

# Lower bound in regression for functional data by representation of small ball probabilities

André Mas

*Université Montpellier 2*

*Institut de Mathématique et de Modélisation de Montpellier*

*e-mail: [mas@math.univ-montp2.fr](mailto:mas@math.univ-montp2.fr)*

**Abstract:** The minimax rate for estimating the regression function  $r(\cdot) = \mathbb{E}(y|X = \cdot)$  when  $y \in \mathbb{R}$  and  $X$  takes values in a function space is the initial motivation of this work. Recent articles underline the major role of the shifted small ball probability  $\mathbb{P}(\|X - x_0\| < \cdot)$  in the variance of classical estimates. The main results are twofold. First, starting from a theorem by [41], we study the small ball probability  $\mathbb{P}(S < \varepsilon)$  when  $\varepsilon \downarrow 0$  with  $S = \sum_{i=1}^{+\infty} \lambda_i Z_i$  where the  $Z_i$ 's are i.i.d. positive and  $(\lambda_i)_{i \in \mathbb{N}}$  a positive nonincreasing sequence such that  $\sum \lambda_i < +\infty$ . It is shown 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. Second this approach allows to derive minimax lower bounds for the risk at a fixed point  $x_0$  when  $X \in \mathcal{H}$  some Hilbert space of functions. Denoting this minimax risk:

$$\mathcal{R}_n^* = \inf_{T_n} \sup_{r \in \mathcal{E}} \mathbb{E} |T_n - r(x_0)|^2$$

where  $T_n$  is any estimate of  $r(x_0)$  and  $\mathcal{E}$  is some class of smooth functions from  $\mathcal{H}$  to  $\mathbb{R}$  it turns out that, in a general framework,  $n^\tau \mathcal{R}_n^* \rightarrow +\infty$  for any  $\tau > 0$ . This negative result may pave the way towards new approaches for modeling regression with functional data.

**AMS 2000 subject classifications:** Primary 62G08, 60K35; secondary 62G20.

**Keywords and phrases:** Functional data, small ball problems, nonparametric regression, regular variation, Gaussian random elements, lower bound.

Received April 2011.

## Contents

1	Preliminaries . . . . .	1746
1.1	The nonparametric regression model for functional data . . . . .	1746
1.2	About non-shifted and shifted small ball problems . . . . .	1749
1.3	The class $\Gamma_0$ . . . . .	1752
2	Main results . . . . .	1754
2.1	Small ball probabilities of random functions are Gamma-varying . . . . .	1755
2.2	Gaussian framework . . . . .	1757
2.3	Upper and lower bound in regression for functional data . . . . .	1758

2.3.1	Upper bound . . . . .	1759
2.3.2	Lower bound . . . . .	1760
3	Complementary facts . . . . .	1763
3.1	Conclusion and perspectives . . . . .	1764
4	Proofs . . . . .	1766
4.1	Proofs of results of sections 2.1, 2.2 and 3 . . . . .	1766
4.2	Proofs of results of section 2.3 . . . . .	1772
	Acknowledgements . . . . .	1775
	References . . . . .	1775

## 1. Preliminaries

The three following subsections are independent. The first introduces the non-parametric regression model for functional data and simply raises the problems attached to obtaining sharp bounds for the quadratic risk at a fixed point. The second gives some basic material about small ball probability. The third collects classical results from extreme value theory as well as the definition of the class  $\Gamma_0$  which is then briefly described. 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. Proofs are given in the last section.

### 1.1. The nonparametric regression model for functional data

Statistics for functional data is a recent domain which has been receiving increasing interests 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 ). FDA extends the classical statistical models designed for vectors to the situation when the data are functions or curves. We refer to the monographs by [47] and [22] for an overview of this topic. Consider the general regression problem with functional data as inputs:

$$y = r(X) + \varepsilon \tag{1.1}$$

where  $y$  and  $\varepsilon$  are real random variables with  $\mathbb{E}\varepsilon = 0$  with variance denoted  $\sigma_\varepsilon^2$ ,  $X$  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  $\mathbf{L}^2(T)$  where  $T$  is a compact set in the Euclidean space or some Sobolev space  $\mathbf{H}^{2,m}(T)$ . 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 from an i.i.d. sample  $(y_i, X_i)_{1 \leq i \leq n}$  and by a classical Nadaraya-Watson approach (see [54] for a general presentation in the finite dimensional setting and [22] for implementation on functional data). This model was studied for instance in [21] and asymptotic results were derived in [19] like a first upper bound for the quadratic risk at a fixed point and in

[18] for a uniform bound under entropy conditions. Apparently no projection-based estimate in this model has been introduced yet, certainly due to a lack of theoretical results on approximation theory for functions defined on a Hilbert space.

When  $X \in \mathbb{R}^d$  model (1.1) was obviously exhaustively studied. It is known since [49, 50] that the optimal rate of convergence in minimax sense depends on smoothness conditions of  $r$  namely the order of the highest achievable smooth derivative of  $r$  say  $m$  (and subject to the choice of an accurate kernel  $K$ ) as well as on the dimension  $d$ . This rate is more precisely of order  $n^{-m/(2m+d)}$  and depletes when the dimension  $d$  increases. This phenomenon is known in nonparametric statistics as ‘curse of dimensionality’. Letting  $d$  go to infinity in the rate above should leave us rather pessimistic about what we could expect from model (1.1) with functional input  $X$ . Indeed the issue of the minimax risk in this model -that is when  $X$  belongs to a function space- is open and will be addressed here. Slightly anticipating we can say that the concerns raised by the curse of dimensionality are grounded and that the rate for any efficient estimate is slower than any power of  $n$ .

We consider here an adapted Nadaraya-Watson estimate. This estimate is asymptotically efficient when  $X \in \mathbb{R}^d$ :

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

where  $x_0$  is a fixed point of the space,  $K$  is a kernel, that is a measurable, unilateral (defined on  $\mathbb{R}^+$ ) positive function with  $\int K = 1$  and  $h = (h_n)_{n \in \mathbb{N}}$  is a nonnegative sequence tending to 0 (the bandwidth). We consider the following additional conditions on the kernel  $K$ :

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

These conditions hold for the naive kernel,  $K(u) = 1$  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. The estimate  $\hat{r}(x_0)$  has already been considered in the papers mentioned above (see also [22, 24]).

Considering the  $L^2$ -risk at a fixed point  $x_0$  leads to a bias-variance decomposition:

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

with

$$\mathcal{B}_n(x_0) = \left\{ \mathbb{E} \frac{\sum_{i=1}^n [r(X_i) - r(x_0)] K(\|X_i - x_0\|/h)}{\sum_{i=1}^n K(\|X_i - x_0\|/h)} \right\}^2 \tag{1.2}$$

$$\mathcal{V}_n(x_0) = \mathbb{E} \left[ \frac{\sum_{i=1}^n \varepsilon_i K(\|X_i - x_0\|/h)}{\sum_{i=1}^n K(\|X_i - x_0\|/h)} \right]^2 \tag{1.3}$$

where  $\varepsilon_i = y_i - r(X_i)$ . It will be shown later in Proposition 5 that:

$$\mathcal{V}_n(x_0) \sim \frac{\sigma_\varepsilon^2 \mathbb{E}K^2(\|X - x_0\|/h)}{n [\mathbb{E}K(\|X - x_0\|/h)]^2} \quad (1.4)$$

$$\mathcal{B}_n(x_0) \sim \frac{1}{\mathbb{E}^2 K(\|X - x_0\|/h)} \left( \sum_{i=1}^{+\infty} \mathbf{b}_i \mathbb{E} \left[ \langle X, e_i \rangle^2 K(\|X\|/h) \right] \right)^2 \quad (1.5)$$

where the  $\mathbf{b}_i$ 's are positive and non random constants given in Proposition 5.

The sequence  $(\mathbf{b}_i)_{i \in \mathbb{N}}$  is not unveiled here because it depends on several parameters which will be introduced later, namely the derivatives of  $r$  and other functions depending on the distribution of  $X$ . We keep in mind that the bias-variance decomposition of the risk is essentially based on the computation of two sorts of moments:  $\mathbb{E}K(\|X - x_0\|/h)$  and  $\mathbb{E}[\langle X, e_i \rangle^2 K(\|X\|/h)]$ . Calculation of  $\mathbb{E}K^2(\|X - x_0\|/h)$  is similar with  $\mathbb{E}K(\|X - x_0\|/h)$ . Turning to  $\mathbb{E}[\langle X, e_i \rangle^2 K(\|X\|/h)]$  which appears in the numerator of (1.5) (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  $\mathbb{E}[\|X\|^2 K(\|X\|/h)]$  which will be treated like  $\mathbb{E}K(\|X - x_0\|/h)$  at (1.7) below. We ask the reader to accept for the moment that the computation of  $\mathcal{R}_n(x_0)$  critically depends on moments of the form  $\mathbb{E}[\|X\|^p K(\|X\|/h)]$  for some integer  $p$ .

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  $x_0$  simple calculus leads in many situations to:

$$\mathbb{E}K(\|X - x_0\|/h) \sim c_d f_X(x_0) h^d \quad (1.6)$$

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 or should be revisited. A major issue is then to compute the preceding expectation without assuming that  $f_X(x_0)$  exists. Applying Fubini's Theorem is sufficient to get rid of the density. Denoting  $\mathbb{P}(\|X - x_0\| < h) = F_{x_0}(h)$  we obtain:

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

and the evaluation of the expectation above essentially depends again on the small ball probability  $F_{x_0}(\cdot)$ . Assume that  $F_{x_0}$  is regularly varying at zero with index  $d$  (which is usually true when  $X$  is finite dimensional) then  $F_{x_0}(h) = C(x_0) h^{d_l}(h)$  where  $C(x_0)$  is a constant depending on  $x_0$ ,  $l$  is a slowly varying function at 0 and  $F_{x_0}(hs)/F_{x_0}(h) \rightarrow s^d$  when  $h \rightarrow 0$  for  $s > 0$  which yields  $\mathbb{E}K(\|X - x_0\|/h) \sim c_d h^{d_l}(h)$  where  $c_d$  depends only on  $d$  and  $K$ . Unfortunately when  $X$  lies in a function space, the most classical examples of  $F_{x_0}(h)$  are not

regularly varying as will be seen below. However we notice for further purpose that the theory of regular variation is of some help in this important special case.

**Remark 1.** It is straightforward to see that the same method may yield the value of such integrals as:

$$\mathbb{E} \left[ \|X - x_0\|^p K \left( \frac{\|X - x_0\|}{h} \right) \right] = h^p F_{x_0}(h) \left[ K(1) + \int_0^1 \tilde{K}_p(s) \frac{F_{x_0}(hs)}{F_{x_0}(h)} ds \right] \quad (1.7)$$

with  $\tilde{K}_p(s) = -[s^p K'(s) + ps^{p-1} K(s)]$

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 expectations in (1.4) and (1.5) -what statisticians are truly interested in. The behaviour of this small ball probability has been an issue hard to circumvent. Usually the authors express their final results in terms of  $F_{x_0}(\cdot)$  and cannot derive more explicit formulae in a general framework (see again [19] or [18]). Along the past decade some authors turned their attention to the question of modeling probability distribution for curve-data with applications in statistics: [11, 31, 12] in a general setting then. But a unified and general result for representing  $F_{x_0}(\cdot)$  efficiently is missing.

**Remark 2.** Some variants of the Nadaraya-Watson estimate above were studied when the norm of the function space  $\|\cdot\|$  is replaced by a semi-norm  $\nu(\cdot)$ . This usually allows to fix the problem of the small ball probability. Indeed taking projection semi-norms like  $\nu(x) = \|\mathcal{P}_d x\|$  where  $\mathcal{P}_d$  is the projection on some  $d$ -dimensional space it is easily seen that  $\mathbb{P}(\nu(X - x_0) < h) \sim c_d h^d$ . However the semi-norm estimate is usually not consistent. It suffices to take  $x_0$  in the null space of  $\nu$  to see that  $\nu(X - x_0) = \nu(X)$  and the estimate cannot converge. This route may be attractive for numerical and practical reasons but will not be considered in this formal setting. The case of increasing  $d = d_n$  has not been investigated, at least up to the author's knowledge.

A special case of (1.1) is the linear regression model  $y = \int X(s) \beta(s) + \varepsilon$ . It has been extensively investigated in the last years. We refer for instance to [31, 10, 9] (see also references therein these works). The optimal rate for the risk is now fully determined in this linear model conversely to the non-linear one investigated here. A comparison is given later at Remark 8.

## 1.2. About non-shifted and shifted small ball problems

Small ball problems could generally be stated in the following way: consider a random variable  $X$  with values in a general normed space  $(E, \|\cdot\|)$  (which may be infinite-dimensional) and estimate  $\mathbb{P}(\|X\| < \varepsilon)$  for small values of  $\varepsilon$ . 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 [39] is a complete state of the art, introducing the main concepts and providing numerous references. Another reference is Chapter 18 of [40] 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 Brownian motion, Brownian sheet, etc). The case of stable random elements was also investigated (see for instance [38, 3]). 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 -certainly more attractive in statistics- could be the shifted small ball probability  $\mathbb{P}(\|X - x_0\| < \varepsilon)$  for a fixed  $x_0$ . A concern 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 [7] or [36] 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  $\varepsilon$ . When  $f_{x_0}$  is regular enough in a neighborhood of zero:

$$F_{x_0}(h) = \int_{B(0,\varepsilon)} \mathbb{P}_{X-x_0}(dx) = \int_{B(0,\varepsilon)} f_{x_0}(x) \mathbb{P}_X(dx) \sim f_{x_0}(0) F(h) \quad (1.8)$$

About this fact see Proposition 2.1 in [1]. 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 explicitly described. Quite often lower and upper bounds are computed so that:

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

where  $\varphi_{x_0}$  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 [37] or to the degree of compactness of linear operators generating  $X$  (see [38]). 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 [53, 34] or the rate of escape of the Brownian motion (see [15]). They even surprisingly provide a sufficient condition for the CLT (see [35], Theorem 10.13 p.289). However small ball problems remained until nowadays a matter essentially reserved to probability theory. For example [55, 56] found applications of small ball techniques to Bayesian statistics. Recently [27] introduced them in conditional extremal quantile estimation. It turns out that this topic may be also of interest in functional data analysis. As shown in the previous section, since Lebesgue's density of an infinite-dimensional random  $X$  does not exist, all the inferential techniques based on the density cannot hold anymore. In this framework, the small ball probabilities appear as natural counterparts and should be investigated with much care.

First we introduce the  $l_2$  framework. Consider  $X$  a random variable defined the following way:

$$X = \left( \sqrt{\lambda_1}x_1, \sqrt{\lambda_2}x_2, \dots \right) \quad (1.9)$$

where  $(\lambda_i)_{1 \leq i \leq n}$  is a real positive sequence arranged in a non-increasing order such that  $\sum_{i=1}^{+\infty} \lambda_i < +\infty$  and  $(x_i)_{1 \leq i \leq n}$  is a 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  $\lambda_i$ 's. About this fact we refer to [14].

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 referred to as Principal Orthogonal Decomposition in other areas 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  $\lambda_i$  the associated eigenvalues

$$X = \sum_{i=1}^{+\infty} \sqrt{\lambda_i} x_i e_i \quad (1.10)$$

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 (1.9) 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 [52]. However we borrow the notations from [41] who extended Sytaya's results to the non-Gaussian framework. First in order to alleviate notations set once and for all:

$$S = \sum_{i=1}^{+\infty} \lambda_i Z_i \quad (1.11)$$

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:

$$\mathbb{P}(S < r) \underset{r \rightarrow 0}{\sim} \frac{1}{\sqrt{2\pi}} \frac{1}{\gamma\sigma} \exp(\gamma r) \Lambda(\gamma) \quad (1.12)$$

where  $\gamma$  and  $\sigma$  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  $\gamma$  and  $\sigma$  are implicit. Let  $S_\gamma$  be the Esscher transform of  $S$  that is the random variable with distribution  $\exp(-\gamma x) \mathbb{P}_S(dx) / \Lambda(\gamma)$ . Then set:

$$r = \mathbb{E}[S_\gamma] = -\frac{\partial \log \Lambda(\gamma)}{\partial \gamma}, \quad (1.13)$$

$$\sigma^2 = \mathbb{V}[S_\gamma] = \frac{\partial^2 \log \Lambda(\gamma)}{\partial \gamma^2}. \quad (1.14)$$

where  $\mathbb{V}$  denotes variance. We note for further purpose that the implicit function theorem ensures the existence and smoothness of functions  $\gamma(r)$  and  $\gamma(\sigma^2)$  derived from (1.13) and (1.14). Without further assumption on the  $\lambda_i$ 's  $\mathbb{P}(S < r)$  cannot be made more explicit. This is done for instance in [14] where these authors considered the case of  $\lambda_i$  with polynomial and exponential decay.

### 1.3. The class $\Gamma_0$

As a last part of this long introduction we shift from functional data analysis and small ball problems to extreme value theory. 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 [25] and [28]). This classical result assesses that whenever  $U_1, \dots, U_n$  is an i.i.d. sample of real random variables,  $M_n = \max\{U_1, \dots, U_n\}$  belongs to the domain of attraction of  $G$ , where  $G$  has same type as one of the three distributions Gumbel, Fréchet and Weibull. The Gumbel law, also

named double exponential distribution, with cumulative distribution function  $R(x) = \exp(-\exp(-x))$  defines the domain of attraction of the third type. Laurens de Haan in [29] characterized the (cumulative) distribution functions of  $U$  such that  $M_n$  belongs to the domain of attraction of  $R$ . We give this result below.

**Theorem ([29]).** *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  $\rho : \mathbb{R} \rightarrow \mathbb{R}^+$ , 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 initially introduced by de Haan as a Form of Regular Variation (see the title of his article). This class of distribution functions is referred to as de Haan's Gamma class in the book by [5] and within this article. Surprisingly, in their book as well as in de Haan's article no examples of functions belonging to  $\Gamma$  is given. The cumulative distribution 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 and not at infinity of the cumulative distribution function function of a real valued random variable we have to modify again slightly the definitions above. We introduce the class  $\Gamma_0$  and feature some of its properties below. We share most of our notations with [5] which differ from those of de Haan.

**Definition 1.** The class  $\Gamma_0$  consists of those functions  $F : \mathbb{R} \rightarrow \mathbb{R}^+$  null over  $(-\infty, 0]$ , non decreasing with  $F(0) = 0$  and right-continuous for which there exists a continuous non decreasing function  $\rho : \mathcal{V}^+ \rightarrow \mathbb{R}^+$ , defined on some a right-neighborhood of zero  $\mathcal{V}^+$  such that  $\rho(0) = 0$  and for all  $x \in \mathbb{R}$ ,

$$\lim_{s \downarrow 0^+} \frac{F(s + x\rho(s))}{F(s)} = \exp(x). \quad (1.15)$$

The function  $\rho$  is called the auxiliary function of  $F$ .

The properties of the auxiliary function are crucial.

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

$$\frac{\rho(s + x\rho(s))}{\rho(s)} \xrightarrow{s \rightarrow 0} 1$$

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

**Remark 3.** When the property in the Proposition above does not hold locally uniformly but only pointwise the function is called Beurling slowly varying. Assuming that  $\rho$  is continuous in Definition 1 yields local uniformity and enables to consider a self-neglecting  $\rho$ .

The class  $\Gamma_0$  is subject to an exponential-integral representation. In fact the following Theorem asserts that the local behaviour at 0 of any  $F$  in  $\Gamma_0$  depends only on the auxiliary mapping  $\rho$ .

**Theorem 1.** *Let  $F$  belong to  $\Gamma_0$  with self-neglecting auxiliary function  $\rho$  then when  $s \rightarrow 0$ :*

$$F(s) = \exp \left\{ \eta(s) - \int_s^1 \frac{1}{\rho(t)} dt \right\} \quad (1.16)$$

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

$$F(\lambda s) / F(s) \rightarrow \begin{cases} \infty & (\lambda > 1) \\ 1 & (\lambda = 1) \\ 0 & (\lambda < 1) \end{cases} \quad \text{as } s \rightarrow 0. \quad (1.17)$$

**Remark 4.** The upper bound 1 in the integral in (1.16) is unimportant and may be replaced by any positive number. Then the function  $\eta$  will change as well.

The proofs of Proposition 1 and of Theorem 1 are closely inspired from the proofs of Lemma 3.10.1, Proposition 3.10.3 and Theorem 3.10.8 in [5] and will be omitted.

Let us also mention that Gaïffas in [26] proposed to model locally the density of sparse data by Gamma-varying functions. This is another statistical application for  $\Gamma_0$ .

It is simple to construct explicit examples of functions in  $\Gamma_0$  by tuning the auxiliary function  $\rho$  and taking  $\eta(\cdot) = 0$  in (1.16). For instance taking  $\rho_1(t) = t^m$  (with  $m > 1$ ) gives  $\mathcal{F}_1(s) = \exp(-1/s^{m-1})$ . Now taking  $\rho_2(t) = -t/\log(t)$  yields  $\mathcal{F}_2(s) = \exp(-[\log(s)]^2)$ . Obviously constants may be added in front of or within the exponential. The next Proposition seems to show a specific feature of functions in the class  $\Gamma_0$  which will be used later: they are very flat at zero.

**Proposition 2.** *Let  $F$  belong to  $\Gamma_0$ . Then for all integer  $p$   $F^{(p)}(0) = 0$  where  $F^{(p)}$  denotes the derivative of order  $p$  of  $F$ .*

## 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 (1.1) belongs to the class  $\Gamma_0$  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  $\Gamma_0$  to derive upper and lower bounds on the risk for (1.1) and at a fixed point. The degeneracy of the lower bound (with respect to the case when  $X$  is finite dimensional) announced earlier 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  $\lim_{u \rightarrow x} f(u)/g(u) \leq c$  for some positive constant  $c$ .

### 2.1. Small ball probabilities 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  $\lambda_j$ 's in a smooth way (which means that  $\lambda(j) = \lambda_j$  for all  $j$  and  $\lambda$  is  $\mathbf{C}^1$ ).

Since our results rely on those of [41] 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  $r < c_3$ :

$$\mathbf{A}_0 : \begin{cases} G(r) \leq c_1 G(br) \\ G(br) \leq c_2 G(r) \\ \mathbb{E}Z^3 < +\infty \end{cases} \quad (2.1)$$

As mentioned in [41] 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 [14]. In particular they show that when  $\lambda_i = i^{-\beta}$  ( $\beta > 1$ ),  $\mathbb{P}(\|X\|^2 < s) \sim F_1(s)$  and that when  $\lambda_i = \exp(-i)$ ,  $\mathbb{P}(\|X\|^2 < s) \sim F_2(s)$  with:

$$F_1(s) = c_1 s^{[1+\beta(2+c_2)]/(2\beta-2)} \exp\left(-c_3 s^{-1/(\beta-1)}\right) \quad (2.2)$$

$$F_2(s) \sim c_4 \left[s^{1/3} \log(1/s)\right]^{-3/4} \exp\left(-[\log(s \log 1/s)]^2/4 + \psi_0(\log(s \log 1/s))\right) \quad (2.3)$$

where  $\psi_0$  is a bounded function. Formula (2.2) is proved as well at page 269 in [40]. Simple algebra proves that both functions on the right hand side of (2.2) and (2.3) have all their derivatives vanishing at 0. We notice that the r.h.s. of (2.2) is always flatter than the r.h.s. of (2.3) 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  $\lambda_i$ '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.

**Proposition 3.** Both functions  $F_1$  and  $F_2$  defined above at (2.2) and (2.3) belong to  $\Gamma_0$  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  $\rho_1$  and  $\rho_2$  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 implicitly at line (1.13). In words it is, up to sign, the inverse of the first order derivative of the log-Laplace transform of  $S$ .

**Theorem 2.** Let  $S = \sum_{i=1}^{+\infty} \lambda_i Z_i$  and set  $\mathbb{P}(S < r) = F(r)$  the small ball probability of  $S$  then  $F \in \Gamma_0$  with auxiliary function:

$$\rho(r) = \frac{1}{\gamma(r)} \tag{2.4}$$

and the representation (1.12) may be rephrased only in terms of  $\gamma(\cdot)$ :

$$\mathbb{P}(S < r) \underset{r \rightarrow 0}{\sim} \frac{1}{\sqrt{2\pi}} \frac{\sqrt{-\gamma'(r)}}{\gamma(r)} \exp \left[ - \int_r^{r_0} \gamma(s) ds \right] \tag{2.5}$$

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

Obviously the r.h.s. of (2.5) is mathematically the same object as the r.h.s. of (1.12). The Gamma-varying version of the right hand side is  $\sqrt{\rho'(r)} \exp[- \int_r^{r_0} ds / \rho(s)]$ . We believe however that this new version is slightly more explicit and maybe more suited for statistical purposes. We will take advantage of the properties of the class  $\Gamma_0$  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  $\mu$  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  $\mu$  must allocate a constant mass of 1 to a space whose dimension increases. Then  $\mu$  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  $\rho$ .

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

$$\mathbf{RV}_+ : \alpha > 1 \text{ or } \mathbf{RV}_1 : \rho(s) = sl(s) \text{ with } l(s) \succeq \log(1/s). \tag{2.6}$$

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

This property of the small ball probability has to be connected with property (1.17). It 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 (1.7). 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  $\rho_1$  and  $\rho_2$  appearing at Proposition 3 and arising from [14] we get  $\rho_1 \in \mathbf{RV}_+$  and  $\rho_2 \in \mathbf{RV}_1$ .

**Remark 5.** For the sake of completeness we point out the following fact which may be misleading: indeed we started from  $\mathbb{P}(\|X\|^2 < r)$  and the properties of this function may differ from the true small ball probability  $\mathbb{P}(\|X\|^2 < r^2)$ . It is simple to show that if  $F \in \Gamma_0$  with auxiliary function  $\rho_F$  then  $G(r) = F(r^2)$  belongs to  $\Gamma_0$  as well with auxiliary function  $\rho_G$  defined by  $\rho_G(r) = \rho_F(r^2)/(2r)$ .

### 2.2. Gaussian framework

Assuming that  $X$  is Gaussian, hence that  $x_i$  in (1.10) 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:

$$r = -\frac{\partial \log \Lambda(\gamma)}{\partial \gamma} = \sum_j \frac{\lambda_j}{1 + 2\gamma\lambda_j} \tag{2.7}$$

which is the initial equation linking  $r$  and  $\gamma$ . We derive below an explicit link between the  $\lambda_j$ 's and  $\gamma(\cdot)$  or equivalently  $\rho(\cdot)$ . Under rather general assumptions on the rate of decrease of the  $\lambda_j$ '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  $l \in [\frac{2}{3}, \frac{3}{2}]$  such that for any  $\varepsilon > 0$  and large enough  $x$*

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

Besides when  $\lambda(x) \asymp_\infty \exp(-x^\alpha)$  for some  $\alpha > 0$ ,  $F(s) \prec_0 \exp[-(\log 1/r)^{1+1/\alpha}]$ . When  $\lambda(\cdot)$  is explicitly 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  $l = \int_0^{+\infty} du / (2 + u^{1+\nu})$ . When  $\lambda_g(x) = c \exp(-\nu x)$ , with again  $c, \nu > 0$ ,  $\gamma_g(s) \sim_0 \log(1/s) / (\nu s)$ .

**Remark 6.** The auxiliary functions  $\rho_a = 1/\gamma_a$  and  $\rho_g = 1/\gamma_g$  match respectively  $\rho_1$  and  $\rho_2$  in Proposition 3. Besides letting  $\nu$  go to infinity we see that, in a way  $\rho_g$  may be viewed as a limit of  $\rho_a$ . In fact  $1/\log(1/s)$  echoes the degeneracy of  $1/(\beta - 1)$ . The second part of Proposition 4 modestly rediscovers the results of [14]. The upper bound of  $F(\cdot)$  is close to the one obtained in Theorem 2. No assumptions are needed on  $\rho$  here but the distribution of  $X$  is Gaussian.

### 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  $g''(x_0)$  (Hessian operator) 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 (1.10) 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  $f''(x)$  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 (1.10). Besides we assume that for all  $i$  the density of the margins  $\langle X, e_i \rangle$  is symmetric.

**Discussion and examples.** Let for instance  $X$  be Gaussian. Chapter 9 and 10 in [40] 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  $\sum_{i \in \mathbb{N}} m_i^2 / \lambda_i < +\infty$ . Then for any  $u = (u_i)_{1 \leq i}$  in  $\mathcal{H}$   $f_{x_0}(u) = \exp(-\sum_{i \geq 1} m_i^2 / 2\lambda_i + \sum_{i \geq 1} u_i m_i / 2\lambda_i)$  and  $\partial_i f_{x_0}(u) = (m_i / 2\lambda_i) f_{x_0}(u)$ . We are interested in the smoothness of these functions at 0. From  $|f_{x_0}(u) - f_{x_0}(0)| \leq \|u\| (\sum_{i \geq 1} m_i^2 / \lambda_i^2)^{1/2}$  the major point is the finiteness of the latter series. This is subject to a condition of decay on the coefficients  $m_i$ 's with respect to the  $\lambda_i$ 's hence of smoothness of the fixed function  $x_0$ .

It turns out that the Gaussian framework may be generalized to some other stable distributions though somewhat restricted. Skorohod ([48]) gives conditions for non Gaussian distributions to check the condition  $P_{X-x_0} \ll P_X$ . We refer to the characterization of stable distributions on Hilbert spaces by Jurek ([33]) for further reading. When these formal conditions hold and  $X = (\lambda_1 Z_1, \lambda_2 Z_2 \dots)$  with  $Z_i$  centered with unit variance and independent, conditions on  $m_i$  and  $\lambda_i$  and similar to the Gaussian case may be formulated from simple calculations, all linking the smoothness of  $f(x_0)$  with the relative decay rate of the Fourier coefficients of  $x_0$  with respect to the eigenvalues.

**Assumptions on the regression function.** Assume that  $r''(x)$  exists for  $x$  in a neighborhood of  $x_0$ . We denote  $\partial_i r_{x_0} = \langle r'(x_0), e_i \rangle$  and  $\partial_{ii}^2 r_{x_0} = \langle r''(x_0)(e_i), e_i \rangle$  and assume as well that:

$$\sum_{i=1}^{+\infty} \lambda_i \left( \frac{\partial_{ii}^2 r_{x_0} f_{x_0}(0)}{2} + \partial_i r_{x_0} \partial_i f_{x_0} \right) = \eta \neq 0 \quad (2.8)$$

**Discussion and examples:** If we except the technical assumption given at the line above, the conditions on  $r$  given here are classical. They hold for instance when  $r(x) = r_0(\omega(x))$  where  $r_0 : \mathbb{R} \rightarrow \mathbb{R}$  has second order derivative around  $\omega(x_0)$  and  $\omega : H \rightarrow \mathbb{R}$  is a simple functional such as  $\omega(x) = \sum_{i=1}^D \langle x, g_i \rangle$  where  $D \in \mathbb{N}$  and  $g_i \in H$  or even  $\omega(x) = \|x\|$  for  $x_0 \neq 0$ . But the functional

setting forces us to consider other kinds of possible regression functions. We briefly comment two of them that may produce irregularity: evaluation and derivation-based functionals. In [17] the regression function  $r$  involves sums and products of evaluation functionals. For instance  $r_{FHV}(x) = x(t_1)x(t_2)$  where  $t_1$  and  $t_2$  are two fixed points in  $[0, 1]$ . In [20]  $r_{FKV}(x) = \int t \cos(t) [X'(t)]^2 dt$ . We recall that the linear mapping  $d_{t_0}(x) = x(t_0)$  is not continuous when  $H = L^2([0, 1])$ . If we require that  $H$  is a reproducing kernel Hilbert space -typically a Sobolev space with boundary conditions such as Cameron–Martin space  $\mathbf{H}_0^2 = \{f \in L^2([0, 1]), \int f'^2 < +\infty, f(0) = 0\}$ - then  $d_{t_0}(\cdot) = \langle k_{t_0}, \cdot \rangle$  where  $k_{t_0} \in H$  is bounded hence continuous (by Riesz representation theorem). Then we derive  $d'_{t_0}(x) = k_{t_0}$  for all  $x$  and  $d''_{t_0} = 0$  and the condition (2.8) on  $r_{FHV}$  may be expressed in terms of coefficients of the kernel  $k_{t_1}$  and  $k_{t_2}$  in the basis  $e_i$ . For the second example, considering smooth spaces of functions will here again ensure the smoothness of the operator. Some calculations show that  $\nabla r_{FKV}(x)(t) = 2 \int_0^t s \cos(s) x'(s) dt$ . Here too the assumption will hold if the coordinates of  $\nabla r_{FKV}(x_0) = r'(x_0)$  and of the diagonal of  $r''(x_0)$  in the basis  $e_i$  tend rapidly to zero.

2.3.1. Upper bound

In view of the results of the preceding section we are in a position to simplify some computations. Turning to the local moments defined at (1.7), from properties of functions in  $\Gamma_0$  and specifically (1.17) we get:

$$\mathbb{E}[\|X - x_0\|^p] K\left(\frac{\|X - x_0\|}{h}\right) \sim K(1) h^p F_{x_0}(h) \sim K(1) h^p f_{x_0}(0) F(h) \quad (2.9)$$

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 formula above leads as well to:

$$\mathbb{E} K^2\left(\frac{\|X - x_0\|}{h}\right) \sim K^2(1) F_{x_0}(h) \quad (2.10)$$

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

$$\mathcal{M}_{K,1}(x_0) = \mathbb{E}\left[(X - x_0) K\left(\frac{\|X - x_0\|}{h}\right)\right] \quad (2.11)$$

$$\mathcal{M}_{K,2}(x_0) = \mathbb{E}\left[(X - x_0) \otimes (X - x_0) K\left(\frac{\|X - x_0\|}{h}\right)\right]. \quad (2.12)$$

Note that  $\mathcal{M}_{K,1}$  belongs to  $\mathcal{H}$  since  $(X - x_0) K(\|X - x_0\|/h)$  does. Formula (2.12) may be explicitated. First let  $u$  and  $v$  be two points in the vector space then  $u \otimes v$  is a linear operator defined by  $[u \otimes v](x) = \langle v, x \rangle y$  and so is  $(X - x_0) \otimes (X - x_0) K(\|X - x_0\|/h)$ . Erasing  $x_0$  and  $K(\|X - x_0\|/h)$  gives the usual covariance operator of  $X$  for a centered  $X$  which is trace-class by definition whenever  $\mathbb{E}\|X\|^2 < +\infty$ . The special covariance operator  $\mathcal{M}_{K,2}(x_0)$  is

obtained by shifting and smoothing  $X$  around  $x_0$  and  $\mathcal{M}_{K,2}$  is also a linear trace-class operator acting from and onto  $\mathcal{H}$  whenever  $\mathbb{E}[\|X - x_0\|^2 K(\|X - x_0\|/h)]$  is finite. We refer to [44] for some statistical results on local moments for finite-dimensional random variables and to [42] for some related results dealing with (2.12) and where random functions and small ball problems appear.

The next Proposition bounds above and below the bias part of the risk of our kernel estimate.

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

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

with  $\mathbf{b}_i = \frac{\partial_{ii}^2 r_{x_0} f_{x_0}(0)}{2} + \partial_i r_{x_0} \partial_i f_{x_0}$  as announced in (1.5). From the line above we derive 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  $\eta^2$ . Hence 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  $\Gamma_0$ ) motivates to prove the reader that  $\mathcal{B}_n(x_0)$  does not reach an unusual rate of decrease to 0 (namely exponential). And the bound  $\mathbf{c}^- \rho^6(h) \leq \mathcal{B}_n(x_0)$  will justify the conditions under which the minimax lower bound for the risk is going to be derived.

**Remark 7.** The role of the auxiliary function  $\rho$  is crucial. The question of its estimation is quite simple indeed. From [5] Corollary 3.10.5(b) p.177 we know that  $\rho$  may be taken as  $F/F'$ . A natural estimator of  $\hat{\rho}$  may be  $\hat{F}/\hat{f}$  where  $\hat{f}$  (resp.  $\hat{F}$ ) is a kernel estimator of the density (resp. of the cumulative distribution function) of  $\|X\|$ . A simple procedure for the practitioner consists here in considering the sample made of the positive real valued random variables  $(z_i = \|X_i\| \mathbb{1}_{\|X_i\| < s})_{1 \leq i \leq n}$  where  $s$  is some threshold close to zero and then to perform the estimation of  $\hat{f}$  and  $\hat{F}$  classically on this sample.

### 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  $\mathcal{B}_n(x_0) \asymp F^\kappa(h)$  for some  $\kappa > 0$ . Then the optimal bandwidth would stem from  $n^{-1} \asymp F^{1+\kappa}(h)$  leading to a  $\mathcal{R}_n(x_0) \preceq n^{-\kappa/(1+\kappa)}$  which would contradict the initial claim of

degenerate rate for the risk. This explains why we derived the lower bound on  $\mathcal{B}_n(x_0)$  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 polynomial rate. What we mean by classical approximation features is explicated now.

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

$$\sup_{r \in \mathcal{E}_p} \mathcal{B}_n(x_0) \preceq h^{2p} \tag{2.13}$$

For instance  $\mathcal{E}_p$  may be the class of Hölder functions of order  $p \in ]0, 1[$ . When  $\mathcal{E}_p$  is the class of functions which have two derivatives at  $x_0$  we see from Proposition 5 that we have in addition

$$h^{6m} \preceq \sup_{r \in \mathcal{E}_p} \mathcal{B}_n(x_0) \preceq h^4$$

when  $\rho(h) \succeq h^m, m > 1$ . Optimizing the bias-variance trade-off in the risk leads to choosing an  $h$  such that  $\sup_{r \in \mathcal{E}_p} \mathcal{B}_n(x_0) = \mathcal{V}_n(x_0)$ . The next Lemma deals with this issue.

**Lemma 1.** *Assume that  $X$  is gaussian and  $\lambda(x) \succ_\infty \exp(-x^\alpha)$  for some  $\alpha > 0$ . Let  $c^*$  be some constant and  $h^*$  be the solution of the functional equation:*

$$\frac{1}{n} = c^* h^{2p} F(h) \tag{2.14}$$

*then  $n^\beta / (nF(h^*)) \rightarrow +\infty$  for any  $\beta > 0$ . When  $X$  is non Gaussian but satisfies the assumptions (2.1) and (2.6) 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  $F(h) \prec_0 \exp[-(\log 1/h)^{1+1/\alpha}]$  holds. When  $X$  is not gaussian and  $\mathbf{RV}_1$  holds the proof of Corollary 1 shows that  $\beta \log n + 2p \log h^* > \beta \zeta(h^*) \log(1/h^*) - 2p \log h^*$  where  $\zeta(h^*)$  tends to  $+\infty$  when  $h^*$  tends to 0. When  $\mathbf{RV}_+$  holds the proof is the same with  $c_\alpha (h^*)^{1-\alpha}$  instead of  $\zeta(h^*) \log(1/h^*)$ .  $\square$

Now our approach to derive lower bounds for the minimax risk follows Tsybakov’s scheme (see [54]): 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  $p_\varepsilon$  stand for the density of  $\varepsilon$ . Assume that for some constant  $p_*$  and for all  $y \in \mathbb{R}$ ,

$$\int_{\mathbb{R}} \left[ \sqrt{p_\varepsilon(t)} - \sqrt{p_\varepsilon(t+y)} \right]^2 dt \leq p_* y^2. \tag{2.15}$$

This assumption is general and appears in Tsybakov’s book. It holds under smoothness assumptions on  $p_\varepsilon$ . We comment it briefly. If  $\Lambda(y)$  denotes the left hand side in the inequality above  $\Lambda(y) \leq 2$  for all  $y$  and we just need to study  $\Lambda$  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  $p_\varepsilon$  is smooth enough hence  $\Lambda'(0) = 0$ . Under accurate conditions on  $\Lambda''$ ,  $\Lambda(y) \leq p_* y^2$  will hold around 0 hence everywhere.

**Theorem 3.** *We consider two situations depending on the design  $X$ :*

**Part I:** *Assume that  $X$  is Gaussian,  $\lambda(\cdot)$  is a convex decreasing function with  $\lambda(x) \asymp_\infty \exp(-x^\alpha)$  for some  $\alpha > 0$  and that (2.15) holds. Denote  $T_n$  any estimator of the regression function at a fixed point  $r(x_0) = E(y|X = x_0)$  and  $\mathcal{R}_n(x_0)$  ( $\equiv \mathcal{R}_n$  for short) the minimax risk over the class  $\mathcal{E}_p$  defined in (2.13):*

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

*then  $\mathcal{R}_n \asymp \exp[-2p(\log n)^{\alpha/(\alpha+1)}]$  which implies  $n^\beta \mathcal{R}_n \rightarrow +\infty$  for any  $\beta > 0$  but  $(\log n)^\beta \mathcal{R}_n \rightarrow 0$  for any  $\beta > 0$ . Strengthening the assumptions on  $\lambda$  and taking  $\lambda(x) \asymp_\infty x^{-\alpha}$  for some  $\alpha > 1$  then  $\mathcal{R}_n \asymp (\log n)^{-(\alpha-1)}$ .*

**Part II:** *Let  $X$  be non gaussian but satisfy the conditions (2.1). Let  $\rho$  be the auxiliary function of the small ball probability of  $X$ . Assume that  $\rho$  is regularly varying at 0 with index  $\alpha \geq 1$  with either  $\alpha > 1$  or  $\alpha = 1$  and  $\rho(s)/s \asymp \log(1/s)$  then again  $n^\beta \mathcal{R}_n \rightarrow +\infty$  for any  $\beta > 0$ .*

In Part II we recall for the sake of completeness the conditions  $\mathbf{RV}_+$  and  $\mathbf{RV}_1$  introduced earlier. 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 a logarithmic power.

**Remark 8.** An interesting comparison is possible with the linear regression model as studied in [8] since these authors focus on the risk at a fixed point too. Their Theorem 4.2 p. 2168 coupled with the rate given at page 2167 gives a lower bound for a linear  $r$ . This lower bound is possibly parametric (that is  $O(1/n)$ ) for very regular designs. Otherwise it depends on the relative smoothness of  $x_0$  (positively) and of  $X$  through the  $\lambda$ 's (negatively) as well as on the regularity of the unknown slope function but remains polynomial. Obviously the lower bound is damaged when shifting to the general regression model.

The next corollary is stated for completeness because the Theorem above gives only a lower bound for  $\mathcal{R}_n$ . Its proof is omitted since it stems from the derivation of Theorem 3.

**Corollary 2.** *We get in fact  $\mathcal{R}_n(x_0) \asymp 1/(nF(h^*)) \asymp (h^*)^{2p}$  where  $h^*$  is defined at (2.14) and the kernel estimator reaches this optimal rate up to constants.*

These 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 9.** 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  $r$  belongs to a class  $\mathcal{E}_p$  for which:

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

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.

### 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  $\Gamma_0$ , depending solely on its auxiliary function  $\rho$ ? 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 (2.5) denoting  $F \in \Gamma_0$  with auxiliary function  $\rho$ . Simple calculus shows that any function  $\phi F$  where  $\phi(x + t\rho(x))/\phi(x) \rightarrow 1$  when  $x \rightarrow 0$  belongs to  $\Gamma_0$  with exactly the same auxiliary function  $\rho$ . 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  $\Gamma_0$ .

Indeed, pick a function  $F_0$  in the class  $\Gamma_0$ . 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  $\rho$  mentioned just under (1.16) 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  $F_1(s) = \exp\{\eta(s) - \int_s^1 1/\rho_1(t) dt\}$  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  $\rho$  be a self-neglecting function. A measurable function  $\phi$  is called  $\rho$ -self-neglecting if:

$$\frac{\phi(x + t\rho(x))}{\phi(x)} \xrightarrow{x \rightarrow 0} 1.$$

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

**Definition 3.** Pick a  $\rho_0$  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  $\rho_0$  by the relationship  $\Delta$  defined for all  $G$  in  $\Gamma_0$  by:

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

Remind that  $\varphi(t) = t\gamma(t)$  is defined within Proposition 4.

**Theorem 4.** Let  $F_0 \in \Gamma_0$  with auxiliary function  $\rho_0 = 1/\gamma_0$ . Assume that  $\rho_0$  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  $\Gamma_0 \setminus \Delta$  say  $\mathbf{F}_0$ . Then one may pick  $F_0^* \in \mathbf{F}_0$  such that  $F_0^*(\cdot) \sim_0 \mathbb{P}(S < \cdot)$  where  $S = \sum_{i=1}^n \lambda_i Z_i$ , the  $Z_i$ 's follow a  $\chi^2(1)$  distribution and:

$$\lambda_i = \frac{1}{\gamma_0(\varphi^{-1}(i))} = \rho_0(\varphi^{-1}(i))$$

**Remark 10.** Once again we encounter a regularly-varying condition on  $\rho_0$ . Here it echoes in a way the assumption  $\mathbf{A}_0$  (necessary to derive (1.12)) 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  $F \Delta G$  if  $F = \phi G$  where  $\phi$  is  $\rho$ -self-neglecting. The forthcoming Theorem represents these functions  $\phi$ .

**Theorem 5.** Let  $\rho$  be self-neglecting at 0 which does not vanish in a neighborhood of 0. A function  $\phi$  is  $\rho$ -self-neglecting if and only if:

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

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

This theorem slightly generalizes the representation Theorem 2.11.3 for self-neglecting functions p.121 in [5] initially due to [6]. If one take  $\phi = \rho$  the representation above coincides with the one announced in this theorem.

### 3.1. Conclusion and perspectives

The first main result 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 initial formula (1.12) of [41]. We hope that this new formulation will be more convenient for modeling 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 formulae (1.12) and followings seem to be intimately suited to the space  $l_2$ . 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 initially 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-neglecting 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 transforms (see Theorem 4.12.9 in [5]). 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 fact is, as an application of the previous, that the optimal risk in nonparametric regression for functional data is slow in the sense that we cannot expect to obtain polynomial rates in the reasonable setting used in this work. It is obviously interpretable in terms of curse of dimensionality. In the past such similar negative results were obtained in different contexts. We have in mind the optimal rates in deconvolution with supersmooth noise as obtained by [16]. The lower bounds were  $O(1/\log^\tau n)$ ,  $\tau > 0$  and this situation is close to the one encountered here. Fortunately Fan's theorems did not prevent people from making deconvolution with very smoothly corrupted data, but they pointed out the formal limits of the approach.

Here one may hope that this negative result will raise new challenges since it will shift attention to new, more restricted and parsimonious non-linear models like the one introduced in [45]:

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

where the  $r_i$  are functions defined on  $\mathbb{R}$  and estimated from one-dimensional projections of the data  $X$ . It is known since [51] that this model is not subject to the curse of dimensionality when  $X$  is valued in  $\mathbb{R}^k$  for a fixed  $k$ . Letting  $k$  increase with the sample size would be a possible track to introduce non-linearity in regression models for functional data and avoiding some redhibitory features of a general model. A work was carried out by [23] in this direction.

Finally the rather negative result in Theorem 3 should be tempered since it relies heavily on the use of the theoretical norm. Other strategies based on data-driven semi-norms coupled with automatic bandwidth selection provide very satisfactory performances. The reader interested by these novel methods may refer to [46] or to [2]. Like in other areas of statistics, purely numerical techniques may sooner or later overcome the flaws and stumbling stones of theory.

#### 4. Proofs

Considerations about the smoothness at 0 of  $F$  and  $\rho$  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 functions which may not be invertible or have smooth derivatives everywhere. For example the auxiliary function  $\rho$  defined on  $\mathbb{R}^+$  for which we always have  $\rho'(0) = 0$  has no smooth inverse on  $[0, c]$  for  $c > 0$ . But we may frequently use the smoothness of, say,  $\rho$  and  $\rho^{-1}$  on sets  $]a, b[$  for  $0 < a < b$  without always justifying it. This section is split in two subsections. In the first are collected the derivations of results related to small ball probability representation by de Haan's Gamma class. The second is devoted to the upper and lower bounds for the risk at a fixed point in the regression model for functional data.

##### 4.1. Proofs of results of sections 2.1, 2.2 and 3

*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  $\rho$  is self-neglecting, follows the lines of the proof of Proposition 3.10.6 in [5].  $\square$

*Proof of Proposition 2.* Suppose that for some  $p$   $F^{(p)}(0) \neq 0$  and take  $p^* = \inf\{p \in \mathbb{N} : F^{(p)}(0) \neq 0\}$ . 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 a Taylor development shows that  $F(s) \sim_0 cs^{p^*}$ . 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 equation above tends to  $\exp(1)$  whereas the right hand side tends to 1.

Second if  $F^{(p^*)}(0) = +\infty$  we clearly have  $F(s)/s^{p^*} \rightarrow +\infty$  when  $s \rightarrow 0$ . Take  $\varepsilon$  such that  $1/\varepsilon > p^* + 2$ . Since  $\rho'(0) = 0$  and  $\rho$  is positive we may pick an  $s_0$  such that  $\sup_{0 \leq u \leq s_0} \rho'(u) \leq \varepsilon$ . From (1.16) we get:

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

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

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

At last  $F(s)/s^p \leq C'' [\rho(s)]^{\frac{1}{\varepsilon} - p^* - 2}$  which contradicts the fact that  $F(s)/s^{p^*} \rightarrow +\infty$ .  $\square$

We start the proof of Theorem 2.

*Proof of Theorem 2.* From Definition 1 and (1.12) we see that Theorem 2 holds whenever for all  $x \in \mathbb{R}$ :

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

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

$$\gamma(s+x\rho(s))/\gamma(s)\sigma(s+x\rho(s))/\sigma(s) \rightarrow 1, \tag{4.1}$$

$$\exp((s+x\rho(s))\gamma(s+x\rho(s)) - s\gamma(s) - x) \frac{\Lambda(\gamma(s+x\rho(s)))}{\Lambda(\gamma(s))} \rightarrow 1. \tag{4.2}$$

The two next lemmas are dedicated to showing that, in the line above the fraction as well as the exponential both tend to 1 when  $s$  goes to zero and  $\rho$  is chosen as in the Theorem. We just have to clarify formula (2.5) within the Theorem. This stems directly from (1.12). Indeed from (1.13) and (1.14) 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 (1.13) at  $\gamma = 0$  we notice that  $\gamma(r_0) = 0 = \log \Lambda(\gamma(r_0))$  and we conclude.  $\square$

**Lemma 2.** Take  $\rho(s) = 1/\gamma(s)$ , then:

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

**Remark 11.** Obviously  $\gamma$  has at least two (we do not need more) continuous derivatives on a neighborhood of infinity (here  $]1, +\infty[$  for instance). It is also straightforward to see that  $\gamma$ , which is strictly decreasing on  $]1, +\infty[$ , is also a  $C^1$  diffeomorphism on this set. Clearly  $\lim_{s \rightarrow 0} \rho(s) = 0$  but from Lemma 2 p.431 in [41] 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  $s\gamma(s) \rightarrow +\infty$ .

*Proof of Lemma 2.* Denote  $I(s) = s\gamma(s) + \log \Lambda(\gamma(s))$ . We should prove that:

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

Taylor’s formula gives:

$$I(s + x\rho(s)) - I(s) = x\rho(s)I'(s) + \frac{x^2}{2}\rho^2(s)I''(c_{s,x}) \tag{4.3}$$

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 (1.14) we see that:

$$I'(s) = \gamma(s) + s\gamma'(s) + \gamma'(s) \frac{\partial \log \Lambda(\gamma(s))}{\partial \gamma} = \gamma(s).$$

Hence (4.3) 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

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

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  $\rho'$  at 0 and its nullity at 0 (see Remark 11) implies on a one hand that the line above is bounded above for fixed  $x$  and  $s$  (hence  $d$ ) going to zero and also that:

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

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

**Lemma 3.** *We have:*

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

*Proof of Lemma 3.* Once again Taylor’s formula leads to:

$$\frac{\gamma(s + x\rho(s))\sigma(s + x\rho(s)) - \gamma(s)\sigma(s)}{\gamma(s)\sigma(s)} = \frac{x\rho(s)}{\gamma(s)\sigma(s)} [\gamma'(c)\sigma(c) + \gamma(c)\sigma'(c)] \tag{4.4}$$

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

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

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

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

We see in [41], Lemma 2 p.431 that  $\lim_{s \rightarrow 0} \gamma(s)\sigma(s) = +\infty$  and that  $\sigma(s) \leq sc_{13}^{-1}$  where  $c_{13}$  is some constant from which it is plain that  $\sup_{\xi \in \mathcal{V}_0} |\sigma'(\xi)| < +\infty$  where  $\mathcal{V}_0$  is any neighborhood of 0. We deduce that  $\sigma'(\xi)/[\gamma(s)\sigma(s)]$  tends to zero which finally yields

$$\frac{\rho(s)}{\gamma(s)\sigma(s)}\gamma'(c)\sigma(c) \rightarrow 0.$$

We turn to the second term in (4.4):  $\rho(s)\gamma(c)\sigma'(c)/\gamma(s)\sigma(s)$ . We rewrite it:

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

As shown above from Lisfhts' work:  $\gamma\sigma \rightarrow +\infty$ ,  $\sup_{c \in \mathcal{V}_0} |\sigma'(c)| < +\infty$  and  $\gamma(c)/\gamma(s)$  is bounded above and this second term also decays to zero. This finishes the proof of Lemma 3.  $\square$

*Proof of Corollary 1.* We focus on the right hand side of (2.5). First from  $\sqrt{-\gamma'(r)}/\gamma(r) = \sqrt{2\rho'(s)}$  and the properties of the auxiliary function  $\rho$  at Proposition 1 we have that  $\rho'(s) \rightarrow 0$ . Hence  $\mathbb{P}(S < r) \leq \exp[-\int_r^{r_0} 1/\rho(s) ds]$  for  $r$  tending to 0. From the direct part of Karamata's theorem's (see [5], p.26)  $\int_r^{r_0} 1/\rho(s) ds \sim c_\alpha r^{1-\alpha}$  when  $\alpha > 1$  and  $\int_r^{r_0} 1/\rho(s) ds = \zeta(r) \log(1/r)$  with  $\zeta(r) \rightarrow +\infty$  when  $r \rightarrow 0$  (see equation (1.5.8) in [5]). Finally when  $\mathbf{RV}_+$  or  $\mathbf{RV}_1$  hold  $-p \log r - \int_r^{r_0} 1/\rho(s) ds$  always tends to  $-\infty$  whatever the choice of  $p$ .  $\square$

Now we turn to the proof of the converse part of Theorem 2, namely Theorem 4. It takes two steps.

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

**Lemma 4.** *When  $\lambda_i = \rho(\varphi^{-1}(i))$ ,  $\sum \lambda_i < +\infty$ .*

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

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

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

$$\int_0 \rho(u) \varphi'(u) du.$$

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

$$\int_0^{\infty} \rho'(u)\varphi(u)du = \int_0^{\infty} u \frac{\rho'(u)}{\rho(u)} du < +\infty$$

Now we are in a position to apply Karamata’s theorem to  $\rho'$ : since  $\rho$  is regularly varying at 0 with index  $d \geq 1$  (since  $\rho'(0) = 0$ ), and monotone in a right neighborhood of zero,  $\rho'$  is also regularly varying with positive index (see Theorem 1.7.2.(b) p.39 in [5]). Then we can apply the direct part of Karamata’s Theorem to  $\rho'$  (see *ibid.* Theorem 1.5.11 (i) p.28 where the limit should be taken here at zero) and  $\lim_{t \rightarrow 0} t\rho'(t)/\rho(t)$  is finite which ensures that the integral above converges and finally that  $\sum \lambda_i < +\infty$ . This completes the proof of Lemma 4.  $\square$

*Proof of Proposition 4.* We start from (2.7) 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  $a(J_\gamma - 1) \leq \gamma \leq a(J_\gamma)$

$$\begin{aligned} \frac{J_\gamma}{3\gamma} + \sum_{j \geq J_\gamma+1} \frac{1}{a(j) + 2\gamma} &\leq \sum_{j \geq 1} \frac{1}{a(j) + 2\gamma} \leq \frac{J_\gamma}{2\gamma} + \sum_{j \geq J_\gamma+1} \frac{1}{a(j) + 2\gamma} \\ \frac{J_\gamma}{3\gamma} + \frac{1}{3} \sum_{j \geq J_\gamma} \frac{1}{a(j)} &\leq r \leq \frac{J_\gamma}{2\gamma} + \sum_{j \geq J_\gamma} \frac{1}{a(j)} \\ \frac{2}{3} &\leq \frac{r\gamma}{J_\gamma} \leq (0.5 + (J_\gamma^{-1} + 1)). \end{aligned}$$

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

$$\frac{2}{3} \leq \frac{r\gamma}{J_\gamma} \leq \frac{3}{2}. \tag{4.5}$$

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

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

is negative for  $r \downarrow 0$  that is for large  $\gamma$ . Finally we get  $r\gamma/a^{-1}(\gamma) \rightarrow l \in [\frac{2}{3}, \frac{3}{2}]$  when  $\gamma$  tends to  $+\infty$ . The first statement of the Theorem is resulting 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  $c$

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

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

*Proof of Theorem 4.* Pick an  $F_0$  in  $\Gamma_0$  with auxiliary function  $\rho_0$  and consider the function

$$F_0^*(r) = \sqrt{\rho_0'(r)/\pi} \exp \left[ - \int_r^{r_0} ds/\rho_0(s) \right]$$

with  $r_0 = \sum_i \rho(\varphi^{-1}(i))$ . Note that  $\sqrt{\rho'(\cdot)}$  hence  $\rho'(\cdot)$  are  $\rho$ -self-neglecting because:

$$\frac{\rho'(r + x\rho(r))}{\rho'(r)} \rightarrow_{r \rightarrow 0} 1.$$

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

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

This proves that  $F_0^* \Delta F_0$ . It remains to show that  $F_0^* \sim_0 \mathbb{P}(S < \cdot)$ . Like above  $\gamma_0 = 1/\rho_0$ . Start from (2.7) 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 equation (4.5), 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  $\mathbb{P}(S < \cdot) \sim F_0^*$ .  $\square$

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 [5].

**Lemma 5.** *Let  $\rho$  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  $\rho(s)/s \rightarrow 0$  when  $s \rightarrow 0$ ) that for a sufficiently 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  $\rho$  is a non decreasing function  $\rho(x_k) \geq \rho(l)$  for all  $k$ . At last

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

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

*Proof of Theorem 5.* Let  $x_n$  be as in the preceding Lemma. Let  $p$  be a  $C^\infty$  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  $x_n$ ,  $\phi(x_n) = \exp(\int_{x_n}^1 \frac{\varepsilon(u)}{\rho(u)} du)$ . In fact we may always define  $\varepsilon(u)$ ,  $x_0 \leq u \leq 1$  such that  $\phi(x_0) = \exp(\int_{x_0}^1 \frac{\varepsilon(u)}{\rho(u)} du)$ . Then assume that  $\phi(x_k) = \exp(\int_{x_k}^1 \frac{\varepsilon(u)}{\rho(u)} du)$  for  $k = 0, 1, \dots, n$ . 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  $x_{n+1} \leq x \leq x_n$   $\lim_{x \rightarrow 0} \phi(x) / \phi(x_n) = 1$ . We note that  $x = x_n - \lambda_x \rho(x_n)$  where  $\lambda_x \in [0, 1]$  hence  $\lim_{x \rightarrow 0} \phi(x_n - \lambda_x \rho(x_n)) / \phi(x_n) = 1$  uniformly with respect to  $\lambda_x \in [0, 1]$ .

The third and last step is devoted to proving that  $|\varepsilon(u)| \rightarrow 0$  when  $u \rightarrow 0$ . 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  $\rho(x_n - \lambda_u \rho(x_n)) / \rho(x_n) \rightarrow 1$  since  $\rho$  is self-neglecting. Finally by the definition of  $\phi$  we get  $\ln \phi(x_n) / \phi(x_n - \rho(x_n)) \rightarrow 0$  which finishes the proof of the Theorem.  $\square$

#### 4.2. Proofs of results of section 2.3

*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 [19] 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 (2.9))  $\mathcal{V}_n(x_0) \sim \frac{\sigma_\varepsilon^2}{n} \frac{\mathbb{E}K^2(\|X - x_0\|/h)}{[\mathbb{E}K(\|X - x_0\|/h)]^2} \sim \frac{\sigma_\varepsilon^2}{nF_{x_0}(h)}$  which yields the desired result by (1.8).

We should now deal with  $\mathcal{B}_n(x_0)$  defined at (1.2). Ferraty, Mas, Vieu show [19] that  $\mathcal{B}_n(x_0)$  is well approximated by:

$$\left[ \frac{\mathbb{E}(r(X) - r(x_0)) K(\|X - x_0\|/h)}{\mathbb{E}K(\|X_1 - x_0\|/h)} \right]^2. \tag{4.6}$$

A Taylor development at order 2 shows that:

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

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:

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

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

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

where  $g_i$  is 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  $\mathbb{E}[\langle X, e_i \rangle \langle X, e_j \rangle K(\frac{\|X\|}{h})]$  for  $i \neq j$  and:

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

Similar calculations show that:

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

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

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

We can confine now to studying  $\mathbf{v}_i(h)$  in the right hand side just above. We seek an upper bound. Note that whenever  $\|X\| \leq h$  we can assume that for all  $i$   $\langle X, e_i \rangle \leq c\sqrt{\lambda_i}h$  for some constant  $c$ . Then  $\mathbf{v}_i(h) \leq c^2 \lambda_i h^2 F(h)$ . We turn to a lower bound for  $\mathbf{v}_i(h)$ . For simplicity we will make calculations with  $K$  the naive kernel and with a modified norm. In fact we will take  $\|X\| = |\langle X, e_i \rangle| + |\sum_{j \neq i} \langle X, e_j \rangle| = |\langle X, e_i \rangle| + Z_i$ . Let  $0 < \underline{c} < \bar{c}$  be two constants:

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

where the probabilities were split because  $|\langle X, e_i \rangle|$  and  $Z_i$  are independent. Consider first  $\mathbb{P}(|\langle X, e_i \rangle|/\sqrt{\lambda_i} \in \rho(h) [\sqrt{\underline{c}}, \sqrt{\bar{c}}]) \geq \mathbf{c} \rho(h)$  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:

$$F(h - \sqrt{\bar{c}} \lambda_i \rho(h)) \geq F(h - \sqrt{\bar{c}} \lambda_1 \rho(h))$$

which yields:

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

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

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

since  $\eta = \sum_{i=1}^{+\infty} \lambda_i \left( \frac{\partial_{ii}^2 r_{x_0} f_{x_0}(0)}{2} + \partial_i r_{x_0} \partial_i f_{x_0} \right) \neq 0$ . □

*Proof of Theorem 3.* We deal with Part I. The proof comes down to adapting [54], 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  $\mathcal{K} \in \mathcal{E}_p$ , compactly supported and  $\mathcal{K}(0) \geq 1$ . Here  $h^*$  is the solution of the equation (2.14). It is plain that  $|r_0(x_0) - r_1(x_0)| \geq 2(h^*)^p$ . 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  $(z_1^0, \dots, z_n^0)$  (resp.  $(z_1^1, \dots, z_n^1)$ ) 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  $\theta_1(U) = 2(h^*)^p \mathcal{K}(U)$ :

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

by Assumption (2.15). For  $n$  large enough and by (2.10) we deduce that:

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

where  $\mathbf{c}^*$  is some constant and we used (2.14). The decomposition of Hellinger distance for product measures (see [54], 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  $\tau = 2(1 - \exp(-\mathbf{c}^*/4))$ . This proves that  $\mathcal{R}_n(x_0) \geq (h^*)^{2p}$  where  $h^*$  is defined within Lemma 1. And (2.14) ensures that  $\mathcal{R}_n \succ \exp[-2p(\log n)^{\alpha/(\alpha+1)}]$  by straightforward calculations. Part II is treated the same way since the proof relies only on the specific  $h^*$  defined at (2.14).  $\square$

### Acknowledgements

This work benefited from the valuable comments of two anonymous referees. I am also grateful to Professor Mikhail Lifshits for pointing out relevant references in small ball literature and for reading and improving an earlier version of this paper. I initiated this work after some fruitful discussions with the STAPH Group in Toulouse. I wish to thank them all for their support.

### 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. [MR0682802](#)

- [2] AIT-SAAD A., FERRATY F., KASSA R., VIEU P. (2008): Cross-validated estimations in the single-functional index model, *Statistics*, **42**, 475–494. [MR2465129](#)
- [3] AURZADA F., LIFSHITS, M., LINDE W. (2009): Small deviations of stable processes and entropy of associated random operators, *Bernoulli*, **15**, 1305–1334. [MR2597594](#)
- [4] BERLINET A., ELAMINE A., MAS A. (2012): Local linear regression for functional data, *Ann. Instit. Statist. Math.*, **63**, 1047–1075. [MR2822967](#)
- [5] BINGHAM N.H., GOLDIE C.M. AND TEUGELS J.L. (1987): *Regular Variations*. Encyclopedia of Mathematics and Its Applications, Cambridge University Press. [MR0898871](#)
- [6] BLOOM S. (1976): A characterization of B-slowly varying functions, *Proc. Amer. Math. Soc.*, **54**, 243–250. [MR0393375](#)
- [7] BORELL C. (1976): Gaussian Radon measures on locally convex spaces, *Math. Scand.*, **38**, 265–285. [MR0436303](#)
- [8] CAI T., HALL P. (2006): Prediction in functional linear regression, *Ann. Statist.*, **34**, 2159–2179. [MR2291496](#)
- [9] CARDOT H., JOHANNES J. (2010): Thresholding projection estimators in functional linear models, *J. Multivariate Anal.*, **101**, 395–408. [MR2564349](#)
- [10] CRAMBES C., KNEIP A., SARDA P. (2009): Smoothing splines estimators for functional linear regression, *Ann. Statist.*, **37**, 35–72. [MR2488344](#)
- [11] DABO-NIANG S. (2002): Estimation de la densité dans un espace de dimension infinie: Application aux diffusions, *C. R. Math. Acad. Sci. Paris.*, **334**, 213–216. [MR1891061](#)
- [12] DELAIGLE A., HALL P. (2010): Defining probability for a distribution of random functions, *Ann. Statist.*, **38**, 1171–1193. [MR2604709](#)
- [13] DEMBO A., MEYER-WOLF E., ZEITOUNI O. (1995): Exact behavior of Gaussian semi-norms, *Statist. Probab. Lett.*, **23**, 275–280. [MR1340163](#)
- [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. [MR1652320](#)
- [15] ERICKSON K. (1980): Rates of escape of infinite-dimensional Brownian motion, *Ann. Probab.*, **8**, 325–338. [MR0566597](#)
- [16] FAN, J. (1991): On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, **19**, 1257–1272. [MR1126324](#)
- [17] FERRATY F., HALL P., VIEU P. (2010): Most-predictive design points for functional data predictors, *Biometrika.*, **97** n°4, 807–824. [MR2746153](#)
- [18] FERRATY F., LAKSACI A., TADJ A., VIEU P. (2010): Rate of uniform consistency for nonparametric estimates with functional variables. *J. Statist. Plann. Inference*, **140**, no. 2, 335–352. [MR2558367](#)
- [19] FERRATY F., MAS A., VIEU P. (2007): Advances on nonparametric regression from functional data, *Aust. New Zealand J. Stat.*, **49** n°3, 1–20. [MR2396496](#)

- [20] FERRATY F., VAN KEILEGOM I., VIEU P. (2010): On the validity of the bootstrap in the nonparametric functional regression, *Scand. J. Stat.*, **37**, 286–306. [MR2682301](#)
- [21] FERRATY F., VIEU P. (2004): Nonparametric models for functional data, with applications in regression, time series prediction and curve discrimination. *J. Nonparametr. Stat.*, **16**, 111–125. [MR2053065](#)
- [22] FERRATY F., VIEU P. (2006): *Nonparametric Functional Data Analysis: Theory and Practice*. Springer, Berlin. [MR2229687](#)
- [23] FERRATY F., VIEU P. (2009): Additive prediction and boosting for functional data, *Comput. Statist. Data Anal.*, **53**, 1400–1413. [MR2657100](#)
- [24] FERRATY F., VIEU P. (2010): Kernel regression estimation for functional data. In *The Oxford Handbook of Functional data Analysis*, Ferraty and Romain Eds, 72–129. [MR2908020](#)
- [25] FISHER R.A., 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.
- [26] GAIFFAS S. (2005): Convergence rates for pointwise curve estimation with a degenerate design, *Math. Methods Statist.*, **14**, 1–27. [MR2158069](#)
- [27] GARDES L., GIRARD S., LEKINA A. (2010): Functional nonparametric estimation of conditional extremal quantiles, *J. Multivariate Anal.*, **101**, 419–433. [MR2564351](#)
- [28] GNEDENKO B.V. (1943): Sur la distribution limite du terme maximum d’une série aléatoire, *Ann. of Math.*, **44**, 423–453. [MR0008655](#)
- [29] 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. [MR0283854](#)
- [30] DE HAAN L. (1974): Equivalence classes of regularly varying functions, *Stochastic Process. Appl.*, **2**, 243–259. [MR0365664](#)
- [31] HALL P., HECKMAN N. (2002): Estimating and depicting the structure of a distribution of random functions, *Biometrika*, **89**, 145–158. [MR1888371](#)
- [32] HALL P., HOROWITZ J.L.N. (2007): Methodology and convergence rate for functional linear regression. *Ann. Statist.*, **35**, 70–91. [MR2332269](#)
- [33] JUREK Z.J. (1979): A representation of the characteristic functions of stable distributions in Hilbert spaces. *Banach Center Publications*, **5** PWN Warsaw, 151–158. [MR0561477](#)
- [34] KUELBS J., LI W.V., LINDE W. (1994): Lim inf results for Gaussian samples and Chung’s functional LIL, *Ann. Probab.*, **22**, 1879–1903. [MR1331209](#)
- [35] LEDOUX M., TALAGRAND M. (1991): *Probability in Banach Spaces. Isoperimetry and Processes*, Ergebnisse der Mathematik und ihrer Grenzgebiete, **23**, Springer-Verlag. [MR1102015](#)
- [36] LI W.V., LINDE W. (1993): Small ball problems for non centered Gaussian measures, *Probab. Math. Statist.*, **14**, 231–251. [MR1321763](#)
- [37] LI W.V., LINDE W. (1999): Approximation, metric entropy and small ball estimates for Gaussian measures. *Ann. Probab.*, **27**, 1556–1578. [MR1733160](#)
- [38] LI W.V., LINDE W. (2004): Small deviations of stable processes via metric entropy, *J. Theoret. Probab.*, **17**, 261–284. [MR2054588](#)

- [39] LI, W.V., 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. [MR1861734](#)
- [40] LIFSHITS M.A. (1995): Gaussian random functions. *Mathematics and its Applications*, 322. Kluwer Academic Publishers, Dordrecht. [MR1472736](#)
- [41] LIFSHITS M.A. (1997): On the lower tail probabilities of some random series, *Ann. Probab.*, **25**, 424–442. [MR1428515](#)
- [42] MAS A. (2008): Local functional principal component analysis, *Complex Analysis and Operator Theory*, **2**, 135–167. [MR2390677](#)
- [43] MEYER-WOLF E., ZEITOUNI O. (1993): The probability of small Gaussian ellipsoids, *Ann. Probab.*, **21**, 14–24. [MR1207213](#)
- [44] MÜLLER H.G., YAN X. (2001): On local moments, *J. Multivariate Anal.*, **76**, 90–109. [MR1811827](#)
- [45] MÜLLER H.G., YAO F. (2008): Functional additive models, *J. Amer. Statist. Assoc.*, **103**, 1534–1544. [MR2504202](#)
- [46] RACHDI M., VIEU P. (2007): Nonparametric regression for functional data: Automatic smoothing parameter selection, *Journal of Statistical Planning and Inference* **137**, 2784–2801. [MR2323791](#)
- [47] RAMSAY, J.O. AND SILVERMAN, B.W. (2005): *Functional data analysis* (Second Ed.). Springer. [MR2168993](#)
- [48] SKOROHOD, A.V. (1967): On the densities of probability measures in functional spaces. Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66) Vol. II: Contributions to Probability Theory, Part 1, p. 163–182 Univ. California Press, Berkeley, Calif. [MR0214128](#)
- [49] STONE C.J. (1980): Optimal rates of convergence for nonparametric estimators, *Ann. Statist.*, **8**, 1348–1360. [MR0594650](#)
- [50] STONE C.J. (1982): Optimal global rates of convergence for nonparametric regression. *Ann. Statist.*, **10**, 1040–1053. [MR0673642](#)
- [51] STONE C.J. (1985): Additive regression and other nonparametric models, *Ann. Statist.*, **13**, 689–705. [MR0790566](#)
- [52] SYTAYA G.N. (1974): On some asymptotic representations of the Gaussian measure in a Hilbert space, *Theory of Stochastic Processes*, **2**, 93–104.
- [53] TALAGRAND M. (1992): On the rate of convergence in Strassen’s LIL, *Progr. Probab.*, Birkhauser, Boston, 339–351. [MR1227629](#)
- [54] TSYBAKOV A. (2004): Introduction à l’estimation non-paramétrique. *Mathématiques et applications*. Springer. [MR2013911](#)
- [55] VAN DER VAART, A.W., VAN ZANTEN, J.H. (2007): Bayesian inference with rescaled Gaussian process priors *Electronic J. Statist.*, **1**, 433–448. [MR2357712](#)
- [56] VAN 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. [MR2418663](#)