Article

Statistical Estimation of the Kullback–Leibler Divergence

Alexander Bulinski ^{1,*} and Denis Dimitrov ² ¹ Steklov Mathematical Institute of Russian Academy of Sciences, 119991 Moscow, Russia² Faculty of Mechanics and Mathematics, Lomonosov Moscow State University, 119234 Moscow, Russia; denis.dimitrov@math.msu.ru

* Correspondence: bulinski@mech.math.msu.su; Tel.: +7-495-939-14-03

Abstract: Asymptotic unbiasedness and L^2 -consistency are established, under mild conditions, for the estimates of the Kullback–Leibler divergence between two probability measures in $\mathbb{R}^d$, absolutely continuous with respect to (w.r.t.) the Lebesgue measure. These estimates are based on certain k -nearest neighbor statistics for pair of independent identically distributed (i.i.d.) due vector samples. The novelty of results is also in treating mixture models. In particular, they cover mixtures of nondegenerate Gaussian measures. The mentioned asymptotic properties of related estimators for the Shannon entropy and cross-entropy are strengthened. Some applications are indicated.

Keywords: Kullback–Leibler divergence; Shannon differential entropy; statistical estimates; k -nearest neighbor statistics; asymptotic behavior; Gaussian model; mixtures

MSC: 60F25; 62G20; 62H12

Citation: Bulinski, A.; Dimitrov, D. Statistical Estimation of the Kullback–Leibler Divergence. *Mathematics* **2021**, *9*, 544. <https://doi.org/10.3390/math9050544>

Academic Editors: Irina Shevtsova and Victor Korolev

Received: 25 January 2021

Accepted: 28 February 2021

Published: 4 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The Kullback–Leibler divergence introduced in [1] is used for quantification of similarity of two probability measures. It plays important role in various domains such as statistical inference (see, e.g., [2,3]), metric learning [4,5], machine learning [6,7], computer vision [8,9], network security [10], feature selection and classification [11–13], physics [14], biology [15], medicine [16,17], finance [18], among others. It is worth to emphasize that mutual information, widely used in many research directions (see, e.g., [19–23]), is a special case of the Kullback–Leibler divergence for certain measures. Moreover, the Kullback–Leibler divergence itself belongs to a class of f -divergence measures (with $f(t) = \log t$). For comparison of various f -divergence measures see, e.g., [24], their estimates are considered in [25,26].

Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on a measurable space $(S, \mathcal{B})$. The Kullback–Leibler divergence between $\mathbb{P}$ and $\mathbb{Q}$ is defined, according to [1], by way of

$$D(\mathbb{P}||\mathbb{Q}) := \begin{cases} \int_S \log\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right) d\mathbb{P} & \text{if } \mathbb{P} \ll \mathbb{Q}, \\ \infty & \text{otherwise,} \end{cases} \quad (1)$$

where $\frac{d\mathbb{P}}{d\mathbb{Q}}$ stands for the Radon–Nikodym derivative. The integral in (1) can take values in $[0, \infty]$. We employ the base e of logarithms since a constant factor is not essential here.

If $(S, \mathcal{B}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, where $d \in \mathbb{N}$, and (absolutely continuous) $\mathbb{P}$ and $\mathbb{Q}$ have densities, $p(x)$ and $q(x)$, $x \in \mathbb{R}^d$, w.r.t. the Lebesgue measure μ , then (1) can be expressed as

$$D(\mathbb{P}||\mathbb{Q}) = \int_{\mathbb{R}^d} p(x) \log\left(\frac{p(x)}{q(x)}\right) dx, \quad (2)$$

where we write dx instead of $\mu(dx)$ to simplify notation. One formally sets $0/0 := 0$, $a/0 := \infty$ if $a > 0$, $\log 0 := -\infty$, $\log(\infty) := \infty$ and $0 \log 0 := 0$. Then $\log(\frac{p(x)}{q(x)})$ is a measurable function with values in $[-\infty, \infty]$. So, the right-hand sides of (1) and (2) coincide. Formula (2) is justified by Lemma A1, see Appendix A.

Denote by $S(f) := \{x \in \mathbb{R}^d : f(x) > 0\}$ the support of a (version of) probability density f . The integral in (2) is taken over $S(p)$ and does not depend on the choice of p and q versions.

The following two functionals are closely related to the Kullback–Leibler divergence. For probability measures $\mathbb{P}$ and $\mathbb{Q}$ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ having densities $p(x)$ and $q(x)$, $x \in \mathbb{R}^d$, w.r.t. the Lebesgue measure μ , one can introduce, according to [27], p. 35, entropy H (also called the Shannon differential entropy) and cross-entropy C as follows

$$H(\mathbb{P}) := - \int_{\mathbb{R}^d} p(x) \log p(x) dx, \quad C(\mathbb{P}, \mathbb{Q}) := - \int_{\mathbb{R}^d} p(x) \log q(x) dx.$$

In view of (2), $D(\mathbb{P}||\mathbb{Q}) = C(\mathbb{P}, \mathbb{Q}) - H(\mathbb{P})$ whenever the right-hand side is well defined.

Usually one constructs statistical estimates of some characteristics of a stochastic model under consideration relying on a collection of observations. In the pioneering paper [28] the estimator of the Shannon differential entropy was proposed, based on the nearest neighbor statistics. In a series of papers this estimate was studied and applied. Moreover, estimators of the Rényi entropy, mutual information and the Kullback–Leibler divergence have appeared (see, e.g., [29–31]). However, the authors of [32] indicated the occurrence of gaps in the known proofs concerning the limit behavior of such statistics. Almost all of these flaws refer to the lack of proved correctness of using the (reversed) Fatou lemma (see, e.g., [28], inequality after the statement (21), or [31], inequality (91)) or the generalized Helly–Bray lemma (see, e.g., [30], page 2171). One can find these lemmas in [33], p. 233, and [34], p. 187. Paper [32] has attracted our attention and motivated study of the declared asymptotic properties. Furthermore, we would like to highlight the important role of the papers [28,30–32]. Thus, in a recent work [35] the new functionals were introduced to prove asymptotic unbiasedness and L^2 -consistency of the Kozachenko–Leonenko estimators of the Shannon differential entropy. We used the criterion of uniform integrability, for different families of functions, to avoid employment of the Fatou lemma since it is not clear whether one could indicate the due majorizing functions for those families. The present paper is aimed at extension of our approach to grasp the Kullback–Leibler divergence estimation. Instead of the nearest neighbor statistics we employ the k -nearest neighbor statistics (on order statistics see, e.g., [36]) and also use more general forms of the mentioned functionals.

Note in passing that there exist investigations treating important aspects of the entropy, Kullback–Leibler divergence and mutual information estimation. The mixed models and conditional entropy estimation are studied, e.g., in [37,38]. The central limit theorem (CLT) for the Kozachenko–Leonenko estimates is established in [39]. In [40], deep analysis of efficiency of functional weighted estimates was performed (including CLT). The limit theorems for point processes on manifolds are employed in [41] to analyze behavior of the Shannon and the Rényi entropy estimates. The convergence rates for the Shannon entropy (truncated) estimates are obtained in [42] for one-dimensional case, see also [43] for multidimensional case. A kernel density plug-in estimator of the various divergence functionals is studied in [25]. The principal assumptions of that paper are the following: the densities are smooth and have common bounded support S , they are strictly lower bounded on S , moreover, the set S is smooth with respect to the employed kernel. Ensemble estimation of various divergence functionals is studied in [25]. Profound results for smooth bounded densities are established in recent work [44]. The mutual information estimation by the local Gaussian approximation is developed in [45]. Note that various deep results (including the central limit theorem) were obtained for the Kullback–Leibler estimates under certain conditions imposed on derivatives of unknown densities (see, e.g., the recent papers [25,46]). In a series of papers the authors demand boundedness of densities to prove

L^2 -consistency for the Kozachenko–Leonenko estimates of differential Shannon entropy (see, e.g., [47]).

Our goal is to provide wide conditions for the asymptotic unbiasedness and L^2 -consistency of the specified Kullback–Leibler divergence estimates without such smoothness and boundedness hypotheses. Furthermore, we do not assume that densities have bounded supports. As a byproduct we obtain new results concerning Shannon differential entropy and cross-entropy.

We employ probabilistic and analytical techniques, namely, weak convergence of probability measures, conditional expectations, regular probability distributions, k -nearest neighbor statistics, probability inequalities, integration by parts in the Lebesgue–Stieltjes integral, analysis of integrals depending on certain parameters and taken over specified domains, criterion of the uniform integrability of various families of functions, slowly varying functions.

The paper is organized as follows. In Section 2, we introduce some notation. In Section 3 we formulate main results, i.e., Theorems 1 and 2. Their proofs are provided in Sections 4 and 5, respectively. Section 6 contains concluding remarks and perspectives of future research. Proofs of several lemmas are given in Appendix A.

2. Notation

Let X and Y be random vectors taking values in $\mathbb{R}^d$ and having distributions P_X and P_Y , respectively, (below we will take $\mathbb{P} = P_X$ and $\mathbb{Q} = P_Y$). Consider random vectors $X_1, X_2, \dots$ and $Y_1, Y_2, \dots$ with values in $\mathbb{R}^d$ such that $\text{law}(X_i) = \text{law}(X)$ and $\text{law}(Y_i) = \text{law}(Y)$, $i \in \mathbb{N}$. Assume also that $\{X_i, Y_i, i \in \mathbb{N}\}$ are independent. We are interested in statistical estimation of $D(P_X||P_Y)$ constructed by means of observations $\mathbb{X}_n := \{X_1, \dots, X_n\}$ and $\mathbb{Y}_m := \{Y_1, \dots, Y_m\}$, $n, m \in \mathbb{N}$. All the random variables under consideration are defined on a probability space $(\Omega, \mathcal{F}, P)$, each measure space is assumed complete.

For a finite set $E = \{z_1, \dots, z_N\} \subset \mathbb{R}^d$, where $z_i \neq z_j$ ($i \neq j$), and a vector $v \in \mathbb{R}^d$, renumerate points of E as $z_{(1)}(v), \dots, z_{(N)}(v)$ in such a way that $\|v - z_{(1)}\| \leq \dots \leq \|v - z_{(N)}\|$, $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^d$. If there are points $z_{i_1}, \dots, z_{i_s}$ having the same distance from v then we numerate them according to the indexes $i_1, \dots, i_s$ increase. In other words, for $k = 1, \dots, N$, $z_{(k)}(v)$ is the k -nearest neighbor of v in a set E . To indicate that $z_{(k)}(v)$ is constructed by means of E we write $z_{(k)}(v, E)$. Fix $k \in \{1, \dots, n-1\}$, $l \in \{1, \dots, m\}$ and (for each $\omega \in \Omega$) put

$$R_{n,k}(i) := \|X_i - X_{(k)}(X_i, \mathbb{X}_n \setminus \{X_i\})\|, \quad V_{m,l}(i) := \|X_i - Y_{(l)}(X_i, \mathbb{Y}_m)\|, \quad i = 1, \dots, n.$$

We assume that X and Y have densities $p = \frac{dP_X}{d\mu}$ and $q = \frac{dP_Y}{d\mu}$. Then with probability one all the points in $\mathbb{X}_n$ are distinct as well as points of $\mathbb{Y}_m$.

Following [31] (see Formula (17) there) introduce an estimate of $D(P_X||P_Y)$

$$\tilde{D}_{n,m}(\mathcal{K}_n, \mathcal{L}_n) := \frac{1}{n} \sum_{i=1}^n (\psi(k_i) - \psi(l_i)) + \log\left(\frac{m}{n-1}\right) + \frac{d}{n} \sum_{i=1}^n \log\left(\frac{V_{m,l_i}(i)}{R_{n,k_i}(i)}\right), \quad (3)$$

where $\psi(t) = \frac{d}{dt} \log \Gamma(t) = \frac{\Gamma'(t)}{\Gamma(t)}$ is the digamma function, $t > 0$, $\mathcal{K}_n := \{k_i\}_{i=1}^n$, $\mathcal{L}_n := \{l_i\}_{i=1}^n$ are collections of integers and, for some $r \in \mathbb{N}$ and all $i \in \mathbb{N}$, $k_i \leq r$, $l_i \leq r$. Note that (3) is well-defined for $n \geq \max_{i=1, \dots, n} k_i + 1$, $m \geq \max_{i=1, \dots, n} l_i$. If $k_i = k$ and $l_i = l$, $i = 1, \dots, n$, then, for $n \geq k+1$ and $m \geq l$, we write

$$\hat{D}_{n,m}(k, l) := \psi(k) - \psi(l) + \log\left(\frac{m}{n-1}\right) + \frac{d}{n} \sum_{i=1}^n \log\left(\frac{V_{m,l}(i)}{R_{n,k}(i)}\right). \quad (4)$$

If $k = l$ then

$$\hat{D}_{n,m}(k) = \log\left(\frac{m}{n-1}\right) + \frac{d}{n} \sum_{i=1}^n \log\left(\frac{V_{m,k}(i)}{R_{n,k}(i)}\right)$$

and we come to formula (5) in [31]. For an intuitive background of the proposed estimates one can address [31] (Introduction, Parts B and C).

We write $B(x, r) := \{y \in \mathbb{R}^d : \|x - y\| \leq r\}$ for $x \in \mathbb{R}^d$, $r > 0$, and $V_d = \mu(B(0, 1))$ is the volume of the unit ball in $\mathbb{R}^d$. Similar to (3) with the same notation and the same conditions for k_i and l_i , $i = 1, \dots, n$, one can define the Kozachenko - Leonenko type estimates of $H(P_X)$ and $C(P_X, P_Y)$, respectively, by formulas

$$\tilde{H}_n(\mathcal{K}_n) := -\frac{1}{n} \sum_{i=1}^n \psi(k_i) + \log V_d + \log(n-1) + \frac{d}{n} \sum_{i=1}^n \log R_{n,k_i}(i), \quad (5)$$

$$\tilde{C}_{n,m}(\mathcal{L}_n) := -\frac{1}{n} \sum_{i=1}^n \psi(l_i) + \log V_d + \log m + \frac{d}{n} \sum_{i=1}^n \log V_{m,l_i}(i). \quad (6)$$

In [28], an estimate (5) was proposed for $k_i = 1$, $i = 1, \dots, n$. If $k_i = k$, $l_i = l$, $i = 1, \dots, n$, $n \geq k+1$ and $m \geq l$, then one has

$$\hat{H}_n(k) := \frac{1}{n} \sum_{i=1}^n \log \left(\frac{V_d R_{n,k}^d(i)(n-1)}{e^{\psi(k)}} \right), \quad \hat{C}_{n,m}(l) := \frac{1}{n} \sum_{i=1}^n \log \left(\frac{V_d V_{m,l}^d(i)m}{e^{\psi(l)}} \right). \quad (7)$$

Remark 1. All our results are valid for statistics (3). To simplify notation we consider estimates (4) since the study of $\tilde{D}_{n,m}(\mathcal{K}_n, \mathcal{L}_n)$ follows the same lines. For the same reason, as in the case of Kullback–Leibler divergence, we will only deal with (7) since (5) and (6) can be analyzed in quite the same way.

Some extra notation is necessary. As in [35], given a probability density f in $\mathbb{R}^d$, we consider the following functions of $x \in \mathbb{R}^d$, $r > 0$ and $R > 0$, that is, define integral functionals (depending on parameters)

$$I_f(x, r) := \frac{\int_{B(x,r)} f(y) dy}{r^d V_d}, \quad (8)$$

$$M_f(x, R) := \sup_{r \in (0, R]} I_f(x, r), \quad m_f(x, R) := \inf_{r \in (0, R]} I_f(x, r). \quad (9)$$

Some properties of function $\int_{B(x,r)} f(y) dy$ are demonstrated in [48]. By virtue of Lemma 2.1 [35], for each probability density f , the function $I_f(x, r)$ introduced above is continuous in (x, r) on $\mathbb{R}^d \times (0, \infty)$. Hence on account of Theorem 15.84 [49] the functions $m_f(\cdot, R)$ and $M_f(\cdot, R)$ for any $R > 0$ have to be upper semicontinuous and lower semicontinuous, respectively. Therefore, Borel measurability of these nonnegative functions ensues from Proposition 15.82 [49]. On the other hand, the function $m_f(x, \cdot)$ is evidently nonincreasing whereas $M_f(x, \cdot)$ is nondecreasing for each x in $\mathbb{R}^d$. Notably, changing $\sup_{r \in (0, R]}$ to $\sup_{r \in (0, \infty)}$ transforms the function $M_f(x, R)$ into the famous Hardy–Littlewood maximal function $M_f(x)$ well-known in Harmonic analysis.

Set $e_{[-1]} := 0$ and $e_{[N]} := \exp\{e_{[N-1]}\}$, $N \in \mathbb{Z}_+$. Introduce a function $\log_{[1]}(t) := \log t$, $t > 0$. For $N \in \mathbb{N}$, $N > 1$, set $\log_{[N]}(t) := \log(\log_{[N-1]}(t))$. Evidently, this function (for each $N \in \mathbb{N}$) is defined if $t > e_{[N-2]}$. For $N \in \mathbb{N}$, consider the continuous nondecreasing function $G_N : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, given by formula

$$G_N(t) := \begin{cases} 0, & t \in [0, e_{[N-1]}], \\ t \log_{[N]}(t), & t \in (e_{[N-1]}, \infty). \end{cases} \quad (10)$$

In other words we employ the function having the form $t r(t)$ where a function $r(t)$, taken as N iterations of $\log t$, is slowly varying for large t .

For probability densities p, q in $\mathbb{R}^d$, $N \in \mathbb{N}$ and positive constants ν, t, ε, R , introduce the functionals taking values in $[0, \infty]$

$$K_{p,q}(\nu, N, t) := \int_{x,y \in \mathbb{R}^d, \|x-y\| > t} G_N(\|\log\|x-y\|^\nu) p(x)q(y) dx dy, \quad (11)$$

$$Q_{p,q}(\varepsilon, R) := \int_{\mathbb{R}^d} M_q^\varepsilon(x, R) p(x) dx, \quad (12)$$

$$T_{p,q}(\varepsilon, R) := \int_{\mathbb{R}^d} m_q^{-\varepsilon}(x, R) p(x) dx. \quad (13)$$

Set $K_{p,q}(\nu, N) := K_{p,q}(\nu, N, e_{[N]})$.

Remark 2. We have stipulated that $1/0 := \infty$ (thus $m_q^{-\varepsilon}(x, R) := \infty$ whenever $m_q(x, R) = 0$). One can write in (12), (13) the integrals over the support $S(p)$ instead of integrating over $\mathbb{R}^d$, whatever the versions of p and q are taken.

3. Main Results

Theorem 1. Let, for some positive ε, R and $N \in \mathbb{N}$, the functionals $K_{p,f}(1, N)$, $Q_{p,f}(\varepsilon, R)$, $T_{p,f}(\varepsilon, R)$ be finite if $f = p$ and $f = q$. Then $D(P_X||P_Y) < \infty$ and

$$\lim_{n,m \rightarrow \infty} E\widehat{D}_{n,m}(k, l) = D(P_X||P_Y). \quad (14)$$

Consider 3 kinds of conditions (labeled A,B,C, possibly with indices, and involving parameters indicated in parentheses) on probability densities.

(A; p, f, ν) For probability densities p, f in $\mathbb{R}^d$ and some positive ν

$$L_{p,f}(\nu) := \int_{\mathbb{R}^d \times \mathbb{R}^d} \|\log\|x-y\|^\nu p(x)f(y) dx dy < \infty. \quad (15)$$

As usual, $\int_A g(z)Q(dz) = 0$ whenever $g(z) = \infty$ (or $-\infty$) for $z \in A$ and $Q(A) = 0$, where Q is a σ -finite measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Condition (15) with $\nu > 1$ is used, e.g., in [28,31,47].

(B₁; f) A version of f is upper bounded by a positive number $M(f) \in (0, \infty)$:

$$f(x) \leq M(f), \quad x \in \mathbb{R}^d.$$

(C₁; f) A version of f is lower bounded by a positive number $m(f) \in (0, \infty)$:

$$f(x) \geq m(f), \quad x \in S(f).$$

Corollary 1. Let, for some $\nu > 1$, condition (A; p, f, ν) be satisfied when $f = p$ and $f = q$. Then the statements of Theorem 1 are true, provided that (B₁; f) and (C₁; f) are both valid for $f = p$ and $f = q$. Moreover, if the latter assumption involving (B₁; f) and (C₁; f) holds then conditions of Theorem 1 are satisfied whenever p and q have bounded supports.

Next we formulate conditions to guarantee L^2 -consistency of estimates (4).

Theorem 2. Let the requirement $K_{p,f}(1, N) < \infty$ in conditions of Theorem 1 be replaced by $K_{p,f}(2, N) < \infty$, given $f = p$ and $f = q$. Then $D(P_X||P_Y) < \infty$ and, for any fixed $k, l \in \mathbb{N}$, the estimates $\widehat{D}_{n,m}(k, l)$ are L^2 -consistent, i.e.,

$$\lim_{n,m \rightarrow \infty} E\left(\widehat{D}_{n,m}(k, l) - D(P_X||P_Y)\right)^2 = 0. \quad (16)$$

Corollary 2. For some $\nu > 2$, let condition (A; p, f, ν) be satisfied if $f = p$ and $f = q$. Assume that (B₁; f) and (C₁; f) are both valid for $f = p$ and $f = q$. Then the statements of Theorem 2

are true. Moreover, if the latter assumption involving $(B_1; f)$ and $(C_1; f)$ holds then conditions of Theorem 2 are satisfied whenever p and q have bounded supports.

Currently we dwell on a modification of condition $(C_1; f)$ introduced in [35] that allows us to work with densities that need not have bounded supports.

$(C_2; f)$ There exist a version of density f and $R > 0$ such that, for some $c > 0$,

$$m_f(x, R) \geq cf(x), \quad x \in \mathbb{R}^d.$$

Remark 3. If, for some positive ε , R and c , condition $(C_2; q)$ is true and

$$\int_{\mathbb{R}^d} q(x)^{-\varepsilon} p(x) dx < \infty, \quad (17)$$

then $T_{p,q}(\varepsilon, R)$ is finite. Hence we could apply, for $f = p$ and $f = q$ in Theorems 1 and 2, condition $(C_2; f)$ and presume, for some $\varepsilon > 0$, validity of (17) and finiteness of $\int_{\mathbb{R}^d} p^{1-\varepsilon}(x) dx$ instead of the corresponding assumptions $T_{p,q}(\varepsilon, R) < \infty$ and $T_{p,p}(\varepsilon, R) < \infty$. An illustrative example to this point is provided with a density having unbounded support.

Corollary 3. Let X, Y be Gaussian random vectors in $\mathbb{R}^d$ with $EX = \mu_X$, $EY = \mu_Y$ and nondegenerate covariance matrices Σ_X and Σ_Y , respectively. Then relations (14) and (16) hold where

$$D(P_X || P_Y) = \frac{1}{2} \left(\text{tr}(\Sigma_Y^{-1} \Sigma_X) + (\mu_Y - \mu_X)^T \Sigma_Y^{-1} (\mu_Y - \mu_X) - d + \log \left(\frac{\det \Sigma_Y}{\det \Sigma_X} \right) \right).$$

The latter formula can be found, e.g., in [2], p. 147, example 6.3. The proof of Corollary 3 is discussed in Appendix A.

Similarly to condition $(C_2; f)$ let us consider the following one.

$(B_2; f)$ There exist a version of density f and $R > 0$ such that, for some $C > 0$,

$$M_f(x, R) \leq Cf(x), \quad x \in S(f).$$

Remark 4. If, for some positive ε , R and c , condition $(B_2; q)$ is true and

$$\int_{\mathbb{R}^d} q(x)^\varepsilon p(x) dx < \infty, \quad (18)$$

then obviously $Q_{p,q}(\varepsilon, R) < \infty$. Thus, in Theorems 1 and 2 one can employ, for $f = p$ and $f = q$, condition $(B_2; f)$ and exploit, for some $\varepsilon > 0$, the validity of (18) and finiteness of $\int_{\mathbb{R}^d} p^{1+\varepsilon}(x) dx$ instead of the assumptions $Q_{p,q}(\varepsilon, R) < \infty$ and $Q_{p,p}(\varepsilon, R) < \infty$, respectively.

Remark 5. D.Evans applied “positive density condition” in Definition 2.1 of [48] assuming the existence of constants $\beta > 1$ and $\delta > 0$ such that $\frac{r^d}{\beta} \leq \int_{B(x,r)} q(y) dy \leq \beta r^d$ for all $0 \leq r \leq \delta$ and $x \in \mathbb{R}^d$. Consequently $m_q(x, \delta) \geq \frac{1}{\beta V_d} := m > 0$, $x \in \mathbb{R}^d$. Then $T_{p,q}(\varepsilon, \delta) \leq m^{-\varepsilon} \int_{\mathbb{R}^d} p(x) dx = m^{-\varepsilon} < \infty$ for all $\varepsilon > 0$. Analogously, $M_q(x, \delta) \leq \frac{\beta}{V_d} := M, M > 0, x \in \mathbb{R}^d$, and $Q_{p,q}(\varepsilon, \delta) \leq M^\varepsilon \int_{\mathbb{R}^d} p(x) dx = M^\varepsilon < \infty$ for all $\varepsilon > 0$. The above mentioned inequalities from Definition 2.1 of [48] are valid, provided that density f is smooth and its support in $\mathbb{R}^d$ is a convex closed body, see proof in [50]. Therefore, if p and q are smooth and their supports are compact convex bodies in $\mathbb{R}^d$, the relations (14) and (16) are valid.

Moreover, as a byproduct of Theorems 1 and 2, we obtain the new results indicating both the asymptotic unbiasedness and L^2 -consistency of the estimates (7) for the Shannon differential entropy and cross-entropy.

Theorem 3. Let $Q_{p,q}(\varepsilon, R) < \infty$ and $T_{p,q}(\varepsilon, R) < \infty$ for some positive ε and R . Then $C(P_X, P_Y)$ is finite and the following statements hold for any fixed $l \in \mathbb{N}$.

- (1) If, for some $N \in \mathbb{N}$, $K_{p,q}(1, N) < \infty$, then $E\hat{C}_{n,m}(l) \rightarrow C(P_X, P_Y)$, $n, m \rightarrow \infty$.
 (2) If, for some $N \in \mathbb{N}$, $K_{p,q}(2, N) < \infty$, then $E(\hat{C}_{n,m}(l) - C(P_X, P_Y))^2 \rightarrow 0$, $n, m \rightarrow \infty$.

In particular, one can employ $L_{p,q}(\nu)$ with $\nu > 1$ instead of $K_{p,q}(1, N)$, and with $\nu > 2$ instead of $K_{p,q}(2, N)$, where $N \in \mathbb{N}$.

The first claim of this Theorem follows from the proof of Theorem 1. In a similar way one can infer the second statement from the proof of Theorem 2. If we take $q = p$ in conditions of Theorem 3 then we get the statement concerning the entropy since $C(P_X, P_X) = H(P_X)$.

Now we consider the case when p and q are mixtures of some probability densities. Namely,

$$p(x) := \sum_{i=1}^I a_i p_i(x), \quad q(x) := \sum_{j=1}^J b_j q_j(x), \quad (19)$$

where $p_i(x)$, $q_j(x)$ are probability densities (w.r.t. the Lebesgue measure μ), positive weights a_i , b_j are such that $\sum_{i=1}^I a_i = 1$, $\sum_{j=1}^J b_j = 1$, $i = 1, \dots, I$, $j = 1, \dots, J$, $x \in \mathbb{R}^d$. Some applications of models described by mixtures are treated, e.g., in [51].

Corollary 4. Let random vectors X and Y have densities of the form (19). Assume that, for some positive ε , R and $N \in \mathbb{N}$, the functionals $K_{f,g}(1, N)$, $Q_{f,g}(\varepsilon, R)$, $T_{f,g}(\varepsilon, R)$ are finite, whenever $f \in \{p_1, \dots, p_I\}$ and $g \in \{p_1, \dots, p_I, q_1, \dots, q_J\}$. Then $D(P_X || P_Y) < \infty$ and, for any fixed $k, l \in \mathbb{N}$, (14) holds. Moreover, if the requirement $K_{f,g}(1, N) < \infty$ is replaced by $K_{f,g}(2, N) < \infty$ then (16) is true.

The proof of this Corollary is given in Appendix A. Thus, due to Corollaries 3 and 4 one can guarantee the validity of (14) and (16) for any mixtures of nondegenerate Gaussian densities. Note also that in a similar way we can claim the asymptotic unbiasedness and L^2 -consistency of estimates (7) for mixtures satisfying conditions of Corollary 4.

Remark 6. Let us compare our new results with those established in [35]. Developing the approach of [35] to analysis of asymptotic behavior of the Kozachenko–Leonenko estimates of the Shannon differential entropy we encounter new complications due to dealing with k -nearest neighbor statistics for $k \in \mathbb{N}$ (not only for $k = 1$). Accordingly, in the framework of the Kullback–Leibler divergence estimation, we propose a new way to bound the function $1 - F_{m,l,x}(u)$ playing the key role in the proofs (see Formula (28)). Furthermore, instead of the function $G(t) = t \log t$ (for $t > 1$), used in [35] for the Shannon entropy estimates, we employ a regularly varying function $G_N(t) = t \log_{[N]}(t)$ where (for t large enough) $\log_{[N]}(t)$ is the N -fold iteration of the logarithmic function and $N \in \mathbb{N}$ can be large. Whence in the definition of integral functional $K_{p,q}(\nu, N, t)$ by formula (11) one can take a function $G_N(z)$ having, for $z > 0$, the growth rate close to that of function z . Moreover, this entails a generalization of paper [35] results. Now we invoke convexity of G_N (see Lemma 6) to provide more general conditions for asymptotic unbiasedness and L^2 -consistency of the Shannon differential entropy as opposed to [35].

4. Proof of Theorem 1

Note that the general structure of this proof, as well as that of Theorem 2, is similar to the one originally proposed in [28] and later used in various papers (see, e.g., [30,31,47]). Nevertheless in order to prove both theorems correctly we employ new ideas and conditions (such as uniform integrability of a family of random variables) in our reasoning.

Remark 7. In the proof, for certain random variables $\alpha, \alpha_1, \alpha_2, \dots$ (depending on some parameters), we will demonstrate that $E\alpha_n \rightarrow E\alpha$, as $n \rightarrow \infty$ (and that all these expectations are finite). To this end, for a fixed $\mathbb{R}^d$ -valued random vector τ and each $x \in A$, where A is a specified subset of $\mathbb{R}^d$, we will prove that

$$E(\alpha_n | \tau = x) \rightarrow E(\alpha | \tau = x), \quad n \rightarrow \infty. \quad (20)$$

It turns out that $E(\alpha_n|\tau = x) = E(\alpha_{n,x})$ and $E(\alpha|\tau = x) = E\alpha_x$, where the auxiliary random variables $\alpha_{n,x}$ and α_x can be constructed explicitly for each $x \in \mathbb{R}^d$. Moreover, it is possible to show that, for each $x \in A$, one has $\alpha_{n,x} \xrightarrow{\text{law}} \alpha_x$, $n \rightarrow \infty$. Thus, to prove (20) the Fatou lemma is not used, it is not evident whether there exists a random variable majorizing those under consideration. Instead we verify, for each $x \in A$, the uniform integrability (w.r.t. measure P) of a family $(\alpha_{n,x})_{n \geq n_0(x)}$. Here we employ the necessary and sufficient conditions of uniform integrability provided by de la Vallée–Poussin theorem (see, e.g., Theorem 1.3.4 in [52]). After that, to prove the desired relation $E\alpha_n \rightarrow E\alpha$, $n \rightarrow \infty$, we have a new task. Namely, we check the uniform integrability of a family $(E(\alpha_n|\tau = x))_{n \geq k_0}$, where $x \in A$, w.r.t. the measure P_τ , i.e., the law of τ , and k_0 does not depend of x . Then we can prove that

$$E\alpha_n = \int_A E(\alpha_n|\tau = x)P_\tau(dx) \rightarrow \int_A E(\alpha|\tau = x)P_\tau(dx) = E\alpha, \quad n \rightarrow \infty.$$

Further we will explain a number of nontrivial details concerning the proofs of uniform integrability of various families, the choice of the mentioned random variables (vectors), the set A , $n_0(x)$ and k_0 .

The first auxiliary result explains why without loss of generality (w.l.g.) we can consider the same parameters ε, R, N for different functionals in conditions of Theorems 1 and 2.

Lemma 1. Let p and q be any probability densities in $\mathbb{R}^d$. Then the following statements are valid.

- (1) If $K_{p,q}(\nu_0, N_0) < \infty$ for some $\nu_0 > 0$ and $N_0 \in \mathbb{N}$ then $K_{p,q}(\nu, N) < \infty$ for any $\nu \in (0, \nu_0]$ and each $N \geq N_0$.
- (2) If $Q_{p,q}(\varepsilon_1, R_1) < \infty$ for some $\varepsilon_1 > 0$ and $R_1 > 0$ then $Q_{p,q}(\varepsilon, R) < \infty$ for any $\varepsilon \in (0, \varepsilon_1]$ and each $R > 0$.
- (3) If $T_{p,q}(\varepsilon_2, R_2) < \infty$ for some $\varepsilon_2 > 0$ and $R_2 > 0$ then $T_{p,q}(\varepsilon, R) < \infty$ for any $\varepsilon \in (0, \varepsilon_2]$ and each $R > 0$.

In particular one can take $q = p$ and the statements of Lemma 1 still remain valid. The proof of Lemma 1 is given in Appendix A.

Remark 8. The results of Lemma 1 allow us to ensure (14) by demanding the finiteness of the functionals $K_{p,q}(1, N_1)$, $Q_{p,q}(\varepsilon_1, R_1)$, $T_{p,q}(\varepsilon_2, R_2)$, $K_{p,p}(1, N_2)$, $Q_{p,p}(\varepsilon_3, R_3)$, $T_{p,p}(\varepsilon_4, R_4)$, for some $\varepsilon_i > 0$, $R_j > 0$ and $N_j \in \mathbb{N}$, where $i = 1, 2, 3, 4$ and $j = 1, 2$. Moreover, if we assume the finiteness of $K_{p,q}(2, N_3)$ and $K_{p,p}(2, N_4)$, for some $N_3 \in \mathbb{N}$, $N_4 \in \mathbb{N}$, instead of the finiteness of $K_{p,q}(2, N_1)$ and $K_{p,p}(2, N_2)$ then (16) holds.

According to Remark 2.4 of [35] if, for some positive ε, R , the integrals $Q_{p,q}(\varepsilon, R)$, $T_{p,q}(\varepsilon, R)$, $Q_{p,p}(\varepsilon, R)$, $T_{p,p}(\varepsilon, R)$ are finite then

$$\int_{\mathbb{R}^d} p(x)|\log q(x)|dx < \infty, \quad \int_{\mathbb{R}^d} p(x)|\log p(x)|dx < \infty. \quad (21)$$

Therefore $D(p||q) < \infty$ (and thus $P_X \ll P_Y$ in view of Lemma A1).

For $n \in \mathbb{N}$ such that $n > 1$, for fixed $k \in \mathbb{N}$ and $m \in \mathbb{N}$, where $1 \leq k \leq n-1$, $1 \leq l \leq m$ and $i = 1, \dots, n$, set $\phi_{m,l}(i) := mV_{m,l}^d(i)$, $\zeta_{n,k}(i) := (n-1)R_{n,k}^d(i)$. Then we can rewrite the estimate $\widehat{D}_{n,m}(k, l)$ as follows:

$$\widehat{D}_{n,m}(k, l) = \psi(k) - \psi(l) + \frac{1}{n} \sum_{i=1}^n (\log \phi_{m,l}(i) - \log \zeta_{n,k}(i)). \quad (22)$$

It is sufficient to prove the following two assertions.

Statement 1. For each fixed l , all m large enough and any $i \in \mathbb{N}$, $E|\log \phi_{m,l}(i)|$ is finite.

Moreover,

$$\mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n \log \phi_{m,l}(i) \right) = \mathbb{E} \log \phi_{m,l}(1) \rightarrow \psi(l) - \log V_d - \int_{\mathbb{R}^d} p(x) \log q(x) dx, \quad m \rightarrow \infty. \quad (23)$$

Statement 2. For each fixed k , all n large enough and any $i \in \mathbb{N}$, $\mathbb{E} |\log \zeta_{n,k}(i)|$ is finite.

Moreover,

$$\mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n \log \zeta_{n,k}(i) \right) = \mathbb{E} \log \zeta_{n,k}(1) \rightarrow \psi(k) - \log V_d - \int_{\mathbb{R}^d} p(x) \log p(x) dx, \quad n \rightarrow \infty. \quad (24)$$

Then in view of (2) and (21)–(24)

$$\mathbb{E} \widehat{D}_{n,m}(k, l) \rightarrow - \int_{\mathbb{R}^d} p(x) \log q(x) dx + \int_{\mathbb{R}^d} p(x) \log p(x) dx = D(P_X || P_Y), \quad n, m \rightarrow \infty,$$

and Theorem 1 will be proved.

Recall that, as explained in [35], for a nonnegative random variable V (thus $0 \leq EV \leq \infty$) and any random $\mathbb{R}^d$ -valued vector, one has

$$EV = \int_{\mathbb{R}^d} \mathbb{E}(V|X=x) P_X(dx). \quad (25)$$

This signifies that both sides of (25) coincide, being finite or infinite simultaneously. Let $F(u, \omega)$ be a regular conditional distribution function of a nonnegative random variable U given X where $u \in \mathbb{R}$ and $\omega \in \Omega$. Let h be a measurable function such that $h : \mathbb{R} \rightarrow [0, \infty)$. It was also explained in [35] that, for P_X -almost all $x \in \mathbb{R}^d$, it follows (without assuming $Eh(U) < \infty$)

$$\mathbb{E}(h(U)|X=x) = \int_{[0, \infty)} h(u) dF(u, x). \quad (26)$$

This means that both sides of (26) are finite or infinite simultaneously and coincide.

By virtue of (25) and (26) one can establish that $\mathbb{E} |\log \phi_{m,l}(i)| < \infty$, for all m large enough, fixed l and for all i , and that (23) holds. To perform this take $U = \phi_{m,l}(i)$, $X = X_i$, $h(u) = |\log u|$, $u > 0$ (we use $h(u) = \log^2 u$ in the proof of Theorem 2) and $V = h(U)$. If $h : \mathbb{R} \rightarrow \mathbb{R}$ and $\mathbb{E} |h(U)| < \infty$ then (26) is true as well. To avoid increasing the volume of this paper we will only examine the evaluation of $\mathbb{E} \log \phi_{m,l}(i)$ as all the steps of the proof will be the same when treating $\mathbb{E} |\log \phi_{m,l}(i)|$.

The proof of Statement 1 is partitioned into 4 steps. The first three demonstrate that there is a measurable $A \subset S(p)$, depending on p and q versions, such that $P_X(S(p) \setminus A) = 0$ and, for any $x \in A$, $i \in \mathbb{N}$, the following relation holds:

$$\mathbb{E}(\log \phi_{m,l}(i) | X_i = x) = \mathbb{E}(\log \phi_{m,l}(1) | X_1 = x) \rightarrow \psi(l) - \log V_d - \log q(x), \quad m \rightarrow \infty. \quad (27)$$

The last Step 4 justifies the desired result (23). Finally Step 5 validates Statement 2.

Step 1. Here we establish the distribution convergence for the auxiliary random variables. Fix any $i \in \mathbb{N}$ and $l \in \{1, \dots, m\}$. To simplify notation we do not indicate the dependence of functions on d . For $x \in \mathbb{R}^d$ and $u \geq 0$, we identify the asymptotic behavior (as $m \rightarrow \infty$) of the function

$$\begin{aligned} F_{m,l,x}^i(u) &:= P(\phi_{m,l}(i) \leq u | X_i = x) = P(mV_{m,l}^d(i) \leq u | X_i = x) \\ &= 1 - P\left(V_{m,l}(i) > \left(\frac{u}{m}\right)^{\frac{1}{d}} \middle| X_i = x\right) = 1 - P\left(\|x - Y_{(l)}(x, \mathbb{Y}_m)\| > \left(\frac{u}{m}\right)^{\frac{1}{d}}\right) \\ &= 1 - \sum_{s=0}^{l-1} \binom{m}{s} (W_{m,x}(u))^s (1 - W_{m,x}(u))^{m-s} := P(\zeta_{m,l,x} \leq u), \end{aligned} \quad (28)$$

where

$$W_{m,x}(u) := \int_{B(x,r_m(u))} q(z) dz, \quad r_m(u) := \left(\frac{u}{m}\right)^{\frac{1}{d}}, \quad \xi_{m,l,x} := m \|x - Y_{(l)}(x, \mathbb{Y}_m)\|^d. \quad (29)$$

We take into account in (28) that random vectors $Y_1, \dots, Y_m, X_i$ are independent and condition that $Y_1, \dots, Y_m$ have the same law as Y . We also noted that an event $\left\{\|x - Y_{(l)}(x, \mathbb{Y}_m)\| > r_m(u)\right\}$ is a union of pair-wise disjoint events $A_s, s = 0, \dots, l-1$. Here A_s means that exactly s observations among $\mathbb{Y}_m$ belong to the ball $B(x, r_m(u))$ and other $m-s$ are outside this ball (probability that Y belongs to the sphere $\{z \in \mathbb{R}^d : \|z - x\| = r\}$ equals 0 since Y has a density w.r.t. the Lebesgue measure μ). Formulas (28) and (29) show that $F_{m,l,x}^i(u)$ is the regular conditional distribution function of $\phi_{m,l}(i)$ given $X_i = x$. Moreover, (28) means that $\phi_{m,l}(i), i \in \{1, \dots, n\}$ are identically distributed and we may omit the dependence on i . So, one can write $F_{m,l,x}(u)$ instead of $F_{m,l,x}^i(u)$.

According to the Lebesgue differentiation theorem (see, e.g., [49], p. 654) if $q \in L^1(\mathbb{R}^d)$, for μ -almost all $x \in \mathbb{R}^d$, one has

$$\lim_{r \rightarrow 0+} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} |q(z) - q(x)| dz = 0. \quad (30)$$

Let $\Lambda(q)$ denote the set of Lebesgue points of a function q , namely the points in $\mathbb{R}^d$ satisfying (30). Evidently it depends on the choice of version within the class of functions in $L^1(\mathbb{R}^d)$ equivalent to q , and, for an arbitrary version of q , we have $\mu(\mathbb{R}^d \setminus \Lambda(q)) = 0$.

Clearly, for each $u \geq 0$, $r_m(u) \rightarrow 0$ as $m \rightarrow \infty$, and $\mu(B(x, r_m(u))) = V_d(r_m(u))^d = \frac{V_d u}{m}$. Therefore by virtue of (30), for any fixed $x \in \Lambda(q)$ and $u \geq 0$,

$$W_{m,x}(u) = \frac{V_d u}{m} (q(x) + \alpha_m(x, u)),$$

where $\alpha_m(x, u) \rightarrow 0$, $m \rightarrow \infty$. Hence, for $x \in \Lambda(q) \cap S(q)$ (thus $q(x) > 0$), due to (28)

$$F_{m,l,x}(u) \rightarrow 1 - \sum_{s=0}^{l-1} \frac{(V_d q(x) u)^s}{s!} e^{-V_d q(x) u} := F_{l,x}(u), \quad m \rightarrow \infty. \quad (31)$$

Relation (31) means that

$$\xi_{m,l,x} \xrightarrow{law} \xi_{l,x}, \quad x \in \Lambda(q) \cap S(q), \quad m \rightarrow \infty,$$

where $\xi_{l,x}$ has the Gamma distribution $\Gamma(\alpha, \lambda)$ with parameters $\alpha = V_d q(x)$ and $\lambda = l$.

For any $x \in S(q)$, one can assume w.l.g. that the random variables $\xi_{l,x}$ and $\{\xi_{m,l,x}\}_{m \geq l}$ are defined on a probability space $(\Omega, \mathcal{F}, P)$. Indeed, by the Lomnicki–Ulam theorem (see, e.g., [53], p. 93) the independent copies of $Y_1, Y_2, \dots$ and $\{\xi_{l,x}\}_{x \in S(q)}$ exist on a certain probability space. The convergence in distribution of random variables survives under continuous mapping. Thus, for any $x \in \Lambda(q) \cap S(q)$, we see that

$$\log \xi_{m,l,x} \xrightarrow{law} \log \xi_{l,x}, \quad m \rightarrow \infty. \quad (32)$$

We have employed that $\xi_{l,x} > 0$ a.s. for each $x \in \Lambda(q) \cap S(q)$ and Y has a density, so it follows that $P(\xi_{m,l,x} > 0) = P(\|x - Y_{(l)}(x, \mathbb{Y}_m)\| > 0) = 1$. More precisely, we take strictly positive versions of $\xi_{l,x}$ and $\xi_{m,l,x}$ for each $x \in \Lambda(q) \cap S(q)$.

Step 2. Now we show that, instead of (27) validity, one can verify the following assertion. For μ -almost every $x \in \Lambda(q) \cap S(q)$

$$E \log \xi_{m,l,x} \rightarrow E \log \xi_{l,x}, \quad m \rightarrow \infty. \quad (33)$$

Note that if $\eta \sim \Gamma(\alpha, \lambda)$, where $\alpha > 0$ and $\lambda > 0$, then $E \log \eta = \psi(\lambda) - \log \alpha$, where ψ is a digamma function. Set $\alpha = V_d q(x)$ for $x \in S(q)$ (then $\alpha > 0$) and $\lambda = l$. Hence $E \log \xi_{l,x} = \psi(l) - \log(V_d q(x)) = \psi(l) - \log V_d - \log q(x)$. By virtue of (26), for each $x \in \mathbb{R}^d$,

$$\begin{aligned} E \log \xi_{m,l,x} &= \int_{(0,\infty)} \log u \, dF_{m,l,x}(u) = \int_{(0,\infty)} \log u \, dP(\phi_{m,l}(1) \leq u | X_1 = x) \\ &= E(\log \phi_{m,l}(1) | X_1 = x). \end{aligned}$$

Hence, for $x \in \Lambda(q) \cap S(q)$, the relation $E(\log \phi_{m,l}(1) | X_1 = x) \rightarrow \psi(l) - \log V_d - \log q(x)$ holds if and only if (33) is true.

According to Theorem 3.5 [54] we would have established (33) if relation (32) could be supplemented, for μ -almost all $x \in \Lambda(q) \cap S(q)$, by the condition of uniform integrability of a family $\{\log \xi_{m,l,x}\}_{m \geq m_0(x)}$. Note that, for each $N \in \mathbb{N}$, a function $G_N(t)$ introduced by (10) is nondecreasing on $(0, \infty)$ and $\frac{G_N(t)}{t} \rightarrow \infty$, as $t \rightarrow \infty$. By the de la Vallée–Poussin theorem (see, e.g., Theorem 1.3.4 [52]), to ensure, for μ -almost every $x \in \Lambda(q) \cap S(q)$, the uniform integrability of $\{\log \xi_{m,l,x}\}_{m \geq m_0(x)}$, it suffices to prove the following statement. For the indicated x , a positive $C_0(x)$ and $m_0(x) \in \mathbb{N}$, one has

$$\sup_{m \geq m_0(x)} EG_N(|\log \xi_{m,l,x}|) \leq C_0(x) < \infty, \quad (34)$$

where G_N appears in conditions of Theorem 1. Moreover, it is possible to find $m_0 \in \mathbb{N}$ that does not depend on $x \in \mathbb{R}^d$ as we will show further.

Step 3. This step is devoted to proving validity of (34). It is convenient to divide this step into its own parts (3a), (3b), etc. For any $N \in \mathbb{N}$, set

$$g_N(t) = \begin{cases} -\frac{1}{t} \left(\log_{[N]}(-\log t) + \frac{1}{\prod_{j=1}^{N-1} \log_{[j]}(-\log t)} \right), & t \in \left(0, \frac{1}{e_{[N]}}\right], \\ 0, & t \in \left(\frac{1}{e_{[N]}}, e_{[N]}\right], \\ \frac{1}{t} \left(\log_{[N]}(\log t) + \frac{1}{\prod_{j=1}^{N-1} \log_{[j]}(\log t)} \right), & t \in (e_{[N]}, \infty), \end{cases}$$

where the product over empty set (when $N = 1$) is equal to 1.

The proof of the following result is placed at Appendix A.

Lemma 2. Let $F(u)$, $u \in \mathbb{R}$, be a distribution function such that $F(0) = 0$. Then, for each $N \in \mathbb{N}$, one has

$$\begin{aligned} (1) \quad & \int_{\left(0, \frac{1}{e_{[N]}}\right]} G_N(|\log u|) dF(u) = \int_{\left(0, \frac{1}{e_{[N]}}\right]} F(u) (-g_N(u)) du, \\ (2) \quad & \int_{(e_{[N]}, \infty)} G_N(|\log u|) dF(u) = \int_{(e_{[N]}, \infty)} (1 - F(u)) g_N(u) du. \end{aligned}$$

Fix N appearing in conditions of Theorem 1. Observe that, for $u \in \left(\frac{1}{e_{[N]}}, e_{[N]}\right]$, one has $G_N(|\log u|) = 0$. Therefore, according to Lemma 2, for $x \in \Lambda(q) \cap S(q)$ and $m \geq l$, we get $EG_N(|\log \xi_{m,l,x}|) := I_1(m, x) + I_2(m, x)$ where

$$I_1(m, x) := \int_{\left(0, \frac{1}{e_{[N]}}\right]} F_{m,l,x}(u) (-g_N(u)) du, \quad I_2(m, x) := \int_{(e_{[N]}, \infty)} (1 - F_{m,l,x}(u)) g_N(u) du.$$

For convenience sake we write $I_1(m, x)$ and $I_2(m, x)$ without indicating their dependence on N, l and d (these parameters are fixed).

Part (3a). We provide bounds for $I_1(m, x)$. Take $R > 0$ appearing in conditions of Theorem 1 and any $u \in \left(0, \frac{1}{e_{[N]}}\right]$. Introduce $m_1 := \max\left\{\left\lceil \frac{1}{e_{[N]} R^d} \right\rceil, l\right\}$, where, for $a \in \mathbb{R}$, $[a] := \inf\{m \in \mathbb{Z} : m \geq a\}$. Then $r_m(u) = \left(\frac{u}{m}\right)^{1/d} \leq \left(\frac{1}{e_{[N]} m}\right)^{1/d} \leq R$ if $m \geq m_1$. Note also that we can consider only $m \geq l$ everywhere below, because the size of sample $\mathbb{Y}_m$ is not

less than the number of neighbors l (see, e.g., (28)). Thus, for $R > 0$, $u \in \left(0, \frac{1}{e_{[N]}}\right]$, $x \in \mathbb{R}^d$ and $m \geq m_1$,

$$\frac{W_{m,x}(u)}{\mu(B(x, r_m(u)))} = \frac{\int_{B(x, r_m(u))} q(y) dy}{r_m^d(u) V_d} \leq \sup_{r \in (0, R]} \frac{\int_{B(x, r)} q(y) dy}{r^d V_d} = M_q(x, R),$$

and we arrive at the inequality

$$W_{m,x}(u) \leq M_q(x, R) \mu(B(x, r_m(u))) = \frac{M_q(x, R) V_d u}{m}. \quad (35)$$

If $\gamma \in (0, 1]$ and $t \in [0, 1]$ then, for all $m \geq 1$, invoking the Bernoulli inequality, one has

$$1 - (1 - t)^m \leq (mt)^\gamma. \quad (36)$$

Recall that we assume $Q_{p,q}(\varepsilon, R) < \infty$ for some $\varepsilon > 0$, $R > 0$. By virtue of Lemma 1 one can take $\varepsilon < 1$. So, due to (36) and since $W_{m,x}(u) \in [0, 1]$ for all $x \in \mathbb{R}^d$, $u > 0$ and $m \geq l$, we get

$$1 - (1 - W_{m,x}(u))^m \leq (mW_{m,x}(u))^\varepsilon. \quad (37)$$

Thus in view of (28), (35) and (37) we have established that, for all $x \in \Lambda(q) \cap S(q)$, $u \in (0, \frac{1}{e_{[N]}}]$ and $m \geq m_1$,

$$\begin{aligned} F_{m,l,x}(u) &= 1 - \sum_{s=0}^{l-1} \binom{m}{s} (W_{m,x}(u))^s (1 - W_{m,x}(u))^{m-s} \\ &\leq 1 - (1 - W_{m,x}(u))^m \leq \left(m \frac{M_q(x, R) V_d u}{m}\right)^\varepsilon = (M_q(x, R))^\varepsilon V_d^\varepsilon u^\varepsilon. \end{aligned} \quad (38)$$

Therefore, for any $x \in \Lambda(q) \cap S(q)$ and $m \geq m_1$, one can write

$$\begin{aligned} I_1(m, x) &\leq (M_q(x, R))^\varepsilon V_d^\varepsilon \int_{\left(0, \frac{1}{e_{[N]}}\right]} u^\varepsilon (-g_N(u)) du \\ &\leq (M_q(x, R))^\varepsilon V_d^\varepsilon \int_{\left(0, \frac{1}{e_{[N]}}\right]} \frac{\log_{[N]}(-\log u) + 1}{u^{1-\varepsilon}} du = U_1(\varepsilon, N, d) (M_q(x, R))^\varepsilon, \end{aligned} \quad (39)$$

where $U_1(\varepsilon, N, d) := V_d^\varepsilon L_N(\varepsilon)$, $L_N(\varepsilon) := \int_{[e_{[N]}, \infty)} (\log_{[N]}(t) + 1) e^{-\varepsilon t} dt < \infty$. We took into account that $(-g_N(u)) \leq \frac{1}{u} (\log_{[N]}(-\log u) + 1)$ whenever $u \in \left(0, \frac{1}{e_{[N]}}\right]$.

Part (3b). We give bounds for $I_2(m, x)$. Since $g_N(u) \leq \frac{\log_{[N+1]}(u) + 1}{u}$ if $u \in (e_{[N]}, \infty)$, we can write, for $m \geq \max\{e_{[N]}^2, l\}$,

$$\begin{aligned} I_2(m, x) &\leq \int_{(e_{[N]}, \sqrt{m}]} (1 - F_{m,l,x}(u)) \frac{\log_{[N+1]}(u) + 1}{u} du \\ &\quad + \int_{(\sqrt{m}, m^2]} (1 - F_{m,l,x}(u)) \frac{\log_{[N+1]}(u) + 1}{u} du + \int_{(m^2, \infty)} (1 - F_{m,l,x}(u)) g_N(u) du \\ &:= J_1(m, x) + J_2(m, x) + J_3(m, x). \end{aligned}$$

Evidently,

$$1 - F_{m,l,x}(u) = \sum_{r=m-l+1}^m \binom{m}{r} (P_{m,x}(u))^r (1 - P_{m,x}(u))^{m-r} = \mathbb{P}(Z \geq m - l + 1),$$

where $P_{m,x}(u) = 1 - W_{m,x}(u)$ and $Z \sim \text{Bin}(m, P_{m,x}(u))$.

By Markov's inequality $\mathbb{P}(Z \geq t) \leq e^{-\lambda t} \mathbb{E} e^{\lambda Z}$ for any $\lambda > 0$ and $t > 0$. One has

$$\begin{aligned} \mathbb{E}e^{\lambda Z} &= \sum_{j=0}^m e^{\lambda j} \binom{m}{j} (P_{m,x}(u))^j (1 - P_{m,x}(u))^{m-j} \\ &= \sum_{j=0}^m \binom{m}{j} (P_{m,x}(u)e^{\lambda})^j (1 - P_{m,x}(u))^{m-j} = (1 - P_{m,x}(u) + e^{\lambda} P_{m,x}(u))^m. \end{aligned}$$

Consequently, for each $\lambda > 0$,

$$\begin{aligned} 1 - F_{m,l,x}(u) &\leq e^{-\lambda(m-l+1)} (1 - P_{m,x}(u) + e^{\lambda} P_{m,x}(u))^m \\ &= e^{-\lambda(m-l+1)} (W_{m,x}(u) + e^{\lambda}(1 - W_{m,x}(u)))^m = e^{\lambda(l-1)} \left(1 - \left(1 - \frac{1}{e^{\lambda}}\right) W_{m,x}(u)\right)^m. \end{aligned} \quad (40)$$

To simplify bounds we take $\lambda = 1$ and set $S_1 = S_1(l) := e^{l-1}$, $S_2 := 1 - \frac{1}{e}$ (recall that l is fixed). Thus, $S_1 \geq 1$ and $S_2 < 1$. Therefore,

$$1 - F_{m,l,x}(u) \leq S_1(1 - S_2 W_{m,x}(u))^m \leq S_1 \exp\{-S_2 m W_{m,x}(u)\}, \quad (41)$$

where we have used simple inequality $1 - t \leq e^{-t}$, $t \in [0, 1]$.

For $R > 0$ appearing in conditions of the Theorem and any $u \in (e_{[N]}, \sqrt{m}]$, one can choose $m_2 := \max\left\{\left\lceil \frac{1}{R^{2d}} \right\rceil, \left\lceil e_{[N]}^2 \right\rceil, l\right\}$ such that if $m \geq m_2$ then $r_m(u) = \left(\frac{u}{m}\right)^{1/d} \leq \left(\frac{1}{\sqrt{m}}\right)^{1/d} \leq R$. Due to (29) and (41), for $u \in (e_{[N]}, \sqrt{m}]$ and $m \geq m_2$, one has

$$\begin{aligned} 1 - F_{m,l,x}(u) &\leq S_1 \exp\left\{-S_2 m \frac{V_d u}{m} \frac{W_{m,x}(u)}{\frac{V_d u}{m}}\right\} \\ &= S_1 \exp\left\{-S_2 V_d u \frac{\int_{B(x, r_m(u))} q(z) dz}{\mu(B(x, r_m(u)))}\right\} \leq S_1 \exp\{-S_2 V_d u m_q(x, R)\}, \end{aligned} \quad (42)$$

by definition of m_f (for $f = q$) in (9). Now we use the following Lemma 3.2 of [35].

Lemma 3. For a version of a density q and each $R > 0$, one has $\mu(S(q) \setminus D_q(R)) = 0$ where $D_q(R) := \{x \in S(q) : m_q(x, R) > 0\}$ and $m_q(\cdot, R)$ is defined according to (9).

It is easily seen that, for any $t > 0$ and each $\delta \in (0, e]$, one has $e^{-t} \leq t^{-\delta}$. Thus, for $x \in D_q(R)$, $m \geq m_2$, $u \in (e_{[N]}, \sqrt{m}]$ and $\varepsilon > 0$, we deduce from conditions of the Theorem (in view of Lemma 1 one can suppose that $\varepsilon \in (0, e]$) that

$$1 - F_{m,l,x}(u) \leq S_1 (S_2 V_d u m_q(x, R))^{-\varepsilon}. \quad (43)$$

We also took into account that $m_q(x, R) > 0$ for $x \in D_q(R)$ and applied relation (42). Thus, for all $x \in \Lambda(q) \cap S(q) \cap D_q(R)$ and any $m \geq m_2$,

$$\begin{aligned} J_1(m, x) &\leq \frac{S_1}{(S_2 V_d)^{\varepsilon} (m_q(x, R))^{\varepsilon}} \int_{(e_{[N]}, \infty)} \frac{\log_{[N+1]}(u) + 1}{u^{1+\varepsilon}} du \\ &= U_2(\varepsilon, N, d, l) (m_q(x, R))^{-\varepsilon}, \end{aligned} \quad (44)$$

where $U_2(\varepsilon, N, d, l) := S_1(l) L_N(\varepsilon) (S_2 V_d)^{-\varepsilon}$.

Part (3c). We provide the bound for $J_2(m, x)$. For all $x \in \Lambda(q) \cap S(q) \cap D_q(R)$ and any $m \geq m_2$, in view of (43), it holds $1 - F_{m,l,x}(\sqrt{m}) \leq S_1 (S_2 V_d m_q(x, R) \sqrt{m})^{-\varepsilon}$. Hence (as $m_2 \geq 2$)

$$\begin{aligned}
J_2(m, x) &\leq \int_{(\sqrt{m}, m^2]} (1 - F_{m,l,x}(u)) \frac{\log_{[N+1]}(u) + 1}{u} du \\
&\leq (1 - F_{m,l,x}(\sqrt{m})) \int_{(\sqrt{m}, m^2]} (\log_{[N+1]}(u) + 1) d \log u \\
&\leq S_1(S_2 V_d)^{-\varepsilon} (m_q(x, R))^{-\varepsilon} m^{-\frac{\varepsilon}{2}} \left(\log_{[N]}(2 \log m) + 1 \right) \frac{3}{2} \log m.
\end{aligned}$$

Then, for all $x \in \Lambda(q) \cap S(q) \cap D_q(R)$ and any $m \geq m_2$,

$$J_2(m, x) \leq U_3(m, \varepsilon, N, d, l) (m_q(x, R))^{-\varepsilon}, \quad (45)$$

where $U_3(m, \varepsilon, N, d, l) := \frac{3}{2} S_1(l) (S_2 V_d)^{-\varepsilon} m^{-\frac{\varepsilon}{2}} \log m \left(\log_{[N]}(2 \log m) + 1 \right) \rightarrow 0, m \rightarrow \infty$.

Part (3d). To indicate bounds for $J_3(m, x)$ we employ several auxiliary results.

Lemma 4. For each $N \in \mathbb{N}$ and any $\nu > 0$, there are $a := a(d, \nu) \geq 0$, $b := b(N, d, \nu) \geq 0$ such that, for arbitrary $x, y \in \mathbb{R}^d$,

$$G_N(|\log \|x - y\|^d|^\nu) \leq a G_N(|\log \|x - y\|^\nu) + b.$$

The proof is given in Appendix A.

On the one hand, by (29), for any $w \geq 0$, we get

$$W_{m,x}(mw) = \int_{B(x, w^{1/d})} q(z) dz = W_{1,x}(w).$$

On the other hand, by (28), one has $F_{1,1,x}(w) = 1 - (1 - W_{1,x}(w)) = W_{1,x}(w)$. Consequently, for any $m \in \mathbb{N}$, $w \geq 0$ and all $x \in \mathbb{R}^d$,

$$W_{m,x}(mw) = F_{1,1,x}(w). \quad (46)$$

Moreover, $F_{1,1,x}(w) = P(\|Y - x\|^d \leq w)$. So, $\xi_{1,1,x} \stackrel{\text{law}}{=} \|Y - x\|^d$. Thus, due to Lemmas 2 and 4 (for $\nu = 1$)

$$\begin{aligned}
&\int_{(e_{[N]}, \infty)} (1 - F_{1,1,x}(w)) g_N(w) dw = \int_{(e_{[N]}, \infty)} G_N(\log w) dF_{1,1,x}(w) \\
&= \mathbb{E} \left[G_N(\log \xi_{1,1,x}) \mathbb{I} \left\{ \xi_{1,1,x} > e_{[N]} \right\} \right] = \mathbb{E} [G_N(\log \|Y - x\|^d) \mathbb{I} \{ \|Y - x\|^d > e_{[N]} \}] \\
&= \int_{y \in \mathbb{R}^d, \|x-y\| > (e_{[N]})^{1/d}} G_N(\log \|x - y\|^d) q(y) dy \\
&\leq a(d, 1) \int_{y \in \mathbb{R}^d, \|x-y\| > (e_{[N]})^{1/d}} G_N(|\log \|x - y\||) q(y) dy + b(N, d, 1) \\
&= a(d, 1) \int_{y \in \mathbb{R}^d, \|x-y\| > e_{[N]}} G_N(\log \|x - y\|) q(y) dy + b(N, d, 1),
\end{aligned} \quad (47)$$

since $G_N(t) = 0$ for $t \in [0, e_{[N-1]}]$, $N \in \mathbb{N}$.

Now we will estimate $1 - F_{m,l,x}(u)$ in a way different from (40). Fix any $\delta > 0$. Note that, for all $m \geq (l-1) \left(1 + \frac{1}{\delta}\right)$ and $s \in \{0, \dots, l-1\}$, it holds $\frac{m}{m-s} \leq \frac{m}{m-l+1} \leq 1 + \delta$. Then, for all $x \in \mathbb{R}^d$, $u \geq 0$ and $m \geq \max\{l, (l-1) \left(1 + \frac{1}{\delta}\right)\}$, in view of (28) one can write

$$\begin{aligned}
1 - F_{m,l,x}(u) &= (1 - W_{m,x}(u)) \sum_{s=0}^{l-1} \binom{m-1}{s} \frac{m}{m-s} (W_{m,x}(u))^s (1 - W_{m,x}(u))^{(m-1)-s} \\
&\leq (1 + \delta) (1 - W_{m,x}(u)) \sum_{s=0}^{l-1} \binom{m-1}{s} (W_{m,x}(u))^s (1 - W_{m,x}(u))^{(m-1)-s} \\
&\leq (1 + \delta) (1 - W_{m,x}(u)).
\end{aligned} \quad (48)$$

We are going to employ the following statement as well.

Lemma 5. For each $N \in \mathbb{N}$, a function $\log_{[N]}(t)$, $t > e_{[N-1]}$, is slowly varying at infinity.

The proof is elementary and thus is omitted.

Part (3e). Now we are ready to get the bound for $J_3(m, x)$. Set $u = mw$. Then one has

$$\begin{aligned} J_3(m, x) &= \int_{(m^2, \infty)} (1 - F_{m,l,x}(u)) \frac{1}{u} \left(\log_{[N]}(\log u) + \frac{1}{\prod_{j=1}^{N-1} \log_{[j]}(\log u)} \right) du \\ &= \int_{(m, \infty)} (1 - F_{m,l,x}(mw)) \frac{1}{w} \left(\log_{[N+1]}(mw) + \frac{1}{\prod_{j=2}^N \log_{[j]}(mw)} \right) dw. \end{aligned}$$

Given $w > m$, Lemma 5 implies that $\log_{[N+1]}(mw) \leq \log_{[N+1]}(w^2) = \log_{[N]}(2 \log w) \leq 2 \log_{[N+1]}(w)$ for w large enough, namely for all $w \geq W$, where $W = W(N)$. Take $\delta > 0$ and set $m_3 := \max \left\{ l, \left\lceil (l-1) \left(1 + \frac{1}{\delta}\right) \right\rceil, \lceil W(N) \rceil, \lceil e_{[N]} \rceil \right\}$. Let further $m \geq m_3$. Then

$$J_3(m, x) \leq 2 \int_{(m, \infty)} (1 - F_{m,l,x}(mw)) \frac{1}{w} \left(\log_{[N+1]}(w) + \frac{1}{\prod_{j=2}^N \log_{[j]}(w)} \right) dw.$$

By virtue of (46) and (48) one has

$$1 - F_{m,l,x}(mw) \leq (1 + \delta)(1 - W_{m,x}(mw)) = (1 + \delta)(1 - F_{1,1,x}(w)). \quad (49)$$

Hence it can be seen that

$$J_3(m, x) \leq 2(1 + \delta) \int_{(m, \infty)} (1 - F_{1,1,x}(w)) g_{N_1}(w) dw. \quad (50)$$

Introduce

$$R_N(x) := \int_{y \in \mathbb{R}^d, \|x-y\| > e_{[N]}} G_N(\log \|x-y\|) q(y) dy, \quad A_p(G_N) := \{x \in S(p) : R_N(x) < \infty\}.$$

Let us note that (1) $P_X(S(p) \setminus A_p(G_N)) = 0$ as $K_{p,q}(1, N) < \infty$;

(2) $P_X(S(p) \setminus S(q)) = 0$ as $P_X \ll P_Y$ (see Lemma A1);

(3) $\mu(S(q) \setminus (\Lambda(q) \cap D_q(R))) = 0$ due to Lemma 3.

Since $P_X \ll \mu$ we conclude that $P_X(S(q) \setminus (\Lambda(q) \cap D_q(R))) = 0$. Hence, one has $P_X(S(p) \setminus (\Lambda(q) \cap D_q(R))) = 0$ in view of 2) and because $B \setminus C \subset (B \setminus A) \cup (A \setminus C)$ for any $A, B, C \subset \mathbb{R}^d$. Set further $A := \Lambda(q) \cap S(q) \cap D_q(R) \cap S(p) \cap A_p(G_N)$. It follows from (1), (2) and (3) that $P_X(S(p) \setminus A) = 0$, so $P_X(A) = 1$. We are going to consider only $x \in A$.

Then, by virtue of (47) and (50), for all $m \geq m_3$ and $x \in A$, we come to the inequality

$$J_3(m, x) \leq 2(1 + \delta)(a(d, 1)R_N(x) + b(N, d, 1)) = A(\delta, d)R_N(x) + B(\delta, d, N), \quad (51)$$

where $A(\delta, d) := 2(1 + \delta)a(d, 1)$, $B(\delta, d, N) := 2(1 + \delta)b(N, d, 1)$.

Part (3f). Here we get the upper bound for $\text{EG}_N(|\log \xi_{m,l,x}|)$. For $m \geq \max\{m_1, m_2, m_3\}$ and each $x \in A$, taking into account (39), (44), (45) and (51) we can claim that

$$\begin{aligned} \text{EG}_N(|\log \xi_{m,l,x}|) &\leq I_1(m, x) + J_1(m, x) + J_2(m, x) + J_3(m, x) \\ &\leq U_1(\varepsilon, N, d)(M_q(x, R))^\varepsilon + U_2(\varepsilon, N, d, l)(m_q(x, R))^{-\varepsilon} \\ &\quad + U_3(m, \varepsilon, N, d, l)(m_q(x, R))^{-\varepsilon} + (A(\delta, d)R_N(x) + B(\delta, d, N)). \end{aligned} \quad (52)$$

For any $\kappa > 0$, one can take $m_4 = m_4(\kappa, \varepsilon, N, d, l) \in \mathbb{N}$ such that $U_3(m, \varepsilon, N, d, l) \leq \kappa$ if $m \geq m_4$. Then by virtue of (52), for each $x \in A$ and $m \geq m_0 := \max\{m_1, m_2, m_3, m_4\}$,

$$\begin{aligned} \mathbb{E}G_N(|\log \xi_{m,l,x}|) &\leq U_1(\varepsilon, N, d)(M_q(x, R))^\varepsilon \\ &+ (U_2(\varepsilon, N, d, l) + \kappa)(m_q(x, R))^{-\varepsilon} + (A(\delta, d)R_N(x) + B(\delta, d, N)) := C_0(x) < \infty. \end{aligned} \quad (53)$$

Hence, for each $x \in A$, we have established uniform integrability of the family $\{\log \xi_{m,l,x}\}_{m \geq m_0}$.

Step 4. Now we verify (23). It was checked, for each $x \in A$ (thus, for P_X -almost every x belonging to $S(p)$) that $\mathbb{E}(\log \phi_{m,l}(1)|X_1 = x) \rightarrow \psi(l) - \log V_d - \log q(x)$, $m \rightarrow \infty$. Set $Z_{m,l}(x) := \mathbb{E}(\log \phi_{m,l}(1)|X_1 = x) = \mathbb{E} \log \xi_{m,l,x}$. Consider $x \in A$ and take any $m \geq m_0$. We use the following property of G_N which is shown in Appendix A.

Lemma 6. For each $N \in \mathbb{N}$, a function G_N is convex on $\mathbb{R}_+$.

By the Jensen inequality a function G_N is nondecreasing and convex.

$$G_N(|Z_{m,l}(x)|) = G_N(|\mathbb{E} \log \xi_{m,l,x}|) \leq G_N(\mathbb{E} |\log \xi_{m,l,x}|) \leq \mathbb{E}G_N(|\log \xi_{m,l,x}|).$$

Relation (53) guarantees that, for all $m \geq m_0$,

$$\begin{aligned} \int_{\mathbb{R}^d} G_N(|Z_{m,l}(x)|) p(x) dx &\leq U_1(\varepsilon, N, d) Q_{p,q}(\varepsilon, R) \\ &+ (U_2(\varepsilon, N, d, l) + \kappa) T_{p,q}(\varepsilon, R) + A(\delta, d) K_{p,q}(1, N) + B(\delta, d, N) < \infty. \end{aligned}$$

Now we know that the family $\{Z_{m,l}(x)\}_{m \geq m_0}$, $x \in A$, is uniformly integrable w.r.t. measure P_X . Thus, for $i \in \mathbb{N}$,

$$\begin{aligned} \mathbb{E} \log \phi_{m,l}(i) &= \int_{\mathbb{R}^d} \mathbb{E}(\log \phi_{m,l}(1)|X_1 = x) P_{X_1}(dx) = \int_{\mathbb{R}^d} Z_{m,l}(x) p(x) dx \\ &\rightarrow \psi(l) - \log V_d - \int_{\mathbb{R}^d} p(x) \log q(x) dx, \quad m \rightarrow \infty, \end{aligned}$$

and we come to relation (23) establishing Statement 1.

Step 5. Here we prove Statement 2. Similar to $F_{m,l,x}(u)$, one can introduce, for $n, k \in \mathbb{N}$, $n \geq k + 1$, $x \in \mathbb{R}^d$ and $u \geq 0$, the following function

$$\begin{aligned} \tilde{F}_{n,k,x}(u) &:= P(\zeta_{n,k}(i) \leq u | X_i = x) = 1 - P\left(\|x - X_{(k)}(x, \mathbb{X}_n \setminus \{x\})\| > r_{n-1}(u)\right) \\ &= 1 - \sum_{s=0}^{k-1} \binom{n-1}{s} (V_{n-1,x}(u))^s (1 - V_{n-1,x}(u))^{n-1-s} := P(\tilde{\zeta}_{n,k,x} \leq u), \end{aligned} \quad (54)$$

where $r_n(u)$ was defined in (29), and

$$V_{n,x}(u) := \int_{B(x, r_n(u))} p(z) dz, \quad \tilde{\zeta}_{n,k,x} := (n-1) \left\| x - X_{(k)}(x, \mathbb{X}_n \setminus \{x\}) \right\|^d. \quad (55)$$

Formulas (54) and (55) show that $\tilde{F}_{n,k,x}(u)$ is the regular conditional distribution function of $\zeta_{n,k}(i)$ given $X_i = x$. Moreover, for any fixed $u \geq 0$ and $x \in \Lambda(p) \cap S(p)$ (thus $p(x) > 0$),

$$\tilde{F}_{n,k,x}(u) \rightarrow 1 - \sum_{s=0}^{k-1} \frac{(V_d p(x) u)^s}{s!} e^{-V_d p(x) u} := \tilde{F}_{k,x}(u), \quad n \rightarrow \infty.$$

Hence, $\tilde{\zeta}_{n,k,x} \xrightarrow{law} \tilde{\zeta}_{k,x}$, $x \in \Lambda(p) \cap S(p)$, $n \rightarrow \infty$. Set $\tilde{A}_p(G_N) := \{x \in S(p) : \tilde{R}_N(x) < \infty\}$, where $N \in \mathbb{N}$ and

$$\tilde{R}_N(x) := \int_{y \in \mathbb{R}^d, \|x-y\| > e_{[N]}} G_N(\log \|x-y\|) p(y) dy.$$

Take $\tilde{A} := \Lambda(p) \cap S(p) \cap D_p(R) \cap \tilde{A}_p(G_N)$. Then $P_X(\tilde{A}) = 1$ and, for $x \in \tilde{A}$, one can verify that $\mathbb{E}G_N(|\log \tilde{\zeta}_{n,k,x}|) \leq \tilde{C}_0(x) < \infty$, for all $n \geq n_0$, and therefore $\mathbb{E} \log \tilde{\zeta}_{n,k,x} \rightarrow \mathbb{E} \log \tilde{\zeta}_{k,x}$

as $n \rightarrow \infty$. Thus, $E(\log \zeta_{n,k}(1) | X_1 = x) \rightarrow \psi(k) - \log V_d - \log p(x)$, $n \rightarrow \infty$. Set $\tilde{Z}_{n,k}(x) := E(\log \zeta_{n,k}(1) | X_1 = x)$. One can see that, for all $n \geq n_0$, $\int_{\mathbb{R}^d} G_N(|\tilde{Z}_{n,k}(x)|) p(x) dx < \infty$. Hence similar to Steps 1–4 we come to relation (24).

So, (14) holds and the proof of Theorem 1 is complete.

5. Proof of Theorem 2

We will follow the general scheme described in Remark 7. However now this scheme is more involved.

First of all note that, in view of Lemma 1, the finiteness of $K_{p,q}(2, N)$ and $K_{p,p}(2, N)$ implies the finiteness of $K_{p,q}(1, N)$ and $K_{p,p}(1, N)$, respectively. Thus, the conditions of Theorem 2 entail validity of Theorem 1 statements. Consequently under the conditions of Theorem 2, for n and m large enough, one can claim that $\hat{D}_{n,m}(k, l) \in L^1(\Omega)$ and $E\hat{D}_{n,m}(k, l) \rightarrow D(P_X || P_Y)$, as $n, m \rightarrow \infty$.

We will show that $\hat{D}_{n,m}(k, l) \in L^2(\Omega)$ for all n and m large enough. Then one can write

$$E\left(\hat{D}_{n,m}(k, l) - D(P_X || P_Y)\right)^2 = \text{var}\left(\hat{D}_{n,m}(k, l)\right) + \left(E\hat{D}_{n,m}(k, l) - D(P_X || P_Y)\right)^2.$$

Therefore to prove (16) we will demonstrate that $\text{var}\left(\hat{D}_{n,m}(k, l)\right) \rightarrow 0$, $n, m \rightarrow \infty$.

Due to (28) the random variables $\log \phi_{m,l}(1), \dots, \log \phi_{m,l}(n)$ are identically distributed (and $\log \zeta_{n,k}(1), \dots, \log \zeta_{n,k}(n)$ are identically distributed as well). The variables $\phi_{m,l}(i)$ and $\zeta_{n,k}(i)$ are the same as in (22). We will demonstrate that $\log \phi_{m,l}(1)$ and $\log \zeta_{n,k}(1)$ belong to $L^2(\Omega)$. Hence (22) yields

$$\begin{aligned} \text{var}\left(\hat{D}_{n,m}(k, l)\right) &= \frac{1}{n^2} \sum_{i,j=1}^n \text{cov}\left(\log \phi_{m,l}(i) - \log \zeta_{n,k}(i), \log \phi_{m,l}(j) - \log \zeta_{n,k}(j)\right) \\ &= \frac{1}{n} \text{var}(\log \phi_{m,l}(1)) + \frac{2}{n^2} \sum_{1 \leq i < j \leq n} \text{cov}(\log \phi_{m,l}(i), \log \phi_{m,l}(j)) + \frac{1}{n} \text{var}(\log \zeta_{n,k}(1)) \\ &\quad + \frac{2}{n^2} \sum_{1 \leq i < j \leq n} \text{cov}(\log \zeta_{n,k}(i), \log \zeta_{n,k}(j)) - \frac{2}{n^2} \sum_{i,j=1}^n \text{cov}(\log \phi_{m,l}(i), \log \zeta_{n,k}(j)). \end{aligned} \quad (56)$$

We mainly follow the notation employed in the above proof of Theorem 1, except the possibly different choice of the sets $A \subset \mathbb{R}^d$, $\tilde{A} \subset \mathbb{R}^d$, positive $U_j, C_j(x), \tilde{C}_j(x)$ and integers m_j, n_j , where $j \in \mathbb{Z}_+$ and $x \in \mathbb{R}^d$. The following Theorem 2 proof is also subdivided in 5 parts. Steps 1–3 deal with the demonstration of relation $\frac{1}{n} \text{var}(\log \phi_{m,l}(1)) \rightarrow 0$ as $n, m \rightarrow \infty$. Step 4 validates the relation $\frac{2}{n^2} \sum_{1 \leq i < j \leq n} \text{cov}(\log \phi_{m,l}(i), \log \phi_{m,l}(j)) \rightarrow 0$ as $n, m \rightarrow \infty$. At Step 5 we establish that

$$\frac{2}{n^2} \sum_{1 \leq i < j \leq n} \text{cov}(\log \zeta_{n,k}(i), \log \zeta_{n,k}(j)) \rightarrow 0, \quad n \rightarrow \infty,$$

This step is rather involved. Step 6 justifies the desired statement $\text{var}\left(\hat{D}_{n,m}(k, l)\right) \rightarrow 0$, $n, m \rightarrow \infty$.

Step 1. We study $E \log^2(\phi_{m,l}(1))$, as $m \rightarrow \infty$. For $x \in \mathbb{R}^d$ and $N \in \mathbb{N}$, introduce

$$R_{N,2}(x) := \int_{\|x-y\| \geq e_{[N]}} G_N(\log^2 \|x-y\|) q(y) dy. \quad (57)$$

Set $A_{p,2}(G_N) := \{x \in S(p) : R_{N,2}(x) < \infty\}$. Then $P_X(S(p) \setminus A_{p,2}(G_N)) = 0$ since $K_{p,q}(2, N) < \infty$. Consider

$$A := \Lambda(q) \cap S(q) \cap D_q(R) \cap S(p) \cap A_{p,2}(G_N), \quad (58)$$

where the first four sets appeared in Theorem 1 proof, R and N are indicated in conditions of Theorem 2. It is easily seen that $P_X(A) = 1$. The reasoning is exactly the same as in the proof of Theorem 1.

Recall that, for each $x \in A$, one has $\log \xi_{m,l,x} \xrightarrow{law} \log \xi_{l,x}$, $m \rightarrow \infty$, where $\xi_{m,l,x} := m \left\| x - Y_{(l)}(x, \mathbb{Y}_m) \right\|^d$ and $\xi_{l,x}$ has $\Gamma(V_d q(x), l)$ distribution. Convergence in law of random variables is maintained by continuous transformations. Thus, for each $x \in A$, we get

$$\log^2 \xi_{m,l,x} \xrightarrow{law} \log^2 \xi_{l,x}, \quad m \rightarrow \infty. \quad (59)$$

For any $x \in A$, according to (28),

$$\begin{aligned} \mathbb{E} \log^2 \xi_{m,l,x} &= \int_{(0,\infty)} \log^2 u \, dF_{m,l,x}(u) = \int_{(0,\infty)} \log^2 u \, dP(\phi_{m,l}(1) \leq u | X_1 = x) \\ &= \mathbb{E}(\log^2 \phi_{m,l}(1) | X_1 = x). \end{aligned} \quad (60)$$

Note that if $\eta \sim \Gamma(\alpha, \lambda)$, where $\alpha > 0$ and $\lambda > 0$, then it is not difficult to verify that

$$\mathbb{E} \log^2 \eta = \frac{\Gamma''(\lambda)}{\Gamma(\lambda)} - 2\psi(\lambda) \log \alpha + \log^2 \alpha.$$

Since $\xi_{l,x} \sim \Gamma(V_d q(x), l)$, for $x \in S(q)$, one has

$$\mathbb{E} \log^2 \xi_{l,x} = \frac{\Gamma''(l)}{\Gamma(l)} - 2\psi(l) \log(V_d q(x)) + \log^2(V_d q(x)) = \log^2 q(x) + h_1 \log q(x) + h_2, \quad (61)$$

where $h_1 := h_1(l, d)$ and $h_2 := h_2(l, d)$ depend only on fixed l and d .

We prove now that, for $x \in A$, one has

$$\mathbb{E}(\log^2 \phi_{m,l}(1) | X_1 = x) \rightarrow \log^2 q(x) + h_1 \log q(x) + h_2, \quad m \rightarrow \infty. \quad (62)$$

Taking into account (60) and (61) we can claim that relation (62) is equivalent to the following one: $\mathbb{E} \log^2 \xi_{m,l,x} \rightarrow \mathbb{E} \log^2 \xi_{l,x}$, $m \rightarrow \infty$. So, in view of (59) to prove (62) it is sufficient to show that, for each $x \in A$, a family $\{\log^2 \xi_{m,l,x}\}_{m \geq m_0(x)}$ is uniformly integrable for some $m_0(x) \in \mathbb{N}$. Then, following Theorem 1 proof, one can certify that, for all $x \in A$ and some nonnegative $C_0(x)$,

$$\sup_{m \geq m_0(x)} \mathbb{E} G_N(\log^2 \xi_{m,l,x}) \leq C_0(x) < \infty. \quad (63)$$

Step 2. Now we will prove (63). For each $N \in \mathbb{N}$, introduce $\rho(N) := \exp\{\sqrt{e_{[N-1]}}\}$ and

$$h_N(t) := \begin{cases} 0, & t \in \left(\frac{1}{\rho(N)}, \rho(N)\right], \\ \frac{2 \log t}{t} \left(\log_{[N]}(\log^2 t) + \frac{1}{\prod_{j=1}^{N-1} \log_{[j]}(\log^2 t)} \right), & t \in \left(0, \frac{1}{\rho(N)}\right] \cup (\rho(N), \infty). \end{cases}$$

As usual, a product over an empty set (if $N = 1$) equals to 1. To show (63) we refer to the next lemma.

Lemma 7. Let $F(u)$, $u \in \mathbb{R}$, be a distribution function such that $F(0) = 0$. Fix an arbitrary $N \in \mathbb{N}$. Then

$$\begin{aligned} (1) \quad & \int_{\left(0, \frac{1}{\rho(N)}\right]} G_N(\log^2 u) dF(u) = \int_{\left(0, \frac{1}{\rho(N)}\right]} F(u) (-h_N(u)) du, \\ (2) \quad & \int_{(\rho(N), \infty)} G_N(\log^2 u) dF(u) = \int_{(\rho(N), \infty)} (1 - F(u)) h_N(u) du. \end{aligned}$$

The proof of this lemma is omitted, being quite similar to one of Lemma 2. By Lemma 7 and since $G_N(\log^2 u) = 0$, for $u \in \left(\frac{1}{\rho(N)}, \rho(N)\right]$, one has

$$\begin{aligned} \mathbb{E}G_N(\log^2 \xi_{m,l,x}) &= \int_{\left(0, \frac{1}{\rho(N)}\right]} F_{m,l,x}(u)(-h_N(u))du + \int_{(\rho(N), \infty)} (1 - F_{m,l,x}(u))h_N(u)du \\ &:= I_1(m, x) + I_2(m, x). \end{aligned}$$

To simplify notation we do not indicate the dependence of $I_i(m, x)$ ($i = 1, 2$) on fixed N, l and d .

For clarity, further implementation of Step 2 is divided into several parts.

Part (2a). At first we consider $I_1(m, x)$. As in Theorem 1 proof, for fixed $R > 0$ and $\varepsilon > 0$ appearing in the conditions of Theorem 2, an inequality $F_{m,l,x}(u) \leq (M_q(x, R))^\varepsilon V_d^\varepsilon u^\varepsilon$ holds for any $x \in A, u \in \left(0, \frac{1}{\rho(N)}\right]$ and $m \geq m_1 := \max\left\{\left\lceil \frac{1}{\rho(N)R^d} \right\rceil, l\right\}$. Taking into account that $0 \leq (-h_N(u)) \leq \frac{(-2\log u)(\log_{[N]}(\log^2 u) + 1)}{u}$ if $u \in \left(0, \frac{1}{\rho(N)}\right]$, we get, for $m \geq m_1$,

$$\begin{aligned} I_1(m, x) &\leq (M_q(x, R))^\varepsilon V_d^\varepsilon \int_{\left(0, \frac{1}{\rho(N)}\right]} \frac{(-2\log u)(\log_{[N]}(\log^2 u) + 1)}{u^{1-\varepsilon}} du \\ &= U_1(\varepsilon, N, d)(M_q(x, R))^\varepsilon. \end{aligned} \quad (64)$$

Here $U_1(\varepsilon, N, d) := V_d^\varepsilon L_{N,2}(\varepsilon)$, $L_{N,2}(\varepsilon) := \int_{[\sqrt{e_{[N-1]}}, \infty)} 2t(\log_{[N]}(t^2) + 1)e^{-\varepsilon t} dt < \infty$ for each $\varepsilon > 0$ and any $N \in \mathbb{N}$.

Part (2b). Consider $I_2(m, x)$. Following the previous theorem proof we at first observe that $h_N(u) \leq \frac{2\log u}{u}(\log_{[N]}(\log^2 u) + 1)$ for $u \in (\rho(N), \infty)$. So, for all $m \geq \max\{\rho^2(N), l\}$,

$$\begin{aligned} I_2(m, x) &\leq \int_{(\rho(N), \sqrt{m}]} (1 - F_{m,l,x}(u)) \frac{2\log u (\log_{[N]}(\log^2 u) + 1)}{u} du \\ &\quad + \int_{(\sqrt{m}, m^2]} (1 - F_{m,l,x}(u)) \frac{2\log u (\log_{[N]}(\log^2 u) + 1)}{u} du \\ &\quad + \int_{(m^2, \infty)} (1 - F_{m,l,x}(u))h_N(u) du := J_1(m, x) + J_2(m, x) + J_3(m, x), \end{aligned}$$

where we do not indicate the dependence of $J_j(m, x)$ ($j = 1, 2, 3$) on N, l and d .

For $R > 0$ and $\varepsilon > 0$ appearing in the conditions of Theorem 2, one can show (see Theorem 1 proof), that inequality

$$1 - F_{m,l,x}(u) \leq S_1(S_2 V_d u m_q(x, R))^{-\varepsilon} \quad (65)$$

holds for any $x \in A, u \in (\rho(N), \sqrt{m}]$ and all $m \geq m_2 := \max\left\{\left\lceil \frac{1}{R^{2d}} \right\rceil, \lceil \rho^2(N) \rceil, l\right\}$. Here $S_1 := S_1(l)$ and S_2 are the same as in the proof of Theorem 1. For all $x \in A$ and $m \geq m_2$, we come to the relations

$$\begin{aligned} J_1(m, x) &\leq \frac{S_1}{(S_2 V_d)^\varepsilon (m_q(x, R))^\varepsilon} \int_{(\rho(N), \infty)} \frac{2\log u (\log_{[N]}(\log^2 u) + 1)}{u^{1+\varepsilon}} du \\ &= U_2(\varepsilon, N, d, l)(m_q(x, R))^{-\varepsilon}, \end{aligned} \quad (66)$$

where $U_2(\varepsilon, N, d, l) := 2S_1(l) L_{N,2}(\varepsilon)(S_2 V_d)^{-\varepsilon}$.

Part (2c). Let us consider $J_2(m, x)$. Take $\delta > 0$. Then, due to (65), for all $x \in A$ and any $m \geq m_2$,

$$\begin{aligned}
J_2(m, x) &\leq 2(1 - F_{m,l,x}(\sqrt{m})) \int_{(\sqrt{m}, m^2]} \log u \left(\log_{[N]}(\log^2 u) + 1 \right) d \log u \\
&\leq 4S_1(S_2 V_d)^{-\varepsilon} m^{-\frac{\varepsilon}{2}} (m_q(x, R))^{-\varepsilon} \left(\log_{[N]}(4 \log^2 m) + 1 \right) \log^2 m \\
&= U_3(m, \varepsilon, N, d, l) (m_q(x, R))^{-\varepsilon},
\end{aligned} \tag{67}$$

where $U_3(m, \varepsilon, N, d, l) := 4S_1(S_2 V_d)^{-\varepsilon} m^{-\frac{\varepsilon}{2}} (\log^2 m) (\log_{[N]}(4 \log^2 m) + 1) \rightarrow 0, m \rightarrow \infty$.

Part (2d). Now we turn to $J_3(m, x)$. Take $u = mw$. Then $J_3(m, x)$ has the form

$$\int_{(m, \infty)} (1 - F_{m,l,x}(mw)) \frac{2 \log(mw)}{w} \left(\log_{[N]}(\log^2(mw)) + \frac{1}{\prod_{j=1}^{N-1} \log_{[j]}(\log^2(mw))} \right) dw.$$

Due to Lemma 5 there exists $T(N) > \rho(N)$ such that

$$\log_{[N]}(\log^2(w^2)) = \log_{[N]}(4 \log^2 w) \leq 2 \log_{[N]}(\log^2 w), \quad w \geq T(N). \tag{68}$$

Pick some $\delta > 0$ and set $m_3 := \max\left\{l, \left\lceil (l-1)\left(1 + \frac{1}{\delta}\right) \right\rceil, \lceil T(N) \rceil, \lceil \rho(N) \rceil\right\}$, where $T(N)$ was introduced in (68). Consider $m \geq m_3$. In view of Lemma 4 (for $\nu = 2$), (49), (68) and since $w > m$,

$$\begin{aligned}
J_3(m, x) &\leq \int_{(m, \infty)} (1 - F_{m,l,x}(mw)) \frac{2 \log(w^2)}{w} \left(\log_{[N]}(\log^2(w^2)) + \frac{1}{\prod_{j=1}^{N-1} \log_{[j]}(\log^2 w)} \right) dw \\
&\leq 4(1 + \delta) \int_{(m, \infty)} (1 - F_{1,1,x}(w)) \frac{2 \log w}{w} \left(\log_{[N]}(\log^2 w) + \frac{1}{\prod_{j=1}^{N-1} \log_{[j]}(\log^2 w)} \right) dw \\
&= 4(1 + \delta) \int_{(m, \infty)} (1 - F_{1,1,x}(w)) h_N(w) dw \leq 4(1 + \delta) \int_{(\rho(N), \infty)} (1 - F_{1,1,x}(w)) h_N(w) dw \\
&= 4(1 + \delta) \int_{(\rho(N), \infty)} G_N(\log^2 w) dF_{1,1,x}(w) = 4(1 + \delta) \mathbb{E}[G_N(\log^2 \xi_{1,1,x}) \mathbb{I}\{\xi_{1,1,x} > \rho(N)\}] \\
&= 4(1 + \delta) \mathbb{E}[G_N((\log \|Y - x\|^d)^2) \mathbb{I}\{\|Y - x\|^d > \rho(N)\}] \\
&= 4(1 + \delta) \int_{y \in \mathbb{R}^d, \|x-y\| > (\rho(N))^{1/d}} G_N((\log \|x - y\|^d)^2) q(y) dy \\
&\leq 4(1 + \delta) \left(a(d, 2) \int_{y \in \mathbb{R}^d, \|x-y\| > (\rho(N))^{1/d}} G_N(\log^2 \|x - y\|) q(y) dy + b(N, d, 2) \right) \\
&\leq 4(1 + \delta) \left(a(d, 2) \left(R_{N,2}(x) + G_N(e_{[N-1]}^2) \right) + b(N, d, 2) \right) \\
&= A(\delta, d) R_{N,2}(x) + B(\delta, d, N),
\end{aligned} \tag{69}$$

$A(\delta, d) := 4(1 + \delta)a(d, 2)$, $B(\delta, d, N) := 4(1 + \delta)(a(d, 2)G_N(e_{[N-1]}^2) + b(N, d, 2))$, $R_{N,2}(x)$ is defined in (57). Here we have also used, for any $N \in \mathbb{N}$, $\nu, t, u > 0$, $t < u$, the following estimates

$$K_{p,q}(\nu, N, u) \leq K_{p,q}(\nu, N, t) \leq K_{p,q}(\nu, N, u) + \max\{G_N(|\log t|^\nu), G_N(|\log u|^\nu)\}.$$

Part (2e). Examine $\mathbb{E}G_N(\log^2 \xi_{m,l,x})$. Thus, for each $x \in A$ and $m \geq \max\{m_1, m_2, m_3\}$, taking into account (64), (66), (67) and (69), we can claim that

$$\begin{aligned}
\mathbb{E}G_N(\log^2 \xi_{m,l,x}) &\leq I_1(m, x) + J_1(m, x) + J_2(m, x) + J_3(m, x) \\
&\leq U_1(\varepsilon, N, d)(M_q(x, R))^\varepsilon + U_2(\varepsilon, N, d, l)(m_q(x, R))^{-\varepsilon} + U_3(m, \varepsilon, N, d, l)(m_q(x, R))^{-\varepsilon} \\
&\quad + A(\delta, d)R_{N,2}(x) + B(\delta, d, N).
\end{aligned} \tag{70}$$

Moreover, for any $\kappa > 0$, one can choose $m_4 := m_4(\kappa, \varepsilon, N, d, l) \in \mathbb{N}$ such that, for $m \geq m_4$, it holds $U_3(m, \varepsilon, N, d, l) \leq \kappa$. Then by (70), for each $x \in A$ and $m \geq m_0 := \max\{m_1, m_2, m_3, m_4\}$,

$$\begin{aligned} \mathbb{E} G_N(\log^2 \xi_{m,l,x}) &\leq U_1(\varepsilon, N, d)(M_q(x, R))^\varepsilon \\ &+ (U_2(\varepsilon, N, d, l) + \kappa)(m_q(x, R))^{-\varepsilon} + A(\delta, d)R_{N,2}(x) + B(\delta, d, N) := C_0(x) < \infty. \end{aligned} \quad (71)$$

Hence we have proved the uniform integrability of the family $\{\log^2 \xi_{m,l,x}\}_{m \geq m_0}$ for each $x \in A$. Therefore, for any $x \in A$ (thus for P_X -almost every $x \in S(p)$), relation (62) holds.

Step 3. Now we can return to $\mathbb{E} \log^2 \phi_{m,l}(1)$. Set $\Delta_{m,l}(x) := \mathbb{E}(\log^2 \phi_{m,l}(1) | X_1 = x) = \mathbb{E} \log^2 \xi_{m,l,x}$. Consider $x \in A$ and take any $m \geq m_0$. A function G_N is nondecreasing and convex according to Lemma 6. Due to the Jensen inequality

$$G_N(\Delta_{m,l}(x)) = G_N(\mathbb{E} \log^2 \xi_{m,l,x}) \leq \mathbb{E} G_N(\log^2 \xi_{m,l,x}). \quad (72)$$

Relation (72) guarantees that, for each $x \in A$ and all $m \geq m_0$,

$$\begin{aligned} \int_{\mathbb{R}^d} G_N(\Delta_{m,l}(x)) p(x) dx &\leq U_1(\varepsilon, N, d) Q_{p,q}(\varepsilon, R) + (U_2(\varepsilon, N, d, l) + \kappa) T_{p,q}(\varepsilon, R) \\ &+ A(\delta, d) K_{p,q}(2, N) + B(\delta, d, N) < \infty. \end{aligned}$$

Uniform integrability of the family $\{\Delta_{m,l}(\cdot)\}_{m \geq m_0}$ (w.r.t measure P_X) is thus established. Hence one can claim that

$$\mathbb{E} \log^2 \phi_{m,l}(1) \rightarrow \int_{\mathbb{R}^d} p(x) \log^2 q(x) dx + h_1 \int_{\mathbb{R}^d} p(x) \log q(x) dx + h_2, \quad m \rightarrow \infty.$$

It is easily seen that finiteness of integrals $Q_{p,q}(\varepsilon, R)$, $T_{p,q}(\varepsilon, R)$ implies that

$$\int_{\mathbb{R}^d} p(x) \log^2 q(x) dx < \infty, \quad \int_{\mathbb{R}^d} p(x) |\log q(x)| dx < \infty.$$

Thus, $\mathbb{E} \log^2 \phi_{m,l}(1) \rightarrow \tau_2 < \infty$ and $\text{var}(\log \phi_{m,l}(1)) = \mathbb{E} \log^2 \phi_{m,l}(1) - (\mathbb{E} \log \phi_{m,l}(1))^2 \rightarrow \tau_2 - \tau_1^2 < \infty$, $m \rightarrow \infty$, where $\tau_1 := \psi(l) - \log V_d - \int_{\mathbb{R}^d} p(x) \log q(x) dx$ according to (23). Consequently, $\frac{1}{n} \text{var}(\log \phi_{m,l}(1)) \rightarrow 0$ as $n, m \rightarrow \infty$.

Step 4. Now we consider $\text{cov}(\log \phi_{m,l}(i), \log \phi_{m,l}(j))$ for $i \neq j$, where $i, j \in \{1, \dots, n\}$. For $x, y \in \mathbb{R}^d$, define conditional distribution function

$$\Phi_{m,l,x,y}^{i,j}(u, w) := P(\phi_{m,l}(i) \leq u, \phi_{m,l}(j) \leq w | X_i = x, X_j = y), \quad u, w \geq 0.$$

For $x, y \in \mathbb{R}^d$, $u, w \geq 0$, $i \neq j$,

$$\begin{aligned} \Phi_{m,l,x,y}^{i,j}(u, w) &= 1 - P(\phi_{m,l}(i) > u | X_i = x, X_j = y) \\ &- P(\phi_{m,l}(j) > w | X_i = x, X_j = y) + P(\phi_{m,l}(i) > u, \phi_{m,l}(j) > w | X_i = x, X_j = y) \\ &= 1 - P\left(\|x - Y_{(l)}(x, \mathbb{Y}_m)\| > r_m(u)\right) - P\left(\|y - Y_{(l)}(y, \mathbb{Y}_m)\| > r_m(w)\right) \\ &+ P\left(\|x - Y_{(l)}(x, \mathbb{Y}_m)\| > r_m(u), \|y - Y_{(l)}(y, \mathbb{Y}_m)\| > r_m(w)\right). \end{aligned} \quad (73)$$

Here $r_m(a) = \left(\frac{a}{m}\right)^{\frac{1}{d}}$ for all $a \geq 0$, as previously. One can write $\Phi_{m,l,x,y}(u, w)$ instead of $\Phi_{m,l,x,y}^{i,j}(u, w)$, because the right-hand side of (73) does not depend on i and j .

Set $A_1 := \{(x, y) : x \in A, y \in A, x \neq y\}$ and $A_2 := \{(x, y) : x \in A, y \in A, x = y\}$, where A is introduced in (58). Evidently, $(P_X \otimes P_X)(A_1) = 1$ and $(P_X \otimes P_X)(A_2) = 0$. Consider $(x, y) \in A_1$. Obviously, for any $a > 0$, $r_m(a) \rightarrow 0$, as $m \rightarrow \infty$. For $(x, y) \in A_1$ we take $m_5 = m_5(u, w, \|x - y\|) := 1 + \left\lceil \left(\frac{2}{\|x - y\|}\right)^d \max\{u, w\} \right\rceil$. Then $r_m(u) < \frac{\|x - y\|}{2}$

and $r_m(w) < \frac{\|x-y\|}{2}$ for all $m \geq m_5$. Thus, $B(x, r_m(u)) \cap B(y, r_m(w)) = \emptyset$ if $m \geq m_5$. Consequently, for $m \geq m_6(u, w, \|x - y\|) := \max \{m_5, 2(l-1)\}$,

$$\begin{aligned} & \mathbb{P}\left(\|x - Y_{(l)}(x, \mathbb{Y}_m)\| > r_m(u), \|y - Y_{(l)}(y, \mathbb{Y}_m)\| > r_m(w)\right) \\ &= \sum_{s_1=0}^{l-1} \sum_{s_2=0}^{l-1} \frac{m!}{s_1!s_2!(m-s_1-s_2)!} (W_{m,x}(u))^{s_1} (W_{m,y}(w))^{s_2} (1 - W_{m,x}(u) - W_{m,y}(w))^{m-s_1-s_2}. \end{aligned} \quad (74)$$

In view of (28), (73) and (74), one has for $\Phi_{m,l,x,y}(u, w)$ the following representation

$$\begin{aligned} & 1 - \sum_{s_1=0}^{l-1} \binom{m}{s_1} (W_{m,x}(u))^{s_1} (1 - W_{m,x}(u))^{m-s_1} - \sum_{s_2=0}^{l-1} \binom{m}{s_2} (W_{m,y}(w))^{s_2} (1 - W_{m,y}(w))^{m-s_2} \\ &+ \sum_{s_1=0}^{l-1} \sum_{s_2=0}^{l-1} \frac{m!}{s_1!s_2!(m-s_1-s_2)!} (W_{m,x}(u))^{s_1} (W_{m,y}(w))^{s_2} (1 - W_{m,x}(u) - W_{m,y}(w))^{m-s_1-s_2}. \end{aligned} \quad (75)$$

For any fixed $(x, y) \in A_1$ and $u, w \geq 0$, we get, as $m \rightarrow \infty$,

$$\begin{aligned} & \frac{m!}{s_1!s_2!(m-s_1-s_2)!} (W_{m,x}(u))^{s_1} (W_{m,y}(w))^{s_2} \rightarrow \frac{(V_d u q(x))^{s_1}}{s_1!} \frac{(V_d w q(y))^{s_2}}{s_2!}, \\ & (1 - W_{m,x}(u) - W_{m,y}(w))^{m-s_1-s_2} \rightarrow e^{-V_d(uq(x)+wq(y))}. \end{aligned} \quad (76)$$

Then, according to (31), (75) and (76), for all fixed $u, w \geq 0$, $(x, y) \in A_1$, one has

$$\begin{aligned} \Phi_{m,l,x,y}(u, w) &\rightarrow 1 - \sum_{s_1=0}^{l-1} \frac{(V_d u q(x))^{s_1}}{s_1!} e^{-V_d u q(x)} - \sum_{s_2=0}^{l-1} \frac{(V_d w q(y))^{s_2}}{s_2!} e^{-V_d w q(y)} \\ &+ \sum_{s_1=0}^{l-1} \sum_{s_2=0}^{l-1} \frac{(V_d u q(x))^{s_1}}{s_1!} \frac{(V_d w q(y))^{s_2}}{s_2!} e^{-V_d(uq(x)+wq(y))} \\ &= \left(1 - \sum_{s_1=0}^{l-1} \frac{(V_d u q(x))^{s_1}}{s_1!} e^{-V_d u q(x)}\right) \left(1 - \sum_{s_2=0}^{l-1} \frac{(V_d w q(y))^{s_2}}{s_2!} e^{-V_d w q(y)}\right) \\ &= F_{l,x}(u) F_{l,y}(w) := \Phi_{l,x,y}(u, w), \quad m \rightarrow \infty. \end{aligned}$$

Thus, $\Phi_{l,x,y}(\cdot, \cdot)$ is identified as a distribution function of a vector $\eta_{l,x,y} := (\xi_{l,x}, \xi_{l,y})$ having independent components such that $\xi_{l,x} \sim \Gamma(V_d q(x), l)$, $\xi_{l,y} \sim \Gamma(V_d q(y), l)$. Observe also that $\Phi_{m,l,x,y}(\cdot, \cdot)$ is a distribution function of a random vector $\eta_{m,l,x,y} := (\xi_{m,l,x}, \xi_{m,l,y})$. Consequently, we have shown that $\eta_{m,l,x,y} \xrightarrow{law} \eta_{l,x,y}$ as $m \rightarrow \infty$. Hence, for any $(x, y) \in A_1$,

$$\log \xi_{m,l,x} \log \xi_{m,l,y} \xrightarrow{law} \log \xi_{l,x} \log \xi_{l,y}, \quad m \rightarrow \infty.$$

Here we take strictly positive versions of random variables under consideration. Note that, for all $i, j \in \mathbb{N}$, $i \neq j$,

$$\begin{aligned} \mathbb{E}(\log \xi_{m,l,x} \log \xi_{m,l,y}) &= \int_{(0,\infty) \times (0,\infty)} \log u \log w d\Phi_{m,l,x,y}(u, w) \\ &= \mathbb{E}(\log \phi_{m,l}(i) \log \phi_{m,l}(j) | X_i = x, X_j = y). \end{aligned} \quad (77)$$

One has $\mathbb{E}(\log \xi_{l,x} \log \xi_{l,y}) = \mathbb{E} \log \xi_{l,x} \mathbb{E} \log \xi_{l,y} = a_{l,d}(x) a_{l,d}(y)$ because $\xi_{l,x}$ and $\xi_{l,y}$ are independent, here $a_{l,d}(z) := \psi(l) - \log V_d - \log q(z)$, $z \in \mathbb{R}^d$.

Now we intend to verify that, for any $(x, y) \in A_1$,

$$\mathbb{E}(\log \phi_{m,l}(1) \log \phi_{m,l}(2) | X_1 = x, X_2 = y) \rightarrow a_{l,d}(x) a_{l,d}(y). \quad (78)$$

Equivalently, one can prove that $\mathbb{E}(\log \xi_{m,l,x} \log \xi_{m,l,y}) \rightarrow \mathbb{E}(\log \xi_{l,x} \log \xi_{l,y})$ for each $(x, y) \in A_1$, as $m \rightarrow \infty$.

Part (4a). We will prove the uniform integrability of a family $\{\log \xi_{m,l,x} \log \xi_{m,l,y}\}_{m \geq m_0}$ for $(x, y) \in A_1$. The convex function $G_N(\cdot)$ is nondecreasing. Thus, following the proof of Step 2 for any $(x, y) \in A_1$, one can find m_0 (the same as in the proof of Step 2 such that, for all $m \geq m_0$,

$$\begin{aligned} \mathbb{E} G_N(|\log \xi_{m,l,x} \log \xi_{m,l,y}|) &\leq \mathbb{E} G_N\left(\frac{1}{2} \log^2 \xi_{m,l,x} + \frac{1}{2} \log^2 \xi_{m,l,y}\right) \\ &\leq \frac{1}{2} \left(\mathbb{E} G_N(\log^2 \xi_{m,l,x}) + \mathbb{E} G_N(\log^2 \xi_{m,l,y}) \right) \leq \frac{U_1}{2} \left((M_q(x, R))^\varepsilon + (M_q(y, R))^\varepsilon \right) \\ &+ \frac{U_2 + \kappa}{2} \left((m_q(x, R))^{-\varepsilon} + (m_q(y, R))^{-\varepsilon} \right) + \frac{A}{2} (R_{N,2}(x) + R_{N,2}(y)) + B := \tilde{C}_0(x, y). \end{aligned} \quad (79)$$

Here we used (71). It is essential that U_1, U_2, κ, A, B do not depend on x and y . Hence, for any $(x, y) \in A_1$, a family $\{\log \xi_{m,l,x} \log \xi_{m,l,y}\}_{m \geq m_0}$ is uniformly integrable. So, we establish (78) for $(x, y) \in A_1$.

Part (4b). We return to $\text{cov}(\log \phi_{m,l}(i), \log \phi_{m,l}(j))$ for $i \neq j, i, j \in \{1, \dots, n\}$. Set $T_{m,l}(x, y) := \mathbb{E}(\log \phi_{m,l}(1) \log \phi_{m,l}(2) | X_1 = x, X_2 = y)$ where $(x, y) \in A_1$. Then (78) means that $T_{m,l}(x, y) = \mathbb{E}(\log \xi_{m,l,x} \log \xi_{m,l,y}) \rightarrow a_{l,d}(x) a_{l,d}(y)$ for any $(x, y) \in A_1$, as $m \rightarrow \infty$. Note that

$$\begin{aligned} G_N(|T_{m,l}(x, y)|) &= G_N(|\mathbb{E} \log \xi_{m,l,x} \log \xi_{m,l,y}|) \\ &\leq G_N(\mathbb{E} |\log \xi_{m,l,x} \log \xi_{m,l,y}|) \leq \mathbb{E} G_N(|\log \xi_{m,l,x} \log \xi_{m,l,y}|). \end{aligned} \quad (80)$$

As $(P_X \otimes P_X)(A_1) = 1$, one can conclude due to (79) and (80) that, for all $m \geq m_0$,

$$\begin{aligned} \int_{\mathbb{R}^d \times \mathbb{R}^d} G_N(|T_{m,l}(x, y)|) p(x) p(y) dx dy &= \int_{(x,y) \in A_1} G_N(|T_{m,l}(x, y)|) p(x) p(y) dx dy \\ &\leq U_1 \int_{\mathbb{R}^d} M_q^\varepsilon(x, R) p(x) dx + (U_2 + \kappa) \int_{\mathbb{R}^d} m_q^{-\varepsilon}(x, R) p(x) dx + A \int_{\mathbb{R}^d} R_{N,2}(x) p(x) dx + B \\ &= U_1 Q_{p,q}(\varepsilon, R) + (U_2 + \kappa) T_{p,q}(\varepsilon, R) + AK_{p,q}(2, N) + B < \infty. \end{aligned}$$

Therefore, for $(x, y) \in A_1$, the family $\{T_{m,l}(x, y)\}_{m \geq m_0}$ is uniformly integrable w.r.t. $P_X \otimes P_X$. Consequently,

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} T_{m,l}(x, y) p(x) p(y) dx dy \rightarrow \int_{\mathbb{R}^d \times \mathbb{R}^d} a_{l,d}(x) a_{l,d}(y) p(x) p(y) dx dy, \quad m \rightarrow \infty.$$

Thus

$$\mathbb{E} \log \phi_{m,l}(1) \log \phi_{m,l}(2) \rightarrow \left(\psi(l) - \log V_d - \int_{\mathbb{R}^d} \log q(x) p(x) dx \right)^2, \quad m \rightarrow \infty. \quad (81)$$

On the other hand, taking also into account (23), we come to the relation

$$\mathbb{E} \log \phi_{m,l}(1) \mathbb{E} \log \phi_{m,l}(2) \rightarrow \left(\psi(l) - \log V_d - \int_{\mathbb{R}^d} \log q(x) p(x) dx \right)^2. \quad (82)$$

Hence (81) and (82) imply that

$$\frac{2}{n^2} \sum_{1 \leq i < j \leq n} \text{cov}(\log \phi_{m,l}(i), \log \phi_{m,l}(j)) = \frac{n-1}{n} \text{cov}(\log \phi_{m,l}(1), \log \phi_{m,l}(2)) \rightarrow 0, \quad n, m \rightarrow \infty.$$

Step 5. Now we consider $\text{cov}(\log \zeta_{n,k}(i), \log \zeta_{n,k}(j))$ for $i \neq j$, where $i, j \in \{1, \dots, n\}$.

Similarly to Step 4, for $x, y \in \mathbb{R}^d$ and $u, w \geq 0$, introduce a conditional distribution function

$$\tilde{\Phi}_{n,k,x,y}^{i,j}(u, w) := \mathbb{P}(\zeta_{n,k}(i) \leq u, \zeta_{n,k}(j) \leq w | X_i = x, X_j = y)$$

$$= \mathbb{P}\left(\left\|x - X_{(k)}(x, \mathbb{X}_n^{i,j} \cup \{y\})\right\| \leq r_{n-1}(u), \left\|y - X_{(k)}(y, \mathbb{X}_n^{i,j} \cup \{x\})\right\| \leq r_{n-1}(w)\right) \\ := \mathbb{P}(\tilde{\eta}_{n,k,x}^{y,i,j} \leq u, \tilde{\eta}_{n,k,y}^{x,i,j} \leq w), \quad u, w \geq 0,$$

where $\mathbb{X}_n^{i,j} = \mathbb{X}_n \setminus \{X_i, X_j\}$, $\tilde{\eta}_{n,k,x}^{y,i,j} := (n-1) \left\|x - X_{(k)}(x, \mathbb{X}_n^{i,j} \cup \{y\})\right\|^d$. We write $\tilde{\Phi}_{n,k,x,y}(u, w)$, $\tilde{\eta}_{n,k,x}^y$ and $\tilde{\eta}_{n,k,y}^x$ instead of $\tilde{\Phi}_{n,k,x,y}^{i,j}(u, w)$, $\tilde{\eta}_{n,k,x}^{y,i,j}$, $\tilde{\eta}_{n,k,y}^{x,i,j}$, respectively, (because $X_1, X_2, \dots$ are i.i.d. random vectors). Moreover, $\tilde{\Phi}_{n,k,x,y}(u, w)$ is the distribution function of a random vector $\tilde{\eta}_{n,k,x,y} := (\tilde{\eta}_{n,k,x}^y, \tilde{\eta}_{n,k,y}^x)$ and the regular conditional distribution function of a random vector $(\zeta_{n,k}(i), \zeta_{n,k}(j))$ given $(X_i, X_j) = (x, y)$. One has

$$\tilde{\Phi}_{n,k,x,y}(u, w) = 1 - \mathbb{P}\left(\left\|x - X_{(k)}(x, \mathbb{X}_n^{i,j} \cup \{y\})\right\| > r_{n-1}(u)\right) \\ - \mathbb{P}\left(\left\|y - X_{(k)}(y, \mathbb{X}_n^{i,j} \cup \{x\})\right\| > r_{n-1}(w)\right) \\ + \mathbb{P}\left(\left\|x - X_{(k)}(x, \mathbb{X}_n^{i,j} \cup \{y\})\right\| > r_{n-1}(u), \left\|y - X_{(k)}(y, \mathbb{X}_n^{i,j} \cup \{x\})\right\| > r_{n-1}(w)\right).$$

Introduce

$$\tilde{R}_{N,2}(x) := \int_{\|x-y\| \geq e_{[N]}} G_N(\log^2 \|x-y\|) p(y) dy,$$

$\tilde{A}_{p,2}(G_N) := \{x \in S(p) : \tilde{R}_{N,2}(x) < \infty\}$ and $\tilde{A} := \Lambda(p) \cap S(p) \cap D_p(R) \cap \tilde{A}_{p,2}(G_N)$, where the first three sets appeared in Theorem 1 proof (Step 5) Then $\mathbb{P}_X(S(p) \setminus \tilde{A}_{p,2}(G_N)) = 0$ since $K_{p,p}(2, N) < \infty$. It is easily seen that $\mathbb{P}_X(\tilde{A}) = 1$.

Take $\tilde{A}_1 := \{(x, y) : x \in \tilde{A}, y \in \tilde{A}, x \neq y\}$ and $\tilde{A}_2 := \{(x, y) : x \in \tilde{A}, y \in \tilde{A}, x = y\}$. Evidently, $(\mathbb{P}_X \otimes \mathbb{P}_X)(\tilde{A}_1) = 1$ and $(\mathbb{P}_X \otimes \mathbb{P}_X)(\tilde{A}_2) = 0$. For any $a > 0$, $r_m(a) \rightarrow 0$, as $m \rightarrow \infty$. Hence, for $(x, y) \in \tilde{A}_1$, one can find $\tilde{n}_5 = \tilde{n}_5(u, w, \|x-y\|) = 2 + \left\lceil \left(\frac{2}{\|x-y\|}\right)^d \max\{u, w\} \right\rceil$ such that $r_{n-1}(u) < \frac{\|x-y\|}{2}$, $r_{n-1}(w) < \frac{\|x-y\|}{2}$ if $n \geq \tilde{n}_5$. Then $B(x, r_{n-1}(u)) \cap B(y, r_{n-1}(w)) = \emptyset$ if $n \geq \tilde{n}_5(u, w, \|x-y\|)$. Thus, for $n \geq \tilde{n}_6 := \max\{\tilde{n}_5, 2k\}$, one has

$$\tilde{\Phi}_{n,k,x,y}(u, w) = 1 - \sum_{s_1=0}^{k-1} \binom{n-2}{s_1} (V_{n-1,x}(u))^{s_1} (1 - V_{n-1,x}(u))^{n-2-s_1} \\ - \sum_{s_2=0}^{k-1} \binom{n-2}{s_2} (V_{n-1,y}(w))^{s_2} (1 - V_{n-1,y}(w))^{n-2-s_2} \\ + \sum_{s_1=0}^{k-1} \sum_{s_2=0}^{k-1} \frac{(n-2)!}{s_1! s_2! (n-2-s_1-s_2)!} (V_{n-1,x}(u))^{s_1} (V_{n-1,y}(w))^{s_2} (1 - V_{n-1,x}(u) - V_{n-1,y}(w))^{n-2-s_1-s_2}.$$

Therefore, for each fixed $(x, y) \in \tilde{A}_1$, $u, w \geq 0$, we get, as $n \rightarrow \infty$,

$$\tilde{\Phi}_{n,k,x,y}(u, w) \rightarrow 1 - \sum_{s_1=0}^{k-1} \frac{(V_d u p(x))^{s_1}}{s_1!} e^{-V_d u p(x)} - \sum_{s_2=0}^{k-1} \frac{(V_d w p(y))^{s_2}}{s_2!} e^{-V_d w p(y)} \\ + \sum_{s_1=0}^{k-1} \sum_{s_2=0}^{k-1} \frac{(V_d u p(x))^{s_1}}{s_1!} \frac{(V_d w p(y))^{s_2}}{s_2!} e^{-V_d (u p(x) + w p(y))} \\ = \left(1 - \sum_{s_1=0}^{k-1} \frac{(V_d u p(x))^{s_1}}{s_1!} e^{-V_d u p(x)}\right) \left(1 - \sum_{s_2=0}^{k-1} \frac{(V_d w p(y))^{s_2}}{s_2!} e^{-V_d w p(y)}\right) = \tilde{F}_{k,x}(u) \tilde{F}_{k,y}(w) \\ := \tilde{\Phi}_{k,x,y}(u, w).$$

Here $\tilde{\Phi}_{k,x,y}(\cdot, \cdot)$ denotes the distribution function of a vector $\tilde{\eta}_{k,x,y} := (\tilde{\zeta}_{k,x}, \tilde{\zeta}_{k,y})$. The components of $\tilde{\eta}_{k,x,y}$ are independent, $\tilde{\zeta}_{k,x} \sim \Gamma(V_d p(x), k)$ and $\tilde{\zeta}_{k,y} \sim \Gamma(V_d p(y), k)$. Consequently, for each fixed $(x, y) \in \tilde{A}_1$, we have shown that $\tilde{\eta}_{n,k,x,y} \xrightarrow{law} \tilde{\eta}_{k,x,y}$ as $n \rightarrow \infty$. Therefore, for such (x, y) ,

$$\log \tilde{\eta}_{n,k,x}^y \log \tilde{\eta}_{n,k,y}^x \xrightarrow{law} \log \tilde{\zeta}_{k,x} \log \tilde{\zeta}_{k,y}, \quad n \rightarrow \infty. \quad (83)$$

Here we take strictly positive versions of the random variables under consideration. In a way similar to (77), for $i, j \in \{1, \dots, n\}$, $i \neq j$, we write

$$\begin{aligned} \mathbb{E}(\log \tilde{\eta}_{n,k,x}^y \log \tilde{\eta}_{n,k,y}^x) &= \int_{(0,\infty) \times (0,\infty)} \log u \log w d\tilde{\Phi}_{n,k,x,y}(u, w) \\ &= \mathbb{E}(\log \zeta_{n,k}(i) \log \zeta_{n,k}(j) | X_i = x, X_j = y). \end{aligned}$$

Since $\tilde{\zeta}_{k,x}$ and $\tilde{\zeta}_{k,y}$ are independent, write $\mathbb{E}(\log \tilde{\zeta}_{k,x} \log \tilde{\zeta}_{k,y}) = \mathbb{E} \log \tilde{\zeta}_{k,x} \mathbb{E} \log \tilde{\zeta}_{k,y} = b_{k,d}(x)b_{k,d}(y)$, where $b_{k,d}(z) := \psi(k) - \log V_d - \log p(z)$, $z \in \mathbb{R}^d$.

For any fixed $M > 0$, consider $\tilde{A}_{1,M} := \{(x, y) \in \tilde{A}_1 : \|x - y\| > M\}$. Now our aim is to verify that, for each $(x, y) \in \tilde{A}_{1,M}$,

$$\mathbb{E}(\log \zeta_{n,k}(1) \log \zeta_{n,k}(2) | X_1 = x, X_2 = y) \rightarrow b_{k,d}(x)b_{k,d}(y). \quad (84)$$

Equivalently, we can prove, for each $(x, y) \in \tilde{A}_{1,M}$, that

$$\mathbb{E} \log \tilde{\eta}_{n,k,x}^y \log \tilde{\eta}_{n,k,y}^x \rightarrow \mathbb{E} \log \tilde{\zeta}_{k,x} \log \tilde{\zeta}_{k,y}, \quad n \rightarrow \infty. \quad (85)$$

The idea that we consider only $(x, y) \in \tilde{A}_{1,M}$ is principal for the further proof.

Part (5a). We are going to establish that, for $(x, y) \in \tilde{A}_{1,M}$, a family $\{\log \tilde{\eta}_{n,k,x}^y \log \tilde{\eta}_{n,k,y}^x\}_{n \geq \tilde{n}_0}$ is uniformly integrable, where $\tilde{n}_0 \in \mathbb{N}$ is independent of x, y , but might depend on M . Then, due to (83), the relation (85) would be valid for such (x, y) as well. As we have seen, the function $G_N(\cdot)$ is nondecreasing and convex. Hence

$$\mathbb{E}G_N(|\log \tilde{\eta}_{n,k,x}^y \log \tilde{\eta}_{n,k,y}^x|) \leq \frac{1}{2} \left(\mathbb{E}G_N(\log^2 \tilde{\eta}_{n,k,x}^y) + \mathbb{E}G_N(\log^2 \tilde{\eta}_{n,k,y}^x) \right). \quad (86)$$

Let us consider, for instance, $\mathbb{E}G_N(\log^2 \tilde{\eta}_{n,k,x}^y)$. Alike Step 2 we can write

$$\begin{aligned} \mathbb{E}G_N(\log^2 \tilde{\eta}_{n,k,x}^y) &= \int_{(0, \frac{1}{\rho(N)})} \tilde{F}_{n,k,x}^y(u) (-h_N(u)) du + \int_{(\rho(N), \infty)} (1 - \tilde{F}_{n,k,x}^y(u)) h_N(u) du \\ &:= I_1(n, x, y) + I_2(n, x, y), \end{aligned}$$

where

$$\begin{aligned} \tilde{F}_{n,k,x}^y(u) &:= \mathbb{P}(\tilde{\eta}_{n,k,x}^y \leq u) = 1 - \mathbb{P}(\|x - X_{(k)}(x, \mathbb{X}_n^{i,j} \cup \{y\})\| > r_{n-1}(u)) \\ &= \mathbb{I}\{\|x - y\| > r_{n-1}(u)\} \left(1 - \sum_{s=0}^{k-1} \binom{n-2}{s} (V_{n-1,x}(u))^s (1 - V_{n-1,x}(u))^{n-2-s}\right) \\ &\quad + \mathbb{I}\{\|x - y\| \leq r_{n-1}(u)\} \left(1 - \sum_{s=0}^{k-2} \binom{n-2}{s} (V_{n-1,x}(u))^s (1 - V_{n-1,x}(u))^{n-2-s}\right), \end{aligned} \quad (87)$$

As usual a sum over empty set is equal to 0 (for $k = 1$).

If $u \in (0, \frac{1}{\rho(N)})$, where $\rho(N) := \exp\{\sqrt{e_{[N-1]}}\}$ and $n \geq \tilde{n}_1 := \left\lceil \frac{1}{\rho(N)M^d} \right\rceil + 1$, then $r_{n-1}(u) \leq M$. Thus, $r_{n-1}(u) < \|x - y\|$ if $(x, y) \in \tilde{A}_{1,M}$. In view of (87) and similarly to (38), one has

$$\tilde{F}_{n,k,x}^y(u) \leq \left(\frac{n-2}{n-1}\right)^\varepsilon (M_p(x, R) V_d u)^\varepsilon \leq (M_p(x, R))^\varepsilon V_d^\varepsilon u^\varepsilon$$

for $(x, y) \in \tilde{A}_{1,M}$, $u \in \left(0, \frac{1}{\rho(N)}\right]$, $n \geq \max\{\tilde{n}_1(M), \tilde{n}_2(R)\}$, here $\tilde{n}_2(R) := \max\left\{\left\lceil \frac{1}{\rho(N)R^d} \right\rceil, k+1\right\}$. So, $I_1(n, x, y) \leq U_1(\varepsilon, N, d)(M_p(x, R))^\varepsilon$ for $(x, y) \in \tilde{A}_{1,M}$ and $n \geq \max\{\tilde{n}_1(M), \tilde{n}_2(R)\}$. Moreover, for all $u \geq 0$, in view of (87) it holds

$$1 - \tilde{F}_{n,k,x}^y(u) \leq \sum_{s=0}^{k-1} \binom{n-2}{s} (V_{n-1,x}(u))^s (1 - V_{n-1,x}(u))^{n-2-s}.$$

The same reasoning as was used in Theorem 1 proof (Step 3, Part (3b)) leads to the inequalities

$$\begin{aligned} 1 - \tilde{F}_{n,k,x}^y(u) &\leq S_1(k)(1 - S_2 V_{n-1,x}(u))^{n-2} \leq S_1 \exp\{-S_2(n-2)V_{n-1,x}(u)\} \\ &\leq S_1 \exp\left\{-\frac{n-2}{n-1} S_2 V_d u m_p(x, R)\right\} \leq S_1 \left(\frac{S_2}{2} V_d u m_p(x, R)\right)^{-\varepsilon} \end{aligned}$$

for all $n \geq \max\{\tilde{n}_3(R), 3\}$. Then similarly to (70), the relation

$$\text{EG}_N(\log^2 \tilde{\eta}_{n,k,x}^y) \leq U_1(M_p(x, R))^\varepsilon + (\tilde{U}_2 + \kappa)(m_p(x, R))^{-\varepsilon} + A \tilde{R}_{N,2}(x) + B := \tilde{C}_1(x) < \infty$$

is valid for all $(x, y) \in \tilde{A}_{1,M}$ and $n \geq \tilde{n}_0(M) := \max\{\tilde{n}_1(M), \tilde{n}_2, \tilde{n}_3, \tilde{n}_4(\kappa), 3\}$. Here $U_1, \tilde{U}_2, \kappa, A, B$ do not depend on x and y . The term $\text{EG}_N(\log^2 \tilde{\eta}_{n,k,y}^x)$ can be treated in the above manner. Thus, in view of (86), one has

$$\begin{aligned} \text{EG}_N(|\log \tilde{\eta}_{n,k,x}^y \log \tilde{\eta}_{n,k,y}^x|) &\leq \frac{U_1}{2} \left((M_p(x, R))^\varepsilon + (M_p(y, R))^\varepsilon \right) \\ &+ \frac{U_2 + \kappa}{2} \left((m_p(x, R))^{-\varepsilon} + (m_p(y, R))^{-\varepsilon} \right) + \frac{A}{2} \left(\tilde{R}_{N,2}(x) + \tilde{R}_{N,2}(y) \right) + B := \tilde{C}_1(x, y). \end{aligned} \quad (88)$$

Therefore, for any $(x, y) \in \tilde{A}_{1,M}$, a family $\{\log \tilde{\eta}_{n,k,x}^y \log \tilde{\eta}_{n,k,y}^x\}_{n \geq \tilde{n}_0}$ is uniformly integrable. Thus, we come to (84) for $(x, y) \in \tilde{A}_{1,M}$.

Part (5b). Now we return to upper bound for $\text{cov}(\log \zeta_{n,k}(1), \log \zeta_{n,k}(2))$. Set

$$\tilde{T}_{n,k}(x, y) := \mathbb{E}(\log \zeta_{n,k}(1) \log \zeta_{n,k}(2) | X_1 = x, X_2 = y) = \mathbb{E} \log \tilde{\eta}_{n,k,x}^y \log \tilde{\eta}_{n,k,y}^x$$

for all $(x, y) \in \tilde{A}_1$. Validity of (84) is equivalent to the following relation: for any $(x, y) \in \tilde{A}_{1,M}$, $\tilde{T}_{n,k}(x, y) \rightarrow b_{k,d}(x)b_{k,d}(y)$, as $n \rightarrow \infty$. Take any $(x, y) \in \tilde{A}_1$. For each $M > 0$, it was shown that

$$\tilde{T}_{n,k}(x, y) \mathbb{I}\{\|x - y\| > M\} \rightarrow b_{k,d}(x)b_{k,d}(y) \mathbb{I}\{\|x - y\| > M\}, \quad n \rightarrow \infty.$$

Note that

$$\begin{aligned} G_N(|\tilde{T}_{n,k}(x, y)| \mathbb{I}\{\|x - y\| > M\}) &\leq G_N(|\tilde{T}_{n,k}(x, y)|) = G_N(|\mathbb{E} \log \tilde{\eta}_{n,k,x}^y \log \tilde{\eta}_{n,k,y}^x|) \\ &\leq G_N(\mathbb{E} |\log \tilde{\eta}_{n,k,x}^y \log \tilde{\eta}_{n,k,y}^x|) \leq \text{EG}_N(|\log \tilde{\eta}_{n,k,x}^y \log \tilde{\eta}_{n,k,y}^x|). \end{aligned} \quad (89)$$

Due to (88) and (89) one can conclude that, for all $n \geq \tilde{n}_0$,

$$\begin{aligned} &\int_{\mathbb{R}^d \times \mathbb{R}^d} G_N(|\tilde{T}_{n,k}(x, y)| \mathbb{I}\{\|x - y\| > M\}) p(x) p(y) dx dy \\ &\leq U_1 \int_{\mathbb{R}^d} M_p^\varepsilon(x, R) p(x) dx + (\tilde{U}_2 + \kappa) \int_{\mathbb{R}^d} m_p^{-\varepsilon}(x, R) p(x) dx + A \int_{\mathbb{R}^d} \tilde{R}_{N,2}(x) p(x) dx + B \\ &= U_1 Q_{p,p}(\varepsilon, R) + (\tilde{U}_2 + \kappa) T_{p,p}(\varepsilon, R) + A K_{p,p}(2, N) + B < \infty. \end{aligned}$$

Therefore, for $(x, y) \in \tilde{A}_1$, the family $\{\tilde{T}_{n,k}(x, y)\mathbb{I}\{\|x - y\| > M\}\}_{n \geq \tilde{n}_0}$ is uniformly integrable w.r.t. $P_X \otimes P_X$. Hence, by virtue of (84), for each $M > 0$,

$$\int_{D(M)} \tilde{T}_{n,k}(x, y)p(x)p(y) dx dy \rightarrow \int_{D(M)} b_{k,d}(x)b_{k,d}(y)p(x)p(y) dx dy, \quad n \rightarrow \infty, \quad (90)$$

where $D(M) := \{x, y \in \mathbb{R}^d, \|x - y\| > M\}$. Now we turn to the case $\|x - y\| \leq M$. One has $\bigcap_{s=1}^{\infty} \{\|X_1 - X_2\| \leq \frac{1}{s}\} = \{X_1 = X_2\}$ and $P(X_1 = X_2) = 0$ as X_1 and X_2 are independent and have a density $p(x)$ w.r.t. the Lebesgue measure μ . Then in view of continuity of a probability measure it holds that $P(\|X_1 - X_2\| \leq M) \rightarrow 0$, as $M \rightarrow 0$. Taking into account that, for an integrable function h , $\int_C h dP \rightarrow 0$ as $P(C) \rightarrow 0$, we get

$$E(\log \zeta_{n,k}(1) \log \zeta_{n,k}(2) \mathbb{I}\{\|X_1 - X_2\| \leq M\}) \rightarrow 0, \quad M \rightarrow 0,$$

since $E \log \zeta_{n,k}(1) \log \zeta_{n,k}(2) \leq \frac{1}{2} (E \log^2 \zeta_{n,k}(1) + E \log^2 \zeta_{n,k}(2)) < \infty$ (the proof is similar to the establishing that $E \log^2 \phi_{m,l}(1) < \infty$). Thus, for any $\gamma > 0$, one can find $M_1 = M_1(\gamma) > 0$ such that, for all $M \in (0, M_1]$ and $n \geq \tilde{n}_0$,

$$\left| \int_{\mathbb{R}^{2d} \setminus D(M)} \tilde{T}_{n,k}(x, y)p(x)p(y) dx dy \right| = |E \log \zeta_{n,k}(1) \log \zeta_{n,k}(2) \mathbb{I}\{\|X_1 - X_2\| \leq M\}| < \frac{\gamma}{3}. \quad (91)$$

Also there exists $M_2 = M_2(\gamma) > 0$ such that, for all $M \in (0, M_2]$,

$$\left| \int_{\mathbb{R}^{2d} \setminus D(M)} b_{k,d}(x)b_{k,d}(y)p(x)p(y) dx dy \right| < \frac{\gamma}{3}. \quad (92)$$

Take $M(\gamma) := \min\{M_1(\gamma), M_2(\gamma)\}$. Due to (90) there is $\tilde{n}_7(M(\gamma), \gamma)$ such that $n \geq \max\{\tilde{n}_0, \tilde{n}_7(M(\gamma), \gamma)\}$ entails the following inequality

$$\left| \int_{D(M)} \tilde{T}_{n,k}(x, y)p(x)p(y) dx dy - \int_{D(M)} b_{k,d}(x)b_{k,d}(y)p(x)p(y) dx dy \right| < \frac{\gamma}{3}. \quad (93)$$

So, in view of (91)–(93), for any $\gamma > 0$, there is $M(\gamma) > 0$ such that, for all n large enough, i.e., $n \geq \max\{\tilde{n}_0, \tilde{n}_7(M(\gamma), \gamma)\}$, one has

$$\left| \int_{\mathbb{R}^d \times \mathbb{R}^d} \tilde{T}_{n,k}(x, y)p(x)p(y) dx dy - \int_{\mathbb{R}^d \times \mathbb{R}^d} b_{k,d}(x)b_{k,d}(y)p(x)p(y) dx dy \right| < \gamma. \quad (94)$$

By virtue of the formula

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} b_{k,d}(x)b_{k,d}(y)p(x)p(y) dx dy = \left(\psi(k) - \log V_d - \int_{\mathbb{R}^d} (\log p(x))p(x) dx \right)^2,$$

and taking into account (94) we deduce the limit relation, for $n \rightarrow \infty$,

$$E \log \zeta_{n,k}(1) \log \zeta_{n,k}(2) \rightarrow \left(\psi(k) - \log V_d - \int_{\mathbb{R}^d} (\log p(x))p(x) dx \right)^2.$$

Moreover, in view of (24) (see Step 5 of Theorem 1 proof), it follows that

$$E \log \zeta_{n,k}(1) E \log \zeta_{n,k}(2) \rightarrow \left(\psi(k) - \log V_d - \int_{\mathbb{R}^d} (\log p(x))p(x) dx \right)^2.$$

Therefore,

$$\frac{2}{n^2} \sum_{1 \leq i < j \leq n} \text{cov}(\log \zeta_{n,k}(i), \log \zeta_{n,k}(j)) = \frac{n-1}{n} \text{cov}(\log \zeta_{n,k}(1), \log \zeta_{n,k}(2)) \rightarrow 0, \quad n \rightarrow \infty.$$

Step 6. Here we complete the analysis of summands in (56). Reasoning as at Steps 1–3 shows that $\frac{1}{n^2} \sum_{i=1}^n \text{var}(\log \zeta_{n,k}(i)) = \frac{1}{n} \text{var}(\log \zeta_{n,k}(1)) \rightarrow 0$ since

$$\text{var}(\log \zeta_{n,k}(i)) = \text{var}(\log \zeta_{n,k}(1)) \rightarrow v_k < \infty$$

for each $i \in \mathbb{N}$, as $n \rightarrow \infty$. It remains to prove that $\frac{2}{n^2} \sum_{i,j=1}^n \text{cov}(\log \phi_{m,l}(i), \log \zeta_{n,k}(j)) \rightarrow 0$, as $n, m \rightarrow \infty$.

For $i = 1, \dots, n$, one has $|\text{cov}(\log \phi_{m,l}(i), \log \zeta_{n,k}(i))| \leq (\text{var}(\log \phi_{m,l}(1)) \text{var}(\log \zeta_{n,k}(1)))^{\frac{1}{2}} < \infty$ for all n, m large enough. So, it suffices to show that

$$\frac{1}{n^2} \sum_{i,j=1, \dots, n; i \neq j} \text{cov}(\log \phi_{m,l}(i), \log \zeta_{n,k}(j)) \rightarrow 0, \quad n, m \rightarrow \infty.$$

For $i, j = 1, \dots, n, i \neq j, u, w \geq 0, x, y \in \mathbb{R}^d$, let us introduce a conditional distribution function

$$\begin{aligned} & \mathbb{P}(\phi_{m,l}(i) \leq u, \zeta_{n,k}(j) \leq w | X_i = x, X_j = y) \\ &= \mathbb{P}\left(\|x - Y_{(l)}(x, \mathbb{Y}_m)\| \leq r_m(u), \|y - X_{(k)}(y, \mathbb{X}_n^{i,j} \cup \{x\})\| \leq r_{n-1}(w)\right) \\ &= \mathbb{P}\left(\|x - Y_{(l)}(x, \mathbb{Y}_m)\| \leq r_m(u)\right) \mathbb{P}\left(\|y - X_{(k)}(y, \mathbb{X}_n^{i,j} \cup \{x\})\| \leq r_{n-1}(w)\right) \\ &= \left(1 - \sum_{s_1=0}^{l-1} \binom{m}{s_1} (W_{m,x}(u))^{s_1} (1 - W_{m,x}(u))^{m-s_1}\right) \\ &\quad \cdot \left(\mathbb{I}\{\|x - y\| > r_{n-1}(w)\} \left(1 - \sum_{s=0}^{k-1} \binom{n-2}{s} (V_{n-1,y}(w))^s (1 - V_{n-1,y}(w))^{n-2-s}\right)\right. \\ &\quad \left.+ \mathbb{I}\{\|x - y\| \leq r_{n-1}(w)\} \left(1 - \sum_{s=0}^{k-2} \binom{n-2}{s} (V_{n-1,y}(w))^s (1 - V_{n-1,y}(w))^{n-2-s}\right)\right). \end{aligned}$$

We used that $\{\mathbb{X}_n, \mathbb{Y}_m\}$ is a collection of independent vectors. Now we combine the estimates obtained at Steps 4 and 5 of Theorem 2 proof to verify that, for $i, j \in \{1, \dots, n\}$ and $i \neq j$, $\text{cov}(\log \phi_{m,l}(i), \log \zeta_{n,k}(j)) = \text{cov}(\log \phi_{m,l}(1), \log \zeta_{n,k}(2)) \rightarrow 0, n, m \rightarrow \infty$.

Thus, we have established that $\text{var}(\hat{D}_{n,m}(k, l)) \rightarrow 0$ as $n, m \rightarrow \infty$, hence (16) holds. The proof of Theorem 2 is complete.

6. Conclusions

The aim of this paper is to provide wide conditions ensuring the asymptotic unbiasedness and mean square consistency for statistical estimates of the Kullback–Leibler divergence proposed in [31]. We do not impose restrictions on the smoothness of the densities under consideration and do not assume that the densities have bounded supports. Thus, in particular one can apply our results to various mixtures of distributions, for instance, to mixture of any nondegenerate normal laws in $\mathbb{R}^d$ (Corollary 4). As a byproduct we relax conditions in our recent analysis of the Kozachenko–Leonenko type estimators for the Shannon differential entropy [55] and use these conditions in estimating the cross-entropy as well. Observe that the integral functional $K_{p,q}$ appearing in Theorems 1–3 involves the function $G_N(t)$ which is close to a function t when parameter N is large enough. Thus, we impose essentially less restrictive condition than one requiring a function $G(t) = t^{1+\nu}$ for some $\nu > 0$ instead of $G_N(t)$. Even for the latter choice of G our results provide the first valid proof without appealing to the Fatou lemma (the long standing problem to obtain correct proofs was discussed in Introduction). An interesting and hard problem for future research is to find the class of functions $\varphi: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that one can replace $G_N(t)$ in expression of $K_{p,q}$ by $G(t) = t\varphi(t)$, where $\varphi(t) \rightarrow \infty$, as $t \rightarrow \infty$, and keep the validity of established theorems. Here one can see the analogy with investigation of fluctuations of sums of random variables or Brownian motion by

G.H.Hardy, H.D.Steinhaus, A.Ya.Khinchin, A.N.Kolmogorov, I.G.Petrovski, W.Feller and other researchers. The increasing precision on the way of description of the upper and lower functions has led to the law of the iterated logarithm and its generalizations. Another deep problem is to provide sharp conditions of CLT validity for estimates of the Kullback - Leibler divergence.

Beside pure theoretical aspects the estimates of entropy and related functionals have diverse applications. In [5], the estimates of the Kullback - Leibler divergence are applied to the change-point detection in time series. That issue is important, e.g., in analysis of stochastic financial models. Moreover, it is interesting to study the spatial variant of this problem. Namely, in [55,56] statistical estimates of entropy and scan-statistics (see, e.g., [57]) were employed for identification of inhomogeneities of fiber materials. In [58], the Kullback–Leibler divergence estimators are used to identify multivariate spatial clusters in the Bernoulli model. A modification of the latter paper idea can also be applied to analysis of the fiber structures. Such structures in $\mathbb{R}^d$ can be modeled by spatial point stochastic process to specify the locations of the centers of fibers (segments). A certain law on the unit sphere of $\mathbb{R}^d$ can be used to model their directions. The length of fibers can be fixed or follow some distribution on $\mathbb{R}_+$. Since various scan domains could contain random number of observations the development of present results will be applied along with the theory of random sums of random variables. The latter theory (see, e.g., [59]) is essential in this case. Moreover, we intend to employ the studied estimators in the feature selection theory, actively used in Genome-wide association studies (GWAS), see, e.g., [16,17,22]. In this regard statistical estimates of the mutual information were proposed, see, e.g., [12]. Note also an important problem of stability analysis of constructing, by means of statistical estimates of the mutual information, a sub-collection of relevant (in a sense) factors determining a random response. The above mentioned applications will be considered in separate publications, supplemented with computer simulations and illustrative graphs.

Author Contributions: Conceptualization, A.B. and D.D.; validation, A.B. and D.D.; writing—original draft preparation, A.B. and D.D.; writing—review and editing, A.B. and D.D.; supervision, A.B.; project administration, A.B.; funding acquisition, A.B. All authors have read and agreed to the published version of the manuscript.

Funding: The work of the first author is supported by the Russian Science Foundation under grant 14-21-00162 and performed at the Steklov Mathematical Institute of Russian Academy of Sciences.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors are grateful to Professor A.Tsybakov for useful discussions. We also thank the Reviewers for remarks and suggestions improving the exposition.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Proofs of Lemmas 1–3 are similar to the proofs of Lemma 2.5, 3.1 and 3.2 in [35]. We provide them for the sake of completeness.

Proof of Lemma 1. (1) Note that $\log \|x - y\| > e_{[N-1]} \geq 1$ if $\|x - y\| > e_{[N]}$ and $N \in \mathbb{N}$. Hence, for such x, y , one has $(\log \|x - y\|)^v \leq (\log \|x - y\|)^{v_0}$ if $v \in (0, v_0]$. If $N \geq N_0$ then $G_N(u) \leq G_{N_0}(u)$ for $u \geq e_{[N-1]} \geq e_{[N_0-1]}$. Thus, $K_{p,q}(v, N) \leq K_{p,q}(v_0, N_0) < \infty$ for $v \in (0, v_0]$ and any integer $N \geq N_0$.

(2) Assume that $Q_{p,q}(\varepsilon_1, R_1) < \infty$. Consider $Q_{p,q}(\varepsilon_1, R)$ where $R > 0$. If $0 < R \leq R_1$ then, for each $x \in \mathbb{R}^d$, in accordance with the definition of M_q one has $M_q(x, R) \leq M_q(x, R_1)$. Consequently, $Q_{p,q}(\varepsilon_1, R) \leq Q_{p,q}(\varepsilon_1, R_1) < \infty$. Let now $R > R_1$. One has

$$\begin{aligned} M_q(x, R) &\leq \max \left\{ M_q(x, R_1), \sup_{R_1 < r \leq R} \frac{\int_{B(x, R_1)} q(x) dx + \int_{B(x, r) \setminus B(x, R_1)} q(x) dx}{\mu(B(x, r))} \right\} \\ &\leq \max \left\{ M_q(x, R_1), M_q(x, R_1) + \frac{1}{\mu(B(x, R_1))} \right\} = M_q(x, R_1) + \frac{1}{\mu(B(x, R_1))}. \end{aligned}$$

Therefore,

$$\begin{aligned} Q_{p,q}(\varepsilon_1, R) &= \int_{\mathbb{R}^d} (M_q(x, R))^{\varepsilon_1} p(x) dx \leq \int_{\mathbb{R}^d} \left(M_q(x, R_1) + \frac{1}{R_1^d V_d} \right)^{\varepsilon_1} p(x) dx \\ &\leq \max\{1, 2^{\varepsilon_1-1}\} \left(Q_{p,q}(\varepsilon_1, R_1) + (R_1^d V_d)^{-\varepsilon_1} \right) < \infty. \end{aligned}$$

Suppose now that $Q_{p,q}(\varepsilon_1, R) < \infty$ for some $\varepsilon_1 > 0$ and $R > 0$. Then, for each $\varepsilon \in (0, \varepsilon_1]$, the Lyapunov inequality leads to the estimate $Q_{p,q}(\varepsilon, R) \leq (Q_{p,q}(\varepsilon_1, R))^{\frac{\varepsilon}{\varepsilon_1}} < \infty$.

(3) Let $T_{p,q}(\varepsilon_2, R_2) < \infty$. Take $0 < R \leq R_2$. Then, for any $x \in \mathbb{R}^d$, according to the definition of m_q we get $0 \leq m_q(x, R_2) \leq m_q(x, R)$. Hence $T_{p,q}(\varepsilon_2, R) \leq T_{p,q}(\varepsilon_2, R_2) < \infty$. Consider $R > R_2$. For any $x \in \mathbb{R}^d$ and every $a > 0$, the function $I_q(x, r)$ is continuous in r on $(0, a]$. Next fix an arbitrary $x \in S(q) \cap \Lambda(q)$. We see that there exists $\lim_{r \rightarrow 0+} I_q(x, r) = q(x)$. For such x , set $I_q(x, 0) := q(x)$. Thus, $I_q(x, \cdot)$ is continuous on any segment $[0, a]$. Hence, one can find $\tilde{R}_2$ in $[0, R_2]$ such that $m_q(x, R_2) = I_q(x, \tilde{R}_2)$ and there exists R_0 in $[0, R]$ such that $m_q(x, R) = I_q(x, R_0)$. If $R_0 \leq R_2$ then $m_q(x, R) = m_q(x, R_2)$ (since $m_q(x, R) \leq m_q(x, R_2)$ for $R > R_2$ and $m_q(x, R) = I_q(x, R_0) \geq m_q(x, R_2)$ as $R_0 \in [0, R_2]$). Assume that $R_0 \in (R_2, R]$. Obviously $R_0 > 0$ as $R_2 > 0$. One has

$$\begin{aligned} m_q(x, R) &= I_q(x, R_0) = \frac{\int_{B(x, R_2)} q(y) dy + \int_{B(x, R_0) \setminus B(x, R_2)} q(y) dy}{\mu(B(x, R_0))} \\ &\geq \frac{\int_{B(x, R_2)} q(y) dy}{\mu(B(x, R_0))} = \frac{\mu(B(x, R_2))}{\mu(B(x, R_0))} I_q(x, R_2) \geq \frac{\mu(B(x, R_2))}{\mu(B(x, R_0))} m_q(x, R_2) \\ &= \left(\frac{R_2}{R_0} \right)^d m_q(x, R_2) \geq \left(\frac{R_2}{R} \right)^d m_q(x, R_2). \end{aligned}$$

Thus, in any case ($R_0 \in [0, R_2]$ or $R_0 \in (R_2, R]$) one has $m_q(x, R) \geq \left(\frac{R_2}{R} \right)^d m_q(x, R_2)$ as $R_2 < R$. Taking into account that $\mu(S(q) \setminus (S(q) \cap \Lambda(q))) = 0$ we deduce the inequality

$$T_{p,q}(\varepsilon_2, R) \leq \left(\frac{R}{R_2} \right)^{\varepsilon_2 d} T_{p,q}(\varepsilon_2, R_2) < \infty.$$

Assume now that $T_{p,q}(\varepsilon_2, R) < \infty$ for some $\varepsilon_2 > 0$ and $R > 0$. Then, for any $\varepsilon \in (0, \varepsilon_2]$, the Lyapunov inequality entails $T_{p,q}(\varepsilon, R) \leq (T_{p,q}(\varepsilon_2, R))^{\frac{\varepsilon}{\varepsilon_2}} < \infty$. This completes the proof. $\square$

Proof of Lemma 2. Begin with relation (1). Observe that if a function g is measurable and bounded on a finite interval $(a, b]$ and ν is a finite measure on the Borel subsets of $(a, b]$

then $\int_{(a,b]} g(x)v(dx)$ is finite. So applying the integration by parts formula (see, e.g., [33], p. 245), for each $a \in \left(0, \frac{1}{e_{[N]}}\right]$, we get

$$\begin{aligned} \int_{\left(a, \frac{1}{e_{[N]}}\right]} F(u)(-g_N(u)) du &= \int_{\left(a, \frac{1}{e_{[N]}}\right]} F(u)d(-G_N(-\log u)) \\ &= G_N(-\log a)F(a) + \int_{\left(a, \frac{1}{e_{[N]}}\right]} G_N(-\log u) dF(u). \end{aligned} \quad (\text{A1})$$

Assume now that $\int_{\left(0, \frac{1}{e_{[N]}}\right]} G_N(-\log u) dF(u) < \infty$. Then by the monotone convergence theorem

$$\lim_{a \rightarrow 0+} \int_{(0,a]} G_N(-\log u) dF(u) = 0. \quad (\text{A2})$$

Given $a > 0$ the following lower bound is obvious

$$\begin{aligned} \int_{(0,a]} G_N(-\log u) dF(u) &\geq G_N(-\log a) \int_{(0,a]} dF(u) \\ &= G_N(-\log a)(F(a) - F(0)) = G_N(-\log a)F(a) \geq 0. \end{aligned}$$

Therefore (A2) implies that

$$G_N(-\log a)F(a) \rightarrow 0, \quad a \rightarrow 0+. \quad (\text{A3})$$

By the Lebesgue monotone convergence theorem letting $a \rightarrow 0+$ in (A1) yields the desired relation (1) of our Lemma. Now we assume that

$$\int_{\left(0, \frac{1}{e_{[N]}}\right]} F(u)(-g_N(u)) du < \infty. \quad (\text{A4})$$

Hence from the equality $\int_{\left(0, \frac{1}{e_{[N]}}\right]} F(u)(-g_N(u)) du = \int_{\left(0, \frac{1}{e_{[N]}}\right]} F(u)d(-G_N(-\log u))$ we get $\lim_{b \rightarrow 0+} \int_{(0,b]} F(u) d(-G_N(-\log u)) = 0$ by monotone convergence theorem. Therefore, for any $c \in (0, b)$, we come to the inequalities

$$\begin{aligned} \int_{(0,b]} F(u)d(-G_N(-\log u)) &\geq \int_{(c,b]} F(u)d(-G_N(-\log u)) \\ &= -F(b)G_N(-\log b) + F(c)G_N(-\log c) + \int_{(c,b]} G_N(-\log u) dF(u) \\ &\geq F(c)G_N(-\log c) - F(b)G_N(-\log b) + (F(b) - F(c))G_N(-\log b) \\ &= F(c)G_N(-\log c) \left(1 - \frac{G_N(-\log b)}{G_N(-\log c)}\right). \end{aligned}$$

Let $c = b^2$ ($b \leq \frac{1}{e_{[N]}} < 1$). Then, for all positive b small enough,

$$1 - \frac{G_N(-\log b)}{G_N(-\log c)} = 1 - \frac{G_N(-\log b)}{G_N(-2\log b)} = 1 - \left(\frac{1}{2}\right) \frac{\log_{[N]}(-\log b)}{\log_{[N]}(-2\log b)} \geq \frac{1}{2}.$$

Thus $\int_{(0,b]} F(u)d(-G_N(-\log u)) \geq \frac{1}{2}F(b^2)G_N(-\log(b^2)) \geq 0$, so $F(b^2)G_N(-\log b^2) \rightarrow 0$ as $b \rightarrow 0$. Consequently we come to (A3) taking $a = b^2$. Then (A1) implies relation (1).

When one of the (nonnegative) integrals in (1) turns infinite while the other one is finite we come to a contradiction. Thus, (1) is established. In quite the same manner one can verify validity of relation (2), therefore further details can be omitted. $\square$

Proof of Lemma 3. Take $x \in S(q) \cap \Lambda(q)$ and $R > 0$. Suppose that $m_q(x, R) = 0$. Since the function $I_q(x, r)$ defined in (8) is continuous in $(x, r) \in \mathbb{R}^d \times (0, \infty)$, there exists $\tilde{R} \in [0, R]$ ($\tilde{R} = \tilde{R}(x, R)$) such that $m_q(x, R) = I_q(x, \tilde{R})$ ($I_q(x, 0) := \lim_{r \rightarrow 0+} I_q(x, r) = q(x)$ for any $x \in \Lambda(q)$ by continuity). If $\tilde{R} = 0$ then $m_q(x, r) = q(x) > 0$ as $x \in S(q) \cap \Lambda(q)$. Hence we have to deal with $\tilde{R} \in (0, R]$. If $I_q(x, \tilde{R}) = 0$ then $\int_{B(x, r)} q(y) dy = 0$ for any $0 < r \leq \tilde{R}$. Thus, (30) ensures that $q(x) = 0$. However, $x \in S(q) \cap \Lambda(q)$. So $m_q(x, R) > 0$ for $x \in S(q) \cap \Lambda(q)$. Thus, $S(q) \cap \Lambda(q) \subset D_q(R) := \{x \in S(q) : m_q(x, R) > 0\}$. It remains to note that $S(q) \setminus \Lambda(q) \subset \mathbb{R}^d \setminus \Lambda(q)$ and $\mu(\mathbb{R}^d \setminus \Lambda(q)) = 0$. Therefore $\mu(S(q) \setminus D_q(R)) = 0$. $\square$

Proof of Lemma 4. We will check that, for given $N \in \mathbb{N}$ and $\tau > 0$, there exist $a := a(\tau) \geq 0$ and $b := b(N, \tau) \geq 0$ such that, for any $c \geq 0$,

$$G_N(\tau c) \leq a G_N(c) + b. \quad (\text{A5})$$

For $c = 0$ the statement is obviously true. Let $c > 0$. It easily seen that $\frac{\log_{[N]}(\tau c)}{\log_{[N]}(c)} \rightarrow 1$ as $c \rightarrow \infty$. Hence one can find $c_0(N, \tau)$ such that, for all $c \geq c_0(N, \tau)$, the inequality $\frac{\log_{[N]}(\tau c)}{\log_{[N]}(c)} \leq 2$ is valid. Consequently, for $c \geq c_0(N, \tau)$,

$$\frac{G_N(\tau c)}{G_N(c)} = \frac{\tau c \log_{[N]}(\tau c)}{c \log_{[N]}(c)} \leq 2\tau := a(\tau).$$

For all $0 \leq c \leq c_0(N, \tau)$ we write $G_N(\tau c) \leq G_N(\tau c_0(N, \tau)) := b(N, \tau)$. Therefore, for any $c \geq 0$, we come to (A5). Thus, for any $\nu > 0$ and $x, y \in \mathbb{R}^d$, $x \neq y$, one has

$$G_N(|\log(\|x - y\|^d)|^\nu) = G_N(d^\nu |\log(\|x - y\|)|^\nu) \leq a(d^\nu) G_N(|\log(\|x - y\|)|^\nu) + b(N, d^\nu).$$

$\square$

Proof of Lemma 6. For $t \in [0, e_{[N-1]}]$, a function $G_N(t) \equiv 0$ is convex. We show that G_N is convex on $(e_{[N-1]}, \infty)$. Consider $t > e_{[N-1]}$. Write $\prod_{\emptyset} := 1$ and $\sum_{\emptyset} := 0$. Then, for $N \in \mathbb{N}$,

$$(G_N(t))' = \log_{[N]}(t) + \prod_{j=1}^{N-1} \frac{1}{\log_{[j]}(t)}.$$

Obviously, $\left(\frac{1}{\log_{[k]}(t)}\right)' = -\frac{1}{t \log_{[k]}^2(t)} \prod_{s=1}^{k-1} \frac{1}{\log_{[s]}(t)}$, $k \in \mathbb{N}$. Thus, for $t > e_{[N-1]}$, we get

$$\begin{aligned} (G_N(t))'' &= \frac{1}{t} \prod_{j=1}^{N-1} \frac{1}{\log_{[j]}(t)} + \sum_{k=1}^{N-1} \left(-\frac{1}{t \log_{[k]}^2(t)} \prod_{s=1}^{k-1} \frac{1}{\log_{[s]}(t)} \prod_{j \in \{1, \dots, N-1\}, j \neq k} \frac{1}{\log_{[j]}(t)} \right) \\ &= \frac{1}{t} \left(\prod_{j=1}^{N-1} \frac{1}{\log_{[j]}(t)} \right) \left(1 - \sum_{k=1}^{N-1} \prod_{s=1}^k \frac{1}{\log_{[s]}(t)} \right). \end{aligned}$$

For $N = 1$ and $t > 0$, we have $(G_1(t))'' = \frac{1}{t} > 0$. Take now $N > 1$. Clearly, for $t > e_{[N-1]}$, one has $\frac{1}{t} \prod_{j=1}^{N-1} \frac{1}{\log_{[j]}(t)} > 0$ because $\log_{[j]}(t) > \log_{[j]}(e_{[N-1]}) = e_{[N-1-j]} \geq 1 > 0$ when $1 \leq j \leq N-1$. Observe also that

$$\sum_{k=1}^{N-1} \prod_{s=1}^k \frac{1}{\log_{[s]}(t)} < \sum_{k=1}^{N-1} \prod_{s=1}^k \frac{1}{e_{[N-1-s]}} \leq \sum_{k=1}^{N-1} \frac{1}{e_{[N-2]}} = \frac{N-1}{e_{[N-2]}} \leq 1. \quad (\text{A6})$$

The last inequality is established by induction in N . Thus, in view of (A6), we have proved that, for all $t > e_{[N-1]}$ and $N \in \mathbb{N}$, the inequality $(G_N(t))'' > 0$ holds. Hence, the function $G_N(t)$ is (strictly) convex on $(e_{[N-1]}, \infty)$.

Let $h : [a, \infty) \rightarrow \mathbb{R}$ be a continuous nondecreasing function. If the restrictions of h to $[a, b]$ and (b, ∞) (where $a < b$) are convex functions then, in general, it is not true that h is convex on $[a, \infty)$. However, we can show that G_N is convex on $[0, \infty)$. Note that a function G_N is convex on $[e_{[N-1]}, \infty)$ since it is convex on $(e_{[N-1]}, \infty)$ and continuous on $[e_{[N-1]}, \infty)$. Take now any $z \in [0, e_{[N-1]}]$, $y \in (e_{[N-1]}, \infty)$ and $s \in [0, 1]$. Then

$$\begin{aligned} G_N(sz + (1-s)y) &\leq G_N(se_{[N-1]} + (1-s)y) \leq sG_N(e_{[N-1]}) + (1-s)G_N(y) \\ &= (1-s)G_N(y) = sG_N(z) + (1-s)G_N(y) \end{aligned}$$

as $G_N(z) = 0$. Thus, for each $N \in \mathbb{N}$, a function $G_N(\cdot)$ is convex on $\mathbb{R}_+$. $\square$

Proof of Corollary 3. The proof (i.e., checking the conditions of both Theorem 1 and 2) is quite similar to the proof of Corollary 2.11 in [35]. $\square$

Proof of Corollary 4. Take $f(x) = \sum_{i=1}^k \gamma_i f_i(x)$, where $f_i(x)$ is a density, $\gamma_i > 0$, $i = 1, \dots, k$, $\sum_{i=1}^k \gamma_i = 1$, $x \in \mathbb{R}^d$. Then according to (9) and (10), for any $x \in \mathbb{R}^d$, $r > 0$ and $R > 0$, one has $I_f(x, r) = \sum_{i=1}^k \gamma_i I_{f_i}(x, r)$, $M_f(x, R) \leq \sum_{i=1}^k \gamma_i M_{f_i}(x, R)$, $m_f(x, R) \geq \sum_{i=1}^k \gamma_i m_{f_i}(x, R)$. We will apply these relations for $f = p$ and $f = q$. It is well-known that, for any $\varepsilon > 0$, $c_i \geq 0$, $i = 1, \dots, k$, $k \in \mathbb{N}$, the following inequality is valid $(\sum_{i=1}^k c_i)^\varepsilon \leq \max\{1, k^{\varepsilon-1}\} \sum_{i=1}^k c_i^\varepsilon$. Moreover, this inequality is obviously satisfied for all $\varepsilon \in \mathbb{R}$ as for $\varepsilon \leq 0$ it holds $(\sum_{i=1}^k c_i)^\varepsilon \leq \sum_{i=1}^k c_i^\varepsilon$. Therefore

$$Q_{p,q}(\varepsilon, R) \leq \max\{1, J^{\varepsilon-1}\} \sum_{i=1}^I \sum_{j=1}^J a_i b_j^\varepsilon Q_{p_i, q_j}(\varepsilon, R) < \infty,$$

$$T_{p,q}(\varepsilon, R) \leq \sum_{i=1}^I \sum_{j=1}^J a_i b_j^{-\varepsilon} T_{p_i, q_j}(\varepsilon, R) < \infty.$$

The same reasoning leads to bounds $Q_{p,p}(\varepsilon, R) < \infty$ and $T_{p,p}(\varepsilon, R) < \infty$. Now in view of (13), for $v > 0$, $t > 0$ and $N \in \mathbb{N}$, we can write $K_{p,q}(v, N, t) = \sum_{i=1}^I \sum_{j=1}^J a_i b_j K_{p_i, q_j}(v, N, t)$. In this manner we can also represent $K_{p,p}(v, N, t)$. $\square$

Lemma A1. Let probability measures $\mathbb{P}$, $\mathbb{Q}$ and a σ -finite measure μ (e.g., the Lebesgue measure) be defined on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Assume that $\mathbb{P}$ and $\mathbb{Q}$ have densities $p(x)$ and $q(x)$, $x \in \mathbb{R}^d$, w.r.t. the measure μ . Then the following statements are true.

- (1) $\mathbb{P} \ll \mathbb{Q}$ if and only if $\mathbb{P}(S(p) \setminus S(q)) = 0$;
- (2) formula (2) holds.

Proof of Lemma A1. (1) Let $\mathbb{P} \ll \mathbb{Q}$. Obviously $\mathbb{Q}(\mathbb{R}^d \setminus S(q)) = 0$. Therefore $\mathbb{P}(\mathbb{R}^d \setminus S(q)) = 0$. Since $S(p) \setminus S(q) \subseteq \mathbb{R}^d \setminus S(q)$, one has $\mathbb{P}(S(p) \setminus S(q)) = 0$.

Now let $\mathbb{P}(S(p) \setminus S(q)) = 0$. Assume that $\mathbb{P}$ is not absolutely continuous w.r.t. $\mathbb{Q}$. Then there exists a set A such that $\mathbb{Q}(A) = 0$ and $\mathbb{P}(A) > 0$. Consequently $\mu(A) > 0$ as $\mathbb{P} \ll \mu$. We can write $A = A_1 \cup A_2$, where $A_1 := A \cap (\mathbb{R}^d \setminus S(q))$, $A_2 := A \cap S(q)$. We get $\mathbb{Q}(A) = \mathbb{Q}(A_1) + \mathbb{Q}(A_2)$ as $A_1 \cap A_2 = \emptyset$. Note that $\mathbb{Q}(A_1) = 0$ since $q \equiv 0$ on A_1 , so $\mathbb{Q}(A_2) = 0$. Relation $\mathbb{Q}(A_2) = \int_{A_2} q(x) \mu(dx)$ yields $\mu(A_2) = 0$ ($q > 0$ on A_2 and μ is a σ -finite measure). One has $\mathbb{P}(A_2) = 0$ because $\mathbb{P} \ll \mu$. Thus, $\mathbb{P}(A) = \mathbb{P}(A_1) + \mathbb{P}(A_2) = \mathbb{P}(A_1) > 0$. Clearly, $A_1 \subset \mathbb{R}^d \setminus S(q)$. Hence $\mathbb{P}(S(p) \setminus S(q)) = \mathbb{P}(S(p) \cap (\mathbb{R}^d \setminus S(q))) \geq \mathbb{P}(S(p) \cap A_1) = \mathbb{P}(A_1) > 0$. We come to the contradiction. Therefore $\mathbb{P} \ll \mathbb{Q}$.

In such a way we have proved that if $\mathbb{P} \ll \mu$ and $\mathbb{Q} \ll \mu$, the relation $\mathbb{P} \ll \mathbb{Q}$ holds if and only if $\mathbb{P}(S(p) \setminus S(q)) = 0$. Obviously we can take as p and q any versions of $\frac{d\mathbb{P}}{d\mu}$ and $\frac{d\mathbb{Q}}{d\mu}$.

(2) Suppose that $\mathbb{P} \ll \mathbb{Q}$. We know that $\mathbb{P}, \mathbb{Q}$ are probability measures, $\mathbb{Q} \ll \mu$ where μ is a σ -finite measure. Then, in view of [33], statement (b) of Lemma on p. 273, the following equality $\frac{d\mathbb{P}}{d\mathbb{Q}} = \frac{d\mathbb{P}}{d\mu} / \frac{d\mathbb{Q}}{d\mu}$ holds $\mathbb{Q}$ -a.s. and consequently $\mathbb{P}$ -a.s. too (on the set $B := \{x : \frac{d\mathbb{Q}}{d\mu} = 0\}$ having $\mathbb{Q}(B) = 0$ a density $\frac{d\mathbb{P}}{d\mathbb{Q}}$ can be taken equal to zero). So, $\frac{d\mathbb{P}}{d\mathbb{Q}}(x) = \frac{p(x)}{q(x)}$ for $\mathbb{P}$ -almost all $x \in \mathbb{R}^d$. One has

$$\int_{\mathbb{R}^d} p(x) \log\left(\frac{p(x)}{q(x)}\right) \mu(dx) = \int_{\mathbb{R}^d} \log\left(\frac{p(x)}{q(x)}\right) d\mathbb{P} = \int_{\mathbb{R}^d} \log\left(\frac{d\mathbb{P}}{d\mathbb{Q}}\right) d\mathbb{P},$$

where all integrals converge or diverge simultaneously. Indeed, if h is a measurable function with values in $[-\infty, \infty]$ then $\int_A h(x) \nu(dx) := 0$, whenever $\nu(A) = 0$ (ν is a finite or a σ -finite measure). We also employed [33], statement (a) of Lemma on p.273, when we changed the integration by μ to integration by $\mathbb{P}$.

Now assume that $\mathbb{P}$ is not absolutely continuous w.r.t. $\mathbb{Q}$, i.e., $\mathbb{P}(S(p) \setminus S(q)) > 0$ in view of part (1) of the present Lemma. As usual, for any measurable $B \subset \mathbb{R}^d$, $\int_B 0 \mu(dx) = 0$. Then

$$\int_{\mathbb{R}^d} p(x) \log\left(\frac{p(x)}{q(x)}\right) \mu(dx) = \int_{S(p) \setminus S(q)} p(x) \log\left(\frac{p(x)}{q(x)}\right) \mu(dx) + \int_{S(p) \cap S(q)} p(x) \log\left(\frac{p(x)}{q(x)}\right) \mu(dx).$$

Evidently

$$\int_{S(p) \setminus S(q)} p(x) \log\left(\frac{p(x)}{q(x)}\right) \mu(dx) = \int_{S(p) \setminus S(q)} \log\left(\frac{p(x)}{q(x)}\right) \mathbb{P}(dx) = \infty \cdot \mathbb{P}(S(p) \setminus S(q)) = \infty$$

as $\mathbb{P}(S(p) \setminus S(q)) > 0$. Since $-\log t \geq 1 - t$ if $t > 0$, we write, for all $x \in S(p) \cap S(q)$, $\log \frac{p(x)}{q(x)} = -\log \frac{q(x)}{p(x)} \geq 1 - \frac{q(x)}{p(x)}$. Thus, $\int_{S(p) \cap S(q)} p(x) \log\left(\frac{p(x)}{q(x)}\right) \mu(dx) \geq \int_{S(p) \cap S(q)} p(x) \left(1 - \frac{q(x)}{p(x)}\right) \mu(dx) = \int_{S(p) \cap S(q)} p(x) \mu(dx) - \int_{S(p) \cap S(q)} q(x) \mu(dx) = \mathbb{P}(S(p) \cap S(q)) - \mathbb{Q}(S(p) \cap S(q)) \geq 0 - 1 = -1$. Consequently $\int_{\mathbb{R}^d} p(x) \log\left(\frac{p(x)}{q(x)}\right) \mu(dx) = \infty$. The proof is complete. $\square$

Remark A1. Note that formula (2) can give an infinite value of $D(\mathbb{P}||\mathbb{Q})$ also when $\mathbb{P} \ll \mathbb{Q}$. It is enough to take $p(x) = \frac{1}{\pi(1+x^2)}$ and $q(x) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{x^2}{2}\}$, $x \in \mathbb{R}$.

References

1. Kullback, R.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [\[CrossRef\]](#)
2. Moulin, P.; Veeravalli, V.V. *Statistical Inference for Engineers and Data Scientists*; Cambridge University Press: Cambridge, UK, 2019.
3. Pardo, L. New developments in statistical information theory based on entropy and divergence measures. *Entropy* **2019**, *21*, 391. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Ji, S.; Zhang, Z.; Ying, S.; Wang, L.; Zhao, X.; Gao, Y. Kullback–Leibler divergence metric learning. *IEEE Trans. Cybern.* **2020**, 1–12. [\[CrossRef\]](#)
5. Noh, Y.K.; Sugiyama, M.; Liu, S.; du Plessis, M.C.; Park, F.C.; Lee, D.D. Bias reduction and metric learning for nearest-neighbor estimation of Kullback–Leibler divergence. *Neural Comput.* **2018**, *30*, 1930–1960. [\[CrossRef\]](#)
6. Claiici, S.; Yurochkin, M.; Ghosh, S.; Solomon, J. Model Fusion with Kullback–Leibler Divergence. In Proceedings of the 37th International Conference on Machine Learning, Online, 12–18 July 2020; Daumé, H., III, Singh, A., Eds.; PMLR: Brookline, MA, USA, 2020; Volume 119, pp. 2038–2047.

7. Póczos, B.; Xiong, L.; Schneider, J. Nonparametric Divergence Estimation with Applications to Machine Learning on Distributions. In Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence, Barcelona, Spain, 14–17 July 2011; AUAI Press: Arlington, VA, USA, 2011; pp. 599–608.
8. Cui, S.; Luo, C. Feature-based non-parametric estimation of Kullback–Leibler divergence for SAR image change detection. *Remote Sens. Lett.* **2016**, *11*, 1102–1111. [\[CrossRef\]](#)
9. Deledalle, C.-A. Estimation of Kullback–Leibler losses for noisy recovery problems within the exponential family. *Electron. J. Stat.* **2017**, *11*, 3141–3164. [\[CrossRef\]](#)
10. Yu, X.-P.; Chen, S.-X.; Peng, M.-L. Application of partial least squares algorithm based on Kullback–Leibler divergence in intrusion detection. In Proceedings of the International Conference on Computer Science and Technology (CST2016), Shenzhen, China, 8–10 January 2016; Cai, N., Ed.; World Scientific: Singapore, 2017; pp. 256–263.
11. Li, J.; Cheng, K.; Wang, S.; Morstatter, F.; Trevino, R.P.; Tang, J.; Liu, H. Feature Selection: A Data Perspective. *ACM Comput. Surv.* **2017**, *50*, 1–45. [\[CrossRef\]](#)
12. Peng, H.; Long, F.; Ding, C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226–1238. [\[CrossRef\]](#)
13. Vergara, J.R.; Estévez, P.A. A review of feature selection methods based on mutual information. *Neural Comput. Appl.* **2014**, *24*, 175–186. [\[CrossRef\]](#)
14. Granero-Belinchón, C.; Roux, S.G.; Garnier, N.B. Kullback–Leibler divergence measure of intermittency: Application to turbulence. *Phys. Rev. E* **2018**, *97*, 013107, 1–10. [\[CrossRef\]](#)
15. Charzyńska, A.; Gambin, A. Improvement of the k -NN entropy estimator with applications in systems biology. *Entropy* **2016**, *18*, 13. [\[CrossRef\]](#)
16. Wang, M.; Jiang, J.; Yan, Z.; Alberts, I.; Ge, J.; Zhang, H.; Zuo, C.; Yu, J.; Rominger, A.; Shi, K.; Alzheimer’s Disease Neuroimaging Initiative. Individual brain metabolic connectome indicator based on Kullback–Leibler Divergence Similarity Estimation predicts progression from mild cognitive impairment to Alzheimer’s dementia. *Eur. J. Nucl. Med. Mol. Imaging* **2020**, *47*, 2753–2764. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Zhong, J.; Liu, R.; Chen, P. Identifying critical state of complex diseases by single-sample Kullback–Leibler divergence. *BMC Genom.* **2020**, *21*, 87. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Li, J.; Shang, P. Time irreversibility of financial time series based on higher moments and multiscale Kullback–Leibler divergence. *Phys. A Stat. Mech. Appl.* **2018**, *502*, 248–255. [\[CrossRef\]](#)
19. Beraha, M.; Betelli, A.M.; Papini, M.; Tirinzoni, A.; Restelli, M. Feature selection via mutual information: New theoretical insights. *arXiv* **2019**, arXiv:1907.07384v1.
20. Carrara, N.; Ernst, J. On the estimation of mutual information. *Proceedings* **2019**, *33*, 31. [\[CrossRef\]](#)
21. Lord, W.M.; Sun, J.; Bollt, E.M. Geometric k -nearest neighbor estimation of entropy and mutual information. *Chaos Interdiscip. J. Nonlinear Sci.* **2018**, *28*, 033114. [\[CrossRef\]](#)
22. Moon, K.R.; Sricharan, K.; Hero, A.O., III. Ensemble estimation of generalized mutual information with applications to Genomics. *arXiv* **2019**, arXiv:1701.08083v2.
23. Suzuki, J. *Estimation of Mutual Information*; Springer: Singapore, 2021.
24. Sason, I.; Verdú, S. F-divergence inequalities. *IEEE Trans. Inf. Theory* **2016**, *62*, 5973–6006. [\[CrossRef\]](#)
25. Moon, K.R.; Sricharan, K.; Greenewald, K.; Hero, A.O., III. Ensemble estimation of information divergence. *Entropy* **2018**, *20*, 560. [\[CrossRef\]](#)
26. Rubenstein, P.K.; Bousquet, O.; Djolonga, J.; Riquelme, C.; Tolstikhin, I. Practical and Consistent Estimation of f -Divergences. In Proceedings of the NeurIPS 2019, 33rd Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; Advances in Neural Information Processing Systems; Wallach, H., Larochelle, H., Beygelzimer, A., d’Alché-Buc, F., Fox, E., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2019; Volume 32, pp. 4070–4080.
27. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423; 623–656. [\[CrossRef\]](#)
28. Kozachenko, L.F.; Leonenko, N.N. Sample estimate of the entropy of a random vector. *Probl. Inf. Transm.* **1987**, *23*, 9–16.
29. Kraskov, A.; Stögbauer, H.; Grassberger, P. Estimating mutual information. *Phys. Rev. E* **2004**, *69*, 066138. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Leonenko, N.N.; Pronzato, L.; Savani, V. A class of Rényi information estimations for multidimensional densities. *Ann. Stat.* **2008**, *36*, 2153–2182. Correction: *Ann. Stat.* **2010**, *38*, 3837–3838. [\[CrossRef\]](#)
31. Wang, Q.; Kulkarni, S.R.; Verdú, S. Divergence estimation for multidimensional densities via k -nearest-neighbor distances. *IEEE Trans. Inf. Theory* **2009**, *55*, 2392–2405. [\[CrossRef\]](#)
32. Pál, D.; Póczos, B.; Szepesvári, C. Estimation of Rényi Entropy and Mutual Information Based on Generalized Nearest-Neighbor Graphs. In Proceedings of the NIPS 2010 Proceedings of the 23rd International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 6–9 December 2010; Advances in Neural Information Processing Systems; Lafferty, J., Williams, C., Shawe-Taylor, J., Zemel, R., Culotta, A., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2010; Volume 23, pp. 1849–1857.
33. Shiryaev, A.N. *Probability—1*, 3rd ed.; Springer: New York, NY, USA, 2016.
34. Loève, M. *Probability Theory*, 4th ed.; Springer: New York, NY, USA, 1977.
35. Bulinski, A.; Dimitrov, D. Statistical estimation of the Shannon entropy. *Acta Math. Sin. Ser.* **2019**, *35*, 17–46. [\[CrossRef\]](#)
36. Biau, G.; Devroye, L. *Lectures on the Nearest Neighbor Method*; Springer: Cham, Switzerland, 2015.
37. Bulinski, A.; Kozhevin, A. Statistical estimation of conditional Shannon entropy. *ESAIM Probab. Stat.* **2019**, *23*, 350–386. [\[CrossRef\]](#)

38. Coelho, F.; Braga, A.P.; Verleysen, M. A mutual information estimator for continuous and discrete variables applied to feature selection and classification problems. *Int. J. Comput. Intell. Syst.* **2016**, *9*, 726–733. [[CrossRef](#)]
39. Delattre, S.; Fournier, N. On the Kozachenko-Leonenko entropy estimator. *J. Stat. Plan. Inference* **2017**, *185*, 69–93. [[CrossRef](#)]
40. Berrett, T.B.; Samworth R.J. Efficient two-sample functional estimation and the super-oracle phenomenon. *arXiv* **2019**, arXiv:1904.09347.
41. Penrose, M.D.; Yukich, J.E. Limit theory for point processes in manifolds. *Ann. Appl. Probab.* **2013**, *6*, 2160–2211. [[CrossRef](#)]
42. Tsybakov, A.B.; Van der Meulen, E.C. Root- n consistent estimators of entropy for densities with unbounded support. *Scand. J. Stat.* **1996**, *23*, 75–83.
43. Singh, S.; Pószoc, B. Analysis of k -nearest neighbor distances with application to entropy estimation. *arXiv* **2016**, arXiv:1603.08578v2.
44. Ryu, J.J.; Ganguly, S.; Kim, Y-H.; Noh, Y-K.; Lee, D.D. Nearest neighbor density functional estimation from inverse Laplace transform. *arXiv* **2020**, arXiv:1805.08342v3.
45. Gao, S.; Steeg, G.V.; Galstyan A. Efficient Estimation of Mutual Information for Strongly Dependent Variables. In Proceedings of the 18th International Conference on Artificial Intelligence and Statistics, San Diego, CA, USA, 9–12 May 2015; Lebanon, G., Vishwanathan, S. V. N., Eds.; PMLR: Brookline, MA, USA, 2015; Volume 38, pp. 277–286.
46. Berrett, T.B.; Samworth R.J.; Yuan M. Efficient multivariate entropy estimation via k -nearest neighbour distances. *Ann. Stat.* **2019**, *47*, 288–318. [[CrossRef](#)]
47. Gorla, M.N.; Leonenko, N.N.; Mergel, V.V.; Novi Inverardi, P.L. A new class of random vector entropy estimators and its applications in testing statistical hypotheses. *J. Nonparametr. Stat.* **2005**, *17*, 277–297. [[CrossRef](#)]
48. Evans, D. A computationally efficient estimator for mutual information. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2008**, *464*, 1203–1215. [[CrossRef](#)]
49. Yeh, J. *Real Analysis: Theory of Measure and Integration*, 3rd ed.; World Scientific: Singapore, 2014.
50. Evans, D.; Jones, A.J.; Schmidt, W.M. Asymptotic moments of near-neighbour distance distributions. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2002**, *458*, 2839–2849. [[CrossRef](#)]
51. Bouguila, N.; Wentao F. (Eds.) *Mixture Models and Applications*; Springer: Cham, Switzerland, 2020.
52. Borkar, V.S. *Probability Theory. An Advanced Course*; Springer: New York, NY, USA, 1995.
53. Kallenberg, O. *Foundations of Modern Probability*; Springer: New York, NY, USA, 1997.
54. Billingsley, P. *Convergence of Probability Measures*, 2nd ed.; Wiley: New York, NY, USA, 1999.
55. Alonso Ruiz, P.; Spodarev, E. Entropy-based inhomogeneity detection in fiber materials. *Methodol. Comput. Appl. Probab.* **2018**, *20*, 1223–1239. [[CrossRef](#)]
56. Dresvyanskiy, D.; Karaseva, T.; Makogin, V.; Mitrofanov, S.; Redenbach, C.; Spodarev, E. Detecting anomalies in fibre systems using 3-dimensional image data. *Stat. Comput.* **2020**, *30*, 817–837. [[CrossRef](#)]
57. Glaz, J.; Naus, J.; Wallenstein, S. *Scan Statistics*; Springer: New York, NY, USA, 2009.
58. Walther, G. Optimal and fast detection of spatial clusters with scan statistics. *Ann. Stat.* **2010**, *38*, 1010–1033. [[CrossRef](#)]
59. Gnedenko, B.V.; Korolev, V.Yu. *Random Summation: Limit Theorems and Applications*; CRC Press: Boca Raton, FL, USA, 1996.