KYBERNETIKA — VOLUME 20 (1984), NUMBER 4

CONSISTENCY OF D-ESTIMATORS

IGOR VAJDA

In the previous paper [11] published in Kybernetika three classes of D-estimators have been introduced. In the present paper Fisher's and strong consistency of these estimators are established. It is shown in particular that the standard D-estimators estimate the parameters from compact spaces strongly consistently for all discrete sample-generating families with the probabilities continuously depending on the parameter. Strong consistency of weak and directed D-estimators of location and scale is established for a wide variety of sample-generating models including irregular ones. Analogical results concerning abstract parametric spaces are presented too.

1. PRELIMINARIES

This paper is a continuation of the paper [11]. It is assumed that the reader is familiar with preliminaries, basic definitions and examples presented there.

In this paper we consider Θ-measurable parametric families $\mathscr{P}_\Theta$, $\mathscr{Q}_\Theta \subset \mathscr{P}$ only. We say that an estimator T: $\mathscr{P}(T) \to \Theta$ is *Fisher consistent* for a generating family $\mathscr{P}_\Theta \subset \mathscr{P}$ if

$$(1.1) \qquad \mathscr{P}_\Theta \subset \mathscr{P}(T) \quad \text{and} \quad \mathbb{T}(P_\theta) = \{\theta\} \quad \text{for all} \quad \theta = \Theta\,.$$

We say that T is *consistent* or *strongly consistent* for a generating family $\mathscr{P}_\Theta$ if it is well-defined and

$$(1.2) \qquad T(P_n) \xrightarrow{P_\theta \text{ or } [P_\theta]} \theta \quad \text{for all} \quad \theta \in \Theta$$

respectively, where $\xrightarrow{P}$ denotes the convergence in P^∞-probability for $n \to \infty$ and $\xrightarrow{[P]}$ denotes the convergence with P^∞-probability 1 for $n \to \infty$, for any $P \in \mathscr{P}$.

Let us consider arbitrary extended real valued functions $D_Q(\theta)$ on Θ, $Q \in \mathscr{P}$, and an estimator T: $\mathscr{P}(T) \to \Theta$ well-defined by the criterion

$$(1.3) \qquad T(Q) \text{ minimizes (maximizes) } D_Q(\theta) \text{ on } \Theta\,.$$

In the proofs that follow $T(Q)$ is supposed to minimize $D_Q(\theta)$ – in the other case the proofs are analogical.

Lemma 1.1. For every $P, Q \in \mathscr{P}(T)$

$$|D_Q(T(P)) - D_Q(T(Q))| \leqq 2\|D_Q - D_P\|_\infty = 2 \sup_{\theta\in\Theta} |D_Q(\theta) - D_P(\theta)| .$$

Proof. If $T(Q)$ minimizes $D_Q(\theta)$ then

$$0 \leqq D_Q(T(P)) - D_Q(T(Q)) \leqq D_Q(T(P)) - D_Q(T(Q)) + D_P(T(Q)) - D_P(T(P))$$

so that

$$|D_Q(T(P)) - D_Q(T(Q))| \leqq |D_Q(T(P)) - D_P(T(P))| + |D_Q(T(Q)) - D_P(T(Q))|$$

and the rest is clear. □

Lemma 1.2. If for each open neighborhood $U(\theta_0)$ of a point $\theta_0 \in \Theta$

$$\inf_{\theta\notin U(\theta_0)} D_Q(\theta) > D_Q(\theta_0) \quad (\sup_{\theta\notin U(\theta_0)} D_Q(\theta) < D_Q(\theta_0))$$

then $Q \in \mathscr{P}(T)$, $T(Q) = \{\theta_0\}$, and $\lim_{n\to\infty} D_Q(\theta_n) = D_Q(\theta_0)$ for some $\theta_n \in \Theta$ implies $\lim_{n\to\infty} \theta_n = \theta_0$.

Proof. Since the Hausdorff space Θ enables to separate any two different points, θ_0 is obviously a unique point of minima of D_Q. If $\lim_{n\to\infty} D_Q(\theta_n) = D(\theta_0)$ then, for every open $U(\theta_0)$ and every

$$0 < \varepsilon \leqq \inf_{\theta\notin U(\theta_0)} D_Q(\theta) - D_Q(\theta_0) ,$$

there exists n_0 such that, for $n > n_0$, $D_Q(\theta_n) < D_Q(\theta_0) + \varepsilon$, i.e. $\theta_n \in U(\theta_0)$. □

Lemmas 1.1, 1.2 obviously imply the following result.

Lemma 1.3. If $Q \in \mathscr{P}(T)$ and for each open neighborhood $U(T(Q))$ of $T(Q) \in \Theta$

$$\inf_{\theta\notin U(T(Q))} D_Q(\theta) > D_Q(T(Q)) \quad (\sup_{\theta\notin U(T(Q))} D_Q(\theta) < D_Q(T(Q)))$$

then either of the conditions $\|D_Q - D_{P_n}\|_\infty \xrightarrow{Q,[Q]} 0$, $D_Q(T(P_n)) \xrightarrow{Q,[Q]} D_Q(T(Q))$ implies $T(P) \xrightarrow{Q,[Q]} T(Q)$. □

2. CONSISTENCY OF STANDARD D-ESTIMATORS

In this section we consider projection familes $\mathscr{P}_\Theta \subset \mathscr{P}$ on a discrete sample space $\mathscr{X}$, extended real valued functions $D_Q(\theta) = D_f(P_\theta, Q)$ on $\Theta \times \mathscr{P}$, and well-defined standard D-estimators $T \triangleq \mathscr{P}_\Theta/D_f$ (cf. Sec. 3 in [11]). For arbitrary sample generating families $\mathcal{Q}_\Theta \subset \mathcal{P}$, the estimates $T(P_n)$ as well as the functions $D_{P_n}(\theta), \theta \in \Theta$, can thus be considered as r.v.'s defined on $(\mathcal{X}^n, \exp \mathcal{X}^n, Q_\theta^n)$ for $\theta \in \Theta$. This conclusion is equivalent with the statement that both $T(P_n)$, $D_{P_n}(\theta)$ are $\mathcal{B}^n$-measurable functions of a random vector $X = (X_1, \ldots, X_n)$ the sample probability space of which is $(\mathcal{X}^n, \exp \mathcal{X}^n, Q_\theta^n)$, $\theta \in \Theta$.

We say that a generating family $\mathcal{P}_\Theta$ is (*strongly*) *D_f-regular* if $D_{P_n}(\theta) \xrightarrow{P_\theta,\, [P_\theta]} 0$ on Θ (here and in the sequel the convergence $\xrightarrow{[P_\theta]}$ pertains to the strong alternative in the brackets; the notation $\xrightarrow{P_\theta([P_\theta])}$ would be perhaps more convenient but we want to avoid too many brackets in our notation).

Lemma 2.1. (A Glivenko theorem.) The whole family $\mathcal{P}$ is strongly D_f-regular for any f or any semibounded f depending on whether $\mathcal{X}$ is finite or countable respectively.

Proof. Let $P \in \mathcal{P}$ be arbitrary. If $\mathcal{X}$ is finite then the desired result follows from the strong law of large numbers, from the explicit formula for $D_{P_n}(\theta)$ given in Sec. 3 of [11], and from the continuity of f. If $\mathcal{X}$ is infinite then, by (2.5) in [11],

$$\chi^1(P, P_n) = \mathsf{E}_\lambda |p - p_n| = \mathsf{E}_\lambda\, 2(p - p_n)\, 1_{\{p > p_n\}} \,.$$

Since $0 \leqq 2(p - p_n)\, 1_{\{p > p_n\}} \leqq 2p$, the strong law of large numbers together with the Lebesgue dominated convergence theorem yield $\chi^1(P, P_n) \xrightarrow{[P_\theta]} 0$. The rest follows from the assumption $\|f\| < \infty$ and from the right inequality in Lemma 2.8 of [11]. □

A family $\mathcal{Q}_\Theta$ is said *D_f-compatible* with $\mathcal{P}_\Theta$ if for every open neighborhood $U(\theta)$ of $\theta \in \Theta$

$$\inf_{\tilde\theta \notin U(\theta)} D_{Q_\theta}(\tilde\theta) > D_{Q_\theta}(\theta) \quad \text{for all} \quad \theta \in \Theta \,. \tag{2.1}$$

A parameter (or a parametric space Θ) is said *identifiable* in $\mathcal{P}_\Theta$ if there exists a semibounded f such that $\mathcal{P}_\Theta$ is D_f-compatible with itself.

Lemma 2.2 A compact Θ is identifiable in each Θ-continuous family $\mathcal{P}_\Theta$.

Proof. Since $D_{P_\theta}(\tilde\theta) = D_f(P_{\tilde\theta}, P_\theta)$ is non-negative, the identifiability condition

$$\inf_{\tilde\theta \in \Theta - U(\theta)} D_f(P_{\tilde\theta}, P_\theta) > 0 \quad \text{for all} \quad \theta \in \Theta \quad \text{(cf. (2.1))} \tag{2.2}$$

is not satisfied only if, for some $\theta \in \Theta$ and $U(\theta) \subset \Theta$, the infimum is zero. For compact Θ this means $\lim_{n\to\infty} D_f(P_{\theta_n}, P_\theta) = 0$ for $\theta_n \in \Theta - U(\theta)$ with $\lim_{n\to\infty} \theta_n = \tilde\theta \in \Theta - U(\theta)$. By Θ-continuity of $\mathcal{P}_\Theta$, $\lim_{n\to\infty} p_{\theta_n}(x) = p_{\tilde\theta}(x)$ for all $x \in \mathcal{X}$. Similar argument as that employed in the proof of Lemma 2.1 yields $\lim_{n\to\infty} \chi^1(P_{\tilde{\theta}}, P_{\theta_n}) = 0$. The above stated convergence together with Lemma 2.8 in [11] implies $\lim_{n\to\infty} \chi^1(P_{\theta_n}, P_\theta) = 0$. Hence, by the triangle inequality

$$\chi^1(P_{\tilde{\theta}}, P_\theta) \leqq \chi^1(P_{\tilde{\theta}}, P_{\theta_n}) + \chi^1(P_{\theta_n}, P_\theta),$$

$\chi^1(P_{\tilde{\theta}}, P_\theta) = 0$ for $\theta \in \Theta$, $\tilde{\theta} \in \Theta - U(\theta)$ which obviously contradicts the assumption $P_{\tilde{\theta}} \neq P_\theta$ for $\tilde{\theta} \neq \theta$ (cf. Sec. 1 in [11] and Lemma 2.1 in [11]). □

Lemma 2.3. If Θ is compact and $\mathscr{P}_\Theta$ Θ-continuous and if $T \triangleq \mathscr{P}_\Theta/D_f$ for $\|f\| < \infty$ is Fisher consistent for $\mathscr{Q}_\Theta$ then $\mathscr{Q}_\Theta$ is D_f-compatible with $\mathscr{P}_\Theta$.

Proof. Analogically as in the proof above, if the assumptions hold and $\mathscr{Q}_\Theta$ is not D_f-compatible, there exist θ_n, $\tilde{\theta} \in \Theta - U(\theta)$ such that $\lim_{n\to\infty} D_f(P_{\theta_n}, Q_\theta) = D_f(P_\theta, Q_\theta)$, $\lim_{n\to\infty} \theta_n = \tilde{\theta}$. The Θ-continuity of $\mathscr{P}_\Theta$ implies $\lim_{n\to\infty} p_{\theta_n}(x) = p_{\tilde{\theta}}(x)$. We shall prove that this implies $\lim_{n\to\infty} D_f(P_{\theta_n}, Q_\theta) = D_f(P_{\tilde{\theta}}, Q_\theta)$ which obviously contradicts the Fisher consistency assumption since in this case $\{\theta, \tilde{\theta}\} \subset \mathbb{T}(Q_\theta)$ for $\tilde{\theta} \neq \theta$. The desired statement follows from the fact that, by (2.7), (2.8) in [11],

$$D_f(P_{\theta_n}, Q_\theta) = \sum_{\mathscr{X}} \Delta(p_{\theta_n}, q_\theta)$$

where

$$\Delta(p_{\theta_n}, q_\theta) = \frac{p_{\theta_n} + q_\theta}{2} \Psi\left(\left|\frac{p_{\theta_n} - q_\theta}{p_{\theta_n} + q_\theta}\right|\right)$$

and $0 < \Psi(u) \leqq 2u\|f\|$. Since this implies

$$0 \leqq \Delta(p_{\theta_n}, q_\theta) \leqq |p_{\theta_n} - q_\theta| \, \|f\| \leqq 2q_\theta\|f\| \quad \text{(cf. proof of Lemma 2.1)},$$

the desired assertion follows from the Lebesque dominated convergence theorem. □

Theorem 2.1. If Θ is identifiable in $\mathscr{P}_\Theta$ then $T \triangleq \mathscr{P}_\Theta/D_f$ is Fisher consistent for $\mathscr{P}_\Theta$. If, moreover, $\mathscr{P}_\Theta$ is (strongly) D_f-regular then T is (strongly) consistent for $\mathscr{P}_\Theta$.

Proof. (I) The identifiability means that (2.2) holds for a semibounded f, i.e. $D_f(P_{\tilde{\theta}}, P_\theta) \geqq D_f(P_\theta, P_\theta) = 0$ with the equality iff $\tilde{\theta} = \theta$. By Lemma 2.1 in [11], this statement remains true for any f. Therefore $P_\theta \in \mathscr{P}(T)$, $\mathbb{T}(P_\theta) = \{\theta\}$, $T(P_\theta) = \theta$ and the Fisher consistency condition (1.1) is proved.

(II) Let now $\mathscr{P}_\Theta$ be (strongly) D_f-regular. Since $D_{P_n}(T(P_n)) = D_{P_n}(\theta) + (D_{P_n}(T(P_n)) - D_{P_n}(\theta))$, where the first term tends to zero by the regularity assumption and the second one is non-positive by (3.1) in [11], it holds $D_{P_n}(T(P_n)) \xrightarrow{P_\theta, [P_\theta]} 0$. By Lemma 2.9 in [11] this implies $\tilde{D}_{P_n}(T(P_n)) \xrightarrow{P, [P_\theta]} 0$ for $\tilde{D}_Q(\theta) = D_{\tilde{f}}(P_\theta, Q)$, $\tilde{f}(u) = |1 - u|$ with $\|\tilde{f}\| = 2$. By Lemma 2.1 it holds $\tilde{D}_{P_n}(\theta) \xrightarrow{[P_\theta]} 0$ on Θ. By the triangle inequality

$$\tilde{D}_{P_\theta}(T(P_n)) \leqq \tilde{D}_{P_n}(\theta) + \tilde{D}_{P_n}(T(P_n))$$

and by the last two convergences it holds $\tilde{D}_{P_\theta}(T(P_n)) \xrightarrow{P_\theta, [P_\theta]} 0 = \tilde{D}_{P_\theta}(\theta) = \tilde{D}_{P_\theta}(T(P_\theta))$ on Θ. Using Lemma 2.8 in [11] this result can be extended from $\tilde{f}(u) = |1 - u|$ to all semibounded functions f. The identifiability implies that the assumptions of Lemma 1.3 hold for $D_Q(\tilde{\theta}) = D_{\tilde{f}}(P_{\tilde{\theta}}, Q)$, $Q = P_\theta$, for all $\tilde{\theta} \in \Theta$ and arbitrary fixed $\theta \in \Theta$. Hence, by Lemma 1.3, the above established convergence $\tilde{D}_{P_\theta}(T(P_n)) \xrightarrow{P_\theta, [P_\theta]} \tilde{D}_{P_\theta}(T(P_\theta))$ implies $T(P_n) \xrightarrow{P_\theta, [P_\theta]} T(P_\theta) = \theta$ for all $\theta \in \Theta$. □

Corollary 2.1. If Θ is identifiable in $\mathscr{P}_\Theta$ (e.g. if Θ is compact and $\mathscr{P}_\Theta$ is Θ-continuous) then all estimators $T \triangleq \mathscr{P}_\Theta/D_f$ with $\|f\| < \infty$ are strongly consistent for $\mathscr{P}_\Theta$. This strong consistency extends to all estimators with $\|f\| = \infty$ provided the support of $\mathscr{P}_\Theta$ is finite.

Example 2.1. The second statement of Corollary 2.1 may not be true without the restriction on support of $\mathscr{P}_\Theta$ as demonstrated by Hannan's [4] example of inconsistent MLE (cf. Example 3.2. in [11]). In Basu's [1] example of inconsistent MLE the support of $\mathscr{P}_\Theta$ is finite but the identifiability is violated ($\Theta = [0, 1] \subset R$ is compact but $\mathscr{P}_\Theta$ is Θ-discontinuous). Note that the parameter in all usual discrete families is identifiable so that the respective D-estimators are strongly consistent for these families.

The next theorem extends previous results to sample generating families $\mathscr{Q}_\Theta$ different from $\mathscr{P}_\Theta$.

Theorem 2.2. If $\mathscr{Q}_\Theta$ is D_f-compatible with $\mathscr{P}_\Theta$ then $T \triangleq \mathscr{P}_\Theta/D_f$ is Fisher consistent for $\mathscr{Q}_\Theta$. If, moreover, D_f is a metric divergence on $\mathscr{P}$ and $\mathscr{Q}_\Theta$ is (strongly) D_f-regular then T is (strongly) consistent for $\mathscr{Q}_\Theta$.

Proof. The Fisher consistency can be established analogically as in part (I) of the proof of Theorem 2.1. As to the strong consistency of T, by the triangle inequality

$$D_{Q_\theta}(T(P_n)) \leqq D_{P_n}(T(P_n)) + D^*_{P_n}(\theta), \quad \text{where} \quad D^*_{P_n}(\theta) = D_f(Q_\theta, P_n),$$

and by (2.1) it holds $D_{Q_\theta}(\theta) \leqq D_{Q_\theta}(T(P_n)) \leqq D_{P_n}(T(P_n)) + D^*_{P_n}(\theta)$. By the triangle inequality

$$D_{P_n}(\theta) \leqq D_{Q_\theta}(\theta) + D^*_{P_n}(\theta)$$

and (3.1) in [11] it holds $D_{P_n}(T(P_n)) \leqq D_{Q_\theta}(\theta) + D^*_{P_n}(\theta)$. Combining this and the previous result we get $D_{Q_\theta}(\theta) \leqq D_{Q_\theta}(T(P_n)) \leqq D_{Q_\theta}(\theta) + 2D^*_{P_n}(\theta)$. Since, by the D_f-regularity of $\mathscr{Q}_\Theta$, it holds $D^*_{P_n}(\theta) \xrightarrow{Q_\theta, [Q]} 0$ we get from here $D_{Q_\theta}(T(P_n)) \xrightarrow{Q_\theta, [Q_\theta]} D_{Q_\theta}(\theta) = D_{Q_\theta}(T(Q_\theta))$. Further, by (2.1), the assupmtions of Lemma 1.3 hold for $Q = Q_\theta$ with $\theta \in \Theta$ arbitrary fixed. Hence the last convergence together with Lemma 1.3 imply $T(P_n) \xrightarrow{Q_\theta, [Q_\theta]} T(Q_\theta) = \theta$ for all $\theta \in \Theta$. □

Corollary 2.2. The estimators $T \triangleq \mathscr{P}_\Theta / D_f$ with metric D_f are strongly consistent for all finite-support families $\mathscr{Q}_\Theta$ D_f-compatible with $\mathscr{P}_\Theta$. If $\|f\| < \infty$ then the restriction on support of $\mathscr{Q}_\Theta$ can be dropped out.

Using Lemma 2.3 together with the fact that χ^α-divergences defined in (2.5) of [11] are metrics with $\|f\| = 2$ for all $\alpha \in (0, 1]$, we get the following statement.

Corollary 2.3. If Θ is compact and $\mathscr{P}_\Theta$ is Θ-continuous and if, for some $\alpha \in (0, 1]$, the estimator $T \triangleq \mathscr{P}_\Theta / \chi^\alpha$ is Fisher consistent for $\mathscr{Q}_\Theta$, i.e.

$$\sum_{\mathscr{X}} |p_{\tilde\theta}^\alpha - q_\theta^\alpha|^{1/\alpha} \geqq \sum_{\mathscr{X}} |p_\theta^\alpha - q_\theta^\alpha|^{1/\alpha} \quad \text{for all} \quad \theta, \tilde\theta \in \Theta ,$$

then it is strongly consistent for $\mathscr{Q}_\Theta$.

Example 2.2. For the standard Hellinger-distance estimator $T \triangleq \mathscr{P}_\Theta / \chi^{1/2}$ (standard $D^{1/2}$-estimator, cf. (2.4), (2.5) in [11]) the above stated Fisher consistency condition reduces to

$$\sum_{\mathscr{X}} (p_{\tilde\theta} q_\theta)^{1/2} \leqq \sum_{\mathscr{X}} (p_\theta q_\theta)^{1/2} \quad \text{for all} \quad \theta, \tilde\theta \in \Theta . \tag{2.3}$$

This as well as the above stated general condition obviously holds for each pair $\mathscr{P}_\Theta$, $\mathscr{Q}_\Theta = \mathscr{P}_\Theta$. Therefore all estimators $T \triangleq \mathscr{P}_\Theta / \chi^\alpha$, $\alpha \in (0, 1]$, are strongly consistent for Θ-continuous $\mathscr{P}_\Theta$ with compact Θ. This result, however, follows from Corollary 2.1 too.

Example 2.2. (*A method of confidential supports.*) To employ specific resolving power of Corollaries 2.2, 2.3, we introduce a class of estimators based on "confidential supports" of probabilities $Q_\theta \in \mathscr{Q}_\Theta$. They can be used for quick-and-rough localization of parameters of all usual discrete families (through data-grouping, they can be applied to continuous families as well).

Let Θ be finite with discrete topology (usually a subset of an original parametric space) and let $\frac{1}{2} < c \leqq 1$ be a fixed confidence-level. For any $\mathscr{Q}_\Theta$ a c-level family of confidential supports $\mathscr{S}_\Theta = \{S_\theta : \theta \in \Theta\}$ is defined by conditions $S_\theta \subset \mathscr{X}$, card $S_\theta < \infty$ for all $\theta \in \Theta$ and

$$Q_\theta(S_\theta) = c,\ Q_{\tilde\theta}(S_\theta) \leqq c \quad \text{for all} \quad \theta, \tilde\theta \in \Theta \tag{2.4}$$

(conditions of existence of $\mathscr{S}_\Theta$ are described by the well-known variant of Neyman--Pearson lemma for finite number of hypotheses). For every $\mathscr{Q}_\Theta$, $\mathscr{S}_\Theta$, we define $\mathscr{P}_\Theta$ as a family of uniform probabilities $P_\theta = U(S_\theta)$ supported by S_θ, $\theta \in \Theta$. Let us suppose $P_{\tilde\theta} \neq P_\theta$ for $\tilde\theta \neq \theta$, which is a typical case when $c > 0$.

It follows from (2.4) that (2.3) holds for $\mathscr{P}_\Theta$, $\mathscr{Q}_\Theta$ under consideration. Hence, by Corollary 2.3, the standard Hellinger-distance estimator $T \triangleq \mathscr{P}_\Theta/\chi^{1/2}$ is strongly consistent for $\mathscr{Q}_\Theta$. T is adaptive in the sense that it is strongly consistent for various families satisfying (2.4) with a fixed $\mathscr{S}_\Theta$. Note that the MLE with this projection family $\mathscr{P}_\Theta$ is inconsistent for any $\mathscr{Q}_\Theta$ under consideration unless $c = 1$ which is quite a rare case (supports S_θ of generating probabilities Q_θ are finite and distinct for distinct θ's from Θ).

3. CONSISTENCY OF WEAK D-ESTIMATORS

In this section we consider projection familes $\mathscr{P}_\Theta \subset \mathscr{P}$ on an arbitrary sample space $(\mathscr{X}, \mathscr{B})$ with a class $\mathscr{E}$ sufficient for $\mathscr{X}$ (cf. Sec. 4 in [11]). Further we consider extended real valued functions $D_\Theta(\theta) = WD_f(P_\theta, Q) = \mathsf{E}_{\tilde{W}_\theta} d_f(F_\theta, G)\, \varphi(F_\theta, G)$ corresponding to factorfunctions φ and to families of factorweights $\tilde{\mathscr{W}}_\Theta$ dominated by a σ-finite measure λ on $(\mathscr{X}, \mathscr{B})$ with densities $\tilde{w}_\theta$. Finally, we consider well-defined weak D-estimators $T \triangleq \mathscr{P}_\Theta/\varphi\, \tilde{\mathscr{W}}_\Theta D_f$ (cf. Sec. 4 in [11]).

For arbitrary sample generating family $\mathscr{Q}_\Theta \subset \mathscr{P}$, the estimates $T(P_n)$ can be considered as r.v.'s defined on $(\mathscr{X}^n, \mathscr{P}^n, Q_\theta^n)$, $Q_\theta \in \mathscr{Q}_\Theta$. Further, due to the assumed Θ-measurability of $\mathscr{P}_\Theta$, $\mathscr{Q}_\Theta$, $D_Q(T(P_n))$ and $D_{P_n}(\theta)$ can be considered as r.v.'s defined on $(\mathscr{X}^n, \mathscr{B}^n, Q_\theta^n)$ too. In this section we employ the notation $\mathscr{W}_\Theta \triangleq \varphi\, \tilde{\mathscr{W}}_\Theta$ introduced in (2.12) of [11] so that e.g. $w_\theta(x) = \varphi(F_\theta(x), G(x))\, \tilde{w}_\theta(x)$ for $x \in \mathscr{X}$.

We say that $\mathscr{P}_\Theta$ is *(strongly) WD_f-regular* if $D_{P_n}(\theta) \xrightarrow{P_\theta,\,[P_\theta]} 0$ on Θ.

Lemma 3.1. (A Glivenko theorem.) Let $\mathscr{X} = \mathbb{R}$. (a) If $\|f\| < \infty$, $\varphi(u, v) = \tilde{\varphi}(u)$ for $(u, v) \in [0, 1]^2$, and for all $\theta \in \Theta$ either (ai) $W_\theta(\mathbb{R}) < \infty$, or (aii) $\mathscr{P}_\Theta \subset \bigcup_{\beta>1} \mathscr{P}_\beta$ and w_θ is essentially bounded on $\mathbb{R}$, or (aiii) $\mathscr{P}_\Theta \subset \bigcup_{\beta>1} \mathscr{P}_\beta$ and $\mathsf{E}_\lambda w_\theta^\beta < \infty$ for some $\beta > 1$, then $\mathscr{P}_\Theta$ is strongly WD_f-regular. (b) If $f(u) = |1 - u|^\alpha/\alpha$, $\alpha \geqq 1$ (cf. (2.5) in [11]), $\varphi(u, v) = [v(1 - v)]^{\alpha-1}/[v^{\alpha-1} + (1 - v)^{\alpha-1}]$ for $(u, v) \in [0, 1]^2$ and for all $\theta \in \Theta$ either (bi) $\tilde{W}_\theta(\mathbb{R}) < \infty$, or (bii) $\mathscr{P}_\Theta \subset \bigcup_{\beta>1} \mathscr{P}_\beta$ and $\tilde{w}_\theta$ is essentially bounded on $\mathbb{R}$, or (biii) $\mathscr{P}_\Theta \subset \bigcup_{\beta>1} \mathscr{P}_\beta$ and $\mathsf{E}_\lambda \tilde{w}_\theta^\beta < \infty$ for some $\beta > 1$, then $\mathscr{P}_\Theta$ is strongly $W\chi^\alpha$-regular.

Proof. (a) By Lemma 2.8 of [11] and by the definition of d_f in Corollary 2.1 of [11], $d_f(F, G) \leqq \|f\| \,.\, 2|F - G|$ so that, for every $P_n \in \mathscr{P}_e$,

$$D_{P_n}(\theta) = \mathsf{E}_{W_\theta} d_f(F_\theta, F_n) \leqq 2\|f\|\, \mathsf{E}_{W_\theta}|F_\theta - F_n|\,.$$

Further, by Hölder's inequality,

$$\mathsf{E}_{W_\theta}|F - F_n| \leqq (\mathsf{E}_\lambda w_\theta^\beta)^{1/\beta}\, (\mathsf{E}_\lambda |F_\theta - F_n|^\alpha)^{1/\alpha}$$

for all $\alpha, \beta \in [1, \infty]$, $\alpha^{-1} + \beta^{-1} = 1$. In particular, for $\beta = 1$ or $\alpha = 1$,

$$\mathsf{E}_{W_\theta}|F_\theta - F_n| \leqq \begin{cases} W_\theta(\mathbb{R}) \operatorname{ess\,sup}_{[\lambda]} |F_\theta - F_n| \leqq W_\theta(\mathbb{R}) \sup_{\mathbb{R}} |F_\theta - F_n| \\ \operatorname{ess\,sup}_{[\lambda]} w_\theta \,.\, \mathsf{E}_\lambda |F_\theta - F_n| \,. \end{cases}$$

Therefore

$$D_{P_n}(\theta) \leqq 2\|f\| \begin{cases} W(\mathbb{R}) \sup_{\mathbb{R}} |F_\theta - F_n| \\ (\mathsf{E}_\lambda w^\beta)^{1/\beta} (\mathsf{E}_\lambda |F_\theta - F_n|)^{(\beta-1)/\beta} \quad \text{for} \quad 1 < \beta < \infty \,. \\ \operatorname{ess\,sup}_{[\lambda]} w_\theta \,.\, \mathsf{E}_\lambda |F_\theta - F_n| \end{cases}$$

By Glivenko theorem $\sup_{\mathbb{R}} |F_\theta - F_n| \xrightarrow{[P\theta]} 0$ for every $P_\theta \in \mathscr{P}$ and, by Boos [2], p. 644, $\mathsf{E}_\lambda |F_\theta - F_n| \xrightarrow{[P\theta]} 0$ for every $P_\theta \in \mathscr{P}_\beta$, $\beta > 1$. Hence (a) holds.

(b) By (2.11) in [11] and by the assumptions considered in (b), $D_{P_n}(\theta) = \mathsf{E}_{\tilde{W}_\theta} |F_\theta - F_n|^\alpha$. Since $\alpha \geqq 1$ and $|F_\theta - F_n| \leqq 1$, it holds

$$D_{P_n}(\theta) \leqq \mathsf{E}_{\tilde{W}_\theta} |F_\theta - F_n|$$

and the rest of proof is the same as in (a). □

Using known generalizations of Glivenko theorem, Lemma 3.1 can be extended to $\mathscr{X} = \mathbb{R}^k$ or to even more general metric spaces $\mathscr{X}$.

A family $\mathscr{Q}_\Theta$ is said *WD_f-compatible* with $\mathscr{P}_\Theta$ if, for every open neighborhood $U(\theta)$, $D_{Q_\theta}(\tilde{\theta})$ satisfies (2.1). Θ is said *weakly identifiable* in $\mathscr{P}_\Theta$ if there exists a semi-bounded f such that $\mathscr{P}_\Theta$ is WD_f-compatible with itself.

Theorem 3.1. If Θ is weakly identifiable in $\mathscr{P}_\Theta$ then $T \triangleq \mathscr{P}_\Theta | \varphi \tilde{\mathscr{W}}_\Theta D_f$ is Fisher consistent for $\mathscr{P}_\Theta$. If, moreover, $\mathscr{P}_\Theta$ is (strongly) WD_f-regular then T is (strongly) consistent for $\mathscr{P}_\Theta$.

The proof is analogical to the proof of Theorem 2.1 and is thus omitted. □

Corollary 3.1. Let Θ be structural w.r.t. the real line $\mathscr{X} = \mathbb{R}$ and let $\tilde{\mathscr{W}}_\Theta$ be generated by a parent $\tilde{W} \ll \lambda$ and $\mathscr{P}_\Theta$ by a parent $P \in \mathscr{P}_\beta$ for some $\beta > 1$. If $\varphi(u, v) = \tilde{\varphi}(u)$ on $[0, 1]^2$ then $\mathscr{W}_\Theta$ is generated by the parent $W = \tilde{\varphi}\tilde{W}$ and, for the semi-bounded $\tilde{f}(u) = |1 - u|$,

$$\tilde{D}_{P_\theta}(\tilde{\theta}) = W_{\tilde{\theta}} D_f(P_{\tilde{\theta}}, P_\theta) = W_{\tilde{\theta}} \chi^1(P_{\tilde{\theta}}, P_\theta) =$$

$$= 2\mathsf{E}_{W_{\tilde{\theta}}} |F[\tilde{\theta}]^{-1} - F[\theta]^{-1}| = 2\mathsf{E}_W |F - F[\theta^{-1}\tilde{\theta}]| \; (\text{here } F[\theta] = F([\theta])) \,.$$

Thus Θ is identifiable in $\mathscr{P}_\Theta$ if for every neighborhood $U(e)$ of the group-unit $e \in \Theta$

$$\inf_{\theta \notin U(e)} \mathsf{E}_W |F - F[\theta]^{-1}| > 0 \,. \tag{3.1}$$

If this condition holds and $W(\mathbb{R}) < \infty$, or $w = \tilde{\varphi}(F)\,\tilde{w} < \infty$ on $\mathbb{R}$, or $\mathsf{E}_\lambda w^\beta = \mathsf{E}_\lambda(\tilde{\varphi}(F)\,\tilde{w})^\beta < \infty$ for some $\beta > 1$, then every weak D_f-estimator $T \triangleq P/WD_f$ wit $\|f\| < \infty$ is strongly consistent for $\mathscr{P}_\Theta$.

Example 3.1. For the location parameter (3.1) reduces to

$$\inf_{|\theta| \geqq \varepsilon} \mathsf{E}_W|F - F[\theta]^{-1}| > 0 \quad \text{for every} \quad \varepsilon > 0$$

and, for the location and scale parameter $\theta = (\mu, \sigma)$, (3.1) reduces to

$$\inf_{\substack{|\mu| \geqq \varepsilon_1 \\ \sigma \notin (\varepsilon_2, \varepsilon_2^{-1})}} \mathsf{E}_W|F - F[\mu, \sigma]^{-1}| > 0 \quad \text{for every} \quad \varepsilon_1 > 0,\ 0 < \varepsilon_2 < 1\,.$$

Since both these conditions hold under mild restrictions on parents W, P, strong consistency of weak D_f-estimators with $\|f\| < \infty$ seems to be satisfactorily answered by Corollary 3.1.

Weak χ^α-estimators of structural parameters with $\|f\| = \infty$ have not been mentioned in Corollary 3.1 in spite that Theorem 3.1 applies to this class of estimators too. It is so because these estimators belong to a wider class of weak D_f-estimators for which more universal form of consistency can be established.

Theorem 3.2. If $WD_f(P, Q) = \mathsf{E}_{\tilde{W}} d_f(F, G)\,\varphi(F, G)$ is a metric on $\mathscr{P}$ and $\mathscr{Q}_\Theta$ is WD_f-compatible with $\mathscr{P}_\Theta$, then $T \triangleq \mathscr{P}_\Theta/\varphi\tilde{\mathscr{W}}_\Theta D_f$ is Fisher consistent for $\mathscr{Q}_\Theta$. If, moreover, $\mathscr{Q}_\Theta$ is (strongly) WD_f-regular then T is (strongly) consistent for $\mathscr{Q}_\Theta$.

Proof is analogical to the proof of Theorem 2.2 and is thus omitted. □

Corollary 3.2. Let Θ be structural w.r.t. $\mathscr{X} = \mathbb{R}$, let $\mathscr{W}_\Theta$ be generated by a parent $W \ll \lambda$, and let $\mathscr{Q}_\Theta$ with a parent $Q \in \mathscr{P}_\beta$ for some $\beta > 1$ be $W\chi^\alpha$-compatible with $\mathscr{P}_\Theta$ generated by $P \in \mathscr{P}$ for some $\alpha \in (0, 1]$ (a compatibility condition analogical to (3.1) can be derived from (2.5), (2.11) in [11]). If $W(\mathbb{R}) < \infty$, or $w < \infty$ on $\mathbb{R}$, or $\mathsf{E}_\lambda w^\beta < \infty$ for some $\beta > 1$, then the corresponding estimator $T \triangleq P/W\chi^\alpha$ is strongly consistent for $\mathscr{Q}_\Theta$.

Corollary 3.3. Let Θ, $\mathscr{X}$, $\mathscr{P}_\Theta$, $\mathscr{Q}_\Theta$ be as in Corollary 3.2, let φ be as in part (b) of Lemma 3.1, let $\tilde{\mathscr{W}}_\Theta$ be generated by a parent $\tilde{W} \ll \lambda$ on $\mathbb{R}$, and let $\mathscr{Q}_\Theta$ be $W\chi^\alpha$-compatible with $\mathscr{P}_\Theta$ for some $\alpha \in [1, \infty)$, i.e.

$$\inf_{\theta \notin U(e)} \mathsf{E}_{\tilde{W}}|F - G[\theta]|^\alpha > \mathsf{E}_W|F - G|^\alpha \tag{3.2}$$

for every open neighborhood $U(e)$ of the unit $e \in \Theta$. If $\tilde{W}(\mathbb{R}) < \infty$, or $\tilde{w} < \infty$ on $\mathbb{R}$, or $\mathsf{E}_\lambda \tilde{w}^\beta < \infty$ for some $\beta > 1$ then $T \triangleq P/\varphi\tilde{W}\chi^\alpha$ is tsrongly consistent for $\mathscr{Q}_\Theta$.

Example 3.2. For the location parameter (3.2) holds whenever $\mathsf{E}_W|F - G[\theta]^{-1}|^\alpha$ is an increasing function of $|\theta|$ in a neighborhood of $\theta = 0$ and non-decreasing elsewhere. Obviously, under mild restrictions on $\tilde{W}$, F, G such a condition can be satisfied uniformly for all $\alpha \in [1, \infty)$. This in accordance with Corollary 3.3 implies strong consistency of all estimators of location $T \triangleq P/\varphi \tilde{W} \chi^\alpha$, $\alpha \in [1, \infty)$. For $\alpha = 2$ we get in this way strong consistency of weak χ^2-estimators of location considered by Boos [2] under little more general conditions than considered in Theorem 2.1 of [2] (there are infinite weights $\tilde{W} \ll \lambda$ with unbounded densities $\tilde{w}$ and bounded integrals $\mathsf{E}_\lambda w^\beta$ for some $\beta > 1$).

4. CONSISTENCY OF DIRECTED D-ESTIMATORS

In this section we consider well-defined directed D-estimators $T^\alpha \triangleq \mathscr{P}_\Theta/W$, $\alpha \in (0, 1]$, with arbitrary projection families $\mathscr{P}_\Theta \subset \mathscr{P}$ dominated by σ-finite measures W on a sample space $(\mathscr{X}, \mathscr{B})$ (cf. Sec. 5 in [11]). We exclude the MLE's from our considerations since their consistency has extensively been studied in the literature (see Sec. 5.3 of Zacks [12] and references given there when projection families are parametrized by $\Theta \subset \mathbb{R}$ and Theorem 1.1 on p. 240 of Ibragimov and Chasminskij [6] when Θ is abstract). We consider functions $D_Q(\theta) = \mathsf{E}_Q p_\theta^\alpha$, $\alpha \in (0, 1]$, on $\Theta \times \mathscr{P}$ where $p_\theta = \mathrm{d}P_\theta/\mathrm{d}W$. The estimates $T(P_n)$ as well as the functions $D_{P_n}(\theta)$, $\theta \in \Theta$, are obviously $\mathscr{B}^n$-measurable on $\mathscr{X}^n$. Therefore, for arbitrary sample generating family $\mathscr{Q}_\Theta \subset \mathscr{P}$, they can be considered as r.v.'s defined on $(\mathscr{X}^n, \mathscr{B}^n, Q_\theta^n)$ for $Q_\theta \in \mathscr{Q}_\Theta$.

$\mathscr{Q}_\Theta$ is said *α-compatible* with $\mathscr{P}_\Theta$ if for every open neighborhood $U(\theta)$ of θ

(4.1) $$\sup_{\tilde{\theta} \notin U(\theta)} D_{Q_\theta}(\tilde{\theta}) < D_{Q_\theta}(\theta) \quad \text{on} \quad \Theta .$$

Contrary to the D_f-compatibility in Section 2 or WD_f-compatibility in Section 3, the function figuring in (4.1) is neither f-divergence nor weak f-divergence. The fact that D is not a distance-like function significantly complicates verification of α-compatibility and the following space, up to Theorem 4.1, is devoted to various technical aspects of this verification.

We shall consider the following necessary and more easily verifiable conditions of α-compatibility

(4.2) $$D_{Q_\theta}(\tilde{\theta}) \leqq D_{Q_\theta}(\theta) \quad \text{for every} \quad \theta, \tilde{\theta} \in \Theta ,$$

(4.3) $$D_{Q_\theta}(\tilde{\theta}) \neq D_{Q_\theta}(\theta) \quad \text{for every} \quad \tilde{\theta} \neq \theta .$$

(Altogether, these conditions represent the Fisher consistency of the respective estimator T^α for $\mathscr{Q}_\Theta$. In the context of Sections 2 and 3 analogical conditions have been true for every $\mathscr{Q}_\Theta = \mathscr{P}_\Theta$ – it is not so here.) The next analogue of Lemma 2.3 summarizes conditions on $\mathscr{P}_\Theta$, under which the Fisher consistency conditions (4.2), (4.3) are sufficient for α-compatibility of the respective $\mathscr{Q}_\Theta$ with $\mathscr{P}_\Theta$.

Lemma 4.1. Let the densities $p_\theta(x)$ be continuous in θ and uniformly bounded a.e. $[W]$. If either (a) Θ is compact or (b) Θ is σ-compact and $\lim p_{\theta_j}(x) = 0$ a.e. $[W]$

for every $\theta_j \in \Theta - \Theta_j$ (cf. (1.2) in [11]), then (4.2), (4.3) are sufficient for α-compatibility of any $\mathcal{Q}_\Theta \ll W$ with $\mathcal{P}_\Theta$.

Proof. (a) Let $p_\theta(x) < K \in \mathbb{R}$ on $\Theta \times S_W$ where $S_W \in \mathcal{B}$ denotes a support of W. If $\theta_j \to \theta$ then $p_{\theta_j}(x)^\alpha \to p_\theta(x)^\alpha$ and $0 \leqq p_{\theta_j}(x)^\alpha$, $p_\theta(x)^\alpha \leqq K$ for $x \in S_W$, $\alpha \in (0, 1]$. Since any $\mathcal{B}$-measurable function equal K^α on S_W is absolutely Q-integrable for every $Q \in \mathcal{P}$, $Q \ll W$, the Lebesque dominated convergence theorem can be applied to $D_Q(\theta_j) = \mathsf{E}_Q p^\alpha_{\theta_j}$. It yields $D_Q(\theta_j) \to D_Q(\theta)$. Hence if $\mathcal{Q}_\Theta \ll W$ then $D_{Q_\theta}(\tilde{\theta})$ is continuous in $\tilde{\theta}$ on the compact Θ for each $\theta \in \Theta$. Since, by (4.2), (4.3), $D_{Q_\theta}(\tilde{\theta})$ is maximized on Θ at $\tilde{\theta} = \theta$ only, (4.1) obviously holds.

(b) In view of (a) it will suffice to prove that $\limsup_{j\to\infty} D_{Q_\theta}(\theta_j) < D_{Q_\theta}(\theta)$ for every $\theta_j \in \Theta - \Theta_j$, $\theta \in \Theta$. Since for every $\alpha \in (0, 1]$ $\lim_{j\to\infty} p_{\theta_j}(x)^\alpha = 0$ a.e. $[W]$, the Lebesque dominated convergence theorem yields $\lim_{j\to\infty} D_{Q_\theta}(\theta_j) = 0$. It further follows from (4.2), (4.3) that $D_{Q_\theta}(\theta) > 0$ which completes the proof. □

It is clear from the proof that if "a.e. W" is replaced by "everywhere on $\mathcal{X}$", then the restriction $\mathcal{Q}_\Theta \ll W$ can be dropped out in the lemma.

The next our aim is to prove that (4.2) holds for location families $\mathcal{Q}_\Theta = \mathcal{P}_\Theta$. Let $\mathcal{Y}$ be an at most countable non-empty set and π a permutation $\mathcal{Y} \to \mathcal{Y}$. We shall say that two functions $\varphi, \psi : \mathcal{Y} \to \mathbb{R}$ are *similar* if the set of those $(y, y') \in \mathcal{Y}^2$ for which $\varphi(y) > \varphi(y')$, $\psi(y) < \psi(y')$ is empty.

Lemma 4.2. (Auxiliary.) If φ is upper-bounded and p is a probability density w.r.t. counting λ on $\mathcal{Y}$ then

$$\max_\pi \sum_{\mathcal{Y}} \varphi p(\pi)$$

is attained on a non-empty class of permutations π for which φ and $p(\pi)$ are similar.

Proof. (I) We first prove that the class of permutations π for which φ and $p(\pi)$ are similar is non-empty. Since both φ and p attain maxima on $\mathcal{Y}$, the set $\mathcal{Y}$ can be well-ordered by descending values of φ and p respectively. Let $o(y), \tilde{o}(y) \in \{1, 2, \ldots\}$ be the respective natural orders of $y \in \mathcal{Y}$. Clearly, $o(y)^{-1}$ and $\varphi(y)$ as well as $\tilde{o}(y)^{-1}$ and $p(y)$ are similar. Denoting by $(y, \pi(y))$ the pair $(y, \tilde{y}) \in \mathcal{Y}^2$ for which $o(y) = \tilde{o}(y)$, we obviously define a permutation $\pi : \mathcal{Y} \to \mathcal{Y}$ (π need not be unique). We shall argue by contradiction that φ and $p(\pi)$ are similar. If $\varphi(y) > \varphi(y')$, $p(\pi(y)) < p(\pi(y'))$ for $(y, y') \in \mathcal{Y}^2$ then $o(y) < o(y')$, $p(\pi(y)) < p(\pi(y'))$,i.e. by the definition of π, $\tilde{o}(\pi(y)) < \tilde{o}(\pi(y'))$, $p(\pi(y)) < p(\pi(y'))$. Using the similarity of $\tilde{o}(y)^{-1}$ and $p(y)$ we obtain $p(\pi(y)) > p(\pi(y'))$, $p(\pi(y)) < p(\pi(y'))$.

(II) Now we prove that if φ and p are not similar then there exists a permutation π such that

$$\sum_{\mathcal{Y}} \varphi p(\pi) > \sum_{\mathcal{Y}} \varphi p\,.$$

Let $(y, y') \in \mathscr{Y}^2$ has the property $\varphi(y) > \varphi(y')$, $p(y) < p(y')$. We define $\pi : \mathscr{Y} \to \mathscr{Y}$ as an identity on $\mathscr{Y} - \{y, y'\}$ and $\pi(y) = y'$, $\pi(y') = y$. It holds

$$\sum_{\mathscr{Y}} \varphi(p(\pi) - p) = \varphi(y)[p(\pi(y)) - p(y)] + \varphi(y')[p(\pi(y')) - p(y')]$$
$$= \varphi(y)[p(y') - p(y)] + \varphi(y')[p(y) - p(y')]$$
$$= [\varphi(y) - \varphi(y')][p(y') - p(y)] > 0 .$$

(III) Lemma 4.2 follows from (I) and (II). □

Lemma 4.3. If $\mathscr{P}_R = \{P[\theta]^{-1} : \theta \in R\}$ is a location family equivalent with the Lebesque λ on $\mathscr{X} = R$ and if $p = dP/d\lambda$ is essentially $[\lambda]$ bounded on R then (4.2) holds for $\mathscr{Q}_\Theta = \mathscr{P}_\Theta$.

Proof. Since λ is invariant w.r.t. the shift $[\theta]$ on R, it holds $D_{P[\theta]^{-1}}(\tilde{\theta}) = D_{P[\theta - \tilde{\theta}]^{-1}}(0)$ and it will suffice to prove the inequality

(4.4) $$\mathsf{E}_{P[\theta]^{-1}} p^\alpha \leqq \mathsf{E}_P p^\alpha \quad \text{for} \quad \theta \in R, \alpha \in (0, 1] .$$

Suppose without loss of generality $\theta > 0$ and consider a net of decompositions $\mathscr{D}^{(j)} = \{E_i^{(j)} : i = 0, \pm 1, \ldots\}$, $j = 1, 2, \ldots$, of R into semiclosed intervals of equal length θ/j with $E_1^{(j)} = [0, \theta/j)$. It obviously holds

(4.5) $$\int_{E_{j+1}^{(j)}} p(x - \theta)\, dx = \int_{E_i^{(j)}} p(x)\, dx \quad \text{for all} \quad i = 0, \pm 1, \ldots$$

Let $p^{(j)} = \mathsf{E}_P(p/\mathscr{B}^{(j)})$ be the conditional expectation of p under σ-algebra $\mathscr{B}^{(j)}$ generated by $\mathscr{D}^{(j)}$, let $\mathscr{Y} = \mathscr{D}^{(j)}$, and let $p^{(j)}(y)$ denotes the a.e. $[\lambda]$ constant value of $p^{j)}(y)$ on the interval $y = E_i^{(j)}$, $y \in \mathscr{Y}$. Then

$$\mathsf{E}_P(p^{(j)})^\alpha = \sum_{\mathscr{Y}} (p^{(j)})^\alpha p^{(j)}, \quad \mathsf{E}_{P[\theta]^{-1}}(p^{(j)})^\alpha = \sum_{\mathscr{Y}} (p^{(j)})^\alpha p^{(j)}(\pi_j)$$

where, in accordance with (4.5), π_j is a permutation of $\mathscr{Y}$ defined by the condition $\pi_j(E_i^{(j)}) = E_{i-j}^{(j)}$ for all $i = 0, \pm 1, \ldots$ Since for every $j = 1, 2, \ldots$ the function $\varphi(y) = p^{(j)}(y)^\alpha \leqq p^{(j)}(y)$ is upper-bounded on $\mathscr{Y}$ and $p^{(j)}$ is a probability density on $\mathscr{Y}$, the assumptions of Lemma 4.2 hold for $\varphi = (p^{(j)})^\alpha$, $p = p^{(j)}$, $j = 1, 2, \ldots$ Thus by Lemma 4.2

$$\mathsf{E}_{P[\theta]^{-1}}(p^{(j)})^\alpha \leqq \mathsf{E}_P(p^{(j)}) \quad \text{for} \quad j = 1, 2, \ldots$$

yields $p^{(j)} \to p$ a.s. $[P]$ and, consequently, a.s. $[P[\theta]^{-1}]$ for every $\theta \in R$. The assumed boundedness of p implies a uniform boundedness of all p^α, $(p^{(j)})^\alpha$, $\alpha \in (0, 1]$. Hence, by the Lebesque dominated convergence theorem,

$$\lim_{j \to \infty} \mathsf{E}_{P[\theta]^{-1}}(p^{(j)})^\alpha = \mathsf{E}_{P[\theta]^{-1}} p^\alpha \quad \text{for all} \quad \theta \in R, \quad \alpha \in (0, 1] .$$

This identity together with the last inequality yield (4.4). □

Theorem 4.1. Let the topology of $\mathscr{X}$ be defined by a pseudo-metric d and let $\mathscr{P}_\Theta$ be arbitrary with densities p_θ, $\theta \in \Theta$, uniformly bounded on $\mathscr{X}$ and uniformly continuous in the sense that for every $\varepsilon > 0$ there exists $\delta > 0$ such that $d(x, x') < \delta$ implies $|p_\theta(x) - p_\theta(x')| < \varepsilon$ for all $\theta \in \Theta$. Then the estimators $T^\alpha \triangleq \mathscr{P}_\Theta / W$, $\alpha \in (0, 1]$, are strongly consistent for all generating families $\mathscr{Q}_\Theta \ll W$ α-compatible with $\mathscr{P}_\Theta$.

Proof. (I) Let $\theta \in \Theta$ and $\varepsilon > 0$ be arbitrary fixed. We shall prove that there exists n_0 such that $\|D_{Q_\theta} - D_{P_n}\|_\infty < \varepsilon$ a.s. $[Q_\theta]$ for all $n > n_0$. The second axiom of countability in the pseudo-metric space $\mathscr{X}$ implies the existence of a countable subset $\mathscr{Y} \subset \mathscr{X}$ dense in $\mathscr{X}$. Let $\mathscr{D}^{(j)} = \{E_y^{(j)} : y \in \mathscr{Y}\}$, $j = 1, 2, \ldots$, be measurable decompositions generating increasing σ-algebras $\mathscr{B}^{(j)}$ the union of which generates $\mathscr{B}$ and suppose that the diameter of all $E_y^{(j)}$ is less than j^{-1} (such decompositions obviously exist). Without loss of generality suppose $W(E_y^{(j)}) > 0$ for all j, y since, otherwise, $\mathscr{X}$ can be replaced by $\mathscr{X} - E_y^{(j)}$ and $E_y^{(j)}$ deleted from $\mathscr{D}^{(j)}$. Let $P_n^{(j)}$, $Q^{(j)}$, $W^{(j)}$ be restrictions of P_n, Q, W on $\mathscr{B}^{(j)}$. Since $P_n^{(j)}$ $Q^{(j)} \ll W^{(j)}$, let $p_n^{(j)} = \mathrm{d}P_n^{(j)}/\mathrm{d}W^{(j)}$, $q^{(j)} = \mathrm{d}Q^{(j)}/\mathrm{d}W^{(j)}$ and denote by $\tilde{P}_n^{(j)}$, $\tilde{Q}_n^{(j)}$ extensions of $P_n^{(j)}$, $Q^{(j)}$ back to $\mathscr{B}$ with Radon-Nikodym densities $\mathrm{d}\tilde{P}_n^{(j)}/\mathrm{d}W = p_n^{(j)}$, $\mathrm{d}\tilde{Q}^{(j)}/\mathrm{d}W = q^{(j)}$. It holds for every $j = 1, 2, \ldots$

$$|\mathsf{E}_{Q_\theta} p_{\tilde{\theta}}^\alpha - \mathsf{E}_{P_n} p_{\tilde{\theta}}^\alpha| \leqq |\mathsf{E}_{Q_\theta} p_{\tilde{\theta}}^\alpha - \mathsf{E}_{\tilde{Q}_\theta^{(j)}} p_{\tilde{\theta}}^\alpha + \\ + |\mathsf{E}_{\tilde{P}_n^{(j)}} p_{\tilde{\theta}}^\alpha - \mathsf{E}_{P_n} p_{\tilde{\theta}}^\alpha| + |\mathsf{E}_{\tilde{Q}_\theta^{(j)}} p_{\tilde{\theta}}^\alpha - \mathsf{E}_{P_n^{(j)}} p_{\tilde{\theta}}^\alpha| ,$$

i.e.

$$(4.6) \qquad |D_Q(\tilde{\theta}) - D_{P_n}(\tilde{\theta})| \leqq |\mathsf{E}_W p_{\tilde{\theta}}^\alpha (q_\theta - q_\theta^{(j)})| + \\ + \frac{1}{n} \sum_{i=1}^n \left| W(D_{y_i}^{(j)})^{-1} \int_{D_{y_i}^{(j)}} p_{\tilde{\theta}}^\alpha \,\mathrm{d}W - p_{\tilde{\theta}}(x_i)^\alpha \right| + \sup_{\mathscr{X}} p_{\tilde{\theta}} \chi^1(\tilde{Q}_\theta^{(j)}, P_n^{(j)}) ,$$

where $D_{y_i}^{(j)}$, $y_i \in \mathscr{Y}$, are defined by the conditions $x_i \in D_{y_i}^{(j)} \in \mathscr{D}^{(j)}$, $i = 1, 2, \ldots, n$.

Let us first consider the obvious inequalities

$$|\mathsf{E}_W p_{\tilde{\theta}}^\alpha (q_\theta - q_\theta^{(j)})| \leqq \sup_{\mathscr{X}} p_\theta \, \mathsf{E}_W |q_\theta - q_\theta^{(j)}| \leqq K \, \mathsf{E}_W |q_\theta - q_\theta^{(j)}| .$$

The σ-finite measure W can be decomposed into mutually singular probabilities P_s on $\mathscr{X}$ as follows

$$W = \sum_{s=1}^\infty u_s P_s , \quad u_s > 0 .$$

Further, for every $s = 1, 2, \ldots$, $q_\theta^{(j)} = \mathsf{E}_W(q_\theta / \mathscr{B}^{(j)}) = u_s \, \mathsf{E}_{P_s}(q_\theta / \mathscr{B}^{(j)})$ a.s. $[P_s]$ where q_θ is P_s-integrable. Hence the above mentioned theorem of Lévy implies $|q_\theta^{(j)} - q_\theta| \xrightarrow{[P_s]} 0$ for every $s = 1, 2, \ldots$, and, consequently, $|q_\theta^{(j)} - q_\theta| \to 0$ a.e. $[W]$. Thus the Scheffé theorem yields $\mathsf{E}_W |q^{(j)} - q| \to 0$ and there exists j_0 such that for all $j > j_0$

$$|\mathsf{E}_W p_{\tilde{\theta}}^\alpha (q_\theta - q_\theta^{(j)})| < \varepsilon/3 \quad \text{for all} \quad \tilde{\theta} \in \Theta .$$

As to the second right-hand term in (4.6), notice that the uniform continuity of

p_θ implies that for every fixed $\alpha \in (0, 1]$ there exists j_1 such that $j > j_1$ (i.e. the diameter of all $D_{y_i}^{(j)}$ less than j_1^{-1}) implies

$$|p_{\tilde\theta}(x_i)^\alpha - \sup_{x \in D_{y_i}^{(j)}} p_{\tilde\theta}(x)^\alpha| < \varepsilon/3 \,, \quad |p_{\tilde\theta}(x_i) - \inf_{x \in D_{y_i}^{(j)}} p_{\tilde\theta}(x)^\alpha| < \varepsilon/3$$

for all $\tilde\theta \in \Theta$ and $i = 1, 2, \ldots, n$, so that

$$\frac{1}{n} \sum_{i=1}^{n} \left| W(D_{y_i}^{(j)})^{-1} \int_{D_{y_i}^{(j)}} p_{\tilde\theta}^\alpha \, \mathrm{d}W - p_{\tilde\theta}(x_i)^\alpha \right| < \frac{\varepsilon}{3} \quad \text{for all} \quad \tilde\theta \in \Theta \,.$$

Let us now consider the third right-hand term in (4.6) with arbitrary $j > \max\{j_0, j_1\}$. The r.v. $\chi^1(\tilde{Q}_\theta^{(j)}, \tilde{P}_n^{(j)})$ defined on $(\mathscr{X}^n, \mathscr{B}^n, Q_\theta^n)$ has the same distribution as the r.v. $\chi^1(Q_\theta^*, P_n^*)$ defined on $(\mathscr{Y}^n, (\exp \mathscr{Y})^n, (Q_\theta^*)^n)$ where $Q_\theta^*(y) = Q_\theta(E_y^{(j)})$ for all $y \in \mathscr{Y}$ and P_n^* denotes an empirical probability defined by (1.1) in [11] for a random sample vector $Y = (Y_1, \ldots, Y_n)$ with sample probability space $(\mathscr{Y}^n, (\exp \mathscr{Y})^n, (Q_\theta^*)^n)$. Since by Lemma 2.1 $\mathscr{Q}_\Theta^* = \{Q_\theta^* : \theta \in \Theta\}$ is strongly χ^1-regular, we see that there exits n_0 such that for all $n > n_0$

$$\chi^1(Q_\theta^*, P_n^*) < \frac{\varepsilon}{3K} \quad \text{a.s.} \quad [Q_\theta^*]$$

for a fixed constant K satisfying the condition

$$\sup_{\mathscr{X}} p_{\tilde\theta} < K \,.$$

Therefore the third right-hand term in (4.6) can be, for $j > \max\{j_0, j_1\}$ and $n > n_0$, upper-estimated as follows

$$\sup_{\mathscr{X}} p_{\tilde\theta} \chi^1(\tilde{Q}^{(j)}, \tilde{P}_n^{(j)}) < \varepsilon/3 \quad \text{a.s.} \quad [Q_\theta] \quad \text{for all} \quad \tilde\theta \in \Theta \,.$$

Combining now for $j > \max\{j_0, j_1\}$ the upper bounds above established for all three right-hand terms in (4.6), we see from (4.6) that for all $n > n_0$

$$|D_{Q_\theta}(\tilde\theta) - D_{P_n}(\tilde\theta)| < \varepsilon \quad \text{a.s.} \quad [Q_\theta] \quad \text{for all} \quad \tilde\theta \in \Theta \,.$$

which implies the desired result.

(II) It follows from (4.1) that, for $\mathscr{Q}_\Theta$ α-compatible with $\mathscr{P}_\Theta$, it holds $\mathscr{Q}_\Theta \subset \mathscr{P}(T^\alpha)$ and $T^\alpha(Q_\theta) = \{\theta\}$ i.e. $T^\alpha(Q_\theta) = \theta$ for all $\theta \in \Theta$ (the Fisher consistency of $T^\alpha \triangleq \mathscr{P}_\Theta/W$ for $\mathscr{Q}_\Theta$).

(III) It follows from (II) and from the assumption of α-compatibility that the assumptions of Lemma 1.3 hold for $Q = Q_\theta$ and arbitrary fixed $\theta \in \Theta$. Since by (I) $\|D_{Q_\theta} - D_{P_n}\|_\infty \xrightarrow{[Q_\theta]} 0$ for all $\theta \in \Theta$, it follows from Lemma 1.3 $T(P_n) \xrightarrow{[Q_\theta]} \theta$ for all $\theta \in \Theta$. □

The next assertion follows directly from Lemma 4.1, Theorem 4.1 and from the mean value theorem on the real line.

Corollary 4.1. If $\mathscr{X} = \mathbb{R}$, if $p_\theta(x)$ are bounded and continuous on $\Theta \times \mathbb{R}$ and differentiable in x with a derivative $\dot{p}_\theta(x) = \mathrm{d}p_\theta(x)/\mathrm{d}x$ bounded on $\Theta \times \mathbb{R}$ and, finally, if Θ is σ-compact with $\lim_{n\to\infty} p_{\theta_j}(x) = 0$ on $\mathbb{R}$ for any $\theta_j \in \Theta - \Theta_j$ (cf. (1.1) in [11]), then $T^\alpha \triangleq \mathscr{P}_\Theta // \lambda$ is strongly consistent for all generating families $\mathscr{Q}_\Theta \ll \lambda$ satisfying the inequality

$$\int_{\mathbb{R}} p_{\tilde{\theta}}(x)^\alpha \, q_\theta(x) \, \mathrm{d}x \leqq \int_{\mathbb{R}} p_\theta(x)^\alpha \, q_\theta(x) \, \mathrm{d}x \quad (\text{cf. } (4.2), (4.3))$$

with equality iff $\tilde{\theta} = \theta$. □

From this Corollary and Lemma 4.3 one could obtain strong consistency of estimators of location T^α with boundedly differentiable projection parent densities p for generating families coinciding with the projection families provided

$$(4.7) \qquad \int_{\mathbb{R}} p(x)^\alpha \, p(x - \theta) \, \mathrm{d}x \neq \int_{\mathbb{R}} p(x)^{\alpha+1} \, \mathrm{d}x \quad \text{for} \quad \theta \neq 0 \quad (\text{cf. } (4.3)) .$$

However, the bounded differentiability of p everywhere on $\mathbb{R}$ is too restrictive (it is not satisfied e.g. by the doubly exponential parent density). Thus we prefer to draw from Lemmas 4.1, 4.3 and from Theorem 4.1 a more specific corollary based on a uniform continuity of the shift $[\theta]$ on $\mathbb{R}$ and on the fact that this uniform continuity is in an obvious sense uniform for all $\theta \in \mathbb{R}$.

Corollary 4.2. If a parent density p of a location family $\mathscr{P}_R$ is uniformly continuous and bounded on $\mathbb{R}$ and either (4.7) or

$$(4.8) \qquad \int_{\mathbb{R}} p(x)^\alpha \, q(x - \theta) \, \mathrm{d}x \leqq \int_{\mathbb{R}} p(x)^\alpha \, q(x) \, \mathrm{d}x$$

with equality iff $\theta = 0$ holds, then $T^\alpha \triangleq P // \lambda$ is strongly consistent for $\mathscr{Q}_R = \mathscr{P}_R$ or for $\mathscr{Q}_R$ with the parent density q respectively.

Example 4.1. The normal, doubly exponential, Cauchy, etc. parent density p is uniformly continuous and bounded on $\mathbb{R}$ and it satisfies (4.7) for all $\alpha \in (0, 1]$. Therefore all estimators of location T^α, $\alpha \in (0, 1]$, with these projection families are strongly consistent for these generating families analogically as their maximum likelihood limits T^0 (cf. Le Cam [9] and Daniels [3]). Since however not only (4.7) but also (4.8) holds for the densities under consideration, we can argue for example that all estimators $T^\alpha \triangleq \mathrm{No}\,(0, 1)//\lambda$, $\alpha \in (0. 1]$, are strongly consistent not only for normal but also for doubly exponential, Cauchy etc. generating families. This property of the estimators under consideration does not extend to the MLE T^0. It is well known for example that the sample mean $T^0 \triangleq \mathrm{No}\,(0, 1)//\lambda$ (see Example 5.2 in [11]) is inconsistent for the Cauchy generating family. To illustrate the fact that new properties of "old" estimators can also be drawn from the results of this section, let us point out that the skipped mean (see Example 5.2 in [11]) is a strongly consistent estimator of location for all generating families the parent density of which satisfies the inequality

$$\int_{-1/2}^{1/2} (1 - 4x^2)\, q(x - \theta)\, \mathrm{d}x \leqq \int_{-1/2}^{1/2} (1 - 4x^2)\, q(x)\, \mathrm{d}x$$

with equality iff $\theta = 0$ (cf. Corollary 4.2.).

Unfortunately the use of Corollary 4.1 with more general structural parameter spaces Θ is limited by conditions (4.2), (4.3). It turns out to be impossible to extend Lemma 4.3 to such spaces (for $\Theta = \mathbb{R} \times (0, \infty)$ and $\mathcal{Q}_\Theta = \mathcal{P}_\Theta$ with parents $P = Q = \mathrm{No}\,(0, 1)$, $D_{Q_\theta}(\tilde{\theta})$ is minimized by $\tilde{\theta} = (\tilde{\mu}, \tilde{\sigma})$ equal $(\mu, \sigma\sqrt{(1 - \alpha)}) \neq \theta)$. Thus we develop in the next section an alternative approach to consistency of estimators T of structural parameters. This approach is specific in the sense that it takes into account the equivariance of these estimators discussed in Section 6 of [11].

5. CONSISTENCY OF DIRECTED D-ESTIMATORS OF STRUCTURAL PARAMETERS

In this section we continue considerations of Section 4. We restrict ourselves to the directed D^α-estimators $T^\alpha \triangleq P//W$, $\alpha \in (0, 1]$, of structural parameters. The projection families $\mathcal{P}_\Theta = \{P_\theta = P[\theta]^{-1} : \theta \in \Theta\} \ll W$ of these estimators are supposed to satisfy the following conditions:

(i) Θ is σ-compact and $\mathcal{X}$ pseudo-metric (there is a metric d on $\mathcal{X}$ free of the property $d(x, \tilde{x}) = 0$ implies $x = \tilde{x}$).

(ii) The mappings $[\theta] : \mathcal{X} \to \mathcal{X}$ are uniformly continuous in the sense that for every $\varepsilon > 0$ there exists $\delta > 0$ such that $d(x, x') < \delta$ implies $d([\theta](x), [\theta](x')) < \varepsilon$ for all $\theta \in \Theta$.

(iii) The projection parent density $p = \mathrm{d}P/\mathrm{d}W$ is uniformly continuous and bounded on $\mathcal{X}$.

(iv) It holds $W[\theta] \ll W$ and the Jacobians $J(\theta) = \mathrm{d}W[\theta]/\mathrm{d}W$ are continuous and bounded on $\mathcal{X}$.

(v) $\lim_{j \to \infty} p([\theta_j]^{-1}(x)) = 0$ for every $\theta_j \in \Theta - \Theta_j$ (cf. (1.2) in [11]).

(vi) T^α, $\alpha \in (0, 1]$, are equivariant (cf. Theorem 6.3 in [11]).

Note that "on $\mathcal{X}$" can be replaced in what follows by "a.e. W" without any impact on the results. If Θ is compact, the condition (v) is not considered. Along with projection families $\mathcal{P}_\Theta$ of above described properties we consider arbitrary generating families $\mathcal{Q}_\Theta = \{Q_\theta = Q[\theta]^{-1} : \theta \in \Theta\} \ll W$ with parent densities $q = \mathrm{d}Q/\mathrm{d}W$.

Lemma 5.1. T^α is strongly consistent for $\mathcal{Q}_\Theta$ iff $T^\alpha(P_n) \xrightarrow{[Q]} e$ where e is the unit of the group Θ.

Proof. Let $\Theta^* \subset \Theta$ and $\theta \in \Theta$ be arbitrary fixed where Θ^* is supposed to be measurable. It holds

$$Q^n_\theta(\{x : T^\alpha(P_n) \in \Theta^*\}) = Q^n([\theta]^{-1}(\{x : T^\alpha(P_n) \in \Theta^*\}))$$
$$\text{(cf. (1.3) in [11])}.$$

Since

$$P_n(E) = \frac{1}{n}\sum_{i=1}^n 1_E(x_i) = \frac{1}{n}\sum_{i=1}^n 1_{[\theta]^{-1}(E)}([\theta]^{-1}(x_i)),$$

it holds

$$[\theta]^{-1}(\{x : T^\alpha(P_n) \in \Theta^*\}) = \{x : T^\alpha(P_n[\theta]^{-1}) \in \Theta^*\} =$$
$$= \{x : \theta T^\alpha(P_n) \in \Theta^*\},$$

where the last equality follows from (vi) and from (6.1) in [11]. Thus it holds

$$Q^n_\theta(\{x : T^\alpha(P_n) \in \Theta^*\}) = Q^n(\{x : \theta T^\alpha(P_n) \in \Theta^*\})$$

i.e. the distribution of the r.v. $T^\alpha(P_n)$ defined on $(\mathscr{X}^n, \mathscr{B}_n, Q^n_\theta)$ is the same as the distribution of the r.v. $\theta T^\alpha(P_n)$ defined on $(\mathscr{X}^n, \mathscr{B}^n, Q^n)$. Therefore $T^\alpha(P_n) \xrightarrow{[Q_\theta]} \theta$ iff $\theta T^\alpha(P_n) \xrightarrow{[Q]} \theta$ i.e. iff (cf. (iv) in Sec. 1 of [11]) $T^\alpha(P_n) \xrightarrow{[Q]} e$. □

Lemma 5.2. If $\mathbb{T}^\alpha(Q)$ is a one-point set then $T^\alpha(P_n) \xrightarrow{[Q]} T^\alpha(Q)$.

Proof. By the assumptions (ii), (iii), (iv), the densities $p_\theta = \mathrm{d}P_\theta/\mathrm{d}W$ satisfying the relation $p_\theta(x) = J([\theta]^{-1})\, p([\theta]^{-1}(x))$ (cf. (6.3) in [11]) are uniformly bounded and uniformly continuous in the sense of Theorem 4.1. Thus by (I) in the proof of Theorem 4.1, $\|D_Q - D_{P_n}\|_\infty \xrightarrow{[Q]} 0$.

By definition of $\mathbb{T}^\alpha$, $\mathbb{T}^\alpha(Q) = \{T^\alpha(Q)\}$ implies the inequality $D_Q(\theta) \leqq D_Q(T^\alpha(Q))$ for all $\theta \in \Theta$ with the equality iff $\theta = T^\alpha(Q)$.

By (i)–(v) the assumptions of Lemma 4.1 hold and, using an argument similar to that used in the proof of Lemma 4.1, we see that the right-hand inequality*) in Lemma 1.3 follows from the result of the preceding paragraph. Combining Lemma 1.3 with the results of the preceding two paragraphs we obtain the desired result. □

Lemmas 5.1, 5.2 together with (6.1) in [11] yield the following result.

*) For equivalent and strictly unambiguous (in the sense $\mathbb{T}^\alpha(Q) = \{T^\alpha(Q)\}$) estimators T^α the following three conditions are equivalent: (a) $T^\alpha(Q) = e$, (b) $\mathscr{Q}_\Theta$ is α-compatible with $\mathscr{P}_\Theta$, (c) T^α is Fisher consistent for $\mathscr{Q}_\Theta$. To see this notice that if $T^\alpha(Q) = e$ then the above established right-hand inequality in Lemma 1.3 means nothing but the α-compatibility of $\mathscr{Q}_\Theta$ with $\mathscr{P}_\Theta$ which in turn obviously implies the Fisher consistency; $T^\alpha(Q) = e$ evidently follows from this consistency. This is why we replace here rather cumbersome α-compatibility introduced in Sec. 4 by a simple condition $T^\alpha(Q) = e$.

Theorem 5.1. T^α is strongly consistent for $\mathcal{Q}_\Theta$ if $\mathbb{T}^\alpha(Q) = \{e\}$ where e is the unit of Θ. A right-modified version $\tilde{T}^\alpha = T^\alpha \,.\, T^\alpha(Q)^{-1}$ of every estimator T^α under consideration is equivariant and strongly consistent for every family $\mathcal{Q}_\Theta$ the parent of which yields one-point $\mathbb{T}^\alpha(Q)$.

Note that the left-modified version $\tilde{T}^\alpha = T^\alpha(Q)^{-1} \,.\, T^\alpha$ is strongly consistent for $\mathcal{Q}_\Theta$ satisfying the conditions of Theorem 5.1 too, but this version is not equivariant unless Θ is commutative.

Theorem 5.1 appears to be a satisfactory general solution of consistency of estimators T^α of structural parameters. It is however desirable to establish additional conditions under which $Q \in \mathcal{P}(T^\alpha)$ yield one-point sets $\mathbb{T}^\alpha(Q)$ and to find out effective methods of evaluation of $T^\alpha(Q)$. In what follows we consider, in addition to (i)–(vi), the following conditions.

(vii) Θ is a subset of $\mathbb{R}^m$ with the Euclidean-norm topology containing a non-empty interior Θ° and no isolated points.

(viii) The derivatives

$$\psi_\theta(x) = \frac{\mathrm{d}}{\mathrm{d}\theta}\, p_\theta(x)^\alpha \in \mathbb{R}^m\,, \quad \psi'_\theta(x) = \left(\frac{\mathrm{d}}{\mathrm{d}\theta}\right)^{\mathrm{T}} \circ \psi_\theta(x) \in (\mathbb{R}^m)^m$$

exist for all $(\theta, x) \in \Theta^\circ \times \mathcal{X}$ with the components continuous and bounded on $\Theta^\circ \times \mathcal{X}$ (here $\mathrm{d}/\mathrm{d}\theta$ denotes the $1 \times m$ matrix $(\partial/\partial\theta_1, \ldots, \partial/\partial\theta_m)$, $\circ$ denotes the usual matrix product, and $^{\mathrm{T}}$ denotes the transposition).

We employ the following notation

$$D'_Q(\theta) = \mathsf{E}_Q\psi_\theta,\ D''_Q(\theta) = \mathsf{E}_Q\psi'_\theta \quad \text{for all} \quad \theta \in \Theta^\circ,\ Q \in \mathcal{P}\,.$$

Theorem 5.2. If $D''_Q(\theta)$ is negative definite on Θ° and there is a root $\theta = \theta(Q)$ of the equation $D'_Q(\theta) = 0$ on Θ°, then the root is unique, $Q \in \mathcal{P}(T^\alpha)$, and $\mathbb{T}^\alpha(Q) = \{\theta(Q)\}$.

Proof. By the assumed continuity and uniform boundedness of ψ_θ, ψ'_θ on $\mathcal{X}$ and by the Lebesque dominated convergence theorem

$$D'_Q(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta}\, D_Q(\theta)\,, \quad D''_Q(\theta) = \left(\frac{\mathrm{d}}{\mathrm{d}\theta}\right)^{\mathrm{T}} \circ \left(\frac{\mathrm{d}}{\mathrm{d}\theta}\, D_Q(\theta)\right) \quad \text{for all} \quad \theta \in \Theta^\circ\,, \quad Q \in \mathcal{P}\,.$$

The rest follows from the definition of $\mathbb{T}^\alpha(Q)$, $T^\alpha(Q)$ in Sec. 5 of [11] and from the standard results of mathematical analysis in $\mathbb{R}^m$. □

Example 5.1. Consider T^α-estimators of the parameter of location and scale $\theta = (\mu, \sigma) \in \Theta = \mathbb{R} \times (0, \infty)$ with the normal projection parent No(0, 1) and with the Lebesque directing measure $W = \lambda$ on $\mathcal{X} = \mathbb{R}$. Standard calculations yield

$$\psi_{\mu,\sigma}\left([\mu, \sigma](x)\right) = \frac{\alpha p(x)^\alpha}{\sigma^{1+\alpha}}\,[x, x^2 - 1]\,,$$

$$\psi'_{\mu,\sigma}([\mu, \sigma](x)) = \frac{\alpha p(x)}{\sigma^{2+\alpha}} \begin{bmatrix} \alpha x^2 - 1 , & \alpha x^3 - (2 + \alpha) x \\ \alpha x^3 - (2 + \alpha) x , & \alpha x^4 - (2\alpha + 3) x^2 + 1 + \alpha \end{bmatrix},$$

$$D'_P(\mu, \sigma) = \frac{(2\pi)^{-\alpha/2} \exp(\alpha\mu^2/2(\alpha + \sigma^2))}{(\alpha + \sigma^2)^{3/2}} [-\mu\sigma^{1-\alpha}, \sigma^{-\alpha}(\mu^2 + 1 - \alpha - \sigma^2)] ,$$

Differentiating $D'_P(\mu, \sigma)$ we find that $D''_P(\mu, \sigma)$ is negative definite in $\mathbb{R} \times (0, \infty)$. Obviously $(\mu(P), \sigma(P)) = (0, \sqrt{(1 - \alpha)})$ is a unique root of the equation $D'_P(\mu, \sigma) = 0$ in the domain $\mathbb{R} \times (0, \infty)$.

It is clear that in this situation assumptions (i), (iii), (vii) hold and assumption (vi) follows from Theorem 6.3 in [11]. All assumptions (i)–(viii) hold provided the group $\mathbb{R} \times (0, \infty)$ is replaced by an arbitrary subgroup $\Theta_\varepsilon = \mathbb{R} \times [\varepsilon, \varepsilon^{-1}]$, $\varepsilon \in (0, 1)$. (More exactly, the assumptions of Theorem 5.2 hold for $\alpha \in (0,1 - \varepsilon^2)$ only since for $\alpha \in [1 - \varepsilon^2, 1]$ there is no root of the equation $D'_P(\mu, \sigma) = 0$ in Θ_ε.)

Thus let us consider for arbitrary fixed $\varepsilon \in (0, 1)$ the estimators of location and scale $T^\alpha = (M^\alpha, S^\alpha) \triangleq \mathscr{P}_{\Theta_\varepsilon}//\lambda$, $\alpha \in (0, 1 - \varepsilon^2)$, with $\mathscr{P}_{\Theta_\varepsilon}$ generated by the normal parent $P = \text{No}(0, 1)$.

By Theorem 5.2, $T^\alpha(P)$ are one-point sets for all $\alpha \in (0, 1 - \varepsilon^2)$, namely $T^\alpha(P) = (M^\alpha(P), S^\alpha(P)) = (0, \sqrt{(1 - \alpha)})$. As $(0, \sqrt{(1 - \alpha)})$ differs from the unit $e = (0, 1)$ of the group Θ_ε, for each α under consideration, it follows from Lemmas 5.1, 5.2 that (M^α, S^α) is inconsistent for $\mathscr{P}_{\Theta_\varepsilon}$. But, by Theorem 5.1, the right-modified version

$$(\tilde{M}^\alpha, \tilde{S}^\alpha) = (M^\alpha, S^\alpha/\sqrt{(1 - \alpha)}) \tag{5.1}$$

is strongly consistent for $\mathscr{P}_{\Theta_\varepsilon}$.

Theorems 5.1, 5.2 can be employed to yield strongly consistent versions of the estimators (M^α, S^α) with non-normal projection families as well. In this respect the differentiability of projection parent density required by (viii) might become too restrictive (for example, the skipped mean – skipped deviation estimator (M^1, S^1) yield by the projection parent density (5.7) in [11] cannot be treated by this manner). In order to extend the above described method to the estimators $T^\alpha \triangleq \mathscr{P}_\Theta//W$ with non-regular projection parent densities p, we shall replace (viii) by the following alternative.

(viii*) The derivatives

$$\varphi_\theta(x) = \frac{\mathrm{d}}{\mathrm{d}\theta} (J(\theta^{-1})\, q_{\theta^{-1}}) \in \mathbb{R}^m , \quad \varphi'_\theta(x) = \left(\frac{\mathrm{d}}{\mathrm{d}\theta}\right)^{\mathrm{T}} \circ \varphi_\theta(x) \in (\mathbb{R}^m)^m$$

exist on $\Theta^\circ \times \mathscr{X}$ with the components continuous and bounded.

Since it is easy to verify that

$$\tilde{D}_Q(\theta) = \mathsf{E}_W p^\alpha J(\theta^{-1})^\alpha q_{\theta^{-1}} , \quad \tilde{D}'_Q(\theta) = \mathsf{E}_W p^\alpha \varphi_\theta , \quad \tilde{D}''_Q(\theta) = \mathsf{E}_W p^\alpha \varphi'_\theta$$

are identical on Θ° with

$$D_Q(\theta)\,, \quad \frac{\mathrm{d}}{\mathrm{d}\theta} D_Q(\theta)\,, \quad \left(\frac{\mathrm{d}}{\mathrm{d}\theta}\right)^{\mathrm{T}} \circ \left(\frac{\mathrm{d}}{\mathrm{d}\theta} D_Q(\theta)\right)$$

considered above for every $Q \in \mathscr{P}$, the following analogue of Theorem 5.2 holds when (i)–(vii) and (viii*) take place.

Theorem 5.2*. If $\tilde{D}''_Q(\theta)$ is negative definite on Θ° and there is a root $\theta = \theta(Q)$ of the equation $\tilde{D}'_Q(\theta) = 0$ on Θ°, then the root is unique, it coincides with the root defined by Theorem 5.2 provided the conditions of Theorem 5.2 hod, and $Q \in \mathscr{P}(T^\alpha)$, $\mathbb{T}^\alpha(Q) = \{\theta(Q)\}$.

Example 5.2. Using Theorems 5.1, 5.2* one can specify right-modified versions $(M^1, S^1/\sigma(Q))$ of the skipped mean–skipped deviation estimator which are strongly consistent for normal, Cauchy and other generating familes $\mathscr{Q}_{\Theta_\varepsilon}$ with the differentiable parent densities q provided $(0, \sigma(Q))$ is a solution to $\tilde{D}'_Q(\mu, \sigma) = 0$.

Note that Theorems 5.1, 5.2 or 5.1, 5.2* solve the problem of consistency of M-estimators of location and scale with the loss functions

$$M(x, \mu, \sigma) = C - p_{\mu,\sigma}(x)^\alpha = C - p((x - \mu)/\sigma)^\alpha\,, \quad C \in \mathbb{R}\,.$$

The ψ-functions of these estimators

$$\psi_{\mu,\sigma}(x) = \psi_{0,1}((x - \mu)/\sigma) \quad \text{(cf. (viii))} \tag{5.2}$$

may be either differentiable (then the consistency of the respective estimators is covered by Theorems 5.1, 5.2, see Example 5.1) or discontinuous (then the consistency is covered by Theorems 5.1, 5.2*, see Example 5.2) or even non-existing everywhere on $\mathbb{R}$. Theorems of Jurečková [7, 8] on the asymptotic representation of M-estimators of location yield in [7, 8] results on consistency of M-estimators with continuous as well as discontinuous monotone ψ-functions. Our results of Sections 4, 5 applied to the particular location model overlap with these results of [7, 8] only slightly because the typical ψ-functions (5.2) of estimators considered in Sections 4, 5 are redescending in the sense of Hampel [5] and thus not monotone as supposed in [7, 8].

(Received September 9, 1983.)

REFERENCES

[1] D. Basu: An inconsistency of the method of maximum likelihood. Ann. Math. Statist. *26* (1955), 144–145.

[2] D. D. Boos: Minimum distance estimators for location and goodness of fit. J. Amer. Statist. Assoc. *76* (1981), 663–670.

[3] H. E. Daniels: The asymptotic efficiency of a maximum likelihood estimator. In: Proc. 4th Berkeley Symp. Math. Statist. Prob. Vol. 1, 151–164. Berkeley Univ. Press, Berkeley 1961.

[4] J. Hannan: Consistency of maximum likelihood estimation of discrete distributions. In:

Contributions to Probability and Statistics (Essays in Honor of Harold Hotelling). Stanford Univ. Press, Stanford 1960, 249—257.

[5] F. R. Hampel: The change of variance curve and optimal redescending *M*-estimators. J. Amer. Statist. Assoc. *76* (1981), 643—648.

[6] I. A. Ibragimov and R. Z. Chasminskij: Asymptotic Theory of Estimation (in Russian). Nauka, Moskva 1979.

[7] J. Jurečková: Asymptotic representation of *M*-estimators of location. Math. Operationsforsch. Statist. Ser. Statist. *11* (1980), 61—73.

[8] J. Jurečková: Asymptotic behavior of *M*-estimators in nonregular cases. Statistics and Decisions *1* (1983), 323—340.

[9] L. Le Cam: On the asymptotic theory of estimation and testing. In: Proc. 3rd Berkeley Symp. Math. Statist. Prob. Vol. 1, 129—156. Berkeley Univ. Press, Berkeley 1965.

[10] R. Š. Lipcer and A. N. Širjajev: Statistics of Random Process (in Russian). Nauka, Moskva 1974.

[11] I. Vajda: Motivation, existence and equivariance of *D*-estimators. Kybernetika *20* (1984), 3, 189—208.

[12] S. Zacks: The Theory of Statistical Inference. J. Wiley, New York 1971.

Ing. Igor Vajda, CSc., Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation — Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 182 08 Praha 8. Czechoslovakia.