Chapter 2

The signal-in-white-noise model

2.1 Definition of the model

To explore some of the issues that arise in many high-dimensional or nonparametric statistical models it is useful to have a 'canonical' model which on the one hand really exhibits these nonparametric features in a non-trivial way, and on the other hand is tractable enough to allow for a detailed mathematical analysis. In these notes this role is played by the so-called *signal-in-(Gaussian)-white-noise model*.

In this model it is assumed that we observe a sample path $(X_t : t \in [0, 1])$ of a stochastic process X that satisfies the stochastic differential equation (SDE)

$$dX_t = f(t)\, dt + \frac{1}{\sqrt{n}}\, dW_t, \qquad X_0 = 0.$$

Here W is a standard Brownian motion (we recall the definition in the next section) and f is an unknown square integrable function on $[0, 1]$. The SDE is just short-hand notation for the corresponding integral form of the equation:

$$X_t = \int_0^t f(s)\, ds + \frac{1}{\sqrt{n}} W_t, \qquad t \in [0, 1]. \tag{2.1}$$

So indeed, the model postulates that the observed data is a deterministic signal, corrupted with additive Gaussian noise. The noise is called 'white' because the increments of the Brownian motion are uncorrelated.

The unkown parameter in the white noise model is the function $f \in L^2[0, 1]$, so it is a genuinely nonparametric model. The natural number n quantifies the signal-to-noise ratio. It plays the role of sample size in this model, in the sense that it should become easier to recover f as n gets larger. We will be interested in the asymptotic behaviour of statistical procedures as $n \to \infty$.

Figure 2.1 shows a simulated data example. The true underlying signal f is the function on the left. The right-hand panel shows a realization of the process X given by (2.1). The statistical problem is to recover the function f from this noisy observation of its primitive function.

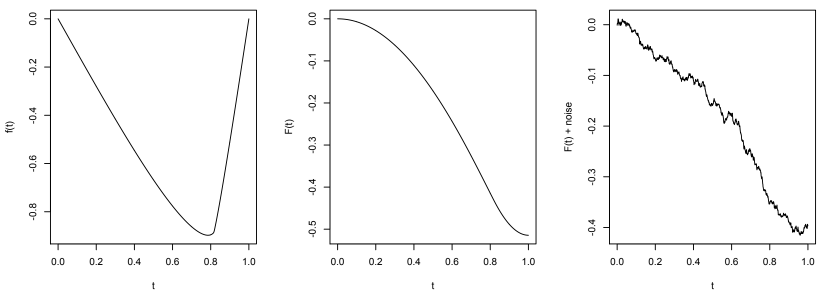

Figure 2.1: From left to right: a function f, its primitive $F(t) = \int_0^t f(s)\,ds$ and the noisy version $F + n^{-1/2}W$.

Clearly the signal-in-white-noise model is somewhat academic. It assumes for instance that we observe a continuous sample path, which is of course not possible in reality. Also, the assumption of purely Gaussian noise can be restrictive in realistic applied settings. Nevertheless, it is a very useful model to study. It allows a detailed analysis that gives insight into some important general phenomena that occur in some form in all high-dimensional or nonparametric models. It can in fact be shown that the model is in some sense very close to some of the more realistic models considered in Chapter 1, as we will briefly discuss in Section 2.5.

2.2 Brownian motion and Wiener integrals

We recall some basic facts about Brownian motion that we need to analyze the signal-in-white-noise model.

Definition 2.2.1. A *(standard) Brownian motion* is a stochastic process $W = (W_t : t \geq 0)$, defined on some probability space $(\Omega, \mathcal{F}, \mathrm{P})$, such that

(i) $W_0 = 0$;

(ii) For all $t \geq s$, the increment $W_t - W_s$ is independent of $(W_u : u \leq s)$;

(iii) For all $t \geq s$, we have $W_t - W_s \sim N(0, t-s)$;

(iv) Almost all sample paths $t \mapsto W_t$ are continuous functions.

Items (ii) and (iii) show that the Brownian motion is a process with *independent and stationary increments*. It is also a *Gaussian process*, meaning that for every finite number of time points $t_1, \ldots, t_n$, the vector $(W_{t_1}, \ldots, W_{t_n})$

has a multivariate normal distribution. Indeed, such a vector has a $N(0, \Sigma)$-distribution, where Σ is the covariance matrix with entries $\Sigma_{ij} = t_i \wedge t_j$. It is not at all immediately clear that the Brownian motion process exists, this is a non-trivial theorem. See for instance Karatzas and Shreve (1991), Revuz and Yor (1991), or Mörters and Peres (2010) for a proof of this theorem and for many of the interesting properties of Brownian motion.

One of the properties of Brownian motion is that its sample paths $t \mapsto W_t$ are very rough. They are nowhere differentiable functions that are not of bounded variation on finite intervals. This implies that integrals of the form $\int f \, dW$ can not be defined simply ω-by-ω as Stieltjes integrals. There exists however a natural alternative way of defining integrals of deterministic functions in $L^2[0, \infty)$ with respect to Brownian motion. Such integrals are called *Wiener integrals*.

For concreteness, let $(\Omega, \mathcal{F}, \mathrm{P})$ be the underlying probability space. For an indicator function f of the form $f = 1_{(s,t]}$ for some $t > s \geq 0$ we simply define

$$I(f) = W_t - W_s.$$

We extend this definition by linearity to simple functions of the form $f = \sum a_i 1_{(s_i,t_i]}$. It is straightforward to verify that this is a proper definition and that such integrals of simple functions are centered Gaussian random variables that satisfy the isometry relation

$$\mathrm{E}I(f)I(g) = \int_0^\infty f(t)g(t)\, dt \tag{2.2}$$

(see Exercise 2.1). Since the simple functions are dense in $L^2[0, \infty)$, this allows us to extend the definition of the integrals to all of $L^2[0, \infty)$ (see Exercise 2.1 again). We end up with a linear map $f \mapsto I(f)$ from $L^2[0, \infty)$ to $L^2(\Omega, \mathcal{F}, \mathrm{P})$ that satisfies the isometry relation for all f and g and such that $I(f) \sim N(0, \int f^2(t)\, dt)$. We note the 'integral' $I(f)$ is defined as an L^2-limit, not ω-by-ω on the underlying probability space. In particular, it is only unique almost surely.

The following theorem summarizes the properties of the Wiener integral.

Theorem 2.2.2. *Let W be a Brownian motion on $(\Omega, \mathcal{F}, \mathrm{P})$. There exists a map $I : L^2[0, \infty) \to L^2(\mathrm{P})$ such that*

(i) If $f = \sum a_{i-1} 1_{(t_{i-1},t_i]}$ for $a_0, \ldots, a_n \in \mathbb{R}$ and $0 \leq t_0 \leq \cdots \leq t_n$, then

$$I(f) = \sum a_{i-1}(W_{t_i} - W_{t_{i-1}});$$

(ii) The map I is linear;

(iii) For all $f \in L^2[0, \infty)$, the random variable $I(f)$ is centered and Gaussian;

(iv) For all $f, g \in L^2[0, \infty)$,

$$\mathrm{E}I(f)I(g) = \int f(t)g(t)\, dt. \tag{2.3}$$

For $f \in L^2[0,\infty)$ and $t \geq 0$ we write $\int f\, dW$ for $I(f)$ and we use the notation

$$\int_0^t f(s)\, dW_s = \int f 1_{(0,t]}\, dW.$$

If X is a stochastic process of the form $X = A + cW$, with A a function, or stochastic process, of bounded variation, W a Brownian motion and $c \in \mathbb{R}$, then we define, for $f \in L^2[0,\infty)$ a function that is integrable with respect to A and $t \geq 0$,

$$\int_0^t f(s)\, dX_s = \int_0^t f(s)\, dA_s + c\int_0^t f(s)\, dW_s.$$

We note that the integral notation is justifiable, in the sense that this stochastic integral has many properties similar to usual integrals. In particular, in view of the definition of the Wiener integrals and the well-known properties of Lebesgue integrals, the integral $\int_0^t f(s)\, dX_s$ can be obtained as an L^2-limit of Riemann-Stieltjes-type sums. Moreover, of course, the map $f \mapsto \int_0^t f(s)\, dX_s$ is linear.

2.3 Sequence formulation of the model

In the signal-in-white-noise model (2.1) we assume that we observe a noisy version of the primitive of a function $f \in L^2[0,1]$. It turns out that equivalently, we can assume that we observe an infinite sequence of noisy Fourier coefficients of f, relative to an arbitrary orthonormal basis of $L^2[0,1]$.

The space $L^2[0,1]$ of (equivalence classes of) square integrable functions on the unit interval is a separable Hilbert space with inner product

$$\langle f, g\rangle = \int_0^1 f(s)g(s)\, ds$$

and corresponding norm $\|\cdot\|_2$ given by $\|f\|_2^2 = \langle f, f\rangle$. Let $e_1, e_2, \ldots$ be an orthonormal basis of the space, for instance the classical Fourier basis. Then every $f \in L^2[0,1]$ can be expanded as $f = \sum f_j e_j$, with convergence in L^2-sense, where $f_j = \langle f, e_j\rangle$. Slightly abusing terminology we call the f_j the Fourier coefficients of f relative to the basis e_j. By the Plancherel relation the Fourier coefficients form a sequence in ℓ^2, with squared ℓ^2-norm given by $\sum f_j^2 = \|f\|_2^2$. (We use the usual notation $\ell^2 = \{(a_1, a_2, \ldots) \in \mathbb{R}^\infty : \sum a_i^2 < \infty\}$ for the Hilbert space of square summable sequences.)

If we have a process X satisfying (2.1) we can define the random variables

$$Y_j = \int_0^1 e_j(s)\, dX_s, \qquad j = 1, 2, \ldots,$$

where the stochastic integral is defined as in the preceding section. If we then set $Z_j = \int_0^1 e_j(s)\, dW_s$ and let f_j be the Fourier coefficients of f relative to the basis e_j, we obtain

$$Y_j = f_j + \frac{1}{\sqrt{n}} Z_j, \qquad j = 1, 2, \ldots. \tag{2.4}$$

Moreover, the properties of the Wiener integral and the fact that the e_j form an orthonormal basis imply that the Z_j form a sequence of independent, standard normal random variables.

Conversely, suppose that (2.4) holds for independent standard normal variables Z_j and a sequence f_j in ℓ^2. Let e_j be an arbitrary orthonormal basis of $L^2[0,1]$. Then it can be shown that $W_t = \sum Z_j \int_0^t e_j(s)\,ds$ defines a Brownian motion (see Exercise 2.2). Since $\sum f_j \int_0^t e_j(s)\,ds = \int_0^t f(s)\,ds$, for $f = \sum f_j e_j$, we then have that $X_t = \sum Y_j \int_0^t e_j(s)\,ds$ is a well-defined stochastic process that satisfies (2.1).

We conclude that the statistical problem of inferring the function $f \in L^2[0,1]$ from observations $(X_t : t \in [0,1])$ satisfying (2.1) is equivalent to the problem of inferring the sequence $(f_j) \in \ell^2$ from the observations Y_j satisfying (2.4).

2.4 Expressions for the likelihood

A stochastic process $X = (X_t : t \in [0,1])$ satisfying (2.1) has continuous sample paths. Hence, it can be seen as a random element in the space $C[0,1]$ of continuous functions on $[0,1]$. More precisely, we view $C[0,1]$ as a Banach space with the uniform norm $\|f\|_\infty = \sup_{t\in[0,1]} |f(t)|$ and a we endow it with the corresponding Borel σ-algebra $\mathcal{B}(C[0,1])$. Then if $(\Omega, \mathcal{F}, \mathrm{P})$ is the probability space on which X is defined, the process X defines a measurable map

$$\begin{aligned}(\Omega, \mathcal{F}) &\to (C[0,1], \mathcal{B}(C[0,1]))\\ \omega &\mapsto (t \mapsto X_t(\omega)).\end{aligned}$$

The *distribution*, or *law* of the process X is the image measure of the probability measure P under this map. For the process X defined by (2.1), we denote this distribution by $\mathrm{P}_{f,n}$. Concretely, for a Borel set $B \subset C[0,1]$ we have

$$\mathrm{P}_{f,n}(B) = \mathrm{P}(X \in B),$$

where X is the process given by (2.1). In particular, the measure $\mathrm{P}_{0,n}$ is the distribution of $W/\sqrt{n}$, for W the standard Brownian motion. The measure $\mathrm{P}_{0,1}$, i.e. the distribution of Brownian motion, is commonly called the *Wiener measure*. When there is no risk of confusion we omit the index n and simply write P_f instead of $\mathrm{P}_{f,n}$.

The following theorem, which is a consequence of the more general *Girsanov theorem*, asserts that for every n, all the laws $\mathrm{P}_{f,n}$, $f \in L^2[0,1]$, are equivalent, i.e. have the same null sets. Moreover, it gives an expression for the Radon-Nikodym derivatives, or likelihood ratios $d\mathrm{P}_{f,n}/d\mathrm{P}_{f_0,n}$. These derivatives are measurable maps on $C[0,1]$ and hence can be viewed as random variables defined on the probability space $(C[0,1], \mathcal{B}(C[0,1]), \mathrm{P}_{g,n})$ for every $g \in L^2[0,1]$. The theorem describes the distribution of the likelihoods when viewed in this manner.

Theorem 2.4.1. *For every $n \in \mathbb{N}$ the laws $\mathrm{P}_{f,n}$, $f \in L^2[0,1]$, are all equivalent measures on $C[0,1]$. Moreover,*

$$\frac{d\mathrm{P}_{f,n}}{d\mathrm{P}_{f_0,n}} = \exp\Big(\sqrt{n}\int_0^1 (f(t) - f_0(t))\,dW_t - \frac{1}{2}n\|f - f_0\|_2^2\Big),$$

$\mathrm{P}_{f_0,n}$-almost surely, where W is a $\mathrm{P}_{f_0,n}$-Brownian motion.

**Proof.* The result is a straightforward consequence of Girsanov's theorem. See for instance Section 3.5 of Karatzas and Shreve (1991). We show that for every pair $f, f_0 \in L^2[0,1]$ we have $\mathrm{P}_{f,n} \ll \mathrm{P}_{f_0,n}$ and that the expression for the Radon-Nikodym derivative given in the statement of the theorem holds. By reversing the roles of f and f_0 we then obtain the statement of equivalence.

Let X be the coordinate process on $C[0,1]$, i.e. $X_t(\omega) = \omega(t)$ for all $\omega \in C[0,1]$ and $t \in [0,1]$. Then the process W defined by

$$dX_t = f_0(t)\,dt + \frac{1}{\sqrt{n}}\,dW_t \tag{2.5}$$

is a $\mathrm{P}_{f_0,n}$-Brownian motion. The process Z defined by

$$Z_t = \exp\Big(\sqrt{n}\int_0^t (f(s) - f_0(s))\,dW_s - \frac{1}{2}n\int_0^t (f(s) - f_0(s))^2\,ds\Big)$$

is a $\mathrm{P}_{f_0,n}$-martingale (check!). Hence, by Girsanov's theorem, the process $\tilde{W}$ defined by

$$d\tilde{W}_t = dW_t - \sqrt{n}(f(t) - f_0(t))\,dt$$

is a Brownian motion under the new measure P defined by $d\mathrm{P} = Z_1\,d\mathrm{P}_{f_0,n}$. Observe that X satisfies $dX = f\,dt + n^{-1/2}d\tilde{W}$. Hence, since $\tilde{W}$ is a P-Brownian motion, we have that $\mathrm{P} = \mathrm{P}_{f,n}$ (check!). This shows that $\mathrm{P}_{f,n} \ll \mathrm{P}_{f_0,n}$ and that $d\mathrm{P}_{f,n}/d\mathrm{P}_{f_0,n} = Z_1$, almost surely. □

If X is the coordinate process on $C[0,1]$, then, by construction, under P_{f_0} it is a process which satisfies (2.1) with f_0 instead of f. Hence, another way to formulate the statement of this theorem is to say that for p_f the density of the law P_f with respect to the law P_0 of the normalized Brownian motion and X a process satisfying (2.1) with f_0 instead of f, the densities a.s. satisfy

$$\frac{p_f}{p_{f_0}}(X) = \exp\Big(\sqrt{n}\int_0^1 (f(t) - f_0(t))\,dW_t - \frac{1}{2}n\|f - f_0\|_2^2\Big), \tag{2.6}$$

where W is a P_{f_0}-Brownian motion. By setting $f_0 = 0$ and recalling (2.5) we find that if X satisfies (2.1), then for the density p_f we have

$$p_f(X) = \exp\Big(n\int_0^1 f(t)\,dX_t - \frac{1}{2}n\int_0^1 f^2(t)\,dt\Big) \tag{2.7}$$

a.s. (check!). Hence if we consider a statistical model where we observe a process X that satisfies (2.1) for some unknown signal f in a collection $\mathcal{F}$, then the MLE for f, if it exists, is the maximizer of the likelihood (2.7) over $\mathcal{F}$. See Exercise 2.3 for a simple example.

2.5 Asymptotic equivalence of nonparametric models

We have already seen that the signal-in-white-noise model is equivalent to the sequence model (2.4). The latter is obviously very similar to the normal means model considered in Example 1.2.2.

In fact, the model is asymptotically close to a number of other nonparametric models as well. This can be made very precise in the context of Le Cam's theory of limits of experiments. See for instance the papers Brown and Low (1996) and Nussbaum (1996) on the equivalence of the signal in white model to nonparametric regression and density estimation. These results are very technical and we will not go into the details here. We just indicate very briefly that there are indeed intimate connections between the models, in order to motivate the fact that the signal-in-white-noise model serves as a 'benchmark model' in this text. The idea is simply that two statistical models are (asymptotically) very similar if the corresponding likelihoods are (asymptotically) very similar.

Example 2.5.1 (Regression)**.** Let $Y_1, \ldots, Y_n$ be observations satisfying

$$Y_i = f(i/n) + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{2.8}$$

where $f : [0,1] \to \mathbb{R}$ is a continuous function and $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. random variables with mean 0 and variance 1. Observing the Y_i is clearly equivalent to observing the process $X^{(n)}$ defined by

$$X_t^{(n)} = \frac{1}{n} \sum_{i \le \lfloor nt \rfloor} Y_i, \qquad t \in [0,1].$$

This process can be decomposed as

$$X_t^{(n)} = \frac{1}{n} \sum_{i \le \lfloor nt \rfloor} f(i/n) + \frac{1}{n} \sum_{i \le \lfloor nt \rfloor} \varepsilon_i.$$

The first term is a Riemann sum that converges uniformly in t to the integral $\int_0^t f(s)\,ds$ if f is for instance Hölder continuous with some exponent $\alpha > 0$. By Donsker's theorem we have the uniform weak convergence

$$\frac{1}{\sqrt{n}} \sum_{i \le \lfloor n\cdot \rfloor} \varepsilon_i \Rightarrow W,$$

where W is a Brownian motion on $[0,1]$. We conclude that for large n, the statistical problem of inferring the regression function f in the regression context (2.8) is very similar to inferring f in the white noise model (2.1).

For a precise result in the context of Le Cam's theory of limits of experiments, see for instance Brown and Low (1996) or Giné and Nickl (2016). The main assumption under which this result is derived is that f is Hölder continuous with some exponent $\alpha > 1/2$. ■

***Example 2.5.2** (Density estimation)**.** Let $X_1, \ldots, X_n$ be a sample from a density f on $[0,1]$. If F is the corresponding distribution function, then we can represent the observations as $X_i = F^{-1}(U_i)$, where $U_1, \ldots, U_n$ are independent, uniform variables on $(0,1)$. Hence, the likelihood ratio can be written as

$$\prod \frac{f}{f_0}(X_i) = e^{\sum \log \frac{f}{f_0}(F_0^{-1}(U_i))}, \tag{2.9}$$

with the U_i i.i.d. and uniform $[0,1]$ under the true distribution P_{f_0}. Let $\mathbb{G}_n(t) = \sqrt{n}(\frac{1}{n}\sum_{i=1}^n 1_{(0,t]}(U_i) - t)$ be the empirical process of the U_i (the normalised empirical distribution function). Then for an arbitrary measurable function h we have

$$\sqrt{n}\int h\, d\mathbb{G}_n = \sum_{i=1}^n h(U_i) - n\int h(t)\, dt.$$

Hence, setting $h = \log \frac{f}{f_0}(F_0^{-1}(\cdot))$ we can rewrite the likelihood ratio as

$$e^{\sqrt{n}\int h\, d\mathbb{G}_n + n\int h(t)\, dt}.$$

Observe that

$$\int h(t)\, dt = \int f_0(t) \log \frac{f}{f_0}(t)\, dt = -\mathrm{KL}(f_0, f).$$

Moreover, by Donsker's theorem the empirical process $\mathbb{G}_n$ converges weakly (for instance in the Skorohod space $D[0,1]$) to a standard Brownian bridge B. Hence, by the continuous mapping theorem, we have the weak convergence

$$\sqrt{n}\int h\, d\mathbb{G}_n \Rightarrow \sqrt{n}\int h(t)\, dB_t.$$

Since B can be represented as $B_t = W_t - tW_1$ for a Brownian motion W, the right-hand side can be written as

$$\begin{aligned}\sqrt{n}\int h(t)\, d(W_t - tW_1) &= \sqrt{n}\int h(t)\, dW_t - W_1 \int h(t)\, dt \\ &= \sqrt{n}\int_0^1 (h(t) - \mathrm{KL}(f_0, f))\, dW_t.\end{aligned}$$

So asymptotically the likelihood ratio looks like a constant times $\exp(\sqrt{n}\int_0^1 (h(t) + \mathrm{KL}(f_0, f))\, dW_t)$. After normalizing we obtain

$$\exp\Big(\sqrt{n}\int_0^1 (h(t) + \mathrm{KL}(f_0, f))\, dW_t - \frac{1}{2}n\int_0^1 (h(t) + \mathrm{KL}(f_0, f))^2\, dt\Big)$$

as approximation to (2.9). By Theorem 2.4.1 this is the likelihood ratio for the model in which we have observations $(X_t : t \in [0,1])$ satisfying

$$dX_t = \Big(\log \frac{f}{f_0}(F_0^{-1}(t)) + \mathrm{KL}(f_0, f)\Big)\,dt + \frac{1}{\sqrt{n}}\,dW_t. \tag{2.10}$$

The derivations above indicate that the statistical problem of inferring the density f from an i.i.d. sample is, for large n, similar to the problem of inferring f from the observation of a stochastic process satisfying (2.10).

If we know in advance that f is close to f_0 we can obtain further approximations. Indeed, in that case $\mathrm{KL}(f_0, f)$ is negligible and

$$h(t) = 2\log\sqrt{\frac{f}{f_0}}(F_0^{-1}(t)) \approx 2\Big(\sqrt{\frac{f}{f_0}}(F_0^{-1}(t)) - 1\Big).$$

It follows that the stochastic integral in the likelihood is then approximately equal to

$$2\int_0^1 \Big(\sqrt{\frac{f}{f_0}}(F_0^{-1}(t)) - 1\Big)\,dW_t = 2\int_0^1 \Big(\sqrt{f(t)} - \sqrt{f_0(t)}\Big)\,dZ_t,$$

where Z is given by

$$Z_t = \int_0^t \frac{1}{\sqrt{f_0(s)}}\,dW_{F_0(s)} = \int_0^{F_0(t)} \frac{1}{\sqrt{f_0(F_0^{-1}(s))}}\,dW_s.$$

The process Z is a continuous martingale with quadratic variation

$$\langle Z\rangle_t = \int_0^{F_0(t)} \frac{1}{f_0(F_0^{-1}(s))}\,ds = \int_0^t \frac{1}{f_0(s)}\,dF_0(s) = t,$$

which implies that Z is in fact a Brownian motion. So for f close to f_0, the likelihood ratio is well approximated by a constant times $\exp(2\int_0^1(\sqrt{f(t)} - \sqrt{f_0(t)})\,dW_t$, where W is a Brownian motion. By Theorem 2.4.1 again, this is the likelihood ratio for the statistical problem of inferring f from the observation of a process X satisfying the SDE

$$dX_t = \sqrt{f(t)}\,dt + \frac{1}{2\sqrt{n}}\,dW_t. \tag{2.11}$$

With (considerably) more mathematical work the local asymptotic equivalence of the statistical problems of i.i.d. density estimation and the signal-in-white-noise model (2.11) can be proved rigorously. Moreover, the local result can be extended to a global result. The main assumptions that are needed to derive the result are that the densities are Hölder continuous of some order $\alpha > 1/2$ and that they are bounded away from 0 on $[0,1]$. See Nussbaum (1996) for details.

Exercise 2.4 briefly explores one more example of asymptotic equivalence, relating drift estimation for a certain class of SDE's to estimating a signal in Gaussian white noise.

2.6 Exercises

Exercise 2.1 (Isometry relation for Wiener integrals)**.** Show that the relation (2.2) holds for simple functions f and g of the form $f = \sum a_i 1_{(t_{i-1},t_i]}$, $g = \sum b_j 1_{(s_{j-1},s_j]}$. Next, show that these simple functions are dense in $L^2[0,\infty)$ and use this to prove that (2.2) holds for all $f, g \in L^2[0,\infty)$.

Exercise 2.2 (Series expansions of Brownian motion)**.** Let e_j be the standard Fourier basis. Show that if Z_j are independent standard normal variables, then the process $W = (W_t : t \in [0,1])$ given by

$$W_t = \sum Z_j \int_0^t e_j(s)\, ds, \qquad t \in [0,1],$$

is well defined and satisfies items (i)–(iii) of Definition 2.2.1. (It can be shown that W satisfies the continuity (iv) of the definition as well and that this is true for *any* orthonormal basis e_j of $L^2[0,1]$, but this is outside the scope of this lecture.)

Exercise 2.3 (Parametric signal-in-white-noise)**.** Suppose we observe a stochastic process $X = (X_t : t \in [0,1])$ satisfying

$$dX_t = \theta\, dt + \frac{1}{\sqrt{n}}\, dW_t,$$

where W is a Brownian motion and $\theta \in \mathbb{R}$ is an unknown parameter. Give an expression for the MLE for θ and derive its asymptotic behaviour.

Exercise 2.4 (Local equivalence of periodic drift estimation and signal-in-white-noise)**.** Suppose that we observe a process $X = (X_t : t \in [0,T])$ satisfying the SDE

$$dX_t = b(X_t)\, dt + dW_t,$$

where b is belongs to the space $C^1_\circ$ of 1-periodic, continuously differentiable functions. The statistical goal is to estimate the function b. For such a process we have a law of large numbers which asserts that if $f \in C^1_\circ$, then as $T \to \infty$

$$\frac{1}{T}\int_0^T f(X_t)\, dt \to \int_0^1 f(x)\rho(x)\, dx,$$

almost surely, where ρ is the probability density on $[0,1]$ given by

$$\rho(x) = Ce^{2\int_0^x b(y)\, dy},$$

where C is the appropriate normalizing constant. Moreover, we have a central limit theorem which asserts that for $f \in C^1_\circ$

$$\frac{1}{\sqrt{T}}\int_0^T f(X_t)\,dW_t \xrightarrow{\mathrm{d}} N\Big(0, \int_0^1 f^2(x)\rho(x)\,dx\Big)$$

as $T \to \infty$.

Let $\mathrm{P}_{b,T}$ be the law that the process X induces on the space $C[0,T]$ of continuous functions on $[0,T]$. By a slight extension of Theorem 2.4.1, the likelihood ratio $d\mathrm{P}_{b,T}/d\mathrm{P}_{b_0,T}$ for two functions $b, b_0 \in C^1_\circ$ is given by

$$\frac{d\mathrm{P}_{b,T}}{d\mathrm{P}_{b_0,T}} = \exp\Big(\int_0^T (b-b_0)(X_t)\,dW_t - \frac{1}{2}\int_0^T (b-b_0)^2(X_t)\,dt\Big),$$

where W is a $\mathrm{P}_{b_0,T}$-Brownian motion. Use this in combination with the LLN and CLT given above to convince yourself that if we know in advance that b is close to another function b_0 in $C^1_\circ$, then for $T \to \infty$, the statistical problem under consideration is close to a signal-in-white-noise problem. Which one exactly?