Gaussian Random Functions

Armin Straub
2 April 2006

Abstract

This document is based on my notes taken at lectures about *Gaussian Random Functions* which were held by visiting Prof. Dr. M. Lifshits (University of St. Petersburg) at TU Darmstadt from 11 to 22 July 2005. These pages only reflect what I picked up from this fine lecture but not necessarily what is mathematically true. I therefore appreciate any hint about things I messed up. –Armin Straub, 23 August 2005

For this second version I have fixed a lot of typos and corrected some minor errors. Nonetheless, there should be lots of them still to be found. –Armin Straub, 7 March 2006

Many thanks to Prof. Dr. M. Lifshits for reading these notes. His corrections have been incorporated in this third version. –Armin Straub, 2 April 2006

Contents

1 Definition of Gaussian Objects

This first section will be fundamental for the rest of these lectures. We will introduce the general notion of Gaussian random vectors that will be used throughout this text.

In this context "objects" mean random variables, random vectors or measures.

1.1 One Dimensional Case

Consider the space $\mathbb{R}$. Recall that a random variable X is normally distributed with mean a and variance σ^2 – usually denoted as $X \sim N(a, \sigma^2)$ – if its density with respect to the Lebesgue measure is

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-a)^2}{2\sigma^2} \right).$$

Let's collect some well known properties of such random variables:

- $X \sim N(0,1) \implies a + \sigma X \sim N(a, \sigma^2)$.
- $X_j \sim N(a_j, \sigma_j^2)$, X_j independent $\implies X_1 + X_2 \sim N(a_1 + a_2, \sigma_1^2 + \sigma_2^2)$.

  This property is called *stability*.
- For $X \sim N(a, \sigma^2)$ the characteristic function and the Laplace transform evaluate to

$$\begin{aligned} \mathbb{E}(\exp(i\,t\,x)) &= \exp\left(i\,a\,t - \frac{\sigma^2 t^2}{2} \right) \\ \mathbb{E}(\exp(\lambda\,x)) &= \exp\left(a\,\lambda + \frac{\sigma^2 \lambda^2}{2} \right). \end{aligned}$$

1.2 Finite Dimensional Case

Now, let our space of interest be $\mathbb{R}^n$. A random vector $X = (X_j)$ is called standard normal if its components X_j are independent and standard normally distributed.

Following the first property in the one dimensional case, we could now define a random vector Y to be Gaussian iff it can be written as $Y = a + L\,X$ where X is standard normal and $L\colon \mathbb{R}^n \to \mathbb{R}^n$ is a linear mapping. As we will notice later, this approach is not well suited for generalization. So we take an alternative one.

In general, there is no way to define Gaussian random vectors via densities as we were able to in the one dimensional case. Just think of the situation were L is degenerated.

Definition 1.1. *A random vector $Y \in \mathbb{R}^n$ is called Gaussian iff (y, Y) is a Gaussian random variable for every $y \in \mathbb{R}^n$.*

Let's check that this definition coincides with the one mentioned before.

$$(a + L\,X, y) = (a, y) + (X, L'\,y) = (a, y) + \sum_{j=1}^{n} X_j\,(L'y)_j,$$

and, clearly, the right-hand side is normally distributed and thus Gaussian.

Notation 1.2. *Let $a = \mathbb{E}\,X$ be the expectation (which may be understood component-wise) and K the covariance matrix of X defined by $K_{j,k} = \operatorname{cov}(X_j, X_k)$. Then, we write $X \sim N(a, K)$ iff X is Gaussian.*

Let X be standard normal – that is $X \sim N(0, I_n)$ – and consider $Y = a + L\,X$. Then $\mathbb{E}(Y) = a$ and $\operatorname{cov}(Y) = \operatorname{cov}(L\,X) = L\,L'$.

Now we have lots of Gaussian random vectors, but there are some questions remaining:

(a) Does $N(a, K)$ exist for all choices a and K?

(b) Is $N(a, K)$ unique?

(c) Are there any Gaussian laws that are not covered by $N(a, K)$?

Let's answer these questions bottom up: the last one reduces to the existence of second moments of random vectors. For uniqueness, we look at the characteristic function. Let X be a Gaussian vector with $\mathbb{E}\,X = a$ and $\operatorname{cov}(X) = K$. Then $\mathbb{E}\,(y, X) = (y, a)$ and $\operatorname{Var}\,(y, X) = (y, K\,y)$. Therefore, the characteristic function is

$$\mathbb{E}\exp(i\,(y, X)) = \exp\left(i(y, a) - \frac{(y, K\,y)}{2}\right),$$

which just depends on a and K. But as one knows from Probability Theory, this means that the distribution of Gaussian X is uniquely determined given a and K.

For the question of existence, we reason that $N(a, K)$ exists iff $N(0, K)$ exists because we may shift Gaussian random vectors. In contrast to a we cannot choose the covariance matrix K arbitrarily. Indeed, K has to be symmetric and positive semi-definite:

$$\begin{aligned} (y, K\,x) &= (K\,y, x) \\ (x, K\,x) &\geqslant 0. \end{aligned}$$

Fortunately, these conditions are sufficient as well. We can see this as follows: they are equivalent to $K\,e_j = \lambda_j\,e_j$ for e_j being an orthonormal basis of $\mathbb{R}^n$ and $\lambda_j \geqslant 0$. But then we can define L via $L\,e_j = \sqrt{\lambda_j}\,e_j$, which makes $L\,X \sim N(0, K)$ for a standard normal X.

In comparison to the one dimensional case, let's write down some properties of Gaussian random vectors:

- $X \sim N(0, I_n) \quad \Longrightarrow \quad a + L\,X \sim N(a, L\,L')$.
- $X_j \sim N(a_j, K_j)$, X_j independent $\quad \Longrightarrow \quad X_1 + X_2 \sim N(a_1 + a_2, K_1 + K_2)$.

  Again, this property is called *stability*.

- For $X \sim N(a, K)$ the characteristic function, resp. the Laplace transform evaluate to

$$\begin{aligned} \mathbb{E}\exp(i\,(y, X)) &= \exp\left(i\,(y, a) - \frac{(y, K\,y)}{2}\right) \\ \mathbb{E}\exp((y, X)) &= \exp((y, a) + \frac{(y, K\,y)}{2}\Big). \end{aligned}$$

Exercise 1.1. Let X_j be Gaussian random variables. Is it true, that $X = (X_1, ..., X_n)$ is a Gaussian random vector?

1.3 General Case

Let $\mathcal{X}$ be a linear space enhanced with a topology (we will clarify meaningful properties of the space and its topology later). Denote with $\mathcal{X}'$ the dual space of $\mathcal{X}$, i.e. the space of continuous linear functionals. Let X be a random vector taking values in $\mathcal{X}$. For $f \in \mathcal{X}'$ and $x \in \mathcal{X}$ we will often write (f, x) instead of $f(x)$.

In complete analogy to the finite dimensional case, we define Gaussian random vectors.

Definition 1.3. *X is called Gaussian iff (f, X) is a Gaussian random variable for every $f \in \mathcal{X}'$.*

In a similar fashion, we define the expectation and covariance of random vectors:

Definition 1.4. *$a \in \mathcal{X}$ is an expectation of X iff $\mathbb{E}\,(f, X) = (f, a)$ for all $f \in \mathcal{X}'$.*

Further, $K\colon \mathcal{X}' \to \mathcal{X}$ is called covariance (operator) of X iff $\operatorname{cov}((f, X), (g, X)) = (f, K\,g)$ for all $f, g \in \mathcal{X}'$.

In the case of existence, we will write $\mathbb{E}(X) = a$ and $\operatorname{cov}(X) = K$.

Remark 1.5. At the beginning of this section we just took $\mathcal{X}$ to be a linear space with a topology. As you can e.g. grasp from the definition of an expectation, we need to have $\mathcal{X}'$ to be sufficiently large (in particular it should be non-degenerate, i.e. larger than just $\{0\}$), so that there are enough continuous linear mappings. Otherwise there could be lots of expectations making our definition loosing sense.

Thus, we restrict ourselves to special spaces $\mathcal{X}$. Three possibilities where generality increases top-down are:

(1) $\mathcal{X}$ a separable Banach space, e.g. $C[0,1]$, $L^p[0,1]$.

(2) $\mathcal{X}$ a locally convex topological linear space that is metrisable, e.g. $C[0,\infty)$, $\mathbb{R}^{\mathbb{N}}$.

(3) $\mathcal{X}$ a locally convex topological linear space such that the law of X is a Radon measure.

We will silently assume that we are in one of these cases (i.e. we are in the last case) and only refer to this as the "usual conditions".

Again, just imitating what we have done before in finite dimensions:

Notation 1.6. *Suppose that $a = \mathbb{E}\,X$ and $K = \operatorname{cov}(X)$ exist. We write $X \sim N(a, K)$ iff X is Gaussian.*

Having defined these notions, we have to ask us the following questions afresh:

(a) Does $N(a, K)$ exist for all choices a and K?

(b) Is $N(a,K)$ unique?

(c) Are there any Gaussian distributions that are not given by $N(a,K)$?

Under the "usual conditions" one can show that for Gaussian X expectation and covariance exist, which answers the last question. The second point is handled exactly as in the finite dimensional case: looking at the characteristic function, which we recognize to depend on a and K only, we can still conclude that the distribution of X is uniquely determined by these two.

Again, $N(a,K)$ exists iff $N(0,K)$ does and we have the following necessary conditions for K:

$$\begin{aligned}(f, Kg) &= (g, Kf)\\ (f, Kf) &\geqslant 0.\end{aligned}$$

But in contrast to the finite dimensional situation, these conditions don't suffice. Unfortunately sufficient conditions are in general difficult and depend on the space $\mathcal{X}$ (comfortingly the question of existence is not of high practical importance – things interesting enough to study tend to exist). See exercise 2.1 for a surprising example.

2 Examples of Gaussian Objects

2.1 Main Examples

Example 2.1. (Standard Gaussian Measure in $\mathbb{R}^{\mathbb{N}}$) Let $\mathcal{X} = \mathbb{R}^{\mathbb{N}}$ endowed with the topology induced by the Fréchet metric

$$\rho(x) = \sum_{j=1}^{\infty} 2^{-j} (|x_j| \wedge 1).$$

The dual space $\mathcal{X}'$ can be interpreted as the space of all "finite real sequences" denoted by $\mathcal{X}' = c_0$. Here finiteness means that for $(f_j)_{j\in\mathbb{N}} \in \mathcal{X}'$ there is an index $j_0 \in \mathbb{N}$ such that $f_j = 0$ for $j \geqslant j_0$. The action of f on $\mathcal{X}$ is given by

$$(f, x) = \sum_{j=1}^{\infty} f_j x_j,$$

which, actually, is just a finite sum.

Now we take a sequence of standard normal random variables $X = (X_j)_{j\in\mathbb{N}}$ with X_j i.i.d. and $X_j \sim N(0,1)$. Then X is a random variable variable taking values in $\mathcal{X}$. Further, X is centered, i.e. $\mathbb{E}\, X = 0$, because for $f \in \mathcal{X}'$ (recall its finiteness)

$$\mathbb{E}\,(f, X) = \mathbb{E}\left(\sum_{j=1}^{\infty} f_j X_j \right) = \sum_{j=1}^{\infty} f_j\, \mathbb{E}\, X_j = 0.$$

The covariance operator $K\colon c_0 \to \mathbb{R}^{\mathbb{N}}$ is just the inclusion mapping $K\colon f \mapsto f$. This can be verified as follows

$$\begin{aligned}\operatorname{cov}((f,X),(g,X)) &= \mathbb{E}((f,X)(g,X)) = \mathbb{E}\Big(\Big(\sum f_j X_j \Big) \Big(\sum g_j X_j \Big) \Big)\\ &= \sum_{j\in\mathbb{N}\times\mathbb{N}} f_{j_1} g_{j_2}\, \mathbb{E}(X_{j_1} X_{j_2}) = \sum f_j g_j = (f, Kg).\end{aligned}$$

The law of X is called the standard Gaussian measure in $\mathbb{R}^{\mathbb{N}}$.

Example 2.2. (Gaussian Measure in Hilbert Spaces) Let $\mathcal{X}$ be a Hilbert space (endowed with the topology induced from the inner product – such obvious choices won't be commented in the sequel). Due to Riesz representation theorem, the continuous linear mappings arise from the inner product with the first argument fixed and the dual space is just identified as $\mathcal{X}' = \mathcal{X}$.

Let $\{e_j : j \in \mathbb{N}\}$ be an orthonormal (Schauder) basis and i.i.d. standard normal random variables (ξ_j) as well as real numbers $\sigma_j \geqslant 0$ such that $\sum \sigma_j^2 < \infty$. We define the random element

$$X = \sum_{j=1}^{\infty} \sigma_j \xi_j e_j$$

in $\mathcal{X}$; the sum converges almost surely in norm and is thus almost surely well defined (so it may be redefined arbitrarily on the exceptional set of measure zero if one feels bothered – we will feel just fine in the sequel with such almost sure well defined elements, at least if we are only interested in their law).

As $\xi_j \sim N(0,1)$ for $j \in \mathbb{N}$ the expectation of X is again $a = \mathbb{E}(X) = 0$ because

$$\mathbb{E}(f, X) = \mathbb{E}\Big(\sum \sigma_j \xi_j (f, e_j) \Big) = \sum \sigma_j \, \mathbb{E}(\xi_j) \, (f, e_j) = 0.$$

To get the covariance operator $K : \mathcal{X} \to \mathcal{X}$ we look at

$$\begin{aligned} \operatorname{cov}((f,X),(g,X)) &= \operatorname{cov}\Big(\sum \sigma_j \xi_j (f, e_j), \sum \sigma_j \xi_j (g, e_j) \Big) \\ &= \sum \operatorname{cov}(\xi_{j_1}, \xi_{j_2}) \, \sigma_{j_1} \sigma_{j_2} (f, e_{j_1}) \, (f, e_{j_2}) \\ &= \sum \sigma_j^2 f_j g_j = (f, Kg) \end{aligned}$$

and from the last equality we read that

$$K : g \mapsto \sum \sigma_j^2 g_j e_j.$$

K can be thought of as an infinite diagonal matrix with entries σ_j^2. Computing the expected square norm of X we get

$$\mathbb{E}\|X\|^2 = \mathbb{E}\Big(\sum X_j^2 \Big) = \sum \sigma_j^2 \, \mathbb{E}\big(\xi_j^2\big) = \sum \sigma_j^2 < \infty.$$

Exercise 2.1. Let $\mathcal{X}$ be an infinite dimensional Hilbert space. Show that there is no standard Gaussian measure on $\mathcal{X}$, i.e. there is no Gaussian random variable X whose covariance operator is the identity.

Example 2.3. (Brownian Motion) Let $\mathcal{X} = C[0,1]$ be the set of all continuous functions on the unit interval and $X = W$ be a Wiener process (i.e. a Brownian motion) on $\mathcal{X}$. The only properties of such a process W we will need here are

$$\mathbb{E}\, W_t = 0 \quad \text{and} \quad \mathbb{E}(W_s W_t) = s \wedge t.$$

We will clarify, what a Gaussian process is, in the following example 2.4. We will see that $s \wedge t$ is the covariance kernel of W which indeed defines W completely.

You can see a typical trajectory of a Brownian motion in figure 2.1.

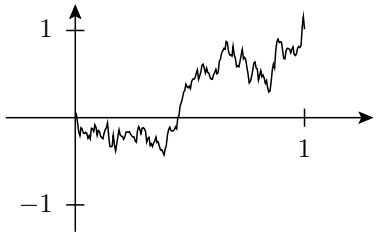

Figure 2.1. A trajectory of Brownian motion

The dual space $\mathcal{X}'$ is the set of all "signed measures" on the unit interval, denoted by $M[0,1]$. A signed measure $\mu \in M[0,1]$ is the difference of two finite measures on $[0,1]$, i.e. $\mu = \nu_1 - \nu_2$. The action of μ on $C[0,1]$ is given by

$$(\mu, f) = \int_0^1 f \,\mathrm{d}\mu.$$

We are now interested in expectation and covariance operator of the Wiener process W. Because of

$$\mathbb{E}\,(\mu, W) = \mathbb{E} \int W_t \,\mu(\mathrm{d}t) = \int \mathbb{E}\, W_t \,\mu(\mathrm{d}t) = 0,$$

its expectation is $\mathbb{E}\, W = 0$. As usual, we determine $K\colon M[0,1] \to C[0,1]$ by

$$\begin{aligned} \operatorname{cov}((\mu, W), (\nu, W)) &= \mathbb{E} \int W_t \,\mu(\mathrm{d}t) \int W_t \,\nu(\mathrm{d}t) = \mathbb{E} \int\int W_s W_t \,\nu(\mathrm{d}t)\, \mu(\mathrm{d}s) \\ &= \int\int \mathbb{E}(W_s W_t)\, \nu(\mathrm{d}t)\, \mu(\mathrm{d}s) = \int\int s \wedge t \,\nu(\mathrm{d}t)\, \mu(\mathrm{d}s) = (\mu, K\nu) \end{aligned}$$

and it follows that

$$(K\nu)(t) = \int_0^1 s \wedge t \,\nu(\mathrm{d}s).$$

Example 2.4. (Gaussian Processes) Let T be a compact metric space. X is called a Gaussian process on T iff X is a Gaussian random vector taking values in $\mathcal{X} = C(T)$. As in the case $T = [0,1]$ for the Wiener process we get $\mathcal{X}' = M(T)$, i.e. the space of all signed measures on T.

For a random vector $X \in C(T)$ to be Gaussian, it is sufficient that for any $t_1, \ldots, t_n \in T$ the random vector $(X_{t_1}, \ldots, X_{t_n})$ is Gaussian. It is clear that this condition is necessary as well, because the measures concentrated in the points $t_1, \ldots, t_n$ are contained in $M(T)$.

Let X be a Gaussian process. Its expectation $\mathbb{E}\, X = a \in C(T)$ can, again, be understood componentwise, since

$$a(t) = \mathbb{E}\, X_t.$$

Its covariance operator $K\colon M(T) \to C(T)$ is given by

$$(K\nu)(t) = \int_T \operatorname{cov}(X_s, X_t)\, \nu(\mathrm{d}s) = \int_T k(s,t)\, \nu(\mathrm{d}s),$$

where $k(s, t) = \operatorname{cov}(X_s, X_t)$ is called *covariance kernel*. It is left as an exercise to thoroughly check these statements.

In particular, a Gaussian process X is defined completely by its expectation and covariance kernel.

2.2 Special Examples of Gaussian Processes

Example 2.5. (Fractional Brownian Motion) Let $T = [0,1]$ be the unit interval and $\alpha \in (0, 2]$. A Gaussian process $W^{(\alpha)}$ is called fractional Brownian motion iff

$$\mathbb{E}\, W^{(\alpha)} = 0 \quad \text{and} \quad \operatorname{cov}\Big(W_s^{(\alpha)}, W_t^{(\alpha)}\Big) = k(s,t) := \frac{1}{2}(|s|^\alpha + |t|^\alpha - |t-s|^\alpha).$$

$W^{(1)}$ is just a classical Wiener process. It may seem that the choice of k looks unmotivated, but note that

$$\mathbb{E}\left|W_t^{(\alpha)} - W_s^{(\alpha)}\right|^2 = |t-s|^\alpha.$$

In figure 2.2 you can see the impact of the parameter α on the look of a "typical" trajectory.

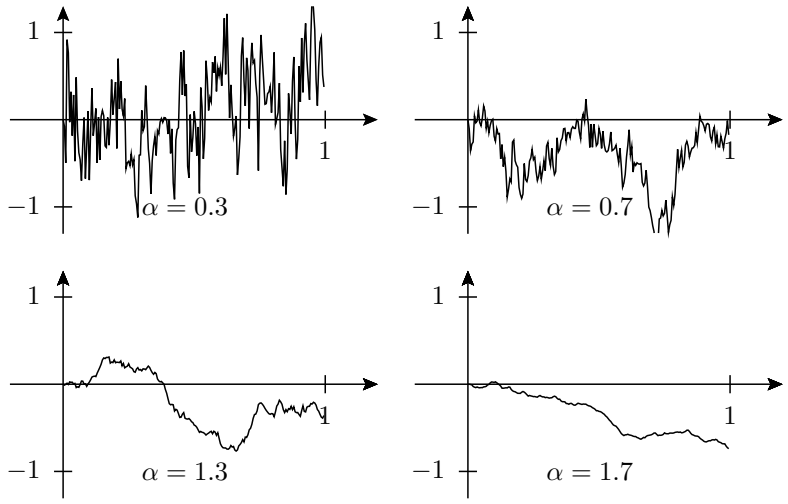

Figure 2.2. Trajectories of fractional Brownian motions

Let's look at the extremal values $\alpha = 2$ and α tending to 0. For $\alpha = 2$ we get $k(s,t) = s\,t$ so that we can define $W_t^{(2)} = t\, W_1^{(2)}$ receiving a fractional Brownian motion $W^{(2)}$ that just depends on the single random variable $W_1^{(2)}$.

Let $\alpha \to 0$. For $s = t$ we have $|t-s|^\alpha \to 0$ which motivates to define the Gaussian process $W^{(0)}$ via

$$\mathbb{E}\, W^{(0)} = 0 \quad \text{and} \quad \operatorname{cov}\Big(W_s^{(0)}, W_t^{(0)}\Big) = k(s,t) = \begin{cases} \frac{1}{2} & \text{for } s \neq t \\ 1 & \text{for } s = t \end{cases}.$$

This makes every $W_t^{(0)}$ a standard normal random variable and reflects a great amount of independence between $W_s^{(0)}$ and $W_t^{(0)}$ for $s \neq t$. It is an exercise to investigate this case more closely.

Example 2.6. (Brownian Sheet) The Brownian sheet is a Gaussian process W on $T = [0,1]^d$ defined by

$$\mathbb{E}\, W_t \quad \text{and} \quad \operatorname{cov}(W_s, W_t) = \prod_{i=1}^{d} s_i \wedge t_i.$$

The covariance kernel can be interpreted geometrically as the volume of intersections as shown in figure 2.3 for $d = 2$.

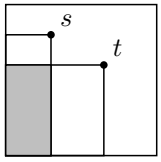

Figure 2.3. Geometric interpretation of Brownian sheet's covariance kernel

Example 2.7. (Lévy's Brownian Function) Let $T = \mathbb{R}^d$. In analogy to the definition of the fractional Brownian motion we define Lévy's Brownian motion $W^{(L)}$ via

$$\mathbb{E}\, W^{(L)} = 0 \quad \text{and} \quad \operatorname{cov}\Big(W_s^{(L)}, W_t^{(L)}\Big) = \frac{1}{2}(\|s\| + \|t\| - \|t-s\|).$$

It's therefore no surprise that

$$\mathbb{E}\Big\|W_t^{(L)} - W_s^{(L)}\Big\|^2 = \|t-s\|.$$

Example 2.8. (Fractional Brownian Sheet) The fractional Brownian sheet with parameter $\alpha \in (0,2]$ is a Gaussian process $W^{(\alpha)}$ on $T = \mathbb{R}^d$ given by

$$\mathbb{E}\, W^{(\alpha)} = 0 \quad \text{and} \quad \operatorname{cov}\Big(W_s^{(\alpha)}, W_t^{(\alpha)}\Big) = \frac{1}{2}\prod_{i=1}^{d}(\|s\|^\alpha + \|t\|^\alpha - \|t-s\|^\alpha).$$

Example 2.9. (Lévy's Fractional Brownian Function) You may guess how Lévy's fractional Brownian function $W^{(L,\alpha)}$ (again a Gaussian process on $T = \mathbb{R}^d$ and $\alpha \in (0,2]$) will be defined. Surely, you get it from

$$\mathbb{E}\, W^{(L,\alpha)} = 0 \quad \text{and} \quad \operatorname{cov}\Big(W_s^{(L,\alpha)}, W_t^{(L,\alpha)}\Big) = \frac{1}{2}(\|s\|^\alpha + \|t\|^\alpha - \|t-s\|^\alpha).$$

2.3 Gaussian White Noise and Integration

Definition 2.10. *Let $(A, \mathfrak{A}, \nu)$ be a measure space and ν a finite measure. A mapping $X\colon \mathfrak{A} \times \Omega \to \mathbb{R}$, $(B,\omega) \mapsto X_B(\omega)$ is called* Gaussian white noise *iff for any $B, B_j \in \mathfrak{A}$ and B_j disjoint*

$$\begin{aligned} X_B &\sim N(0, \nu(B)) \\ X_{B_1 \cup B_2} &= X_{B_1} + X_{B_2} \quad a.s. \\ X_{B_1}, X_{B_2} &\quad \textit{independent}. \end{aligned}$$

Gaussian white noise X is like a random measure on A lacking only σ-additivity (which makes it a random content in fact).

We will now define the integral over Gaussian white noise X for integrands $f \in L^2(A,\nu)$. Just as the Lebesgue integral is commonly defined stepwise – starting with indicator functions, extending it to linear combinations of indicator functions and then to arbitrary functions – we will introduce the integral over white noise.

- Let $f = 1_B$ be an indicator function and $B \in \mathfrak{A}$. We define

$$\int f \,\mathrm{d}X = \int 1_B \,\mathrm{d}X := X_B.$$

- Now set $f = \sum_{j=1}^n c_j 1_{B_j}$ for $c_j \in \mathbb{R}$ and $B_j \in \mathfrak{A}$. The integral of f is

$$\int f \,\mathrm{d}X = \int \sum_{j=1}^n c_j 1_{B_j} \,\mathrm{d}X := \sum_{j=1}^n c_j X_{B_j}.$$

  Note that this is well defined due to the properties of the Gaussian white noise X.

- For $f_n \to f$ in the L^2 sense we define the integral of f as the L^2 limit of the integrals of f_n

$$\int f_n \,\mathrm{d}X \to \int f \,\mathrm{d}X.$$

  It is left to the reader to check all the details.

The integral over Gaussian white noise has a property called *isometry* that proves to be very useful later on. For $f, g \in L^2(A, \nu)$

$$\int_A f g \,\mathrm{d}\nu = \mathbb{E}\Big(\int f \,\mathrm{d}X \int g \,\mathrm{d}X \Big)$$

and in particular

$$\int_A f^2 \,\mathrm{d}\nu = \mathbb{E}\Big(\int f \,\mathrm{d}X \Big)^2.$$

The reason we are introducing Gaussian white noise is that many Gaussian processes including Brownian motion can be written as integrals over Gaussian white noise.

Example 2.11. (Brownian motion) Let $(A,\ \mathfrak{A},\ \nu) = ([0,\ 1],\ \mathfrak{B},\ \lambda)$ where λ denotes the Lebesgue measure over the Borel sets $\mathfrak{B}$ and let X be Gaussian white noise on this space. For $t \in [0, 1]$ define

$$W_t = \int 1_{[0,t]} \,\mathrm{d}X = X_{[0,t]}.$$

We claim that W is indeed a Brownian motion. Let's check

$$\begin{aligned} \mathbb{E}\, W_t &= \mathbb{E}\, X_{[0,t]} = 0 \\ \text{and} \quad \operatorname{cov}(W_s, W_t) &= \mathbb{E}\Big(\int 1_{[0,s]} \,\mathrm{d}X \int 1_{[0,t]} \,\mathrm{d}X \Big) = \int_0^1 1_{[0,s]} 1_{[0,t]} \,\mathrm{d}\lambda = s \wedge t \end{aligned}$$

where we used the isometry property of the integral over X.

Exercise 2.2. Do the same for the Brownian sheet.

Remark 2.12. In literature (see e.g. [Liptser, 1997, section 10]), you will also find another notion of Gaussian white noise. Although the trajectories of a Brownian motion W are almost surely nowhere differentiable, one is able to define a generalized derivative. The resulting Gaussian process (the meaning of process is a bit more relaxed then, because the trajectories are distributions) W' is also referred to as Gaussian white noise. W' is a centered process defined by its covariance kernel

$$k(s, t) = \delta(t - s)$$

where δ denotes the Dirac distribution. This is not surprising if you keep in mind that the Brownian motion W is a martingale. A local change should be independent from other local changes and the average change should be zero.

You might see a link to our notion of Gaussian white noise if you recall example 2.11. There we showed that Brownian motion arises as an integral $\int_0^t \dots$ over white noise. More exactly, we have $X_B = \int 1_B \, \mathrm{d}W$ where the last integral is a stochastic one.

3 Kernels of Gaussian Measures

Throughout this section let $\mathcal{X}$ be a linear space, X be a Gaussian random vector taking values in this space and P be the law of X. For simplicity we assume X to be centered, i.e. $\mathbb{E}(X) = 0$. Therefore we have $P = N(0, K)$, where $K : \mathcal{X}' \to \mathcal{X}$ is the covariance operator of X.

We now want to tag a subspace $H \subseteq \mathcal{X}$ that contains most information about the measure P. This subspace will depend on P and we will call it *kernel*.

3.1 Definition and Basic Properties

In order to define the kernel, we will introduce the notion of measurable linear functionals which is pretty straightforward.

The dual space $\mathcal{X}'$ of $\mathcal{X}$ consists of all the continuous linear functionals on $\mathcal{X}$. We are now going to enlarge this space obtaining the so called measurable linear functionals $\mathcal{X}'_P$ which depend on the measure P.

First, we observe that every $f \in \mathcal{X}'$ is square-integrable, that is $(f, \cdot) \in L^2(\mathcal{X}, P)$, due to

$$\int (f, x)^2 \, \mathrm{d}P(x) = \mathbb{E}\,(f, X)^2 = (f, K f) < \infty.$$

Thus we have the injection

$$I' : \mathcal{X}' \to L^2(\mathcal{X}, P).$$

Just as measurable functions arise as limits of continuous functions, we define the measurable linear functionals to be the limits (in the L^2 sense) of continuous ones. In other words $\mathcal{X}'_P = \text{closure}(I'(\mathcal{X}'))$.

An element $z \in \mathcal{X}'_P$ can therefore be written as $z = \lim_{n\to\infty} (f_n, \cdot)$, where we mean $\mathbb{E}(z(X) - (f_n, X))^2 \to 0$. Note that $z(X)$ is well defined on a set of measure 1.

As a closed subspace of the Hilbert space $L^2(\mathcal{X}, P)$ the space $\mathcal{X}'_P$ is a Hilbert space itself reusing the existing inner product

$$(z_1, z_2)_{\mathcal{X}'_P} = (z_1, z_2)_{L^2(\mathcal{X}, P)} = \mathbb{E}(z_1(X)\, z_2(X)).$$

In particular, $\|z\|^2_{\mathcal{X}'_P} = \mathbb{E}(z(X))^2$.

We suggestively denoted the injection of continuous linear functionals into the square integrable functions by I', because we are interested in its adjoint operator $I : \mathcal{X}'_P \to \mathcal{X}$ (recall that $\mathcal{X}'_P$ is a Hilbert space and thus may be identified with its dual space).

Note 3.1. As I' is not continuous in general we don't automatically have the existence of its adjoint operator. But we won't track this here.

So I is defined via $(f, Iz) = (I'f, z)_{\mathcal{X}'_P}$ for $f \in \mathcal{X}'$ and $z \in \mathcal{X}'_P$.

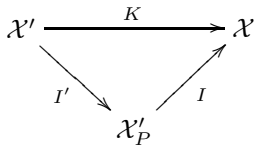

Figure 3.1. Decomposition of covariance operator $K = II'$

We claim that I is in some sense the square root of K, that is $K = II'$ making diagram 3.1 commutative. Indeed, for every $f, g \in \mathcal{X}'$ we have

$$(g, II'f) = (I'g, I'f)_{\mathcal{X}'_P} = \mathbb{E}((g,X)(f,X)) = \operatorname{cov}((g,X),(f,X)) = (g, Kf).$$

Furthermore, I is an injection: if $Iz = 0$ then $(g, Iz) = (I'g, z) = 0$ for all $g \in \mathcal{X}'$. Because $\{I'g : g \in \mathcal{X}'\}$ is dense in $\mathcal{X}'_P$ this implies that $z = 0$.

Definition 3.2. *We define the kernel of P to be $H_P := H := I(\mathcal{X}'_P)$ endowed with the inner product induced from $\mathcal{X}'_P$*

$$(h_1, h_2)_H := (z_1, z_2)_{\mathcal{X}'_P} = \mathbb{E}(z_1(X)\, z_2(X)),$$

where $h_i = Iz_i$, that is z_i is the preimage of h_i (well defined since I is an injection).

Thus I maps the measurable linear functionals $\mathcal{X}'_P$ bijectively onto the kernel. Let's collect some properties of the kernel.

- $K(\mathcal{X}') \subset H \subset \mathcal{X}$, where $K(\mathcal{X}') \subsetneq H$ whenever there are measurable non-continuous linear functionals. This is because $K(\mathcal{X}') = I(I'(\mathcal{X}'))$ which is the image of all continuous linear functionals whereas H is the image of all measurable linear functionals (I is an injection).
- If $\dim(H) = \infty$ then $P(H) = 0$. This is perhaps somewhat astonishing as H shall in some sense contain all important information about P.
- But $\operatorname{supp}(P) = \operatorname{closure}(H)$ which softens the above comment about H being a zero set. Recall that $\operatorname{supp}(P)$ denotes the smallest closed set of measure 1.
- H is separable.

3.2 Examples

Example 3.3. (Standard Gaussian Measure in $\mathbb{R}^{\mathbb{N}}$) Choose $\mathcal{X} = \mathbb{R}^{\mathbb{N}}$ and let P be the standard Gaussian measure on $\mathcal{X}$ – see example 2.1. If X is a random vector with law P, e.g. $X = (X_j)_{j \in \mathbb{N}}$ such that X_j i.i.d. and $X_j \sim N(0,1)$. What is the kernel H_P of P?

We claim that

$$H_P = l_2.$$

We know that $\mathcal{X}' = c_0$ is the set of all finite sequences and $(f, x) = \sum f_j x_j$ for $f \in \mathcal{X}'$. I' – the interpretation of $\mathcal{X}'$ as a subset of $L^2(\mathcal{X}, P)$ – does the same in a different space

$$(I'f)(x) = \sum f_j x_j.$$

The inner product in $\mathcal{X}'_P$ is just borrowed from $L^2(\mathcal{X}, P)$ and thus

$$\|I'f\|_{\mathcal{X}'_P} = \mathbb{E}\, f(X)^2 = \mathbb{E}\Big(\sum f_j x_j\Big)^2 = \sum f_j^2 = \|f\|_2^2$$

where we did not really feel that f was a finite sequence. The set of the measurable linear functionals $\mathcal{X}'_P$ is in some sense enriched with the L^2 limits of $\mathcal{X}'$ and thus consists of all

$$z(x) = \sum z_j x_j \quad \text{where} \quad \sum z_j^2 < \infty.$$

Note that z is well defined on a set $A \subset \mathcal{X}$ of measure $P(A) = 1$. While $\mathcal{X}' = c_0$ we thus have $\mathcal{X}'_P = l_2$ with the standard inner products.

Immediate calculation (just check $(f, Iz) = (I'f, z)_{\mathcal{X}'_P}$ for $f \in \mathcal{X}'$ and $z \in \mathcal{X}'_P$) provides us with

$$I\colon \mathcal{X}'_P \to \mathcal{X}, \quad z \mapsto (z_j)_{j\in\mathbb{N}},$$

verifying that

$$H_P = I\big(\mathcal{X}'_P\big) = l_2,$$

which denotes equality of the spaces including the norm.

Let's check some of the mentioned properties of the kernel. We have $l_2 \subsetneq \mathbb{R}^{\mathbb{N}}$ and as $\dim\big(\mathbb{R}^{\mathbb{N}}\big) = \infty$ we indeed have $P(l_2) = 0$.

Example 3.4. (Gaussian Measure in Hilbert Spaces) Following example 2.2, let $\mathcal{X}$ be a Hilbert space, $\{e_j\colon j \in \mathbb{N}\}$ an orthonormal basis, ξ_j i.i.d. and $\xi_j \sim N(0,1)$ as well as $\sigma_j \geqslant 0$ and $\sum \sigma_j^2 < \infty$. Let P be the law of $X = \sum \sigma_j \xi_j e_j$. What is the kernel H_P of P?

As $\mathcal{X}$ is a Hilbert space we have $\mathcal{X}' = \mathcal{X}$ and for $f \in \mathcal{X}'$ we get $(f, x) = \sum f_j (x, e_j)$ which is not changed by I'

$$(I'f)(x) = \sum f_j (x, e_j)$$

as we are still talking about the same object under I'. The norm in $\mathcal{X}'_P$ comes from L^2 and thus

$$\|I'f\|_{\mathcal{X}'_P} = \mathbb{E}(I'f, X)^2 = \mathbb{E}\Big(\sum \sigma_j \xi_j f_j\Big)^2 = \sum f_j^2 \sigma_j^2. \tag{3.1}$$

It is therefore not astonishing that $\mathcal{X}'_P$ consists of all $z \in L^2(\mathcal{X}, P)$ such that

$$z(x) = \sum z_j (x, e_j) \quad \text{where} \quad \sum z_j^2 \sigma_j^2 < \infty.$$

Again, z is well defined only on a set $A \subset \mathcal{X}$ of measure $P(A) = 1$. The norm in $\mathcal{X}'_P$ is given by (see 3.1)

$$\|z\|_{\mathcal{X}'_P} = \sum z_j^2 \sigma_j^2.$$

We claim that $I\colon \mathcal{X}'_P \to \mathcal{X}$ is defined by

$$Iz = \sum z_j \sigma_j^2 e_j,$$

which is verified by

$$(f, Iz) = \sum f_j z_j \sigma_j^2 = (I'f, z)_{\mathcal{X}'_P}$$

for $f \in \mathcal{X}'$ and $z \in \mathcal{X}_P'$. Therefore, the kernel $H_P = I\big(\mathcal{X}_P'\big)$ is made up by all $h \in \mathcal{X}$ that can be written as $h = Iz$ for $z \in \mathcal{X}_P'$. Then $h_j = z_j \sigma_j^2$ and thus

$$\|h\|_{H_P}^2 = \|z\|_{\mathcal{X}_P'}^2 = \sum z_j^2 \sigma_j^2 = \sum \frac{h_j^2}{\sigma_j^2}$$

(if certain σ_j should be zero than h_j has to be zero as well) and we get

$$H_P = \left\{ h \in \mathcal{X} : \sum \frac{h_j^2}{\sigma_j^2} < \infty \right\}$$

with inner product

$$(g, h)_{H_P} = \sum \frac{g_j h_j}{\sigma_j^2},$$

which does not conincide with the inner product of $\mathcal{X}$.

Example 3.5. (Finite Dimensional Case) Let $\mathcal{X} = \mathbb{R}^n$ and $P = N(0, K)$ a centered Gaussian measure such that $K : \mathbb{R}^n \to \mathbb{R}^n$ is not degenerated, i.e. $K(\mathbb{R}^n) = \mathbb{R}^n$.

Then – as we can conclude from the Hilbert space case – the kernel H_P is just the whole space $H_P = \mathbb{R}^n$ endowed with the inner product

$$(g, h)_{H_P} = (g, K^{-1} h) = \left(K^{-\frac{1}{2}} g, K^{-\frac{1}{2}} h \right).$$

3.3 Factorization Theorem

Let's recall the construction of the kernel H of the Gaussian measure P. First we decomposed K into $K = I\, I'$ where I' is the inclusion of the continuous linear functionals into the square integrable functions. Then, the kernel of P is defined $H = I\big(\mathcal{X}_P'\big)$ as the image of all measurable linear functionals under I.

Unfortunately, the space $\mathcal{X}_P'$ of all measurable functionals is quite difficult in general. This way it would e.g. be very hard to get the kernel of the Wiener process. Therefore we are looking for an alternative replacing $\mathcal{X}_P'$ by something more convenient.

Theorem 3.6. (Factorization Theorem) *Let $\mathcal{H}$ be a Hilbert space and $J : \mathcal{H} \to \mathcal{X}$ an injective linear mapping such that $K = J\, J'$. Then the kernel H is $J(\mathcal{H})$ with inner product $(J z_1, J z_2)_H = (z_1, z_2)_{\mathcal{H}}$.*

This theorem essentially tells us that we can take an arbitrary Hilbert space $\mathcal{H}$ instead of the special maybe difficult $\mathcal{X}_P'$. Figure 3.2 shows this situation.

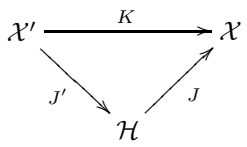

Figure 3.2. Factorization theorem

Example 3.7. (Brownian Motion) Having the factorization theorem as a tool we can go for the kernel of the Wiener measure. Let $\mathcal{X}$ be $C[0,1]$ and X be the Wiener process with law P. Then $\mathbb{E}(X)=0$ and the covariance operator $K\colon M[0,1]\to C[0,1]$ is determined by $(K\,\mu)(t)=\int s\wedge t\,\mu(\mathrm{d}s)$.

We take our Hilbert space to be $\mathcal{H}=L^2[0,1]$ and the operator $J\colon L^2[0,1]\to C[0,1]$ to be the integral operator $(J\,z)(s)=\int_0^s z(t)\mathrm{d}t$. Now we have to check that this is what we need to apply the factorization theorem.

The adjoint operator $J'\colon M[0,1]\to L^2[0,1]$ of J is $(J'\,\mu)(t)=\mu[t,1]$. It is left as an exercise to check this. As J is an injection we just have to validate $J\,J'=K$:

$$\begin{aligned}(J\,J'\,\mu)(s) &= \int_0^s (J'\,\mu)(t)\mathrm{d}t=\int_0^s \mu[t,1]\mathrm{d}t=\int_1^t\int_0^s \mathrm{d}t\,\mu(\mathrm{d}u)\\ &= \int_0^1\int_0^1 1_{t\leqslant s}\,1_{t\leqslant u}\,\mathrm{d}t\,\mu(\mathrm{d}u)=\int_0^1 s\wedge u\,\mu(\mathrm{d}u)=(K\,\mu)(s).\end{aligned}$$

We can therefore use the factorization theorem to obtain $H_P=J\big(L^2[0,1]\big)$ which tells us that $h\in H_P$ iff $h(s)=\int_0^s z(t)\mathrm{d}t$ for some $z\in L^2[0,1]$. This condition is equivalent to h being absolutely continuous with square integrable derivative and $h(0)=0$. Thus our kernel is

$$H_P=\big\{h\in\mathrm{AC}[0,1]\colon h(0)=0, h'\in L^2[0,1]\big\}$$

with inner product

$$(g,h)_{H_P}=(g',h')_{L^2[0,1]}=\int_0^1 g'(s)\,h'(s)\mathrm{d}s.$$

We recognize that H_P is a Sobolev space usually denoted by $\dot{W}^{\frac{1}{2}}$ or $H_0^{1,2}$. In this context, its is also called *Cameron-Martin* space. Because this was the first really important kernel studied, this name is sometimes used for other kernels as well.

As a generalization of what we did for the Brownian motion we now want to look at processes that arise as integrals over Gaussian white noise (and this class of processes is very rich).

We take a measure space $(\mathcal{A},\nu)$ and Z Gaussian white noise on $\mathcal{A}$. Let $\mathcal{X}$ be $C(T)$ where T is compact. We want to obtain the kernel of (the law of) the process $X_t=\int_{\mathcal{A}} m_t(u)\,Z(\mathrm{d}u)$ which is well defined if $m_t\in L^2(\mathcal{A},\nu)$.

Our approach is now fairly the same as in the case of the Wiener process (recall that we get it by setting $(\mathcal{A},\nu)=([0,1],\lambda)$ and $m_t=1_{[0,t]}$). We take the Hilbert space $\mathcal{H}=L^2(\mathcal{A},\nu)$ and the operator $J\colon\mathcal{H}\to C(T)$ defined by $(J\,f)(t)=\int_{\mathcal{A}} m_t(u)\,f(u)\,\nu(\mathrm{d}u)$.

Then the adjoint operator $J'\colon M(T)\to\mathcal{H}$ is $(J'\,\mu)(u)=\int_T m_t(u)\,\mu(\mathrm{d}t)$. Its again an exercise to verify this. Let's check if $K=J\,J'$

$$\begin{aligned}(J\,J'\,\mu)(t) &= \int_{\mathcal{A}} m_t(u)\,(J'\,\mu)(u)\,\nu(\mathrm{d}u)=\int_{\mathcal{A}} m_t(u)\int_T m_s(u)\,\mu(\mathrm{d}s)\,\nu(\mathrm{d}u)\\ &= \int_T\int_{\mathcal{A}} m_t(u)\,m_s(u)\,\nu(\mathrm{d}u)\,\mu(\mathrm{d}s)=\int_T k(s,t)\,\mu(\mathrm{d}s)=(K\,\mu)(t)\end{aligned}$$

where

$$\begin{aligned}k(s,t) &= \int_{\mathcal{A}} m_t(u)\,m_s(u)\,\nu(\mathrm{d}u)\\ &= \mathbb{E}\Big(\int_{\mathcal{A}} m_t(u)\,Z(\mathrm{d}u)\int_{\mathcal{A}} m_s(u)\,Z(\mathrm{d}u)\Big)=\mathbb{E}(X_t\,X_s)=\mathrm{cov}(X_t,X_s).\end{aligned}$$

Example 3.8. (Brownian Sheet) By setting $\mathcal{X} = C(T)$ with $T = [0,1]^d$ and choosing $(\mathcal{A}, \nu) = (T, \lambda)$ as our white noise space with Gaussian white noise Z, we get the Brownian sheet X as the integral

$$X_t = \int_T 1_{[0,t]} \mathrm{d}Z.$$

So we can use the above observations by setting $m_t(u) = 1_{[0,t]}(u)$. The kernel of the Brownian Sheet then consists of all functions f that can be written as

$$f(t) = \int_{[0,t]} z(u)\mathrm{d}u$$

for some $z \in L^2\big([0,1]^d\big)$. To write the kernel more explicitly you can interpret z as the mixed derivative of f, that is $z = \frac{\partial^d}{\partial x_1 \dots \partial x_d} f$ and imitate what we did in the case of the Wiener measure.

Example 3.9. (Brownian Bridge) This example is quite important for statistics. You get the Brownian bridge $\dot{W}$ from $\dot{W}_t = W_t - t\,W_1$ where W is a standard Brownian motion and $t \in [0,1]$. Thus the Brownian bridge starts and ends in 0. Just as a good bridge spans a river the Brownian bridge spans the unit interval, see figure 3.3.

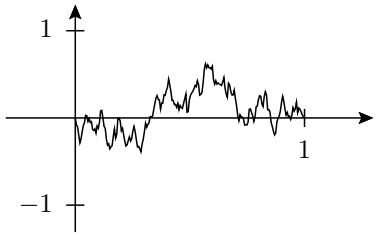

Figure 3.3. A trajectory of Brownian bridge

The Brownian bridge could also be defined by its covariance kernel $k(s,t) = s \wedge t - s\,t$. To represent it as an integral over white noise we can choose $(\mathcal{A}, \nu) = \big([0,1]^2, \lambda\big)$ and $m_t(u) = 1_{B_t}(u)$ where $B_t = [0,t] \times [0, 1-t]$. You can think of B_t as a rectangle in the unit square $[0,1]^2$ with its lower left vertex in $(0,0)$ and its upper right on the diagonal just as in figure 3.4. This ingenious representation indeed does its job

$$\int m_t(u)\, m_s(u)\mathrm{d}u = \int 1_{B_t \cap B_s}(u)\mathrm{d}u = (s \wedge t)\,((1-s) \wedge (1-t)) = k(s,t).$$

Its left as an exercise to complete the construction of the kernel (or wait for example 3.11).

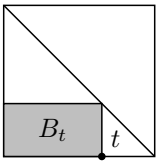

Figure 3.4. Visualization of B_t

Example 3.10. (Fractional Brownian Motion) $W_t^{(\alpha)}$ was defined for $\alpha \in (0,2)$ and $t \in \mathbb{R}$ by its covariance kernel $k(s,t) = \frac{1}{2}(\|s\|^\alpha + \|t\|^\alpha - \|s-t\|^\alpha)$ and $\mathbb{E}\, W_t^{(\alpha)} = 0$.

Let $(\mathcal{A},\nu)=(\mathbb{R},\lambda)$ and

$$m_t(u)=(t-u)_+^{\frac{\alpha-1}{2}}-(-u)_+^{\frac{\alpha-1}{2}}$$

(the second term has to added for staying in L^2). This way, we receive the fractional Brownian motion $W^{(\alpha)}$ as an integral over white noise. For checking this, we recall that the cumbersome definition of k can be replaced by

$$\mathbb{E}\Big(W_s^{(\alpha)}-W_t^{(\alpha)}\Big)^2=|s-t|^\alpha.$$

But then we understand that our definition of m_t is usable up to a constant factor c_α depending on the choice of α:

$$\int_{\mathbb{R}}(m_t(u)-m_s(u))^2\mathrm{d}u=\int_{\mathbb{R}}(t-u)_+^{\frac{\alpha-1}{2}}-(s-u)_+^{\frac{\alpha-1}{2}}\mathrm{d}u=c_\alpha\,|s-t|^\alpha$$

where the last equality can be established by a linear transform.

So we again succeeded in finding the kernel H, but our current representation is not very handily. $f\in H$ iff it can be written as

$$f(t)=\int_{\mathbb{R}}\left((t-u)_+^{\frac{\alpha-1}{2}}-(-u)_+^{\frac{\alpha-1}{2}}\right)z(u)\mathrm{d}u$$

for some $z\in L^2(\mathbb{R})$.

You get a nicer presentation of H if you recognize the above integrand to be the derivative of f of fractional order $\beta=\frac{\alpha+1}{2}$, i.e. setting $\beta-1=\frac{\alpha-1}{2}$.

3.4 Kernel in Literature

In literature you will often find a different approach to the kernel of a Gaussian measure under the name *reproducing kernel Hilbert space*. Nevertheless these concepts coincide as we will see shortly.

Starting with a positive definite function $k\colon T^2\to\mathbb{R}$ (you can think of this as the kernel of a covariance function) the reproducing kernel Hilbert space $H\subset\{f\colon T\to\mathbb{R}\}$ is constructed as follows:

(a) First all sections $k(s,\cdot\,)$ are added to H and the inner product is defined to be

$$(k(s,\cdot\,),k(t,\cdot\,))=k(s,t).$$

(b) Then all linear combinations of elements of H are added.

(c) And finally H is taken to be the completion.

How is this related to our notion of a kernel? Let's assume $(X_t)_{t\in T}$ to be a continuous process (otherwise one has to go into topological details) and $\mathcal{X}=C(T)$. Then $\mathcal{X}'$ equals $M(T)$ containing in particular the Dirac measures δ_t and $h=I\,I'\,\delta_t\in H_P$ is an element of our kernel. Writing h explicitly, we get (here, k is the kernel of the covariance function of X)

$$h(s)=(\delta_s,h)=(\delta_s,I\,I'\,\delta_t)=(I'\,\delta_s,I'\,\delta_t)_{\mathcal{X}_P'}=\mathbb{E}((\delta_s,X)\,(\delta_t,X))=k(t,s),$$

i.e. $h=k(t,\cdot\,)$. The inner product in H_P for such elements is

$$(I\,I'\,\delta_s,I\,I'\,\delta_t)_{H_P}=(I'\,\delta_s,I'\,\delta_t)_{\mathcal{X}_P'}=\mathbb{E}((\delta_s,X)\,(\delta_t,X))=k(t,s).$$

We have therefore shown that the reproducing kernel Hilbert space H for k is a subspace of our kernel H_P.

The two spaces even coincide, which we won't show here just indicating that equality comes from the fact that arbitrary measures in $\mathcal{X}'$ can be approximated by Dirac measures.

3.5 Linear Transformations

Once again, let $\mathcal{X}$ be a linear space and X a Gaussian random vector taking values in $\mathcal{X}$ with covariance function K_X. If $\mathcal{Y}$ is another linear space and $L\colon \mathcal{X} \to \mathcal{Y}$ a linear mapping, then $Y = L\,X$ is Gaussian, too, with covariance function $K_Y = L\,K_X\,L'$. Denote with P the law of X and with Q the law of Y. What do we know about the kernel of Q?

It is reasonable to hope, that $H_Q = L\,H_P$. This is indeed true and the inner product of H_Q is defined by

$$\|y\|_{H_Q} = \inf_{x \in H_P,\, Lx = y} \|x\|_{H_P}.$$

Proof. For simplicity we assume that L is an injection. Using the usual notation, we have the injection I such that $I\,I' = K_X$ and $H_P = I\big(\mathcal{X}_P'\big)$.

Let's use the factorization theorem and choose $\mathcal{H} = \mathcal{X}_P'$ to be the Hilbert space and $J\colon \mathcal{H} \to \mathcal{Y}$ defined by $J = L\,I$ the injective linear mapping. We can choose J this way because

$$J\,J' = L\,(I\,I')\,L' = L\,K_X\,L' = K_Y.$$

Applying the factorization theorem, we then get

$$H_Q = J(\mathcal{H}) = L\big(I\,\mathcal{X}_P'\big) = L\,H_P$$

which was to be shown. Furthermore (because we assumed L to be an injection), for $y = L\,x$

$$\|y\|_{H_Q} = \|x\|_{H_P}.$$

□

Example 3.11. (Brownian Bridge) In example 3.9 we defined the Brownian bridge to be $\dot{W}_t = W_t - t\,W_1$ which is nothing but $\dot{W} = L\,W$ for the linear mapping $L\colon C[0,1] \to C[0,1]$ defined by $(L\,x)(t) = x(t) - t\,x(1)$.

We can therefore use our theorem about linear transformations (but note that L is not an injection) to get the kernel H_Q of the Brownian bridge from the Cameron-Martin kernel H_P of the Brownian motion. $g \in H_Q$ iff it can be written as $g(t) = (L\,h)(t) = h(t) - t\,h(1)$ for some $h \in H_P$. It is easy to see that $g \in H_P$ also, thus $H_Q \subset H_P$. The necessary condition $g(1) = 0$ is sufficient as we can then choose $h = g \in H_P$ to get $g = L\,h$. The kernel of Brownian bridge is thus given as

$$H_Q = \{h \in H_P\colon\ h(1) = 0\} = \big\{h \in \mathrm{AC}[0,1]\colon\ h(0) = h(1) = 0, h' \in L^2[0,1]\big\}$$

endowed with the norm

$$\|h\|_{H_Q}^2 = \|h\|_{H_P}^2 = \int_0^1 (h'(u))^2\,\mathrm{d}u,$$

where $\|h\|_{H_Q} \leqslant \|h\|_{H_P}$ follows from our theorem about linear transformations and equality is left as an exercise.

4 Cameron-Martin Theorem

This section and those to follow try to show the importance of the notion of a kernel for a Gaussian measure.

4.1 Absolute Continuity of Shifted Measures

Let P be a measure defined on the linear space $\mathcal{X}$ and $h \in \mathcal{X}$. We define the shifted measure P_h via

$$P_h(A) = P(A - h).$$

Under what circumstances is P_h absolutely continuous with respect to P, i.e. $P_h \ll P$? We will discuss this question for Gaussian measures P. Thus if $P = N(a, K)$ then $P_h = N(a + h, K)$ which relates to a change of variable for a Gaussian measure.

Example 4.1. (One Dimensional Case) Let $\mathcal{X} = \mathbb{R}$ and $P = N(0, 1)$, $h \in \mathbb{R}$. Then the density of P is

$$p(x) = \frac{1}{\sqrt{2\,\pi}} \exp\left(-\frac{x^2}{2} \right)$$

and thus $q(x) = p(x - h)$ is the density of P_h. It is easy to see that $P_h \ll P$ and

$$\frac{\mathrm{d}P_h}{\mathrm{d}P}(x) = \exp\left(h\,x - \frac{h^2}{2} \right).$$

Therefore, in the one dimensional case we always have absolute continuity, which will change in the general case. But keep an eye on the density of P_h with respect to P in the case of existence.

Theorem 4.2. (Cameron-Martin) *Let P be a centered Gaussian measure on the linear space $\mathcal{X}$. Then $P_h \ll P$ for $h \in \mathcal{X}$ iff $h \in H_P$ where H_P is the kernel of P.*

In the case $h \in H_P$

$$\frac{\mathrm{d}P_h}{\mathrm{d}P}(x) = \exp\left(z(x) - \frac{\|h\|_{H_P}^2}{2} \right) =: q(x) \tag{4.1}$$

where $h = I\,z$ and $I : \mathcal{X}_P' \to \mathcal{X}$ as in definition 3.2.

Equation 4.1 is called *Cameron-Martin formula*.

Proof. The proof can be done in two steps. First suppose $h \in H_P$ and show that equation 4.1 holds. This can be done by calculating the characteristic functions for P_h and $q(x)\,P(\mathrm{d}x)$. This is left as an exercise.

The second step is to show that if $P_h \ll P$ then $h \in H_P$. This is quite difficult and may not be too interesting and is thus left out here. □

Example 4.3. (Brownian Motion) This example exposes the special case of P being the law of a Brownian motion, which is what Cameron and Martin studied. So $\mathcal{X} = C[0, 1]$ and X a Brownian motion with law P and kernel H_P. We already know

$$\|h\|_{H_P}^2 = \int_0^1 (h'(s))^2 \mathrm{d}s.$$

We find $z \in \mathcal{X}'_P = L^2[0,1]$ such that $h = I\,z$ via $z = h'$. It is not very surprising that

$$z(x) = \int_0^1 h'(s)\mathrm{d}x(s),$$

which is evident in the case where x is differentiable, but otherwise left as an exercise. The Cameron-Martin formula now yields

$$\frac{\mathrm{d}P_h}{\mathrm{d}P}(x) = \exp\left(\int_0^1 h'(s)\mathrm{d}x(s) - \frac{1}{2}\int_0^1 (h'(s))^2\mathrm{d}s \right).$$

4.2 Borell Inequality for Shifted Sets

An interesting application of the Cameron-Martin theorem is Borell's inequality for shifted sets. Starting with a symmetric set A (i.e. $A = -A$) with known probability $P(A)$ we can ask how much will the probability change if we consider $A_h = A + h$ instead.

Theorem 4.4. (Borell Inequality for Shifted Sets) *Let P be a centered Gaussian measure on the linear space $\mathcal{X}$, H_P the kernel of P and $h \in H_P$. For any symmetric set A,*

$$P(A+h) \geqslant P(A)\exp\left(-\frac{\|h\|_{H_P}^2}{2} \right). \tag{4.2}$$

This theorem is very interesting because of its generality.

Proof. We have $P(A + h) = P(-A - h) = P(A - h)$. The first equality comes from the centeredness of P and the second from the symmetry of A. We thus conclude

$$\begin{aligned} P(A+h) &= \frac{1}{2}(P(A+h) + P(A-h)) \\ &= \frac{1}{2}\int_A \exp\left(z(x) - \frac{\|h\|_{H_P}^2}{2} \right) + \exp\left(z(-x) - \frac{\|-h\|_{H_P}^2}{2} \right) P(\mathrm{d}x) \\ &= \exp\left(-\frac{\|h\|_{H_P}^2}{2} \right) \int_A \frac{\exp(z(x)) + \exp(-z(x))}{2} P(\mathrm{d}x) \\ &\geqslant \exp\left(-\frac{\|h\|_{H_P}^2}{2} \right) P(A). \end{aligned}$$

□

It is additionally true that $P(A) \geqslant P(A + h)$ for convex symmetric sets A which will be shown later on – see theorem 7.3.

5 Isoperimetric Inequalities

5.1 Introduction to Isoperimetric Inequalities

Isoperimetric problems are like that: of all sets with the same volume find sets with minimal surface. It is very well known that in $\mathbb{R}^n$ it's the balls that minimize their surface.

So let's start with $\mathbb{R}^n$ and the Lebesgue measure λ. If B is a ball and A another "nice" set (i.e. compact and smooth boundary) then the so called *isoperimetric inequality* can be written as

$$\lambda(A) = \lambda(B) \quad \Longrightarrow \quad |\partial A| \geqslant |\partial B|,$$

where the left hand side denotes the surface measure. We don't want to be more specific about the special style of the surface measure. The generalization into infinite dimensions would not be very transparent anyway.

Our goal is to rewrite this inequality in more general ways. So, how can we get rid of the surface measure? Think about what the surface measure actually measures. If we enlarge the set A just a "little bit", then the measure of this enlargement is approximately proportional to the surface measure. Define $A^r = A + B_r$ for any set A. B_r denotes the closed ball of radius r and $A + B_r$ is often referred to as the Minkowski sum. Then

$$\lambda(A^r \backslash A) \geqslant \lambda(B^r \backslash B)$$

is a replacement for $|\partial A| \geqslant |\partial B|$. As $\lambda(A) = \lambda(B)$ we can rewrite it to get the equivalent inequality

$$\lambda(A) = \lambda(B) \quad \Longrightarrow \quad \lambda(A^r) \geqslant \lambda(B^r).$$

Although we had small values of r in mind when constructing this replacement, this inequality turns out to be true for all $r \geqslant 0$, and some important insights even arise from considering large values for r. In this formulation the isoperimetric inequality is applicable to all measurable sets A and not just to "nice" ones.

In a next step, we further want to get rid of the ball B in our inequality, too, to get a representation of this inequality just in terms of the measure. We know that $\lambda(B_r) = c_n\, r^n$ for constants c_n. If B is a ball of radius ρ, then B^r is a ball of radius $\rho + r$ and we get

$$\lambda(A^r) \geqslant \lambda(B^r) = c_n\,(\rho + r)^n = c_n \left(\left(\frac{\lambda(A)}{c_n} \right)^{\frac{1}{n}} + r \right)^n$$

which leads to the inequality

$$\sqrt[n]{\frac{\lambda(A^r)}{c_n}} \geqslant \sqrt[n]{\frac{\lambda(A)}{c_n}} + r.$$

The function $\varphi(v) = \sqrt[n]{v/c_n}$ involved in this representation is called *isoperimetric function*. Given a volume it returns the radius of a ball with just this volume. Try to interpret the inequality in this way.

From $\mathbb{R}^n$ to the sphere $\mathbb{S}^n$. We want to extend our results to other measures, specifically the surface measure σ_n on the sphere $\mathbb{S}^n$ which is characterized by σ-finiteness and invariance under rotations and shifts. The sphere is of special interest in this context, because you can get the Gaussian isoperimetric inequality in some sense as the limit of the spherical isoperimetric inequality, see [Ledoux, 1996] for details.

Using the natural (geodesic) metric on the sphere we define A^r in the same way as before, i.e. $A^r = A + U_r$. Our isoperimetric inequality then writes as

$$\sigma_n(A) = \sigma_n(B) \quad \Longrightarrow \quad \sigma_n(A^r) \geqslant \sigma_n(B^r)$$

with the unknown optimal (in the sense of minimizing their surface) sets B.

It turns out that these sets B with minimal surface are again balls (w.r.t. the natural metric), which can also be interpreted as intersections of half-spaces in $\mathbb{R}^{n+1}$ with the sphere $\mathbb{S}^n$ as shown in figure 5.1.

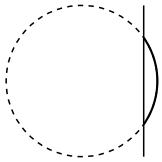

Figure 5.1. A ball on the sphere $\mathbb{S}^1$

5.2 Gaussian Isoperimetric Inequality

Now we go back to our Gaussian measures and start with a standard normal random vector $X = (X_1, ..., X_n)$ with law $P = \mathbb{P}_X = N(0, E_n)$ in $\mathbb{R}^n$ endowed with the natural metric. We again ask for the solution of the isoperimetric inequality

$$P(A) = P(B) \quad \Longrightarrow \quad P(A^r) \geqslant P(B^r)$$

where again $A^r = A + B_r$. The solutions are no more ordinary balls but half-spaces, which can be considered as very large balls, see figure 5.2.

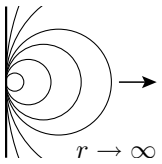

Figure 5.2. Half-spaces as large balls

Just as we did in the case of the Lebesgue measure, we want to rewrite this inequality in such a way that the half-spaces disappear. If B is a half-space, say $B = \{x \in \mathbb{R}^n \colon (f, x) \leqslant \rho\}$ for some $f \in \mathbb{R}^n$, then

$$P(B) = \mathbb{P}(X \in B) = \mathbb{P}((f, X) \leqslant \rho) = \Phi\left(\frac{\rho}{\|f\|}\right),$$

where Φ is the distribution function of a standard normal random variable. The enlargement of B is again a half-space $B^r = \{x \in \mathbb{R}^n \colon (f, x) \leqslant \rho + \|f\| r\}$ (check this as an exercise) and

$$P(B^r) = \mathbb{P}((f, X) \leqslant \rho + \|f\| r) = \Phi\left(\frac{\rho + \|f\| r}{\|f\|}\right) = \Phi\big(\Phi^{-1}(P(B)) + r\big).$$

As in the case of the Lebesgue measure, we then obtain (for $P(A^r) \geqslant P(B^r)$) the analytical version of the isoperimetric inequality

$$\Phi^{-1}(P(A^r)) \geqslant \Phi^{-1}(P(A)) + r.$$

Remark 5.1. It is not crucial for this inequality, that P is standard normal. Sure, if P is not standard normal then no more all half-spaces minimize their surface. But for establishing the isoperimetric inequality in its analytical version we did only use the fact that for each set A there exists a optimal set B which is a half-space such that $P(A) = P(B)$.

We now want to generalize this inequality, that does not depend on the dimension, to the infinite dimensional case. How is A^r defined then?

Theorem 5.2. (Gaussian Isoperimetric Inequality) *Let P be a centered Gaussian measure on a linear space $\mathcal{X}$ and $\mathcal{E}_1$ the unit ball of the kernel H_P, i.e. $\mathcal{E}_1 = \{h \in H_P \colon \|h\|_{H_P} \leqslant 1\} \subseteq \mathcal{X}$. Define $A^r = A + r\,\mathcal{E}_1$. Then for any measurable set A the following is true*

$$\Phi^{-1}(P_\star(A^r)) \geqslant \Phi^{-1}(P(A)) + r.$$

$P_\star$ is given by $P_\star(A) = \sup_{B \subseteq A} P(B)$ where the supremum is taken over all measurable subsets of A.

The formula states that there exists a measurable subset of A^r such that the isoperimetric inequality of the finite case almost holds. This may be necessary as A^r does not need to be measurable itself. The isoperimetric inequality 5.2 was obtained and published independently in 1974-1975 by C. Borell in Sweden and by V. Sudakov and B. Tsirelson in USSR. In the literature it is sometimes erroneously associated only with C. Borell's name.

The unit ball of the kernel $\mathcal{E}_1$ is sometimes referred to as *ellipsoid of distortion* of P. For an explanation, think of P as a centered Gaussian measure on $\mathbb{R}^2$. Suppose that the covariance operator K is not degenerated and has a decomposition $K = I' I$. Then $\mathcal{E}_1 = I\,B$ is the linear image of the unit ball B of $\mathcal{X}$ and thus indeed an ellipsoid.

5.3 Concentration Principle

Let X be a centered Gaussian random vector on the linear space $\mathcal{X}$. Our goal in this subsection is to evaluate probabilities of the kind

$$\mathbb{P}(f(X) \geqslant r) \quad \text{and} \quad \mathbb{P}(f(X) \leqslant -r)$$

for $r \to \infty$ where $f \colon \mathcal{X} \to \mathbb{R}$ is a functional. We will be able to give bounds for such probabilities for magically many functionals.

Definition 5.3. *A functional f is said to be H_P-Lipschitzian with Lipschitz constant σ, i.e. $f \in \mathrm{Lip}(\sigma)$, iff for all $x \in \mathcal{X}$ and $h \in H_P$*

$$|f(x+h) - f(x)| \leqslant \sigma \|h\|_{H_P}.$$

Let m be the median of $f(X)$, which means that $\mathbb{P}(f(X) \geqslant m) \geqslant \frac{1}{2}$ and $\mathbb{P}(f(X) \leqslant m) \geqslant \frac{1}{2}$.

Theorem 5.4. (Concentration Principle) *If $f \in \mathrm{Lip}(\sigma)$ then for all $r > 0$ the following statements are true*

$$\begin{aligned} \mathbb{P}(f(X) \geqslant m + r) &\leqslant 1 - \Phi\Big(\frac{r}{\sigma}\Big) \\ \mathbb{P}(f(X) \leqslant m - r) &\leqslant 1 - \Phi\Big(\frac{r}{\sigma}\Big). \end{aligned}$$

The concentration principle tells us that the random variable $f(X)$ is concentrated around its median as strong as a normally distributed variable (which is super-exponential).

Proof. Set $A = \{x \in \mathcal{X}\colon f(x) \leqslant m\}$ and use the isoperimetric inequality 5.2 as well as $P(A) \geqslant \frac{1}{2}$ to achieve

$$\Phi^{-1}(P_\star(A^r)) \geqslant \Phi^{-1}(P(A)) + r \geqslant r,$$

which implies $P_\star(A^r) \geqslant \Phi(r)$. Hence for $\delta > 0$ we find $B_\delta \subseteq A^r$ measurable so that $P(B_\delta) \geqslant \Phi(r) - \delta$. Given $y \in A^r$, there is $h \in H_P$ with $\|h\|_{H_P} \leqslant r$ and $x \in A$ such that

$$f(y) = f(x+h) \leqslant f(x) + \sigma r \leqslant m + \sigma r.$$

Then

$$\mathbb{P}(f(X) \leqslant m + \sigma r) \geqslant \mathbb{P}(X \in B_\delta) = P(B_\delta) \geqslant \Phi(r) - \delta.$$

As δ can be chosen arbitrarily, we get

$$\mathbb{P}(f(X) > m + r) \leqslant 1 - \Phi\Big(\frac{r}{\sigma}\Big),$$

and as the right-hand side is continuous our claim is proven. For the second inequality consider $-f$ instead of f. □

Example 5.5. (Gaussian Processes) Let $\mathcal{X} = C(T)$ with T compact metric and X a centered Gaussian process on T. Consider the functional

$$f\colon \mathcal{X} \to \mathbb{R}, \quad x \mapsto \sup_{t \in T} x(t)$$

and define σ via

$$\sigma^2 = \sup_{t \in T} \operatorname{Var}(X_t).$$

We will show that $f \in \operatorname{Lip}(\sigma)$. This will give us e.g. the estimate

$$\mathbb{P}\Big(\sup_{t \in T} X_t \geqslant r\Big) \leqslant 1 - \Phi\Big(\frac{r-m}{\sigma}\Big) \leqslant \exp\Big(-\frac{(r-m)^2}{2\sigma^2}\Big), \tag{5.1}$$

where m is the median of $\sup_{t \in T} X_t$. This will be continued in example 8.5.

First,

$$|f(x+h) - f(x)| = \Big|\sup_{t \in T}(x(t) + h(t)) - \sup_{t \in T} x(t)\Big| \leqslant \sup_{t \in T} |h(t)|,$$

and since $h \in H_P$, we have $z \in \mathcal{X}_P'$ such that (recall that $\mathcal{X}' = M(T)$)

$$h(t) = (\delta_t, h) = (\delta_t, I\,z) = (I'\,\delta_t, z)_{\mathcal{X}_P'}.$$

The Cauchy-Schwarz inequality therefore yields

$$|h(t)| \leqslant \|I'\,\delta_t\|_{\mathcal{X}_P'}\,\|z\|_{\mathcal{X}_P'} = \|I'\,\delta_t\|_{\mathcal{X}_P'}\,\|h\|_{H_P},$$

and

$$\|I'\,\delta_t\|_{\mathcal{X}_P'}^2 = \mathbb{E}(\delta_t, X)^2 = \mathbb{E}\,X_t^2 = \operatorname{Var} X_t \leqslant \sigma^2$$

implies that $f \in \operatorname{Lip}(\sigma)$.

We are now equipped with a tool for estimating large deviations of sup-like functions.

6 Large Deviations

6.1 Introduction

The theory of large deviations exists in a more general context than the Gaussian one. However, we will exercise this theory quite restricted here.

Let $\mathcal{X}$ be a topological space, not necessary linear, and $Y_n \in \mathcal{X}$ a sequence of random elements that converges in probability to a point $a \in \mathcal{X}$, i.e.

$$Y_n \xrightarrow{\mathbb{P}} a.$$

A measurable set $A \subset \mathcal{X}$ that is separated from a is called a *large deviation*. Since A is separated from a we clearly have

$$P(Y_n \in A) \longrightarrow 0.$$

In the sequel we will be interested in the precise speed of this convergence.

Example 6.1. Let $\mathcal{X} = \mathbb{R}$, X_i i.i.d. random variables and $Y_n = \frac{1}{n}\sum_{j=1}^{n} X_j$. Then

$$Y_n \xrightarrow{\text{a.s.}} a = \mathbb{E}\, X$$

and typical large deviation probabilities are given by

$$P(Y_n \geqslant a + \rho) \quad \text{or} \quad P(Y_n \leqslant a - \rho)$$

which are bounded exponentially if X_j have finite exponential moments.

Our goal is to describe the convergence of $P(Y_n \in A)$ in a way like

$$P(Y_n \in A) \approx \exp(-\nu_n I(A)) \tag{6.1}$$

where $\nu_n \to \infty$ is the *rate*, $I\colon \mathcal{X} \to [0, \infty]$ the *deviation function* and $I(A) = \inf_{h \in A} I(h)$. Note that the rate is to be the same for all sets A. $I(h)$ can be seen as kind of a distance between h and a, but it lacks properties like the triangle inequality. 6.1 then states an exponential decrease.

We won't require 6.1 for all large deviations A. If the sequence Y_n fulfils

$$\begin{aligned} \limsup_{n\to\infty} \frac{\log P(Y_n \in A)}{\nu_n} &\leqslant -I(A) \quad \text{for all closed } A \\ \text{and} \quad \liminf_{n\to\infty} \frac{\log P(Y_n \in A)}{\nu_n} &\geqslant -I(A) \quad \text{for all open } A, \end{aligned}$$

we say that the *strong large deviation principle*, abbreviated as LDP, is valid for Y_n. If we replace closed by compact in our conditions, we get the notion of the weak large deviation principle. Literature sometimes refers to the strong principle as the narrow one, and the weak principle is known as the vague one.

If Y_n obeys the LDP, then for any measurable set A

$$-I\Big(\mathring{A}\Big) \leqslant \liminf_{n\to\infty} \frac{\log P(Y_n \in A)}{\nu_n} \leqslant \limsup_{n\to\infty} \frac{\log P(Y_n \in A)}{\nu_n} \leqslant -I\big(\bar{A}\,\big).$$

In particular for regular sets, meaning that $I(\mathring{A}) = I(\bar{A}\,)$,

$$\lim_{n\to\infty} \frac{\log P(Y_n \in A)}{\nu_n} = -I(A)$$

which is exactly what we had in mind when writing down 6.1.

6.2 Gaussian Large Deviations

Let $X \sim N(0, K)$ be a Gaussian centered random vector on a linear space $\mathcal{X}$ and define $Y_n = \frac{1}{n} X$. Obviously $Y_n \longrightarrow 0$ in probability. Our goal will be to show that LDP is valid for Y_n and to determine the appropriate rate and deviation function.

Since

$$\mathbb{P}(Y_n \in A) = \mathbb{P}\left(\frac{1}{n} X \in A\right) = \mathbb{P}(X \in n\,A) = P(n\,A),$$

we have a representation of the large deviation probabilities where the sets $n\ A$ are changing instead of the measures $\mathbb{P}_{Y_n}$. Figure 6.1 shows how these changing sets can look like. For the sake of generality and because it does not make a difference, we will examine $P(r\,A)$ for $r \in \mathbb{R}^+$ and $r \to \infty$.

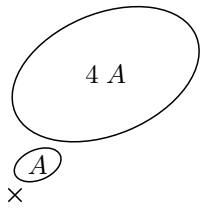

Figure 6.1. Changing sets $r\,A$

Theorem 6.2. (Gaussian Large Deviation) *Let P be a centered Gaussian measure on $\mathcal{X}$. Then*

$$\begin{aligned} \limsup_{r\to\infty} \frac{\log P(r\,A)}{r^2} &\leqslant -I(A) \quad \textit{for all closed } A \\ \textit{and} \quad \liminf_{r\to\infty} \frac{\log P(r\,A)}{r^2} &\geqslant -I(A) \quad \textit{for all open } A, \end{aligned}$$

where $I(A) = \inf_{h\in A} I(h)$ and

$$I(h) = \begin{cases} \frac{1}{2}\,||h||^2_{H_P} & \textit{for } h \in H_P \\ \infty & \textit{otherwise.} \end{cases}$$

In other words, the sequence $Y_n = \frac{1}{n} X_n$ obeys the LDP with rate n^2 and deviation function I.

Proof. Set H be the kernel of X. Let A be closed and set $\rho = \inf_{h \in A \cap H} \|h\|_H$. Then $I(A) = \frac{\rho^2}{2}$, and the case $\rho \leqslant 0$ is trivial. So assume $\rho > 0$ and take any $\bar{\rho} \in (0, \rho)$. The ball

$$B = \{h \in H : \|h\|_H \leqslant \bar{\rho}\}$$

in H of radius $\bar{\rho}$ has no intersection with A, and one can show that B is compact. We can therefore separate A and B with open sets. In particular, we find an open neighborhood V of 0 such that

$$(B + V) \cap A = \emptyset$$

and thus for $r > 0$

$$(r\,B + r\,V) \cap r\,A = \emptyset.$$

We use this for

$$\begin{aligned} P(r\,A) &\leqslant P((r\,B + r\,V)^c) \\ &\leqslant 1 - \Phi\big(\Phi^{-1}(P(r\,V)) + r\,\bar{\rho}\,\big) \quad \text{isop. inequality 5.2} \\ &\leqslant 1 - \Phi(r\,\bar{\rho}\,) \quad \text{for } r \text{ big enough, because } P(r\,V) \to 1 \\ &\leqslant \exp\Big(-\frac{r^2\,\bar{\rho}^{\,2}}{2}\Big), \end{aligned}$$

which implies

$$\frac{\log P(r\,A)}{r^2} \leqslant -\frac{\bar{\rho}^{\,2}}{2}$$

for r big enough. Since $\bar{\rho} < \rho$ was arbitrary we conclude

$$\limsup_{r \to \infty} \frac{\log P(r\,A)}{r^2} \leqslant -\frac{\rho^2}{2} = -\,I(A).$$

Now let A be open and $h \in A \cap H$. If we don't find such an h, then $I(A) = \infty$ and the assertion is trivial. Since A is open we can choose a neighborhood V of 0 such that $V + h \subset A$ and V symmetric (if V is not symmetric we can migrate to $V \cap -V$).

$$\begin{aligned} P(r\,A) &\geqslant P(r\,(V + h)) = P(r\,V + r\,h) \\ &\geqslant P(r\,V)\exp\Big(-\frac{\|r\,h\|_H^2}{2}\Big) \quad \text{Borell shift inequality 4.2} \\ &\geqslant \frac{1}{2}\exp\Big(-\frac{r^2\,\|h\|_H^2}{2}\Big) \quad \text{for } r \text{ big enough since } P(r\,V) \to 1 \end{aligned}$$

and therefore

$$\frac{\log P(r\,A)}{r^2} \geqslant \frac{\log\left(\frac{1}{2}\right)}{r^2} - \frac{\|h\|_H^2}{2}$$

for r big enough. We thus achieved

$$\liminf_{r \to \infty} \frac{\log P(r\,A)}{r^2} \geqslant -\frac{\|h\|_H^2}{2}$$

for any $h \in A \cap H$. □

Example 6.3. (Brownian Bridge) This is somewhat a toy example but it may help to see how things work. Let X_t be a Brownian bridge as defined in example 3.9, i.e. $T = [0,1]$ and $\mathcal{X} = C(T)$. We want to estimate

$$\mathbb{P}\left(\sup_{t \in T} |X_t| \geqslant r \right)$$

for big r which has some importance in statistics. Using the closed set

$$A = \left\{ x \in \mathcal{X} \colon \sup_{t \in T} |x(t)| \geqslant 1 \right\}$$

we can write this probability as

$$\mathbb{P}\left(\sup_{t \in T} |X_t| \geqslant r \right) = P(r\,A).$$

It can be shown that A is regular in the sense that $I(\mathring{A}) = I(\bar{A})$ and thus

$$\lim_{r \to \infty} \frac{\log P(r\,A)}{r^2} = -\,I(A) = -\,\frac{1}{2} \inf_{h \in A \cap H} \|h\|_H^2.$$

We now claim that

$$\inf_{h \in A \cap H} \|h\|_H^2 = 4$$

and therefore $I(A) = 2$, which provides the asymptotic

$$\mathbb{P}\left(\sup_{t \in T} |X_t| \geqslant r \right) \sim \exp\bigl(-\,2\,r^2\bigr).$$

To show this, we recall how $\|h\|_H^2$ was defined. We have to determine the infimum value of

$$\|h\|_H^2 = \int_0^1 h'(t)^2 \,\mathrm{d}t$$

where h is an absolutely continuous function over $[0,1]$ such that $h(0) = h(1) = 0$ and additionally $\sup_{t \in [0,1]} |h(t)| \geqslant 1$. We claimed that this infimum equals 4 and this value is actually attained by the "hat" function g shown in figure 6.2

$$g(x) = \begin{cases} 2\,x & \text{for } x \in \left[0, \frac{1}{2}\right] \\ 2 - 2\,x & \text{for } x \in \left(\frac{1}{2}, 1\right] \end{cases}$$

because $g'(x)^2 = 4$.

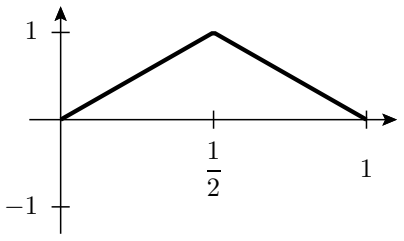

Figure 6.2. Hat function g

It therefore remains to show $\|h\|_H^2 \geqslant 4$ for any allowed h. Choose $s \in [0, 1]$ such that $|h(s)| \geqslant 1$. Then due to the Hoelder inequality,

$$1 \leqslant |h(s)| = \left| \int_0^s h'(t)\,\mathrm{d}t \right| \leqslant \sqrt{s} \sqrt{\int_0^s h'(t)^2\,\mathrm{d}t},$$

and therefore

$$\int_0^s h'(t)^2\,\mathrm{d}t \geqslant \frac{1}{s}.$$

With the same argument

$$\int_s^1 h'(t)^2\,\mathrm{d}t \geqslant \frac{1}{1-s},$$

which yields our claim

$$\|h\|_H^2 \geqslant \frac{1}{s} + \frac{1}{1-s} = \frac{1}{s\,(1-s)} \geqslant 4.$$

So you have to be ready to solve difficult optimization problems when considering large deviation problems.

Example 6.4. (Gaussian Processes) Here we will just state results for general Gaussian processes. If X is an arbitrary centered Gaussian process, we are interested in the probability

$$\mathbb{P}\left(\sup_{t \in T} |X_t| \geqslant r \right) = P(r\,A),$$

where A is defined just as in example 6.3. From example 5.5 we know that for

$$\sigma^2 = \sup_{t \in T} \operatorname{Var}(X_t)$$

we have

$$\mathbb{P}\left(\sup_{t \in T} X_t \geqslant r \right) \leqslant \exp\left(-\frac{r^2}{2\,\sigma^2} \right)$$

which is just inequality 5.1. It is therefore not too striking that

$$\lim_{r \to \infty} \frac{\log \mathbb{P}(\sup_{t \in T} X_t \geqslant r)}{r^2} = -\frac{1}{2\,\sigma^2},$$

but this is indeed just what we get when looking at the supremum over the absolute values

$$\lim_{r \to \infty} \frac{\log \mathbb{P}(\sup_{t \in T} |X_t| \geqslant r)}{r^2} = \lim_{r \to \infty} \frac{\log P(r\,A)}{r^2} = -\frac{1}{2\,\sigma^2}.$$

We can use this inequality to reproduce our result from example 6.3. If X is a Brownian bridge, then $\operatorname{Var} X_t = t - t^2$ which is obtained from the covariance kernel of X given in example 3.9. Thus

$$\sigma^2 = \sup_{t \in T} \operatorname{Var}(X_t) = \sup_{t \in T} \left(t - t^2 \right) = t - t^2 \big|_{t = \frac{1}{2}} = \frac{1}{4}$$

which yields again $I(A) = \frac{1}{2\,\sigma^2} = 2$.

Exercise 6.1. Let $(\mathcal{X}, \|\cdot\|)$ be a Banach space and $X \sim N(0, K)$ a centered Gaussian random vector taking values in $\mathcal{X}$. Show that

$$\lim_{r\to\infty} \frac{\mathbb{P}(\|X\| \geqslant r)}{r^2} = -\frac{1}{2\,||K||}.$$

7 Convexity and Other Inequalities

7.1 Concavity of Measures

Let V be a linear space. A function $\varphi\colon V \to \mathbb{R}$ is called *concave* if

$$\varphi(\alpha\, x + (1-\alpha)\, y) \geqslant \alpha\, \varphi(x) + (1-\alpha)\, \varphi(y)$$

for all $x, y \in V$ and all $\alpha \in [0,1]$. We now want to introduce the notion of concavity for measures μ. Unfortunately, the first try via

$$\mu(\alpha\, A + (1-\alpha)\, B) \geqslant \alpha\, \mu(A) + (1-\alpha)\, \mu(B)$$

is of no use because there are no interesting measures having this property. Instead we allow different notions of concavity depending on some function Q

$$\mu(\alpha\, A + (1-\alpha)\, B) \geqslant Q(\alpha, \mu(A), \mu(B))$$

again for all measurable A, B and all $\alpha \in [0, 1]$. Two concrete types of concavity are due to Borell and Ehrhard.

Example 7.1. (Borell concavity) Concavity due to Borell is given by the condition

$$\log \mu(\alpha\, A + (1-\alpha)\, B) \geqslant \alpha \log \mu(A) + (1-\alpha) \log \mu(B), \tag{7.1}$$

and is therefore also called *log-concavity*. Equivalently,

$$\mu(\alpha\, A + (1-\alpha)\, B) \geqslant \mu(A)^{\alpha}\, \mu(B)^{1-\alpha}.$$

It turns out that any Gaussian measure P is log-concave, but 7.1 is never sharp.

Example 7.2. (Ehrhard concavity) For the particular case of Gaussian measures

$$\Phi^{-1}\, \mu(\alpha\, A + (1-\alpha)\, B) \geqslant \alpha\, \Phi^{-1}\, \mu(A) + (1-\alpha)\, \Phi^{-1}\, \mu(B)$$

defines concavity due to Ehrhard.

In contrast to log-concavity, this inequality is sharp e.g. for parallel half-spaces.

Theorem 7.3. (Anderson Inequality) *Let P be a centered Gaussian measure on $\mathcal{X}$ and $A \subset \mathcal{X}$ a measurable symmetric convex set. Then for any $h \in \mathcal{X}$*

$$P(A) \geqslant P(A + h).$$

Proof. Since A is symmetric

$$A - h = -A - h = -(A + h),$$

and we can apply log-concavity 7.1 to the sets $A' = A + h$ and $B' = A - h$ to get

$$\log P\left(\frac{A' + B'}{2}\right) \geqslant \frac{1}{2}\log P(A') + \frac{1}{2}\log P(B') = \log P(A')$$

as $P(A') = P(B')$ due to the symmetry of P. Because A is convex

$$\frac{A' + B'}{2} = \frac{(A + h) + (A - h)}{2} = \frac{A + A}{2} = A,$$

which implies

$$P(A) \geqslant P(A') = P(A + h). \qquad \square$$

Example 7.4. (Gaussian Processes) Let X be a centered Gaussian process on T. Since the set

$$A = \{x \in C(T) \colon |x(t)| \leqslant \varepsilon(t)\}$$

with $\varepsilon \in C(T)$ and $\varepsilon > 0$ is symmetric convex, we can apply the Anderson inequality 7.3 to get

$$P(|X_t - h(t)| \leqslant \varepsilon(t) \text{ for all } t \in T) \quad \leqslant \quad P(|X_t| \leqslant \varepsilon(t) \text{ for all } t \in T).$$

For ε chosen to be constant this means that the event that the process X is in a tube of width ε around a function h, can be replaced by the much simpler event that X is bounded by ε.

7.2 Correlation Conjecture

If X_1 and X_2 are independent Gaussian vectors on $\mathcal{X}$, then

$$\mathbb{P}(X_1 \in A_1, X_2 \in A_2) = \mathbb{P}(X_1 \in A_1)\,\mathbb{P}(X_2 \in A_2).$$

But often we have to deal with dependent vectors. Experience says that still

$$\mathbb{P}(X_1 \in A_1, X_2 \in A_2) \geqslant \mathbb{P}(X_1 \in A_1)\,\mathbb{P}(X_2 \in A_2) \tag{7.2}$$

might be true for reasonable A_1, A_2 (though it is not true for all measurable sets). Anyhow 7.2 is yet not proved and is known as the *correlation conjecture*. In particular, it is conjectured that 7.2 is true e.g. for X_j centered Gaussian and A_j symmetric convex (which is similar to a ball).

There is another form of the correlation conjecture. Let X be Gaussian and $\mathcal{A}$, $\mathcal{B}$ symmetric convex sets. Then it is believed that

$$\mathbb{P}(X \in \mathcal{A} \cap \mathcal{B}) \geqslant \mathbb{P}(X \in \mathcal{A})\,\mathbb{P}(X \in \mathcal{B}) \tag{7.3}$$

which looks like some "independence of X from itself". You can see the relation if you set $X = (X_1, X_2)$ on $\mathcal{X} \times \mathcal{X}$ and define the symmetric convex sets

$$\mathcal{A} = \{(x_1, x_2) \in \mathcal{X} \times \mathcal{X} \colon x_1 \in A\}, \quad \mathcal{B} = \{(x_1, x_2) \in \mathcal{X} \times \mathcal{X} \colon x_2 \in B\}.$$

Conjecture 7.3 is proved for some special cases.

- It is known to be true for Gaussian measures on $\mathcal{X} = \mathbb{R}^2$,
- or if one of the sets $\mathcal{A}$, $\mathcal{B}$ is a symmetric strip, i.e.

  $$\{x \in \mathcal{X} : -\alpha \leqslant (f, x) \leqslant \alpha\}$$

  for some $f \in \mathcal{X}'$ and $\alpha \in \mathbb{R}$.
- Using the last fact step by step e.g. for $\mathcal{B} = \mathcal{B}_1 \cap \mathcal{B}_2$

  $$\mathbb{P}(X \in \mathcal{A} \cap \mathcal{B}) \geqslant \mathbb{P}(X \in \mathcal{A} \cap \mathcal{B}_1)\, \mathbb{P}(X \in \mathcal{B}_2) \geqslant \mathbb{P}(X \in \mathcal{A})\, \mathbb{P}(X \in \mathcal{B}_1)\, \mathbb{P}(X \in \mathcal{B}_2),$$

  you get for a finite intersection $\mathcal{B}$ of strips $\mathcal{B}_j$, i.e. $\mathcal{B} = \bigcap_{j=1}^n \mathcal{B}_j$, the following variant of conjecture 7.3

  $$\mathbb{P}(X \in \mathcal{A} \cap \mathcal{B}) \geqslant \mathbb{P}(X \in \mathcal{A}) \prod_{j=1}^n \mathbb{P}(X \in \mathcal{B}_j)$$

  which can be extended to countable intersections as well. This is called Šidak inequality.

Although the variant of the correlation conjecture 7.3 is not yet proved there is a weaker but proved version that suffices for some applications.

Theorem 7.5. (Weak Correlation Inequality) *For any $\varepsilon > 0$ there exists a constant $K_\varepsilon > 0$ such that for any Gaussian vector X and A, B symmetric convex*

$$\mathbb{P}(X \in A \cap B) \geqslant \mathbb{P}(X \in (1-\varepsilon)\, A)\, \mathbb{P}\left(X \in \frac{B}{K_\varepsilon}\right).$$

Example 7.6. Lets give an application of the weak correlation inequality to small deviation theory. If X is a Gaussian vector on $(\mathcal{X}, \|\cdot\|)$ which is composed as $X = X_1 + X_2$ where X_1 and X_2 are possibly dependent Gaussian vectors. If roughly speaking X_1 is a nice process and X_2 a small one, then we can estimate

$$\begin{aligned}
\mathbb{P}(\|X\| \leqslant r) &= \mathbb{P}(\|X_1 + X_2\| \leqslant r) \\
&\geqslant \mathbb{P}(\|X_1\| \leqslant (1-\varepsilon)\, r, \|X_2\| \leqslant \varepsilon\, r) \quad \text{triangle inequality} \\
&\geqslant \mathbb{P}\big(\|X_1\| \leqslant (1-\varepsilon)^2\, r\big)\, \mathbb{P}\left(\|X_2\| \leqslant \frac{\varepsilon\, r}{K_\varepsilon}\right) \quad \text{weak correlation inequality}
\end{aligned}$$

where the first factor is the one of asymptotic relevance.

7.3 Shift-isoperimetric Inequalities and S-conjecture

We already considered isoperimetric inequalities that arise when we consider an event A and swap to a neighborhood $A^r = A + B_r$. For Gaussian measures P we obtained the isoperimetric inequality 5.2

$$\Phi^{-1}(P_\star(A^r)) \geqslant \Phi^{-1}(P(A)) + r.$$

Now instead of enlargements A^r we will be interested in shifts $A + h$ and dilations $r\, A$.

For $h \in H$ in the kernel of P we have the shift-isoperimetric inequality (which is due to J. Kuelbs and Wenbo Li)

$$\Phi^{-1}(P(A)) - \|h\|_H \leqslant \Phi^{-1}(P(A+h)) \leqslant \Phi^{-1}(P(A)) + \|h\|_H$$

where equality holds for half-spaces.

Switching to dilations $r\,A$ for $r > 0$ we define

$$G(r) = \Phi(r) - \Phi(-r) = 2\,\Phi(r) - 1 = \mathbb{P}(|X| \leqslant r) \quad \text{for } X \sim N(0,1)$$

which thus can be interpreted as the Gaussian measure of a strip centered in 0. Let X be a centered Gaussian measure and A convex symmetric. The "S-conjecture", stated by Kwapien and Sawa, and proved later by Latala and Oleszkiewitz claims that for ρ such that $G(\rho) = P(A)$

$$\begin{aligned} P(r\,A) &\leqslant G(r\,\rho) \quad \text{for } r \leqslant 1 \\ P(r\,A) &\geqslant G(r\,\rho) \quad \text{for } r \geqslant 1. \end{aligned}$$

In other words – the measure of $P(r\,A)$ decreases faster than $P(r\,S)$ for $r \leqslant 1$ where S is a strip of the same measure as A. If $r \geqslant 1$ then the measure $P(r\,A)$ decreases (when $r \downarrow 0$) faster then $P(r\,S)$.

Unfortunately, this conjecture is not too useful as the shrink of $P(r\,A)$ for $r < 1$ is bounded exponentially in most cases whereas the "conjecture" just provides the almost linear bound $G(r\,\rho)$.

8 Metric Entropy and Sample Paths

8.1 General Metric Entropy

Let (T, ρ) be a metric space. We define the following covering/packing numbers

$N(\varepsilon)$	minimal number of sets of diameter $\leqslant \varepsilon$ needed to cover T
$N'(\varepsilon)$	minimal number of balls of diameter $\leqslant \varepsilon$ needed to cover T
$M(\varepsilon)$	maximal number of points with pairwise distance $> \varepsilon$.

Example 8.1. If ρ is the discrete metric on T, i.e. every two points have distance 1, then

$$\begin{aligned} N(\varepsilon) &= \begin{cases} 1 & \text{for } \varepsilon \geqslant 1 \\ |T| & \text{for } \varepsilon < 1 \end{cases} \\ N'(\varepsilon) &= \begin{cases} 1 & \text{for } \varepsilon \geqslant 2 \\ |T| & \text{for } \varepsilon < 2 \end{cases} \\ M(\varepsilon) &= \begin{cases} 1 & \text{for } \varepsilon \geqslant 1 \\ |T| & \text{for } \varepsilon < 1 \end{cases}. \end{aligned}$$

From the definition it follows that

$$N(\varepsilon) \leqslant M(\varepsilon) \leqslant N'(\varepsilon) \leqslant N\Big(\frac{\varepsilon}{2}\Big). \tag{8.1}$$

$N(\varepsilon) \leqslant N'(\varepsilon)$ is trivially true, because a covering with balls is always a covering with arbitrary sets, too. A set of diameter ε is always contained in the ball of diameter $2\,\varepsilon$ around any of its points and thus $N'(\varepsilon) \leqslant N\left(\frac{\varepsilon}{2}\right)$.

Definition 8.2. (Metric Entropy) *The function*

$$\mathbb{H}(\varepsilon) = \log N(\varepsilon)$$

is called metric entropy *of* (T, ρ) *and*

$$C(\varepsilon) = \log M(\varepsilon)$$

metric capacity*of* (T, ρ) .

The interesting property of the metric entropy H is its behavior for $\varepsilon \to 0$. A typical graph is shown in figure 8.1.

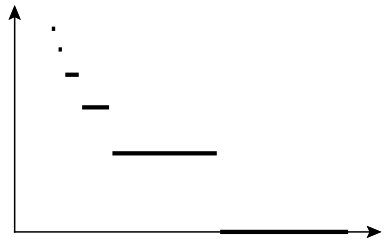

Figure 8.1. Typical graph of a metric entropy

8.2 Metric Entropy of Gaussian Processes

Every centered Gaussian process X on T generates a semi-metric ρ (also known as the Dudley metric) on T via

$$\rho(s,t)^2 = \mathbb{E}(X_t - X_s)^2 = \operatorname{Var}(X_t - X_s).$$

The metric entropy $\mathbb{H}(X, \varepsilon)$ on (T, ρ) is called the metric entropy of X. The capacity $C(X, \varepsilon)$ is defined analogously. For applications it is important that the metric entropy of a process X is in some cases easy to evaluate.

The integral

$$\Psi(X, u) = \int_0^u \sqrt{\mathbb{H}(X, \varepsilon)}\, \mathrm{d}\varepsilon$$

is called the *Dudley integral* of X. It's an increasing function that is either everywhere finite or everywhere infinite. If you recall that the metric entropy $\mathbb{H}(\varepsilon)$ is defined as $\log N(\varepsilon)$ you can see that

$$\sqrt{\mathbb{H}} = \sqrt{\log N} \quad \text{is the inverse of} \quad \exp\bigl(N^2\bigr)$$

which makes the Dudley integral perhaps a bit less arbitrarily chosen.

Theorem 8.3. (Continuity Condition) *If the Dudley integral* $\Psi(X, u) < \infty$ *is finite for a centered Gaussian process* X*, then* X *is continuous (in the sense that a continuous version of* X *exists).*

It is therefore sufficient for a process X to be continuous that

$$\mathbb{H}(X, \varepsilon) \leqslant \frac{c}{\varepsilon^{2-\delta}}$$

for some constant c and $\delta > 0$. Or equivalently,

$$N(X, \varepsilon) \leqslant \exp\Big(\frac{c}{\varepsilon^{2-\delta}}\Big).$$

Note that these conditions are optimal in the sense that counter-examples exist for $\delta = 0$.

Theorem 8.3 makes it easy to see that a Brownian motion W is continuous. Remember that $N(W, \varepsilon)$ is the metric entropy on $T = [0, 1]$ equipped with the metric

$$\rho(s, t) = \sqrt{\mathbb{E}\,(X_t - X_s)^2} = \sqrt{|t - s|}.$$

An ε-ball w.r.t. the metric ρ is just an ε^2-ball w.r.t. the standard metric. Therefore we need approximately ε^{-2} balls w.r.t. ρ to cover the unit interval $[0, 1]$, i.e. for ε small enough

$$N(W, \varepsilon) \leqslant \frac{c}{\varepsilon^2}$$

and the Dudley integral is trivially finite.

The continuity condition $\Psi(X, u) < \infty$ is not necessary in general, but a (very deep) result of Fernique states that it is necessary for continuity and boundedness of centered Gaussian stationary processes, see also [Ledoux, 1996]. For centered stationary processes the distribution of $X_t - X_s$ only depends on $|t - s|$. The Dudley metric ρ is therefore induced by a Fréchet metric p

$$\rho(s, t) = \sqrt{\mathbb{E}\,(X_t - X_s)^2} = p(|t - s|).$$

Theorem 8.4. (Dudley Theorem) *Let X be a centered Gaussian process and*

$$\sigma^2 = \sup_{t \in T} \mathbb{E}\, X_t^2.$$

Then

$$\mathbb{E} \sup_{t \in T} X_t \leqslant 4\sqrt{2}\, \Psi\Big(\frac{\sigma}{2}\Big).$$

Example 8.5. (Estimate for Supremum) We can use Dudley's theorem to get an estimate for the supremum of a centered Gaussian process X. From example 5.5 we know that

$$\mathbb{P}\Big(\sup_{t \in T} X_t \geqslant r\Big) \leqslant 1 - \Phi\Big(\frac{r - m}{\sigma}\Big)$$

where m is the median of $\sup_{t \in T} X_t$. But what is m in concrete situations? From the Jensen inequality we know that

$$m \leqslant \mathbb{E} \sup_{t \in T} X_t,$$

which implies using Dudley's theorem that

$$\mathbb{P}\Big(\sup_{t \in T} X_t \geqslant r\Big) \leqslant 1 - \Phi\Big(\frac{r - m}{\sigma}\Big) \leqslant 1 - \Phi\Big(\frac{r - 4\sqrt{2}\, \Psi\big(\frac{\sigma}{2}\big)}{\sigma}\Big).$$

Let's look at some application of the capacity $C(X, \varepsilon)$ of a centered Gaussian process X. The Sudakov minoration

$$\mathbb{E} \sup_{t \in T} X_t \geqslant c \sqrt{C(X, \varepsilon)}\, \varepsilon \quad \text{for all } \varepsilon > 0$$

is the counterpart to Dudley's theorem. Note that

$$\mathbb{E} \sup_{t \in T} X_t = \infty \quad \Longrightarrow \quad \sup_{t \in T} X_t = \infty \quad \text{a.s.}$$

due to the concentration principle 5.4 (this implication is in general not true).

Just to mention two other results,

$$\begin{aligned} X \text{ path-wise bounded} &\implies C(X, \varepsilon) \leqslant \frac{c}{\varepsilon^2} \quad \text{and} \\ X \text{ continuous} &\implies C(X, \varepsilon) = o\left(\frac{1}{\varepsilon^2}\right). \end{aligned} \tag{8.2}$$

8.3 Metric Entropy of an Operator

Let X be a centered Gaussian on $(\mathcal{X}, \|\cdot\|)$. We are now interested in properties of $\|X\|$. Using

$$B' = \{f \in \mathcal{X}' \colon \|f\| \leqslant 1\}$$

and

$$\|x\| = \sup_{f \in B'} (f, x)$$

for $x \in \mathcal{X}$, we can rewrite $\|X\|$ as

$$\|X\| = \sup_{f \in B'} (f, X) = \sup_{f \in B'} Y(f)$$

where $Y(f) = (f, X)$ is a Gaussian process on B'.

Therefore, we are interested in the metric entropy $\mathbb{H}(Y, \varepsilon)$ of the new process Y. This was defined as the metric entropy of (B', ρ) where

$$\rho(f, g)^2 = \mathbb{E}((f, X) - (g, X))^2 = \mathbb{E}(f - g, X)^2 = \|I'(f - g)\|_{\mathcal{X}_P'}^2.$$

The last equality gives rise to the definition of the metric entropy for an operator.

Let $L \colon \mathcal{X} \to \mathcal{Y}$ be an operator between two normed spaces $\mathcal{X}$ and $\mathcal{Y}$. We define the metric entropy $\mathbb{H}(L, \varepsilon)$ of the operator L to be the metric entropy $\mathbb{H}(B, \varepsilon)$ of the metric space

$$B = \{Lx \colon x \in \mathcal{X}, \|x\| \leqslant 1\} \subset \mathcal{Y}.$$

For compact operators L this is well defined.

With this definition it is easy to see that the metric entropy $\mathbb{H}(Y, \varepsilon)$ is just the metric entropy $\mathbb{H}(I', \varepsilon)$ of the operator I'. When studying properties of $\|X\|$ one is thus interested in the metric entropy $\mathbb{H}(I', \varepsilon)$.

It is not too surprising that the metric entropy $\mathbb{H}(I, \varepsilon)$ is of interest as well. This entropy is the metric entropy of $(\mathcal{B}_1, \|\cdot\|_{\mathcal{X}})$ where $\mathcal{B}_1$ is the compact unit ball of the kernel H_P

$$\mathcal{B}_1 = \{h \in H_P \colon \|h\|_{H_P} \leqslant 1\} \subset \mathcal{X}.$$

Whenever we have an operator $L: \mathcal{X} \to \mathcal{Y}$ and its adjoint operator $L': \mathcal{Y}' \to \mathcal{X}'$ there seems to be a relation between their metric entropy. It is conjectured that

$$N(L, \varepsilon) \approx N(L', \varepsilon) \quad \text{for } \varepsilon \to 0,$$

i.e. the metric entropys of L and L' are of the same order.

A result due to J. Kuelbs and W. Li shows that small ball probabilities of a Gaussian measure P are connected to the metric entropy $\mathbb{H}(I, \varepsilon)$.

Theorem 8.6. (J. Kuelbs, W. Li) *Let X be a centered Gaussian on $(\mathcal{X}, \|\cdot\|)$. For $\alpha \in [0, 2)$*

$$\mathbb{H}(I, \varepsilon) \approx \varepsilon^{-\alpha} \quad \Longleftrightarrow \quad \log \mathbb{P}(\|X\| < r) \approx -r^{-\frac{2\alpha}{2-\alpha}}.$$

Note that for $\alpha > 2$ the process X would be unbounded (compare equation 8.5). E.g. for the Wiener process $\alpha = 1$.

Proof. Let's proof " $\Longrightarrow$ ". So we assume that there are c_1, $c_2 > 0$ such that for $\varepsilon > 0$ small enough

$$c_1 \varepsilon^{-\alpha} \leqslant \log N(I, \varepsilon) \leqslant c_2 \varepsilon^{-\alpha}.$$

We will now try to show upper and lower bounds for $\log \mathbb{P}(\|X\| < r)$ for r small enough.

Upper Bound. Using $M(I, \varepsilon) \geqslant N(I, \varepsilon)$ from 8.1 we get

$$\log N(I, \varepsilon) \geqslant c_1 \varepsilon^{-\alpha} \quad \Longrightarrow \quad \log M(I, \varepsilon) \geqslant c_1 \varepsilon^{-\alpha}$$

and the left side is assumed. Recall that $\mathbb{H}(I, \varepsilon)$ is the metric entropy of the space $(\mathcal{E}_1, \|\cdot\|_{\mathcal{X}})$ with $\mathcal{E}_1$ being the closed unit ball of the kernel H. The definition of $M_\varepsilon = M(I, \varepsilon)$ gives us points

$$x_1, \ldots, x_{M_\varepsilon} \in \mathcal{E}_1 \quad \text{such that each} \quad \|x_k - x_j\| > \varepsilon.$$

Thus for $\lambda > 0$

$$\lambda x_1, \ldots, \lambda x_{M_\varepsilon} \in \mathcal{E}_\lambda \quad \text{such that each} \quad \|\lambda x_k - \lambda x_j\| > \lambda \varepsilon.$$

Let B denote the unit ball of $\mathcal{X}$. Therefore the balls

$$\lambda x_j + \frac{\lambda \varepsilon}{2} B, \quad j = 1, \ldots, M_\varepsilon$$

are disjoint which implies

$$\sum_{j=1}^{M_\varepsilon} P\left(\lambda x_j + \frac{\lambda \varepsilon}{2} B\right) \leqslant 1.$$

Applying Borell's shift inequality 4.4, we get

$$\sum_{j=1}^{M_\varepsilon} P\left(\frac{\lambda \varepsilon}{2} B\right) \exp\left(-\frac{\lambda^2 \|x_j\|_H^2}{2}\right) \leqslant 1,$$

and since $\|x_j\|_H \leqslant 1$

$$M_\varepsilon P\Big(\frac{\lambda\,\varepsilon}{2}B\Big) \leqslant \exp\Big(\frac{\lambda^2}{2}\Big).$$

Setting $r = \frac{\lambda\,\varepsilon}{2}$ this rewrites to

$$P(r\,B) \leqslant \frac{1}{M_\varepsilon}\exp\Big(\frac{2\,r^2}{\varepsilon^2}\Big).$$

Using our assumption we have for any representation $\varepsilon = \delta\,r^u$ and $\delta, u > 0$

$$\log P(r\,B) \leqslant \frac{2\,r^2}{\varepsilon^2} - c_1\,\varepsilon^{-\alpha} = \frac{2}{\delta^2}\,r^{2-2u} - c_1\,\delta^{-\alpha}\,r^{-\alpha\,u}.$$

It is straightforward to choose u so that $2-2u = -\,\alpha\,u$, i.e. $u = \frac{2}{2-\alpha}$, yielding

$$\log P(r\,B) \leqslant -\,K\,r^{-\frac{2\,\alpha}{2-\alpha}}$$

for $K = c_1\,\delta^{-\alpha} - 2\,\delta^{-2}$. Choosing $\delta > 0$ appropriately such that $K > 0$ yields the asserted upper bound.

Lower Bound. Let $\varepsilon > 0$ and set $N_\varepsilon = N'(I,\varepsilon)$. We can cover the unit ball $\mathcal{E}_1$ of the kernel with N_ε balls of radius ε. Again, B denotes the unit ball of $\mathcal{X}$. For $\lambda > 0$

$$\begin{aligned} \mathcal{E}_1 &\subset \bigcup_{j=1}^{N_\varepsilon} (x_j + \varepsilon\,B) \\ \Longrightarrow\quad \lambda\,\mathcal{E}_1 &\subset \bigcup_{j=1}^{N_\varepsilon} (\lambda\,x_j + \lambda\,\varepsilon\,B) \\ \Longrightarrow\quad \lambda\,\mathcal{E}_1 + \lambda\,\varepsilon\,B &\subset \bigcup_{j=1}^{N_\varepsilon} (\lambda\,x_j + 2\,\lambda\,\varepsilon\,B). \end{aligned}$$

Using the isoperimetric inequality 5.2 and the Anderson inequality 7.3 yields

$$\Phi\big(\lambda + \Phi^{-1}(P(\lambda\,\varepsilon\,B))\big) \leqslant P(\lambda\,\mathcal{E}_1 + \lambda\,\varepsilon\,B) \leqslant N_\varepsilon\,P(2\,\lambda\,\varepsilon\,B)$$

for any $\varepsilon > 0$. Replacing ε by $\frac{\varepsilon}{\lambda}$ this implies that for all $\varepsilon > 0$

$$\Phi\big(\lambda + \Phi^{-1}(P(\varepsilon\,B))\big) \leqslant N_{\varepsilon/\lambda}\,P(2\,\varepsilon\,B).$$

We want to choose $\lambda > 0$ so that the left side is $\geqslant \frac{1}{2}$. This is e.g. established for

$$\lambda = \sqrt{2\,|\log P(\varepsilon\,B)|}$$

since

$$\begin{aligned} \Phi\big(\lambda + \Phi^{-1}(P(\varepsilon\,B))\big) &\geqslant \frac{1}{2} \\ \Longleftrightarrow\quad \Phi^{-1}(P(\varepsilon\,B)) &\geqslant -\lambda \\ \Longleftrightarrow\quad \Phi(-\lambda) &\leqslant P(\varepsilon\,B) \end{aligned}$$

and

$$\Phi(-\lambda) \leqslant \exp\Big(-\frac{\lambda^2}{2}\Big).$$

So far we reached

$$\frac{1}{2} \leqslant N_{\varepsilon/\lambda}\, P(2\,\varepsilon\, B)$$

for all $\varepsilon > 0$. Thus

$$-\log 2 \leqslant \log N_{\varepsilon/\lambda} + \log P(2\,\varepsilon\, B),$$

and using $N_\varepsilon = N'(I,\varepsilon) \leqslant N\big(I, \frac{\varepsilon}{2}\big)$ from 8.1 we get for $c = 2^\alpha\, c_2$ and ε small enough

$$c\,(\varepsilon/\lambda)^{-\alpha} \geqslant \log N_{\varepsilon/\lambda} \geqslant -\log 2 + |\log P(2\,\varepsilon\, B)|$$

since we assumed $\log N(I,\varepsilon) \leqslant c_2\, \varepsilon^{-\alpha}$. Replacing ε by $\frac{\varepsilon}{2}$, this yields

$$c\,\varepsilon^{-\alpha}\Big(2\,\Big|\log P\Big(\frac{\varepsilon}{2}\,B\Big)\Big|\Big)^{\alpha/2} \geqslant -\log 2 + |\log P(\varepsilon\, B)|.$$

We now set $\varphi(\varepsilon) = |\log P(\varepsilon\, B)|$ to get

$$\varphi(\varepsilon) \leqslant \log 2 + \tilde{c}\,\varepsilon^{-\alpha}\,\varphi\Big(\frac{\varepsilon}{2}\Big)^{\frac{\alpha}{2}}$$

for a new constant $\tilde{c} > 0$.

We have to show that there is some constant $K > 0$ such that for ε small enough

$$\log P(\varepsilon\, B) \geqslant -K\,\varepsilon^{-\frac{2\alpha}{2-\alpha}}$$

or, equivalently, that

$$\varepsilon^{\frac{2\alpha}{2-\alpha}}\,\varphi(\varepsilon) = \varepsilon^{\frac{2\alpha}{2-\alpha}}\,|\log P(\varepsilon\, B)| \leqslant K.$$

We already have

$$\varepsilon^{\frac{2\alpha}{2-\alpha}}\,\varphi(\varepsilon) \leqslant \varepsilon^{\frac{2\alpha}{2-\alpha}}\log 2 + \tilde{c}\,\varepsilon^{\frac{\alpha^2}{2-\alpha}}\,\varphi\Big(\frac{\varepsilon}{2}\Big)^{\frac{\alpha}{2}}$$

and it is left as an exercise to complete the argument.

□

Theorem 8.7. (Talagrand's Lower Bound) *Let X be a centered Gaussian process on T. If*

$$N(X,\varepsilon) \leqslant c\,\varepsilon^{-\beta}$$

for some constant c and $\beta > 0$ then there is a constant $K > 0$ such that

$$\mathbb{P}\Big(\sup_{s,t\in T} |X_t - X_s| \leqslant \varepsilon\Big) \geqslant \exp\big(-K\,\varepsilon^{-\beta}\big).$$

A slightly more general result is proved in [Ledoux, 1996] using Šidak's inequality (correlation inequality for strips).

9 Expansions

In this section we will deal with two closely related problems.

1. **Problem.** Given a linear space $\mathcal{X}$ and a Gaussian measure $P = N(0, K)$ on $\mathcal{X}$. Construct

$$X = \sum_{j=1}^{\infty} \xi_j x_j$$

where ξ_j are independent Gaussian random variables and $x_j \in \mathcal{X}$ such that the sum converges and P is the law of X.

2. **Problem.** Given a linear space $\mathcal{X}$ and a Gaussian vector X on $\mathcal{X}$. Construct a representation

$$X = \sum_{j=1}^{\infty} \xi_j x_j \tag{9.1}$$

of X where ξ_j are Gaussian random variables and $x_j \in \mathcal{X}$. 9.1 is called *expansion* of X.

Note that the second problem is a bit harder than the first one, because we have to construct a sum that reproduces X almost surely. These problems are especially important for modelling a Gaussian vector X.

9.1 General Series of Independent Vectors

Let $(\mathcal{X}, \|\cdot\|)$ be a normed linear space and X_j random vectors taking values in $\mathcal{X}$. Then

$$S_n = \sum_{j=1}^{n} X_j$$

is a random vector on $\mathcal{X}$, too. But what exactly does

$$S = \sum_{j=1}^{\infty} X_j$$

mean? What type of convergence can we expect? In general there are three important basic types of convergence.

(a) Convergence in distribution, i.e. weak convergence of the laws of S_n.

(b) Convergence in probability, meaning that for some S and all $\varepsilon > 0$

$$\lim_{n \to \infty} \mathbb{P}(\|S_n - S\| > \varepsilon) = 0.$$

(c) Almost sure convergence.

Nevertheless these three types of convergence agree in a special case.

Theorem 9.1. *Let X_j be symmetric independent random vectors in $(\mathcal{X}, \|\cdot\|)$ such that their partial sums S_n converge in distribution. Then this convergence is almost sure convergence.*

The proof is not easy and thus omitted.

Now let's turn towards the solution of our first problem. Unsurprisingly, the kernel of the Gaussian measure P plays a key role again.

Theorem 9.2. *Let $P = N(0, K)$ be a Gaussian measure on a linear space $\mathcal{X}$ and $\{h_j : j \in \mathbb{N}\}$ an orthonormal basis of the kernel H. Then for $\xi_j \sim N(0,1)$ i.i.d. the law of*

$$X = \sum_{j=1}^{\infty} \xi_j h_j$$

is P.

Proof. We have to check that the laws of the partial sums converge weakly to P. We do this by looking at the characteristic functions. We show that for $f \in \mathcal{X}'$

$$\mathbb{E} \exp\left(i \sum_{j=1}^{n} \xi_j (f, h_j) \right) \longrightarrow \int_{\mathcal{X}} \exp(i (f, x)) \, P(\mathrm{d}x).$$

For fixed f both sides may be thought to arise from Gaussian random variables, and since their expectations vanish we have to check only their variances. For the right side we get

$$\int_{\mathcal{X}} (f, x)^2 \, P(\mathrm{d}x) = \| I' f \|^2_{\mathcal{X}'_P},$$

and for the left one

$$\begin{aligned} \operatorname{Var}(f, S_n) &= \operatorname{Var}\left(\sum_{j=1}^{n} \xi_j (f, h_j) \right) = \sum_{j=1}^{n} (f, h_j)^2 = \sum_{j=1}^{n} (f, I z_j)^2 \\ &= \sum_{j=1}^{n} (I' f, z_j)^2_{\mathcal{X}'_P} \quad \longrightarrow \quad \| I' f \|^2_{\mathcal{X}'_P}, \end{aligned}$$

which was to be shown. The second last term is a sum of Fourier coefficients w.r.t. the basis $\{z_j\}$. □

Now we are ready to turn to the second problem. Given X, we have to construct an expansion $X = \sum \xi_j x_j$.

Theorem 9.3. *Let X be a centered Gaussian random vector on $\mathcal{X}$ and H be the kernel of the law of X. Choose an orthonormal basis $\{z_j : j \in \mathbb{N}\}$ of $\mathcal{X}'_P$. Then $\{h_j : j \in \mathbb{N}\}$, where $h_j = I z_j$, is an orthonormal basis of H.*

$$Y = \sum_{j=1}^{\infty} z_j(X) \, h_j$$

is well defined and $X = Y$ almost surely.

Proof. $z_j(X)$ is well defined almost surely. We will to show $X = Y$ a.s. or, equivalently,

$$\begin{aligned} \mathbb{P}_{Y-X} &= \delta_0 \quad \text{Dirac measure in } 0 \\ \mathbb{E} \exp(i (f, Y - X)) &= 1 \quad \text{for all } f \in \mathcal{X}' \end{aligned}$$

by checking characteristic functions. It is sufficient to show that almost surely $(f, Y - X) = 0$, or $(f, X) = (f, Y)$, for all $f \in \mathcal{X}'$. By definition we have almost surely

$$(f, Y) = \lim_{n \to \infty} \left(f, \sum_{j=1}^{n} z_j(X) \, h_j \right) = \sum_{j=1}^{\infty} z_j(X) \, (f, h_j) = \sum_{j=1}^{\infty} z_j(X) \, (I' f, z_j).$$

On the other hand we have $L^2(\mathcal{X}, P)$ convergence for $n \to \infty$ of

$$\sum_{j=1}^{n} z_j(\,\cdot\,)\,(I'\,f, z_j) \longrightarrow I'\,f(\,\cdot\,)$$

since $\mathcal{X}_P'$ is a Hilbert space with orthonormal basis $\{z_j\}$ and convergence in its metric means L^2 convergence. Because $I'\,f(X) = (f, X)$, we get

$$(f, Y) = \sum_{j=1}^{\infty} z_j(X)\,(I'\,f, z_j) = I'\,f(X) = (f, X)$$

almost surely from theorem 9.1. □

Example 9.4. (Brownian Motion) To write down an expansion W for a Brownian motion, all we have to do is to construct an orthonormal basis of the kernel. We achieve this by

- choosing an orthonormal basis $\{z_j : j \in \mathbb{N}\}$ of $L^2[0,1]$,
- which generates via

  $$h_j(t) = \int_0^t z_j(s)\,\mathrm{d}s$$

  an orthonormal basis $\{h_j : j \in \mathbb{N}\}$ of the kernel.
- Now the expansion

  $$W = \sum_{j=1}^{\infty} \xi_j\, h_j,$$

  where $\xi_j \sim N(0,1)$ are i.i.d., defines a Brownian motion.

This way, we will try to get some nice expansions.

(a) Choose the orthonormal basis

$$z_0 \equiv 1, \quad z_n(s) = \sqrt{2}\cos n\,\pi\,s \quad \text{for } n \geqslant 1.$$

of $L^2[0,1]$. Then

$$h_0(t) = t, \quad h_n(t) = \sqrt{2}\,\frac{\sin n\,\pi\,t}{n\,\pi}$$

yielding the expansion

$$W_t = \xi_0\,t + \sqrt{2}\sum_{n=1}^{\infty} \xi_n\,\frac{\sin n\,\pi\,t}{n\,\pi}.$$

In example 3.9 the Brownian bridge $\dot{W}$ was defined by $\dot{W}_t = W_t - t\,W_1$, and therefore

$$\dot{W}_t = \sqrt{2}\sum_{n=1}^{\infty} \xi_n\,\frac{\sin n\,\pi\,t}{n\,\pi}$$

is an expansion that defines a Brownian bridge.

(b) Another orthonormal basis of $L^2[0,1]$ is

$$z_n(s) = \sqrt{2}\sin n\,\pi\,s$$

which implies

$$h_n(t) = \sqrt{2}\,\frac{1-\cos n\,\pi\,t}{n\,\pi}$$

providing the expansion

$$W_t = \sqrt{2}\sum_{n=1}^{\infty}\xi_n\,\frac{1-\cos n\,\pi\,t}{n\,\pi}.$$

(c) The following expansion is known as Karhunen-Loève expansion. Starting with

$$z_n(s) = \sqrt{2}\cos\Big(\Big(n-\frac{1}{2}\Big)\,\pi\,s\Big),$$

we get

$$h_n(t) = \sqrt{2}\,\frac{\sin\Big(\Big(n-\frac{1}{2}\Big)\,\pi\,s\Big)}{\Big(n-\frac{1}{2}\Big)\,\pi}$$

which are also orthogonal in $L^2[0,1]$. The expansion

$$W_t = \sqrt{2}\sum_{n=1}^{\infty}\xi_n\,\frac{\sin\Big(\Big(n-\frac{1}{2}\Big)\,\pi\,s\Big)}{\Big(n-\frac{1}{2}\Big)\,\pi}$$

is based on an eigenbasis for W in $L^2[0,1]$ instead of a basis of $C[0,1]$. You can find a quick deduction of this eigenbasis e.g. in [Vanden-Eijnden, 2003, 4.4].

(d) For the Haar-Schauder expansion we will use the Haar basis of $L^2[0,1]$.

$$\Psi(s) = \begin{cases} 1 & \text{for } s\in\Big[0,\frac{1}{2}\Big) \\ -1 & \text{for } s\in\Big(\frac{1}{2},1\Big] \\ 0 & \text{otherwise} \end{cases}$$

defines the so called *Haar function* which is shown in figure 9.1.

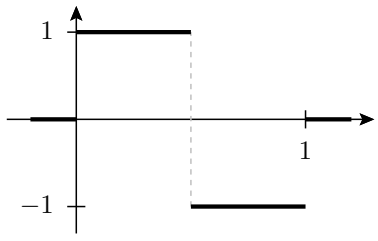

Figure 9.1. Haar function

The Haar basis of $L^2[0,1]$ is then given by

$$z_o(s)\equiv 1,\quad z_{j,k}(s) = 2^{\frac{j}{2}}\,\Psi\big(2^j\,s-k\big)\quad \text{for } j=0,1,\ldots \text{ and } k=0,\ldots,2^j-1,$$

see figure 9.2.

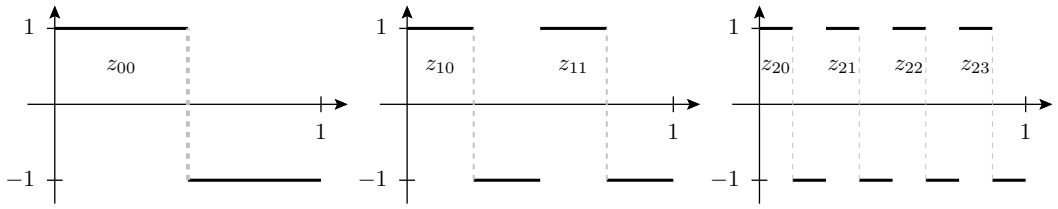

Figure 9.2. Haar basis

Note that we get a basis of $L^2(\mathbb{R})$ if we consider $z_{j,k}$ for $j, k \in \mathbb{Z}$. Nevertheless, this provides the Schauder basis

$$h_o(t) = t, \quad h_{j,k}(t) = 2^{-\frac{j}{2}} \tilde{h}\big(2^j t - k\big) \quad \text{for } j = 0, 1, \ldots \text{ and } k = 0, \ldots, 2^j - 1$$

where $\tilde{h}$ is a roof on $[0,1]$ defined by

$$\tilde{h}(t) = \begin{cases} t & \text{for } t \in \left[0, \frac{1}{2}\right] \\ 1 - t & \text{for } t \in \left(\frac{1}{2}, 1\right] \\ 0 & \text{otherwise.} \end{cases}$$

The appendant expansion is

$$W_t = \xi_0 t + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j - 1} \xi_{j,k} h_{j,k}(t)$$

which again contains an expansion for the Brownian bridge, too.

Example 9.5. (Complex Brownian Motion) As another example we will take a short look at the complex Brownian motion which arises as a Gaussian W on $\mathcal{X} = C([0,1], \mathbb{C})$ whose real and imaginary parts are independent real Brownian motions $W^{(1)}$ and $W^{(2)}$

$$W_t = W_t^{(1)} + i\, W_t^{(2)}.$$

It is not surprising – and left as an exercise to show – that everything works as in the real case. Instead of $L^2[0,1]$ we now have to deal with complex valued square integrable functions and so on. An orthonormal basis of $L^2([0,1], \mathbb{C})$ is

$$z_n(s) = \exp(2\, n\, \pi\, i\, s) \quad \text{for } n \in \mathbb{Z}.$$

This way we get $h_0(t) = t$ and

$$h_n(t) = \frac{\exp(2\, n\, \pi\, i\, t) - 1}{2\, n\, \pi\, i}, \quad n \neq 0,$$

as an orthonormal basis of the kernel and the expansion

$$W_t = \sum_{n=-\infty}^{\infty} \xi_n \frac{\exp(2\, n\, \pi\, i\, t) - 1}{2\, n\, \pi\, i}.$$

It is maybe worth to mention that this expansion can be extended to the case of the complex fractional Brownian motion $W^{(\alpha)}$ which again is the sum

$$W_t^{(\alpha)} = W_t^{(\alpha,1)} + i\, W_t^{(\alpha,2)}$$

of two independent real fractional Brownian motions. Due to a result of K. Dzhaparidze and H. van Zanten

$$W_t^{(\alpha)} = \sum_{n=-\infty}^{\infty} \xi_n \frac{\exp(2\omega_n i t) - 1}{2\omega_n i} \sigma_n$$

is an expansion. Here ω_n are real zeros of the Bessel functions $J_{1-\frac{\alpha}{2}}$. In particular, for $\alpha = 1$ we have

$$J_{\frac{1}{2}}(z) = \sqrt{\frac{2\pi}{z}} \sin z \quad \text{with zeros } \pi n \text{ for } n \in \mathbb{Z}$$

which is just what we found in the case of the Brownian motion. σ_n are constants which behave like

$$\frac{\sigma_n}{\omega_n} \sim c_\alpha n^{-\frac{\alpha}{2}} \quad \text{for } n \to \infty$$

for a constant c_α.

9.2 Linear Operators and Gaussian Vectors

Given an operator

$$L: \mathcal{H} \to \mathcal{X} \quad \text{for } \mathcal{H} \text{ Hilbert and } \mathcal{X} \text{ linear}$$

we will try to construct a Gaussian vector X on $\mathcal{X}$ corresponding to the operator L.

Let $\{l_j : j \in \mathbb{N}\}$ be an orthonormal basis of $\mathcal{H}$ and ξ_j i.i.d. standard normal random variables. We try to define X using the formal expansion

$$X = \sum_{j=1}^{\infty} \xi_j L l_j,$$

and say that X was generated by L which surely depends on the basis chosen in $\mathcal{H}$.

Theorem 9.6. *The following is true.*

(i) *Convergence of the series used to define X happens either with probability 0 or 1.*

(ii) *Convergence does not depend on the basis $\{l_j\}$.*

(iii) *In the case of convergence we have*

$$K = L L'$$

where K is the covariance operator of X. Consequently, the law P of X does not depend on the choice of $\{l_j\}$.

(iv) *Furthermore, in the case of convergence the kernel H of P is given by*

$$H = L\mathcal{H} \quad \text{and} \quad B_1 = \{h \in H : \|h\| \leqslant 1\} = L\{l \in \mathcal{H} : \|l\| \leqslant 1\}.$$

The third point of theorem 9.6 states that the linear operator L actually generates a Gaussian distribution P.

This enables us to define a new norm for the operator L via

$$\|L\|_l = \sqrt{\mathbb{E}\,||X||_{\mathcal{X}}^2} \asymp \sqrt[p]{\mathbb{E}\,||X||_{\mathcal{X}}^p}.$$

Finally, we define the so called l-numbers $l_n(L)$ which serve as a measure of approximability by finite rank operators

$$\begin{aligned} l_n(L) &= \inf\{\|L-F\|_l\colon \operatorname{rank}(F)<n\} \\ &\asymp \inf\left\{\sqrt{\mathbb{E}\left\|X-\sum_{j=1}^{n-1}\xi_j x_j\right\|^2}\colon x_j\in\mathcal{X}, \xi_j \text{ normal}\right\}. \end{aligned}$$

It follows from the Anderson inequality 7.3 that the last infimum does not change if it is just taken over independent random variables ξ_j.

10 Strassen's Law

10.1 Scalar Laws of Iterated Logarithm

The law of the iterated logarithm comes in two main versions:

- for sums of independent random vectors,
- for the Wiener process.

Let X_j be i.i.d. standardized random vectors, i.e. $\mathbb{E}\,X_j=0$ and $\mathbb{E}\,X_j^2=1$. Define

$$S_n=\sum_{j=1}^{n} X_j.$$

Then the Hartman-Wintner law states that almost surely

$$\begin{aligned} \limsup_{n\to\infty}\frac{S_n}{\sqrt{2\,n\log\log n}} &= 1 \\ \liminf_{n\to\infty}\frac{S_n}{\sqrt{2\,n\log\log n}} &= -1. \end{aligned}$$

The analogous result for a Wiener process W on $[0,\infty)$ due to Lévy reads

$$\begin{aligned} \limsup_{t\to\infty}\frac{W_t}{\sqrt{2\,t\log\log t}} &= 1 \\ \liminf_{t\to\infty}\frac{W_t}{\sqrt{2\,t\log\log t}} &= -1 \end{aligned} \tag{10.1}$$

10.2 Functional Law

We will now extend the scalar law for the Wiener process to the so called functional law which looks at whole trajectories $(W_s)_{s\in[0,T]}$. Let W be a Brownian motion on $[0,\infty)$. We define a new, properly scaled, Brownian motion

$$W_s^{(T)}=\frac{W_{Ts}}{\sqrt{T}}$$

for $T>0$. Eventually, T will tend to infinity. In the sequel we look at $W^{(T)}$ as a Brownian motion on $[0,1]$. Then $W^{(T)}$ contains all information of $(W_s)_{s\in[0,T]}$.

Similarly to the scalar law, we set

$$Y_s^{(T)} = \frac{W_s^{(T)}}{\sqrt{2\log\log T}} = \frac{W_{Ts}}{\sqrt{2\,T\log\log T}}.$$

The scalar law of the iterated logarithm states that

$$\operatorname*{limsup}_{T\to\infty} Y_1^{(T)} = 1 \quad \text{and} \quad \operatorname*{liminf}_{T\to\infty} Y_1^{(T)} = -1.$$

We have to define the notion of convergence to a set.

Definition 10.1. *Let $x = (x_t)_{t\geqslant 0}$ be a sequence in a metric linear space $\mathcal{X}$. x is said to converge to the set $A \subset \mathcal{X}$, written as*

$$x_t \hookrightarrow A \quad \text{for } t\to\infty,$$

iff A is the set of all limit points of x, i.e.

$$\begin{aligned} \lim_{t\to\infty} d(x_t, A) &= 0 \quad \text{(there are no other limit points)} \\ (\forall a \in A) \quad \operatorname*{liminf}_{t\to\infty} d(x_t, a) &= 0 \quad \text{(all points are involved).} \end{aligned}$$

The functional law states that $Y^{(T)}$ converges to the unit ball of the kernel when $T\to\infty$.

Theorem 10.2. (Strassen's Functional Law of the Iterated Logarithm) *Let $Y^{(T)}$ be defined as above and H the kernel of the associated Brownian motion. For $T\to\infty$, a.s.*

$$Y^{(T)} \hookrightarrow \mathcal{E}_1 := \{h\in H\colon\ \|h\|_H \leqslant 1\} \subset C[0,1].$$

More verbosely,

$$\left.\frac{W_{Ts}}{\sqrt{2\,T\log\log T}}\right|_{s\in[0,1]} \hookrightarrow \mathcal{E}_1 \quad \text{for } T\to\infty.$$

The following corollary is an immediate but pretty nice consequence.

Corollary 10.3. *Let $\mathcal{F}\colon C[0,1]\to\mathbb{R}$ be continuous. Then, a.s.*

$$\operatorname*{limsup}_{T\to\infty} \mathcal{F}\Big(Y^{(T)}\Big) = \sup_{h\in\mathcal{E}_1} \mathcal{F}(h)$$

Example 10.4. Take e.g. $\mathcal{F}(x) = x(1)$ to get Lévy's scalar law of the iterated logarithm.

Other interesting choices are

$$\mathcal{F}(x) = \sup_{t\in[0,1]} x(t), \quad \inf_{t\in[0,1]} x(t) \quad \text{or} \quad \int_0^1 |x(t)|^p\,\mathrm{d}t$$

yielding e.g.

$$\operatorname*{limsup}_{T\to\infty} \frac{\sup_{t\in[0,T]} W_t}{\sqrt{2\,T\log\log T}} = \operatorname*{limsup}_{T\to\infty} \sup_{t\in[0,1]} Y_t^{(T)} = \sup_{h\in\mathcal{E}_1} \sup_{t\in[0,1]} h(t) = 1, \tag{10.2}$$

since the supremum on the right is attained for $h(s)=s$.

Note that 10.2 is a remarkable extension of Lévy's law 10.1 where W_t was considered instead of $\sup_{t\in[0,T]} W_t$. This demonstrates what was mentioned as "looking at whole trajectories" in the introduction.

10.3 Proof of Strassen's Law

Proof. Let $\gamma>1$. We will decompose $Y^{(T)}$ as

$$Y^{(T)}=\hat{Y}^{(T)}+\dot{Y}^{(T)}$$

depending on γ. Define $\hat{Y}^{(T)}$ by stopping $Y^{(T)}$ at the time $\frac{1}{\gamma}$, i.e.

$$\hat{Y}_s^{(T)}=Y_{s\wedge\frac{1}{\gamma}}^{(T)}.$$

$\hat{Y}^{(T)}$ depends on W_t for $t\in\left[0,\frac{T}{\gamma}\right]$ only, and is therefore independent from $\dot{Y}^{(T)}$ which depends on $W_t-W_{T/\gamma}$ for $t\in\left[\frac{T}{\gamma},T\right]$ only since

$$\dot{Y}_s^{(T)}=Y_s^{(T)}-\hat{Y}_s^{(T)}=\begin{cases}0 & s\leqslant\frac{1}{\gamma}\\ Y_s^{(T)}-Y_{\frac{1}{\gamma}}^{(T)} & s\in\left[\frac{1}{\gamma},1\right]\end{cases}.$$

Set $T_n=\gamma^n$. d is used to denote the metric induced by the sup-norm used in $C[0,1]$. Then the following is claimed to be true.

First Claim. For all $\gamma>1$ and for all $\varepsilon>0$

$$\sum_n \mathbb{P}\Big(d\Big(Y^{(T_n)},\mathcal{E}_1\Big)>\varepsilon\Big)<\infty.$$

Second Claim. When $\gamma\downarrow 1$, we find $\varepsilon_1(\gamma)\to 0$ such that

$$\sum_n \mathbb{P}\left(\sup_{T\in[T_n,T_{n+1}]} d\Big(Y^{(T)},Y^{(T_{n+1})}\Big)>\varepsilon_1(\gamma)\right)<\infty.$$

Third Claim. Let $h\in H$ be in the kernel and $\|h\|_H<1$. Then for all $\varepsilon>0$

$$\sum_n \mathbb{P}\Big(\Big\|Y^{(T_n)}-h\Big\|<\varepsilon\Big)=\infty.$$

Note that the $Y^{(T_n)}$ are dependent so we can't use Borel-Cantelli directly.

Fourth Claim. For $\gamma\to\infty$ we find $\varepsilon_2(\gamma)$ such that $\lim_{\gamma\to\infty}\varepsilon_2(\gamma)=0$ and

$$\sum_n \mathbb{P}\Big(\Big\|\hat{Y}^{(T_n)}\Big\|>\varepsilon_2(\gamma)\Big)<\infty.$$

These claims suffice to prove Strassen's law since the first two claims say that (a) there are no other limit points of $Y^{(T)}$ while the last two state that (b) all points of $\mathcal{E}_1$ are actually limit points. We can see these implications like this:

(a) Using Borel-Cantelli, the first claim provides

$$\limsup_{n\to\infty} d\Big(Y^{(T_n)}, \mathcal{E}_1\Big) = 0 \quad \text{a.s.}$$

for all $\gamma > 0$, while the second one yields

$$\limsup_{n\to\infty} \sup_{T\in[T_n,T_{n+1}]} d\Big(Y^{(T)}, Y^{(T_{n+1})}\Big) \leqslant \varepsilon_1(\gamma) \quad \text{a.s.}$$

Due to the triangular inequality, for every $n \in \mathbb{N}$

$$d\Big(Y^{(T)}, \mathcal{E}_1\Big) \leqslant d\Big(Y^{(T)}, Y^{(T_{n+1})}\Big) + d\Big(Y^{(T_{n+1})}, \mathcal{E}_1\Big)$$

and therefore

$$\limsup_{T\to\infty} d\Big(Y^{(T)}, \mathcal{E}_1\Big) \leqslant \varepsilon_1(\gamma) \quad \text{a.s.}$$

It remains to let $\gamma \downarrow 1$ and use $\varepsilon_1(\gamma) \to 0$.

(b) We have to show that all points of $\mathcal{E}_1$ are limit points, i.e. for all $h \in \mathcal{E}_1$

$$\liminf_{T\to\infty} \Big\|Y^{(T)} - h\Big\| = 0 \quad \text{a.s.}$$

It suffices to show this for h such that $\|h\|_H < 1$ (using approximation and standard diagonal arguments). Let's fix such an h. Combining the last two claims we get

$$\sum_n \mathbb{P}\Big(\Big\|Y^{(T_n)} - h\Big\| < \varepsilon, \Big\|\hat{Y}^{(T_n)}\Big\| \leqslant \varepsilon_2(\gamma)\Big) = \infty.$$

Since $\varepsilon > 0$ was arbitrary, we can in particular choose $\varepsilon = \varepsilon_2(\gamma)$ yielding

$$\Big\|\dot{Y}^{(T_n)} - h\Big\| \leqslant \Big\|\dot{Y}^{(T_n)} - Y^{(T_n)}\Big\| + \Big\|Y^{(T_n)} - h\Big\| \leqslant \varepsilon_2(\gamma) + \varepsilon \leqslant 2\,\varepsilon_2(\gamma).$$

Therefore

$$\sum_n \mathbb{P}\Big(\Big\|\dot{Y}^{(T_n)} - h\Big\| \leqslant 2\,\varepsilon_2(\gamma)\Big) = \infty.$$

Remember that $\dot{Y}^{(T)}$ only depends on $W_t - W_{T/\gamma}$ for $t \in \left[\frac{T}{\gamma}, T\right]$ and that $T_n = \gamma^n$. So the $\dot{Y}^{(T_n)}$ are independent which allows us to apply Borel-Cantelli to get a.s.

$$\liminf_{n\to\infty} \Big\|\dot{Y}^{(T_n)} - h\Big\| \leqslant 2\,\varepsilon_2(\gamma)$$

On the other hand, we get directly from the last claim that a.s.

$$\limsup_{n\to\infty} \Big\|\hat{Y}^{(T_n)}\Big\| \leqslant \varepsilon_2(\gamma)$$

Since

$$\left\| Y^{(T_n)} - h \right\| = \left\| \hat{Y}^{(T_n)} + \dot{Y}^{(T_n)} - h \right\| \leqslant \left\| \dot{Y}^{(T_n)} - h \right\| + \left\| \hat{Y}^{(T_n)} \right\|$$

combining these two yields

$$\liminf_{n\to\infty} \left\| Y^{(T_n)} - h \right\| \leqslant 3\,\varepsilon_2(\gamma) \quad \text{a.s.}$$

It remains to let $\gamma \to \infty$ which means $\varepsilon_2(\gamma) \to 0$.

We will now prove the first and third claim. The second and fourth claims can be proved using similar techniques as were used for the first one.

Proof of First Claim. The first claim states that for all $\gamma > 0$ and for all $\varepsilon > 0$

$$\sum_n \mathbb{P}\Big(d\Big(Y^{(T_n)}, \mathcal{E}_1 \Big) > \varepsilon \Big) < \infty.$$

Remember that $W^{(T)}$ was a Brownian motion on $[0,\,1]$ for all $T > 0$. Thus for $L_n = 2\log\log T_n$

$$Y^{(T_n)} = \frac{W^{(T_n)}}{\sqrt{2\log\log T_n}} \quad \text{is distributed as is} \quad \frac{W}{\sqrt{L_n}}.$$

Therefore

$$\begin{aligned}
\mathbb{P}\Big(d\Big(Y^{(T_n)}, \mathcal{E}_1 \Big) > \varepsilon \Big) &= \mathbb{P}\Big(d\Big(\frac{W}{\sqrt{L_n}}, \mathcal{E}_1 \Big) > \varepsilon \Big) = \mathbb{P}\big(d\big(W, \mathcal{E}_{\sqrt{L_n}} \big) > \varepsilon \sqrt{L_n} \big) \\
&= \mathbb{P}\big(W \notin \mathcal{E}_{\sqrt{L_n}} + \varepsilon \sqrt{L_n}\, B \big) \quad (B \text{ denotes the unit ball}) \\
&\leqslant 1 - \Phi\big(\Phi^{-1}\big(\mathbb{P}\big(W \in \varepsilon \sqrt{L_n}\, B \big) \big) + \sqrt{L_n} \big) \quad (\text{isop. ineq. } 5.2) \\
&\leqslant 1 - \Phi\big(1 + \sqrt{L_n} \big) \quad \text{for } n \text{ big enough} \\
&\leqslant \exp\Big(-\frac{\big(1+\sqrt{L_n}\big)^2}{2} \Big) \\
&\leqslant \exp\Big(-\frac{L_n}{2} \Big) \exp\big(-\sqrt{L_n} \big).
\end{aligned}$$

We can estimate

$$\exp(-x) \leqslant c\, x^{-4}$$

for all $x > 0$ and a fixed constant $c > 0$. Thus

$$\begin{aligned}
\mathbb{P}\Big(d\Big(Y^{(T_n)}, \mathcal{E}_1 \Big) > \varepsilon \Big) &\leqslant \exp\Big(-\frac{L_n}{2} \Big) \exp\big(-\sqrt{L_n} \big) \\
&\leqslant \exp\Big(-\frac{L_n}{2} \Big) c\, L_n^{-2} \\
&\leqslant \frac{1}{\log T_n}\, c\, L_n^{-2} \quad (L_n = 2\log\log T_n) \\
&= \frac{c}{n\log\gamma} (\log(n\log\gamma))^{-2} \quad (T_n = \gamma^n) \\
&\leqslant \frac{c_1}{n\,(\log n)^2} \quad \text{for a new constant } c_1.
\end{aligned}$$

This was to show since $\sum \frac{1}{n\,(\log n)^2} < \infty$.

Proof of Third Claim. Let $\varepsilon > 0$ and $h \in H$ such that $\|h\|_H < 1$. We have to show that

$$\sum_n \mathbb{P}\Big(\Big\| Y^{(T_n)} - h \Big\| < \varepsilon \Big) = \infty.$$

We use that $Y^{(T_n)}$ is distributed as $\frac{W}{\sqrt{L_n}}$, see the proof of the first claim, to get

$$\begin{aligned}
\mathbb{P}\Big(\Big\| Y^{(T_n)} - h \Big\| < \varepsilon \Big) &= \mathbb{P}\Big(\Big\| \frac{W}{\sqrt{L_n}} - h \Big\| < \varepsilon \Big) = \mathbb{P}\big(\| W - h\sqrt{L_n} \| < \varepsilon \sqrt{L_n} \big) \\
&= \mathbb{P}\big(W \in \varepsilon \sqrt{L_n}\, B + h\sqrt{L_n} \big) \quad (B \text{ again the unit ball}) \\
&\geqslant \mathbb{P}\big(W \in \varepsilon \sqrt{L_n}\, B \big) \exp\Big(-\frac{\|h\|_H^2 L_n}{2} \Big) \quad \text{(Borell shift ineq. 4.4)} \\
&\geqslant \frac{1}{2} \exp\Big(-\frac{\|h\|_H^2 L_n}{2} \Big) \quad \text{for } n \text{ big enough} \\
&\geqslant \frac{1}{2} (\log T_n)^{-\|h\|_H^2} \quad (L_n = 2 \log\log T_n) \\
&\geqslant c\, n^{-\|h\|_H^2} \quad \text{for some constant } c \quad (T_n = \gamma^n).
\end{aligned}$$

This estimate suffices since $\sum n^{-\alpha} = \infty$ for $\alpha < 1$.

□

10.4 Extensions of Strassen's Law

There are several possibilities to extend Strassen's law 10.2. We will just mention some directions.

Other norms. The space $C[0, 1]$ endowed with the sup-norm can be replaced with other normed spaces, e.g. L^p spaces. As long as the norm $\| \cdot \|$ is "reasonable" and a.s.

$$\|W\| < \infty,$$

this norm can be used in Strassen's law as well. Note that the weaker the norm, the stronger the result will be, as we have more continuous functionals then.

Rate of convergence. From Strassen's law we know that a.s.

$$\lim_{T \to \infty} d\Big(Y^{(T)}, \mathcal{E}_1 \Big) = 0.$$

But how does this convergence take place? One is able to show that in fact a.s.

$$d\Big(Y^{(T)}, \mathcal{E}_1 \Big) \asymp \frac{c}{(\log\log T)^{\frac{2}{3}}}$$

for a constant c. Strassen's law further says that for any $h \in \mathcal{E}_1$ a.s.

$$\liminf_{T \to \infty} \Big\| Y^{(T)} - h \Big\| = 0.$$

For h such that $\|h\|_H < 1$ one can show that a.s.

$$\liminf_{T \to \infty} \Big\| Y^{(T)} - h \Big\| \log\log T = \frac{c}{\sqrt{1 - \|h\|_H^2}}$$

for another constant c. Therefore the rate of convergence to inner points of $\mathcal{E}_1$ is

$$\frac{1}{\log\log T},$$

but this term explodes on the surface. Nevertheless the rate of convergence is generally between (every value is actually attained)

$$\frac{1}{\log\log T} \quad \text{and} \quad \frac{1}{(\log\log T)^{\frac{2}{3}}}$$

depending on the smoothness of h (in the spirit of Hölder).

Note that we are talking about large deviations here since roughly speaking

$$Y^{(T)} \approx h \quad \Longleftrightarrow \quad \frac{W}{\sqrt{2\log\log T}} \approx h \quad \Longleftrightarrow \quad W \approx h\sqrt{2\log\log T} \to \infty.$$

Here, $\approx$ means "is approximately".

Multivariate Case. Instead of a Brownian motion W, one might want to consider a vector

$$W = \left(W^{(1)}, \dots W^{(n)}\right)$$

of independent Brownian motions. However, most things won't change. You just have to replace $\mathcal{E}_1$ by the unit ball of the kernel of W which is equipped with the norm

$$\|h\|_H^2 = \int_0^1 \|h'(s)\|_{\mathbb{R}^n}^2 \,\mathrm{d}s.$$

Multiparametric Case. Strassen's law can also be extended to the case of a Brownian sheet W on $[0,\infty)^n$. The scaled version $W^{(T)}$

$$W_s^{(T)} = W_{s_1,\dots,s_n}^{(T)} = \frac{W_{Ts}}{T^{n/2}} = \frac{W_{Ts_1,\dots,Ts_n}}{T^{n/2}}$$

is again a Brownian sheet, and

$$Y^{(T)} = \frac{W^{(T)}}{\sqrt{2\log\log T}}$$

converges to the unit ball of the kernel of Brownian sheet.

Fractional Brownian Motion. If $W^{(\alpha)}$ is a fractional Brownian motion, then

$$W_s^{(T,\alpha)} = \frac{W_{Ts}^{(\alpha)}}{T^{\alpha/2}}$$

defines a Brownian motion of fraction α as well, and

$$Y^{(T)} = \frac{W^{(T,\alpha)}}{\sqrt{2\log\log T}}$$

converges to the unit ball of the kernel of fractional Brownian motion with fraction α. Unfortunately, there is no easy description of this kernel. Remember that we had to deal with fractional derivatives.

Conventions

$(\Omega, \mathfrak{A}, \mathbb{P})$

is always our underlying probability space. If we say that X is a random vector in a linear space $\mathcal{X}$ we mean, that the mapping $X: \Omega \to \mathcal{X}$ is $\mathfrak{A}$-measurable.

$A \subset B$

is used to denote that A is a (not necessarily strict) subset of B. A strict subset is written as $A \subsetneq B$.

$f \approx g$

Roughly speaking $f \approx g$ if f and g are asymptotically of the same order, i.e.

$$0 < \liminf \frac{f}{g} \leqslant \limsup \frac{f}{g} < \infty.$$

From this it is easily seen that $\approx$ is an equivalence relation.

$f = o(g)$ and $f = O(g)$

These are the usual notations introduced by E. Landau:

$$f = o(g) \quad \Longleftrightarrow \quad \lim \frac{f}{g} = 0$$

and

$$f = O(g) \quad \Longleftrightarrow \quad \limsup \left| \frac{f}{g} \right| < \infty.$$

(y, x)

is often used to denote the appropriate "duality product", meaning that

$$(y, x) := y(x),$$

where e.g. $x \in X$ and $y \in X'$ is a continuous linear functional.

This notation is used to emphasize, that the term may be interpreted both as a function in x and a function in y. In the above example, we might read x as a functional over X'.

In a Hilbert space this notation is used for the inner product as well.

Bibliography

[Bogachev, 1998] Bogachev, V. I. (1998). *Gaussian Measures*. American Mathematical Society.

[Ledoux, 1996] Ledoux, M. (1996). Isoperimetry and Gaussian analysis. In *Lectures on Probability Theory and Statistics*, volume 1648 of *Lecture Notes in Math.*, pages 165–294. Springer.

[Ledoux and Talagrand, 1991] Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces: Isoperimetry and Processes*. Springer, New York.

[Lifshits, 1995] Lifshits, M. A. (1995). *Gaussian Random Functions*. Kluwer.

[Liptser, 1997] Liptser, R. (1997). Stochastic processes. Available at http://www.eng.tau.ac.il/ liptser/.

[Vanden-Eijnden, 2003] Vanden-Eijnden, E. (2003). Stochastic calculus. Available at http://www.cims.nyu.edu/ eve2/.

Index

Armin Straub
math@arminstraub.de