The Annals of Statistics
1997, Vol. 25, No. 2, 435–477

M-ESTIMATION, CONVEXITY AND QUANTILES[1]

BY V. I. KOLTCHINSKII

University of New Mexico

The paper develops a class of extensions of the univariate quantile function to the multivariate case (M-quantiles), related in a certain way to M-parameters of a probability distribution and their M-estimators. The spatial (geometric) quantiles, recently introduced by Koltchinskii and Dudley and by Chaudhuri as well as the regression quantiles of Koenker and Basset, are the examples of the M-quantile function discussed in the paper. We study the main properties of M-quantiles and develop the asymptotic theory of empirical M-quantiles. We use M-quantiles to extend L-parameters and L-estimators to the multivariate case; to introduce a bootstrap test for spherical symmetry of a multivariate distribution, and to extend the notion of regression quantiles to multiresponse linear regression models.

1. Introduction. The univariate quantile function has been extensively used in statistical inference. One of the reasons of such a wide applicability is that many important robust statistics (like, for instance, L- and R-estimators) are defined as functionals of the empirical quantile function. Our goal is to extend the notion of quantiles to a broader context, specifically, to probability measures on finite-dimensional spaces. Since there is no really natural ordering of vectors in $\mathbb{R}^d$ for $d > 1$, such an extension of quantiles is a hard problem in the statistics of multivariate data. A brief outline of some of the known approaches for the study of order statistics and quantiles in the multivariate case is as follows:

1. Tukey (1975), Barnett (1976), Eddy (1985) and Reiss (1989) suggested different ways of ordering multivariate data (by peeling the convex hulls of data points, using auxillary real-valued functions, etc.).
2. Pyke (1975, 1985) considered extensions of the empirical quantile process. Einmahl and Mason (1992) developed a class of extensions of the univariate quantile function to the multivariate case. Their quantile function $U_{\mathscr{C}}$ is defined as

$$U_{\mathscr{C}}(t) := \inf\{\lambda(C)\colon P(C) \geq t,\ C \in \mathscr{C}\} \quad \text{for } t \in (0,1),$$

where P is a probability measure on $\mathbb{R}^d$, $\mathscr{C} \subset \mathbb{R}^d$ is a class of Borel subsets and λ is a real-valued function defined on $\mathscr{C}$. They studied asymptotics of the corresponding empirical quantile processes.

Received January 1996; revised October 1996.

[1]Research partially supported by a Humboldt Research Fellowship while on leave to the University of Giessen and by a Sandia National Laboratories Grant AR-2011.

AMS 1991 *subject classifications*. Primary 60F05; secondary 62E20, 60F17.

Key words and phrases. Quantile function, spatial quantiles, empirical quantiles, regression quantiles, M-parameter, M-estimator, empirical processes.

3. Koltchinskii and Dudley (1996) ([K-D]) [see also Koltchinskii (1994a, b, 1996)], and Chaudhuri (1996) ([Ch]) suggested an extension of the distribution function F_P (of a probability measure P) to the multivariate case such that the inverse function F_P^{-1} can be viewed as an extension of the quantile function. These functions possess many properties known and used in the one-dimensional case. For instance, F_P characterizes P, F_P^{-1} is equivariant with respect to translations (as well as with respect to orthogonal transformations), and the asymptotic properties of their empirical versions are the same as in the univariate case [see, e.g., Shorack and Wellner (1986)]. The corresponding "median" $F_P^{-1}(0)$ is nothing but the well-known Haldane's spatial L_1-median [see Haldane (1948)].

Our approach in this paper is rather close to Koltchinskii (1996). It allows us, for example, to define L-parameters of the distribution P as integral functionals of the form $L(P) := \int h(F_P^{-1}(t))\,dm(t)$, and to define L-estimators as L-parameters of the empirical measure P_n, avoiding ordering of multivariate data [see Section 4 of this paper and Koltchinskii (1995a)]. A multivariate extension of R-estimators can be also defined in a similar way, as a certain functional of the empirical quantile function.

In the one-dimensional case, it is well known that the set of all medians of P coincides with the set of all minimal points of the function $\int(|s-x|-|x|)P(dx)$. However, it is less known that quantiles of P can be also characterized as minimal points of a certain function. For instance, let

$$f_P(s) := 0.5\int_{\mathbb{R}^1}(|s-x|-|x|+s)P(dx) \quad \text{for } s \in \mathbb{R}^1. \tag{1.1}$$

Then for any $t \in (0,1)$, the set of all tth quantiles of P is exactly the set of all minimal points of the function $f_{P,t}(s) := f_P(s) - st,\ s \in \mathbb{R}^1$. Note that the function f_P is convex, the distribution function F_P is a subgradient of f_P and the quantile function F_P^{-1} is a subgradient of the *Young–Fenchel conjugate* of f_P, defined as

$$f_P^*(t) := \sup_{s\in\mathbb{R}^1}[st - f_P(s)] \quad \text{for } t \in \mathbb{R}^1$$

and known in smooth analysis as the Legendre transformation of f_P. In what follows we extend this construction to the multivariate case and define *M-distribution* and *M-quantile functions* of a probability measure P on an arbitrary measurable space. These functions are related to the class of M-parameters of P, defined by convex minimization. The relationship is essentially the same as between the distribution and quantile functions and the median, so that the M-parameter can be viewed as a "median," whereas the corresponding M-estimator is a "sample median."

We use some notations, as introduced in the Appendix. Let $(\mathbf{X}, \mathscr{A}, P)$ be a probability space, and let $f\colon \mathbb{R}^d \times \mathbf{X} \mapsto \mathbb{R}^1$. Suppose that for any $s \in \mathbb{R}^d$, $f(s,\cdot)$ is an integrable function. We define the integral transformation $(P, s) \mapsto f_P(s) \in$

$\mathbb{R}^1$ of the probability measure P with the kernel f:

$$f_P(s) := \int_{\mathbf{X}} f(s, x)P(dx) \quad \text{for } s \in \mathbb{R}^d.$$

If there exists a point s_0, which minimizes f_P on $\mathbb{R}^d$, it is called an *M-parameter* of P with respect to the kernel f. Under some smoothness, the M-parameters are the solutions of the equation $F_P(s) = 0$, where $F_P(s) := \int_{\mathbf{X}} \nabla f(s; x)P(dx)$.

Given a sample $(X_1, \dots, X_n)$ from the distribution P, P_n denotes the empirical measure based on this sample; that is, $P_n := n^{-1}\sum_1^n \delta_{X_k}$, where δ_x is the unit point mass at $x \in \mathbf{X}$. An M-parameter of P_n is known as an *M-estimator*.

DEFINITION 1.1. We call a minimal point of the functional

$$f_{P,t}(s) := f_P(s) - \langle s, t\rangle, \qquad s \in \mathbb{R}^d$$

an (M, t)*-parameter* of P with respect to f. An (M, t)-parameter of the empirical measure P_n will be called an (M, t)*-estimator* (with respect to f) based on a sample $(X_1, \dots, X_n)$.

Of course, M-parameters of P coincide with $(M, 0)$-parameters, and M-estimators are just $(M, 0)$-estimators.

Huber (1967) developed the asymptotic theory of M-estimators. Pollard (1988), Haberman (1989), Niemiro (1992) and others have emphasized the role of convexity of the kernel $f(s, x)$ in s in the asymptotic study of M-estimators (as well as some other estimators defined by minimization of statistical functionals). They showed that in this case the theory can be substantially simplified and the asymptotic normality of M-estimators can be established under minimal and very natural assumptions.

We use convex analysis rather extensively in this paper. The relevant facts can be found in Rockafellar (1970) ([Ro]). Some frequently used notions are given in the Appendix.

In what follows, we assume that $f(\cdot, x)$ is a convex function for P almost all x. We call such a kernel f *P-convex*. If there exists $A \in \mathscr{A}$ such that $f(\cdot, x)$ is strictly convex for all $x \in A$ and $P(A) > 0$, then f will be called a *P-strictly convex kernel*.

Given a P-convex kernel f, we call the subdifferential map ∂f_P of the convex function f_P the *M-distribution map* of P with respect to f. Its inverse, $\partial f_P^{-1} = \partial f_P^*$, is called the *M-quantile map* of P with respect to f. For the convex function f_P given by (1.1), ∂f_P and ∂f_P^{-1} are multivalued versions of the usual distribution and quantile functions of P (see Example 2.2), which motivates our definitions. For any $t \in \mathbb{R}^d$, $(\partial f_P)^{-1}(t)$ is the set of all (M, t)-parameters of P. If, in addition, f is a P-strictly convex kernel, then ∂f_P^{-1} is a single-valued map, so that the (M, t)-parameter is unique (for all t such that it exists). In this case we call a subgradient F_P (i.e., a map F_P from $\mathbb{R}^d$ into $\mathbb{R}^d$ such that $F_P(s) \in \partial f_P(s)$ for all $s \in \mathbb{R}^d$) the *M-distribution function* of P with respect to f, and we call its inverse F_P^{-1} from $F_P(\mathbb{R}^d)$ onto $\mathbb{R}^d$, the *M-quantile function*. In particular, if $0 \in F_P(\mathbb{R}^d)$, then the M-parameter of P is unique and it coincides with the "median" $F_P^{-1}(0)$ and M-estimators are "sample medians" $F_{P_n}^{-1}(0)$. Using some basic facts of convex analysis, it is easy to describe the range $F_P(\mathbb{R}^d)$ of the map F_P (see Section 2). We also give some conditions when F_P charaterizes P (in the sense that $F_P = F_Q$ implies $P = Q$). In Section 3, we study equivariance of M-distribution and M-quantile maps with respect to the groups of translations and orthogonal transformations and obtain representations of these functions for spherically symmetric probability distributions.

Section 5 develops the asymptotic theory of empirical M-distribution and M-quantile functions. Specifically, we prove that the sequences $n^{1/2}(F_{P_n} - F_P)$ and $n^{1/2}(F_{P_n}^{-1} - F_P^{-1})$ converge in distribution to Gaussian processes.

We apply these results to two main examples. First of all, we consider a class of extensions of quantiles to the multivariate case. To define such spatial quantiles we use the kernel $f(s, x) := \varphi(|s - x|) - \varphi(|s_0 - x|),\ s, x \in \mathbb{R}^d$, where $s_0 \in \mathbb{R}^d$ is a fixed point and φ is a convex nondecreasing function on $[0, +\infty)$, differentiable in $(0, +\infty)$. In this case F_P is defined by

$$F_P(s) := \int_{\{x \neq s\}} \Phi(|s - x|) \frac{s - x}{|s - x|} P(dx) \quad \text{for } s \in \mathbb{R}^d,$$

where $\Phi := \varphi'$. If φ is strictly increasing and P is not concentrated in a line, then the range $V_P := F_P(\mathbb{R}^d)$ is a ball (which, in particular, could coincide with the whole space $\mathbb{R}^d$) with a number of spherical "holes" in it (one "hole" for each atom of P). The map F_P is one-to-one from $\mathbb{R}^d$ onto V_P, so that F_P^{-1} is well defined (see Proposition 2.6). In particular, if $\varphi(u) \equiv u$ and $\Phi \equiv 1$, we get an extension of the distribution and quantile functions to the multivariate case as introduced by [K-D] and [Ch]. We study the main properties of these M-distribution and M-quantile functions (analytical properties, characterization of the distribution P, asymptotics of their empirical counterparts, etc.). We show also their equivariance with respect to translations and orthogonal transformations and obtain representations of F_P and F_P^{-1} for spherically symmetric distributions. These representations are used in Section 4 to develop tests for spherical symmetry of multivariate distributions.

Another example is related to the notion of regression quantiles, introduced by Koenker and Basset (1978). Consider a sample $(X_i, Y_i),\ i = 1, \dots, n$ of i.i.d. observations, where $X_i,\ i \geq 1$ take values in $\mathbb{R}^m$ and $Y_i,\ i \geq 1$ take values in $\mathbb{R}^1$. Suppose that our observations satisfy the following linear regression model with random design:

$$Y_i = \langle s_0, X_i \rangle + \xi_i, \qquad 1 \leq i \leq n,$$

where $s_0 \in \mathbb{R}^m$ and $\xi_i,\ i \geq 1$ are i.i.d. errors, independent of $X_i\ i \geq 1$. Koenker and Basset (1978) define (empirical) regression α-quantiles for $\alpha \in [0, 1]$ as minimal points of the function in s,

$$\sum_{1}^{n}\big[|Y_i - \langle s, X_i\rangle| - (2\alpha - 1)\langle s, X_i\rangle\big].$$

One of the reasons to introduce this notion was to develop a technique of robust estimation for regression models. Koenker and Portnoy (1987) studied L-estimators for regression, defined in terms of regression quantiles. Chaudhuri (1991) considered nonparametric estimates of regression quantiles. See also Gutenbrunner and Jurečková (1992) and Gutenbrunner, Jurečková, Koenker and Portnoy (1993) for applications of this notion.

[Ch] suggested extending the notion to the case of multiresponse regression, using the version of spatial quantiles introduced by him and by [K-D]. In this paper, we define regression quantiles for multiresponse regression models and show that they can be viewed and studied as a special case of the M-quantile function with respect to a certain convex kernel.

2. *M*-distribution and *M*-quantile functions with respect to convex kernels.

Let f be a P-convex kernel such that for all $s \in \mathbb{R}^d$ $f(s, \cdot)$ is a P-integrable function. Then f_P is a convex function from $\mathbb{R}^d$ into $\mathbb{R}^1$. Moreover, if f is a P-strictly convex kernel, then f_P is a strictly convex function. Let $df(s; v; x)$ denote the directional derivative of the function $f(\cdot, x)$ at a point $s \in \mathbb{R}^d$ in the direction $v \in \mathbb{R}^d$,

$$df(s; v; x) := \inf_{\lambda \in (0, +\infty)} \Delta f(s; v; \lambda; x) = \lim_{\lambda \downarrow 0} \Delta f(s; v; \lambda; x),$$

where $\Delta f(s; v; \lambda; x) := \lambda^{-1}(f(s + \lambda v; x) - f(s; x))$. The directional derivative exists for any x such that $f(\cdot, x)$ is convex; thus, it exists P a.e. Since the function $\Delta f(s; v; \lambda; x)$ is increasing in $\lambda > 0$, one can use the monotone convergence theorem to show that for all $s \in \mathbb{R}^d$,

$$df_P(s; v) = \lim_{\lambda \downarrow 0} \Delta f_P(s; v; \lambda) = \lim_{\lambda \downarrow 0} \int_{\mathbf{X}} \Delta f(s; v; \lambda; x) P(dx) = \int_{\mathbf{X}} df(s; v; x) P(dx).$$

Clearly,

$$\partial f_P(s) = \bigcap_{v \in \mathbb{R}^d,\, v \neq 0} \left\{ t\colon \int_{\mathbf{X}} df(s; v; x) P(dx) \geq \langle v, t\rangle \right\}.$$

Given $s \in \mathbb{R}^d$, $x \mapsto \partial f(s; x)$ is a closed convex multivalued map. Given a multivalued map $x \mapsto C(x) \subset \mathbb{R}^d$, an integral $\int_{\mathbf{X}} C(x) P(dx)$ is defined as a set

$$\left\{ \int_{\mathbf{X}} c(x) P(dx)\colon c : \mathbf{X} \mapsto \mathbb{R}^d \text{ is } P\text{-integrable, } c(x) \in C(x) \text{ for } P\text{-almost all } x \right\}.$$

The next statement is a special case of Theorem 8.3.4 of Ioffe and Tihomirov (1974). It shows that one can subdifferentiate under the integral sign and gives an integral representation of the multivalued map ∂f_P.

THEOREM 2.1. *Suppose f is a P-convex kernel, and for all $s \in \mathbb{R}^d$ $f(s;\cdot)$ is P-integrable. Then, for all $s \in \mathbb{R}^d$,*

$$\partial f_P(s) = \int_{\mathbf{X}} \partial f(s;x) P(dx).$$

A map F_P: $\mathbb{R}^d \mapsto \mathbb{R}^d$ is a subgradient of the convex function f_P if and only if there exists a map F: $\mathbb{R}^d \times \mathbf{X} \mapsto \mathbb{R}^d$, such that $F(s,x) \in \partial f(s,x)$ P a.e. [*so, $F(s,x)$ is a subgradient of $f(s,x)$ P a.e.*], *$F(s,\cdot)$ is P-integrable and*

$$F_P(s) := \int_{\mathbf{X}} F(s;x) P(dx). \tag{2.1}$$

EXAMPLE 2.2. Let $\mathbf{X} = \mathbb{R}^1$, and let $f(s,x) := 0.5(|s-x|-|x|+s)$ for $s, x \in \mathbb{R}^1$. Then

$$f_P(s) := 0.5 \int_{\mathbb{R}^1} (|s-x| - |x|) P(dx) + 0.5s \quad \text{for } s \in \mathbb{R}^1.$$

In this case for any $x \in \mathbb{R}^1$, a function $F(\cdot;x)$, defined by $F(s;x) := \mathbf{I}_{(-\infty,s]}(x)$ for $x, s \in \mathbb{R}^1$, is a subgradient of the convex function $f(\cdot;x)$. Therefore, a function F_P, defined by

$$F_P(s) := \int_{\mathbb{R}^1} \mathbf{I}_{(-\infty,\, s]}(x) P(dx) = P((-\infty, s]) \quad \text{for } s \in \mathbb{R}^1,$$

is a subgradient of f_P. Clearly, F_P is the distribution function of the probability measure P. The inverse F_P^{-1} of the function F_P is the quantile function of P. It's easy to show that the subdifferential of f_P is a multivalued map ∂f_P, defined by $\partial f_P(s) = [F_P(s-), F_P(s)]$ for $s \in \mathbb{R}^1$. On the other hand, the inverse of the multivalued map ∂f_P,

$$\partial f_P^{-1}(t) = \begin{cases} [F_P^{-1}(t), F_P^{-1}(t+)], & \text{for } t \in (0,1), \\ (-\infty, F_P^{-1}(0+)], & \text{for } t = 0, \\ [F_P^{-1}(1), +\infty), & \text{for } t = 1, \\ \varnothing, & \text{for } t \notin [0,1], \end{cases}$$

is the subdifferential ∂f_P^* of the conjugate f_P^*. In particular, $F_P^{-1}(t)$ is a subgradient of the convex function $f_P^*(t)$, $t \in (0,1)$ [for $t = 0$ and $t = 1$, $F_P^{-1}(0+)$ and $F_P^{-1}(1)$ are also subgradients of f_P^*, if they are finite]. The set of (M,t)-parameters of P coincides with the set $[F_P^{-1}(t), F_P^{-1}(t+)]$ of t-quantiles of P. Clearly, ∂f_P and $\partial f_P^* = \partial f_P^{-1}$ can be viewed as multivalued versions of the usual distribution and quantile functions, respectively.

Motivated by this example, we introduce the following definition in the general case.

DEFINITION 2.3. The subdifferential map ∂f_P of the function f_P will be called the *M-distribution map* of P with respect to f. Respectively, the subdifferential map $\partial f_P^* = \partial f_P^{-1}$ of the conjugate map f_P^* will be called the *M-quantile map* of P with respect to f. Any point of the set $\partial f_P^{-1}(t)$ is called an *(M,t)-quantile* of P with respect to f. If ∂f_P (respectively, ∂f_P^{-1}) is a single-valued map, we call it the M-distribution (respectively, the M-quantile) function of P with respect to f.

Define

$$r_f(v; x) := \sup_{s\in\mathbb{R}^d} df(s; v; x) \quad \text{for } v \in \mathbb{R}^d \text{ and } x \in \mathbf{X},$$

so that $r_f(\cdot, x)$ is the recession function of $f(\cdot, x)$. It follows from the properties of the recession function (see [Ro]), that for any $s \in \mathbb{R}^d$ and for P almost all x

$$r_f(v;\, x) = \lim_{\lambda\uparrow+\infty} \Delta f(s; v; \lambda; x).$$

Therefore, by the monotone convergence theorem, we have the following representation for the recession function of f_P:

$$\begin{aligned} r_P(v) := r_{f_P}(v) &:= \sup_{s\in\mathbb{R}^d} df_P(s; v) = \lim_{\lambda\uparrow+\infty} \Delta f_P(s; v; \lambda) \\ &= \lim_{\lambda\uparrow+\infty} \int_{\mathbf{X}} \Delta f(s; v; \lambda; x) P(dx) = \int_{\mathbf{X}} r_f(v; x) P(dx) \quad \text{for } v \in \mathbb{R}^d. \end{aligned}$$

We define a convex set

$$B_P := B_{f_P} := \bigcap_{v\in\mathbb{R}^d,\, v\neq 0} \{t\colon \langle v, t\rangle < r_P(v)\}.$$

The following properties of M-distribution and M-quantile maps easily follow from the general facts of convex analysis (see [Ro], especially Sections 23–27). They are quite similar to the properties of the multivalued versions of the usual distribution and quantile functions [such as their monotonicity, semicontinuity, relationship with (M, t)-parameters (quantiles) and M-parameters (medians), etc.; see Example 2.2].

2.1. *Properties of M-distribution and M-quantile maps.* Let f be a P-convex kernel. Then we have the following properties.

1. The M-distribution map ∂f_P of P with respect to f is a monotone upper semicontinuous multivalued map. Moreover, if f is P-strictly convex, then ∂f_P is strictly monotone.
2. The M-quantile map of P with respect to f is the inverse of the M-distribution map: $\partial f_P^* = (\partial f_P)^{-1}$. It is a monotone upper semicontinuous multivalued map.

   Let $D_P := D_{f_P}$ be the set of all points $s \in \mathbb{R}^d$ such that $df_P(s; v)$ is linear in v.
3. f_P is differentiable at points $s \in D_P$ and $\partial f_P(s) = \{\nabla f_P(s)\},\ s \in D_P$. D_P is a dense subset of $\mathbb{R}^d$ and its complement has Lebesgue measure 0. The gradient ∇f_P is continuous on D_P.
4. For any $t \in \mathbb{R}^d$, the set of all (M, t)-parameters of P with respect to f coincides with $\partial f_P^*(t) = \partial f_P^{-1}(t)$ [the set of all (M, t)-quantiles of P with respect to f]. Thus, an (M, t)-parameter of P exists iff f_P^* is subdifferentiable at t. The set of (M, t)-parameters is nonempty and bounded iff $t \in B_P$. The set of (M, t)-parameters consists of the unique point s iff f_P^* is differentiable at t and $s = \nabla f_P^*(t)$.

Let now f be a P-strictly convex kernel. Denote

$$U_P := U_{f_P} := B_P \Big\backslash \bigcup_{s \in \mathbb{R}^d \setminus D_P} \partial f_P(s).$$

Then:

5. The convex function f_P^* is continuously differentiable in B_P with the gradient (and subgradient) ∇f_P^*. Moreover, ∇f_P^* is the M-quantile function of P with respect to f.
6. For any $t \in B_P$, the unique (M, t)-parameter of P coincides with $\nabla f_P^*(t)$. The gradient ∇f_P of the function f_P is a homeomorphism of D_P and U_P. For any $t \in U_P$, $\nabla f_P^*(t) = (\nabla f_P)^{-1}(t)$.

To be more specific, assume that $\mathbf{X} = \mathbb{R}^d$, and that P is a Borel probability on $\mathbb{R}^d$. Define a function

$$f_P(s) := \int_{\mathbb{R}^d} (h(s-x) - h(s_0 - x))P(dx) \quad \text{for } s \in \mathbb{R}^d,$$

where $h\colon \mathbb{R}^d \mapsto \mathbb{R}^1$ is convex, $s_0 \in \mathbb{R}^d$ is a fixed point, and $h(s-\cdot) - h(s_0 - \cdot)$ is P-integrable. Using Theorem 2.1, we get the following representation of the M-distribution function:

$$\partial f_P(s) = \int_{\mathbb{R}^d} \partial h(s-x)P(dx) \quad \text{for } s \in \mathbb{R}^d. \tag{2.2}$$

For all $x \in \mathbb{R}^d$ and $v \in \mathbb{R}^d$,

$$r_f(v; x) = \lim_{\lambda \uparrow +\infty} \frac{h(\lambda v - x) - h(-x)}{\lambda} = r_h(v).$$

Therefore

$$r_P(v) := r_{f_P}(v) = \int_{\mathbb{R}^d} r_f(v; x)P(dx) = r_h(v) \quad \text{for } v \in \mathbb{R}^d,$$

so the recession function of f_P does not depend on P. It follows that B_P coincides with a convex set

$$B_h := \bigcap_{v \in \mathbb{R}^d,\, v \neq 0} \{t\colon \langle v, t\rangle < r_h(v)\}.$$

Assume that h is continuously differentiable in $\mathbb{R}^d \setminus \{0\}$. It means that $D_h = \mathbb{R}^d \setminus \{0\}$, unless h is differentiable at 0 too, and then $D_h = \mathbb{R}^d$. Let $t_0 \in \partial h(0)$. Then the map $H\colon \mathbb{R}^d \mapsto \mathbb{R}^d$, defined by

$$H(s) := \begin{cases} \nabla h(s), & \text{for } s \neq 0, \\ t_0, & \text{for } s = 0, \end{cases}$$

is a subgradient of the convex function h. Therefore the map F_P, defined by

$$F_P(s) := \int_{\mathbb{R}^d} H(s-x)P(dx),$$

is a subgradient of f_P.

Let A_P be the set of all atoms of the probability measure P, and let, for $s \in A_P$, $p(s) := P(\{s\})$. Since

$$\partial h(s-x) = \begin{cases} \partial h(0), & \text{for } x = s, \\ \{H(s-x)\}, & \text{for } x \neq s, \end{cases}$$

we have

$$\partial f_P(s) = \begin{cases} \{F_P(s)\}, & \text{for } s \notin A_P, \\ F_P(s) + p(s)(\partial h(0) - H(0)), & \text{for } s \in A_P. \end{cases}$$

Thus f_P is continuously differentiable in $\mathbb{R}^d \setminus A_P$. If h is not differentiable at 0, then $D_P = \mathbb{R}^d \setminus A_P$, otherwise f_P is continuously differentiable in $\mathbb{R}^d$ and $D_P = \mathbb{R}^d$. We also define

$$U_P := B_h \Big\backslash \bigcup_{s \in A_P} \big(F_P(s) + p(s)(\partial h(0) - H(0))\big).$$

The following statement immediately follows from the main properties of M-distribution and M-quantile maps.

PROPOSITION 2.4. (i) *For all* P, F_P *is a monotone map, continuous in* $\mathbb{R}^d \setminus A_P$. *If* $f(s;x) = h(s-x) - h(s_0 - x)$ *is a* P-*strictly convex kernel, then*

(ii) F_P *is a strictly monotone one-to-one map from* $\mathbb{R}^d$ *onto* $U_P \cup F_P(A_P)$ *and it is a homeomorphism of* $\mathbb{R}^d \setminus A_P$ *and* U_P. *If* P *is nonatomic, or if* h *is differentiable at* 0, *then* F_P *is a homeomorphism of* $\mathbb{R}^d$ *and* B_h.

We extend the map F_P^{-1} (*defined in a natural way on the set* U_P) *to the whole set* B_h *assuming that for* $t \in B_h \setminus U_P$ $F_P^{-1}(t)$ *coincides with the unique* $s \in A_P$ *such that* $t \in (F_P(s) + p(s)(\partial h(0) - F_P(0)))$. *Then the following holds.*

(iii) *For all* $t \in B_h$ *the unique* (M, t)-*parameter of* P *coincides with* $F_P^{-1}(t)$.

Note that if h is a strictly convex function, then $f(s;\cdot) = h(s-\cdot) - h(s_0 - \cdot)$ is a P-strictly convex kernel. It also holds, if P is not concentrated in a line, h is convex and for all $s_1, s_2 \in \mathbb{R}^d$, such that $s_1 \neq \alpha s_2$ for all $\alpha \geq 0$, we have

$$h(\lambda s_1 + (1-\lambda)s_2) < \lambda h(s_1) + (1-\lambda)h(s_2) \quad \text{for } \lambda \in (0,1).$$

It is well known that in the one-dimensional case the distribution function F_P characterizes P, that is, $F_P \equiv F_Q$ implies $P = Q$. Now we extend this importatant property to the class of M-distribution functions defined above [see Koldobskii (1990) and Mattner (1992) for more general conditions on the kernels of convolution integrals, characterizing probability measures]. Denote $\mathscr{S}'(\mathbb{R}^d)$ the space of tempered generalized functions (distributions). Given $g \in \mathscr{S}'(\mathbb{R}^d)$, denote $\tilde{g}$ the Fourier transform of g [see Gelfand and Shilov (1964) or Reed and Simon (1975)]. Since Borel probabilities on $\mathbb{R}^d$ can be also viewed as generalized functions, we use the same notation for their Fourier transforms (characteristic functions):

$$\tilde{P}(\lambda) := \int_{\mathbb{R}^d} e^{i\langle \lambda, x\rangle} P(dx) \quad \text{for } \lambda \in \mathbb{R}^d.$$

Given h, let $\mathscr{P}_h$ be the set of all Borel probability measures P on $\mathbb{R}^d$, such that the following hold:

1. $h(s-x) - h(-x)$ is a P-strictly convex kernel;
2. $h(s-\cdot) - h(-\cdot)$ is P-integrable for all $s \in \mathbb{R}^d$;
3. for some $\alpha > 0$ $\int_{\mathbb{R}^d} |H(s-x)|P(dx) = O(|s|^\alpha)$ as $s \to \infty$.

THEOREM 2.5. *Suppose that the Fourier transform $\tilde{h}$ is given by a Borel function on $\mathbb{R}^d$ such that $\tilde{h} \neq 0$ a.e. in $\mathbb{R}^d$. Then, for any two probability measures $P, Q \in \mathscr{P}_h$, the equality $F_P \equiv F_Q$ implies that $P = Q$.*

Let us consider a special case of the previous kernel. Namely, assume that $h(x) := \varphi(|x|)$ for $x \in \mathbb{R}^d$, where φ is a convex nondecreasing function on $[0, +\infty)$. We also suppose that the functions $\varphi(|s-\cdot|) - \varphi(|s_0 - \cdot|),\ s \in \mathbb{R}^d$, are P-integrable, and define

$$f_P(s) := \int_{\mathbb{R}^d} (\varphi(|s-x|) - \varphi(|s_0 - x|))P(dx) \quad \text{for } s \in \mathbb{R}^d.$$

Denote

$$\rho_\varphi := \lim_{\varepsilon \downarrow 0} \frac{\varphi(\varepsilon) - \varphi(0)}{\varepsilon} \quad \text{and} \quad R_\varphi := \lim_{\lambda \uparrow +\infty} \frac{\varphi(\lambda)}{\lambda}.$$

The use of Theorem 2.1 and a bit of subdifferential calculus allow us to get the following representation of the M-distribution map ∂f_P for all $s \in \mathbb{R}^d$:

$$\begin{aligned}\partial f_P(s) &= \int_{\mathbb{R}^d} \partial(\varphi(|s-x|))P(dx) \\ &= \int_{\mathbb{R}^d \setminus \{s\}} \partial\varphi(|s-x|)\frac{s-x}{|s-x|}P(dx) + p(s)\ \bar{B}(0, \rho_\varphi).\end{aligned}$$

We assume for simplicity that φ is differentiable in $(0, +\infty)$. Let $\Phi(\lambda) := \varphi'(\lambda),\ \lambda > 0$. Then $\partial f_P(s) = F_P(s) + p(s)\bar{B}(0, \rho_\varphi) = \bar{B}(F_P(s), p(s)\rho_\varphi),\ s \in \mathbb{R}^d$, where

$$F_P(s) := \int_{\mathbb{R}^d \setminus \{s\}} \Phi(|s-x|)\frac{s-x}{|s-x|}P(dx) \quad \text{for } s \in \mathbb{R}^d.$$

It follows that F_P is a subgradient of f_P. For $h(x) := \varphi(|x|)$ we have

$$r_h(v) := \lim_{\lambda \uparrow +\infty} \frac{h(\lambda v) - h(0)}{\lambda} = \lim_{\lambda \uparrow +\infty} \frac{\varphi(\lambda|v|) - \varphi(0)}{\lambda} = R_\varphi\ |v|,$$

and it follows that $B_P = B_h = B(0; R_\varphi)$. Let

$$U_P := B(0; R_\varphi) \Big\backslash \bigcup_{s \in A_P} \bar{B}(F_P(s); p(s)\rho_\varphi).$$

In this case we get the following special version of Proposition 2.4.

PROPOSITION 2.6. (i) *For all* P, F_P *is a monotone continuous in* $\mathbb{R}^d \setminus A_P$. *For any* $s \in A_P$,

$$F_P(s+\varepsilon v) \to F_P(s) + p(s)\rho_\varphi v \quad \textit{as } \varepsilon \to 0 \textit{ for } v \in S^{d-1};$$

$$\text{(ii)} \qquad F_P(\lambda v) \to \begin{cases} R_\varphi v, & \textit{for } R_\varphi < +\infty \\ \infty, & \textit{otherwise} \end{cases} \qquad \textit{as } \lambda \to \infty \textit{ for } v \in S^{d-1}.$$

If either φ *is strictly convex, or it is strictly increasing and* P *is not concentrated in a line, then*:

(iii) F_P *is a strictly monotone one-to-one map from* $\mathbb{R}^d$ *onto* $U_P \cup F_P(A_P)$ *and it is a homeomorphism of* $\mathbb{R}^d \setminus A_P$ *and* U_P. *If* P *is nonatomic, then* F_P *is a homeomorphism of* $\mathbb{R}^d$ *and* $B(0; R_\varphi)$.

We extend the map F_P^{-1} (*defined on the set* U_P) *to the whole ball* $B(0; R_\varphi)$ *assuming that for* $t \in B(0; R_\varphi) \setminus U_P$ $F_P^{-1}(t)$ *coincides with the unique* $s \in A_P$ *such that* $t \in \bar{B}(F_P(s); p(s)\rho_\varphi)$. *Then*:

(iv) *For any* $t \in B(0; R_\varphi)$, *the unique* (M, t)*-parameter of* P *coincides with* $F_P^{-1}(t)$.

The next example was considered by [K-D] and [Ch].

EXAMPLE 2.7. Let $\mathbf{X} = \mathbb{R}^d$, $d \geq 1$, and let $f(s; x) = |s-x| - |x|$. Consider the functional

$$f_P(s) := \int_{\mathbb{R}^d} (|s-x| - |x|) P(dx) \quad \text{for } s \in \mathbb{R}^d.$$

In this case we have $\varphi(\lambda) := \lambda,\ \lambda > 0$, and $\Phi(\lambda) := \varphi'(\lambda) = 1,\ \lambda > 0$, and the map F_P, defined by

$$F_P(s) := \int_{\{x \neq s\}} \frac{s-x}{|s-x|} P(dx),$$

is a subgradient of f_P. In the case $d = 1$ we have $(s-x)/|s-x| = \operatorname{sign}(s-x)$, and

$$F_P(s) = P((-\infty, s]) + P((-\infty, s)) - 1 \quad \text{for } s \in \mathbb{R}^1,$$

so F_P is just a simple modification of the distribution function of P, while F_P^{-1} is a simple modification of the quantile function of P.

For $d > 1$, Proposition 2.6 (with $\rho_\varphi = R_\varphi = 1$) gives the main properties of the maps F_P and F_P^{-1}, first studied by [K-D] and [Ch].

Note that properties (iii) and (iv) in Proposition 2.6 immediately imply the existence and the uniqueness of the Haldane's spatial median (if P is not concentrated in a line). See also Milasevic and Ducharme (1987).

EXAMPLE 2.8. Let $\mathbf{X} = \mathbb{R}^d$, $d \geq 1$, and let $f(s; x) := |s-x|^r - |x|^r$ for some $r > 1$. We consider the functional

$$f_P(s) := \int_{\mathbb{R}^d} (|s-x|^r - |x|^r) P(dx) \quad \text{for } s \in \mathbb{R}^d,$$

which is well defined for all P with $\int_{\mathbb{R}^d} |x|^{r-1} P(dx) < +\infty$. In this case $\varphi(\lambda) = \lambda^r,\ \lambda \geq 0$. We have $R_\varphi = +\infty,\ \rho_\varphi = 0$, and $\Phi(\lambda) = r\lambda^{r-1}$, $\lambda \geq 0$. A map F_P, defined by

$$F_P(s) := r \int_{\{x \neq s\}} (s - x)|s - x|^{r-2} P(dx),$$

is a subgradient of f_P. Proposition 2.6 implies the following:

(i) F_P is a strictly monotone homeomorphism from $\mathbb{R}^d$ onto $\mathbb{R}^d$;
(ii) $F_P(\lambda v) \to \infty$ as $\lambda \to \infty$ for $v \in S^{d-1}$;
(iii) For any $t \in \mathbb{R}^d$, the unique (M, t)-parameter of P coincides with $F_P^{-1}(t)$.

In Examples 2.7, 2.8 we have $h(x) = |x|^r,\ r \geq 1$. The Fourier transform of this function is given by the following formula [see Gelfand and Shilov (1964)]:

$$\tilde{h}(\lambda) = 2^{r+d} \pi^{d/2} \frac{\Gamma((r+d)/2)}{\Gamma(-r/2)} |\lambda|^{-r-d}.$$

If $r \neq 2k,\ k = 1, 2, \ldots$, we have $\tilde{h}(\lambda) \neq 0$ for all $\lambda \neq 0$.
Given $r \geq 1$, denote $\mathscr{P}_r$ the set of all Borel probabilities with

$$\int_{\mathbb{R}^d} |x|^{r-1} P(dx) < +\infty.$$

Theorem 2.5 implies the following statement.

COROLLARY 2.9. *If $r \geq 1$ and $r \neq 2k,\ k = 1, 2, \ldots$, then, for any two probability measures $P, Q \in \mathscr{P}_r$, the equality $F_P = F_Q$ implies $P = Q$.*

3. Equivariance and symmetry. In this section we study equivariance of M-distribution and M-quantile functions with respect to certain groups of transformations.

3.1. *Group of translations.* Given $\theta \in \mathbb{R}^d$, denote by P_θ the θ-translation of the Borel probability measure P in $\mathbb{R}^d$.

PROPOSITION 3.1. *Let $h: \mathbb{R}^d \mapsto \mathbb{R}^1$ be a convex function. Suppose that for some $s_0 \in \mathbb{R}^d$ and for all $s \in \mathbb{R}^d$ functions $h(s - \cdot) - h(s_0 - \cdot)$ are P-integrable. Let*

$$f_P(s) := \int_{\mathbb{R}^d} (h(s - x) - h(s_0 - x)) P(dx) \quad \text{for } s \in \mathbb{R}^d.$$

Then for all $\theta \in \mathbb{R}^d$, (i) *$\partial f_{P_\theta}(s) = \partial f_P(s - \theta)$ for $s \in \mathbb{R}^d$;* (ii) *$\partial f_{P_\theta}^{-1}(t) = \theta + \partial f_P^{-1}(t)$ for $t \in \mathbb{R}^d$.*

If h is continuously differentiable in $\mathbb{R}^d \setminus \{0\}$, we can choose a point $t_0 \in \partial h(0)$, and define a map

$$H(s) := \begin{cases} \nabla h(s), & \text{for } s \neq 0, \\ t_0, & \text{for } s = 0. \end{cases}$$

Then the map F_P, defined by

$$F_P(s) := \int_{\mathbb{R}^d} H(s-x) P(dx) \quad \text{for } s \in \mathbb{R}^d,$$

is a subgradient of f_P, and Proposition 2.4 gives its properties. In this case we have the following proposition.

PROPOSITION 3.2. (i) *For all* $\theta \in \mathbb{R}^d$, $F_{P_\theta}(s) = F_P(s-\theta)$, $s \in \mathbb{R}^d$. (ii) *For all* $\theta \in \mathbb{R}^d$, *the* (M, t)*-parameter of* P_θ *is the* θ*-translation of the* (M, t)*-parameter of* P:

$$F_{P_\theta}^{-1}(t) = \theta + F_P^{-1}(t) \quad \textit{for } t \in B_h.$$

3.2. *Orthogonal group.* Let us consider the function

$$f_P(s) := \int_{\mathbb{R}^d} (\varphi(|s-x|) - \varphi(|s_0 - x|)) P(dx) \quad \text{for } s \in \mathbb{R}^d,$$

where φ is a convex function on $[0, +\infty)$, $s_0 \in \mathbb{R}^d$ is a fixed point and the functions $\varphi(|s - \cdot|) - \varphi(|s_0 - \cdot|)$ are P-integrable for all $s \in \mathbb{R}^d$.

PROPOSITION 3.3. *The following hold for all orthogonal transformations* O *of* $\mathbb{R}^d$:

(i) $\partial f_{P\circ O^{-1}}(s) = O \partial f_P(O^{-1}s)$ *for* $s \in \mathbb{R}^d$;
(ii) $\partial f_{P\circ O^{-1}}^{-1}(t) = O \partial f_P^{-1}(O^{-1}t)$ *for* $t \in \mathbb{R}^d$.

Denote

$$\rho(x; \lambda) := \sqrt{(\lambda - x^{(1)})^2 + (x^{(2)})^2 + \cdots + (x^{(d)})^2} \quad \text{for } x \in \mathbb{R}^d \text{ and } \lambda \geq 0.$$

PROPOSITION 3.4. *Let* P *be a Borel probability measure in* $\mathbb{R}^d$, *spherically symmetric with respect to the center* $a \in \mathbb{R}^d$. *Then we have the following*:

(i)
$$\partial f_P(s) = \begin{cases} \psi_P(|s-a|) \dfrac{s-a}{|s-a|}, & \textit{for } s \neq a, \\ \bar{B}(0; \rho_\varphi), & \textit{for } s = a, \end{cases}$$

where ψ_P *is a monotone multivalued map from* $(0, +\infty)$ *into* $\mathbb{R}^1$, *defined by*

$$\psi_P(\lambda) := \int_{\mathbb{R}^d} \frac{\lambda + a^{(1)} - x^{(1)}}{\rho(x-a; \lambda)} (\partial \varphi)\big(\rho(x-a; \lambda)\big)\, P(dx);$$

(ii)
$$\partial f_P^{-1}(t) = \begin{cases} a + (\psi_P)^{-1}(|t|) \dfrac{t}{|t|}, & \textit{for } |t| > \rho_\varphi, \\ a, & \textit{otherwise}. \end{cases}$$

If φ is differentiable in $(0, +\infty)$, and $\Phi(\lambda) = \varphi'(\lambda)$, $\lambda > 0$, then the map F_P, defined by

$$F_P(s) := \int_{\{x \neq s\}} \Phi(|s-x|) \frac{s-x}{|s-x|} P(dx) \quad \text{for } s \in \mathbb{R}^d,$$

is a subgradient of f_P. Proposition 2.6 gives the properties of this map. In this case we have

PROPOSITION 3.5. (i) *For any orthogonal transformation O of $\mathbb{R}^d$, $F_{P \circ O^{-1}}(s) = OF_P(O^{-1}s)$ for $s \in \mathbb{R}^d$.*

(ii) *If φ is strictly convex, or it is strictly increasing and P is not concentrated in a line, then $F^{-1}_{P \circ O^{-1}}(t) = OF^{-1}_P(O^{-1}t)$ for $t \in B(0; R_\varphi)$.*

PROPOSITION 3.6. *Let P be a Borel probability measure in $\mathbb{R}^d$, spherically symmetric with respect to the center $a \in \mathbb{R}^d$. Then we have the following*:

(i)
$$F_P(s) = \begin{cases} \psi_P(|s-a|) \dfrac{s-a}{|s-a|}, & \text{for } s \neq a, \\ 0, & \text{for } s = a, \end{cases}$$

where ψ_P is a nondecreasing function from $(0, +\infty)$ *into* $\mathbb{R}^1$, *defined by*

$$\psi_P(\lambda) := \int_{\mathbb{R}^d} \frac{\lambda + a^{(1)} - x^{(1)}}{\rho(x-a;\lambda)} \varphi'(\rho(x-a;\lambda))\, P(dx);$$

(ii)
$$F^{-1}_P(t) = \begin{cases} a + (\psi_P)^{-1}(|t|) \dfrac{t}{|t|}, & \text{for } |t| > \rho_\varphi, \\ a, & \text{otherwise}. \end{cases}$$

Note that, given $a \in \mathbb{R}^d$, the transformation $s \mapsto 2a - s$ of $\mathbb{R}^d$ (reflection of the space $\mathbb{R}^d$ with respect to the center a) is the superposition of a translation and the orthogonal transformation $s \mapsto -s$. Suppose that P is symmetric with respect to the reflection about a (which means that $X - a$ and $a - X$ are identically distributed, given a random vector X with distribution P). In this case, Propositions 3.2 and 3.5 imply that

$$F_P(s) = -F_P(2a - s) \quad \text{for } s \in \mathbb{R}^d$$

and

$$F^{-1}_P(-t) = 2a - F^{-1}_P(t) \quad \text{for } t \in B(0; R_\varphi).$$

3.3. *Affine group.* Unfortunately, the M-distribution and M-quantile functions described above are not affine equivariant. In order to define affine equivariant versions of these functions, one should use in their definitions the kernel f, which depends on P. One of the possible choices of such a kernel is based on standardization of P using the square root of its covariance matrix [Rao (1988) suggested doing this in order to define an affine equivariant version of Hadane's median]. Namely, one can use the kernel

$$f(s, x) := \varphi(|s - x|_P) - \varphi(|s_0 - x|_P),$$

where $|s|_P := \left|\mathrm{inv}(\Sigma_P^{1/2})s\right| = \langle \mathrm{inv}(\Sigma_P)s, s\rangle^{1/2}$, $s \in \mathbb{R}^d$, Σ_P being the covariance operator of measure P.

4. Statistical applications of *M*-distributions and *M*-quantiles. In this section we consider some statistical applications of M-distribution and M-quantile functions, introduced by [K-D] and [Ch]; see our Example 2.7. In this case $\Phi(\lambda) \equiv 1$, so we have

$$F_P(s) := \int_{\{x \neq s\}} \frac{s-x}{|s-x|} P(dx). \tag{4.1}$$

In addition to the properties of the map F_P, summarized in Sections 2 and 3, we need the following proposition. Denote $\mathbf{I}_d$ the identity operator in $\mathbb{R}^d$.

PROPOSITION 4.1. *Suppose that P has a bounded density in $\mathbb{R}^d$ with $d \geq 2$. Then F_P is continuously differentiable in $\mathbb{R}^d$ with derivative*

$$F'_P(s) = \int_{\{x \neq s\}} \frac{1}{|s-x|}\left[\mathbf{I}_d - \frac{(s-x)\otimes(s-x)}{|s-x|^2}\right] P(dx).$$

Moreover, F'_P is uniformly continuous in $\mathbb{R}^d$, and $F'_P(s)$ is positively definite for all $s \in \mathbb{R}^d$.

All asymptotic results below are based on limit theorems for empirical M-distribution and M-quantile functions, obtained in Section 5.

4.1. *L-estimators.* In the one-dimensional case, L-estimators (also called L-statistics), based on a sample $(X_1, \dots, X_n)$, are defined as linear combinations of order statistics $X_{[1]} \leq X_{[2]} \leq \cdots \leq X_{[n]}$, or, more generally, $L_n := \sum_1^n m_{n,j} h(X_{[j]})$ where $m_{n,j}$ are certain weights, and h is a transformation function. It is hard to extend this definition to the multivariate case, since there is no really natural ordering in $\mathbb{R}^d$, $d \geq 1$. But in many cases an L-estimator could be represented as a functional of the empirical measure, $L_n = L(P_n)$, where the functional $L(P)$ is defined as $L(P) := \int_0^1 h(F_P^{-1}(t))m(t)\,dt$ with a certain weight function m. Here F_P is the distribution function of P, and F_P^{-1} is its quantile function. It is natural to call $L(P)$ an L-parameter of P. [Ch] suggested using a similar representation to define L-parameters and L-estimators in the multivariate case. His suggestion was to replace F_P^{-1} for the inverse of the map F_P, defined by (4.1). In fact, any M-quantile function considered in previous sections could be used as well, so one could relate L-parameters and L-estimators to any M-parameter defined by convex minimization. To be specific, we assume that F_P is defined by (4.1). In this case

$$L(P) := L_{h,\mu}(P) := \int_{B_d} h(F_P^{-1}(t))\mu(dt),$$

where μ is a signed measure on B_d with finite total variation, and $h\colon \mathbb{R}^d \mapsto \mathbb{R}^m$ is a vector-valued Borel function, such that the integral in the right-hand side is well defined.

For instance, one could be interested in the uniform distribution μ on $B(0, r)$ for a certain value $0 < r < 1$. Such an L-parameter could be viewed as a multivariate extension of the trimmed mean. Similarly, linearly trimmed means could be considered.

THEOREM 4.2. *Suppose that we have the following conditions*:

(i) *P has a bounded density in $\mathbb{R}^d$, $d \geq 2$;*
(ii) *μ is supported in a ball $B(0, r)$ with some $0 < r < 1$;*
(iii) *h is continuously differentiable in a neighborhood of $F_P^{-1}(B(0, r))$.*

Then an L-estimator $L_n := \int_{B_d} h(F_{P_n}^{-1}(t))\mu(dt)$ is asymptotically normal. The limit distribution of the sequence $n^{1/2}(L_n - L(P))$ is that of the random vector $-\int_{\mathbb{R}^d} h'(s)\,\mathrm{inv}(F'_P(s))\xi_P(s)(\mu \circ F_P)(ds)$. Moreover, the following asymptotic representation holds:

$$L_n = L(P) - \int_{\mathbb{R}^d} h'(s)\,\mathrm{inv}(F'_P(s))(F_{P_n}(s) - F_P(s))(\mu \circ F_P)(ds) + o_p(n^{-1/2}) \quad \text{as } n \to \infty.$$

The proof follows directly from Theorem 5.7. Assume that $h(s) \equiv s$, so that $L(P) := \int_{B_d} F_P^{-1}(t)\mu(dt)$. Assume also that $\mu(B_d) = 1$. Let P be a symmetric distribution with center $a \in \mathbb{R}^d$ in the sense that the random vectors $X - a$ and $a - X$ are identically distributed, given X has the distribution P. If $\mu(A) = \mu(-A)$ for any Borel subset A of B_d, then it is easy to show that $L(P) = a$. In particular, if P is spherically symmetric about $a \in \mathbb{R}^d$, then $L(P) = a$. Thus L-statistics described above could be used to estimate the center of symmetry of the distribution P.

In the one-dimensional case, μ is often absolutely continuous on $[0, 1]$: $\mu(dt) = m(t)\,dt$. Note that in this case, if P is nonatomic, then $P \circ F_P^{-1}$ is the uniform distribution on $[0, 1]$. It means that, in fact, an L-parameter can be expressed as

$$L(P) = \int_{[0,1]} h(F_P^{-1}(t))m(t)(P \circ F_P^{-1})(dt).$$

Koltchinskii (1995a) extended this to the multivariate case, defining

$$L(P) := \int_{B_d} h(F_P^{-1}(t))m(t)(P \circ F_P^{-1})(dt) = \int_{\mathbb{R}^d} h(s)m(F_P(s))P(ds),$$

and studied the asymptotic properties of the corresponding L-estimators.

4.2. *Statistical tests.* Suppose that F_P is an M-distribution function, defined by (4.1). Let π be a Borel probability on $\mathbb{R}^d$. In order to test the hypothesis $P = \pi$ against the alternative $P \neq \pi$ one could use the test statistics $D_n := n^{1/2}\|F_{P_n} - F_\pi\|_{\mathbb{R}^d}$ or $W_n^2 := n\int_{\mathbb{R}^d} |F_{P_n}(s) - F_\pi(s)|^2\pi(ds)$, which could be viewed as multivariate extensions of Kolmogorov's and ω^2-tests, respectively. If π is nonatomic, then Theorem 5.6 implies that the limit distribution of D_n

coincides with the distribution of $\|\xi_\pi\|_{\mathbb{R}^d}$, while the limit distribution of W_n^2 is that of $\int|\xi_\pi(s)|^2\pi(ds)$. One could use bootstrap [which is justified in view of the theorem of Giné and Zinn (1990)] in order to determine the rejection region.

Suppose now we have a sample $(X_1,\dots,X_n)$ from a probability distribution P in $\mathbb{R}^d$, and our goal is to test the hypothesis that P is spherically symmetric. Many authors suggested different tests for symmetry [see, for example, Arcones and Giné (1991) for some tests in the one-dimensional case and Beran (1979), Baringhaus (1991) for the multivariate case].

If P is spherically symmetric and $a \in \mathbb{R}^d$ is the center of symmetry, then, by Proposition 3.6, we have the following representation of the M-distribution function F_P:

$$F_P(s) = \begin{cases} \psi_P(|s-a|)\dfrac{s-a}{|s-a|}, & \text{for } s \neq a,\\ 0, & \text{for } s = a,\end{cases}$$

where $\psi_P(\lambda) := \psi_P(a;\lambda)$ is defined by

$$\psi_P(\lambda) := \int_{\mathbb{R}^d} \frac{\lambda + a^{(1)} - x^{(1)}}{\sqrt{(\lambda+a^{(1)}-x^{(1)})^2 + (a^{(2)}-x^{(2)})^2 + \cdots + (a^{(d)}-x^{(d)})^2}}\, P(dx).$$

It's easy to see that $\psi_P(a;\lambda) := \langle F_P(\lambda e_1 + a), e_1\rangle$, where e_1 is the first vector of the canonical basis of $\mathbb{R}^d$.

For an arbitrary probability distribution P in $\mathbb{R}^d$, define

$$\tilde{\psi}_P(\lambda) := \int_{S^{d-1}} \langle F_P(\lambda v + F_P^{-1}(0)), v\rangle m(dv),$$

where m denotes the uniform distribution on S^{d-1}. Note that for a spherically symmetric P $\tilde{\psi}_P \equiv \psi_P$.

We introduce a measure of asymmetry of P as follows:

$$\gamma(P) := \sup_{s\in\mathbb{R}^d}\left| F_P(s + F_P^{-1}(0)) - \tilde{\psi}_P(|s|)\frac{s}{|s|}\right|.$$

It's easy to see that the functional $\gamma(P)$ is invariant with respect to all orthogonal transformations of $\mathbb{R}^d$ [i.e., $\gamma(P\circ O^{-1}) = \gamma(P)$], and $\gamma(P) = 0$ iff P is spherically symmetric. Therefore, a test for spherical symmetry of P can be based on the measure of asymmetry $\gamma(P_n)$ of the empirical distribution.

THEOREM 4.3. *If P has a bounded spherically symmetric density in $\mathbb{R}^d$ with $d \geq 2$, then the sequence of distributions of $n^{1/2}\gamma(P_n)$ converges weakly to the distribution of the random variable $\|\delta_P\|_{\mathbb{R}^d}$, where*

$$\delta_P(s) := \begin{cases} \zeta_P(s) - \displaystyle\int_{S^{d-1}} \langle \xi_P(|s|v + F_P^{-1}(0)), v\rangle m(dv)\frac{s}{|s|}, & \textit{for } s \neq 0,\\ 0, & \textit{for } s = 0,\end{cases}$$

$$\zeta_P(s) := \xi_P(s + F_P^{-1}(0)) - F'_P(s + F_P^{-1}(0))\,\mathrm{inv}(F'_P(F_P^{-1}(0)))\xi_P(F_P^{-1}(0)).$$

If P is not spherically symmetric, then for all $t \geq 0$,

$$\lim_{n\to\infty} \Pr\{n^{1/2}\gamma(P_n) \leq t\} = 0.$$

We describe a bootstrap version of this test. Any spherically symmetric distribution P is completely characterized by its center $a \in \mathbb{R}^d$ and the distribution π of the random variable $|X-a|$, given X with distribution P. We call (a, π) parameters of P. Given a sample $(X_1, \dots, X_n)$ from the unknown distribution P in $\mathbb{R}^d$, denote by π_n the empirical distribution based on the sample $(|X_1 - F_{P_n}^{-1}(0)|, \dots, |X_n - F_{P_n}^{-1}(0)|)$. Let P_n^s be the spherically symmetric distribution in $\mathbb{R}^d$ with parameters $(F_{P_n}^{-1}(0), \pi_n)$. Define on a probability space $(\Omega_s, \Sigma_s, \Pr_s)$ an i.i.d. sample $(\hat{X}_1, \dots, \hat{X}_n)$ from the distribution P_n^s. Let $\hat{P}_n$ be the empirical distribution of this sample. It could be shown that if P has a bounded density in $\mathbb{R}^d$ with $d \geq 2$, then the sequence of distributions (with respect to $\Pr_s$) of the random variables $n^{1/2}\gamma(\hat{P}_n)$ converges weakly to the distribution of $\|\delta_{P^s}\|_{\mathbb{R}^d}$ in probability Pr. Here P^s denotes the spherically symmetric distribution in $\mathbb{R}^d$ with parameters $(F_P^{-1}(0), \pi)$, where π is the distribution of the random variable $|X - F_P^{-1}(0)|$, given X with distribution P. In particular, if P is spherically symmetric, then $P^s = P$. This allows us to construct a bootstrap test for spherical symmetry, based on the test statistic $n^{1/2}\gamma(P_n)$. The proofs go beyond this paper [see Koltchinskii and Lang Li (1996)].

4.3. *Regression quantiles.* Let (X, Y) be a vector in the space $\mathbb{R}^m \times \mathbb{R}^d = \mathbb{R}^{m+d}$, satisfying the following model:

$$Y = S_0 X + \xi, \tag{4.2}$$

where random vectors X and ξ are independent, and S_0 is a linear operator from $\mathbb{R}^m$ into $\mathbb{R}^d$. In the case $d = 1$, of course, $S_0 x = \langle s_0, x\rangle$ for some $s_0 \in \mathbb{R}^m$.

Denote by R the distribution of the random vector (X, Y), and let R_X, R_Y, P be the distributions of X, Y, ξ, respectively.

The space of all linear operators from $\mathbb{R}^m$ into $\mathbb{R}^d$ could be identified with $\mathbb{R}^{md}$ (if $u_i,\ i = 1, \dots, d$ is the canonical basis in $\mathbb{R}^d$ and $v_j,\ j = 1, \dots, m$ is the canonical basis in $\mathbb{R}^m$, then we take $\{u_i \otimes v_j,\ i = 1, \dots, d,\ j = 1, \dots, m\}$ as a basis in $\mathbb{R}^{md}$).

We introduce a kernel $g\colon \mathbb{R}^{md} \times \mathbb{R}^m \times \mathbb{R}^d \mapsto \mathbb{R}^1$, $g(S; x, y) := |Sx - y| - |y|$. Clearly, for all $x \in \mathbb{R}^m$, $y \in \mathbb{R}^d$ g is convex in $S \in \mathbb{R}^{md}$. Assuming that $\mathbb{E}|X| = \int_{\mathbb{R}^m} |x| R_X(dx) < +\infty$, define a convex function $g_R\colon \mathbb{R}^{md} \mapsto \mathbb{R}^1$,

$$g_R(S) := \int_{\mathbb{R}^m \times \mathbb{R}^d} g(S; x, y) R(dx, dy).$$

For $t \in B_d$, we define *a regression t-quantile* as a minimal point of the following functional in $S \in \mathbb{R}^{md}$:

$$\mathbb{E}(|SX - Y| - |Y| - \langle SX, t\rangle) = g_R(S) - \langle S, t \otimes \mathbb{E}X\rangle.$$

This is an extension of the original definition of Koenker and Basset (1978) for $d = 1$.

A subgradient $G_R\colon \mathbb{R}^{md} \mapsto \mathbb{R}^{md}$ of the function g_R is given by:

$$G_R(S) := \int_{\mathbb{R}^{m+d}} G(S; x, y)\ R(dx, dy),$$

where $G\colon \mathbb{R}^{md} \times \mathbb{R}^m \times \mathbb{R}^d$ is defined by

$$G(S; x, y) := \begin{cases} \dfrac{Sx - y}{|Sx - y|} \otimes x, & \text{for } Sx \neq y, \\ 0, & \text{otherwise.} \end{cases}$$

Under condition (4.2),

$$G_R(S) = \mathbb{E}\frac{(S - S_0)X - \xi}{|(S - S_0)X - \xi|} \otimes X = \mathbb{E}F_P((S - S_0)X) \otimes X. \tag{4.3}$$

Calculations show that

$$\partial g(S; x, y) = \begin{cases} \left\{ \dfrac{Sx - y}{|Sx - y|} \otimes x \right\}, & \text{if } Sx \neq y, \\ \{u \otimes x\colon |u| \leq 1\}, & \text{otherwise,} \end{cases}$$

which, in view of Theorem 2.1, implies

$$\partial g_R(S) = G_R(s) + \left\{ u \otimes \int_{\{(x, y)\colon\ Sx = y\}} xR(dx, dy)\colon |u| \leq 1 \right\}.$$

We need some properties of g_R and G_R (most of them follow from the general properties of M-distribution and M-quantile functions; see Section 2). Relationship (4.4) below follows from (4.3) and Proposition 4.1.

Denote

$$B_R := \bigcap_{V \in \mathbb{R}^{md},\, V \neq 0} \{T\colon \langle V, T \rangle < \mathbb{E}|VX|\}.$$

Then B_R is an open convex subset of the ball in $\mathbb{R}^{md}$ with center 0 and radius $E|X|$.

We introduce the following conditions:

1. X has an absolutely continuous distribution with $\mathbb{E}|X|^2 < +\infty$;
2. P has a bounded density p in $\mathbb{R}^d$;
3. if $d = 1$, assume, in addition, that p is continuous and $R_X(\{x\colon p(\langle s, x \rangle) > 0\}) > 0$ for all $s \in \mathbb{R}^m$.

Given $S_1, S_2 \in \mathbb{R}^{md}$ with $S_1 \neq S_2$, denote

$$\mathscr{L}(S_1, S_2; x) := \{S_1 x + \lambda(S_2 - S_1)x\colon \lambda \in \mathbb{R}^1\} \quad \text{for } x \in \mathbb{R}^m.$$

PROPOSITION 4.4. *For all $T \in B_R$ there exists an (M, T)-parameter of R. If $d > 1$ and for all $S_1, S_2 \in \mathbb{R}^{md}$ with $S_1 \neq S_2$ we have $R(\{(x, y)\colon y \notin \mathscr{L}(S_1, S_2; x)\}) > 0$, then g_R is strictly convex and the (M, T)-parameter is unique for all $T \in B_R$. If conditions (1)–(3) hold, then g_R is a strictly convex continuously differentiable function with gradient G_R, which is a strictly* *monotone one-to-one map from* $\mathbb{R}^{md}$ *onto the open convex set* B_R. *Moreover*, G_R *is a homeomorphism of* $\mathbb{R}^{md}$ *and* B_R, *and it is continuously differentiable in* $\mathbb{R}^{md}$ *with nonsingular and strictly positive derivative*

$$G'_R(S) = \mathbb{E}F'_P\big((S - S_0)X\big) \otimes X \otimes X. \tag{4.4}$$

For all $T \in B_R$, *the unique* (M, T)*-parameter of* R *coincides with* $G_R^{-1}(T)$.

Since for $t \in B_d$,

$$\langle V, t \otimes \mathbb{E}X\rangle = \langle t, \mathbb{E}VX\rangle < \mathbb{E}|VX|,$$

we have $t \otimes \mathbb{E}X \in B_R$, and it is clear that under conditions (1)–(3) the regression t-quantile exists, is unique and coincides with $G_R^{-1}(t \otimes \mathbb{E}X)$.

For $S = S_0$, we have $G_R(S_0) = F_P(0) \otimes \mathbb{E}X$, which implies $S_0 = G_R^{-1}(F_P(0) \otimes \mathbb{E}X)$. If either $F_P(0) = 0$ (which means that the spatial median of ξ is 0), or $\mathbb{E}X = 0$, then $S_0 = G_R^{-1}(0)$.

Consider a sample (X_i, Y_i), $i = 1, \dots, n$ from the distribution R, and let R_n be the empirical distribution based on this sample. Denote $\bar{X}_n := (X_1 + \cdots + X_n)/n$. Since for $t \in B_d$,

$$\langle V, t \otimes \bar{X}_n\rangle = \langle t, n^{-1}(VX_1 + \cdots + VX_n)\rangle < n^{-1}(|VX_1| + \cdots + |VX_n|),$$

we have

$$t \otimes \bar{X}_n \in B_{R_n} := \bigcap_{V \in \mathbb{R}^{md},\, V \neq 0} \big\{T\colon \langle V, T\rangle < n^{-1}\big(|VX_1| + \cdots + |VX_n|\big)\big\}.$$

Thus, for all $t \in B_d$, the empirical regression t-quantiles exist. Moreover, for $d > 1$, under condition (2) we have $R_n\big(\{(x, y)\colon y \in \mathscr{L}(S_1, S_2; x)\}\big) = 0$ for all $S_1, S_2 \in \mathbb{R}^{md}$ with $S_1 \neq S_2$ a.s. Thus, due to Proposition 4.4, for all $t \in B_d$ the empirical regression t-quantile exists, is unique and coincides with $G_{R_n}^{-1}(t \otimes \bar{X}_n)$.

Let $Q\colon B_d \mapsto \mathbb{R}^{md}$ be the regression quantile function [which is uniquely defined under the conditions (1)–(3)], and let $Q_n\colon B_d \mapsto \mathbb{R}^{md}$ be a version of the empirical regression quantile function [which always exists and, for $d > 1$, under conditions (1), (2) is a.s. unique]. Define

$$A(t) := \operatorname{inv}(G'_R \circ G_R^{-1})(t \otimes \mathbb{E}X) \quad \text{for } t \in B_d.$$

THEOREM 4.5. *If conditions* (1)–(3) *hold*, *then the sequence of empirical regression quantile processes* $\{n^{1/2}(Q_n - Q)(t)\}_{n\geq 1}$ *converges weakly locally in* B_d *to the Gaussian process*

$$A(t)\int_{R^{m+d}} \big(t \otimes x - G(Q(t); x, y)\big) W_R^{\circ}(dx, dy), \quad t \in B_d.$$

Moreover, *locally uniformly in* $t \in B_d$

$$Q_n(t) = Q(t) + A(t)(t \otimes (\bar{X}_n - \mathbb{E}X)) - A(t)(G_{R_n} - G_R) \circ Q(t) + o_p(n^{-1/2}). \tag{4.5}$$

It is worth mentioning that the definition of the regression quantiles and their asymptotic theory can be easily extended to the case of a discrepancy function $g(S; x, y) := h(Sx - y) - h(-y)$ with an arbitrary convex h: $\mathbb{R}^d \mapsto \mathbb{R}^1$. This extension is also based on the theory of M-quantiles developed in this paper. See Bai, Rao and Wu (1992) for the asymptotic properties of M-estimators of linear regression parameters under such a discrepancy function.

5. Asymptotics of empirical M-distribution and M-quantile functions. First we consider asymptotics of general empirical M-quantile functions at a given point. Let f: $\mathbb{R}^d \times \mathbf{X} \mapsto \mathbb{R}^1$ be a P-convex kernel. Assume that it is P-integrable for all $s \in \mathbb{R}^d$. Consider a convex function f_P, defined by

$$f_P(s) := \int_{\mathbf{X}} f(s, x) P(dx) \quad \text{for } s \in \mathbb{R}^d.$$

Let F: $\mathbb{R}^d \times \mathbf{X} \mapsto \mathbb{R}^d$ be such that $F(s, x) \in \partial f(s, x)$ for all $s \in \mathbb{R}^d$ and P-almost all $x \in \mathbf{X}$. Then

$$F_P(s) := \int_{\mathbf{X}} F(s, x) P(dx) \quad \text{for } s \in \mathbb{R}^d$$

is a subgradient of f_P (see Theorem 2.1).

Let $\{C_n\}_{n \geq 1}$ be a sequence of random sets $\omega \mapsto C_n(\omega) \subset \mathbb{R}^d$, and let η_n be a sequence of random vectors in $\mathbb{R}^d$. Given a sequence $\{\delta_n\}_{n \geq 1}$ of positive numbers, we write

$$C_n = \eta_n + o_p(\delta_n) \quad \text{as } n \to \infty$$

iff for all $\varepsilon > 0$,

$$\limsup_{n \to \infty} \Pr^*(\{C_n \not\subset B(\eta_n; \varepsilon\delta_n)\}) = 0.$$

THEOREM 5.1. *Let $t_0 \in \mathbb{R}^d$. Suppose that we have the following*:

(i) *P has the unique (M, t_0)-parameter with respect to f (which means that f_P^* is differentiable at the point t_0). We denote it by s_0 [then $F_P(s_0) = t_0$];*

(ii) *$\int_{\mathbf{X}} |F(s, x)|^2 P(dx) < +\infty$ for all s in a neighborhood of s_0;*

(iii) *F_P is continuously differentiable at the point s_0 and the derivative $F'_P(s_0)$ is a positively definite operator in $\mathbb{R}^d$.*

Then the following representation holds for empirical (M, t_0)-quantiles

$$\partial f_{P_n}^{-1}(t_0) = s_0 - \operatorname{inv}(F'_P(s_0))(F_{P_n}(s_0) - F_P(s_0)) + o_p(n^{-1/2}) \quad \text{as } n \to \infty.$$

If s_n is an (M, t_0)-quantile of P_n, then $n^{1/2}(s_n - s_0)$ is asymptotically normal. Its limit distribution is that of the Gaussian random vector $-\operatorname{inv}(F'_P(s_0)) \int_{\mathbf{X}} F(s_0, x) W_P^{\circ}(dx)$.

The following corollary is a result of Haberman (1989) [see also Niemiro (1992)].

COROLLARY 5.2. *Suppose that P has the unique M-parameter s_0 with respect to f (which means that f_P^* is differentiable at point* 0) *and conditions* (ii) *and* (iii) *of Theorem* 5.1 *hold. Then the following representation holds for any M-estimator s_n*:

$$s_n = s_0 - \operatorname{inv}(F'_P(s_0))F_{P_n}(s_0) + o_p(n^{-1/2}) \quad \text{as } n \to \infty.$$

In particular, $n^{1/2}(s_n - s_0)$ is asymptotically normal. Its limit distribution is that of the Gaussian random vector $-\operatorname{inv}(F'_P(s_0)) \int_{\mathbf{X}} F(s_0, x) W_P^\circ(dx)$.

We consider now uniform asymptotic results for empirical M-distribution and M-quantile functions. To do this, we need some facts on empirical processes and on weak convergence (see the Appendix).

Suppose that F: $\mathbb{R}^d \times \mathbf{X} \mapsto \mathbb{R}^d$ is a function, such that for all $s \in \mathbb{R}^d$ and for all $1 \le j \le d$ the jth component $F^{(j)}(s;\cdot)$ of $F(s;\cdot)$ is P-square integrable: $F^{(j)}(s;\cdot) \in L_2(\mathbf{X}; dP)$. Consider

$$F_P(s) := \int_{\mathbf{X}} F(s;x)P(dx)$$

and

$$F_{P_n}(s) := \int_{\mathbf{X}} F(s;x)P_n(dx) = n^{-1}\sum_1^n F(s;X_k) \quad \text{for } s \in \mathbb{R}^d.$$

We call F *P-Glivenko-Cantelli* iff the class of functions $\mathscr{F} := \bigcup_{j=1}^n \{F^{(j)}(s;\cdot) : s \in \mathbb{R}^d\} \in GC(P)$. We call F *P-Donsker* iff, for any $1 \le j \le d$, $F^{(j)}$ is ρ_P-continuous on $\mathbb{R}^d$ and $\mathscr{F} \in CLT(P)$.

Let ϕ be a nondecreasing continuous function from $[0, +\infty)$ into itself, and let $\psi(s) := \phi(|s|)$, $s \in \mathbb{R}^d$. The following facts follow from the properties of empirical processes (see the Appendix and further references there).

THEOREM 5.3. *If F/ψ is P-Glivenko–Cantelli, then a.s. $\|F_{P_n} - F_P\|_{\mathbb{R}^d, \psi} \to 0$ as $n \to \infty$. In particular, a.s. for all compacts $K \subset \mathbb{R}^d$, $\|F_{P_n} - F_P\|_K \to 0$ as $n \to \infty$. If ψ is bounded and F is P-Glivenko–Cantelli, then a.s. $\|F_{P_n} - F_P\|_{\mathbb{R}^d} \to 0$ as $n \to \infty$. If F/ψ is P-Donsker, then*

$$n^{1/2}(F_{P_n} - F_P) \to_w \xi_P \text{ in } l_\psi^\infty(\mathbb{R}^d) \quad \text{as } n \to \infty,$$

where $\xi_P(s) := \int_{\mathbf{X}} F(s,x) W_P^\circ(dx)$ for $s \in \mathbb{R}^d$. In particular,

$$n^{1/2}(F_{P_n} - F_P) \to_{w,\,\mathrm{loc}} \xi_P \quad \text{as } n \to \infty.$$

If ψ is bounded and F is P-Donsker, then

$$n^{1/2}(F_{P_n} - F_P) \to_w \xi_P \quad \text{as } n \to \infty.$$

Let now f: $\mathbb{R}^d \times \mathbf{X} \mapsto \mathbb{R}^1$ be a P-strictly convex kernel. Assume that it is P-integrable for all $s \in \mathbb{R}^d$. Consider a convex function f_P, defined by

$$f_P(s) := \int_{\mathbf{X}} f(s,x)P(dx) \quad \text{for } s \in \mathbb{R}^d.$$

Let $F: \mathbb{R}^d \times \mathbf{X} \mapsto \mathbb{R}^d$ be such that $F(s, x) \in \partial f(s, x)$ for all $s \in \mathbb{R}^d$ and P-almost all $x \in \mathbf{X}$. Then

$$F_P(s) := \int_{\mathbf{X}} F(s, x)P(dx) \quad \text{for } s \in \mathbb{R}^d$$

is a subgradient of f_P (see Theorem 2.1). Let $\delta_f(s; x)$ be the diameter of the bounded convex set $\partial f(s, x) : \delta_f(s; x) := \mathrm{Diam}(\partial f(s; x))$.

Denote $Q := F_P^{-1} = (\nabla f_P)^{-1}: U_P \mapsto D_P$ [see properties (5), (6) in Section 2.1].

THEOREM 5.4. *Let S be an open subset of $\mathbb{R}^d$ and let $T := F_P(S)$. Suppose the following*:

(i) *F_P is continuously differentiable in S with nonsingular derivative $F'_P(s)$, $s \in S$;*
(ii) *for some $c > 0$, either $\psi \equiv c$, or for all $s \in \mathbb{R}^d$ with $|s|$ sufficiently large, $F_P(s) \geq c\psi(s)$;*
(iii) *F/ψ is P-Donsker*;
(iv) $\sup_{s\in\mathbb{R}^d} \int_{\mathbf{X}} \delta_f(s; x)P_n(dx) = o_P(n^{-1/2})$ *as $n \to \infty$;*
(v) *for any compact set $K \subset B_P$* $\limsup_{n\to\infty} \Pr^*(\{K \not\subset B_{P_n}\}) = 0$.

Then for any $Q_n: \Omega \times T \mapsto \mathbb{R}^d$, such that $Q_n(t) \in \partial f_{P_n}^{-1}(t)$ for $t \in T$ [so that $Q_n(t)$ is an (M, t)-estimator], we have

$$n^{1/2}(Q_n - Q) \to_{w,\,\mathrm{loc}} \eta_P \quad \text{in } S \text{ as } n \to \infty,$$

where $\eta_P := -\,\mathrm{inv}(F'_P \circ Q)(\xi_P \circ Q)$. *Moreover,*

$$n^{1/2}(Q_n - Q + \mathrm{inv}(F'_P \circ Q)(F_{P_n} - F_P) \circ Q) \to_{p,\,\mathrm{loc}} 0 \quad \text{in } S \text{ as } n \to \infty.$$

Now we consider in some detail asymptotics of empirical spatial distributions and quantiles. Let φ be a convex increasing function on $[0, +\infty)$, continuously differentiable in $(0, +\infty)$. Let $\Phi := \varphi'$. Suppose that for some $C > 0$,

$$\Phi(\lambda_1 + \lambda_2) \leq C[\Phi(\lambda_1) + \Phi(\lambda_2)] \quad \text{for } \lambda_1,\ \lambda_2 > 0 \tag{5.1}$$

and for all $c > 0$,

$$\frac{\Phi(\lambda + c)}{\Phi(\lambda)} \to 1 \quad \text{as } \lambda \to \infty. \tag{5.2}$$

Denote $\psi(s) := \Phi(|s|) \bigvee 1$ for $s \in \mathbb{R}^d$. We consider an M-distribution function

$$F_P(s) := \int_{\{x\neq s\}} \Phi(|s - x|)\frac{s - x}{|s - x|}P(dx) \quad \text{for } s \in \mathbb{R}^d,$$

which is a subgradient of the convex function

$$f_P(s) := \int_{\mathbb{R}^d} \big(\varphi(|s - x|) - \varphi(|x|)\big)P(dx) \quad \text{for } s \in \mathbb{R}^d.$$

THEOREM 5.5. *If*

$$\int_{\mathbb{R}^d} \Phi(|x|)P(dx) < +\infty, \tag{5.3}$$

then a.s.

$$\|F_{P_n} - F_P\|_{\mathbb{R}^d,\psi} \to 0 \quad \textit{as } n \to \infty. \tag{5.4}$$

In particular, a.s. for all compacts $K \subset \mathbb{R}^d$,

$$\|F_{P_n} - F_P\|_K \to 0 \quad \textit{as } n \to \infty. \tag{5.5}$$

If $\Phi(+\infty) < +\infty$, *then a.s.*

$$\|F_{P_n} - F_P\|_{\mathbb{R}^d} \to 0 \quad \textit{as } n \to \infty. \tag{5.6}$$

THEOREM 5.6. *If* P *is nonatomic and*

$$\int_{\mathbb{R}^d} \Phi^2(|x|)P(dx) < +\infty, \tag{5.7}$$

then

$$n^{1/2}(F_{P_n} - F_P) \to_w \xi_P \quad \textit{in } l_\psi^\infty(\mathbb{R}^d) \textit{ as } n \to \infty, \tag{5.8}$$

where

$$\xi_P(s) := \int_{\{x \neq s\}} \Phi(|s-x|)\frac{s-x}{|s-x|} W_P^\circ(dx) \quad \textit{for } s \in \mathbb{R}^d.$$

In particular,

$$n^{1/2}(F_{P_n} - F_P) \to_{w,\,\mathrm{loc}} \xi_P \quad \textit{as } n \to \infty. \tag{5.9}$$

If $\Phi(+\infty) < +\infty$, *then*

$$n^{1/2}(F_{P_n} - F_P) \to_w \xi_P \quad \textit{as } n \to \infty. \tag{5.10}$$

Suppose that P is nonatomic and it is not concentrated in a line. Then, by Proposition 2.6, F_P^{-1} is a well defined map from $B(0; R_\varphi)$ into $\mathbb{R}^d$. Denote $Q := F_P^{-1}$.

THEOREM 5.7. *Let* S *be an open subset of* $\mathbb{R}^d$ *and let* $T := F_P(S)$. *If* F_P *is continuously differentiable in* S *with nonsingular derivative* $F'_P(s)$, $s \in S$, *and if condition* (5.7) *holds, then, for any* Q_n *from* $\Omega \times T$ *into* $\mathbb{R}^d$, *such that* $Q_n(t) \in \partial f_{P_n}^{-1}(t)$ *for all* $t \in T$ [*so that* $Q_n(t)$ *is an* (M, t)-*parameter of* P],

$$n^{1/2}(Q_n - Q) \to_{w,\,\mathrm{loc}} \eta_P \quad \textit{in } S \textit{ as } n \to \infty,$$

where $\eta_P := -\operatorname{inv}(F'_P \circ F_P^{-1})(\xi_P \circ F_P^{-1})$. *Moreover,*

$$n^{1/2}(Q_n - Q + \operatorname{inv}(F'_P \circ F_P^{-1})(F_{P_n} - F_P) \circ Q) \to_{p,\,\mathrm{loc}} 0 \quad \textit{in } S \textit{ as } n \to \infty.$$

REMARKS. Since P is nonatomic and it is not concentrated in a line, P_n is not concentrated in a line either with probability tending to 1 as $n \to \infty$. By Proposition 2.6, $F_{P_n}^{-1}$ is well defined on $B(0; R_\varphi)$ and $Q_n \equiv F_{P_n}^{-1}$ with probability tending to 1 as $n \to \infty$. If, moreover, $P(L) = 0$ for all straight lines L in $\mathbb{R}^d$, then P_n is not concentrated in a line and $Q_n \equiv F_{P_n}^{-1}$ for almost all samples $(X_1, \dots, X_n)$ from P with size $n > 2$.

Note that the proof of Theorem 5.7 (see Theorems 5.4 and A.4) includes, in fact, a differentiation of the map $R \mapsto F_R^{-1}$. Suppose that P is nonatomic. Given $R \in \mathscr{P}(\mathbb{R}^d)$, denote $P_\varepsilon := (1 - \varepsilon)P + \varepsilon R$, where $\varepsilon \in [0, 1]$. Let $t \in B(0; R_\varphi)$ be a point such that F_P is continuously differentiable in a neighborhood of the point $F_P^{-1}(t)$ with nonsingular derivative. Then

$$\frac{F_{P_\varepsilon}^{-1}(t) - F_P^{-1}(t)}{\varepsilon} \to -\operatorname{inv}(F'_P \circ F_P^{-1}(t))(F_P - F_R) \circ F_P^{-1}(t) \quad \text{as } \varepsilon \to 0.$$

It follows, in particular, that the influence curve of the functional $P \mapsto F_P^{-1}(t)$ is given by the following expression:

$$IC(x; P; F_P^{-1}(t)) = -\operatorname{inv}(F'_P \circ F_P^{-1}(t))(F_P - F_{\delta_x}) \circ F_P^{-1}(t)$$
$$\text{for all } t \neq F_P(x),\ x \in \mathbb{R}^d.$$

Koltchinskii (1994a, b) studied Bahadur–Kiefer representations of empirical spatial quantiles.

APPENDIX

Notation. $\langle \cdot, \cdot \rangle$ denotes the canonical inner product in the finite-dimensional space $\mathbb{R}^d$; $|\cdot| = \langle \cdot, \cdot \rangle^{1/2}$ is the corresponding norm; $\|\cdot\|$ (sometimes with indices) is used for the norms of functions or linear transformations (matrices). If $x \in \mathbb{R}^d$, then $x^{(j)}$ denotes the jth coordinate of x. Similarly, if F is an $\mathbb{R}^d$-valued function, $F^{(j)}$ is its jth coordinate.

We denote the ball in $\mathbb{R}^d$ with center $s \in \mathbb{R}^d$ and radius $R > 0$ by $B(s; R)$. In particular, $B_d := B(0; 1)$. S^{d-1} denotes the unit sphere in $\mathbb{R}^d$: $S^{d-1} := \{s \in \mathbb{R}^d\colon |s| = 1\}$. Given a set $A \subset \mathbb{R}^d$, $\bar{A}$ denotes its closure.

We use the notation $\mathscr{L}(\mathbb{R}^d)$ for the space of all linear transformations of $\mathbb{R}^d$. Given two vectors $u \in \mathbb{R}^d$ and $v \in \mathbb{R}^m$, $u \otimes v$ denotes the linear transformation $x \mapsto \langle v, x \rangle u$ from $\mathbb{R}^m$ into $\mathbb{R}^d$.

Regularly, we denote by ∇f the gradient of a function f on an open subset of $\mathbb{R}^d$, and by F' the derivative of a vector-valued map F from an open subset of $\mathbb{R}^d$ into $\mathbb{R}^d$. So $\nabla f(s) \in \mathbb{R}^d$ and $F'(s) \in \mathscr{L}(\mathbb{R}^d)$.

Given an invertible $A \in \mathscr{L}(\mathbb{R}^d)$, the notation $\operatorname{inv}(A)$ will be used for its inverse. Given an $\mathscr{L}(\mathbb{R}^d)$-valued function Ψ, defined on a subset of $\mathbb{R}^d$, $\operatorname{inv}(\Psi)$ denotes the function $s \mapsto \operatorname{inv}(\Psi(s))$ (assuming, of course, that $\Psi(s)$ is invertible). If ζ is a function from a subset of $\mathbb{R}^d$ into $\mathbb{R}^d$, and if Ψ is an $\mathscr{L}(\mathbb{R}^d)$-valued function defined on the domain of ζ, then the "product" $\Psi\zeta$ is the function $s \mapsto \Psi(s)\zeta(s)$.

We denote by $(\Omega, \Sigma, \mathrm{Pr})$ the main probability space which carries all random elements under consideration; $\mathbb{E}$ denotes the expectation with respect to Pr. Define

$$\mathrm{Pr}^*(A) := \inf\{\mathrm{Pr}(B): B \in \Sigma,\ B \supset A\},$$
$$\mathbb{E}^*\xi := \inf\{\mathbb{E}\eta: \eta - \Sigma\text{-measurable},\ \eta \geq \xi\}.$$

Given a measurable space $(\mathbf{X}, \mathscr{A})$, $\mathscr{P}(\mathbf{X})$ denotes the set of all probability measures on $\mathbf{X}$.

Facts on convex analysis. The function $f: \mathbb{R}^d \mapsto \mathbb{R}^1$ is called *convex* iff $f(\alpha s_1 + (1-\alpha)s_2) \leq \alpha f(s_1) + (1-\alpha) f(s_2)$ for all $s_1, s_2 \in \mathbb{R}^d$ and $\alpha \in [0,1]$, and it is called *strictly convex* iff the above inequality is strict for all $s_1 \neq s_2$, $\alpha \in (0,1)$. Any convex function is continuous on $\mathbb{R}^d$ ([Ro], Theorem 10.1).

For $s \in \mathbb{R}^d$ and $v \in \mathbb{R}^d$, $v \neq 0$, denote $\Delta f(s; v; \lambda) := \lambda^{-1}(f(s+\lambda v) - f(s))$. The directional derivative of f at the point s in the direction v is defined by

$$df(s;v) := \inf_{\lambda \in (0, +\infty)} \Delta f(s; v; \lambda) = \lim_{\lambda \downarrow 0} \Delta f(s; v; \lambda).$$

A set $\partial f(s)$ of all points $t \in \mathbb{R}^d$ such that $f(s') \geq f(s) + \langle s' - s, t\rangle$ for all $s' \in \mathbb{R}^d$ is called a *subdifferential* of f at a point $s \in \mathbb{R}^d$. A function f is called subdifferentiable at a point $s \in \mathbb{R}^d$ iff $\partial f(s) \neq \varnothing$. Any point $t \in \partial f(s)$ is called a *subgradient* of f at point s. If f is differentiable at a point s, we denote $\nabla f(s)$ the *gradient* of f at s.

Given a convex function f, define a set $D_f := \{s \in \mathbb{R}^d: df(s;v) \text{ is linear in } v\}$. Then f is differentiable at any point $s \in D_f$ and $\partial f(s) = \{\nabla f(s)\}$, $s \in D_f$. D_f is a dense subset of $\mathbb{R}^d$ and its complement $\mathbb{R}^d \setminus D_f$ has Lebesgue measure 0. The gradient $\nabla f: s \mapsto \nabla f(s)$ is continuous on D_f ([Ro], Theorems 25.2, 25.5).

A map $F: \mathbb{R}^d \mapsto \mathbb{R}^d$ is called a *subgradient* of f iff $F(s) \in \partial f(s)$ for all $s \in \mathbb{R}^d$.

A map G from $\mathbb{R}^d$ into $2^{\mathbb{R}^d}$ is called a *multivalued map*. For instance, $s \mapsto \partial f(s)$ is a multivalued map. The *inverse* of a multivalued map G is the multivalued map G^{-1} such that $G^{-1}(t) = \{s \in \mathbb{R}^d: t \in G(s)\}$ for all $t \in \mathbb{R}^d$.

A multivalued map G is called *single-valued* iff for any $s \in \mathbb{R}^d$ the set $G(s)$ contains only one element. Of course, any such map could be identified with a regular map from $\mathbb{R}^d$ into $\mathbb{R}^d$.

A multivalued map G is called *monotone* iff for any $s_1, s_2 \in \mathbb{R}^d$ and for any $t_1 \in G(s_1)$, $t_2 \in G(s_2)$ $\langle s_2 - s_1, t_2 - t_1\rangle \geq 0$. It is called *strictly monotone* iff for any $s_1, s_2 \in \mathbb{R}^d$ such that $s_1 \neq s_2$ and for any $t_1 \in G(s_1)$, $t_2 \in G(s_2)$, $\langle s_2 - s_1, t_2 - t_1\rangle > 0$.

In particular, a map $F: \mathbb{R}^d \mapsto \mathbb{R}^d$ is called monotone (strictly monotone) iff the multivalued map $\mathbb{R}^d \ni s \mapsto \{F(s)\}$ is monotone (strictly monotone).

A multivalued map G is said to be *upper semicontinuous* at a point $s \in \mathbb{R}^d$ iff for any neighborhood V of the set $G(s)$ there exists a neighborhood U of the point s such that for any $s' \in U$ we have $G(s') \subset V$.

A function $f^*: \mathbb{R}^d \mapsto \mathbb{R}^1 \cup \{+\infty\}$, defined by $f^*(t) := \sup_{s \in \mathbb{R}^d}[\langle s, t\rangle - f(s)]$, is called the *Young–Fenchel conjugate* of f. It is a convex function and $\partial f^* = (\partial f)^{-1}$ ([Ro], Theorem 23.5).

The *recession function* r_f of a convex function f on $\mathbb{R}^d$ is defined as

$$r_f(v) := \sup_{s\in\mathbb{R}^d}[f(s+v) - f(s)] \quad \text{for } v \in \mathbb{R}^d.$$

Empirical processes and weak convergence. Let S be a set, and let $(L, |\cdot|)$ be a separable Banach space. Denote by $l^\infty(S) = l^\infty(S; L)$ the normed space of all uniformly bounded functions $Y\colon S \mapsto L$ with norm

$$\|Y\|_S := \sup_{s\in S}|Y(s)| \quad \text{for } Y \in l^\infty(S).$$

More generally, if ψ is a positive function on S, we define a norm

$$\|Y\|_{S,\psi} := \sup_{s\in S}\frac{|Y(s)|}{\psi(s)}$$

and denote by $l^\infty_\psi(S)$ the normed space of all functions Y with $\|Y\|_{S,\psi} < +\infty$.

Let $\{\xi_n\}_{n\geq 1}$ be a sequence of random functions $\xi_n\colon S \mapsto L$. We say that $\{\xi_n\}_{n\geq 1}$ *converges weakly* in the space $l^\infty(S)$ to a random function $\xi\colon S \mapsto L$, or

$$\xi_n \to_w \xi \quad \text{as } n \to \infty,$$

iff there exists a separable subspace $\mathscr{N} \subset l^\infty(S)$ such that $\xi \in \mathscr{N}$ a.s. and for any bounded continuous functional $\Phi\colon l^\infty(S) \mapsto \mathbb{R}$ we have $\mathbb{E}^*\Phi(\xi_n) \to \mathbb{E}\Phi(\xi)$ as $n \to \infty$.

The sequence $\{\xi_n\}_{n\geq 1}$ converges weakly in $l^\infty_\psi(S)$ to a random function ξ iff

$$\frac{\xi_n}{\psi} \to_w \frac{\xi}{\psi} \quad \text{as } n \to \infty.$$

It's easy to check that if $\xi_n \to_w \xi$ in $l^\infty_\psi(S)$, then the same holds in $l^\infty_{\psi_1}(S)$ for any ψ_1 such that $c^{-1}\psi \leq \psi_1 \leq c\psi$ for some $c > 0$.

Let S be an open subset of a metric space. The sequence $\{\xi_n\}_{n\geq 1}$ *converges weakly locally* in S to a random function ξ, or $\xi_n \to_{w,\,\mathrm{loc}} \xi$ as $n \to \infty$, iff ξ is a.s. continuous in S and for any compact subset K of S, $\{\xi_n\}_{n\geq 1}$ converges weakly to ξ in the space $l^\infty(K)$.

Let $(\mathbf{X}, \mathscr{A}, P)$ be a probability space. Denote $(\Omega, \Sigma, \Pr)$ a countable product of copies of $(\mathbf{X}, \mathscr{A}, P)$ times the unit interval $[0, 1]$ with Borel σ-algebra and Lebesgue measure. Let X_i, $i \geq 1$ be coordinates on $(\Omega, \Sigma, \Pr)$. We define the empirical measure P_n, based on a sample $(X_1, \ldots, X_n)$, as

$$P_n := n^{-1}\sum_1^n \delta_{X_i},$$

where δ_x is a unit point mass at x. A signed random measure $Z_n := n^{1/2}(P_n - P)$ is called an empirical process. Usually Z_n is considered as a stochastic process $Z_n(f)$, indexed by $\mathscr{A}$-measurable functions on $\mathbf{X}$. Here $\nu(f) := \int_{\mathbf{X}} f\, d\nu$ for a signed measure ν on $\mathscr{A}$.

The limit process (in the sense of finite-dimensional distributions) of the sequence $Z_n(f)$, $f \in L_2(\mathbf{X}, dP)$ of empirical processes is the so called *P-Brownian bridge*, that is, a Gaussian process $W_P^\circ(f)$, $f \in L_2(\mathbf{X}, dP)$ with mean 0 and covariance

$$\mathbb{E}W_P^\circ(f)W_P^\circ(g) = P(fg) - P(f)P(g).$$

The P-Brownian bridge can be also viewed as a stochastic integral $W_P^\circ(f) = \int_{\mathbf{X}} f\, dW_P^\circ$.

A class $\mathscr{F} \subset L_1(\mathbf{X}, dP)$ is called P-Glivenko–Cantelli, or $\mathscr{F} \in GC(P)$, iff the sequence of random variables $\{\|P_n - P\|_{\mathscr{F}}\}$ tends to 0 as $n \to \infty$ a.s.

We introduce the following metric in $L_2(\mathbf{X}, dP)$:

$$\rho_P(f, g) := (P(f-g)^2 - (P(f-g))^2)^{1/2}.$$

A class $\mathscr{F} \subset L_2(\mathbf{X}, dP)$ is called P-Donsker, or $\mathscr{F} \in CLT(P)$, iff

$$Z_n \to_w W_P^\circ \text{ in } l^\infty(\mathscr{F}) \quad \text{as } n \to \infty.$$

A measurable real-valued function E on $\mathbf{X}$ is called an *envelope* of the class $\mathscr{F}$ iff $|f(x)| \le E(x)$ for all $x \in \mathbf{X}$, $f \in \mathscr{F}$.

Given a (pseudo)metric space (S, d), $N_d(S; \varepsilon)$ denotes the minimal number of balls of radiuses ε, covering S. Given $Q \in \mathscr{P}(\mathbf{X})$ and $1 \le r < +\infty$, denote $d_{Q,r}$ the metric of the space $L_r(\mathbf{X}; dQ)$. We define

$$N_r(\mathscr{F}; \varepsilon) := N_{r,E}(\mathscr{F}; \varepsilon) := \sup_{Q \in \mathscr{P}(\mathbf{X})} N_{d_{Q,r}}\left(\mathscr{F}; \varepsilon\left(\int_{\mathbf{X}} E^r\, dQ\right)^{1/r}\right) \quad \text{for } \varepsilon > 0.$$

We set $H_r(\mathscr{F}; \varepsilon) := \log N_r(\mathscr{F}; \varepsilon)$.

Under some measurability, the conditions $H_1(\mathscr{F}; \varepsilon) < +\infty$ for all $\varepsilon > 0$ and $\int_{\mathbf{X}} E\, dP < +\infty$ imply $\mathscr{F} \in GC(P)$. The conditions

$$\int_0^\infty H_2^{1/2}(\mathscr{F}; \varepsilon)\, d\varepsilon < +\infty$$

and $\int_{\mathbf{X}} E^2\, dP < +\infty$ imply $\mathscr{F} \in CLT(P)$. For the so-called VC-subgraph classes $\mathscr{F}$, there exist constants $V = V(\mathscr{F})$ and $C = C(V; r)$ such that

$$N_r(\mathscr{F}; \varepsilon) \le C\varepsilon^{-V} \quad \text{for all } \varepsilon > 0.$$

In particular, if there exists a finite-dimensional space $\mathscr{H}$ of real-valued functions on $\mathbf{X} \times \mathbb{R}^1$, such that any set of the form $\{(x, t)\colon f(x) \ge t \ge 0\}$, or $\{(x, t)\colon f(x) \le t \le 0\}$, with some $f \in \mathscr{F}$, can be represented as $\{h \ge 0\}$ for some $h \in \mathscr{H}$, then $\mathscr{F}$ is a VC-subgraph class with $V(\mathscr{F})$ depending only on the dimension of $\mathscr{H}$. See Dudley (1984), Pollard (1990) and van der Vaart and Wellner (1996) for these and further facts on empirical processes and VC-subgraph classes.

Proofs of the main results.

PROOF OF THEOREM 2.1. The first statement follows immediately from Theorem 8.3.4 of Ioffe and Tihomirov (1974). If $F_P(s)$ is a subgradient of f_P at a point $s \in \mathbb{R}^d$, then $F_P(s) \in \partial f_P(s)$. The first statement along with the definition of the integral of multivalued maps implies that there exists a map F: $\mathbb{R}^d \times \mathbf{X} \mapsto \mathbb{R}^d$, such that $F(s, x) \in \partial f(s, x)$ P a.e., $F(s, \cdot)$ is P-integrable, and (2.1) holds. Then it is easy to check that the map F_P: $\mathbb{R}^d \mapsto \mathbb{R}^d$, defined by (2.1), is a subgradient of the convex function f_P. Indeed, we have for all $s \in \mathbb{R}^d$, $v \in S^{d-1}$ $\langle v, F(s, x)\rangle \le df(s; v; x)$ P a.e. Therefore

$$\langle v, F_P(s)\rangle = \int_{\mathbf{X}} \langle v, F(s, x)\rangle P(dx) \le \int_{\mathbf{X}} df(s; v; x) P(dx) = df_P(s; v),$$

and the second statement follows. □

PROOF OF THEOREM 2.5. First note, that $H^{(j)} = (\partial/\partial x_j)h$ (in the sense of generalized functions). Using the properties of the Fourier transform, we get

$$\tilde{H}^{(j)}(\lambda) = -i\lambda_j \tilde{h}(\lambda) \quad \text{for } \lambda \in \mathbb{R}^d. \tag{A.1}$$

We show that for $P \in \mathscr{P}_h$,

$$\tilde{F}_P^{(j)} = \tilde{H}^{(j)}\tilde{P}, \qquad 1 \le j \le d. \tag{A.2}$$

Indeed, assume first that P has a density p belonging to the space $\mathscr{S}(\mathbb{R}^d)$. Then $F_P^{(j)} = H^{(j)} * p$, which implies $\tilde{F}_P^{(j)} = \tilde{H}^{(j)}\tilde{p}$, and (A.2) follows.

Next, assume that P is any Borel probability measure with bounded support. Given a random vector X with distribution P, let Y be a random vector independent of X, with density in $\mathscr{S}(\mathbb{R}^d)$. Define $Z_n := X + n^{-1}Y$. Let Π_n be the distribution of Z_n. Then Π_n has the density p_n belonging to $\mathscr{S}(\mathbb{R}^d)$, and we have

$$\tilde{F}_{\Pi_n}^{(j)} = \tilde{H}^{(j)}\tilde{\Pi}_n, \qquad 1 \le j \le d. \tag{A.3}$$

Using continuity of F_P in $\mathbb{R}^d \setminus A_P$ (see Proposition 2.4), it is easy to show that for all $s \notin A_P$,

$$F_{\Pi_n}(s) = \int_{\mathbb{R}^d} H(s-x)\Pi_n(dx) = \mathbb{E}H(s-X-n^{-1}Z) = \mathbb{E}F_P(s-n^{-1}Z) \to F_P(s),$$

which implies that $F_{\Pi_n}^{(j)}$ converges to $F_P^{(j)}$ in $\mathscr{S}'(\mathbb{R}^d)$ for all $1 \le j \le d$. Due to continuity of the Fourier transform, we also have that $\tilde{F}_{\Pi_n}^{(j)}$ converges to $\tilde{F}_P^{(j)}$ for all $1 \le j \le d$ in $\mathscr{S}'(\mathbb{R}^d)$. On the other hand, $\tilde{\Pi}_n$ converges to $\tilde{P}$. Passing to the limit in (A.3) as $n \to \infty$, we get (A.2) for an arbitrary P with bounded support.

In the general case, given an X with distribution $P \in \mathscr{P}_h$, denote X_n the random vector obtained by conditioning of X on $|X| \le n$, and let Π_n be the distribution of X_n. Then (A.3) holds for such a Π_n. Under the condition $P \in \mathscr{P}_h$, we have for all $s \in \mathbb{R}^d$,

$$F_{\Pi_n}(s) = \frac{1}{P\{x\colon |x| \le n\}} \int_{\{x\colon |x|\le n\}} H(s-x)P(dx) \to F_P(s) \quad \text{as } n \to \infty.$$

It follows that for all $1 \le j \le d$ $F_{\Pi_n}^{(j)}$ converges to $F_P^{(j)}$ in $\mathscr{S}'(\mathbb{R}^d)$ as $n \to \infty$, which implies that $\tilde{F}_{\Pi_n}^{(j)}$ converges to $\tilde{F}_P^{(j)}$ in $\mathscr{S}'(\mathbb{R}^d)$. Since we also have $\tilde{\Pi}_n$ converges to $\tilde{P}$, by passing to the limit in (A.3), we get (A.2) in the general case.

Now, if we have two Borel probabilities P and Q such that $F_P \equiv F_Q$, we get, for all $1 \le j \le d$, $\tilde{F}_P^{(j)} = \tilde{F}_Q^{(j)}$. Therefore, by (A.1) and (A.2), $\lambda_j \tilde{h}(\lambda)(\tilde{P}(\lambda) - \tilde{Q}(\lambda)) \equiv 0$. Under the condition $\tilde{h} \neq 0$ a.e. in $\mathbb{R}^d$, we have $\tilde{P} = \tilde{Q}$ a.e., and since $\tilde{P}$ and $\tilde{Q}$ are both continuous, we can conclude that $\tilde{P} = \tilde{Q}$ and $P = Q$. □

PROOF OF PROPOSITION 2.6. Almost everything follows from Proposition 2.4. One need prove just the second statement of (i), and (ii). Let $s \in A_P$. Clearly, there exists a probability measure Q on $\mathbb{R}^d$ such that $Q(\{s\}) = 0$ and $P = p(s)\delta_s + (1 - p(s))Q$, where δ_s is the unit mass at the point s. We have $F_P = p(s)F_{\delta_s} + (1 - p(s))F_Q$. It follows that

$$F_P(s+\varepsilon v) = p(s)\Phi(\varepsilon)v + (1-p(s))F_Q(s+\varepsilon v).$$

Since F_Q is continuous at the point s and $\Phi(0+) = \rho_\varphi$, we get

$$F_P(s+\varepsilon v) \to p(s)\rho_\varphi v + (1-p(s))F_Q(s).$$

It remains to note that since $F_{\delta_s}(s) = 0$, we have $F_P(s) = (1 - p(s))F_Q(s)$, which proves (i).

Part (ii) follows from the representation

$$F_P(\lambda v) = \int_{\{x \neq \lambda v\}} \Phi\left(\lambda \left| v - \frac{x}{\lambda} \right|\right) \frac{v - x/\lambda}{|v - x/\lambda|} P(dx),$$

the fact that $\Phi(+\infty) = R_\varphi$ and the dominated convergence theorem. □

PROOF OF PROPOSITION 3.1. Indeed, for all $s \in \mathbb{R}^d$,

$$\begin{aligned} f_{P_\theta}(s) &= \int_{\mathbb{R}^d} (h(s-x) - h(s_0 - x))P_\theta(dx) \\ &= \int_{\mathbb{R}^d} (h(s-\theta-x) - h(s_0-\theta-x))P(dx) \\ &= \int_{\mathbb{R}^d} (h(s-\theta-x) - h(s_0-x))P(dx) \\ &\quad + \int_{\mathbb{R}^d} (h(s_0-x) - h(s_0-\theta-x))P(dx) \\ &= f_P(s-\theta) + f_{P_\theta}(s_0+\theta). \end{aligned}$$

To prove (i), note that $t \in \partial f_{P_\theta}(s)$ iff for all $v \in \mathbb{R}^d$, $v \neq 0$ we have

$$\langle v, t\rangle \le df_{P_\theta}(s; v) = df_P(s-\theta; v).$$

It means that $t \in \partial f_P(s-\theta)$. Then we have for all $t \in \mathbb{R}^d$,

$$\begin{aligned} f^*_{P_\theta}(t) &= \sup_{s\in\mathbb{R}^d}[\langle s, t\rangle - f_{P_\theta}(s)] = \sup_{s\in\mathbb{R}^d}[\langle s, t\rangle - f_P(s-\theta) - f_{P_\theta}(s_0+\theta)] \\ &= \sup_{s\in\mathbb{R}^d}[\langle s+\theta, t\rangle - f_P(s)] - f_{P_\theta}(s_0+\theta) \\ &= \langle \theta, t\rangle + \sup_{s\in\mathbb{R}^d}[\langle s, t\rangle - f_P(s)] - f_{P_\theta}(s_0+\theta) \\ &= \langle \theta, t\rangle + f^*_P(t) - f_{P_\theta}(s_0+\theta). \end{aligned}$$

Finally, $s \in \partial f^*_{P_\theta}(t) = \partial f^{-1}_{P_\theta}(t)$ iff for all $v \in \mathbb{R}^d$, $v \neq 0$

$$\langle s, v\rangle \leq df^*_{P_\theta}(t; v) = \langle \theta, v\rangle + df^*_P(t; v),$$

which is equivalent to

$$\langle s-\theta, v\rangle \leq df^*_P(t; v).$$

It means that $s-\theta \in \partial f^*_P(t) = \partial f^{-1}_P(t)$, so (ii) holds. $\square$

PROOF OF PROPOSITION 3.3. Note that $\partial f_P(s)$ does not depend on the choice of s_0 in the definition of f_P. So, we could assume $s_0 = 0$ and $f_P(s) := \int_{\mathbb{R}^d}(\varphi(|s-x|) - \varphi(|x|))P(dx)$. For an orthogonal transformation O and $s \in \mathbb{R}^d$, we get

$$\begin{aligned} f_{P\circ O^{-1}}(s) &= \int_{\mathbb{R}^d}\big(\varphi(|s-x|) - \varphi(|x|)\big)P\circ O^{-1}(dx) \\ &= \int_{\mathbb{R}^d}\big(\varphi(|s-Ox|) - \varphi(|Ox|)\big)P(dx) \\ &= \int_{\mathbb{R}^d}\big(\varphi(|O^{-1}s-x|) - \varphi(|x|)\big)P(dx) = f_P(O^{-1}s). \end{aligned}$$

Next, for $t \in \mathbb{R}^d$, we have

$$\begin{aligned} f^*_{P\circ O^{-1}}(t) &= \sup_{s\in\mathbb{R}^d}[\langle s, t\rangle - fP\circ O^{-1}(s)] = \sup_{s\in\mathbb{R}^d}[\langle s, t\rangle - f_P(O^{-1}s)] \\ &= \sup_{s\in\mathbb{R}^d}[\langle O^{-1}s, O^{-1}t\rangle - f_P(O^{-1}s)] \\ &= \sup_{s\in\mathbb{R}^d}[\langle s, O^{-1}t\rangle - f_P(s)] = f^*_P(O^{-1}t). \end{aligned}$$

Relationships (i) and (ii) now follow by subdifferentiation. $\square$

PROOF OF PROPOSITION 3.4. Without loss of generality we can and do assume that $s_0 = a = 0$ (since subdifferentials ∂f_P and ∂f^*_P do not depend on the choice of s_0). Since P is invariant with respect to any orthogonal transformation O, it follows from the proof of Proposition 3.3 that $f_{P\circ O^{-1}}(s) = f_P(O^{-1}s) = f_P(s)$, $s \in \mathbb{R}^d$, so the convex function f_P is invariant with respect to all orthogonal transformations. It means (see [Ro], page 110) that there exists a lower semicontinuous convex nondecreasing function φ_P on $[0, +\infty)$ such that

$f_P(s) = \varphi_P(|s|)$. Let $(e_1, \ldots, e_d)$ be an orthonormal basis in $\mathbb{R}^d$. Given a vector $x \in \mathbb{R}^d$, let $(x_1, \ldots, x_d)$ be its coordinates in this basis. Then for $\lambda \geq 0$,

$$\varphi_P(\lambda) = f_P(\lambda e_1) = \int_{\mathbb{R}^d} \varphi(|\lambda e_1 - x|)P(dx) = \int_{\mathbb{R}^d} \varphi\big(\rho(x;\lambda)\big)P(dx).$$

Subdifferentiating this relationship, we get (i).

Now we use the formula for the conjugate of the convex function $f_P(s) = \varphi_P(|s|)$ (see [Ro], page 110) to get $f_P^*(t) = \varphi_P^+(|t|),\ t \in \mathbb{R}^d$, where the function φ_P^+ is a *monotone conjugate* of φ_P:

$$\varphi_P^+(\lambda) := \sup_{\alpha \geq 0}[\lambda\alpha - \varphi_P(\alpha)] \quad \text{for } \lambda \geq 0.$$

It is a lower semicontinuous convex nondecreasing function such that for all $\lambda \geq 0$, $\partial\varphi_P^+(\lambda) = (\partial\varphi_P)^{-1}(\lambda)$. Subdifferentiating the relationship for f_P^* gives (ii). $\square$

In order to prove Theorem 5.1 we need a couple of facts about convex functions and their minimal points. The next lemma is just a stochastic version of a well-known fact of convex analysis (see [Ro], Theorem 10.8).

LEMMA A.1. *Let $f_n\colon \mathbb{R}^d \times \Omega \mapsto \mathbb{R}^1$ be a sequence of random functions, convex on $\mathbb{R}^d$ for* Pr-*almost all $\omega \in \Omega$. Suppose that for all $s \in \mathbb{R}^d$, $f_n(s) \to f(s)$ as $n \to \infty$ in probability. Then for all compacts $K \subset \mathbb{R}^d$ $\|f_n - f\|_K \to 0$ as $n \to \infty$ in probability.*

Recall that $\partial f^{-1}(0)$ coincides with the set of all minimal points of f.

LEMMA A.2. *Let $f\colon \mathbb{R}^d \mapsto \mathbb{R}^1$ be a convex function and let $g\colon \mathbb{R}^d \mapsto \mathbb{R}^1$ be a quadratic function, defined by $g(s) := \langle s, a\rangle + \frac{1}{2}\langle As, s\rangle,\ s \in \mathbb{R}^d$, where $a \in \mathbb{R}^d$ is a vector, and A is a symmetric positively definite operator in $\mathbb{R}^d$ such that $\langle As, s\rangle \geq c|s|^2$ for all $s \in \mathbb{R}^d$ with a constant $c > 0$. Let $b := -\operatorname{inv}(A)a$. Take a number $r > |b|$. Denote $\delta := \big(6c^{-1}\|f - g\|_{\bar{B}(0;r)}\big)^{1/2}$. If $\delta \leq r - |b|$, then $\partial f^{-1}(0)$ belongs to the ball $\bar{B}(b;\delta)$.*

PROOF. The unique minimum of g is at the point b. By the Taylor expansion,

$$g(s) = g(b) + \tfrac{1}{2}\langle A(s-b), s-b\rangle \quad \text{for } s \in \mathbb{R}^d.$$

If $|s - b| = \delta$, then

$$g(s) \geq g(b) + \frac{c}{2}\delta^2 = g(b) + 3\|f - g\|_{\bar{B}(0,r)}.$$

Since for all s with $|s - b| = \delta$ $\bar{B}(b;\delta) \subset \bar{B}(0;r)$, $g(s) \leq f(s) + \|f - g\|_{\bar{B}(0,r)}$ and $g(b) \geq f(b) - \|f - g\|_{\bar{B}(0,r)}$. Therefore, for all s on the sphere $|s - b| = \delta$, we have $f(s) \geq f(b) + \|f - g\|_{\bar{B}(0,r)}$. Since f is convex, any minimal point of f must be in the ball $\bar{B}(b,\delta)$. $\square$

The next statement is very useful in the study of asymptotics of statistical estimators defined by convex minimization. In fact, its versions have already been used by Pollard (1988) and Niemiro (1992). Let $f_n: \mathbb{R}^d \times \Omega \mapsto \mathbb{R}^1$ be a sequence of random functions. Assume that $f_n(s, \omega)$ is convex in s on $\mathbb{R}^d$ for Pr-almost all $\omega \in \Omega$. Suppose that $\mathbb{E}|f_n(s)| < +\infty$, and there exists a function $f: \mathbb{R}^d \mapsto \mathbb{R}^1$ such that $\mathbb{E}f_n(s) = f(s)$, for all $s \in \mathbb{R}^d$ and all $n \geq 1$. Let $F_n: \mathbb{R}^d \times \Omega \mapsto \mathbb{R}^d$ be a subgradient of f_n for Pr-almost all $\omega \in \Omega$. Then (see Theorem 2.1) $F(s) := \mathbb{E}F_n(s)$ is a subgradient of f for all $s \in \mathbb{R}^d$.

Given $s_0 \in \mathbb{R}^d$, denote

$$r_n(s_0; u) := f_n(s_0 + u) - f_n(s_0) - \langle F_n(s_0), u\rangle \quad \text{for } u \in \mathbb{R}^d,$$
$$r(s_0; u) := f(s_0 + u) - f(s_0) - \langle F(s_0), u\rangle \quad \text{for } u \in \mathbb{R}^d.$$

THEOREM A.3. *Suppose the following*:

(i) *the function f has its unique minimum at the point $s_0 \in \mathbb{R}^d$*;

(ii) *if is twice continuously differentiable at this point, so that the first derivative $f'(s_0) = F(s_0) = 0$ and the second derivative $f''(s_0) = F'(s_0)$ is a positively definite operator in $\mathbb{R}^d$*;

(iii) $r_n(s_0; u/n^{1/2}) - r(s_0; u/n^{1/2}) = o_p(n^{-1})$ *as $n \to \infty$ for all $u \in \mathbb{R}^d$*;

(iv) *the sequence $n^{1/2}F_n(s_0)$ is stochastically bounded.*

Then the following asymptotic representation holds for the minimal set of f_n:

$$\partial f_n^{-1}(0) = \partial f_n^*(0) = s_0 - \operatorname{inv}(F'(s_0))F_n(s_0) + o_p(n^{-1/2}) \quad \textit{as } n \to \infty.$$

PROOF. Since f is twice continuously differentiable at the point s_0, we have for all $u \in \mathbb{R}^d$,

$$nr\left(s_0; \frac{u}{n^{1/2}}\right) = \frac{1}{2}\langle F'(s_0)u, u\rangle + o(1) \quad \text{as } n \to \infty. \tag{A.4}$$

It follows from conditions (ii), (iii) and (A.4) that for all $u \in \mathbb{R}^d$,

$$n\left(f_n\left(s_0 + \frac{u}{n^{1/2}}\right) - f_n(s_0) - \left\langle F_n(s_0), \frac{u}{n^{1/2}}\right\rangle\right) = \frac{1}{2}\langle F'(s_0)u, u\rangle + o_p(1) \quad \text{as } n \to \infty.$$

Moreover, Lemma A.1 implies that convergence in the last relationship is uniform over compacts in $\mathbb{R}^d$. Denote

$$\tilde{f}_n(u) := n\left(f_n\left(s_0 + \frac{u}{n^{1/2}}\right) - f_n(s_0)\right) \quad \text{for } u \in \mathbb{R}^d$$

and

$$\tilde{g}_n(u) := n^{1/2}\langle F_n(s_0), u\rangle + \tfrac{1}{2}\langle F'(s_0)u, u\rangle \quad \text{for } u \in \mathbb{R}^d.$$

Take $\varepsilon > 0$. By condition (iv) there exists $r > 0$ such that

$$|n^{1/2} \operatorname{inv}(F'(s_0))F_n(s_0)| < r/2$$

with probability at least $1-\varepsilon$. On the other hand, we have

$$\delta_n^2 := \|\tilde{f}_n - \tilde{g}_n\|_{\bar{B}(0;r)} = o_p(1) \quad \text{as } n \to \infty.$$

By Lemma A.2 the set $\partial \tilde{f}_n^{-1}(0)$ of all minimal points of $\tilde{f}_n$ belongs (with probability tending to 1) to a ball with center $-n^{1/2}\,\mathrm{inv}(F'(s_0))F_n(s_0)$ and with radius of the order δ_n. It remains to observe that $\partial f_n^{-1}(t_0) = s_0 + n^{-1/2}\partial \tilde{f}_n^{-1}(0)$ and the result follows. $\square$

Proof of Theorem 5.1. We apply Theorem A.3 to the following functions:

$$f_n(s) := f_{P_n}(s) - \langle s, t_0\rangle, \qquad F_n(s) := F_{P_n}(s) - t_0,$$
$$f(s) := f_P(s) - \langle s, t_0\rangle, \qquad F(s) := F_P(s) - t_0 \quad \text{for } s \in \mathbb{R}^d.$$

Clearly, s_0 is the unique minimal point of f, and conditions (i) and (ii) of Theorem A.3 hold. Due to the central limit theorem, condition (iv) of Theorem A.3 follows from condition (ii) of Theorem 5.1, since $n^{1/2}F_n(s_0) = n^{1/2}(F_{P_n}(s_0) - F_P(s_0))$. To check condition (iii) of Theorem A.3, note that

$$\mathbb{E}r_n\left(s_0; \frac{u}{n^{1/2}}\right) = r\left(s_0; \frac{u}{n^{1/2}}\right),$$

and

$$\begin{aligned}
&\mathrm{Var}\left(nr_n\left(s_0; \frac{u}{n^{1/2}}\right)\right) \\
&\quad = \mathrm{Var}\left(\sum_1^n\left[f\left(s_0 + \frac{u}{n^{1/2}}, X_i\right) - f(s_0, X_i) - \left\langle F(s_0, X_i), \frac{u}{n^{1/2}}\right\rangle\right]\right) \\
&\quad = \sum_1^n \mathrm{Var}\left(f\left(s_0 + \frac{u}{n^{1/2}}, X_i\right) - f(s_0, X_i) - \left\langle F(s_0, X_i), \frac{u}{n^{1/2}}\right\rangle\right) \\
&\quad \le n\int_{\mathbf{X}}\left(f\left(s_0 + \frac{u}{n^{1/2}}, x\right) - f(s_0, x) - \left\langle F(s_0, x), \frac{u}{n^{1/2}}\right\rangle\right)^2 P(dx).
\end{aligned} \tag{A.5}$$

By the definition of the subgradient of convex functions, we have the inequalities

$$f(s_0 + u, x) - f(s_0, x) \ge \langle F(s_0, x), u\rangle$$

and

$$f(s_0 + u, x) - f(s_0, x) \le \langle F(s_0 + u, x), u\rangle$$

which imply

$$0 \le f(s_0 + u, x) - f(s_0, x) - \langle F(s_0, x), u\rangle \le \langle F(s_0 + u, x) - F(s_0, x), u\rangle.$$

Thus

$$\begin{aligned}
0 &\le f\left(s_0 + \frac{u}{n^{1/2}}, x\right) - f(s_0, x) - \left\langle F(s_0, x), \frac{u}{n^{1/2}}\right\rangle \\
&\le n^{-1/2}\left\langle F\left(s_0 + \frac{u}{n^{1/2}}, x\right) - F(s_0, x), u\right\rangle
\end{aligned}$$

and it follows from (A.5) that

$$\text{(A.6)} \qquad \operatorname{Var}\left(nr_n\left(s_0; \frac{u}{n^{1/2}}\right)\right) \le \mathbb{E}\left\langle F\left(s_0 + \frac{u}{n^{1/2}}, x\right) - F(s_0, x), u\right\rangle^2.$$

By the monotonicity of the subgradient of a convex function, the sequence of nonnegative functions

$$\gamma_n(\cdot) = \left\langle F\left(s_0 + \frac{u}{n^{1/2}}, \cdot\right) - F(s_0, \cdot), u\right\rangle, \qquad n \ge 1$$

is nonincreasing. Denote γ: $\mathbf{X} \mapsto [0, +\infty)$ its limit. Then

$$\text{(A.7)} \qquad \begin{aligned} \int_{\mathbf{X}} \gamma(x) P(dx) &= \lim_{n\to\infty} \int_{\mathbf{X}} \left\langle F\left(s_0 + \frac{u}{n^{1/2}}, x\right) - F(s_0, x), u\right\rangle P(dx) \\ &= \lim_{n\to\infty} \left\langle F_P\left(s_0 + \frac{u}{n^{1/2}}\right) - F_P(s_0), u\right\rangle = 0, \end{aligned}$$

since F_P is continuous at the point 0. But γ is nonnegative, so $\gamma = 0$ P-a.e. Condition (ii) implies, by dominated convergence, that $\int_{\mathbf{X}} \gamma_n^2 dP \to 0$, $n \to \infty$. Thus (A.6) and (A.7) imply condition (iii) of Theorem A.3. Since $\partial f_n^{-1}(0) = \partial f_{P_n}^{-1}(t_0)$ and $F_n(s_0) = F_{P_n}(s_0) - F_P(s_0)$, this completes the proof. □

We need one fact on asymptotics of inverses of random functions. Let $S \subset \mathbb{R}^d$ and H: $S \mapsto \mathbb{R}^d$. Given $\varepsilon > 0$, a map J from a subset $T \subset \mathbb{R}^d$ into S will be called an *ε-inverse* of H on T iff $|t - H(J(t))| < \varepsilon$ for all $t \in T$. Let $\mathscr{J}_H(T; S; \varepsilon)$ be the set of all ε-inverses on T of the map H from S into $\mathbb{R}^d$. The following statement was proved by Koltchinskii (1995b) (see Theorems 2.5 and 2.7 there).

THEOREM A.4. *Let $\{a_n\}_{n\ge1}$ be a sequence of positive real numbers with $a_n \to \infty$, $n \to \infty$. Let S and T be open subsets of $\mathbb{R}^d$, and G be a one-to-one map from S onto T, continuously differentiable in S with nonsingular derivative $G'(s)$, $s \in \mathbb{R}^d$. Let $\{\varepsilon_n\}_{n\ge1}$ be a sequence of positive real numbers with $\varepsilon_n = o(a_n^{-1})$ as $n \to \infty$. Let G_n be functions from $\Omega \times S$ into $\mathbb{R}^d$. Suppose there exists a stochastic process ξ: $S \mapsto \mathbb{R}^d$ such that*

$$a_n(G_n - G) \to_{w,\,\text{loc}} \xi \quad \text{in } S \text{ as } n \to \infty.$$

Let J_n be a function from $\Omega \times T$ into $\mathbb{R}^d$, satisfying the following condition: for any compact $K \subset T$ there exists a compact $C \subset S$ such that

$$\limsup_{n\to\infty} \Pr{}^*\big(\{J_n \notin \mathscr{J}_{G_n}(K; C; \varepsilon_n)\}\big) = 0.$$

Then

$$a_n(J_n - G^{-1} + \operatorname{inv}(G' \circ G^{-1})(G_n - G) \circ G^{-1}) \to_{p,\,\text{loc}} 0 \text{ in } T \text{ as } n \to \infty.$$

PROOF OF THEOREM 5.4. We apply Theorem A.4 to the maps $G := F_P$, $G_n := F_{P_n}$ and $J_n := Q_n$. By the definitions of subdifferential and subgradient, we have $F_{P_n}(Q_n(t)) \in \partial f_{P_n}(Q_n(t))$ and, since $Q_n(t) \in \partial f_{P_n}^{-1}(t)$, we have $t \in \partial f_{P_n}(Q_n(t))$. Thus, for all $t \in B_{P_n} \cap T$,

$$
\begin{aligned}
|F_{P_n}(Q_n(t)) - t| &\le \mathrm{Diam}(\partial f_{P_n}(Q_n(t))) \le \int_{\mathbf{X}} \mathrm{Diam}(\partial f(Q_n(t), x)) P_n(dx) \\
&= \int_{\mathbf{X}} \delta_f(Q_n(t); x) P_n(dx) \le \sup_{s \in \mathbb{R}^d} \int_{\mathbf{X}} \delta_f(s; x) P_n(dx).
\end{aligned} \tag{A.8}
$$

Let K be a compact subset of T. Then, clearly, $K \subset B_P$. Take $\varepsilon > 0$. Denote

$$
\begin{aligned}
A := \{\omega\colon K \subset B_{f_{P_n}}\} &\cap \left\{\omega\colon \sup_{s \in \mathbb{R}^d} \frac{|F_{P_n}(s) - F_P(s)|}{\psi(s)} < \varepsilon\right\} \\
&\cap \left\{\omega\colon \sup_{s \in \mathbb{R}^d} \int_{\mathbf{X}} \delta_f(s; x) P_n(dx) < \varepsilon n^{-1/2}\right\}.
\end{aligned} \tag{A.9}
$$

It follows from Theorem 5.3 and conditions (iv), (v) that for large n, $\Pr^*(\Omega \setminus A) \le \varepsilon$.

If $\psi \equiv c$, we have for all $\omega \in A$ and $t \in K$,

$$
|F_{P_n}(Q_n(t)) - F_P(Q_n(t))| \le c\varepsilon. \tag{A.10}
$$

Otherwise, by condition (ii), we get for large $a > 0$,

$$
\inf_{|s| \ge a} \frac{|F_P(s)|}{\psi(s)} \ge c. \tag{A.11}
$$

If $\omega \in A$, we have, by (A.9) and (A.11), that for all s with $|s| \ge a$,

$$
|F_{P_n}(s)| \ge (c - \varepsilon)\phi(|s|). \tag{A.12}
$$

We have either $|Q_n(t)| \le a$ or $|Q_n(t)| > a$ and then, by inequality (A.12),

$$
|F_{P_n}(Q_n(t))| \ge (c - \varepsilon)\phi(|Q_n(t)|).
$$

It follows from (A.8) and (A.9) that for all $\omega \in A$ and $t \in K$ $|F_{P_n}(Q_n(t))| \le |t| + \varepsilon n^{-1/2}$. Therefore, for all $\omega \in A$ and $t \in K$,

$$
|Q_n(t)| \le \phi^{-1}\left(\frac{|t| + \varepsilon n^{-1/2}}{c - \varepsilon}\right) \bigvee a.
$$

Since K is bounded, there exists $r > 0$ such that $|t| \le r$ for all $t \in K$. It follows that for all $\omega \in A$,

$$
\sup_{t \in K} |Q_n(t)| \le \phi^{-1}\left(\frac{r + \varepsilon}{c - \varepsilon}\right) \bigvee a. \tag{A.13}
$$

Since, by (i), $S \subset D_P$, it follows that F_P is a homeomorphism of S and T (see the properties of M-distribution function). Therefore, there exists a compact

$C \subset S$ and a positive number δ such that $F_P^{-1}(K_\delta) \subset C$. Using (A.13), we get for all $\omega \in A$ and $t \in K$,

$$|F_{P_n}(Q_n(t)) - F_P(Q_n(t))| \le \varepsilon\phi(|Q_n(t)|) \le \varepsilon\left[\frac{r+\varepsilon}{c-\varepsilon} \bigvee \phi(a)\right]. \tag{A.14}$$

It follows from (A.8), (A.10) and (A.14) that for $\omega \in A$ and $t \in K$,

$$|t - F_P(Q_n(t))| \le \varepsilon\left[\frac{r+\varepsilon}{c-\varepsilon} \bigvee \phi(a) \bigvee c\right] + \varepsilon. \tag{A.15}$$

Clearly, $\varepsilon > 0$ can be choosen in such a way that (A.15) yields $|t - F_P(Q_n(t))| < \delta$. It means that $F_P(Q_n(t)) \in K_\delta$, which implies $Q_n(t) \in C$. Thus we proved that for all $\omega \in A$, $Q_n(K) \subset C$. By condition (iv) and (A.8), there exists a sequence $\{\varepsilon_n\}$ of positive numbers such that $\varepsilon_n = o(n^{-1/2})$ and for all large n, $\sup_{t\in K} |F_{P_n}(Q_n(t)) - t| \le \varepsilon_n$. It shows that

$$\limsup_{n\to\infty} \Pr{}^*\big(\{Q_n \notin \mathscr{J}(K; C; \varepsilon_n)\}\big) = 0. \tag{A.16}$$

Under condition (iii), Theorem 5.3 implies that

$$n^{1/2}(F_{P_n} - F_P) \to_{w,\,\mathrm{loc}} \xi_P \quad \text{as } n \to \infty \text{ in } \mathbb{R}^d,$$

and the result follows from Theorem A.4. $\square$

PROOF OF THEOREMS 5.5 AND 5.6. Let $\Delta_k := (\Phi^{-1}(2^{k-1}), \Phi^{-1}(2^k)]$ for $k \ge 1$, and $\Delta_0 := [0, \Phi^{-1}(1)]$. Define

$$\tilde\psi(s) := \sum_{k=0}^{\infty} 2^k I_{\{|s|\in\Delta_k\}} \quad \text{for } s \in \mathbb{R}^d,$$

and let

$$F(s,x) := \begin{cases} \dfrac{\Phi(|s-x|)}{\tilde\psi(s)} \dfrac{s-x}{|s-x|}, & \text{for } x \ne s, \\ 0, & \text{otherwise.} \end{cases}$$

Let $\mathscr{F} := \bigcup_1^d \mathscr{F}^{(j)}$, where $\mathscr{F}^{(j)} := \{F^{(j)}(s,\cdot): s \in \mathbb{R}^d\}$. Since $\tilde\psi(s) \ge 1$, condition (5.1) implies that the function $E_{\mathscr{F}}(x) := C_1 + C_2\Phi(|x|)$ is an envelope of the class $\mathscr{F}$ for some constants $C_1 > 0,\ C_2 > 0$. Consider the functions

$$F_1(s,x) := \begin{cases} \dfrac{\Phi(|s-x|)}{\tilde\psi(s)}, & \text{for } x \ne s, \\ 0, & \text{otherwise} \end{cases}$$

and

$$F_2(s,x) := \begin{cases} E_{\mathscr{F}}(x) \dfrac{s-x}{|s-x|}, & \text{for } x \ne s, \\ 0, & \text{otherwise.} \end{cases}$$

Let

$$\mathscr{F}_1 := \{F_1(s,\cdot): s \in \mathbb{R}^d\},$$
$$\mathscr{F}_{1,k} := \{F_1(s,\cdot): s \in \Delta_k\} = \{2^{-k}\Phi(|s-\cdot|): s \in \Delta_k\},$$
$$\mathscr{F}_2^{(j)} := \{F_2^{(j)}(s,\cdot): s \in \mathbb{R}^d\}, \qquad 1 \le j \le d.$$

Clearly, $\mathscr{F}_1 = \bigcup_0^\infty \mathscr{F}_{1,k}$. It's also obvious that $E_{\mathscr{F}}$ is an envelope of $\mathscr{F}_1$ and $\mathscr{F}_2^{(j)}$, $1 \le j \le d$. The inequality

$$\begin{aligned}
&\big|F^{(j)}(s_1,x) - F^{(j)}(s_2,x)\big| \\
&\quad= \left|\left(\frac{\Phi(|s_1-x|)}{\tilde\psi(s_1)} - \frac{\Phi(|s_2-x|)}{\tilde\psi(s_2)}\right)\frac{s_1^{(j)}-x^{(j)}}{|s_1-x|}\right. \\
&\qquad\left. + \frac{\Phi(|s_2-x|)}{\tilde\psi(s_2)}\left(\frac{s_1^{(j)}-x^{(j)}}{|s_1-x|} - \frac{s_2^{(j)}-x^{(j)}}{|s_2-x|}\right)\right| \\
&\quad\le \left|\frac{\Phi(|s_1-x|)}{\tilde\psi(s_1)} - \frac{\Phi(|s_2-x|)}{\tilde\psi(s_2)}\right| + E_{\mathscr{F}}(x)\left|\frac{s_1^{(j)}-x^{(j)}}{|s_1-x|} - \frac{s_2^{(j)}-x^{(j)}}{|s_2-x|}\right| \\
&\quad\le |F_1(s_1,x) - F_1(s_2,x)| + \big|F_2^{(j)}(s_1,x) - F_2^{(j)}(s_2,x)\big|
\end{aligned}$$

implies that for any $1 \le r < +\infty$ and $Q \in \mathscr{P}(\mathbb{R}^d)$ with $\int_{\mathbb{R}^d} E_{\mathscr{F}}^r(x)Q(dx) < +\infty$, the following estimate holds for the metrics $d_{Q,r}$:

$$\begin{aligned}
&d_{Q,r}(F^{(j)}(s_1,\cdot), F^{(j)}(s_2,\cdot)) \\
&\qquad\le d_{Q,r}(F_1(s_1,\cdot), F_1(s_2,\cdot)) + d_{Q,r}(F_2^{(j)}(s_1,\cdot), F_2^{(j)}(s_2,\cdot)).
\end{aligned}$$

It follows that

$$H_{d_{Q,r}}(\mathscr{F}^{(j)};\varepsilon) \le H_{d_{Q,r}}(\mathscr{F}_1;\varepsilon/2) + H_{d_{Q,r}}(\mathscr{F}_2^{(j)};\varepsilon/2)$$

and

$$H_r(\mathscr{F}^{(j)};\varepsilon) \le H_r(\mathscr{F}_1;\varepsilon/2) + H_r(\mathscr{F}_2^{(j)};\varepsilon/2). \tag{A.17}$$

In order to estimate the entropies, note that for all $k \ge 0$ $\mathscr{F}_{1,k}$ is a VC-subgraph class with $V(\mathscr{F}_{1,k})$ depending only on d. The proof follows the lines of Pollard (1984), Chapter 2, Example 26. Indeed, for all $s \in \Delta_k$,

$$\begin{aligned}
\big\{(x,t): 2^{-k}\Phi(|s-x|) \ge t\big\} &= \big\{(x,t): |s-x|^2 \ge (\Phi^{-1}(2^k t))^2\big\} \\
&= \left\{(x,t): \sum_1^d (s^{(j)} - x^{(j)})^2 \ge (\Phi^{-1}(2^k t))^2\right\} \\
&= \big\{(x,t): P(x) - (\Phi^{-1}(2^k t))^2 \ge 0\big\},
\end{aligned}$$

where $P(x)$ is a polynomial of $x = (x^{(1)},\dots,x^{(d)})$ of the second degree. Denote by $\mathscr{H}$ the linear span of the set of all such polynomials and a function $(\Phi^{-1}(2^k t))^2$. Then for any $s \in \Delta_k$ the set $\{(x,t): 2^{-k}\Phi(|s-x|) \ge t\}$ can be represented as $\{(x,t): h(x,t) \ge 0\}$ for some $h \in \mathscr{H}$. Thus, $\mathscr{F}_{1,k}$ is a VC-subgraph class and $V(\mathscr{F}_{1,k})$ depends only on d. It follows that for $1 \le r < +\infty$

$N_r(\mathscr{F}_{1,k};\varepsilon) \le C\varepsilon^{-V}$, $\varepsilon > 0$ with constants $C > 0, V > 0$ depending only on d and r. Since the function $2^{-k}E_{\mathscr{F}}(\cdot)$ is an envelope of $\mathscr{F}_{1,k}$, we have for all $Q \in \mathscr{P}(\mathbb{R}^d)$,

$$N_{d_{Q,r}}\left(\mathscr{F}_{1,k};\varepsilon 2^{-k}\left(\int_{\mathbf{X}} E^r_{\mathscr{F}}\,dQ\right)^{1/r}\right) \le C\varepsilon^{-V}, \qquad \varepsilon > 0,$$

or

$$N_{d_{Q,r}}\left(\mathscr{F}_{1,k};\varepsilon\left(\int_{\mathbf{X}} E^r_{\mathscr{F}}\,dQ\right)^{1/r}\right) \le C2^{-kV}\varepsilon^{-V}, \qquad \varepsilon > 0.$$

Hence, for all $Q \in \mathscr{P}(\mathbb{R}^d)$ and for all $\varepsilon > 0$,

$$\begin{aligned} N_{d_{Q,r}}\left(\mathscr{F}_1;\varepsilon\left(\int_{\mathbf{X}} E^r_{\mathscr{F}}\,dQ\right)^{1/r}\right) &\le \sum_{k=0}^{\infty} N_{d_{Q,r}}\left(\mathscr{F}_{1,k};\varepsilon\left(\int_{\mathbf{X}} E^r_{\mathscr{F}}\,dQ\right)^{1/r}\right) \\ &\le C\sum_{1}^{\infty} 2^{-kV}\varepsilon^{-V}. \end{aligned}$$

Therefore,

$$H_r(\mathscr{F}_1;\varepsilon) = O\left(\log\frac{1}{\varepsilon}\right) \quad \text{as } \varepsilon \to 0. \tag{A.18}$$

Now we show that all classes $\mathscr{F}_2^{(j)}$, $1 \le j \le d$ are *VC*-subgraph, too. We have for any $s \in \mathbb{R}^d$,

$$\begin{aligned} &\{(x,t)\colon F_2^{(j)}(s,x) \ge t \ge 0\} \\ &\quad = \left\{(x,t)\colon s^{(j)} - x^{(j)} \ge \frac{t|s-x|}{E_{\mathscr{F}}(x)},\ x \ne s,\ t \ge 0\right\} \bigcup \{(s,0)\} \\ &\quad = \left\{(x,t)\colon (s^{(j)} - x^{(j)})^2 - \frac{t^2\sum_{k=1}^d (s^{(k)} - x^{(k)})^2}{E^2_{\mathscr{F}}(x)} \ge 0,\right. \\ &\qquad\qquad \left. x \ne s,\ t \ge 0\right\} \bigcup \{(s,0)\}. \end{aligned} \tag{A.19}$$

Let $\mathscr{H}$ be the linear span of functions 1, $x^{(j)}$, $(x^{(j)})^2$, t^2, $t^2(x^{(j)}/E^2_{\mathscr{F}}(x))$ and $t^2(x^{(j)}/E^{\mathscr{F}}(x))^2$. Then the class of sets $\mathscr{C} := \{\{(x,t)\colon h(x,t) \ge 0\}\colon h \in \mathscr{H}\}$ is Vapnik–Chervonenkis, and so is the class

$$\tilde{\mathscr{C}} := \{C \cap \{(x,t)\colon x \ne s,\ t \ge 0\}\} \cup \{(s,0)\}\colon C \in \mathscr{C},\ s \in \mathbb{R}^d\}.$$

By (A.19), for all $s \in \mathbb{R}^d$, $\{(x,t)\colon F_2^{(j)}(s,x) \ge t \ge 0\} \in \tilde{\mathscr{C}}$. Therefore $\{\{(x,t)\colon F_2^{(j)}(s,x) \ge t \ge 0\}\colon s \in \mathbb{R}^d\}$ is a Vapnik–Chervonenkis class. Quite similarly, the class $\{\{(x,t)\colon F_2^{(j)}(s,x) \le t \le 0\}\colon s \in \mathbb{R}^d\}$ is Vapnik–Chervonenkis, too. It follows that for all $1 \le j \le d$, classes of functions $\mathscr{F}_2^{(j)}$ are VC-subgraph, and we have

$$H_r(\mathscr{F}_2^{(j)};\varepsilon) = O\left(\log\frac{1}{\varepsilon}\right) \quad \text{as } \varepsilon \to 0. \tag{A.20}$$

It follows from (A.17), (A.18) and (A.20) that for all $1 \le j \le d$,

$$H_r(\mathscr{F}^{(j)}; \varepsilon) = O\biggl(\log \frac{1}{\varepsilon}\biggr) \quad \text{as } \varepsilon \to 0,$$

which implies

$$H_r(\mathscr{F}; \varepsilon) = O\biggl(\log \frac{1}{\varepsilon}\biggr) \quad \text{as } \varepsilon \to 0. \tag{A.21}$$

Condition (5.3) implies that $\int_{\mathbb{R}^d} E_{\mathscr{F}}(x)P(dx) < +\infty$, which along with (A.21) implies that $\mathscr{F} \in GC(P)$, and (5.4) follows. It, in turn, implies (5.5) and (5.6).

Under condition (5.7) we have $\int_{\mathbb{R}^d} E^2_{\mathscr{F}}(x)P(dx) < +\infty$, which implies [again, along with (A.21)] that $\mathscr{F} \in CLT(P)$. It follows that the process

$$n^{1/2}\frac{F_{P_n}(s) - F_P(s)}{\tilde{\psi}(s)} = n^{1/2}(P_n - P)(F(s, \cdot)), \qquad s \in \mathbb{R}^d$$

converges weakly in $l^\infty(\mathbb{R}^d)$ to

$$\frac{\xi_P(s)}{\tilde{\psi}(s)} = \int_{\mathbb{R}^d} F(s, x)W^\circ_P(dx), \;\; s \in \mathbb{R}^d.$$

By the definition of the function $\tilde{\psi}$, we have for all $s \in \mathbb{R}^d$ $1/2 \le (\psi(s)/\tilde{\psi}(s)) \le 1$. Therefore, (5.8) also holds. The kernel

$$G(s, x) := \frac{\Phi(|s - x|)}{\psi(s)}\frac{s - x}{|s - x|}\mathbf{I}_{\{s \ne x\}}$$

is continuous in s at any point $s_0 \ne x$. Since P is nonatomic, it is, in fact, continuous in $s \in \mathbb{R}^d$ for P-almost all x. This, by dominated convergence, implies continuity of $G(s, \cdot)$ in $s \in \mathbb{R}^d$ with respect to the metric ρ_P. Since $\xi_P(s) = \psi(s)W^\circ_P(G(s, \cdot))$, W°_P is a.s. ρ_P-continuous on $\{G(s; \cdot): s \in \mathbb{R}^d\}$, and ψ is continuous in $\mathbb{R}^d$, we can conclude that the process ξ_P is continuous on $\mathbb{R}^d$. This fact and (5.8) imply (5.9) and (5.10). □

PROOF OF THEOREM 5.7. We use Theorem 5.4. Condition (i) obviously holds. Condition (ii) with $\psi(s) := \Phi(|s|) \vee 1$ is a consequence of the following fact (which is rather easy to prove): under conditions (5.1) and (5.2), $F_P(s) = \Phi(|s|)(s/|s|) + o(\Phi(|s|))$ as $s \to \infty$. Condition (iii) follows from Theorem 5.6. Next, since

$$\delta_f(s; x) := \operatorname{Diam}(\partial f(s; x)) = \begin{cases} \rho_\varphi, & \text{for } s = x, \\ 0, & \text{otherwise,} \end{cases}$$

we have

$$\sup_{s \in \mathbb{R}^d} \int_{\mathbb{R}^d} \delta_f(s; x)P_n(dx) = \frac{\rho_\varphi}{n} = o(n^{-1/2}).$$

So, condition (iv) also holds. Finally, condition (v) is obvious, since $B_{P_n} = B_P = B(0; R_\varphi)$, and the result follows from Theorem 5.4. □

PROOF OF THEOREM 4.5. We apply Theorem 5.4 to $F := G$, and R, R_n instead of P, P_n. First note that G_R is continuously differentiable in $\mathbb{R}^{md}$ with nonsingular derivative $G'_R(S)$, $S \in \mathbb{R}^{md}$ given by (4.4). This means that condition (i) of Theorem 5.4 holds. Assume that $\psi \equiv 1$, so that (ii) holds. Note that G is R-Donsker, since the class of functions $\bigcup_{j=1}^{m+d}\{G^{(j)}(S;\cdot,\cdot): S \in \mathbb{R}^{md}\}$ is a Vapnik–Chervonenkis subgraph with a square integrable envelope, and for all $1 \le j \le d$, the functions $G^{(j)}(S;\cdot,\cdot)$ are ρ_R-continuous in $S \in \mathbb{R}^{md}$. Thus (iii) holds. Next we have

$$\delta_g(S; x, y) = \begin{cases} |x|, & \text{if } Sx = y,\\ 0, & \text{otherwise.}\end{cases}$$

Therefore,

$$\sup_{S\in\mathbb{R}^{md}} \int_{\mathbb{R}^{m+d}} \delta_g(S; x, y) R_n(dx; dy) = \sup_{S\in\mathbb{R}^{md}} \sum_{i=1}^{n} |X_i|\ \mathbf{I}_{\{SX_i = Y_i\}}.$$

Under conditions (i), (ii), we have $\mathbb{E}\mathbf{I}_{\{SX=Y\}}|X| = 0$, and rather standard empirical processes argument shows that

$$\sup_{S\in\mathbb{R}^{md}} \sum_{i=1}^{n} |X_i|\ \mathbf{I}_{\{SX_i=Y_i\}} = o_p(n^{-1/2}),$$

so (iv) holds. By the law of large numbers in separable Banach spaces, the sequence of stochastic processes $n^{-1}(|VX_1| + \cdots + |VX_n|)$ converges to the function $\mathbb{E}|VX|$ uniformly in $V \in S^{dm-1}$. This implies condition (v).

Thus, we can use Theorem 5.4 to get the following asymptotic representation of any J_n: $\Omega \times B_R \mapsto \mathbb{R}^{md}$, such that $J_n(T) \in \partial g_{R_n}^{-1}(T)$ for all $T \in B_R$:

$$J_n = G_R^{-1} - \operatorname{inv}(G'_R \circ G_R^{-1})(G_{R_n} - G_R) \circ G_R^{-1} + o_p(n^{-1/2}),$$

which holds locally uniformly in B_R.

Take $r < 1$. Using consistency of $\bar{X}_n$, it is easy to show that there exists a compact $K \subset B_R$ such that with probability close to 1, $t \otimes \bar{X}_n \in K$ for all t with $|t| \le r$ and for all n large enough. It follows that uniformly in $|t| \le r$,

$$\begin{aligned} J_n(t\otimes \bar{X}_n) = G_R^{-1}(t\otimes\bar{X}_n) - \operatorname{inv}(G'_R \circ G_R^{-1}(t\otimes\bar{X}_n))(G_{R_n} - G_R)\\ \circ G_R^{-1}(t\otimes\bar{X}_n) + o_p(n^{-1/2}).\end{aligned}$$

Since G_R^{-1} is uniformly continuous in K, $n^{1/2}(G_{R_n} - G_R)$ is asymptotically equicontinuous and $\bar{X}_n \to \mathbb{E}X$ as $n \to \infty$, we have uniformly in $|t| \le r$,

$$\begin{aligned} J_n(t\otimes\bar{X}_n) = [G_R^{-1}(t\otimes\bar{X}_n) - G_R^{-1}(t\otimes\mathbb{E}X)] + G_R^{-1}(t\otimes\mathbb{E}X)\\ - \operatorname{inv}(G'_R\circ G_R^{-1}(t\otimes\mathbb{E}X))(G_{R_n} - G_R)\circ G_R^{-1}(t\otimes\mathbb{E}X) + o_p(n^{-1/2}).\end{aligned}$$

Note that there exists J_n: $\Omega \times B_R \mapsto \mathbb{R}^{md}$, such that $J_n(T) \in \partial g_{R_n}^{-1}(T)$ for all $T \in B_R$ and $Q_n(t) = J_n(t\otimes\bar{X}_n)$ for all $t \in B_d$. Using differentiability of G_R^{-1} and the condition $|\bar{X}_n - \mathbb{E}X| = O_P(n^{-1/2})$, we now easily get asymptotic representation (4.5), and complete the proof. □

Acknowledgments. I thank R. Dudley for several interesting conversations on the topic of the paper. I also thank M. Arcones, S. Efromovich, P. Pathak and W. Stute for several helpful comments, A. Koldobskii and L. Mattner for suggesting the way to prove Theorem 2.6, and an anonymous referee and an Associate Editor for their suggestions, which substantially improved the paper.

REFERENCES

ARCONES, M. and GINÉ, E. (1991). Some bootstrap tests of symmetry for univariate distributions. *Ann. Statist.* **19** 1496–1511.

BAI, Z. D., RAO, C. R. and WU, Y. (1992). M-estimation of a multivariate linear regression parameters under a convex discrepancy function. *Statist. Sinica* **2** 237–254.

BARINGHAUS, L. (1991). Testing for spherical symmetry of a multivariate distribution. *Ann. Statist.* **19** 899–917.

BARNETT, V. (1976). The ordering of multivariate data (with comments). *J. Roy. Statist. Soc. Ser. A* **139** 318–354.

BERAN, R. (1979). Testing for ellipsoidal symmetry of a multivariate density. *Ann. Statist.* **7** 150–162.

CHAUDHURI, P. (1991). Nonparametric estimates of regression quantiles and their local Bahadur representation. *Ann. Statist.* **19** 760–777.

CHAUDHURI, P. (1996). On a geometric notion of quantiles for multivariate data. *J. Amer. Statist. Assoc.* **91** 862–872.

DUDLEY, R. M. (1984). A course on empirical processes. *École d'ete de Probabilités de Saint-Flour XII. Lecture Notes in Math.* **1097** 1–142. Springer, Berlin.

EDDY, W. F. (1985). Ordering of multivariate data. In *Computer Science and Statistics: The Interface* (L. Billard, ed.) 25–30. North-Holland, Amsterdam.

EINHMAHL, J. H. J. and MASON, D. (1992). Generalized quantile process. *Ann. Statist.* **20** 1062–1078.

GELFAND, I. M. and SHILOV, G. E. (1964). *Generalized Functions* **1**. *Properties and Operations*. Academic Press, New York.

GINÉ, E. and ZINN, J. (1990). Bootstrapping general empirical measures. *Ann. Probab.* **18** 851–869.

GUTENBRUNNER, C. and JUREČKOVÁ, J. (1992). Regression quantile and regression rank score process in the linear model and derived statistics. *Ann. Statist.* **20** 305–330.

GUTENBRUNNER, C., JUREČKOVÁ, J., KOENKER, R. and PORTNOY, S. (1993). Tests of linear hypotheses based on regression rank scores. *J. Nonparametr. Statist.* **2** 307–331.

HABERMAN, S. J. (1989). Concavity and estimation. *Ann. Statist.* **17** 1631–1661.

HALDANE, J. B. S. (1948). Note on the median of a multivariate distribution. *Biometrika* **35** 414–415.

HUBER, P. J. (1967). The behavior of maximum likelihood estimates under non-standard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 221–233. Univ. California Press, Berkeley.

IOFFE, A. D. and TIHOMIROV, V. M. (1974). *The Theory of Extremal Problems*. Nauka, Moscow. (In Russian.)

KOENKER, R. and BASSET, G. (1978). Regression quantiles. *Econometrica* **46** 33–50.

KOENKER, R. and PORTNOY, S. (1987). L-estimators for linear models. *J. Amer. Statist. Assoc.* **82** 851–857.

KOLDOBSKII, A. (1990). Inverse problem for potentials of measures in Banach spaces. In *Probability Theory and Mathematical Statistics* (B. Grigelionis and J. Kubilius, eds.) **1** 627–637. VSP/Mokslas.

KOLTCHINSKII, V. (1994a). Bahadur–Kiefer approximation for spatial quantiles. In *Probability in Banach Spaces* (J. Hoffmann-Jorgensen, J. Kuelbs and M. B. Marcus, eds.) 394–408. Birkhäuser, Boston.

KOLTCHINSKII, V. (1994b). Nonlinear transformations of empirical processes: functional inverses and Bahadur-Kiefer representations. In *Probability Theory and Mathematical Statistics. Proceedings of the Fifth International Vilnius Conference* (B. Grigelionis, J. Kubilius, et al., eds.) 423–445. VSP-TEV, The Netherlands.

KOLTCHINSKII, V. (1995a). Spatial quantiles and L-estimators. In *Exploring Stochastic Laws* (A. Skorohod and Yu. Borovskih, eds.) 183–200. VSP Scientific Publishers, The Netherlands.

KOLTCHINSKII, V. (1995b). Differentiability of inverse operators and limit theorems for inverse functions. Unpublished manuscript.

KOLTCHINSKII, V. (1996). M-estimation and spatial quantiles. In *Robust Statistics, Data Analysis and Computer Intensive Methods. Lecture Notes in Statist.* **109**. Springer, New York.

KOLTCHINSKII, V. and DUDLEY, R. M. (1996). On spatial quantiles. Unpublished manuscript.

KOLTCHINSKII, V. and LANG, L. (1996). A bootstrap test for spherical symmetry of a mutivariate distribution. Unpublished manuscript.

MATTNER, L. (1992). Completness of location families, translated moments, and uniqueness of charges. *Probab. Theory Related Fields* **92** 137–149.

MILASEVIC, P. and DUCHARME, G. R. (1987). Uniqueness of the spatial median. *Ann. Statist.* **15** 1332–1333.

NIEMIRO, W. (1992). Asymptotics for M-estimators defined by convex minimization. *Ann. Statist.* **20** 1514–1533.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

POLLARD, D. (1988). Asymptotics for least absolute deviation regression estimators. *Econometric Theory* **7** 186–199.

POLLARD, D. (1990). *Empirical Processes: Theory and Applications*. IMS, Hayward, CA.

PYKE, R. (1975). Multidimensional empirical processes: some comments. In *Stochastic Processes and Related Topics, Proceedings of the Summer Research Institute on Statistical Inference for Stochastic Processes* **2** 45–58. Academic Press, New York.

PYKE, R. (1985). Opportunities for set-indexed empirical and quantile processes in inference. In *Bulletin of the International Statistical Institute: Proceedings of the 45th Session, Invited Papers* **51** 25.2.1–25.2.12. ISI, Voozburg, The Netherlands.

RAO, C. R. (1988). Methodology based on the L_1 norm in statistical inference. *Sankhyā Ser. A* **50** 289–313.

REED, M. and SIMON, B. (1975). *Methods of Modern Mathematical Physics. II. Fourier Analysis, Self-Adjointness*. Academic Press, New York.

REISS, R.-D. (1989). *Approximate Distribution of Order Statistics*. Springer, New York.

ROCKAFELLAR, T. R. (1970). *Convex Analysis*. Princeton Univ. Press.

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

TUKEY, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians* (R. D. James, ed.) **2** 523–531. Canadian Math. Congress, Montreal.

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

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