The Annals of Statistics
2000, Vol. 28, No. 2, 591–629

EMPIRICAL GEOMETRY OF MULTIVARIATE DATA: A DECONVOLUTION APPROACH[1]

BY V. I. KOLTCHINSKII

University of New Mexico

Let $\{Y_j:\ j = 1,\dots,n\}$ be independent observations in $\mathbb{R}^m,\ m \geq 1$ with common distribution Q. Suppose that $Y_j = X_j + \xi_j,\ j = 1,\dots,n$, where $\{X_j,\ \xi_j,\ j = 1,\dots,n\}$ are independent, $X_j,\ j = 1,\dots,n$ have common distribution P and $\xi_j,\ j = 1,\dots,n$ have common distribution μ, so that $Q = P * \mu$. The problem is to recover hidden geometric structure of the support of P based on the independent observations Y_j. Assuming that the distribution of the errors μ is known, deconvolution statistical estimates of the fractal dimension and the hierarchical cluster tree of the support that converge with exponential rates are suggested. Moreover, the exponential rates of convergence hold for adaptive versions of the estimates even in the case of normal noise ξ_j with unknown covariance. In the case of the dimension estimation, though, the exponential rate holds only when the set of all possible values of the dimension is finite (e.g., when the dimension is known to be integer). If this set is infinite, the optimal convergence rate of the estimator becomes very slow (typically, logarithmic), even when there is no noise in the observations.

1. Introduction. The goal of this paper is to suggest a new approach to the problems of statistical recovery of geometric properties of the support of multivariate data. The support of a Borel probability measure P in $\mathbb{R}^m$ is defined as

$$\mathrm{supp}(P) := \bigcap \big\{F:\ P(F) = 1,\ F \subset \mathbb{R}^m,\ F \text{ is closed}\big\}.$$

It is easy to see that $P(\mathrm{supp}(P)) = 1$. In what follows, we assume that $\mathrm{supp}(P)$ is a compact set. The problems of nonparametric estimation of the set $\mathrm{supp}(P)$, based on an i.i.d. sample from P (i.e., in the case of *direct* observations), have been thoroughly studied by Korostelev and Tsybakov (1993), Korostelev, Simar and Tsybakov (1995), Mammen and Tsybakov (1995), Polonik (1995), Cuevas and Fraiman (1997). In this paper, we are interested in estimation only of some geometric parameters of the support (dimension, number of clusters, etc.), but in the case of *indirect* observations.

Received June 1998; revised December 1999.
[1]Supported in part by NSA Grant MDA904-99-1-0031, by Sandia National Laboratories grant AU-4005, and by a Humboldt Research Fellowship.
AMS 1991 *subject classifications*. Primary 62H30, 62H12, 62G07; secondary 62F17.
Key words and phrases. support of probability distribution, metric entropy, entropy dimension, clusters, deconvolving estimators.

The first problem will be to estimate *the entropy dimension* of the support. The entropy dimension of a set $E \subset \mathbb{R}^m$ is defined as

$$\dim(E) := \lim_{\varepsilon\to 0} H(E;\varepsilon)/\log(\varepsilon^{-1}), \tag{1.1}$$

where $H(E;\varepsilon)$ denotes the ε-entropy of the set E with respect to an arbitrary norm in $\mathbb{R}^m$. The problem of dimension estimation often occurs in applications when one deals with high-dimensional data with functionally dependent components, contaminated by random noise, the number of functionally independent components being relatively small. The case of nonlinear relationships have been discussed (often at a heuristic level) in pattern recognition literature, where the problem is often called the estimation of intrinsic dimension [see Fukunaga (1990), Jain and Dubes (1988), Goldfarb (1985), Wyse, Dubes and Jain (1980) among others]. We explore the case when the unknown dimension is not necessarily integer, which could have applications to the problem of estimation of dimension of fractals [see Falconer (1997), Edgar (1997)].

The second problem will be to estimate the number of clusters (connected components) and the hierarchical cluster tree of the support, which is a theoretical version of empirical hierarchical trees (describing families of hierarchically nested partitions of multivariate data) that are frequently used in cluster analysis and taxonomy [Hartigan (1975), Jambu (1978), Grenander (1981), Gordon (1996), Bock (1996); see Section 3 for precise definitions]. The related problem of estimation of the order of mixture models has been studied recently, for example, in a paper of Dacunha-Castelle and Gassiat (1997). See also Pollard (1982) for an application of empirical processes in cluster analysis.

We suggest in Section 2 a simple convolution model, which provides a probabilistic framework for these two problems as well as for some other statistical problems of support geometry. We refer to this circle of problems (in the context of the convolution model) as to *empirical geometry of multivariate data.* By this we mean in this paper the whole circle of problems related to statistical estimation of geometric features of the support of a probability distribution (especially in the context of the convolution model). Section 3 deals with the problems of dimension estimation, Section 4 is about the estimation of the cluster structure. In the next two sections (5 and 6) we discuss the problem of adaptive estimation of hidden geometric features in the case when the distribution of the noise is not completely known. The proofs are given in Section 7.

2. Convolution model and deconvolution estimates: preliminaries.

Let $Y_j,\ j = 1, 2, \ldots, n$ be i.i.d. random vectors in $\mathbb{R}^m$ with $m \geq 1$. Let Q denote the common distribution of Y_j. Suppose that

$$Y_j = X_j + \xi_j, \qquad j = 1, 2, \ldots, n, \tag{2.1}$$

where $X_j,\ j = 1, 2, \ldots, n$ are i.i.d. random vectors in $\mathbb{R}^m$ with common distribution P; $\xi_j,\ j = 1, 2, \ldots, n$ are i.i.d. random vectors in $\mathbb{R}^m$ with common distribution μ; X_j: $1 \leq j \leq n$ and ξ_j: $1 \leq j \leq n$ are independent (so that

$Q = P * \mu$, where $*$ denotes the convolution). It is supposed that P is the distribution of interest and its support has certain structure (for instance, it is a smooth manifold in $\mathbb{R}^m$ of unknown dimension, or it consists of an unknown finite number of mutually disjoint clusters, etc.) Since the sample comes from the population with the distribution $Q = P * \mu$, the geometric structure of P is "hidden" in the observations $(Y_1, \dots, Y_n)$. Assuming that μ is known, our goal is to recover hidden geometric characteristics of the support of P.

We outline a deconvolution method of estimation of $P(A)$ for a given subset $A \subset \mathbb{R}^m$, which is used later in both problems. The methods of nonparametric deconvolution have been developed intensively in the recent years. We refer to such authors as Carroll and Hall (1988), Stefanski and Carroll (1990), Zhang (1990), Fan (1991, 1992), Efromovich (1997).

We assume in what follows, for simplicity, that the distribution μ is symmetric; that is, $\mu(A) = \mu(-A)$ for all Borel sets $A \subset \mathbb{R}^m$. Given a signed Borel measure ν on $\mathbb{R}^m$, $\tilde{\nu}$ will denote its Fourier transform: $\tilde{\nu}(t) = \int_{\mathbb{R}^m} \exp\{i(t, x)\}\nu(dx)$, $t \in \mathbb{R}^m$. Suppose there exist a symmetric Borel probability measure Ψ and a signed measure of bounded total variation $\mathscr{K} := \mathscr{K}_\Psi := \mathscr{K}_{\Psi,\mu}$ on $\mathbb{R}^m$ such that

$$\mathscr{K} * \mu = \Psi. \tag{2.2}$$

Define

$$\hat{P}_n(A) := \hat{P}_{n,\Psi}(A) := \hat{P}_{n,\Psi,\mu}(A) := n^{-1} \sum_{j=1}^{n} \mathscr{K}(A - Y_j), \qquad A \subset \mathbb{R}^m, A \text{ is Borel.} \tag{2.3}$$

It easily follows from (2.2) that for all Borel sets $A \subset \mathbb{R}^m$,

$$\mathbb{E}\hat{P}_n(A) = (\mathscr{K} * Q)(A) = (\mathscr{K} * \mu * P)(A) = (\Psi * P)(A) =: P_\Psi(A).$$

In what follows, we call $\hat{P}_n$ *a deconvolving empirical measure.* If $\mu := \delta_0$ is the probability measure concentrated at the point $0 \in \mathbb{R}^m$ (i.e., there are no errors, $\xi_j \equiv 0$), then one can take $\mathscr{K} = \Psi = \delta_0$ and $\hat{P}_n$ becomes the empirical measure P_n based on the sample $(Y_1, \dots, Y_n) = (X_1, \dots, X_n)$. In the presence of the errors, it is reasonable to choose a measure Ψ supported in a small neighborhood of the point 0, or, more generally, such that for some small $\varepsilon > 0$, $\delta > 0$,

$$\Psi\{x\colon\ |x| \ge \varepsilon\} \le \delta, \tag{2.4}$$

where $|\cdot|$ is a norm in $\mathbb{R}^m$. Then the measure P_Ψ is close to P, so that $\hat{P}_n$ can be used as an estimator of P.

For instance, assume that $\tilde{\mu}(t) \ne 0$ on $\mathbb{R}^m$, and consider a symmetric Borel probability measure Ψ on $\mathbb{R}^m$, such that the functions $\tilde{\Psi}/\tilde{\mu}$ and $x \mapsto K(x) := K_\Psi(x) := K_{\Psi,\mu}(x)$, where

$$K(x) = (2\pi)^{-m} \int_{\mathbb{R}^m} \cos\{(t, x)\}\tilde{\Psi}(t)/\tilde{\mu}(t)\, dt, \qquad x \in \mathbb{R}^m \tag{2.5}$$

both belong to the space $L_1(\mathbb{R}^m)$. Define a signed Borel measure $\mathscr{K} := \mathscr{K}_\Psi := \mathscr{K}_{\Psi,\mu}$ on $\mathbb{R}^m$ with density K: $\mathscr{K}(A) := \int_A K(x)dx$, $A \subset \mathbb{R}^m$, A is Borel. Then it's easy to check that the condition (2.2) holds. Moreover, even if $K \notin L_1(\mathbb{R}^m)$, but $\tilde{\Psi}/\tilde{\mu} \in L_1(\mathbb{R}^m)$, we can still get, by a simple computation using Fubini's theorem, that for all *bounded* sets $A \subset \mathbb{R}^m$, $\mathbb{E}\hat{P}_n(A) = (P * \Psi)(A)$, which will be enough for our purposes.

We assume that $\tilde{\Psi}/\tilde{\mu} \in L_1(\mathbb{R}^m)$ *throughout the paper.*

One can also consider quite similarly the case when $\operatorname{supp}(P), \operatorname{supp}(\mu) \subset \mathbb{T}^m$, where $\mathbb{T}^m$ is the m-dimensional torus [the quotient group $\mathbb{R}^m/(2\pi\mathbb{Z})^m$]. In this case deconvolving empirical measures are based on Fourier series instead of integrals [see, e.g., Efromovich (1997)].

The next very easy fact shows that the best possible convergence rate in such problems as the dimension or the number of connected components (clusters) estimation is exponential. Let $\mathscr{P}$ be a class of probability distributions in $\mathbb{R}^m$ with compact supports, and let τ be a function from $\mathscr{P}$ into the set of all nonnegative integers $\mathbb{Z}_+$. For instance, $\tau(P)$ could be the dimension or the number of clusters of the set $\operatorname{supp}(P)$. Given two absolutely continuous probability distributions on $\mathbb{R}^m$ ν_1 and ν_2, we denote $I[\nu_1|\nu_2]$ the Kullback–Leibler information deviation between ν_1 and ν_2: $I[\nu_1|\nu_2] := \int_{\mathbb{R}^m} \log(d\nu_1/d\nu_2)\, d\nu_1$. Given a probability measure ν on $\mathbb{R}^m$ and a vector $\theta \in \mathbb{R}^m$, let ν_θ be the θ-shift of ν: $\nu_\theta(A) := \nu(A - \theta)$.

PROPOSITION 2.1. *Suppose that* $\operatorname{card}(\tau(\mathscr{P})) \geq 2$. *Suppose also that* μ *is absolutely continuous with a uniformly bounded density such that* $\int (d\mu/dx) \cdot \log(d\mu/dx)\, dx > -\infty$, $\tilde{\mu}(t) \neq 0$ *for all* $t \in \mathbb{R}^m$, *and* $\sup_{\theta \in K} I[\mu_\theta|\mu] < +\infty$ *for all bounded sets* $K \subset \mathbb{R}^m$. *Then there exists* $q \in (0, 1)$ *such that for all large enough* n,

$$\inf_{\check{\tau}_n} \sup_{P \in \mathscr{P}} \mathbb{P}\{\check{\tau}_n \neq \tau(P)\} \geq q^n,$$

where the inf *is taken over all the estimators* $\check{\tau}_n$ *of* $\tau(P)$, *based on the sample* $(Y_1, \ldots, Y_n)$ *with distribution* $Q = P * \mu$.

Thus, a reasonable goal is to construct the estimates of the dimension, the number of clusters and the hierarchical cluster tree of the support, which converge with exponential rates to the true parameters or structures.

Let $\mathscr{N} := \mathscr{N}(\varepsilon)$ be a finite set of balls (with respect to the norm $|\cdot|$) of radius ε, such that $\operatorname{supp}(P) \subset \bigcup_{B \in \mathscr{N}(\varepsilon)} B$. Note that $\mathscr{N}$ does not have to be a minimal covering of the set $\operatorname{supp}(P)$; it could cover a much larger subset that is known to contain the support. Our estimates of various geometric characteristics of the support of P are based on the geometric properties of such sets as $\bigcup\{B\colon B \in \mathscr{N}(\varepsilon),\ \hat{P}_n(B) \geq \gamma\}$ (with some parameters $\varepsilon, \gamma > 0$), consisting of the balls with large enough deconvolution "mass." We will slightly expand such sets (say, by adding the neighboring balls, or by increasing the radius of the balls) and define the estimates of the dimension, the number of clusters, etc. in terms of various covering numbers and numerical characteristics of the cluster structure of the expanded sets.

3. Estimation of the dimension of the support. Given a nonempty bounded set $E \subset \mathbb{R}^m$, denote $N(E;\varepsilon)$ the minimal number of balls (with respect to the norm $|\cdot|$) of radius $\varepsilon > 0$, covering E. The number $H(E;\varepsilon) := \log N(E;\varepsilon)$ is called the ε-entropy of the set E and the function $\varepsilon \mapsto H(E;\varepsilon)$ is the metric entropy of E. The *entropy dimension* of the set E is defined by (1.1), provided that the limit there exists [see Kolmogorov and Tikhomirov (1959)]. In fractal geometry this number is also called box (or box-counting) dimension of E [see Edgar (1997), Falconer (1997)]. It does not depend on the choice of the norm in $\mathbb{R}^m$. We give here a short list of well-known facts about this dimension [see Falconer (1997) for more details]:

1. If E is a finite set, then $\dim(E) = 0$.
2. If E is a nonempty open set in $\mathbb{R}^m$, then $\dim(E) = m$.
3. If E is a d-dimensional compact manifold in $\mathbb{R}^m$, then $\dim(E) = d$ (here d is a nonnegative integer).
4. The dimension $\dim(E)$ does not have to be integer. It could take noninteger values, for instance, for fractals in $\mathbb{R}^m$. For example, the dimension of the Cantor set in $\mathbb{R}^1$ is $\log 2/\log 3$; the dimension of the Sierpinski gasket in $\mathbb{R}^2$ is $\log 3/\log 2$; the dimension of the Koch curve in $\mathbb{R}^2$ is $\log 4/\log 3$, etc.
5. If $E_1 \subset E_2$, then $\dim(E_1) \le \dim(E_2)$.
6. If $f : E \mapsto \mathbb{R}^m$ is Lipschitz, then $\dim(f(E)) \le \dim(E)$.
7. If f is bi-Lipschitz (i.e., it is one-to-one and both f and f^{-1} are Lipschitz), then $\dim(f(E)) = \dim(E)$. In particular, this applies to the so-called similarity transformations [see Falconer (1997)].
8. For any bounded open set V in $\mathbb{R}^m$ and for any $d \in [0, m]$, one can find a set $E \subset V$ of dimension d. For instance, one could take the attractor of a family of similarity transforms with properly choosen similarity ratios [see Falconer (1997), Theorem 2.7]. If ν is the d-dimensional Hausdorff measure, then, for such an E with some constants $c_1, c_2 > 0$, $c_1\varepsilon^d \le \nu(B \cap E) \le c_2\varepsilon^d$ for all the balls B of radius $\varepsilon > 0$ with the centers in E [Falconer (1997), page 40].

In what follows, we denote $\dim(P) := \dim(\operatorname{supp}(P))$. Given a ball B, denote B^+ (resp., B^-) the ball with the same center as B, having twice larger (resp., twice smaller) radius.

For any given $\varepsilon > 0$, we fix a minimal collection $\mathscr{N}(\varepsilon)$ of balls of radius ε, covering $B(0;1)$. Given a set $\mathscr{D} \subset [0, m]$ and constants $\Theta > 0$, $C > 0$, we define the set $\mathscr{P} := \mathscr{P}(\mathscr{D}, \Theta, C)$ of all Borel probability measures P on $\mathbb{R}^d$ such that:

1. $\operatorname{supp}(P) \subset B(0, 1)$.
2. $\dim(P) \in \mathscr{D}$.
3. For $d := \dim(P)$.

$$\operatorname{card}\big(\{B \in \mathscr{N}(\varepsilon):\ \operatorname{dist}(B^+,\ \operatorname{supp}(P)) \le \varepsilon\}\big) \le \Theta\varepsilon^{-d}. \tag{3.1}$$

4. For any $\varepsilon > 0$ and for all balls B of radius ε,

(3.2) $$P(B) \le C\varepsilon^d.$$

Given $\mathscr{D} \subset [0, m]$, for all large enough $\Theta > 0$, $C > 0$, $\mathscr{P}(\mathscr{D}, \Theta, C) \neq \emptyset$. Moreover, for all $d_1, d_2 \in \mathscr{D}$ one can find $P_1, P_2 \in \mathscr{P}(\mathscr{D}, \Theta, C)$ with $\dim(P_1) = d_1$, $\dim(P_2) = d_2$ and $\operatorname{supp}(P_1) \cap \operatorname{supp}(P_2) = \emptyset$.

We assume that the set $\mathscr{D}$ is closed and denote

$$D := \max(\mathscr{D}), \qquad \delta := \delta(\mathscr{D}) := \min\{|d_1 - d_2|: \ d_1, d_2 \in \mathscr{D}, \ d_1 \neq d_2\}.$$

Given a Borel probability measure Ψ on $\mathbb{R}^m$ and numbers $\varepsilon > 0$, $\gamma > 0$, define

$$\hat{\mathscr{N}}_n := \hat{\mathscr{N}}_n(\varepsilon; \gamma; \Psi) := \Big\{B \in \mathscr{N}(\varepsilon): \ \hat{P}_{n,\Psi}(B^+) \ge 2\gamma\Big\},$$

$$\hat{N}_n := \hat{N}_n(\varepsilon; \gamma; \Psi) := \operatorname{card}\big(\hat{\mathscr{N}}_n(\varepsilon; \gamma; \Psi)\big).$$

Now we are in a position to define the estimator $\hat{d}_n := \hat{d}_n(\varepsilon; \gamma; \Psi)$ of the unknown dimension $d := \dim(P)$ as the smallest number $a \in \mathscr{D}$, that minimizes the quantity $|\log \ \hat{N}_n / \log(\varepsilon^{-1}) - a|$.

First we assume that $\delta(\mathscr{D}) > 0$, and hence $\mathscr{D}$ is a finite set. For instance, if P is an absolutely continuous distribution on a d-dimensional compact manifold $\mathscr{M} \subset \mathbb{R}^m$, such that its density p satisfies the condition $0 < \inf_{x \in \mathscr{M}} \ p(x) \le \sup_{x \in \mathscr{M}} \ p(x) < +\infty$, then $\mathscr{D} = \{0, 1, \dots, m\}$ and

(3.3) $$\hat{d}_n := \left[\frac{\log \ \hat{N}_n}{\log(\varepsilon^{-1})} + 1/2\right].$$

Various models of fractals [e.g., attractors of families of similarity transforms, Falconer (1997)] could lead to many different choices of the set $\mathscr{D}$.

Define

(3.4) $$\varepsilon(\mathscr{D}; \Theta; C) := \Big(\Theta^{-1} \wedge (2C)^{-1}\Big)^{2/\delta(\mathscr{D})} \text{ and } \gamma(\mathscr{D}, \varepsilon; \Theta) := (\varepsilon^D/(12\Theta)) \wedge (1/2).$$

THEOREM 3.1. *Suppose that* $\varepsilon < \varepsilon(\mathscr{D}; \Theta; C)$, $\gamma < \gamma(\mathscr{D}; \varepsilon; \Theta)$ *and* Ψ *satisfies the condition*

(3.5) $$\Psi\{x: \ |x| \ge \varepsilon\} \le \gamma.$$

Then there exist $\Lambda > 0$ *and* $q \in (0, 1)$ *such that*

(3.6) $$\sup_{P \in \mathscr{P}(\mathscr{D}; \Theta; C)} \mathbb{P}\{\hat{d}_n(\varepsilon; \gamma; \Psi) \neq \dim(P)\} \le \Lambda q^n.$$

The next statement shows that for a given μ one can always choose universal sequences $\{\varepsilon_n\}$ and $\{\gamma_n\}$ of parameters and $\{\Psi_n\}$ of measures such that the convergence rate of the estimator $\hat{d}_n$ is arbitrarily close to the exponential one. To this end, let $\{q_n\}$ be a sequence of positive numbers such that $q_n \uparrow 1$ as $n \to \infty$ (arbitrarily slow). Choose a sequence $\{\sigma_n\}$ of positive real numbers such that $\sigma_n \to 0$ and $\sigma_n / \log(q_n^{-1}) \to \infty$ as $n \to \infty$. Denote

$$\rho(R) := \inf_{|t|_\infty \le R} |\tilde{\mu}(t)|, \qquad R > 0,$$

where $|t|_\infty := \max_{1\le j\le m} |t_j|$ and suppose that $\rho(R) > 0$ for all $R \ge 0$. Define ε_n as the solution of the equation

$$\frac{\varepsilon^{m(m+1)}\rho(\varepsilon^{-m-1}(\log\ \varepsilon^{-1})^2)}{(\log\ \varepsilon^{-1})^{2m+1}} = \sigma_n^{1/2}. \tag{3.7}$$

The solution exists for all large n and is unique, since the function in the left side of the equation tends to 0 monotonically as $\varepsilon \to 0$.

For $a > 0$ and an even number $l \ge 2$, define

$$\psi_{a,l}(u) := ac(l)\left|\frac{\sin\ au}{au}\right|^l,\ \ u \in \mathbb{R}^1,\ \ \text{where } c(l) = \left(\int_{\mathbb{R}^1}\left|\frac{\sin\ u}{u}\right|^l du\right)^{-1}.$$

Let $\Psi_{a,l}$ be the measure in $\mathbb{R}^1$ with density $\psi_{a,l}$. Define $\Psi_n := \Psi_{a_n/2,2} \times \cdots_m \times \Psi_{a_n/2,2}$ with $a_n := \varepsilon_n^{-m-1}(\log\ \varepsilon_n^{-1})^2$. Let $\gamma_n := \varepsilon_n^m/(\log\ \varepsilon_n^{-1})$.

THEOREM 3.2. *Suppose that* $\rho(R) > 0$ *for all* $R \ge 0$. *Then, for all* $\Theta > 0$, $C > 0$ *and* $\mathscr{D} \subset [0,m]$, *the estimate* $\hat{d}_n := \hat{d}_n(\varepsilon_n, \gamma_n, \Psi_n)$ *satisfies the following bound*:

$$\sup_{P\in\mathscr{P}(\mathscr{D},\Theta,C)} \mathbb{P}\{\hat{d}_n \ne \dim(P)\} = O(q_n^n) \quad \text{as } n \to \infty. \tag{3.8}$$

Next we show that, in the case when the set $\mathscr{D}$ of all possible dimensions of the support of P is infinite, the typical minimax rates of convergence of *any* dimension estimate could become very slow, for instance logarithmic, even when there are no errors, that is, $\xi_j \equiv 0$.

Consider any closed set $\mathscr{D} \subset [0,m]$ with $D := \sup(\mathscr{D})$. Given a nondecreasing nonnegative function φ on $[0,1]$ such that $\varphi(\delta) \le \delta,\ \delta \in [0,1]$ let us call the set $\mathscr{D}$ *φ-rich* if, for all $\delta \in [0,1]$, there exist $d_1, d_2 \in \mathscr{D}$ such that $\varphi(\delta) \le d_2 - d_1 \le \delta$. The set $\mathscr{D}$ will be called *rich* if it is φ-rich for $\varphi(\delta) = \rho\delta$ with some constant $\rho \in (0,1]$. For example, if $\mathscr{D}$ is dense in an interval, then it is rich; the sets $\{n^{-1}:\ n \ge 1\}$ and $\{2^{-n}:\ n \ge 1\}$ are also rich, but the set $\{2^{-2^n}:\ n \ge 1\}$ is not; this set is φ-rich with $\varphi(\delta) = 0.5\delta^2$. More generally, a set $\mathscr{D} := \{d_n:\ n \ge 1\}$, where $d_n := \sum_{k=n}^\infty b_k,\ \{b_n\}$ is a decreasing sequence of positive numbers, such that

$$b_n \to 0 \text{ as } n \to \infty, \qquad \sum_{n=1}^\infty b_n \le m \text{ and } b_n \ge \varphi(b_{n-1}) \text{ for all } n \ge 1$$

is φ-rich. In particular, if $\varphi := \sum \varphi(d_n) I_{[d_n, d_{n-1})}$ (i.e., φ is a step function) with $d_n - d_{n+1} \ge \varphi(d_{n-1})$, then $\mathscr{D}$ is φ-rich.

Let us call the set $\mathscr{D}$ *φ-poor* if for all $\delta \in [0,1]$ there exists a δ-separated set $\mathscr{D}' \subset \mathscr{D}$ (i.e., $|d_2 - d_1| \ge \delta$ for all $d_1, d_2 \in \mathscr{D}'$ such that $d_1 \ne d_2$) such that $\mathscr{D} \subset (\mathscr{D}')^{\varphi(\delta)}$. For example, any subset of $[0,m]$ is φ-poor with $\varphi(\delta) = \delta$ (in this case, take a maximal δ-separated subset of $\mathscr{D}$ as the set $\mathscr{D}'$). Consider a set $\mathscr{D} := \{d_n:\ \ n \ge 1\}$ and a step function $\varphi := \sum \varphi(d_n) I_{[d_n, d_{n-1})}$, where

$d_n := \sum_{k=n}^{\infty} b_k,\ \{b_n\}$ is a decreasing sequence of positive numbers such that

$$b_n \to 0 \text{ as } n \to \infty, \qquad \sum_{n=1}^{\infty} b_n \le m,\ d_{n+1} \le b_n \text{ and } d_{n+1} \le \varphi(d_n) \text{ for all } n \ge 1.$$

Then $\mathscr{D}$ is φ-poor.

In particular, it is easily seen that the set $\mathscr{D} := \{d_n : \ n \ge 1\}$, where d_1 is a small enough number and $d_{n+1} = d_n^{\alpha},\ n \ge 1,\ \alpha > 1$ is φ-rich with $\varphi := \frac{1}{2}\sum d_n^{\alpha} I_{[d_n, d_{n-1})}$ and it is φ-poor with $\varphi := \sum d_n^{\alpha} I_{[d_n, d_{n-1})}$.

Denote $\mathscr{P}(\mathscr{D})$ the set of all Borel probability measures supported in the ball $B(0;1)$ and such that $\dim(P) \in \mathscr{D}$. Let $\mathscr{N} = \mathscr{N}(\varepsilon)$ be a minimal class of balls of radius $\varepsilon > 0$, covering $B(0;1)$. Given $\epsilon > 0$, let $\mathscr{P}_{\epsilon}(\mathscr{D}, \Theta, C)$ be the class of all Borel probability measures on $\mathbb{R}^m$, satisfying, for all $\varepsilon > \epsilon$, the conditions 1–4 of the definition of the set $\mathscr{P}(\mathscr{D}, \Theta, C)$ above. Given $\alpha > 0$, define $\varepsilon_n := n^{-\alpha}$ and $\mathscr{P}_n(\mathscr{D}) := \mathscr{P}_{\varepsilon_n}(\mathscr{D}, \Theta, C)$.

THEOREM 3.3. (i) *Suppose the set $\mathscr{D}$ of all possible dimensions is φ-rich. For any distribution μ of the noise, for large enough $C > 0$, $\Theta > 0$, for any $\alpha, \beta > 0$ such that $\alpha\beta < \log\ 2$, for all $n \ge 6$ and for all estimates $\check{d}_n$ of* $\dim(P)$, *based on the observations* $(Y_1, \dots, Y_n)$,

$$\begin{aligned} &\sup_{P \in \mathscr{P}(\mathscr{D})} \mathbb{P}\left\{\left|\check{d}_n - \dim(P)\right| \ge \frac{1}{2}\varphi\left(\frac{\beta}{\log\ n}\right)\right\} \\ &\quad \ge \sup_{P \in \mathscr{P}_n(\mathscr{D})} \mathbb{P}\left\{\left|\check{d}_n - \dim(P)\right| \ge \frac{1}{2}\varphi\left(\frac{\beta}{\log\ n}\right)\right\} \ge 1/4. \end{aligned} \tag{3.9}$$

(ii) *Suppose that the set $\mathscr{D}$ of all possible dimensions is φ-poor and that with some $A, B > 0$,*

$$|\tilde{\mu}(t)| \ge B|t|^{-A} \quad \text{for all } |t| \ge 1. \tag{3.10}$$

Then there exist a constant $\alpha(D;m;A) > 0$ and an estimator $\hat{d}_n$, based on the observations $(Y_1, \dots, Y_n)$, such that for all $0 < \alpha < \alpha(D;m;A)$, for all $\beta > 3(|\log\ \Theta| \vee |\log\ C|)/\alpha$, and for all $\Delta > 0$,

$$\sup_{P \in \mathscr{P}_n(\mathscr{D})} \mathbb{P}\left\{\left|\hat{d}_n - \dim(P)\right| \ge \varphi\left(\frac{\beta}{\log\ n}\right)\right\} = O(n^{-\Delta}). \tag{3.11}$$

Specifically, we choose in part (ii) of Theorem 3.3,

$$\alpha(D;m;A) := \tfrac{1}{2}[A(D+1) + (m+1)D]^{-1} \tag{3.12}$$

and we use there the estimator $\hat{d}_n$, which minimizes the function $a \mapsto |(\log\ \hat{N}_n / \log(\varepsilon_n^{-1})) - a|$ on the δ-separated set $\mathscr{D}_n \subset \mathscr{D}$ with $\delta := \beta/\log\ n$ and such that $\mathscr{D} \subset (\mathscr{D}_n)^{\varphi(\delta)}$ (see the definition of φ-poor sets above). The estimate $\hat{d}_n$ is based on the measure $\Psi := \Psi_n := \Psi_{a_n, 2} \times \cdots \times \Psi_{a_n, 2}$ (see the definition before Theorem 3.2) with the sequence $a_n := n^{\sigma}$. It is supposed that $\sigma > \alpha(D+1)$ and $\sigma(m+A) - \alpha(m+D) < 1/2$. Such a choice of σ is possible if α satisfies the condition $\alpha < \alpha(D; m; A)$. We use $\varepsilon_n := n^{-\alpha}$ and $\gamma := \gamma_n := \varepsilon_n^D/(\log\ n)^{1/2}$ in the estimator $\hat{d}_n$.

Note also that, for a rich set $\mathscr{D} \subset [0, m]$ [which is, at the same time, φ-poor with $\varphi(\delta) \equiv \delta$, as any subset of $[0, m]$ is], Theorem 3.3 gives the convergence rate of the order $(\log\ n)^{-1}$, even in the case when there is no noise (i.e., $\tilde{\mu} \equiv 1$). Similarly, in the case of the set $\mathscr{D} := \{2^{-2^n}:\ n \geq 1\}$ (that is not rich) the convergence rate is of the order $(\log\ n)^{-2}$ both in the upper and in the lower bounds.

4. Estimation of the hierarchical cluster tree and the number of clusters. We start this section with precise definition of the hierarchical cluster tree of the support $\operatorname{supp}(P)$. Let S be a set and $\mathscr{V}$ a *finite* family of subsets of S such that $S \in \mathscr{V}$, and $A, B \in \mathscr{V}$ implies that either $A \subset B$ or $B \subset A$ or $A \cap B = \varnothing$. Consider $\mathscr{V}$ as the set of all vertices of a graph. Let us connect two vertices A and B with an edge iff $A \supset B$ and there is no $C \in \mathscr{V}$ such that $A \supset C \supset B$. Denote this graph by $\mathscr{G}$. Clearly, $\mathscr{G}$ is a connected graph which does not have cycles, so it is a tree with root S [see, e.g., Bolobás (1979)]. In cluster analysis, it is often convenient to assign to each vertex of the tree $\mathscr{G}$ a number, $\mathscr{V} \ni V \mapsto \chi(V) \in \mathbb{R}$, such that $V_1 \subset V_2$ implies $\chi(V_1) < \chi(V_2)$. Such a function will be called *a height* on the tree $\mathscr{G}$. Given two trees $\mathscr{G}_1$ and $\mathscr{G}_2$ with heights χ_1, χ_2 defined on them, we write $(\mathscr{G}_1, \chi_1) \cong (\mathscr{G}_2, \chi_2)$ if the trees are isomorphic (i.e., there exists a one-to-one mapping between their vertices such that any two adjacent vertices in one graph correspond to the adjacent vertices in another one) and the relationship $\chi_1(V_1) < \chi_1(U_1)$ between any two vertices V_1, U_1 in $\mathscr{G}_1$ is equivalent to the relationship $\chi_2(V_2) < \chi_2(U_2)$ between their images under the mapping V_2, U_2 in $\mathscr{G}_2$. This defines an isomorphism between the trees with heights. We write $(\mathscr{G}_1, \chi_1) \ncong (\mathscr{G}_2, \chi_2)$ if there is no such an isomorphism. Most often, the actual values of the height χ are irrelevant, but it's rather important to know the order of the values for different vertices of the tree [see Gordon (1996)], so the tree $\mathscr{G}$ with the height χ on it should be recovered up to isomorphism $\cong$. Often, when it is clear which heights are in mind, we write simply $\mathscr{G}_1 \cong \mathscr{G}_2$, or $\mathscr{G}_1 \ncong \mathscr{G}_2$.

We are interested in the estimation of the hierarchical cluster tree of the set $\operatorname{supp}(P)$. We proceed now to the definition of this tree. Any connected component of the support will be called a *cluster* of P. Under the assumption of compactness of the set $\operatorname{supp}(P)$, the clusters are disjoint compact sets; the distance between any two clusters is strictly positive. The number of different clusters will be denoted $\nu(P)$. We assume now that the number of clusters of $\operatorname{supp}(P)$ is finite: $\nu := \nu(P) < +\infty$. Each cluster is supposed to be a closure of an open set in $\mathbb{R}^m$. Let $\mathscr{C} := \{C_j,\ \ j = 1, \ldots, \nu\}$ be the set of all clusters.

The minimal distance between two subsets in $\mathbb{R}^m$ is defined as

$$\operatorname{dist}(A, B) := \inf\{|x - y|:\ x \in A,\ \ y \in B\};$$

the Hausdorff distance is

$$h(A, B) := \inf\{\delta > 0:\ A \subset B^{\delta} \text{ and } B \subset A^{\delta}\}.$$

Note that dist is not a (pseudo)metric, whereas h is.

Let ρ be a "dissimilarity" (ρ does not have to be a metric) of two compact sets in $\mathbb{R}^m$, such that $\rho(D, D) = 0$ and $\rho(D_1, D_2) = \rho(D_2, D_1)$ for all compact sets $D, D_1, D_2 \subset \mathbb{R}^m$. We also assume that ρ satisfies the following properties: $\rho(D_1, D_2) > 0$ for all compact disjoint sets $D_1, D_2 \subset \mathbb{R}^m$ and

$$\left|\rho(D_1, D_2) - \rho(D_1', D_2')\right| \leq h(D_1, D_1') + h(D_2, D_2').$$

For instance, ρ could be the Hausdorff metric h, or it could be the minimal "distance" dist between the sets.

Given a finite class of compact disjoint sets $\mathscr{K} := \{K_1, \dots, K_r\}$, denote

$$\delta(\mathscr{K}) := \delta_{\rho}(\mathscr{K}) := \min\{\rho(K_i, K_j):\ i \neq j\}.$$

Clearly, $\delta(\mathscr{K}) > 0$. Given $\varepsilon \geq 0$, an ordered subset $\mathscr{M} := \{M_1, \dots, M_s\} \subset \mathscr{K}$ will be called an ε-chain iff for all $k = 1, \dots, s-1$ we have $\rho(M_k, M_{k+1}) \leq \delta(\mathscr{K}) + \varepsilon$. A 0-chain will be called simply a chain. A chain (an ε-chain) is called maximal iff it is not contained in any other chain (ε-chain). Let $\mathscr{K}^*$ (resp., $\mathscr{K}_{\varepsilon}^*$) be the class of all unions of maximal chains (resp., ε-chains) from $\mathscr{K}$.

Let $D \subset \mathbb{R}^m$ be any set which consists of ν compact connected components, say, $D_1, D_2, \dots, D_{\nu}$. Denote $\mathscr{D} := \{D_1, \dots, D_{\nu}\}$. Now, we construct a *hierarchical cluster ε-tree* of the set D. First, we define the classes of sets $\mathscr{D}_j(\varepsilon),\ j = 1, \dots, J$ and the numbers $0 =: \delta_0(\varepsilon) < \delta_1(\varepsilon) < \dots < \delta_J(\varepsilon)$ by the following recursive procedure:

$j := 1$;

$\mathscr{D}_j(\varepsilon) := \mathscr{D}$;

$\delta_j(\varepsilon) := \delta(\mathscr{D})$;

while $\operatorname{card}(\mathscr{D}_j(\varepsilon)) > 1$ **do**

begin

$\qquad\mathscr{D}_{j+1}(\varepsilon) := (\mathscr{D}_j(\varepsilon))_{\varepsilon}^*$;

$\qquad j := j + 1$;

$\qquad\delta_j(\varepsilon) := \delta(\mathscr{D}_j(\varepsilon))$;

end;

$J := j$.

Define $\mathscr{V}(\varepsilon) := \bigcup_{j=1}^{J} \mathscr{D}_j(\varepsilon)$. We consider $\mathscr{V}(\varepsilon)$ as the set of all vertices of a graph $\mathscr{G}_{\varepsilon} = \mathscr{G}_{\varepsilon}(D)$. Two vertices V_1 and V_2 are connected with an edge iff $V_1 \supset V_2$ and there is no $V \in \mathscr{V}(\varepsilon)$ such that $V_1 \supset V \supset V_2$. Clearly, $\mathscr{G}_{\varepsilon}$ is a tree. We assign to each vertex V of the graph $\mathscr{G}_{\varepsilon}$ its *ε-height,* defined as the value of $\delta_j(\varepsilon)$ for the maximal j such that $V \in \mathscr{D}_j(\varepsilon)$. The ε-height of V will be denoted $\chi_\varepsilon(V)$. Clearly, $V_1 \subset V_2$ implies $\chi_\varepsilon(V_1) < \chi_\varepsilon(V_2)$.

If $\varepsilon = 0$, we call $\mathscr{G}(D) := \mathscr{G}_0(D)$ the *hierarchical cluster tree* of the set D. In this case we also skip ε in other notations introduced above (say, $\delta_j := \delta_j(0)$, $\chi := \chi_0$, etc.).

In what follows, a tree $\mathscr{G}(\operatorname{supp}(P))$ will be called the hierarchical cluster tree of the distribution P. It will be denoted simply by $\mathscr{G}(P)$. We also introduce the notation

$$\delta(P) := \min_{1\le j\le J}(\delta_j - \delta_{j-1}), \tag{4.1}$$

where $\delta_j,\ j = 0, \dots, J$ are the numerical characteristics of the cluster structure of the support of P, defined above.

Given a nondecreasing nonnegative function $\varepsilon \mapsto \tau(\varepsilon)$ with $\tau(0+) = 0$ and a number $\Delta > 0$, we define the class $\mathscr{P}^{\tau,\Delta}$ of all Borel probability measures P on $\mathbb{R}^m$ such that:

1. $\operatorname{supp}(P) \subset B(0;1)$.
2. For all $\varepsilon \in (0,1)$ and for any ball B of radius ε, the condition $B \cap \operatorname{supp}(P) \neq \emptyset$ implies $P(B^+) \ge \tau(2\varepsilon)$.
3. The set $\operatorname{supp}(P)$ consists of a finite number of connected components and $\delta(P) \ge \Delta$.

We denote $B^{--} := (B^-)^-$. Let $\mathscr{N}^{--} := \mathscr{N}^{--}(\varepsilon)$ be a minimal family of balls of radius $\varepsilon/4$, covering $B(0;1)$, and let $\mathscr{N} := \mathscr{N}(\varepsilon) := \{B : B^{--} \in \mathscr{N}^{--}(\varepsilon)\}$.

Given ε, γ and Ψ, define

$$\begin{aligned}\hat{\mathscr{S}}_n &:= \hat{\mathscr{S}}_n(\varepsilon, \gamma, \Psi)\\ &:= \bigcup\bigl\{\bar{B}\colon\ B \in \mathscr{N}(\varepsilon)\ \exists\ B' \in \mathscr{N}(\varepsilon)\colon\ \hat{P}_{n,\Psi}(B') \ge 2\gamma \text{ and } \operatorname{dist}(B, B') \le 2\varepsilon\bigr\}.\end{aligned}$$

Now we can define an estimator $\hat{\nu}_n := \hat{\nu}_n(\varepsilon, \gamma, \Psi)$ of the number of clusters $\nu := \nu(P)$ of the support as the number of clusters of the set $\hat{\mathscr{S}}_n$. Let $\hat{\mathscr{C}}_n := \{\hat{C}_j^{(n)} :\ j = 1, \dots, \hat{\nu}_n\}$ be the class of all clusters of the set $\hat{\mathscr{S}}_n$. We also define the *empirical hierarchical cluster tree,* based on the observations $(Y_1, \dots, Y_n)$, as the tree $\hat{\mathscr{G}}_n := \mathscr{G}_{7\varepsilon}(\hat{\mathscr{S}}_n)$. The *empirical height* on this tree is defined as 7ε-height; it will be denoted $\hat{\chi}_n$. In what follows, we use $(\hat{\mathscr{G}}_n, \hat{\chi}_n)$ as the estimator of the tree $\mathscr{G} := \mathscr{G}(P)$ with the height $\chi = \chi_0$ on it. The definition of the empirical cluster tree above leads to an agglomerative algorithm of hierarchical clustering similar to the ones described in the literature [see Gordon (1996) for the review on the subject]. The main difference is that these algorithms should be now applied not to the original data points $(Y_1, \dots, Y_n)$, but rather to the centers of the "massive" balls selected via deconvolution method.

THEOREM 4.1. *Suppose that* $\varepsilon < (\Delta/14)$, $\gamma < \tau(\varepsilon/2)/6$ *and*

$$\Psi\{x : |x| \ge \varepsilon/2\} \le \gamma. \tag{4.2}$$

Then there exist $\Lambda > 0$ *and* $q \in (0,1)$ *such that*

$$\sup_{P\in\mathscr{P}^{\tau,\Delta}} \mathbb{P}\{(\hat{\mathscr{G}}_n, \hat{\chi}_n) \not\approxeq (\mathscr{G}, \chi)\} \le \Lambda q^n. \tag{4.3}$$

The next statement shows that with no knowledge of the parameters characterizing the geometry of the support, one can still construct an estimate of the hierarchical cluster tree that converges with a rate arbitrarily close to the exponential. As in Section 3, let $\{q_n\}$ be a sequence of positive numbers such that $q_n \uparrow 1$ as $n \to \infty$. Recall the definition of the function ρ: $\rho(R) := \inf\{|\tilde{\mu}(t)|: |t|_\infty \le R\}$.

THEOREM 4.2. *Suppose that* $\rho(R) > 0$ *for all* $R > 0$. *Let* $\varepsilon = \varepsilon_n$, $\gamma = \gamma_n$ *and* $\Psi = \Psi_n$ *be the same as in Theorem* 3.2. *Then, for all* τ *such that* $\varepsilon^m = O(\tau(\varepsilon))$ *and for all* $\Delta > 0$,

$$\sup_{P\in\mathscr{P}^{\tau,\Delta}} \mathbb{P}\{(\hat{\mathscr{G}}_n, \hat{\chi}_n) \not\approxeq (\mathscr{G}, \chi)\} = O(q_n^n) \quad \text{as } n \to \infty.$$

COROLLARY 4.3. *Under the conditions of Theorem* 4.1, *there exist* $\Lambda > 0$ *and* $q \in (0,1)$ *such that*

$$\sup_{P\in\mathscr{P}^{\tau,\Delta}} \mathbb{P}\{\hat{\nu}_n \neq \nu(P)\} \le \Lambda q^n. \tag{4.4}$$

Under the conditions of Theorem 4.2, *we have, for all* τ *such that* $\varepsilon^m = O(\tau(\varepsilon))$ *and for all* $\Delta > 0$,

$$\sup_{P\in\mathscr{P}^{\tau,\Delta}} \mathbb{P}\{\hat{\nu}_n \neq \nu(P)\} = O(q_n^n) \quad \text{as } n \to \infty. \tag{4.5}$$

A compact set $D \subset \mathbb{R}^m$ is called *k-connected* iff its boundary ∂D consists of k connected components. Suppose that D is k-connected for some $k \ge 1$, and denote $v(D) := k$ and $v(P) := v(\text{supp}(P))$. Our next goal is to suggest an estimation procedure for the quantity $v(P)$, based on the sample $(Y_1, \dots, Y_n)$.

Denote $\mathscr{P}^{\tau}_{\Delta,\delta}$ the class of all Borel probability measures P on $\mathbb{R}^m$ such that:

1. $\text{supp}(P) \subset B(0;1)$.
2. For all $\varepsilon > 0$ and for any ball B of radius ε, the condition $B \cap \text{supp}(P) \neq \varnothing$ implies $P(B^+) \ge \tau(2\varepsilon)$.
3. $v(P) < +\infty$ and

$$\min_{i\neq j} \text{dist}(\Gamma_i; \Gamma_j) \ge \Delta, \tag{4.6}$$

   where $\{\Gamma_i: 1 \le i \le v(P)\}$ is the set of all connected components of the boundary $\partial(\text{supp}(P))$.
4. For all $\varepsilon \le \delta$ and for any $x \in \partial(\text{supp}(P))$, there exists a ball B of radius 2ε such that $\text{dist}(B; \text{supp}(P)) \ge \varepsilon$ and $\text{dist}(x; B) \le 2\varepsilon$.

We define the set

$$\hat{\partial}_n(\hat{P}_n) := \bigcup\{\bar{B}:\ B \in \mathscr{N}(\varepsilon):\ \exists\ B',\ B_1, B_2 \in \mathscr{N}(\varepsilon):\ \operatorname{dist}(B, B') \le 6\varepsilon,$$

$$\operatorname{dist}(B', B_1) \le 3\varepsilon,\ \operatorname{dist}(B', B_2) \le 3\varepsilon,\ \hat{P}_n(B_1) \ge 2\gamma,\ \hat{P}_n(B_2) < 2\gamma\}.$$

Let $\hat{\upsilon}_n$ denote the number of the clusters of $\hat{\partial}_n(\hat{P}_n)$.

THEOREM 4.4. *Suppose that* $\varepsilon < (\Delta/40) \wedge \delta$, $\gamma < \tau(\varepsilon/2)/6$ *and the condition* (4.2) *holds. Then there exist* $\Lambda > 0$ *and* $q \in (0, 1)$ *such that*

$$\sup_{P \in \mathscr{P}^{\tau}_{\Delta, \delta}} \mathbb{P}\{\hat{\upsilon}_n \ne \upsilon(P)\} \le \Lambda q^n.$$

5. Adaptive estimation of geometric characteristics of the support. The development of adaptive versions of the estimation procedures considered in the previous sections poses a number of problems. One can think about at least two sides of the adaptation: the adaptation of the estimates to the unknown geometry of the support and the adaptation to the unknown distribution of the noise. We give below only brief comments on the first side of the problem; the second one will be developed in some detail in this and, especially, in the next section.

The choice of the parameters of the estimates of geometric characteristics in convolution models (such as ε, γ and the measure Ψ in the procedures of dimension and cluster structure estimation) requires the knowledge of the minimal size of "geometric features" that are to be recovered in the process of deconvolution. For instance, in the case of the estimation of the hierarchical cluster tree one of the important quantities to know is $\delta(P)$, defined by (4.1). Theorems 3.2 and 4.2 show that there are estimates of the dimension and of the hierarchical cluster tree of the support that converge to the true parameters with probability of the error decreasing almost exponentially fast and the estimates are not using the information about the geometry of the support. However, this does not allow one to determine the values of the tunable parameters of the estimates that provide the recovery of the geometric characteristics for a given finite sample. One of the aspects of the problem is to find the minimal size of geometric features of the support that can be recovered for a given sample size with a given level of confidence. The bounds obtained in the proofs of the main results of Sections 3 and 4 can be used, in principle, to address such kind of questions, since the constants involved in the bounds could be written explicitly. Consider, for instance, the problem of estimation of hierarchical cluster tree of the support in the setting of Section 4. Suppose that $|\cdot| = |\cdot|_\infty$ and assume, for simplicity, that P is a uniform distribution on its support and that there exists a number $\rho(P) > 0$ such that for all $\varepsilon \le \rho(P)$ and for all $x \in \operatorname{supp}(P)$ there exists a ball B of radius ε such that $B \subset \operatorname{supp}(P)$, $B \ni x$. For example, if $\operatorname{supp}(P)$ is a finite union of disjoint balls, then $\rho(P)$ is the smallest radius of the balls. Note that the condition (2) of the definition of the set $\mathscr{P}^{\tau, \Delta}$ holds for such a P and for $\varepsilon \le \rho(P)$ with $\tau(2\varepsilon) = \varepsilon^m$.

Define for a given $\alpha \in (0,1)$ $\bar{\varepsilon} := \inf\Big\{\varepsilon > 0\colon\ \exists\ \Psi\ \zeta_\alpha(\Psi;\varepsilon;m) < C_m\Big\}$, where

$$\zeta_\alpha(\Psi;\varepsilon;m) := \|K_\Psi\|_\infty(\|K_\Psi\|_{L_1} \vee 1)\frac{\log(2/\alpha) + m\ \log(1/\varepsilon)}{n} \vee \frac{\Psi\{x\colon\ |x| \geq \varepsilon/2\}}{\varepsilon^m}.$$

Here K_Ψ is the kernel of the deconvolution estimate defined by (2.5). It can be derived from the proof of Theorem 4.1 (using Remark 7.5 below) that with some constant $C_m > 0$ (that could be written explicitly) and under the assumption $\delta(P)/14 \bigwedge \rho(P) > \bar{\varepsilon}$ we have $\mathbb{P}\big\{(\hat{\mathscr{G}}_n, \hat{\chi}_n) \not\cong (\mathscr{G}, \chi)\big\} \leq \alpha$. The estimates of the hierarchical cluster tree $(\hat{\mathscr{G}}_n, \hat{\chi}_n)$ in the last bound use the values of the parameters $\varepsilon > \bar{\varepsilon}$, $\gamma = C_m\varepsilon^m$ and Ψ such that $\zeta_\alpha(\Psi;\varepsilon;m) < C_m$. In general, we do not know whether the condition $\delta(P)/14 \bigwedge \rho(P) > \bar{\varepsilon}$ is satisfied, or not. But the deconvolution estimates still provide a partial recovery of the cluster structure of the support even in the case when such a condition fails. [If $S_\varepsilon(P)$ denotes the union of all the clusters of supp(P) that do not contain a ball of radius ε (i.e., that are too small to be recovered at the level of resolution ε), then it follows from the proof of Theorem 4.1 that for $\varepsilon = \bar{\varepsilon}$ we have with probability at least $1-\alpha$ $\operatorname{supp}(P)\setminus S_\varepsilon(P) \subset \hat{\mathscr{S}}_n \subset (\operatorname{supp}(P))^{7\varepsilon}$, so, the estimate $\hat{\mathscr{S}}_n$ still allows one to recover large enough clusters]. Thus, the quantity $\bar{\varepsilon}$ characterizes the size of the features of the support that can be recovered with confidence $1-\alpha$.

To be more specific, let ν be the distribution of a random vector in $\mathbb{R}^m$ with independent components that have bilateral exponential distribution (i.e., their density is $\frac{1}{2}e^{-|x|}$ and their characteristic function is $(1+t^2)^{-1}$). Suppose that the noise $\xi_j = \sigma\eta_j$, where $\sigma > 0$ and η_j, $j = 1,\dots,n$ are i.i.d. symmetric vectors in $\mathbb{R}^m$ with common distribution ν. Then one can show by a simple computation that for some constant $C(m) > 0$,

$$\bar{\varepsilon} \leq \tilde{\varepsilon} := \inf\bigg\{\varepsilon > 0\colon \bigg[1 + \Big(\frac{\sigma}{\varepsilon}\Big)^{4m}\Big(\log\frac{1}{\varepsilon}\Big)^{2m}\bigg] \times \Big(\frac{1}{\varepsilon}\Big)^m\Big(\log\frac{1}{\varepsilon}\Big)^{m/2}\frac{\log(2/\alpha) + m\ \log(1/\varepsilon)}{n} \leq C(m)\bigg\}.$$

We will use $\varepsilon = \tilde{\varepsilon}$ and $\Psi = N(0;\delta I)$ with $\delta \asymp \varepsilon(\log(1/\varepsilon))^{-1/2}$ in the estimates of the hierarchical cluster tree. If the noise is small (say, under the assumption that $\sigma \leq \varepsilon$), we find that the minimal size of the features that can be recovered with probability at least $1-\alpha$ is of the order $\big(\frac{\log(2/\alpha)}{n}\big)^{1/m}\big(\log(n/\log(2/\alpha))\big)^{7/2}$ (which can be shown to be an optimal bound up to a logarithmic factor). In the case of large noise (say, for $\sigma = 1$) the same quantity is of the order $\big(\frac{\log(2/\alpha)}{n}\big)^{1/5m}\big(\log(n/\log(2/\alpha))\big)^{7/10}$ Despite the fact that the constants in the above bounds can be written explicitly, they are currently very far from being best possible and the bounds are rather conservative. Computations based on the bounds show that the sample sizes needed to achieve reasonable level of "resolution" in dimension estimation and cluster analysis problems can be very large. The development of efficient adaptive estimates of geometric characteristics in convolution models would require substantial improvements of the inequalities (based on the combination of the methods of probability and combinatorial geometry, e.g., sharper bounds on the covering numbers) and the study of asypmtotic behavior of the statistics estimating geometric structure of the support, which could be accomplished only for much more specialized models than the ones considered in the current paper.

We now turn to the deconvolution problems in the case of unknown distribution of the noise. We first assume that the probability distribution μ of the random noise ξ_j is unknown, but there is an estimator $\hat{\mu} := \hat{\mu}_n$ of μ, based on the data $(Y_1, \dots, Y_n)$. We use this estimator instead of μ in the deconvolution procedures of Section 2 and show (under some assumptions) that the exponential rates of convergence for dimension estimates and cluster structure estimates are preserved. In the next section, we define an estimate $\hat{\mu}$ in the case when μ is $N(0;\Sigma_0)$ with unknown covariance Σ_0. We assume in both sections that $|\cdot|$ is the standard Euclidean norm in $\mathbb{R}^m$.

Denote Λ_μ the set of all Borel probability measures λ on $\mathbb{R}^m$ such that $\lambda * \lambda' = \mu$ for some Borel probability measure λ' on $\mathbb{R}^m$. Clearly, $\mu \in \Lambda_\mu$. If $\tilde{\mu}(t) \neq 0$ for all $t \in \mathbb{R}^m$ (which is assumed throughout the paper), then the measure λ' defined by the relationship $\lambda * \lambda' = \mu$ is unique. Given $\lambda \in \Lambda_\mu$, we set $T_\mu \lambda = \lambda'$. Clearly, $T_\mu(T_\mu \lambda) \equiv \lambda$.

We start with the problem of the dimension estimation in the setting of Section 3 and under the assumption that the set $\mathscr{D}$ of all possible dimensions is finite. Consider a class of probability measures $\mathscr{P} \subset \mathscr{P}(\mathscr{D}, \Theta, C)$.

We denote $\hat{d}_n(\varepsilon;\gamma;\Psi;\nu)$ the dimension estimator of Section 3 with μ replaced by a probability distribution ν on $\mathbb{R}^m$.

THEOREM 5.1. *Suppose that* $\varepsilon < \varepsilon(\mathscr{D};\Theta;C)$, $\gamma < \gamma(\mathscr{D};\varepsilon;\Theta)$ *and* Ψ *satisfies the condition*

$$\Psi\{x\colon\ |x| \geq \varepsilon/2\} \leq \gamma/2. \tag{5.1}$$

Suppose also that $\hat{\mu} := \hat{\mu}_n$ *is an estimator of* μ, *based on the sample* $(Y_1, \dots, Y_n)$, *such that for some* $\Lambda > 0$ *and* $q \in (0,1)$,

$$\sup_{P\in\mathscr{P}} \mathbb{P}\Big\{\hat{\mu} \notin \Lambda_\mu \ \textit{or}\ (T_\mu \hat{\mu})\{x\colon\ |x| \geq \varepsilon/2\} \geq \gamma/2\Big\} \leq \Lambda q^n. \tag{5.2}$$

Let $\tilde{d}_n := \hat{d}_n(\varepsilon;\gamma;\Psi;\hat{\mu})$. *Then there exist* $\Lambda > 0$ *and* $q \in (0,1)$ *such that*

$$\sup_{P\in\mathscr{P}} \mathbb{P}\{\tilde{d}_n \neq \dim(P)\} \leq \Lambda q^n. \tag{5.3}$$

Next we consider clustering problems in the context and under the notations of Section 4. We assume that $\mathscr{P} \subset \mathscr{P}^{\tau,\Delta}$. Given ε, γ and a probability distribution ν on $\mathbb{R}^m$, one can define an empirical cluster tree $\hat{\mathscr{G}}_n(\varepsilon;\gamma;\Psi;\nu)$ and the height $\hat{\chi}_n(\varepsilon;\gamma;\Psi;\nu)$ on it as in Section 3, based on the distribution ν instead of μ. Similarly, we define the estimates of the number of clusters $\hat{\nu}_n(\varepsilon;\gamma;\Psi;\nu)$ and of the number of connected components of the boundary $\hat{\upsilon}_n(\varepsilon;\gamma;\Psi;\nu)$.

THEOREM 5.2. *Suppose that* $\varepsilon < (\Delta/14)$, $\gamma < \tau(\varepsilon/2)$ *and*

$$\Psi\{x\colon\ |x| \geq \varepsilon/4\} \leq \gamma/2. \tag{5.4}$$

Suppose also that $\hat{\mu} := \hat{\mu}_n$ is an estimator of μ, based on the sample $(Y_1, \dots, Y_n)$, such that for some $\Lambda > 0$ and $q \in (0, 1)$,

$$\sup_{P\in\mathscr{P}} \mathbb{P}\Big\{\hat{\mu} \notin \Lambda_\mu \text{ or } (T_\mu \hat{\mu})\{x\colon\ |x| \geq \varepsilon/4\} \geq \gamma/2\Big\} \leq \Lambda q^n. \tag{5.5}$$

We set $\tilde{\mathscr{G}}_n := \hat{\mathscr{G}}_n(\varepsilon; \gamma; \Psi; \hat{\mu})$ and $\tilde{\chi}_n := \hat{\chi}_n(\varepsilon; \gamma; \Psi; \hat{\mu})$. Then there exist $\Lambda > 0$ and $q \in (0, 1)$ such that

$$\sup_{P\in\mathscr{P}} \mathbb{P}\big\{(\tilde{\mathscr{G}}_n, \tilde{\chi}_n) \not\cong (\mathscr{G}, \chi)\big\} \leq \Lambda q^n. \tag{5.6}$$

COROLLARY 5.3. *Under the conditions of Theorem 5.2, there exist $\Lambda > 0$ and $q \in (0, 1)$ such that*

$$\sup_{P\in\mathscr{P}} \mathbb{P}\big\{\tilde{\nu}_n \neq \nu(P)\big\} \leq \Lambda q^n, \tag{5.7}$$

where $\tilde{\nu}_n := \hat{\nu}_n(\varepsilon; \gamma; \Psi; \hat{\mu})$.

6. The case of normal noise with unknown covariance. Assume now that the random variables ξ_j, $j = 1, 2, \dots$ in the model (2.1) have normal distribution in $\mathbb{R}^m$ with mean 0 and unknown covariance Σ_0. Also assume that the matrix Σ_0 satisfies the condition $\|\Sigma_0\| \leq M$ with a constant $M > 0$.

We suggest below an estimate $\hat{\Sigma} := \hat{\Sigma}_{n,R}$ of the covariance matrix Σ_0 such that the normal distribution $\hat{\mu}$ with parameters 0 and $\hat{\Sigma}$ satisfies the conditions of the theorems of the previous section and thus it can be used in the procedures of recovery of unknown dimension and cluster structure of the support. It should be mentioned that a related problem of estimation of the order of finite mixtures of distributions in the case of translation mixtures *with unknown scale parameter* has been studied recently (using different methods) by Dacunha-Castelle and Gassiat (1997).

We start with introducing some notations and imposing the conditions on the unknown distribution P. Given $R > 0$, λ_R denote the uniform distribution on the ball $B(0; R)$ in $\mathbb{R}^m$. Let $p \in [1, +\infty]$. Denote

$$\delta_p(P; R) := R^{-2} \|\log|\tilde{P}|\|_{L_p(B(0,R);\lambda_R)}, \qquad R > 0. \tag{6.1}$$

In what follows, we assume that

$$\delta_p(P; R) \to 0 \quad \text{as } R \to \infty. \tag{6.2}$$

Let U be the uniform distribution on the unit sphere S^{m-1}. Note that in the case of the normal distribution $\mu = N(0; \Sigma_0)$ we have, by a change to polar coordinates, for $p \in (1, \infty)$,

$$\begin{aligned}\|\log|\tilde{\mu}|\|^p_{L_p(B(0;R);\lambda_R)} &= 2^{-p}\lambda(B(0;R))^{-1}\int_{B(0;R)} (\Sigma_0 t, t)^p\, dt \\ &= 2^{-p}\frac{m}{2p+m} R^{2p} \int_{S^{m-1}} (\Sigma_0 v, v)^p U(dv),\end{aligned}$$

which implies that

$$\delta_p(\mu; R) = R^{-2}\|\log|\tilde{\mu}|\|_{L_p(B(0;R);\lambda_R)} \tag{6.3}$$
$$= 2^{-1}\left(\frac{m}{2p+m}\right)^{1/p}\left(\int_{S^{m-1}}(\Sigma_0 v, v)^p U(dv)\right)^{1/p}.$$

For $p = \infty$ we get

$$\delta_\infty(\mu; R) = R^{-2}\|\log|\tilde{\mu}|\|_{L_\infty(B(0;R);\lambda_R)} = 2^{-1}\|\Sigma_0\|. \tag{6.4}$$

Thus for normal distributions the condition (6.2) does not hold. Note that if P is normal, it becomes unidentifiable in the case of convolution model with normal noise. This possibility is excluded due to the condition (6.2). On the other hand, condition (6.2) holds for non-normal stable distributions and for many distributions with compact support.

We denote in what follows the empirical measure based on the sample $(Y_1, \dots, Y_n)$ by Q_n. Given $R > 0$, define an estimate $\check{\Sigma} := \check{\Sigma}_{n,R,M}$ of Σ_0 as a nonnegatively definite matrix Σ with $\|\Sigma\| \le M$ that minimizes the functional

$$\Sigma \mapsto \left\|\log|\tilde{Q}_n(\cdot)|I_{\{\log|\tilde{Q}_n(\cdot)|\le 2MR^2\}} + \tfrac{1}{2}(\Sigma\cdot, \cdot)\right\|_{L_p(B(0;R);\lambda_R)}. \tag{6.5}$$

The idea behind this definition is that the theoretical version of this empirical functional can be written as

$$\Sigma \mapsto \left\|\log|\tilde{Q}(\cdot)| + \tfrac{1}{2}(\Sigma\cdot, \cdot)\right\|_{L_p(B(0;R);\lambda_R)} \tag{6.6}$$
$$= \left\|\log|\tilde{P}(\cdot)| + \tfrac{1}{2}((\Sigma - \Sigma_0)\cdot, \cdot)\right\|_{L_p(B(0;R);\lambda_R)}.$$

If $R > 0$ is large enough, then, under the condition (6.2), the term $\frac{1}{2}((\Sigma - \Sigma_0)\cdot, \cdot)$ in the functional will dominate the first term and the minimum will be attained for Σ close to Σ_0. To study the situation more precisely, we prove the following theorem. Given $S \subset \mathbb{R}^m$, let $\mathscr{P}(S)$ denote the set of all Borel probability measures P with $\mathrm{supp}(P) \subset S$.

THEOREM 6.1. *There exists a universal constant $A = A(m; p)$ such that for some $\Lambda \ge 1$ and $q \in (0, 1)$,*

$$\sup_{\|\Sigma_0\|\le M}\ \sup_{P\in\mathscr{P}(B(0;1))} \mathbb{P}\{\|\check{\Sigma}_{n,R,M} - \Sigma_0\| \ge A\delta_p(P; R)\} \le \Lambda q^n. \tag{6.7}$$

Moreover, Λ and q can be choosen to satisfy, with some constant $C > 0$, the conditions

$$\log\ \Lambda \le CMR^2 \text{ and } \left|\log\ \log\frac{1}{q}\right| \le CMR^2 \tag{6.8}$$

for all large enough $R > 0$ and $M > 0$.

The following statement easily follows from the bound (6.8).

COROLLARY 6.2. *Suppose that* $0 < q_n \uparrow 1$ *as* $n \to \infty$. *Suppose also that* $R_n, M_n \to \infty$ *as* $n \to \infty$ *and*

$$M_n R_n^2 = o(|\log\ \log(q_n^{-1})| \wedge \log\ n) \quad as\ n \to \infty.$$

Then, for all $M > 0$,

$$\sup_{\|\Sigma_0\| \le M} \sup_{P \in \mathscr{P}(B(0;1))} \mathbb{P}\{\|\check{\Sigma}_{n, R_n, M_n} - \Sigma_0\| \ge A\delta_p(P; R_n)\} = o(q_n^n) \quad as\ n \to \infty. \tag{6.9}$$

We show now how to construct an estimate $\hat{\Sigma}$ such that the normal distribution $\hat{\mu} := N(0; \hat{\Sigma})$ satisfies condition (5.2) or (5.5).

We assume, in addition, that Σ_0 is positively definite, so that, for some $\beta > 0$ and for all $v \in S^{m-1}$,

$$(\Sigma_0 v, v) > \beta. \tag{6.10}$$

Define, for a sufficiently small number $\delta > 0$, the estimate $\hat{\Sigma} := \hat{\Sigma}_{n, R, M, \delta} := (1 - \delta)\check{\Sigma}_{n, R, M}$. Let $\mathscr{P}$ be a set of Borel probability measures on $\mathbb{R}^m$ and let $\delta_p(\mathscr{P}; R) := \sup_{P \in \mathscr{P}} \delta_p(P; R)$. Let B denote the maximal positive number such that $\mathbb{P}\{|Z| \ge u\} \le \exp(-Bu^2)$, $u \ge 0$, Z being the standard normal vector in $\mathbb{R}^m$ ($B = B_m$ is a constant, depending only on m).

PROPOSITION 6.3. *Suppose that* $\delta_p(\mathscr{P}; R) \to 0$ as $R \to \infty$. *Let* $\varepsilon > 0$ *and* $\gamma > 0$. *Suppose that* δ *is sufficiently small* (*specifically*,

$$\delta \le \delta(\gamma; M; \beta; \varepsilon) := 2 \wedge \frac{B}{\log(\gamma^{-1})}(M + \beta/4)^{-1}\varepsilon\Big). \tag{6.11}$$

Choose $R > 0$ *such that* $\delta_p(\mathscr{P}; R) \le (\beta/(4A))\delta$. *Then there exists* $\Lambda > 0$ *and* $q := q \in (0, 1)$ *such that for* $\hat{\mu} := N(0; \hat{\Sigma})$,

$$\sup_{P \in \mathscr{P}} \mathbb{P}\Big\{\hat{\mu} \notin \Lambda_\mu\ or\ (T_\mu \hat{\mu})\{x : |x| \ge \varepsilon\} \ge \gamma\Big\} \le \Lambda q^n. \tag{6.12}$$

Moreover, Λ *and* q *satisfy the condition* (6.8).

PROPOSITION 6.4. *Suppose that for some* $\alpha > 0$, $\delta_p(\mathscr{P}; R) = O(R^{-\alpha})$ as $R \to \infty$. *Let* $0 < q_n \uparrow 1$ *as* $n \to \infty$ *and* $q_n \le e^{-1/n}$ *for all* $n \ge 1$. *Choose sequences* $\{\varepsilon_n\}$ *and* $\{\gamma_n\}$ *such that* $\varepsilon_n \to 0$, $\gamma_n \to 0$,

$$\log(\varepsilon_n^{-1}) + \log\ \log(\gamma_n^{-1}) = o(\log|\log\ \log(q_n^{-1})|) \quad as\ n \to \infty.$$

Choose also sequences $\{R_n\}$, $\{M_n\}$ *and* $\{\delta_n\}$ *such that* $R_n, M_n \to \infty$, $\delta_n \to 0$ *as* $n \to \infty$ *and*

$$\log\ R_n = o(\log|\log\ \log(q_n^{-1})|),\ \log(\delta_n^{-1}) = o(\log\ R_n),$$

$$\log\ M_n + \log(\varepsilon_n^{-1}) + \log\ \log(\gamma_n^{-1}) = o(\log(\delta_n^{-1})) \quad as\ n \to \infty.$$

Let $\hat{\Sigma} := \hat{\Sigma}_{n, R_n, M_n, \delta_n}$ *and* $\hat{\mu} := \hat{\mu}_n := N(0; \hat{\Sigma})$. *Then*

$$\sup_{P \in \mathscr{P}} \mathbb{P}\Big\{\hat{\mu} \notin \Lambda_\mu \text{ or } (T_\mu \hat{\mu})\{x\colon\ |x| \geq \varepsilon_n\} \geq \gamma_n\Big\} = o(q_n^n).$$

Propositions 6.3 and 6.4 along with Theorems 5.1 and 5.2 allow one to get versions of the results of Sections 3 and 4 in the case of normal noise with unknown covariance. For instance, the statement of Theorem 3.2 takes in this case the following form.

THEOREM 6.5. *Let* $\Theta > 0$, $C > 0$ *and* $\mathscr{D} \subset [0, m]$ *be a finite subset. Suppose that* $\mathscr{P} \subset \mathscr{P}(\mathscr{D}, \Theta, C)$ *and for some* $\alpha > 0$, $\delta_p(\mathscr{P}; R) = O(R^{-\alpha})$ as $R \to \infty$. *Let* $0 < q_n \uparrow 1$ *as* $n \to \infty$ *and* $q_n \leq e^{-1/n}$ *for all* $n \geq 1$. *Choose a sequence* $\{\varepsilon_n\}$ *such that* $\varepsilon_n \to 0$ *and*

$$\log(\varepsilon_n^{-1}) = o(\log|\log\ \log(q_n^{-1})|) \quad \textit{as } n \to \infty.$$

Let $\gamma_n := \varepsilon_n^m / (\log\ \varepsilon_n^{-1})$. *Define the estimate* $\hat{\mu}_n$ *as in Proposition* 6.4. *Define* Ψ_n *as in Theorem* 3.2. *Then, the estimate* $\tilde{d}_n := \hat{d}_n(\varepsilon_n, \gamma_n, \Psi_n, \hat{\mu}_n)$ *satisfies the following bound*:

$$\sup_{P \in \mathscr{P}} \mathbb{P}\{\tilde{d}_n \neq \dim(P)\} = O(q_n^n) \quad \textit{as } n \to \infty.$$

Note that the estimate $\tilde{d}_n$ depends neither on unknown parameters of the class $\mathscr{P}$ (such as $\Theta, C, \alpha > 0, \mathscr{D}$) nor on the numbers M, β related to the unknown covariance Σ_0; it depends only on the sequence $\{q_n\}$.

7. Proofs of the main results. We give below an exponential bound for large deviations of the deconvolving empirical measures, which is frequently used in what follows. Denote

$$\beta_m(\Psi) := \beta_m(\Psi; \mu) := (2\pi)^{-m} \int_{\mathbb{R}^m} \left| \frac{\tilde{\Psi}(t)}{\tilde{\mu}(t)} \right| dt.$$

We will denote by "mes" the Lebesgue measure.

LEMMA 7.1. *Suppose that* $\mathscr{C}$ *is a class of bounded Borel subsets in* $\mathbb{R}^m$. *Then for all* $\delta > 0$ *and for all Borel probability measures* P *on* $\mathbb{R}^m$ *the following bound holds*:

$$\begin{aligned} &\mathbb{P}\Big\{\sup\nolimits_{\lambda \in \Lambda_\mu} \sup\nolimits_{C \in \mathscr{C}} \big|\hat{P}_{n,\Psi,\lambda}(C) - \mathbb{E}\hat{P}_{n,\Psi,\lambda}(C)\big| \geq \delta\Big\} \\ &\quad \leq 4\ \exp\Big\{-\frac{n\delta^2}{32\beta_m^2(\Psi;\mu)\sup_{C \in \mathscr{C}}\ \text{mes}^2(C)}\Big\}. \end{aligned} \tag{7.1}$$

PROOF. Introduce the notation

$$\varphi(t) := \mathbb{E}e^{i(t, Y)} \quad \text{and} \quad \varphi_n(t) := n^{-1} \sum_{j=1}^n \exp(i(t, Y_j)).$$

For a bounded Borel set $C \subset \mathbb{R}^m$ and for $\lambda \in \Lambda_\mu$, we have (with $\lambda' := T_\mu \lambda$)

$$K_{\Psi, \lambda}(C) = (2\pi)^{-m} \int_{\mathbb{R}^m} \int_C \exp(-i(t, x))\, dx \frac{\tilde{\Psi}(t)}{\tilde{\lambda}(t)}\, dt$$
$$= (2\pi)^{-m} \int_{\mathbb{R}^m} \int_C \exp(-i(t, x))\, dx \frac{\tilde{\Psi}(t)}{\tilde{\mu}(t)} \tilde{\lambda}'(t)\, dt$$

and we get

$$\hat{P}_{n, \Psi, \lambda}(C) = n^{-1} \sum_{j=1}^{n} K_\lambda(C - Y_j)$$
$$= (2\pi)^{-m} \int_{\mathbb{R}^m} n^{-1} \sum_{j=1}^{n} \int_{C - Y_j} \exp(-i(t, x))\, dx \frac{\tilde{\Psi}(t)}{\tilde{\mu}(t)} \tilde{\lambda}'(t)\, dt$$
$$= (2\pi)^{-m} \int_{\mathbb{R}^m} n^{-1} \sum_{j=1}^{n} \exp(i(t, Y_j)) \int_C \exp(-i(t, x))\, dx \frac{\tilde{\Psi}(t)}{\tilde{\mu}(t)} \tilde{\lambda}'(t)\, dt$$
$$= (2\pi)^{-m} \int_{\mathbb{R}^m} \varphi_n(t) \int_C \exp(-i(t, x))\, dx \frac{\tilde{\Psi}(t)}{\tilde{\mu}(t)} \tilde{\lambda}'(t)\, dt.$$

We also have

$$\mathbb{E}\hat{P}_{n, \Psi, \lambda}(C) = (2\pi)^{-m} \int_{\mathbb{R}^m} \varphi(t) \int_C \exp(-i(t, x))\, dx \frac{\tilde{\Psi}(t)}{\tilde{\mu}(t)} \tilde{\lambda}'(t)\, dt.$$

Therefore

$$\hat{P}_{n, \Psi, \lambda}(C) - \mathbb{E}\hat{P}_{n, \Psi, \lambda}(C)$$
$$= (2\pi)^{-m} \int_{\mathbb{R}^m} [\varphi_n(t) - \varphi(t)] \int_C \exp(-i(t, x))\, dx \frac{\tilde{\Psi}(t)}{\tilde{\mu}(t)} \tilde{\lambda}'(t)\, dt,$$

which implies

$$\sup_{\lambda \in \Lambda_\mu} \sup_{C \in \mathscr{C}} \left| \hat{P}_{n, \Psi, \lambda}(C) - \mathbb{E}\hat{P}_{n, \Psi, \lambda}(C) \right| \le \beta_m(\Psi; \mu) \sup_{C \in \mathscr{C}} \operatorname{mes}(C) \int_{\mathbb{R}^m} |\varphi_n(t) - \varphi(t)| \nu(dt), \tag{7.2}$$

where $\nu(dt) := \beta_m^{-1}(\Psi; \mu) |\tilde{\Psi}(t)/\tilde{\mu}(t)| dt$ is a Borel probability measure on $\mathbb{R}^m$.

Obviously, we have

$$n \int_{\mathbb{R}^m} |\varphi_n(t) - \varphi(t)| \nu(dt) \le \left\| \sum_{j=1}^{n} \xi_j \right\|_{L_1(\mathbb{R}^m; d\nu)} + \left\| \sum_{j=1}^{n} \eta_j \right\|_{L_1(\mathbb{R}^m; d\nu)}, \tag{7.3}$$

where

$$\xi_j(t) = \cos(t, Y_j) - \mathbb{E}\ \cos(t, Y), \qquad \eta_j(t) = \sin(t, Y_j) - \mathbb{E}\ \sin(t, Y).$$

The rest of the proof is based on the standard Bernstein–Hoeffding-type bounds. We get, using Jensen's inequality,

$$\begin{aligned}
\mathbb{E}\ \exp\Bigl\{\theta\Bigl\|\sum_{j=1}^{n}\xi_j\Bigr\|_{L_1(\mathbb{R}^m;d\nu)}\Bigr\} &= \mathbb{E}\ \exp\Bigl\{\theta\int_{\mathbb{R}^m}\Bigl|\sum_{j=1}^{n}\xi_j(t)\Bigr|\nu(dt)\Bigr\}\\
&\le \int_{\mathbb{R}^m}\mathbb{E}\ \exp\Bigl\{\theta\Bigl|\sum_{j=1}^{n}\xi_j(t)\Bigr|\Bigr\}\nu(dt)\\
&\le \int_{\mathbb{R}^m}\mathbb{E}\ \exp\Bigl\{\theta\sum_{j=1}^{n}\xi_j(t)\Bigr\}\nu(dt)\\
&\quad + \int_{\mathbb{R}^m}\mathbb{E}\ \exp\Bigl\{-\theta\sum_{j=1}^{n}\xi_j(t)\Bigr\}\nu(dt).
\end{aligned}\tag{7.4}$$

Next we use a standard approach [see Ledoux and Talagrand (1991), page 31] to bound the expectation $\mathbb{E}\ \exp(\theta\xi(t))$. Since $|\xi(t)|\le 2$ and $\mathbb{E}\xi(t)=0$, we have $\mathbb{E}\ \exp\{\theta\xi(t)\}\le\exp\{2\theta^2\}$.

Note. More specifically, we used above an elementary inequality

$$e^{\theta x}\le \operatorname{ch}(2\theta)+\frac{x}{2}\operatorname{sh}(2\theta)\le\exp(2\theta^2)+\frac{x}{2}\operatorname{sh}(2\theta),$$

which follows from the convexity of the function $x\mapsto e^{\theta x}$ and from the identity

$$\theta x = 2\theta\frac{1+x/2}{2}-2\theta\frac{1-x/2}{2}.$$

Hence

$$\mathbb{E}\ \exp\Bigl\{\theta\sum_{j=1}^{n}\xi_j(t)\Bigr\}=\Bigl(\mathbb{E}\ \exp\{\theta\xi(t)\}\Bigr)^n\le\exp\{2\theta^2 n\}.\tag{7.5}$$

The similar bound holds with θ replaced with $-\theta$. Thus, it easily follows from (7.4) and (7.5) that

$$\mathbb{E}\ \exp\Bigl\{\theta\Bigl\|\sum_{j=1}^{n}\xi_j\Bigr\|_{L_1(\mathbb{R}^m;d\nu)}\Bigr\}\le 2\ \exp\{2\theta^2 n\}.$$

Therefore,

$$\mathbb{P}\Bigl\{\Bigl\|\sum_{j=1}^{n}\xi_j\Bigr\|_{L_1(\mathbb{R}^m;d\nu)}\ge x\Bigr\}\le 2\ \exp\{2\theta^2 n-\theta x\}.$$

Plugging in the last bound $x=n\varepsilon$ and $\theta=\varepsilon/4$, yields

$$\mathbb{P}\Bigl\{\Bigl\|\sum_{j=1}^{n}\xi_j\Bigr\|_{L_1(\mathbb{R}^m;d\nu)}\ge n\varepsilon\Bigr\}\le 2\ \exp\Bigl\{-\frac{\varepsilon^2 n}{8}\Bigr\}.\tag{7.6}$$

Since a similar bound holds for $\sum_1^n\eta_j$, we get from (7.3) and (7.6) that

$$\mathbb{P}\Bigl\{\|\varphi_n-\varphi\|_{L_1(\mathbb{R}^m;d\nu)}\ge n\varepsilon\Bigr\}\le 4\ \exp\Bigl\{-\frac{\varepsilon^2 n}{32}\Bigr\}.$$

Now the last bound and (7.2) imply (7.1). □

REMARK 7.2. Note that Bernstein's inequality implies the following bound for any distribution of the noise μ with a density uniformly bounded by a constant $C_\mu > 0$ and for any ball $B \subset \mathbb{R}^m$ of radius $\varepsilon > 0$,

$$\mathbb{P}\Big\{\big|\hat{P}_{n,\Psi,\mu}(B) - \mathbb{E}\hat{P}_{n,\Psi,\mu}(B)\big| \geq \delta\Big\}$$
$$\leq 2\ \exp\left\{-\frac{n\delta}{4c_m 2^{-m}\varepsilon^m \|K\|_\infty}\left[\frac{\delta}{c_m C_\mu \varepsilon^m \|K\|_{L_1}} \wedge 1\right]\right\},$$

where $K := K_{\Psi,\mu}$ is defined by (2.5) and $c_m := 2^m v_m$, v_m denoting the volume of the unit ball in $\mathbb{R}^m$ with respect to a choosen norm [for instance, $v_m := 2\pi^{m/2}/(m\Gamma(m/2))$ in the case of the Euclidean norm]. Indeed, since $B - u \cap B - v \neq \emptyset$ implies $|u - v| \leq 2\varepsilon$, we have

$$\begin{aligned}
\mathbb{E}|\mathscr{K}(B-Y)|^2 &= \mathbb{E}\int_{\mathbb{R}^m}\int_{\mathbb{R}^m} K(u)K(v)I_{B-Y}(u)I_{B-Y}(v)\,du\,dv \\
&= \int_{\mathbb{R}^m}\int_{\mathbb{R}^m} K(u)K(v)\mathbb{P}(Y \in (B-u)\cap(B-v))\,du\,dv \\
&= \int_{\mathbb{R}^m}\int_{\mathbb{R}^m} K(u)K(v)Q((B-u)\cap(B-v))\,du\,dv \\
&\leq \int_{\{|u-v|\leq 2\varepsilon\}} |K(u)||K(v)|Q(B-u)\,du\,dv \\
&= \int_{\mathbb{R}^m} |K(u)|Q(B-u)\int_{\{|v-u|\leq 2\varepsilon\}} |K(v)|\,dv\,du \\
&\leq \|K\|_\infty c_m \varepsilon^m \int_{\mathbb{R}^m} |K(u)|Q(B-u)\,du \\
&= c_m\varepsilon^m \|K\|_\infty \int_B \int_{\mathbb{R}^m} q(x-u)|K(u)|\,du\,dx \\
&\leq c_m^2 2^{-m}\varepsilon^{2m}\|q\|_\infty \|K\|_\infty \|K\|_{L_1},
\end{aligned}$$

where q is the density of Q. Since also

$$|\mathscr{K}(B-Y)| = \Big|\int_{B-Y} K(u)\,du\Big| \leq \int_{B-Y} |K(u)|\,du \leq \|K\|_\infty c_m 2^{-m}\varepsilon^m,$$

and $\|q\|_\infty \leq C_\mu$, Bernstein's inequality does imply the bound. Note that $\|K\|_\infty \leq \beta_m(\Psi;\mu)$.

PROOF OF PROPOSITION 2.1. Indeed, by the assumption $\operatorname{card}(\tau(\mathscr{P})) \geq 2$, there exist $P_1, P_2 \in \mathscr{P}$ such that $P_1 \neq P_2$. This implies $\tilde{P}_1 \not\equiv \tilde{P}_2$. Since $\tilde{\mu} \not\equiv 0$, we also have $\tilde{P}_1\tilde{\mu} \not\equiv \tilde{P}_2\tilde{\mu}$, which yields $P_1 * \mu \not\equiv P_2 * \mu$. Both measures $P_1 * \mu$ and $P_2 * \mu$ are absolutely continuous (since μ is). It follows that the information deviation,

$$I[P_1 * \mu | P_2 * \mu] := \int_{\mathbb{R}^m} \log\frac{d(P_1 * \mu)}{d(P_2 * \mu)}\, d(P_1 * \mu) > 0.$$

Let m be the density of μ. By the assumptions, there exists a constant C such that $m(x) \le C,\ x \in \mathbb{R}^m$. Since $\int m \log m\, dx > -\infty$ and $I[\mu_\theta|\mu] < +\infty$ locally uniformly in θ and the sets $\mathrm{supp}(P_1)$, $\mathrm{supp}(P_2)$ are compact, we get

$$\inf_{y_1 \in \mathrm{supp}(P_1)} \inf_{y_2 \in \mathrm{supp}(P_2)} \int_{\mathbb{R}^m} \left[\log \frac{m(x-y_1)}{C}\right] m(x-y_2)\, dx > -\infty.$$

Note. The condition $I[\mu_\theta|\mu] < +\infty$ locally uniformly in θ implies that

$$\inf_{y_1 \in \mathrm{supp}(P_1)} \inf_{y_2 \in \mathrm{supp}(P_2)} \int_{\mathbb{R}^m} \left[\log \frac{m(x-y_1)}{C} - \log \frac{m(x-y_2)}{C}\right] m(x-y_2)\, dx > -\infty.$$

Since $\int m \log m\, dx > -\infty$, we also get

$$\int_{\mathbb{R}^m} \left[\log \frac{m(x-y_2)}{C}\right] m(x-y_2)\, dx = \int_{\mathbb{R}^m} \left[\log \frac{m(x)}{C}\right] m(x)\, dx > -\infty,$$

which implies the bound.

By the Fubini theorem, it follows that

$$\int_{\mathbb{R}^m} \int_{\mathbb{R}^m} \log \frac{m(x-y_1)}{C} P_1(dy_1) \int_{\mathbb{R}^m} m(x-y_2) P_2(dy_2)\, dx > -\infty$$

and by Jensen's inequality,

$$\int_{\mathbb{R}^m} \left[\log \int_{\mathbb{R}^m} \frac{m(x-y)}{C} P_1(dy)\right] \int_{\mathbb{R}^m} m(x-y) P_2(dy)\, dx > -\infty.$$

Since

$$\int_{\mathbb{R}^m} m(x-y) P_1(dy) \le C,$$

the last bound implies that for any P_1, P_2 with compact supports we have

$$\int_{\mathbb{R}^m} \left|\log \int_{\mathbb{R}^m} m(x-y) P_1(dy)\right| \int_{\mathbb{R}^m} m(x-y) P_2(dy)\, dx < +\infty.$$

Therefore $I[P_1 * \mu | P_2 * \mu] < +\infty$.

Next we get

$$\inf_{\check{\tau}_n} \sup_{P \in \mathscr{P}} \mathbb{P}\{\check{\tau}_n \ne \tau(P)\} \ge \inf_{T_n} \alpha_n,$$

where the last inf is taken over all the tests of the hypothesis $P = P_1$ against the alternative $P = P_2$ and α_n denotes the probability of the error of the first kind. It is well known [see, e.g., Chencov (1972), page 138] that, for any $q < \exp\{-I[P_1 * \mu | P_2 * \mu]\}$ and for large enough n, $\inf_{T_n} \alpha_n \ge q^n$, which implies the statement. □

PROOF OF THEOREM 3.1. In what follows $\mathscr{P} := \mathscr{P}(\mathscr{D}; \Theta; C)$. Using Lemma 7.1, we get

$$\sup_{P \in \mathscr{P}} \mathbb{P}\left\{\sup_B |\hat{P}_{n,\Psi}(B) - P_\Psi(B)| \ge \gamma\right\} \le 4q^n,$$

where the supremum inside the probability is taken over the set of all balls of radius 2ε and

$$q := \exp\left\{-\frac{1}{32}\gamma^2/(c_m^2\beta_m^2(\Psi)\varepsilon^{2m})\right\}.$$

Consider the event

$$A := \left\{\max_{B\in\mathscr{N}(\varepsilon)} |\hat{P}_{n,\Psi}(B^+) - P_\Psi(B^+)| < \gamma\right\}.$$

Then

$$\sup_{P\in\mathscr{P}} \mathbb{P}(A^c) \le 4q^n. \tag{7.7}$$

Consider a measure $P \in \mathscr{P}$ and let $d := \dim(P)$. Given $\omega \in A$, we claim that, for all $B \in \hat{\mathscr{N}_n}$, we have $\operatorname{dist}(B^+, \operatorname{supp}(P)) \le \varepsilon$. Indeed, suppose that, on the contrary, there exists a $B \in \hat{\mathscr{N}_n}$ such that $\operatorname{dist}(B^+, \operatorname{supp}(P)) > \varepsilon$. Since $\hat{P}_{n,\Psi}(B^+) \ge 2\gamma$ and $\omega \in A$, we get $P_\Psi(B^+) > \gamma$. On the other hand, since $\operatorname{dist}(B^+, \operatorname{supp}(P)) > \varepsilon$, for all $x \in \operatorname{supp}(P)$, we have $B^+ - x \subset \{y\colon\ |y| \ge \varepsilon\}$, and, by the condition (3.5),

$$\Psi(B^+ - x) \le \Psi\{y\colon\ |y| \ge \varepsilon\} = \Psi\{y\colon\ |y| \ge \varepsilon\} \le \gamma.$$

Thus, we also have

$$P_\Psi(B^+) = \int_{\operatorname{supp}(P)} \Psi(B^+ - x)P(dx) \le \gamma.$$

This contradiction shows that $\operatorname{dist}(B^+, \operatorname{supp}(P)) \le \varepsilon$. By the condition (3.1), $\omega \in A$ implies that $\hat{N}_n \le \Theta\varepsilon^{-d}$, which yields $\log\ \hat{N}_n/\log(\varepsilon^{-1}) \le d + (\log\ \Theta/\log(\varepsilon^{-1}))$. Since we have $\varepsilon < \varepsilon(\mathscr{D};\Theta;C)$ with $\varepsilon(\mathscr{D};\Theta;C)$ defined by (3.4), it follows that $\log\ \Theta/\log(\varepsilon^{-1}) < \delta/2$, and we obtain that on the event A

$$\frac{\log\ \hat{N}_n}{\log(\varepsilon^{-1})} \le d + \delta/2. \tag{7.8}$$

Note now that if $\omega \in A$ and $B \notin \hat{\mathscr{N}_n}$, then $\hat{P}_{n,\Psi}(B^+) < 2\gamma$ and hence $P_\Psi(B^+) < 3\gamma$. On the other hand, since B is the ball of radius ε having the same center as B^+ has, we get

$$\begin{aligned} P_\Psi(B^+) &= (P * \Psi)(B^+) = (P \times \Psi)\{(x, y)\colon\ x + y \in B^+\} \\ &\ge (P \times \Psi)\{(x, y)\colon\ x \in B,\ |y| \le \varepsilon\} = P(B)\Psi\{y\colon\ |y| \le \varepsilon\}, \end{aligned}$$

which implies $P(B) \le 3(\Psi\{y\colon\ |y| \le \varepsilon\})^{-1}\gamma \le 3\gamma(1-\gamma)^{-1} \le 6\gamma$.

Note that the number of the balls from $\mathscr{N}$, covering $\operatorname{supp}(P)$, is less than $\Theta\varepsilon^{-d}$ [by the condition (3.1)]. Denote $\mathscr{N}_P$ such a covering set. Since

$$\operatorname{supp}(P) \subset \bigcup_{B\in\hat{\mathscr{N}_n}\cap\mathscr{N}_P} B \ \cup \bigcup_{B\in\mathscr{N}_P,\ B\notin\hat{\mathscr{N}_n}} B$$

and, for $\omega \in A$ and $B \notin \hat{\mathscr{N}}_n$, $P(B) \le 6\gamma$, we can write, using (3.2),

$$1 = P(\mathrm{supp}(P)) \le \sum_{B \in \hat{\mathscr{N}}_n \cap \mathscr{N}_P} P(B) + \sum_{B \in \mathscr{N}_P, B \notin \hat{\mathscr{N}}_n} P(B)$$
$$\le \hat{N}_n C \varepsilon^d + 6\gamma\Theta\varepsilon^{-d},$$

which implies

$$\hat{N}_n \ge \frac{1 - 6\gamma\Theta\varepsilon^{-d}}{C}\varepsilon^{-d}.$$

Therefore

$$\frac{\log\ \hat{N}_n}{\log(\varepsilon^{-1})} \ge d + \frac{\log\ C^{-1}(1 - 6\gamma\Theta\varepsilon^{-d})}{\log(\varepsilon^{-1})}.$$

Under the conditions $\varepsilon < \varepsilon(\mathscr{D};\Theta;C)$ and $\gamma < \gamma(\mathscr{D};\varepsilon;C)$ with $\varepsilon(\mathscr{D};\Theta;C)$ and $\gamma(\mathscr{D};\varepsilon;C)$ defined by (3.4), we get on the event A $(\log\ \hat{N}_n/\log(\varepsilon^{-1})) \ge d - \delta/2$, which, in view of (7.7), (7.8), the definition of the estimate $\hat{d}_n$ and the fact that the set $\mathscr{D}$ is δ-separated, concludes the proof of (3.6). □

REMARK 7.3. Using Remark 7.2, we get

$$\sup_{P \in \mathscr{P}} \mathbb{P}\left\{ \sup_{B \in \mathscr{N}(\varepsilon)} \left| \hat{P}_{n,\Psi}(B^+) - P_\Psi(B^+) \right| \ge \gamma \right\} \le 2\ \mathrm{card}(\mathscr{N}(\varepsilon)) q^n,$$

where

$$q := \exp\left\{ -\frac{\gamma}{4c_m \varepsilon^m \|K\|_\infty} \left[\frac{\gamma}{c_m 2^m C_\mu \varepsilon^m \|K\|_{L_1}} \wedge 1 \right] \right\}.$$

This gives different values of Λ and q in (3.6).

PROOF OF THEOREM 3.2. By the definition of the sequences $\{\varepsilon_n\}$ and $\{\gamma_n\}$, we have $\varepsilon_n \to 0$, $\gamma_n \to 0$, $\gamma_n = o(\varepsilon_n^m)$ as $n \to \infty$. Thus, if n is large enough, the conditions $\varepsilon < \varepsilon(\mathscr{D};\Theta;C)$ and $\gamma < \gamma(\mathscr{D};\varepsilon;\Theta)$ hold for $\varepsilon := \varepsilon_n$ and $\gamma := \gamma_n$. The following easy bound for $\Psi := \Psi_{a,l} \times \cdots m \times \Psi_{a,l}$:

$$\Psi\{x\colon\ |x| \ge \varepsilon\} \le m\Psi_{a,l}\{u\colon\ |u| \ge \varepsilon/m^{1/2}\} \le ca^{-(l-1)} m^{(l+1)/2} \varepsilon^{-(l-1)}$$

implies that

$$\Psi_n\{x\colon\ |x| \ge \varepsilon_n\} = O\left(\varepsilon_n^{-1} a_n^{-1}\right) = O\left(\frac{\varepsilon_n^m}{(\log\ \varepsilon_n^{-1})^2} \right) = o(\gamma_n) \quad \text{as } n \to \infty.$$

Therefore, the condition (3.5) also holds for $\varepsilon := \varepsilon_n$, $\gamma := \gamma_n$ and $\Psi := \Psi_n$ with large enough n. By the proof of Theorem 3.1, we can conclude that

$$\sup_{P \in \mathscr{P}} \mathbb{P}\{\hat{d}_n \ne d\} \le 4\ \exp\left\{ -\frac{1}{32} \frac{\gamma_n^2 n}{c_m^2 \beta_m^2(\Psi_n) \varepsilon_n^{2m}} \right\}.$$

Note that $\tilde{\Psi}_{a,l} = 2\pi a c(l) U_{[-a,a]} * \cdots {}_l * U_{[-a,a]} =: 2\pi a c(l) U^{*l}_{[-a,a]}$, where $U_{[-a,a]}$ denotes the uniform density on $[-a, a]$ [see, e.g., Bhattacharya and Ranga Rao (1976)]. Thus, $\tilde{\Psi}_{a,l}$ is a finite function (supported in the interval $[-la, la]$). Then $\tilde{\Psi}_n(t) = \tilde{\Psi}_{a,l}(t_1) \cdots \tilde{\Psi}_{a,l}(t_m)$ is a finite function on $\mathbb{R}^m$. Using the formulas for Ψ_n and $\tilde{\Psi}_n$, it's easy to show that

$$\beta_m(\Psi_n) = O\left(\frac{a_n^m}{\rho(a_n)}\right) \quad \text{as } n \to \infty$$

and the definitions of ε_n, a_n allows one to represent $\rho(a_n)$ as

$$\rho(a_n) := \sigma_n^{1/2} \varepsilon_n^{-m(m+1)} (\log(\varepsilon_n^{-1}))^{2m+1}.$$

Combining the above results, one can get by a simple calculation that

$$\sup_{P \in \mathscr{P}} \mathbb{P}\{\hat{d}_n \neq d\} \leq 4 \ \exp\{-\theta n \delta_n\}$$

with some constants $\theta > 0$, which implies that $\sup_{P \in \mathscr{P}} \mathbb{P}\{\hat{d}_n \neq d\} = O(q_n^n)$ as $n \to \infty$. □

PROOF OF THEOREM 3.3. To prove (i), note that, under the condition that the set $\mathscr{D}$ is φ-rich, there exist $P_1, P_2 \in \mathscr{P}_n(\mathscr{D})$ with disjoint supports and such that $\dim(P_2) = \dim(P_1) + \delta$, where $\varphi(\beta/\log\ n) \leq \delta \leq \beta/\log\ n$. Denote $d_1 := \dim(P_1)$, $d_2 := \dim(P_2)$. We can also assume that for all $\varepsilon \geq \varepsilon_n$, $P_1(B) \leq (C/2)\varepsilon^{d_1}$ and $P_2(B) \leq (C/2)\varepsilon^{d_2}$ [since for large enough constants such classes of measures as $\mathscr{P}(\mathscr{D}; \Theta; C/2)$ are nonempty]. Consider a probability measure $P_0 := (1-n^{-1})P_1 + n^{-1}P_2$. Obviously, P_0 is supported in $B(0; 1)$ and $\dim(P_0) = d_2$. We also have for all the balls with radius $\varepsilon \geq \varepsilon_n$,

$$P_1(B) \leq (C/2)\varepsilon^{d_1} = (C/2)\varepsilon^{-\delta}\varepsilon^{d_2}$$

and, since

$$\varepsilon^{-\delta} = \exp\{\delta\ \log(1/\varepsilon)\} \leq \exp\{\beta(\log\ n)^{-1}\ \log(1/\varepsilon_n)\} = \exp\{\alpha\beta\},$$

we get (under the assumption that $\alpha\beta < \log\ 2$) that $P_1(B) \leq C\varepsilon^{d_2}$, and, hence, $P_0(B) \leq C\varepsilon^{d_2}$. This means that $P_0 \in \mathscr{P}_n(\mathscr{D})$. For any estimate $\check{d}_n$, based on the sample $(Y_1, \ldots, Y_n)$, we have

$$\sup_{P \in \mathscr{P}_n(\mathscr{D})} \mathbb{P}_P\left\{\left|\check{d}_n - \dim(P)\right| \geq \frac{1}{2}\varphi\left(\frac{\beta}{\log\ n}\right)\right\}$$

$$\geq \max_{j=0,1} \mathbb{P}_{P_j}\left\{\left|\check{d}_n - \dim(P)\right| \geq \frac{1}{2}\varphi\left(\frac{\beta}{\log\ n}\right)\right\}.$$

A sample $(X_1, \ldots, X_n)$ from the distribution P_0 can be represented as

$$X_j := \begin{cases} X'_j, & \text{if } \theta_j = 1, \\ X''_j, & \text{if } \theta_j = 0, \end{cases} \qquad j = 1, \ldots, n,$$

where the random variables X'_j, X''_j, θ_j, $j = 1, \ldots, n$ are independent, X'_j has the distribution P_1, X''_j has the distribution P_2 and θ_j takes value 1 with probability $1 - n^{-1}$ and 0 with probability n^{-1}. Suppose that for some p with $0 < p < 1$,

$$\mathbb{P}_{P_1}\left\{\left|\check{d}_n - d_1\right| \geq \frac{1}{2}\varphi\left(\frac{\beta}{\log\ n}\right)\right\} < p.$$

Then, since $d_2 - d_1 \geq \varphi(\beta/\log\ n)$, we get for all $n \geq 6$,

$$\begin{aligned}
&\mathbb{P}_{P_0}\left\{\left|\check{d}_n - d_2\right| \geq \frac{1}{2}\varphi\left(\frac{\beta}{\log\ n}\right)\right\}\\
&\quad\geq \mathbb{P}_{P_0}\left\{\left|\check{d}_n - d_2\right| \geq \frac{1}{2}\varphi\left(\frac{\beta}{\log\ n}\right) | \theta_1 = 1, \dots, \theta_n = 1\right\}\\
&\qquad\times \mathbb{P}_{P_0}\{\theta_1 = 1, \dots, \theta_n = 1\}\\
&\quad\geq \mathbb{P}_{P_1}\left\{\left|\check{d}_n - d_2\right| \geq \frac{1}{2}\varphi\left(\frac{\beta}{\log\ n}\right)\right\}(1 - n^{-1})^n\\
&\quad\geq \frac{1}{3}\mathbb{P}_{P_1}\left\{\left|\check{d}_n - d_1\right| \leq \frac{1}{2}\varphi\left(\frac{\beta}{\log\ n}\right)\right\} \geq (1 - p)/3.
\end{aligned}$$

Thus

$$\max_{j=0,1}\mathbb{P}_{P_j}\left\{\left|\check{d}_n - \dim(P)\right| \geq \frac{1}{2}\varphi\left(\frac{\beta}{\log\ n}\right)\right\} \geq \min(p, (1 - p)/3),$$

which implies

$$\sup_{P \in \mathscr{P}_n(\mathscr{D})}\mathbb{P}_P\left\{\left|\check{d}_n - \dim(P)\right| \geq \frac{1}{2}\varphi\left(\frac{\beta}{\log\ n}\right)\right\} \geq 1/4,$$

and (i) follows.

To prove (ii), note that for our choice of $\Psi = \Psi_n$ and under the assumption (3.10), we have, by a simple computation, $\beta_m(\Psi_n) = O(a_n^{m+A}) = O(n^{\sigma(m+A)})$ as $n \to \infty$.

If $\alpha > 0$ satisfies the condition $\alpha < \alpha(D; m; A)$, where $\alpha(D; m; A)$ is defined by (3.12), and $\sigma > \alpha(D + 1)$, $\sigma(m + 1) - \alpha(m - D) < 1/2$, then it is easy to check that

$$\beta_m(\Psi_n) = o(n^{1/2}\varepsilon_n^{D-m}(\log\ n)^{-1}).$$

Using the bound (7.7), we get for the set

$$A := A_n := \left\{\max_{B \in \mathscr{N}(\varepsilon_n)}\left|\hat{P}_{n,\Psi}(B) - P_\Psi(B)\right| < \varepsilon_n^D/(\log\ n)^{1/2}\right\},$$

that $\sup_{P \in \mathscr{P}_n(\mathscr{D})}\mathbb{P}(A_n^c) = o(n^{-\Delta})$ as $n \to \infty$ for all $\Delta > 0$. Let $P \in \mathscr{P}_n(\mathscr{D})$ and $d := \dim(P)$. On the event A_n, we still have the bounds $\log\ \hat{N}_n/\log(\varepsilon_n^{-1}) \leq d + (\log\ \Theta/\log(\varepsilon_n^{-1}))$ and

$$\frac{\log\ \hat{N}_n}{\log(\varepsilon_n^{-1})} \geq d + \frac{\log\ C^{-1}(1 - 6\Theta\varepsilon_n^{D-d}/(\log\ n)^{1/2})}{\log(\varepsilon_n^{-1})}.$$

Since $\log(\varepsilon_n^{-1}) = \alpha \ \log \ n$, these bounds imply that for all large n,

$$\left|\frac{\log \ \hat{N}_n}{\log(\varepsilon_n^{-1})} - d\right| \leq \frac{1}{3}\frac{\beta}{\log \ n}.$$

Since the set $\mathscr{D}$ is φ-poor, the definition of the estimator $\hat{d}_n$ implies that on the event A_n,

$$|\hat{d}_n - d| \leq \varphi\left(\frac{\beta}{\log \ n}\right),$$

and (3.11) follows, proving (ii). □

Proof of Theorem 4.1. We use the notation

$$\hat{\mathscr{R}}_n := \hat{\mathscr{R}}_n(\varepsilon) := \left\{B \in \mathscr{N}(\varepsilon): \ \hat{P}_n(B) \geq 2\gamma\right\},$$

$$\hat{\mathscr{T}}_n := \hat{\mathscr{T}}_n(\varepsilon) = \left\{B \in \mathscr{N}(\varepsilon): \ \exists \ B' \in \hat{\mathscr{R}}_n: \ \mathrm{dist}(B, B') \leq 2\varepsilon\right\}$$

and

$$\hat{\mathscr{V}}_n := \bigcup\{\bar{B}: \ B \in \hat{\mathscr{R}}_n\}.$$

Clearly,

$$\hat{\mathscr{S}}_n = \bigcup\{\bar{B}: \ B \in \hat{\mathscr{T}}_n\}.$$

As in the proof of Theorem 3.1, consider the event $A := \left\{\max_{B\in\mathscr{N}(\varepsilon)} |\hat{P}_n(B) - P_\Psi(B)| < \gamma\right\}$. Then we have, similarly to (7.7),

$$\sup_{P\in\mathscr{P}^{\tau,\Delta}} \mathbb{P}(A^c) \leq 4q^n \tag{7.9}$$

with

$$q := \exp\left\{-\gamma^2/(32c_m^2 2^{-2m}\beta_m^2(\Psi)\varepsilon^{2m})\right\}. \tag{7.10}$$

Let us consider a measure $P \in \mathscr{P}^{\tau,\Delta}$.

Claim 1. *For $\omega \in A$ and for a ball $B \in \mathscr{N}$, the condition* $\mathrm{dist}(B;\ \mathrm{supp}(P)) \geq \varepsilon$ *implies $B \notin \hat{\mathscr{R}}_n$. Moreover, the condition* $\mathrm{dist}(B; \mathrm{supp}(P)) \geq 5\varepsilon$ *implies $B \notin \hat{\mathscr{T}}_n$.*

Indeed, consider $B \in \mathscr{N}$ such that $\mathrm{dist}(B; \mathrm{supp}(P)) \geq \varepsilon$. Since for $x \in \mathrm{supp}(P)$, $B - x \subset \{y: \ |y| \geq \varepsilon\}$, we get [using (4.2)]

$$P_\Psi(B) = \int_{\mathrm{supp}(P)} \Psi(B - x)P(dx) \leq \Psi(\{y : |y| \geq \varepsilon\}) \leq \gamma. \tag{7.11}$$

If $\omega \in A$, this implies that $\hat{P}_n(B) \leq 2\gamma$, which means that $B \notin \hat{\mathscr{R}}_n$. Quite similarly, if $B \in \mathscr{N}$ is such that $\mathrm{dist}(B; \mathrm{supp}(P)) \geq 5\varepsilon$, then, for $\omega \in A$, $B \notin \hat{\mathscr{T}}_n$ [otherwise, there would exist $B' \in \hat{\mathscr{R}}_n$ with $\mathrm{dist}(B'; \mathrm{supp}(P)) \geq \varepsilon$, which has been proved to be wrong].

CLAIM 2. *For* $\omega \in A$ *and for a cluster* $C \subset \operatorname{supp}(P)$, *the conditions* $B \in \mathcal{N}$ *and* $B^{--} \cap C \neq \emptyset$ *imply* $B \in \hat{\mathcal{R}}_n$. *As a consequence, we have* $C \subset \hat{\mathcal{V}}_n$.

Indeed, if, for some cluster $C \subset \operatorname{supp}(P)$ and for a ball $B \in \mathcal{N}$, we have $B^{--} \cap C \neq \emptyset$, then, according to the definition of the class $\mathcal{P}^{\tau,\Delta}$, we have $P(B^-) \geq \tau(\varepsilon/2) > 6\gamma$. For all $x \in B^-$, $B - x \supset B(x;\varepsilon/2) - x = \{y : |y| \leq \varepsilon/2\}$, and we get [again using (4.2)]

$$
\begin{aligned}
P_\Psi(B) &= \int_{\operatorname{supp}(P)} \Psi(B-x)P(dx) \geq \int_{B^-} \Psi(B-x)P(dx) \\
&\geq \Psi\Big(\{y\colon\ |y| \leq \varepsilon/2\}\Big)P(B^-) \geq 3\gamma.
\end{aligned}
\tag{7.12}
$$

Thus, if $\omega \in A$ and $B^{--} \cap C \neq \emptyset$, then $\hat{P}_n(B) \geq 2\gamma$, which means that $B \in \hat{\mathcal{R}}_n$. Since the balls from $\mathcal{N}^{--}$ cover the set C, we get $C \subset \hat{\mathcal{V}}_n$.

CLAIM 3. *On the event* A, *the random sets*

$$\hat{D}_j^{(n)} := \bigcup\{\bar{B}\colon\ B \in \hat{\mathcal{T}}_n,\ \operatorname{dist}(B;C_j) \leq 5\varepsilon\}, \qquad j = 1,\dots,\nu$$

are connected. Moreover, they are disjoint and $\bigcup_{j=1}^{\nu} \hat{D}_j^{(n)} = \hat{\mathcal{S}}_n$.

Indeed, if $\omega \in A$, then $\hat{D}_j^{(n)} \supset C_j$, $j = 1,\dots,\nu$ and the sets $\hat{D}_j^{(n)}$, $j = 1,\dots,\nu$ are disjoint [by Claims 1, 2 and the condition $\varepsilon < \Delta/14$ along with the definition of the class $\mathcal{P}^{\tau,\Delta}$ (see the condition $\delta(P) > \Delta$)]. Thus, it is enough to show that these sets are connected. Given $C = C_j$ for some $j = 1,\dots,\nu$, consider the set $\mathcal{N}_C$ of all balls $B \in \mathcal{N}$ such that there exists a ball $B' \in \mathcal{N}$, $(B')^{--} \cap C \neq \emptyset$ and $\operatorname{dist}(B;B') \leq 2\varepsilon$. On the event A, any of these balls belongs to $\hat{\mathcal{T}}_n$, by Claim 2. Moreover, their union is a connected set. Indeed, let $B_1, B_2 \in \mathcal{N}_C$. Then there exist $B_1', B_2' \in \mathcal{N}$, $(B_1')^{--} \cap C \neq \emptyset$, $(B_2')^{--} \cap C \neq \emptyset$ and $\operatorname{dist}(B_1, B_1') \leq 2\varepsilon$, $\operatorname{dist}(B_2, B_2') \leq 2\varepsilon$. Since C is connected, there exist points $x_1 \in B_1' \cap C$, $x_2 \in B_2' \cap C$ and a continuous curve in C between these points. The curve is covered by a set $\mathcal{B}^{--}$ of balls from $\mathcal{N}^{--}$. Let $\mathcal{B} := \{B\colon B^{--} \in \mathcal{B}^{--}\}$. Then, obviously, $\mathcal{B} \subset \mathcal{N}_C$. Now consider the balls from $\mathcal{N}$, covering the straight line between the centers of B_1 and B_1'. All of them are within the distance 2ε from B_1', so they are also in $\mathcal{N}_C$. A similar remark applies to the balls from $\mathcal{N}$, covering the straight line between B_2 and B_2'. Thus, we constructed a "connected chain" of balls from $\mathcal{N}_C$ between any two balls in this set, proving that their union is connected. Denote the union $\tilde{C}$ and let $\hat{C}$ be the connected component of $\hat{D}_j^{(n)}$, which contains $\tilde{C}$. Let B be any ball such that $B \in \hat{\mathcal{T}}_n$ and $B \subset \hat{D}_j^{(n)}$. By the definition of $\hat{\mathcal{T}}_n$, there exists a ball $B' \in \hat{\mathcal{R}}_n$ such that $\operatorname{dist}(B;B') \leq 2\varepsilon$. The straight line, connecting the centers of the balls B and B' is covered by balls from $\mathcal{N}$. Clearly, the distance from each of these balls to B' is less than or equal to 2ε. Thus, all these balls must be in the set $\hat{\mathcal{T}}_n$. Moreover, each of them belongs to $\hat{D}_j^{(n)}$ (since they are within the distance 5ε from $C = C_j$). As to the ball B', we have (by Claim 1)

$\operatorname{dist}(B', C) \le \varepsilon$. It follows that there exists a point $x \in C$ within the distance ε from the ball B'. This point is covered by a ball $(B'')^{--} \in \mathscr{N}^{--}$, which means that $\operatorname{dist}(B'; B'') \le 2\varepsilon$ and at the same time $(B'')^{--} \cap C \ne \emptyset$. Therefore, $B' \in \mathscr{N}_C$. Thus, there exists a chain of the balls from $\hat{\mathscr{T}}_n$, connecting the ball B with the set $\hat{C}$. This implies that $\hat{D}_j^{(n)}$ is a connected set.

It follows from Claim 3 that $\hat{\nu}_n = \nu$ on the event A, and there exists a permutation π of the set of the numbers $1, 2, \dots, \nu$ such that $\hat{C}^{(n)}_{\pi(j)} = \hat{D}_j^{(n)}$. It follows from Claim 1 that, for all $\omega \in A$ and for all $B \in \hat{\mathscr{T}}_n$, $\operatorname{dist}(B; \operatorname{supp}(P)) \le 5\varepsilon$, which implies that $\hat{\mathscr{S}}_n \subset [(\operatorname{supp}(P))^{7\varepsilon}]$. On the other hand, by the Claim 2, for all clusters $C \subset \operatorname{supp}(P)$, we have $C \subset \hat{\mathscr{S}}_n$, which implies that, for $\omega \in A$, $\operatorname{supp}(P) \subset \hat{\mathscr{S}}_n$. Therefore, on the event A, $h(\hat{\mathscr{S}}_n; \operatorname{supp}(P)) \le 7\varepsilon$.

Next we use the following lemma, which can be easily proved by induction, using the recursive definition of the cluster tree.

LEMMA 7.4. *Suppose that $D, D' \subset \mathbb{R}^m$ are compact sets and*

$$D = \bigcup_{j=1}^{\nu} D_j, \qquad D' = \bigcup_{j=1}^{\nu} D'_j,$$

where D_j, D'_j, $1 \le j \le \nu$ are connected compact sets and $D_i \cap D_j = D'_i \cap D'_j = \emptyset$ for all $i \ne j$. Suppose also that

$$h(D; D') < \varepsilon < (1/2) \min_{1 \le j \le J} (\delta_j - \delta_{j-1}),$$

where

$$\delta_j := \delta_j(\mathscr{D}), \qquad j = 1, \dots, J, \qquad J := J(\mathscr{D}), \qquad \mathscr{D} := \{D_1, \dots, D_\nu\}$$

are the numbers recursively defined in Section 4. Then $(\mathscr{G}(D), \chi) \cong (\mathscr{G}_\varepsilon(D'), \chi_\varepsilon)$.

Under the conditions $P \in \mathscr{P}^{\tau, \Delta}$ and $\varepsilon < \Delta/14$, one can apply Lemma 7.1 to the sets $D_j := C_j$ and $D'_j = \hat{C}^{(n)}_{\pi(j)}$, which gives the isomorphism of the trees $(\hat{\mathscr{G}}_n, \hat{\chi}_n)$ and $(\mathscr{G}, \chi)$ on the event A. Now the bound (7.9) yields (4.3). □

REMARK 7.5. Other values of constants Λ, q in Theorem 4.1 can be obtained using Remark 7.2. Namely, we get

$$\sup_{P \in \mathscr{P}} \mathbb{P}\Big\{ \sup_{B \in \mathscr{N}(\varepsilon)} |\hat{P}_{n, \Psi}(B) - P_\Psi(B)| \ge \gamma \Big\} \le 2 \operatorname{card}(\mathscr{N}(\varepsilon)) q^n,$$

where

$$q := \exp\left\{ -\frac{\gamma}{4 c_m 2^{-m} \varepsilon^m \|K\|_\infty} \left[\frac{\gamma}{c_m C_\mu \varepsilon^m \|K\|_{L_1}} \wedge 1 \right] \right\}.$$

The proof of Theorem 4.2 is similar to the proof of Theorem 3.2 above. The proof of Corollary 4.3 is obvious.

PROOF OF THEOREM 4.4. Consider a measure $P \in \mathscr{P}^{\tau}_{\Delta,\delta}$. We use the following notations:

$$\partial_{-}\hat{\mathscr{R}}_n := \big\{B \in \mathscr{N}:\ \exists\ B_1, B_2, B_1 \in \hat{\mathscr{R}}_n,\ B_2 \notin \hat{\mathscr{R}}_n,\ \operatorname{dist}(B, B_1) \le 3\varepsilon,\ \operatorname{dist}(B, B_2) \le 3\varepsilon\big\},$$

$$\partial\hat{\mathscr{R}}_n := \big\{B \in \mathscr{N}:\ \exists\ B' \in \partial_{-}\hat{\mathscr{R}}_n:\ \operatorname{dist}(B, B') \le 6\varepsilon\big\}.$$

Then $\hat{\partial}_n(\hat{P}_n) := \bigcup\{\bar{B}:\ B \in \partial\hat{\mathscr{R}}_n\}$. Note that, on the event A (see the notations of the proof of Theorem 4.1), we have

$$\{B \in \mathscr{N}:\ B^{--} \cap \partial(\operatorname{supp}(P)) \ne \varnothing\} \subset \partial_{-}\hat{\mathscr{R}}_n. \tag{7.13}$$

Indeed, the condition $B^{--} \cap \partial(\operatorname{supp}(P)) \ne \varnothing$ implies, by Claim 2 in the proof of Theorem 4.1, that $B \in \hat{\mathscr{R}}_n$. On the other hand, property 4 in the definition of the class $\mathscr{P}^{\tau}_{\Delta,\delta}$ implies that there exists a ball $B' \in \mathscr{N}$ such that $\operatorname{dist}(B, B') \le 3\varepsilon$ and $\operatorname{dist}(B'; \operatorname{supp}(P)) \ge \varepsilon$ [indeed, first one can find a ball $\tilde{B}$ of radius 2ε such that for a point $x \in B^{--} \cap \partial(\operatorname{supp}(P))$ we have $\operatorname{dist}(x, \tilde{B}) \le 2\varepsilon$ and $\operatorname{dist}(\tilde{B}, \operatorname{supp}(P)) > \varepsilon$; then one can find a ball $B' \in \mathscr{N}$ covering the center of $\tilde{B}$; B' possesses the required properties]. By Claim 1 in proof of Theorem 4.1, we conclude that, on the event A, $B' \notin \hat{\mathscr{R}}_n$. Thus, on the event A, $B \in \partial_{-}\hat{\mathscr{R}}_n$, which implies (7.13).

Let $\Gamma := \Gamma_j$ be a cluster of $\operatorname{supp}(P)$. Since Γ is connected, the set $\bigcup\{\bar{B}:\ B^{--} \cap \Gamma \ne \varnothing\}$ is also connected. Also, note that if $\operatorname{dist}(B; \partial(\operatorname{supp}(P))) > 6\varepsilon$, then, on the event A, the ball $B \notin \partial_{-}\hat{\mathscr{R}}_n$. It follows again from Claims 1 and 2 of the proof of Theorem 4.1. Indeed, for all the balls $B' \in \mathscr{N}$, such that $\operatorname{dist}(B, B') \le 3\varepsilon$, we either have $\operatorname{dist}(B'; \operatorname{supp}(P)) \ge \varepsilon$, or $B' \subset \operatorname{supp}(P)$. Given that the event A occurs, we have in the first case, by Claim 1, that all the balls $B' \notin \hat{\mathscr{R}}_n$, while in the second case, by Claim 2, all the balls $B' \in \hat{\mathscr{R}}_n$. In both cases, we get $B \notin \partial_{-}\hat{\mathscr{R}}_n$. By the definition of the set $\partial\hat{\mathscr{R}}_n$, it immediately follows that if $\operatorname{dist}(B; \partial(\operatorname{supp}(P))) > 14\varepsilon$, then, on the event A, $B \notin \partial\hat{\mathscr{R}}_n$.

Suppose now that $B \in \partial\hat{\mathscr{R}}_n$. We claim that, on the event A, there exists a connected chain of the balls from $\partial\hat{\mathscr{R}}_n$, starting with B and ending with a ball B' such that $(B')^{--} \cap \operatorname{supp}(P) \ne \varnothing$. Indeed, there exists a ball $B_1 \in \partial_{-}\hat{\mathscr{R}}_n$ such that $\operatorname{dist}(B, B_1) \le 6\varepsilon$. By the definition of $\partial\hat{\mathscr{R}}_n$, all the balls from $\mathscr{N}$, covering the straight line between the centers of B and B_1 are still in $\partial\hat{\mathscr{R}}_n$. Since $B_1 \in \partial_{-}\hat{\mathscr{R}}_n$, we have $\operatorname{dist}(B_1, \partial(\operatorname{supp}(P))) \le 6\varepsilon$, which implies that there exists a ball $B' \in \mathscr{N}$ such that $(B')^{--} \bigcap \operatorname{supp}(P) \ne \varnothing$ and $\operatorname{dist}(B', B_1) \le 6\varepsilon$. Clearly, $B' \in \partial_{-}\hat{\mathscr{R}}_n$, and all the balls from $\mathscr{N}$, covering the straight line between the centers of B' and B_1 are in the set $\partial\hat{\mathscr{R}}_n$.

Finally, we define

$$\hat{\Gamma}^{(n)}_j := \bigcup\big\{\bar{B}:\ B \in \partial\hat{\mathscr{R}}_n,\ \operatorname{dist}(B; \Gamma_j) \le 14\varepsilon\big\}, \qquad j = 1, \dots, v(P).$$

By condition (iii) of the definition of the class $\mathscr{P}^{\tau}_{\Delta,\delta}$, the minimal distance between the sets Γ_j, $j = 1, \dots, v(P)$ is at least 40ε. We also know that any ball $B \in \mathscr{N}$, such that $\operatorname{dist}(B;\ \operatorname{supp}(P)) \ge 14\varepsilon$, does not belong to $\partial\hat{\mathscr{R}}_n$. It follows that the sets $\hat{\Gamma}_j^{(n)}$, $j = 1, \ldots, v(P)$ are disjoint. Also, each ball from $\hat{\Gamma}_j^{(n)}$ can be connected by a chain of balls from $\partial\hat{\mathscr{R}}_n$ to a ball B such that $B^{--} \bigcap \Gamma_j \neq \emptyset$. It follows that the sets $\hat{\Gamma}_j^{(n)}$, $j = 1, \ldots, v(P)$ are connected, and since their union is equal to the set $\partial\hat{\mathscr{R}}_n$, they are the clusters of this set. It follows that, on the event A, $\hat{v}_n = v(P)$. Similarly to (7.9), we get the bound $\sup_{P \in \mathscr{P}^{\tau}_{\Delta,\delta}} \mathbb{P}(A^c) \leq 4q^n$, which immediately implies the proof. □

PROOF OF THEOREM 5.1. Denote

$$A := \left\{ \hat{\mu} \in \Lambda_\mu,\ \sup_{B \in \mathscr{N}(\varepsilon)} \left| n^{-1} \sum_{j=1}^n \mathscr{K}(B^+ - Y_j) - (P * \Psi * T_\mu \hat{\mu})(B^+) \right| \leq \gamma, \right.$$
$$\left. (T_\mu \hat{\mu})\{x\colon\ |x| \geq \varepsilon/2\} \leq \gamma/2 \right\}.$$

Then, by Lemma 7.1 and the condition (5.2), $\sup_{P \in \mathscr{P}} \mathbb{P}(A^c) \leq \Lambda q^n$ with some $\Lambda > 0$, $q \in (0, 1)$. Next we follow the lines of the proof of Theorem 3.1. Given $\omega \in A$, we claim that, for all $B \in \hat{\mathscr{N}}_n$, we have $\operatorname{dist}(B^+, \operatorname{supp}(P)) \leq \varepsilon$. To this end, it is enough to show that $\hat{P}(B^+) := (P * \Psi * T_\mu \hat{\mu})(B^+) \leq \gamma$, which (see the proof of Theorem 3.1) would follow from the bound

$$(\Psi * T_\mu \hat{\mu})(B^+ - x) \leq (\Psi * T_\mu \hat{\mu})\{y\colon\ |y| \geq \varepsilon\} \leq \gamma. \tag{7.14}$$

To establish (7.14), note that

$$\begin{aligned}(\Psi * T_\mu \hat{\mu})\{y\colon\ |y| \geq \varepsilon\} &= (\Psi \times T_\mu \hat{\mu})\{(y, z)\colon\ |y + z| \geq \varepsilon\} \\ &\leq \Psi\{y\colon\ |y| \geq \varepsilon/2\} + (T_\mu \hat{\mu})\{y\colon\ |y| \geq \varepsilon/2\},\end{aligned}$$

which is $\leq \gamma$ on the event A. Similarly to the proof of Theorem 3.1, this yields on the event A $(\log\ \hat{N}_n / \log(\varepsilon^{-1})) \leq d + \delta/2$.

It remains to show that if $\omega \in A$ and $B \notin \hat{\mathscr{N}}_n$, then $P(B) \leq 6\gamma$. This will imply (again, as in Theorem 3.1) that on the event A $(\log\ \hat{N}_n / \log(\varepsilon^{-1})) \geq d - \delta/2$ and the result will follow. We have $\hat{P}_n(B^+) < 2\gamma$, which yields $\hat{P}(B^+) < 3\gamma$. On the other hand, since B is the ball of radius ε having the same center as B^+ has, we get

$$\begin{aligned}\hat{P}(B^+) &= (P * \Psi * T_\mu \hat{\mu})(B^+) = (P \times \Psi \times T_\mu \hat{\mu})\{(x, y, z)\colon\ x + y + z \in B^+\} \\ &\geq 1(P \times \Psi \times T_\mu \hat{\mu})\{(x, y, z)\colon\ x \in B, |y| \leq \varepsilon/2,\ |z| \leq \varepsilon/2\} \\ &= P(B)\Psi\{y\colon\ |y| \leq \varepsilon/2\}(T_\mu \hat{\mu})(\{z\colon\ |z| \leq \varepsilon/2\}),\end{aligned}$$

which implies

$$P(B) \leq 3(\Psi\{y\colon\ |y| \leq \varepsilon/2\})^{-1}(T_\mu \hat{\mu})(\{z\colon\ |z| \leq \varepsilon/2\})^{-1}\gamma \leq 3\gamma(1 - \gamma/2)^{-2} \leq 6\gamma.$$

Now the proof can be completed exactly as in the case of Theorem 3.1. □

PROOF OF THEOREM 5.2. First, as in the proof of Theorem 5.1, we define

$$A := \left\{ \hat{\mu} \in \Lambda_\mu,\ \sup_{B \in \mathcal{N}(\varepsilon)} \left| n^{-1} \sum_{j=1}^{n} \mathcal{K}(B - Y_j) - (P * \Psi * T_\mu \hat{\mu})(B) \right| \le \gamma, \right.$$

$$\left. (T_\mu \hat{\mu})\{x\colon\ |x| \ge \varepsilon/4\} \le \gamma/2 \right\}$$

and show that $\sup_{P \in \mathcal{P}} \mathbb{P}(A^c) \le \Lambda q^n$ with some $\Lambda > 0,\ q \in (0, 1)$.

Next we follow the proof of Theorem 4.1. In order to prove Claim 1, it is enough to show that

$$\hat{P}(B) := (P * \Psi * T_\mu \hat{\mu})(B) \le \gamma. \tag{7.15}$$

Note that (under the assumptions of Claim 1), we have for all $x \in \operatorname{supp}(P)$,

$$\begin{aligned} (\Psi * T_\mu \hat{\mu})(B - x) &\le (\Psi * T_\mu \hat{\mu})(\{y\colon\ |y| \ge \varepsilon\}) \\ &= (\Psi \times T_\mu \hat{\mu})(\{(y, z)\colon\ |y + z| \ge \varepsilon\}) \\ &\le \Psi(\{y\colon\ |y| \ge \varepsilon/2\}) + (T_\mu \hat{\mu})(\{z\colon\ |z| \ge \varepsilon/2\}), \end{aligned}$$

which, under the condition (5.4) and on the event A, is less than or equal to γ. Since

$$\hat{P}(B) = \int_{\operatorname{supp}(P)} (\Psi * T_\mu \hat{\mu})(B - x) P(dx),$$

we get (7.15), and Claim 1 follows.

Similarly, to establish Claim 2, it's enough to show that $\hat{P}(B) \ge 3\gamma$. To this end, we write

$$\begin{aligned} \hat{P}(B) &= (P * \Psi * T_\mu \hat{\mu})(B) = (P \times \Psi \times T_\mu \hat{\mu})(\{(x, y, z)\colon\ x + y + z \in B\}) \\ &\ge P(B^-)\Psi(\{y\colon\ |y| \le \varepsilon/4\})(T_\mu \hat{\mu})(\{z\colon\ |z| \le \varepsilon/4\}), \end{aligned}$$

which, under the condition (5.4) and on the event A, is $\ge 3\gamma$. The rest of the proof of Theorem 4.1 goes through with no changes. □

The proof of Corollary 5.3 is rather straightforward. We proceed to the results of Section 6.

PROOF OF THEOREM 6.1. We have

$$\begin{aligned} &\left\| \log |\tilde{Q}(\cdot)| + \tfrac{1}{2}(\Sigma\cdot, \cdot) \right\|_{L_p(B(0;R);\lambda_R)} \\ &\quad = \left\| \log |\tilde{P}(\cdot)| + \tfrac{1}{2}((\Sigma - \Sigma_0)\cdot, \cdot) \right\|_{L_p(B(0;R);\lambda_R)} \\ &\quad \ge \tfrac{1}{2} \left\| ((\Sigma - \Sigma_0)\cdot, \cdot) \right\|_{L_p(B(0;R);\lambda_R)} - \left\| \log |\tilde{P}(\cdot)| \right\|_{L_p(B(0;R);\lambda_R)} \\ &\quad = \tfrac{1}{2} \left\| ((\Sigma - \Sigma_0)\cdot, \cdot) \right\|_{L_p(B(0;R);\lambda_R)} - \delta_p(P;R)R^2. \end{aligned}$$

A simple computation (using polar coordinates) shows that

$$\lambda(B(0;R))^{-1}\int_{B(0;R)}\left|((\Sigma-\Sigma_0)t,t)\right|^p\,dt$$
$$=\frac{m}{2p+m}R^{2p}\int_{S^{m-1}}\left|((\Sigma-\Sigma_0)v,v)\right|^pU(dv).$$

Therefore,

$$\left\|\log|\tilde{Q}(\cdot)|+\frac{1}{2}(\Sigma\cdot,\cdot)\right\|_{L_p(B(0;R);\lambda_R)}$$
$$\geq\frac{1}{2}\left(\frac{m}{2p+m}\right)^{1/p}\|\Sigma-\Sigma_0\|_pR^2-\delta_p(P;R)R^2,$$

where $\|A\|_p:=\|(A\cdot,\cdot)\|_{L_p(S^{m-1};dU)}$. It is easy to check that for a symmetric nonnegatively definite A $(1/c_p)\|A\|\leq\|A\|_p\leq\|A\|$ with some absolute constant $c_p>0$. Hence

$$\left\|\log|\tilde{Q}(\cdot)|+\frac{1}{2}(\Sigma\cdot,\cdot)\right\|_{L_p(B(0;R);\lambda_R)}$$
$$\geq\left[\frac{1}{2c_p}\left(\frac{m}{2p+m}\right)^{1/p}\|\Sigma-\Sigma_0\|-\delta_p(P;R)\right]R^2,$$

which implies

$$\|\Sigma-\Sigma_0\|\leq 2c_p\left(\frac{2p+m}{m}\right)^{1/p}$$
$$\times\left[\delta_p(P;R)+R^{-2}\left\|\log|\tilde{Q}(\cdot)|+\tfrac{1}{2}(\Sigma\cdot,\cdot)\right\|_{L_p(B(0;R);\lambda_R)}\right]. \tag{7.16}$$

Define

$$\Delta_n:=\left\|\log|\tilde{Q}_n(\cdot)|I_{\{|\log|\tilde{Q}_n(\cdot)||\leq 2MR^2\}}-\log|\tilde{Q}(\cdot)|\right\|_{L_p(B(0;R);\lambda_R)}.$$

Then

$$\left\|\log|\tilde{Q}(\cdot)|+\tfrac{1}{2}(\check{\Sigma}\cdot,\cdot)\right\|_{L_p(B(0;R);\lambda_R)}$$
$$\leq\left\|\log|\tilde{Q}_n(\cdot)|I_{\{|\log|\tilde{Q}_n(\cdot)||\leq 2MR^2\}}+\tfrac{1}{2}(\check{\Sigma}\cdot,\cdot)\right\|_{L_p(B(0;R);\lambda_R)}+\Delta_n.$$

Since $\check{\Sigma}$ minimizes the functional (6.5), we can further get

$$\begin{aligned}
&\Big\|\log|\tilde{Q}(\cdot)| + \tfrac{1}{2}(\check{\Sigma}\cdot,\cdot)\Big\|_{L_p(B(0;R);\lambda_R)} \\
&\quad\le \Big\|\log|\tilde{Q}_n(\cdot)| I_{\{|\log|\tilde{Q}_n(\cdot)||\le 2MR^2\}} + \tfrac{1}{2}(\Sigma_0\cdot,\cdot)\Big\|_{L_p(B(0;R);\lambda_R)} + \Delta_n \\
&\quad\le \Big\|\log|\tilde{Q}(\cdot)| + \tfrac{1}{2}(\Sigma_0\cdot,\cdot)\Big\|_{L_p(B(0;R);\lambda_R)} + 2\Delta_n \\
&\quad= \Big\|\log|\tilde{P}(\cdot)|\Big\|_{L_p(B(0;R);\lambda_R)} + 2\Delta_n \\
&\quad= R^2\delta_p(P;R) + 2\Delta_n.
\end{aligned} \tag{7.17}$$

It follows from (7.16) and (7.17) that

$$\|\check{\Sigma} - \Sigma_0\| \le 4\left(\frac{2p+m}{m}\right)^{1/p} c_p\big[\delta_p(P;R) + R^{-2}\Delta_n\big]. \tag{7.18}$$

Now we get an exponential bound for the random variable Δ_n. Define the event

$$\begin{aligned}
E := E(R) := \Big\{&\sup_{|t|\le R}|\tilde{Q}_n(t) - \tilde{Q}(t)| \le (\exp(-MR^2) - \exp(-2MR^2)) \\
&\wedge \tfrac{1}{2}\exp(-2MR^2) \wedge \tfrac{1}{2}R^2\delta_p(P;R)\exp(-2MR^2)\Big\}.
\end{aligned}$$

On the event E, the condition $|\tilde{Q}_n(t)| \le \exp(-2MR^2)$ implies that for $|t| \le R$ $|\tilde{Q}(t)| \le \exp(-MR^2)$. Therefore, we have

$$\begin{aligned}
&\Big(\lambda(B(0;R))^{-1}\int_{B(0;R)} |\log|\tilde{Q}(t)||^p I_{\{|\log|\tilde{Q}_n(t)||\ge 2MR^2\}}\,dt\Big)^{1/p} \\
&\quad\le \Big(\lambda(B(0;R))^{-1}\int_{B(0;R)} |\log|\tilde{Q}(t)||^p I_{\{|\log|\tilde{Q}(t)||\ge MR^2\}}\,dt\Big)^{1/p}.
\end{aligned}$$

Since $\log|\tilde{Q}(t)| = \log|\tilde{P}(t)| - \frac{1}{2}(\Sigma_0 t, t)$ and $\|\Sigma_0\| \le M$, the condition $|\log|\tilde{Q}(t)|| \ge MR^2$ implies $|\log|\tilde{P}(t)|| \ge MR^2/2$. Hence, we get

$$\begin{aligned}
&\Big(\lambda(B(0;R))^{-1}\int_{B(0;R)} |\log|\tilde{Q}(t)||^p I_{\{|\log|\tilde{Q}_n(t)||\ge 2MR^2\}}\,dt\Big)^{1/p} \\
&\quad\le \Big(\lambda(B(0;R))^{-1}\int_{B(0;R)} |\log|\tilde{P}(t)||^p I_{\{|\log|\tilde{P}(t)||\ge MR^2/2\}}\,dt\Big)^{1/p} \\
&\quad\quad+\Big(\lambda(B(0;R))^{-1}\int_{B(0;R)} \big|\tfrac{1}{2}(\Sigma_0 t, t)\big|^p I_{\{|\log|\tilde{P}(t)||\ge MR^2/2\}}\,dt\Big)^{1/p} \\
&\quad\le \delta_p(P;R)R^2 \\
&\quad\quad+\tfrac{1}{2}\|\Sigma_0\|\Big(\lambda(B(0;R))^{-1}\int_{B(0;R)} |t|^{2p} I_{\{|\log|\tilde{P}(t)||\ge MR^2/2\}}\,dt\Big)^{1/p}
\end{aligned} \tag{7.19}$$

$$\le \delta_p(P;R)R^2 + \frac{\|\Sigma_0\|}{2M}\Big(\lambda(B(0;R))^{-1}\int_{B(0;R)} |\log|\tilde{P}(t)||^p\, dt\Big)^{1/p}$$

$$\le \delta_p(P;R)\left[1 + \frac{\|\Sigma_0\|}{M}\right] \le 2\delta_p(P;R)R^2.$$

Also, on the event E, the condition $|\log|\tilde{Q}_n(t)|| \le 2MR^2$ implies that $|\tilde{Q}(t)| \ge \frac{1}{2}\exp(-2MR^2)$, and hence,

$$(7.20)\qquad \begin{aligned} &\Big|\log|\tilde{Q}_n(t)| I_{\{|\log|\tilde{Q}_n(t)||\le 2MR^2\}} - \log|\tilde{Q}(t)|\Big| \\ &\quad\le \frac{|\tilde{Q}_n(t) - \tilde{Q}(t)|}{|\tilde{Q}_n(t)| \wedge |\tilde{Q}(t)|} \le 2\exp(2MR^2)|\tilde{Q}_n(t) - \tilde{Q}(t)| \le R^2\delta_p(P;R). \end{aligned}$$

It follows from (7.19) and (7.20) that, on the event E, $\Delta_n \le 3R^2\delta_p(P;R)$, which implies

$$(7.21)\qquad \mathbb{P}\{\Delta_n \ge 3R^2\delta_p(P;R)\} \le \mathbb{P}(E(R)^c).$$

The bounds (7.18) and (7.21) imply that

$$(7.22)\qquad \mathbb{P}\left\{\|\check{\Sigma} - \Sigma_0\| \ge 16\left(\frac{2p+m}{m}\right)^{1/p} c_p\delta_p(P;R)\right\} \le \mathbb{P}(E(R)^c).$$

To bound the probability $\mathbb{P}(E(R)^c)$, one can use standard exponential bounds for empirical processes. We use basic definitions and notations of this theory [see, e.g., van der Vaart and Wellner (1996)]. Consider the classes of functions $\mathscr{F}_c := \{\cos(t,\cdot):\ |t| \le R\}$, $\mathscr{F}_s := \{\sin(t,\cdot):\ |t| \le R\}$. The bound $|\cos(t,y) - \cos(s,y)| \le |t-s||y|$ for all $t, s, y \in \mathbb{R}^m$ and the fact that the number of the balls of radius $\delta > 0$ needed to cover the ball $B(0;R)$ is bounded by $c_m R^m \delta^{-m}$ (with a constant c_m depending only on m) imply the following bound for the minimal number of balls of radius δ with respect to the random metric $d_{Q_n,1}$ of the space $L_1(\mathbb{R}^m; dQ_n)$, covering the class $\mathscr{F}_c$:

$$N_{d_{Q_n,1}}(\mathscr{F}_c;\delta) \le c_m\left(\frac{R}{\delta}\right)^m \left(n^{-1}\sum_{j=1}^n |Y_j|\right)^m.$$

Given a sequence $\{\varepsilon_n\}$ of i.i.d. Rademacher random variables (independent of $\{Y_n\}_{n\ge 1}$), this allows one to get the bound

$$\begin{aligned} \mathbb{P}_\varepsilon\left\{\Big\|\sum_{j=1}^n \varepsilon_j\delta_{Y_j}\Big\|_{\mathscr{F}_c} \ge 2n\delta\right\} &\le 2N_{d_{Q_n,1}}(\mathscr{F}_c;\delta)\exp\{-2n\delta^2\} \\ &\le 2c_m\left(\frac{R}{\delta}\right)^m\left(n^{-1}\sum_{j=1}^n |Y_j|\right)^m \exp\{-2n\delta^2\}, \end{aligned}$$

which implies

$$\mathbb{P}\left\{\left\|\sum_{j=1}^{n} \varepsilon_j \delta_{Y_j}\right\|_{\mathscr{F}_c} \geq 2n\delta\right\} \leq 2c_m\left(\frac{R}{\delta}\right)^m \mathbb{E}|Y|^m \ \exp\{-2n\delta^2\}. \tag{7.23}$$

Using standard symmetrization inequalities [see, e.g., van der Vaart and Wellner (1996), Lemma 2.3.7], we get from (7.23) for all $n > 1/(2\delta^2)$,

$$\mathbb{P}\left\{\left\|Q_n - Q\right\|_{\mathscr{F}_c} \geq 8\delta\right\} \leq 4c_m\left(\frac{R}{\delta}\right)^m \mathbb{E}|Y|^m \ \exp\{-2n\delta^2\}.$$

A similar bound holds for the class $\mathscr{F}_s$, which allows us to write

$$\mathbb{P}\left\{\sup_{t\in B(0;R)} |\tilde{Q}_n(t) - \tilde{Q}(t)| \geq 16\delta\right\} \leq 8c_m\left(\frac{R}{\delta}\right)^m \mathbb{E}|Y|^m \ \exp\{-2n\delta^2)\}. \tag{7.24}$$

The bound $\mathbb{P}(E(R)^c) \leq \Lambda q^n$ with the conditions (6.8) on Λ and q now follows by plugging in (7.24),

$$\begin{aligned}\delta := 16^{-1}\Big[&(\exp(-MR^2) - \exp(-2MR^2)) \\ &\wedge \tfrac{1}{2}\ \exp(-2MR^2) \wedge \tfrac{1}{2}\ R^2\delta_p(P;R)\exp(-2MR^2)\Big].\end{aligned}$$

In view of (7.22), the proof is complete. □

PROOF OF COROLLARY 6.2 AND PROPOSITION 6.3. It follows from (6.8) that with some constant $C > 0$ for all large n,

$$\begin{aligned}\Lambda q^n &= \exp\left\{-n\ \log\frac{1}{q} + \log\ \Lambda\right\} \leq \exp\{-n\ \exp(-CM_nR_n^2) + CM_nR_n^2\} \\ &\leq \exp\left\{-\frac{1}{2}n\ \exp(-CM_nR_n^2)\right\} = \exp\{-n\ \log(q_n^{-1})\} = o(q_n^n) \quad \text{as } n \to \infty.\end{aligned}$$

To prove Proposition 6.3, note that under the assumption

$$\|\check{\Sigma}_{n,R} - \Sigma_0\| \leq (\beta/4)\delta, \tag{7.25}$$

we have

$$\begin{aligned}((\Sigma_0 - \hat{\Sigma}_{n,R,\delta})v, v) &= ((\Sigma_0 - \check{\Sigma}_{n,R})v, v) + \delta(\Sigma_0 v, v) + \delta((\check{\Sigma}_{n,R} - \Sigma_0)v, v) \\ &\geq \beta\delta - (\beta/4)\delta\Big(1 + \delta\Big) \geq (\beta/2)\delta.\end{aligned} \tag{7.26}$$

Under condition (7.25), we also have that

$$\|\hat{\Sigma}_{n,R} - \Sigma_0\| \leq \Big(M + \beta/4\Big)\delta. \tag{7.27}$$

To see that (6.12) holds for $\hat{\mu} := \hat{\mu}_n := N(0; \hat{\Sigma}_{n,R})$, note that under the assumption (7.25) $\hat{\mu}_n \in \Lambda_\mu$ and $T_\mu \hat{\mu} = N(0; (\Sigma_0 - \hat{\Sigma}_{n,R}))$. Moreover, if (7.25) holds, then (7.27) implies that

$$\begin{aligned} |x| &\le \|(\Sigma_0 - \hat{\Sigma}_{n,R})^{1/2}\| \left|(\Sigma_0 - \Sigma_{n,R})^{-1/2} x\right| \\ &\le (M + \beta/4)^{1/2} \delta^{1/2} \left|(\Sigma_0 - \Sigma_{n,R})^{-1/2} x\right| \end{aligned}$$

and we have

$$\begin{aligned} (T_\mu \hat{\mu})\bigl\{x\colon\ |x| \ge \varepsilon\bigr\} &\le (T_\mu \hat{\mu})\left\{x\colon\ |(\Sigma_0 - \hat{\Sigma}_{n,R})^{-1/2} x| \ge \frac{\varepsilon}{(M + \beta/4)^{1/2} \delta^{1/2}}\right\} \\ &\le \exp\left\{-B \frac{\varepsilon^2}{(M + \beta/4)\delta}\right\}. \end{aligned}$$

Since δ satisfies the condition (6.11), we get $(T_\mu \hat{\mu})\bigl\{x\colon\ |x| \ge \varepsilon\bigr\} \le \gamma$, which, along with Theorem 6.1, is enough to check (6.12). □

The proofs of Proposition 6.4 and Theorem 6.5 easily follow from previous results.

Acknowledgment. I am thankful to the Associate Editor and to the referee for numerous comments and suggestions that improved the paper.

REFERENCES

BHATTACHARYA, R. N. and RANGA RAO, R. (1976). *Normal Approximation and Asymptotic Expansions*. Wiley, New York.

BOCK, H-H. (1996). Probability models and hypothesis testing in partitioning cluster analysis. In *Clustering and Classification* (P. Arabie, L. J. Hubert and G. De Soete, eds.) 377–453, World Scientific, River Edge, NJ.

BOLLOBÁS, B. (1979). *Graph Theory: An Introductory Course.* Springer, New York.

CARROLL, R. J. and HALL, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83** 1184–1186.

CHENCOV, N. N. (1972). *Statistical Decision Rules and Optimal Inference*. Nauka, Moscow. (In Russian).

CUEVAS, A. and FRAIMAN, R. (1997). A plug-in approach to support estimation. *Ann. Statist.* **25** 2300–2312.

DACUNHA-CASTELLE, D. and GASSIAT, E. (1997). The estimation of the order of a mixture model. *Bernoulli* **3** 279–299.

EDGAR, G. A. (1997). *Integral, Probability and Fractal Measures*. Springer, New York.

EFROMOVICH, S. (1997). Density estimation for the case of supersmooth measurement error. *J. Amer. Statist. Assoc.* **92** 526–535.

FAN, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** 1257–1272.

FAN, J. (1992). Deconvolution with supersmooth distributions. *Canad. J. Statist.* **20** 155–169.

FALCONER, K. (1997). *Techniques in Fractal Geometry*. Wiley, New York.

FUKUNAGA, K. (1990). *Introduction to Statistical Pattern Recognition*, 2nd ed. Academic Press, Boston.

GOLDFARB, L. (1985). A new approach to pattern recognition. In *Progress in Pattern Recognition* **2** (L. N. Kanal and A. Rosenfeld, eds.). North-Holland, Amsterdam.

GORDON, A. D. (1996). Hierarchical classification In *Clustering and Classification* (P. Arabie, L. J. Hubert and G. De Soete, eds.) 65–121. World Scientific, River Edge, NJ.

GRENANDER, U. (1981). *Lectures in Pattern Theory*, **3**. *Regular Structures*. Springer, New York.

HARTIGAN, J. A. (1975). *Clustering Algorithms*. Wiley, New York.

JAIN, A. K. and DUBES, R. C. (1988). *Algorithms for Clustering Data*. Prentice-Hall, Englewood, New Jersey.

JAMBU, M. (1978). Classification automatique pour l'analyse des données **1**. Dunod, Paris.

KOLMOGOROV, A. N. and TIKHOMIROV, V. M. (1959). ε-entropy and ε-capacity of the sets in functional spaces. *Uspehi Matem. Nauk* **14** 3–86.

KOROSTELEV, A. P. and TSYBAKOV, A. B. (1993). Minimax theory of image reconstruction. *Lecture Notes in Statist.* **82**. Springer, New York.

KOROSTELEV, A. P., SIMAR, L. and TSYBAKOV, A. B. (1995). Efficient estimation of monotone boundaries. *Ann. Statist.* **23** 476–489.

LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces.* Springer, New York.

MAMMEN, E. and TSYBAKOV, A. B. (1995). Asymptotical minmax recovery of sets with smooth boundaries. *Ann. Statist.* **23** 502–524.

POLLARD, D. (1982). A central limit theorems for k-means clustering. *Ann. Probab.* **10** 919–926.

POLONIK, W. (1995). Measuring mass concentrations and estimating density contour clusters: an excess mass approach. *Ann. Statist.* **23** 855–882.

STEFANSKI, L. and CARROLL, R. J. (1990). Deconvoluting kernel density estimators. *Statistics* **2** 169–184.

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes. With Applications to Statistics*. Springer, New York.

WYSE, N., DUBES, R. and JAIN, A. K. (1980). A critical evaluation of intrinsic dimensionality algorithms. In *Pattern Recognition in Practice* (E. S. Gelsema and L. N. Kanal, eds.). North-Holland, Amsterdam.

ZHANG, C. H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18** 806–831.

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF NEW MEXICO
ALBUQUERQUE, NEW MEXICO 87131-1141
E-MAIL: vlad@math.unm.edu