ON STABLE LAWS FOR ESTIMATING FUNCTIONS AND DERIVED ESTIMATORS

by

Pranab Kumar Sen

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1477

December 1984

ON STABLE LAWS FOR ESTIMATING FUNCTIONS AND DERIVED ESTIMATORS *

BY PRANAB KUMAR SEN

University of North Carolina, Chapel Hill

Stable laws for M-estimators, maximum likelihood and other estimators are obtained through parallel results for the estimating functions and relative compactness of some related estimating functional processes.

1. Introduction. Let $\{P_\theta\}_{\theta\varepsilon\Theta}$ be a family of probability measures on $(\mathcal{X},\mathcal{A})$, indexed by the parameter $\theta \varepsilon \Theta$, where the parameter space Θ is a subset of the p-dimensional Euclidean space R^p, for some $p \geq 1$. Let $\{X_i, i \geq 1\}$be a sequence of independent random vectors (r.v.) [not necessarily identically distributed (i.d.)], such that under P_θ , X_i has a probability density function (p.d.f.) $f_i(x,\theta)$, for $i \geq 1$. Let $\eta_i(x,\theta)$, $i \geq 1$ be R^p-valued functions on $\mathcal{X} \times \Theta$. Then, an *estimating function* (p-vector) may be defined [viz., Huber (1967), Hájek (1970), Inagaki(1970,1973), and others] as

$$(1.1) \qquad S_n(t) = \Sigma_{i=1}^n \eta_i(X_i, t) \ , \quad t \ \varepsilon R^p \ ,$$

and if $T_n = T(X_1,\ldots,X_n)$ be a R^p-valued r.v., such that for some given $\alpha > 0$,

$$(1.2) \qquad n^{-\alpha} \ S_n(T_n) \ \to 0, \text{ in probability, as } n \to \infty \ ,$$

then, T_n is termed a *derived estimator* of θ ; the *maximum likelihood estimator (MLE)*, Huber's *M-estimator* and some others all belong to this class of estimators derived from suitable estimating functions. In the literature, the specific case of $\alpha = \frac{1}{2}$ has been treated in detail [see the references cited above], where the asymptotic normality of $n^{-\frac{1}{2}}S_n(\theta)$ plays a vital role. It is quite

AMS Subject Classification Numbers: 60F05, 62E20, 62F12

Key Words and Phrases: Asymptotic distribution, estimating functions, M-estimators, maximum likelihood estimators, relative compactness, stable laws.

* Work supported by the Office of Naval Research, Contract No. N00014-83-K-0387

conceivable that in a general setup, for some $\alpha \geq \frac{1}{2}$, $n^{-\alpha}S_n(\theta)$ may have (asymptotically) a (multivariate) *stable distribution* (note that the *characteristic exponent of the stable law* in our notation corresponds to α^{-1}, and the particular cases of normal and Cauchy distributions correspond to $\alpha = \frac{1}{2}$ and 1, respectively) . In the normal case, the asymptotic (multi-)normality of the estimating function $n^{-\frac{1}{2}}S_n(\theta)$ and of the derived estimator (i.e., $n^{\frac{1}{2}}(T_n - \theta)$) are equivalent. The question may therefore arise whether a similar equivalence result holds when, in general, $n^{-\alpha}S_n(\theta)$ has asymptotically a stable law, for some $\alpha \geq \frac{1}{2}$, and the present note provides a (distributional-) invariance result in this direction. It is shown that under suitable regularity conditions, an asymptotic stable law for $n^{-\alpha}S_n(\theta)$, for some $\alpha \in [\frac{1}{2},1]$, ensures an equivalent stable law for $n^{1-\alpha}(T_n - \theta)$.

Along with the preliminary notions, the main theorem is presented in Section 2, and its proof is outlined in Section 3. The last section deals with M-estimators of location in the general multivariate case.

2. The main theorem. In the regular case of $\alpha = \frac{1}{2}$ (i.e., asymptotically normal law), it has been assumed [c.f. Inagaki (1973)] that $E\eta_i(X_i,\theta)$, $i \geq 1$ all exist and are continuously differentiable (with respect to θ); the differential coefficient matrices play the dominant role in the asymptotic equivalence results. In the general case, though $E||\eta_i(X_i,\theta)||^k$ may exist for every $k < \alpha^{-1}$, we may not be in a position to assume that $E\eta_i(X_i,\theta)$ is finite (particularly, when $\alpha = 1$) or the variance of $\eta_i(X_i,\theta) - \eta_i(X_i,\theta')$ is finite (when $\alpha > \frac{1}{2}$). Hence, to justify (1.2), in the general case, we assume that

(2.1) $n^{-\alpha}S_n(\theta)$ has asymptotically a stable law G with centering parameter 0, where we confine ourselves to

(2.2) $$\tfrac{1}{2} \leq \alpha \leq 1 .$$

We may note that for $\alpha > 1$, $n^{1-\alpha}$ converges to 0 as $n \to \infty$, so that even if

$n^{1-\alpha}(T_n - \theta)$ has asymptotically a non-degenerate distribution, T_n may fail to be a consistent estimator of θ, and hence, we may not have much interest in the asymptotic properties of $\{T_n\}$.

For every $d \in R^p$, we let

(2.3) $$U_i(X_i, \theta, d) = \eta_i(X_i, \theta + d) - \eta_i(X_i, \theta), \quad i \geq 1.$$

For an arbitrary block $B = (u,v]$ (where $u < v$ and both belong to R^p), we define the *increment functions* as

(2.4) $$U_i(B) = \Sigma_{\{j_k=0,1;1\leq k\leq p\}} (-1)^{p-\Sigma_k j_k} U_i(X_i, \theta, u + J(v-u)), \quad i \geq 1,$$

where $J = \mathrm{Diag}(j_1, \ldots, j_p)$, and let $\lambda(B)$ be the Lebesgue measure of B. Then, we assume that for every compact set $\Theta_c \subset \Theta$, there exists positive numbers d_o, D and an r (> 1), such that for every $i \geq 1$,

(2.5) $$E_\theta || U_i(X_i, \theta, d)||^r \leq D||d||^r, \ \forall d : \ ||d|| \leq d_o,$$
$$E_\theta || U_i(B)||^r \leq D[\lambda(B)]^r, \ \forall u,v : ||u|| \leq d_o, \ ||v|| \leq d_o.$$

Let us also denote by

(2.6) $$-\Gamma_i = \lim_{d \downarrow 0} \{d^{-1} E_\theta [U_i(X_i, \theta, de_1), \ldots, U_i(X_i, \theta, de_p)]\}, \quad i \geq 1,$$

where the e_j are p-vectors and e_j has 1 in the jth position and 0 elsewhere, for $j = 1, \ldots, p$. Let then

(2.7) $$\bar{\Gamma}_n = n^{-1} \Sigma_{i=1}^n \Gamma_i, \quad n \geq 1.$$

We assume that there exists a positive integer n_o, such that

(2.8) $\bar{\Gamma}_n$ is nonsingular (ns) for every $n \geq n_o$.

Note that if in (2.2), $\alpha = 1$, $n^{1-\alpha}$ is also equal to 1. In this case, we may need to assume that (2.5) holds for every $d_o > 0$, and this will be referred to as (2.5'). Also, in this case, we may need to strengthen (2.6)-(2.8) to :

(2.9) $$\lim_{n\to\infty} ||n^{-1}\Sigma_{i=1}^n E_\theta U_i(X_i, \theta, d) + \bar{\Gamma}_n d|| = 0, \ \forall \text{(fixed) } d \in R^p,$$

where $\bar{\Gamma}_n$ does not depend on d and it satisfies (2.8). Then, we have the following.

Theorem 1 . *If in (2.2),* $\alpha \in [\frac{1}{2},1)$, *then under (2.1), (2.5), (2.6) and (2.8)*

(2.10) $n^{1-\alpha}\,\bar{\Gamma}_n(T_n - \theta)$ *has asymptotically the stable law G ,*

where G is defined in (2.1). For $\alpha = 1$, *under (2.1), (2.5') and (2.9),*

(2.11) $\bar{\Gamma}_n(T_n - \theta)$ *has asymptotically the stable law G .*

The proof of the theorem is provided in the next section. Note that for $\alpha = \frac{1}{2}$, in (2.5), we may take $r = 2$, although the second moment may not exist for $\alpha > \frac{1}{2}$; for our purpose, $r > 1$ suffices. Also, our (2.5) is more easily verifiable than the sup-norm moment condition in Inagaki (1973) or others.

3. Proof of the main theorem. First, we proceed to construct a sequence of *estimating functional processes* and establish its *tightness (or relative compactness);* these are then incorporated in the proof of the main theorem.

For some arbitrary positive $K (< \infty)$, let $C = [-K,K]^p$ be a compact subset of R^p . For every $n (\geq 1)$ and (fixed) $\theta \in \Theta$, we consider a (vector-valued) stochastic process $W_n = \{ W_n(u) \; ; \; u \in C \}$ (belonging to the space $D^p[C]$, endowed with the (extended) Skorokhod J_1-topology) , where

$$(3.1) \qquad W_n(u) = n^{-\alpha} \Sigma_{i=1}^n [\, U_i(X_i,\theta,n^{\alpha-1}u) - E_\theta U_i(X_i,\theta,n^{\alpha-1}u)] \;, \quad u \in C.$$

Then, we have the following.

Lemma 3.1. For every $\alpha \in [\frac{1}{2},1)$, under (2.5) , or under (2.5') for $\alpha = 1$,

$$(3.2) \qquad \sup_{u \in C} ||W_n(u)|| \rightarrow 0 \text{ , in probability, as } n \rightarrow \infty .$$

Proof. Without any loss of generality, in (2.5) [or (2.5')], we let $r \in (1,2]$. Note that by (2.3) and (3.1), for every (fixed) $u \in C$, $W_n(u)$ involves (n) independent summands. Thus, using a version of the L^p-*convergence theorem* [viz., Chatterjee (1969)], we have, for every $r \in (1,2]$,

$$(3.3) \qquad E_\theta || W_n(u)||^r \leq 4\, n^{-r\alpha} \Sigma_{i=1}^n E_\theta ||U_i(X_i,\theta,n^{\alpha-1}u)||^r .$$

Now, by (2.5) (for $\alpha < 1$), for every $d_o > 0$, there exists an n_o , such that

$n^{\alpha-1}||u|| \le d_o$, for every $u \varepsilon C$ and $n \ge n_o$; for $\alpha = 1$, (2.5') ensures the same. Hence, for $n \ge n_o$, the right hand side (rhs) of (3.3) is bounded from above by

(3.4) $\quad 4D\, n^{-r\alpha}\, n^{-r(1-\alpha)}\;\; n\; ||u||^r \;\le\; c_r^* K^r n^{-(r-1)}$,

where c_r^* (< ∞) is a positive number independent of $u \varepsilon C$. Since the rhs of (3.4) converges to 0 as $n \to \infty$, by using the Markov inequality, we obtain that for finitely many (say, m) (arbitrary) points $u_1,\ldots,u_m$ (all belonging to C),

(3.5) $\quad [\, W_n(u_1),\ldots,W_n(u_m)] \to (0,\ldots,0)$, in probability, as $n \to \infty$.

Thus, to establish (3.2), it suffices to verify the tightness of $\{W_n\}$. Towards this, we define a block B = (u,v] (for u,v ε C) as in before (2.4), so that as in (2.4), the increment of W_n over the block B is given by

$$
\begin{aligned}
(3.6) \qquad W_n(B) &= \Sigma_{\{j_k=0,1;1\le k\le p\}} \;(-1)^{p-\Sigma_k j_k}\; W_n(u+J(v-u)) \\
&= n^{-\alpha}\, \Sigma_{i=1}^{n}\;\; \Sigma_{\{j_k=0,1;1\le k\le p\}}\; (-1)^{p-\Sigma_k j_k} . \\
&\qquad [\, U_i(X_i,\theta,n^{\alpha-1}(u+J(v-u))) - E_\theta U_i(X_i,\theta,n^{\alpha-1}(u+J(v-u)))] \\
&= n^{-\alpha}\, \Sigma_{i=1}^{n}[\, U_{ni}(B) - E_\theta U_{ni}(B)\,] \text{ , say,}
\end{aligned}
$$

where the $U_{ni}(B)$ are independent r.v., so that proceeding as in (3.3) and (3.4), we obtain that under (2.5) (for $n \ge n_o$) (when α <1) or (2.5') (when $\alpha = 1$),

(3.7) $\quad E_\theta||W_n(B)||^r \;\le\; c_r^*\, n^{-(r-1)}\;[\lambda(B)]^r$, for some r > 1,and every $B \subset C$.

This in accordance with the multiparameter version of the classical Billingsley (1968) inequality ensure the tightness of W_n . Q.E.D.

Next, we note that by (2.5), (2.6) and (2.7), for every $\alpha \varepsilon [\tfrac{1}{2},1)$,

(3.8) $\quad ||n^{-\alpha}\Sigma_{i=1}^{n}\; E_\theta U_i(X_i\;,\theta,\; n^{\alpha-1}u) \;+\; \bar{\Gamma}_n u\;|| \to 0$, as $n \to \infty$,

uniformly in u ε C, while, (2.9) is just the same result for $\alpha = 1$. Therefore, from (3.2) and (3.8) (or (2.9) for $\alpha = 1$), we obtain that as $n \to \infty$, (3.9) $\sup_{u \in C} || n^{-\alpha}\Sigma_{i=1}^{n} [\eta_i(X_i,\theta+n^{\alpha-1}u) - \eta_i(X_i,\theta) + \bar{\Gamma}_n u || \to 0$, in probability.

Now, by (2.1), $n^{-\alpha}\Sigma_{i=1}^{n}\eta_i(X_i,\theta) = n^{-\alpha}S_n(\theta)$ has asymptotically a stable law G (with centering parameter 0) , and without any loss of generality, we assume that G is nondegenerate (otherwise, the results will be trivial). For nondegenerate G , $n^{-\alpha}S_n(\theta)$ is $O_p(1)$ and is nondegenerate too. On the other hand, using (2.8) and (2.9), we claim that for every $\varepsilon > 0$, there exists a compact set C_ε (in R^p) such that on writing $T_n = \theta + n^{-(1-\alpha)}u_n$ (with T_n defined as in (1.2)),

(3.10) $P\{ u_n \in C_\varepsilon \mid \theta \} \geq 1 - \varepsilon$, for every $n \geq n_o$.

It may be noted that by virtue of (2.8), though u_n may not be unique, all such solutions are convergent-equivalent , in probability, and hence, for the asymptotic distribution of T_n , any one of these would be usable. Consequently, from (3.9) and (3.10), we obtain that as $n \to \infty$,

(3.11) $n^{-\alpha}S_n(T_n) - n^{-\alpha}S_n(\theta) + n^{1-\alpha}\bar{\Gamma}_n(T_n - \theta) \to 0$, in probability,

so that using (1.2) and (3.11), we arrive at (2.10) and (2.11). This completes the proof of the theorem.

We may note that the asymptotic linearity result in (3.9) has been used as the main tool in the proof of the asymptotic normality of $n^{\frac{1}{2}}(T_n - \theta)$ for the regular case of $\alpha = \frac{1}{2}$; it plays the same role in the general case of $\alpha \in [\frac{1}{2},1]$. One of the advantages of using the estimating function in (1.1) (instead of the usual likelihood function) is that it may be used to study the asymptotic behaviour of the MLE when the model may be incorrect. For example, if $h_i(x,\theta)$, $i \geq 1$ be the assumed p.d.f.'s for the r.v.'s X_i, $i \geq 1$, while the true p.d.f.'s are the $f_i(x,\theta)$, $i \geq 1$, the $\eta_i(.,\theta)$ will be defined in terms of the $h_i(.,\theta)$ whereas (2.5) through (2.9) can be verified with respect to the ture p.d.f.'s $f_i(.,\theta)$. This will reveal the robustness of the MLE for departures from the assumed model. A classical case relates to normal $h_i(.,\theta)$ against Cauchy $f_i(.,\theta)$, $i \geq 1$, where , we would have $\alpha = 1$ and a stable law of the Cauchy type.

4. Stable laws for M-estimators of location. Let $\{X_i = (X_{i1},\ldots,X_{ip})'\ ,\ i \geq 1\}$ be a sequence of independent i.d.r.v. having a p($\geq$ 1)-variate continuous distribution function (d.f.) F, such that the jth marginal d.f. is symmetric about a location θ_j , for j =1,...,p, $\theta = (\theta_1,\ldots,\theta_p)$ is the location vector. To estimate θ , we use a vector ψ of score functions $\psi_j(u)$, $u \in R$, j=1,...,p, where we assume that for each j, ψ_j is nondecreasing and skew-symmetric . Then, we have the set of estimating functions

$$(4.1)\qquad S_n(t) = \Sigma_{i=1}^{n} [\ \psi_1(X_{i1} - t_1)\ ,\ldots,\ \psi_p(X_{ip} - t_p)\ ]\ ,\ t \in R^p\ .$$

Note that marginally, each $\Sigma_{i=1}^{n}\psi_j(X_{ij} - \theta_j)$ has a distribution symmetric about 0, and $S_n(t)$ is nonincreasing in each of its p arguments. Hence, for (1.2), we may locate a closed rectangle for which $S_n(t) = 0$, and the centre of gravilty of this closed rectangle may be taken as the M-estimator of θ . The sample mean, median and MLE (vectors) are all particular cases of these M-estimators.

Whenever the ψ_j are continuous and satisfy a Lipschitz condition, it is easy to verify that (2.5) holds; we do not need the ψ_j to be bounded in this context. Moreover, if the ψ_j have continuous first order derivatives almost everywhere (a.e.) (i.e., the set of points for the discontinuities of the derivatives is of measure 0) or if the ψ_j are continuous while the marginal densities have all finite Fisher information, then (2.6) holds; we do not need (2.7)-(2.8)[as we are dealing with i.i.d.r.v.'s here]. (2.9) is of course more restrictive and demands some sort of linearity of the expected values of $\psi_j(X_{ij}-\theta_j-d_j) - \psi_j(X_{ij} - \theta_j)$. Thus, (2.10) holds under quite general regularity conditions, while for (2.11), we need a more precise linearity result on the expected score-differences. In this context, it may not be necessary to assume that the ψ_j are monotone. But, then one needs to assume that $S_n(\theta)$ has location 0 (in some meaningful way), and one needs to verify (1.2) also (as there may be multiple (non-equivalent) roots). Finally, for $\alpha < 3/4$, one may also allow jump discontinuities for the ψ_j (finitely often).For some related linearity results, we may refer to Jurečková and Sen (1981 a,b).

REFERENCES

BILLINGSLEY, P. (1968). Convergence of Probability Measures. John Wiley, New York.

CHATTERJEE, S.D. (1969). An L^p- convergence theore. Ann. Math. Statist. 40, 1068-1070.

HÁJEK, J. (1970). A characterization of limiting distributions of regular estimates. Zeitschrift Wahrsch. verw. Geb. 14, 323-330.

HUBER, P.J. (1967). The behaviour of maximum likelihood estimators under nonstandard conditions. Proc. Fifth Berkeley Symp. Math. Statist. Prob. 1, 221-233.

INAGAKI, N. (1970). On the limiting distribution of a sequence of estimators with uniformity property. Ann. Inst. Statist. Math. 22, 1-13.

INAGAKI, N. (1973). Asymptotic relations between the likelihood estimating function and the maximum likelihood estimator. Ann. Inst. Statist. Math. 25 , 1-26.

JUREČKOVÁ, J. AND SEN, P.K. (1981a). Invariance principles for some stochastic processes related to M-estimators and their role in sequential statistical inference. Sankhyā , Ser.A 43 , 191-210.

JUREČKOVÁ, J. AND SEN, P.K. (1981b). Sequential procedures based on M-estimators with discontinuous score functions. Jour. Statist. Plan. Infer. 5, 253-266.