Representation of small ball probabilities in Hilbert space and lower bound in regression for functional data

André Mas

Representation of small ball probabilities in Hilbert space and lower bound in regression for functional data.

André Mas

March 1, 2011

Abstract

Let $S = \sum_{i=1}^{+\infty} \lambda_i Z_i$ where the Z_i's are i.d.d. positive with $\mathbb{E}|Z|^3 < +\infty$ and $(\lambda_i)_{i \in \mathbb{N}}$ a positive nonincreasing sequence such that $\sum \lambda_i < +\infty$. We study the small ball probability $\mathbb{P}(S < \varepsilon)$ when $\varepsilon \downarrow 0$. We start from a result by Lifshits (1997) who computed this probability by means of the Laplace transform of S. We prove that $\mathbb{P}(S < \cdot)$ belongs to a class of functions introduced by de Haan, well-known in extreme value theory, the class of Gamma-varying functions, for which an exponential-integral representation is available. This approach allows to derive bounds for the rate in nonparametric regression for functional data at a fixed point x_0 : $\mathbb{E}(y|X = x_0)$ where $(y_i, X_i)_{1 \leq i \leq n}$ is a sample in $(\mathbb{R}, \mathcal{F})$ and $\mathcal{F}$ is some space of functions. It turns out that, in a general framework, the minimax lower bound for the risk is of order $(\log n)^{-\tau}$ for some $\tau > 0$ depending on the regularity of the data and polynomial rates cannot be achieved.

Keywords : Small ball problems, functional data, regular variation, nonparametric regression, lower bound, Gaussian random elements.

1 Preliminaries

The three following susbections are independent. The first gives some basic material about small ball probability. The second collects classical results from extreme value theory as well as the definition of the class Γ_0 which is then briefly described. The third introduces the nonparametric regression model for functional data and simply raises the problems attached to obtaining sharp bounds for the quadratic risk at a fixed point. The notions encountered in this long introduction though intially distinct from each other merge in the sequel of this work and give birth to the main results. Some proofs are given in the last section.

1.1 About non-shifted and shifted small ball problems

Small ball problems could generally be stated the following way : consider a random variable X with values in a general normed space $(E, \|\cdot\|)$ (which may not be finite-dimensional) and estimate $\mathbb{P}(\|X\| < \varepsilon)$ for small values of ε. This issue may be viewed as a counterpart of the large deviations or concentration problems (where $\mathbb{P}(\|X\| > M)$ is studied for large M) and the terms "small deviations" or "lower tail behaviour" are sometimes encountered to name small ball problems. The core of the literature on small ball problems focuses on Gaussian random variables. The survey by Li and Shao (2001) is a complete state of the art, introducing the main concepts and providing numerous references. Another reference is Chapter 18 of Lifshits (1995) entirely devoted to Gaussian random functions. Much attention has been given to Brownian motion (when $(E, \|\cdot\|) = (C(0.1), |\cdot|_\infty)$) or its relatives (fractional Bronwian motion, Bronwian sheet, etc). The case of stable random elements was also investigated (see for instance Li, Linde (2004), Aurzada, Lifshits, Linde (2009)). Another issue is related to the norm. Indeed in infinite dimensional spaces, norms or metrics are not equivalent and this may influence the local behaviour of $\mathbb{P}(\|X\| < \varepsilon)$.

A more general question could be the shifted small ball probability $\mathbb{P}(\|X - x_0\| < \varepsilon)$ for a fixed x_0. A stumbling stone arises from the shift x_0. It turns out that, in general, computations cannot be carried out for any x_0. Several works focus on expliciting the set of those x_0 for which the shifted small ball probability may be computed from the non-shifted one (when $x_0 = 0$). We refer to Borell (1976) or Li and Linde (1993) for instance. A classical example stems from the situation where $\mathbb{P}_{X-x_0} \ll \mathbb{P}_X$ where $\mathbb{P}_X$ denotes the probability distribution induced by the random element X. The classical Cameron-Martin's theorem for Brownian motion illustrates this case for instance. Absolute regularity yields :

$$\mathbb{P}(\|X - x_0\| < \varepsilon) = \int_{B(0,\varepsilon)} \mathbb{P}_{X-x_0}(dx) = \int_{B(0,\varepsilon)} f_{x_0}(x) \mathbb{P}_X(dx)$$

where $f_{x_0} = d\mathbb{P}_{X-x_0}/d\mathbb{P}_X$ and $B(0,\varepsilon)$ stands for the ball centered at 0 with radius ε. When f_{x_0} is regular enough in a neighborhood of zero :

$$F_{x_0}(h) = \int_{B(0,\varepsilon)} \mathbb{P}_{X-x_0}(dx) = \int_{B(0,\varepsilon)} f_{x_0}(x) \mathbb{P}_X(dx) \sim f_{x_0}(0) F(h) \tag{1}$$

About this fact see Proposition 2.1 in de Acosta (1983). In general the sharpness of existing results may vary, depending on the triplet $((E, \|\cdot\|), \mathbb{P}_X, x_0)$ under consideration. In fact there are only few spaces for which the local behaviour of $\mathbb{P}(\|X - x_0\| < \varepsilon)$ is explicitely described. Quite often lower and upper bounds are computed so that :

$$\mathbb{P}(\|X - x_0\| < \varepsilon) \asymp \varphi_{x_0}(\varepsilon)$$

where φ_{x_0} is known and $f \asymp g$ means here that for some positive constants c^- and c^+ the positive functions f and g satisfy :

$$0 < c^- \leq \liminf_0 \frac{f}{g} \leq \limsup_0 \frac{f}{g} \leq c^+.$$

Sometimes only one of these bounds is accessible or needed.

It is worth noting or recalling a few crucial features of small deviations techniques. The Laplace transform, as well as in large deviations problems, is a major tool when coupled with the saddlepoint method. Small deviations are intimately connected with the entropy of the unit ball of the reproducing kernel Hilbert space associated with X, with the l-approximation numbers of X (i.e. the rate of approximation of X by a finite dimensional random variable, see Li, Linde (1999)) or to the degree of compactness of linear operators generating X (see Li, Linde (2004)). All these notions are clearly connected to the regularity of the process X, when X is a process.

Applications of small ball probabilities are numerous : they appear when studying rates of convergence in the Law of the Iterated Logarithm (see Talagrand (1992), Kuelbs, Li, Linde (1994)) or the rate of escape of the Brownian motion (see Erickson (1980)). They even surprisingly provide a sufficient condition for the CLT (see Ledoux, Talagrand (1991), Theorem 10.13 p.289). However small ball problems remained until nowadays a matter essentially reserved to probability theory. However Van der Vaart and van Zanten (2007 and 2008) found applications of small ball techniques to Bayesian statistics It turns out that this topic may be also of interest in another area of statistics : functional data analysis. FDA for short extends classical statistical models designed for vectors to the situation when the data are functions or curves. One of the concern may be summarized this way : since Lebesgue's density of an infinite-dimensional random X does not exist, all the inference techniques based on the density cannot hold anymore. In this framework, the small ball probabilities appear as a natural counterpart and should be investigated with much care. We illustrate this fact by pointing out an elementary example -kernel methods- in the next subsection below.

First let us precise the l_2 framework. Consider X a random variable defined the following way :

$$X = \left(\sqrt{\lambda_1}x_1, \sqrt{\lambda_2}x_2, ...\right) \tag{2}$$

where $(\lambda_i)_{1\leq i\leq n}$ is a real positive sequence arranged in a non-decreasing order such that $\sum_{i=1}^{+\infty}\lambda_i < +\infty$ and $(x_i)_{1\leq i\leq n}$ is sequence of real independent and identically distributed random variables with null expectation. From Kolmogorov's $0-1$ law it is straightforward to see that X exists as a l_2-valued random element. The square norm of X is $S = \sum_{i=1}^{+\infty}\lambda_i x_i^2$.

The small ball problem consists here in estimating for different choices of the sequence $(\lambda_i)_{i\in\mathbb{N}}$ and $(x_i)_{i\in\mathbb{N}}$ the probability $\mathbb{P}(S < r)$ when r tends to zero. The latter probability is expected to depend on the λ_i's. About this fact we refer to Dunker, Lifshits, Linde (1998).

The inspection of the case $E = l_2$ is motivated by the application to functional statistics mentioned in the paragraph above. Indeed random functions are often reconstructed by interpolation techniques, like splines or wavelets, in Hilbert spaces such as $L^2([0,T])$ or the Sobolev space $W^{m,2}([0,T])$, $m \in \mathbb{N}$. Then the random element X is valued in a separable Hilbert space $\mathcal{H}$ and all these Hilbert spaces of functions are isometrically isomorphic to l_2. In this framework a useful tool is the so-called Karhunen-Loève decomposition (sometimes refered to as Principal Orthogonal Decomposition in other area of mathematics such as PDEs). Any centered random function X will be represented by its coordinates in a basis of eigenvectors of the covariance operator $\mathcal{C}_X = \mathbb{E}[X \otimes X]$. When e_i's are the eigenvectors of $\mathcal{C}_X$ and λ_i the associated eigenvalues

$$X = \sum_{i=1}^{+\infty}\sqrt{\lambda_i}x_i e_i \tag{3}$$

where the x_i's are uncorrelated real random variables. The x_i's are actually always independent when X is Gaussian and are assumed to be in most settings. The l_2 random element defined in (2) is formally identifiable with this Karhunen-Loève decomposition familiar in Functional Data Analysis.

Historically the description of the exact behaviour of Gaussian small ball probability in Hilbert space is due to Sytaya (1974). However we borrow the notations from Lifshits (1997) who extendend Sytaya's results in several directions amongts which the non-Gaussian framework. First in order to alleviate notations set once and for all :

$$S = \sum_{i=1}^{+\infty}\lambda_i Z_i \tag{4}$$

where $\lambda_i > 0$ are arranged in decreasing order with $\sum_{i=1}^{+\infty}\lambda_i < +\infty$ and Z_i are positive random variables (they stand for the x_i^2's above). For the sake of completeness and since the main theorems of this work heavily rely on his results we recall them. In the previously mentioned article Lifshits proved that :

$$\mathbb{P}(S < r) \underset{r\to 0}{\sim} \frac{1}{\sqrt{2\pi}}\frac{1}{\gamma\sigma}\exp(\gamma r)\Lambda(\gamma) \tag{5}$$

where γ and σ are functions of r defined below and $\Lambda(\gamma) = \mathbb{E}\exp(-\gamma S)$ is the Laplace transform of S evaluated at $\gamma(r)$. The definitions of γ and σ are implicit. Let S_γ be the Esscher transform of S that is the random variable with distribution $\exp(-\gamma x)\mathbb{P}_S(dx)/\Lambda(\gamma)$. Then set :

$$r = \mathbb{E}[S_\gamma] = -\frac{\partial \log\Lambda(\gamma)}{\partial\gamma}, \tag{6}$$

$$\sigma^2 = \mathbb{V}[S_\gamma] = \frac{\partial^2 \log\Lambda(\gamma)}{\partial\gamma^2}. \tag{7}$$

where $\mathbb{V}$ denotes variance. Without further assumption on the λ_i's $\mathbb{P}(S < r)$ cannot be made more explicit. This is done for instance in Dunker, Lifshits, Linde (1998) where these author considered the case of λ_i with polynomial and exponential decay. Due to the remark below (3) we will sometimes refer to $\mathbb{P}(S < r)$ as a small ball probability for an l_2-valued random element.

The article is organized as follows. The next subsection develops some aspects of mathematical statistics which motivate this approach on small ball problems. Then a class of functions which appears in extremes value theory -the class Γ_0- is introduced in the next section. Our main theorem shows that small ball probabilities of l_2 random elements (hence of random functions belonging to a Hilbert space) belong to the class Γ_0. We then show how this result may be used to solving the statistical issues mentioned earlier. In particular we prove the the optimal rate of convergence in nonparametic regression for functional variables is always slower than any power of n. The derivations of the main results are collected in the last part of the article.

1.2 The class Γ_0

The theory of extremes is another well-known topic connecting probability theory, mathematical statistics and real analysis through regular variation and Karamata's theory. The foundations of extreme value theory may be illustrated by the famous Fisher-Tippett theorem (see Fisher, Tippett (1928) and Gnedenko (1943)). This classical result assesses that whenever $U_1, ..., U_n$ is an i.id. sample of real random variables, $M_n = \max\{U_1, ..., U_n\}$ belongs to the domain of attraction of G, where G has same type as one of the three distributions Gumbel, Frechet and Weibull. The Gumbel law, also named double exponential distribution, with cumulative distribution function $\Lambda(x) = \exp(-\exp(-x))$ defines the so-called "domain of attraction of the third type". Laurens de Haan (1971) characterized the (cumulative) distribution functions of U such that M_n belongs to the domain of attraction of Λ. We give this result below.

Theorem (de Haan, 1971) : *If F is the cumulative distribution function of a real random variable X which belongs to the domain of attraction of the third type (Gumbel) there exists a measurable function $\rho : \mathbb{R} \to \mathbb{R}^+$, called the auxiliary function of F, such that :*

$$\lim_{s \uparrow x_+} \frac{\overline{F}(s + x\rho(s))}{\overline{F}(s)} = \exp(-x)$$

where $\overline{F}(s) = 1 - F(s)$, $x_+ = \sup\{x : F(x) < 1\}$.

This property was intially introduced by de Haan as a "Form of Regular Variation" (see the title of his article). This class of distribution function is referred to as de Haan's Gamma class in the book by Bingham, Goldie and Teugels (1987) and within this article. In the latter book the definition is slightly different from the one given above. Gamma-variation is defined at infinity and for non-decreasing functions which comes down to taking $x_+ = +\infty$ and taking $\exp(x)$ instead of $\exp(-x)$ in the display above. Surprisingly, in their book as well as in de Haan's article no examples of functions belonging to Γ is given. The cumulative distribution function function of the Gaussian distribution belongs to this class with $x_+ = +\infty$ and $\rho(s) = 1/s$.

Since we focus on the local behaviour at zero of the cumulative distribution function function of a real valued random variable we have to modifiy again slightly the definitions above. We introduce the class Γ_0 and feature some of its properties below. We borrow most of our notations from Bingham, Goldie and Teugels (1987) which differ from those of de Haan.

Definition 1 *The class Γ_0 consists of those functions $F : \mathbb{R} \to \mathbb{R}^+$ null over $(-\infty, 0]$, non decreasing with $F(0) = 0$ and right-continuous for which there exists a continuous non decreasing function $\rho : \mathcal{V}^+ \to \mathbb{R}^+$, defined on some a right-neighborhood of zero $\mathcal{V}^+$ such that $\rho(0) = 0$ and for all $x \in \mathbb{R}$,*

$$\lim_{s \downarrow 0^+} \frac{F(s + x\rho(s))}{F(s)} = \exp(x) \tag{8}$$

The function ρ is called the auxiliary function of F.

The properties of the auxiliary function are crucial.

Proposition 1 *From Definition 1 above we deduce that :* $\rho(s)/s \to 0$ *as* $s \to 0$ *and* ρ *is self-neglecting which means that :*

$$\frac{\rho(s + x\rho(s))}{\rho(s)} \overset{s \to 0}{\to} 1$$

locally uniformly in $x \in \mathbb{R}$.

Remark 1 *When the property in the proposition above does not hold locally uniformly but only pointwise the function is called Beurling slowly varying. Assuming that* ρ *is continuous in Definition 1 yields local uniformity and enables to consider a self-neglecting* ρ.

The class Γ_0 is subject to an exponential-integral representation. In fact the following Theorem asserts that the local behaviour at 0 of any F in Γ_0 depends only on the auxiliary mapping ρ.

Theorem 1 *Let* F *belong to* Γ_0 *with self-neglecting auxiliary function* ρ *then when* $s \to 0$ *:*

$$F(s) = \exp\left\{\eta(s) - \int_s^1 \frac{1}{\rho(t)} dt\right\} \tag{9}$$

with $\eta(s) \to c \in \mathbb{R}$ *and the auxiliary function* ρ *is unique up to asymptotic equivalence and may be taken as* $\int_0^s F(t)\, dt/F(s)$. *Besides*

$$F(\lambda s)/F(s) \to \begin{cases} \infty & (\lambda > 1) \\ 1 & (\lambda = 1) \\ 0 & (\lambda < 1) \end{cases} \qquad \text{as } s \to 0. \tag{10}$$

Remark 2 *The upper bound* 1 *in the integral in display (9) is unimportant and may be replaced by any positive number. Then the function* η *will change as well.*

The proof of Proposition 1 as well as Theorem 1 are inspired from the proofs of Lemma 3.10.1, Proposition 3.10.3 and Theorem 3.10.8 in Bingham et al (1987) and will be omitted. Let us also mention that Gaïffas (2005) proposed to model locally the density of sparse data by gamma-varying functions. This is another statistical application for Γ_0. It is simple to construct explicit examples of functions in Γ_0 by tuning the auxiliary function ρ and taking $\eta(\cdot) = 0$ in (9). For instance taking $\rho_1(t) = t^m$ (with $m > 1$) gives $\mathcal{F}_1(s) = \exp\left(-1/s^{m-1}\right)$. Now taking $\rho_2(t) = -t/\log(t)$ yields $\mathcal{F}_2(s) = \exp\left(-[\log(s)]^2\right)$. Obviously constants may be added in front of or within the exponential. The next Proposition seems to show a specific feature of the class Γ_0.

Proposition 2 *Let* F *belong to* Γ_0. *Then for all integer* p $F^{(p)}(0) = 0$ *where* $F^{(p)}$ *denotes the derivative of order* p *of* F.

1.3 The nonparametric regression model for functional data

As a last part of this introduction we shift from small ball problems and extreme theory to statistics for functional data. This recent domain of statistics has been receiving increasing interest and was boosted by computational advances. We briefly recall that the main purpose of functional data analysis (FDA) is to model and study datasets where observations are of functional nature (usually observed on a grid then smoothed, approximated and reconstructed by projection on accurate basis). We refer to the monographs by Ramsay and Silverman (2005) and Ferraty and Vieu (2006) for an overview of this topic. Along the past decade some authors turned their attention to the question of modelizing probability distribution for curve-data with applications in statistics : Dabo-Niang (2002), Hall and Heckman (2002) Delaigle and Hall

(2010) in a general setting then Dabo-Niang, Ferraty and Vieu (2004 and 2006), Ferraty, Goïa and Vieu (2007) with applications to classifications through modal curves for instance. Consider the regression problem with functional data as inputs :

$$y = r(X) + \varepsilon \tag{11}$$

where y, ε are real with ε centered whose variance is denoted σ_ε^2, X belongs to the Hilbert space $\mathcal{H}$ and r is a function from $\mathcal{H}$ to $\mathbb{R}$. The space $\mathcal{H}$ may be chosen to be $L^2(T)$ where T is a compact set in the Euclidean space or som Sobolev space $\mathbf{H}^{2,m}$. It is endowed with an inner product $\langle \cdot, \cdot \rangle$ inducing a norm $\|\cdot\|$. Estimating the regression function at a fixed point x_0 namely $r(x_0) = \mathbb{E}(y|X = x_0)$ is possible by a classical Nadarya-Watson approach (see Tsybakov (2004) for a general presentation in the finite dimensional setting and Ferraty Vieu (2006) for implementation on functional data). This model was studied for instance in Ferraty, Vieu (2004) and asymptotic results were derived in Ferraty, Mas, Vieu (2007) like a first upper bound for the quadratic risk. It seems that an equivalent of projection-based estimate in this model has not been introduced yet, certainly due to a lack of theoretical results on approximation theory for functions defined on a Hilbert space. The linear regression model $y = \int X(s)\beta s + \varepsilon$ has been extensively investigated in the last years and several authors proved optimality results like for instance Hall and Horowitz (2007), Crambes, Kneip, Sarda (2009) or Cardot and Johannes (2010) (see also references therein these works). It seems that the optimal (in minimax sense) asymptotic risk has not been obtained yet in the more general model (11). The behaviour of the small ball probability was a stumbling stone hard to circumvent.

An adapted Nadaraya-Watson estimate reads :

$$\widehat{r}(x_0) = \frac{\sum_{i=1}^n y_i K(\|X_i - x_0\|/h)}{\sum_{i=1}^n K(\|X_i - x_0\|/h)}$$

where x_0 is a fixed point of the space, K is a kernel, that is a mesurable, unilateral (defined on $\mathbb{R}^+$) positive function with $\int K = 1$ and h is a nonnegative number tending to 0 (the bandwidth). Considering the L^2-risk at a fixed point x_0 leads to a bias-variance decomposition :

$$\mathcal{R}_n(x_0) = \mathbb{E}\left(\widehat{r}(x_0) - r(x_0)^2\right) = \mathcal{B}_n(x_0) + \mathcal{V}_n(x_0)$$

with

$$\mathcal{B}_n(x_0) = \left\{\mathbb{E}\frac{\sum_{i=1}^n [r(X_i) - r(x_0)] K(\|X_i - x_0\|/h)}{\sum_{i=1}^n K(\|X_i - x_0\|/h)}\right\}^2 \tag{12}$$

$$\mathcal{V}_n(x_0) = \mathbb{E}\left[\frac{\sum_{i=1}^n \varepsilon_i K(\|X_i - x_0\|/h)}{\sum_{i=1}^n K(\|X_i - x_0\|/h)}\right]^2 \tag{13}$$

where $\varepsilon_i = y_i - r(X_i)$.

Lemma 1 *The following holds for the two components of the risk at a fixed point x_0 of the kernel estimator $\widehat{r}(x_0)$:*

$$\mathcal{V}_n(x_0) \sim \frac{\sigma_\varepsilon^2}{n} \frac{\mathbb{E}K^2(\|X - x_0\|/h)}{[\mathbb{E}K(\|X - x_0\|/h)]^2} \tag{14}$$

$$\mathcal{B}_n(x_0) \sim \frac{1}{\mathbb{E}^2 K(\|X - x_0\|/h)} \left(\sum_{i=1}^{+\infty} \mathbf{b}_i \mathbb{E}\left[\langle X, e_i\rangle^2 K(\|X\|/h)\right]\right)^2 \tag{15}$$

where the $\mathbf{b}_i$'s are positive and non random constants.

The sequence $(\mathbf{b}_i)_{i\in\mathbb{N}}$ is not given here because it depends on several parameters which will be introduced later. The proof of this lemma will not be explicitely carried out. It will be encapsulated in the proof of Proposition 5 which is more precise about the bounds (14) and (15). We keep in mind that the bias-variance decomposition of the risk is essentially based on the computation of two sorts of moments : $\mathbb{E}K\left(\|X-x_0\|/h\right)$ and $\mathbb{E}\left[\langle X,e_i\rangle^2 K\left(\|X\|/h\right)\right]$. Calculation of $\mathbb{E}K^2\left(\|X-x_0\|/h\right)$ is similar with $\mathbb{E}K\left(\|X-x_0\|/h\right)$. In a multivariate setting, when X is an $\mathbb{R}^d$ valued random variable, and the density of X f_X is smooth enough at x_0 computations lead in many situations to :

$$\mathbb{E}K\left(\|X-x_0\|/h\right)\sim c_d f_X\left(x_0\right)h^d \tag{16}$$

where c_d denotes the volume of the unit ball in the space $\mathbb{R}^d$. The r.h.s. of the formula above may vary, depending on the support of the distribution of X. However neither Lebesgue's measure or a counterpart to f_X may be defined when X is valued in a Hilbert space for instance. The classical notion of volume of a ball cannot be generalized to such spaces. As a consequence when X is a process, the density of X at x_0 does not make sense anymore. A major issue is then to compute the preceding expectation without assuming that $f_X\left(x_0\right)$ exists. We consider the following conditions on the kernel K :

K has compact support (say $[0,1]$), is absolutely continuous and bounded above and below with $K\left(1\right)>0$

These conditions hold for the naive kernel, $K\left(u\right)=1$ if and only if $u\in[0,1]$. We do not seek minimal conditions on the kernel here and the assumption above could certainly be alleviated but is sufficient to carry out computations. Applying Fubini's Theorem is sufficient to get rid of the density. Denoting $\mathbb{P}\left(\|X-x_0\|<h\right)=F_{x_0}\left(h\right)$ we obtain :

$$\mathbb{E}K\left(\frac{\|X-x_0\|}{h}\right)=F_{x_0}\left(h\right)\left[K\left(1\right)-\int_0^1 K'\left(s\right)\frac{F_{x_0}\left(hs\right)}{F_{x_0}\left(h\right)}ds\right]$$

It is straightforward to see that the same method may yield the value of such integrals as :

$$\mathbb{E}\left[\|X-x_0\|^p K\left(\frac{\|X-x_0\|}{h}\right)\right]=h^p F_{x_0}\left(h\right)\left[K\left(1\right)+\int_0^1 \widetilde{K}_p\left(s\right)\frac{F_{x_0}\left(hs\right)}{F_{x_0}\left(h\right)}ds\right] \tag{17}$$

with $\widetilde{K}_p\left(s\right)=-\left[s^pK'\left(s\right)+ps^{p-1}K\left(s\right)\right]$ and the evaluation of the expectation above essentially depends again on the small ball probability $F_{x_0}\left(\cdot\right)$. When X is a random function the behaviour of F_{x_0} at 0 is crucial and determines the rate of convergence to zero of the above expectation -what statisticians are truly interested in.

Assume that F_{x_0} is regularly varying at zero with index d (which is usually true when X is finite dimensional) then by definition $F_{x_0}\left(h\right)=Ch^d l\left(h\right)$ where C is a constant, l is a slowly varying function at 0 and $F_{x_0}\left(hs\right)/F_{x_0}\left(h\right)\to s^d$ when for $s>0$ and $h\to 0$ which yields $\mathbb{E}K\left(\|X-x_0\|/h\right)\sim c_d h^d l\left(h\right)$ where c_d depends only on d and K. Unfortunately when X lies in a function space, the most classical examples of $F\left(h\right)$ are not reguarly varying as will be seen below. But however we notice for further purpose that the theory of regular variation is of some help in this important special case.

Turning to $\mathbb{E}\left[\langle X,e_i\rangle^2 K\left(\|X\|/h\right)\right]$ which appears in the numerator of (15) (note that here x_0 does not appear anymore) is more tricky and will not be done at this stage. This expectation is bounded above by $\mathbb{E}\left[\|X\|^2 K\left(\|X\|/h\right)\right]$ similar to (17) with $p=2$ but this boudn is not sharp and no other equivalent could be derived from the previous considerations.

2 Main results

We are ready to give the main results. This section is split in three parts. In the first it is shown that the function $F_{x_0}(\cdot)$ which is crucial for evaluating the risk in model (11) belongs to the class Γ_0 of Gamma-varying functions in a quite general framework. In the second we focus on the case of a Gaussian design. In the third we use the properties of the class Γ_0 to derive upper and lower bounds on the risk for (11) and at a fixed point. A notable fact is that the lower bound is degenerate : it is slower than any negative power of n. This may be seen as an ultimate symptom of the curse of dimensionality. If f and g are two positive functions the notation $f \preceq_x g$ means that $\lim_{u\to x} f(u)/g(u) \leq c$ for some positive constant c.

2.1 Small ball probability of random functions are Gamma-varying

This sub-section connects the two apparently distinct notions of probability seen before : the class of small ball probabilities in l_2 and de Haan's Gamma class of functions. Both families of functions are defined by their local behaviour around 0. In what follows, the exponent -1 is strictly reserved to denoting the generalized inverse of a function f denoted f^{-1}. Consequently in general $f^{-1} \neq 1/f$. Let us introduce the function $\lambda(\cdot)$ which interpolates the λ_j's a smooth way (which means that $\lambda(j) = \lambda_j$ for all j and λ is $\mathbf{C}^1$).

Since our results rely on those of Lisfhits (1997) we recall now the assumptions needed in this article. Let G denote the (cumulative) distribution function of Z then we assume that there exists $b \in (0,1)$, $c_1 > 1$, $c_2 \in (0,1)$ and $c_3 > 0$ such that for $r < c_3$:

$$\mathbf{A}_0 : \begin{cases} G(r) \leq c_1 G(br) \\ G(br) \leq c_2 G(r) \\ \mathbb{E}Z^3 < +\infty \end{cases} \tag{18}$$

As mentioned in Lifhsits (1997) assumption $\mathbf{A}_0$ states that the local behaviour at 0 of G is polynomial and $\mathbf{A}_0$ holds whenever the density g of Z is regularly varying at 0 with index $\alpha > -1$. We also note that the assumption above holds for a large class of classical positive distributions of Z itself (Gamma, Beta...) or when $Z = X^2$ with X Gaussian, X Laplace, Uniform or Student distributions for instance. These considerations are of interest for the statistician in order not to limit the approach to Gaussian models. Note that the assumption on the convergence of the third order moment of Z was alleviated in some recent papers. We keep it here since it is general enough for our purpose.

When $(Z_i)_{i\in\mathbb{N}}$ is a sequence of random variables whose cumulative distribution function G is regularly varying at 0 with strictly positive index, the explicit form of the small ball probability was derived for explicit sequences of log convex $\lambda(\cdot)$ by Dunker, Lifshits, Linde (1998). In particular they show that when $\lambda_i = i^{-\beta}$ $(\beta > 1)$, $\mathbb{P}\left(\|X\|^2 < s\right) \sim F_1(s)$ and that when $\lambda_i = \exp(-i)$ $\mathbb{P}\left(\|X\|^2 < s\right) \sim F_2(s)$ with :

$$F_1(s) = c_1 s^{[1+\beta(2+c_2)]/(2\beta-2)} \exp\left(-c_3 s^{-1/(\beta-1)}\right) \tag{19}$$

$$F_2(s) \sim c_4 \left[s^{1/3}\log(1/s)\right]^{-3/4} \exp\left(-\left[\log(s\log 1/s)\right]^2/4 + \psi_0(\log(s\log 1/s))\right) \tag{20}$$

where ψ_0 is a bounded function. Formula (19) is proved as well at page 269 in Lifshits (1995). Simple algebra proves that both functions on the right hand side of (19) and (20) have all their derivatives vanishing at 0. We notice that the r.h.s. of (19) is always flatter than the r.h.s. of (20) which in turn will always be flatter at 0 than any polynomial function (like $c_d s^d$). However we notice that the degree of flatness is directly connected with the rate of decrease of the λ_i's which quantifies, exactly like the l-numbers, the accuracy of a finite-dimensional approximation of X. We emphasize the following Proposition, which will not be proved, on purpose.

Proposition 3 *Both functions F_1 and F_2 defined above at (19) and (20) belong to Γ_0 with respective auxiliary functions $\rho_1(s) \sim s^{\beta/(\beta-1)}$ and $\rho_2(s) \sim s \log(1/s)$ which both match Proposition 1.*

The auxiliary functions ρ_1 and ρ_2 could be more precisely computed but we only need equivalencies at this stage.
We are ready to extend this fact to general sequences $(\lambda_i)_{i\in\mathbb{N}}$. Remind that the function $\gamma(\cdot)$ was defined implicitely at display (6). In words it is, up to sign, the inverse of the first order derivative of the log-Laplace transform of $S = \sum_{i=1}^{+\infty} \lambda_i Z_i$.

Theorem 2 *Let S be defined by (4) and set $\mathbb{P}(S < s) = F(s)$ the small ball probability of S then $F \in \Gamma_0$ with auxiliary function :*

$$\rho(s) = \frac{1}{\gamma(s)} \tag{21}$$

and the representation (5) may be rephrased only in terms of $\gamma(\cdot)$:

$$\mathbb{P}(S < r) \underset{r\to 0}{\sim} \frac{1}{\sqrt{2\pi}} \frac{\sqrt{-\gamma'(r)}}{\gamma(r)} \exp\left[-\int_r^{r_0} \gamma(s)\, ds\right] \tag{22}$$

where $r_0 = \mathbb{E}Z \cdot \sum_{j=1}^{+\infty} \lambda_j$.

Obviously the r.h.s. of (22) is mathematically the same object as the r.h.s. of (5). The "Gamma-varying version" of the r.h.s. is $\sqrt{\rho'(r)/\pi} \exp\left[-\int_r^{r_0} ds/\rho(s)\right]$. We believe however that this new version is slightly more explicit and maybe more suited for statistical purposes. We can take advantage as well of the properties of the class Γ_0 listed earlier.

The Theorem may be intuitively explained in view of Proposition 2. Indeed when X lies in $\mathbb{R}^d$ and in a general context $F(s) \sim_0 p_d(s) = c_d s^d$. The function p_d has the following property : $p_d^{(k)}(0) = 0$ whenever $k \neq d$. Consequently in an infinite dimensional space we can expect that all the derivatives at 0 should be null and this property is recovered through Proposition 2. A more geometric way to understand this consists in considering the problem of the concentration of a probability measure. Let μ be the measure associated with the random variable X. Once again starting from $\mathbb{R}^d$ and letting d increase -even if this approach is not really fair- we see that μ must allocate a constant mass of 1 to a space whose dimension increases. Then μ gets more and more diffuse, allowing fewer mass to balls and visiting rarely fixed points such as x_0 (and their neighborhoods), resulting in a very flat small ball probability function.

The following corollary provides some information about the rate of decrease to zero of $F(\cdot)$ when an additional assumption is made on ρ.

Corollary 1 *Assume that $\rho(s) = s^\alpha l(s)$ with $l(\cdot)$ slowly varying at 0 (which just means that ρ is regularly varying at 0 with index $\alpha \geq 1$ and set :*

$$\mathbf{RV}_+ : \alpha > 1 \text{ or } \mathbf{RV}_1 : \rho(s) = s l(s) \text{ with } l(s) \succeq \log(1/s). \tag{23}$$

If $\mathbf{RV}_+$ holds $\log \mathbb{P}(S < r) \preceq c_\alpha r^{1-\alpha}$ and when $\mathbf{RV}_1$ holds $\log \mathbb{P}(S < r) \preceq -\varsigma(r) \log(1/r)$ for some $\varsigma(r) \to +\infty$ when $r \to 0$. In both preceding cases for all integer p $\lim_{r\to 0} r^{-p}\mathbb{P}(S < r) = 0$.

This property fo the small ball probability has to be connected with property (10), is referred to as "rapid variation" at 0 in the literature on regular variations and may be compared or opposed with the regularly varying situation discussed below (17). The assumptions $\mathbf{RV}_+$ and $\mathbf{RV}_1$ will be encountered again when addressing the case of nonparametric regression. At last, note that for the auxiliary functions ρ_1 and ρ_2 appearing at Proposition 3 and arising from Dunker, Lifshits, Linde (1998) work we get $\rho_1 \in \mathbf{RV}_+$ and $\rho_2 \in \mathbf{RV}_1$.

Proof of Corollary 1: We focus on the right hand side of (22). First from $\sqrt{-\gamma'(r)}/\gamma(r) = \sqrt{2\rho'(s)}$ and the properties of the auxiliary function ρ at Proposition 1 we have that $\rho'(s) \to 0$. Hence $\mathbb{P}(S < r) \leq \exp\left[-\int_r^{r_0} 1/\rho(s)\,ds\right]$ for r tending to 0. From the direct part of Karamata's theorem's (see Bingham, Goldie Teugels (1991), p.26) $\int_r^{r_0} 1/\rho(s)\,ds \sim c_\alpha r^{1-\alpha}$ when $\alpha > 1$ and $\int_r^{r_0} 1/\rho(s)\,ds = \varsigma(r)\log(1/r)$ with $\varsigma(r) \to +\infty$ when $r \to 0$ (see display (1.5.8) in Bingham, Goldie Teugels (1991)). Finally when $\mathbf{RV}_+$ or $\mathbf{RV}_1$ hold $-p\log r - \int_r^{r_0} 1/\rho(s)\,ds$ always tend to $-\infty$ whatever the choice of p.

Remark 3 *For the sake of completeness we point out the following fact which may be misleading : indeed we started from* $\mathbb{P}\left(\|X\|^2 < r\right)$ *and the properties of this function may differ from those of what may be intended as the true small ball probability* $\mathbb{P}\left(\|X\|^2 < r^2\right)$. *It is not difficult to show that if* $F \in \Gamma_0$ *with auxiliary function* ρ_F *then* $G(r) = F\left(r^2\right)$ *belongs to* Γ_0 *as well with auxiliary function* ρ_G *defined by* $\rho_G(r) = \rho_F\left(r^2\right)/(2r)$.

2.2 Gaussian framework

Assuming that X is Gaussian, hence that x_i are $\mathcal{N}(0,1)$ distributed provides a critical amount of extra information. Indeed it is then possible to compute in a more explicit form :

$$r = -\frac{\partial \log \Lambda(\gamma)}{\partial \gamma} = \sum_j \frac{\lambda_j}{1 + 2\gamma\lambda_j} \tag{24}$$

which is the seminal equation linking r and γ. We derive below an explicit link between the λ_j's and $\gamma(\cdot)$ or equivalently $\rho(\cdot)$. Under rather general assumptions on the rate of decrease of the λ_j's we obtain as well an upper bound for the small ball probability which will be exploited in the next subsection when investigating a lower bound for the regression.

Proposition 4 *Assume that* X *is Gaussian, that* $\lambda(\cdot)$ *is a convex decreasing function and set* $\varphi(t) = t\gamma(t)$. *We have the following : There exists a fixed constant* $l \in \left[\frac{2}{3}, \frac{3}{2}\right]$ *such that for any* $\varepsilon > 0$ *and large enough* x

$$\lambda(x(l+\varepsilon)) \leq \frac{1}{\gamma(\varphi^{-1}(x))} = \rho\left(\varphi^{-1}x\right) \leq \lambda(x(l-\varepsilon))$$

Besides when $\lambda(x) \succ_\infty \exp(-x^\alpha)$ *for some* $\alpha > 0$, $F(s) \prec_0 \exp\left[-(\log 1/r)^{1+1/\alpha}\right]$. *When* $\lambda(\cdot)$ *is explicitely known more precise relationships may be derived. For instance when* $\lambda_a(x) = cx^{-1-\nu}$ *with* $c, \nu > 0$, $\gamma_a(s) \sim_0 s^{-1-1/\nu}c^{1/\nu}l^{1+1/\nu}$ *and* $l = \int_0^{+\infty} du/\left(2 + u^{1+\nu}\right)$. *When* $\lambda_g(x) = c\exp(-\nu x)$, *with again* $c, \nu > 0$, $\gamma_g(s) \sim_0 \log(1/s)/(\nu s)$.

Remark 4 *The auxiliary fucntions* $\rho_a = 1/\gamma_a$ *and* $\rho_g = 1/\gamma_g$ *match respectively* ρ_1 *and* ρ_2. *Besides letting* ν *go to infinity we see that, in a way* ρ_g *may be viewed as a limit of* ρ_a. *In fact* $1/\log(1/s)$ *echoes the degeneracy of* $1/(\beta - 1)$. *Proposition 4 modestly rediscover the results of Dunker, Lifshits, Linde (1998). The upper bound of* $F(\cdot)$ *is close to the one obtained in Proposition 1. No asumptions are needed on* ρ *here but the distribution of* X *is Gaussian.*

Proof of Proposition 4 :
We start from (24) and denote $a(\cdot) = 1/\lambda(\cdot)$ which may be rewritten :

$$r = \sum_{j \geq 1} \frac{1}{a(j) + 2\gamma}$$

Let us set $J_\gamma = \inf\{j : a(j) \geq \gamma\}$ so that $a(J_\gamma - 1) \leq \gamma \leq a(J_\gamma)$

$$\frac{J_\gamma}{3\gamma}+\sum_{j\geq J_\gamma+1}\frac{1}{a(j)+2\gamma}\leq\sum_{j\geq 1}\frac{1}{a(j)+2\gamma}\leq\frac{J_\gamma}{2\gamma}+\sum_{j\geq J_\gamma+1}\frac{1}{a(j)+2\gamma}$$
$$\frac{J_\gamma}{3\gamma}+\frac{1}{3}\sum_{j\geq J_\gamma}\frac{1}{a(j)}\leq r\leq\frac{J_\gamma}{2\gamma}+\sum_{j\geq J_\gamma}\frac{1}{a(j)}$$
$$\frac{2}{3}\leq\frac{r\gamma}{J_\gamma}\leq\left(0.5+\left(J_\gamma^{-1}+1\right)\right)$$

The convexity of λ, hence of a yields $\sum_{j\geq J}1/a(j)\leq(J_\gamma+1)/a(J_\gamma)$ (see Cardot, Mas, Sarda (2007)) hence for $r\downarrow 0$ that is for large $\gamma(r)$:

$$\frac{2}{3}\leq\frac{r\gamma}{J_\gamma}\leq\frac{3}{2}\tag{25}$$

Now consider the function (of the variable γ) : $d(\gamma)=\gamma r(\gamma)/a^{-1}(\gamma)$ is decreasing at least when $\gamma\uparrow+\infty$ since :

$$d'(\gamma)=\frac{\left(r(\gamma)-\gamma\sigma^2\right)}{a^{-1}(\gamma)}-\frac{\gamma r(\gamma)}{\left[a^{-1}(\gamma)\right]^2\cdot a'\left(a^{-1}(\gamma)\right)}$$

is negative for $r\downarrow 0$ that is for large γ. Finally we get $r\gamma/a^{-1}(\gamma)\to l\in\left[\frac{2}{3},\frac{3}{2}\right]$ when γ tends to $+\infty$. The first statement of the Theorem is a consequence of the latter limit. As a consequence when $\lambda(x)\succ_\infty\exp(-x^\alpha)$, $a(x)\prec_\infty\exp(x^\alpha)$ and finally $r\gamma\succ_0(\log\gamma)^{1/\alpha}\geq(\log 1/r)^{1/\alpha}$. At last we get $\gamma(r)\succ_0\frac{1}{r}(\log 1/r)^{1/\alpha}$ which finally yields for some constant c

$$\exp\left[-\int_r^{r_0}\gamma(s)\,ds\right]\leq c\exp\left[\int_r^{r_0}(\log 1/s)^{1/\alpha}\,d(\log 1/s)\right]$$
$$=c\exp\left[-(\log 1/r)^{1+1/\alpha}\right]$$

The last sentence of the Theorem , when the function λ is known, is easily derived by noting that $\sum_{j\geq 1}1/(a(j)+2\gamma)\sim_{\gamma\to+\infty}\int_0^{+\infty}\frac{dx}{a(x)+2\gamma}$.

2.3 Upper and lower bound in regression for functional data

We fix once and for all the assumptions considered in what follows. These assumptions appear in addition to those considered in the previous sections. Remind that if g is some function defined on $\mathcal{H}$ and with values in $\mathbb{R}$ the first order Fréchet-derivative of g at x_0 (its infinite-dimensional gradient) may be identified with an element of $\mathcal{H}$. The second order derivative $g''(x_0)$ is identified with a symmetric operator from $\mathcal{H}$ to $\mathcal{H}$.

Assumptions on the distribution of X. The random element X is centered and in the development (3) the x_i's are independent. We have $P_{X-x_0}\ll P_X$ with $f_{x_0}=dP_{X-x_0}/dP_X$ such that $f_{x_0}(0)>0$, $f'(x_0)\in\mathcal{H}$ exists and the second order derivative of f_{x_0} denoted $f''_{x_0}(x)$ is for all x in a neighborhood of x_0 a bounded linear operator from $\mathcal{H}$ to $\mathcal{H}$. Denote $\partial_i f_{x_0}=\langle f'(x_0),e_i\rangle$ where e_i is one of the eigenvectors appearing in (3). Besides we assume that for all i the density of the margins $\langle X,e_i\rangle$ is symmetric.

Assumptions on the regression function. Assume that r has first and second order derivative at x_0. We denote $\partial_i r_{x_0}=\langle r'(x_0),e_i\rangle$ and $\partial^2_{ii}r_{x_0}=\langle r'(x_0)(e_i),e_i\rangle$ and assume as well that :

$$\sum_{i=1}^{+\infty}\lambda_i\partial_i r_{x_0}\partial_i f_{x_0}\neq 0$$

At this point a discussion on the assumptions related to the distribution of X is needed. Take for instance the case of a gaussian X. Chapter 9 and 10 in Lifshits (1995) are clear about these issues (see more specifically p.102-107.) It is possible to shift the assumptions on the regularity of X to conditions on the regularity of x_0. First in order to define $f'(x_0)$ we need to assume that $x_0 = (m_i)_{1\leq i}$ belongs to the kernel of X that is $\sum m_i^2/\lambda_i < +\infty$. Then for any $u = (u_i)_{1\leq i}$ in $\mathcal{H}$ $f_{x_0}(u) = \exp\left(-\sum_{i\geq 1} m_i^2/2\lambda_i + \sum_{i\geq 1} u_i m_i/2\lambda_i\right)$and $\partial_i f_{x_0} = \partial_i f_{x_0}(u) = (m_i/2\lambda_i) f_{x_0}(u)$. We are interested in th smoothness of these functions at 0. From $|f_{x_0}(u) - f_{x_0}(0)| \preceq \|u\| \left(\sum_{i\geq 1} m_i^2/\lambda_i^2\right)^{1/2}$. And the finiteness of the latter series is subject to a condition of decay on the coefficients m_i's.

It turns out that the gaussian framework may be generalized to some other distributions. It suffices to consider a symmetric random variable U such that $\mathbb{E}U = 0$ and $\mathbb{V}U = 1$ with smooth density function at 0.Then introduce the scaled $Z_i = \lambda_i U^2$ in order to derive a new development (3). Examples for U are the uniform distribution on $[-a, a]$, $a > 0$, shifted Beta distributions...

2.3.1 Upper bound

The reader was left with Lemma 1. In view of the results of this section we are in a position to simplify some computations. Turning to the local moments defined at display (17), from properties of functions in Γ_0 and specifically (10) we get

$$\mathbb{E}\left[\|X - x_0\|^p\right] K\left(\frac{\|X - x_0\|}{h}\right) \sim K(1) h^p F_{x_0}(h) \sim K(1) h^p f_{x_0}(0) 8F(h) \qquad (26)$$

We see again that the representation theorem of the preceding section is of some help to simplify our calculations. We mention for immediate purpose that the derivation of both fromula above leads as well to :

$$\mathbb{E}K^2\left(\frac{\|X - x_0\|}{h}\right) \sim K^2(1) F_{x_0}(h) \qquad (27)$$

Let the local moments of order 1 and 2 of X at x_0 be respectively defined by :

$$\mathcal{M}_{K,1}(x_0) = \mathbb{E}\left[(X - x_0) K\left(\frac{\|X - x_0\|}{h}\right)\right] \qquad (28)$$

$$\mathcal{M}_{K,2}(x_0) = \mathbb{E}\left[(X - x_0) \otimes (X - x_0)\right] K\left(\frac{\|X - x_0\|}{h}\right). \qquad (29)$$

Formula (29) may be explicited. First let u and v be two points in the vector space then $u \otimes v$ is a linear operator defined by $[u \otimes v](x) = \langle v, x\rangle y$. Now erasing x_0 and $K(\|X - x_0\|/h)$ gives the usual covariance operator of X for a centered X. The special covariance operator $\mathcal{M}_{K,2}(x_0)$ is obtained by shifting and smoothing X around x_0. Note that $\mathcal{M}_{K,1}$ belongs to $\mathcal{H}$ and $\mathcal{M}_{K,2}$ is a linear trace-class operator acting from and onto $\mathcal{H}$. We refer to Müller and Yan (2001) for some statistical results on local moments for finite-dimensional random variables and to Mas (2008) for some related results dealing with (29) and where random functions and small ball problems appear.

The next Proposition completes Lemma 1.

Proposition 5 *For the variance part of the risk the equivalence holds $\mathcal{V}_n(x_0) \sim \sigma_\varepsilon^2/\left[f_{x_0}(0) nF(h)\right]$. For the bias part we just provide an approximate rate :*

$$\mathbf{c}^- \rho^6(h) \leq \mathcal{B}_n(x_0) \leq \mathbf{c}^+ h^4$$

where $\mathbf{c}^+$ and $\mathbf{c}^-$ depend only on $f_{x_0}(0)$, $f'(x_0)$, $r'(x_0)$ and $r''(x_0)$. When$\rho(h) \succeq h^m$ for some m $\mathcal{B}_n(x_0)$ decreases to 0 at most and at least at a polynomial rate.

The problem here is to ensure a rough control of $\mathcal{B}_n(x_0)$. As will be seen soon $\rho^6(h)$ turns out to be regularly varying in most cases and decays to zero at a polynomial rate. The unusual framework (namely with distributions in the class Γ_0) motivates to prove the reader that $\mathcal{B}_n(x_0)$ does not reach an unusual rate of decreaes to 0 (namely exponential). And the lower bound we obtain for this specific estimator justifies the conditions under which the minimax lower bound is going to be derived.

Proof of Proposition (5) :

We start with $\mathcal{V}_n(x_0)$. It is simple to see that $\mathcal{V}_n(x_0) = n\sigma_\varepsilon^2 \mathbb{E}\omega_{1,n}^2$with :

$$\omega_{1,n} = \frac{K(\|X_1 - x_0\|/h)}{\sum_{i=1}^n K(\|X_i - x_0\|/h)}$$

Computations like those carried out in Ferraty, Mas, Vieu (2007) show that :

$$\mathbb{E}\omega_{1,n}^2 \sim \mathbb{E}K^2(\|X - x_0\|/h) / [n\mathbb{E}K(\|X - x_0\|/h)]^2$$

hence that (see (26)) $\mathcal{V}_n(x_0) \sim \frac{\sigma_\varepsilon^2}{n} \frac{\mathbb{E}K^2(\|X-x_0\|/h)}{[\mathbb{E}K(\|X-x_0\|/h)]^2} \sim \frac{\sigma_\varepsilon^2}{nF_{x_0}(h)}$which yields the desired result by (1).

We should now deal with $\mathcal{B}_n(x_0)$ defined at (12). Ferraty, Mas, Vieu (2007 show that $\mathcal{B}_n(x_0)$ is well approximated by :

$$\left[\frac{\mathbb{E}(r(X) - r(x_0)) K(\|X - x_0\|/h)}{\mathbb{E}K(\|X_1 - x_0\|/h)}\right]^2 \tag{30}$$

A Taylor development at order 2 shows that :

$$\mathbb{E}(r(X) - r(x_0)) K(\|X - x_0\|/h) = \langle r'(x_0), \mathcal{M}_{K,1}(x_0)\rangle + \mathrm{tr}\left[r''(\xi)\mathcal{M}_{K,2}(x_0)\right]/2$$

where r' and r'' stands for the first and second order Gâteaux derivative of r at x_0 and $\xi = \theta X + (1-\theta)x_0$ for some random $\theta \in (0,1)$. We first deal with the first order term $\langle r'(x_0), \mathcal{M}_{K,1}(x_0)\rangle$. From $\mathbb{P}_{X-x_0} \ll \mathbb{P}_X$ we see that :

$$\begin{aligned}\mathbb{E}\left[(X - x_0) K\left(\frac{\|X - x_0\|}{h}\right)\right] &= \mathbb{E}\left[X f_{x_0}(X) K\left(\frac{\|X\|}{h}\right)\right] \\ &= \mathbb{E}\left[X[f_{x_0}(X) - f_{x_0}(0)] K\left(\frac{\|X\|}{h}\right)\right] + f_{x_0}(0)\mathbb{E}\left[XK\left(\frac{\|X\|}{h}\right)\right]\end{aligned}$$

We assumed that f_i the density of x_i is symmetric. This yields for all i :

$$\mathbb{E}\left[\langle X, e_i\rangle K\left(\frac{\|X\|}{h}\right)\right] = \mathbb{E}\left[x_i K\left(\frac{\|X\|}{h}\right)\right] = \int\left(\int tK\left(\sqrt{t^2+s^2}/h\right) f_i(t)\,dt\right) g_i(s)\,ds = 0$$

where g_i si the density of $\sum_{j=1, j\neq i}^{+\infty} \lambda_j x_j^2$. Now

$$\mathbb{E}\left[X[f_{x_0}(X) - f_{x_0}(0)] K\left(\frac{\|X\|}{h}\right)\right] = \mathbb{E}\left[X\langle f'(x_0), X\rangle K\left(\frac{\|X\|}{h}\right)\right] + R_n$$

where R_n involves the second order derivative of f_{x_0} and will be neglected. Then denoting $\partial_i r_{x_0} = \langle r'(x_0), e_i\rangle$

$$\begin{aligned}\langle r'(x_0), \mathcal{M}_{K,1}(x_0)\rangle &\sim \sum_{i=1}^{+\infty} \partial_i r_{x_0} \mathbb{E}\left[\langle X, e_i\rangle\langle f'(x_0), X\rangle K\left(\frac{\|X\|}{h}\right)\right] \\ &= \sum_{i=1}^{+\infty} \partial_i r_{x_0} \mathbb{E}\left[\langle X, e_i\rangle\left(\sum_{j=1}^{+\infty} \partial_j f_{x_0}\langle X, e_j\rangle\right) K\left(\frac{\|X\|}{h}\right)\right]\end{aligned}$$

Arguments based on the symmetry of the density of the $\langle X, e_i\rangle$ lead to cancelling $\mathbb{E}\left[\langle X, e_i\rangle \langle X, e_j\rangle K\left(\frac{\|X\|}{h}\right)\right]$ for $i \neq j$ and :

$$\left\langle r'(x_0), \mathcal{M}_{K,1}(x_0)\right\rangle \sim \sum_{i=1}^{+\infty} \partial_i r_{x_0} \partial_i f_{x_0} \mathbb{E}\left[\langle X, e_i\rangle^2 K\left(\frac{\|X\|}{h}\right)\right]$$

Similar calculations show that :

$$\operatorname{tr}\left[r''(x_0)\mathcal{M}_{K,2}(x_0)\right] \sim f_{x_0}(0)\sum_{i=1}^{+\infty} \partial_{ii}^2 r_{x_0} \mathbb{E}\left[\langle X, e_i\rangle^2 K\left(\frac{\|X\|}{h}\right)\right]$$

where $\partial_{ii}^2 r_{x_0} = \langle r''(x_0) e_i, e_i\rangle$ and finally denoting $\mathbb{E}\left[\langle X, e_i\rangle^2 K(\|X\|/h)\right] = \mathbf{v}_i(h)$

$$\mathbb{E}(r(X) - r(x_0)) K(\|X - x_0\|/h) \sim \sum_{i=1}^{+\infty}\left(\frac{\partial_{ii}^2 r_{x_0} f_{x_0}(0)}{2} + \partial_i r_{x_0}\partial_i f_{x_0}\right)\mathbf{v}_i(h)$$

We can confine now derive an upper bound. Indeed for $h \downarrow 0$,

$$\begin{aligned}
|\mathbb{E}(r(X) - r(x_0)) K(\|X - x_0\|/h)| &\leq 2\sup_i\left\{\frac{\partial_{ii}^2 r_{x_0} f_{x_0}(0)}{2} + |\partial_i r_{x_0}\partial_i f_{x_0}|\right\}\sum_{i=1}^{+\infty}\mathbf{v}_i(h) \\
&\leq 2\sup_i\left\{\frac{\partial_{ii}^2 r_{x_0} f_{x_0}(0)}{2} + |\partial_i r_{x_0}\partial_i f_{x_0}|\right\}\mathbb{E}\left[\|X\|^2 K(\|X\|/h)\right] \\
&\leq 2K(1)\sup_i\left\{\frac{\partial_{ii}^2 r_{x_0} f_{x_0}(0)}{2} + |\partial_i r_{x_0}\partial_i f_{x_0}|\right\} h^2 F(h)
\end{aligned}$$

This together with (30) and (26) leads to the upper bound of the Proposition with $\mathbf{c}^+ = 8\left[\sup_i\left\{\partial_{ii}^2 r_{x_0} f_{x_0}(0)/2 + |\partial_i r_{x_0}\partial_i f_{x_0}|\right\}/f_{x_0}(0)\right]^2$.

We turn to the lower bound. Since $r''(\xi)\mathcal{M}_{K,2}(x_0)$ is a positive operator we may confine ourselves to the first term. For simplicity we will make calculations with the naive kernel for K and with a modified norm. In fact we will take $\|X\| = |\langle X, e_i\rangle| + \left|\sum_{j\neq i}\langle X, e_j\rangle\right| = |\langle X, e_i\rangle| + Z_i$. Let $0 < \underline{c} < \overline{c}$ be two constants :

$$\begin{aligned}
\mathbb{E}\left[\langle X, e_i\rangle^2 K(\|X\|/h)\right] &\geq \lambda_i\underline{c}\rho^2(h)\,\mathbb{P}\left(\sqrt{\lambda_i\underline{c}}\rho(h) \leq |\langle X, e_i\rangle| \leq \sqrt{\lambda_i\overline{c}}\rho(h), |\langle X, e_i\rangle| + Z_i \leq h\right) \\
&\geq \lambda_i\underline{c}\rho^2(h)\,\mathbb{P}\left(|\langle X, e_i\rangle| \in \sqrt{\lambda_i}\rho(h)\left[\sqrt{\underline{c}}, \sqrt{\overline{c}}\right]\right)\mathbb{P}\left(Z_i \leq h - \sqrt{\overline{c}\lambda_i}\rho(h)\right) \\
&\geq \lambda_i\underline{c}\rho^2(h)\,\mathbb{P}\left(|\langle X, e_i\rangle| \in \sqrt{\lambda_i}\rho(h)\left[\sqrt{\underline{c}}, \sqrt{\overline{c}}\right]\right)\mathbb{P}\left(\|X\| \leq h - \sqrt{\overline{c}\lambda_i}\rho(h)\right)
\end{aligned}$$

where the probabilities were split because $|\langle X, e_i\rangle|$ and Z_i are independent. Consider first $\mathbb{P}\left(|\langle X, e_i\rangle|/\sqrt{\lambda_i} \in \rho(h)\left[\sqrt{\underline{c}}, \sqrt{\overline{c}}\right]\right) \geq \mathbf{c}\rho(h)$ where $\mathbf{c}$ is some constant independent of i if the distribution of all $|\langle X, e_i\rangle|/\sqrt{\lambda_i}$ is bounded below in a neighborhood of 0 which will be assumed here (it is true when X is gaussian). Then lastly :

$$F\left(h - \sqrt{\overline{c}\lambda_i}\rho(h)\right) \geq F\left(h - \sqrt{\overline{c}\lambda_1}\rho(h)\right)$$

which yields :

$$\mathbb{E}\left[\langle X, e_i\rangle^2 K(\|X\|/h)\right] \geq \lambda_i\underline{c}\mathbf{c}\rho^3(h)\,\mathbb{P}\left(\|X\| \leq h - \sqrt{\overline{c}\lambda_1}\rho(h)\right)$$

From $F\left(h-\sqrt{\overline{c}\lambda_1}\rho\left(h\right)\right)/F\left(h\right)\rightarrow\exp\left(-\sqrt{\overline{c}\lambda_1}\right)$ we get that for a h close enough to 0 $\mathbb{E}\left[\langle X,e_i\rangle^2 K\left(\|X\|/h\right)\right]\geq\lambda_i\mathbf{c}''\rho^3\left(h\right)F\left(h\right)$ where $\mathbf{c}''$ does not depend on n or on i. Finally we get :

$$\begin{aligned}&\left[\frac{\mathbb{E}\left(r\left(X\right)-r\left(x_0\right)\right)K\left(\|X-x_0\|/h\right)}{\mathbb{E}K\left(\|X_1-x_0\|/h\right)}\right]^2\\&\geq\rho^6\left(h\right)\left[\mathbf{c}''\sum_{i=1}^{+\infty}\lambda_i\partial_i r_{x_0}\partial_i f_{x_0}/f_{x_0}\left(0\right)\right]^2\\&=\mathbf{c}^-\rho^6\left(h\right)\end{aligned}$$

since $\sum_{i=1}^{+\infty}\lambda_i\partial_i r_{x_0}\partial_i f_{x_0}\neq 0$.

2.3.2 Lower Bound

From the preceding subsection the optimal risk for the kernel estimate is obtained by selecting an h balancing the trade-off between variance and bias. Imagine that we had found in Proposition 5 a result such as $\mathcal{B}_n\left(x_0\right)\asymp F^{\kappa}\left(h\right)$ for some $\kappa>0$. Then the optimal bandwidth would stem from $n^{-1}\asymp F^{1+\kappa}\left(h\right)$ leading to a $\mathcal{R}_n\asymp n^{-\kappa/(1+\kappa)}$ which would contradict the initial claim of degenerate rate for the risk. This explains why we spend some energy in delivering the lower bound on $\mathcal{B}_n\left(x_0\right)$ in Proposition 5. As will be seen now when r belongs to a class large enough to inherit classical approximation features, $\mathcal{R}_n$ cannot decrease at a ploynomial rate. What we mean by classical approximation features is explicited now.

Let $\mathcal{E}_p$ denote any class of $\mathbb{R}$-valued functions defined on $\mathcal{H}$ such that :

$$\sup_{r\in\mathcal{E}_p}\mathcal{B}_n\left(x_0\right)\preceq h^{2p}\tag{31}$$

For instance $\mathcal{E}_p$ may be the class of Hölder functions of order $p\in\left]0,1\right[$. When $\mathcal{E}_p$ is the class of function which have two derivatives at x_0 we see from Proposition 5 that (31) holds for some $p>0$ when $\rho\left(h\right)\geq h^m$ $m>1$. Optimizing the bias-variance trade-off in the risk leads to choosing an h such that $\sup_{r\in\mathcal{E}_p}\mathcal{B}_n\left(x_0\right)=\mathcal{V}_n\left(x_0\right)$. The next Lemma deals with this issue.

Lemma 2 *Assume that X is gaussian and $\lambda\left(x\right)\succ_{\infty}\exp\left(-x^{\alpha}\right)$ for some $\alpha>0$. Let c^* be some constant and h^* be the solution of the functional equation :*

$$\frac{1}{n}=c^*h^{2p}F\left(h\right)\tag{32}$$

then $n^{\beta}/\left(nF\left(h^*\right)\right)\rightarrow+\infty$ for any $\beta>0$. When X is non Gaussian but satisfies the assumptions (18) and (23) the same conclusion holds.

Proof of the Lemma : Only the case $0<\beta<1$ has to be investigated. When X is Gaussian the lemma is easily derived from Proposition 4 since it was proved that $F\left(h\right)\prec_0\exp\left[-\left(\log 1/h\right)^{1+1/\alpha}\right]$ holds. When X is not gaussian and $\mathbf{RV}_1$ holds the proof of Corollary 1 shows that $\beta\log n+2p\log h^*>\beta\varsigma\left(h^*\right)\log\left(1/h^*\right)-2p\log h^*$where $\varsigma\left(h^*\right)$ tends to $+\infty$ when h^* tends to 0. When $\mathbf{RV}_+$ holds the proof is the same with $c_{\alpha}\left(h^*\right)^{1-\alpha}$ instead of $\varsigma\left(h^*\right)\log\left(1/h^*\right)$.

Now our approach to derive lower bounds for the minimax risk follows Tsybakov's scheme (see Tsybakov (2004)) : we construct two models r_0 and r_1 far enough from each other but such that the Hellinger distance between the two models is bounded. Let p_{ε} stand for the density of ε. Assume that for some constant p_* and for all $y\in\mathbb{R}$,

$$\int_{\mathbb{R}}\left[\sqrt{p_{\varepsilon}\left(t\right)}-\sqrt{\left[p_{\varepsilon}\left(t+y\right)\right]}\right]^2dt\leq p_*y^2\tag{33}$$

This assumption is general and appears in Tsybakov's book. It holds under smoothness assumptions on p_ε. We comment it briefly. If $\Lambda(y)$ denotes the left hand side in the display above $\Lambda(y) \leq 2$ for all y and we just need to study Λ on a compact neighborhood around 0 (up to a rescaling through the constant p_*). We see that $\Lambda(0) = 0$ and $\Lambda'(y) = -\int p'_\varepsilon(t+y)\sqrt{p_\varepsilon(t)/p_\varepsilon(t+y)}dt$ whenever p_ε is smooth enough hence $\Lambda'(0) = 0$. Under accurate conditions on Λ'', $\Lambda(y) \leq p_* y^2$ will hold around 0 hence everywhere.

Theorem 3 ***Part I :*** *Assume that X is Gaussian, $\lambda(\cdot)$ is a convex decreasing function with $\lambda(x) \succ_\infty \exp(-x^\alpha)$ for some $\alpha > 0$ and that 33 holds. Denote T_n any estimator of the regression function at a fixed point $r(x_0) = E(y|X = x_0)$ and $\mathcal{R}_n$ the minimax risk over the class $\mathcal{E}_p$ defined in (31) :*

$$\mathcal{R}_n = \min_{T_n} \sup_{r\in\mathcal{E}_p} \mathbb{E}\left[T_n - r(x_0)\right]^2$$

then $\mathcal{R}_n \succ \exp\left[-c(\log n)^{1-1/(\alpha+1)}\right]$ which imples $n^\beta \mathcal{R}_n \to +\infty$ for any $\beta > 0$ but $(\log n)^\beta \mathcal{R}_n \to 0$ for any $\beta > 0$. Strengthening the assumptions on λ and taking $\lambda(x) \succ_\infty x^{-\alpha}$ for some $\alpha > 1$ then $\mathcal{R}_n \succ (\log n)^{-(\alpha-1)}$.
Part II : *Let X be non gaussian but satisfy the conditions (18). Let ρ be the auxiliary function of the sall ball probability of X. Assume that ρ is regularly varying at 0 with index $\alpha \geq 1$ with either $\alpha > 1$ or $\alpha = 1$ and $\rho(s)/s \succeq \log(1/s)$ then again $n^\beta \mathcal{R}_n \to +\infty$ for any $\beta > 0$.*

In Part II we recall for the sake of completeness the conditions **RV**. introduced sooner. The theorem above shows that it is not possible to estimate the regression function in a nonparametric model with functional inputs at a polynomial rate. The rates may be considered as degenerate even when the functional variable X is very smooth (case $\lambda(x) = \exp(-x^\alpha)$ for some $\alpha > 0$) and the data concentrated close to a finite-dimensional space. In the classical situations of polynomial decay, $\lambda(x) \simeq x^{-\alpha}$ for some $\alpha > 1$ the situation gets even worse and the optimal rate we may recover is logarithmic. These negative results are clearly connected with the complexity of the setting : the general nonparametric model coupled with the sparsity of functional spaces already mentioned in the paragraph below Proposition 2.

Remark 5 *Other classes of regression functions could be considered. Here $\mathcal{E}_p$ was considered because calculations are possible when looking for an upper bound. However the theorem above holds, up to a change of constants when r blongs to a class $\mathcal{E}_p$ for which :*

$$\sup_{r\in\mathcal{E}_p} \mathbb{E}\left(r(X) - r(x_0)\right) K\left(\|X - x_0\|/h\right) \asymp h^p F(h)$$

Like in a finite-dimensional framework, obtaining large values of p switches the problem to defining higher order kernels designed for functional data. This issue is out of the scope of this work. Yet, because of the degeneracy of the convergence rate we are not sure it deserves much attention in this setting.

Proof of Theorem 3:

The proof comes down to adapting Tsybakov (2004, Chapter 2, p.81) to our framework. We consider two distant hypotheses : $r_0 \equiv 0$ and $r_1(x) = 2(h^*)^p \mathcal{K}(\|x - x_0\|/h)$ with $\mathcal{K} \in \mathcal{E}_p$ and compactly supported. Here h^* is the solution of the equation (32). It is plain that $|r_0(x_0) - r_1(x_0)| = 2(h^*)^p$. Set $z_i^0 = (y_i^0, X_i)$ and $z_i^1 = (y_i^1, X_i)$, denote $\mathbb{P}_0$ (resp. $\mathbb{P}_1$) the distribution of the vector $(z_1^0, ..., z_n^0)$ (resp $(z_1^1, ..., z_n^1)$) when the regression function is r_0 (resp. r_1) and $\mathbb{P}_{0,i}$ (resp. $\mathbb{P}_{1,i}$) the distribution of the margin z_i^0 (resp. z_i^1). We are going to prove that the Hellinger-distance between $\mathbb{P}_0$ and $\mathbb{P}_1$ $\mathbf{H}(\mathbb{P}_0, \mathbb{P}_1)$ is less than a given $\tau < +\infty$. Let f stand for the density of $U = \|X - x_0\|/h$. The function f is nothing but the first order derivative of the small ball probability F. We first compute the Hellinger distance between the margins of $\mathbb{P}_0$ and $\mathbb{P}_1$ by conditioning with respect to U. Let $\theta_1(U) = 2(h^*)^p \mathcal{K}(U)$:

$$\begin{aligned}\mathbf{H}^2(\mathbb{P}_{0,i}, \mathbb{P}_{1,i}) &\equiv \int\int \left[p_\varepsilon^{1/2}(t) - p_\varepsilon^{1/2}(t-\theta_1(u))\right]^2 f(u)\,dtdu \\ &\leq p_* \int \theta_1^2(u) f(u)\,du = 4p_*(h^*)^{2p}\mathbb{E}\mathcal{K}^2(U)\end{aligned}$$

by Assumption (33). For n large enough and by (27) we deduce that :

$$\begin{aligned}\mathbf{H}^2(\mathbb{P}_{0,i}, \mathbb{P}_{1,i}) &\leq 8p_*(h^*)^{2p} F(h^*)\mathcal{K}^2(1) \\ &\leq \mathbf{c}^*/n\end{aligned}$$

where $\mathbf{c}^*$ is some constant and we used (32). The decomposition of Hellinger distance for product measures (see Tsybakov (2004) p. 69). gives

$$\begin{aligned}\mathbf{H}^2(\mathbb{P}_0, \mathbb{P}_1) &= 2\left(1 - \left(1 - \frac{\mathbf{H}^2(\mathbb{P}_{0,1}, \mathbb{P}_{1,1})}{2}\right)^n\right) \\ &\leq 2\left(1 - \left(1 - \frac{\mathbf{c}^*}{2n}\right)^n\right) \leq 2\left(1 - \exp\left(\frac{\mathbf{c}^*}{4}\right)\right)\end{aligned}$$

and $\mathbf{H}^2(\mathbb{P}_0, \mathbb{P}_1) \leq \tau$ with $\tau = 2(1 - \exp(\mathbf{c}^*/4))$ which almost finishes the proof of the Theorem. The last sentence is proved with the same techniques and in view of Proposition 4.

3 Complementary facts

In this short section are collected results of secondary interest. They complete however the precedings by underlining some facts about the non-unicity and the limits of the representation obtained above. Indeed the preceding theorems lead to the following question : is it possible to obtain a one to one representation, in a general framework, of the small ball probabilities of random elements in l_2 -characterized by the sequence $(\lambda_i)_{i\in\mathbb{N}}$- by a function in Γ_0, depending solely on its auxiliary function ρ ? The answer is negative for at least two reasons. First it is plain that two series S and S' built from different sequences $(\lambda_i, Z_i)_{i\in\mathbb{N}}$ may have equivalent (at 0) small ball probabilities. Second, imagine that we confine to Gaussian small ball probabilities and consider again the r.h.s. of (22) denoted $F \in \Gamma_0$ with auxiliary function ρ. It is plain to see that any function ϕF where $\phi(x + t\rho(x))/\phi(x) \to 1$ when $x \to 0$ belongs to Γ_0 with exactly the same auxiliary function ρ. Consequently even fixing the distribution of the sequence Z_i is not sufficient to obtain a one to one mapping between small ball probabilities and the set Γ_0.

Indeed, pick a function F_0 in the class Γ_0. This function is essentially defined by its auxiliary $\rho_0(\cdot)$ and Theorem 1 is not precise enough for us to identify it with a small ball probability. This is due to the non-unicity of ρ mentioned just under (9) by the words "up to asymptotic equivalence". If $\rho_1 \sim_0 \rho_0$ $\lim_{s\downarrow 0^+} F_0(s + x\rho_1(s))/F_0(s) = \exp(x)$ as well. But the local behaviour at 0 of $F_1(s) = \exp\left\{\eta(s) - \int_s^1 1/\rho_1(t)\,dt\right\}$ may differ from $F_0(s)$ and F_0 may not be equivalent with F_1. What we show below is that if F_0 is accurately scaled we may deduce from F_0 a new function F_0^* which has the same auxiliary function as F_0 (but which may not be equivalent to F_0) and such that for a well-chosen sequence $(\lambda_i)_{i\in\mathbb{N}}$ and the Gaussian small ball probability $\mathbb{P}(S < r)$ is such that $\mathbb{P}(S < r) \sim_0 F_0^*(r)$

We start with a definition which seems to be new.

Definition 2 *Let ρ be a self-neglecting function. A function ϕ is called ρ-self-neglecting if :*

$$\frac{\phi(x + t\rho(x))}{\phi(x)} \underset{x\to 0}{\to} 1.$$

It is obvious that, if ϕ is ρ-self-neglecting it is ρ^*-self-neglecting whenever $\rho^* \sim_0 \rho$. We propose below in Theorem (5) a representation theorem for ρ-self-neglecting functions.

Definition 3 *Pick a ρ_0 in the class of self-neglecting functions at 0 such that $\rho_0(0) = 0$. We define the equivalence class of a function $F_0 \in \Gamma_0$ with auxiliary function ρ_0 by the relationship $\triangle$ defined for all G in Γ_0 by :*

$$F_0 \triangle G \Leftrightarrow \frac{F_0}{G} \text{ is } \rho - \textit{self-neglecting for some } \rho \sim_0 \rho_0$$

Remind that $\varphi(t) = t\gamma(t)$.

Theorem 4 *Let $F_0 \in \Gamma_0$ with auxiliary function $\rho_0 = 1/\gamma_0$. Assume that ρ_0 is regularly varying at 0 with index $\kappa > 1$ and $\mathbf{C}^1$ in a neighborhood of 0. Consider the equivalence class of F_0 in $\Gamma_0 \backslash \triangle$ say $\mathbf{F}_0$. Then one may pick $F_0^* \in \mathbf{F}_0$ such that $F_0^*(\cdot) \sim_0 \mathbb{P}(S < \cdot)$ were $S = \sum \lambda_i Z_i$, the Z_i's follow a $\chi^2(1)$ distribution and :*

$$\lambda_i = \frac{1}{\gamma(\varphi^{-1}(i))} = \rho\left(\varphi^{-1}(i)\right)$$

Remark 6 *Once again we encounter a regularly-varying condition on ρ. Here it echoes in a way the assumption $\mathbf{A}_0$ (necessary to derive (5)) which claims that the cdf of Z is itself regularly varying at 0. An interesting open question would consist in finding examples of auxiliary functions which are not regularly varying with positive index, whenever it is possible.*

For the sake of completeness we obtain a last result, complementing and illustrating Proposition 3. From this Proposition we see that $F \triangle G$ if $F = \phi G$ where ϕ is ρ-self-neglecting. The forthcoming Theorem represents these functions ϕ.

Theorem 5 *Let ρ be self-neglecting at 0 which does not vanish in a neighborhood of 0. A function ϕ is ρ-self-neglecting iff :*

$$\phi(x) = c(x) \exp\left(\int_x^1 \frac{\varepsilon(u)}{\rho(u)} du\right)$$

where $c(u) \to c \in]0, +\infty)$ and $\varepsilon(u) \to 0$ when $u \to 0$ and ε has the same regularity as ρ.

This theorem generalizes the representation Theorem 2.11.3 for self-neglecting functions p.121 in Bingham et al. (1987) initially due to Bloom (1976). If one take $\phi = \rho$ the representation above coincides with the one announced in this theorem.

3.1 Conclusion and perspectives

The first main results of this article identifies small ball probabilities in l_2 with a class of rapidly varying functions involved in extreme value theory and whose derivatives at all orders vanish at zero. This representation was obtained through previous works especially the seminal and precious formula (5) of Lifshits (1997). We hope that this new formulation will be more convenient for modelizing the small ball probabilities with some applied -especially statistical-purposes in mind. However many other questions arise. For instance the generalization to random elements with values in l_p or in more general Banach spaces is certainly an intricate matter since the starting fomulas (5) and followings seem to be intimately suited to the space l_2.

A more promising track could be to explore the links between the auxiliary function ρ, which inherits all the information on the regularity of X, with the metric entropy of the unit ball of the reproducing kernel Hilbert space of X as explored in Li, Linde (1999) or with the degree of compactness of the operator v in Li, Linde (2004) for instance, the latter operator v being obviously close to the covariance operator of X hence in connection with the a_i's (or λ_i's) of this article.

A surprising fact is the parallel that can be drawn between large deviations on a one hand and extreme value theory on the other hand. Both were intially introduced to model and explore large values of sequences of random elements. It turns out that both provide an accurate setting to study small deviations as well : Laplace transform for the classical approach and methods around the domain of attraction of the third type (Gamma class, self-negclecting functions...) as outlined here. However the connections between regular variations and small ball probabilities have been known since de Bruin in 1959, and his theorem on Laplace transfoms (see Theorem 4.12.9 in Bingham et al. (1987)). This work confirms that both Tauberian and extreme value theory may provide tools complementing large deviations techniques to derive new results in this area.

The other result shows, as an application of the previous, that the optimal risk in nonparametric regression for functional data is degenerate in the sense that we cannot expect obtain polynomial rates in the reasonable setting used in this work. It is obviously interpretable in terms of curse of dimensionality. A work is in progress to study the additive regression namely the model :

$$y = \sum_{i=1}^{k} r_i \left(\langle X, e_i \rangle \right) + \varepsilon$$

where the r_i are functions defined on $\mathbb{R}$ and estimated from one-dimensional projections of the data X. It is known since Stone(1985) that this model is not subject to the curse of dimensionality when X is valued in $\mathbb{R}^d$. It would be a possible track to introduce non-linearity in regression models for functional data and avoiding some redhibitory features of a general model. The role of the auxiliary function ρ is major. The question of its estimation is quite simple indeed. From Bingham et al. (1987) Corollary 3.10.5(b) p.177 we know that ρ may be taken as F/F'. A natural estimator of $\widehat{\rho}$ may be $\widehat{F}/\widehat{f}$ where $\widehat{f}$ (resp. $\widehat{F}$) is a kernel estimator of the density (resp. of the cumulative distribution function) of $\|X\|$. This is a simple procedure to check some of the needed properties of ρ such as its rate of decrease to 0.

4 Proofs

Considerations about the smoothness at 0 of F and ρ are not the matter in this work and we will take it for granted that both functions are smooth enough. Besides along the proofs we may sometimes consider generalized or local inverses of some fonctions which may not be invertible or have smooth derivatives everywhere. For example the auxiliary function ρ defined on $\mathbb{R}^+$ for which we always have $\rho'(0) = 0$ has no inverse on $[0, c]$ for $c > 0$. But we may frequently use the smoothness of, say, ρ and ρ^{-1} on sets $]a, b[$ for $0 < a < b$ without always justifying it. We start with the proof of Proposition 1.

Proof of Proposition 1 :

Suppose that $\rho(s)/s$ does not tend to zero when s does. Then we may pick an $\varepsilon > 0$ such that for infinitely many $s_k \downarrow 0$ when $k \uparrow +\infty$, $\rho(s_k)/s_k > \varepsilon$. Now fix $x < -\varepsilon^{-1}$ then $s_k + x\rho(s_k) < 0$ and $F(s_k + x\rho(s_k)) = 0$ for all k and $F(s_k + x\rho(s_k))/F(s_k)$ cannot converge to $\exp(x)$. The second part of the proof, namely ensuring the ρ is self-neglecting, follows the lines of the proof of Proposition 3.10.6 in Bingham et al. (1987). ■

Proof of Proposition 2 : Suppose that for some p $F^{(p)}(0) \neq 0$ and take $p^* = \inf\left\{p \in \mathbb{N} : F^{(p)}(0) \neq 0\right\}$. It is plain that $F^{(p^*)}(0) > 0$ since F is positive. Then we should consider two cases. First if $F^{(p^*)}(0) = c < +\infty$ then $F(s) \sim cs^{p^*}$. Taking :

$$\frac{F(s+\rho(s))}{F(s)} = \frac{F(s+\rho(s))}{(s+\rho(s))^{p^*}} \frac{s^{p^*}}{F(s)} \frac{(s+\rho(s))^{p^*}}{s^{p^*}}$$

we see that the left hand side of the display above tends to $\exp(1)$ whereas the right hand side tends to 1.

Second if $F^{(p^*)}(0) = +\infty$ we clearly have $F(s)/s^{p^*} \to +\infty$ when $s \to 0$. Take ε such that $1/\varepsilon > p^* + 2$. Since $\rho'(0) = 0$ and ρ is positive we may pick an s_0 such that $\sup_{0 \le u \le s_0} \rho'(u) \le \varepsilon$. From (9) we get :

$$\begin{aligned}\frac{F(s)}{s^p} &\le \frac{C}{s^p \rho^2(s)} \exp\left\{-\int_s^1 \frac{1}{\rho(t)} dt\right\} \le \frac{C}{\rho^{2+p}(s)} \exp\left\{-\int_s^1 \frac{1}{\rho(t)} dt\right\} \\ &\le \frac{C'}{\rho^{2+p}(s)} \exp\left\{-\int_s^{s_0} \frac{1}{\rho(t)} dt\right\}\end{aligned}$$

where we assume that $s \le s_0$. Then we have

$$\begin{aligned}\exp\left\{-\int_s^{s_0} \frac{1}{\rho(t)} dt\right\} &= \exp\left\{-\int_s^{s_0} \frac{\rho'(t)}{\rho(t)} \frac{1}{\rho'(t)} dt\right\} \le \exp\left\{-\frac{1}{\varepsilon}\int_s^{s_0} \frac{\rho'(t)}{\rho(t)} dt\right\} \\ &= \exp\left\{\frac{1}{\varepsilon}\ln\rho(s) - \frac{1}{\varepsilon}\ln\rho(s_0)\right\}.\end{aligned}$$

At last

$$\frac{F(s)}{s^p} \le C'' [\rho(s)]^{\frac{1}{\varepsilon} - p^* - 2}$$

which contradicts the fact that $F(s)/s^{p^*} \to +\infty$. $\blacksquare$

We start the proof of Theorem 2

Proof of Theorem 2 :

From Definition 1 and (5) we see that Theorem 2 holds whenever for all $x \in \mathbb{R}$:

$$\lim_{s\to 0} \frac{\gamma(s)\sigma(s)}{\gamma(s + x\rho(s))\sigma(s + x\rho(s))} \exp\left((s + x\rho(s))\gamma(s + x\rho(s)) - s\gamma(s)\right) \frac{\Lambda(\gamma(s + x\rho(s)))}{\Lambda(\gamma(s))} = \exp x.$$

We will more specifically prove below that when s decays to 0 :

$$\begin{aligned}\gamma(s + x\rho(s))/\gamma(s)\, \sigma(s + x\rho(s))/\sigma(s) &\to 1 \\ \exp\left((s + x\rho(s))\gamma(s + x\rho(s)) - s\gamma(s) - x\right) \frac{\Lambda(\gamma(s + x\rho(s)))}{\Lambda(\gamma(s))} &\to 1\end{aligned}$$

The two next lemmas are dedicated to showing that, in the above display the fraction as well as the exponential both tend to 1 when s goes to zero and ρ is chosen as in the Theorem.We just have to clarifiy formula (22) within the Theorem. This stem directly from (5). Indeed from (6) and (7) we see that $\sigma^2 = -\partial r/\partial\gamma$ and we just have to show that $\gamma r + \log\Lambda(\gamma) = \int_{r_0}^r \gamma(s)\, ds$. Elementary calculations yield :

$$\frac{\partial(\gamma r + \log\Lambda(\gamma))}{\partial r} = \gamma(r).$$

Let $r_0 = \mathbb{E}Z \cdot \sum_{j=1}^n \lambda_j$. Applying formula (6) at $\gamma = 0$ we notice that $\gamma(r_0) = 0 = \log\Lambda(\gamma(r_0))$ and we conclude.

Lemma 3 *Take $\rho(s) = 1/\gamma(s)$, then :*

$$\lim_{s\to 0} \exp\left((s + x\rho(s))\gamma(s + x\rho(s)) - s\gamma(s) - x\right) \frac{\Lambda(\gamma(s + x\rho(s)))}{\Lambda(\gamma(s))} = 1.$$

Remark 7 *Obviously γ has at least two (we do not need more) continuous derivatives on a neighborhood of infinity (here $]1,+\infty)$ for instance). It is also strightforward to see that γ, which is strictly decreasing on $]1,+\infty)$, is also a C^1 diffeomorphism on this set. Clearly $\lim_{s\to 0}\rho(s) = 0$ but from (**??**) it is plain that $\rho(s)/s$ also tends to zero when s does which implies that $\rho'(0) = 0$. Indeed proving that $\rho(s)/s$ tends to zero comes down to proving that $s\gamma(s) \to +\infty$.*

Proof of Lemma 3 :

Denote $I(s) = s\gamma(s) + \log\Lambda(\gamma(s))$. We should prove that :

$$\lim_{s\to 0} I(s + x\rho(s)) - I(s) - x = 0$$

Taylor's formula gives :

$$I(s + x\rho(s)) - I(s) = x\rho(s) I'(s) + \frac{x^2}{2}\rho^2(s) I''(c_{s,x}) \tag{34}$$

where $c_{s,x} = c$ lies somewhere in $[s, s + x\rho(s)]$ if $x \geq 0$ and in $[s + x\rho(s), x]$ if $x < 0$. From (7) we see that :

$$\begin{aligned} I'(s) &= \gamma(s) + s\gamma'(s) + \gamma'(s)\frac{\partial \log\Lambda(\gamma(s))}{\partial\gamma} \\ &= \gamma(s) \end{aligned}$$

Hence (34) may be rewritten :

$$\begin{aligned} I(s + x\rho(s)) - I(s) &= x + \frac{x^2}{2}\rho^2(s)\gamma'(c_{s,x}) \\ &= x + \frac{x^2}{2}\frac{\gamma'(c_{s,x})}{\gamma^2(s)} = x - \frac{x^2}{2}\frac{d(1/\gamma)}{ds}(c_{s,x}) \cdot \frac{\gamma^2(c_{s,x})}{\gamma^2(s)} \end{aligned}$$

We first show that $\gamma^2(c_{s,x})/\gamma^2(s) = \rho^2(s)/\rho^2(c_{s,x})$ is bounded above. We may always write $c_{s,x} = s + t_x(s)\rho(s)$ where $-x \leq t_x(s) \leq x$ for all s. Taylor's formula yields

$$\rho(s + t_x(s)\rho(s)) = \rho(s) + t_x(s)\rho(s)\rho'(d) = \rho(s)\left(1 + t_x(s)\rho'(d)\right)$$

where d lies between s and $s + t_x(s)\rho(s)$. Hence :

$$\frac{\rho(s)}{\rho(c_{s,x})} = \frac{1}{1 + t_x(s)\rho'(d)} \leq \frac{1}{1 - |x|\rho'(d)}$$

The continuity of ρ' at 0 and its nullity at 0 (see Remark 7) implies on a one hand that the display above is bounded above for fixed x and s (hence d) going to zero and also that :

$$\frac{d(1/\gamma)}{ds}(c_{s,x}) = \rho'(c_{s,x}) \to 0.$$

At last, $I(s + x\rho(s)) - I(s) \to x$ which finishes the proof of the Lemma. $\blacksquare$.

Lemma 4 *We have :*

$$\lim_{s\to 0}\frac{\gamma(s + x\rho(s))\,\sigma(s + x\rho(s))}{\gamma(s)\,\sigma(s)} = 1$$

Proof :
Once again Taylor's formula leads to :

$$\frac{\gamma\left(s+x\rho\left(s\right)\right)\sigma\left(s+x\rho\left(s\right)\right)-\gamma\left(s\right)\sigma\left(s\right)}{\gamma\left(s\right)\sigma\left(s\right)}=\frac{x\rho\left(s\right)}{\gamma\left(s\right)\sigma\left(s\right)}\left[\gamma'\left(c\right)\sigma\left(c\right)+\gamma\left(c\right)\sigma'\left(c\right)\right] \tag{35}$$

where $c\in\left(s,s\pm x\rho\left(s\right)\right)$. We will prove that $\left[\gamma'\left(c\right)\sigma\left(c\right)+\gamma\left(c\right)\sigma'\left(c\right)\right]/\gamma^{2}\left(s\right)\sigma\left(s\right)$ tends to zero. We cut the latter into two terms. First consider

$$\frac{\rho\left(s\right)}{\gamma\left(s\right)\sigma\left(s\right)}\gamma'\left(c\right)\sigma\left(c\right)=\frac{\gamma'\left(c\right)}{\gamma^{2}\left(c\right)}\frac{\gamma^{2}\left(c\right)}{\gamma^{2}\left(s\right)}\frac{\sigma\left(c\right)}{\sigma\left(s\right)}$$

We proved above within the proof of the previous Lemma (3) that $\gamma^{2}\left(c\right)/\gamma^{2}\left(s\right)$ is bounded above. We proved as well that $\gamma'\left(c\right)/\gamma^{2}\left(c\right)$ tends to zero when c does. Finally we should just control $\sigma\left(c\right)/\sigma\left(s\right)$. We have $\sigma\left(c\right)=\sigma\left(s\right)+\left(c-s\right)\sigma'\left(\xi\right)$ where $\xi\in\left[s,c\right]$ hence

$$0\leq\frac{\sigma\left(c\right)}{\sigma\left(s\right)}=1+\frac{c-s}{\sigma\left(s\right)}\sigma'\left(\xi\right)\leq1+\frac{x}{\gamma\left(s\right)\sigma\left(s\right)}\sigma'\left(\xi\right).$$

We see in Lifshits (1997, Lemma 2 p.431) that $\lim_{s\to0}\gamma\left(s\right)\sigma\left(s\right)=+\infty$ and that $\sigma\left(s\right)\leq sc_{13}^{-1}$ where c_{13} is some constant from which it is plain that $\sup_{\xi\in\mathcal{V}_{0}}\left|\sigma'\left(\xi\right)\right|<+\infty$ where $\mathcal{V}_{0}$ is any neighborhood of 0. We deduce that $\sigma'\left(\xi\right)/\gamma\left(s\right)\sigma\left(s\right)$ tends to zero which finally yields

$$\frac{\rho\left(s\right)}{\gamma\left(s\right)\sigma\left(s\right)}\gamma'\left(c\right)\sigma\left(c\right)\to0.$$

We turn to the second term in (35) : $\rho\left(s\right)\gamma\left(c\right)\sigma'\left(c\right)/\gamma\left(s\right)\sigma\left(s\right)$. We rewrite it :

$$\frac{\gamma\left(c\right)\sigma'\left(c\right)}{\gamma^{2}\left(s\right)\sigma\left(s\right)}=\frac{1}{\gamma\left(s\right)\sigma\left(s\right)}\frac{\gamma\left(c\right)}{\gamma\left(s\right)}\sigma'\left(c\right)$$

As shown above from Lisfhits' work : $\gamma\sigma\to+\infty$, $\sup_{c\in\mathcal{V}_{0}}\left|\sigma'\left(c\right)\right|<+\infty$ and $\gamma\left(c\right)/\gamma\left(s\right)$ is bounded above and this second term also decays to zero. This finishes the proof of Lemma 4.Now we turn to the proof of the converse part, Theorem 4. It takes two steps.

First we should make sure that when $\lambda_{i}=\rho\left(\varphi^{-1}\left(i\right)\right),\sum\lambda_{i}<+\infty$ which will ensure that the random element defined by $S=\sum\lambda_{i}Z_{i}$ is well-defined.

Lemma 5 *When* $\lambda_{i}=\rho\left(\varphi^{-1}\left(i\right)\right)$, $\sum\lambda_{i}<+\infty$.

Proof : It is easily seen that φ^{-1} is non decreasing in a neighborhood of $+\infty$. Indeed it suffices to prove that φ is, which may be deduced from its definition by studying its derivative. By the way one may also see that φ is concave. Now since φ^{-1} is non decreasing it is enough to prove that :

$$\int^{+\infty}\rho\left(\varphi^{-1}\left(x\right)\right)dx<+\infty$$

where the notation above means "the improper integral converges at infinity". Set $u=\varphi^{-1}\left(x\right)$ above then we should examine :

$$\int_{0}\rho\left(u\right)\varphi'\left(u\right)du.$$

Integrating by part this comes down to ensuring first that $\rho\left(u\right)\varphi\left(u\right)=u$ tends to a finite limit as u tends to 0 which is plain and that

$$\int_{0}\rho'\left(u\right)\varphi\left(u\right)du=\int_{0}u\frac{\rho'\left(u\right)}{\rho\left(u\right)}du<+\infty$$

Now we are in a position to apply Karamata's theorem to ρ' : since ρ is regularly varying at 0 with index $d \geq 1$ (since $\rho'(0) = 0$), and monotone in a right neighborhood of zero, ρ' is also regularly varying with index ≥ 0 (see Theorem 1.732.b p.39 in Bingham et al.(1987)). Then we can apply the direct part of Karamata's Theorem to ρ' (see ibid. Theorem 1.5.11 (i) p.28 where the limit should be taken here at zero) and

$$\lim_{t \to 0} \frac{t\rho'(t)}{\rho(t)} < +\infty$$

which ensures that the integral above converges and finally that $\sum \lambda_i < +\infty$. this completes the proof of Lemma 5. ■

Proof of Theorem 4 :

Pick an F_0 in Γ_0 with auxiliary function ρ_0 and consider the function $F_0^*(r) = \sqrt{\rho_0'(r)/\pi}\exp\left[-\int_r^{r_0} ds/\rho_0(s\right.$ with $r_0 = \sum_i \rho\left(\varphi^{-1}(i)\right)$. Note that $\sqrt{\rho'(\cdot)}$ hence $\rho'(\cdot)$ are ρ-self-neglecting because :

$$\frac{\rho'(r + x\rho(r))}{\rho'(r)} \to_{r \to 0} 1$$

Indeed $\rho'(r + x\rho(r)) = \rho'(r(1 + x\rho(r)/r))$, ρ' is regularly varying with positive index since ρ is itself regularly varying with index $\kappa > 1$, and $\rho(r)/r \to 0$ lead to

$$\lim_{r \to 0} \rho'(r(1 + x\rho(r)/r))/\rho'(r) = \lim(1 + x\rho(r)/r)^{\kappa - 1} = 1$$

This proves that $F_0^* \bigtriangleup F_0$. It remains to show that $F_0^* \sim_0 \mathbb{P}(S < \cdot)$. Like above $\gamma_0 = 1/\rho_0$. Start from (24) that is $r = \sum_j \lambda_j/(1 + 2\gamma_0\lambda_j)$. Now, following the proof of Proposition 4 we set $J(r) = r/\rho_0(r)$ (we just make use of display (25), fix $J(r)\rho_0(r)/r = 1$ instead of bounding it above and below) and take $a(\cdot) = J^{-1}(\cdot)$ then finally $S = \sum_{i=1}^{+\infty} Z_i/a(i)$. By construction $\mathbb{P}(S < \cdot) \sim F_0^*$.

Finally we turn to the proof of Theorem 5 and start with a Lemma. This Lemma, its proof and the subsequent proof of the theorem adapt the derivation of Lemma 2.11.2 and Theorem 2.11.3 of Bingham et al. (1987).

Lemma 6 *Let ρ be self-neglecting at 0. For $x_0 > 0$ sufficiently small the sequence $x_n = x_{n-1} - \rho(x_{n-1})$ tends to 0.*

Proof : First note that the sequence x_n is decreasing since $\rho \geq 0$ and notice from the properties of self-neglecting functions (namely $\rho(s)/s \to 0$ when $s \to 0$) that for a sufficently small $x_0 > 0$, $x_n \geq 0$ for all n. The limit of x_n exists, is denoted l. Suppose that $l > 0$. Then $\rho(l) > 0$ and since ρ is a non decreasing function $\rho(x_k) \geq \rho(l)$ for all k. At last

$$\begin{aligned} x_n = x_{n-1} - \rho(x_{n-1}) &= x_0 - \sum_{k=0}^{n-1} \rho(x_k) \\ &\leq x_0 - n\rho(l). \end{aligned}$$

Letting n go to infinity x_n goes to $-\infty$ which contradicts $x_n \geq 0$ hence the Lemma. ■

Proof of Theorem 5:

Let x_n be as in the preceding Lemma. Let p be a C^∞ probability density on $[0, 1]$ and set for $x_{n+1} \leq u \leq x_n$

$$\varepsilon(u) = \frac{\ln\phi(x_{n+1}) - \ln\phi(x_n)}{x_n - x_{n+1}} p\left(\frac{x_n - u}{x_n - x_{n+1}}\right)\rho(u).$$

The proof takes three steps.

We prove first that for all x_n, $\phi(x_n) = \exp\left(\int_{x_n}^1 \frac{\varepsilon(u)}{\rho(u)} du\right)$. In fact we may always define $\varepsilon(u)$, $x_0 \leq u \leq 1$ such that $\phi(x_0) = \exp\left(\int_{x_0}^1 \frac{\varepsilon(u)}{\rho(u)} du\right)$. Then assume that $\phi(x_k) = \exp\left(\int_{x_k}^1 \frac{\varepsilon(u)}{\rho(u)} du\right)$ for $k = 0, 1, .., n$. We have :

$$\begin{aligned}
\int_{x_{n+1}}^1 \frac{\varepsilon(u)}{\rho(u)} du &= \int_{x_{n+1}}^{x_n} \frac{\varepsilon(u)}{\rho(u)} du + \int_{x_n}^1 \frac{\varepsilon(u)}{\rho(u)} du \\
&= \ln \phi(x_n) + \frac{\ln \phi(x_{n+1}) - \ln \phi(x_n)}{x_n - x_{n+1}} \int_{x_{n+1}}^{x_n} p\left(\frac{x_n - u}{x_n - x_{n+1}}\right) du \\
&= \ln \phi(x_n) - (\ln \phi(x_{n+1}) - \ln \phi(x_n)) \int_1^0 p(t)\, dt \\
&= \ln \phi(x_{n+1})
\end{aligned}$$

Second we prove that for $x_{n+1} \leq x \leq x_n$ $\lim_{x\to 0} \phi(x)/\phi(x_n) = 1$. We note that $x = x_n - \lambda_x \rho(x_n)$ where $\lambda_x \in [0, 1]$ hence

$$\lim_{x\to 0} \frac{\phi(x_n - \lambda_x \rho(x_n))}{\phi(x_n)} = 1$$

uniformly with respect to $\lambda_x \in [0, 1]$.

The third and last step is devoted to proving that $|\varepsilon(u)| \to 0$ when $u \to 0$. Indeed for all $x_{n+1} \leq u \leq x_n$,

$$|\varepsilon(u)| \leq |p|_\infty \left|\frac{\ln \phi(x_{n+1}) - \ln \phi(x_n)}{x_n - x_{n+1}}\right| \rho(u)$$

We focus on

$$\begin{aligned}
\left|\frac{\ln \phi(x_{n+1}) - \ln \phi(x_n)}{x_n - x_{n+1}}\right| \rho(u) &= \frac{\rho(u)}{\rho(x_n)} \ln \frac{\phi(x_n)}{\phi(x_{n+1})} \\
&= \frac{\rho(x_n - \lambda_u \rho(x_n))}{\rho(x_n)} \ln \frac{\phi(x_n)}{\phi(x_n - \rho(x_n))}
\end{aligned}$$

Just like above $\rho(x_n - \lambda_u \rho(x_n))/\rho(x_n) \to 1$ since ρ is self-neglecting. Finally by the definition of ϕ we get

$$\ln \frac{\phi(x_n)}{\phi(x_n - \rho(x_n))} \to 0$$

which finishes the proof of the Theorem. ■

References

[1] de Acosta A. (1983) : Small deviations in the functional central limit theorem with applications to functional laws of the iterated logarithm, *Ann. Probab.*, **11**, 78-101.

[2] Aurzada F., Lifshits, M., Linde W. (2009) : Small deviations of stable processes and entropy of associated random opeartors, *Bernoulli*, **15**, 1305-1334.

[3] Berlinet A., Elamine A., Mas A. (2011) : Local linear regression for functional data, to appear in AISM.

[4] Bingham N.H., Goldie C.M. and Teugels J.L. (1987) : *Regular Variations.* Encyclopedia of Mathematics and Its Applications, Cambridge University Press.

[5] Bloom S. (1976) : A characterization of B-slowly varying functions, *Proc. Amer. Math. Soc.*, **54**, 243-250.

[6] Borell C. (1976) : Gaussian Radon measures on locally convex spaces, *Math. Scand.*, **38**, 265-285.

[7] Cardot H., Johannes J. (2010) : Thresholding projection estimators in functional linear models, *J. Multivariate Anal.*,.**101**, 395-408.

[8] Crambes C., Kneip A., Sarda P. (2009) : Smoothing splines estimators for functional linear regression, *Ann. Statist.*, **37**, 35-72.

[9] Dabo-Niang S. (2002) : Estimation de la densité dans un espac ed dimension infinie: Application aux diffusions, *C. R. Math. Acad. Sci. Paris.*, **334**, 213-216.

[10] Dabo-Niang S., Ferraty F., Vieu P. (2004) Nonparametric unsupervised classification of satellite wave altimeter forms. In COMPSTAT 2004 - *Proceedings in Computational statistics*, 879-886. Physica, Heidelberg.

[11] Dabo-Niang, Ferraty F., Vieu P. (2006) Mode estimation for functional random variable and its application for curves classification. *Far East J. Theor. Stat.*, **18**, 93-119.

[12] Delaigle A., Hall P. (2010) : Defining probability for a distribution of random functions, *Ann. Statist.*, **38**, 1171-1193.

[13] Dembo A., Meyer-Wolf E. and Zeitouni O. (1995) : Exact behavior of Gaussian semi-norms, *Statist. Probab. Lett.*, **23**, 275-280.

[14] Dunker T., Lifshits M.A., Linde W. (1998) : Small deviation probabilities of sums of independent random variables. In High dimensional probability (Oberwolfach, 1996), 59–74, Progr. Probab., 43, Birkhäuser, Basel.

[15] Erickson K. (1980) : Rates of escape of infinite-dimensional Brownian motion, *Ann. Probab.*, **8**, 325-338.

[16] Ferraty F., Goia A., Vieu P. (2007) On the use of modal curves for radar waveform classification. *Comp. Stat. Data Analysis*, **51**, 4878-4890.

[17] Ferraty F., Mas A., Vieu P. (2007). Advances on nonparametric regression from functional data, *Aust. New Zealand J. Stat.*, **49** n°3, 1-20.

[18] Ferraty F., Vieu P. (2004) Nonparametric models for functional data, with applications in regression, time series prediction and cruve discrimination. *J. Nonparametr. Stat.*, **16**, 111-125.

[19] Ferraty F., Vieu P. (2006) *Nonparametric Functional Data Analysis: Theory and Practice.* Springer, Berlin.

[20] Fisher R.A. and Tippett L.H.C. (1928) : Limiting forms of the frequency distribution of the largest or smaller member of a sample, *Proceedings of the Cambridge Philosophical Society,* **24**, 180-190.

[21] Gaiffas S. (2005) : Convergence rates for pointwise curve estimation with a degenerate design, *Math. Methods Statist.*, **14**, 1-27.

[22] Gnedenko B.V.(1943) : Sur la distribution limite du terme maximum d'une série aléatoire, *Ann. of Math.*, **44**, 423-453.

[23] de Haan L. (1971) : A form of regular variation and its application to the domain of attraction of the double exponential distribution, *Z. Wahrscheinlichkeitstheorie. verw. Geb.,* **17**, 241-258.

[24] de Haan L. (1974) : Equivalence classes of regularly varying functions, *Stochastic Process. Appl.*, **2**, 243-259.

[25] Hall P, Heckman N. (2002) : Estimating and depicting the structure of a distribution of random functions, *Biometrika*, **89**, 145-158.

[26] Hall P, Horowitz J.L. N. (2007) : Methodology and convergence rate for functional linear regression. *Ann. Statist.*, **35**, 70-91.

[27] Kuelbs J., Li W.V. and Talagrand M. (1994) : Lim inf results for Gaussian samples and Chung's functional LIL, *Ann. Probab.*, **22**, 1879-1903.

[28] Ledoux M. and Talagrand M. (1991) : *Probability in Banach Spaces. Isoperimetry and Processes,* Ergebnisse der Mathematik und ihrer Grenzgebiete, **23**, Springer-Verlag.

[29] Li W.V. and Linde W. (1993) : Small ball problems for non centered Gaussian measures, *Probab. Math. Statist.*, **14**, 231-251.

[30] Li W.V. and Linde W. (1999) : Approximation, metric entropy and small ball estimates for Gaussian measures. *Ann. Probab.*, **27**, 1556–1578.

[31] Li W.V. and Linde W. (2004) : Small deviations of stable processes via metric entropy, *J. Theoret. Probab.*, **17**, 261-284.

[32] Li, W.V. and Shao Q.-M. (2001) : Gaussian processes : Inequalities, small ball probabilities and applications, In Rao and Shanbhag (eds.), Handbook of Statistics, **19**, Elsevier, New-York, 533-597.

[33] Lifshits M.A.(1995) Gaussian random functions. Mathematics and its Applications, 322. Kluwer Academic Publishers, Dordrecht.

[34] Lifshits M.A. (1997) : On the lower tail probabilities of some random series, *Ann. Probab.*, **25**, 424-442.

[35] Mas A. (2008) : Local functional principal component analysis, *Complex Analysis and Operator Theory*, **2**, 135-167.

[36] Meyer-Wolf E. and Zeitouni O (1993) : The probability of small Gaussian ellipsoïds, *Ann. Probab.*, **21**, 14-24.

[37] Muller H.G. and Yan X. (2001) : On local moments, *J. Multivariate Anal.*, **76**, 90-109.

[38] Stone C.J. (1985) : Additive regression and other nonparametric models, *Ann. Statist.*, **13**, 689–705.

[39] Sytaya G.N. (1974) : On some asymptotic representations of the Gaussian measure in a Hilbert space, *Thoery of Stochastic Processes*, **2**, 93-104.

[40] Talagrand M. (1992) : On the rate of convergence in Strassen's LIL, *Progr. Probab.*, Birkhauser, Boston, 339-351.

[41] Tsybakov A. (2004) : Introduction à l'estimation non-paramétrique . Mathématiques et applications. Springer.

[42] van der Vaart, A.W., van Zanten, J.H. (2007) : Bayesian inference with rescaled Gaussian process priors *Electronic J. Statist*, **1**, 433-448.

[43] an der Vaart, A.W., van Zanten, J.H. (2008) : Rates of contraction of posterior distributions based on Gaussian process priors. *Ann. Statist.*, **36**, 1435-1463.