_{j=1}^{\infty} \delta_{\lambda_j}(.) \cdot \mu_j, \quad \lambda_j \in \mathbb{R}^d, \quad \mu_j > 0, \quad j \in \mathbb{N}.$$

Consider the set of functions $\left\{e^{i<s,\lambda_j>}\right\}_{j\in\mathbb{N}}$ as a "basis" in Mercer's theorem (which formally does not hold here, together with its generalizations to non-compact spaces; cf. [16]). Then, the formally written Karhunen-Loéve expansion of X would yield

$$X(t) = \sum_{j=1}^{\infty} \sqrt{\mu_j}\xi_j e^{i<t,\lambda_j>}, \quad t \in \mathbb{R}^d$$

where $\{\xi_j\}$ are uncorrelated random variables with $\mathbf{E}\xi_j = 0$, $\mathbf{E}\xi_j^2 = 1$, $j \in \mathbb{N}$. This expansion coincides however with the spectral representation (2.4.4) of X with random measure

$$W(.) = \sum_{j=1}^{\infty} \sqrt{\mu_j}\xi_j \delta_{\lambda_j}(.).$$

Remark 2.5.7 (Simulation of Gaussian random functions):
If the Karhunen-Loéve expansion of a Gaussian random function $X = \{X(t) : t \in T\}$ with a.s. continuous realizations on a compact T is known, it can be used to simulate X on T as follows. By Theorem 2.5.2,

$$X(t) \overset{a.s.}{=} \sum_{j=1}^{\infty} \sqrt{\lambda_j}\xi_j\psi_j(t), \quad t \in T$$

where $\{\xi_j\}$ are i.i.d. $N(0,1)$ random variables, $\lambda_j > 0$ and $\psi_j(.)$ some known functions. Then, one can simulate $X(t)$ by taking its approximation

$$\hat{X}_N(t) = \sum_{j=1}^{N} \sqrt{\lambda_j}\xi_j\psi_j(t), \quad t \in T;$$

it holds

$$\hat{X}_N(t) \xrightarrow[N\to\infty]{a.s} X(t)$$

uniformly in $t \in T$. For instance, this method can be used to simulate the Brownian sheet; cf. Exercise 2.5.1. However, the convergence of $\hat{X}_N$ to X may be rather slow, so that the performance of the simulation is poor: indeed, since ξ_j are i.i.d. $N(0,1)$, we get

$$\delta_N(t) = \|X(t) - \hat{X}_N(t)\|^2_{L^2(\Omega,\mathcal{F},\mathcal{P})} = \| \sum_{j=N+1}^{\infty} \sqrt{\lambda_j}\xi_j\psi_j(t)\|^2_{L^2(\Omega,\mathcal{F},\mathcal{P})}$$
$$= \mathbf{E}| \sum_{j=N+1}^{\infty} \sqrt{\lambda_j}\xi_j\psi_j(t)|^2 = \sum_{j=N+1}^{\infty} \lambda_j\psi_j^2(t), \quad t \in T.$$

The rate of convergence of this sum to zero as $N \to \infty$ depends heavily on the choice of $\{\psi_j\}_{j\in\mathbb{N}}$. For instance, in case of the Brownian motion (cf. Example 2.5.3) for $t \in [0,1]$

$$\delta_N(t) = \frac{2}{\pi^2} \sum_{j=N+1}^{\infty} \frac{\sin^2((j+\frac{1}{2})\pi t)}{(j+\frac{1}{2})^2} \leq \frac{2}{\pi^2} \sum_{j=N+1}^{\infty} \frac{1}{j^2} \leq \frac{2}{\pi^2} \int_{N+1}^{\infty} \frac{1}{x^2} dx$$
$$= \frac{2}{\pi^2}(-1)\left.\frac{1}{x}\right|_{N+1}^{\infty} = \frac{2}{\pi^2(N+1)} = 0(N^{-1}), \quad N \to \infty.$$

Exercise 2.5.4
Show that for the Wiener process $X = \{X(t) : t \in [0,1]\}$ the following (alternative to (2.5.12)) Karhunen-Loéve representation holds:

$$X(t) = t \cdot \xi_0 + \sqrt{2} \sum_{n=1}^{\infty} \frac{sin(\pi n t)}{\pi n} \xi_n, \quad t \in [0,1], \tag{2.5.13}$$

where $\{\xi_n\}_{n=0,1,\ldots}$ is an i.i.d. sequence of $N(0,1)$-distributed random variables. To do this, consider the *Brownian bridge* $\varphi = \{\varphi(t) : t \in [0,1]\}$ defined as a centered Gaussian process with covariance function $C(s,t) = \min\{s,t\} - st$ and a.s. continuous paths. It can be proven (please check this!) that

$$\varphi(t) \stackrel{d}{=} X(t) - t \cdot X(1), \quad t \in [0,1],$$

so that $\varphi(0) = \varphi(1) = 0$ a.s. which justifies the name "Bridge". Then,

$$X(t) \stackrel{d}{=} \varphi(t) + t \cdot X(1) = \varphi(t) + t \cdot \xi_0, \quad t \in [0,1],$$

where $\xi_0 \sim N(0,1)$. Use this representation of X to prove (2.5.13) by finding the Karhunen -Loéve representation of φ. For that, show as in Example 2.5.1, 1) that

$$\lambda_j = (\pi^2 j^2)^{-1}, \quad \psi_j(t) = \sqrt{2}\sin(\pi_j t), \quad j \in \mathbb{N}$$

Exercise 2.5.5 (Ornstein-Uhlenbeck-process):
The *Ornstein-Uhlenbeck-process* $X = \{X(t) : t \in \mathbb{R}\}$ is a stationary centered Gaussian process with covariance function

$$C(s,t) = e^{-|s-t|}, \quad s,t \in \mathbb{R}$$

1. Show that
$$X(t) \stackrel{d}{=} e^{-t} w(e^{2t}), \quad t \in \mathbb{R},$$
where $w = \{w(t) : t \geq 0\}$ is a Wiener process.

2. Show that the spectral density f of X coincides with the Cauchy density:
$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R}$$

3. Find the Karhunen-Loéve expansion for X on $[-a,a], a > o$ using representation 1).

2.6 Additional exercises

Exercise 6 Prove that the characteristic function of any random variable is a valid covariance function, i.e. it is positive definite.

Exercise 7 Prove *Pólya's criterion:* If $C : [0;\infty) \to \mathbb{R}$ is such that $C(0) = 1$, $C(t)$ is continuous and convex, and $\lim_{t\to\infty} C(t) = 0$, then C is a positive definite function.

Exercise 8 Are the following functions $C : \mathbb{R} \to \mathbb{R}$ valid covariance functions?

a) $C(t) \equiv 0$

b) $C(t) \equiv 1$

c) *Nugget effect:* $C(t) = \begin{cases} 1, & t = 0 \\ 0, & t \neq 0 \end{cases}$

d) $C(t) = \begin{cases} 1, & t = 0 \\ 1/2, & t \neq 0 \end{cases}$

e) $C(t) = \sin(t)$

f) $C(t) = \begin{cases} \exp(-|t|), & t \in \mathbb{Z} \\ 0, & \text{otherwise} \end{cases}$

Exercise 9 Prove the following statement: A function $C : \mathbb{R} \to \mathbb{C}$ is real, continuous, and positive definite, if and only if it is the cosine transform of measure F on $[0;\infty)$, i.e.

$$C(x) = \int_{[0,\infty)} \cos(xt)\, dF(t) \qquad \text{for all } x \in \mathbb{R}.$$

Compute the spectral density f for the following covariance functions $C : \mathbb{R} \to \mathbb{R}$:

1. $C(x) = \exp(-x^2)$
2. $C(x) = \exp(-|x|)$
3. $C(x) = \begin{cases} 1 - \frac{|x|}{2} & , -2 \leq x \leq 2 \\ 0, & \text{otherwise} \end{cases}$

Exercise 10 Prove that $\cos(x)$ is a valid covariance function on $\mathbb{R}$.

Exercise 11 There are several definitions for a stable random variable. Namely, a random variable X is said to have a stable distribution, if one of the following properties holds:

1. For each $n \in \mathbb{N}$ there exist constants $c_n > 0$, $d_n \in \mathbb{R}$ such that $X_1 + \ldots + X_n \stackrel{d}{=} c_n X + d_n$, where $X_1, \ldots, X_n$ are independent copies of X.

2. There are parameters $\alpha \in (0,2]$, $\sigma \geq 0$, $\beta \in [-1,1]$ and $\mu > 0$ such that the characteristic function $\phi(t) = \mathbf{E}\exp(itX)$ of X has the form

$$\phi(t) = \begin{cases} \exp\left(-\sigma^\alpha |t|^\alpha (1 - i\beta \mathrm{sgn}(t) \tan\frac{\pi\alpha}{2}) + i\mu t\right) & \text{if } \alpha \neq 1 \\ \exp\left(-\sigma|t|(1 + i\beta\frac{2}{\pi}\mathrm{sgn}(t)\log|t|) + i\mu t\right) & \text{if } \alpha = 1. \end{cases}$$

Show that 2) implies 1). (In fact the two definitions are equivalent.)

Exercise 12 *(*The parameters of a stable random variable) According to the upper definition we can denote stable distributions by $S_\alpha(\sigma,\beta,\mu)$. The parameter α is often called the *index of stability* because it can be shown that the norming constants c_n in Exercise 1 are of the form $c_n = n^{1/\alpha}$ with $\alpha \in (0,2]$ (see Feller, 1967).

1. The parameter μ is a *shift parameter*. This is backed by the following property. Let $X \sim S_\alpha(\sigma,\beta,\mu)$ and let $c \in \mathbb{R}$ be a constant. Show that $X + c \sim S_\alpha(\sigma,\beta,\mu+c)$.

2. The parameter σ is called the *scale parameter*. The reason lies in the following: Let $X \sim S_\alpha(\sigma,\beta,\mu)$ and let $c \in \mathbb{R}\setminus\{0\}$. Show that

$$cX \sim \begin{cases} S_\alpha(|c|\sigma, \beta\mathrm{sgn}(c), c\mu) & \text{if } \alpha \neq 1, \\ S_1(|c|\sigma, \beta\mathrm{sgn}(c), c\mu - \frac{2}{\pi}c\log|c|\sigma\beta) & \text{if } \alpha = 1. \end{cases}$$

3. The parameter β is a *skewness parameter*. Show that, for any $0 < \alpha < 2$ it holds

$$X \sim S_\alpha(\sigma,\beta,0) \text{ if and only if } -X \sim S_\alpha(\sigma,-\beta,0).$$

Exercise 13

1. Show that $X \sim S_\alpha(\sigma,\beta,\mu)$ is symmetric about zero if and only if $\beta = 0$ and $\mu = 0$.

2. Show that the random variable $X \sim N(\mu,\sigma^2)$ is stable with $\alpha = 2$.

3. Show that the Cauchy random variable X with probability density function $f_X(x) = \frac{\gamma}{\pi(x^2+\gamma^2)}$, $\gamma > 0$ is stable. What is the value of α?

Exercise 14

1. Find a simple formula for the variogram $\gamma(s,t)$ of a random field $X = \{X(t),\, t \in \mathbb{R}^d\}$ with mean value function $\mu(t)$ and covariance function $C(s,t)$, $s,t \in \mathbb{R}^d$, $d \geq 1$.

2. Now assume that X, defined on a compact $T \subset \mathbb{R}^d$ and $\mathbf{diam}T := \sup_{s,t\in T} d(s,t) < \infty$ with the canonical pseudo-metric $d(s,t) := \sqrt{2\gamma(s,t)} < \infty$. The smoothness of the random field X is closely related to the behavior of $\gamma(s,t)$ for points $s,t \in \mathbb{R}^d$ with infinitesimal small distance. We now try to find a sufficient condition for the a.s. continuity of the field X. It is known that Gaussian random fields are a.s. continuous and bounded with probability one if there exists a $\delta > 0$ such that

$$\int_\delta^\infty p(e^{-u^2})\, du < \infty,$$

where $p^2(u) = \sup_{|s-t|\le u} 2\gamma(s,t)$. Show that the existence of a constant $K \in (0,\infty)$ and $\alpha, \eta > 0$ such that

$$\gamma(s,t) \le \frac{K}{|\log|s-t||^{1+\alpha}}$$

for all s, t with $|s-t| < \eta$ is a sufficient condition for the a.s. continuity of X.

Exercise 15 Let ν be a finite measure, given on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $d \ge 1$. Let $Y, X_1, X_2, \dots$ be independent random elements on a probability space $(\Omega, \mathcal{F}, ¶)$ such that $Y \sim \mathrm{P}oi(\nu(\mathbb{R}^d))$ and $X_1, X_2, \dots$ i.i.d. d-dimensional $\mathcal{F}|\mathcal{B}(\mathbb{R}^d)$-measurable random vectors such that $¶_{X_1}(B) = \nu(B)/\nu(\mathbb{R}^d)$, $B \in \mathcal{B}(\mathbb{R}^d)$. We introduce the *Poisson random measure* $W : \mathcal{B}(\mathbb{R}^d) \times \Omega$ by

$$W(B,\omega) = \sum_{j=1}^{Y(\omega)} \mathbf{1}_B(X_j(\omega)), \quad B \in \mathcal{B}(\mathbb{R}^d).$$

Show that the Poisson random measure given on a semiring of bounded Borel sets in $\mathbb{R}^d$ is an orthogonal non-centered random measure.

Exercise 16 Let $\{W_t,\, t \in \mathbb{R}\}$ be a complex-valued L^2-process such that
i) $\mathbf{E}|W_s - W_t|^2 \to 0$ for any $s \in \mathbb{R}$ with $s \downarrow t$,
ii) it has independent increments, i.e. $\mathbf{E}(W_{t_2} - W_{t_1})\overline{(W_{t_3} - W_{t_2})} = 0$ for any $t_1 < t_2 < t_3$.
We introduce the family of random variables $W((a,b]) := W(b) - W(a)$ on the semiring $\mathcal{K} = \{(a,b],\, -\infty < a \le b < \infty\}$, where $(a,a] = \emptyset$. Show that W is an orthogonal random measure on $\mathcal{K}$.

Exercise 17 Let E be a measurable space, and ν a σ-finite measure given on a semiring $\mathcal{K}(E)$ of subsets of E. Show that simple functions of the form $f : E \to \mathbb{C}$, $f = \sum_{i=1}^m c_i \mathbf{1}_{B_i}$, where $c_i \in \mathbb{C}$, $B_i \in \mathcal{K}(E)$, $i = 1, \dots, m$, $\bigcup_{i=1}^m B_i = E$, $B_i \cap B_j = \emptyset$ for $i \ne j$, are dense in $L^2(E,\nu)$.

Exercise 18 Show Lemma 2.3.4 4): Let E be a measurable space, and ν a σ-finite measure given on a semiring $\mathcal{K}(E)$ of subsets of E. Let W be a centered orthogonal random measure, i.e. $\mathbf{E}W(A) = 0$, $A \in \mathcal{K}(E)$. Show that $\mathbf{E}I(f) = 0$ for $f \in L^2(E,\nu)$.

Exercise 19 *An illustrative example of the Karhunen Theorem (1)*

1. Show that the Wiener process $W = \{W(t),\, t \in [0, 2\pi]\}$ has the representation

$$W(t) = \frac{1}{\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} \frac{1 - e^{-ikt}}{ik} z_k, \quad t \in [0, 2\pi],$$

where the z_k are uncorrelated centered random variables with unit variance and the series converges in the mean-square sense for every $t \in [0, 2\pi]$
(for $k = 0$ we set $(1 - e^{-ikt})/ik = -t$). Use the following steps:

Step 1: Determine the values $f(t,k)$, $t \in [0, 2\pi]$, $k \in \mathbb{Z}$ of the function f in the representation of the covariance function of the stochastic process, which are given as the coefficients of the Fourier series of $\mathbf{1}_{[0,t]}(u)$.

Step 2: Determine the value of the covariance function $C(s,t)$, $s, t \in [0, 2\pi]$, which arises by taking the space $E = \mathbb{Z}$, the counting measure ν on $\mathbb{Z}$, i.e. $\nu(\{k\}) = 1$, $k \in \mathbb{Z}$ and the function f from step 1. *Hint: Apply the Parseval equality.*

Step 3: Apply the Karhunen Theorem.

Exercise 20 *An illustrative example of the Karhunen Theorem (2)*
Show that the Wiener process $W = \{W(t),\ t \in [0,1]\}$ has the representation

$$W(t) = \sum_{k=1}^{\infty} S_k(t) z_k, \quad t \in [0,1],$$

where the $S_k(t)$, $t \in [0,1]$, $k \geq 1$ are the *Schauder functions* and the z_k are uncorrelated centered random variables with unit variance and the series converges in the mean-square sense for every $t \in [0,1]$. Use the same steps as in the preceding exercise.

Exercise 21
Let W be a Gaussian white noise based on Lebesgue measure, and use it to define a random field on $\mathbb{R}_+^d = \{(t_1, \ldots, t_d) : t_i \geq 0\}$ by setting

$$W(t) = W([0,t]),$$

where $[0,t]$ is the rectangle $\prod_{i=1}^d [0,t_i]$. W_t is called the Brownian sheet on $\mathbb{R}_+^d$, or *multiparameter Brownian motion.* If $d = 1$, it is the standard Brownian motion.

(a) Show that W is a centered Gaussian field on $\mathbb{R}_+^d$ with covariance

$$\mathbf{E}(W_s W_t) = \min(s_1, t_1) \times \cdots \times \min(s_d, t_d).$$

(b) Suppose $d > 1$, and fix $d-k$ of the index variables t_i. Show that W is a scaled k-parameter Brownian sheet in the remaining variables.

(c) Find the Karhunen-Loève expansion for W on $[0,1]^d$.

Exercise 22
Let $T \in \mathbb{N}$ and $X = \{X(t),\ t \in [0,T]\}$ be a real-valued process on $[0,T]$ with $\mathbf{E}X(t) = 0$ and $\mathbf{E}X(t)^2 < \infty$ for all $t \in [0,T]$.

(a) Suppose that $C(s,t) = \cos(2\pi(t-s))$. Show that the Karhunen-Loève expansion has only two summands and determine the terms of the expansion.

(b) Suppose that $C(s,t) = (1 - |t-s|)\mathbf{1}\{0 \leq t-s \leq 2T\}$. Determine the Karhunen-Loève expansion.

(c) Suppose that $C(s,t) = \sum_{n=0}^{\infty} \frac{1}{1+n^2} \cos\left(n \frac{2\pi}{T}(t-s)\right)$. Find the eigenvalues and eigenfunctions of the Fredholm operator A from Exercise 1.

Exercise 23 Let $\{X(t),\ t \in [0,1]\}$ be the Brownian motion with covariance function $C(s,t) = \min\{s,t\}$ for $s,t \in [0,1]$. Show that

$$\mathcal{H} = \{g : [0,1] \to \mathbb{R} \text{ cont.} : \exists\, g'(t) \text{ for almost all } t \in [0,1],\ g(s) = \int_0^s g'(t)\,dt,\ \int_0^1 (g'(t))^2\,dt < \infty\}$$

is the corresponding reproducing kernel Hilbert space.

Exercise 24 Let $X = \{X_t,\ t \in \mathbb{R}\}$ be a random polynomial with $X_t = Y_0 + Y_1 t + \ldots + Y_n t^n$, $Y_i \sim \mathcal{N}(0,1)$ i.i.d., $i = 0, \ldots, n$. Determine the expected value, the variance, the covariance function and the characteristic function of X_t.

Exercise 25

(a) Calculate the spectral density of a random process $X = \{X_t,\, t \in \mathbb{R}\}$ with
- triangular covariance function $C(h) = (1 - |h|)\mathbf{1}\{|h| \le 1\}$,
- covariance function $C(h) = \frac{1}{8}e^{-|h|}\,(3\cos(h) + \sin|h|)$.

(b) Show that the spectral density of the Ornstein-Uhlenbeck process with covariance function $C(s,t) = \exp\{-\alpha|t-s|\}$, $s,t \in \mathbb{R}$, $\alpha > 0$, is given by

$$f(u) = \frac{1}{\pi}\frac{\alpha}{1+\alpha^2u^2}, \quad u \in \mathbb{R},$$

and verify that it can be interpreted as a Gamma(1/2)-mixture of Gaussian distributions, i. e.

$$f(u) = \int_0^\infty \left(\frac{t\alpha^2}{\pi}\right)^{\frac{1}{2}} e^{-t\alpha^2u^2}\frac{1}{\sqrt{\pi}}e^{-t}t^{\frac{1}{2}-1}dt, \quad u \in \mathbb{R}.$$

(c) Calculate the spectral density of a random field $X = \{X_t,\, t \in \mathbb{R}^2\}$ with Gaussian covariance function $C(h) = e^{-\|h\|_2^2}$.

Exercise 26
Let T be a separable and compact Hausdorff space. Let $\mathcal{H}$ be a Hilbert space of functions $g : T \to \mathbb{R}$ which admits a reproducing kernel. Show that the reproducing kernel is uniquely determined by the Hilbert space $\mathcal{H}$.

Exercise 27
Show that under the same assumptions of Exercise 3, the reproducing kernel of a reproducing kernel Hilbert space is positive semi-definite.

Exercise 28
A stochastic process $\{B(t), t \in [0,1]\}$ is called a *Brownian bridge* if

- the joint distribution of $B(t_1), B(t_2), ..., B(t_n)$, $t_1, ..., t_n \in [0,1]$, $n \in \mathbb{N}$ is Gaussian with $\mathbf{E}B(t) = 0$ for all $t \in [0,1]$,
- the covariance function of $B(t)$ is given by

$$C(s,t) = \min\{s,t\} - st,$$

- the sample path function of $B(t,w)$ is continuous in t with probability one.

(a) Find the Karhunen-Loève expansion of the Brownian bridge.

(b) Show that $B(t) = W(t) - tW(1) = (1-t)W\left(\frac{t}{1-t}\right)$, $t \in [0,1]$, where $\{W(t),\, t \in [0,1]\}$ is the Brownian motion.

(c) Find a representation of the Brownian bridge based on the Karhunen-Loève expansion of the Brownian motion. Compare this representation with the Karhunen-Loève expansion of the Brownian bridge.

Exercise 29
A stochastic process $\{X(t),\, t \ge 0\}$ is called *Ornstein-Uhlenbeck process* if it is a centered Gaussian process with covariance function $C(s,t) = \exp\{-\alpha|t-s|\}$, $s,t \ge 0$, $\alpha > 0$. Show that X is a transformed Brownian motion and determine the corresponding (transformed) Karhunen-Loève expansion.

3 Models of random fields

Starting with a more detailed treatment of the basics of Gaussian random functions, we shall introduce more general classes of random functions such as stable, infinitely divisible and self-similar (fractal) random functions.

3.1 Gaussian random fields

In Definition 1.2.2, Gaussian random functions were introduced by giving all their finite-dimensional distributions, which are multivariate normal ones. It has been shown that the probability law of Gaussian random functions is fixed by their mean value and covariance functions. In this section, we are going to consider some of the properties of Gaussian random functions such as a.s. boundedness, a.s. continuity of realisations, differentiability, etc. in more detail. However, we shall not be able to cover many important aspects of their theory, on which excellent specialized monographs already exist (cf. [1], [15], [37], [44]).
First of all, consider the properties of paths of Gaussian random functions. Then, let us deepen some of the examples of Gaussian random functions which we already encountered in Chapters 1 and 2.

3.1.1 Properties of paths of Gaussian random functions

Let $X = \{X(t) : t \in T\}$ be a real-valued random function on a topological space T.

Definition 3.1.1
X is called *separable*, if there exists a countable dense subset $D \subset T$ and a fixed event A with $P(A) = 0$ such that

$$\{\omega \in \Omega : X(t) \in B, t \in I \cap D\} \setminus \{\omega \in \Omega : X(t) \in B, t \in I\} \subset A$$

for any closed subset $B \subset \mathbb{R}$ and open $I \subset T$.

It can be proven (cf. [43]) that any random function X with index space T being a metrical space and values in a compact set $K \subset \mathbb{R}$ has a separable modification. Later on, we assume that X be always separable. Separability is introduced in order to study the properties of $\sup_{t \in I} X(t)$, the set I being an open subset of T. Indeed, we first need to prove that $\sup_{t \in I} X(t)$ is a random variable, i.e. is $\mathcal{F}|\mathcal{B}_{\mathbb{R}}$-measurable. This holds true, since

$$\sup_{t \in I} X(t) \stackrel{a.s.}{=} \sup_{t \in I \cap D} X(t)$$

due to separability of X, and $\sup_{t \in I \cap D}$ is measurable, since this supremum is taken over a countable set of random variables $X(t)$, $t \in I \cap D$. Now assume that T is a compact metric space with distance function Δ on T. Then, for separable X, $\sup_{t \in T} X(t) \stackrel{a.s.}{=} \sup_{t \in D} X(t)$ is a valid random

variable. Let X be a (real-valued separable) Gaussian random function on T. Considering a.s. boundedness and continuity of X, first let us show that the choice of the metric Δ is irrelevant, since it can be always replaced by the canonical metric

$$\rho(s,t) = \sqrt{\mathbf{E}(X(s) - X(t))^2}, \quad s,t \in T.$$

Lemma 3.1.1
Let X be a random function as above with $\sup\limits_{t\in T} \mathbf{E}X^2(t) < \infty$, which is a.s. continuous:

$$\forall t \in T, \quad \lim_{s \overset{\Delta}{\to} t} X(s) = X(t) \quad \text{a.s.}$$

Then, X is mean square continuous as well. In other words, ρ is a continuous function.

$$\lim_{s \overset{\Delta}{\to} t} \rho(s,t) = 0, \quad t \in T.$$

Hence $s \overset{\Delta}{\to} t$ means that $\rho(s,t) \to 0$.

To prove this lemma, we need the following two facts from the basic probability course:

Definition 3.1.2
A family of random variables $\{\xi_n\}_{n\in\mathbb{N}}$ is uniformly integrable, if

$$\sup_{n\in\mathbb{N}} \mathbf{E}\left[|\xi_n| I(|\xi_n| > x)\right] \xrightarrow[x\to\infty]{} 0.$$

Lemma 3.1.2
A family of random variables $\{\xi_n\}_{n\in\mathbb{N}}$ is uniformly integrable iff

1. Uniform boundedness:
$$\sup_{n\in\mathbb{N}} \mathbf{E}|\xi_n| < \infty$$
2. Uniform absolute continuity:
$$\sup_{n\in\mathbb{N}} \mathbf{E}\left[|\xi_n| I(A_m)\right] \to 0,$$
if $P(A_m) \to 0$

Without Proof; see e.g. [9], p. 190.

Theorem 3.1.1
If $\{\xi_n\}_{n\in\mathbb{N}}$ is uniformly integrable and

$$\xi_n \xrightarrow[n\to\infty]{a.s.} \xi, \quad \text{then} \quad \mathbf{E}|\xi| < \infty \quad \text{and} \quad \mathbf{E}\xi_n \xrightarrow[n\to\infty]{} \mathbf{E}\xi, \quad \mathbf{E}|\xi_n - \xi| \xrightarrow[n\to\infty]{} 0.$$

Without Proof; see [9], p. 188.

Proof of Lemma 3.1.1 Fix a $t \in T$. Introduce a random field $Y(s) = (X(s) - X(t))^2$, $s \in T$. It holds

$$\sup_{s\in T} \mathbf{E}Y(s) \le 4 \sup_{s\in T} \mathbf{E}X^2(s) < \infty.$$

To show that $\{Y(s) : s \in T\}$ is uniformly integrable, it is sufficient (by Lemma 3.1.2) to prove that

$$\sup_{s\in T} \mathbf{E}\left[Y(s)I(A_m)\right] \to 0, \quad \text{if} \quad P(A_m) \xrightarrow[m\to\infty]{} 0.$$

By the inequality of Cauchy-Schwarz, we have

$$\sup_{s\in T} \mathbf{E}[Y(s)I(A_m)] \le \sup_{s\in T} \sqrt{EY^2(s)} P(A_m) \to 0,$$

since

$$\sup_{s\in T} \sqrt{\mathbf{E}Y^2(s)} < \infty$$

due to the Gaussianity of X. Then, $\{Y(s) : s \in T\}$ is uniformly integrable and then by Theorem 3.1.1 it holds

$$\mathbf{E}Y(s) \xrightarrow[s\to t]{} 0 \ \text{ since } \ Y(s) \xrightarrow[s\to t]{a.s.} 0;$$

□

In the sequel, assume that ρ is always continuous with respect to the Δ-convergence.

Lemma 3.1.3
Let X be a centered separable real-valued Gaussian process on a compact metric space (T, Δ). Then X is a.s. continuous w.r.t. the topology induced by Δ, iff it is a.s. continuous w.r.t. the topology induced by ρ.

Proof We have to show that

$$\lim_{s,\Delta(s,t)\to 0} |X(s) - X(t)| \overset{a.s.}{=} 0 \quad \Leftrightarrow \lim_{s:\rho(s,t)\to 0} |X(s) - X(t)| = 0, \quad t \in T.$$

Fix any $t \in T$. Since ρ is Δ-continuous, we have

$$\Delta(s,t) \to 0 \quad \Rightarrow \rho(s,t) \to 0$$

and then

$$\lim_{s:\rho(s,t)\to 0} |X(s) - X(t)| \overset{a.s.}{=} 0 \Rightarrow \lim_{s:\Delta(s,t)\to 0} |X(s) - X(t)| \overset{a.s.}{=} 0.$$

Let us prove the reverse statement. Let X be a.s. Δ-continuous. For any $\delta > 0$, introduce

$$A_\delta = \{(s,t) \in T \times T : \rho(s,t) \le \delta\} \subset T \times T.$$

Since T is compact, so is $T \times T$. Since ρ is Δ-continuous A_δ is Δ-closed in $T \times T$ (and hence a compact in $T \times T$). Take

$$A_0 = \bigcap_{\delta>0} A_\delta.$$

Then A_0 is compact as well, and by the properties of compactness $\forall \varepsilon > 0$ $\exists$ a finite ε-network of balls with centers in A_0 covering A_0:

$$A_0 \subset \bigcup_{(s',t')\in B} \{(s,t) \in T \times T : \Delta(s,s') \le \varepsilon, \quad \Delta(t,t') \le \varepsilon\}$$

for a finite subset $B \subset A_0$. By Δ-continuity of ρ, there exists $\delta(\varepsilon) \to 0$ such that $A_{\delta(\varepsilon)}$ is covered by this finite ε-network as well, i.e., for any

$$(s,t) \in A_{\delta(\varepsilon)} \quad \exists \quad (s',t') \in B : \quad \Delta(s,s') \le \varepsilon, \quad \Delta(t,t') \le \varepsilon.$$

Since

$$|X(s) - X(t)| \le |X(s) - X(s')| + \underbrace{|X(s') - X(t')|}_{\overset{*}{=}0} + |X(t') - X(t)|$$

*a.s., since $s', t' \in B \subset A_0$
we get

$$\sup_{s,t\in T:\rho(s,t)\le\delta(\varepsilon)} |X(s) - X(t)| \le 2 \sup_{s,t\in T:\Delta(s,t)\le\varepsilon} |X(s) - X(t)|,$$

hence X is a.s. ρ-continuous as well. □

Remark 3.1.1
The assertion of this lemma is obviously true for σ-compact metric spaces T.

Definition 3.1.3
Let X be a centered Gaussian random field on a ρ-compact T. For any $\varepsilon > 0$, let $N(\varepsilon)$ be the smallest number of ρ-balls of radius ε covering $T : N(\varepsilon) = N(T, \rho, \varepsilon)$. Then $N(\varepsilon)$ is called the *(metric) entropy* of T (or X). $H(\varepsilon) = \log N(\varepsilon)$ is *a log-entropy* of T. Since T is ρ-compact, it follows that $N(\varepsilon) < \infty$, $\varepsilon > 0$. Introduce

$$\mathbf{diam}_\rho(T) = \sup_{s,t\in T} \rho(s,t).$$

For $\varepsilon \ge \mathbf{diam}_\rho(T)$, we obviously have $N(\varepsilon) = 1$. Introduce the modulus of continuity of X by

$$\omega_{X,\rho}(\delta) = \sup_{s,t\in T:\rho(s,t)\le\delta} |X(s) - X(t)|, \quad \delta > 0.$$

We would like to state the following results without proof. For proofs, see [1], p.14 ff.

Theorem 3.1.2
Let $X = \{X(t) : t \in T\}$ be a centered separable Gaussian random field on a ρ-compact T with the metric entropy $H(\varepsilon)$. Then,

$$\mathbf{E} \sup_{t\in T} X(t) \le K \cdot \int_0^{\frac{\mathbf{diam}(T)}{2}} \sqrt{H(\varepsilon)}\, d\varepsilon,$$

where $K \ge 0$ is a constant.

Theorem 3.1.3
Under conditions of Theorem 3.1.2 there exists a random variable η such that

$$\omega_{X,\rho}(\delta) \le K \cdot \int_0^\delta \sqrt{H(\varepsilon)}\, d\varepsilon$$

for all $\delta \le \eta$, where $K \ge 0$ is a constant.

Theorem 3.1.4
Let X be a centered Gaussian random field with

$$P(\sup_{t\in T} X(t) < \infty) = 1.$$

X is a.s. uniformly ρ-continuous on T iff $\varphi(\delta) \to 0$ as $\delta \to 0$ with

$$\varphi(\delta) = \mathbf{E} \sup_{\rho(s,t)\le\delta} (X(s) - X(t)).$$

Furthermore, if $\varphi(\delta) \to 0$ as $\delta \to 0$, then for all $\varepsilon > 0$ $\exists$ an a.s. finite random variable $\eta > 0$ such that

$$\omega_{X,\rho}(\varepsilon) \le \varphi(\delta) \cdot |\ln \varphi(\delta)|^\varepsilon,$$

for all $\delta \le \eta$.

Now consider the special case of T being a compact subset of $\mathbb{R}^d$. Introduce

$$p(u) = \sqrt{\sup_{|s-t|\le u} 2\gamma(s,t)} = \sqrt{\sup_{|s-t|\le u} \mathbf{E}|X(s) - X(t)|^2},$$

where $|.|$ is the Euclidean metric. For stationary X, it holds

$$p(u) = \sqrt{2 \sup_{|t|\le u} |C(0) - C(t)|}.$$

Let $\omega_X(\delta)$ be the modulus of continuity of a random field X with respect to the Euclidean norm.

Corollary 3.1.1
If either

$$\int_0^\delta \sqrt{-\log u}\, dp(u) < \infty \quad \text{or} \quad \int_\delta^\infty p(e^{-u^2})\, du < \infty$$

for some $\delta > 0$ then the centered Gaussian random field X on T is a.s. continuous and bounded, i.e., $P(\sup_{t\in T} X(t) < \infty) = 1$. Furthermore, there exists a constant $K = K(d)$ and a random variable η such that

$$\omega_X(\delta) \le K \cdot \int_0^{p(\delta)} \sqrt{-\log u}\, dp(u)$$

for all $\delta < \eta$.

Proof : Let us show that our assertion follows from Theorem 3.1.3, if we prove that □

$$\int_0^\delta \sqrt{H(\varepsilon)}\, d\varepsilon \le K \cdot \int_0^{p(\delta)} \sqrt{-\log u}\, dp(u)$$

for small δ. First of all, since $p(u)$ is non-decreasing, the Riemann integral

$$\int_0^\delta \sqrt{-\log u}\, dp(u)$$

is well-defined.
Let L be the side length of the smallest cube c_L containing a compact T. Since

$$p(u) = \sqrt{\sup_{|s-t|\le u} \rho^2(s,t)} = \sup_{|s-t|\le u} \rho(s,t),$$

c_L (and hence T) can be covered by at least $\left(\frac{L}{\Delta}\right)^d$ Euclidean balls of radius $\Delta > 0$;
if $p(\Delta) = \varepsilon \Rightarrow \Delta = p^{-1}(\varepsilon)$ where $p^{-1}(\varepsilon) = \sup\{u : p(u) \ge \varepsilon\}$ is a generalised inverse of a non-decreasing function p. Hence,

$$N(\varepsilon) \le \left(\frac{L}{\Delta}\right)^d = \left(\frac{L}{p^{-1}(\varepsilon)}\right)^d \Rightarrow H(\varepsilon) = \log N(\varepsilon) \le d \log\left(\frac{L}{p^{-1}(\varepsilon)}\right)$$

and

$$\int_0^\delta \sqrt{H(\varepsilon)}\, d\varepsilon \le \sqrt{d} \int_0^\delta \sqrt{\log L - \log p^{-1}(\varepsilon)} d\varepsilon \overset{u=p^{-1}(\varepsilon),\varepsilon=p(u)}{=} \sqrt{d} \int_0^{p(\delta)} \sqrt{\log L - \log u}\, dp(u)$$

$$\le \underbrace{2\sqrt{d}}_{K} \int_0^{p(\delta)} \sqrt{-\log u} dp(u), \tag{3.1.1}$$

for small enough δ because

$$\exists \delta > 0 : \quad \log L \le -3 \log u \quad for \quad |u| \le p(\delta),$$

and hence

$$\sqrt{\log L - \log u} \le \sqrt{-4 \log u} = 2\sqrt{-\log u}.$$

Remark 3.1.2
A sufficient condition for the integrals in Corollary 3.1.1 to be finite is

$$\rho(s,t) \le \frac{c}{|\log|s-t||^{\frac{1}{2}+\alpha}}, \quad c, \alpha > 0, \quad \forall s,t : |s-t| < \beta.$$

Proof :
If

$$\rho(s,t) \le \frac{c}{|\log|s-t||^{\frac{1}{2}+\alpha}}, \quad s,t : |s-t| \le \beta,$$

then

$$p(u) \le \frac{c}{|\log u|^{\frac{1}{2}+\alpha}}, \quad 0 \le u \le \beta,$$

and

$$\int_0^\delta \sqrt{-\log u}\; dp\;(u) = p(u)\cdot\sqrt{-\log u}\Big|_0^\delta - \int_0^\delta p(u)d(-\log u)^{\frac{1}{2}} \overset{p(0)=0}{=} p(\delta)\sqrt{-\log\delta}$$
$$+\int_0^\delta p(u)\frac{1}{2}\frac{1}{\sqrt{-\log u}}d\log u \le p(\delta)\sqrt{-\log\delta} + \int_0^\delta \frac{c\cdot d\log u}{|\log u|^{1+\alpha}}$$
$$= p(\delta)\sqrt{-\log\delta} + c\int_{-\infty}^{\log\delta}\frac{dx}{|x|^{1+\alpha}} < \infty \text{ for } \delta \le \beta.$$

□

Exercise 3.1.1
Show that conditions of Corollary 3.1.1 are satisfied for the Brownian motion $X = \{X(t) : t \in [0;1]\}$.

The above necessary conditions are quite sharp in the stationary case, as it can bee seen from the following.

Proposition 3.1.1
If $X = \{X(t) : t \in T\}$ is a stationary centered Gaussian random field, $T \subset \mathbb{R}^d$ is open and for the covariance function C of X the inequality

$$\frac{K_1}{(-\log|t|)^{1+\alpha_1}} \le C(0) - C(t) \le \frac{K_2}{(-\log|t|)^{1+\alpha_2}}$$

holds for $|t|$ small enough. Then X has a.s. continuous realizations if $\alpha_2 > 0$, and a.s. discontinuous realizations, if $\alpha_1 < 0$.
(Without proof; see the proof in [1], p. 47-48)

Conditions for a.s. continuity can be also formulated in terms of the spectral measure of X:

Proposition 3.1.2
Let $X = \{X(t) : t \in T\}$ be a stationary, centered Gaussian process on a compact $T \subset \mathbb{R}$ with spectral measure μ. If

$$\int_0^\infty (\log(1+\lambda))^{1+\alpha}\mu(d\lambda) < \infty$$

for some $\alpha > 0$ then X has a.s. continuous paths. If this integral diverges for some $\alpha < 0$ then X is a.s. discontinuous.
(Without proof; see [4])

Now suppose that X is a centered Gaussian random field on $T \subset \mathbb{R}^d$ and examine the question of a.s. differentiability of its realisations.

Definition 3.1.4
Let $t \in \mathbb{R}^d$. For any $k \in \mathbb{N}$ and directions $h_1, \ldots, h_k \in \mathbb{R}^d$, introduce the L^2-partial derivative of X in t in direction $h = (h_1, \ldots, h_k)$ (or mean-square partial derivative) of order k by

$$\frac{\partial^k X(t)}{\partial h} = \lim_{\delta_1,\ldots,\delta_k\to 0}\frac{\Delta^k X(t, \langle\delta, h\rangle)}{\delta_1\cdot\ldots\cdot\delta_k},$$

where $\delta = (\delta_1, \ldots, \delta_k) \in \mathbb{R}^k$, $\langle\delta, h\rangle = \sum_{i=1}^k \delta_i h_i \in \mathbb{R}^d$,

$$\Delta^k X(t,y) = \sum_{s\in\{0,1\}^k,\ s=(s_1,\dots,s_k)} (-1)^{k-\sum_{i=1}^k s_i} X(t+\sum_{i=1}^k s_i y_i), \quad y=(y_1,\dots,y_k),$$

is the symmetrical difference of X and the limit is understood sequentially in the mean square sense: first $\delta_1 \to 0$, then $\delta_2 \to 0$ and so on.

Furthermore:

Lemma 3.1.4 1. A random field X is mean square differentiable of order k in all directions in a region $T \subset \mathbb{R}^d$, if

$$\lim_{\delta_1,\dots,\delta_k,\delta_1',\dots,\delta_k'\to 0} \frac{1}{\delta_1\dots\delta_k\cdot\delta_1'\dots\delta_k'}\mathbf{E}(\Delta^k X(t,\sum_{i=1}^k \delta_i h_i)\cdot\Delta^k X(s,\sum_{i=1}^k \delta_i' h_i')) \tag{3.1.2}$$

exists for all $s,t\in T$ and directions $h,h'\in(\mathbb{R}^d)^k$.

2. If X is Gaussian then $\frac{\partial^k X}{\partial h}$ is Gaussian as well for any $h\in(\mathbb{R}^d)^k$.

Proof Property 2 is obvious since a L^2-limit of Gaussian random variables is Gaussian.
1. For any random variables $\{Y_n\}_{n\in\mathbb{N}}$ with $\mathbf{E}Y_n^2<\infty$, $n\in\mathbb{N}$ it holds that $\{Y_n\}$ converges in L^2-sense as $n\to\infty$ iff $\mathbf{E}(Y_nY_m)$ converges as $n,m\to\infty$. Indeed,

$$\mathbf{E}|Y_n-Y_m|^2 \xrightarrow[n,m\to\infty]{} 0 \quad \text{iff} \quad \mathbf{E}(Y_n\cdot Y_m)\xrightarrow[n,m\to\infty]{} c,$$

since

$$\mathbf{E}|Y_n-Y_m|^2 = \mathbf{E}Y_n^2 - 2\mathbf{E}(Y_n\cdot Y_m)+\mathbf{E}Y_m^2, \quad \forall n,m\in\mathbb{N}.$$

Then

$$\frac{\Delta^k X(t,(\delta_1h_1,\dots,\delta_kh_k))}{\delta_1\cdot\ldots\cdot\delta_k}$$

converges to a limit as $\delta_1,\dots,\delta_k\to 0$ in L^2-sense if condition (3.1.2) holds true. □

Now take the space $\mathbb{R}^d\times(\mathbb{R}^d)^k$ with the norm $\|.\|_{d,k}$ given by $\|(t,y)\|_{d,k}=|t|_{\mathbb{R}^d}+|y|_{\mathbb{R}^{dk}}$, where $|x|_{\mathbb{R}^n}$ is the Euclidean norm in $\mathbb{R}^n$. Let $B_r(t,y)$ be the ball of radius $r>0$ with center at (t,y) in some space $\mathbb{R}^d\times(\mathbb{R}^d)^k$. Introduce the set $T_\varepsilon = T\times\{y: 1-\varepsilon<|y|_{\mathbb{R}^{dk}}<1+\varepsilon\}$

Theorem 3.1.5
Let X be a centered Gaussian random field given on a open index set $T\subset\mathbb{R}^d$ which has all kth-order L^2-partial derivatives in all directions everywhere in T. X is k times continuously differentiable with probability one, i.e., $X\in C^k(T)$ a.s., if $\exists C\in(0;\infty)$ and $\varepsilon,\gamma,r_0>0$ such that

$$\mathbf{E}(\Delta^k X(t,\delta_1\cdot y_1)-\Delta^k X(s,\delta_2\cdot y_2))^2 \le C\log(\|(t,y_1)-(s,y_2)\|_{d,k}+|\delta_1-\delta_2|)^{-(1+\gamma)} \tag{3.1.3}$$

for all $((t,y_1),(s,y_2))\in T_\varepsilon\times T_\varepsilon : (s,y_2)\in B_r(t,y_1)$ and some $0<\delta_1,\delta_2,r<r_0$.

Proof Define a new Gaussian random field $\widetilde{X}=\{\widetilde{X}(t,y,\delta):(t,y,\delta)\in\widetilde{T}\}$, where $\widetilde{T}=T_\varepsilon\times(-r;r)$ is an open subset of $\mathbb{R}^d\times(\mathbb{R}^d)^k\times\mathbb{R}$ with the norm $\|(t,y,\delta)\|_{d,k,1}=\|(t,y)\|_{d,k}+|\delta|$ and

$$\widetilde{X}(t,y,\delta) = \begin{cases} \frac{\Delta^k X(t,\delta y)}{\delta^k}, & \delta \neq 0 \\ \frac{\partial^k X(t)}{\partial y}, & \delta = 0 \end{cases} \quad (t,y,\delta) \in \widetilde{T}.$$

$X \in C^k(T)$ a.s. iff $\widetilde{X} \in C(\widetilde{T})$. Moreover, it is clearly sufficient if $\widetilde{X} \in C$ (hyperplane $\delta = 0$ in $\widetilde{T}$). But condition (3.1.3) being true, our assertion follows from Corollary 3.1.1 and Remark 3.1.2. □

3.1.2 Gaussian random polynomials

Let $X = \{X(t) : t \in \mathbb{R}^d\}$ be a random field given by

$$X(t) = \sum_{|\alpha| \leq n} a_\alpha \cdot t^\alpha, \quad t = (t_1, \ldots, t_d)^T \in \mathbb{R}^d,$$

where $\alpha = (\alpha_1, \cdots, \alpha_d)^T$ is a multiindex with $\alpha_i \in \mathbb{N} \cup \{0\}$, $i = 1, \ldots, d$, $|\alpha| = \alpha_1 + \cdots + \alpha_d$, $t^\alpha = t_1^{\alpha_1} \cdot t_2^{\alpha_2} \cdot \ldots \cdot t_d^{\alpha_d}$, and $\{a_\alpha\}$ is a sequence of i.i.d. $N(0,1)$-distributed random variables. This is a random polynomial of degree $n \in \mathbb{N}$ with Gaussian coefficients a_α. (For $d = 1$, we get $X(t) = a_0 + a_1 t + \ldots + a_{n-1} t^{n-1} + a_n t^n$, $t \in \mathbb{R}$). Such polynomials form a special case of Gaussian linear functions introduced in Example 2.2.1, 2):

$X(t) =< \vec{a}, \vec{t} >$, where $\vec{a} = (a_\alpha : |\alpha| \leq n)^T$ and $\vec{t} = (t^\alpha : |\alpha| \leq n)^T$ are the vectors of coefficients and of variables, respectively. It holds obviously $X(t) \sim N(0, |\vec{t}|^2)$, with

$$|\vec{t}|^2 = \sum_{|\alpha| \leq n} t^{2\alpha}, \quad t^{2\alpha} = t_1^{2\alpha_1} \cdot t_2^{2\alpha_2} \cdot \ldots \cdot t_d^{2\alpha_d}$$

and covariance function

$$C(s,t) = \sum_{|\alpha| \leq n} s^\alpha \cdot t^\alpha.$$

Evidently, $X(t)$ has an integral representation of Example 1.2.1.

$$X(t) = \int_{\mathbb{R}^d} g(t,x) W(dx), \quad t \in \mathbb{R}^d, \tag{3.1.4}$$

where

$$g(t,x) = \sum_{|\alpha| \leq n} t^\alpha I(x \in A_\alpha),$$

and $\{A_\alpha\}$ is a family of pairwise disjoint subsets of $\mathbb{R}^d$ with unit volume $\lambda_d(A_\alpha) = 1$, $\forall \alpha$. W is a Gaussian white noise with Lebesgue control measure; cf. Sections 2.3.1 and 2.3.2 for the definition of W and the integral in (3.1.4).

One of the classical questions for random Gaussian polynomials is the problem about the mean number of real (or complex) zeros of X in a compact domain $T \subset \mathbb{R}^d$. If $d = 1$, then the classical result of M. Kac (1943) says that

$$\mathbf{E} N_n = \frac{2}{\pi} \log n (1 + o(1)), \quad n \to \infty, \tag{3.1.5}$$

where N_n is the number of real roots of $X(t) = 0$, $degX = n$. Moreover,

$$\mathbf{E}N_n(a,b) = \frac{1}{\pi}\int_a^b \left(1 - \left(\frac{(n+1)t^n(1-t^2)}{1-t^{2n+2}}\right)^2\right)^{1/2} \frac{dt}{|1-t^2|}, \tag{3.1.6}$$

where $N_n(a,b)$ is the number of real zeros of X in $(a,b), -\infty \le a < b \le +\infty$.

In the Special case $(a,b) = \mathbb{R}$ we have

$$\mathbf{E}N_n = \frac{1}{\pi}\int\limits_{\mathbb{R}} \sqrt{\frac{1}{(t^2-1)^2} - \frac{(n+1)^2\, t^{2n}}{(t^{2n+2}-1)^2}}\, dt \overset{(*)}{=} \frac{4}{\pi}\int\limits_0^1 \left[1 - \left(\frac{(n+1)\, x^n(1-x^2)}{(1-x^{\,2n+2})}\right)^2\right]^{1/2} \frac{dx}{1-x^2},$$

the latter is the original formula by M. Kac (1943).
The equality $(*)$ can be proven by noting that by symmetry

$$\mathbf{E}N_n = \frac{2}{\pi}\int\limits_0^\infty \left(\frac{1}{(t^2-1)^2} - \frac{(n+1)^2t^{2n}}{(t^{2n+2}-1)^2}\right)^{1/2} dt = \frac{2}{\pi}\int\limits_0^1 (...)^{1/2}\, dt + \frac{2}{\pi}\int\limits_1^\infty (...)^{1/2}\, dt = \frac{4}{\pi}\int\limits_0^1 (...)^{1/2}\, dt$$

where we used the substitution $x = 1/t$ in the last integral.
For $n = 1$, we have $\mathbf{E}N_1 = \frac{4}{\pi}\int\limits_0^1 \frac{\sqrt{1-4\,\frac{x^2(1-x^2)^2}{(1-x^4)^2}}}{1-x^2}\, dx = 1$.

Remark 3.1.3
The asymptotic of the mean number of real roots heavily depends on that form of the joint distribution of all coefficients

$$\vec{a} = (a_\alpha : |\alpha| \le n)^T.$$

If all a_α are independent and id. distributed the asymptotic (3.1.5) holds essentially for distributions of a_0 which belongs to the domain of attraction of α-stable laws (see [23]) with another constant factor c on the place of $\frac{2}{\pi}$. It is proven in [41] that if Θ is a class of all non-degenerated distributions of a_0 then

$$\inf_\Theta \sup_{n\in\mathbb{N}} \mathbf{E}(N_n | X(t) \not\equiv 0) = 1.$$

Let $\Theta_{a,b}$ be the class of distributions of a_0 such that

$$P(a_0 > 0) = a, \quad P(a_0 < 0) = b, \quad a, b \ge 0 : 0 < a + b \le 1.$$

Then Nazarov and Zaporozhets showed that

$$\inf_{\Theta_{a,b}} \sup_{n\in\mathbb{N}} \mathbf{E}(N_n | X(t) \not\equiv 0) = 1 + \frac{1-|a-b|}{a+b}.$$

For instance, for symmetric distributions $(a = b = \frac{1}{2})$ we get

$$\inf_{\Theta_{\frac{1}{2},\frac{1}{2}}} \sup_{n\in\mathbb{N}} \mathbf{E}(N_n | X(t) \not\equiv 0) = 2.$$

In order to prove formula (3.1.6), let us do it in a more general setting:

Theorem 3.1.6
Let

$$X(t) = a_0 + a_1 t + a_2 t^2 + \ldots + a_n t^n, \quad t \in \mathbb{R},$$

be a random polynomial, where the random coefficients $a_0, \ldots, a_n$ have a joint density $p(x_0, \ldots, x_n)$. Let $g(t)$ be any smooth function on $\mathbb{R}$, $g \in C^1(\mathbb{R})$. Denote by $N_n(g; a, b)$ the number of real solutions of the equation $X(t) = g(t)$, $t \in [a; b]$. Then

$$\mathbf{E}N_n(g; a, b) = \int_a^b \int_{\mathbb{R}^n} p(g(t) - x_n t^n - \ldots - x_1 t, x_1, \ldots, x_n) |n x_n t^{n-1} + \ldots + x_1 - g'(t)| dx_1 \ldots dx_n dt$$

Proof If X is a polynomial of degree n, then the probability measure on such polynomials is introduced by

$$dX = p(x_0, x_1, \ldots, x_n) dx_0 dx_1 \ldots dx_n.$$

Let us try to pass to more convenient coordinates. If X intersects the graph of g at a point with an abscissa s, then we can write

$$0 = X(s) = a_0 + a_1 s + \ldots + a_n s^n - g(s) \quad \Rightarrow \quad a_0 = g(s) - a_1 s^1 - \ldots - a_n s^n;$$

hence, perform a change of coordinates

$$\begin{cases} x_0 = g(s) - y_1 s - \ldots - y_n s^n \\ x_1 = y_1 \\ \vdots \\ x_n = y_n \end{cases}$$

We get

$$dX = p(x_0, x_1, \ldots, x_n) dx_0 \ldots dx_n = p(g(s) - y_1 s - \ldots - y_n s^n, y_1, \ldots, y_n)$$
$$|g'(t) - y_1 - \ldots - n y_n t^{n-1}| dt dy_1 \ldots dy_n$$

Let $H = \{$all polynomials X of degree n: graph of X intersects the graph of g within $[a; b] \times \mathbb{R}\}$
Then integrate $I(X \in H)$ with respect to dX:
we get

$$\mathbf{E}N_n(g; a, b) = \int_{X \in H} N_n(g; a, b) dX = \int_a^b \int_{\mathbb{R}^n} p(g(s) - y_1 s - \ldots - y_n s^n, y_1, \ldots, y_n)$$
$$|g'(s) - y_1 - \ldots - n y_n s^{n-1}| dy_1 \ldots dy_n ds,$$

since any polynomial is counted as many times on the left-hand side as often it intersects the graph of g. □

Corollary 3.1.2
For i.i.d. $N(0, 1)$-distributed coefficients a_i, $i = 0, \ldots, n$, the above theorem yields formula (3.1.6).

Proof Take

$$g(t) \equiv 0, \quad p(x_0, x_1, \ldots, x_n) = \frac{1}{(2\pi)^{\frac{n+1}{2}}} \exp(-\frac{1}{2}(x_0^2 + \ldots + x_n^2))$$

Then

$$p(g(t) - x_n t^n - \ldots - x_1 t, x_1, \ldots, x_n) = \frac{1}{(2\pi)^{\frac{n+1}{2}}} \exp(-\frac{1}{2}((x_1 t_1 + \ldots + x_n t^n)^2 + x_1^2 + \ldots + x_n^2))$$

If we can prove that

$$\frac{1}{(2\pi)^{\frac{n+1}{2}}} \int_{\mathbb{R}^n} \exp(-\frac{1}{2}((x_1 t + \ldots + x_n t^n)^2 + x_1^2 + \ldots + x_n^2) \cdot |nt^{n-1}x_n + \ldots + x_1| dx_1 \ldots dx_n$$
$$= \frac{1}{\pi} \frac{1}{|1-t^2|} \left(1 - \left(\frac{(n+1)t^n(1-t^2)}{1-t^{2n+2}}\right)^2\right)^{1/2}, \quad t \in (a; b),$$

we are done. □

Exercise 3.1.2
Show this!

Another possibility to get formula (3.1.6) is by means of the so-called *Rice's* formula. Let

$$N_u^+(0,T) = \#\{t \in [0;T] : X(t) = u, X'(t) > 0\}$$

be the number of up-crossings of the smooth stochastic process X of level $u \in \mathbb{N}$ over $[0;T]$. Then, Rice's formula states, that (under some assumptions on X that will be given later) it holds

$$\mathbf{E}N_u^+(0,T) = \int_0^T \int_0^\infty x p_t(u,x) dx dt,$$

where p_t is the probability density of the distribution of random vector $(X(t), X'(t))^T$. Actually, the name "Rice's formula" attributes to the whole class of similar integral formulae for the moments of N_u^+ and related quantities

$$N_u(0,T) = \#\{t \in [0;T] : X(t) = u\},$$
$$N_u^-(0,T) = \#\{t \in [0;T] : X(t) = u, X'(t) < 0\} \quad \text{(number of downcrossings)},$$

and so on. These formulae go back to Rice and Kac in the 1940s who proved them for stationary Gaussian processes. (cf. [1], p. 264-265, [3], p. 69 ff for discussion). The classical Rice's formula (1944, 1945) states that for a stationary centered Gaussian process $X = \{X(t) : t \in \mathbb{R}\}$ with unit variance and second spectral moment λ_2 it holds

$$\mathbf{E}N_u(0,T) = \frac{\sqrt{\lambda_2}}{\pi} e^{-\frac{u^2}{2}} T, \quad \mathbf{E}N_u^+(0,T) = \frac{\mathbf{E}N_u(0,T)}{2},$$

here λ_2 is defined as

$$\lambda_2 = \int_{\mathbb{R}^d} x^2 \mu_X(dx),$$

where μ_X is the spectral measure of X.
Now let us formulate and prove Rice's formula for the case of Gaussian processes. The same formula holds for a much more general class of processes: see [3], p. 74-79. Generalizations for the case of vector-valued random fields are also available; see [3], Chapter 6 and [1], 11.2.

Theorem 3.1.7 (Gaussian Rice formula):
Let $X = \{X(t) : t \in I\}$ be a Gaussian process with C^1-smooth paths, and I be an interval on $\mathbb{R}$. Let $u \in \mathbb{N}$, $t_1, \ldots, t_k \in I$ be any pairwise distinct points and the distribution of $(X(t_1), \ldots, X(t_k))$ be not degenerate with density $p_{X(t_1),\ldots,X(t_k)}$. Denote by $\mathbf{E}N_u^{[k]}$ the k-th factorial moment of $N_u = N_u(I)$:

$$\mathbf{E}N_u^{[k]} = \mathbf{E}[N_u(N_u - 1) \cdot \ldots \cdot (N_u - k + 1)I(N_u \geq k)].$$

Then it holds

$$\mathbf{E}N_u^{[k]} = \int_{I^k} \mathbf{E}[|X'(t_1) \cdot \ldots \cdot X'(t_k)| \; |X(t_1) = u, \ldots, X(t_k) = u] p_{X(t_1)\ldots X(t_k)}(u, \ldots, u) dt_1 \ldots dt_k \tag{3.1.7}$$

Before we begin to prove this Theorem, some remarks are in order.

1. An equivalent form of (3.1.7) is

$$\mathbf{E}N_u^{[k]} = \int_{I^k} dt_1 \ldots dt_k \int_{\mathbb{R}^k} |x'_1 \cdot x'_2 \cdot \ldots \cdot x'_k| p_{X(t_1)\ldots X(t_k)X'(t_1)\ldots X'(t_k)}(u, \ldots, u, x'_1, \ldots, x'_k) dx'_1 \ldots dx'_k,$$

   where $p_{X(t_1)\ldots X(t_k)X'(t_1)\ldots X'(t_k)}$ is the joint probability density of $X(t_j)$, $i = 1, \ldots, k$ and $X'(t_j)$, $j = 1, \ldots, k$.

2. In particular, for $k = 1$ it holds

$$\mathbf{E}N_u(I) = \int_I \int_{\mathbb{R}} |x| p_{X(t),X'(t)}(u, x) dx dt \tag{3.1.8}$$

   It can be shown that this formula holds also for non-Gaussian processes X satisfying the following conditions:

   a) Function $(t, x) \mapsto p_{X(t)}(x)$ is continuous for $t \in T$, x in a neighborhood of u.

   b) $(t, x, x') \mapsto p_{X(t),X'(t)}(x, x')$ is continuous for $t \in I$ in a neighborhood of u and $x' \in \mathbb{R}$.

   c) If $\omega_{X'}(\delta)$ is the modulus of continuity of X' then

$$\mathbf{E}[\omega_{X'}(\delta)] \xrightarrow[\delta \to 0]{} 0.$$

3. It follows from formula (3.1.8) that $\mathbf{E}N_u(I)$ is finite for Gaussian processes, since the right-hand side is always finite. Unfortunately, this may not be the case for $\mathbf{E}(N_u^{[k]}(I)) < \infty$, $k > 1$.

4. It is in general not easy to show that the distribution of $X(t_1)\dots X(t_k)$ is non-degenerate. However, in the case of stationary Gaussian X, an easy sufficient condition for that is that there exists no countable set $A : \mu_X(A^c) = 0$, where μ_X is the spectral measure of X.

For the proof of Theorem 3.1.7 we need the following lemma:

Lemma 3.1.5 (Kac's counting formula):
Let $f \in C^1(I)$ be a real-valued function defined on an interval $I = [t_1, t_2]$, $f(t_1) \neq u$, $f(t_2) \neq u$, such that there are no critical points of f at level $u : \nexists t \in I : f(t) = u,\ f'(t) = 0$. Then

$$N_u(f, I) = \lim_{\delta \to 0} \frac{1}{2\delta} \int_{t \in I : |f(t)-u| < \delta} |f'(t)| dt \tag{3.1.9}$$

Proof Since $f \in C^1(I)$, $N_u(f, I)$ is finite. Let $N_u(f, I) = n \in N \cup \{0\}$. If $n = 0$, then both sides of (3.1.9) are zero, since the integration set in (3.1.9) is empty for $\delta > 0$ small enough. Let $n \geq 1$ and $s_1, \dots, s_n \in I$ be the points where $f(s_i) = u$, $i = 1, \dots, n$. Since $f'(s_i) \neq 0$, $i = 1, \dots, n$, we have for small enough $\delta > 0$ that

$$f^{-1}(u - \delta, u + \delta) = \bigcup_{j=1}^{n} J_j$$

where J_j are disjoint intervals and $s_i \in J_i$, $i = 1, \dots, n$. $f|_{J_i}$ is a one-to-one mapping, a diffeomorphism, so that one gets

$$\int_{J_i} |f'(t)| dt = 2\delta$$

by an exchange of variables. Then for small enough $\delta > 0$, it holds

$$\frac{1}{2\delta} \int_{t \in I : |f(t)-u| < \delta} |f'(t)| dt = \frac{1}{2\delta} \sum_{i=1}^{n} \int_{J_i} |f'(t)| dt = \frac{n \cdot 2\delta}{2\delta} = n.$$

□

Proof of Theorem 3.1.7. We prove only the case $k = 1$. For $k > 1$ see [3], p. 73. Without loss of generality we assume that $I = [0, 1]$. Let $X^{(n)}$ be the dyadic approximation of X, i.e. the graph of $X^{(n)}$ is a polygonal line with vertices $\{(\frac{k}{2^n}, X(\frac{k}{2^n})),\ k = 0, \dots, 2^n\}$. One can easily show that

$$\sup_{t \in [0,1]} |X^{(n)}(t) - X(t)| \xrightarrow[n \to \infty]{a.s.} 0,$$

$$\left| X^{(n)}(t) - X(t) \right| \overset{\text{a.s.}}{\leq} 2 \sup_{t \in I} |X(t)|, \quad \forall n \in \mathbb{N}, t \in I,$$

whereas

$$\mathbf{E} \left[\sup_{t \in I} |X(t)| \right]^m < \infty$$

for all $m \in \mathbb{N}$, since X is a Gaussian process with a.s. C^1-paths. By Lebesgue's theorem on dominated convergence

$$\mathbf{var}X^{(n)} \xrightarrow[n\to\infty]{} \mathbf{var}X(t)$$

uniformly on $t \in I$. Hence, for large n

$$\mathbf{var}(X^{(n)}(t)) \le b > 0$$

for all $t \in I$ and some $b > 0$, since X is non-degenerate process. For such large n, it holds

$$X^{(n)}\left(\frac{k}{2^n}\right) \neq u \quad \text{a.s.}, \quad k = 0, \dots, 2^n,$$

since $X^{(n)}(t)$ has a (Gaussian) density, $t \in I$. Since the result of Lemma 3.1.5 holds also for polygonal f (although they are only piecewise from $C^1(I)$) by additivity of left and right-hand side of formula (3.1.9) we get

$$N_u(X^{(n)}, I) \overset{\text{a.s.}}{=} \lim_{\delta\to 0} \frac{1}{2\delta} \int_{t\in I: |X^{(n)}(t)-u|<\delta} |X^{(n)'}(t)|dt \tag{3.1.10}$$

for any n large enough, where the right-hand side of (3.1.10) is obviously bounded by the total number of segments in $X^{(n)}$, which is 2^n. Applying dominated convergence theorem, we get

$$\begin{aligned}\mathbf{E}N_u(X^{(n)}, I) &= \lim_{\delta\to 0} \frac{1}{2\delta} \int_I \underbrace{\mathbf{E}(|X^{(n)'}(t)| \cdot I\{|X^{(n)}(t) - u| < \delta\})}\\ &\quad \mathbf{E}(I\{|X^{(n)}(t) - u| < \delta\}(|X^{(n)'}(t)||X^{(n)}(t)))dt\\ &= \lim_{\delta\to 0} \int_I \frac{1}{2\delta} \int_{u-\delta}^{u+\delta} \mathbf{E}(|X^{(n)'}(t)| \mid X^{(n)}(t) = x) p_{X^{(n)}(t)}(x)dxdt.\end{aligned}$$

Since $X, X^{(n)}$ have continuous sample paths and are Gaussian, their expectation and covariance functions are continuous as well. Since for $t \in \left(\frac{k}{2^n}, \frac{k+1}{2^n}\right]$ it holds

$$\begin{aligned}X^{(n)}(t) &= (k + 1 - 2^n t)X\left(\frac{k}{2^n}\right) + (2^n t - k)X\left(\frac{k+1}{2^n}\right)\\ &= \underbrace{2^n\left(X\left(\frac{k+1}{2^n}\right) - X\left(\frac{k}{2^n}\right)\right)}_{\xi(k)} t + \underbrace{(k+1)X\left(\frac{k}{2^n}\right) - kX\left(\frac{k+1}{2^n}\right)}_{\eta(k)},\end{aligned}$$

we get $X^{(n)'}(t) = \xi(k)$ and hence

$$\mathbf{E}(|X^{(n)'}(t)| \mid X^{(n)} = x) = \mathbf{E}(|\xi(k)| \mid \xi(k)t + \eta(k) = x) = \mathbf{E}\left(\left|\frac{x - \eta(k)}{t}\right|\right).$$

Thus,

$$\mathbf{E}(|X^{(n)'}(t)| \mid X^{(n)}(t) = x) p_{X^{(n)}(t)}(x)$$

is a continuous function of (t, x) on $\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right] \times [u - \delta_0, u + \delta_0]$ for some $\delta_0 > 0$, and hence

$$\mathbf{E}N_u(X^{(n)}, I) = \int_I \lim_{\delta\to 0} \frac{1}{2\delta} \int_{u-\delta}^{u+\delta} \mathbf{E}(|X^{(n)'}(t)| \mid X^{(n)}(t) = x) p_{X^{(n)}(t)}(x) dx dt$$
$$= \int_I \mathbf{E}(|X^{(n)'}(t)| \mid X^{(n)}(t) = u) p_{X^{(n)}(t)}(u) dt$$

by the mean value theorem. Now take the limit as $n \to \infty$ on both sides of the last equality. Since

$$N_u(X^{(n)}, I) \uparrow N_u(X, I)$$

as $n \to \infty$, we get

$$\mathbf{E}N_u(X^{(n)}, I) \xrightarrow[n\to\infty]{} \mathbf{E}N_u(X, I)$$

by the theorem on monotone convergence. The monotone convergence $N_u(X^{(n)}, I)$ to $N_u(X, I)$ takes place, since with probability one there are no extremal points of X at level u. [Ylvisaker Theorem]. As $n \to \infty$, the expectation and the covariance matrix of $(X^{(n)}(t), X^{(n)'}(t))^T$ converge to the corresponding expressions of X, that is why

$$\int_I \mathbf{E}(|X^{(n)'}(t)| \mid X^{(n)}(t) = u) p_{X^{(n)}(t)}(u) dt \xrightarrow[n\to\infty]{} \int_I \mathbf{E}(|X'(t)| \mid X(t) = u) p_{X(t)}(u) dt.$$

□

Exercise 3.1.3
Show that (3.1.6) holds using Rice's formula!

Another problem of interest is the location of all complex zeros of

$$X = \{X(t) : t \in \mathbb{C}\}, \quad X(t) = \sum_{i=0}^{n} a_i t^i$$

on the complex plane $\mathbb{C}$. Let $R_n(a, b) = \#\{z \in \mathbb{C} : a \le |z| \le b, X(z) = 0\}$, $0 \le a < b < \infty$ be the number of zeros of X on the ring. Let $S_n(\alpha, \beta) = \#\{z \in \mathbb{C} : \alpha \le \arg z \le \beta, X(z) = 0\}$, $-\pi \le \alpha < \beta \le \pi$ be the number of zeros of X in the sector. The following result (see [25]) can be proven for arbitrary distributions of i.i.d. coefficients a_i, $i = 0, \dots, n$ of X:

Theorem 3.1.8 1. It holds $\frac{1}{n} R_n(1-\delta, 1+\delta) \xrightarrow[n\to\infty]{a.s.} 1$ for any $\delta \in (0, 1)$ iff

$$\mathbf{E}\log(1 + |a_0|) < \infty. \tag{3.1.11}$$

2. For any distribution of a_0, it holds

$$\frac{1}{n} S_n(\alpha, \beta) \xrightarrow[n\to\infty]{a.s.} \frac{\beta - \alpha}{2\pi}.$$

This result shows that all complex zeros of random polynomials satisfying conditions of Theorem 3.1.8 asymptotically lie (with probability one) around the unit circle $|z| = 1$. Moreover, they are spread over $|z| = 1$ uniformly.

For Gaussian coefficients a_0, a more exact result can be proven (see [51]): for any $\delta > 0$

$$\frac{1}{n}\mathbf{E}R_n\left(e^{-\frac{\delta}{n}}, e^{\frac{\delta}{n}}\right) \xrightarrow[n\to\infty]{} \frac{1+e^{2\delta}}{1-e^{-2\delta}} - \frac{1}{\delta}.$$

This asymptotic was extended to distributions of a_0 belonging to domains of attraction of α-stable laws by [26]. It is worth noting that the behavior of complex roots can be very different from that stated in Theorem 3.1.8, 1) if condition (3.1.11) does not hold. There are examples of polynomials such that on average $\frac{n}{2} + o(1)$ of its n roots concentrate near the origin, whereas another half goes to infinity as $n \to \infty$ (see [60]).

Now let us consider the case of dimension d of index space greater than one, i.e.

$$X(t) = \sum_{|\alpha|\le n} a_\alpha \cdot t^\alpha$$

is a polynomial of d variables ($t \in \mathbb{R}^d$). Let M_n be the random hypersurface defined by the equation $X(t) = 0$; let $\mathcal{H}^{d-1}$ be the Hausdorff measure of dimension $d-1$.

Theorem 3.1.9
If a_α are i.i.d. $N(0,1)$-distributed random variables, then for any compact set $K \in R^d$ it holds

$$\mathbf{E}\mathcal{H}^{d-1}(M_n \cap K) = \frac{\Gamma\left(\frac{d+1}{2}\right)}{2\pi^{\frac{d+1}{2}}} \int_K \int_0^\pi \dots \int_0^{2\pi} \sqrt{\sum_{i=1}^d \frac{g(x_i)}{(1-x_i^2)} v_i^2(\varphi)} \cdot D(\varphi) d\varphi dx,$$

where $dx = dx_1 \dots dx_d$, $d\varphi = d\varphi_1 \dots d\varphi_{d-1}$,

$$D(\varphi) = \sin^{d-2}\varphi_1 \sin^{d-3}\varphi_2 \cdot \ldots \cdot \sin\varphi_{d-2}$$
$$g(x_i) = 1 - \left(\frac{(n+1)x_i^n(1-x_i^2)}{1-x_i^{2n+2}}\right)^2, \quad i = 1, \dots, d$$

$$\begin{cases} v_1(\varphi) = \cos\varphi_1 \\ v_2(\varphi) = \sin\varphi_1 \cdot \cos\varphi_2 \\ \vdots \\ v_{d-1}(\varphi) = \sin\varphi_1 \cdot \sin\varphi_2 \cdot \ldots \cdot \sin\varphi_{d-2} \cdot \cos\varphi_{d-1} \\ v_d(\varphi) = \sin\varphi_1 \cdot \sin\varphi_2 \cdot \ldots \cdot \sin\varphi_{d-2} \sin\varphi_{d-1} \end{cases}$$

Moreover, it holds

$$\mathbf{E}\mathcal{H}^{d-1}(M_n \cap K) = \frac{\log n}{\pi} \mathcal{H}_{d-1}(K \cap K_0)(1 + o(1)), \quad n \to \infty,$$

where

$$K_0 = \bigcup_{i=1}^d \{x = (x_1, \dots, x_d) \in \mathbb{R}^d : |x_i| = 1\}$$

The proof of this result can be found in [24].
The asymptotic result above means that almost the whole surface M_n concentrates around K_0 for $n \to \infty$. More generally, it holds $\mathbf{E}\mathcal{H}^{d-1}(M_n \cap K) = \mathcal{O}(1)$, as $n \to \infty$ for polynomials X

with i.i.d. coefficients a_α that have a density p with $\sup\limits_{x\in\mathbb{R}} p(x) < \infty$ and $\mathbf{E}a_0 < \infty$, if compact K is chosen so that $\mathbf{dist}(K, K_0) > 0$. Here $\mathbf{dist}(.,.)$ is the Euclidean distance between two sets (see [22]).
In the same paper, the statement of Theorem 3.1.9 is generalized to symmetric strictly α-stable distributions of a_α.

3.1.3 Large deviations for Gaussian random functions

In this Section, we investigate the asymptotic behavior of the probability $P(\sup\limits_{t\in T} X(t) \geq u),\ u \to +\infty$, for a Gaussian random function $X = \{X(t),\ t \in T\}$. Probabilities of that kind play an important role e.g. in risk theory (insurance mathematics) and financial mathematics, as well as in reliability theory and theory of extreme values. This asymptotic behaviors depends heavily on the behavior of the variance of X.
See the book Piterbarg [44] (1996) for a complete treatment of this subject. We state and prove the following powerful and elegant result by M. Talagrand (1988) [53].

Theorem 3.1.10
Let $X = \{X(X),\ t \in T\}$ be a centered a.s. bounded real separable Gaussian random function on a compact T with continuous covariance function and a unique maximum point of variance, i.e.,

$$\exists!\ t_0 \in T\ :\ \sigma_T^2 := \ E\ X^2(t_0) = \sup_{t\in T}\ E\ X^2(t). \tag{3.1.12}$$

Suppose that

$$\lim_{\delta\to 0}\ \frac{1}{\delta}\ \left(E\ \sup_{t\in T_\delta}\ X\ (t)\right)\ =\ 0, \tag{3.1.13}$$

where the set $T_\delta\ =\ \{t \in T\ :\ E\ (X(t)X(t_0)) \geq\ \sigma_T^2\ -\ \delta^2\}$. Then

$$P\ \left(\sup_{t\in T}\ X(t)\ \geq\ u\right)\ \underset{u\to\to\infty}{\sim}\ \Psi(u/\sigma_T), \tag{3.1.14}$$

where $\Psi(x)\ =\ \frac{1}{\sqrt{2\pi}}\ \int\limits_x^\infty e^{-t^2/2}dt, x \in \mathbb{R}$ is the tail probability function of standard normal law. Also the converse statement is true: if (3.1.12) holds true then property (3.1.13) follows from (3.1.14).

Remark 3.1.4
Relation (3.1.14) does not imply both relations (3.1.12) and (3.1.13). For instance, let $X = \{X(t), t \in T\}, X(t) \equiv X_0 \sim N(0,1),\ \forall t \in T$. It holds $P(\sup_{t\in T}\ X(t) \geq\ u)\ =\ P\ (X_0 \geq u)\ = \Psi(u)$, i.e., (3.1.14) holds, however, $\mathbf{Var}X(t)\ = 1\ \ \forall t \in T$, which means that condition (3.1.12) is not true.

Proof Let us prove only the sufficiency of (3.1.12) and (3.1.13) for (3.1.14). Since $\{X(t_0) \geq u\}\ \subseteq\ \{\sup_{t\in T_0}\ X(t) \geq\ u\}$ then $\forall u \in \mathbb{R}\ P(\sup_{t\in T} X(t) \geq u)\ \geq\ P(X(t_0) \geq u)\ =\ \Psi(u/\sigma_T)$. Hence, it is sufficient to show that $\limsup_{u\to+\infty}\ P(\sup_{t\in T}\ X(t) \geq\ u)/\Psi(u/\sigma_T)\ \leq 1$, because

$$1 \leq \liminf_{u\to+\infty}\ P(\sup_{t\in T}\ X(t) \geq\ u)\ /\ \Psi(u/\sigma_T)$$

$$\leq \limsup_{u\to+\infty} P(\sup_{t\in T} X(t) \geq u) \ / \ \Psi(u/\sigma_T) \ \leq 1$$

yields relation (3.1.14).
It follows from (3.1.13) that $\forall \varepsilon > 0 \ \exists \delta_0 > 0$ small enough such that for any $\delta \leq 2\delta_0$ it holds

$$\mathbf{E}(\sup_{t\in T_\delta} X(t)) \ \leq \ \varepsilon^2 \cdot \delta. \tag{3.1.15}$$

Assume that $\delta_0 \leq \sigma_T \varepsilon^2$. By assumption on $t_0, \sup_{t\notin T_\delta} \mathbf{E}\, X^2(t) < \sigma_T^2$.
Without loss of generality, we may put $\sigma_T = 1$, since otherwise we may scale $X(t)$ by σ_T : $\widetilde{X}(t) = X(t)/\sigma_T$.
We need the following Lemmata:

Lemma 3.1.6
Let $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, $x \in \mathbb{R}$ be the density if the standard normal law. Then for all $x > 0$ it holds $\left(\frac{1}{x} - \frac{1}{x^3}\right)\varphi(x) \leq \Psi(x) \leq \frac{1}{x}\varphi(x)$.

To prove the upper bound, notice that

$$\sqrt{2\pi}\ \psi(x) = \int_x^\infty e^{-y^2/2} dy = \int_{\{y\geq x\}} e^{-y^2/2} dy \underset{1\leq y/x}{\leq} \int_x^\infty \frac{y}{x} e^{-y^2/2} dy = \frac{1}{x}\,\frac{1}{2}\int_x^\infty e^{-y^2/2} dy^2$$
$$= \frac{1}{x} e^{-x^2/2}.$$

For the lower bound, compute $\sqrt{2\pi}\ \psi(x) = \int_x^{+\infty} e^{-y^2/2} dy \ = |$ Substitution $y = x + t/x|$
$$= \frac{1}{x}\int_0^\infty e^{-\frac{1}{2}(x^2+2t+t^2/x^2)} dt = \frac{1}{x}\, e^{-x^2/2} \int_0^\infty e^{-(2t+t^2/x^2)/2}\, dt \underbrace{\geq}_{\text{Since } e^{-z}\geq 1-z} \frac{1}{x}\, e^{-x^2/2} \int_0^\infty e^{-t}\left(1 - \frac{1}{2}\frac{t^2}{x^2}\right) dt$$
$$\geq \frac{1}{x} e^{-x^2/2} \cdot \left(1 - \frac{1}{x^2}\int_0^\infty \frac{t^2}{2} e^{-t}\, dt\right) = \frac{1}{x}\left(1 - \frac{1}{x^2}\right) e^{-x^2/2}, \text{since } \int_0^\infty e^{-t}\,\frac{t^2}{2}\, dt = -\int_0^\infty \frac{t^2}{2}\, de^{-t} = \int_0^\infty e^{-t}\, dt = 1.$$

Remark 3.1.5
In particular, it holds $\psi(x) \geq \kappa_1 \cdot \frac{1}{x}$ for all $x \geq 1$ and some $\kappa_1 > 0$.

Lemma 3.1.7 (Inequality of Borell-Tsirelson-Sudakov, 1975):
Let $Y = \{Y(t), t \in T\}$ be a separable Gaussian random function, a.s. bounded on T and centered. Then, for $w \in \mathbb{R}$ such that $P(\sup_{t\in T} Y(t) \geq w) \leq 1/2$ and any $u \geq w$ we have $P(\sup_{t\in T} Y(t) \geq u) \leq \psi\left(\frac{u-w}{\sigma_T}\right)$ where $\sigma_T^2 = \sup_{t\in T} \mathbf{E}Y^2(t)$. (without proof).

By Lemma 3.1.7 it holds

$$\lim_{u\to+\infty} \frac{P(\sup_{t\in T\setminus T_{\delta_0}} Y(t) \geq u)}{\psi(u)} = 0, \tag{3.1.16}$$

since $T \setminus T_{\delta_0} = \{t \in T : \mathbf{E}\, X(t)X(t_0) < 1 - \delta_0^2\}$, and for large u

$$\frac{P(\sup_{t\in T\setminus T_{\delta_0}} Y(t) \geq u)}{\psi(u)} \overset{L.3.1.7}{\leq} \frac{\psi\left(\frac{u-w}{\theta}\right)}{\psi(u)} \overset{L.3.1.6}{\leq} \frac{\theta u}{u-w}\, \frac{e^{-(u-w)^2/2}}{\kappa_1 e^{-u^2/2}}$$

$$\sim \exp\left(-u^2/2 \underbrace{\left(1/\theta^2-1\right)}_{>0,\ \text{since } \theta<1}\right) e^{2uw/\theta} \underset{u\to+\infty}{\to} 0.$$

Since $P\left(\sup_{t\in T_{\delta_0}\cup(T\setminus T_{\delta_0})=T} X(t) \geq u\right) \leq P\left(\left\{\sup_{t\in T_{\delta_0}} X(t)\geq u\right\}\cup\left\{\sup_{t\in T\setminus T_{\delta_0}} X(t) \geq u\right\}\right)$ $\leq P\left(\sup_{t\in T_{\delta_0}} X(t) \geq u\right) + P\left(\sup_{t\in T\setminus T_{\delta_0}} X(t) \geq u\right)$ for all $u>0$, and in view of (3.1.16), it is enough to show that $\limsup_{u\to+\infty} P\left(\sup_{t\in T_{\delta_0}} X(t) \geq u\right)/\psi(u) \leq 1$, i.e., for all sufficiently large $u>0$

$$P\left(\sup_{t\in T_{\delta_0}} X(t) \geq u\right) \leq \psi(u)\,(1+K\varepsilon) \tag{3.1.17}$$

for some constant $K>0$.
Now fix $u>0$ and set $\alpha = 1/(\varepsilon\, u)$. Define the following non-decreasing sequence of sets: $V_{-1}=\emptyset$, $V_k = T_{2^k\alpha}$, for $k\geq 0$, and let $U_\kappa = V_k\setminus V_{k-1} = \{t\in T : 1-(2^k\alpha)^2 \leq \mathbf{E}(X(t)X(t_0)) < 1-(2^{k-1}\alpha)^2\}, \forall k\geq 1, U_0 = V_0$. Let $p=\inf\{k\in\mathbb{N} : 2^k\alpha\geq\delta_0\}$. It follows then $T_{\delta_0}\subset\bigcup_{0\leq k\leq p} U_k$. Let us prove the bound on each "ring" U_k, and then combine them.
Set $\mu_k = E\left(\sup_{t\in V_\kappa} X(t)\right)$. For $k\leq p$ it holds by the bound (3.1.15) that $\mu_k \leq \alpha 2^k\varepsilon^2$.

Lemma 3.1.8
Let $Y=\{Y(t), t\in T\}$ be a centered a.s. continuous Gaussian process on T, and Z a Gaussian random variable such that $EZ=0$ and the family $\{Z, Y(t), t\in T\}$ is jointly Gaussian. Assume that

$$\mathbf{E}((Y(t)-Z)Z)\leq 0 \tag{3.1.18}$$

for each $t\in T$. Set

$$\mu = \mathbf{E}\sup_{t\in T} Y(t),\ \sigma^2 = E\ Z^2,\ a = \left(\sup_{t\in T}\mathbf{E}(Y(t)-Z)^2\right)^{1/2}.$$

If $a<\sigma$ and $u>\max\{2\mu, \mu+\sigma\}$ then

$$P\left(\sup_{t\in T} Y(t)\geq u\right) \leq \Psi(u/\sigma)\ \left[1+K\cdot\frac{au}{\sigma^2}\cdot e^{a^2u^2/(2\sigma^4)}\right] e^{2u\mu/\sigma^2},$$

where $K>0$ is a constant.

See the proof in [Adler Taylor], p. 80-83.
Use this Lemma with

$$T=U_0=V_0 \text{ and } Z=X(t_0),\ Y(t)=X(t),\ t\in U_0,\ \sigma=1,\ \mu=\mu_0\leq\alpha\cdot\varepsilon^2=\varepsilon/u.$$

Since (3.1.12) holds, (3.1.18) is true by Cauchy-Schwarz inequality. Then we have for each u large enough

$$P\left(\sup_{t\in U_0} X(t)\geq u\right) \leq \Psi(u)\ \left[1+K\cdot a\cdot u\cdot e^{a^2u^2/2}\right]\underbrace{e^{2u\cdot\varepsilon/u}}_{e^{2\varepsilon}}. \tag{3.1.19}$$

Calculate $a = \omega_0$, where $\omega_k = \sup_{t\in V_k} (E(X(t) - X(t_0))^2)^{1/2}$ for all $k \geq 0$. We have for any Gaussian $Z \sim N(0, \sigma^2)$

$$\mathbf{E}|Z| = 2\int\limits_0^\infty x\ \varphi_Z(x)dx = \frac{2}{\sqrt{2\pi}\sigma}\int\limits_0^\infty xe^{-x^2/(2\sigma^2)}dx = |y = \frac{x}{\sigma}|$$

$$= \sigma\frac{2}{\sqrt{2\pi}}\underbrace{\int\limits_0^\infty ye^{-y^2/2}\ dy}_{1} = \sigma\sqrt{\frac{2}{\pi}} = \sqrt{\frac{2}{\pi}}(\mathbf{E}|Z|^2)^{1/2},$$

hence

$$\omega_k = \sqrt{\frac{\pi}{2}}\ \sup_{t\in V_k}\mathbf{E}|X(t) - X(t_0)| \leq \sqrt{\frac{\pi}{2}}\mathbf{E}\sup_{t\in V_k}|X(t) - X(t_0)| \underbrace{\leq}_{\text{since } t_0\leftarrow V_k\ \forall k}$$

$$\leq\ 2\sqrt{\frac{\pi}{2}}\mathbf{E}\sup_{t\in V_k}|X(t)| \leq \underbrace{\sqrt{2\pi}}_{K_1}\ E\sup_{t\in V_k}X(t) =\ K_1\cdot\mu_k \leq K_1\cdot\alpha\cdot 2^k\varepsilon^2,$$

since by symmetry of $N(0, \sigma^2)$ it holds

$$P(\sup_{t\in T}|X(t)| > u)\ \leq\ 2\ P(\sup_{t\in T}X(t) > u)$$

and hence

$$\mathbf{E}\ \sup_{t\in T}\ |X(t)| = \int\limits_0^\infty P(\sup_{t\in T}|X(t)| > u)du \leq 2\int\limits_0^\infty P(\sup_{t\in T}X(t) > u)du =\ 2\mathbf{E}\sup_{t\in T}X(t).$$

Then we insert $a \leq K_1\cdot\mu_0 \leq K_1\frac{\varepsilon}{u}$ in (3.1.19) to get

$$P(\sup_{t\in U_0}X(t) \geq u) \leq \Psi(u)\cdot\left[1 + K\cdot K_1\cdot\varepsilon\cdot e^{K_1^2\varepsilon^2/2}\right]\ e^{2\varepsilon} \leq \Psi(n)\left[1 + K_2\varepsilon\right]$$

for $\varepsilon > 0$ small enough, since then $e^\varepsilon = 1 + \varepsilon + o(\varepsilon)$ as $\varepsilon \to +0$.

A similar argument shall be applied to all U_k for $p \geq k \geq 1$. Set $Z = (1 - (\alpha\cdot 2^{k-1})^2)X(t_0)$ in Lemma 3.1.8. Check that inequality (3.1.18) holds: for $t \in U_k,\ 1 \leq k\ \leq p$

$$\mathbf{E}\ \left[\left(X(t) - X(t_0)\ +\ (\alpha\cdot 2^{k-1})^2\ X(t_0)\right)X(t_0)\right]\ \left(1 - (\alpha\cdot 2^{\kappa-1})^2\right)$$

$$= \underbrace{\left(1 - (\alpha\cdot 2^{k-1})^2\right)}_{\geq 0}\underbrace{\left(E\left[X(t)X(t_0)\right] - 1 + \underbrace{(\alpha\cdot 2^{k-1})^2}_{\leq\sigma^2}\right)}_{\leq 0,\ \text{ if } t\in U_k.} \leq 0.$$

Then Lemma 3.1.8 yields

$$P\ (\sup_{t\in U_k}X(t) \geq u) \leq \Psi(u/\sigma)\ \left(1 + K\frac{au}{\sigma^2}\ e^{a^2u^2/(2\sigma^4)}\right)\cdot e^{2u\mu/\sigma^2}, \tag{3.1.20}$$

where

$$\sigma^2 = \mathbf{E}Z^2 = \left(1-(\alpha\cdot 2^{k-1})^2\right)^2 \ \mathbf{E}X^2(t_0) = \left(1-(\alpha\cdot 2^{k-1})^2\right)^2,$$

$$\mu = \mathbf{E}\sup_{t\in U_k} X(t) \le \mathbf{E}\sup_{t\in V_k} X(t) = \mu_k \le \alpha\cdot 2^k\cdot\varepsilon^2 = \left|\alpha = \frac{1}{\varepsilon u}\right| = \frac{2^k\cdot\varepsilon}{u},$$

$$a = \sup_{t\in U_k}\left(\mathbf{E}\ (X(t)-(1-(\alpha\cdot 2^{k-1})^2)X(t_0))^2\right)^{1/2}$$

$$\le \sup_{t\in V_k} \parallel X(t)-(1-(\alpha\cdot 2^{k-1})^2)X(t_0)\|_2,$$

where $\|\cdot\|_2$ is the $L_2(\Omega,\mathcal{F},P)-$norm, and $K>0$ is a constant. By triangle inequality,

$$\sup_{t\in V_\kappa} \parallel X(t)-Z\|_2 \le \underbrace{\sup_{t\in V_k} \parallel X(t)-X(t_0)\|_2}_{\omega_k} + \sup_{t\in V_k} \parallel X(t_0)-Z\|_2$$

$$\le K_1\cdot\alpha\cdot 2^k\ \varepsilon^2 \ + \ (\alpha\cdot 2^{k-1})^2\cdot\underbrace{\|X(t_0)\|_2}_{1},$$

since $1\le k\le p$ and hence $\delta_0\ge\alpha\cdot 2^{k-1}$, where $\delta_0\le\varepsilon^2$ by the choice of δ_0, we have

$$(\alpha\cdot 2^{k-1})^2 \le \alpha\cdot 2^{k-1}\cdot\delta_0 \le \alpha\cdot 2^k\cdot\varepsilon^2.$$

Then

$$a \ \le \ K_1\cdot\alpha\cdot 2^k\ \varepsilon^2 + \alpha\cdot 2^k\ \varepsilon^2 = K_2\cdot\alpha\cdot 2^k\cdot\varepsilon^2$$

for some constant $K_2>0$. Since $0<\varepsilon<1$, inequality (3.1.20) reads

$$P\ (\sup_{t\in U_k} X(t)\ge u) \le \ \Psi\left(\frac{u}{1-(\alpha\cdot 2^{k-1})^2}\right)$$

$$\times\left[1+K\varepsilon^2 2^k\frac{u}{\varepsilon u}K_2\frac{1}{\left(1-(\alpha 2^{k-1})^2\right)^2}\ \exp\left(K_2^2\frac{u^2}{\varepsilon^2u^2}2^{2k}\frac{\varepsilon^4}{2}\left(1-(\alpha\cdot 2^{k-1})^2\right)^{-4}\right)\right]$$

$$\times\exp\left(2u2^k\frac{\varepsilon}{u}\frac{1}{\left(1-(\alpha\cdot 2^{k-1})^2\right)^2}\right)$$

$$\le \ \Psi\left(\frac{u}{1-(\alpha\cdot 2^{k-1})^2}\right)\ \left[1+K_3\cdot\varepsilon\ 2^k\cdot\ e^{K_4\ \varepsilon^2 2^{2k}}\right]e^{K_5\cdot\varepsilon 2^k}$$

$$\le \ \Psi\ \left(\frac{u}{1-(\alpha\cdot 2^{k-1})^2}\right)\ e^{K_6\cdot\varepsilon\cdot 2^{2k}},$$

since for large $u>0$ $\frac{1}{1-(\alpha\cdot 2^{k-1})^2}\Big|_{\alpha=\frac{1}{\varepsilon u}}$ is bounded. Using the inequality $\frac{1}{1-x}\ge 1+x,\ 0\le x<1$, we get

$$\frac{u}{1-(\alpha\cdot 2^{\kappa-1})^2}\ \ge u+u\cdot\ (\alpha\cdot 2^{\kappa-1})^2.$$

Since $\forall y \geq 0$

$$\Psi(x+y) = \int\limits_{x+y}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2} dt = |z = t - y| = \int\limits_{x}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-(z+y)^2/2} dz$$
$$= \frac{e^{-y^2/2}}{\sqrt{2\pi}} \int\limits_{x}^{\infty} e^{-z^2/2}\, e^{-zy} dz \leq \frac{1}{\sqrt{2\pi}} \cdot e^{-xy} \int\limits_{x}^{\infty} e^{-z^2/2}\, dz \;= e^{-xy}\Psi(x),$$

we have $\Psi\left(\frac{u}{1-(\alpha\cdot 2^{k-1})^2}\right) \underbrace{\leq}_{\Psi\downarrow} \Psi\left(u + u(\alpha\cdot 2^{k-1})^2\right) \leq \exp\underbrace{\left(-u^2(\alpha\cdot 2^{\kappa-1})^2\right)}_{2^{2(k-1)}/\varepsilon^2} \Psi(u)$ and hence

$$\sum_{1\leq k\leq p} P(\sup_{t\in U_k} X(t) \geq u) \leq \Psi(u) \sum_{1\leq k\leq p} \exp\left(-\frac{2^{2k}}{4\varepsilon^2} + K_6\varepsilon\cdot 2^{2k}\right)$$
$$\underbrace{\leq}_{\varepsilon\in(0,1)} \Psi(u) \sum_{1\leq k\leq p} \exp\left(-2^{2k}\left(\frac{1}{4\varepsilon} - K_6\varepsilon\right)\right) \underset{\varepsilon\to+0}{\longrightarrow} 0.$$

To summarize, we proved that

$$P\left(\sup_{t\in T_{\delta_0}} X(t) \geq u\right) \leq \sum_{k=0}^{p} P\left(\sup_{t\in U_k} X(t) \geq u\right) \leq \left|\text{ Note that } T_{\delta_0} \subset \bigcup_{k=0}^{p} U_k\right|$$
$$\leq \Psi(u)(1 + K_2\cdot\varepsilon) + \Psi(u) \underbrace{\sum_{1\leq k\leq p} e^{-2^{2k}\left(\frac{1}{4\varepsilon} - K_6\varepsilon\right)}}_{\leq K_7\cdot\varepsilon} \leq \Psi(u)(1 + K_8\cdot\varepsilon)$$

for large u and $\varepsilon > 0$ arbitrary small. Theorem 3.1.10 is proved. □

The Borell-Tsirelson-Ibragimov-Sudakov inequality is a mighty tool in the investigation of Gaussian random functions. It has a number of important corollaries, for instance, the follows one.

Corollary 3.1.3
Let $X = \{X(t), t \in T\}$ be a centered separable Gaussian random function. The following statements are equivalent:

1. $P\left(\sup\limits_{t\in T} X(t) < \infty\right) = 1$
2. $\mathbf{E} \sup\limits_{t\in T} X(t) < \infty$
3. $\mathbf{E} \exp\left(s(\sup\limits_{t\in T} X(t))^2\right) < \infty$ for all sufficiently small $s > 0$.

It is known for any random variables $Z \geq 0$ a.s. that $\mathbf{E}\, e^{sZ} < \infty$, $s > 0 \;\Rightarrow\; \mathbf{E}Z < \infty \;\Rightarrow\; P(Z < \infty) = 1$. Hence, the assertions of Corollary 3.1.3 that are specific to the Gaussian case, are 1) ⇒ 2) ⇒ 3).

Proof of Corollary 3.1.3 1) $\Rightarrow$ 2). By Lemma 3.1.7, the tail probability function of $Z = \sup_{t\in T} X(t)$ is exponentially bounded for large argument values, which yields
$\mathbf{E}Z = \int_0^\infty P(Z > u)du \le \int_0^w \underbrace{P(Z > u)}_{\le 1}\, du + \int_w^\infty \Psi(\frac{u-w}{\sigma_T})du \le w + \mathbf{E}(\sigma_T Y) < \infty$, where w as in Lemma 3.1.7 and $Y \sim N(0,1)$.
2) $\Rightarrow$ 3). For any $s > 0$ it holds $\mathbf{E}e^{sZ^2} = \int_0^\infty P(e^{sZ^2} > u)\, du = \int_0^\infty P(Z > \sqrt{1/s\ \log\ u})\, du$ since $\mathbf{E}Z < \infty$, it follows Z a.s. bounded, and hence for $w \in \mathbb{R}$ as in Lemma 3.1.7 and $u_0 = e^{sw^2}$

$$\mathbf{E}\ e^{s\ Z^2} \overset{L.\ 3.1.7}{\le} \int_0^{u_0} P\ (Z > \sqrt{\log\ u^{1/s}})\ du\ +\ \int_{u_0}^\infty \Psi\ \left(\frac{\sqrt{(\log u)1/s} - w}{\sigma_T}\right) du$$

$$\le \Big|\ \text{Substitution } u = e^{s(\sigma_T t + w)^2}\Big| u_0 + \int_0^\infty \Psi(t)\ 2s\ (\sigma_T t + + w)\ \sigma_T\ e^{s\sigma_T^2 t^2 + 2s\sigma_T tw + sw^2} dt$$

$$= u_0 + \ 2s\sigma_T\ e^{sw^2} \int_0^\infty (\sigma_T t + w)\ \Psi(t)\ e^{s\sigma_T^2 t^2 + 2s\sigma_T w\cdot t}\ dt$$

$$\overset{L.\ 3.1.6}{\le} u_0 + \ 2s\sigma_T\ e^{sw^2} \int_0^\infty (\sigma_T + w/t)\ \varphi(t)\ e^{s\sigma_T^2 t^2 + 2s\sigma_T wt}\ dt < \infty$$

if $s > 0$ is small enough so that $-1/2 + s\sigma_T^2 < 0$, where $\varphi(t) = \frac{1}{\sqrt{2\pi}}\ e^{-t^2/2}$. □

3.1.4 Comparison of Gaussian random functions

Let $X = \{X(t), t \in T\}$ and $Y = \{Y(t), t \in T\}$ be two a.s. bounded separable Gaussian processes. Now can they be compared?

Definition 3.1.5 (Stochastic order):
Let X_1 and X_2 be two random variables on a probability space $(\Omega, \mathcal{F}, P)$. One says that X_1 is stochastically smaller than or equal to X_2 ($X_1\ \le_{st}\ X_2$) if $P(X_1 > u)\ \le\ P(X_2 > u)\ \forall u \in \mathbb{R}$.

Theorem 3.1.11 (Slepian's inequality (1963)):
Let X and Y be Gaussian random functions as above, centered and such that $Var\ X(t)\ =\ Var\ Y(t),\ t \in T$,

$$\gamma_X(s,t)\ \le\ \gamma_Y(s,t),\ \forall s,t \in T, \tag{3.1.21}$$

where γ_X and γ_Y are variograms of X and Y, respectively. Then,

$$\sup_{t\in T} X(t) \le_{st} \sup_{t\in T}\ Y(t), \tag{3.1.22}$$

and

$$\mathbf{E}\ \sup_{t\in T} X(t) \le \mathbf{E}\ \sup_{t\in T}\ Y(t). \tag{3.1.23}$$

Remark 3.1.6
1) Since $Var\ X(t)\ =\ \mathbf{E}X^2(t)\ =\ Var\ Y(t)\ =\ \mathbf{E}Y^2(t), t\ \in\ T$ the condition $\gamma_X(s,t)\ =$

$\frac{1}{2}\mathbf{E}\ (X(s)-X(t))^2 \le\ \frac{1}{2}\ \mathbf{E}(Y(s)-Y(t))^2\ =\ \gamma_Y(s,t),\ s,t\in T$ is equivalent to $C_X(s,t)\ge C_Y(s,t),\ s,t\in T$, where $C_X(s,t)=\mathbf{E}(X(s)X(t))$ and $C_Y(s,t)=\mathbf{E}Y(s)Y(t)$ are covariance functions of X and Y.

2) Inequality (3.1.23) is a simple corollary of (3.1.22), which is an intrinsic property of stochastic order: if $X_1\ \le_{st}\ X_2$ then $\mathbf{E}X_1\ =\ \int\limits_0^\infty\ P(X_1\ >\ u)du\ -\ \int\limits_{-\infty}^0\ P(X_1\ \le\ u)du$ $\le\ \int\limits_0^\infty\ P(X_2>u)du-\int\limits_{-\infty}^0\ P(X_2\ \le u)du=\mathbf{E}X_2.$

3) It is essential in inequality (3.1.22) that X and Y are without an absolute value. Namely, inequality $\sup\limits_{t\in T}\ |X(t)|\ \le_{st}\ \sup\limits_{t\in T}\ |Y(t)|$ does not follow from condition (3.1.21).

Exercise 3.1.4
Give a counterexample!

To prove Theorem 3.1.11, we need the following Lemma.

Lemma 3.1.9
Let $X=(X_1,\dots,X_n)^T\sim N(0,C)$, $C=(c_{ij})_{i,j=1}^n$. Let $h:\mathbb{R}^n\to\mathbb{R}$ satisfy $h\in\ C^2(\mathbb{R}^n)$, $\frac{\partial^2h(x)}{\partial x_{i_0}\partial x_{j_0}}\ge 0$ for all $x\in\mathbb{R}^n$ and some $1\le i_0,j_0\le n$,

$$\frac{h(x)}{|x|^d}=\underset{|x|\to\infty}{o(1)}\ ,\quad \frac{1}{|x|^d}\frac{\partial h(x)}{\partial x_i}=o(1),\ |x|\to\infty,\quad \frac{1}{|x|^d}\ \frac{\partial^2h(x)}{\partial x_i\partial x_j}=o(1),\ |x|\to\infty \tag{3.1.24}$$

for some $d>0$ and all $i,j=1,...,n$.

Then $H(C)=\mathbf{E}h(X_1,\dots,X_n)$ is a non-decreasing function of c_{i_0,j_0}.

Proof Show that $\frac{\partial H(C)}{\partial c_{i_0,j_0}}\ge 0$ if $\frac{\partial^2h(x)}{\partial x_{i_0}\partial x_{j_0}}\ge 0$. Assume that $\det(C)\ne 0$. Let φ_X be the density of $X=(X_1,\dots,X_n)^T\sim N(0,C)$: $\varphi_X(x)=\frac{1}{(2\pi)^{n/2}\sqrt{\det C}}\ e^{-x^TC^{-1}x/2},x\in\mathbb{R}^n$. It can be easily seen that

$$\frac{\partial\varphi_X}{\partial c_{ii}}=\frac{1}{2}\ \frac{\partial^2\varphi_X}{\partial x_i^2},\qquad \frac{\partial\varphi_X}{\partial c_{ij}}=\frac{\partial^2\varphi_X}{\partial x_i\partial x_j},i\ne j,\ i=1,\dots,n,\ j=1,\dots,n. \tag{3.1.25}$$

To calculate $\frac{\partial\varphi_X}{\partial c_{ij}}$, write $\frac{\partial\varphi_X}{\partial c_{ij}}=\frac{\partial\varphi_X}{\partial\kappa_{ij}}\ \frac{\partial\kappa_{ij}}{\partial c_{ij}}$, where $C^{-1}=(\kappa_{ij})_{i,j=1}^n$. Integrating by parts twice, we get

$$\begin{aligned}\frac{\partial H(C)}{\partial c_{i_0j_0}}&=\frac{\partial}{\partial c_{i_0j_0}}\int\limits_{\mathbb{R}^n}h(x)\varphi_X(x)dx=\int\limits_{\mathbb{R}^n}h(x)\frac{\partial\varphi_X(x)}{\partial c_{i_0j_0}}dx\\&\overset{(3.1.25)}{=}\int\limits_{\mathbb{R}^n}h(x)\ \frac{\partial^2\varphi_X(x)}{\partial x_{i_0}\partial x_{j_0}}\ dx\ \overset{(3.1.24)}{=}\int\limits_{\mathbb{R}^n}\underbrace{\frac{\partial^2h(x)}{\partial x_{i_0}\partial x_{j_0}}}_{\ge 0}\ \varphi_X(x)dx\ \ge 0.\end{aligned}$$

For a singular C, approximate it by a sequence of C_k with $\det(C_k)\ne 0$, $C_k\to\ C,k\to\infty$. □

Proof of Theorem 3.1.11 Since X and Y are separable, it is enough to prove this theorem for a finite T. Indeed, in the case of general T, let T_n be a sequence of finite subsets of T such that $T_n\subset T_{n+1}$, and $T_n\to\widetilde{T}$, where $\widetilde{T}\subset T$ is a dense subset of T. Then

$\sup_{t\in T_n} X(t) \overset{a.s.}{\underset{n\to\infty}{\uparrow}} \sup_{t\in\widetilde{T}} X(t) = \sup_{t\in T} X(t)$, and by monotonicity of this convergence, it holds $P(\sup_{t\in T_n} X(t) > u) \underset{n\to\infty}{\to} P(\sup_{t\in T} X(t) > u)$, $\mathbf{E}\sup_{t\in T_n} X(t) \underset{n\to\infty}{\to} \mathbf{E}\sup_{t\in T} X(t)$.
(The same convergence holds for the random field Y).
So let $T = \{t_1, \ldots, t_n\}$. Take $h(x) = \prod_{i=1}^{n} h_i(x_i),\ x \in \mathbb{R}^n$, where $h_i : \mathbb{R} \to \mathbb{R}_+$ is non-increasing, $h_i \in C^2(\mathbb{R})$ satisfying all assumptions of Lemma 3.1.9 Notice that $\frac{\partial^2 h(x)}{\partial x_i \partial x_j} = h_i'(x_i)\ h_j'(x_j) \times \prod_{k\neq i,j} h_k(x_k) \geq 0$. Since $C_X(t_i, t_j) \geq C_Y(t_i, t_j)$. $i, j = 1, \ldots, n$ by Remark 3.1.6, 1), using Lemma 3.1.9 we get $\mathbf{E}\prod_{i=1}^{n} h_i(X(t_i)) \geq \mathbf{E}\prod_{i=1}^{n} h_i(Y(t_i))$. Take now $\{h_i^{(k)}\}_{k=1}^{\infty}$ to be a sequence of positive non-increasing C^2-smooth functions as above s.t. $h_i^{(\kappa)}(x) \underset{\kappa\to\infty}{\to} \mathbf{1}(x \leq u)$. It follows that

$$P(\sup_{t\in T} X(t) \leq u) = \mathbf{E}\mathbf{1}(X(t_i) \leq u,\ i = 1, \ldots, n) = \mathbf{E}\prod_{i=1}^{n} \mathbf{1}(X(t_i) \leq u)$$
$$= \lim_{\kappa\to\infty} \mathbf{E}\prod_{i=1}^{n} h_i^{(\kappa)}(X(t_i)) \geq \lim_{\kappa\to\infty} \mathbf{E}\prod_{i=1}^{n} h_i^{(\kappa)}(Y(t_i)) = \ldots = P(\sup_{t\in T} Y(t) \leq u),$$

$\forall u \in \mathbb{R}$, which implies $P(\sup_{t\in T} X(t) > u) = 1 - P(\sup_{t\in T} X(t) \leq u) \leq 1 - P(\sup_{t\in T} Y(t) \leq u)$
$= P(\sup_{t\in T} Y(t) > u),\ u \in \mathbb{R}$. □

One of the most important extensions of Slepian's inequality is the following Sudakov-Fernique inequality:

Theorem 3.1.12
Let X and Y be two random functions as in the beginning of Section 3.1.4. Assume that $\mathbf{E}X(t) \leq \mathbf{E}Y(t),\ t \in T$ and that $\gamma_X(s,t) \leq \gamma_Y(s,t),\ s,t \in T$. Then $\mathbf{E}\sup_{t\in T} X(t) \leq \mathbf{E}\sup_{t\in T} Y(t)$.

Without proof. Notice that the weaker ordering of expectations (and not $\leq_{st}$) is obtained under more general conditions (we do not require that X and Y are centered with identical variance).

3.1.5 Entropy bounds

By Lemma 3.1.7, we have the inequality $P(\sup_{t\in T} X(t) \geq u) \leq C_1\ e^{-(u-w)^2/(2\sigma_T^2)},\ u \geq w$ for a separable a.s. bounded Gaussian random function $X = \{X(t),\ t \in T\}$, where $C_1 > 0$ is a constant and $\sigma_T^2 = \sup_{t\in T} \mathbf{E}X^2(t)$. Our goal is to improve this tail bound by imposing some extra conditions on the entropy of X.

Theorem 3.1.13
Let $X = \{X(t), t \in T\}$ be a centered a.s. continuous Gaussian random function with entropy $N(\varepsilon), \varepsilon > 0$. If $N(\varepsilon) \leq \kappa\varepsilon^{-\alpha}$ for $\varepsilon > 0$ and some $\alpha > 0$ then $P(\sup_{t\in T} X(t) \geq u)$ $\leq C_\alpha \cdot\ u^{\alpha+\beta} e^{-u^2/(2\sigma_T^2)}$ for sufficiently large $u > 0$, every $\beta > 0$ and $C_\alpha = C(\kappa, \alpha, \sigma_T^2) > 0$.

Proof Let $\rho(\cdot,\cdot)$ be the canonical metric of X. For any $\varepsilon > 0$ introduce $\mu(t, \varepsilon) = \mathbf{E}\sup_{s\in B_\varepsilon(t)} X(s)$,

$\mu(\varepsilon) = \sup_{t\in T} \mu(t,\varepsilon)$, where $B_\varepsilon(t)$ is the ball with center t and radius $\varepsilon > 0$ in the metric ρ.
Since $T \subseteq \bigcup_{i=1}^{N(\varepsilon)} B_\varepsilon(t_i)$ for some $t_1, \ldots, t_{N(\varepsilon)} \in T$, Lemma 3.1.7 yields

$$P(\sup_{t\in T} X(t) \geq u) \leq \sum_{i=1}^{N(\varepsilon)} P(\sup_{t\in B_\varepsilon(t_i)} X(t) \geq u) \leq N(\varepsilon) C_1 e^{-(u-\mu(\varepsilon))^2/(2\sigma_T^2)} \tag{3.1.26}$$

for $u \geq \omega = \mu(\varepsilon)$. By Theorem 3.1.2, we get $\mu(t,\varepsilon) \leq \widetilde{C}_2 \cdot \int_0^\varepsilon \sqrt{\log N(\delta)} d\delta \overset{\text{by assum. on } N(\varepsilon)}{\leq}$
$\widetilde{C}_2 \int_0^\varepsilon (\log \kappa - \alpha \log \delta)^{1/2} d\,\delta \overset{\text{for small } \varepsilon > 0}{\leq} \widetilde{C}_3 \sqrt{\alpha} \int_0^\varepsilon (\log 1/\delta)^{1/2} d\delta \leq \widetilde{C}_4 \varepsilon (\log 1/\varepsilon)^{1/2}$, since

$$\int_0^\varepsilon (\log 1/\delta)^{1/2} d\delta = \delta \sqrt{\log 1/\delta}\,|_0^\varepsilon - \int_0^\varepsilon \delta \frac{1}{2} \frac{1}{\sqrt{-\log \delta}} \left(-\frac{1}{\delta}\right) d\delta$$
$$= \varepsilon \sqrt{\log 1/\varepsilon} + \frac{1}{2} \int_0^\varepsilon \frac{d\,\delta}{\sqrt{-\log \delta}} \leq \varepsilon \sqrt{\log 1/\varepsilon} + \frac{1}{2} \frac{\varepsilon}{\sqrt{-\log \varepsilon}} \leq 2\varepsilon \sqrt{\log(1/\varepsilon)}.$$

Set $\varepsilon = 1/u$ and chose u large so that $u > \widetilde{C}_4 \varepsilon (\log 1/\varepsilon)^{1/2}$. Then $u - \mu(\varepsilon) \geq 0$, and $\mu(\varepsilon) \leq \widetilde{C}_4 u^{-1} \sqrt{\log u}$, $N(\varepsilon) \leq K u^\alpha$. Applying this to inequality (3.1.26), we get

$$P(\sup_{t\in T} X(t) \geq u) \leq \widetilde{C}_5 u^\alpha \exp\left(\frac{(u - \widetilde{C}_4 u^{-1} \sqrt{\log u})^2}{2\delta_T^2}\right)$$
$$= \widetilde{C}_5 u^\alpha \exp\left(-\frac{u^2}{2\sigma_T^2} + \frac{1}{\sigma_T^2} \widetilde{C}_4 \sqrt{\log u} - \widetilde{C}_4^2 u^{-2} \log \frac{u}{2\sigma_T^2}\right)$$
$$= \widetilde{C}_5 \underbrace{e^{-\widetilde{C}_4^2/(2\sigma_T^2)} \cdot u^{-2 \log u}}_{\leq 2} u^\alpha \underbrace{e^{\widetilde{C}_6 \sqrt{\log u}}}_{\leq u^\beta} \cdot e^{-u^2/(2\sigma_T^2)} \leq \widetilde{C}_7 u^{\alpha+\beta} u^{-u^2/(2\sigma_T^2)},$$

for $\forall \beta > 0$ and $u \to \infty$. □

Several extensions of the bound in Theorem 3.1.13 are thinkable. We formulate just one of those due to M. Talagrand (1994).

Theorem 3.1.14
Le X be as in Theorem 3.1.13. If $N(\varepsilon) \leq (A/\varepsilon)^\alpha$ for some $A > \sigma_T$, $\alpha > 0$ and $\varepsilon \in (0, \varepsilon_0)$, where $\varepsilon_0 \in [0, \sigma_T]$ then $P(\sup_{t\in T} X(t) \geq u) \leq \frac{(C_1 \cdot A\ u)^\alpha}{(\sqrt{\alpha}\sigma_T^2)^\alpha} \Psi(u/\sigma_T)$ for all $u \geq \sigma_T^2(1 + \sqrt{\sigma})/\varepsilon_0$ where $C_1 > 0$ is a constant.

(Without proof).

3.1.6 Logarithmic large deviations

Going to the logarithmic scale, the following asymptotic result for large durations holds for **any** a.s. bounded Gaussian random functions:

Theorem 3.1.15
Let $X = \{X(t),\ t \in T\}$ be an a.s. bounded separable Gaussian random function with $\sigma_T^2 = \sup\limits_{t\in T} Var\, X(t)$. Then there exists $\mu \in \mathbb{R}$ s.t. $\lim\limits_{u\to+\infty} \frac{1}{u}(\log\ P\,(\sup\limits_{t\in T}\ X(t) > u) + (u+\mu)^2_{/(2\sigma_T^2)}) = 0$.

Before proving this theorem, a remark and a lemma are in order:

Remark 3.1.7
It follows from Lemma 3.1.6 that for a random variable $\xi \sim\ N(-\mu, \sigma_T^2)$ holds $\lim\limits_{u\to+\infty} \frac{1}{u}\ (\log P(\xi > u) + (u+\mu)^2/(2\sigma_T^2)) = 0$, which means that the tails of $\sup\limits_{t\in T}\ X(t)$ and ξ are equivalent at the logarithmic scale.

Lemma 3.1.10
Let $F(x) = P(\sup\limits_{t\in T} X(t) \le x),\ x \in \mathbb{R}$ be the c.d.f. of $\sup\limits_{t\in T} X(t)$. Then the function $f(u) = \sigma_T \Phi^{-1}(F(u)) - u$ is non-decreasing, and there exists $\mu = \lim\limits_{u\to+\infty} f(u)$. Here $\Phi(\cdot)$ is the c.d.f. of $N(0,1)$-law.

Proof of Theorem 3.1.15 Introduce the notation $g(u) = \Phi^{-1}(F(u))$. Then, $P(\sup\limits_{t\in T}\ X(t) > u) = 1 - \Phi(g(u)),\ u \in \mathbb{R}$. For $u \to +\infty$, it holds $g(u) \to +\infty$. By Remark 3.1.7, we have for the $N(0,1)$-law $\lim\limits_{r\to+\infty}\ \frac{1}{r}\ (\log(1-\Phi(r)) + r^2/2) = 0$. Making the substitution $r \longmapsto g(u),\ u \to +\infty$, we get $0 = \lim\limits_{u\to+\infty} \frac{1}{g(u)}(\log \underbrace{(1 - \Phi(g(u)))}_{P(\sup\limits_{t\in T}\ X(t)>u)} + \frac{g^2(u)}{2})$.

By Lemma 3.1.10, it holds $\mu = \lim\limits_{u\to+\infty}\ (\sigma_T g(u) - u)$, which yields $g(u) \underset{u\to+\infty}{\sim} \frac{u+\mu}{\sigma_T}$ and hence

$$\lim_{u\to+\infty} \frac{\sigma_T}{u} \left(\log\ P(\sup_{t\in T} X(t) > u) + \frac{(u+\mu)^2}{2\sigma_T^2}\right) = 0. \qquad \square$$

To prove Lemma 3.1.10, we need some tools from convex geometry.

3.1.7 Convexity and isoperimetric inequality for Gaussian measures

Let L be a linear topological space.

Definition 3.1.6
A functional $f : L \to (-\infty, +\infty]$ is
1) *lower convex*, if $\forall\ x, y \in L$ and $\forall\ \gamma \in [0,1]\ f(\gamma\ x + (1-\gamma)y)\ \le\ \gamma f\ (x) + (1-\gamma)f(y)$.
2) *lower semicontinuous*, if $\liminf\limits_{x\to y}\ f(x)\ \ge\ f(y).\ \forall\ y \in L$.

For lower semicontinuous functions f, the sets $\Pi_{f,u} = \{x \in L : f(x) \le u\}$ are closed in L. Let $\nu(L)$ be the class of all lower semicontinuous lower convex functionals on L. For $f \in \nu(L)$, it holds that they are measurable w.r.t. the Borel σ-algebra in L.

Example 3.1.1
Let $L = \mathbb{R}^T$, T be arbitrary space. Then $f(x) = \sup\limits_{t\in T}\ x(t)\ \ \in \nu(L)$.

Definition 3.1.7
Let P be a probability measure on (L, B_L), and $f \in \nu(L)$.
1) The distribution of f is $P_f = P \circ f^{-1}$.
2) The distribution function of f is $F_f(u) = P(\Pi_{f,u}) = P(\{x \in L : f(x) \le u\}),\ u \in \mathbb{R}$.

Clearly, P_f is a probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.
Let $X = (X_1, \dots, X_n)^T \sim N(\mu,\ K)$ be an $n-$dim. Gaussian random vector on a probability space $(\Omega, \mathcal{F}, P)$. Then it is known that $X \stackrel{d}{=} \mu + K^{1/2}Y$, where $Y \sim N(0, I)$ is a standard normal random vector. Define the *deviation ellipsoid* of X by $\mathcal{E}_r = K^{1/2} \cdot B_r(0)$, where $B_r(0)$ is the r- neighborhood of the origin, $r > 0$. Then the parallel set of any $A \subset \mathbb{R}^n$ is defined by $A_r = A + \mathcal{E}_r$. The following *isoperimetric inequality* for the Gaussian measure $P_X(\cdot) = P(X \in \cdot)$ of X can be proven by symmetrization techniques (see [Lifshits], Theorem. 1.5):

Theorem 3.1.16
For any $A \in \mathcal{B}_{\mathbb{R}^n}$ and $X \sim N(\mu, K), r > 0$ it holds $\Phi^{-1}(P_X(A_r)) \ \geq\ \Phi^{-1}(P_X(A)) + r$.

The measure of any Borel set A is minimal among all "larger" sets A_r.

The next inequality states convexity of Gaussian measures:

Theorem 3.1.17 (Ehrhard inequality):
Let $X \sim N(\mu, K),\ \gamma \in [0, 2)$, and A, B be two non-empty convex subsets of $\mathbb{R}^n$. Then $\Phi^{-1}\ (P_X(\gamma A + (1-\gamma)B)) \geq \gamma\ \Phi^{-1}\ (P_X(A)) + (1-\gamma)\ \Phi^{-1}(P_X(B))$.

Now let us transfer these inequalities to Gaussian measures on general *linear topological spaces (LTS).*

Definition 3.1.8
A LTS L is called *locally convex* if Γ the there exists the base of topology in L which contains only convex sets. That is, for any open $V \subset L$ and $x \in V$ exists convex open $U \subset X$ s.t. $x \in U \subset V$.

From now is, we assume L to be a locally convex LTS.

Definition 3.1.9
1. The *algebra of cylindric* sets C_0 of L is the family of subsets of L of the form
$\{x \in L : (f_1(x), \dots, f_n(x) \in A\}$, $A \in \mathcal{B}_{\mathbb{R}^n}, n \in \mathbb{N},\ f_1, \dots,\ f_n \in L^*$, where L^* is the dual space to L (the space of all linear continuous functionals on L). Such sets are celled cylinders.
2. $C = \sigma(C_0)$ is celled *cylindric σ-Algebra.*

Let B_L be the Borel σ-Algebra of L, and $Z \subset B_L$ any algebra.

Definition 3.1.10
1. A set function $M : Z \to \mathbb{R}_+$ which is monotone w.r.t. inclusion (i.e., $A, B \in Z, A \subset B \ \Rightarrow\ M(A) \leq M(B)$) is called *Radon*, if $\forall A \in Z \mu(A) = \sup\{M(B) : B \subset A, B \in Z - \text{compact}\}$. If μ is a measure then M is called a *Radon measure.*
2. The *outer measure* for M is $M^*(A) = \inf\{M(B) : A \supset B,\ B \in Z\}$, which is defined for all $A \subseteq L$. The *inner measure* of M is $M_*\ (A) = \sup\ \{M(B) : B \subset A,\ B \in Z\},\ A \subseteq L$.

It can be proven that the values of a Radon measure on C_0 completely define if, i.e., for two Radon measures M and Λ on B_L "$M(B) = \Lambda(B),\ B \in C_0$" $\Rightarrow$ "$M(B) = \Lambda(B),\ B \in B_L$".

Let now P be a probability measure on (L, C).

Definition 3.1.11
1. The *mean* of P is a $\mu \in L$ s.t. $f(\mu) = \int\limits_L f(x)P(dx)\ \forall f \in L^*$. If $\mu = 0$ then P is celled centered.
2. A linear operator $K : L^* \to L$ is called *covariance operator* of P if $\forall f, g \in L^*$ it holds $f(K_g) = \int\limits_L f(x-\mu)g(x-\mu)P(dx)$. It is clear that K is *self-adjoint*, i.e., $f(K_g) = g(K_f),\ \forall f, g \in L^*$.

3. The *characteristic functional* of P is a function $\varphi_P : L^* \to \mathbb{R}$ given by $\varphi_P(f) = \int\limits_L e^{if(x)} P(dx)$, $f \in L^*$. It can be shown that for Radon measures M and Λ on B_L $\varphi_M = \varphi_\Lambda$ yields $M = \Lambda$.

Definition 3.1.12
1) A probability measure P on a σ-Algebra $\mathcal{F}$, s.t. $C \subset \mathcal{F}$ is called *Gaussian* if for all $f \in L^*$ the distribution P_f is univariate Gaussian.
2) If P is Gaussian with mean $\mu \in \mathbb{R}$ and covariance operator K we write $P = N(\mu, K)$. The class of all Radon Gaussian measures on B_L will be denoted by $G(L)$.

Lemma 3.1.11
If $P = N(\mu, K)$ then its characteristic functional is given by $\varphi_P(f) = e^{if(\mu) - \frac{1}{2} f(K_f)}$, $f \in L^*$.

Proof It is clear that $\forall f \in L^*$ $P_f = N(f(\mu), f(K_f))$ by Definitions 3.1.11 and 3.1.12, since $E_{P_f} = \int\limits_{\mathbb{R}} x dP_f(x) = \int\limits_L f(y) P(dy) = f(\mu)$, $Var_{P_f} = \int\limits_{\mathbb{R}} (x - E_{P_f})^2 dP_f(x) = \int\limits_L (f(y - E_{P_f}))^2 P(dy) = f(K_f)$. Since $\varphi_P(f) = \int\limits_L e^{if(y)} P(dy) = \int\limits_{\mathbb{R}} e^{ix} P_f(dx) = exp\{i\ E_{P_f} - \frac{1}{2}\ Var_{P_f}\}$, we are done. □

Since characteristic functionals define their measures in a unique way, the mean $\mu \in \mathbb{R}$ and K (covariance operator) define $N(\mu, K)$ uniquely. Now it is possible to generalize Ehrhard inequality to Gaussian measures on locally convex LTS L:

Theorem 3.1.18 (Ehrhard inequality):
Let $P \in G(L)$, and let A, B be non-empty convex Borel subsets of L, $\gamma \in (0, 1)$. Then

$$\Phi^{-1}(P_*(\gamma A + (1-\gamma)B)) \geq \gamma\ \Phi^{-1}(P(A)) + (1-\gamma)\Phi^{-1}(P(B)).$$

In particular, if A and B are closed then

$$\Phi^{-1}\ (P(\overline{\gamma A + (1-\gamma)B})) \geq \gamma\ \Phi^{-1}(P(A)) + (1-\gamma)\Phi^{-1}(P(B)). \tag{3.1.27}$$

(without proof).
Similarity, the isoperimetric inequality of Theorem 3.1.16 rewrites

Theorem 3.1.19 (Isoperimetric inequality):
Let $P \in G(L)$, $A \in B_L$, $r > 0$. Then $\Phi^{-1}(P(A_r)) \geq \Phi^{-1}(P(A)) + r$, where $A_r = A + r\mathcal{E}$ and $\mathcal{E}$ is the derivation ellipsoid of P.

Lemma 3.1.12
Let $f \in \nu(L)$, $P \in G(L)$. Then the function $\Psi(u) = \Phi^{-1}\ (F_f(u))$, $u \in \mathbb{R}$ is concave.

Proof For any $u_1, u_2 \in \mathbb{R}$ and $\gamma \in [0, 1]$, we have to show that

$$\psi(\gamma u_1 + (1-\gamma)u_2) \geq \gamma\psi(u_1) + (1-\gamma)\psi(u_2).$$

Since $f \in \nu(L)$, sets $\Pi_{f,u}$ are closed and convex in L. Moreover, it follows from the convexity of f that $(\gamma\Pi_{f,u_1} + (1-\gamma)\Pi_{f,u_2}) \subset \Pi_{f,\gamma u_1 + (1-\gamma)u_2}$, since $f(\gamma x + (1-\gamma)y) \leq \gamma f(x) + (1-\gamma)f(y) \leq \gamma u_1 + (1-\gamma)u_2$, if $x \in \Pi_{t,u_1}, y \in \Pi_{t,u_2}$, hence $\gamma\{x \in L : f(x) \leq u_1\} + (1-\gamma)\{x \in L : f(x) \leq u_2\} \subseteq \{x \in L :\ f(x) \leq \gamma u_1 + (1-\gamma)u_2\}$. By Theorem 3.1.18, it holds

$$\underbrace{\Phi^{-1}\ (P(\Pi_{f,\gamma u_1 + (1-\gamma)u_2}))}_{\psi(\gamma u_1 + (1-\gamma)u_2)} \geq \underbrace{\gamma\ \Phi^{-1}\ (P(\Pi_{f,u_1}))}_{\psi(u_1)} + (1-\gamma)\ \underbrace{\Phi^{-1}\ (P(\Pi_{f,u_2}))}_{\psi(u_2)},$$

which means that ψ is concave. □

Proof of Lemma 3.1.10 By the above Lemma, $\psi(u) = \Phi^{-1}(F(u))$ is concave, then $f(u) = \sigma_T \Phi^{-1}(F(u)) - u$ is concave either. Since $\sigma_T > 0$ and a linear function is both convex and concave. Then, $f \nearrow$ and $\exists \lim_{u \to +\infty} f(u) = \mu \in \overline{\mathbb{R}}$. Let us show that $\mu < \infty$. □

Bibliography

[1] R. Adler and J. Taylor. *Random Fields and Geometry.* Springer, New York, 2007.

[2] R. Askey. Refinements of abel summability for jacobi series, in: Harmonic analysis on homogeneous spaces, ed. by C.C. More. *Proc. Symp. Pure Math. XXVI*, pages 335–338, 1973.

[3] J.-M. Azais and M. Wschebor. *Level Sets and Extrema of Random Processes and Fields.* Wiley, New York, 2009.

[4] N. H. Bingham, C. M. Goldie, and J. L Tengels. *Regular variable.* Cambridge Univ. Press, 1987.

[5] S. Bochner. *Lectures on Fourier integrals.* Princeton Univ. Press, 1959.

[6] V. I. Bogachev. *Gaussian Measures.* AMS, 1998.

[7] P. Bremaud. *Markov Chains, Gibbs Fields, Monte Carlo Simulation and Queues.* Springer, 1999.

[8] A. V. Bulinski and A. P. Shashkin. *Limit theorems for associated random fields and related systems.* World Scientific, Singapore, 2007.

[9] A. V. Bulinski and A. N. Shiryaev. *Theory of stochastic processes.* Fizmatlit, 2005. (in Russian).

[10] J.-P. Chiles and P. Delfiner. *Geostatistics: Modeling Spatial Uncertainty.* Wiley, New York, 1999.

[11] H. Cramér and M. Leadbetter. *Stationary and Related Stochastic Processes.* Wiley, New York, 1967.

[12] N. Cressie. *Statistics for spatial data.* World Scientific, Singapore, 2007.

[13] J. L. Doob. *Stochastic Processes.* Wiley, 1953.

[14] R. M. Dudley. *Uniform Central Limit Theorems.* Cambridge Univ. Pr., 1999.

[15] X. Fernique. *Fonctions aléatoires gaussiennes, vecteurs aléatoires gaussiens.* CRM, Montreal, 1997.

[16] J.C. Ferreira and V. A. Menegatto. *"Eigenvalues of integral operators defined by smooth positive definite kernels" Integral and operator Theory.* Birkhäuser Basel, Volumne 64, No. 1, p. 61-61, 2009.

[17] H.-O. Georgii. *Gibbs Measures and Phase Transitions.* de Gruyter, Berlin, 1988.

[18] H.-O. Georgii. *Random Fields on a Network*. Springer, 1995.

[19] I. I. Gikhman and A. V. Skorokhod. *The Theory of Stochastic Processes*. Springer, Berlin, 1974.

[20] T. Gneiting. Radial positive definite functions generated by Euclid's hat. *J. Multivar. Anal.*, 69:88–119, 1999.

[21] T. Gneiting and Z. Sasvári. The characterization problem for isotropic covariance functions. *Math. Geol.*, 31:105–111, 1999.

[22] I. Ibragimov, F. Götze, and D. Zaporozhets. *On a random algebraic surfaces*. Preprint, 2003.

[23] I. Ibragimov and N.B. Maslova. *The mean number of real zeros of random polynomials, I. Coefficients with zero mean*, volume 16. Theory of Probability and its Applications, 1971.

[24] I. Ibragimov and S. Podkorytov. *On random real algebraic surfaces*. Dokl. Akad. Nauk, 343 edition, 1995.

[25] I. Ibragimov and D. Zaporozhets. *On the distribution of roots of random polynomials*. Preprint, The, Bielefeld, 2006.

[26] I. Ibragimov and O. Zeitouni. *On roots of random polynomials*. Trans. AMS, 347 edition, 1995.

[27] K. Itô and M. Nisio. *On the convergence of sums of independent Banach space valued random variables*. Osaka J. Math. 5, 35-48, 1968.

[28] A. V. Ivanov and N. N. Leonenko. *Statistical Analysis of Random Fields*. Kluwer, 1989.

[29] H. S. Janzer, F. Raudies, H. Neumann, and F. Steiner. Image processing and feature extraction from a perspective of computer vision and physical cosmology. *to be published in Mathematical Analysis of Evolution, Information and Complexity*, 2009.

[30] M. Kac. *On the average number of real roots of a random algebraic equation*. Bull. Amer. Math. Sve. 49, 1943.

[31] D. Khoshnevisan. *Multiparameter Processes: An Introduction to Random Fields*. Springer, 2002.

[32] A. N. Kolmogorov and S. V. Fomin. *Elements of the Theory of Functions and Functional Analysis*. Graylock Press, Rochester, 1961.

[33] H. König. *Eigenvalue distribution of compact operators*. Birkhäuser, 1986.

[34] C. Lantuejoul. *Geostatistical Simulation - Models and Algorithms*. Springer, 2002.

[35] M. Ledoux and M. Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes*. Springer, 1991.

[36] M. Leonenko. *Limit Theorems for Random Fields with Singular Spectrum*. Kluwer, 1999.

[37] M. A. Lifshits. *Gaussian Random Functions*. Kluwer, 1995.

[38] W. Magnus, F. Oberhettinger, and R. P. Soni. *Stable non-Gaussian random processes.* Chapman & Hall/CRC, 1994.

[39] V. A. Malyshev and R. A. Minlos. *Gibbs Random Fields: Cluster Expansions.* Kluwer, 1991.

[40] B. Matern. *Spatial Variation.* Springer, second edition, 1986.

[41] A. Nazarov and D. Zaporozhets. *What is the least expected numer of real root of a random polynomial?*, volume 53. Theory of Probability and Applications, 2009.

[42] H. H. Ngy. Semiinvariants in the theory of characterization of probability distributions. *J. Math. Sci.*, 35(2):2437–2438, 1986.

[43] K. Norio. *Recent development on random fields and their sample paths. Part I:sample path continuity of random fields*, volume 16. Soochow J. Math., 1990.

[44] V. I. Piterbarg. *Asymptotic Methods in the Theory of Gaussian Processes and Fields.* AMS, 1996.

[45] A. Ramm. *Random Fields Estimation.* World Scientific, 2005.

[46] G. Samorodnitsky and M. Taqqu. *Stable non-Gaussian random processes.* Chapman & Hall/CRC, 1994.

[47] A. Sasvári. *Positive definite and definitizable functions.* Akademie-Verlag, 1994.

[48] M. Schlather. Construction of positive definite functions. Lecture at 15^{th} Workshop on Stoch. Geometry, Stereology and Image Analysis, Blaubeuren, 22.-27.03.2009.

[49] M. Schlather. Introduction to positive definite functions and to unconditional simulation of random fields. Technical Report ST-99-10, Lancaster University, UK, 1999.

[50] I. J. Schoenberg. Metric spaces and complete monotone functions. *Anna. Math.*, 39(4):811–841, 1938.

[51] L. Shepp and R. J. Vanderbei. *The complex zeros of random polynomials.* Trans. AMS, 347 edition, 1995.

[52] D. Stoyan, W. Kendall, and J. Mecke. *Stochastic Geometry and its Applications.* Wiley and Sons, second edition, 1995.

[53] M. Talagrand. Small tails for the supremum of a Gaussian process. *Ann. Inst. H. Poincaré Probab. Statist.*, 24(2):307–315, 1988.

[54] H. Wackernagel. *Multivariate Geostatistics.* Springer, second edition, 1998.

[55] A.D. Wentzel. *Theorie zufälliger Prozesse.* Birkhäuser, Basel, 1979.

[56] D. V. Widder. *Necessary ans sufficient conditions for the representation of function as Laplace integral.* Trans. of the American Math. Soc., vol. 33 edition, 1932.

[57] G. Winkler. *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods.* Springer, 2003.

[58] A. M. Yaglom. *Correlation Theory of Stationary and Related Random Functions, Volume I.* Springer, 1987.

[59] A. M. Yaglom. *Correlation Theory of Stationary and Related Random Functions, Volume II.* Springer, 1987.

[60] D. Zaporoshets. *An example of a random polynomial with unusual zeros behaviour*, volume 50. Theory Probability Application, 2006.