THE PINSKER BOUND IN MIXED GAUSSIAN WHITE NOISE

SYLVAIN DELATTRE AND MARC HOFFMANN

Université Paris XII Val de Marne and Université Paris VII

Abstract. We study the problem of estimating a signal f from noisy data under a certain squared-error type loss. We assume that f belongs to a certain Sobolev class. The noise process is represented by $t \to \frac{1}{\sqrt{n}} \int_0^t \sqrt{V_s} dW_s$, where V is a random process independent of the driving Brownian motion W. Thus, conditional on V, the function f is observed in Gaussian white noise. This setup generalizes the traditional 'ideal signal + noise' framework adopted in nonparametric estimation.

We establish upper and lower bounds for the asymptotic minimax risk (as $n \to \infty$) up to constants. Since V is observable and its law does not depend on f, it is an ancillary statistic. Links to Fisher's ancillary principle are discussed. In particular, we characterize the influence of the law of V on the optimal constants.

Key words and phrases. Gaussian white noise, mixed normality, nonparametric L_2 efficiency, Pinsker bound, linear filtering, minimax estimation, Sobolev ellipsoids, ancillary statistics

AMS Classification: 62B15, 62G05.

Correspondence should be addressed to: M. Hoffmann, Laboratoire PMA, CNRS UMR 7599 et Université Paris VII, 13 rue Clisson, 75013 Paris.

1. Mixed Gaussian white noise

1.1. **Gaussian white noise.** The problem of *signal recovery in Gaussian white noise* has served for long as a representation for signal recovery in nonparametric statistics: for $n \geq 1$, we observe the stochastic process $(X_t^n)_{t\in[0,1]}$ defined by

$$dX_t^n = f(t)dt + \sqrt{\frac{v(t)}{n}}\, dW_t, \tag{1}$$

where $W = (W_t)_{t\in[0,1]}$ is a Wiener process started at $W_0 = 0$ and f is the unknown parameter of interest, belonging to a compact subset $\Sigma \subset L_2([0,1])$. The variance function $v(t)$ is bounded below, integrable, and measures the local fluctuations of the noise intensity, whose order of magnitude is roughly $1/\sqrt{n}$. The goal is to estimate f as $n \to \infty$, i.e. when the noise level vanishes asymptotically.

1.1.1. *Gaussian white noise as limit experiment for nonparametric estimation.* The cognitive value of model (1), even in the simplest case when $v(t)$ is constant goes back to Ibragimov and Khasminski (1977). Due to its mathematical simplicity, it is a natural candidate for an ideal, or test model, for curve estimation. Indeed, under some mild assumptions on the smoothness of f, the regression model, when one observes a signal f contaminated by Gaussian noises of variance v and sampled on a uniform grid with mesh $1/n$ is asymptotically "close" to the homogeneous white noise model, i.e. when $v(t)$ is constant for all t in (1). These heuristics were extensively developed by the Russian school in the 1980's, see the references below.

The precise mathematical relationship between white noise and curve estimation was eventually established by Brown and Low (1996) who proved the asymptotic equivalence (see e.g. Le Cam and Yang, 1990) of model (1) and nonparametric regression. An analogous result for asymptotic equivalence of density estimation and homogeneous white noise was obtained by Nussbaum (1996).

More generally, introducing some inhomogeneity in the variance function significantly enriches the curve estimation modelling. For instance, Brown and Low (1996) proved that model (1) is asymptotically equivalent to the following experiment: for $i = 1,\dots,n$, we observe (X_i^n, Y_i^n), where

$$Y_i^n = f(X_i^n) + \sigma(X_i^n)\, \varepsilon_i^n. \tag{2}$$

The random variables ε_i^n are i.i.d. $\mathcal{N}(0,1)$ and the X_i^n are independent of the ε_i^n and are i.i.d. with density μ. The equivalence between (1) and (2) is obtained under the calibration $v(t) = \sigma^2(t)/\mu(t)$. Other improvements in this direction were further obtained by Grama and Nussbaum (1996) for generalized linear models, Milstein and Nussbaum (1998) for diffusion processes with small variance.

1.1.2. *Gaussian white noise and parametrically LAN models.* The white noise model (1) has another fundamental property: it may serve as well as a local approximation model for parametric experiments. Indeed, regular parametric models can be asymptotically identified with Gaussian shift experiments, a statement which we know in modern terms as the LAN property (local asymptotic normality, see e.g. Ibragimov and Khasminski, 1980). In the nonparametric case, the white noise model (1) is precisely the nonparametric extension of a Gaussian shift experiment. This concept together with its applications for risk bounds in L_2-loss were investigated by Golubev (1992); see also Efromovitch, (1996).

We summarize the two above paragraphs by stating the following broad guiding principle:

The Gaussian white noise model (1):

(i) *has a rich enough structure to encompass various models such as density estimation or signal estimation in various contexts (heteroscedastic, irregular design),*
(ii) *appears as a natural limit model for nonparametric experiments which are parametrically LAN.*

1.2. **Mixed Gaussian white noise.** The goal of this paper is to get to the logical next step, by allowing a random structure in the variance of the Gaussian white noise. A first pilot model is obtained by replacing the function $v(t)$ in (1) by a random process $(V_t)_{t\in[0,1]}$ independent of the driving Brownian motion W: we observe the stochastic process $X^n = (X^n_t)_{t\in[0,1]}$ on $\mathbb{R}$, with now

$$dX^n_t = f(t)dt + \sqrt{\frac{V_t}{n}}\, dW_t. \tag{3}$$

We have a noise process which is Gaussian conditional on V: we call it a *mixed white noise experiment* (MWN). Following (ii) of our guiding principle, the MWN appears as a natural candidate for a limit model for nonparametric experiments with random information which are parametrically *locally asymptotically mixed normal*, or have the LAMN property (see again Ibragimov and Hasminski, 1980, or Jeganathan, 1982 and 1983). Investigating directly MWN experiments may presumably enlarge the scope of models where fine nonparametric asymptotic results can be obtained.

Let us give some examples of application and a conjecture for the use of MWN as limit experiments; two of them are borrowed from the statistic of random processes.

1.2.1. *Example 1: nonparametric regression with mixed normal noise.* Consider the following regression model: for $i = 1, \dots, n$, we observe

$$Y_i^n = f(i/n) + \sqrt{G}\, \varepsilon_i^n \tag{4}$$

where the ε_i^n are i.i.d. $\mathbf{N}(0,1)$ and are independent of the random variable G, which we observe. In the same line as Theorem 4 in Brown and Low (1996), Model (4) is asymptotically equivalent to the mixed white noise model. We outline the proof in the appendix.

1.2.2. *Example 2: a null recurrent diffusion model.* We consider a continuous process Y of the form

$$Y_s = x_0 + \int_0^s f(Y_u)du + B_s, \quad s \in \mathbb{R}_+ \tag{5}$$

where $x_0 \in \mathbb{R}$ and $(B_s)_{s\in\mathbb{R}_+}$ is a standard Brownian motion. The unknown 1-dimensional function f is assumed to be smooth and compactly supported in $[-r, r]$. In Delattre and Hoffmann (2001) we prove that the experiment generated by the observation $(X_t)_{t\in[0,n]}$ for $n \geq 1$ is locally asymptotically equivalent to the mixed Gaussian white noise experiment

$$dX_t^n = (f - f_0)(t)dt + n^{-1/4}\Lambda^{-1/2} m_{f_0}(t)^{-1/2} dW_t, \tag{6}$$

where $\Lambda = |\mathcal{N}(0,1)|$ in law, and

$$m_f(t) = \frac{2}{\exp\big(-2\int_{-r}^t f(y)dy\big) + \exp\big(2\int_t^r f(y)dy\big)}.$$

The local approximation holds for f in a neighbourhood around f_0 in a proper topology, shrinking to 0 at some rate as $n \to \infty$. Thus, the information is measured by the amount $n^{-1/4}$ times the random variable $\Lambda^{-1/2} m_{f_0}(t)^{-1/2}$ which somehow depends on the local time of the process Y.

1.2.3. *Example 3: estimating the diffusion coefficient from discrete observations.* We again consider a continuous adapted 1-dimensional process Y^σ of the form

$$Y_s^\sigma = x_0 + \int_0^s b_u du + \int_0^s \sigma(Y_u^\sigma)dB_u, \ s \in [0,1] \tag{7}$$

where $x_0 \in \mathbb{R}$ and $(B_s)_{s\in[0,1]}$ is a standard Brownian motion. We observe the process Y^σ sampled at times i/n for $i = 0, \dots, n$. The unknown parameter is the diffusion coefficient (the volatility function) $\sigma^2(s)$ and the drift b_s is a nuisance parameter. This volatility model has some importance in financial mathematics.

Nussbaum [personal communication] suggested that under some restrictions, the model generated by discrete data $Y_{i/n}^\sigma, i = 0, \dots, n$ is asymptotically equivalent to a mixed white noise experiment. A candidate for a local equivalence would be

$$dX_t^n = \frac{1}{\sqrt{2}} \log \frac{\sigma^2(t)}{\sigma_0(t)} dt + \frac{1}{\sqrt{n}} \frac{1}{\sqrt{L_1^t(Y^{\sigma_0})}} dW_t \tag{8}$$

for $\inf_{s \in [0,1]} Y_s^{\sigma_0} < t < \sup_{s \in [0,1]} Y_s^{\sigma_0}$. Here W is a two-sided Brownian motion, independent of Y^{σ_0} and $L_1^t(Y^{\sigma_0})$ denotes the local time of Y^{σ_0} at time 1 and level t.

We see that the information of the model is given by the classical $n^{-1/2}$ times the random process $1/\sqrt{L_1^t(Y^{\sigma_0})}$, which precisely measures the time asymptotically spent by the process at level t, and thus reveals its importance for the nonparametric estimation of $\sigma^2(t)$.

1.3. **Formal definition.** A general mixed white noise experiment can be described as follows: we observe $X^n = (X_t^n)_{t \in [a,b]}$ on $\mathbb{R}$, with

$$dX_t^n = f(t)dt + \sqrt{\frac{V_t}{n}}\, dW_t. \tag{9}$$

Here $W = (W_t)_{t \in [a,b]}$ is a standard Brownian motion. Note that the observation of X^n implicitely yields the observation of V, since V is the quadratic variation of X^n. Namely, for all interval $[s,t] \subset [a,b]$:

$$\int_s^t n^{-1} V_u \, du = \lim_{p \to \infty} \sum_{t_i^p \in \tau_p^{s,t}} (X_{t_i^p}^n - X_{t_{i-i}^p}^n)^2, \tag{10}$$

where $\tau_p^{s,t}$ is a sequence of nested subdivisions of the interval $[s,t]$ with mesh going to 0 as $p \to \infty$ and the limit is taken for the convergence in probability.

Let $\rho : [a,b] \to (0,\infty)$ be a continuous weight function. We define the weighted Sobolev class associated to ρ as

$$\mathbf{W}(\beta,\rho) = \big\{ f : \mathbb{R} \to \mathbb{R} \text{ such that } \int_a^b |f^{(\beta)}(t)|^2 \, \rho(t) \, dt \le 1 \big\} \tag{11}$$

where $\beta \ge 1$ is an integer and $f^{(\beta)}$ denotes the derivative of order β of f in the distributional sense.

We will further consider the experiment generated by X^n with parameter f and space parameter $\mathbf{W}(\beta,\rho)$.

1.4. **Objectives: the Pinsker bound.** We have so far defined the MWN as an extension of the classical Gaussian white noise model, based on the principle that it appears as an ideal nonparametric model for signal estimation with random information. We now define a precise statistical program within this new setup.

Our goal is to solve the so-called Pinsker problem. Actually, the Pinsker program, even in its formulation, significantly differs from the nonrandom case, i.e. when $V(t) = v(t)$ is deterministic. The Pinsker bound (see Nussbaum (1999), for a review of the concept) describes the exact asymptotics of the minimax risk in a class of nonparametric estimation problems where root-n consistent estimators do not exist. Accordingly, the Pinsker bound provides not only the optimal rate of convergence for estimators, but also the optimal constants. The optimal constants are well known for estimation in regular parametric models, and are usually given by the Fisher information. To that extent, the Pinsker bound is a nonparametric analogue of the Fisher bound in parametric estimation.

2. The Pinsker bound

2.1. **Pinsker's program and the minimax theory.** Given an estimator F of the unknown parameter f and a risk function $R_n(\cdot, f)$ we define *the maximal or worst risk of F relative to R_n* as

$$\mathbf{R}_n(F) = \sup_{f \in \mathbf{W}(\beta,\rho)} R_n(F, f).$$

An estimator is said to attain an optimal rate of convergence $r_n \to 0$ (relative to R_n) if for some constant c_1:

$$\mathbf{R}_n(F) \leq r_n\, c_1(1 + o(1))$$

and no estimator can attain a better rate: for a $c_2 > 0$

$$\inf_F \mathbf{R}_n(F) \geq r_n\, c_2(1 + o(1))$$

where the infimum is taken over all estimators. For reasonable choice of ρ, β and v, it is well-known that the optimal rate relative to the mean integrated squared-error risk

$$R_n(F, f) = E_f^n\Big[\int_a^b |F(t) - f(t)|^2 dt\Big] \tag{12}$$

is $r_n = n^{-2\beta/(1+2\beta)}$, where E_f^n denotes expectation w.r.t. the law of the observation. In a certain sense, the speed of convergence measures the complexity of the model. However, rate optimality is unsatisfactory if finer structures are considered: it only involves β and rules out the influence of ρ and v. Theferore, the next level of analysis is to find the optimal bound $c_1 = c_2$ if it exists.

2.1.1. *Pinsker's result.* For the homogeneous white noise model (1) and if $\rho(t)$ is constant for all t, the value of the optimal constant c_1 is known since 1980 with the celebrated paper of Pinsker (1980) on *"Optimal filtering of square integrable signals in Gaussian white noise"* . Progressively, results about optimal constants in L_2 appeared in various models. Among others: Efromovitch and Pinsker (1982) in density estimation, Nussbaum (1985) for nonparametric regression with uniform design, Golubev and Nussbaum (1990) for regression with random design, Golubev (1992, 1993) for Gaussian stationary sequences, Belitser and Levit (1995) for inhomogeneous variance functions.

2.1.2. *Relation to the minimax theory.* In his original approach, by using the term "optimal filtering" Pinsker related his result to a Bayesian concept, namely that a minimax problem is equivalent to a worst-case Bayesian problem. More precisely, we can rewrite

$$\inf_F \mathbf{R}_n(F) = \inf_F \sup_\mu \int_{\mathbf{W}(\beta,\rho)} \mu(df) R_n(F,f), \tag{13}$$

where the supremum is taken over all prior probabilities μ with support concentrated on $\mathbf{W}(\beta,\rho)$. Under some restriction on the set of estimators, we can interchange inf and sup and (13) is equal to

$$\sup_\mu \left(\inf_F \int_{\mathbf{W}(\beta,\rho)} \mu(df) R_n(F,f) \right).$$

Thus establishing the Pinsker bound consists in finding the asymptotically -up to constants- worst Bayesian prior. "Evil" nature competes with the statistician as follows: it chooses a "bad" sequence of priors μ_n^+ concentrated on the parameter space $\mathbf{W}(\beta,\rho)$. The statistician then picks the best possible sequence of estimators against μ_n^+. The optimal (minimax) estimator is designed to fight against the least favourable sequence μ_n^+, i.e. can compete efficiently in the following situation: a referee generates the realization (f,ω) of a random variable distributed as $\mu_n^+(df) \otimes P_f^n(d\omega)$. In this acceptance of asymptotic minimaxity:

It is essential to separate the structure of the prior μ_n^+ and the measurement of the outcome ω in the statistical experiment,

since ω is distributed according to $P_f^n(d\omega)$ and $f \sim \mu_n^+(df)$. This remark will be of importance in the next section.

2.2. **The choice of a risk function for mixed Gaussian white noise.** If trying to extend the Pinsker bound to MWN, it seems natural at first glance to pick the mean integrated squared error $R_n(F,f)$ defined in (12) as a risk function. However, the structure of the MWN raises a delicate issue:

2.2.1. *Fisher's Ancillary Principle and conditional squared-error risk.* According to the ancillary principle established by R.A. Fisher in the late 1920's -see e.g. the paper of Kiefer, (1972), in the Encyclopedia of Statistical Sciences- repeated sampling criteria should be applied to conditional experiments defined by setting ancillary statistics at their observed value.

An ancillary statistic is an observable random variable which distribution does not depend on the unknown f. In the MWN, the variance process V is observed and its law does not depend on f: it is an ancillary statistic. According to Fisher's principle, one should consider

$$R_n^{\texttt{conditional}}(f, F, v) = E_f^n\Big[\int_a^b |F(t) - f(t)|^2 dt \;\Big|\; V = v\Big] \qquad (14)$$

as risk function. Indeed, one can hardly deny that a sensible measure of the risk is the conditional risk $R_n^{\texttt{conditional}}(f, F, v)$ given $V = v$. Of course, we could object that in practice, we do not observe the limiting variance process V of the ideal MWN model. However, for the sake of coherence, we must consider the ancillarity issue even in the limit. In that case, we are back to Pinsker's original result: the worst sequence of priors $\mu_n^+(v)$ is allowed to depend on v and we find the same Pinsker constant as the one obtained for the Gaussian white noise (1) with variance function v. The flexibility of the model due to a random structure in the variance is apparently lost, at least within the context of Pinsker's problem.

2.2.2. *The ancillary principle within the minimax theory.* However, if we abruptly apply the ancillary principle without care, and use (14) to solve our minimization problem, we are led to the following mathematical paradox:

Admittedly, the statistic V is a function of X and thus the observed value $v = v(\omega)$ of V depends on the observed value ω of X. But this comes into conflict with the minimax paradigm contained in Pinsker's program: the worst sequence of priors $\mu_n^+(df)$, which generates the function f, which itself determines the law $P_f^n(d\omega)$ selected to pick the observation ω, cannot depend on $v(\omega)$, i.e. on the observation itself! This is implicitly stated in (14) when we solve the associated Pinsker problem.

We can try to highlight this paradox: the Pinsker constant, in the acceptance of its Bayesian aspects, is an index of complexity of a statistical model, and *not an index of precision of estimation.* Clearly, the minimax criterion tries to minimize the worst case performance, and the *charasteristics of the problem* have to be set prior to the experiment. In this context, one shall take advantage of the random structure of V, which is precisely left out by the ancillarity principle.

2.2.3. *The unconditional risk.* Another understanding of the preceding point can be given as follows. Suppose that, given an estimator F and prior to the experiment, the statistician wonders how many observations are required to build a confidence set (in integrated squared norm) of the unknow parameter, given a certain confidence level. Such a scenario is rather hypothetic, but proves to be helpful for the overall understanding of our setting. In that case, the quantity

$$\sup_{f \in \mathbf{W}(\beta,\rho)} E_f^n\Big[\int_a^b |F(t) - f(t)|^2 dt \;\Big|\; V = v\Big]$$

is not available at this stage and the statistician may legitimately choose to work rather with

$$\sup_{f\in\mathbf{W}(\beta,\rho)}\int \mathcal{L}(V)(dv)\; E_f^n\Big[\int_a^b |F(t)-f(t)|^2 dt \;\Big|\; V=v\Big],$$

where $\mathcal{L}(V)$ denotes the law of V, i.e. average over all possible scenarios. But this last quantity is precisely $\mathbf{R}_n(F)$ defined in (2.1). This latter choice is just the radical opposite point of view and suggests to take the unconditional risk instead. The unconditional risk has the advantage to be compatible with the minimax theory. It is not satisfactory though, since the ancillary issue is left out.

2.3. **A unifying approach: a functional risk.** We wish to define an appropriate risk function for treating the MWN model which possesses the following properties:

(i) The risk function should be compatible with the minimax theory, and thus lead to a consistent Pinsker problem (Section 2.2.2.)
(ii) It should integrate Fisher's ancillary principle (Section 2.2.1.).
(iii) It should take into account the possibility to consider the unconditional risk (Section 2.2.3.).

In particular, Pinsker's original result should be retrieved when V is deterministic. We propose the following compromise: let

$$\Phi\;:[a,b]\times\mathcal{C}([a,b],\mathbb{R})\rightarrow\mathbb{R}_+$$

be an arbitrary positive functional on the space of real-valued continuous functions on $[a,b]$. Given an estimator F of f, we propose to consider the following *penalized risk*

$$R_n(F,f)[\Phi] := E_f^n\Big[\int_a^b |F(t)-f(t)|^2\Phi[t,(V_s)_{s\in[a,b]}]dt\Big]. \qquad (15)$$

Accordingly, we define a *penalized maximal risk* by taking

$$\mathbf{R}_n(F)[\Phi] = \sup_{f\in\mathbf{W}(\beta,\rho)} R_n(F,f)[\Phi]. \qquad (16)$$

Solving Pinsker's problem corresponds then to finding a functional

$$\Phi\mapsto P^\star(\beta,\rho,\mathcal{L}(V))[\Phi]$$

such that:

$$\lim_{n\to\infty} n^{2\beta/(2\beta+1)}\inf_F \mathbf{R}_n(F)[\Phi] = P^\star(\beta,\rho,\mathcal{L}(V))[\Phi], \qquad (17)$$

for all positive Φ defined as above and where the infimum is taken over all estimators. Let us briefly discuss the functional penalized risk (15): the function Φ models the influence of the ancillary statistic V we wish to take into account in the risk function. In the context of Section 2.2.3, that is prior to any experiment, one can take $\Phi=1$, retrieve the unconditional risk and agree with (iii) of our requirement. The influence of the law of V, if any, will be maximal in this case. From an opposite point of view, choosing a sequence Φ_n such that $\Phi_n(t,V) \equiv \Phi_n(V)$ and

$$\Phi_n(u)\,\mathcal{L}(V)(du) \to \delta_v \text{ weakly as } n \to \infty,$$

where δ_v is the Dirac mass in v one retrieves the conditional risk given $V = v$ and Fisher's ancillary principle in its simplest form. The possible influence of $\mathcal{L}(V)$ on the problem is ruled out. We meet in this case point (ii) of our requirement.

More interestingly, the flexibility gained by considering an integrated risk versus trial functions Φ agrees with the minimax theory: the "worst" sequence μ_n^+ of prior probabilities for which we (asymptotically solve) the minimization problem

$$\inf_F \int \mu_n^+(df) R_n(F,f)[\Phi] \tag{18}$$

does not depend on the ancillary outcome $V(\omega) = v$, strictly speaking, but rather on the test function Φ set prior to the experiment. In other words

Our statistical conflict about the ancillarity issue can be solved by our capacity to express such quantities (18), depending on Φ.

2.4. **Main results.** In this paper, we compute the Pinsker bound for the mixed white noise model (9) and study the influence of the law $\mathcal{L}(V)$ of the variance process V on the minimax constants for the penalized risk (15) under the general Sobolev constraint $\mathbf{W}(\beta,\rho)$.

Generally speaking, we can assess the influence of $\mathcal{L}(V)$ on the Pinsker bound by studying

$$\Phi \mapsto P^\star(\beta,\rho,\mathcal{L}(V))[\Phi] = \lim_{n\to\infty} n^{2\beta/(1+2\beta)} \inf_F \mathbf{R}_n(F)[\Phi].$$

If $\mathcal{L}(V)$ plays no role, the mapping $\Phi \mapsto P^\star(\beta,\rho,\mathcal{L}(V))[\Phi]$ should be linear in the homogeneous case when $V_t \equiv V$. This becomes transparent if we consider instead a simpler finite dimensional problem: let us temporarily restrict ourselves to the class of constant functions

$$\mathbf{W}_0 := \{f(t) = m \in [m_1, m_2]\},$$

for some interval $[m_1,m_2] \subset \mathbb{R}$. Then the vector (X_1^n, V) is a sufficient statistic, and we can equivalently consider the observation $X_1^n = m + \sqrt{V} n^{-1/2}\, W_1$. In that case, we know that

$$\lim_{n\to\infty} n \inf_F \sup_{f\in\mathbf{W}_0} R_n(F,f) = \int v\, \mathcal{L}(dv), \tag{19}$$

as shown by considering, first, the conditional risk $R_n^{\texttt{conditional}}(f, F, v)$ given $V = v$ for which the limit is given by the inverse of the conditional Fisher information $1/v$ and, then, integrating w.r.t. $\mathcal{L}(V)$. In particular, for any positive test function Φ:

$$\lim_{n\to\infty} n \inf_F \sup_{f\in\mathbf{W}_0} R_n(F, f)[\Phi] = \int v\ \Phi(v)\ \mathcal{L}(dv),$$

and, as we can see, the map $\Phi \mapsto n \inf_F \sup_{f\in\mathbf{W}_0} R_n(F, f)[\Phi]$ is linear. There is no influence of $\mathcal{L}(V)$ at the level of constants: the optimal bound for either the conditional or the unconditional risk is attained by the same estimator $F_0(t) = X_1^n$. In particular, a discussion about ancillarity is unnecessary in this parametric framework.

Surprisingly, the situation changes in a nonparametric context: there is a dimension effect which affects the Pinsker bound: Theorem 5 below shows that the transform $\Phi \mapsto P^\star(\beta, \rho, \mathcal{L}(V))[\Phi]$ becomes nonlinear as soon as $\mathcal{L}(V)$ is nondegenerate.

Even in the simplest case where $\Phi \equiv 1$, a simple modification of the asymptotically optimal estimator in the deterministic case cannot lead to an optimal bound. More precisely, if $F_n^\star(v, t)$ denotes the optimal Pinsker estimator in the deterministic case (i.e. when $\mathcal{L}(V) = \delta_v$), then the conditional Pinsker estimator $F_n^\star(V, t)$ obtained by plugging V in $F_n^\star(\cdot, t)$ is asymptotically suboptimal: under some restrictions, Theorem 1 shows that if $\mathcal{L}(V)$ is nondegenerate

$$\lim_{n\to\infty} \frac{\inf_F \mathbf{R}_n(F)}{\mathbf{R}_n(F_n^\star(V, \cdot)} < 1.$$

Not only the asymptotically minimax estimator must depend on the realization of the variance process V, but also on $\mathcal{L}(V)$.

More precisely, if we consider for instance the homogeneous case with $[a, b] = [0, 1]$, $\rho(t) = \rho$ and $V_t = V$ are constant in time, Lemma 1 belows shows that for nice enough Φ, the number $a_\Phi^\star(x)$ that solves the equation

$$E\left[\frac{V^2}{xV + a_\Phi^\star(x)}\Phi(V)\right] = E\left[\Phi(V)\right] \tag{20}$$

is well defined. Theorem 5 then shows that

$$\lim_{n\to\infty} n^{2\beta/(2\beta+1)}\mathbf{R}_n(F)[\Phi] = P^\star(\beta, \rho, \mathcal{L}(V))[\Phi],$$

where the asymptotic constant is given by

$$P^\star(\beta, \rho, \mathcal{L}(V))[\Phi] = C_\beta^\star \frac{E[\Phi(V)]}{\rho^{1/2\beta+1}} \left(\int_0^1 a_\Phi^\star(x) x^{1/\beta} dx\right)^{2\beta/(2\beta+1)}$$

for an explicit $C_\beta^\star > 0$. A extention of this type for nonhomogeneous V and arbitrary ρ is given in Theorem 6. The nonlinear character of the Pinsker functional is contained in the transform $a^\star_\Phi(x)$.

In particular, if the law of V is degenerate, i.e. $\mathcal{L}(V) = v$, then (20) does not depend on Φ and $a^\star(x) = v(1-x)$. It follows that

$$P^\star(\beta,\rho,\mathcal{L}(V))[\Phi] = \frac{v^{2\beta/(2\beta+1)}}{\rho^{1/(2\beta+1)}} P^\star_\beta \cdot \Phi(v),$$

where now $P^\star_\beta$ is the classical Pinsker constant, specified in (24) below. In that latter case, the map $\Phi \mapsto P^\star(\beta,\rho,\mathcal{L}(V))[\Phi]$ is linear and we retrieve Pinsker's original result of 1980.

As we see, in a regular parametric model with random information, we can apply Fisher's principle and adopt a minimax approach simultaneously: both points of view agree. This seems no longer true for estimating an infinite dimensional parameter: the analogy between Fisher's information and the Pinsker bound (Nussbaum, 1999) reaches its limit in this context.

2.5. **Organization of the paper.** In Section 3.1, we consider the case of a homogeneous variance process $V_t \equiv V$ when $\Phi \equiv 1$ and $E(V)$ is finite. We restrict here the definition of the signal f to the interval $[a,b] = [0,1]$.

We first treat the case of linear estimators over periodic Sobolev classes with constant weight function $\rho(t) \equiv \rho$. Evaluation of the minimax risk among linear estimators shows that the conditional Pinsker estimator can be beaten in this context. We exhibit the optimal linear estimator $F_n^{\star\star}$ by means of the $a^\star(x)$ transform, defined in Lemma 1.

We can extend this result to a periodic Sobolev class of bounded functions in Theorem 2, which enables us to relax the integrability assumption on V. Proposition 2 and 40 study some qualitative properties of the transform $a^\star(x)$. Theorem 3 shows the optimality of the linear estimator $F_n^{\star\star}$ among all estimators.

Section 3.2 carries over those results to nonperiodic bounded Sobolev classes in Theorem 4, under additional assumptions on the integrability of V. Section 3.3 extends further to the case of f defined on an arbitrary interval $[a,b] \subset \mathbb{R}$ and introduces arbitrary penalty functions $\Phi(V)$ in the risk function. We have a a first characterization of the Pinsker functional $P^\star(\beta,\rho,\mathcal{L}(V))[\Phi]$ in this homogeneous context in Theorem 5.

A final generalization in the nonhomogeneous case (simultaneously for V and ρ) is given in Theorem 6 of Section 4. Some proofs are delayed until Section 5.

3. The homogeneous case

In this section, we consider the case

$$V_t \equiv V, \quad \Phi(t,V) \equiv \Phi(V) \quad \text{and} \quad \rho_t \equiv \rho > 0, \tag{21}$$

where Φ is the test function in the penalized risk (15) and ρ is the weight function in the definition of the class $\mathbf{W}(\beta,\rho)$ in (11). Note that we can consider X^n as the canonical process on the canonical space, and thus the dependence on n may be dropped. All the information about the asymptotics is transferred into P_f^n.

3.1. **The periodic case.** Assume further that $\Phi(V) = 1$, $[a,b] = [0,1]$ and let us restrict ourselves to $\widetilde{\mathbf{W}}(\beta,\rho)$, the subset of $\mathbf{W}(\beta,\rho)$ consisting of functions f such that $f^{(k)}(0) = f^{(k)}(1)$ for $k = 0,\dots,\beta-1$. In this setting, the worst risk reads

$$\mathbf{R}_n(F, \widetilde{\mathbf{W}}(\beta,\rho)) = \sup_{f \in \widetilde{\mathbf{W}}(\beta,\rho)} E_f^n\left[\int_0^1 \left|F(t) - f(t)\right|^2 dt\right]. \tag{22}$$

The trigonometrical othonormal basis in $L_2([0,1])$ can be used to represent a Sobolev class as an ellipsoid. For $\lambda \in \mathbb{Z}$, define the Fourier coefficient $\hat f(\lambda)$ of f as $\hat f(\lambda) = \int_0^1 e^{i2\pi\lambda t} f(t)\,dt$. We thus have the representation

$$f \in \widetilde{\mathbf{W}}(\beta,\rho) \Leftrightarrow \forall \lambda \in \mathbb{Z},\ \hat f(\lambda) = \overline{\hat f(-\lambda)} \text{ and } \sum_{\lambda \geq 1} \lambda^{2\beta}\, |\hat f(\lambda)|^2 \leq \frac{1}{2\rho(2\pi)^{2\beta}}.$$

For $n \geq 1$, a linear filter is a sequence $(\varphi_n(\lambda,V))_{\lambda\in\mathbb{Z}} \in l^2(\mathbb{Z})$ such that $0 \leq \varphi_n(\lambda,V) \leq 1$ almost surely for all λ. For such a φ_n, a linear estimate F_n of f is given by

$$F_n(t) = \sum_{\lambda\in\mathbb{Z}} \varphi_n(\lambda,V) Y_\lambda\, e^{-i2\pi\lambda t}, \quad \text{where} \quad Y_\lambda = \int_0^1 e^{i2\pi\lambda t}\, dX_t$$

denotes the empirical Fourier coefficient. The worst risk of the linear estimator F_n over $\widetilde{\mathbf{W}}(\beta,\rho)$ reads

$$\mathbf{R}_n(F_n) = \sup_{f \in \widetilde{\mathbf{W}}(\beta,\rho)} E_f^n\left[\sum_{\lambda\in\mathbb{Z}} \left|\hat f(\lambda) - \varphi_n(\lambda,V) Y_\lambda\right|^2\right].$$

3.1.1. *Pinsker's result.* If the law of V is degenerate, i.e. $V \equiv v > 0$, we find back the framework of Pinsker (1980), which corresponds to model (1) in its simplest form. Define

$$C_\beta = \left((2\pi)^{2\beta} 2\beta/[(\beta+1)(2\beta+1)]\right)^{\beta/(2\beta+1)}$$

and let

$$\varphi_n^\star(\lambda,v) = \left(1 - |\lambda|^\beta C_\beta (v\rho)^{\beta/(1+2\beta)} n^{-\beta/(1+2\beta)}\right)_+ \tag{23}$$

where $x_+ = \max(0, x)$. Let

$$P^\star_\beta = (2\beta+1)^{1/(2\beta+1)}\Big(\frac{\beta}{\pi(\beta+1)}\Big)^{2\beta/(2\beta+1)}. \tag{24}$$

Then, Pinsker's result takes the remarkable form

Proposition 1. *(Pinsker, 1980)*

$$\lim_{n\to\infty} n^{2\beta/(2\beta+1)} \inf_F \mathbf{R}_n(F, \widetilde{\mathbf{W}}(\beta,\rho)) = \frac{P^\star_\beta}{\rho^{1/(2\beta+1)}}\, v^{2\beta/(2\beta+1)} \tag{25}$$

where the infimum is taken over all estimators. Moreover, this bound is attained by the linear estimator $F^\star_n(v,t) = \sum_{\lambda\in\mathbb{Z}} \varphi^\star_n(\lambda, v) Y_\lambda\, e^{-i2\pi\lambda t}$.

The proof of the lower bound can be found in Pinsker (1980), or Nussbaum (1985) in a dense form. See also the review paper of Nussbaum (1999) about the Pinsker bound. The upper bound can be readily checked from (23) together with the decomposition (28) below.

3.1.2. *The random case for linear estimators.* In the case of a random V, a first guess to achieve the optimal bound is to consider the *conditional Pinsker estimator*, defined by the plug-in rule

$$F^\star_n(V,t) = \sum_{\lambda\in\mathbb{Z}} \varphi^\star_n(\lambda, V) Y_\lambda\, e^{-i2\pi\lambda t}. \tag{26}$$

If $E(V) < +\infty$, the estimator $F^\star_n(V,\cdot)$ certainly achieves the bound

$$\frac{P^\star_\beta}{\rho^{1/(2\beta+1)}}\, E\left(V^{2\beta/(2\beta+1)}\right) \tag{27}$$

which is simply obtained from the deterministic case by considering $F^\star_n(v,\cdot)$ for the minimax risk conditional on $V = v$ and by integrating with respect to the law of V, see Section 2.4 in the Introduction. Surprisingly, it turns out that this procedure is not optimal if the whole parameter space is $\widetilde{\mathbf{W}}(\beta,\rho)$ and that the bound (27) can be improved. For an arbitrary linear estimator F_n constructed through the linear filter $(\varphi_n(\lambda, V))_{\lambda\in\mathbb{Z}}$, the usual bias-variance of the risk, conditional on V, reads

$$E^n_f\Big[\int_0^1 [F_n(t) - f(t)]^2\, dt \,\Big|\, V\Big] =$$
$$= \sum_{\lambda\in\mathbb{Z}} \big|1 - \varphi_n(\lambda, V)\big|^2\, |\hat f(\lambda)|^2 + \frac{V}{n}\sum_{\lambda\in\mathbb{Z}} |\varphi_n(\lambda, V)|^2. \tag{28}$$

In order to minimize this risk, it suffices to consider the filters satisfying $\varphi_n(\lambda, v) = \varphi_n(-\lambda, v)$ for all λ (for which the corresponding estimator F_n is real valued). Since there is no constraint on $\hat f(0)$, we have $\mathbf{R}_n(F_n) = +\infty$ if $\varphi_n(0, V) \neq 1$. Hence we set $\varphi_n(0, V) = 1$ and the worst risk becomes

$$\mathbf{R}_n(F_n) = \frac{1}{n}E[V] + \frac{1}{(2\pi)^{2\beta}\rho}\, S^2_{F_n} + \frac{2}{n}\sum_{\lambda\geq 1} E\Big[V|\varphi_n(\lambda, V)|^2\Big]$$

where $S_{F_n}^2 = \sup_{\lambda \geq 1} \lambda^{-2\beta} E\Big[\big|1-\varphi_n(\lambda,V)\big|^2\Big]$. (If $E(V)=+\infty$, the above equality is still meaningful but of no interest: this will lead us to impose restrictions on $\mathcal{L}(V)$ in Theorem 1 below.) Given $\tau>0$, our task is to minimize each term in the above sum in φ_n, under the constraint $\sup_\lambda |\lambda|^{-2\beta} E\Big[\big|1-\varphi_n(\lambda,V)\big|^2\Big] \leq \tau$, and then minimize in τ. For this, we need the following fundamental lemma, which proof is delayed until Section 5.

Lemma 1. *Let V be a positive random variable.*

(1) *For all $x \in (0,1)$, for all random variable B such that $E(B^2) \leq x^2$, we have*

$$E\Big[V(1-B)^2\Big] \geq G(x) := E\Big[V(1-B_x^\star)^2\Big], \tag{29}$$

$$\textit{with} \quad B_x^\star = \frac{xV}{xV + a^\star(x)} \tag{30}$$

where $a^\star(x)$ is the unique positive number such that $E[B_x^{\star 2}] = x^2$. Moreover, equality holds in (29) iff $B = B_x^\star$ a.s..

(2) *The functions G and $a^\star$ are decreasing, differentiable and $G'(x) = -2a^\star(x)$.*

This enables us to define the following estimator. Let

$$\varphi_n^{\star\star}(\lambda,V) = \begin{cases} 1 - B^\star_{|\lambda|^\beta S_n} & \text{if } |\lambda|^\beta S_n < 1 \\ 0 & \text{if } |\lambda|^\beta S_n \geq 1 \end{cases} \tag{31}$$

where $S_n = K/n^{\beta/(2\beta+1)}$ for a constant K specified in (33) below, so that $S_{F_n^\star} = S_n$. Put

$$F_n^{\star\star}(t) = \sum_{\lambda \in \mathbb{Z}} \varphi_n^{\star\star}(\lambda,V) Y_\lambda e^{-i2\pi\lambda t}.$$

Note that the new procedure $F_n^{\star\star}(t)$ depends simultaneously on the realization $V(\omega)$ and on $\mathcal{L}(V)$, through $B^\star_{|\lambda|^\beta S_n}(\omega)$ which is determined by $V(\omega)$ and $a^\star$ (see 1. of Lemma 1 above), whereas the conditional Pinsker estimator $F_n^\star(V(\omega),t)$ only depends on V through its value $V(\omega)$. We obtain

$$\begin{aligned}
\mathbf{R}_n(F_n^{\star\star}) &= \frac{1}{n}E[V] + \frac{1}{(2\pi)^{2\beta}\rho} S_n^2 + (2/n) \sum_{1 \leq \lambda < S_n^{-1/\beta}} G(\lambda^\beta S_n), \\
&= \frac{1}{n}E[V] + \frac{1}{(2\pi)^{2\beta}\rho} S_n^2 + (2/nS_n^{1/\beta}) S_n^{1/\beta} \sum_{1 \leq \lambda < S_n^{-1/\beta}} G((\lambda S_n^{1/\beta})^\beta), \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (32) \\
&= \frac{1}{n}E[V] + n^{-2\beta/(2\beta+1)} \Big[\frac{K^2}{\rho(2\pi)^{2\beta}} + \frac{2}{K^{1/\beta}} \int_0^1 G(x^\beta)\,dx + o(1)\Big].
\end{aligned}$$

Since G is decreasing, the convergence of the above Riemann sum holds. Minimizing with respect to K yields

$$K = \left(\beta^{-1}\rho(2\pi)^{2\beta}\int_0^1 G(x^\beta)\,dx\right)^{\beta/(2\beta+1)} \tag{33}$$

thus

$$\lim_{n\to\infty} n^{2\beta/(2\beta+1)}\mathbf{R}_n(F_n^{\star\star}) = \frac{2\beta+1}{\rho^{1/(2\beta+1)}(2\pi\beta)^{2\beta/(2\beta+1)}}\left(\int_0^1 G(x^\beta)\,dx\right)^{2\beta/(2\beta+1)}. \tag{34}$$

In view of 2 of Lemma 1, integration by parts yields for all $\varepsilon \in (0,1)$:

$$\begin{aligned}\int_\varepsilon^1 G(x^\beta)\,dx &= 2\beta\int_\varepsilon^1 a^\star(x^\beta)x^\beta\,dx - \varepsilon G(\varepsilon^\beta),\\ &\geq 2\beta\int_\varepsilon^1 a^\star(x^\beta)x^\beta\,dx - \int_0^\varepsilon G(x^\beta)\,dx.\end{aligned}$$

Therefore

$$\int_0^1 G(x^\beta)\,dx = 2\beta\int_0^1 a^\star(x^\beta)x^\beta\,dx = 2\int_0^1 a^\star(x)x^{1/\beta}dx. \tag{35}$$

Remark 1. Remark that, since G decreases, this equality alway holds whatever the law of V is (even if $G(0) = E[V] = +\infty$), and for the same reason, the Riemann sum in (32) always converges to $\int_0^1 G(x^\beta)\,dx$, two facts that we will use later.

We summarize our result in the following theorem.

Theorem 1. *Suppose that $E(V) < +\infty$. Consider the worst risk over $\widetilde{\mathbf{W}}(\beta,\rho)$ defined by (22).*

(1) *Among the class of linear estimators, $F_n^{\star\star}$ defined by (31) is minimax and one has*

$$\lim_{n\to\infty} n^{\frac{2\beta}{2\beta+1}}\mathbf{R}_n(F_n^{\star\star}, \widetilde{\mathbf{W}}(\beta,\rho)) = P^\star(\beta,\rho,\mathcal{L}(V)) \tag{36}$$

with

$$P^\star(\beta,\rho,\mathcal{L}(V)) = \frac{P_\beta^\star}{\rho^{1/(2\beta+1)}}\left(\tfrac{(\beta+1)(2\beta+1)}{\beta^2}\int_0^1 a^\star(x)x^{1/\beta}\,dx\right)^{\frac{2\beta}{2\beta+1}}, \tag{37}$$

where $P_\beta^\star$ is given by (24) and $a^\star(x)$ is such that

$$E\left[\frac{V^2}{(xV + a^\star(x))^2}\right] = 1. \tag{38}$$

(2) *If the law of V is non degenerate and $\int_0^1 a^\star(x)x^{1/\beta}\,dx < \infty$, we have*

$$\lim_{n\to\infty}\frac{\mathbf{R}_n(F_n^{\star\star}(\cdot), \widetilde{\mathbf{W}}(\beta,\rho))}{\mathbf{R}_n(F_n^\star(V,\cdot), \widetilde{\mathbf{W}}(\beta,\rho))} < 1,$$

where $F_n^\star(V,\cdot)$ is the conditional Pinsker estimator defined in (26).

Remark 2. If $\mathcal{L}(V) = \delta_v$ then $a^\star(x) = v(1-x)$ and since $\int_0^1 (1-x)x^{1/\beta}\,dx = \beta^2/[(\beta+1)(2\beta+1)]$, we find back Proposition 1, i.e. Pinsker's result.

Remark 3. Unless $\mathcal{L}(V) = \delta_v$, we see from 2. that the optimal linear filter must depend on V, not only through its realization $V(\omega)$ but also through its law $\mathcal{L}(V)$. The influence of $\mathcal{L}(V)$ is characterized by the function $a^\star$.

The proof of (36) follows from (34) and (35). The proof of 2. is a consequence of the following proposition.

Proposition 2. *We have*

$$\int_0^1 a^\star(x)x^{1/\beta}\,dx \le \tfrac{\beta^2}{(\beta+1)(2\beta+1)} E(V^{2\beta/(2\beta+1)})^{(2\beta+1)/2\beta}. \tag{39}$$

Moreover the equality occurs if and only if V is deterministic or both sides are infinite.

Proof. Without loss of generality, we may assume that $E(V^{2\beta/(2\beta+1)}) < \infty$, otherwise (39) is proved. Let $D = V^{\beta/(2\beta+1)}/E(V^{2\beta/(2\beta+1)})^{1/2}$ and $B_x = Dx\mathbb{I}_{\{Dx\le 1\}}$. Since $E(B_x^2) \le x^2$, Lemma 1 implies that $G(x) \le E\big[V(1-B_x)^2\big]$. Thus

$$\begin{aligned}\int_0^1 G(x)x^{-1+1/\beta}dx &\le E\Big(V\int_0^{1/D}(1-Dx)^2x^{-1+1/\beta}dx\Big)\\ &= E\Big[\frac{V}{D^{1/\beta}}\Big]\int_0^1(1-x)^2x^{-1+1/\beta}dx\\ &= E(V^{2\beta/(2\beta+1)})^{(2\beta+1)/2\beta}\frac{2\beta^3}{(\beta+1)(2\beta+1)}.\end{aligned}$$

and inequality (39) follows from (35).

Assume that equality holds in (39). Set $\widetilde{G}(x) = E\big[V(1-B_x)^2\big]$. By (29) of Lemma 1, we have $G(x) \le \widetilde{G}(x)$, hence, by continuity, we have $G(x) = \widetilde{G}(x)$ everywhere. By 1 of Lemma 1, it follows that $B_x = B_x^\star$ a.s.. In particular, $D\mathbb{I}_{\{D\le 1\}} = B_1 = B_1^\star = 1$ a.s., hence $D = 1$ a.s. and the proposition is proved. □

3.1.3. *The case where V has not a moment of order* 1. The restriction $E(V) < +\infty$ in Theorem 1 is inessential if we assume that the parameter f is bounded by some constant M. This is a classical hypothesis in nonparametric estimation that we shall make from now on. So, we further restrict the parameter space to

$$\widetilde{\mathbf{W}}(\beta,\rho,M) = \Big\{f \in \widetilde{\mathbf{W}}(\beta,\rho) : \sup_t |f(t)| \le M\Big\}.$$

In order to build an asymptotically minimax estimator, we only need to modify the value of $\varphi_n^{\star\star}(0,V)$ in the preceding construction. Define $\varphi_n^{\star\star}(\lambda,V)$ by (31) for $\lambda \neq 0$ and $\varphi_n^{\star\star}(0,V) = 1 - B_y^\star$ where $y = y(n,\mathcal{L}(V))$ minimizes $y^2 + G(y)/n$. (In view of Lemma 1 such a y does exist.)

Theorem 2. *Suppose that $\int_0^1 a^\star(x)x^{1/\beta}dx < +\infty$. Consider $M > 0$ and the worst risk (22) with $\widetilde{\mathbf{W}}(\beta,\rho,M)$ in place of $\widetilde{\mathbf{W}}(\beta,\rho)$. Let $F_n^{\star\star}$ be the estimator associated to $\varphi_n^{\star\star}$ with the modification of $\varphi_n^{\star\star}(0,V)$ defined above. Then $F_n^{\star\star}$ is asymptotically minimax among linear estimators and we have*

$$\lim_{n\to\infty} n^{\frac{2\beta}{2\beta+1}}\mathbf{R}_n(F_n^{\star\star},\widetilde{\mathbf{W}}(\beta,\rho,M)) = P^\star(\beta,\rho,\mathcal{L}(V))$$

where $P^\star(\beta,\rho,\mathcal{L}(V))$ is defined in Theorem 1.

Proof. In view of the proof of Theorem 1, we have

$$\begin{aligned}\mathbf{R}_n(F_n^{\star\star},\widetilde{\mathbf{W}}(\beta,\rho,M)) &\le M^2\, E\Big[\big|1-\varphi_n(0,V)\big|^2\Big]+(1/n)E\Big[V|\varphi_n(0,V)|^2\Big]+\\ &\quad + n^{-2\beta/(2\beta+1)}(P^\star(\beta,\rho,\mathcal{L}(V))+o(1)),\\ &= M^2y(n)^2+\frac{1}{n}G(y(n))+n^{-2\beta/(2\beta+1)}(P^\star(\beta,\rho,\mathcal{L}(V))+o(1))\end{aligned}$$

where $y(n) = \mathrm{argmin}_x(x^2+G(x)/n)$. Moreover we have $\varepsilon G(\varepsilon^\beta) \le \int_0^\varepsilon G(x^\beta)dx \to 0$ as $\varepsilon \to 0$ since G is decreasing and $\int_0^1 G(x^\beta)dx = 2\int_0^1 a^\star(x)x^{1/\beta}dx < +\infty$. It follows that for all $\eta > 0$:

$$\begin{aligned}M^2y(n)^2+\frac{1}{n}G(y(n)) &\le n^{-2\beta/(2\beta+1)}\Big(M^2\eta^2+n^{-1/(2\beta+1)}G(\eta n^{-\beta/(2\beta+1)})\Big),\\ &= M^2\eta^2 O(n^{-2\beta/(2\beta+1)}).\end{aligned}$$

This proves the upper bound.

Now we prove the lower bound. Consider an arbitrary filter φ_n. Since the function

$$t \mapsto \frac{1}{\big(\lambda^{2\beta}2\rho(2\pi)^{2\beta}\big)^{1/2}}\big(e^{i2\pi\lambda t}+e^{-i2\pi\lambda t}\big)$$

belongs to $\widetilde{\mathbf{W}}(\beta,\rho,M)$ for $\lambda \ge \lambda_0 := \big(\frac{1}{2}M^2\rho(2\pi)^{2\beta}\big)^{1/2\beta}$, we have

$$\mathbf{R}_n(F_n) \ge \frac{1}{(2\pi)^{2\beta}\rho}\,S_{F_n}^2 + \frac{2}{n}\sum_{\lambda\ge 1}E\Big[V|\varphi_n(\lambda,V)|^2\Big]$$

where $S_{F_n}^2 = \sup_{\lambda\ge\lambda_0}\lambda^{-2\beta}E\Big[\big|1-\varphi_n(\lambda,V)\big|^2\Big]$. As in the proof of Theorem 1, minimizing the right-hand side in φ and taking the limit as $n\to\infty$, yields the lower bound. □

3.1.4. *Optimality of linear estimators.* Our next result shows that indeed, $F_n^{\star\star}$ is asymptotically optimal not only among linear filters but among all estimators. Therefore, the constant $P^\star(\beta, \rho, \mathcal{L}(V))$ is optimal.

Theorem 3. *For all $M > 0$ we have*

$$\lim_{n\to\infty} n^{\frac{2\beta}{2\beta+1}} \inf_F \mathbf{R}_n(F, \widetilde{\mathbf{W}}(\beta, \rho, M)) = P^\star(\beta, \rho, \mathcal{L}(V))$$

where $P^\star(\beta, \rho, \mathcal{L}(V))$ is defined in Theorem 1 and the infimum is taken over all estimators.

The proof of the lower bound is delayed until Section 5.

We end this section by sharpening a little the analysis of the optimal bound versus the performance of the conditional Pinsker estimator $F_n^\star(V, \cdot)$. This can be measured by the ratio

$$P^\star(\beta, \rho, \mathcal{L}(V)) / \int \mathcal{L}(V)(dv) P^\star(\beta, \rho, \delta_v),$$

as follows from (27) and Theorem 3. Clearly, this ratio is invariant under dilation $V \mapsto aV$, so, in some sense, it only involves the "shape" of the distribution $\mathcal{L}(V)$. Actually, one can check that for $v_0 > 1$, the choice

$$\mathcal{L}(V)(dv) = C_{\beta, v_0} \, \mathbb{I}_{\{v > v_0\}} \frac{1}{v^{1+2\beta/(2\beta+1)} \log(v)} dv$$

makes the above ratio to be 0. Restricting the above distribution to compacts (up to a suitable normalization), it can still be made arbitrarily close to 0. However, we have the following limitation:

Proposition 3. *For all $\varepsilon, \varepsilon' > 0$, we have*

$$\left(\int_0^1 a^\star(x) x^{1/\beta} \, dx \right)^{2\beta/(2\beta+1)} \geq$$
$$\geq C_{\beta,\varepsilon,\varepsilon'} E\Big[V^{\varepsilon + 2\beta/(2\beta+1)} \mathbb{I}_{\{V \leq 1\}} + V^{-\varepsilon' + 2\beta/(2\beta+1)} \mathbb{I}_{\{V \geq 1\}} \Big]$$

where $C_{\beta,\varepsilon,\varepsilon'}$ is a positive constant which does not depend on $\mathcal{L}(V)$.

Proof. The function $\psi : (0, +\infty) \to (0, 1)$, $z \mapsto \psi(z) = E\Big[V^2/(V+z)^2\Big]$ is a diffeomorphism of class C^∞. Therefore

$$\int_0^1 a^\star(x) x^{1/\beta} \, dx = \int_0^1 \psi^{-1}(x^2) x^{1+1/\beta} \, dx = \int_0^\infty dz \int_0^{\sqrt{\psi(z)}} x^{1+1/\beta} \, dx$$
$$= \tfrac{2\beta}{2\beta+1} \int_0^\infty \psi(z)^{(2\beta+1)/(2\beta)} \, dz.$$

Set $\alpha = 2\beta/(2\beta+1)$. We have for $0 < q < 1 < q' < \infty$:

$$\int_0^\infty \psi(z)^{1/\alpha} dz =$$

$$= \int_0^1 \frac{2}{(1-q)z^q} \Big((\tfrac{1-q}{2})^\alpha z^{q\alpha} \psi(z) \Big)^{1/\alpha} + \int_1^\infty \frac{2}{(q'-1)z^{q'}} \Big((\tfrac{q'-1}{2})^\alpha z^{q'\alpha} \psi(z) \Big)^{1/\alpha},$$

$$\geq \left[\int_0^1 \frac{2}{(1-q)} (\tfrac{1-q}{2})^\alpha z^{q\alpha - q} \psi(z) + \int_1^\infty \frac{2}{(q'-1)} (\tfrac{q'-1}{2})^\alpha z^{q'\alpha - q'} \psi(z) \right]^{1/\alpha}$$

by Jensen's inequality since $\alpha < 1$. We thus need a lower bound for $\int_0^1 z^{q\alpha - q} \psi(z)\, dz$ and $\int_1^\infty z^{q'\alpha - q'} \psi(z)\, dz$.

$$\begin{aligned} \int_0^1 z^{q\alpha-q}\psi(z)\, dz &= E\Big[V^2 \int_0^1 \frac{z^{q\alpha-q}}{(V+z)^2} dz \Big], \\ &= E\Big[V^{1+q\alpha-q} \int_0^{1/V} \frac{y^{q\alpha-q}}{(1+y)^2} dy \Big], \\ &\geq \int_0^1 \frac{y^{q\alpha-q}}{(1+y)^2} dy\, E\Big[V^{\alpha+(1-q)(1-\alpha)} \mathbb{I}_{\{V\leq 1\}} \Big]. \end{aligned}$$

Likewise we have

$$\int_1^\infty z^{q'\alpha-q'}\psi(z)\, dz \geq \int_1^\infty \frac{y^{q'\alpha-q'}}{(1+y)^2} dy\, E\Big[V^{\alpha+(1-q')(1-\alpha)} \mathbb{I}_{\{V\geq 1\}} \Big].$$

□

3.2. **The noncircular case.** We will later restrict the parameter space to

$$\mathbf{W}(\beta, \rho, M) = \left\{ f \in \mathbf{W}(\beta, \rho) \;:\; \sup_t |f(t)| \leq M \right\}.$$

We also need here the following integrability property of V:

Assumption A. For some $\delta > 0$, we have $E\left(V^{\frac{2\beta}{2\beta+1}+\delta} \right) < \infty$. □

3.2.1. *Translation into sequence space.* We exploit the periodic case by noting that the class $\mathbf{W}(\beta, \rho)$ can be expressed as an ellipsoid up to boundary modifications. More precisely, we have

$$\widehat{f}(\lambda) = -\frac{1}{i2\pi\lambda} \widehat{f'}(\lambda) + \frac{f(1) - f(0)}{i2\pi\lambda},$$

and more generally

$$\hat{f}(\lambda) = \frac{\widehat{f^{(\beta)}}(\lambda)}{(-i2\pi\lambda)^\beta} + \sum_{\ell=0}^{\beta-1} \frac{\delta_f^\ell}{(i2\pi\lambda)^{\ell+1}}, \tag{40}$$

$$\text{with} \quad \delta_f^\ell = (-1)^\ell \big[f^{(\ell)}(1) - f^{(\ell)}(0) \big]. \tag{41}$$

Define $q_0(f) = 0$ and for $\lambda \neq 0$:

$$q_\lambda(f) = \sum_{\ell=0}^{\beta-1} \frac{\delta_f^\ell}{(i2\pi\lambda)^{\ell+1}}.$$

We thus have the following characterization of the class $\mathbf{W}(\beta,\rho)$ as a sequence space, namely

$$\mathbf{W}(\beta,\rho) = \Big\{ f \in C^{\beta-1}([0,1]) \ : \ \sum_{\lambda\neq 0} |2\pi\lambda|^{2\beta} \big|\hat{f}(\lambda) - q_\lambda(f)\big|^2 \leq \rho^{-1} \Big\}.$$

Put $Q_0^n = 0$ and for $\lambda \neq 0$, $Q_\lambda^n = \sum_{\ell=0}^{\beta-1} \frac{\Delta^{\ell,n}}{(i2\pi\lambda)^{\ell+1}}$ where $\Delta^{\ell,n}$ is an estimator of δ_f^ℓ to be specified later. We consider estimates of $\hat{f}(\lambda)$ of the form

$$\varphi_n^{\star\star}(\lambda, V)(Y_\lambda - Q_\lambda^n) + Q_\lambda^n$$

where (recall Section 3.1) $Y_\lambda = \int_0^1 e^{i2\pi\lambda t} dX_t$ is the empirical Fourier coefficient. We now show that estimating δ_f^ℓ properly enables us to attain the bound of the periodic case.

3.2.2. *Construction of $\Delta^{\ell,n}$.* Pick up a kernel $K : [0,1] \to \mathbb{R}$ of class C^β satisfying

(i) $K^{(\ell)}(0) = K^{(\ell)}(1) = 0$ for $\ell = 0 \ldots \beta - 1$.
(ii) $\int_0^1 K(t)\, dt = 1$
(iii) $\int_0^1 t^\ell\, K(t)\, dt = 0$ for $\ell = 1, \ldots, \beta - 1$,

a choice which is obviously possible. Introduce the random bandwidth

$$H_n = (V/n)^{1/2\beta}.$$

Put

$$\Delta^{\ell,n} = \int_0^1 \Big[(-1)^\ell K_{H_n}^{(\ell)}(1-t) - K_{H_n}^{(\ell)}(t) \Big] dX_t \tag{42}$$

where we denote $K_h^{(\ell)}(t) = h^{-1} \frac{\partial^\ell}{\partial t^\ell} K(t/h)$. Finally, define

$$F_n^{\star\star}(t) = \sum_{\lambda\in\mathbb{Z}} \big\{ \varphi_n^{\star\star}(\lambda, V)(Y_\lambda - Q_\lambda^n) + Q_\lambda^n \big\} e^{-i2\pi\lambda t}. \tag{43}$$

Theorem 4. *For all $M > 0$, we have*

$$\lim_{n\to\infty} n^{\frac{2\beta}{2\beta+1}} \mathbf{R}_n(F_n^{\star\star}, \mathbf{W}(\beta,\rho,M)) = P^\star(\beta,\rho,\mathcal{L}(V))$$

where $P^\star(\beta,\rho,\mathcal{L}(V))$ is defined in Theorem 1.

Moreover, since the lower bound over $\widetilde{\mathbf{W}}(\beta,\rho,M)$ is also a lower bound over $\mathbf{W}(\beta,\rho,M)$, Theorem 3 shows that $F_n^{\star\star}$ is optimal for the risk $\mathbf{R}_n$.

Proof. In the following $C = C_{\beta,\rho}$ will denote a generic constant, possibly varying from line to line. We first need this auxiliary result:

Lemma 2. *Assume that $E(V^q) < \infty$. Let $0 \le \ell \le \beta - 1$. For all $p \ge 0$ such that $p(\beta-\ell-1/2)/2\beta \le q$, we have*

$$\sup_{f \in \mathbf{W}(\beta,\rho)} E_f^n\big[|\Delta^{\ell,n} - \delta_f^\ell|^p\big] \le C_p\, n^{-p(\beta-\ell-1/2)/2\beta}.$$

Proof. By classical kernel approximation

$$E_f^n\big[|\Delta^{\ell,n} - \delta_f^\ell|^p \mid V\big] \le C_p\Big(H_n^{p(\beta-\ell-1/2)} + (V/nH_n^{2\ell+1})^{p/2}\Big).$$

Hence

$$E_f^n\big[|\Delta^{\ell,n} - \delta_f^\ell|^p\big] \le C_p\, n^{-p(\beta-\ell-1/2)/2\beta} E[V^{p(\beta-\ell-1/2)/2\beta}]$$

and the result follows from the integrability property of V. □

In view of Section 3.1.3 and the translation of $\mathbf{W}(\beta,\rho)$ in terms of sequence space we easily derive that

$$\lim_{n\to\infty} n^{\frac{2\beta}{2\beta+1}} \sup_{f\in\mathbf{W}(\beta,\rho,M)} \sum_{\lambda\in\mathbb{Z}} E_f^n\Big[\big|\varphi_n^{\star\star}(\lambda,V)(Y_\lambda - q_\lambda(f)) - \hat f(\lambda) + q_\lambda(f)\big|^2\Big] = $$
$$= P^\star(\beta,\rho,\mathcal{L}(V)).$$

Moreover we have

$$\Phi_n(\lambda,V) - \hat f(\lambda) = \varphi_n^{\star\star}(\lambda,V)\big\{Y_\lambda - q_\lambda(f)\big\} - \hat f(\lambda) + q_\lambda(f) +$$
$$+ \big\{\varphi_n^{\star\star}(\lambda,V) - 1\big\}\big\{q_\lambda(f) - Q_\lambda^n\big\}.$$

Therefore, since

$$\Big|\sum_\lambda |a_\lambda + b_\lambda|^2 - \sum_\lambda |a_\lambda|^2 - \sum_\lambda |b_\lambda|^2\Big| \le 2\Big(\sum_\lambda |a_\lambda|^2 \sum_\lambda |b_\lambda|^2\Big)^{1/2},$$

in order to prove Theorem 4 it suffices to prove that

$$\sup_{f \in \mathbf{W}(\beta,\rho)} \sum_{\lambda\neq 0} E_f^n\Big[|1-\varphi_n^{\star\star}(\lambda,V)|^2\big|Q_\lambda^n - q_\lambda(f)\big|^2\Big] = o(n^{-2\beta/(2\beta+1)}). \tag{44}$$

Clearly $\big|Q_\lambda^n - q_\lambda(f)\big|^2 \le C\sum_{\ell=0}^{\beta-1} \frac{|\Delta^{\ell,n} - \delta_f^\ell|^2}{|\lambda|^{2(\ell+1)}}$. Let $\ell \in \{0,\dots,\beta-1\}$. Since $\ell + 1/2 < \beta$ we can pick $p \in (0,2)$ such that

$$2\ell + \tfrac{\ell+1/2}{\beta} < p\beta < 2\ell + 1,$$
$$p\beta < 2\ell + \tfrac{\ell+1/2}{\beta} + 2(\beta - \ell - \tfrac12)\big(\tfrac{1}{2\beta/(2\beta+1)} - \tfrac{1}{2\beta/(2\beta+1)+\delta}\big).$$

The last inequality is equivalent to $q := \frac{4}{2-p}\frac{\beta-\ell-1/2}{2\beta} < \frac{2\beta}{2\beta+1} + \delta$, hence we have $E(V^q) < +\infty$ thanks to Assumption A. Using the fact that

$|1-\varphi_n^{\star\star}(\lambda,V)|^p \geq |1-\varphi_n^{\star\star}(\lambda,V)|^2$ since $0 \leq \varphi_n^{\star\star}(\lambda,V) \leq 1$, Hölder inequality yields

$$\begin{aligned}
&E_f\Big[|1-\varphi_n^{\star\star}(\lambda,V)|^2\big|\Delta^{\ell,n}-\delta_f^\ell\big|^2\Big]\\
&\leq \Big(E_f^n\Big[|1-\varphi_n^{\star\star}(\lambda,V)|^2\Big]\Big)^{p/2}\Big(E_f^n\Big[\big|\Delta^{\ell,n}-\delta_f^\ell\big|^{4/(2-p)}\Big]\Big)^{1-p/2},\\
&\leq Cn^{-2(\beta-\ell-1/2)/2\beta}\Big(E_f^n\Big[|1-\varphi_n^{\star\star}(\lambda,V)|^2\Big]\Big)^{p/2} \quad \text{by Lemma 2},\\
&\leq Cn^{-(\beta-\ell-1/2)/\beta}\Big(1\wedge\big\{K^p n^{-p\beta/(2\beta+1)}|\lambda|^{p\beta}\big\}\Big)
\end{aligned}$$

where we used that $E\big[|1-\varphi_n^{\star\star}(\lambda,V)|^2\big] \leq 1\wedge\big\{K^2 n^{-2\beta/(2\beta+1)}|\lambda|^{2\beta}\big\}$ for some constant K. It follows that

$$\begin{aligned}
&\sum_{\lambda\neq 0} E_f^n\left[|1-\varphi_n^{\star\star}(\lambda,V)|^2\frac{|\Delta^{\ell,n}-\delta_f^\ell|^2}{|\lambda|^{2(\ell+1)}}\right]\\
&\leq Cn^{-(\beta-\ell-1/2)/\beta}\Big(n^{-p\beta/(2\beta+1)}K^p\sum_{1\leq\lambda\leq n^{1/(2\beta+1)}/K^{1/\beta}}\lambda^{p\beta-2\ell-2}+\sum_{\lambda>n^{1/(2\beta+1)}/K^{1/\beta}}\lambda^{-2\ell-2}\Big),\\
&\leq CK^{(2\ell+1)/\beta}n^{-(\beta-\ell-1/2)/\beta}\big(n^{-p\beta/(2\beta+1)}+n^{-(2\ell+1)/(2\beta+1)}\big)
\end{aligned}$$

since $p\beta-2\ell-2<-1$. This yields (44) because of

$$\frac{2\ell+1}{2\beta+1}+\frac{\beta-\ell-1/2}{\beta} > \frac{p\beta}{2\beta+1}+\frac{\beta-\ell-1/2}{\beta} > \frac{2\beta}{2\beta+1}.$$

□

3.3. **The general homogeneous case.** We can now relax the restrictions of Section 3.1 and solve the Pinsker problem in the case (21) and when $[a,b]\subset\mathbb{R}$ is an arbitrary compact interval. Thus the worst functional risk over $\mathbf{W}(\beta,\rho,M)$ reads

$$\mathbf{R}_n(F,\mathbf{W}(\beta,\rho,M))[\Phi] = \\ = \sup_{f\in\mathbf{W}(\beta,\rho,M)} E_f^n\left[\Phi(V)\int_a^b\big|F(t)-f(t)\big|^2dt\right].$$

If it exists, let $a_\Phi^\star(x)$ be the solution of

$$E\left[\frac{\Phi(V)V^2}{(xV+a_\Phi^\star(x))^2}\right] = E(\Phi(V)). \tag{45}$$

We thus have a new definition of $a^\star$ which is simply obtained from (38) by the change of probability $P_{\text{old}}\mapsto P_{\text{new}} := \Phi(V)\cdot P/E(\Phi(V))$. Assumption A becomes in this more general setting:

Assumption B. We have

$0<E(\Phi(V))<\infty$ and $E\left(\Phi(V)V^{\frac{2\beta}{2\beta+1}+\delta}\right)<\infty$ for some $\delta>0$. □

Let $T_{a,b}$ denote the increasing affine transform mapping $[0,1]$ onto $[a,b]$ and let $F^{\star\star}_{n,a,b}$ be the estimator defined by

$$F^{\star\star}_{n,a,b}\big[(X_s)_{s\in[a,b]},t\big] = \frac{1}{\sqrt{b-a}}F^{\star\star}_n\Big[\Big(\frac{X_{T_{a,b}(s)}}{\sqrt{b-a}}\Big)_{s\in[0,1]}, T^{-1}_{a,b}(t)\Big] \quad \forall t\in[a,b]$$

where $F^{\star\star}_n$ is the linear estimator on $[0,1]$ defined by (43), the filter $(\varphi^{\star\star}_n(\lambda,V))_{\lambda\in\mathbb{Z}}$ is constructed using the definition of $a^\star_\Phi$ given by (45) and $\rho(b-a)^{-2\beta}$ in place of ρ. From Theorems 3 and 4, using the same arguments under the measure P_{new} together with the affine transformation $T_{a,b}$, we readily obtain the following bound.

Theorem 5. *Grant Assumption B. For all $M>0$, we have*

$$\lim_{n\to\infty} n^{\frac{2\beta}{2\beta+1}} \inf_F \mathbf{R}_n(F,\mathbf{W}(\beta,\rho,M))[\Phi] = P^\star(\beta,\rho,\mathcal{L}(V))[\Phi]$$

where

$$P^\star(\beta,\rho,\mathcal{L}(V))[\Phi] = \frac{P^\star_\beta}{\rho^{1/(2\beta+1)}}E[\Phi(V)]\bigg((b-a)^{\frac{(\beta+1)(2\beta+1)}{\beta^2}}\int_0^1 a^\star_\Phi(x)x^{1/\beta}\,dx\bigg)^{\frac{2\beta}{2\beta+1}}, \quad (46)$$

the constant $P^\star_\beta$ is given by (24), $a^\star_\Phi(x)$ is is defined in (45) and the infimum is taken over all estimators. Moreover, the bound (46) is attained by $F^{\star\star}_{n,a,b}$.

4. The general case

We are ready to deal with the general model (9) and solve the Pinsker problem as stated in the Introduction. For notational simplicity , we will use the notation V for the whole process $(V_t)_{t\in[a,b]}$ in this Section. The assumptions are the following.

Assumption C. We have for some $\delta>0$:

$$E\big(\sup_{t\in[a,b]}\Phi(t,V)\big)<\infty \quad\text{and}\quad E\bigg(\sup_{t\in[a,b]}\Phi(t,V)\sup_{t\in[a,b]}V_t^{\frac{2\beta}{2\beta+1}+\delta}\bigg)<\infty.$$

Moreover, the processes V and $\Phi(\cdot,V)$ have almost surely continuous paths and the weight function $t\to\rho(t)$ is continuous and positive. □

We exploit the result of Section 3.3. using a localization argument. We need some notation. Let $N\geq 1$ be an integer. Introduce an equispaced grid on $[a,b]$ of mesh $\delta_N=(b-a)/N$ defined by $t^N_k=a+k\delta_N$ for $k=1,\dots,N$. Recall from Section 3.1.2. that the optimal estimator defined for the periodic case $F^{\star\star}_n=F^{\star\star}_n(K)$ depends on a parameter $K=S_n n^{\beta/(2\beta+1)}$ introduced below in (31) and which was calibrated by

(33). Likewise, $F^{\star\star}_{a,b} = F^{\star\star}_{a,b}(K)$ in Section 3.3. Take now K as a free parameter and define

$$\mathbb{F}_{n,N}(t) = F^{\star\star}_{t^N_{k-1},t^N_k}(K^N_k)(t) \quad \text{if} \quad t^N_{k-1} \leq t < t^N_k, \tag{47}$$

where $(K^N_k, k = 1, \ldots, N)$ is a sequence of positive numbers to be chosen properly. Indeed, it suffices to evaluate the minimax risk on each $[t^N_{k-1}, t^N_k)$ and optimize in K^N_k. Recall that

$$\mathbf{R}_n(F, \mathbf{W}(\beta, \rho, M))[\Phi] =$$

$$= \sup_{f \in \mathbf{W}(\beta,\rho,M)} E^n_f\left[\int_a^b \big|F(t) - f(t)\big|^2 \Phi[t, (V_s)_{s\in[a,b]}]dt\right].$$

Theorem 6. *Grant Assumption C. For all $M > 0$, we have*

$$\lim_{n\to\infty} n^{2\beta/(2\beta+1)} \inf_F \mathbf{R}_n(F, \mathbf{W}(\beta, \rho, M))[\Phi] =$$

$$= P^\star_\beta\left(\tfrac{(\beta+1)(2\beta+1)}{\beta^2}\int_a^b dt\,\frac{E[\Phi(t,V)]^{\frac{2\beta+1}{2\beta}}}{\rho(t)^{1/2\beta}}\int_0^1 dx\, a^\star_{\Phi,t}(x)x^{1/\beta}\right)^{\frac{2\beta}{2\beta+1}} \tag{48}$$

where $P^\star_\beta$ is given by (24), *$a^\star_t(x)$ is the solution of*

$$E\left[\frac{\Phi(t,V)V_t^2}{(xV_t + a^\star_{\Phi,t}(x))^2}\right] = E[\Phi(t,V)] \quad \textit{if } E[\Phi(t,V)] > 0,$$

$$a^\star_{\Phi,t}(x) = 0 \quad \textit{otherwise,}$$

and the infimum is taken over all estimators. Moreover the bound (48) is attained by the estimator $\mathbb{F}_{n,N}$ up to a remainder term which converges to 0 *as $N \to \infty$, calibrated with $K^N_k = \delta_N^{-\beta} L_N(t^N_k)$ defined in (58) below.*

The proof of Theorem 6 is given in Section 5.

5. Proofs

As usual the notation C will stand for a generic constant, possibly varying from line to line and which is independent of f and n. Whenever needed, any other dependence will be explicitly mentioned.

5.1. **Proof of Lemma 1.** 1. Let B be a random variable such that $E(B^2) \leq x^2$ and set $H = B - B^\star_x$. On one hand we have

$$V(1-B)^2 = V(1-B^\star_x)^2 + VH^2 - 2VH(1-B^\star_x),$$

$$= V(1-B^\star_x)^2 + VH^2 - 2\frac{a^\star(x)}{x}HB^\star_x$$

since

$$1 - B^\star_x = \frac{a^\star(x)}{x}\frac{B^\star_x}{V}.$$

On the other hand

$$x^2 \geq E(B^2) = E\Big[(B_x^\star)^2 + H^2 + 2HB_x^\star\Big] = x^2 + E[H^2] + 2E[HB_x^\star].$$

Therefore

$$E\Big[V(1-B)^2\Big] \geq E\Big[V(1-B_x^\star)^2\Big] + E\Big[VH^2\Big] + \frac{a^\star(x)}{x}E\Big[H^2\Big].$$

Clearly, we have equality iff $H = 0$ a.s. and the second point follows.

2. The function $\psi\ :\ \mathbb{R}_+^* \to (0,1),\ z \mapsto \psi(z) = E\Big[V^2/(V+z)^2\Big]$ is a diffeomorphism of class C^∞. Therefore, since $a^\star(x) = x\psi^{-1}(x^2)$, $a^\star$ is of class C^∞. By definition, we have

$$E\Big[\frac{V^2}{(xV + a^\star(x))^2}\Big] = 1.$$

By differentiating, we obtain

$$E\Big[\frac{V^2(V + a^{\star\prime}(x))}{(xV + a^\star(x))^3}\Big] = 0. \tag{49}$$

Moreover

$$G(x) = a^\star(x)^2 E\Big[\frac{V}{(xV + a^\star(x))^2}\Big]$$

and we can clearly differentiate under the expectation, thus

$$\begin{aligned}
G'(x) &= 2a^{\star\prime}(x)a^\star(x)E\Big[\frac{V}{(xV + a^\star(x))^2}\Big] - 2a^\star(x)^2E\Big[\frac{V(V + a^{\star\prime}(x))}{(xV + a^\star(x))^3}\Big], \\
&= 2a^\star(x)E\Big[\frac{a^{\star\prime}(x)V^2x - a^\star(x)V^2}{(xV + a^\star(x))^3}\Big], \\
&= -2a^\star(x)E\Big[\frac{V^3x + a^\star(x)V^2}{(xV + a^\star(x))^3}\Big] \quad \text{by (49)}, \\
&= -2a^\star(x)E\Big[\frac{V^2}{(xV + a^\star(x))^2}\Big], \\
&= -2a^\star(x)
\end{aligned}$$

which ends the proof of the lemma. □

5.2. **Proof of Theorem 3.** The proof of the lower bound follows a classical route: the minimax risk is bounded from below by the Bayes risk for any prior essentially concentrated on the ellipsoid (see e.g. Nussbaum, 1999). We then choose a prior for which $F_n^{\star\star}$ is the Bayesian estimator.

If ν is a probability measure on $L^2([0,1], dt)$ endowed with its Borel σ-field, define the Bayes risk:

$$\mathbf{B}_n(\nu) = \inf_F \int E_f^n\big[\|F - f\|^2\big]\, \nu(df). \tag{50}$$

Recall that

$$\inf_F \mathbf{R}_n(F, \widetilde{\mathbf{W}}(\beta,\rho,M)) \geq \sup\Big\{\mathbf{B}_n(\nu);\ \nu \text{ such that } \nu[\widetilde{\mathbf{W}}(\beta,\rho,M)] = 1\Big\}.$$

Lemma 3. *If $\|f\| \leq R$ for ν-almost all f, then*

$$\mathbf{B}_n(\nu) \geq \mathbf{B}_n(\mu) - 4R^2\|\nu - \mu\|$$

where $\|\nu - \mu\|$ denotes the total variation of the signed measure $\nu - \mu$.

Proof. For an estimator F, set $\widetilde{F} = R\,F/\|F\|$. Since

$$\|F - f\|^2 = \|\widetilde{F} - f\|^2 + (\|F\| - R)^2 + 2\big(1 - \frac{R}{\|F\|}\big)\big(R\|F\| - (F, f)\big),$$

we have

$$\|F - f\|^2 \geq \|\widetilde{F} - f\|^2 \quad \text{on } \{\|F\| \geq R\} \quad \nu(df)\text{-a.e..}$$

Therefore

$$\begin{aligned}
\mathbf{B}_n(\nu) &= \inf_{F:\|F\|\leq R} \int \nu(df) E_f\big[\|F - f\|^2\big],\\
&= \inf_{F:\|F\|\leq R} \int \nu(df) E_f\big[(2R^2) \wedge \|F - f\|^2\big],\\
&\geq \inf_{F:\|F\|\leq R} \int \mu(df) E_f\big[(2R^2) \wedge \|F - f\|^2\big] - 4R^2\|\mu - \nu\|,\\
&= \inf_{F:\|F\|\leq R} \int \mu(df) E_f\big[\|F - f\|^2\big] - 4R^2\|\mu - \nu\|,\\
&\geq \mathbf{B}_n(\mu) - 4R^2\|\mu - \nu\|,
\end{aligned}$$

which proves the lemma. □

Now, let $(v_\lambda)_{\lambda\geq 1}$ be a sequence of positive number such that $\sum_{\lambda\geq 1} v_\lambda < \infty$ and let μ be the law on $L^2([0,1], dt)$ of the process

$$\Psi(t) = \sqrt{2}\sum_{\lambda\geq 1} \sqrt{v_\lambda}\big[\xi_{2\lambda}\cos(2\pi\lambda t) + \xi_{2\lambda+1}\sin(2\pi\lambda t)\big],$$

where $(\xi_\lambda)_{\lambda\geq 1}$ denotes a sequence of i.i.d. Gaussian standard random variables. Then we have

$$\mathbf{B}_n(\mu) = \sum_{\lambda\in\mathbb{Z}} E^n_\mu\left[\left|E^n_\mu(\hat{\Psi}(\lambda) \mid (Y_k)_{k\in\mathbb{Z}}) - \hat{\Psi}(\lambda)\right|^2\right]$$

where E^n_μ denotes integration w.r.t. the measure $P^n_\mu(d\omega) = \int \mu(df) P^n_f(d\omega)$. Since V is measurable w.r.t. $(Y_k)_{k\in\mathbb{Z}}$ -recall (10) in the Introduction- and since the random vectors $(Y_\lambda, \hat{\Psi}(\lambda))_{\lambda\in\mathbb{Z}}$ are independent conditional on V, we have

$$E_\mu\left[\hat{\Psi}(\lambda) \mid (Y_k)_{k\in\mathbb{Z}}\right] = E_\mu(\hat{\Psi}(\lambda) \mid Y_\lambda, V) = \frac{v_{|\lambda|}}{v_{|\lambda|} + V/n} Y_\lambda$$

with $v_0 = 0$. Therefore

$$E^n_\mu \left[\left| E^n_\mu(\hat\Psi(\lambda) \mid Y_\lambda, V) - \hat\Psi(\lambda) \right|^2 \mid V \right] = v_{|\lambda|} \frac{V}{V + n v_{|\lambda|}}.$$

Integrating w.r.t. the law of V, we readily get

$$\mathbf{B}_n(\mu) = 2 \sum_{\lambda \geq 1} v_\lambda E\Big[\frac{V}{V + n v_\lambda}\Big].$$

Let $0 < \varepsilon < 1$. Set $u_n = K n^{1/(2\beta+1)}$ for a constant K specified in (52) below. If we choose

$$v_\lambda = \begin{cases} \frac{1}{n} a^\star\big((\lambda/u_n)^\beta\big)/(\lambda/u_n)^\beta & \text{if } \varepsilon < \lambda/u_n < 1 \\ 0 & \text{otherwise,} \end{cases}$$

then

$$\mathbf{B}_n(\mu) = \frac{2}{n} \sum_{\lambda\,:\,\varepsilon < \lambda/u_n < 1} a^\star\big((\lambda/u_n)^\beta\big) E\Big[\frac{V}{(\lambda/u_n)^\beta V + a^\star\big((\lambda/u_n)^\beta\big)}\Big]. \quad (51)$$

The above Riemann sum on $[\varepsilon, 1]$ converges by continuity, therefore

$$\begin{aligned} \lim_{n\to\infty} n^{2\beta/(2\beta+1)} \mathbf{B}_n(\mu) &= 2K \int_\varepsilon^1 a^\star\big(x^\beta\big) E\Big[\frac{V}{x^\beta V + a^\star\big(x^\beta\big)}\Big]\, dx, \\ &= 2K \int_\varepsilon^1 \Big(x^\beta a^\star(x^\beta) + G(x^\beta)\Big)\, dx. \end{aligned}$$

where we used the definition of $a^\star$ and G, and the fact that

$$\frac{aV}{xV + a} = \frac{a}{x}\Big(\frac{xV}{xV + a}\Big)^2 + V\Big(1 - \frac{xV}{xV + a}\Big)^2.$$

Define

$$\nu_n(\cdot) = \frac{\mu_n\big(\cdot \cap \widetilde{\mathbf{W}}(\beta, \rho, M)\big)}{\mu_n\big(\widetilde{\mathbf{W}}(\beta, \rho, M)\big)}.$$

We claim that for any K such that

$$K < \left(\frac{1}{2(2\pi)^{2\beta} \int_0^1 a^\star(x^\beta) x^\beta dx}\right)^{1/(2\beta+1)} \quad (52)$$

we have

$$n^{2\beta/(2\beta+1)} \|\mu_n - \nu_n\| \to 0. \quad (53)$$

Applying Lemma 3 ($\|f\| \leq M^2$ for ν_n-almost all f), it follows that

$$\liminf_{n\to\infty} n^{2\beta/(2\beta+1)} \mathbf{B}_n(\nu_n) \geq$$

$$\geq \frac{2}{\big(\rho 2(2\pi)^\beta\big)^{1/(2\beta+1)}} \frac{\int_\varepsilon^1 x^\beta a^\star(x^\beta)\, dx + \int_\varepsilon^1 G(x^\beta)\, dx}{\Big(\int_\varepsilon^1 x^\beta a^\star(x^\beta)\, dx\Big)^{1/(2\beta+1)}}.$$

Since ε is arbitrarily small, this proves Theorem 3 in view of (35).

It remains to show (53). We have

$$\|\Psi^{(\beta)}\|^2 = \sum_{\lambda\geq 1}(2\pi\lambda)^{2\beta} v_\lambda(\xi_{2\lambda}^2 + \xi_{2\lambda+1}^2).$$

We deduce first that

$$\begin{aligned} E\big[\|\Psi^{(\beta)}\|^2\big] &= 2\sum_{\lambda\geq 1}(2\pi\lambda)^{2\beta} v_\lambda \\ &= 2(2\pi)^{2\beta}K^{2\beta+1}u_n^{-1}\sum_{\varepsilon u_n<\lambda<u_n} a^\star\big((\lambda/u_n)^\beta\big)(\lambda/u_n)^\beta \\ &= 2(2\pi)^{2\beta}K^{2\beta+1}\Big(\int_\varepsilon^1 a^\star(x^\beta)x^\beta\,dx + o(1)\Big) \end{aligned}$$

Second, Burckholder inequalities yield, for $p\geq 2$:

$$\begin{aligned} & E\Big[\Big|\|\Psi^{(\beta)}\|^2 - E\big[\|\Psi^{(\beta)}\|^2\big]\Big|^p\Big] \\ &\leq C_p E\Big[\Big|\sum_{\varepsilon u_n<\lambda<u_n}(2\pi\lambda)^{4\beta}v_\lambda^2(\xi_{2\lambda}^2+\xi_{2\lambda+1}^2-2)^2\Big|^{p/2}\Big] \\ &\leq C_{p,\beta}u_n^{p/2-1}\sum_{\varepsilon u_n\leq\lambda<u_n}\lambda^{2\beta p}v_\lambda^p \\ &\leq C_{p,\beta}\frac{u_n^{(2\beta+1/2)p}}{n^p}u_n^{-1}\sum_{\varepsilon u_n\leq\lambda<u_n} a^\star\big((\lambda/u_n)^\beta\big)^p(\lambda/u_n)^{2\beta p} \\ &\leq C_{p,\beta,K}u_n^{-p/2}\int_\varepsilon^1 a^\star\big(x^\beta\big)^p x^{2\beta p}\,dx \end{aligned}$$

Thus, thanks to Chebyshev inequality, for all $\eta>0$, $p\geq 2$, we have

$$P\Big((\|\Psi^{(\beta)}\|^2 > 2(2\pi)^{2\beta}K^{2\beta+1}\int_\varepsilon^1 a^\star(x^\beta)x^\beta\,dx + \eta\Big) \leq \\ \leq C_{p,K,\varepsilon,\eta}\; u_n^{-p/2}. \quad (54)$$

It suffices then to take $p/2 > 2\beta+1$, and we deduce that

$$n^{2\beta/(2\beta+1)}P(\|\Psi^{(\beta)}\|^2 > 1/\rho) \to 0$$

for all K satisfying (52).

All that remains to be proved is $n^{2\beta/(2\beta+1)}P\big(\sup_{t\in[0,1]}|\Psi(t)|>M\big)\to 0$. One easily checks that

$$\sup_{t\in[0,1]}|\Psi(t)| \leq \frac{\sqrt{2}K}{n^{2\beta/(2\beta+1)}}S_n$$

with

$$S_n = \frac{1}{u_n}\sum_{\varepsilon u_n\leq\lambda<u_n} a^\star\big((\lambda/u_n)^\beta\big)(\lambda/u_n)^{-\beta}\big(|\xi_{2\lambda}|+|\xi_{2\lambda+1}|\big)$$

and moreover $\sup_n E(S_n^2) < +\infty$. Chebyshev inequality thus yields that $P\big(\sup_{t\in[0,1]}|\Psi(t)| > M\big) = O(n^{-4\beta/(2\beta+1)})$. This completes the proof of Theorem 3.

5.3. **Proof of Theorem 6 : lower bound.**

Lemma 4. *Let $(\Gamma_m, V_m)_{m\geq 1}$ be a sequence of 2-dimensional random vectors such that $E(\Gamma_m) < \infty$, and $V_m > 0$ a.s. for all m. Assume that $(\Gamma_m, V_m) \to (\Gamma, V)$ in law, with $P(\Gamma > 0) > 0$ and that $(\Gamma_m)_{m\geq 1}$ is uniformly integrable. For $x \in (0,1)$, let $a_m^\star(x)$ be the unique solution of*

$$E\Big[\frac{\Gamma_m V_m^2}{\big(xV_m + a_m^\star(x)\big)^2}\Big] = E(\Gamma_m) \tag{55}$$

and define $a^\star(x)$ analogously, replacing (Γ_m, V_m) by (Γ, V) in (55) above. Then $a_m^\star \to a^\star$ as $m \to \infty$ pointwise on $(0,1)$.

Proof. Let $x \in (0,1)$. For $a \geq 0$ let

$$\varphi_m(a) = E\Big[\frac{\Gamma_m V_m^2}{\big(xV_m + a\big)^2}\Big]$$

and define $\varphi(a)$ analogously replacing (Γ_m, V_m) by (Γ, V). Note that $\varphi_m(a)$ and $\varphi(a)$ are decreasing in a. From the uniform integrability of the Γ_m and the weak convergence of (Γ_m, V_m) it follows that $\varphi_m(a) \to \varphi(a)$ for all $a \geq 0$. Moreover, since φ is continuous, the convergence holds uniformly in $a \geq 0$. Noting that $a_m^\star(x) = \varphi_m^{-1}(E(\Gamma_m))$, $a^\star(x) = \varphi^{-1}(E(\Gamma))$ and $E(\Gamma_m) \to E(\Gamma)$ the conclusion follows. □

Let $h : [a,b] \to \mathbb{R}_+$ be a nonnegative Riemann integrable function such that $\int_a^b h(t)\,dt \leq 1$. Define

$$\bar{h}_k^N = \frac{1}{\delta_N}\int_{t_{k-1}^N}^{t_k^N} h(s)\,ds \quad \text{and} \quad \rho_{N,k}^+ = \sup_{s\in[t_{k-1}^N, t_k^N]} \rho(s)$$

where -recall Section 4- we have set $\delta_N = (b-a)/N$ and let

$$\widetilde{\mathbf{W}}(\beta, h, N, k, M)$$

denote the class of periodic functions on the interval $[t_{k-1}^N, t_k^N]$ which are bounded by M and with Sobolev index β and (constant) Sobolev weight function $\rho_{N,k}^+/\delta_N \bar{h}_k^N$. Clearly if $f \in \widetilde{\mathbf{W}}(\beta, h, N, k, M)$ for all $k = 1, \ldots, N$ then $f \in \mathbf{W}(\beta, \rho, M)$. Thus

$$\sup_{f \in \mathbf{W}(\beta,\rho,M)} E_f^n\Big(\int_a^b \Phi(t,V)\big(F(t) - f(t)\big)^2 dt\Big) \geq$$

$$\geq \sum_{k=1}^N \sup_{f \in \widetilde{\mathbf{W}}(\beta,h,N,k,M)} E_f^n\Big(\Phi_{N,k}^- \int_{t_{k-1}^N}^{t_k^N} \big(F(t) - f(t)\big)^2 dt\Big)$$

where $\Phi^-_{N,k} = \inf_{s\in[t^N_{k-1},t^N_k)} \Phi(s,V)$. For notational simplicity, we further drop the dependence of the parameter space in the risk functional. Using that each summand of (51) is increasing, a straightforward extension of Theorem 3 implies

$$\liminf_{n\to\infty} n^{2\beta/(2\beta+1)} \inf_F \mathbf{R}_n(F)[\Phi] \geq$$
$$\geq C^\star_\beta \sum_{k=1}^N \frac{E(\Phi^-_{N,k})}{\left(\rho^+_{N,k}/\delta_N \bar{h}^N_k\right)^{1/(2\beta+1)}} \left(\delta_N \int_0^1 a^\star_{N,k}(x) x^{1/\beta}\, dx\right)^{2\beta/(2\beta+1)}$$

where $C^\star_\beta = P^\star_\beta \left(\frac{(\beta+1)(2\beta+1)}{\beta}\right)^{2\beta/(2\beta+1)}$ and $a^\star_{N,k}(x)$ is defined by

$$E\left[\frac{\Phi^-_{N,k}(V^-_{N,k})^2}{(xV^-_{N,k}+a^\star_{N,k}(x))^2}\right] = E(\Phi^-_{N,k}) \quad \text{with} \quad V^-_{N,k} = \inf_{s\in[t^N_{k-1},t^N_k]} V_s$$

if $E(\Phi^-_{N,k}) > 0$ and 1 otherwise. We readily derive

$$\liminf_{n\to\infty} n^{2\beta/(2\beta+1)} \inf_F \mathbf{R}_n(F)[\Phi] \geq C^\star_\beta \int_a^b [\bar{h}_N(t)]^{1/(2\beta+1)} \psi_N(t) dt$$

where ψ_N and $\bar{h}_N$ are the stepwise functions defined by

$$\psi_N(t) = \frac{E(\Phi^-_{N,k})}{\left(\rho^+_{N,k}\right)^{1/(2\beta+1)}} \left(\int_0^1 a^\star_{N,k}(x) x^{1/\beta}\, dx\right)^{2\beta/(2\beta+1)},$$
$$\bar{h}_N(t) = \bar{h}^N_k \quad \text{if } t \in [t^N_{k-1}, t^N_k).$$

The stepwise function with value $a^\star_{N,k}(x)$ on $[t^N_{k-1}, t^N_k)$ converges pointwise to $t \to a^\star_t(x)$ on the set $\{t : \Phi(t,V) > 0 \text{ a.s.}\}$ by Lemma 4 and Assumption C. Therefore, for all t such that $\Phi(t,V) > 0$ a.s., we have by Fatou lemma that

$$\liminf_{N\to\infty} \psi_N(t) \geq \psi(t) = \frac{E(\Phi(t,V))}{\rho(t)^{1/(2\beta+1)}} \left(\int_0^1 a^\star_t(x) x^{1/\beta}\, dx\right)^{2\beta/(2\beta+1)}.$$

On the set $\{t : \Phi(t,V) = 0 \text{ a.s.}\}$, we have $\psi_N(t) \to 0$ as $N \to \infty$, therefore we have $\liminf_{N\to\infty} \psi_N(t) \geq \psi(t)$ for all t. Applying again Fatou lemma yields

$$\liminf_{n\to\infty} n^{2\beta/(2\beta+1)} \inf_F \mathbf{R}_n(F)[\Phi] \geq C^\star_\beta \int_0^1 h(t)^{1/(2\beta+1)} \psi(t)\, dt.$$

Maximizing the RHS of the above inequality with respect to h under the constraint $\int_a^b h(t)dt \leq 1$ (apply Hölder inequality) yields the choice

$$h(t) = \psi(t)^{(2\beta+1)/2\beta} \Big/ \int_0^1 \psi(s)^{(2\beta+1)/2\beta}\, ds$$

and the result follows. The proof of the lower bound is complete.

5.4. **Proof of Theorem 6 : upper bound.** One can simply check that under the calibration given in (58), the estimator $\mathbb{F}_{n,N}$ attains the lower bound. We rather give a self containing proof, introducing the quantity in (58) in a more transparent way. Clearly

$$n^{2\beta/(2\beta+1)} E_f^n \left(\int_a^b \Phi(t,V)\big(\mathbb{F}_{n,N}(t) - f(t)\big)^2 dt \right) \leq$$

$$\leq \sum_{k=1}^{N} n^{2\beta/(2\beta+1)} E_f^n \left(\Phi_{N,k}^{+} \int_{t_{k-1}^N}^{t_k^N} \big(\mathbb{F}_{n,N}(t) - f(t)\big)^2 dt \right) \tag{56}$$

where $\Phi_{N,k}^{+} = \sup_{s \in [t_{k-1}^N, t_k^N)} \Phi(s,V)$. Let $a_{N,k}^{\star}(x)$ be the solution of

$$E\left[\frac{\Phi_{N,k}^{+} \sup_{s\in[t_{k-1}^N,t_k^N]} V_s^2}{(x \sup_{s\in[t_{k-1}^N,t_k^N]} V_s + a_{N,k}^{\star}(x))^2} \right] = E\left[\Phi_{N,k}^{+}\right]$$

if $E\left[\Phi_{N,k}^{+}\right] > 0$ and set $a_{N,k}^{\star} = 0$ otherwise. Using that $\mathbb{F}_{n,N} = F^{\star\star}_{t_{k-1}^N, t_k^N}$ on $[t_{k-1}^N, t_k^N)$, using same arguments as those used for the evaluation of the risk in the homogeneous case (in particular the fact that (44) holds for any $K \in (0,\infty)$), together with the affine transform $T_{t_{k-1}^N, t_k^N}$ from $[0,1]$ to $[t_{k-1}^N, t_k^N]$, we obtain that each term of the sum in (56) is less than

$$E(\Phi_{N,k}^{+}) \left(\frac{\delta_N^{2\beta} (K_k^N)^2 I_k^N(f)}{(2\pi)^{2\beta} \rho_{N,k}^{-}} + \frac{4}{(K_k^N)^{1/\beta}} \int_0^1 x^{1/\beta} a_{N,k}^{\star}(x) dx + \varepsilon_{n,N,k} \right), \tag{57}$$

where $I_k^N(f) = \int_{t_{k-1}^N}^{t_k^N} |f^{(\beta)}(s)|^2 \rho(s) ds$, $\rho_{N,k}^{-} = \inf_{s\in[t_{k-1}^N, t_k^N)} \rho(s)$ and $\varepsilon_{n,N,k}$ is a remainder term converging to 0 as $n \to \infty$, uniformly in $f \in \mathbf{W}(\beta,\rho,M)$. Introduce now the following piecewise constant functions:

$$g_N(f,t) = \delta_N^{-1} I_k^N(f), \quad L_N(t) = \delta_N^{\beta} K_k^N,$$
$$\rho_N(t) = \inf_{s\in[t_{k-1}^N,t_k^N]} \rho(s), \quad \Phi_N^{+}(t) = \sup_{s\in[t_{k-1}^N,t_k^N)} \Phi(s,V),$$
$$\text{and} \quad G_N(t) = \int_0^1 x^{1/\beta} a_{N,k}^{\star}(x) dx \quad \text{if } t_{k-1}^N \leq t < t_k^N.$$

Summing in k inequality (57) yields

$$n^{2\beta/(2\beta+1)} \mathbf{R}_n(\mathbb{F}_{n,N}) \leq$$

$$\leq \int_a^b \left(\frac{L_n^2(t) g_N(f,t)}{(2\pi)^{2\beta} \rho_N(t)} + \frac{4}{L_N(t)^{1/\beta}} G_N(t) \right) E[\Phi_N^{+}(t)] dt + N \max_k \varepsilon_{n,N,k}$$

having $\int_a^b g_N(f,t) dt \leq 1$. Thus

$$\limsup_{n\to\infty} n^{2\beta/(2\beta+1)} \mathbf{R}_n(\mathbb{F}_{n,N}) \leq$$

$$\leq \sup_{t\in[a,b]} \frac{L_N^2(t)E[\Phi_N^+(t)]}{(2\pi)^{2\beta}\rho_N(t)} + 4\int_a^b \frac{G_N(t)}{L_N(t)^{1/\beta}}E[\Phi_N^+(t)]dt.$$

Again, we have a similar minimization problem as in Section 3.1.2. Elementary computation yields the choice

$$L_N(t) = \frac{\rho_N(t)^{1/2}}{E[\Phi_N^+(t)]^{1/2}}(2\pi)^\beta\left(\frac{2}{\beta\pi}\int_a^b \frac{G_N(s)E[\Phi_N^+(s)]^{(2\beta+1)/2\beta}}{\rho_N(s)^{1/2\beta}}ds\right)^{\beta/(2\beta+1)}. \tag{58}$$

Some calculus shows that

$$\limsup_{n\to\infty} n^{2\beta/(2\beta+1)}\mathbf{R}_n(\mathbb{F}_{n,N}) \leq$$

$$\leq P_\beta^\star\left[\frac{(\beta+1)(2\beta+1)}{\beta^2}\int_a^b \frac{E[\Phi_N^+(t)]}{\rho_N(t)^{1/(2\beta)}}G_N(t)dt\right]^{2\beta/(2\beta+1)}.$$

Let $a_\infty^\star(x)$ be the solution of

$$E\left[\frac{\sup_{t\in[a,b]}\Phi(t,V)\sup_{t\in[a,b]}V_t^2}{(x\sup_{t\in[a,b]}V_t + a_\infty^\star(x))^2}\right] = E(\sup_{t\in[a,b]}\Phi(t,V)).$$

and let $G_\infty = \int_0^1 x^{1/\beta}a_\infty^\star(x)dx$. Clearly $G_N(t)\leq G_\infty$ and similar arguments as those used for the lower bound, using Lebesgue theorem instead of Fatou lemma show that $G_N(t)\to G(t) = \int_0^1 x^{1/\beta}a_{\Phi,t}^\star(x)dx$ pointwise. Now, applying again Lebesgue theorem we obtain the desired bound. The proof of the upper bound is complete.

6. Appendix

Let $(\Omega,\mathcal{F},P)$ be a probability space on which is defined a standard Brownian motion $W=(W_t)_{t\in[0,1]}$. Up to considering an extension of Ω, we may assume that G is defined on Ω and is independent of W. Define

$$\widetilde{Z}_n^f = \exp\left(\sqrt{\frac{n}{G}}\int_0^1 f(s)dW_s - \frac{n}{2G}\int_0^1 f^2(s)ds\right)$$

and let $\widetilde{P}_n^f = \widetilde{Z}_n^f\cdot P$. The mixed white noise experiment is

$$(\Omega,\mathcal{F},(\widetilde{P}_n^f)_{f\in\mathbf{W}(\beta,\rho)}). \tag{59}$$

Likewise, let

$$Z_n^f = \exp\left(\frac{n}{G}\sum_{i=1}^n f(i/n)(W_{i/n}-W_{(i-1)/n}) - \frac{1}{2G}\sum_{i=1}^n f^2(i/n)\right)$$

and $P_n^f = Z_n^f\cdot P$. The regression model (4) is

$$(\Omega,\mathcal{F},(P_n^f)_{f\in\mathbf{W}(\beta,\rho)}).$$

From standard arguments from the Le Cam theory (see for instance Le Cam and Yang, 1990 or Nussbaum, 1996, for details of the method of proof employed here), the asymptotic equivalence of model (4) and the mixed white noise experiment defined by (59) is implied by the following convergence

$$E(|Z_n^f - \widetilde{Z}_n^f|) \to 0$$

uniformly in $f \in \mathbf{W}(\beta, \rho)$. Writing

$$Z_n^f - \widetilde{Z}_n^f = \widetilde{Z}_n^f \left(\frac{Z_n^f}{\widetilde{Z}_n^f} - 1 \right)$$

and using Scheffe lemma, it is enough to prove $\frac{Z_n^f}{\widetilde{Z}_n^f} \to 1$ in $\widetilde{P}_n^f$-probability, uniformly in $f \in \mathbf{W}(\beta, \rho)$. Moreover

$$\log \frac{Z_n^f}{\widetilde{Z}_n^f} = \sqrt{\frac{n}{G}} \left(\sum_{i=1}^n f(i/n)(\widetilde{W}_{i/n} - \widetilde{W}_{(i-1)/n}) - \int_0^1 f(s) d\widetilde{W}_s \right) +$$

$$+ \frac{n}{2G} \left\{ \left(\sum_{i=1}^n f(i/n) \int_{(i-1)/n}^{i/n} f(s) ds - \int_0^1 f^2(s) ds \right) - \right.$$

$$\left. - \frac{1}{2} \left(\frac{1}{n} \sum_{i=1}^n f^2(i/n) - \int_0^1 f^2(s) ds \right) \right\}$$

where $\widetilde{W}_t = W_t - \sqrt{\frac{n}{G}} \int_0^t f(s) ds$. Under $\widetilde{P}_n^f$, the process $\widetilde{W}$ is a standard Brownian motion. It remains to prove the convergence to 0 in $\widetilde{P}_n^f$-probability of the two terms in the above sum.

Conditional on G, the first term is a sum of centered independent random variables with variance $\frac{n}{G} \int_{(i-1)/n}^{i/n} (f(i/n) - f(s))^2 ds$. From the smoothness property of f it is easily seen that this term has the right order.

The other two terms only involve smoothness properties of f. The computations straightforward so we omit them. The proof of Proposition 1 is complete.

References

[1] Belitser, E. and Levit, B. (1995). On minimax filtering on ellipsoids. *Math. Meth. Statist.* **4**, 259-273.

[2] Brown, L.D. and Low, M.G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24**, 2384-2398.

[3] Delattre, S. and Hoffmann, M. (2001). Asymptotic equivalence for a null recurrent diffusion model. Forthcoming in *Bernoulli.*

[4] Efromovich, S.Yu. (1996). On nonparametric regression for iid observations in a general setting. *Ann. Statist.* **23**, 1376-1392.

[5] Efromovich, S.Yu. and Pinsker, M.S. (1982). Estimating a square integrable spectral density based on a sequence of observations. *Problems Inform. Transmission* **17**, 182-196.

[6] Golubev, G.K. (1991). Local asymptotic normality in problems of nonparametric estimation of functions and lower bound for quadratic risk. *Theory Probab. Appl.* **36**, 1, 152-157.

[7] Golubev, G.K. (1992). Quasilinear filtering of Gaussian stationary sequences. *Theory Probab. Appl.* **37**, 4, 793-800.

[8] Golubev, G.K. (1993). Nonparametric estimation of smooth spectral densities of Gaussian stationary sequences. *Theory Probab. Appl.* **38**, 4, 630-639.

[9] Golubev, G.K. and Nussbaum, M. (1990). A risk bound in Sobolev class regression. *Ann. Statist.* **18**, 758-778.

[10] Grama, I. and Nussbaum, M. (1998). Asymptotic equivalence for nonparametric generalized linear models. *Probab. Theory Relat. Fields.* **111**, 167-214.

[11] Ibragimov, I.A. and Khasminskii, R.Z. (1977). On the estimation of an infinite dimensional parameter in Gaussian white noise. *Soviet Math. Dokl.* **236**, No. 5, 1053-1055.

[12] Ibragimov, I.A. and Khasminskii, R.Z. (1981). *Statistical Estimation: Asymptotic Theory.* Springer, New York.

[13] Jeganathan, P. (1982). On the asymptotic theory of estimation when the limit of the loglikelihood is mixed normal. *Sankhyā, Series A.* **44**, part 2, 173-212.

[14] Jeganathan, P. (1983). Some asymptotic properties of risk functions when the limit of the experiment is mixed normal. *Sankhyā, Series A.* **45**, part 1, 66-87.

[15] Kiefer J. (1972) Conditional inference. In: *Encyclopedia of Statistical Science*, Volume 2, 103-109, John Wiley, New York.

[16] Le Cam, L. and Yang, G.L. (1990). *Asymptotics in Statistics : Some basic concepts*. Springer, New York.

[17] Milstein G. and Nussbaum M, (1998). Autoregression approximation of a nonparametric diffusion model. *Probab. Theory Relat. Fields.* **112**, 535-543.

[18] Nussbaum, M. (1985). Spline smoothing in regression models and asymptotic efficiency in L_2. *Ann. Statist.* **13**, 984-997.

[19] Nussbaum, M. (1996). Asymptotic equivalence of density estimation and white noise. *Ann. Statist.* **24**, 2399-2430.

[20] Nussbaum, M. (1999). Minimax risk, Pinsker bound. In: *Encyclopedia of Statistical Sciences*, Update Volume 3, 451-460, John Wiley, New York.

[21] Pinsker, M.S. (1980). Optimal filtering of square integrable signals in Gaussian white noise (in Russian). *Problemy Peredachi Informatsii* **16**, (2), 52-68. English translation *Problems Infom. Transmission*, (1980), 120-133.

E-mail address: delat@ccr.jussieu.fr, hoffmann@math.jussieu.fr