ON THE KOZACHENKO-LEONENKO ENTROPY ESTIMATOR

SYLVAIN DELATTRE AND NICOLAS FOURNIER

Abstract. We study in details the bias and variance of the entropy estimator proposed by Kozachenko and Leonenko [13] for a large class of densities on $\mathbb{R}^d$. We then use the work of Bickel and Breiman [4] to prove a central limit theorem in dimensions 1 and 2. In higher dimensions, we provide a development of the bias in terms of powers of $N^{-2/d}$. This allows us to use a Richardson extrapolation to build, in any dimension, a root-n consistent entropy estimator satisfying a central limit theorem which allows for explicit (asymptotic) confidence intervals. To our knowledge, all the previous general root-n consistency results were concerning dimension 1.

1. Introduction and main results

1.1. **The setting.** Consider a probability measure F on $\mathbb{R}^d$ with density f. We are interested in its entropy defined by

$$H(f) = -\int_{\mathbb{R}^d} f(x)\log f(x)\mathrm{d}x.$$

For $N \geq 1$ and for $X_1, \dots, X_{N+1}$ an i.i.d. sample of F, we consider, for each $i = 1, \dots, N+1$,

$$R_i^N = \min\{|X_i - X_j| \ : \ j = 1, \dots, N+1, \ j \neq i\} \quad \text{and} \quad Y_i^N = N(R_i^N)^d. \tag{1}$$

Here $|\cdot|$ stands for any norm on $\mathbb{R}^d$. For $x \in \mathbb{R}^d$ and $r \geq 0$, we set $B(x,r) = \{y \in \mathbb{R}^d \ : \ |y-x| \leq r\}$ and we introduce $v_d = \int_{B(0,1)} \mathrm{d}x$. We also denote by $\gamma = -\int_0^\infty e^{-x}\log x \mathrm{d}x \simeq 0.577$ the Euler constant. We finally set

$$H_N = \frac{1}{N+1}\sum_{i=1}^{N+1} \log Y_i^N + \gamma + \log v_d. \tag{2}$$

The estimator H_N of $H(f)$ was proposed by Kozachenko and Leonenko [13]. The object of the paper is to study in details the bias, variance and asymptotic normality of H_N.

1.2. **Heuristics.** Let us explain briefly why H_N should be consistent.

The conditional law of Y_i^N knowing X_i is approximately $\mathrm{Exp}(v_d f(X_i))$ for N large: for $r > 0$, $\Pr(Y_i^N > r \,|\, X_i) = [1 - F(B(X_i, (r/N)^{1/d}))]^N \simeq \exp(-NF(B(X_i,(r/N)^{1/d}))) \simeq \exp(-v_d f(X_i) r)$.

Consequently, we expect that $Y_i^N = \xi_i/(v_d f(X_i))$, for a family $(\xi_i)_{i=1,\dots,N+1}$ of approximately Exp(1)-distributed random variables, hopefully not too far from being independent.

We thus expect that $(N+1)^{-1}\sum_{i=1}^{N+1}\log Y_i^N \simeq \mathbb{E}[\log(\xi_1/(v_d f(X_1)))] = \mathbb{E}[\log \xi_1] - \log v_d - \mathbb{E}[\log f(X_1)] \simeq -\gamma - \log v_d + H(f)$ and thus that $H_N \simeq H(f)$.

2010 *Mathematics Subject Classification.* 62G05.

Key words and phrases. Non-parametric estimation, Entropy, Nearest neighbors.

1.3. **Motivation.** Estimating the entropy given some observations seems to be useful in various applied sciences and engineering. Searching for works on this topic, one finds a considerable number of applied papers that we do not try to summarize. Let us mention various fields such as independent component analysis, image analysis, genetic analysis, speech recognition, manifold learning, kinetic physics, molecular chemistry, computational neuroscience, etc. Beirlant, Dudewicz, Györfi and van der Meulen [1] also mention applications to quantization, econometrics and spectroscopy.

The estimation of the relative entropy (or Kullback-Leibler divergence) of f with respect to some known g is deduced from the entropy estimation, since $H(f|g) = -H(f) - \int_{\mathbb{R}^d} f(x)\log g(x)\mathrm{d}x$ and since $\int_{\mathbb{R}^d} f(x)\log g(x)\mathrm{d}x$ is naturally estimated by the root-N asymptotically normal estimator $N^{-1}\sum_{i=1}^N \log g(X_i)$, at least if $\int_{\mathbb{R}^d} f(x)\log^2 g(x)\mathrm{d}x < \infty$.

Concerning applications to statistics, let us mention a few goodness-of-fit tests based on the entropy estimation: see Vasicek [23] for normality (Gaussian laws maximize the entropy among all distributions with given variance), Dudewicz and van der Meulen [6] for uniformity (uniform laws maximize the entropy among all distributions with given support), Mudholkar and Lin [17] for exponentiality (exponential laws maximize the entropy among all $\mathbb{R}_+$-supported distributions with given mean). Also, Robinson [20] proposed an independence test, based on the fact that $f \otimes g$ maximizes $H(h)$ among all densities h with marginals f and g.

1.4. **Available results.** We now list a few mathematical results. Let us first mention the review paper [1] by Beirlant, Dudewicz, Györfi and van der Meulen.

Levit [16] has shown that $\mathbb{V}\mathrm{ar}(\log f(X_1)) = \int_{\mathbb{R}^d} f(x)\log^2 f(x)\mathrm{d}x - (H(f))^2$ is the smallest possible normalized (by N) asymptotic quadratic risk for entropy estimators in the local minimax sense.

Essentially, there are two types of methods for the entropy estimation. The *plug in* method consists in using an estimator of the form $H_N = -\int_{\mathbb{R}^d} f_N(x)\log f_N(x)\mathrm{d}x$, where f_N is an estimator of f. One then needs to use something like a kernel density estimator and this requires to have an idea of the tail behavior of f. Joe [12] considers the case where f is bounded below on its (compact) support, while Hall and Morton [11] propose some root N and asymptotically normal estimators assuming that $f(x) \sim a|x|^{-\alpha}$ (with α known) or $f(x) \sim a\exp(-b|x|^{-\alpha})$ (with α known).

The second class of methods consists in using *spacings* if $d = 1$, see Vasicek [23], or *neighbors* as proposed by Kozachenko and Leonenko [13]. In [13], a consistency result is proved (for H_N defined by (2)), in any dimension, under rather weak conditions on f and this is generalized to other notions of entropies by Leonenko, Pronzato and Savani [14] and Leonenko and Pronzato [15]. Instead of using *nearest* neighbor, we can use k-th nearest neighbors with either k fixed or $1 \ll k \ll N$ (similarly, in dimension 1, we can use k-spacings).

In dimension 1 and assuming that f is bounded below on its (compact) support, Hall [9, 10] and van Es [22] show some root N consistency and asymptotic normality for the entropy estimator based on k spacings (in both cases where k is fixed or tends to infinity at some suitable rate). Tsybakov and van der Meulen [21] are the first to prove some root N consistency for some entropy estimator for general densities with unbounded support, in dimension 1. They consider a modified version of (2) and assume that f is sufficiently regular, positive and has some sub-exponential tails. Still in dimension 1, El Haje and Golubev [7] prove some root N consistency and asymptotic normality for the entropy estimator based on 1-spacings. Furthermore, their assumptions on f are weaker than those of [21]. In particular, they allow f to have some zeroes and some fat tails.

Bickel and Breiman [4] prove a very general central limit theorem for nonlinear bounded functionals of nearest neighbors in any dimension. Unfortunately, this does not apply to H_N and anyway, they do not study the bias. The work of Bickel and Breiman has been generalized in many directions, see Chatterjee [5], Penrose and Yukich [19] and Biau and Devroye [3], with new ideas of proofs. However, these generalizations do not help us much, either because the moment conditions are too strong or because f is supposed to be bounded from below on its compact support. To prove our central limit theorem, the simplest is thus to start from the results of [4], because they are the closest to our framework.

Let us finally mention the paper of Pál, Póczos and Szepesvári [18]: they study other notions of entropy, work in dimension $d \geq 1$, use estimators based on nearest neighbors and give some bounds on their rates of convergence.

1.5. **Notation.** Let $r_0 > 0$ be fixed (we will assume for simplicity that $r_0 = 1$ in the proofs). We introduce the constant κ and the functions $m, M : \mathbb{R}^d \mapsto [0, \infty)$ defined by

$$\kappa = \sup_{x \in \mathbb{R}^d} F(B(x, r_0)), \quad m(x) = \inf_{\varepsilon \in (0, r_0)} \frac{F(B(x, \varepsilon))}{\varepsilon^d} \quad \text{and} \quad M(x) = \sup_{y \in B(x, 2r_0)} f(y). \tag{3}$$

Observe that we have $m \leq v_d f \leq v_d M$ (at least if f is continuous).

For $\beta > 0$, we say that $f \in \mathcal{D}^\beta(\mathbb{R}^d)$ if $f \in C^k(\mathbb{R}^d)$ for $k = \max\{i \in \mathbb{N} \,:\, i < \beta\}$ and if $D^k f$ is locally Hölder continuous with index $\beta - k$. We then set (with $\sum_1^0 = 0$ when $k = 0$)

$$G_\beta(x) = \sum_{i=1}^k |D^i f(x)| + \sup_{y \in B(x, r_0)} \frac{|D^k f(y) - D^k f(x)|}{|x - y|^{\beta - k}}. \tag{4}$$

1.6. **Important convention.** We write $\int_{\mathbb{R}^d} \phi(x) f(x) \mathrm{d}x$ for $\mathbb{E}[\phi(X_1)] = \int_{\{f>0\}} \phi(x) f(x) \mathrm{d}x$, even if ϕ is not well-defined on $\{f = 0\}$.

1.7. **Raw results.** We state here our results in their widest possible generality. The next subsection contains some more comprehensible corollaries. We start with the asymptotic normality and variance.

Theorem 1. *Assume that f is bounded and continuous and that, for some $q > 0$, some $\theta \in (0, 1)$,*

$$\int_{\mathbb{R}^d} \Big(|x|^q + \log^2 m(x) + \frac{\log^2(2 + |x|)}{(F(B(x, r_0)))^\theta} + \frac{M(x)}{m(x)} (1 + |\log m(x)|) \Big) f(x) \mathrm{d}x < \infty. \tag{5}$$

(i) If $\theta > 1/2$, then $\sqrt{N}(H_N - \mathbb{E}[H_N])$ goes in law to $\mathcal{N}(0, \sigma^2(f))$ as $N \to \infty$, where

$$\sigma^2(f) = \int_{\mathbb{R}^d} f(x) \log^2 f(x) \mathrm{d}x - (H(f))^2 + \chi_d,$$

with

$$\chi_d = 2 \log 2 + \frac{\pi^2}{6} - 1 + \int_0^\infty \int_0^\infty e^{-u-v} T(u, v) \frac{\mathrm{d}u}{u} \frac{\mathrm{d}v}{v},$$

and $T : (0, \infty)^2 \mapsto (0, \infty)$ defined by $T(v_d r^d, v_d s^d) = \int_{B(0, r+s) \setminus B(0, r \vee s)} [\exp(\int_{B(0,r) \cap B(y,s)} \mathrm{d}z) - 1] \mathrm{d}y$.

(ii) If $\theta \in (0, 1/2]$, then there is a constant C such that for all $N \geq 1$, $\mathbb{V}\mathrm{ar}\, H_N \leq C N^{-2\theta}$.

Remark 2. *We recall that $v_d = \pi^{d/2} / \Gamma(d/2 + 1)$ for $|\cdot|_2$ and $v_d = 2^d$ for $|\cdot|_\infty$. The constant χ_d depends on d and on the norm $|\cdot|$. We computed it numerically by a Monte-Carlo method in the following situations:*

	d=1	d=2	d=3	d=4	d=5	d=6	d=7	d=8	d=9	d=10	d=20
χ_d for $\lvert\cdot\rvert_2$	2.14	2.29	2.42	2.52	2.61	2.67	2.7	2.7	2.8	2.9	?
χ_d for $\lvert\cdot\rvert_\infty$	2.14	2.31	2.47	2.60	2.70	2.78	2.84	2.88	2.91	2.94	3.03

We indicate in Section 7 how we obtained these numerical values. The results for $|\cdot|_\infty$ are more reliable because our Monte-Carlo method is unbiased. We did not indicate any approximate value of χ_{20} for $|\cdot|_2$ because we obtain too variable results.

We next study the bias.

Theorem 3. *Let $\beta \in (0,2] \cap (0,d]$ and assume that $f \in \mathcal{D}^\beta(\mathbb{R}^d)$. Put $R = M + G_\beta$. Assume that $\kappa < 1$ and that there are $q > 0$ and $\theta \in (0, \beta/d]$ such that*

$$\int_{\mathbb{R}^d} \Big(|x|^q + \frac{\log(2+|x|)}{(F(B(x,r_0)))^\theta} + \frac{R^{2\theta}(x)}{m^{2\theta}(x)} + \frac{R^{\theta d/\beta}(x)}{m^{\theta(d+\beta)/\beta}(x)} \Big) f(x) \mathrm{d}x < \infty. \tag{6}$$

Then there is a constant C such that for all $N \geq 1$,

$$|\mathbb{E}[H_N] - H(f)| \leq \frac{C}{N^\theta}.$$

This provides at best a bias in $O(N^{-2/d})$ and this is a natural limitation: as seen in Subsection 1.2, we need to approximate $v_d f(x) r$ by $NF(B(x,(r/N)^{1/d}))$, for which the error is of order $N^{-2/d}$. In dimensions $1,2,3$, $N^{-2/d} = o(N^{-1/2})$ and we can hope that the bias is negligible when compared to the standard deviation. But in higher dimension, it will be predominant. To get a smaller bias, one possibility is to use a Richardson extrapolation, which requires a polynomial development.

Theorem 4. *Assume that $d \geq 3$, let $\beta \in (2,d]$ and set $\ell = \max\{i \in \mathbb{N} \,:\, 2i < \beta\} \geq 1$. Assume that $f \in \mathcal{D}^\beta(\mathbb{R}^d)$. Put $R = M + G_\beta$. Assume that $\kappa < 1$ and that there is $q > 0$ such that*

$$\int_{\mathbb{R}^d} \Big(|x|^q + \frac{\log(2+|x|)}{(F(B(x,r_0)))^{\beta/d}} + \frac{R^{2\beta/d}(x)}{m^{2\beta/d}(x)} + \frac{R^{\beta/2}(x)}{m^{\beta/d+\beta/2}(x)} \Big) f(x) \mathrm{d}x < \infty. \tag{7}$$

Then there are $\lambda_1, \dots, \lambda_\ell \in \mathbb{R}$ and $C > 0$ such that for all $N \geq 1$,

$$\Big| \mathbb{E}[H_N] - H(f) - \sum_{i=1}^{\ell} \frac{\lambda_i}{N^{2i/d}} \Big| \leq \frac{C}{N^{\beta/d}}.$$

Remark 5. *The λ_i's can be made explicit. In particular, if the norm is symmetric (that is, $|(x_{\sigma(1)}, \dots, x_{\sigma(d)})| = |(x_1, \dots, x_d)|$ for any permutation σ), there is $c > 0$, depending only on d and on the norm, such that $\lambda_1 = c \int_{\mathbb{R}^d} f^{-2/d-1}(x) |\nabla f(x)|^2 \mathrm{d}x$ provided $\lim_{|x|\to\infty} f^{-2/d}(x)|\nabla f(x)| = 0$.*

To produce some confidence intervals, we need to estimate $\sigma^2(f)$, which is not very difficult.

Proposition 6. *Assume that f is bounded and continuous and that for some $q > 0$, some $\theta > 0$,*

$$\int_{\mathbb{R}^d} \Big(|x|^q + \log^2 m(x) + \frac{\log^2(2+|x|)}{(F(B(x,r_0)))^\theta} \Big) f(x) \mathrm{d}x < \infty. \tag{8}$$

Fix $N \geq 1$, recall (1) and put

$$V_N = \frac{1}{N+1} \sum_{i=1}^{N+1} \log^2 Y_i^N - \Big(\frac{1}{N+1} \sum_{i=1}^{N+1} \log Y_i^N \Big)^2 + \chi_d - \frac{\pi^2}{6}.$$

Then V_N goes in probability to $\sigma^2(f)$, defined in Theorem 1, as $N \to \infty$.

1.8. **Corollaries.** We now sacrifice generality to present a more comprehensible statement with a central limit theorem sufficient to produce explicit (asymptotic) confidence intervals. Weaker results can be derived from the theorems of the previous section under weaker assumptions, we give an example at the end of the subsection.

If $d = 1, 2, 3$, we set $H_N^{(d)} = H_N$ and $a_d = 1$. If $d \geq 4$, it is necessary to proceed to some extrapolation to get a bias in $o(N^{-1/2})$. We consider $\ell = \lfloor d/4 \rfloor$ and the real numbers $\alpha_{0,d}, \dots, \alpha_{\ell,d}$ satisfying $\sum_{k=0}^{\ell} \alpha_{k,d} = 1$ and, for all $i = 1, \dots, \ell$, $\sum_{k=0}^{\ell} \alpha_{k,d} 2^{2ki/d} = 0$. For $N \geq \ell$, we put $n = \lfloor (N + 1 - \ell)/(2^{\ell+1} - 1) \rfloor$, so that $N + 1 \geq \sum_{k=0}^{\ell}(2^k n + 1)$. We thus can split our $(N + 1)$-sample into $\ell + 1$ (independent) sub-samples of sizes $2^\ell n + 1, 2^{\ell-1}n + 1, \dots, n + 1$ and build with these sub-samples the estimators $H^0_{2^\ell n}, H^1_{2^{\ell-1}n}, \dots H^\ell_n$ exactly as in (1)-(2). Finally, we set

$$H_N^{(d)} = \sum_{k=0}^{\ell} \alpha_{k,d} H^k_{2^{\ell-k}n} \tag{9}$$

and put $a_d = (2 - 2^{-\ell}) \sum_{k=0}^{\ell} \alpha_{k,d}^2 2^k$.

Corollary 7. *Fix $\varepsilon \in (0, 1)$. Assume that $f \in \mathcal{D}^\nu(\mathbb{R}^d)$ with $\nu = 1$ if $d = 1$, $\nu = 2$ if $d \in \{2, 3\}$ and $\nu = d/2 + \varepsilon$ if $d \geq 4$. Fix $r_0 > 0$ such that $\kappa < 1$, assume that $R = M + G_\nu$ is bounded and that there is $c > 0$ such that $m(x) \geq cf(x)$ for all $x \in \mathbb{R}^d$. Assume finally that*

$$\int_{\mathbb{R}^d} |x|^{d+\varepsilon} f(x)\mathrm{d}x < \infty \tag{10}$$

and that

$$\int_{\{f>0\}} \sqrt{R(x)} f^{-\varepsilon}(x)\mathrm{d}x < \infty \quad \textit{if } d \in \{1, 2\} \quad \textit{and} \quad \int_{\{f>0\}} R^{d/4}(x) f^{1/2-d/4-\varepsilon}(x)\mathrm{d}x < \infty \quad \textit{if } d \geq 3. \tag{11}$$

Then

$$\sqrt{\frac{N}{V_N}} \Big(H_N^{(d)} - H(f) \Big) \longrightarrow \mathcal{N}(0, a_d) \quad \textit{in law as } N \to \infty.$$

Remark 8. *For $d \in \{4, 5, 6, 7\}$, we have $\ell = 1$, $\alpha_{0,d} = 2^{2/d}/(2^{2/d} - 1)$ and $\alpha_{1,d} = -1/(2^{2/d} - 1)$. This gives $a_4 \simeq 34.97$, $a_5 \simeq 54.97$, $a_6 \simeq 79.65$, $a_7 \simeq 109.01$.*

These values are very large and it is not clear, in practice, that the extrapolation is judicious, except if N is very large. However, this is the only way we have to propose a satisfying *theoretical* result when $d \geq 4$.

Let us discuss our assumptions.

• We assume some regularity on f. This is natural, since we need to use that f is well-approximated by its means of on small balls.

• We suppose that $m \geq cf$ for some constant $c > 0$. This condition does not seem very stringent but is not automatically verified, even in dimension 1 and with f Lipschitz continuous and compactly supported, as shown by the following counter-example. Consider the hat function $\chi(x) = \max\{1 - |x|, 0\}$ and the density $f(x) = a \sum_{n \geq 2} \chi_n(x)$, where $\chi_n(x) = n^{-3}\chi(n^3(x - n^{-1}))$ and where a is a normalization constant. In words, χ_n is the hat function centered at n^{-1} with height n^{-3} and with width $2n^{-3}$. The supports of χ_n and χ_ℓ are disjoint for any $\ell > n \geq 2$. The function f is thus Lipschitz continuous with Lipschitz constant a. And of course, f is compactly supported. Fix any $r_0 > 0$ and take this value to define m as in (3). Then one easily finds $n_0 \geq 2$

such that for all $n \geq n_0$, one has (i) $r_n := n^{-2}/3 < r_0$, (ii) $[n^{-1} - r_n, n^{-1} + r_n]$ does not intercept $\cup_{\ell \neq n} \operatorname{Supp} \chi_\ell$. Hence for all $n \geq n_0$,

$$m(n^{-1}) \leq \frac{F([n^{-1} - r_n, n^{-1} + r_n])}{r_n} = 3n^2 \int_{n^{-1}-r_n}^{n^{-1}+r_n} f(x)\mathrm{d}x \leq 3an^2 \int_0^1 \chi_n(x)\mathrm{d}x = \frac{3an^2}{n^9}.$$

Since $f(n^{-1}) = an^{-3}$, we conclude that $\lim_{n\to\infty} m(n^{-1})/f(n^{-1}) = 0$, so that there cannot exist $c > 0$ such that $m(x) \geq cf(x)$ for all x.

• We have some decay condition: f needs to have a moment of order strictly larger than d. This is not very stringent, but we believe this is a technical condition. Observe that when f is sufficiently regularly varying and positive so that $f \simeq R \simeq m$, in some weak sense to be made precise (we think here of Gaussian distributions or of examples (a), (b), (c) below), then our main conditions are $\int_{\mathbb{R}^d} |x|^{d+\varepsilon} f(x)\mathrm{d}x < \infty$ and $\int_{\mathbb{R}^d} f^{1/2-\varepsilon}(x)\mathrm{d}x < \infty$. And the first condition implies the second one (with actually a smaller $\varepsilon' > 0$), see Remark 28-(i). To summarize, for a non-vanishing and very regularly varying f, our main restriction is (10), which is not very stringent but probably technical.

• When $d \in \{1, 2\}$, f is plainly allowed to vanish: for example, Corollary 7 applies to any compactly supported f, provided it is of class $C^d(\mathbb{R}^d)$ (whence R is bounded), provided $m \geq cf$ and provided there is $\varepsilon > 0$ such that $\int_{\{f>0\}} f^{-\varepsilon}(x)\mathrm{d}x < \infty$. This is rather general. On the contrary, f cannot vanish if $d \geq 10$. Indeed, consider x_0 on the boundary of $\{f > 0\}$, so that $R(x_0) > 0$. Since $f \in \mathcal{D}^{d/2}(\mathbb{R}^d)$ by assumption, we deduce that $f(x_0 + h) \leq C|h|^{d/2}$ for $h \in \mathbb{R}^d$ small enough, so that $R(x_0+h)f^{1/2-d/4-\varepsilon}(x_0+h) \geq c|h|^{d(1/2-d/4-\varepsilon)/2}$, which cannot be integrable around $h = 0$ if $(d/4 - 1/2)/2 \geq 1$, i.e. if $d \geq 10$. For $d \in \{3, \dots, 9\}$, f is allowed to vanish with restrictions. Again, we believe these restrictions are technical.

Finally, let us state a corollary with a bad rate of convergence but with very few assumptions.

Corollary 9. *Assume that $f \in \mathcal{D}^\nu(\mathbb{R}^d)$ with $\nu = \min\{d, 2\}$. Fix $r_0 > 0$ such that $\kappa < 1$, assume that $R = M + G_\nu$ is bounded and that there is $c > 0$ such that $m(x) \geq cf(x)$ for all $x \in \mathbb{R}^d$. Assume finally* (10) *for some $\varepsilon > 0$, that $\int_{\{f>0\}} M(x)|\log f(x)|\mathrm{d}x < \infty$ and that*

$$\int_{\mathbb{R}^d} \sqrt{R(x)}\mathrm{d}x < \infty \quad \textit{if } d = 1 \quad \textit{and} \quad \int_{\mathbb{R}^d} R^{d/(2+d)}(x)\mathrm{d}x < \infty \quad \textit{if } d \geq 2. \tag{12}$$

Then there is a constant $C > 0$ such that for all $N \geq 1$, $\mathbb{E}[(H_N - H(f))^2] \leq CN^{-1}$ if $d = 1$ and $\mathbb{E}[(H_N - H(f))^2] \leq CN^{-4/(d+2)}$ if $d \geq 2$.

Any $f \in C^{\min\{d,2\}}(\mathbb{R}^d)$ with compact support such that $m \geq cf$ and $\int_{\{f>0\}} |\log f(x)|\mathrm{d}x < \infty$ satisfies the assumptions of Corollary 9. It is of course possible, using Theorems 1, 3 and 4 to derive intermediate statements between Corollaries 7 and 9.

1.9. **Examples.** Recall that we have an explicit central limit theorem when Corollary 7 applies. Some arguments are given at the end of the paper.

(a) If $f(x) = (2\pi)^{-d/2}\exp(-|x|^2/2)$ (or any other non-degenerate normal distribution), Corollary 7 applies.

(b) If $f(x) = c_{d,a}e^{-(1+|x|^2)^{a/2}}$ for some $a > 0$, then Corollary 7 applies.

(c) If $f(x) = c_{d,a}(1 + |x|^2)^{-(d+a)/2}$ (Student law) with $a > d$, then Corollary 7 applies.

(d) If $f(x) = c_{d,a}|x|^a e^{-|x|}$ (or $f(x) = c_a x^a e^{-x}\mathbf{1}_{\{x>0\}}$ if $d = 1$), then Corollary 7 applies when $d = 1, a \geq 1$, when $d = 2$, $a \geq 2$, when $d = 3$, $a \in [2, 12)$ and when $d \in \{4, \dots, 9\}$, $a \in (d/2, 4d/(d-2))$. For any $d \geq 3$, $a \geq 2$, Corollary 9 applies.

(e) If $f(x) = c_{a,b} \prod_{i=1}^d x_i^{a_i}(1 - x_i)^{b_i}\mathbf{1}_{\{x_i\in[0,1]\}}$ for some $a = (a_1, \dots, a_d)$ and $b = (b_1, \dots, b_d)$ both in $(0, \infty)^d$, we set $\tau = \min\{a_1, b_1, \dots, a_d, b_d\}$ and $\mu = \max\{a_1, b_1, \dots, a_d, b_d\}$. Corollary 7 applies if $d = 1$, $\tau \geq 1$ or $d = 2$, $\tau \geq 2$ or $d = 3$, $2 \leq \tau \leq \mu < 4$. For any $d \geq 3$, $\tau \geq 2$, Corollary 9 applies.

(f) If $d = 1$ and $f(x) = c_p x^p |\sin(\pi/x)|\mathbf{1}_{\{x\in(0,1)\}}$ for some $p \geq 2$, then Corollary 7 applies.

This last example is of course far-fetched, but it shows that our results apply to densities with many zeroes (in dimensions 1 and 2).

1.10. **Comparison with previous results.** We will use the results of Bickel and Breiman [4]: for $\varepsilon \in (0, 1)$, for $\log_\varepsilon$ a bounded approximation of log and for H_N^ε the corresponding cutoff version of (2), it holds that $\sqrt{N}(H_N^\varepsilon - \mathbb{E}[H_N^\varepsilon]) \to \mathcal{N}(0, \sigma_\varepsilon^2(f))$. This paper is thus very interesting and applies to very general densities f but does not include the entropy as an admissible functional. Furthermore, they do not quantify the bias.

As already mentioned, the results of Hall and Morton [11] apply to densities of which we know quite precisely the tail behavior. This is a rather stringent condition.

The results of Hall [9, 10], van Es [22], Tsybakov and van der Meulen [21] and El Haje and Golubev [7] only concern the one-dimensional case. Furthermore, [9, 10, 22] apply only to densities bounded below on their (compact) support. A lot of regularity is assumed in [21]: in our list of examples, on (a) and (b) (with $a > 1$) are included. Still when $d = 1$, it is difficult to compare our results with those of [7], because in both cases, the assumptions are not very transparent. Let us however mention that their study seems to include example (c) for all $a > 0$ (with a CLT), while we have to assume that $a > 1$. On the contrary, they suppose that $f > 0$ has a finite number of connected components, so that they cannot deal with example (f).

When $d \geq 2$, the only quantified consistency result seems to be that of Pál, Póczos and Szepesvári. They assume that f is compactly supported and study other notions of entropy, but, if we extrapolate, we find some estimate looking like $|H_N - H(f)| \leq CN^{-1/(2d)}$ with high probability. Recall that the bias is actually in $N^{-2/d}$.

As a conclusion, it seems we provide the first root N and asymptotic normality result for a general entropy estimator in dimension $d \geq 2$.

However, our assumptions are not very transparent and probably far from optimal, at least when $d \geq 3$. Also, our proofs are rather tedious: quoting Bickel and Breiman [4], "we believe this is due to the complexity of the problem".

Let us finally mention that Berrett, Samworth and Yuan [2] posted on arxiv, a few months after we did, a paper on the Kozachenko-Leonenko estimator using k-nearest neighbors with $1 \ll k \ll N$. In the first version, they were limited to the case $d \leq 3$, because their study of the bias was not sufficiently precise. This was fixed in a second version. At the end, they are able to provide an *efficient* entropy estimator in any dimension. The assumptions are not directly comparable to ours but resemble much.

1.11. **Plan of the paper.** In the next section, we compute some conditional laws and prove some easy estimates of constant use. Section 3 is devoted to the proof of Theorem 1 (central limit theorem). We prove Proposition 6 in Section 4 (estimation of the variance). In Section 5, we study very precisely how well f is approximated its mean on a small ball. In Section 6, we handle the proofs of Theorems 3 and 4 (concerning the bias). Finally, the corollaries are verified and we discuss the examples and the numerical computation of χ_d in Section 7.

1.12. **Notation.** We recall that we write F for the law of X_1 and f for its density. The functions m, M, G_β are defined in (3)-(4). We introduce some shortened notation:

$$\mathbf{f}_1 = f(X_1), \quad \mathbf{f}_2 = f(X_2), \quad \mathbf{m}_1 = m(X_1), \quad \mathbf{m}_2 = m(X_2), \quad \mathbf{M}_1 = M(X_1), \quad \mathbf{M}_2 = M(X_2),$$
$$\mathbf{a}_1^N(r) = F(B(X_1,(r/N)^{1/d})), \quad \mathbf{a}_2^N(r) = F(B(X_2,(r/N)^{1/d})),$$
$$\mathbf{b}_{12}^N(r,s) = F(B(X_1,(r/N)^{1/d}) \cup B(X_2,(s/N)^{1/d})).$$

We write C for a finite constant used in the upperbounds and c for a positive constant used in the lowerbounds. Their values do never depend on N, but are allowed to change from line to line.

During the whole proof, we assume that $r_0 = 1$ for simplicity.

2. Preliminaries

To start with, we compute some conditional laws.

Lemma 10. *For $N \geq 1$ and $r, s > 0$, we have*

$$\Pr(Y_1^N > r \,|\, X_1) = (1 - \mathbf{a}_1^N(r))^N,$$
$$\Pr(Y_1^N > r, Y_2^N > s \,|\, X_1, X_2) = \mathbf{1}_{\{|X_1 - X_2| > (\frac{r \vee s}{N})^{1/d}\}}(1 - \mathbf{b}_{12}^N(r,s))^{N-1}.$$

Proof. By definition, see (1),

$$\{Y_1^N > r\} = \bigcap_{i=2}^{N+1} \big\{X_i \notin B(X_1,(r/N)^{1/d})\big\}.$$

The first claim follows. Next,

$$\{Y_1^N > r, Y_2^N > s\} = \big\{|X_1 - X_2| > \big(\frac{r \vee s}{N}\big)^{1/d}\big\} \bigcap \bigcap_{i=3}^{N+1} \big\{X_i \notin B(X_1,(r/N)^{1/d}) \cup B(X_2,(s/N)^{1/d})\big\}.$$

This implies the second claim. $\square$

We next verify some estimates of constant use.

Lemma 11. *(i) For all $N \geq 1$, all $r \in [0, N]$, we have $\mathbf{a}_1^N(r) \leq v_d \mathbf{M}_1 r/N$.*

(ii) For all $N \geq 1$, all $r \in [0, N]$, we have $(1 - \mathbf{a}_1^N(r))^N \leq \exp(-\mathbf{m}_1 r)$.

(iii) For all $N \geq 2$, all $r \in [0, N]$, we have $(1 - \mathbf{a}_1^N(r))^{N-1} \leq \exp(-\mathbf{m}_1 r/2)$.

(iv) If $\int_{\mathbb{R}^d} |x|^q f(x) \mathrm{d}x < \infty$, put $g(x) = 1 \vee \mathbb{E}[|X_1 - x|^q]$, which is bounded by $C(1 + |x|^q)$. For all $N \geq 1$, all $r > 0$, we have $1 - \mathbf{a}_1^N(rN) \leq \mathbf{g}_1/r^{q/d}$, where $\mathbf{g}_1 = g(X_1)$.

Proof. For (i), we use that $\sup_{B(X_1,(r/N)^{1/d})} f \leq \mathbf{M}_1$ since $r \leq N$. Consequently, $\mathbf{a}_1^N(r) \leq \mathbf{M}_1 \mathrm{Leb}(B(X_1,(r/N)^d)) = v_d \mathbf{M}_1 r/N$. For (ii), we write $(1 - \mathbf{a}_1^N(r))^N \leq \exp(-N\mathbf{a}_1^N(r))$ and we use that $N\mathbf{a}_1^N(r) \geq N\mathbf{m}_1 r/N = \mathbf{m}_1 r$. Point (iii) is checked similarly, using that $(N-1)/N \geq 1/2$ for all $N \geq 2$. By the Markov inequality, we have $1 - F(B(x, r^{1/d})) = \Pr(|X_1 - x| > r^{1/d}) \leq g(x)/r^{q/d}$ and (iv) follows from the fact that $\mathbf{a}_1^N(rN) = F(B(X_1, r^{1/d}))$. $\square$

3. Variance and central limit theorem

This section is devoted to the proof of Theorem 1. We first cut H_N in pieces.

Lemma 12. *For $\varepsilon \in (0,1]$ and $y>0$, we introduce $\log_\varepsilon y = \log(\varepsilon \vee y \wedge \varepsilon^{-1})$. Then we can write, for any $N \geq 1/\varepsilon$, $H_N = H_N^\varepsilon + K_N^{1,\varepsilon} + K_N^{2,\varepsilon} + K_N^3$, where*

$$H_N^\varepsilon = \frac{1}{N+1}\sum_{i=1}^{N+1} \log_\varepsilon Y_i^N + \gamma + \log v_d, \quad K_N^{1,\varepsilon} = \frac{1}{N+1}\sum_{i=1}^{N+1} \log[(Y_i^N/\varepsilon)\wedge 1],$$

$$K_N^{2,\varepsilon} = \frac{1}{N+1}\sum_{i=1}^{N+1} \log[1 \vee (\varepsilon Y_i^N)\wedge(\varepsilon N)], \quad K_N^3 = \frac{1}{N+1}\sum_{i=1}^{N+1} \log[(Y_i^N/N)\vee 1].$$

Proof. Recall that $N \geq 1/\varepsilon$. It suffices to note that for all $y \in (0,\infty)$, we have

$$\log y = \log_\varepsilon y + \log[(y/\varepsilon)\wedge 1] + \log[1\vee(\varepsilon y)\wedge(\varepsilon N)] + \log[(y/N)\vee 1].$$

This is easily checked separating the cases $y \in (0,\varepsilon]$, $y\in(\varepsilon,1/\varepsilon]$, $y \in (1/\varepsilon, N]$ and $y \in (N,\infty)$. □

We next apply the result of Bickel and Breiman [4].

Proposition 13. *Assume that f is bounded and continuous and fix $\varepsilon \in (0,1]$. We then have $\sup_{N\geq 1/\varepsilon} N\mathbb{V}\mathrm{ar}\, H_N^\varepsilon < \infty$. Furthermore, $\sqrt{N}(H_N^\varepsilon - \mathbb{E}[H_N^\varepsilon])$ goes in law to $\mathcal{N}(0,\sigma_\varepsilon^2(f))$ as $N\to\infty$, where $\sigma_\varepsilon^2(f) = A_\varepsilon + B_\varepsilon + C_\varepsilon$, with*

$$A_\varepsilon = \mathbb{E}\Big[\int_0^\infty \log_\varepsilon^2(r^d/\mathbf{f}_1)\mu_0(\mathrm{d}r)\Big] - \mathbb{E}\Big[\int_0^\infty \log_\varepsilon(r^d/\mathbf{f}_1)\mu_0(\mathrm{d}r)\Big]^2,$$

$$B_\varepsilon = \mathbb{E}\Big[\int_0^\infty\int_0^\infty \log_\varepsilon(r_1^d/\mathbf{f}_1)\log_\varepsilon(r_2^d/\mathbf{f}_2)\mu_1(\mathrm{d}r_1,\mathrm{d}r_2)\Big],$$

$$C_\varepsilon = \mathbb{E}\Big[\int_0^\infty\int_0^\infty \log_\varepsilon(r_1^d/\mathbf{f}_1)\log_\varepsilon(r_2^d/\mathbf{f}_1)\mu_2(\mathrm{d}r_1,\mathrm{d}r_2)\Big],$$

where the finite (signed) measures μ_0 on $[0,\infty)$ and μ_1,μ_2 on $[0,\infty)^2$ are defined by

$$\mu_0([0,r]) = 1-e^{-v_d r^d}, \quad \mu_1([0,r_1]\times[0,r_2]) = e^{-v_d(r_1^d+r_2^d)}[v_d r_1^d + v_d r_2^d - v_d^2 r_1^d r_2^d],$$

$$\mu_2([0,r_1]\times[0,r_2]) = e^{-v_d(r_1^d+r_2^d)}[T(v_d r_1^d, v_d r_2^d) - v_d(r_1\vee r_2)^d],$$

with $T:(0,\infty)^2 \mapsto \mathbb{R}_+$ defined by $T(v_d r^d, v_d s^d) = \int_{B(0,r+s)\setminus B(0,r\vee s)}[\exp(\int_{B(0,r)\cap B(y,s)}\mathrm{d}z)-1]\mathrm{d}y$.

Proof. Since $f:\mathbb{R}^d \mapsto [0,\infty)$ and $\log_\varepsilon : [0,\infty)\mapsto\mathbb{R}$ are bounded and continuous, we can apply [4, Theorems 3.5 and 4.1] (with the notation therein, $D_i = (Y_i^N)^{1/d}$, we thus take $h(x,r) = \log_\varepsilon(r^d)$, whence $\tilde{h}(x,r) = \log_\varepsilon(r^d/f(x))$). This first theorem precisely tells us $\lim_{N\to\infty} N\mathbb{V}\mathrm{ar}\, H_N^\varepsilon = \sigma_\varepsilon^2(f)$ (whence of course $\sup_{N\geq 1/\varepsilon} N\mathbb{V}\mathrm{ar}\, H_N^\varepsilon < \infty$) and the second one tells us that $\sqrt{N}(H_N^\varepsilon - \mathbb{E}[H_N^\varepsilon])$ goes in law to $\mathcal{N}(0,\sigma_\varepsilon^2(f))$. Actually, there is a typo in [4]: $L_0(\mathrm{d}r)$ in (3.6) has to be a non-negative measure, so that $L_0(r)$ (see (3.1)) has to be replaced by $-L_0(r)$ or, as we did, by $1-L_0(r)$. □

Remark 14. *For all $r_1,r_2>0$ and $u_1,u_2>0$,*

$$T(v_d r_1^d, v_d r_2^d) \leq d2^{d-1} v_d (r_1\vee r_2)^{d-1}(r_1\wedge r_2)e^{v_d(r_1\wedge r_2)^d},$$

$$T(u_1,u_2) \leq d2^{d-1}(u_1\vee u_2)^{1-1/d}(u_1\wedge u_2)^{1/d}e^{u_1\wedge u_2}.$$

Proof. Assume $r_1 \geq r_2$. We have $\int_{B(0,r_1)\cap B(y,r_2)} \mathrm{d}z \leq \mathrm{Leb}(B(y,r_2)) = v_d r_2^d$. Since furthermore $\mathrm{Leb}(B(0,r_1+r_2)\setminus B(0,r_1)) = v_d((r_1+r_2)^d - r_1^d)$, $T(v_d r_1^d, v_d r_2^d) \leq v_d((r_1+r_2)^d - r_1^d)e^{v_d r_2^d}$. But $((r_1+r_2)^d - r_1^d) \leq d(r_1+r_2)^{d-1} r_2 \leq d2^{d-1} r_1^{d-1} r_2$, which proves the first inequality. The second inequality follows from the first one applied to $r_1 = (u_1/v_d)^{1/d}$ and $r_2 = (u_2/v_d)^{1/d}$. □

Lemma 15. *Assume that f is bounded and continuous and that $\int_{\mathbb{R}^d} f(x)\log^2 f(x)\mathrm{d}x < \infty$. Then $\lim_{\varepsilon\to 0}\sigma_\varepsilon^2(f) = \sigma^2(f)$, with $\sigma^2(f)$ defined in Theorem 1.*

Proof. For $\xi \sim \mathrm{Exp}(1)$ and independent of X_1, it holds that $\Pr((\xi/v_d)^{1/d} > r) = \exp(-v_d r^d)$. We thus can write $A_\varepsilon = \mathbb{V}\mathrm{ar}\,(\log_\varepsilon(\xi/(v_d\mathbf{f}_1)))$. By dominated convergence, we will have that $\lim_{\varepsilon\to 0} A_\varepsilon = \mathbb{V}\mathrm{ar}\,(\log(\xi/(v_d\mathbf{f}_1))) =: A$, provided $\mathbb{E}[\log^2(\xi/(v_d\mathbf{f}_1))] < \infty$. This is the case, because $\int_{\mathbb{R}^d} f(x)\log^2 f(x)\mathrm{d}x < \infty$ by assumption. Now by independence, $A = \mathbb{V}\mathrm{ar}\,(\log\xi) + \mathbb{V}\mathrm{ar}\,(\log\mathbf{f}_1)$. First, it holds that $\mathbb{V}\mathrm{ar}\,(\log\mathbf{f}_1) = \int_{\mathbb{R}^d} f(x)\log^2 f(x)\mathrm{d}x - (H(f))^2$. We also have $\mathbb{V}\mathrm{ar}\,(\log\xi) = \pi^2/6$, because $\mathbb{E}[\log\xi] = \int_0^\infty(\log r)e^{-r}\mathrm{d}r = \Gamma'(1) = -\gamma$ and because $\mathbb{E}[\log^2\xi] = \int_0^\infty(\log^2 r)e^{-r}\mathrm{d}r = \Gamma''(1) = \gamma^2 + \pi^2/6$ (both equailities are well-known and can be derived from the Weierstrass formula, see [24, page 236]). We have proved that $\lim_{\varepsilon\to 0} A_\varepsilon = \int_{\mathbb{R}^d} f(x)\log^2 f(x)\mathrm{d}x - (H(f))^2 + \pi^2/6$.

An integration by parts (more precisely, writing $\log_\varepsilon r = \log(1/\varepsilon) - \int_r^\infty \mathbf{1}_{\{u\in(\varepsilon,1/\varepsilon)\}}\frac{\mathrm{d}u}{u}$, using the Fubini theorem and that $\mu_1([0,r_1)\times[0,r_2)) = 0$ if $r_1\vee r_2 = \infty$) shows that

$$B_\varepsilon = \mathbb{E}\Big[\int_0^\infty\int_0^\infty \frac{d\mathbf{1}_{\{r_1^d/\mathbf{f}_1\in(\varepsilon,1/\varepsilon)\}}\mathrm{d}r_1}{r_1}\frac{d\mathbf{1}_{\{r_2^d/\mathbf{f}_2\in(\varepsilon,1/\varepsilon)\}}\mathrm{d}r_2}{r_2} e^{-v_d(r_1^d+r_2^d)}[v_d r_1^d + v_d r_2^d - v_d^2 r_1^d r_2^d]\Big].$$

Using now the change of variables $(u_1,u_2) = (v_d r_1^d, v_d r_2^d)$, we find that

$$B_\varepsilon = \mathbb{E}\Big[\int_{v_d\mathbf{f}_1\varepsilon}^{v_d\mathbf{f}_1/\varepsilon}\int_{v_d\mathbf{f}_2\varepsilon}^{v_d\mathbf{f}_2/\varepsilon} e^{-u_1-u_2}(u_1+u_2-u_1u_2)\frac{\mathrm{d}u_2}{u_2}\frac{\mathrm{d}u_1}{u_1}\Big].$$

Similarly (observe that $\mu_2([0,r_1)\times[0,r_2))$ is null as soon as $r_1\vee r_2 = \infty$ by Remark 14),

$$C_\varepsilon = \mathbb{E}\Big[\int_{v_d\mathbf{f}_1\varepsilon}^{v_d\mathbf{f}_1/\varepsilon}\int_{v_d\mathbf{f}_1\varepsilon}^{v_d\mathbf{f}_1/\varepsilon} e^{-u_1-u_2}(T(u_1,u_2) - u_1\vee u_2)\frac{\mathrm{d}u_2}{u_2}\frac{\mathrm{d}u_1}{u_1}\Big].$$

Unfortunately, neither B_ε nor C_ε converge as $\varepsilon\to 0$. We write $B_\varepsilon + C_\varepsilon = I_\varepsilon + J_\varepsilon + K_\varepsilon$, where

$$I_\varepsilon = \mathbb{E}\Big[\int_{v_d\mathbf{f}_1\varepsilon}^{v_d\mathbf{f}_1/\varepsilon}\int_{v_d\mathbf{f}_2\varepsilon}^{v_d\mathbf{f}_2/\varepsilon} e^{-u_1-u_2}(u_1+u_2-u_1\vee u_2 - u_1u_2)\frac{\mathrm{d}u_2}{u_2}\frac{\mathrm{d}u_1}{u_1}\Big],$$

$$J_\varepsilon = \mathbb{E}\Big[\int_{v_d\mathbf{f}_1\varepsilon}^{v_d\mathbf{f}_1/\varepsilon}\int_{v_d\mathbf{f}_1\varepsilon}^{v_d\mathbf{f}_1/\varepsilon} e^{-u_1-u_2}T(u_1,u_2)\frac{\mathrm{d}u_2}{u_2}\frac{\mathrm{d}u_1}{u_1}\Big],$$

$$K_\varepsilon = \mathbb{E}\Big[\int_{v_d\mathbf{f}_1\varepsilon}^{v_d\mathbf{f}_1/\varepsilon}\Big(\int_{v_d\mathbf{f}_2\varepsilon}^{v_d\mathbf{f}_2/\varepsilon} e^{-u_2}(u_1\vee u_2)\frac{\mathrm{d}u_2}{u_2} - \int_{v_d\mathbf{f}_1\varepsilon}^{v_d\mathbf{f}_1/\varepsilon} e^{-u_2}(u_1\vee u_2)\frac{\mathrm{d}u_2}{u_2}\Big)e^{-u_1}\frac{\mathrm{d}u_1}{u_1}\Big].$$

Since $|u_1+u_2-u_1\vee u_2 - u_1u_2| = |u_1\wedge u_2 - u_1u_2| \leq u_1u_2 + \sqrt{u_1u_2} \in L^1(\mathbb{R}_+^2, e^{-u_1-u_2}\frac{\mathrm{d}u_1}{u_1}\frac{\mathrm{d}u_2}{u_2})$, we see that $\lim_{\varepsilon\to 0} I_\varepsilon = \int_0^\infty\int_0^\infty e^{-u_1-u_2}(u_1+u_2-u_1\vee u_2 - u_1u_2)\frac{\mathrm{d}u_2}{u_2}\frac{\mathrm{d}u_1}{u_1} =: I$ by dominated convergence. But $I = 2\int_0^\infty\int_0^{u_1} e^{-u_1-u_2}u_2(1-u_1)\frac{\mathrm{d}u_2}{u_2}\frac{\mathrm{d}u_1}{u_1} = 2\int_0^\infty e^{-u_1}(1-e^{-u_1})(1-u_1)\frac{\mathrm{d}u_1}{u_1} = 2\log 2 - 1$. Indeed, introduce $\varphi(t) = 2\int_0^\infty e^{-u_1}(1-e^{-tu_1})(1-u_1)\frac{\mathrm{d}u_1}{u_1}$: we have $\varphi(0) = 0$ and $\varphi'(t) = 2\int_0^\infty e^{-u_1}e^{-tu_1}(1-u_1)\mathrm{d}u_1 = 2t/(1+t)^2$, so that $\varphi(1) = \int_0^1[2t/(1+t)^2]\mathrm{d}t = 2\log 2 - 1$.

Next, we deduce from Remark 14 that $T \in L^1(\mathbb{R}_+^2, e^{-u_1-u_2}\frac{\mathrm{d}u_1}{u_1}\frac{\mathrm{d}u_2}{u_2})$, whence $\lim_{\varepsilon\to 0} J_\varepsilon = \int_0^\infty\int_0^\infty e^{-u_1-u_2}T(u_1,u_2)\frac{\mathrm{d}u_2}{u_2}\frac{\mathrm{d}u_1}{u_1}$.

Finally, we verify that $\lim_{\varepsilon\to 0} K_\varepsilon = 0$. Using a symmetry argument (and that $\mathbf{f}_1, \mathbf{f}_2$ are i.i.d.),

$$K_\varepsilon = 2\mathbb{E}\Big[\int_{v_d\mathbf{f}_1\varepsilon}^{v_d\mathbf{f}_1/\varepsilon}\Big(\int_{v_d\mathbf{f}_2\varepsilon}^{v_d\mathbf{f}_2/\varepsilon} e^{-u_2}\mathbf{1}_{\{u_2<u_1\}}\frac{\mathrm{d}u_2}{u_2} - \int_{v_d\mathbf{f}_1\varepsilon}^{v_d\mathbf{f}_1/\varepsilon} e^{-u_2}\mathbf{1}_{\{u_2<u_1\}}\frac{\mathrm{d}u_2}{u_2}\Big)e^{-u_1}\mathrm{d}u_1\Big].$$

Since next $\mathbf{f}_1$ and $\mathbf{f}_2$ have the same law, it holds that

$$L_\varepsilon = 2\mathbb{E}\Big[\int_0^\infty\Big(\int_{v_d\mathbf{f}_2\varepsilon}^{v_d\mathbf{f}_2/\varepsilon} e^{-u_2}\mathbf{1}_{\{u_2<u_1\}}\frac{\mathrm{d}u_2}{u_2} - \int_{v_d\mathbf{f}_1\varepsilon}^{v_d\mathbf{f}_1/\varepsilon} e^{-u_2}\mathbf{1}_{\{u_2<u_1\}}\frac{\mathrm{d}u_2}{u_2}\Big)e^{-u_1}\mathrm{d}u_1\Big] = 0.$$

Hence $|K_\varepsilon| = |K_\varepsilon - L_\varepsilon| \le 2\mathbb{E}[\int_0^{v_d\mathbf{f}_1\varepsilon}|F_\varepsilon(u_1,\mathbf{f}_1,\mathbf{f}_2)|e^{-u_1}\mathrm{d}u_1 + \int_{v_d\mathbf{f}_1/\varepsilon}^\infty |F_\varepsilon(u_1,\mathbf{f}_1,\mathbf{f}_2)|e^{-u_1}\mathrm{d}u_1]$, where $F_\varepsilon(u_1,\mathbf{f}_1,\mathbf{f}_2) = \int_{v_d\mathbf{f}_2\varepsilon}^{v_d\mathbf{f}_2/\varepsilon} e^{-u_2}\mathbf{1}_{\{u_2<u_1\}}\frac{\mathrm{d}u_2}{u_2} - \int_{v_d\mathbf{f}_1\varepsilon}^{v_d\mathbf{f}_1/\varepsilon} e^{-u_2}\mathbf{1}_{\{u_2<u_1\}}\frac{\mathrm{d}u_2}{u_2}$. But

$$|F_\varepsilon(u_1,\mathbf{f}_1,\mathbf{f}_2)| \le 2 + |\log(v_d\mathbf{f}_1\varepsilon)| + |\log(v_d\mathbf{f}_2\varepsilon)| \le C(1+|\log\mathbf{f}_1| + |\log\mathbf{f}_2| + |\log\varepsilon|)$$

and we end with

$$K_\varepsilon \le C\mathbb{E}\Big[(1-e^{-v_d\mathbf{f}_1\varepsilon} + e^{-v_d\mathbf{f}_1/\varepsilon})(1+|\log\mathbf{f}_1| + |\log\mathbf{f}_2| + |\log\varepsilon|)\Big].$$

Using that f is bounded, we see that $1-e^{-v_d\mathbf{f}_1\varepsilon} \le C\varepsilon$ so that, since $\mathbb{E}[|\log\mathbf{f}_1|] < \infty$ by assumption, we clearly have $\lim_{\varepsilon\to 0}\mathbb{E}[(1-e^{-v_d\mathbf{f}_1\varepsilon})(1+|\log\mathbf{f}_1| + |\log\mathbf{f}_2| + |\log\varepsilon|)] = 0$. Next, $\lim_{\varepsilon\to 0}\mathbb{E}[e^{-v_d\mathbf{f}_1/\varepsilon}(1+|\log\mathbf{f}_1| + |\log\mathbf{f}_2|)] = 0$ by dominated convergence. At last,

$$\mathbb{E}[e^{-v_d\mathbf{f}_1/\varepsilon}] \le e^{-v_d/\sqrt{\varepsilon}} + \mathrm{Pr}\left(\mathbf{f}_1 < \sqrt{\varepsilon}\right) \le e^{-v_d/\sqrt{\varepsilon}} + \mathrm{Pr}\left(|\log\mathbf{f}_1| > |\log\sqrt{\varepsilon}|\right) \le e^{-v_d/\sqrt{\varepsilon}} + \frac{4\mathbb{E}[\log^2\mathbf{f}_1]}{\log^2\varepsilon}.$$

We conclude that $\lim_{\varepsilon\to 0}|\log\varepsilon|\mathbb{E}[e^{-v_d\mathbf{f}_1/\varepsilon}] = 0$, so that $\lim_{\varepsilon\to 0} K_\varepsilon = 0$.

Recalling that $\sigma_\varepsilon^2(f) = A_\varepsilon + B_\varepsilon + C_\varepsilon = A_\varepsilon + I_\varepsilon + J_\varepsilon + K_\varepsilon$, we have checked that $\lim_{\varepsilon\to 0}\sigma_\varepsilon^2(f) = \int_{\mathbb{R}^d} f(x)\log^2 f(x)\mathrm{d}x - (H(f))^2 + \pi^2/6 + 2\log 2 - 1 + \int_0^\infty\int_0^\infty e^{-u_1-u_2}T(u_1,u_2)\frac{\mathrm{d}u_2}{u_2}\frac{\mathrm{d}u_1}{u_1}$, which is nothing but $\sigma^2(f)$. □

It remains to show that $K_N^{1,\varepsilon}$, $K_N^{2,\varepsilon}$ and K_N^3 are sufficiently small. We first give some expressions of their variances.

Lemma 16. *Recall that $K_N^{1,\varepsilon}$, $K_N^{2,\varepsilon}$ and K_N^3 were defined, for $\varepsilon\in(0,1]$ and $N\ge 1/\varepsilon$, in Lemma 12. For $r,s\ge 0$, put*

$$\Gamma_N(r,s) = \mathbf{1}_{\{|X_1-X_2|>(\frac{r\vee s}{N})^{1/d}\}}(1-\mathbf{b}_{12}^N(r,s))^{N-1} - (1-\mathbf{a}_1^N(r))^N(1-\mathbf{a}_2^N(s))^N.$$

We have $\mathbb{V}\mathrm{ar}\,K_N^{1,\varepsilon} \le W_N^{11,\varepsilon} + W_N^{12,\varepsilon}$, $\mathbb{V}\mathrm{ar}\,K_N^{2,\varepsilon} \le W_N^{21,\varepsilon} + W_N^{22,\varepsilon}$ *and* $\mathbb{V}\mathrm{ar}\,K_N^3 \le W_N^{31} + W_N^{32}$, *where*

$$W_N^{11,\varepsilon} = \frac{2}{N+1}\int_0^1 \mathbb{E}[1-(1-\mathbf{a}_1^N(\varepsilon r))^N]\log\frac{1}{r}\frac{\mathrm{d}r}{r},\quad W_N^{12,\varepsilon} = \frac{2N}{N+1}\int_0^1\int_0^r \mathbb{E}[\Gamma_N(\varepsilon r,\varepsilon s)]\frac{\mathrm{d}s}{s}\frac{\mathrm{d}r}{r},$$

$$W_N^{21,\varepsilon} = \frac{2}{N+1}\int_1^{\varepsilon N} \mathbb{E}[(1-\mathbf{a}_1^N(r/\varepsilon))^N]\log r\frac{\mathrm{d}r}{r},\quad W_N^{22,\varepsilon} = \frac{2N}{N+1}\int_1^{\varepsilon N}\int_1^r \mathbb{E}[\Gamma_N(r/\varepsilon, s/\varepsilon)]\frac{\mathrm{d}s}{s}\frac{\mathrm{d}r}{r},$$

$$W_N^{31} = \frac{2}{N+1}\int_1^\infty \mathbb{E}[(1-\mathbf{a}_1^N(Nr))^N]\log r\frac{\mathrm{d}r}{r},\quad W_N^{32} = \frac{2N}{N+1}\int_1^\infty\int_1^r \mathbb{E}[\Gamma_N(Nr,Ns)]\frac{\mathrm{d}s}{s}\frac{\mathrm{d}r}{r}.$$

Proof. By exhangeability, we have $\mathbb{V}\mathrm{ar}\, K_N^{1,\varepsilon} \le W_N^{11,\varepsilon} + W_N^{12,\varepsilon}$, $\mathbb{V}\mathrm{ar}\, K_N^{2,\varepsilon} \le W_N^{21,\varepsilon} + W_N^{22,\varepsilon}$ and $\mathbb{V}\mathrm{ar}\, K_N^{3} \le W_N^{31} + W_N^{32}$, where

$$\begin{aligned}
W_N^{11,\varepsilon} &= (N+1)^{-1}\mathbb{E}[\log^2[(Y_1^N/\varepsilon)\wedge 1]],\\
W_N^{21,\varepsilon} &= (N+1)^{-1}\mathbb{E}[\log^2[1\vee(\varepsilon Y_1^N)\wedge(\varepsilon N)]]\\
W_N^{31} &= (N+1)^{-1}\mathbb{E}[\log^2[(Y_1^N/N)\vee 1]],\\
W_N^{12,\varepsilon} &= (N+1)^{-1}N\mathbb{C}\mathrm{ov}(\log[(Y_1^N/\varepsilon)\wedge 1],\log[(Y_2^N/\varepsilon)\wedge 1]),\\
W_N^{22,\varepsilon} &= (N+1)^{-1}N\mathbb{C}\mathrm{ov}(\log[1\vee(\varepsilon Y_1^N)\wedge(\varepsilon N)],\log[1\vee(\varepsilon Y_2^N)\wedge(\varepsilon N)]),\\
W_N^{32} &= (N+1)^{-1}N\mathbb{C}\mathrm{ov}(\log[(Y_1^N/N)\vee 1],\log[(Y_2^N/N)\vee 1]).
\end{aligned}$$

Since $\log^2 z = 2\int_1^\infty \mathbf{1}_{\{r<z\}}\log r\frac{\mathrm{d}r}{r} + 2\int_0^1 \mathbf{1}_{\{r\ge z\}}\log\frac{1}{r}\frac{\mathrm{d}r}{r}$ for all $z>0$, it holds that for Z a positive random variable,

$$\mathbb{E}[\log^2 Z] = 2\int_1^\infty \Pr(Z>r)\log r\frac{\mathrm{d}r}{r} + 2\int_0^1 \Pr(Z\le r)\log\frac{1}{r}\frac{\mathrm{d}r}{r}. \tag{13}$$

Using Lemma 10, we easily conclude that

$$\begin{aligned}
W_N^{11,\varepsilon} &= \frac{2}{N+1}\int_0^1 \Pr(Y_1^N\le\varepsilon r)\log\frac{1}{r}\frac{\mathrm{d}r}{r} = \frac{2}{N+1}\int_0^1 \mathbb{E}[1-(1-\mathbf{a}_1^N(\varepsilon r))^N]\log\frac{1}{r}\frac{\mathrm{d}r}{r},\\
W_N^{21,\varepsilon} &= \frac{2}{N+1}\int_1^\infty \Pr((\varepsilon Y_1^N)\wedge(\varepsilon N)>r)\log r\frac{\mathrm{d}r}{r} = \frac{2}{N+1}\int_1^{\varepsilon N} \mathbb{E}[(1-\mathbf{a}_1^N(r/\varepsilon))^N]\log r\frac{\mathrm{d}r}{r},\\
W_N^{31} &= \frac{2}{N+1}\int_1^\infty \Pr(Y_1^N/N>r)\log r\frac{\mathrm{d}r}{r} = \frac{2}{N+1}\int_1^\infty \mathbb{E}[(1-\mathbf{a}_1^N(Nr))^N]\log r\frac{\mathrm{d}r}{r}.
\end{aligned}$$

Also, for Z_1 and Z_2 two positive random variables,

$$\mathbb{C}\mathrm{ov}(\log Z_1,\log Z_2) = \int_0^\infty\int_0^\infty \Big(\Pr(Z_1>r,Z_2>s)-\Pr(Z_1>r)\Pr(Z_2>s)\Big)\frac{\mathrm{d}s}{s}\frac{\mathrm{d}r}{r}. \tag{14}$$

Indeed, it suffices to use that $\log z = \int_0^\infty(\mathbf{1}_{\{r<z\}}-\mathbf{1}_{\{r<1\}})\frac{\mathrm{d}r}{r}$ and the bilinearity of the covariance. For $r,s>0$, we deduce from Lemma 10 (and the independence of $\mathbf{a}_1^N(r)$ and $\mathbf{a}_2^N(s)$) that

$$\begin{aligned}
&\Pr(Y_1^N>r,Y_2^N>s)-\Pr(Y_1^N>r)\Pr(Y_2^N>s)\\
=&\mathbb{E}\Big[\mathbf{1}_{\{|X_1-X_2|>(\frac{r\vee s}{N})^{1/d}\}}(1-\mathbf{b}_{12}^N(r,s))^{N-1}\Big] - \mathbb{E}[(1-\mathbf{a}_1^N(r))^N]\mathbb{E}[(1-\mathbf{a}_2^N(s))^N]\\
=&\mathbb{E}[\Gamma_N(r,s)].
\end{aligned} \tag{15}$$

Consequently, $\Pr((Y_1^N/\varepsilon)\wedge 1>r,(Y_2^N/\varepsilon)\wedge 1>s) - \Pr((Y_1^N/\varepsilon)\wedge 1>r)\Pr((Y_2^N/\varepsilon)\wedge 1>s) = \mathbb{E}[\Gamma_N(\varepsilon r,\varepsilon s)]\mathbf{1}_{\{r,s\in[0,1]\}}$, whence by (14),

$$W_N^{12,\varepsilon} = \frac{N}{N+1}\int_0^1\int_0^1 \mathbb{E}[\Gamma_N(\varepsilon r,\varepsilon s)]\frac{\mathrm{d}s}{s}\frac{\mathrm{d}r}{r} = \frac{2N}{N+1}\int_0^1\int_0^r \mathbb{E}[\Gamma_N(\varepsilon r,\varepsilon s)]\frac{\mathrm{d}s}{s}\frac{\mathrm{d}r}{r}.$$

Similarly, we deduce from (15) the equality $\Pr(1\vee(\varepsilon Y_1^N)\wedge(\varepsilon N)>r, 1\vee(\varepsilon Y_2^N)\wedge(\varepsilon N)>s) - \Pr(1\vee(\varepsilon Y_1^N)\wedge(\varepsilon N)>r)\Pr(1\vee(\varepsilon Y_2^N)\wedge(\varepsilon N)>s) = \mathbb{E}[\Gamma_N(r/\varepsilon,s/\varepsilon)]\mathbf{1}_{\{r,s\in[1,\varepsilon N]\}}$, so that

$$W_N^{22,\varepsilon} = \frac{N}{N+1}\int_1^{\varepsilon N}\int_1^{\varepsilon N} \mathbb{E}[\Gamma_N(r/\varepsilon,s/\varepsilon)]\frac{\mathrm{d}s}{s}\frac{\mathrm{d}r}{r} = \frac{2N}{N+1}\int_1^{\varepsilon N}\int_1^r \mathbb{E}[\Gamma_N(r/\varepsilon,s/\varepsilon)]\frac{\mathrm{d}s}{s}\frac{\mathrm{d}r}{r}.$$

Finally, $\Pr((Y_1^N/N)\vee 1 > r, (Y_2^N/N)\vee 1 > s) - \Pr((Y_1^N/N)\vee 1 > r)\Pr((Y_2^N/N)\vee 1 > s) = \mathbb{E}[\Gamma_N(Nr,Ns)]\mathbf{1}_{\{r,s\geq 1\}}$, whence

$$W_N^{32} = \frac{N}{N+1}\int_1^\infty\int_1^\infty \mathbb{E}[\Gamma_N(Nr,Ns)]\frac{\mathrm{d}s}{s}\frac{\mathrm{d}r}{r} = \frac{2N}{N+1}\int_1^\infty\int_1^r \mathbb{E}[\Gamma_N(Nr,Ns)]\frac{\mathrm{d}s}{s}\frac{\mathrm{d}r}{r}.$$

The proof is complete. □

We now study the terms $W_N^{11,\varepsilon}$, $W_N^{12,\varepsilon}$, $W_N^{21,\varepsilon}$, $W_N^{22,\varepsilon}$, W_N^{31} and W_N^{32}.

Lemma 17. *Assume that f is bounded. Then $NW_N^{11,\varepsilon} \leq C\varepsilon$ for all $\varepsilon\in(0,1]$ and all $N\geq 1/\varepsilon$.*

Proof. Recall that $W_N^{11,\varepsilon} = 2(N+1)^{-1}\mathbb{E}[\int_0^1(1-(1-\mathbf{a}_1^N(\varepsilon r))^N)\log\frac{1}{r}\frac{\mathrm{d}r}{r}]$. But $1-(1-x)^N\leq Nx$ for $x\in[0,1]$, so that $1-(1-\mathbf{a}_1^N(\varepsilon r))^N \leq N\mathbf{a}_1^N(\varepsilon r)\leq v_d\mathbf{M}_1\varepsilon r\leq C\varepsilon r$ by Lemma 11-(i) and since f is bounded. Thus

$$W_N^{11,\varepsilon}\leq \frac{C\varepsilon}{N+1}\int_0^1\log\frac{1}{r}\mathrm{d}r\leq\frac{C\varepsilon}{N}$$

as desired. □

Lemma 18. *Assume that $\int_{\mathbb{R}^d}[\log^2 m(x)]f(x)\mathrm{d}x<\infty$. Then we have $NW_N^{21,1}\leq C$ for all $N\geq 1$ and $\lim_{\varepsilon\to 0}\sup_{N\geq 1/\varepsilon} NW_N^{21,\varepsilon}=0$.*

Proof. Recall that $W_N^{21,\varepsilon} = 2(N+1)^{-1}\mathbb{E}[\int_1^{\varepsilon N}(1-\mathbf{a}_1^N(r/\varepsilon))^N\log r\frac{\mathrm{d}r}{r}]$. Thanks to Lemma 11-(ii) (and since $r/\varepsilon\leq N$),

$$NW_N^{21,\varepsilon}\leq 2\mathbb{E}\Big[\int_1^\infty e^{-\mathbf{m}_1 r/\varepsilon}\log r\frac{\mathrm{d}r}{r}\Big].$$

If $\varepsilon=1$, we use that $2\int_1^\infty e^{-ar}\log r\frac{\mathrm{d}r}{r} = \int_1^\infty ae^{-ar}\log^2 r\mathrm{d}r = \int_a^\infty e^{-t}\log^2(t/a)\mathrm{d}t\leq C(1+\log^2 a)$ and that $\mathbb{E}[\log^2\mathbf{m}_1]<\infty$ by assumption to conclude that $\sup_{N\geq 1}NW_N^{21,1}<\infty$. The fact that $\lim_{\varepsilon\to 0}\sup_{N\geq 1/\varepsilon}NW_N^{21,\varepsilon}=0$ follows from the dominated convergence theorem. □

Lemma 19. *If $\int_{\mathbb{R}^d}(|x|^q+[\log^2(2+|x|)][F(B(x,1))]^{-\theta})f(x)\mathrm{d}x<\infty$ for some $q>0$ and some $\theta>0$, then $NW_N^{31}\leq CN^{-\theta}$ for all $N\geq 1$.*

Proof. Recall that $W_N^{31} = 2(N+1)^{-1}\mathbb{E}[\int_1^\infty(1-\mathbf{a}_1^N(Nr))^N\log r\frac{\mathrm{d}r}{r}]$. For $g(x)=1\vee\mathbb{E}[|X_1-x|^q]$ as in Lemma 11, we write $NW_N^{31}\leq 2A_N+2B_N$, where

$$A_N = \mathbb{E}\Big[\int_1^{(2\mathbf{g}_1)^{d/q}}(1-\mathbf{a}_1^N(Nr))^N\log r\frac{\mathrm{d}r}{r}\Big]\quad\text{and}\quad B_N=\mathbb{E}\Big[\int_{(2\mathbf{g}_1)^{d/q}}^\infty(1-\mathbf{a}_1^N(Nr))^N\log r\frac{\mathrm{d}r}{r}\Big].$$

Using that $(1-\mathbf{a}_1^N(rN))^N\leq(\mathbf{g}_1/r^{q/d})^N$ by Lemma 11-(iv),

$$B_N\leq\mathbb{E}\Big[\int_{(2\mathbf{g}_1)^{d/q}}^\infty\Big(\frac{\mathbf{g}_1}{r^{q/d}}\Big)^N\log r\frac{\mathrm{d}r}{r}\Big]\leq\Big(\frac{1}{2}\Big)^{N-1}\mathbb{E}\Big[\mathbf{g}_1\int_1^\infty\frac{\log r}{r^{1+q/d}}\mathrm{d}r\Big]\leq\frac{C}{2^N}.$$

because $\mathbb{E}[\mathbf{g}_1]<\infty$ (recall that $g(x)\leq C(1+|x|^q)$).

Next, there is a constant C such that $(1-x)^N\leq e^{-Nx}\leq C(Nx)^{-\theta}$ for all $x\in(0,1]$, whence $(1-\mathbf{a}_1^N(rN))^N\leq C(N\mathbf{a}_1^N(rN))^{-\theta}\leq C(N\mathbf{a}_1^N(N))^{-\theta}=C(NF(B(X_1,1)))^{-\theta}$ for all $r\geq 1$. Thus

$$A_N\leq\frac{C}{N^\theta}\mathbb{E}\Big[\int_1^{(2\mathbf{g}_1)^{d/q}}\frac{1}{(F(B(X_1,1)))^\theta}\log r\frac{\mathrm{d}r}{r}\Big]\leq\frac{C}{N^\theta}\mathbb{E}\Big[\frac{\log^2(2\mathbf{g}_1)^{d/q}}{(F(B(X_1,1)))^\theta}\Big]\leq\frac{C}{N^\theta}$$

because $\mathbb{E}[(F(B(X_1,1)))^{-\theta}\log^2(2\mathbf{g}_1)^{d/q}]<\infty$ (observe that $\log^2(2g(x))^{d/q}\le C\log^2(2+|x|))$. All in all, $NW_N^{31}\le C2^{-N}+CN^{-\theta}\le CN^{-\theta}$ as desired. □

Lemma 20. *If* $\int_{\mathbb{R}^d}(|x|^q+[\log^2(2+|x|)][F(B(x,1))]^{-\theta})f(x)\mathrm{d}x<\infty$ *for some* $q>0$ *and some* $\theta>0$, *then* $NW_N^{32}\le CN^{1-2\theta}$ *for all* $N\ge 2$ *(we do not claim that* $|NW_N^{32}|\le CN^{1-2\theta}$*).*

Proof. Recall that $W_N^{32}=2(N/(N+1))\int_1^\infty\int_1^r\mathbb{E}[\Gamma_N(Nr,Ns)]\frac{\mathrm{d}s}{s}\frac{\mathrm{d}r}{r}$ and that we have the obvious bound $\Gamma_N(Nr,Ns)\le(1-\mathbf{b}_{12}^N(Nr,Ns))^{N-1}$. Introducing $g(x)=1\vee\mathbb{E}[|X_1-x|^q]$ as in Lemma 11, we write $W_N^{32}\le 2(A_N+B_N)$, where

$$A_N=\mathbb{E}\Big[\int_1^{(2\mathbf{g}_1)^{d/q}}\int_1^s(1-\mathbf{b}_{12}^N(rN,sN))^{N-1}\frac{\mathrm{d}s}{s}\frac{\mathrm{d}r}{r}\Big],$$
$$B_N=\mathbb{E}\Big[\int_{(2\mathbf{g}_1)^{d/q}}^{\infty}\int_1^s(1-\mathbf{b}_{12}^N(rN,sN))^{N-1}\frac{\mathrm{d}s}{s}\frac{\mathrm{d}r}{r}\Big].$$

First, we use that $1-\mathbf{b}_{12}^N(rN,sN)\le 1-\mathbf{a}_1^N(rN)\le\mathbf{g}_1r^{-q/d}$ by Lemma 11-(iv), whence

$$B_N\le\mathbb{E}\Big[\int_{(2\mathbf{g}_1)^{d/q}}^{\infty}\Big(\frac{\mathbf{g}_1}{r^{q/d}}\Big)^{N-1}\log r\frac{\mathrm{d}r}{r}\Big]\le\frac{C}{2^N}$$

as in the previous proof (here $N\ge 2$).

Next, there is a constant C such that $(1-x)^{N-1}\le e^{-(N-1)x}\le C/(Nx)^{2\theta}$ for all $N\ge 2$, all $x\in(0,1]$. Moreover, for all $r,s\ge 1$,

$$\mathbf{b}_{12}^N(Nr,Ns)\ge F(B(X_1,1))\vee F(B(X_2,1))\ge[F(B(X_1,1))F(B(X_2,1))]^{1/2}.$$

Thus $(1-\mathbf{b}_{12}^N(Nr,Ns))^{N-1}\le CN^{-2\theta}[F(B(X_1,1))F(B(X_2,1))]^{-\theta}$. Since $2\int_1^a\int_1^r\frac{\mathrm{d}s}{s}\frac{\mathrm{d}r}{r}=\log^2 a$,

$$A_N\le\frac{C}{N^{2\theta}}\mathbb{E}\Big[\frac{\log^2(2\mathbf{g}_1)^{d/q}}{[F(B(X_1,1))F(B(X_2,1))]^\theta}\Big]\le\frac{C}{N^{2\theta}}\mathbb{E}\Big[\frac{\log^2(2+|X_1|)}{[F(B(X_1,1))]^\theta}\Big]\mathbb{E}\Big[\frac{1}{[F(B(X_2,1))]^\theta}\Big]\le\frac{C}{N^{2\theta}}.$$

We used that $\log^2(2g(x))^{d/q}\le C\log^2(2+|x|)$, recall that $g(x)\le C(1+|x|^q)$. All in all, $W_N^{32}\le CN^{-2\theta}+C2^{-N}\le CN^{-2\theta}$. □

Finally, we treat $W_N^{12,\varepsilon}$ and $W_N^{22,\varepsilon}$, which are more difficult.

Lemma 21. *If* $\int_{\mathbb{R}^d}(M(x)/m(x))(1+|\log m(x)|)f(x)\mathrm{d}x<\infty$, *then* $N(W_N^{12,1}+W^{22,1})\le C$ *for all* $N\ge 2$ *and* $\limsup_{\varepsilon\to 0}\sup_{N\ge 1/\varepsilon}N(W_N^{12,\varepsilon}+W_N^{22,\varepsilon})\le 0$.

Proof. Recall that $\Gamma_N(r,s)=\mathbf{1}_{\{|X_1-X_2|>(\frac{r\vee s}{N})^{1/d}\}}(1-\mathbf{b}_{12}^N(r,s))^{N-1}-(1-\mathbf{a}_1^N(r))^N(1-\mathbf{a}_2^N(s))^N$ and that, for $N\ge 1/\varepsilon$,

$$W_N^{12,\varepsilon}=\frac{2N}{N+1}\int_0^1\int_0^r\mathbb{E}[\Gamma_N(\varepsilon r,\varepsilon s)]\frac{\mathrm{d}s}{s}\frac{\mathrm{d}r}{r}=\frac{2N}{N+1}\int_0^\varepsilon\int_0^r\mathbb{E}[\Gamma_N(r,s)]\frac{\mathrm{d}s}{s}\frac{\mathrm{d}r}{r},$$
$$W_N^{22,\varepsilon}=\frac{2N}{N+1}\int_1^{\varepsilon N}\int_1^r\mathbb{E}[\Gamma_N(r/\varepsilon,s/\varepsilon)]\frac{\mathrm{d}s}{s}\frac{\mathrm{d}r}{r}=\frac{2N}{N+1}\int_{1/\varepsilon}^N\int_{1/\varepsilon}^r\mathbb{E}[\Gamma_N(r,s)]\frac{\mathrm{d}s}{s}\frac{\mathrm{d}r}{r}.$$

We observe that $|X_1-X_2|>(\frac{r}{N})^{1/d}+(\frac{s}{N})^{1/d}$ implies $B(X_1,(r/N)^{1/d})\cap B(X_2,(s/N)^{1/d})=\emptyset$ and thus $\mathbf{b}_{12}^N(r,s)=\mathbf{a}_1^N(r)+\mathbf{a}_2^N(s)$. Hence if $r>s>0$, we have $\Gamma_N(r,s)=\Gamma_N^1(r,s)+\Gamma_N^2(r,s)$, where

$$\Gamma_N^1(r,s)=\mathbf{1}_{\{|X_1-X_2|>(\frac{r}{N})^{1/d}+(\frac{s}{N})^{1/d}\}}(1-\mathbf{a}_1^N(r)-\mathbf{a}_2^N(s))^{N-1}-(1-\mathbf{a}_1^N(r))^N(1-\mathbf{a}_2^N(s))^N,$$
$$\Gamma_N^2(r,s)=\mathbf{1}_{\{(\frac{r}{N})^{1/d}<|X_1-X_2|\le(\frac{r}{N})^{1/d}+(\frac{s}{N})^{1/d}\}}(1-\mathbf{b}_{12}^N(r,s))^{N-1}-(1-\mathbf{a}_1^N(r))^N(1-\mathbf{a}_2^N(s))^N.$$

Using next that $(1-\mathbf{a}_1^N(r)-\mathbf{a}_2^N(s))\le(1-\mathbf{a}_1^N(r))(1-\mathbf{a}_2^N(s))$, we may write, if $r>s>0$,

$$\begin{aligned}\Gamma_N^1(r,s)\le&\mathbf{1}_{\{|X_1-X_2|>(\frac{r}{N})^{1/d}\}}(1-\mathbf{a}_1^N(r))^{N-1}(1-\mathbf{a}_2^N(s))^{N-1}-(1-\mathbf{a}_1^N(r))^N(1-\mathbf{a}_2^N(s))^N\\ \le&(1-\mathbf{a}_1^N(r))^{N-1}(1-\mathbf{a}_2^N(s))^{N-1}\Big[\mathbf{a}_1^N(r)+\mathbf{a}_2^N(s)-\mathbf{1}_{\{|X_1-X_2|\le(\frac{r}{N})^{1/d}\}}\Big]\\ =&\Gamma_N^{11}(r,s)+\Gamma_N^{12}(r,s)+\Gamma_N^{13}(r,s)+\Gamma_N^{14}(r,s),\end{aligned}$$

where

$$\begin{aligned}\Gamma_N^{11}(r,s)=&(1-\mathbf{a}_1^N(r))^{N-1}(1-\mathbf{a}_2^N(s))^{N-1}\mathbf{a}_2^N(s),\\ \Gamma_N^{12}(r,s)=&(1-\mathbf{a}_1^N(r))^{N-1}[(1-\mathbf{a}_2^N(s))^{N-1}-(1-\mathbf{a}_1^N(s))^{N-1}]\mathbf{a}_1^N(r),\\ \Gamma_N^{13}(r,s)=&(1-\mathbf{a}_1^N(r))^{N-1}(1-\mathbf{a}_1^N(s))^{N-1}[\mathbf{a}_1^N(r)-\mathbf{1}_{\{|X_1-X_2|\le(\frac{r}{N})^{1/d}\}}],\\ \Gamma_N^{14}(r,s)=&(1-\mathbf{a}_1^N(r))^{N-1}[(1-\mathbf{a}_1^N(s))^{N-1}-(1-\mathbf{a}_2^N(s))^{N-1}]\mathbf{1}_{\{|X_1-X_2|\le(\frac{r}{N})^{1/d}\}}.\end{aligned}$$

Step 1. We first study Γ_N^2. For $0<s<r<N$, since $\mathbf{b}_{12}^N(r,s)\ge\mathbf{a}_1^N(r)$,

$$\Gamma_N^2(r,s)\le\mathbf{1}_{\{(\frac{r}{N})^{1/d}<|X_1-X_2|\le(\frac{r}{N})^{1/d}+(\frac{s}{N})^{1/d}\}}(1-\mathbf{a}_1^N(r))^{N-1}\le\mathbf{1}_{\{X_2\in C_N(X_1,r,s)\}}e^{-\mathbf{m}_1r/2}.$$

We introduced the annulus $C_N(x,r,s)=B(x,(r/N)^{1/d}+(s/N)^{1/d})\setminus B(x,(r/N)^{1/d})$ and we used Lemma 11-(iii). Consequently,

$$\mathbb{E}[\Gamma_N^2(r,s)]\le\mathbb{E}\Big[F(C_N(X_1,r,s))e^{-\mathbf{m}_1r/2}\Big]\le\frac{v_d}{N}\mathbb{E}\Big[\mathbf{M}_1e^{-\mathbf{m}_1r/2}\Big]\Big((r^{1/d}+s^{1/d})^d-r\Big),$$

because $\sup_{C_N(X_1,r,s)}f\le\mathbf{M}_1$ (since $(r/N)^{1/d}+(s/N)^{1/d}\le2$) and because $\mathrm{Leb}(C_N(X_1,r,s))=(v_d/N)((r^{1/d}+s^{1/d})^d-r)$. But (recall that $r>s>0$)

$$(r^{1/d}+s^{1/d})^d-r\le ds^{1/d}(r^{1/d}+s^{1/d})^{d-1}\le ds^{1/d}(2r^{1/d})^{d-1}.$$

We have checked that $\mathbb{E}[\Gamma_N^2(r,s)]\le CN^{-1}\mathbb{E}[\mathbf{M}_1e^{-\mathbf{m}_1r/2}]s^{1/d}r^{(d-1)/d}$ for all $0<s<r<N$.

Step 2. Next, $\Pr(|X_1-X_2|\le(r/N)^{1/d}\,|\,X_1)=\mathbf{a}_1^N(r)$, so that $\mathbb{E}[\Gamma_N^{13}(r,s)]=0$.

Step 3. Lemma 11-(i)-(iii) gives us $\mathbb{E}[\Gamma_N^{11}(r,s)]\le v_dN^{-1}\mathbb{E}[\mathbf{M}_2\exp(-\mathbf{m}_1r/2-\mathbf{m}_2s/2)]s$.

Step 4. Since $(1-x)^{N-1}-(1-y)^{N-1}\le\mathbf{1}_{\{x\le y\}}(N-1)(1-x)^{N-2}(y-x)\le Ny$ for all $x,y\in[0,1]$,

$$\Gamma_N^{12}(r,s)\le N(1-\mathbf{a}_1^N(r))^{N-1}\mathbf{a}_1^N(s)\mathbf{a}_1^N(r)\le v_d(1-\mathbf{a}_1^N(r))^{N-1}\mathbf{a}_1^N(r)\mathbf{M}_1s.$$

by Lemma 11-(i). But since $N\ge2$,

$$(1-\mathbf{a}_1^N(r))^{N-1}\mathbf{a}_1^N(r)\le\mathbf{a}_1^N(r)e^{-N\mathbf{a}_1^N(r)/2}\le N^{-1}[\mathbf{m}_1re^{-\mathbf{m}_1r/2}+\mathbf{1}_{\{\mathbf{m}_1r<2\}}].$$

We finally used that $N\mathbf{a}_1^N(r)\ge\mathbf{m}_1r$ and that the function $f(x)=xe^{-x/2}$ is bounded by 1 and decreasing on $[2,\infty)$. Thus $\mathbb{E}[\Gamma_N^{12}(r,s)]\le v_dN^{-1}\mathbb{E}[(\mathbf{m}_1re^{-\mathbf{m}_1r/2}+\mathbf{1}_{\{\mathbf{m}_1r<2\}})\mathbf{M}_1]s$.

Step 5. Since $(1-x)^{N-1}-(1-y)^{N-1}\le Ny$ for all $x,y\in[0,1]$ as in Step 4,

$$\Gamma_N^{14}(r,s)\le N(1-\mathbf{a}_1^N(r))^{N-1}\mathbf{a}_2^N(s)\mathbf{1}_{\{|X_1-X_2|\le(\frac{r}{N})^{1/d}\}}.$$

But $|X_1-X_2|\le(r/N)^{1/d}\le1$ implies that $\sup_{B(X_2,(s/N)^{1/d})}f\le\sup_{B(X_1,2)}f\le\mathbf{M}_1$, whence $\mathbf{a}_2^N(s)\le\mathbf{M}_1\mathrm{Leb}(B(X_2,(s/N)^{1/d}))=v_dN^{-1}\mathbf{M}_1s$. Consequently,

$$\Gamma_N^{14}(r,s)\le v_d(1-\mathbf{a}_1^N(r))^{N-1}\mathbf{M}_1s\mathbf{1}_{\{|X_1-X_2|\le(r/N)^{1/d}\}}.$$

Taking first the expectation knowing X_1, we find $\mathbb{E}[\Gamma_N^{14}(r,s)]\le v_d\mathbb{E}[(1-\mathbf{a}_1^N(r))^{N-1}\mathbf{a}_1^N(r)\mathbf{M}_1]s$. We thus conclude exactly as in Step 4 that $\mathbb{E}[\Gamma_N^{14}(r,s)]\le v_dN^{-1}\mathbb{E}[(\mathbf{m}_1re^{-\mathbf{m}_1r/2}+\mathbf{1}_{\{\mathbf{m}_1r<2\}})\mathbf{M}_1]s$.

Step 6. Gathering the bounds found in the five first steps, we see that $\mathbb{E}[\Gamma_N(r,s)] \le CN^{-1}G(r,s)$ for all $0<s<r<N$, where

$$G(r,s) = \mathbb{E}\Big[\mathbf{M}_1 e^{-\mathbf{m}_1 r/2} s^{1/d} r^{(d-1)/d} + \mathbf{M}_2 e^{-\mathbf{m}_1 r/2 - \mathbf{m}_2 s/2} s + (\mathbf{m}_1 r e^{-\mathbf{m}_1 r/2} + \mathbf{1}_{\{\mathbf{m}_1 r<2\}})\mathbf{M}_1 s\Big].$$

Thus for all $\varepsilon \in (0,1]$, all $N \ge (1/\varepsilon)\vee 2$,

$$N(W_N^{12,\varepsilon} + W_N^{22,\varepsilon}) \le C\int_0^\infty \int_0^r \frac{G(r,s)}{rs}\mathbf{1}_{\{r\le\varepsilon \text{ or } r\ge 1/\varepsilon\}}\mathrm{d}s\mathrm{d}r.$$

If we show that $\int_0^\infty \int_0^r G(r,s)\frac{\mathrm{d}s}{s}\frac{\mathrm{d}r}{r}<\infty$, this will imply that $\sup_{N\ge 2} N(W_N^{12,1}+W_N^{22,1})<\infty$ and also that $\limsup_{\varepsilon\to 0}\sup_{N\ge 1/\varepsilon} N(W_N^{12,\varepsilon}+W_N^{22,\varepsilon}) \le 0$ by dominated convergence. First,

$$\begin{aligned}\int_0^r \frac{G(r,s)}{rs}\mathrm{d}s \le & C\mathbb{E}\Big[\mathbf{M}_1 e^{-\mathbf{m}_1 r/2} + \frac{\mathbf{M}_2}{\mathbf{m}_2 r} e^{-\mathbf{m}_1 r/2}(1-e^{-\mathbf{m}_2 r/2}) + \mathbf{M}_1\mathbf{m}_1 r e^{-\mathbf{m}_1 r/2} + \mathbf{M}_1 \mathbf{1}_{\{\mathbf{m}_1 r<2\}}\Big]\\ \le & C\mathbb{E}\Big[\mathbf{M}_1 e^{-\mathbf{m}_1 r/4} + \frac{\mathbf{M}_2}{\mathbf{m}_2 r}e^{-\mathbf{m}_1 r/2}(1-e^{-\mathbf{m}_2 r/2})\Big].\end{aligned}$$

For the last inequality, we used that $e^{-x/2}+xe^{-x/2}+\mathbf{1}_{\{x<2\}} \le Ce^{-x/4}$ for all $x\ge 0$. Since now $\int_0^\infty e^{-ar/2}(1-e^{-br/2})\frac{\mathrm{d}r}{r} = \log(1+b/a)$ for all $a>0$, $b\ge 0$, which can be checked by differentiating both side in b and by using the value at $b=0$, we conclude that

$$\int_0^\infty\int_0^r \frac{G(r,s)}{rs}\mathrm{d}s\mathrm{d}r \le C\mathbb{E}\Big[\frac{\mathbf{M}_1}{\mathbf{m}_1} + \frac{\mathbf{M}_2}{\mathbf{m}_2}\log\Big(1+\frac{\mathbf{m}_2}{\mathbf{m}_1}\Big)\Big] \le C\mathbb{E}\Big[\frac{\mathbf{M}_1}{\mathbf{m}_1} + \frac{\mathbf{M}_2}{\mathbf{m}_2}(1+|\log \mathbf{m}_1| + |\log \mathbf{m}_2|)\Big].$$

This last quantity is finite by assumption and because $(\mathbf{m}_1,\mathbf{M}_1)$ and $(\mathbf{m}_2,\mathbf{M}_2)$ are independent and have the same law. □

Finally, we give the

Proof of Theorem 1. We assume that f is continuous and bounded and (5) for some $q>0$ and some $\theta\in(0,1]$ (and with $r_0=1$). This implies that $\int_{\mathbb{R}^d} f(x)\log^2 f(x)\mathrm{d}x<\infty$, because $m\le v_d f\le C$, so that $\log^2 f \le C(1+\log^2 m)$. We thus can apply Proposition 13 and all the lemmas of the section.

We first check (ii) and thus assume that $\theta\in(0,1/2]$. We then use Lemma 12 with $\varepsilon=1$ and write $H_N = H_N^1 + K_N^{1,1} + K_N^{2,1} + K_N^3$. But $H_N^1 = 0$ a.s., whence $\mathbb{V}\mathrm{ar}\, H_N \le 3\mathbb{V}\mathrm{ar}\, K_N^{1,1} + 3\mathbb{V}\mathrm{ar}\, K_N^{2,1} + 3\mathbb{V}\mathrm{ar}\, K_N^3 \le 3W_N^{11,1} + 3W_N^{12,1} + 3W_N^{21,1} + 3W_N^{22,1} + 3W_N^{31} + 3W_N^{32}$ by Lemma 16. By Lemmas 17, 18, 19, 20 and 21, we find that $\mathbb{V}\mathrm{ar}\, H_N \le CN^{-1} + CN^{-1-\theta} + CN^{-2\theta}$ for all $N\ge 2$, whence $\mathbb{V}\mathrm{ar}\, H_N \le CN^{-2\theta}$. The case $N=1$ is of course not an issue.

We next prove (i) and thus assume that $\theta\in(1/2,1]$. For each $\varepsilon\in(0,1]$ and $N\ge 1/\varepsilon$, we write $H_N = H_N^\varepsilon + K_N^{1,\varepsilon} + K_N^{2,\varepsilon} + K_N^3$ as in Lemma 12. We then infer from Proposition 13 that for each $\varepsilon\in(0,1]$, $\sqrt{N}(H_N^\varepsilon - \mathbb{E}[H_N^\varepsilon])$ goes in law to $\mathcal{N}(0,\sigma_\varepsilon^2(f))$ as $N\to\infty$. We also know that $\lim_{\varepsilon\to 0}\sigma_\varepsilon^2(f) = \sigma^2(f)$ by Lemma 15. To conclude that $\sqrt{N}(H_N - \mathbb{E}[H_N])$ goes in law to $\mathcal{N}(0,\sigma^2(f))$, it only remains to verify that $\lim_{\varepsilon\to 0}\sup_{N\ge 1/\varepsilon} N\mathbb{V}\mathrm{ar}\,(K_N^{1,\varepsilon}+K_N^{2,\varepsilon}+K_N^3)=0$. But we have $\mathbb{V}\mathrm{ar}\,(K_N^{1,\varepsilon}+K_N^{2,\varepsilon}+K_N^3) \le 3(W_N^{11,\varepsilon}+W_N^{12,\varepsilon}+W_N^{21,\varepsilon}+W_N^{22,\varepsilon}+W_N^{31}+W_N^{32})$, see Lemma 16. By Lemmas 17, 18 and 21, it holds that $\limsup_{\varepsilon\to 0}\sup_{N\ge 1/\varepsilon} N(W_N^{11,\varepsilon}+W_N^{12,\varepsilon}+W_N^{21,\varepsilon}+W_N^{22,\varepsilon}) \le 0$. We then infer from Lemma 19 that $\sup_{N\ge 1/\varepsilon} NW_N^{31} \le C\varepsilon^\theta$ and from Lemma 20 that $\sup_{N\ge 1/\varepsilon} NW_N^{32} \le C\varepsilon^{2\theta-1}$. Both tend to 0 as $\varepsilon\to 0$ because $\theta\in(1/2,1]$. All in all, $\limsup_{\varepsilon\to 0}\sup_{N\ge 1/\varepsilon} N\mathbb{V}\mathrm{ar}\,(K_N^{1,\varepsilon}+K_N^{2,\varepsilon}+K_N^3)\le 0$ as desired. □

4. Estimation of the variance

The goal of this section is to prove Proposition 6. We have already done most of the work.

Proof of Proposition 6. We assume that f is bounded and continuous and satisfies (8) for some $q > 0$ and some $\theta > 0$ (with $r_0 = 1$). We write $V_N = A_N - B_N^2 + \chi_d - \pi^2/6$, where

$$A_N = \frac{1}{N+1}\sum_{i=1}^{N+1} \log^2 Y_i^N \quad \text{and} \quad B_N = \frac{1}{N+1}\sum_{i=1}^{N+1} \log Y_i^N.$$

For $\varepsilon \in (0,1]$, we recall that $\log_\varepsilon y = \log(\varepsilon \vee y \wedge (1/\varepsilon))$ and define, for $N \geq 1/\varepsilon$,

$$A_N^\varepsilon = \frac{1}{N+1}\sum_{i=1}^{N+1} \log_\varepsilon^2 Y_i^N \quad \text{and} \quad B_N^\varepsilon = \frac{1}{N+1}\sum_{i=1}^{N+1} \log_\varepsilon Y_i^N.$$

Step 1. For each fixed $\varepsilon \in (0,1]$, since $f : \mathbb{R}^d \mapsto \mathbb{R}_+$ and $\log_\varepsilon : [0,\infty) \mapsto \mathbb{R}$ are bounded and continuous, we infer from [4, Theorem 3.5] that $\sup_{N\geq 1/\varepsilon} N(\mathbb{V}\mathrm{ar}\,(A_N^\varepsilon) + \mathbb{V}\mathrm{ar}\,(B_N^\varepsilon)) < \infty$.

Step 2. Here we show that Y_1^N goes in law to $\xi/(v_d\mathbf{f}_1)$ as $N \to \infty$, where $\xi \sim \mathrm{Exp}(1)$ is independent of X_1. For each $r > 0$, we have $\Pr(Y_1^N > r \,|\, X_1) = (1 - \mathbf{a}_1^N)^N = (1 - F(B(X_1, (r/N)^{1/d})))^N$ by Lemma 10. Since f is continuous, $NF(B(x,(r/N)^{1/d})) \to v_d f(x) r$ as $N \to \infty$ for all $x \in \mathbb{R}^d$. Thus $\Pr(Y_1^N > r \,|\, X_1)$ a.s. tends to $\exp(-v_d\mathbf{f}_1 r)$ as $N \to \infty$. By dominated convergence, since $\Pr(Y_1^N > r \,|\, X_1)$ is bounded by 1, we conclude that for each $r > 0$, $\lim_{N\to\infty} \Pr(Y_1^N > r) = \lim_{N\to\infty} \mathbb{E}[\Pr(Y_1^N > r \,|\, X_1)] = \mathbb{E}[\exp(-v_d\mathbf{f}_1 r)]$, which equals $\Pr(\xi/(v_d\mathbf{f}_1) > r)$.

Step 3. For each fixed $\varepsilon \in (0,1]$, since $\log_\varepsilon : [0,\infty) \mapsto \mathbb{R}$ is bounded and continuous, we deduce from Step 2 that $\lim_{N\to\infty} \mathbb{E}[A_N^\varepsilon] = \lim_{N\to\infty} \mathbb{E}[\log_\varepsilon^2 Y_1^N] = \mathbb{E}[\log_\varepsilon^2(\xi/(v_d\mathbf{f}_1))]$ and that $\lim_{N\to\infty} \mathbb{E}[B_N^\varepsilon] = \lim_{N\to\infty} \mathbb{E}[\log_\varepsilon Y_1^N] = \mathbb{E}[\log_\varepsilon(\xi/(v_d\mathbf{f}_1))]$.

Step 4. As seen in the first paragraph of the proof of Lemma 15, $\mathbb{E}[\log^2(\xi/(v_d\mathbf{f}_1))] < \infty$ (we have $\int_{\mathbb{R}^d} f(x)\log^2 f(x)\mathrm{d}x < \infty$ because $\int_{\mathbb{R}^d} f(x)\log^2 m(x)\mathrm{d}x < \infty$ and $m \leq v_d f \leq C$ by assumption). Thus $\lim_{\varepsilon\to 0} \mathbb{E}[\log_\varepsilon^2(\xi/(v_d\mathbf{f}_1))] = \mathbb{E}[\log^2(\xi/(v_d\mathbf{f}_1))]$ and $\lim_{\varepsilon\to 0} \mathbb{E}[\log_\varepsilon(\xi/(v_d\mathbf{f}_1))] = \mathbb{E}[\log(\xi/(v_d\mathbf{f}_1))]$ by dominated convergence (because $|\log_\varepsilon x| \leq |\log x|$ for all $\varepsilon \in (0,1)$ and all $x > 0$ and because $\lim_{\varepsilon\to 0}\log_\varepsilon x = \log x$ for all $x > 0$).

Step 5. Here we verify that $\lim_{\varepsilon\to 0}\sup_{N\geq 1/\varepsilon} \mathbb{E}[|A_N^\varepsilon - A_N|] = 0$. It is checked similarly that $\lim_{\varepsilon\to 0}\sup_{N\geq 1/\varepsilon} \mathbb{E}[|B_N^\varepsilon - B_N|] = 0$. We recall that for all $\varepsilon > 0$, all $N \geq 1/\varepsilon$, all $y \in (0,\infty)$, $\log y = \log_\varepsilon y + \log[(y/\varepsilon)\wedge 1] + \log[1 \vee (\varepsilon y) \wedge (\varepsilon N)] + \log[(y/N)\vee 1]$ and (same formula with $\varepsilon = 1$) $\log y = \log[y \wedge 1] + \log[1 \vee y \wedge N] + \log[(y/N) \vee 1]$, see the proof of Lemma 12. Starting from $|\log^2 y - \log_\varepsilon^2 y| \leq 2|\log y||\log y - \log_\varepsilon y|$, we end with

$$\begin{aligned}|\log^2 y - \log_\varepsilon^2 y| \leq & 2\Big|\log[y\wedge 1] + \log[1\vee y \wedge N] + \log[(y/N)\vee 1]\Big| \\ & \times \Big|\log[(y/\varepsilon)\wedge 1] + \log[1\vee(\varepsilon y)\wedge(\varepsilon N)] + \log[(y/N)\vee 1]\Big|.\end{aligned}$$

Hence, by the Cauchy-Schwarz inequality, for $N \geq 1/\varepsilon$,

$$\begin{aligned}\mathbb{E}[|A_N^\varepsilon - A_N|] \leq & \mathbb{E}[|\log^2 Y_1^N - \log_\varepsilon^2 Y_1^N|] \\ \leq & C\mathbb{E}\Big[\log^2[Y_1^N \wedge 1] + \log^2[1\vee Y_1^N \wedge N] + \log^2[(Y_1^N/N)\vee 1]\Big]^{1/2} \\ & \times \mathbb{E}\Big[\log^2[(Y_1^N/\varepsilon)\wedge 1] + \log^2[1\vee(\varepsilon Y_1^N)\wedge(\varepsilon N)] + \log^2[(Y_1^N/N)\vee 1]\Big]^{1/2}.\end{aligned}$$

With the notation of Lemma 12 (see also its proof), this precisely rewrites

$$\mathbb{E}[|A_N^\varepsilon - A_N|] \le C\Big((N+1)W_N^{11,1} + (N+1)W_N^{21,1} + (N+1)W_N^{31}\Big)^{1/2} \times \Big((N+1)W_N^{11,\varepsilon} + (N+1)W_N^{21,\varepsilon} + (N+1)W_N^{31}\Big)^{1/2}$$

We then deduce from Lemmas 17, 18 and 19 that $\lim_{\varepsilon\to 0}\sup_{N\ge 1/\varepsilon}\mathbb{E}[|A_N^\varepsilon - A_N|] = 0$. We can apply these three lemmas thanks to (8) (with $r_0 = 1$) and because f is bounded.

Step 6. Here we conclude that $\lim_{N\to\infty}\mathbb{E}[|A_N - \mathbb{E}[\log^2(\xi/(v_d\mathbf{f}_1))]|] = 0$ as $N\to\infty$. It is checked similarly that $\lim_{N\to\infty}\mathbb{E}[|B_N - \mathbb{E}[\log(\xi/(v_d\mathbf{f}_1))]|] = 0$. We fix $\varepsilon\in(0,1]$ and write

$$\mathbb{E}[|A_N - \mathbb{E}[\log^2(\xi/(v_d\mathbf{f}_1))]|] \le \mathbb{E}[|A_N - A_N^\varepsilon|] + \mathbb{E}[|A_N^\varepsilon - \mathbb{E}[\log_\varepsilon^2(\xi/(v_d\mathbf{f}_1))]|] + |\mathbb{E}[\log_\varepsilon^2(\xi/(v_d\mathbf{f}_1))] - \mathbb{E}[\log^2(\xi/(v_d\mathbf{f}_1))]|.$$

Taking first the limsup as $N\to\infty$ (so that the middle term of the RHS disappears by Steps 1 and 3) and then the limsup as $\varepsilon\to 0$ (using Steps 4 and 5) completes the step.

Step 7. By Step 6, $V_N = A_N - B_N^2 + \chi_d - \pi^2/6$ goes to $\Sigma = \mathbb{V}\mathrm{ar}\,(\log(\xi/(v_d\mathbf{f}_1))) + \chi_d - \pi^2/6$ in probability. But $\mathbb{V}\mathrm{ar}\,(\log(\xi/(v_d\mathbf{f}_1))) = \int_{\mathbb{R}^d} f(x)\log^2 f(x)\mathrm{d}x - (H(f))^2 + \pi^2/6$, see the first paragraph of the proof of Lemma 15. Thus $\Sigma = \int_{\mathbb{R}^d} f(x)\log^2 f(x)\mathrm{d}x - (H(f))^2 + \chi_d = \sigma^2(f)$. □

5. A tedious Taylor approximation

Here we study in detail how well $(1 - F(B(x,(r/N)^{1/d}))^N$ approximates $\exp(-v_d f(x) r)$.

Lemma 22. *Let $\beta > 0$, set $\rho = \min\{\beta, 2\}$, $k = \max\{i\in\mathbb{N} : i < \beta\}$ and $\ell = \max\{i\in\mathbb{N} : 2i < \beta\}$. Assume that $\kappa = \sup_{x\in\mathbb{R}^d} F(B(x,1)) < 1$, that $f\in\mathcal{D}^\beta(\mathbb{R}^d)$ and recall that M and G_β were defined in (3) and (4). For each $N\ge 1$, consider $h_N : \mathbb{R}^d \mapsto (0,\infty)$ such that for all $x\in\mathbb{R}^d$,*

$$h_N(x) \le \min\Big\{N, \frac{N^{\rho/(d+\rho)}}{(G_\beta(x))^{d/(d+\rho)}}, \frac{\sqrt{N}}{M(x)}\Big\}. \tag{16}$$

(i) If $\beta\in(0,2]$, for all $N\ge 1$, all $x\in\mathbb{R}^d$, all $r\in[0,h_N(x)]$,

$$[1 - F(B(x,(r/N)^{1/d}))]^N = e^{-v_d f(x) r}(1 + R_N(x,r)),$$

where R_N satisfies, for some constant $C>0$, for all $N\ge 1$, all $x\in\mathbb{R}^d$, all $r\in[0,h_N(x)]$,

$$|R_N(x,r)| \le C\Big(\frac{r^2M^2(x)}{N} + \Big(\frac{r}{N}\Big)^{\beta/d} rG_\beta(x)\Big).$$

(ii) If $\beta > 2$, for all $N\ge 1$, all $x\in\mathbb{R}^d$, all $r\in[0,h_N(x)]$,

$$[1 - F(B(x,(r/N)^{1/d}))]^N = e^{-v_d f(x) r}\Big(1 + \sum_{i=1}^{\ell}\frac{g_i(x,r)}{N^{2i/d}} + S_N(x,r)\Big),$$

where S_N satisfies, for some constant $C>0$, for all $N\ge 1$, all $x\in\mathbb{R}^d$, all $r\in[0,h_N(x)]$,

$$|S_N(x,r)| \le C\frac{r^2M^2(x)}{N} + C\Big(\frac{r}{N}\Big)^{\beta/d}[rG_\beta(x) + (rG_\beta(x))^{\beta/2}]$$

and where the functions $g_1,\dots,g_\ell : \mathbb{R}^d\times[0,\infty)\mapsto\mathbb{R}$ (not depending on N) satisfy, for some constant $C>0$, for all $x\in\mathbb{R}^d$ and all $r\ge 0$,

$$|g_i(x,r)| \le Cr^{2i/d}[rG_\beta(x) + (rG_\beta(x))^i].$$

Proof. Step 1. For $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d$ a multi-index, we use the standard notation $|\alpha| = \sum_{i=1}^d \alpha_i$, $\alpha! = \alpha_1! \dots \alpha_d!$ and $h^\alpha = h_1^{\alpha_1} \dots h_d^{\alpha_d}$ for $h \in \mathbb{R}^d$. Using the Taylor formula, we write, for $y \in B(x,1)$, with the convention that $\sum_1^0 = 0$ if $k = 0$ (i.e. $\beta \in (0,1]$),

$$f(y) = f(x) + \sum_{|\alpha|=1}^{k} \frac{1}{\alpha!} \partial_\alpha f(x)(y-x)^\alpha + \Delta_\beta(x,y),$$

with $|\Delta_\beta(x,y)| \le C|x-y|^\beta G_\beta(x)$. The function G_β was precisely designed for that purpose.

For $\varepsilon \in (0,1]$, we integrate the above equality on $B(x,\varepsilon)$:

$$F(B(x,\varepsilon)) = v_d \varepsilon^d f(x) + \sum_{|\alpha|=1}^{k} \frac{1}{\alpha!} \partial_\alpha f(x) \int_{B(x,\varepsilon)} (y-x)^\alpha \mathrm{d}y + \int_{B(x,\varepsilon)} \Delta_\beta(x,y) \mathrm{d}y.$$

But one easily checks that $\int_{B(x,\varepsilon)} (y-x)^\alpha \mathrm{d}y = 0$ if $|\alpha|$ is odd, while $\int_{B(x,\varepsilon)} (y-x)^\alpha \mathrm{d}y = c_\alpha \varepsilon^{|\alpha|+d}$ when $|\alpha|$ is even, with $c_\alpha = \int_{B(0,1)} y^\alpha \mathrm{d}y$. We thus may write, using the convention that $\sum_1^0 = 0$ when $\ell = 0$ (i.e. $\beta \in (0,2]$),

$$F(B(x,\varepsilon)) = v_d \varepsilon^d f(x) + \sum_{j=1}^{\ell} \varepsilon^{2j+d} \delta_{2j} f(x) + \varepsilon^{d+\beta} \Gamma_\beta(x,\varepsilon),$$

with $|\Gamma_\beta(x,\varepsilon)| = \varepsilon^{-d-\beta} |\int_{B(x,\varepsilon)} \Delta_\beta(x,y) \mathrm{d}y| \le C G_\beta(x)$ and $\delta_{2j} f(x) = \sum_{|\alpha|=2j} \frac{c_\alpha}{\alpha!} \partial_\alpha f(x)$, which satisfies $|\delta_{2j} f(x)| \le C G_\beta(x)$ (because $2j \le 2\ell < \beta$).

We conclude that for all $r \in [0,N]$, all $x \in \mathbb{R}^d$,

$$NF(B(x,(r/N)^{1/d})) = v_d f(x) r + r \sum_{j=1}^{\ell} \Big(\frac{r}{N}\Big)^{2j/d} \delta_{2j} f(x) + r \Big(\frac{r}{N}\Big)^{\beta/d} G_\beta(x) \theta_{N,1}(x,r), \tag{17}$$

for some uniformly bounded $\theta_{N,1}$, that is, $\sup_{N\ge1} \sup_{x\in\mathbb{R}^d} \sup_{r\in[0,N]} |\theta_{N,1}(x,r)| < \infty$. Also, one easily verifies that there is a constant $C > 0$ such that for all $N \ge 1$, $x \in \mathbb{R}^d$, $r \in [0, h_N(x)]$,

$$r \sum_{j=1}^{\ell} \Big(\frac{r}{N}\Big)^{2j/d} |\delta_{2j} f(x)| + r \Big(\frac{r}{N}\Big)^{\beta/d} G_\beta(x) \le C. \tag{18}$$

It suffices to use that $|\delta_{2j} f(x)| \le C G_\beta(x)$, that $(r/N)^{2j/d} + (r/N)^{\beta/d} \le 2(r/N)^{\rho/d}$ for all $j = 1, \dots, \ell$ (use that $h_N(x) \le N$ whence $r/N \le 1$ and recall that $\rho = \min\{\beta, 2\}$) and that $r(r/N)^{\rho/d} G_\beta(x) \le 1$ because $r \le h_N(x) \le N^{\rho/(d+\rho)}/(G_\beta(x))^{d/(d+\rho)}$.

Step 2. For $0 \le r \le N$, we have $F(B(x,(r/N)^{1/d})) \le \kappa < 1$. Hence

$$\log[1 - F(B(x,(r/N)^{1/d}))] = -F(B(x,(r/N)^{1/d})) + [F(B(x,(r/N)^{1/d}))]^2 \theta_{N,2}(x,r),$$

for some uniformly bounded function $\theta_{N,2}$. And we have $[F(B(x,(r/N)^{1/d}))]^2 \le v_d^2 N^{-2} r^2 M^2(x)$ by definition of M. As a consequence, for some new uniformly bounded function $\theta_{N,3}$,

$$[1 - F(B(x,(r/N)^{1/d}))]^N = \exp\Big(- NF(B(x,(r/N)^{1/d})) + N^{-1} r^2 M^2(x) \theta_{N,3}(x,r) \Big). \tag{19}$$

Now since $h_N(x) \le \sqrt{N}/M(x)$, it holds that for all $N \ge 1$, $x \in \mathbb{R}^d$, $r \in [0, h_N(x)]$,

$$N^{-1} r^2 M^2(x) \le 1. \tag{20}$$

Combining (17) and (19) gives us, for $x \in \mathbb{R}^d$ and $r \in [0, h_N(x)]$,

$$[1 - F(B(x, (r/N)^{1/d}))]^N = e^{-v_d f(x) r} \exp\Big(I_N(x,r) + J_N(x,r)\Big),$$

where

$$I_N(x,r) = -\, r \sum_{j=1}^{\ell} \Big(\frac{r}{N}\Big)^{2j/d} \delta_{2j} f(x),$$

$$J_N(x,r) = -\, r\Big(\frac{r}{N}\Big)^{\beta/d} G_\beta(x)\theta_{N,1}(x,r) + \frac{r^2 M^2(x)}{N}\theta_{N,3}(x,r).$$

Using that I_N and J_N are uniformly bounded (for $r \in [0, h_N(x)]$) by (18) and (20), we may write

$$[1 - F(B(x, (r/N)^{1/d}))]^N = e^{-v_d f(x) r}\Big(1 + \sum_{i=1}^{\ell} \frac{1}{i!}(I_N(x,r))^i + R_N(x,r)\Big), \tag{21}$$

with (since $\ell + 1 \geq \beta/2$ by definition of ℓ)

$$|R_N(x,r)| \leq C(|I_N(x,r)|^{\ell+1} + |J_N(x,r)|) \leq C(|I_N(x,r)|^{\beta/2} + |J_N(x,r)|).$$

Step 3. If $\beta \in (0, 2]$, then $\ell = 0$ and thus $I_N = 0$. Using (21) and noting that

$$|R_N(x,r)| \leq C|J_N(x,r)| \leq Cr\Big(\frac{r}{N}\Big)^{\beta/d} G_\beta(x) + C\frac{r^2M^2(x)}{N}$$

for all $x \in \mathbb{R}^d$ and all $r \in [0, h_N(x)]$ completes the proof of (i).

Step 4. We now suppose that $\beta > 2$, whence $\ell \geq 1$ and $\rho = 2$. First, we have $|I_N(x,r)| \leq Cr(r/N)^{2/d}G_\beta(x)$ for all $r \in [0, h_N(x)]$, because $|\delta_{2j} f(x)| \leq CG_\beta(x)$ and because $r/N \leq 1$. Thus

$$|R_N(x,r)| \leq C\Big(\frac{r}{N}\Big)^{\beta/d}(rG_\beta(x))^{\beta/2} + C\Big(\frac{r}{N}\Big)^{\beta/d} rG_\beta(x) + C\frac{r^2M^2(x)}{N}. \tag{22}$$

Using next the multinomial theorem, we find (here $i_1, \dots, i_\ell$ are non-negative integers)

$$\begin{aligned}\sum_{i=1}^{\ell} \frac{1}{i!}(I_N(x,r))^i =& \sum_{i=1}^{\ell}(-r)^i \sum_{i_1+\dots+i_\ell = i} \Big(\frac{r}{N}\Big)^{2(i_1+2i_2+\dots+\ell i_\ell)/d} \frac{(\delta_2 f(x))^{i_1}(\delta_4 f(x))^{i_2}\cdots(\delta_{2\ell} f(x))^{i_\ell}}{i_1!\cdots i_\ell!}\\ =& \sum_{m=1}^{\ell} \frac{g_m(x,r)}{N^{2m/d}} + T_N(x,r),\end{aligned}$$

where

$$g_m(x,r) = r^{2m/d} \sum_{i_1+2i_2+\dots+\ell i_\ell = m} (-r)^{i_1+\dots+i_\ell} \frac{(\delta_2 f(x))^{i_1}(\delta_4 f(x))^{i_2}\cdots(\delta_{2\ell} f(x))^{i_\ell}}{i_1!\cdots i_\ell!}$$

is well-defined on $\mathbb{R}^d \times [0,\infty)$ and where

$$T_N(x,r) = \sum_{i=1}^{\ell}(-r)^i \sum_{\substack{i_1+\dots+i_\ell = i\\ i_1+2i_2+\dots+\ell i_\ell > \ell}} \Big(\frac{r}{N}\Big)^{2(i_1+2i_2+\dots+\ell i_\ell)/d} \frac{(\delta_2 f(x))^{i_1}(\delta_4 f(x))^{i_2}\cdots(\delta_{2\ell} f(x))^{i_\ell}}{i_1!\cdots i_\ell!}.$$

Recalling (21), we have checked that for all $x \in \mathbb{R}^d$, all $r \in [0, h_N(x)]$,

$$(1-F(B(x,(r/N)^{1/d}))^N = e^{-v_d f(x) r}\Big(1+\sum_{m=1}^{\ell}\frac{g_m(x,r)}{N^{2m/d}} + S_N(x,r)\Big),$$

with $S_N(x,r) = T_N(x,r) + R_N(x,r)$.

Since $|\delta_{2j} f(x)| \le C G_\beta(x)$, we deduce that for all $m = 1, \dots, \ell$, all $x \in \mathbb{R}^d$, all $r \ge 0$,

$$|g_m(x,r)| \le C r^{2m/d} \sum_{i_1+2i_2+\cdots+\ell i_\ell = m} (rG_\beta(x))^{i_1+\cdots+i_\ell} \le C r^{2m/d}[rG_\beta(x) + (rG_\beta(x))^m],$$

because $i_1 + 2i_2 + \cdots + \ell i_\ell = m$ implies $1 \le i_1 + \cdots + i_\ell \le m$.

Similarly, using that $r \le N$,

$$|T_N(x,r)| \le C\Big(\frac{r}{N}\Big)^{2(\ell+1)/d}[rG_\beta(x) + (rG_\beta(x))^\ell] \le C\Big(\frac{r}{N}\Big)^{\beta/d}[rG_\beta(x) + (rG_\beta(x))^{\beta/2}]$$

because $2(\ell+1) \ge \beta$ and $1 \le \ell < \beta/2$. Recalling (22), we find that

$$|S_N(x,r)| \le C\Big(\frac{r}{N}\Big)^{\beta/d}[rG_\beta(x) + (rG_\beta(x))^{\beta/2}] + C\frac{r^2 M^2(x)}{N}$$

as desired. □

6. Bias

The whole section is devoted to the proof of Theorems 3 and 4. We first provide an integral expression of the bias.

Lemma 23. *We have*

$$\mathbb{E}[H_N] = H(f) + \mathbb{E}\Big[\int_0^\infty \Big((1-\mathbf{a}_1^N(r))^N - e^{-v_d \mathbf{f}_1 r}\Big)\frac{\mathrm{d}r}{r}\Big].$$

Proof. Since $\log z = \int_0^\infty (\mathbf{1}_{\{r<z\}} - \mathbf{1}_{\{r<1\}})\frac{\mathrm{d}r}{r}$ for $z > 0$, for any positive random variable Z,

$$\mathbb{E}[\log Z] = \int_0^\infty [\Pr(Z > r) - \mathbf{1}_{\{r<1\}}]\frac{\mathrm{d}r}{r}.$$

Recalling that $\mathbb{E}[H_N] = \gamma + \log v_d + \mathbb{E}[\log Y_1^N]$, see (2), we deduce from Lemma 10 that

$$\mathbb{E}[H_N] = \gamma + \log v_d + \mathbb{E}\Big[\int_0^\infty [(1-\mathbf{a}_1^N(r))^N - \mathbf{1}_{\{r<1\}}]\frac{\mathrm{d}r}{r}\Big]. \tag{23}$$

Also, for $\xi \sim \text{Exp}(1)$ independent of X_1, so that $\Pr(\xi/(v_d\mathbf{f}_1) > r \,|\, X_1) = \exp(-v_d \mathbf{f}_1 r)$,

$$\mathbb{E}\Big[\log\frac{\xi}{v_d\mathbf{f}_1}\Big] = \mathbb{E}\Big[\int_0^\infty [e^{-v_d\mathbf{f}_1 r} - \mathbf{1}_{\{r<1\}}]\frac{\mathrm{d}r}{r}\Big].$$

But $\mathbb{E}[\log(\xi/(v_d\mathbf{f}_1))] = \mathbb{E}[\log \xi] - \log v_d - \mathbb{E}[\log \mathbf{f}_1] = -\gamma - \log v_d + H(f)$, so that we have $H(f) = \gamma + \log v_d + \mathbb{E}[\int_0^\infty [e^{-v_d\mathbf{f}_1 r} - \mathbf{1}_{\{r<1\}}]\frac{\mathrm{d}r}{r}]$. Recalling (23), one easily concludes. □

From now on, we fix $\beta > 0$ and we set $\rho = \min\{\beta, 2\}$ and $\ell = \max\{i \in \mathbb{N} \,:\, 2i < \beta\}$. We assume that $f \in \mathcal{D}^\beta(\mathbb{R}^d)$ and recall that M and G_β were defined in (3) and (4). We assume that $\kappa = \sup_{\mathbb{R}^d} F(B(x,1)) < 1$. We put $R = M + G_\beta$ and introduce, for each $N \ge 1$, the function $h_N : \mathbb{R}^d \mapsto [0, N]$ defined by

$$h_N(x) = \min\Big\{\frac{2\log N}{m(x)}, N, \frac{N^{\rho/(d+\rho)}}{(R(x))^{d/(d+\rho)}}, \frac{\sqrt{N}}{R(x)}\Big\},$$

which of course satisfies (16). We also introduce the shortened notation

$$\mathbf{R}_1 = R(X_1) \quad \text{and} \quad \mathbf{h}_{N1} = h_N(X_1).$$

We observe that $\Pr(\cup_{i=1}^4 \Omega_N^i) = 1$, where

$$\Omega_N^1 = \Big\{\mathbf{h}_{N1} = \frac{2\log N}{\mathbf{m}_1}\Big\}, \quad \Omega_N^2 = \Big\{\mathbf{h}_{N1} = N\Big\}, \quad \Omega_N^3 = \Big\{\mathbf{h}_{N1} = \frac{N^{\rho/(d+\rho)}}{\mathbf{R}_1^{d/(d+\rho)}}\Big\}, \quad \Omega_N^4 = \Big\{\mathbf{h}_{N1} = \frac{\sqrt{N}}{\mathbf{R}_1}\Big\}.$$

We infer from Lemma 23 that $\mathbb{E}[H_N] - H(f) = B_1^N + B_2^N + B_3^N + B_4^N$, where

$$B_1^N = \mathbb{E}\Big[\int_0^{\mathbf{h}_{N1}} \Big((1-\mathbf{a}_1^N(r))^N - e^{-v_d \mathbf{f}_1 r}\Big)\frac{\mathrm{d}r}{r}\Big],$$

$$B_2^N = -\mathbb{E}\Big[\int_{\mathbf{h}_{N1}}^\infty e^{-v_d\mathbf{f}_1 r}\frac{\mathrm{d}r}{r}\Big], \quad B_3^N = \mathbb{E}\Big[\int_{\mathbf{h}_{N1}}^N (1-\mathbf{a}_1^N(r))^N\frac{\mathrm{d}r}{r}\Big], \quad B_4^N = \mathbb{E}\Big[\int_N^\infty (1-\mathbf{a}_1^N(r))^N\frac{\mathrm{d}r}{r}\Big].$$

The two terms B_2^N and B_3^N can be studied together.

Lemma 24. *Assume that* $\int_{\mathbb{R}^d}[R^{2\theta}(x)/m^{2\theta}(x) + R^{\theta d/\rho}(x)/m^{\theta(d+\rho)/\rho}(x)]f(x)\mathrm{d}x < \infty$ *for some* $\theta \in (0,1]$*. Then for all* $N \geq 1$, $|B_2^N| + |B_3^N| \leq CN^{-\theta}$.

Proof. First, $(1-\mathbf{a}_1^N(r))^N \leq \exp(-\mathbf{m}_1 r)$ for all $r \in [0,N]$ by Lemma 11-(ii). Since furthermore $m(x) \leq v_d f(x)$ by definition of m,

$$|B_2^N| + |B_3^N| \leq 2\mathbb{E}\Big[\int_{\mathbf{h}_{N1}}^\infty e^{-\mathbf{m}_1 r}\frac{\mathrm{d}r}{r}\Big] = 2\mathbb{E}\Big[\int_{\mathbf{h}_{N1}\mathbf{m}_1}^\infty e^{-u}\frac{\mathrm{d}u}{u}\Big] \leq 2\mathbb{E}[\Phi(\mathbf{h}_{N1}\mathbf{m}_1)],$$

where $\Phi(x) = \exp(-x)\mathbf{1}_{\{x\geq 1\}} + [1+\log(1/x)]\mathbf{1}_{\{x<1\}}$.

On Ω_N^1, $\mathbf{h}_{N1}\mathbf{m}_1 = 2\log N$, whence $\Phi(\mathbf{h}_{N1}\mathbf{m}_1) \leq CN^{-2}$ and $\mathbb{E}[\Phi(\mathbf{h}_{N1}\mathbf{m}_1)\mathbf{1}_{\Omega_N^1}] \leq CN^{-2}$.

On Ω_N^2, we have $\mathbf{h}_{N1}\mathbf{m}_1 = N\mathbf{m}_1$. Furthermore, there is a constant $C > 0$ such that $\Phi(x) \leq Cx^{-\theta}$ for all $x > 0$. Thus, recalling that $m(x) \leq v_d f(x) \leq v_d R(x)$,

$$\mathbb{E}[\Phi(\mathbf{h}_{N1}\mathbf{m}_1)\mathbf{1}_{\Omega_N^2}] \leq \frac{C}{N^\theta}\mathbb{E}\Big[\frac{1}{\mathbf{m}_1^\theta}\Big] = \frac{C}{N^\theta}\mathbb{E}\Big[\frac{\mathbf{R}_1^{\theta d/\rho}}{\mathbf{m}_1^\theta \mathbf{R}_1^{\theta d/\rho}}\Big] \leq \frac{C}{N^\theta}\mathbb{E}\Big[\frac{\mathbf{R}_1^{\theta d/\rho}}{\mathbf{m}_1^{\theta(d+\rho)/\rho}}\Big] \leq \frac{C}{N^\theta}.$$

On Ω_N^3, we have $\mathbf{h}_{N1}\mathbf{m}_1 = N^{\rho/(d+\rho)}\mathbf{m}_1/\mathbf{R}_1^{d/(d+\rho)}$ and there is $C > 0$ such that $\Phi(x) \leq Cx^{-\theta(d+\rho)/\rho}$ for all $x > 0$. Thus

$$\mathbb{E}[\Phi(\mathbf{h}_{N1}\mathbf{m}_1)\mathbf{1}_{\Omega_N^3}] \leq \frac{C}{N^\theta}\mathbb{E}\Big[\frac{\mathbf{R}_1^{\theta d/\rho}}{\mathbf{m}_1^{\theta(d+\rho)/\rho}}\Big] \leq \frac{C}{N^\theta}.$$

On Ω_N^4, $\mathbf{h}_{N1}\mathbf{m}_1 = \sqrt{N}\mathbf{m}_1/\mathbf{R}_1$ and there is $C > 0$ such that $\Phi(x) \leq Cx^{-2\theta}$ for all $x > 0$, whence

$$\mathbb{E}[\Phi(\mathbf{h}_{N1}\mathbf{m}_1)\mathbf{1}_{\Omega_N^4}] \leq \frac{C}{N^\theta}\mathbb{E}\Big[\frac{\mathbf{R}_1^{2\theta}}{\mathbf{m}_1^{2\theta}}\Big] \leq \frac{C}{N^\theta}.$$

This completes the proof. □

Lemma 25. *Assume that* $\int_{\mathbb{R}^d}[|x|^q + \log(2+|x|)[F(B(x,1))]^{-\theta}]f(x)\mathrm{d}x < \infty$ *for some* $q > 0$ *and some* $\theta \in (0,1]$*. Then for all* $N \geq 1$, $|B_4^N| \leq CN^{-\theta}$.

Proof. Recall that $B_4^N = \mathbb{E}[\int_N^\infty (1-\mathbf{a}_1^N(r))^N \frac{\mathrm{d}r}{r}] = \mathbb{E}[\int_1^\infty (1-\mathbf{a}_1^N(Nu))^N \frac{\mathrm{d}u}{u}]$. We introduce $g(x) = 1 \vee \mathbb{E}[|X_1 - x|^q]$ and $\mathbf{g}_1 = g(X_1)$ as usual. We write $B_4^N = I_N + J_N$, where

$$I_N = \mathbb{E}\Big[\int_1^{(2\mathbf{g}_1)^{d/q}} (1-\mathbf{a}_1^N(Nu))^N \frac{\mathrm{d}u}{u}\Big] \quad \text{and} \quad J_N = \mathbb{E}\Big[\int_{(2\mathbf{g}_1)^{d/q}}^\infty (1-\mathbf{a}_1^N(Nu))^N \frac{\mathrm{d}u}{u}\Big].$$

First, there is $C>0$ such that for all $x \in (0,1]$, all $N \geq 1$, $(1-x)^N \leq e^{-Nx} \leq C(Nx)^{-\theta}$, whence $(1-\mathbf{a}_1^N(Nu))^N \leq C/(N\mathbf{a}_1^N(Nu))^\theta \leq C/(NF(B(X_1,1)))^\theta$ for all $u \geq 1$. Hence

$$I_N \leq \frac{C}{N^\theta}\mathbb{E}\Big[\frac{1}{(F(B(X_1,1)))^\theta}\int_1^{(2\mathbf{g}_1)^{d/q}} \frac{\mathrm{d}r}{r}\Big] \leq \frac{C}{N^\theta}\mathbb{E}\Big[\frac{\log(2+|X_1|)}{(F(B(X_1,1)))^\theta}\Big] \leq \frac{C}{N^\theta}.$$

We used that $|\log(2\mathbf{g}_1)^{d/q}| \leq C\log(2+|X_1|)$ because $1 \leq g(x) \leq C(1+|x|^q)$.

Next, since $1-\mathbf{a}_1^N(Nu) \leq \mathbf{g}_1 u^{-q/d}$ by Lemma 11-(iv),

$$J_N \leq \mathbb{E}\Big[\int_{(2\mathbf{g}_1)^{d/q}}^\infty \Big(\frac{\mathbf{g}_1}{u^{q/d}}\Big)^N \frac{\mathrm{d}u}{u}\Big] \leq \Big(\frac{1}{2}\Big)^{N-1}\mathbb{E}\Big[\mathbf{g}_1\int_1^\infty \frac{\mathrm{d}u}{u^{1+q/d}}\Big] \leq \frac{C}{2^{N-1}}$$

because $\mathbb{E}[\mathbf{g}_1] < \infty$. This completes the proof. □

We next treat B_1^N when $\beta \in (0,2]$.

Lemma 26. *Assume $\beta \in (0,2]$ and $\int_{\mathbb{R}^d}[R^{2\theta}(x)/f^{2\theta}(x) + R^{\theta d/\beta}(x)/f^{\theta(\beta+d)/\beta}(x)]f(x)\mathrm{d}x < \infty$ for some $\theta \in [0, 1\wedge(\beta/d)]$. Then for all $N \geq 1$, $|B_1^N| \leq CN^{-\theta}$.*

Proof. Thanks to Lemma 22-(i), we have $B_1^N = \mathbb{E}[\int_0^{\mathbf{h}_{N1}} e^{-v_d\mathbf{f}_1 r}R_N(X_1,r)\frac{\mathrm{d}r}{r}]$, with furthermore $|R_N(X_1,r)| \leq C(N^{-1}r^2\mathbf{R}_1^2 + N^{-\beta/d}r^{1+\beta/d}\mathbf{R}_1)$. Hence $|B_1^N| \leq C(I_N + J_N)$, with

$$I_N = \frac{1}{N}\mathbb{E}\Big[\mathbf{R}_1^2\int_0^{\mathbf{h}_{N1}} e^{-v_d\mathbf{f}_1 r} r\mathrm{d}r\Big] \quad \text{and} \quad J_N = \frac{1}{N^{\beta/d}}\mathbb{E}\Big[\mathbf{R}_1\int_0^{\mathbf{h}_{N1}} e^{-v_d\mathbf{f}_1 r}r^{\beta/d}\mathrm{d}r\Big].$$

First, for $a,b \geq 0$, we have $\int_0^a e^{-br}r\mathrm{d}r \leq 2(\min\{b^{-1},a\})^2$ because $\int_0^\infty e^{-br}r\mathrm{d}r = 2b^{-2}$ and $\int_0^a r\mathrm{d}r = a^2/2$. Since now $\theta \in [0,1]$, we deduce that $\int_0^a e^{-br}r\mathrm{d}r \leq 2b^{-2\theta}a^{2-2\theta}$. Hence

$$I_N \leq \frac{2}{N}\mathbb{E}\Big[\frac{\mathbf{R}_1^2\mathbf{h}_{N1}^{2-2\theta}}{(v_d\mathbf{f}_1)^{2\theta}}\Big].$$

Since now $h_N(x) \leq \sqrt{N}/R(x)$, we end with $I_N \leq CN^{-\theta}\mathbb{E}[\mathbf{R}_1^{2\theta}\mathbf{f}_1^{-2\theta}] \leq CN^{-\theta}$.

Next, we observe that $\int_0^a e^{-br}r^{\beta/d}\mathrm{d}r \leq C(\min\{b^{-1},a\})^{1+\beta/d} \leq C(b^{-\theta d/\beta}a^{1-\theta d/\beta})^{1+\beta/d}$ because $\theta d/\beta \in [0,1]$. Hence

$$J_N \leq \frac{C}{N^{\beta/d}}\mathbb{E}\Big[\frac{\mathbf{R}_1\mathbf{h}_{N1}^{(1-\theta d/\beta)(1+\beta/d)}}{(v_d\mathbf{f}_1)^{(\theta d/\beta)(1+\beta/d)}}\Big].$$

But we have $\beta \leq 2$ so that $\rho = \beta$ and thus $\mathbf{h}_{N1} \leq N^{\beta/(\beta+d)}/\mathbf{R}_1^{d/(\beta+d)}$. This precisely gives $J_N \leq CN^{-\theta}\mathbb{E}[\mathbf{R}_1^{\theta d/\beta}/\mathbf{f}_1^{\theta(\beta+d)/\beta}] \leq CN^{-\theta}$. □

We finally study B_1^N when $\beta > 2$.

Lemma 27. *Assume $\beta \in (2,d]$, recall that $\ell = \max\{i \in \mathbb{N} : 2i < \beta\}$ and take for granted that $\int_{\mathbb{R}^d}[R^{2\beta/d}(x)/f^{2\beta/d}(x) + R^{\beta/2}(x)/f^{\beta/d+\beta/2}(x)]f(x)\mathrm{d}x < \infty$. There are some constants $\lambda_1,\dots,\lambda_\ell \in \mathbb{R}$ such that for all $N \geq 1$,*

$$\Big|B_1^N - \sum_{i=1}^\ell \frac{\lambda_i}{N^{2i/d}}\Big| \leq \frac{C}{N^{\beta/d}}.$$

Proof. Using the notation of Lemma 22-(ii), we write

$$B_1^N = \mathbb{E}\Big[\int_0^{\mathbf{h}_{N1}} e^{-v_d \mathbf{f}_1 r}\Big(\sum_{i=1}^{\ell} \frac{g_i(X_1,r)}{N^{2i/d}} + S_N(X_1,r)\Big)\frac{\mathrm{d}r}{r}\Big] = \sum_{i=1}^{\ell}\frac{\lambda_i}{N^{2i/d}} + \sum_{i=1}^{\ell}\Delta_N^i + \varepsilon_N,$$

where

$$\lambda_i = \mathbb{E}\Big[\int_0^\infty e^{-v_d\mathbf{f}_1 r} g_i(X_1,r)\frac{\mathrm{d}r}{r}\Big],\quad \Delta_N^i = -\frac{1}{N^{2i/d}}\mathbb{E}\Big[\int_{\mathbf{h}_{N1}}^\infty e^{-v_d\mathbf{f}_1 r} g_i(X_1,r)\frac{\mathrm{d}r}{r}\Big],$$

$$\varepsilon_N = \mathbb{E}\Big[\int_0^{\mathbf{h}_{N1}} e^{-v_d\mathbf{f}_1 r} S_N(X_1,r)\frac{\mathrm{d}r}{r}\Big].$$

It remains to prove that $\lambda_1,\dots,\lambda_\ell$ are well-defined and finite, that $\sum_{i=1}^\ell |\Delta_N^i| \le CN^{-\beta/d}$ and that $|\varepsilon_N| \le CN^{-\beta/d}$, which we do successively in the three following steps.

Step 1. Recalling that $|g_i(x,r)| \le Cr^{2i/d}[rR(x) + (rR(x))^i] \le Cr^{2i/d}[1+(rR(x))^i]$, we see that λ_i is well defined for all $i=1,\dots,\ell$ because

$$\mathbb{E}\Big[\int_0^\infty e^{-v_d\mathbf{f}_1 r} r^{2i/d}[1+(r\mathbf{R}_1)^i]\frac{\mathrm{d}r}{r}\Big] = \mathbb{E}\Big[\int_0^\infty e^{-v_d u}\Big(\frac{u}{\mathbf{f}_1}\Big)^{2i/d}\Big[1+\Big(u\frac{\mathbf{R}_1}{\mathbf{f}_1}\Big)^i\Big]\frac{\mathrm{d}u}{u}\Big] \le C\mathbb{E}\Big[\frac{\mathbf{R}_1^i}{\mathbf{f}_1^{2i/d+i}}\Big]$$

(we used that $\mathbf{R}_1 \ge \mathbf{f}_1$). This is finite because $i \le \ell \le \beta/2$ and because $\mathbb{E}[\mathbf{R}_1^{\beta/2}/\mathbf{f}_1^{\beta/d+\beta/2}] < \infty$.

Step 2. Since $|S_N(x,r)| \le Cr^2R^2(x)/N + C(r/N)^{\beta/d}[rR(x)+(rR(x))^{\beta/2}]$ for all $r\in[0,h_N(x)]$ and since $rR(x)+(rR(x))^{\beta/2} \le 1+2(rR(x))^{\beta/2}$ (recall that $\beta>2$), we have $\varepsilon_N \le \varepsilon_N^1+\varepsilon_N^2$, with

$$\varepsilon_N^1 = \frac{C}{N}\mathbb{E}\Big[\mathbf{R}_1^2\int_0^{\mathbf{h}_{N1}} e^{-v_d\mathbf{f}_1 r} r\mathrm{d}r\Big] \quad\text{and}\quad \varepsilon_N^2 = \frac{C}{N^{\beta/2}}\mathbb{E}\Big[\int_0^{\mathbf{h}_{N1}} e^{-v_d\mathbf{f}_1 r} r^{\beta/d}[1+(r\mathbf{R}_1)^{\beta/2}]\frac{\mathrm{d}r}{r}\Big].$$

First,

$$\varepsilon_N^2 \le \frac{C}{N^{\beta/2}}\mathbb{E}\Big[\int_0^\infty e^{-v_d u}\Big(\frac{u}{\mathbf{f}_1}\Big)^{\beta/d}\Big[1+\Big(\frac{u\mathbf{R}_1}{\mathbf{f}_1}\Big)^{\beta/2}\Big]\Big)\frac{\mathrm{d}u}{u}\Big] \le \frac{C}{N^{\beta/d}}\mathbb{E}\Big[\frac{\mathbf{R}_1^{\beta/2}}{\mathbf{f}_1^{\beta/d+\beta/2}}\Big] \le \frac{C}{N^{\beta/d}}.$$

Next, since $\int_0^a e^{-br} r\mathrm{d}r \le 2(\min\{b^{-1},a\})^2 \le 2b^{-2\beta/d}a^{2-2\beta/d}$ for $a,b>0$ (recall that $\beta/d\in(0,1]$) and since $\mathbf{h}_{N1} \le \sqrt{N}/\mathbf{R}_1$,

$$\varepsilon_N^1 \le \frac{C}{N}\mathbb{E}\Big[\mathbf{R}_1^2\frac{\mathbf{h}_{N1}^{2-2\beta/d}}{\mathbf{f}_1^{2\beta/d}}\Big] \le \frac{C}{N^{\beta/d}}\mathbb{E}\Big[\frac{\mathbf{R}_1^{2\beta/d}}{\mathbf{f}_1^{2\beta/d}}\Big] \le \frac{C}{N^{\beta/d}}.$$

Step 3. For $i=1,\dots,\ell$, using that $|g_i(x,r)| \le Cr^{2i/d}[1+(rR(x))^i]$,

$$|\Delta_N^i| \le \frac{C}{N^{2i/d}}\mathbb{E}\Big[\int_{\mathbf{h}_{N1}}^\infty e^{-v_d\mathbf{f}_1 r} r^{2i/d}[1+(r\mathbf{R}_1)^i]\frac{\mathrm{d}r}{r}\Big] \le \frac{C}{N^{2i/d}}\mathbb{E}\Big[\frac{\mathbf{R}_1^i}{\mathbf{f}_1^{2i/d+i}}\int_{\mathbf{h}_{N1}\mathbf{f}_1}^\infty e^{-v_d u}u^{2i/d}[1+u^i]\frac{\mathrm{d}u}{u}\Big].$$

We used that $f \le R$ for the last inequality. Since now $e^{-v_d u}u^{2i/d}[1+u^i] \le Ce^{-v_d u/2}$, we find

$$|\Delta_N^i| \le \frac{C}{N^{2i/d}}\mathbb{E}\Big[\frac{\mathbf{R}_1^i}{\mathbf{f}_1^{2i/d+i}}e^{-v_d\mathbf{h}_{N1}\mathbf{f}_1/2}\Big] = C\mathbb{E}\Big[\Big(\frac{\mathbf{R}_1}{N^{2/d}\mathbf{f}_1^{2/d+1}}\Big)^i e^{-v_d\mathbf{h}_{N1}\mathbf{f}_1/2}\Big].$$

It follows that $\sum_{i=1}^\ell |\Delta_N^i| \le C\mathbb{E}[\eta_N]$, where

$$\eta_N = \Big[\frac{\mathbf{R}_1}{N^{2/d}\mathbf{f}_1^{2/d+1}} + \Big(\frac{\mathbf{R}_1}{N^{2/d}\mathbf{f}_1^{2/d+1}}\Big)^\ell\Big]e^{-v_d\mathbf{h}_{N1}\mathbf{f}_1/2}.$$

On Ω_N^1, we have $v_d \mathbf{h}_{N1}\mathbf{f}_1/2 = (v_d\mathbf{f}_1/\mathbf{m}_1)\log N \geq \log N$, so that

$$\mathbb{E}[\eta_N \mathbf{1}_{\Omega_N^1}] \leq \frac{1}{N}\mathbb{E}\Big[\frac{\mathbf{R}_1}{N^{2/d}\mathbf{f}_1^{2/d+1}} + \Big(\frac{\mathbf{R}_1}{N^{2/d}\mathbf{f}_1^{2/d+1}}\Big)^\ell\Big] \leq \frac{C}{N}\mathbb{E}\Big[1+\frac{\mathbf{R}_1^{\beta/2}}{\mathbf{f}_1^{\beta/d+\beta/2}}\Big] \leq \frac{C}{N}.$$

We used that $1 \leq \ell \leq \beta/2$.

On Ω_N^2, we have $\mathbf{h}_{N1} = N$. Furthermore, there is a constant $C > 0$ such that for all $x > 0$, $e^{-v_d x/2} \leq Cx^{(2-\beta)/d}$ and $e^{-v_d x/2} \leq Cx^{(2\ell-\beta)/d}$ (recall that $\beta > 2\ell \geq 2$). Hence

$$\eta_N \mathbf{1}_{\Omega_N^2} \leq \Big[\frac{\mathbf{R}_1}{N^{2/d}\mathbf{f}_1^{2/d+1}} + \Big(\frac{\mathbf{R}_1}{N^{2/d}\mathbf{f}_1^{2/d+1}}\Big)^\ell\Big]e^{-v_d N\mathbf{f}_1/2} \leq \frac{C}{N^{\beta/d}\mathbf{f}_1^{\beta/d}}\Big(\frac{\mathbf{R}_1}{\mathbf{f}_1}+\Big(\frac{\mathbf{R}_1}{\mathbf{f}_1}\Big)^\ell\Big).$$

Since $\mathbf{f}_1 \leq v_d\mathbf{M}_1 \leq v_d\mathbf{R}_1$ and since $1 \leq \ell \leq \beta/2$, we conclude that

$$\mathbb{E}[\eta_N \mathbf{1}_{\Omega_N^2}] \leq \frac{C}{N^{\beta/d}}\mathbb{E}\Big[\frac{1}{\mathbf{f}_1^{\beta/d}}\Big(\frac{\mathbf{R}_1}{\mathbf{f}_1}\Big)^{\beta/2}\Big] = \frac{C}{N^{\beta/d}}\mathbb{E}\Big[\frac{\mathbf{R}_1^{\beta/2}}{\mathbf{f}_1^{\beta/d+\beta/2}}\Big] \leq \frac{C}{N^{\beta/d}}.$$

On Ω_N^3, we have $\mathbf{h}_{N1}\mathbf{f}_1 = N^{2/(d+2)}\mathbf{f}_1/\mathbf{R}_1^{d/(d+2)}$ (recall that $\rho = 2$ because $\beta > 2$). Moreover, there is $C > 0$ such that $e^{-v_d x/2} \leq Cx^{(2-\beta)(d+2)/(2d)}$ and $e^{-v_d x/2} \leq Cx^{(2\ell-\beta)(d+2)/(2d)}$. Hence

$$\begin{aligned}\eta_N \mathbf{1}_{\Omega_N^3} \leq& C\frac{\mathbf{R}_1}{N^{2/d}\mathbf{f}_1^{2/d+1}}(\mathbf{h}_{N1}\mathbf{f}_1)^{(2-\beta)(d+2)/(2d)} + C\Big(\frac{\mathbf{R}_1}{N^{2/d}\mathbf{f}_1^{2/d+1}}\Big)^\ell(\mathbf{h}_{N1}\mathbf{f}_1)^{(2\ell-\beta)(d+2)/(2d)}\\ =& C\frac{\mathbf{R}_1^{\beta/2}}{N^{\beta/d}\mathbf{f}_1^{\beta/d+\beta/2}}\end{aligned}$$

and $\mathbb{E}[\eta_N\mathbf{1}_{\Omega_N^3}] \leq CN^{-\beta/d}$.

Finally on Ω_N^4, we have $\mathbf{h}_{N1}\mathbf{f}_1 = \sqrt{N}\mathbf{f}_1/\mathbf{R}_1$. Moreover, there is a constant $C > 0$ such that $e^{-v_d x/2} \leq Cx^{2(2-\beta)/d}$ and $e^{-v_d x/2} \leq Cx^{2(2\ell-\beta)/d}$ for all $x > 0$. Hence

$$\eta_N \mathbf{1}_{\Omega_N^4} \leq C\frac{\mathbf{R}_1}{N^{2/d}\mathbf{f}_1^{2/d+1}}(\mathbf{h}_{N1}\mathbf{f}_1)^{2(2-\beta)/d} + C\Big(\frac{\mathbf{R}_1}{N^{2/d}\mathbf{f}_1^{2/d+1}}\Big)^\ell(\mathbf{h}_{N1}\mathbf{f}_1)^{2(2\ell-\beta)/d}.$$

We now use the Young inequality with $p = \beta/2$ and $p_* = \beta/(\beta-2)$ for the first term and $p = \beta/(2\ell)$ and $p_* = \beta/(\beta-2\ell)$ for the second one:

$$\eta_N \mathbf{1}_{\Omega_N^4} \leq C\Big(\frac{\mathbf{R}_1}{N^{2/d}\mathbf{f}_1^{2/d+1}}\Big)^{\beta/2} + C(\mathbf{h}_{N1}\mathbf{f}_1)^{-2\beta/d} = \frac{C\mathbf{R}_1^{\beta/2}}{N^{\beta/d}\mathbf{f}_1^{\beta/d+\beta/2}} + \frac{C\mathbf{R}_1^{2\beta/d}}{N^{\beta/d}\mathbf{f}_1^{2\beta/d}}.$$

Thus $\mathbb{E}[\eta_N\mathbf{1}_{\Omega_N^4}] \leq CN^{-\beta/d}$, which completes the proof. □

We quickly give the

Proof of Remark 5. Coming back to the proof of Lemma 22, we see that $g_1(x,r) = -r^{2/d+1}\delta_2 f(x)$, with $\delta_2 f(x) = \sum_{|\alpha|=2}(c_\alpha/\alpha!)\partial_\alpha f(x)$ and $c_\alpha = \int_{B(0,1)} y^\alpha \mathrm{d}y$. But for $\alpha = (\alpha_1, \dots, \alpha_d)$ with $|\alpha| = 2$, we see that $c_\alpha = 0$ unless there is i such that $\alpha_i = 2$ and then $c_\alpha > 0$ does not depend on i (because we work with some symmetric norm). Thus $g_1(x,r) = -cr^{2/d+1}\Delta f(x)$ for some constant $c > 0$. Coming back to the proof of Lemma 27, it holds that $\lambda_1 = \mathbb{E}[\int_0^\infty e^{-v_d\mathbf{f}_1}g_1(X_1,r)\frac{\mathrm{d}r}{r}]$. Thus, allowing the value of $c > 0$ to vary,

$$\lambda_1 = -c\mathbb{E}\Big[\int_0^\infty e^{-v_d u}\Delta f(X_1)(u/\mathbf{f}_1)^{2/d+1}\frac{\mathrm{d}u}{u}\Big] = -c\mathbb{E}\Big[\Delta f(X_1)/\mathbf{f}_1^{2/d+1}\Big].$$

As a consequence,

$$\lambda_1 = -c \int_{\mathbb{R}^d} f^{-2/d}(x)\Delta f(x)\mathrm{d}x = c \int_{\mathbb{R}^d} f^{-2/d-1}(x)|\nabla f(x)|^2 \mathrm{d}x,$$

the integration by parts being allowed if $\lim_{|x|\to\infty} f^{-2/d}(x)|\nabla f(x)| = 0$. □

We now have all the weapons to conclude the

Proof of Theorem 3. We fix $\beta \in (0,2] \cap (0,d]$, so that $\rho = \beta$. We assume that $f \in \mathcal{D}^\beta(\mathbb{R}^d)$, we recall that M and G_β were defined in (3) and (4) and that $R = M + G_\beta$. We assume that $\kappa = \sup_{x\in\mathbb{R}^d} F(B(x,1)) < 1$. We assume (6) for some $\theta \in (0, \beta/d]$ and some $q > 0$ (with $r_0 = 1$). We recall that $\mathbb{E}[H_N] - H(f) = B_1^N + B_2^N + B_3^N + B_4^N$. It suffices to use Lemmas 24, 25 and 26 to find that $|B_1^N| + |B_2^N| + |B_3^N| + |B_4^N| \le CN^{-\theta}$. The condition (6) indeed implies that we can apply all these lemmas (this uses that $m \le v_d f$). □

Proof of Theorem 4. We fix $d \ge 3$ and $\beta \in (2, d]$, whence $\rho = 2$ and $\ell = \max\{i \in \mathbb{N} : 2i < \beta\} \ge 1$. We assume that $f \in \mathcal{D}^\beta(\mathbb{R}^d)$, we recall that M and G_β were defined in (3) and (4) and that $R = M + G_\beta$. We assume that $\kappa = \sup_{\mathbb{R}^d} F(B(x,1)) < 1$. We assume (7) for some $q > 0$ (with $r_0 = 1$). We recall that $\mathbb{E}[H_N] - H(f) = B_1^N + B_2^N + B_3^N + B_4^N$. By Lemmas 24 and 25 with $\theta = \beta/d$, we have $|B_2^N| + |B_3^N| + |B_4^N| \le CN^{-\beta/d}$. Lemma 27 tells us that $|B_1^N - \sum_{i=1}^{\ell} \lambda_i N^{-2i/d}| \le CN^{-\beta/d}$. We indeed can apply all these lemmas thanks to (7) (and since $m \le v_d f$). All this shows that $|\mathbb{E}[H_N] - H(f) - \sum_{i=1}^{\ell} \lambda_i N^{-2i/d}| \le CN^{-\beta/d}$. □

7. Corollaries, examples and numerical computation of χ_d

7.1. **Corollaries.** We start with a remark.

Remark 28. *(i) If $\int_{\mathbb{R}^d} |x|^{d+\varepsilon} f(x)\mathrm{d}x < \infty$ for some $\varepsilon \in (0,1)$, then $\int_{\mathbb{R}^d} f^{1/2-\varepsilon'}(x)\mathrm{d}x < \infty$, where $\varepsilon' = \varepsilon/(4d+2)$.*

(ii) If f is bounded, if $m \ge cf$ for some constant $c > 0$ and if $\int_{\mathbb{R}^d} |x|^{d+\varepsilon} f(x)\mathrm{d}x < \infty$ for some $\varepsilon \in (0,1)$, then, with $\theta = 1/2 + \varepsilon/(4d+2)$,

$$\int_{\mathbb{R}^d} \Big(\log^2 m(x) + \frac{\log^2(2+|x|)}{[F(B(x,1))]^\theta} \Big) f(x)\mathrm{d}x < \infty.$$

Proof. For point (i), we write $f^{1/2-\varepsilon'}(x) = f^{1/2-\varepsilon'}(x)(1+|x|)^{(d+\varepsilon)(1/2-\varepsilon')}(1+|x|)^{-(d+\varepsilon)(1/2-\varepsilon')}$ and we use the Hölder inequality with $p = 1/(1/2-\varepsilon')$ and $p_* = 1/(1/2+\varepsilon')$. This gives $\int_{\mathbb{R}^d} f^{1/2-\varepsilon'}(x)\mathrm{d}x \le I^{1/p} J^{1/p_*}$, with $I = \int_{\mathbb{R}^d} f(x)(1+|x|)^{d+\varepsilon}\mathrm{d}x < \infty$ by assumption and $J = \int_{\mathbb{R}^d} (1+|x|)^{-(d+\varepsilon)(1/2-\varepsilon')/(1/2+\varepsilon')}\mathrm{d}x < \infty$ because $(d+\varepsilon)(1/2-\varepsilon')/(1/2+\varepsilon') > d$.

For (ii), since $f \le m/c \le v_d f/c$ and since f is bounded, we can find C such that $\log^2 m \le C(1+\log^2 f) \le Cf^{-\theta}$ and $[F(B(x,1))]^{-\theta} \le m^{-\theta}(x) \le Cf^{-\theta}(x)$. We thus only have to prove that $I = \int_{\mathbb{R}^d} f^{1-\theta}(x)\log^2(2+|x|)\mathrm{d}x < \infty$. Since $1-\theta = 1/2-\varepsilon'$, this is checked as point (i). □

We can now give the

Proof of Corollary 7. We fix $\varepsilon \in (0,1)$ and assume that $f \in \mathcal{D}^\nu(\mathbb{R}^d)$ with $\nu = 1$ if $d = 1$, $\nu = 2$ if $d \in \{2,3\}$ and $\nu = d/2 + \varepsilon$ if $d \ge 4$. We assume that $\kappa < 1$ (with $r_0 = 1$), that $R = M + G_\nu$ is bounded and that there is $c > 0$ such that $m \ge cf$. We finally assume that $\int_{\mathbb{R}^d} |x|^{d+\varepsilon} f(x)\mathrm{d}x < \infty$ and that (a) $\int_{\{f>0\}} \sqrt{R(x)} f^{-\varepsilon}(x)\mathrm{d}x < \infty$ if $d \in \{1,2\}$ or that (b) $\int_{\{f>0\}} R^{d/4}(x) f^{1/2-d/4-\varepsilon}(x)\mathrm{d}x < \infty$ if $d \ge 3$.

Step 1. We can apply Theorem 1 and Proposition 6 with $\theta = 1/2 + \varepsilon/(4d+2)$. Indeed, f is continuous and bounded, so that we have only to check (5) and (8). By Remark (28)-(ii), we only have to verify that $\int_{\mathbb{R}^d}(M(x)/m(x))(1+|\log m(x)|)f(x)\mathrm{d}x < \infty$. But since $(M/m)f \le CM \le CR$ and $|\log m| \le C(1+|\log f|) \le Cf^{-\varepsilon}$ (because $m \le Cf \le C$), $(M/m)(1+|\log m|)f \le CRf^{-\varepsilon}$. If $d \in \{1,2\}$, we use that $R \le C\sqrt{R}$ and conclude with (a). If $d \ge 3$, we write $Rf^{-\varepsilon} \le Rf^{-\varepsilon}(R/f)^{d/4-1/2} = R^{d/4+1/2}f^{1/2-d/4-\varepsilon} \le CR^{d/4}f^{1/2-d/4-\varepsilon}$ and conclude with (b).

Step 2. We now show that we can apply Theorem 3 with $\theta = 1/2+\varepsilon/(4d+2)$ and $\beta = \nu$ when $d \in \{1,2,3\}$. By Remark (28)-(ii), we only have to verify that $I = \int_{\mathbb{R}^d}(R(x)/m(x))^{2\theta}f(x)\mathrm{d}x < \infty$ and $J = \int_{\mathbb{R}^d}(R^{\theta d/\nu}(x)/m^{\theta(d+\nu)/\nu}(x))f(x)\mathrm{d}x < \infty$. First, $(R/m)^{2\theta}f \le CR^{1+\varepsilon/(2d+1)}f^{-\varepsilon/(2d+1)}$. If $d \in \{1,2\}$, we write $(R/m)^{2\theta}f \le \sqrt{R}f^{-\varepsilon}$, so that $I < \infty$ by (a). If $d = 3$, we write $(R/m)^{2\theta}f \le CR^{3/4}f^{-1/4-\varepsilon}$ so that $I < \infty$ by (b). Next, if $d \in \{1,2\}$, so that $\nu = d$, we have $(R^{\theta d/\nu}/m^{\theta(d+\nu)/\nu})f \le CR^{\theta}f^{1-2\theta} \le C\sqrt{R}f^{-\varepsilon}$, whence $J < \infty$ by (a). If $d = 3$, then $(R^{\theta d/\nu}/m^{\theta(d+\nu)/\nu})f \le CR^{3\theta/2}f^{1-5\theta/2} \le CR^{3/4}f^{-1/4-\varepsilon}$, whence $J < \infty$ by (b).

Step 3. We now prove that we can apply Theorem 4 with $\beta = d/2 + \varepsilon d/(4d+2) \in (d/2,\nu)$ when $d \ge 4$. Since $\beta/d = 1/2+\varepsilon/(4d+2)$, we can use Remark 28-(ii) and we only have to check that $I = \int_{\mathbb{R}^d}(R(x)/m(x))^{2\beta/d}f(x)\mathrm{d}x$ and $J = \int_{\mathbb{R}^d}(R^{\beta/2}(x)/m^{\beta/d+\beta/2}(x))f(x)\mathrm{d}x$ are finite. But $(R/m)^{2\beta/d}f \le CRf^{-\varepsilon/(2d+1)} \le CRf^{-\varepsilon} \le CRf^{-\varepsilon}(R/f)^{d/4-1/2} = CR^{d/4+1/2}f^{1/2-d/4-\varepsilon} \le CR^{d/4}f^{1/2-d/4-\varepsilon}$. Next, $(R^{\beta/2}/m^{\beta/d+\beta/2})f \le CR^{d/4}f^{1/2-d/4-\varepsilon(d+2)/(8d+4)} \le CR^{d/4}f^{1/2-d/4-\varepsilon}$. Hence I and J are finite by (b).

Step 4. Here we conclude when $d \in \{1,2,3\}$: by Step 1, we know that $\sqrt{N}(H_N - \mathbb{E}[H_N]) \to \mathcal{N}(0,\sigma^2(f))$ in law and that $V_N \to \sigma^2(f)$ in probability. By Step 2, we know that $|\mathbb{E}[H_N]-H(f)| \le CN^{-1/2-\varepsilon/(4d+2)}$, so that $\sqrt{N}(H_N - \mathbb{E}[H_N]) \to 0$ in probability. We deduce that, as desired, $\sqrt{N/V_N}(H_N - H(f)) \to \mathcal{N}(0,1)$ in law.

Step 5. We now assume that $d \ge 4$ and observe that since $\varepsilon \in (0,1)$, with $\beta = d/2+\varepsilon d/(4d+2)$, we have $\ell = \max\{i \in \mathbb{N} : 2i < \beta\} = \lfloor d/4 \rfloor$. By Step 1, we know that $\sqrt{N}(H_N - \mathbb{E}[H_N]) \to \mathcal{N}(0,\sigma^2(f))$ in law and that $V_N \to \sigma^2(f)$ in probability. By Step 2, we know that there are some numbers $\lambda_1,\dots,\lambda_\ell$ so that $|\mathbb{E}[H_N] - H(f) - \sum_{i=1}^{\ell}\lambda_i N^{-2i/d}| \le CN^{-1/2-\varepsilon/(4d+2)}$.

Recall now (9): we have $H_N^{(d)} = \sum_{k=0}^{\ell}\alpha_{k,d}H^k_{2^{\ell-k}n}$, where $n = \lfloor (N+1-\ell)/(2^{\ell+1}-1)\rfloor$, and where $H^0_{2^\ell n},\dots,H^\ell_n$ are independent. For each $k = 0,\dots,\ell$, we have

$$\sqrt{2^{\ell-k}n}(H^k_{2^{\ell-k}n} - \mathbb{E}[H^k_{2^{\ell-k}n}]) \to \mathcal{N}(0,\sigma^2(f)), \tag{24}$$

$$\Big|\mathbb{E}[H^k_{2^{\ell-k}n}] - H(f) - \sum_{i=1}^{\ell}\lambda_i(2^{\ell-k}n)^{-2i/d}\Big| \le C(2^{\ell-k}n)^{-1/2-\varepsilon/(4d+2)} \le CN^{-1/2-\varepsilon/(4d+2)}. \tag{25}$$

From (24) and since $n \sim (2^{\ell+1}-1)^{-1}N$, we conclude that $\sqrt{N}(H_N^{(d)} - \mathbb{E}[H_N^{(d)}]) \to \mathcal{N}(0,a_d\sigma^2(f))$, where $a_d = (2-2^{-\ell})\sum_{k=0}^{\ell}\alpha_{k,d}^2 2^k$. And the numbers $\alpha_{k,d}$ are such that $\sum_{k=0}^{\ell}\alpha_{k,d} = 1$ and $\sum_{k=0}^{\ell}\alpha_{k,d}2^{2ki/d} = 0$ for all $i = 1,\dots,\ell$. A little computation allows us to deduce from (25) that $|\mathbb{E}[H_N^{(d)}] - H(f)| \le CN^{-1/2-\varepsilon/(4d+2)}$, whence $\sqrt{N}|\mathbb{E}[H_N^{(d)}] - H(f)| \to 0$ in probability. All this proves that $\sqrt{N}(H_N^{(d)} - H(f)) \to \mathcal{N}(0,a_d\sigma^2(f))$ in law. Since finally $V_N \to \sigma^2(f)$ in probability, we conclude that $\sqrt{N/V_N}(H_N^{(d)} - H(f)) \to \mathcal{N}(0,a_d)$ in law as desired. ☐

We next give the

Proof of Corollary 9. We assume that $f \in \mathcal{D}^\nu(\mathbb{R}^d)$ with $\nu = \min\{d, 2\}$, that $\kappa < 1$ (with $r_0 = 1$), that $R = M + G_\nu$ is bounded and that there is $c > 0$ such that $m \geq cf$. Assume finally that $\int_{\mathbb{R}^d} |x|^{d+\varepsilon} f(x)\mathrm{d}x < \infty$ for some $\varepsilon > 0$, that $\int_{\{f>0\}} M(x)|\log f(x)|\mathrm{d}x < \infty$ and that (a) $\int_{\mathbb{R}^d} \sqrt{R(x)}\mathrm{d}x < \infty$ if $d = 1$ and (b) $\int_{\mathbb{R}^d} R^{d/(2+d)}(x)\mathrm{d}x < \infty$ if $d \geq 2$. Let $\theta = 1/2$ if $d = 1$ and $\theta = 2/(d+2)$ if $d \geq 2$.

The only thing we have to verify is that we can apply Theorems 1 and 3 with this θ (and with $\beta = \nu$). We only have to verify (5) and (6). By Remark 28-(ii) and since $(M/m)(1 + |\log m|)f \leq CM + CM|\log f|$ (and M is integrable because e.g. if $d \geq 2$, we have $M \leq R \leq CR^{d/(2+d)}$ and (b)), we conclude that (5) holds with any $\theta \in (0, 1/2 + \varepsilon/(4d+2)]$ and thus in particular with our θ. For (6), we need $\int_{\mathbb{R}^d} (R(x)/m(x))^{2\theta} f(x)\mathrm{d}x$ and $\int_{\mathbb{R}^d} (R^{\theta d/\nu}(x)/m^{\theta(d+\nu)/\nu}(x))f(x)\mathrm{d}x$ to be finite.

If $d = 1$, we have $(R/m)^{2\theta} f = (R/m)f \leq CR \leq C\sqrt{R}$, as well as $(R^{\theta d/\nu}/m^{\theta(d+\nu)/\nu})f = \sqrt{R}m^{-1}f \leq C\sqrt{R}$, whence the result by (a).

If $d \geq 2$, we have $(R/m)^{2\theta} f(x) = (R/m)^{4/(d+2)} f \leq CR^{4/(d+2)} f^{1-4/(d+2)} \leq CR \leq CR^{d/(2+d)}$ and $(R^{\theta d/\nu}/m^{\theta(d+\nu)/\nu})f = R^{d/(d+2)}m^{-1}f \leq CR^{d/(d+2)}$, whence the result by (b). □

7.2. **Examples.** We next verify that the examples of Subsection 1.9 satisfy the announced properties. It is always easily checked that there is $c > 0$ such that $f \geq cm$, so that we omit the proof, except in example (f) where it is rather tedious.

(a) If $f(x) = (2\pi)^{-d/2} \exp(-|x|^2/2)$, Corollary 7 applies: we can take $r_0 = 1/2$ and the only difficulty is to check (11). But there is a constant C such that $R(x) \leq C \sup_{y \in B(x,1)} (1 + |y|^{\lceil \nu \rceil}) f(y) \leq Ce^{2|x|} f(x)$. Consequently, if $d \in \{1, 2\}$, we have $\int_{\mathbb{R}^d} \sqrt{R(x)} f^{-\varepsilon}(x)\mathrm{d}x \leq C \int_{\mathbb{R}^d} e^{|x|} f^{1/2-\varepsilon}(x)\mathrm{d}x$ and if $d \geq 3$, $\int_{\mathbb{R}^d} R^{d/4}(x) f^{1/2-d/4-\varepsilon}(x)\mathrm{d}x \leq C \int_{\mathbb{R}^d} e^{d|x|/2} f^{1/2-\varepsilon}(x)\mathrm{d}x$. These integrals indeed converge if e.g. $\varepsilon = 1/4$.

(b) If $f(x) = c_{d,a} e^{-(1+|x|^2)^{a/2}}$ for some $a > 0$, then Corollary 7 applies: we can take $r_0 = 1/2$ and the only difficulty is to check (11). If $a \in (0, 1]$, there is a constant C such that $R(x) \leq C \sup_{y \in B(x,1)} f(y) \leq Cf(x)$ and (11) follows with e.g. $\varepsilon = 1/4$. If now $a > 1$, we have $R(x) \leq C \sup_{y \in B(x,1)} (1 + |x|^{\lceil \nu \rceil (a-1)}) f(y) \leq Ce^{2a(1+|x|^2)^{(a-1)/2}} f(x)$ and, again, (11) follows with $\varepsilon = 1/4$.

(c) If $f(x) = c_{d,a}(1 + |x|^2)^{-(d+a)/2}$ with $a > d$, then Corollary 7 applies: we can take $r_0 = 1/2$, we have $R \leq Cf$ so that (11) only requires (in any dimension) that $\int_{\mathbb{R}^d} f^{1/2-\varepsilon}(x)\mathrm{d}x < \infty$ and this is the case, for $\varepsilon > 0$ small enough, because $a > d$. Also, (10) with $\varepsilon > 0$ small enough follows from the fact that $a > d$.

(d) If $f(x) = c_{d,a}|x|^a e^{-|x|}$ with $a > 0$, then f belongs to $\mathcal{D}^a(\mathbb{R}^d)$, we can take $r_0 = 1/2$ and we have $M(x) + G_\beta(x) \leq C(1 + |x|^a)e^{-|x|}$ for $\beta \in (0, a]$.

- If $d = 1$, $a \geq 1$ or $d = 2$, $a \geq 2$, we can apply Corollary 7: (11) holds true for $\varepsilon \in (0, (d/a) \wedge (1/2))$.
- If $d = 3$ and $a \in [2, 12)$, we can apply Corollary 7: (11) holds for $\varepsilon \in (0, 1/2)$ such that $(1/4 + \varepsilon)a < 3$.
- If $d \in \{4, \ldots, 9\}$ and $a \in (d/2, 4d/(d-2))$, we can apply Corollary 7: (11) holds true for $\varepsilon \in (0, 1)$ such that $(d/4 - 1/2 + \varepsilon)a < d$.
- If $d \geq 10$, we can never apply Corollary 7.
- But, for any $d \geq 3$, $a \geq 2$, Corollary 9 applies.

(e) If $f(x) = c_{a,b} \prod_{i=1}^d x_i^{a_i}(1-x_i)^{b_i} \mathbf{1}_{\{x_i \in [0,1]\}}$ for some $a = (a_1, \dots, a_d)$ and $b = (b_1, \dots, b_d)$ both in $(0,\infty)^d$, we set $\tau = \min\{a_1, b_1, \dots, a_d, b_d\}$ and $\mu = \max\{a_1, b_1, \dots, a_d, b_d\}$. Then $f \in \mathcal{D}^\tau(\mathbb{R}^d)$ and we can take $r_0 = 1/4$. For any $\beta \in (0, \tau]$, we have $M(x) + G_\beta(x) \le C$ (so that for any powers $\alpha > 0$, $\eta > 0$, $\int_{\{f>0\}}(M(x) + G_\beta(x))^\alpha f^{-\eta}(x)\mathrm{d}x < \infty$ if and only if $\eta < 1/\mu$).

- If $d = 1$, $\tau \ge 1$ or $d = 2$, $\tau \ge 2$, Corollary 7 applies: (11) holds for $\varepsilon \in (0, 1/\mu)$.
- If $d = 3$, $2 \le \tau \le \mu < 4$, Corollary 7 applies: (11) holds for $\varepsilon \in (0,1)$ such that $(1/4 - \varepsilon)\mu < 1$.
- If $d \ge 4$, we can never apply Corollary 7. But for any $d \ge 3$, $\tau \ge 2$, Corollary 9 applies.

(f) If $d = 1$ and $f(x) = c_p x^p |\sin(\pi/x)| \mathbf{1}_{\{x \in (0,1)\}}$ with $p \ge 2$, then $f \in \mathcal{D}^1(\mathbb{R})$ and we can choose $r_0 = 1/6$. To apply Corollary 7, we need to verify that (i) $\int_0^1 f^{-\varepsilon}(x)\mathrm{d}x < \infty$ for some $\varepsilon > 0$ and (ii) there is $c > 0$ such that $m \ge cf$.

We start with (i). By the Cauchy-Schwarz inequality and since $\int_0^1 x^{-p\varepsilon}(x)\mathrm{d}x < \infty$ for $\varepsilon \in (0, 1/p)$, it suffices to verify that $I = \int_0^1 |\sin(\pi/x)|^{-\varepsilon}\mathrm{d}x < \infty$ for $\varepsilon > 0$ small enough. But $I = \int_1^\infty |\sin(\pi u)|^{-\varepsilon} \frac{\mathrm{d}u}{u^2} \le \sum_{k\ge1} k^{-2} \int_k^{k+1} |\sin(\pi u)|^{-\varepsilon}\mathrm{d}u$. Now if $\varepsilon \in (0,1)$, $\int_k^{k+1} |\sin(\pi u)|^{-\varepsilon}\mathrm{d}u$ is finite and does not depend on k, whence $I < \infty$ as desired.

For (ii), we need to verify that there is $c > 0$ such that $D(x,r) := r^{-1}\int_{x-r}^{x+r} f(y)\mathrm{d}y \ge cf(x)$ for all $x \in (0,1)$ and all $r \in (0, 1/6)$. Let us give the main steps.

(A) For all $0 \le a \le b \le a+2$, $\int_a^b |\sin \pi u|\mathrm{d}u \ge c(b-a)(|\sin(\pi a)| + |\sin(\pi b)|)$. This relies on a little study: assume that $0 \le a \le b \le 2$ by periodicity and separate the cases $0 \le a \le b \le 1$, $0 \le a \le 1 \le b \le 2$ and $1 \le a \le b \le 2$.

(B) If $x \in [1/2, 1)$, $D(x,r) \ge r^{-1}\int_{x-r}^x f(y)\mathrm{d}y \ge cr^{-1}\int_{x-r}^x |\sin(\pi/y)|\mathrm{d}y = cr^{-1}\int_a^b |\sin(\pi u)|\frac{\mathrm{d}u}{u^2}$, with $a = x^{-1}$ and $b = (x-r)^{-1}$. Since $1 \le a \le b \le 3$, we deduce from (A) that $D(x,r) \ge cr^{-1}\int_a^b |\sin(\pi u)|\mathrm{d}u \ge cr^{-1}(b-a)|\sin(\pi a)|$. Using again that $1 \le a \le b \le 3$, we see that $r^{-1}(b-a) = ab \ge 1$, so that finally, $D(x,r) \ge c|\sin(\pi/x)| \ge cx^p|\sin(\pi/x)|$ as desired.

(C) If now $x \in (0, 1/2)$, we write $D(x,r) \ge cx^p r^{-1}\int_x^{x+r} |\sin(\pi/y)|\mathrm{d}y = cx^p r^{-1}\int_a^b |\sin(\pi u)|\frac{\mathrm{d}u}{u^2}$, with $a = (x+r)^{-1}$ and $b = x^{-1}$, so that $3/2 \le a \le b$.

- If $\lfloor b \rfloor \in \{\lfloor a \rfloor, \lfloor a \rfloor + 1\}$, which implies that $b - a \le 2$, we use point (A) to write $D(x,r) \ge cx^p r^{-1} b^{-2}(b-a)|\sin(\pi b)| \ge cx^p|\sin(\pi/x)|$ as desired. We used that $r^{-1}b^{-2}(b-a) = ab^{-1} \ge 3/7$ because $3/2 \le a \le b \le a+2$.
- If $\lfloor b \rfloor \ge \lfloor a \rfloor + 2$, we write

$$D(x,r) \ge cx^p r^{-1} \sum_{k=\lfloor a \rfloor+1}^{\lfloor b \rfloor - 1} k^{-2} \int_k^{k+1} |\sin(\pi u)|\mathrm{d}u = cx^p r^{-1} \sum_{k=\lfloor a \rfloor+1}^{\lfloor b \rfloor - 1} k^{-2} \ge cx^p r^{-1} \int_{\lfloor a \rfloor+1}^{\lfloor b \rfloor} t^{-2}dt,$$

whence $D(x,r) \ge cx^p r^{-1}[(\lfloor a \rfloor + 1)^{-1} - \lfloor b \rfloor^{-1}]$. Using that $\lfloor b \rfloor \ge \lfloor a \rfloor + 2 \ge 3$, we easily conclude that $D(x,r) \ge cx^p r^{-1}(a^{-1} - b^{-1}) = cx^p \ge cx^p|\sin(\pi/x)|$.

7.3. **Numerical computation of χ_d.** Recall that χ_d was defined in Theorem 1. If first we use the norm $|\cdot|_\infty$, then we can compute expicitely, for any $r, s > 0$ and $y \in \mathbb{R}^d$,

$$\varphi(r,s,y) := \int_{B(0,r)\cap B(y,s)} \mathrm{d}z = \prod_{k=1}^d [r \wedge (y_k + s) - (-r) \vee (y_k - s)].$$

Then, it is slightly tedious but not difficult, using that $\varphi(r,s,y)=\varphi(s,r,y)$, to see that $\chi_d = 2\log 2 + \pi^2/6 - 1 + \mathbb{E}[Z]$, where Z is defined as follows. Consider, $U, V \sim \text{Exp}(1)$, as well as $X_1,\dots,X_d \sim \mathcal{U}([-1,1])$, all these variables being independent. Set $R = ((U\vee V)/v_d)^{1/d}$, $S = ((U\wedge V)/v_d)^{1/d}$. Consider the vectors $X=(X_1,\dots,X_d)$ and $Y=(R+S)X$. Finally, put

$$Z = \frac{e^{\varphi(R,S,Y)}-1}{UV}\mathbf{1}_{\{|Y|_\infty \geq R\}}.$$

We thus can estimate χ_d by a simple Monte-Carlo method, approximating $\mathbb{E}[Z]$ by $N^{-1}\sum_{i=1}^N Z_i$, where the Z_i's are i.i.d. copies of Z. We used $N = 1000000$.

Next, when using the norm $|\cdot|_\infty$, the situation is more intricate because we found no explicit formula for

$$\varphi(r,s,y) := \int_{B(0,r)\cap B(y,s)} \mathrm{d}z,$$

so that we have to use a double Monte-Carlo method.

We can however note that, denoting by e_1 the vector $(1,0,\dots,0)\in\mathbb{R}^d$, we have $\varphi(r,s,y) = r^d\varphi(1,s/r,|y|e_1/r)$ for all $r,s>0$ and $y\in\mathbb{R}^d$. Using spherical coordinates, one finds, using the abusive notation $r=((u\vee v)/v_d)^{1/d}$ and $s=((u\wedge v)/v_d)^{1/d}$

$$\chi_d = 2\log 2 + \frac{\pi^2}{6} - 1 + \int_0^\infty\int_0^\infty \frac{e^{-u-v}}{uv}\mathrm{d}u\mathrm{d}v\int_r^{r+s}\Big[e^{r^d\varphi(1,s/r,\rho e_1/r)}-1\Big]dv_d\rho^{d-1}\mathrm{d}\rho.$$

Using next the substitution $\rho\mapsto t = [(\rho/r)^d-1]/[(1+s/r)^d-1]$, we get

$$\chi_d = 2\log 2+\frac{\pi^2}{6}-1+\int_0^\infty\int_0^\infty\frac{e^{-u-v}}{uv}\mathrm{d}u\mathrm{d}v\int_0^1\Big[e^{r^d\varphi(1,s/r,[1+t([1+s/r]^d-1)]^{1/d}e_1)}-1\Big]v_d[(r+s)^d-r^d]\mathrm{d}t.$$

We then approximated χ_d, for N and K large, by

$$2\log 2+\frac{\pi^2}{6}-1+\frac{v_d}{N}\sum_{i=1}^N\frac{(R_i+S_i)^d-R_i^d}{U_iV_i}\Big[\exp\Big(\frac{v_dR_i^d}{K}\sum_{j=1}^K\mathbf{1}_{\{|W_j-\tau_ie_1|_2\leq S_i/R_i\}}\Big)-1\Big],$$

where $U_1,\dots,U_N$ and $V_1,\dots,V_N$ are Exp(1)-distributed, $T_1,\dots,T_n$ are $\mathcal{U}([0,1])$-distributed, and $W_1,\dots,W_K$ are $\mathcal{U}(B(0,1))$-distributed, all these variables being independent. We have also set $R_i = ((U_i\vee V_i)/v_d)^{1/d}$, $S_i=((U_i\wedge V_i)/v_d)^{1/d}$ and $\tau_i = (1+T_i[(1+S_i/R_i)^d-1])^{1/d}$. The idea is that for all i fixed, $(v_d/K)\sum_{j=1}^K\mathbf{1}_{\{|W_j-\tau_ie_1|_2\leq S_i/R_i\}}$ approximates $\varphi(1,(S_i/R_i),\tau_ie_1)$. We actually used $K=N=30000$.

References

[1] J. Beirlant, E.J. Dudewicz, L. Györfi, E.C. van der Meulen, *Nonparametric entropy estimation: an overview,* Int. J. Math. Stat. Sci. **6** (1997), 17–39.

[2] T.B. Berrett, R.J. Samworth, M. Yuan, *Efficient multivariate entropy estimation via k-nearest neighbour distances,* arXiv:1606.00304

[3] G. Biau, L. Devroye, *Lectures on the nearest neighbor method,* Springer Series in the Data Sciences. Springer, Cham, 2015.

[4] P.J. Bickel, L. Breiman, *Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test,* Ann. Probab. **11** (1983), 185–214.

[5] S. Chatterjee, *A new method of normal approximation,* Ann. Probab. **36** (2008), 1584–1610.

[6] E.J. Dudewicz, E.C. van der Meulen, *Entropy-based tests of uniformity,* J. Amer. Statist. Assoc. **76** (1981), no. 376, 967–974.

[7] H. El Haje, Y. Golubev, *On entropy estimation by m-spacing method*, J. Math. Sci. **163** (2009), 290–309.

[8] M.N. Goria, N.N. Leonenko, V.V. Mergel, P.L. Novi Inverardi, *A new class of random vector entropy estimators and its applications in testing statistical hypotheses.* J. Nonparametr. Stat. **17** (2005), 277–297.
[9] P. Hall, *Limit theorems for sums of general functions of m-spacings*, Math. Proc. Cambridge Philos. Soc. **96** (1984), no. 3, 517–532.
[10] P. Hall, *On powerful distributional tests based on sample spacings*, J. Multivariate Statist. **19** (1986), 201–225.
[11] P. Hall, S. Morton, *On the estimation of entropy*, Ann. Inst. Statist. Math. **45** (1993), no. 1, 69–88.
[12] H. Joe, *Estimation of entropy and other functionals of a multivariate density,* Ann. Inst. Statist. Math. **41** (1989), no. 4, 683697.
[13] L.F. Kozachenko, N.N. Leonenko, *A statistical estimate for the entropy of a random vector,* Problemy Peredachi Informatsii **23** (1987), 9–16.
[14] N. Leonenko, L. Pronzato, V. Savani, *A class of Rényi information estimators for multidimensional densities,* Ann. Statist. **36** (2008), 2153–2182.
[15] N. Leonenko, L. Pronzato, *Correction: A class of Rényi information estimators for multidimensional densities,* Ann. Statist. **38** (2010), 3837–3838.
[16] B. Ya. Levit, *Asymptotically efficient estimation of nonlinear functionals,* Problems Inform. Transmission **14** (1978), 65–72.
[17] G.S. Mudholkar, C.T. Lin, *On two applications of characterization theorems to goodness-of-fit,* Goodness-of-fit (Debrecen, 1984), 395–414, Colloq. Math. Soc. János Bolyai, 45, North-Holland, Amsterdam, 1987.
[18] D. Pál, B. Póczos, C. Szepesvári, *Estimation of Rényi Entropy and Mutual Information Based on Generalized Nearest-Neighbor Graphs*, Proceedings of NIPS, 2010.
[19] M.D. Penrose, J.E. Yukich, *Limit theory for point processes in manifolds,* Ann. Appl. Probab. 23 (2013), 2161–2211.
[20] P.M. Robinson, *Consistent nonparametric entropy-based testing*, Rev. Econom. Stud. **58** (1991), no. 3, 437–453.
[21] A.B. Tsybakov, E.C. van der Meulen, *Root-n consistent estimators of entropy for densities with unbounded support,* Scand. J. Statist. **23** (1996), 75–83.
[22] B. van Es, *Estimating functionals related to a density by a class of statistics based on spacings,* Scand. J. Statist. **19** (1992), no. 1, 61–72.
[23] O. Vasicek, *A test for normality based on sample entropy,* J. Roy. Statist. Soc. Ser. B **38** (1976), 54–59.
[24] E.T. Whittaker, G.N. Watson, *A course of modern analysis. An introduction to the general theory of infinite processes and of analytic functions; with an account of the principal transcendental functions.* Reprint of the fourth (1927) edition. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1996.

Sylvain Delattre, Laboratoire de Probabilités et Modèles Aléatoires, UMR 7599, Université Paris Diderot, Case courrier 7012, avenue de France, 75205 Paris Cedex 13, France.
E-mail address: `sylvain.delattre@univ-paris-diderot.fr`

Nicolas Fournier, Laboratoire de Probabilités et Modèles Aléatoires, UMR 7599, Université Pierre-et-Marie Curie, Case 188, 4 place Jussieu, F-75252 Paris Cedex 5, France.
E-mail address: `nicolas.fournier@upmc.fr`