M- , L- and R-estimators

Jana Jurečková

1. Introduction

The classical estimation procedures - e.g. the sample mean as an estimator of location and the least squares estimator of regression - are highly sensitive to the outlying observations and to the long-tailed distributions. It was illustrated in a variety of Monte Carlo studies (e.g., Andrews at al.(1972)), in the results on the characterization of the normal law through the admissibility of the sample mean and the least squares estimator with respect to the quadratic loss (Kagan,Linnik and Rao (1965, 1972)); in the studies of tail-behavior of location estimators (Jurečková (1979,1981)), among others.

Among the robust alternatives of the classical estimators, which are less sensitive to the deviations from a specific distribution shape, three broad classes play the most important role : M-estimators, L-estimators and R-estimators.

The aim of the present chapter is to describe these three classes of estimators, their finite-sample properties as well as asymptotic properties, first on the location and then on the regression case. We shall also touch the computionally easier one-step versions of these estimators and the mutual relations of the estimators. This account is far from being exhaustive; various other results concerning the robust estimators may be found in the bibliography.

2. Estimation of location

Let $X_1, X_2, \ldots$ be a sequence of independent observations from the population with the distribution function (d.f.) $F(x-\theta)$. The problem is that of estimating θ after observing $X_1, \ldots, X_n$. We shall assume, unless otherwise stated, that F is absolutely continuous with the symmetric density f. If we do not impose any other special conditions on F, we cannot take the sample mean as a convenient estimator of θ. We must then look for alternative procedures which are robust , i.e. relatively insensitive to the special shape of distribution.

Assume that F is an unknown member of a given family $\mathcal{F}$ of distribution functions. The choice of the estimation procedure then depends on $\mathcal{F}$ which may be as large as the family of all [symmetric] absolutely continuous d.f.'s, it may be a neighborhood of a fixed distribution or a finite set of distribution shapes. Estimating θ over a large $\mathcal{F}$ corresponds to the global point of view; the estimator which is not very poor whatever is $F \in \mathcal{F}$ is then paid for by the lower efficiency. Estimating θ over a small neighborhood $\mathcal{F}$ of a given distribution corresponds to the local point of view; for convenient neighborhoods there often exists an estimator which is minimax over $\mathcal{F}$.

2.1. M-estimators (estimators of maximum likelihood type)

The class of M-estimators was suggested by Huber (1964), who then studied their properties in a series of papers; the results may be also found in Huber's recent monograph (1981).

Let $X_1, X_2, \ldots$ be a sequence of independent observations from the population with the d.f. $F(x-\theta)$ such that F is absolutely continuous and $F(x)+F(-x)=1$, $x \in R^1$. Then <u>M-estimator</u> $M_n=M_n(X_1,\dots,X_n)$ is defined implicitly as a solution of the equation

$$\sum_{i=1}^{n} \psi(X_i-t) = 0 \qquad (2.1)$$

with respect to t, where ψ is an appropriate function attaining positive as well as negative values. If there are more solutions of (2.1) then M_n may be defined as that the nearest to a preliminary consistent estimator T_n of θ (and larger one, if there are two solutions equally distant from T_n; we may put $M_n = 0$, if there is no solution).

The function ψ is often selected nondecreasing and skew-symmetric. In such case may be M_n defined as

$$M_n = \frac{1}{2}(M_n^- + M_n^+) \qquad (2.2)$$

where

$$M_n^- = \sup\left\{t:\ \sum_{i=1}^{n}\psi(X_i-t) > 0\right\}$$
$$M_n^+ = \inf\left\{t:\ \sum_{i=1}^{n}\psi(X_i-t) < 0\right\} \qquad (2.3)$$

or, alternatively, M_n may be defined through the randomization: M_n is equal either to M_n^- or M_n^+, both with probability $1/2$.

If F hapens to be known and smooth, we can put $\psi(x) = -f'(x)/f(x)$, $x \in R^1$ and M_n then coincides with the maximum likelihood estimator (m.l.e.) of θ. Particularly, for $\psi(x) \equiv x$ is $M_n = \bar{X}_n$, which turns out to be the m.l.e. for the normal distribution. The class of M-estimators covers also the sample median (which corresponds to $\psi(x) = \operatorname{sign} x$).

Various ψ-functions lead to various M-estimators; the question is then that of the proper choice of ψ. We intuitively feel that, if M_n is to be resistant to the outliers and to long-tailed distributions, we should take a bounded ψ-function. The most utilized function ψ is that suggested by Huber (1964)

$$\psi(x) = \begin{cases} x & \text{if } |x| \leq c \\ c.\text{sign } x & \text{if } |x| > c \end{cases} \tag{2.4}$$

with given $c > 0$. Various alternative choices of ψ are described, e.g., in Andrews et al.(1972). If we wish to get a better performance of M_n at very long-tailed distributions, we should select a function satisfying

$$\psi(x) = 0 \qquad \text{if } |x| > c \tag{2.5}$$

for some $c > 0$. The pertaining M-estimators, called redescending, are studied by Collins (1977), Portnoy (1977), Collins and Portnoy (1981); see also Huber (1981).

2.1.1. Finite-sample properties of M-estimators

Assume that $X_1, X_2, \dots, X_n$ are i.i.d. random variables distributed according to d.f. $F(x-\theta)$ such that $F(x)+F(-x) = 1$, $x \in R^1$. Let M_n be an M-estimator defined in (2.2) and (2.3) with a nondecreasing nonconstant function ψ such that $\psi(-x) = -\psi(x)$, $x \in R^1$. Then

(i) $M_n(x_1+c, \dots, x_n+c) = M_n(x_1, \dots, x_n) + c$ for $\underset{\sim}{x} \in R^n$, $c \in R^1$

(M_n is translation-equivariant);

(ii) $P_\theta\left\{\sum_{i=1}^n \psi(X_i - t) < 0\right\} \leq P_\theta(M_n \leq t) = P_\theta\left\{\sum_{i=1}^n \psi(X_i - t) \leq 0\right\}$

for $t, \theta \in R^1$;

(iii) $\frac{1}{2} - \frac{\varepsilon}{2} \leq P_\theta(M_n < \theta) \leq P_\theta(M_n \leq \theta) \leq \frac{1}{2} + \frac{\varepsilon}{2}$

for $\theta \in R^1$, with $\varepsilon = P_0(\sum_{i=1}^{n} \psi(X_i) = 0)$.

By (iii), M_n is median unbiased provided $P_0(\sum_{i=1}^{n} \psi(X_i) = 0) = 0$. By (i), M_n is translation-equivariant; however, M_n is generally not scale-equivariant, i.e., it generally does not satisfy $M_n(cx_1,\ldots,cx_n) = c\, M_n(x_1,\ldots,x_n)$, $c > 0$. In practice it means that M-estimators of location should be supplemented by an estimator of scale.

The following theorem, due to Huber (1968), shows that M_n generated by ψ of (2.4) has an interesting minimax property over the Kolmogorov neighborhood of the normal distribution.

THEOREM 2.1. Let $X_1,\ldots,X_n$ be i.i.d. random variables distributed according to the d.f. $F(x-\theta)$ such that $F(x)+F(-x)=1$, $x \in R^1$ and F is an unknown element of the family

$$\mathcal{F} = \left\{ F : \sup_{x \in R^1} |F(x) - \Phi(x)| \leq \varepsilon \right\} \qquad (2.6)$$

with Φ being the d.f. of the standard normal distribution, $\varepsilon > 0$. Let M_n be defined as

$$P(M_n = M_n^+) = P(M_n = M_n^-) = \frac{1}{2} \qquad (2.7)$$

where M_n^-, M_n^+ are defined in (2.3) and ψ is given in (2.4) with $c > 0$ satisfying

$$e^{-2ac}\, \Phi(a-c) - \Phi(-a-c) = \varepsilon(1+e^{-2ac}), \quad a > 0. \qquad (2.8)$$

Then M_n minimizes

$$\sup_{F \in \mathcal{F}} \sup_{\theta \in R^1} \max \left[P_\theta(T_n - \theta < -a), P_\theta(T_n - \theta > a) \right] \qquad (2.9)$$

over the set of estimators T_n of θ.

Proof. The theorem is proved in Huber (1968); see also Huber (1969). Further and more general finite-sample minimax results may be found in Huber and Strassen(1973), Rieder (1977,1980); see also Huber (1981).

2.1.2. Asymptotic efficiency of M-estimators

If $T_n = T_n(X_1,\ldots,X_n)$, $n=1,2,\ldots$, is a sequence of estimators which is asymptotically normally distributed as $n\to\infty$, then the efficiency of T_n is usually measured through the variance of its asymptotic distribution. The M-estimators are asymptotically normally distributed under mild conditions on ψ and F; this was first proved in Huber (1964) and (1965).

If ψ is skew-symmetric and has bounded variation in every interval, i.e., it may be written as $\psi = \psi_1 - \psi_2$ where ψ_1 and ψ_2 are nondecreasing, and if F has an absolutely continuous symmetric density $f(x)$ such that

$$I(F) = \int (f'(x)/f(x))^2 \, dF(x) < \infty \tag{2.10}$$

(finite Fisher's information) and

$$\int \psi^2(x) \, dF(x) < \infty \tag{2.11}$$

then the M-estimator M_n is consistent and asymptotically normally distributed in the sense that

$$\sqrt{n}(M_n - \theta) \xrightarrow{\mathcal{D}} N(0, \sigma^2(\psi, F)) \quad \text{as} \quad n\to\infty \tag{2.12}$$

where

$$\sigma^2(\psi, F) = \int \psi^2(x)\,dF(x) \,.\, \left(\int f(x) \, d\psi(x) \right)^{-2} . \tag{2.13}$$

The more we assume about ψ, the less we need to impose on F to achieve the asymptotic normality of M_n ; for instance, if ψ is a step-function, then the derivative of F should exist only in a neighborhood of the jump-points of ψ .

We see that, for ψ bounded, $\sigma^2(\psi, F)$ is finite for a large class of distributions. The characteristic $\sup_{F \in \mathcal{F}} \sigma^2(\psi, F)$ may be considered as a measure of robustness of the M-estimator generated by ψ over the family $\mathcal{F}$. If $\mathcal{F}$ is a neighborhood of a given distribution, for instance of the normal one, there may exist an optimal ψ which minimizes $\sup_{F \in \mathcal{F}} \sigma^2(\psi, F)$. Let us illustrate one of such minimax results (established by Huber (1964)) corresponding to the case that $\mathcal{F}$ forms a special neighborhood of the normal distribution.

THEOREM 2.2. (Huber (1964)). Let $\mathcal{F}_\varepsilon$ be the family of ε-contaminated normal distributions, i.e.,

$$\mathcal{F}_\varepsilon = \left\{ F : F = (1-\varepsilon)\Phi + \varepsilon H,\ H \in \mathfrak{M} \right\} \tag{2.14}$$

where $\mathfrak{M}$ is the set of all symmetric distribution functions, ε is a fixed number, $0 \leq \varepsilon < 1$, and Φ is the standard normal d.f. Denote $\psi_0(x)$ the function defined in (2.4) with c satisfying

$$2\left[(f^*(c)/c) - 1 + \Phi(c) \right] = \varepsilon/(1-\varepsilon) , \quad f^*(x) = \frac{d\Phi}{dx} . \tag{2.15}$$

Then

$$\sup_{F \in \mathcal{F}_\varepsilon} \sigma^2(\psi_0, F) = \inf_\psi \sup_{F \in \mathcal{F}_\varepsilon} \sigma^2(\psi, F) \tag{2.16}$$

and the supremum on the left-hand side of (2.16) is attained for the d.f. F_0 with the density

$$f_0(x) = \begin{cases} (1-\varepsilon)\, f^*(x) & \text{if } |x| \leq c \\ (1-\varepsilon)(2\pi)^{-1/2} \exp\left\{ \frac{c^2}{2} - c|x| \right\} & \text{if } |x| > c. \end{cases} \tag{2.17}$$

Remark 1. The distribution (2.17) is the least informative one in $\mathcal{F}_\varepsilon$, i.e. $I(F_0) = \inf\{I(F) : F \in \mathcal{F}_\varepsilon\}$; the M-estimator generated by ψ_0 is the maximum likelihood estimator for F_0.

Remark 2. The characteristic $\sup_{F \in \mathcal{F}} \sigma^2(\psi, F)$ is also studied in Collins (1977), Portnoy (1977) and in Collins and Portnoy(1981).

2.1.3. Some further developments

Hampel (1964) introduced the influence curve of an estimator T_n as a measure of the local sensitivity of T_n to the infinitesimal deviations from the underlying distribution. It is the measure of the sensitivity of the functional countrepart $T(F)$ of T_n and is defined as

$$IC(x;F,T) = \lim_{h \to 0} \frac{1}{h} \left[T((1-h)F + h\,\delta_x) - T(F) \right], \qquad (2.18)$$

where δ_x is the degenerate d.f. of the constant x, $x \in R^1$. The influence curve of the M-estimator generated by ψ is

$$IC(x;T,F) = \psi(x - T(F)) \left(\int \psi'(x - T(F))\, dF(x) \right)^{-1}. \qquad (2.19)$$

Field and Hampel (1978) (cf. Field (1978)) developed an Edgeworth-type expansion for $(-g_n'/g_n)$ with g_n being the density of the M-estimator M_n. Their method provides very precise approximations even for small samples.

Boos and Serfling (1980) derived the law of iterated logarithm for M_n (cf. Serfling (1980)). Bahadur's type representations of M_n were established by Carroll (1978) and Jurečková (1980). Jurečková and Sen (1982) proved the moment convergence of M_n and derived the asymptotically risk-efficient sequential versions of M_n with respect to the loss $L(a,c) = a(T_n - \theta)^2 + cn$; $a, c > 0$.

2.2. R-estimators (estimators derived from the signed-rank tests)

The signed-rank test of the hypothesis $H : \theta = \theta_o$ is typically based on the statistic

$$S_n(\underset{\sim}{X} - \theta_o) = \sum_{i=1}^{n} \text{sign}\,(X_i - \theta_o)\,\varphi^+\left(\frac{R^+_{ni}(\theta_o)}{n+1}\right) \qquad (2.20)$$

where $R^+_{ni}(\theta_o)$ is the rank of $|X_i - \theta_o|$ among $|X_1-\theta_o|,\ldots,|X_n-\theta_o|$ and $\varphi^+(t) = \varphi((t+1)/2)$, $0 < t < 1$, where $\varphi(t)$ is nondecreasing and square-integrable function, $\varphi(1-t) = -\varphi(t)$, $0 < t < 1$. The statistic $S_n(\underset{\sim}{X}-t)$ is then nonincreasing in t and attains positive as well as negative values with probability 1 and $E_{\theta_o} S_n(\underset{\sim}{X}-\theta_o) = 0$. The R-estimator of θ is then defined as a solution of the equation $S_n(\underset{\sim}{X}-t)=0$; more precisely, it is defined as

$$R_n = \frac{1}{2}\,(R_n^- + R_n^+) \qquad (2.21)$$

where

$$R_n^- = \sup\left\{t : S_n(\underset{\sim}{X}-t) > 0\right\}$$
$$R_n^+ = \inf\left\{t : S_n(\underset{\sim}{X}-t) < 0\right\}. \qquad (2.22)$$

The R-estimators of location, which are the inversions of the signed-rank tests, were suggested by Hodges and Lehmann (1963). Only some single R-estimators could be given a simple explicit form : besides the sample median (which is the inversion of the sign test), the most well-known is the R-estimator corresponding to the Wilcoxon one-sample test (usually called Hodges-Lehmann's estimator); it can be written as

$$R_n = \text{med}\left\{\frac{X_i + X_j}{2} : 1 \leq i, j \leq n\right\}. \qquad (2.23)$$

The trimmed version of the Hodges-Lehmann's estimator , namely,

$$R_n = \text{med}\left\{\frac{X_i + X_j}{2} : [n\alpha]+1 \leq i, j = n-[n\alpha]\right\}, \qquad (2.24)$$

$0 < \kappa < 1/2$, appears in some contexts (cf. Miura (1981)); this estimator corresponds to the trimmed Wilcoxon test.

2.2.1. Finite-sample properties of R-estimators

Let $X_1,\dots,X_n$ be i.i.d. random variables distributed according to an absolutely continuous d.f. $F(x-\theta)$ such that $F(x)+F(-x)=1$, $x \in R^1$. Let R_n be an R-estimator defined in (2.21) and (2.22), generated by a nondecreasing score function φ such that $\varphi(1-t) = -\varphi(t)$, $0<t<1$. Then

(i) $R_n(x_1+c,\dots,x_n+c) = R_n(x_1,\dots,x_n)+c$ for $\underset{\sim}{x} \in R^n$, $c \in R^1$

(translation-equivariance)

(ii) $R_n(cx_1,\dots,cx_n) = c\, R_n(x_1,\dots,x_n)$, $\underset{\sim}{x} \in R^n$, $c>0$

(scale-equivariance)

(iii) $P_\theta(S_n(\underset{\sim}{X}-t)<0) \le P_\theta(R_n \le t) = P_\theta(S_n(\underset{\sim}{X}-t) \le 0)$, $t,\theta \in R^1$

(iv) $\frac{1}{2} - \frac{\varepsilon}{2} \le P_\theta(R_n < \theta) \le P_\theta(R_n \le \theta) \le \frac{1}{2} + \frac{\varepsilon}{2}$, $\theta \in R^1$,

with $\varepsilon = P_\theta(S_n(\underset{\sim}{X}-\theta)=0) = P_0(S_n(\underset{\sim}{X})=0)$.

Remark. The properties (i), (iii) and (iv) are analogous to these of M-estimators . The value ε in (iv) is independent of F. The property (ii) is the only one which we miss in the case of M-estimators. On the other hand, R-estimators do not have the finite-sample minimax property of Huber's M-estimator (see Theorem 2.1).

2.2.2. Asymptotic efficiency of R-estimators

Hodges and Lehmann (1963) proved that the asymptotic efficiency of the R-estimator coincides with the Pitman efficiency of the corresponding signed-rank test. Thus, to establish the efficiency of an R-estimator, we need to know the asymptotic distribution of the signed-rank statistics under contiguous location alternatives; and this was studied in details in the monograph Hájek-Šidák (1967).

Assume that the score-generating function φ is skew-symmetric, square-integrable and is of bounded variation on every subinterval of (0,1), i.e., $\varphi = \varphi_1 - \varphi_2$ where φ_1 and φ_2 are nondecreasing; then, provided F has an absolutely continuous symmetric density with finite Fisher's information (see (2.10)), then

$$\sqrt{n}\,(R_n - \theta) \xrightarrow{\mathcal{D}} N(0, \sigma^2(\varphi, F)) \tag{2.25}$$

with

$$\sigma^2(\varphi, F) = \int_0^1 \varphi^2(t)\,dt \cdot \left(\int \varphi(F(x))\,f'(x)\,dx\right)^{-2} . \tag{2.26}$$

We see that $0 < \sigma^2(\varphi, F) < \infty$ under general conditions. If we put

$$\varphi(t) = \varphi_f(t) = -f'(F^{-1}(t))/f(F^{-1}(t)) , \quad 0 < t < 1 \tag{2.27}$$

then $\sigma^2(\varphi, F) = 1/I(F)$; it means that the class of R-estimators also contains an asymptotically efficient element. Similarly as in the case of M-estimators, we are interested in the behavior of $\sup_{F \in \mathcal{F}} \sigma^2(\varphi, F)$ over some family $\mathcal{F}$ of distributions, e.g. over the family $\mathcal{F}_\varepsilon$ of contaminated normal distributions (2.14). Then (cf. Jaeckel (1971)), if we put $\varphi_0(t) = \varphi_{f_0}(t)$, $0 < t < 1$, with f_0 being the least informative distribution (2.17), i.e.,

$$\varphi_0(t) = \begin{cases} -c & \text{if} \quad t < \alpha \\ \Phi^{-1}((t-\frac{\varepsilon}{2})/(1-\varepsilon)) & \text{if} \quad \alpha \le t \le 1-\alpha \\ c & \text{if} \quad t > 1-\alpha \end{cases} \tag{2.28}$$

where $\alpha = (\varepsilon/2)+(1-\varepsilon)\Phi(-c)$, we get an R-estimator satisfying

$$\sup_{F\in\mathcal{F}_\varepsilon} \sigma^2(\varphi_0, F) = \inf_{\varphi} \sup_{F\in\mathcal{F}_\varepsilon} \sigma^2(\varphi, F).$$

It could be shown similarly that the trimmed Hodges-Lehmann's estimator (2.24) provides the saddle-point for the family of contaminated logistic distributions (cf. Miura (1981)).

2.2.3. Some further developments

Antille (1974) established the Bahadur's type representation of Hodges-Lehmann's estimator and Hušková and Jurečková (1981) for a more general R-estimator. van Eeden (1970) and Beran(1974) developed asymptotically uniformly efficient (adaptive) R-estimators of location. Sen (1980) proved the moment convergence and developed the asymptotically risk-efficient sequential versions of R-estimators.

2.3. L-estimators (Linear combinations of order statistics)

Let $X_1, X_2, \ldots$ be a sequence of independent random variables, identically distributed according to d.f. $F(x-\theta)$, $F(x)+F(-x) = 1$, $x \in R^1$; let $X_{n:1} \le \ldots \le X_{n:n}$ be the order statistics corresponding to $X_1,\ldots,X_n$. The L-estimator of θ is defined as

$$L_n = \sum_{i=1}^{n} c_{ni} X_{n:i} \tag{2.29}$$

where the coefficients $c_{n1},\ldots,c_{nn}$ satisfy

$$c_{ni} = c_{n,n-i+1} \ge 0, \; i=1,\ldots,n \; ; \quad \sum_{i=1}^{n} c_{ni} = 1 \; . \tag{2.30}$$

This class of estimators covers the sample mean as well as the sample median. The L-estimators are computionally more appealing than M- and R-estimators. If we wish to get a robust L-estimator, insensitive to the extreme observations, we must put $c_{ni}=0$ for $i \le k_n$ and $i \ge n-k_n+1$ with a proper k_n. Typical examples of such estimators are the α-trimmed mean,

$$L_n = (1/(n-2[n\alpha])) \sum_{i=[n\alpha]+1}^{n-[n\alpha]} X_{n:i} \qquad (2.31)$$

and the α-Winsorized mean,

$$L_n = \frac{1}{n}\left\{[n\alpha]\ X_{n:[n\alpha]} + \sum_{i=[n\alpha]+1}^{n-[n\alpha]} X_{n:i} + [n\alpha]\ X_{n:n-[n\alpha]+1}\right\}; \qquad (2.32)$$

$0 < \alpha < 1/2$ and $[x]$ is the largest integer k satisfying $k \le x$.

Most of the L-estimators may be expressed in the following way

$$L_n = \frac{1}{n} \sum_{i=1}^{n} J\left(\frac{i}{n+1}\right) X_{n:i} + \sum_{j=1}^{k} a_j X_{n:[np_j]} \qquad (2.33)$$

where $J(u)$, $0 < u < 1$, is a proper weight function satisfying $J(u) = J(1-u) \ge 0$, $0 < u < 1$ and $p_1,\dots,p_k$; $a_1,\dots,a_k$ are given constants satisfying $0 < p_1 < \dots < p_k < 1$, $p_j=1-p_{k-j+1}$, $a_j=a_{k-j+1}$ $=0$, $j=1,\dots,k$. L_n is then of the form (2.29) with c_{ni} equal to $n^{-1}J(\frac{i}{n+1})$ plus an additional contribution a_j if $i=[np_j]$ for some j $(1 \le j \le k)$. It is usually assumed that J is a fairly smooth function; many L-estimators reduce to just one term of (2.33).

2.3.1. Finite-sample properties of L-estimators

Let L_n be the L-estimator defined in (2.29) and satisfying (2.30); then, provided $X_1, X_2, \dots$ are distributed according to the d.f. $F(x-\theta)$, $F(x)+F(-x)=1$, $x \in R^1$, it holds

(i) $L_n(x_1+c,\ldots,x_n+c) = L_n(x_1,\ldots,x_n) + c$; $\underset{\sim}{x} \in R^n$, $c \in R^1$.

(ii) $L_n(cx_1,\ldots,cx_n) = c\, L_n(x_1,\ldots,x_n)$; $\underset{\sim}{x} \in R^n$, $c > 0$.

(iii) If F is absolutely continuous, then

$$P_\theta(L_n < \theta) = P_\theta(L_n \leq \theta) = 1/2, \quad \theta \in R^1 .$$

2.3.2. Asymptotic efficiency of L-estimators

The asymptotic normality of L-estimators was studied by many authors under various conditions on F and on c_{ni}'s. We may mention Bickel (1965 , 1967), Boos (1979), Boos and Serfling (1980), Chernoff, Gastwirth and Johns (1967), Huber (1969), Shorack (1969,1972), Stigler (1969,1974), among others. A good review of the asymptotic results on L-estimators may be found in Serfling (1980).

Let us first consider the L-estimators of the form (2.33) with vanishing second component. Then, again, the more we assume on J, the less we must assume on F. For robust L-estimators, it is more convenient to put more restrictions on J rather than on F. From the various theorems on asymptotic normality of L_n, let us describe one proved in Stigler (1974; see also Stigler (1979)).

THEOREM 2.3. Let $X_1, X_2, \ldots$ be the sequence of independent observations from the d.f. $F(x-\theta)$ such that $F(x)+F(-x)=1$, $x \in R^1$. Let $J(u)$ be a function such that $J(u)=J(1-u)$, $0<u<1$ and $\int_0^1 J(u)du=1$. Then, under the assumptions

(i) $J(u)=0$ for $0<u<\alpha$ and $1-\alpha<u<1$, is bounded and satisfies a Lipschitz condition of order $>1/2$ (except possibly a finite number of points of F^{-1} measure 0);

(ii) $\int |F(x)(1-F(x)|^{1/2}\, dx < \infty$ and

$$\sigma^2(J,F) = \iint J(F(x))J(F(y))\,|F(\min(x,y))-F(x)F(y)|\,dx\,dy \tag{2.34}$$

is positive.

Then the estimator

$$L_n = \frac{1}{n} \sum_{i=1}^{n} J\left(\frac{i}{n+1}\right) X_{n:i} \tag{2.35}$$

satisfies

$$\sqrt{n}\,(L_n-\theta) \xrightarrow{\mathcal{D}} N(0, \sigma^2(J,F)), \text{ as } n\to\infty . \tag{2.36}$$

If L_n is of the form (2.33) and the second component does not vanish, then, under the assumptions of Theorem 2.3, $\sqrt{n}\,(L_n-\theta)$ is asymptotically normally distributed $N(0, \sigma^2(F))$ with

$$\sigma^2(F) = \operatorname{var}\Big\{-\int (I[X_1 \leq y] - F(y))J(F(y))dy + \sum_{j=1}^{k}[a_j/f(F^{-1}(p_j))](p_j-I[X_1=F^{-1}(p_j)])\Big\}, \tag{2.37}$$

provided $F^{-1}(p)$ has positive derivative at p_j, $j=1,\dots,k$ (cf. Boos (1979)).

Under additional assumptions on F, the asymptotic normality with the variance $\sigma^2(J,F)$ of (2.34) may be established even for J which puts positive weights on the extremes (Stigler (1974), Shorack (1972)).

If F has an absolutely continuous density f and finite Fisher's information $I(F)$, then the L-estimator (2.35) with

$$J(t) = J_F(t) = \psi_F'(F^{-1}(t))/I(F), \quad 0 < t < 1 \tag{2.38}$$

where $\psi_F(x) = -f'(x)/f(x)$, $x \in R^1$, satisfies $\sigma^2(J_F,F) = 1/I(F)$.

It means that the class of L-estimators also contains an asymptotically efficient element.

If we put $J_0(t) = J_{F_0}(t)$, $0 < t < 1$, with F_0 being the d.f. of the least informative distribution (2.17), i.e.,

$$J_0(t) = \begin{cases} 1/(1-2\alpha) & \text{if} \quad \alpha \le t \le 1-\alpha \\ 0 & \text{otherwise} \end{cases} \qquad (2.39)$$

with $\alpha = F_0(-c) = \frac{\varepsilon}{2} + (1-\varepsilon)\Phi(-c)$, we get an L-estimator satisfying

$$\sup_{F\in\mathcal{Y}_\varepsilon} \sigma^2(J_0,F) = \inf_J \sup_{F\in\mathcal{Y}_\varepsilon} \sigma^2(J,F) \qquad (2.40)$$

where $\mathcal{Y}_\varepsilon$ is the family of ε-contaminated normal distributions (2.14). The L-stimator generated by $J_0(t)$ is the α-trimmed mean. We may conclude that the α-trimmed mean is the most recommendable estimator of the center of symmetry of the contaminated normal distribution. It is computionally simple and it is not only translation- but also scale-equivariant. Bickel and Lehmann (1975) proved another attractive property of the α-trimmed mean : its asymptotic efficiency relatively to the sample mean $\bar{X}_n$ cannot get below $(1-2\alpha)^2$ not only for symmetric F but also for every strictly increasing and continuous F.

2.3.3. Some further developments

Berry-Esseen bounds for L-estimators were studied, among others, by Bickel (1967), Bjerve (1977), Boos and Serfling (1979), Helmers (1977,1980,1981); the law of iterated logarithm and almost sure asymptotic results were established by Wellner (1977a, b) and van Zwet (1980). Invariance principles for Lestimators were proved by Sen (1977,1978; see also Sen (1981)). Moment convergence of L-estimators and their asymptotically risk-efficient versions were studied by Jurečková and Sen (1982). The tail behavior of L-estimators in the finite sample case was studied by Jurečková (1979,1981).

3. Estimation of regression

Let $X_n = (X_{n1},\dots,X_{nn})'$ be the vector of independent observations satisfying

$$\underset{\sim}{X}_n = \underset{\sim}{C}_n \underset{\sim}{\theta} + \underset{\sim}{E} \tag{3.1}$$

where $\underset{\sim}{\theta} = (\theta_1,\dots,\theta_p)'$ is the vector of unknown regression parameters, $\underset{\sim}{E} = (E_1,\dots,E_n)'$ is the vector of errors and $\underset{\sim}{C}_n = [c_{ij}^{(n)}]_{i=1,\dots,n}^{j=1,\dots,p}$ is a known design matrix of the rank p. The problem is that of estimating $\underset{\sim}{\theta}$. We shall assume troughout that E_i, $i=1,\dots,n$, are independent and identically distributed according to a common d.f. F which is an unknown member of a family $\mathcal{F}$ of d.f.'s. The coordinates of $\underset{\sim}{X}_n$ and of $\underset{\sim}{C}_n$ depend on n ; we shall not indicate explicitly this dependence unless it could cause a confusion. Let us denote

$$\delta_i(t) = X_i - \sum_{j=1}^{p} c_{ij} t_j, \quad i=1,\dots,n \tag{3.2}$$

the residuals corresponding to the vector $\underset{\sim}{t} = (t_1,\dots,t_p)'$.

What was said about the sensitivity of the sample mean to the outlying observations and to the long-tailed distributions, holds also for the least-squares estimator (l.s.e.); and the outliers are more difficult to track in the linear model. The M-,R- and L-estimators extend, in a more or less straightforward way, to the linear model.

3.1. M-estimators

The M-estimator $\underset{\sim}{M}_n$ of $\underset{\sim}{\theta}$ is defined as a solution of the system of equations

$$\sum_{i=1}^{n} c_{ij}\psi(X_i - \sum_{k=1}^{p} c_{ik}t_k) = 0, \quad j=1,\dots,p \tag{3.3}$$

with respect to $t_1,\dots,t_p$. If there are more solutions of (3.3), then M_n may be defined as that the nearest to some proper preliminary consistent estimator of $\underset{\sim}{\theta}$. If F has an absolutely continuous density f and we put $\psi(x) = -f'(x)/f(x)$, $x \in R^1$, we get the m.l.e. of $\underset{\sim}{\theta}$; $\underset{\sim}{M}_n$ coincides with the l.s.e. if $\psi(x) = x$, $x \in R^1$.

Similarly as in the location case, $\underset{\sim}{M}_n$ is translation-equivariant but generally not scale-equivariant, so that, unless the scale of F is supposed to be known, $\underset{\sim}{M}_n$ should be supplemented by an appropriate estimator of scale.

The asymptotic behavior of $\underset{\sim}{M}_n$ as $n \to \infty$ was studied by Relles (1968), Huber (1972,1973), Yohai and Maronna(1979), among others. Under the assumptions on ψ and on F analogous to these in the location case (besides the assumption of symmetry of F), it was shown that, as $n \to \infty$, $\underset{\sim}{M}_n$ is asymtotically p-dimensionally normally distributed with mean $\underset{\sim}{\theta}$ and with the covariance matrix $\sigma^2(\psi,F)\underset{\sim}{\Sigma}_n^{-1}$ with $\sigma^2(\psi,F)$ given in (2.13) and $\underset{\sim}{\Sigma}_n = \underset{\sim}{C}_n' \underset{\sim}{C}_n$; the matrix $\underset{\sim}{\Sigma}_n$ is assumed to be positive and of the rank p for $n \geq n_0$. We see that, the sequence $\{\underset{\sim}{C}_n\}$ being fixed, the efficiency properties of $\underset{\sim}{M}_n$ depend only on the constant $\sigma^2(\psi,F)$ and are analogous to these in the location case. This further implies that the asymptotic minimax property of M-estimators over the family $\mathcal{F}_\varepsilon$ of ε-contaminated normal distributions (see Section 2.1.2), extends to the linear model (3.1).

Huber (1973) considered the asymptotic behavior of M_n in the case that $p \to \infty$ simultaneously with n. An extension of M-estimators to the multivariate linear model and its asymptotic behavior was studied by Maronna (1976) and Carroll (1978). M-estimators of regression parameters with random design matrix were studied by Maronna, Bustos and Yohai (1979). Bahadur type representation of M-estimators in the linear model was considered by Jurečková and Sen (1981a,b).

3.2. R-estimators

R-estimators of regression parameters are inversions of linear rank tests of regression. The general rank test of the hypothesis $H : \underset{\sim}{\theta} = \underset{\sim}{\theta}_o$ in the model (3.1) is based on the vector of statistics

$$S_{nj}(\underset{\sim}{\theta}_o) = \sum_{i=1}^{n}(c_{ij}-\bar{c}_j)\,\varphi\left(\frac{R_{ni}(\underset{\sim}{\theta}_o)}{n+1}\right), \quad j=1,\dots,p \qquad (3.4)$$

where $R_{ni}(\underset{\sim}{\theta}_o)$ is the rank of the residual $\delta_i(\underset{\sim}{\theta}_o)$ among $\delta_1(\underset{\sim}{\theta}_o), \dots, \delta_n(\underset{\sim}{\theta}_o)$ and $\varphi(t)$ is a nondecreasing square-integrable score function, $0 < t < 1$. Denote $\underset{\sim}{S}_n(\underset{\sim}{t}) = (S_{n1}(\underset{\sim}{t}),\dots,S_{nn}(\underset{\sim}{t}))'$; then, under $\underset{\sim}{\theta}=\underset{\sim}{\theta}_o$, it holds $E_{\underset{\sim}{\theta}_o} S_n(\underset{\sim}{\theta}_o) = \underset{\sim}{0}$ and analogously as in the location case, we may define the R-estimator of $\underset{\sim}{\theta}_o$ as any solution of the system of "equations"

$$S_{nj}(\underset{\sim}{t}) \mathrel{\overset{\sim}{=}} 0, \quad j=1,\dots,p \qquad (3.5)$$

with respect to $\underset{\sim}{t}$.

The statistics (3.4) are invariant to the translation, so that they are not able to estimate the main additive effect (i.e., the component θ_j for which $c_{ij}=1$, $i=1,\dots,n$). The main additive effect should be estimated with the aid of the signed rank statistics on the same line as the location parameter (cf. Jurečková (1971b).

Adichie (1967), following the ideas of Hodges and Lehmann, suggested an estimator of (θ_1,θ_2) in the regression model $X_i = \theta_1+\theta_2 c_i+E_i$, $i=1,\dots,n$, based on the Wilcoxon tests and derived its asymptotic distribution. Jurečková (1971a), Koul(1972) and Jaeckel (1972) then extended the procedure to the p-parameter regression and to the general linear rank tests. The three respective R-estimators are asymptotically equivalent and thus they have the same asymptotic distributions and efficiencies. The estimators differ in the way how they describe the solution of (3.5). Jurečková (1971a) suggested the estimator $\underset{\sim}{R}_n$ as any solution of the minimization problem

$$\sum_{j=1}^{p} \sum_{i=1}^{n} |S_{nj}(\underset{\sim}{t})| \;:=\; \min \tag{3.6}$$

and proved that $\underset{\sim}{R}_n$ is asymptotically normally distributed with the expectation $\underset{\sim}{\theta}$ and with covariance matrix $\sigma^2(\varphi,F)\cdot \underset{\sim}{\Sigma}_n^{-1}$ with $\sigma^2(\varphi,F)$ given in (2.25) and $\underset{\sim}{\Sigma}_n = \underset{\sim}{C}_n'\underset{\sim}{C}_n$. The assumptions on φ and on F are similar to these in the location case while the assumptions on $\underset{\sim}{C}_n$ were rather restrictive in Jurečková(1971a) and some related papers (concordance-discordance condition on the columns of $\underset{\sim}{C}_n$ for $n \geq n_0$). Later on Heiler and Willers (1979) proved that the asymptotic normality of $\underset{\sim}{R}_n$ holds also without concordance-discordance condition.

Jaeckel (1972) suggested R-estimator of $\underset{\sim}{\theta}$ as a solution of the minimization problem

$$\sum_{i=1}^{n} \left[\varphi\left(\frac{R_{ni}(\underset{\sim}{t})}{n+1}\right) - \bar{\varphi}_n\right]\delta_i(\underset{\sim}{t}) \;:=\; \min \tag{3.7}$$

with respect to $\underset{\sim}{t}$; $\overline{\varphi}_n = \frac{1}{n} \sum_{i=1}^{n} \varphi(\frac{i}{n+1})$. The idea is that (3.7) could be considered as a measure of the dispersion of the residuals $\delta_i(\underset{\sim}{t})$, $i=1,\ldots,n$, instead of the proper variance of the residuals which is used in the method of least squares. Jaeckel proved the asymptotic euqivalence of the solution of (3.7) and of (3.6), respectively, as $n \longrightarrow \infty$.

Koul (1971) suggested the R-estimator as a solution, instead of (3.6) and (3.7), of an appropriate quadratic form in the statistics $S_{nj}(\underset{\sim}{t})$, $j=1,\ldots,p$, with respect to $\underset{\sim}{t}$. All three estimators are asymptotically equivalent, as $n \to \infty$.

3.3. L - estimators

While being computionally very appealing in the location case, the L-estimators do not have any straight forward extension to the linear model. Let us mention some of the attempts which appeared in the literature.

Koenker and Bassett (1978) extended the concept of quantiles to the linear model. For a fixed α , $0 < \alpha < 1$, put

$$\psi_\alpha(x) = \alpha - I\left[x < 0\right] \tag{3.8}$$

and

$$\rho_\alpha(x) = x . \psi_\alpha(x), \quad x \in R^1. \tag{3.9}$$

Koenker and Bassett defined the α-th regression quantile as the solution $\underset{\sim}{T}_n(\alpha) = (T_{n1}(\alpha),\ldots,T_{np}(\alpha))'$ of the minimization

$$\sum_{i=1}^{n} \rho_\alpha(X_i - \sum_{j=1}^{p} c_{ij} t_j) := \min \tag{3.10}$$

with respect to $\underset{\sim}{t} = (t_1, \ldots, t_p)'$. They proved that the asymptotic behavior of the regression quantiles is similar to that of the standard sample quantiles and suggested the following α-trimmed least squares estimator :

Remove X_i from the sample if $\delta_i(\underset{\sim}{T}_n(\alpha)) < 0$ (the i-th residual from $\underset{\sim}{T}_n(\alpha)$ is negative) or if $\delta_i(\underset{\sim}{T}_n(1-\alpha)) > 0$, $i=1,\ldots,n$; $0 < \alpha < \frac{1}{2}$; and calculate the least-squares estimator using the remaining observations. The resulting estimator $\underset{\sim}{L}_n^*$ was later studied by Ruppert and Carroll (1980) who proved that it is asymptotically normally distributed with mean $\underset{\sim}{\theta}$ and with the covariance matrix $\sigma^2(\alpha, F) \cdot \underset{\sim}{\Sigma}_n^{-1}$ where $\underset{\sim}{\Sigma}_n = \underset{\sim}{C}_n' \underset{\sim}{C}_n$ and $\sigma^2(\alpha, F)$ is the asymptotic variance of the α-trimmed mean in the location case. The concept of regression quantile seems to provide a basis for an extension of various other L-estimators from the location to the regression model.

Ruppert and Carroll (1980) also suggested another extension of the α-trimmed mean to the linear model. Starting with some reasonable preliminary estimator $\underset{\sim}{L}_o$, one calculates the residuals $\delta_i(\underset{\sim}{L}_o)$ from $\underset{\sim}{L}_o$, $i=1,\ldots,n$, and removes the observations corresponding to $[n\alpha]$ smallest and $[n\alpha]$ largest residuals. The estimator $\underset{\sim}{L}_n^{**}$ is then defined as the least-squares estimator calculated from the remaining observations. The asymptotic behavior of $\underset{\sim}{L}_n^{**}$ depends on $\underset{\sim}{L}_o$ and generally is not similar to that of the trimmed mean; $\underset{\sim}{L}_n^{**}$ is asymptotically equivalent to $\underset{\sim}{L}_n^*$ provided $\underset{\sim}{L}_o = \frac{1}{2}(\underset{\sim}{T}_n(\alpha) + \underset{\sim}{T}_n(1-\alpha))$.

Bickel (1973) proposed a general class of one-step L-estimators of $\underset{\sim}{\theta}$ depending on a preliminary estimate of $\underset{\sim}{\theta}$. The estimators have the best possible efficiency properties, i.e. analogous to those of the corresponding location L-estimators but they are computionally complex and are not invariant under a reparametrization of the vector space spanned by the columns of $\underset{\sim}{C}_n$.

4. Computional aspects-one step versions of the estimators

Besides the L-estimators of location and Hodges-Lehmann estimator, the estimators considered so far are not very computionally appealing. They are generally defined in the implicit form or as a solution of a complex minimization problem.

Thus, it is often convenient to use the one-step versions of the estimators, which are characterized as follows: we start with some reasonably good consistent preliminary estimator $\hat{\underset{\sim}{\theta}}$ and them apply one step of Gauss-Newton method to the equation defining the estimator. Under mild conditions, it could be shown that the result of one-step Gauss-Newton approximation behaves asymptotically as the root of the equation. This idea was applied by Kraft and van Eeden (1972a,b) to the R-estimators of location and regression, respectively. Bickel (1975) studied the one-step versions of the M-estimators in the linear model.

Let us first describe the one-step version of the M-estimator. Let M_n be the M-estimator of θ in the linear model (3.1), defined as the solution of the system of equations (3.3).

Assume that the design matrix $\underset{\sim}{C}_n$ satisfies the condition $n^{-1}\underset{\sim}{C}_n'\underset{\sim}{C}_n \longrightarrow \underset{\sim}{\Sigma}$ as $n\longrightarrow\infty$ where $\underset{\sim}{\Sigma}$ is a positive pxp matrix. Then, provided F has an absolutely continuous density f, $I(F)<\infty$ and ψ has bounded variation on any compact interval,

$$\sqrt{n}(\underset{\sim}{M}_n-\underset{\sim}{\theta}) = (1/\gamma\sqrt{n})\ \underset{\sim}{\Sigma}^{-1}\underset{\sim}{C}_n'\ \underset{\sim}{v}_n(\underset{\sim}{\theta}) + o_p(1) \tag{4.1}$$

with $\underset{\sim}{v}_n(\underset{\sim}{\theta}) = (\psi(\delta_1(\underset{\sim}{\theta})),\ldots,\psi(\delta_n(\underset{\sim}{\theta})))'$ (4.2)

and $\gamma = -\int\psi(x)f(x)dx$ (4.3)

(cf. Bickel (1975), Jurečková (1977). Let $\hat{\underset{\sim}{\theta}}_n$ be a consistent preliminary estimator of $\underset{\sim}{\theta}$ which is shift-equivariant, i.e. $\hat{\underset{\sim}{\theta}}_n(\underset{\sim}{x}+\underset{\sim}{C}_n\underset{\sim}{t}) = \hat{\underset{\sim}{\theta}}_n(\underset{\sim}{x})+\underset{\sim}{t}$, $\underset{\sim}{x}\in R^n$, $\underset{\sim}{t}\in R^p$, and which satisfies

$$\|\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}\| = O_p(1), \quad \text{as} \quad n\longrightarrow\infty \quad . \tag{4.4}$$

The one-step version of $\underset{\sim}{M}_n$ is defined as

$$\underset{\sim}{M}_n^{*} = \hat{\underset{\sim}{\theta}}_n + (1/n\ \hat{\gamma}_n)\ \underset{\sim}{\Sigma}^{-1}\ \underset{\sim}{C}_n'\ \underset{\sim}{v}_n(\hat{\underset{\sim}{\theta}}_n) \tag{4.5}$$

where $\hat{\gamma}_n$ is an appropriate consistent estimator of γ; one of the possible estimators of γ is

$$\hat{\gamma}_n = n^{-1/2}\ \|\underset{\sim}{t}_2-\underset{\sim}{t}_1\|^{-1}\ \|\underset{\sim}{\Sigma}^{-1}\cdot\underset{\sim}{C}_n'(\underset{\sim}{v}_n(\underset{\sim}{t}_2)-\underset{\sim}{v}_n(\underset{\sim}{t}_1))\| \tag{4.6}$$

where $\underset{\sim}{t}_1,\underset{\sim}{t}_2$ is a fixed pair of px1 vectors, $\underset{\sim}{t}_1\neq\underset{\sim}{t}_2$. Then it could be proved that

$$\sqrt{n}\ \|\underset{\sim}{M}_n-\underset{\sim}{M}_n^{*}\| \xrightarrow{P} 0, \quad \text{as} \quad n\longrightarrow\infty \quad . \tag{4.7}$$

Let us briefly describe possible one-step versions of R-estimator $\underset{\sim}{R}_n$ defined in (3.6) or (3.7). Assume that

$$n^{-1}\sum_{i=1}^{n}(c_{ij}-\bar{c}_j)(c_{ik}-\bar{c}_k) \longrightarrow \sigma^*_{jk}, \quad \text{as} \quad n\to\infty \qquad (4.8)$$

for $1\leq j,k\leq p$, where $\underset{\sim}{\Sigma}^* = \left[\sigma^*_{jk}\right]_{j,k=1,\dots,p}$ is a positive matrix. Then, provided φ is of bounded variation on any compact subinterval of $(0,1)$ and square integrable, F has an absolutely continuous density and $I(F)<\infty$; it could be proved (Jurečková(1971a))

$$\underset{\sim}{R}_n = \underset{\sim}{\theta} + (1/n\gamma)\ \underset{\sim}{\Sigma}^{*\,-1}\ \underset{\sim}{S}_n(\underset{\sim}{\theta}) + o_p(n^{-1/2}) \qquad (4.9)$$

with $\underset{\sim}{S}_n(\underset{\sim}{\theta}) = (S_{n1}(\underset{\sim}{\theta}),\dots,S_{np}(\underset{\sim}{\theta}))'$ given in (3.4) and

$$\gamma = -\int\varphi(F(x))f'(x)dx \ . \qquad (4.10)$$

Let $\hat{\underset{\sim}{\theta}}_n$ be the preliminary shift-equivariant estimator satisfying (4.4). Then, if we knew γ, we could consider the one-step version of $\underset{\sim}{R}_n$ in the form

$$\underset{\sim}{R}'_n = \hat{\underset{\sim}{\theta}}_n + (1/n\gamma)\ \underset{\sim}{\Sigma}^{*\,-1}\ \underset{\sim}{S}_n(\hat{\underset{\sim}{\theta}}_n), \qquad (4.11)$$

and $\sqrt{n}\,\|\underset{\sim}{R}_n-\underset{\sim}{R}'_n\| \xrightarrow{P} 0$ as $n\to\infty$. However, γ is generally unknown; Kraft and van Eeden (1972b) suggested to replace γ in (4.11) by $\int_0^1(\varphi(t)-\bar{\varphi})^2dt$, $\bar{\varphi} = \int_0^1\varphi(t)dt$. The resulting estimator (say , $\underset{\sim}{R}''_n$) is generally not asymptotically equivalent to $\underset{\sim}{R}_n$; it could be proved (cf.HUMAK (1982)) that $\sqrt{n}\,\|\underset{\sim}{R}_n-\underset{\sim}{R}''_n\| \xrightarrow{P} 0$ as $n\to\infty$ if and only if either both $\underset{\sim}{R}_n$ and $\underset{\sim}{R}''_n$ are asymptotically efficient (i.e., where $\varphi(t)= -f'(F^{-1}(t))/f(F^{-1}(t)))$ or if $\underset{\sim}{R}_n$ (and thus also $\underset{\sim}{R}''_n$) is asymptotically equivalent to the preliminary estimator $\hat{\underset{\sim}{\theta}}_n$. In order to get an estimator asymptotically equivalent to $\underset{\sim}{R}_n$, we should replace γ in (4.11) by an appropriate estimator $\hat{\gamma}_n$, similarly as in the case of M-estimator. One of such possible estimators is

$$\hat{\gamma}_n = n^{-1/2} \| \underset{\sim}{t}_2 - \underset{\sim}{t}_1 \|^{-1} \cdot \| \underset{\sim}{\Sigma}^{*\,-1} (\underset{\sim}{S}_n(\underset{\sim}{t}_2) - \underset{\sim}{S}_n(\underset{\sim}{t}_1)) \| \qquad (4.12)$$

where $\underset{\sim}{t}_1, \underset{\sim}{t}_2$ are fixed px1 vectors, $\underset{\sim}{t}_1 \neq \underset{\sim}{t}_2$.

5. Asymptotic relations of M-R-L-estimators

We have seen that the three groups of estimators, though being defined in different ways, follow the same idea: to cut-off the influence of outliers, to diminish the sensitivity to the long-tailed distributions. It turns out that these three classes of estimators are even nearer than one would expect; in fact, they become asymptotically equivalent as $n \to \infty$.

The asymptotic relations of M-L-R estimators were studied by Jaeckel (1971), Bickel and Lehmann (1975), Jurečková (1977, 1978, 1981), Hušková and Jurečková (1981), among others. Let us briefly illustrate some of the results on the location submodel.

Let $X_1, X_2, \ldots$ be the sequence of independent observations, identically distributed according to the distribution function $F(x-\theta)$ such that $F(x)+F(-x)=1$, $x \in R^1$. Let M_n be the M-estimator generated by the function $\psi(x)$, $x \in R^1$ and R_n be the R-estimator generated by the function $\varphi(t)$, $0<t<1$. Then, under some regularity conditions, $\sqrt{n}(M_n - R_n) = o_p(1)$ as $n \to \infty$ if and only if

$$\psi(x) = a\,\varphi(F(x)), \qquad a > 0 \qquad (5.1)$$

for almost all $x \in R^1$. The relation (5.1) means that, given the distribution F, there exists an M-estimator to every R-estimator (and vice versa) such that both estimators are asymptotically equivalent. Being dependent on the unknown d.f. F, the relation (5.1) does not enable to determine the value of the M-estimator once we have calculated the value of R-estimator; it rather indicates which type of M-estimators belongs to a given type of R-estimators etc.

Let L_n be the L-estimator (2.34) generated by the function $J(t)$ such that $J(t) = J(1-t) \geq 0$, $0 < t < 1$. Then, under some smoothness conditions on J and F, $\sqrt{n}(L_n - M_n) = o_p(1)$ as $n \to \infty$ for the M-estimator M_n generated by the function

$$\psi(x) = \int_0^1 J(t)(I[F(x) \leq t] - t)\, dF^{-1}(t), \quad x \in R^1. \qquad (5.2)$$

Let L_n be the α-trimmed mean (2.31); then $\sqrt{n}(L_n - M_n) = o_p(1)$ as $n \to \infty$ where M_n is Huber estimator generated by ψ given in (2.4), more precisely,

$$\psi(x) = \begin{cases} F^{-1}(\alpha) & \text{if } x < F^{-1}(\alpha) \\ x & \text{if } F^{-1}(\alpha) \leq x \leq F^{-1}(1-\alpha) \\ F^{-1}(1-\alpha) & \text{if } x > F^{-1}(1-\alpha) . \end{cases} \qquad (5.3)$$

If L_n is a linear combination of single sample quantiles, $L_n = \sum_{j=1}^{k} a_j X_{n:[np_j]}$ (cf.2.33), then $\sqrt{n}(L_n - M_n) = o_p(1)$ where M_n is the M-estimator generated by the function

$$\psi(x) = -\sum_{j=1}^{k} \left[a_j / f(F^{-1}(p_j))\right] (I\left[F(x) \leq p_j\right] - p_j), \qquad (5.4)$$

$x \in R^1$. Especially, the M-estimator countrepart of the α-Winsorized mean is generated by the function

$$\psi(x) = \begin{cases} F^{-1}(\alpha) - \dfrac{}{f(F^{-1}(\alpha))} & \text{for} \quad x < F^{-1}(\alpha) \\ x & \text{for } F^{-1}(\alpha) \leq x \leq F^{-1}(1-\alpha) \\ F^{-1}(1-\alpha) + \dfrac{}{f(F^{-1}(1-\alpha)} & \text{for} \quad x > F^{-1}(1-\alpha). \end{cases} \qquad (5.5)$$

The relations of R-and L-estimators could be derived by combining the relations of M- and R-estimators and these of M- and L-estimators, respectively.

REFERENCES

Adichie,J.N.(1967). Estimate of regression parameters based on rank tests. Ann.Math.Statist.38,894-904.

Andrews,D.F., Bickel,P.J., Hampel,F.R.,Huber,P.J., Rogers,W.H. Tukey,J.W.(1972). Robust Estimates of Location. Survey and Advances. Princeton University Press.

Antille,A.(1974). A linearized version of the Hodges-Lehmann estimator. Ann.Statist.2, 1308-1313.

Azencott,R., Birgé,L., Costa,V., Dacunha-Castelle,D., Deniau,C., Deshayes,J., Huber,C., Jolivaldt,P., Oppenheim,G., Picard,D., Trécourt,P. and Viano,C. (1977). Théorie de la robustesse et estimation d'un parametre. Astérisque 43-44.

Beran,R.(1974). Asymptotically efficient adaptive rank estimates in location models. Ann.Statist.2, 63-74.

Beran,R.(1977a). Robust location estimates. Ann.Statist.5, 431-444.

Beran,R.(1977b). Minimum Hellinger distance estimates for parametric models. Ann.Statist.5, 445-463.

Beran,R.(1978). An efficient and robust adaptive estimator of location. Ann.Statist. 6, 292-313.

Bickel,P.J.(1965). On some robust estimates of lccation. Ann.Math.Statist. 36, 847-858.

Bickel,P.J.(1967). Some contributions to the theory of order statistics.Proc. 5th Berkeley Symp.1, 575-591.

Bickel,P.J.(1975). One-step Huber estimates in the linear model. J.Amer.Statist.Assoc.70, 428-434.

Bickel,P.J.(1976). Another look at robustness: A review of reviews and some new developments. Scand.Journ.of Statistics 3, 145-168.

Bickel,P.J. and Lehmann,E.L.(1975). Descriptive statistics for nonparametric model. I.Introduction, II.Location. Ann.Statist.3, 1038-1044 and 1045-1069.

Birnbaum,A., Laska,E.(1967). Optimal robustness: A general method, with applications to linear estimates of location. J.Amer.Statist.Assoc.62, 1230-1240.

Bjerve,S. (1977). Error bounds for linear combinations of order statistics. Ann.Statist.5, 357-369.

Boos,D.D.(1979). A differential for L-statistics. Ann.Statist.7, 955-959.

Boos,D. and Serfling,R.J.(1979). On Berry-Essen rates for statistical functions, with application to L-estimates. Technical Rep.,Dept.Statist., Florida State Univ.

Boos,D. and Serfling,R.J.(1980). A note on differentials and the CLT and LIL for statistical functions, with application to M-estimates. Ann.Statist.8, 618-624.

Caroll,R.J.(1978). On the asymptotic distribution of multivariate M-estimates. Journ.Multivar.Analysis 8,361-371.

Chernoff,H., Gastwith,J.L. and Johns,M.V.(1967). Asymptotic distribution of linear combinations of order statistics, with application to estimation. Ann.Math.Statist.38,52-72.

Collins,J.R.(1976). Robust estimation of a location parameter in the presence of asymmetry. App.Statist.25, 228-237.

Collins,J.R.(1977). Upper bounds on asymptotic variances of M-estimators of location. *Ann.Statist.*5, 646-657.

Collins,S.L. and Portnoy,J.R.(1981). Maximizing the variance of M-estimators using the generalized method of moment spaces. *Ann.Statist.*9, 567-577.

Dutter,R.(1975a). Robust regression: Different approaches to numerical solution and algorithms. Res.Report No 6/ETH Zürich.

Dutter,R.(1975b). Numerical solution of robust regression problems: Computional aspects, a comparison. Res.Report No 7, ETH Zürich.

Dutter,R.(1977). Numerical solution of robust regression problems: Computional aspects, a comparison. *J.Statist. Comput.Simul.*5, 207-238.

Dutter,R.(1978). Robust regression: LINWDR and NLWDR. COMPSTAT 1978, Proc.,L.C.A. Corsten (ed.),*Physica-Verlag*,*Vienna*.

Field,C.A.(1978). Summary of small size asymptotics for location estimates. Proc.2nd Prague Symp.on Asymptotic Statistics, 173-179, North -Holland.

Field,C.A.and Hampel,F.R.(1978). Small sample asymptotic distributions of M-estimators of location. Res.Rep. 17,Fachgruppe für Statistik, ETH Zürich.

Gastwirth,J. and Rubin,H.(1969). On robust linear estimators. *Ann.Math.Statist.*40, 24-39.

Hájek,J. and Šidák,Z.(1967). Theory of Rank Tests. Academia, Prague.

Hampel,F.R.(1971). A general qualitative definition of robustness. *Ann.Math.Statist.*42, 1887-1896.

Hampel,F.R.(1973). Robust estimation: A condensed partial survey. *Z.Wahrscheinlichkeitstheorie verw.Geb.*27, 87-104.

Hampel,F.R.(1974a). Small sample asymptotics. Proc.Prague Symp. on Asympt.Statistics II, 109-126.

Hampel,F.R.(1974b). The influence curve and its role in robust estimation. J.Amer.Statist.Assoc.62, 1179-1186.

Hampel,F.R., Rousseeuw,P.J. and Ronchetti,E.(1981). The change-of-variance curve and optimal redescending M-estimators. J.Amer.Statist.Assoc.76, 643-648.

Heiler,S. and Willers,R.(1979). Asymptotic normality of R-estimates in the linear model. Forschungsbericht 79/6, Univ.Dortmund.

Helmers,R.(1977). The order of the normal approximation for linear combinations of order statistics with smooth weight functions. Ann.Prob.5, 940-953.

Helmers,R.(1980). Edgeworth expansions for linear combinations of order statistics. Math.Centre Tract 105, Amsterdam.

Helmers,R.(1981). A Berry-Essen theorem for linear combination of order statistics. Ann.Probab.9, 342-347.

Hodges,J.L. and Lehmann,E.L.(1963). Estimates of location based on rank tests. Ann.Math.Statist.34, 598-611.

Hogg,R.V.(1967). Some observations on robust estimation. J.Amer.Statist.Assoc.62, 1179-1186.

Hogg,R.V.(1974). Adaptive robust procedures. A partial review and some suggestions for future. Applications and theory. J.Amer.Statist.Assoc.69, 909-923.

Hogg,R.V.(1979). Statistical Robustness: One View of its use in Aplications today. The American Statistician 33,108-105.

Hollander,M.and Wolfe,D.A.(1973). Nonparametric Statistical Methods. J.Wiley,N.York.

Huber,P.J.(1964). Robust estimation of a location parameter. Ann.Math.Statist.35, 73-101.

Huber,P.J.(1965a). A Robust version of a probability ratio test. Ann.Math.Statist.36, 1753-1758.

Huber,P.J.(1965b). The behaviour of maximum likelihood estimates under nonstandard conditions. Proc.5th Berkeley Symp.1, 221-233.

Huber,P.J.(1968). Robust confidence limits. Z.Wahrscheinlichkeitstheorie verw.Geb.10, 269-278.

Huber,P.J.(1969). Théorie de l'inférence statistique robuste. Presses de l'Université Montréal.

Huber,P.J.(1970). Studentizing robust estimates. In: Nonparametric Techniques in Statistical Inference (M.L.Puri,ed.) Cambridge Univ.Press.

Huber,P.J.(1972). Robust statistics. A review. Ann.Math.Statist. 43, 1041-1067.

Huber,P.J.(1973). Robust regression: Asymptotics, conjectures and Monte Carlo. Ann.Statist.1, 799-821.

Huber,P.J.(1977). Robust Statistical Precedures. Regional Conference Series in Applied Math.No 27, SIAM Philadelphia.

Huber,P.J.(1981). Robust Statistics. J.Wiley, N.York.

Huber,P.J. and Dutter,R.(1977). Numerical solutions of robust regression problems. In: COMPSTAT 1974, Bruckmann (ed.), Physica Verlag, Vienna.

Huber,P.J. and Strassen,V.(1973). Minimax tests and Neyman-Pearson lemma for capacities. Ann.Statist.1, 251-263; 2, 223-224.

K.M.S. HUMAK (1982). Statistische Metoden der Modelbildung. Band II. Akademie-Verlag,Berlin

Hušková,M. and Jurečková,J.(1981). Second order asymptotic relations of M-estimators and L-estimators in two-sample location model. Journ.Statist.Planning Infer.5, 309-328.

Jaeckel,L.A.(1971a). Robust estimates of location: Symmetry and asymmetry contamination. Ann.Math.Statist.42, 1020-1034.

Jaeckel,L.A.(1971b). Some flexible estimates of location. Ann. Math.Statist.42, 1540-1552.

Jaeckel,L.A.(1972). Estimating regression coefficients by minimizing the despersion of the residuals. Ann.Math.Statist.43, 1449-1458.

Jurečková,J.(1969). Asymptotic linearity of a rank statistic in regression parameter. Ann.Math.Statist.40, 1889-1900.

Jurečková,J.(1971a). Nonparametric estimate of regression coefficients. Ann.Math.Statist.42, 1328-1338.

Jurečková,J.(1971b). Asymptotic independence of rank test statistic for testing symmetry on regression. Sankhya A 33,1-18.

Jurečková,J.(1973a). Central limit theorem for Wilcoxon rank statistics process. Ann.Statist.1, 1046-1060,

Jurečková,J.(1973b). Asymptotic behaviour of rank and signed-rank statistics from the point of view of applications. Proc.Prague Symp.on Asymptotic Statistics I, 139-155.

Jurečková,J.(1977). Asymptotic relations of M-estimates and R-estimates in linear regression model. Ann.Statist.5,464-472.

Jurečková,J.(1978). Asymptotic relations of least-squares estimate and of two robust estimates of regression parameter vector. Trans.7th Prague Conf. and European Meeting of Statisticians II, 231-237.

Jurečková,J. (1979). Finite sample comparison of L-estimators of location. Comment.Math.Univ.Carolinae 20, 507-518.

Jurečková,J.(1980). Asymptotic representation of M-estimators of location. Math.Operationsforsch.Statist., Ser.STATISTICS 11, 61-73.

Jurečková, J.(1981). Tail behavior of location estimators. Ann.Statist.9, 578-585.

Jurečková, J.(1981).Robust estimators of location and their second order asymptotic relations. Submitted to ISI Centennial Volume.

Jurečková,J. and Sen.P.K.(1981a). Invariance principles for some stochastic processes related to M-estimators and their role in sequential statistical inference. Sankhya A 43, 191-210.

Jurečková,J. and Sen.P.K.(1981b). Sequential procedures based on M-estimators with discontinuous score functions. Journ. Statist.Planning Infer.5, 253-266.

Jurečková,J. and Sen.P.K.(1982). M-estimators and L-estimators of location: Uniform integrability and asymptotically risk-efficient sequential versions. Commun.Statist.C 1 (to appear).

Kagan,A.M., Linnik,Yu.V. and Rao, C.R.(1965). On a characterization of the normal law based on a property of the sample average. Sankhya A 27, 405-406.

Kagan,A.M., Linnik, Yu.V. and Rao,C.R.(1972). Characteristic Problems of Mathematical Statistics (in Russian). Nauka, Moscow.

Koenker,R.and Bassett,C.(1978). Regression quantiles. Econometrica 40,33-50.

Koul,H.L.(1971). Asymptotic behavior of a class of confidence regions based on ranks in regression. Ann.Math.Statist.42, 466-476.

Koul,H.L.(1977). Behavior of robust estimators in the regression model with dependent errors. Ann.Statist.5, 681-699.

Kraft,C. and van Eeden,C.(1972a). Asymptotic efficiencies of quick methods of computing efficient estimates based on ranks. J.Amer.Statist.Assoc.67, 199-202.

Kraft,C. and van Eeden,C.(1972b). Linearized rank estimates and signed rank estimates for the general linear hypothesis. Ann.Math.Statist.43, 42-57.

Launer,R.L.and Wilkinson,G.N.(eds.)(1979). Robustness in Statistics. Academic Press, New York.

Maronna.R.A.(1976). Robust M-estimates of multivariate location and scatter. Ann.Statist.4, 51-67.

Maronna,R., Bustos,O.and Yohai,V.(1979). Bias- and efficiency robustness of general M-estimators for regression with random carriers. In: Smoothing Techniques for Curve Estimation (T.Gasser and M.Rosenblatt,eds.),91-116. Lecture Notes in Math.757, Springer-Verlag.

Miura,R.(1981). Adaptive confidence intervals for a location parameter. The Keiei Kenkyu XXXI, 197-218.

Nowak,H.and Zentgraf,R.(eds.)(1980). Robuste Verfahren. Medizinische Informatik und Statistik No 20. Springer-Verlag.

Portnoy,S.L.(1977). Robust estimation in dependent situations. Ann.Statist.5, 22-43.

Raoult,J.P.(1980). Statistique non paramétrique asymptotique. Lecture Notes in Math.821, Springer-Verlag.

Relles,D.A.(1968). Robust regression by modified least squares. PhD Thesis, New York.

Relles,D.A. and Rogers,W.H.(1977). Statisticians are fairly robust estimators of location. J-Amer.Statist.Assoc.72,107-111.

Rey,W.J.J.(1978). Robust statistical methods. Lecture Notes in Math.690.Springer-Verlag.

Rieder,H.(1977). Least favorable pairs for special capacities. Ann.Statist.5, 909-921.

Rieder,H.(1980). Estimates derived from robust tests. Ann.Statist.8, 106-115.

Rousseeuw,P.J.(1981). A new infinitesimal approach to robust estimation. Z.Wahrscheinlichkeitstheorie verw.Geb.56,127-132.

Ruppert,D. and Carrol,R.J.(1980). Trimmed least squares estimation in the linear model. J.Amer.Statist.Assoc.75,828-838.

Sen,P.K.(1977). On Wiener process embedding for linear combinations of order statistics. Sankhya A 39, 138-143.

Sen,P.K.(1978). An invariance principle for linear combinations of order statistics. Z.Wahrscheinlichkeitstheorie verw. Geb. 42, 327-40.

Sen,P.K.(1980). On nonparametric sequential point estimation of location based on general rank order statistics. Sankhya A 42, 202-218.

Sen,P.K.(1981). Sequential Nonparametrics. Invariance Principles and Statistical Inference. J.Wiley.

Serfling,R.J.(1980). Approximation Theorems of Mathematical Statistics. J.Wiley.

Shorack,G.R.(1969). Asymptotic normality of linear combinations of functions of order statistics. Ann.Math.Statist.40, 2041-2050.

Shorack,G.R.(1972). Functions of order statistics. Ann.Math. Statist.43, 412-427.

Sievers,G.L.(1978). Estimation of location : A large deviation comparison. Ann.Statist.6, 610-618.

Stigler,S.M.(1969). Linear functions of order statistics. Ann. Math.Statist. 40, 770-788.

Stigler,S.M.(1974). Linear functions of order statistics with smooth weight functions. Ann.Statist.2, 676-693.

Stigler,S.M.(1977). Do robust estimators work with real data? Ann.Statist. 5, 1055-1098.

Torgenson,E.N.(1971). A counterexample on translation invariant estimators. Ann.Math.Statist.42, 1450-1451.

van Eeden, C. (1970). Efficiency-robust estimation of location. *Ann. Math. Statist.* 41, 172-181.

van Zwet, W.R. (1980). A strong law for linear functions of order statistics. *Ann. Probab.* 8, 986-990.

Wegman, E.J. and Carroll, R.J. (1977). A Monte Carlo Study of Robust estimation of location. *Comm. Statist.* A 6, 795-812.

Wellner, J.A. (1977a). A Glivenko-Cantelli-theorem and strong laws of large numbers for functions of order statistics. *Ann. Statist.* 5, 473-480 Correction note Ann. Statist. 6, 1394.

Wellner, J.A. (1977b). A law of iterated logarithm for functions of order statistics. *Ann. Statist.* 5, 481-494.