On a class of estimators of the location parameter based on a weighted sum of the observations

by

Louis-Paul Rivest

On a class of estimators of the location parameter based on a weighted sum of the observations

by

Louis-Paul Rivest

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Department of Mathematics

McGill University

Montreal

July 1978

On a class of estimators of the location parameter based on a weighted sum of the observations

by

Louis-Paul Rivest

ABSTRACT

The asymptotic properties of a new class of estimators of the location parameter are investigated. Each estimator is based on a weighted sum of the observations. An observation is weighted according to its magnitude (as for an M estimator) and according to its rank (as for an L estimator).

These estimators induce a new class of location parameters which is studied as a set of functionals defined from a space of distribution functions into R.

Finally, a new ordering of distribution functions is introduced. Some basic properties are derived using a generalized concept of unimodality.

Department of Mathematics
McGill University
Ph.D. Thesis

On a class of estimators of the location parameter
based on a weighted sum of the observations

by

Louis-Paul Rivest

RESUME

Dans cette thèse on étudie une nouvelle classe d'estimateurs du paramètre de location. Chaque estimateur est basé sur une somme pondérée des observations. La pondération de chaque observation dépend de l'ordre de grandeur de cette observation (comme pour un M estimateur) et de son rang (comme pour un L estimateur).

Ces estimateurs définissent une nouvelle classe de caractéristiques de location. On étudie les différentes propriétés de ces caractéristiques en les considérant comme des fonctionnelles définies d'un ensemble de fonctions de distribution dans R.

Finalement, on présente une nouvelle relation d'ordre pour les fonctions de distribution. On prouve certaines propriétés fondamentales de cette relation en utilisant un concept d'unimodalité généralisé.

Department of Mathematics
McGill University
Ph.D. Thesis

Acknowledgements

I wish to express my gratitude to my supervisor, Prof. Harold Ruben, for his helpful comments and valuable suggestions during the elaboration of this work.

I am also indebted to Prof. Constance van Eeden for some challenging discussions.

I wish to thank all the statistics group in the Department of Mathematics of McGill University, especially Prof. George P.H. Styan, for their encouragement.

I am grateful to the National Research Council Canada and to la Direction générale de l'Enseignement Supérieur du Québec for their financial assistance.

Finally, I want to pay special tribute to Maura Crilly for her skillful typing and for the patience she displayed towards the errors of an inexperienced author.

TABLE OF CONTENTS

Chapter I

Estimation of the location parameter, an historical survey

In this thesis, the following problem is investigated: let X_1 , $X_2,\ldots,X_n$ be a random sample from a distribution F . Suppose $F(x) = G(x-\theta_x)$, the parameter θ_x is to be estimated.

Section I.1 The mean : statistical and intuitive content

Long before there was any statistical concern on this subject, the mean was commonly used. Even now, despite the "a la mode" emphasis on robustness, it still retains its supremacy among all measures of location.

As a consequence, the location parameter and its estimator, the mean, have usually been singled out. This is exemplified by the early robust methods implemented to decrease the influence of extreme observations by rejection of outliers. In this approach, one computes the "mean" without taking the outliers into consideration.

For the non statistician, the mean is the location parameter. In day to day life, there are many examples of this identification, for instance, the final standing of a student is obtained by averaging his marks, the average production of goals by a hockey player is a measure of his ability.

One may tentatively explain such a popularity using the following arguments: the mean is an easily computable statistic and it has an intuitive content which is lacking to its modern challengers; it measures the average performance of the observed random phenomenon. The mean carries intuitive meaning while most "robust" estimators do not.

The statistical model which legitimates this approach is well known: assume that G is the standard normal distribution, then the mean is the minimum variance unbiased estimator.

Section I.2 Early methods

This section is taken out of a paper by Stigler (1973) about the history of robust estimation.

Before the end of the last century, statistics was not a science in its own right. The problem of estimating location was left to the experimenter.

In 1763, James Short (an English astronomer) had estimated the sun's parallax using the average of 3 means (the standard one and 2 trimmed versions).

In 1852, Benjamin Pierce (a mathematician astronomer) introduced a test to find outliers. He estimated the location using the mean of the restricted sample.

The method of least squares was introduced by Gauss and Legendre at the beginning of the nineteenth century. By the end of that century, weighted least squares estimates were commonly used. The weight of each observation depended upon the experimenter's estimate of the probable error.

The median and other simple functions of order statistics were introduced by Laplace and Gauss in the 1810's.

Section I.3 Modern parametric estimation

Traditionally, the estimation of the location parameter has been linked to the estimation of the scale parameter. The problem investigated was the following: $X_1, X_2, \ldots, X_n$ is a random sample from a distribution F, where $F(x) = G((x-\theta_x)/\sigma_x)$ in which G is known, θ_x and σ_x, the location and scale parameter respectively are to be estimated.

The actual approach is however different. Nowadays, when estimating location one assumes that the scale is known. If, however, it is not known, and if it is needed for the estimation, one uses a simple estimator, such as, for example, the interquartile range.

The results discussed in the later part of this section concern only the location problem, they really are special cases of the location-scale results originally found.

Assume that $\sigma_x = 1$, that G is known and has an absolutely continuous density. The maximum likelihood method, foreshadowed by Gauss (and much lately by Edgeworth), estimates θ_x using the value of θ maximizing:

$$\prod_{i=1}^{n} g(X_i - \theta)$$

where g is the density of G. $\hat{\theta}_x$ is a solution of:

$$(1.1) \qquad \sum_{i=1}^{n} \psi\,(X_i - \theta) = 0$$

where $\psi(x) = -\,g'(x)/g(x)$.

This approach requires the use of numerical analysis techniques to solve (1.1).

In order to find suitable and easily computable estimators, Lloyd (1952) used the least squares method in the following way:

Let $X_{(1)} < X_{(2)} <..< X_{(n)}$ be the order statistics from the random sample under consideration.

Let $U_{(k)} = X_{(k)} - \theta_x$. Note that the $U_{(k)}$'s are the order statistics from a random sample with known distribution G . One computes

$$E(U_{(k)}) = \alpha_k \qquad k=1,2,\ldots,n$$

$$\text{Cov }(U_{(k)}, U_{(\ell)}) = v_{k\ell} \qquad k,\ell=1,2,\ldots,n \ .$$

Let $\underset{\sim}{\alpha}$ be the vector of the α_k's, V be the n×n matrix of the $v_{k\ell}$'s and $\underset{\sim}{X}_{(\ )}$ be the vector of the $X_{(i)}$'s.

$E(\underset{\sim}{X}_{(\ )}-\underset{\sim}{\alpha}) = \underset{\sim}{e}\,\theta_x$ where $\underset{\sim}{e}$ is an n×1 vector of 1's and

$$\text{Cov }(\underset{\sim}{X}_{(\ )}-\alpha) = V \ .$$

The least squares estimator of θ_x, $\hat{\theta}_x$ is the one minimizing:

$$(\underset{\sim}{X}_{(\ )} - \underset{\sim}{\alpha} - \underset{\sim}{e}\,\theta)\,'V^{-1}\,(\underset{\sim}{X}_{(\ )} - \underset{\sim}{\alpha} - \underset{\sim}{e}\,\theta) \ .$$

Using the Gauss-Markov theorem, this estimator is the best linear unbiased estimator of θ_x. One computes:

$$\hat{\theta}_x = (\underset{\sim}{e}\,'V^{-1}\,\underset{\sim}{e})^{-1}\,\underset{\sim}{e}'\,V^{-1}\,(\underset{\sim}{X}_{(\ )} - \underset{\sim}{\alpha}) \ .$$

and

$$V(\hat{\theta}_x) = (\underset{\sim}{e}'\,V^{-1}\,\underset{\sim}{e})^{-1} \ .$$

Much time has been spent computing numerically the $v_{\ell k}$'s and α_k's for well known distributions.

Jung (1958) and Bennett (1952) have studied the asymptotic approximation of this estimator which had been introduced by Daniell (1920). They estimated θ_x using the L estimator $\hat{\theta}_x$:

$$\hat{\theta}_x = \frac{1}{n} \sum_{i=1}^{n} J(i/(n+1)) \; X_{(i)}$$

where J(t) is a function defined in [0,1].

If J satisfies

$$J(G(x)) = (\frac{d}{dx} - g'(x)/g(x))/\int_{-\infty}^{\infty}(g'(y))^2(g(y))^{-1}dy \; ,$$

they showed that $\hat{\theta}_x$ is an efficient estimator.

Estimators based on ranks were introduced by Hodges and Lehmann (1963), we shall not here consider such estimators.

Section I.4 Huber's contribution and the actual context

Tukey (1960) initiated the modern preoccupation on robustness. Tukey posed the following question: "In parametric estimation, the underlying distribution is assumed a priori to be G_o let say, then using the statistical theory an optimal estimator for θ_x can be found; are the optimality properties preserved if G , the true underlying distribution, differs from G_o ?"

The answer to this question is no: for example, the mean of a random sample from a contaminated normal behaves very poorly.

Huber (1964) made the first important contribution to the theory of robust estimation of a location parameter. First he defined the M estimator of a location parameter in the following way:

let ρ be a given convex function having derivative ψ . Then the M estimator of θ_x is obtained by minimizing

$$\sum_{i=1}^{n} \rho(X_i-\theta) ,$$

hence, $\hat{\theta}_x$ is a solution of

$$\sum_{i=1}^{n} \psi(X_i-\theta) = 0 .$$

If $\rho(x)=x^2$, the M estimator is the least squares estimator. This estimator is the analogue of the maximum likelihood estimator, instead of maximizing the likelihood function, one is maximizing

$$\prod_{i=1}^{n} e^{-\rho(X_i-\theta)} .$$

Huber then investigated the following problem: if G is unknown but lies in a given "neighborhood" of a known distribution G_o , say; under certain regularity conditions he found a minimax estimator for θ_x , i.e. an estimator minimizing, in this neighborhood of G_o , the maximum of the "error". In the asymptotic theory of the minimax estimation, the variance is used to measure the error. The minimax M estimator is based upon

$$\psi(x) = \begin{cases} -\ k & x \le x_0 \\ -\ \dfrac{g_0'(x)}{g_0(x)} & x_0 < x \le x_1 \\ \ k & x_1 < x \end{cases}$$

where x_0 and x_1 are the end points of the interval $\{x : |g'(x)/g(x)| < k\}$.

Once this work was completed, the following questions arose: "How can order statistics be used in robust estimation?", "Are there L estimators asymptotically analogous to the minimax M estimator?" "What is the asymptotic relation between M and L estimators?"

Some of these questions were answered by Jaeckel (1971). Jaeckel found a minimax L estimator for symmetric g_0, if:

$$J(G_0(x)) = \begin{cases} 0 & x<-M \\ [\frac{d}{dx} - g_0'(x)/g(x)]/\int_{-M}^{M}(g_0'(y))^2(g_0(y))^{-1}dy & -M\leq x\leq M \\ 0 & M<x \end{cases},$$

the L estimator based on J is minimax for the given neighborhood of G_0 .

Note that this L estimator is a trimmed version of the efficient L estimator for G_0 found by Bennett (1952) and Jung (1958).

Besides these questions, the Princeton Robustness study (1972) highlights the fact that minimax M estimators are satisfactory if the contamination of the known distribution was small. In a highly contaminated situation these estimators break down.

To cope with this difficulty highly robust M estimators were introduced.

The content of this thesis lies in the continuation of these results.

Some pending problems will be discussed in the second chapter: the asymptotic relation between M and L estimators, the L counter part of highly robust M estimator will be investigated and a formal theory of the highly robust M estimator will be developed.

Chapter II

Asymptotic properties of a new estimator of the location parameter

Section II.1 Two known estimators of the location parameter

The problem under investigation is the following:

Let G be a distribution function and

$$\mathcal{G}_\theta = \{G(x-\theta) : \theta \in R\} .$$

A random variable X, having distribution F in $\mathcal{G}_\theta$ is sampled and we wish to estimate θ_x such that

$$F(x) = G(x-\theta_x) .$$

Let $X_1,\ldots,X_n$ be a random sample of X and $X_{(1)},\ldots,X_{(n)}$ the corresponding ordered sample.

Definition II.1 Estimator based on linear combination of ordered statistics or L estimator

Let J(t) be an integrable function on [0,1] such that $\int_0^1 J(t)dt = 1$.

Then

$$\hat{T}_n = \frac{1}{n} \sum_{i=1}^{n} J\left(\frac{i}{n+1}\right) X_{(i)}$$

is an L estimator.

Definition II.2 Huber's M estimator

Let $\psi(x)$ be an increasing function such that $\psi(x)$ is positive (negative) for large positive (negative) values of x. The M estimator $\hat{T}_n$ is defined as a solution of

$$\sum_{i=1}^{n} \psi(X_i-\theta) = 0 .$$

Definition II.3

Let F(x) be a distribution function and define

$$F^{-1}(t) = \inf \{x : F(x) \geq t\} \quad .$$

Theorem II.1

If (i) J(t) is a bounded variation function in [0,1] and J(t) = 0 $t \notin [\delta, 1-\delta]$ where $1/2>\delta>0$,

(ii) $F^{-1}(t)$ and J(t) are not discontinuous together,

the estimator

$$\hat{T}_n = \frac{1}{n} \sum_{i=1}^{n} J\left(\frac{i}{n+1}\right) X_{(i)}$$

is such that

$$L(n^{1/2}(\hat{T}_n - \mu)) \xrightarrow[n\to\infty]{} N(0,\sigma^2)$$

where (i) $L(X)$ is the distribution of X

(ii) $\mu = \int_0^1 J(t)\, F^{-1}(t)dt$

(iii) $\sigma^2 = \int_0^1 [A(t)]^2 dt - [\int_0^1 A(t)dt]^2$

and $d\,A(t) = J(t)\, d\, F^{-1}(t)$.

Proof: See Huber (1969) page 129.

Theorem II.2 (Huber (1969) p. 67)

If $\lambda(\xi) = \int_{-\infty}^{\infty} \psi(x-\xi)dF(x)$ is such that

(i) $\lambda(\xi_0) = 0$,

(ii) $\lambda(\xi)$ is continuous monotone in a neighborhood of ξ_0 ,

(iii) $\int_{-\infty}^{\infty} (\psi(x-\xi))^2 dF(x)$ is finite and continuous at ξ_0 ,

the estimator $\hat{T}_n$ solution of:

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

is such that

1) $\mathcal{L}(n^{1/2}\lambda(\hat{T}_n)) \xrightarrow[n \to \infty]{} N(0,\sigma_1^2)$ where $\sigma_1^2 = \int_{-\infty}^{\infty}(\psi(x-\xi_0))^2 dF(x)$,

2) if furthermore λ is differentiable at ξ_0 and if $\lambda'(\xi_0) \in (-\infty,0)$

$$\mathcal{L}(n^{1/2}(\hat{T}_n-\xi_0)) \xrightarrow[n \to \infty]{} N(0,\sigma^2) \quad \text{where} \quad \sigma^2 = \sigma_1^2/(\lambda'(\xi_0))^2 .$$

Section II.2 Introduction of Gaussian processes

This section contains a survey of known results about Gaussian processes which will be used in section II.4.

Lemma II.1

If U is a random variable with a $\mathcal{U}[0,1]$ distribution, $F^{-1}(U)$ is a random variable having distribution F .

Proof: Clear

Therefore if X is a random variable having distribution F, X and $F^{-1}(U)$, where U is $\mathcal{U}[0,1]$, are identically distributed. Hence an ordered sample from X, $X_{(1)},\ldots,X_{(n)}$, can be written as $F^{-1}(U_{(1)}),\ldots,F^{-1}(U_{(n)})$ where $U_{(1)},\ldots,U_{(n)}$ is an ordered sample from U .

Let $U^{(n)}(t)$ be a stochastic process defined in the following way:

$$U^{(n)}(t) = U_{(i)} \qquad t = \frac{i}{n+1} \; i=0,1,\ldots,n+1$$

$U^{(n)}(t)$ = the linear interpolation between the points

$(\frac{i}{n+1} , U_{(i)})$ and $(\frac{i+1}{n+1} , U_{(i+1)})$ $\qquad t \in (\frac{i}{n+1} , \frac{i+1}{n+1}) \qquad i = 0,1,\ldots,n$

where $U_{(0)} = 1$, $U_{(n+1)} = 1$

Let: $Z^{(n)}(t) = n^{1/2}(U^{(n)}(t) - t)$.

Definition II.4 Gaussian process

A stochastic process with continuous path $Z(t)$ is Gaussian if all the vectors $(Z(t_1),\ldots,Z(t_m))$ where m is finite and $t_i \varepsilon [0,1]$ are normally distributed. The distribution of such a process is specified by $E(Z(t))$ for all $t \varepsilon [0,1]$ and $\mathrm{cov}(Z(t),Z(s))$ for s and t in the unit square.

A Gaussian process satisfying:

(i) $E(Z(t)) = 0 \quad t \varepsilon [0,1]$,

(ii) $\mathrm{Cov}(Z(t),Z(s)) = \min(s,t)-st$

for s and t in [0,1] is called a Brownian Bridge.

Definition II.5 Weak convergence of a sequence of stochastic processes

Let $\{Y^{(n)}(t)\}_{n=1}^{\infty}$ be a sequence of processes taking values in C[0,1], the space of continuous functions in [0,1] with the sup norm, $Y^{(n)}(t)$ is said to converge weakly to $Y(t)$ if for any continuous functional q defined on C[0,1]:

$$L(q(Y^{(n)}(.))) \xrightarrow[n \to \infty]{} L(q(Y(.))) .$$

Theorem II.3

The sequence of processes $\{Z^{(n)}(t)\}_{n=1}^{\infty}$ previously defined is such that:

$$Z^{(n)}(t) \xrightarrow[n \to \infty]{} Z(t) \text{ weakly,}$$

where $Z(t)$ is the Brownian Bridge.

Proof: Huber (1969) p. 115.

Theorem II.4

Given $\varepsilon>0$, there exist M_ε in R , n_ε in N such that

$$n > n_\varepsilon \rightarrow \sup_{i=1,\ldots,n} n^{1/2} |U_{(i)} - \frac{i}{n+1}| < M_\varepsilon$$

except on a set of probability ε .

The proof of this theorem is an easy consequence of theorem II.3.

Hájek and Sidák (1967) p. 174-184 and Billingsley (1968) p. 102-108 provide some explanations on this topic.

Lemma II.2

Let $G(t)$ be a bounded variation function in $[0,1]$ then $\int_0^1 Z(t)dG(t)$ is distributed $N(0,s^2)$, where

$$s^2 = \int_0^1 [G(t)]^2 \, dt - [\int_0^1 G(t)dt]^2$$

Proof: see Miller (1964) p. 103-104.

Section II.3 L-M estimator of the location parameter

Definition II.6 L-M estimator of the location parameter

Let: (i) $J(t)$ be a positive bounded variation function defined on $[0,1]$ such that $\int_0^1 J(t) \, dt > 0$,

(ii) $\psi(x)$ be an increasing left continuous function which is positive (negative) for large positive (negative) values of x .

The L-M estimator $\hat{T}_n$ based on J and ψ is defined as a solution of:

(2.0) $$\sum_{i=1}^{n} J\left(\frac{i}{n+1}\right) \psi\left(X_{(i)}-\theta\right) = 0 .$$

Remarks

(i) $\hat{T}_n$ is defined only if there exists $i \in \{1,2,\ldots,n\}$ such that $J(\frac{i}{n+1}) > 0$,

(ii) Since ψ is not assumed to be continuous nor strictly increasing equation (2.0) may not have one and only one solution. In those cases define:

$$\hat{T}_n{}^* = \inf\{\theta : \sum_{i=1}^{n} J\left(\frac{i}{n+1}\right) \psi\left(X_{(i)} - \theta\right) \le 0\} ,$$

$$\hat{T}_n{}^{**} = \sup\{\theta : \sum_{i=1}^{n} J\left(\frac{i}{n+1}\right) \psi\left(X_{(i)} - \theta\right) \ge 0\}$$

and $\hat{T}_n = \alpha\hat{T}_n{}^* + (1-\alpha)\,\hat{T}_n{}^{**}$ where $\alpha\in[0,1]$.

Note that since ψ is left continuous, the LHS of (2.0) is right continuous and

$$\sum_{i=1}^{n} J\left(\frac{i}{n+1}\right) \psi\left(X_{(i)} - \hat{T}_n\right) \le 0 ,$$

(iii) If $J(t) = 1$ for all $t\in[0,1]$, L-M estimators reduce to M estimators,

(iv) If $\psi(x) = x \quad x\in R$

$$\hat{T}_n = \left(\sum_{i=1}^{n} J\left(\frac{i}{n+1}\right)\right)^{-1} \sum_{i=1}^{n} J\left(\frac{i}{n+1}\right) X_{(i)}$$

and, in this case L-M estimators are asymptotically equivalent to L estimators.

Define:

$$x) = \int_0^1 J(t)\psi(F^{-1}(t)-x)dt .$$

The next two lemmas will provide some clues about the behaviour of λ .

Lemma II.3

If there exists $x_0 \varepsilon$ R satisfying $|\lambda(x_0)| < \infty$, then

(i) $\lambda(x)$ is defined for all x in R (eventually $\lambda(x) = \pm \infty$)

(ii) $\lambda(x)$ is decreasing, positive (negative) for large negative (positive) values of x .

This lemma is a straightforward generalization of Huber's (1969) lemma p. 64.

Lemma II.4

If $b > a$ in R satisfy:

$$\int_0^1 J(t)|\psi(F^{-1}(t)-a)|dt < \infty \quad \text{and}$$

$$\int_0^1 J(t)|\psi(F^{-1}(t)-b)|dt < \infty \quad \text{then}$$

$\lambda(x)$ is finite in [a,b], continuous in (a,b).

Proof: Take $x \varepsilon (a,b)$, since ψ is increasing

$$|\psi(F^{-1}(t)-x)| \leq \max\{|\psi(F^{-1}(t)-a)| , |\psi(F^{-1}(t)-b)|\}$$
$$< |\psi(F^{-1}(t)-a)| + |\psi(F^{-1}(t)-b)| .$$

So that:

$$\int_0^1 J(t)|\psi(F^{-1}(t)-x)|dt < \infty ,$$

and

$$|\lambda(x)| < \infty .$$

To prove the continuity take x_0 in (a,b) , for any x in [a,b]:

$$|\lambda(x_0) - \lambda(x)| \leq$$

$$\int_0^1 J(t)\,|\psi(F^{-1}(t)-x_0) - \psi(F^{-1}(t)-x)|\,dt \ .$$

Since $$|\psi(F^{-1}(t)-x_0) - \psi(F^{-1}(t)-x)| \ <$$

$$2\{|\psi(F^{-1}(t)-a)| + |\psi(F^{-1}(t)-b)|\}$$

and $\psi(F^{-1}(t)-x_0)$ is continuous a.e.dt as an increasing function of t :

$$\lim_{x \to x_0} |\lambda(x) - \lambda(x_0)| = 0 \ ,$$

using the Lebesgue monotone convergence theorem.

This ends the proof.

Now on the asymptotic normality of these estimators will be proved. One has investigated many ways to find conditions as mild as possible for this asymptotic normality to hold.

For M estimators, Huber (1964), using the Lindeberg Lévy condition has obtained what one might call the best possible result for this restricted area.

For L estimators satisfying: $J(t) = 0 \ t \notin [\delta, 1-\delta]$ for a $\delta \in (0, 1/2)$, Huber (1969) has again obtained the best possible result using the weak convergence of the $Z^{(n)}(t)$'s, defined in section II.2, to the Brownian bridge.

Several attempts have been made in order to prove the asymptotic normality of general L estimators.

Jung (1958) proved the result under very restrictive conditions on J . Recent authors have focussed their attention mainly in two directions.

First, they were looking for a sum S_n of independent random variables such that:

$$\lim_{n \to \infty} n^{1/2} (\hat{T}_n - S_n) = 0 \text{ in probability,}$$

where $\hat{T}_n$ is the L estimator. Then the asymptotic normality of $n^{1/2} S_n$ is a consequence of the Lindeberg Lévy condition for the Central Limit theorem. Chernoff Gastwirth and Johns (1967) gave such a proof under too many regularity conditions. Stigler (1969) (1974) using Hajek (1968) projection method provided a very elegant proof under reasonable assumptions.

The second method uses the weak convergence of the $Z^{(n)}(t)$'s towards $Z(t)$. Shorack (1969) (1972) proved the result under conditions more restrictive than those of Stigler (1974). Furthermore application of Shorack's (1972) results to this problem requires very stringent assumptions on J , the weight function.

A new method will be introduced now. First Huber's result for L estimators with $J(t) = 0 \; t \notin [\delta, 1-\delta]$ will be generalized to L-M estimators having the same property. Then, using this result it will be shown that $\hat{T}_n$ has the same asymptotic behaviour as a sum S_n of independent random variables.

The following two lemmas will be used in the next two sections:

Lemma II.5

If (i) $\{k_n\}_{n=1}^{\infty}$ is an R-sequence converging to ξ_0 ,

(ii) $\lambda_n(x) = \sum_{i=1}^{n} J\left(\frac{i}{n+1}\right) \psi\left(F^{-1}\left(\frac{i}{n+1}\right) - x\right)$,

(iii) there exists $\delta_{(1)} \varepsilon (0, {}^1/2)$ such that

$$J(t) = 0, t < \delta_{(1)} \quad ,$$

or (iv) $\lim_{t \to 0} \sqrt{t}\, \psi\, (F^{-1}(t) - x) = 0$ for any x in a neighborhood of ξ_0 ,

(v) there exists $\delta_{(2)} \varepsilon (0, {}^1/2)$ such that

$$J(t) = 0, \ t > 1 - \delta_{(2)} ,$$

or (vi) $\lim_{t \to 1} \sqrt{1-t}\, \psi\, (F^{-1}(t) - x) = 0$ for any x in a neighborhood of ξ_0 ,

then $\lim_{n \to \infty} n^{1/2} (\lambda(k_n) - \lambda_n(k_n)) = 0$.

Proof: Without loss of generality, assume $\xi_0 = 0$. The convergence of k_n implies: for any $\varepsilon > 0$, there exists $n_0 = n_0(\varepsilon)$ satisfying:

$$n > n_0 \to |k_n| < \varepsilon .$$

Pick δ in the following way:

a) (iii) holds take $\delta = \delta_{(1)}$

b) (iv) holds and $\lim_{t \to 0} \psi\, (F^{-1}(t) - \varepsilon) > -\infty$ take $\delta = 0$

c) (iv) holds and $\lim_{t \to 0} \psi\, (F^{-1}(t) - \varepsilon) = -\infty$, take δ in $(0, {}^1/2)$ small enough such that: $\psi(F^{-1}(t) + \varepsilon) < 0$ for $t < \delta$.

Let:
$$J^{(1)}(t) = \begin{cases} J(t) & t \varepsilon [\delta, 1] \\ 0 & \text{elsewhere} \end{cases}$$

and $\lambda^{(1)}(x)$, $\lambda_n^{(1)}(x)$ be the corresponding λ and λ_n function with J replaced by $J^{(1)}$. We first prove:

$$(2.1) \qquad \lim_n n^{1/2} (\lambda^{(1)}(k_n) - \lambda_n^{(1)}(k_n)) = 0 \ .$$

If $\delta=0$ take $\delta_1=0$, if $\delta>0$ take δ_1 in $(0,\delta)$.

For t in $[\delta_1,1]$ there exists M_0 in R such that

$$\psi(F^{-1}(t)-\varepsilon) > M_0 \ .$$

If M_0 is positive, $\psi(F^{-1}(t) - k_n)$ is increasing positive for $n>n_0$. If M_0 is negative then $\psi(F^{-1}(t) - k_n)$ can be written as the difference of two positive increasing functions in $[F^{-1}(\delta_1)-\varepsilon \ , \infty)$:

$$\psi(x) = \psi_1(x) - \psi_2(x)$$

where $\psi_2(x) = -M_0, x\varepsilon[F^{-1}(\delta_1)-\varepsilon,\infty)$. Note that if ψ fulfills assumption (iv), ψ_1 and ψ_2 fulfill the same assumption.

Since $J^{(1)}$ is a bounded variation function there exist two increasing functions $J_1^{(1)}$ and $J_2^{(1)}$ such that for $t\varepsilon[\delta_1,1]$:

$$J^{(1)}(t) = J_1^{(1)}(t) - J_2^{(1)}(t) \ .$$

Note that:

$$J^{(1)}(t) = J_1^{(1)}(t) + |c_0| - \{J_2^{(1)}(t) + |c_0|\}$$

where $c_0=\min \{J_1^{(1)}(\delta_1) \ , \ J_2^{(1)}(\delta_1) \ , \ 0\}$. Hence it can be assumed that $J_1^{(1)}$ and $J_1^{(2)}$ are positive increasing.

Therefore it will suffice to prove (2.1) under the following assumptions:

d) $J^{(1)}(t)$ is positive increasing in $[\delta_1,1]$,

e) $\psi(x)$ is positive increasing in $[F^{-1}(\delta_1) - \varepsilon, \infty)$.

If $\delta=0$, take $n_1=n_0$ if $\delta>0$ there exists $n_1>n_0$ such that

$$n_1 > \frac{2}{\delta-\delta_1} \ .$$

For $n>n_1$, consider:

$$\lambda^{(1)}(k_n) = \int_\delta^1 J^{(1)}(t)\ \psi\ (F^{-1}(t)-k_n)\ dt$$

$$\leq \sum_{i=[n\delta]+1}^{n} \int_{\frac{i-1}{n}}^{\frac{i}{n}} J^{(1)}(t)\ \ \psi\,(F^{-1}(t)-k_n)\ dt$$

$$\leq \frac{1}{n} \sum_{i=[n\delta]+1}^{n-1} J^{(1)}(\tfrac{i}{n})\ \psi\ \ (F^{-1}(\tfrac{i}{n}) - k_n)$$

$$+ \int_{1-\frac{1}{n}}^{1} J^{(1)}(t)\ \psi\ \ (F^{-1}(t)-k_n)\ dt$$

using assumptions d) and e).

The Σ in the previous expression is less or equal than:

$$\frac{1}{n} \sum_{i=[n\delta]+1}^{n-1} J^{(1)}(\tfrac{i+1}{n+1})\ \ \psi\ (F^{-1}(\tfrac{i+1}{n+1}) - k_n)$$

$$\leq \frac{1}{n} \sum_{i=[n\delta]+1}^{n} J^{(1)}(\tfrac{i}{n+1})\ \ \psi\ (F^{-1}(\tfrac{i}{n+1}) - k_n)$$

$$= \lambda_n^{(1)}(k_n)\ .$$

In order to prove:

(2.2) $\quad \limsup_n n^{1/2}\ (\lambda^{(1)}(k_n) - \lambda_n^{(1)}(k_n)) \leq 0\ ,$

one must show:

$$\lim_n n^{1/2} \int_{1-\frac{1}{n}}^{1} J^{(1)}(t)\ \psi\ (F^{-1}(t)-k_n)\ dt = 0\ .$$

If (v) holds, that is obvious. Suppose (vi) holds and that ε is small enough such that:

$$\lim_{t\to 1} \sqrt{1-t}\ \psi\ (F^{-1}(t)+\varepsilon) = 0\ .$$

Choose $n_2 > n_1$ in N such that for

$$t > 1 - \frac{1}{n_2} , \quad \sqrt{1-t}\ \psi\,(F^{-1}(t)+\varepsilon) < \varepsilon .$$

Let $M_1 = \sup_{t\varepsilon[\delta_1,1]} J^{(1)}(t)$, for $n > n_2$:

$$n^{1/2} \int_{1-\frac{1}{n}}^{1} J^{(1)}(t)\ \psi\,(F^{-1}(t) - k_n)\ dt$$

$$\leq n^{1/2} M_1\ \varepsilon \int_{1-\frac{1}{n}}^{1} \frac{1}{\sqrt{1-t}}\ dt$$

$$= \varepsilon M_1 .$$

Therefore (2.2) holds. To prove (2.1) it will suffice to prove:

$$\liminf_{n} n^{1/2}(\lambda^{(1)}(k_n) - \lambda_n^{(1)}(k_n)) \geq 0 .$$

For $n > n_2$, consider:

$$\lambda^{(1)}(k_n) = \int_{\delta}^{1} J^{(1)}(t)\ \psi\,(F^{-1}(t) - k_n)\ dt$$

$$\geq \sum_{i=[n\delta]+1}^{n-1} \int_{\frac{i}{n}}^{\frac{i+1}{n}} J^{(1)}(t)\ \psi\,(F^{-1}(t) - k_n)\ dt$$

$$\geq \frac{1}{n} \sum_{i=[n\delta]+1}^{n-1} J^{(1)}(\tfrac{i}{n})\ \psi\,(F^{-1}(\tfrac{i}{n}) - k_n)$$

using assumptions d) and e) ,

$$\geq \frac{1}{n} \sum_{i=[n\delta]+1}^{n-1} J^{(1)}(\tfrac{i}{n+1})\ \psi\,(F^{-1}(\tfrac{i}{n+1}) - k_n)$$

$$= \lambda_n^{(1)}(k_n) - \frac{1}{n} J^{(1)}(\tfrac{n}{n+1})\ \psi\,(F^{-1}(\tfrac{n}{n+1}) - k_n) .$$

Using assumption (v) or (vi), it is easily seen that:

$$\lim_n \frac{1}{\sqrt{n}} J\left(\frac{n}{n+1}\right) \psi\left(F^{-1}\left(\frac{n}{n+1}\right) - k_n\right) \le 0$$

and (2.1) is true.

The lemma is true under assumption a) or b), suppose c) holds: let

$$J^{(2)}(t) = J(t) - J^{(1)}(t) .$$

Let $\lambda^{(2)}$ and $\lambda_n^{(2)}$ be the corresponding expression for λ and λ_n when J is replaced by $J^{(2)}$. Pick δ_2 in $(\delta,1)$, using an argument similar to the preceeding one, it will suffice to show that:

(2.3) $$\lim_n n^{1/2} (\lambda^{(2)}(k_n) - \lambda_n^{(2)}(k_n)) = 0$$

under the following conditions:

f) $J^{(2)}$ is a negative increasing bounded function in $[0, \delta_2]$,

g) ψ is a negative increasing function in $(-\infty , F^{-1}(\delta_2)+\varepsilon)$.

Pick $n_3 > n_2$ in N such that

$$n_3 > \frac{2}{\delta_2-\delta} .$$

For $n > n_3$ consider:

$$\lambda^{(2)}(k_n) = \int_0^\delta J^{(2)}(t) \psi (F^{-1}(t) - k_n)\, dt$$

(2.4) $$\le \sum_{i=1}^{[n\delta]+1} \int_{\frac{i-1}{n}}^{\frac{i}{n}} J^{(2)}(t) \psi (F^{-1}(t) - k_n)\, dt .$$

Using assumptions f) and g) and the fact that the product of two increasing negative functions is a positive decreasing function, (2.4) is less or equal to:

$$\frac{1}{n} \sum_{i=2}^{[n\delta]+1} J^{(2)}\left(\frac{i-1}{n}\right) \psi \left(F^{-1}\left(\frac{i-1}{n}\right) - k_n\right) + \int_0^{\frac{1}{n}} J^{(2)}(t) \psi (F^{-1}(t) - k_n)\, dt$$

$$\leq \frac{1}{n}\sum_{i=1}^{[n\delta]} J^{(2)}(\tfrac{i}{n+1})\ \psi\ (F^{-1}(\tfrac{i}{n+1}) - k_n) + \int_0^{1/n} J^{(2)}(t)\ \psi\ (F^{-1}(t) - k_n)\ dt\ .$$

To prove:

$$\lim_n n^{1/2} \int_0^{1/n} J(t)\ \psi\ (F^{-1}(t) - k_n)\ dt = 0$$

one uses an argument similar to the one used in the proof of (2.2) with assumption (iv). Therefore:

$$\limsup_n n^{1/2}(\lambda^{(2)}(k_n) - \lambda_n^{(2)}(k_n)) \leq 0\ .$$

To end the proof, for $n > n_3$, consider:

$$\lambda^{(2)}(k_n) = \int_0^{\delta} J^{(2)}(t)\ \psi\ (F^{-1}(t) - k_n)\ dt$$

$$\geq \sum_{i=1}^{[n\delta]} \int_{\frac{i-1}{n}}^{\frac{i}{n}} J^{(2)}(t)\ \psi\ (F^{-1}(t) - k_n)\ dt$$

$$\geq \frac{1}{n}\sum_{i=1}^{[n\delta]} J(\tfrac{i+1}{n+1})\ \psi\ (F^{-1}(\tfrac{i+1}{n+1}) - k_n)\ dt$$

using assumptions f) and g),

$$= \lambda_n^{(2)}\ (k_n) - \frac{1}{n} J(\tfrac{1}{n+1})\ \psi\ (F^{-1}(\tfrac{1}{n+1}) - k_n)\ .$$

Using assumption (iv), it is easily seen that:

$$\lim_n n^{-1/2} J(\tfrac{1}{n+1})\ \psi\ (F^{-1}(\tfrac{1}{n+1}) - k_n) \leq 0\ .$$

Hence (2.3) holds and the lemma is proved.

Note that $\int_0^1 \psi^2(F^{-1}(t)-\xi)\ dt < \infty$ in a neighborhood of ξ_0 is a sufficient condition for assumptions (iv) and (vi) to hold.

Lemma II.6

If i) $\lambda(\xi_0) = 0$,

ii) λ is continuous monotone near ξ_0 ,

iii) the asymptotic distribution of $n^{-1/2} \sum_{i=1}^{n} \{J(\frac{i}{n+1}) \psi (X_{(i)}-k_n) - \lambda(k_n)\}$ is continuous where $\{k_n\}_{n=1}^{\infty} = \{k_n(g)\}_{n=1}^{\infty}$, is defined by:

$$n^{1/2} \lambda(k_n) = g ,$$

then $\lim_n P\{n^{1/2}\lambda(\hat{T}_n) > g\} =$

$$\lim_n P\{n^{-1/2} \sum_{i=1}^{n} (J(\frac{i}{n+1}) \psi (X_{(i)}-k_n) - \lambda(k_n)) < - g\} .$$

Proof: Using the assumption on λ ,

$$\lim_n k_n = \xi_0 \quad \text{and}$$

for n big enough:

$$\{n^{1/2} \lambda (\hat{T}_n) \geq g\} = \{\hat{T}_n \leq k_n\} .$$

Using the definition of $\hat{T}_n$, one obtains:

$$\{\sum_{i=1}^{n} J(\frac{i}{n+1}) \psi (X_{(i)} - k_n) < 0\} \subset \{\hat{T}_n \leq k_n\} \quad \text{, for n big enough}$$

and if $\varepsilon > 0$:

$$\{T_n \leq k_n\} \subset \{\sum_{i=1}^{n} J(\frac{i}{n+1}) \psi (X_{(i)} - k_n) < \varepsilon\} \text{ or:}$$

$$\{n^{-1/2} \sum_{i=1}^{n} (J (\frac{i}{n+1}) \psi (X_{(i)} - k_n) - \lambda (k_n)) < -g \}$$

$$\subset \{\hat{T}_n \leq k_n\} \text{ and}$$

$$\{n^{-1/2} \sum_{i=1}^{n} (J (\frac{i}{n+1}) \psi (X_{(i)} - k_n) - \lambda(k_n)) < - g+\varepsilon \}$$

$\supset \{\hat{T}_n \leq k_n\}$ so that if the asymptotic distribution of $n^{-1/2} \sum_{i=1}^{n} (J(\frac{i}{n+1})\ \psi(X_{(i)} - k_n) - \lambda(k_n))$ is continuous the result is true.

Note that for any $\varepsilon > 0$ and for any n big enough:

$$\{n^{-1/2} \sum_{i=1}^{n} (J(i/(n+1))\ \psi(X_{(i)} - k_n) - \lambda(k_n)) < -g\}$$

$$\subset \{n^{1/2} \lambda(\hat{T}_n) \geq g\} \subset$$

$$\{n^{-1/2} \sum_{n=1}^{n} (J(i/(n+1))\ \psi(X_{(i)} - k_n) - \lambda(k_n)) \leq -g + \varepsilon\} .$$

Section II.4 Asymptotic normality of L-M estimators with $J(t) = 0, t \notin [\delta, 1-\delta]$, where $\delta \varepsilon (0, \frac{1}{2})$.

In this section, we assume $J(t) = 0, t \notin [\delta, 1-\delta]$. Note that for $t \varepsilon [\delta, 1-\delta]$:

$$|\psi(F^{-1}(t) - x)| < \max\{|\psi(F^{-1}(\delta)-x)|, |\psi(F^{-1}(1-\delta) - x)|\}$$

and

$$\int_0^1 J(t)\ |\psi(F^{-1}(t) - x)|dt < \infty \qquad x \varepsilon R .$$

Hence the conclusion of lemma II.4 holds and λ is continuous.

Definition II.7

Dealing with asymptotic properties of estimators, the following symbols will be used. A sequence of random variables $\{X_n\}_{n=1}^{\infty}$ is said to be

a) $o_p(1)$ if

$$\lim_n X_n = 0 \text{ in probability} ,$$

b) $O_p(1)$ if

$$\lim_n \sup |X_n| \text{ is bounded in probability.}$$

<u>Theorem II.5</u>

If i) $\psi\,(F^{-1}(t) - x)$ is continuous at ξ_0 a.e.dt where $\lambda\,(\xi_0) = 0$,

ii) $\psi\,(F^{-1}(t) - \xi_0)$ and $J(t)$ are not discontinuous together,

iii) λ is strictly decreasing near ξ_0 ,

then the L-M estimator $\hat{T}_n$ defined as a solution of:

$$\sum_{i=1}^{n} J(\tfrac{i}{n+1})\ \psi\,(X_{(i)} - \theta) = 0$$

is such that:

1) $\mathcal{L}(n^{1/2}\ \lambda\ (\hat{T}_n)) \xrightarrow[n \to \infty]{} N(0,\sigma_1^2)$

where $\sigma_1^2 = \int_0^1 [A(t)]^2\, dt - [\int_0^1 A(t)dt]^2$ and $d(A(t)) = J(t)d\psi\ (F^{-1}(t) - \xi_0)$,

2) if furthermore λ is differentiable at ξ_0 , and $\lambda'(\xi_0) \in (-\infty\ , 0)$:

$$\mathcal{L}(n^{1/2}\ (\hat{T}_n - \xi_0)) \xrightarrow[n \to \infty]{} N\ (0,\sigma^2) \quad \text{where} \quad \sigma^2 = \frac{\sigma_1^2}{[\lambda'(\xi_0)]^2}\ .$$

The following lemma is needed:

<u>Lemma II.7</u>

If (i) $J(t)$ and $\psi\,(F^{-1}(t))$ are not discontinuous together,

(ii) $\psi(F^{-1}(t)-x)$ is continuous at 0 a.e. dt,

(iii) $\{k_n\}_{n=1}^{\infty}$ is an R sequence such that

$$\lim_n k_n = 0\ ,$$

then, $$h_n(x(\cdot)) = n^{-1/2} \sum_{i=1}^{n} J\ (\tfrac{i}{n+1})\ [\psi(F^{-1}(\tfrac{i}{n+1} + n^{-1/2}\ x(\tfrac{i}{n+1})) - k_n)$$

$$-\ \psi\ (F^{-1}(\tfrac{i}{n+1}) - k_n)]$$

$$\text{and } h(x(\cdot)) = \int_0^1 x(t)\, J(t)\, d\psi(F^{-1}(t))$$

are such that:

$$h_n(x(\cdot)) - h(x(\cdot)) \text{ goes to 0 uniformly on the set } S$$

of functions satisfying:

i) $\sup_S ||x|| < \infty$ where $||x|| = \sup_{t\varepsilon[0,1]} |x(t)|$

ii) $\sup_S q_\alpha(x(\cdot)) \longrightarrow 0$ if $\alpha \longrightarrow 0$

where $q_\alpha(x(\cdot)) = \sup_{|t-s|<\alpha} |x(t) - x(s)|$.

<u>Proof</u>: Take δ' in $(0,\delta)$ such that F^{-1} is continuous at $\delta_0 = \frac{\delta+\delta'}{2}$ and $1-\delta_0$ and:

$$\psi(F^{-1}(\delta_0) - x) \quad , \quad \psi(F^{-1}(1-\delta_0) - x)$$

are continuous at 0 . Since F is a distribution function, $\sup_S ||x|| < \infty$ and $\lim_n k_n = 0$, there exist M_0 and n_0 in N such that:

$$n > n_0 \rightarrow F^{-1}(t+n^{-1/2} x(t)) - k_n \ \varepsilon\ [-M_0, M_0]$$

$$\text{and } F^{-1}(t) - k_n \ \varepsilon\ [-M_0, M_0]$$

for all t in $[\delta', 1-\delta']$ and for all $x(\cdot)\ \varepsilon\ S$. Using an argument similar to the one at the beginning of lemma II.5, we may assume:

a) $J(t)$ is a positive increasing bounded function in $[\delta', 1-\delta']$,

b) $\psi(x)$ is a positive increasing function in $[-M_0, M_0]$.

Define: $G(t) = \psi(F^{-1}(t))$

$$G_n(t) = \psi(F^{-1}(t) - k_n) .$$

For any $\varepsilon>0$, there exists $\alpha_1 = \alpha_1(\varepsilon)$, $\alpha_1>0$ such that:

$$\sup_S q_\alpha (x(\cdot)) < \varepsilon \text{ if } \alpha_1 \geq \alpha \quad ,$$

there exist $\alpha_2 = \alpha_2(\varepsilon), \alpha_2 > 0$ and $m_1 = m_1(\varepsilon)$ such that:

$$\left| \sum_{j=0}^{m_1-1} J(t_{j+1}) (G(t_{j+1}) - G(t_j)) - \int_{\delta_0}^{1-\delta_0} J(t) \, d\, G(t) \right| < \frac{\varepsilon}{2}$$

and

$$\left| \sum_{j=0}^{m_1-1} J(t_j) (G(t_{j+1}) - G(t_j)) - \int_{\delta_0}^{1-\delta_0} J(t) \, dG(t) \right| < \frac{\varepsilon}{2} \quad ,$$

for all partitions $\delta_0 = t_0 < t_1 < \ldots < t_{m_1} = 1-\delta_0$ of $[\delta_0, 1-\delta_0]$

satisfying:

$$\max_{0 \leq j < m_1} |t_{j+1} - t_j| < \alpha_2 \quad .$$

This is possible since J and G are not discontinuous together. Let:

$$\alpha_0 = \min\{\alpha_1 , \alpha_2\}$$

$m = m(\varepsilon)$, $m \in N$ and choose:

$$\delta_0 = t_0 < t_1 < \ldots < t_m = 1 - \delta_0$$

such that F^{-1} is continuous at t_j and ψ is continuous at $F^{-1}(t_j)$ for all $j \in \{1,2,\ldots,m\}$ and:

$$\sum_{j=0}^{m-1} (J(t_{j+1}) - J(t_j)) (G(t_{j+1}) - G(t_j)) < \varepsilon$$

$$\max_{0 \leq j < m} |t_{j+1} - t_j| < \alpha.$$

Note that almost all t's in [0,1] satisfy the continuity condition using assumption ii), hence such a choice of t_j's is feasible.

Using ε , m, α_0 and the fixed t_j's we will prove the lemma.

Consider:

$$h_n(x(\cdot)) = n^{-1/2} \sum_{j=0}^{m-1} \{E_{1j}^n + E_{2j}^n + E_{3j}^n\}$$

where:

$$\sum_i = \sum_{t_j \le \frac{i}{n+1} < t_{j+1}}$$

$$E_{1j}^n = \sum_i J(t_j) \{G_n(\frac{i}{n+1} + \gamma_j^n) - G_n(\frac{i}{n+1})\}$$

$$E_{2j}^n = \sum_i J(t_j) \{G_n(\frac{i}{n+1} + n^{-1/2} x(\frac{i}{n+1})) - G_n(\frac{i}{n+1} + \gamma_j^n)\}$$

$$E_{3j}^n = \sum_i \{J(\frac{i}{n+1}) - J(t_j)\} \{G_n(\frac{i}{n+1} + n^{-1/2} x(\frac{i}{n+1})) - G_n(\frac{i}{n+1})\}$$

$$\gamma_j^n = \frac{[n^{-1/2} x(t_j)]}{n+1}$$, note $[x]$ = entire value of x .

The sign of γ_j^n is function of $x(t_j)$ only, therefore assume $\gamma_j^n \ge 0$. For $j = m-1$, one should have:

$$\sum_i = \sum_{t_{m-1} \le \frac{i}{n+1} \le t_m} ,$$

since the three lost terms are bounded, we may omit them.

Let $n_1 \varepsilon N$, $n_1 > n_0$ such that

$$n > n_1 \rightarrow \gamma_j^n < \delta_0 - \delta' \text{ for all } j .$$

Consider:

$$E_{1j}^n = J(t_j) \sum_{t_{j+1} \le \frac{i}{n+1} < t_{j+1} + \gamma_j^n} G_n(\frac{i}{n+1}) - \sum_{t_j \le \frac{i}{n+1} < t_j + \gamma_j^n} G_n(\frac{i}{n+1})$$

for $n > n_1$

$$n^{-1/2} E_{1j}^n \le J(t_j)\, \gamma_j^n\, n^{-1/2} \{G_n(t_{j+1}+\gamma_j^n) - G(t_j)\}$$

using assumptions a), b), and the fact that $\gamma_j^n \ge 0$. Now,

$$\lim_n J(t_j)\, \gamma_j^n\, n^{-1/2} = x(t_j)\, J(t_j) \text{ and}$$

$$\lim_n G_n\,(t_{j+1} + \gamma_j^n) - G_n(t_j) = G(t_{j+1}) - G(t_j)$$

using the continuity assumptions on the t_j's. Bounding similarly $n^{-1/2} E_{1j}^n$ from below, one obtains: there exists $n_2 = n_2(\varepsilon, m, t_1, t_2, \ldots, t_m)$ $n_2 > n_1$ in N such that:

$$n > n_2 \rightarrow |n^{-1/2} E_{1j}^n - J(t_j)\,(G(t_{j+1}) - G(t_j))\, x(t_j)\,| < \frac{\varepsilon}{m}$$

for all $j \in \{1,2,\ldots,m\}$.

Consider E_{2j}^n, since

$$|t_{j+1} - t_j| < \alpha_0, |x(t_j) - x\,(\tfrac{i}{n+1})| < q_{\alpha_0}\,(x(\cdot)) < \varepsilon$$

by the choice of α_0. G_n being increasing:

$$G_n(\tfrac{i}{n+1} + n^{-1/2}\, x\,(\tfrac{i}{n+1})) \le G_n\,(\tfrac{i}{n+1} + n^{-1/2}\,(x(t_j) + \varepsilon))$$

$$\le G_n\,\left(\frac{i+[(n+1)\; n^{-1/2}(x(t_j)+\varepsilon)+1]}{n+1}\right).$$

Let $\xi_j^n = \frac{1}{n+1}\,[(n+1)n^{-1/2}(x(t_j) + \varepsilon) + 1]$ and note:

$$\xi_j^n \ge \gamma_j^n \quad .$$

Using once more assumptions a) and b),

$$E_{2j}^{n} \leq J(t_j) \quad \{\sum_i (G_n(\frac{i}{n+1} + \xi_j^n) - G_n(\frac{i}{n+1} + \gamma_j^n))\}$$

$$= \quad J(t_j) \quad \sum_{t_{j+1} + \gamma_{jn} \leq \frac{i}{n+1} < t_{j+1} + \xi_j^n} G_n\left(\frac{i}{n+1}\right) \quad - \quad \sum_{t_j + \gamma_{jn} \leq \frac{i}{n+1} < t_j + \xi_j^n} G_n\left(\frac{i}{n+1}\right) \; .$$

Since $(n+1)\ (\xi_j^n - \gamma_j^n)$ behaves asymptotically as $n^{1/2}\varepsilon$ and using the continuity properties of the t_j's , $n^{-1/2}\ E_{2j}^n$ is asymptotically bounded by

$$\varepsilon\ J(t_j)\ \{G(t_{j+1}) - G(t_j)\} \;.$$

The same way one can find a lower asymptotic bound for $n^{-1/2}E_{2j}^n$ and:

$$\lim_n \sup |n^{-1/2}\ E_{2j}^n| \leq \varepsilon\ J(t_j)\ (G(t_{j+1}) - G(t_j)) \;.$$

Consider E_{3j}^n , using a) and b):

$$E_{3j}^n \leq (J(t_{j+1}) - J(t_j))\ \{\sum_i\ (G_n(\frac{i}{n+1} + \beta_n) - G_n(\frac{i}{n+1}))\}$$

where $\beta_n = \frac{1}{n+1}\ [(n+1)\ n^{-1/2}||x||+1]$.

Hence:

$$E_{3j}^n \leq (J(t_{j+1}) - J(t_j)) \quad \sum_{t_{j+1} \leq \frac{i}{n+1} < t_{j+1} + \beta_n} G_n\left(\frac{i}{n+1}\right) - \sum_{t_j \leq \frac{i}{n+1} < t_j + \beta_n} G_n\left(\frac{i}{n+1}\right) \; .$$

Therefore $n^{-1/2}\ E_{3j}^n$ is asymptotically bounded by:

$$(J(t_{j+1}) - J(t_j))\ (G(t_{j+1}) - G(t_j))||x||$$

using an argument similar to the preceeding one. Bounding the same way E^n_{3j} from below,

$$\lim_n \sup |n^{-1/2} E^n_{3j}| \leq (J(t_{j+1}) - J(t_j))\ (G(t_{j+1}) - G(t_j))\ ||x|| \ .$$

Hence:

$$|h_n(x(\cdot)) - \sum_{j=0}^{m-1} J(t_j)\ (G(t_{j+1}) - G(t_j))\ x\ (t_j)|$$

is asymptotically bounded by $0(\varepsilon)$.

Consider:

$$|h(x(\cdot)) - \sum_{j=0}^{m-1} J(t_j)\ (G(t_{j+1}) - G(t_j))\ x\ (t_j)|$$

$$= |\sum_{j=0}^{m-1} \{\int_{t_j}^{t_{j+1}} (J(t)x(t) - J(t_j)\ x\ (t_j))\ d\ G(t)\}|$$

$$\leq \sum_{j=0}^{m-1} \{\int_{t_j}^{t_{j+1}} J(t)|x(t) - x(t_j)|d\ G(t)$$

$$+ |x(t_j)|\ \int_{t_j}^{t_{j+1}} (J(t) - J(t_j))d\ G(t)\}$$

$$\leq q_{\alpha_0}\ (x(\cdot))\ \int_{\delta_0}^{1-\delta_0} J(t)\ d\ G(t) + \frac{\varepsilon}{2}\ ||x||$$

$$= 0(\varepsilon)\ .$$

Hence for any $\varepsilon_0 > 0$, we can find $N_{\varepsilon_0} \in N$ such that:

$$n > N_{\varepsilon_0} \rightarrow \sup_S |h_n(x(\cdot)) - h(x(\cdot))| < \varepsilon_0\ .$$

The lemma is proved.

This is a generalization of the lemma used by Huber (1969) to prove theorem II.1 .

<u>Proof of theorem II.5</u> Without loss of generality, assume $\xi_0=0$. Using Lemma II.6, $n^{1/2}\lambda(\hat{T}_n)$ will have the same asymptotic distribution as

$$(2.5) \qquad n^{-1/2} \sum_{i=1}^{n} (J(\tfrac{i}{n+1})\, \psi(X_{(i)} - k_n) - \lambda(k_n))$$

where $n^{1/2}\lambda(k_n) = g$, g fixed in R provided the asymptotic distribution of (2.5) is continuous.

Using lemma II.5 under assumptions iii) and v), lemma II.1 and the definition of the processes $\{Z^{(n)}(t)\}$, one obtains that (2.5) has the same asymptotic distribution as:

$$n^{-1/2} \sum_{i=1}^{n} J(\tfrac{i}{n+1}) \left(\psi\left(F^{-1}(\tfrac{i}{n+1}) + n^{-1/2} Z^{(n)}(\tfrac{i}{n+1})\right) - k_n\right) - \psi\left(F^{-1}(\tfrac{i}{n+1}) - k_n\right))$$

As shown by Rivest (1976) p. 26-27, given $\varepsilon > 0$ there exist α_ε , M_ε , $n_1=n_1(\varepsilon)$ such that, if

$$S_\varepsilon = \{x(\cdot) \in C[0,1] : ||x|| < M_\varepsilon\ q_\alpha(x(\cdot)) < \varepsilon \text{ for all } \alpha \leq \alpha_\varepsilon\} ,$$

$$P(Z^{(n)}(t) \in S_\varepsilon) \geq 1-\varepsilon \text{ for all } n > n_1 .$$

Now, using lemma II.7,

$$(h_n(Z^{(n)}(\cdot)) - h(Z^{(n)}(\cdot)) \text{ is } o_p(1) .$$

Hence $h_n(Z^{(n)}(\cdot))$ and $h(Z^{(n)}(\cdot))$ have the same asymptotic distribution and using theorem II.3

$$L(h(Z^{(n)}(\cdot))) \xrightarrow[n\to\infty]{} L(h(Z(\cdot))) .$$

Therefore the asymptotic distribution of $h_n(Z^{(n)}(\cdot))$ is the one of:

$$\int_0^1 J(t)\ Z(t)\ d\,\psi\ (F^{-1}(t))$$

which, according to lemma II.2, is:

$$N(0,\sigma_1^2)\ .$$

Hence the asymptotic distribution of (2.5) is continuous and the first part of the theorem is proved.

Since λ is differentiable and $\hat{T}_n$ is consistent,

$$\lambda\ (\hat{T}_n) = \lambda(0) + \hat{T}_n\ \lambda'(0) + o\ (\hat{T}_n)$$

hence:

$$\frac{n^{1/2}\lambda(\hat{T}_n)}{n^{1/2}\lambda'(0)\hat{T}_n} \longrightarrow 1 \text{ in probability as } n \to \infty\ .$$

Since $\dfrac{n^{1/2}\lambda(\hat{T}_n)}{\lambda'(0)}$ converges weakly to a $N(0,\sigma^2)$ the last statement implies:

$$L(n^{1/2}\hat{T}_n) \xrightarrow[n \to \infty]{} N(0,\sigma^2)\ .$$

Note that the result of this theorem is still valid if the assumption $J(t) = 0\ t \notin [\delta,1-\delta]$ is replaced by $\psi(x)$ is a bounded function of x or $F^{-1}(t)$ is a bounded function of t .

<u>Corollary II.1</u>

Let $\{\hat{T}_n^*\}$ be a sequence of statistics satisfying:

for all $\varepsilon > 0$ there exist $M_0 = M_0(\varepsilon)$ and $n_0 = n_0\ (\varepsilon, M_0)$ in N such that:

$$n > n_0 \ \to P\ \{n^{1/2}|\hat{T}_n^* - \theta_0| > M_0\} \leq \varepsilon\ .$$

Then assuming:

i) $\psi(F^{-1}(t)-x)$ is continuous at θ_0 a.e.dt,

ii) $\psi(F^{-1}(t)-\theta_0)$ and $J(t)$ are not discontinuous together,

iii) λ is strictly decreasing near θ_0 ,

$$L\left(n^{1/2}\left(\sum_{i=1}^{n} J\left(\frac{i}{n+1}\right) \psi\left(X_{(i)} - \hat{T}_n^*\right) - \lambda\left(\hat{T}_n^*\right)\right)\right) \xrightarrow[n \to \infty]{}$$

$$L\left(\int_0^1 J(t)\, Z(t)\, d\, \psi\left(F^{-1}(t) - \theta_0\right)\right) .$$

Note that θ_0 is not assumed to satisfy:

$$\lambda(\theta_0) = 0 .$$

Proof: Without loss of generality assume $\theta_0 = 0$.

Consider for a fixed $\varepsilon > 0$,

$$P\{n^{1/2} |\lambda(\hat{T}_n^*) - \lambda_n(\hat{T}_n^*)| > \varepsilon\} ,$$

for $n > n_0$ this is less or equal to:

$$\varepsilon + P\{n^{1/2} |\lambda(\hat{T}_n^*) - \lambda_n(\hat{T}_n^*)| > \varepsilon \text{ and } |T_n^*| < n^{-1/2} M_0\}.$$

Using lemma II.5, if $|T_n^*| < n^{-1/2} M_0$ there exists $n_1 = n_1(\varepsilon, M_1)$, $n_1 > n_0$ in N such that

$$n > n_1 \rightarrow n^{1/2} |\lambda(\hat{T}_n^*) - \lambda_n(\hat{T}_n^*)| < \varepsilon ,$$

so that for $n > n_1$

$$P\{n^{1/2} |\lambda(\hat{T}_n^*) - \lambda_n(\hat{T}_n^*)| > \varepsilon\} \leq \varepsilon \text{ and}$$

$$n^{1/2} |\lambda(\hat{T}_n^*) - \lambda_n(\hat{T}_n^*)| \text{ is } o_p(1) .$$

Define:

$$h_n(x(\cdot)) = n^{-1/2} \sum_{i=1}^{n} J(\frac{i}{n+1}) \{ \psi (F^{-1}(\frac{i}{n+1} + n^{-1/2}x(\frac{i}{n+1})) -$$

$$k_n(y_n(\cdot))) - \psi (F^{-1}(\frac{i}{n+1}) - k_n(y_n(\cdot)))\} ,$$

$$h(x(\cdot)) = \int_0^1 J(t)\, x(t)\, d\, \psi (F^{-1}(t))$$

where $y_n(\cdot) \in C_n \subset C\,[0,1]$ and k_n is a functional defined on $C\,[0,1]$. Note that the conclusion of lemma II.7 is true with k_n as a functional provided $k_n(y_n(\cdot))$ converges uniformly to 0 as n goes to ∞ .

To end the proof it suffices to show that $h_n(Z^{(n)}(\cdot))$ and $h(Z^{(n)}(\cdot))$ have the same behaviour as $n \to \infty$ (with $k_n(y_n(\cdot)) = \hat{T}_n^*$).

We want to find $N_0 = N_0(\varepsilon) \in N$ such that

$$P\{|h_n(Z^{(n)}(\cdot)) - h(Z^{(n)}(\cdot))| > \varepsilon\} = 0(\varepsilon) \text{ if } n > N_0 .$$

As in theorem II.5, we can find $M_2=M_2(\varepsilon)$, $\alpha_0=\alpha_0(\varepsilon)$ $n_2=n_2(\varepsilon,M_2)$, $n_2>n_1$ such that the set

$$S_\varepsilon = \{x(\cdot) \in C\,[0,1] : ||x|| < M_2,\ q_\alpha(x(\cdot)) < \varepsilon \text{ for all } \alpha < \alpha_0\}$$

satisfies:

$$m > n_2 \to P\{Z^{(m)}(\cdot) \in S_\varepsilon\} \geq 1 - \varepsilon .$$

For a fixed $m > n_2$ and $n > n_1$,

$$P\{|h(Z^{(m)}(\cdot)) - h_n(Z^{(m)}(\cdot))| > \varepsilon\}$$

$$\leq 2\varepsilon + P\{|h(Z^{(m)}(\cdot)) - h_n(Z^{(m)}(\cdot))| > \varepsilon \text{ and } Z^{(m)}(\cdot) \in S_\varepsilon \text{ and }$$

$$|\hat{T}_n^*| < n^{-1/2}M_0\}$$

For $Z^{(m)}(\cdot) \in S_\varepsilon$ and $|\hat{T}^*_n| < n^{-1/2} M_0$, note that $k_n(Z^{(n)}(\cdot)) = \hat{T}^*_n$ converges uniformly to 0 as $n \to \infty$ hence using lemma II.6 there exists $n_3 = n_3(\varepsilon, S_\varepsilon, M_1)$ in N such that $n > n_3$ implies $\sup_{S_\varepsilon} |h(Z^{(m)}(\cdot)) - h_n(Z^{(m)}(\cdot))| < \varepsilon$.

Therefore for any fixed $m > n_2$ and for $n > n_3$

$$P\{|h(Z^{(m)}(\cdot)) - h_n(Z^{(m)}(\cdot))| > \varepsilon\} \leq 2\varepsilon ,$$

this implies for $n > n_3$:

$$P\{|h(Z^{(n)}(\cdot)) - h_n(Z^{(n)}(\cdot))| > \varepsilon\} \leq 2\varepsilon .$$

And the corollary is proved.

Section II.5 Asymptotic behaviour of L-M estimators .

Suppose $\lambda(\xi_0) = 0$ and define:

$$\lambda_H(x) = \int_0^1 \psi_H(F^{-1}(t) - x)\, dt$$

where:

$$\psi_H(x) = \int_0^x J(F(y+\xi_0))\, d\,\psi(y)$$

$$- \int_0^1 \int_0^{F^{-1}(t)-\xi_0} J(F(y+\xi_0))\, d\,\psi(y)\, dt$$

$$\int_a^b d\psi = \int_{[a,b)} d\psi$$

$$\int_b^a d\psi = - \int_{[a,b)} d\psi \quad \text{if } b > a .$$

Lemma II.8

If $\int_0^1 \psi^2(F^{-1}(t)-x)\, dt$ is finite in $V_0 = \{\xi : |\xi - \xi_0| \leq \varepsilon\}$ where

$\varepsilon > 0$, $\psi_H(x)$ is a well defined increasing function satisfying:

i) $\lambda_H(\xi_0) = 0$

ii) $\int_0^1 \psi_H^2(F^{-1}(t) - x)\ dt$, λ_H and λ are continuous at ξ_0 .

Proof: Let $M = \sup_{t\varepsilon[0,1]} |J(t)|$, note that M is finite since J is a bounded variation function.

Consider:

$$\int_0^1 \int_0^{F^{-1}(t)-\xi_0} J(F(y+\xi_0))\ d\psi(y)\ dt$$

(2.6)

$$\leq \int_0^1 M\{|\psi(F^{-1}(t)-\xi_0)| + |\psi(0)|\}\ dt .$$

Since $\int_0^1 \psi^2(F^{-1}(t) - \xi_0 \pm \varepsilon)\ dt < \infty$, $\int_0^1 |\psi(F^{-1}(t) - \xi_0 \pm \varepsilon)|\, dt < \infty$ so that:

$$\int_0^1 \int_0^{F^{-1}(t)-\xi_0} J(F(y+\xi_0))\ d\psi(y) dt < \infty$$

using (2.6). Hence ψ_H is a well defined increasing function and:

$$\lambda_H(\xi_0) = 0 .$$

The fact that $\int_0^1 |\psi(F^{-1}(t)-\xi_0 \pm \varepsilon)| dt < \infty$ implies $\int_0^1 |\psi_H(F^{-1}(t)-\xi_0 \pm \varepsilon)| dt < \infty$ and, using lemma II.4, λ and λ_H are continuous at ξ_0 .

(2.6) and the assumption on ψ imply:

$$\int_0^1 \psi_H^2(F^{-1}(t) - \xi_0 \pm \varepsilon)\ dt < \infty .$$

We want to prove:

$$\lim_{\xi\to\xi_0} \int_0^1 \psi_H^2(F^{-1}(t) - \xi)dt = \int_0^1 \psi_H^2(F^{-1}(t) - \xi_0)dt \ .$$

Note that:

$$\psi_H^2(F^{-1}(t) - \xi) < \psi_H^2(F^{-1}(t) - \xi_0 - \varepsilon) + \psi_H^2(F^{-1}(t) - \xi_0 + \varepsilon) \text{ for } \xi \text{ in } V_0 \ .$$

The RHS is an integrable function, hence, using domininated convergence, one obtains the continuity of:

$$\int_0^1 \psi_H^2(F^{-1}(t) - \xi) \, dt \text{ at } \xi_0 \ .$$

<u>Theorem II.6</u>

If i) ψ is left continuous and $J(t)$ and $\psi(F^{-1}(t)-\xi_0)$ are not discontinuous together,

ii) $\psi(F^{-1}(t)-x)$ is continuous at ξ_0 a.e. dt ,

iii) $\int_0^1 \psi^2(F^{-1}(t)-x) \, dt < \infty$ in a neighborhood of ξ_0 ,

iv) λ is monotone at ξ_0 ,

v) $\lambda'(\xi) = \frac{d\lambda(\xi)}{d\xi}$ exists and is negative at ξ_0 ,

vi) There exist $\varepsilon > 0$, M_0 in N such that $|J(t) - J(s)|<M_0|t-s|$ for s, t in $[0,\varepsilon]$ or s,t in $[1-\varepsilon,1]$,

vii) There exist M_1, M_2 in N such that F is differentiable in $\{x\varepsilon R: |x|>M_1\}$ and $f(x) = \frac{dF(x)}{dx} < M_2$ if $|x| > M_1$.

Then the estimator $\hat{T}_n$, a solution of

$$\sum_{i=1}^{n} J(\tfrac{i}{n+1})\ \psi\ (X_{(i)} - \theta) = 0$$

is such that:

$$\mathcal{L}(n^{1/2}(\hat{T}_n - \xi_0)) \xrightarrow[n \to \infty]{} N(0,\sigma^2)$$

where $\sigma^2 = \sigma_1^2 \dfrac{1}{[\lambda'(\xi_0)]^2}$

and $\sigma_1^2 = E(\psi_H^2(X_1 - \xi_0))$.

Note that ψ is left continuous implies:

$$\int_a^b d\psi = \psi(b) - \psi(a) .$$

Assumptions (vi) and (vii) may be replaced by $J(t) = 0, t \notin [\delta, 1-\delta]$ for a certain $\delta \varepsilon (0, {}^1/_2)$.

<u>Lemma II.9</u>

Under the assumptions of theorem II.6, for any $\eta>0$

$$\lim_n P\{n^{-1/2} \left| \sum_{i=1}^{n} J(\tfrac{i}{n+1})\ \psi\ (X_{(i)} - k_n(g)) - \lambda(k_n(g)) - \psi_H(X_i - \xi_0) \right| > \eta\} = 0$$

where $\{k_n(g)\}_{n=1}^{\infty}$ is an R sequence satisfying

$$n^{1/2}\ \lambda(k_n(g)) = g \quad \text{for } g \ \varepsilon\ R\ ,\quad n\ \varepsilon\ N\ .$$

<u>Proof</u>: Without loss of generality, assume $\xi_0=0$ and $\int_0^1 \psi^2(F^{-1}(t)\pm\varepsilon)dt < \infty$. Let g be a fixed real number and $k_n = k_n(g)$. The fact that λ is decreasing implies:

$$\lim_n k_n = 0 .$$

We will first prove:

$$n^{-1/2} \{ \sum_{i=1}^{n} (J(\tfrac{i}{n+1}) \psi (X_{(i)} - k_n) - \lambda(k_n)) -$$

$$\sum_{i=1}^{n} \psi_H(X_{(i)} - k_n) - \psi_H(F^{-1}(\tfrac{i}{n+1}) - k_n)\}$$

is $o_p(1)$.

Using lemma II.5 under assumptions (iv) and (vi) and lemma II.1, it suffices to prove:

$$n^{-1/2} \{\sum_{i=1}^{n} J(\tfrac{i}{n+1}) (\psi(F^{-1}(U_{(i)}) - k_n) - \psi(F^{-1}(\tfrac{i}{n+1}) - k_n))$$

(2.7)

$$- \sum_{i=1}^{n} (\psi_H(F^{-1}(U_{(i)}) - k_n) - \psi_H(F^{-1}(\tfrac{i}{n+1}) - k_n))\}$$

is $o_p(1)$.

Using theorem II.4, given $\eta > 0$, there exist $M_3 = M_3(\eta)$ and $n_0 = n_0(M_3 , \eta)$ in N such that:

(2.8) $\max_{i\varepsilon\{1,2,\ldots n\}} |\frac{i}{n+1} - U_{(i)}| < M_3 n^{-1/2}$ for $n > n_0$ except on a set of probability at most $\eta/3$.

Consider:

$$\lambda'(0) = \lim_n \lambda(k_n)/k_n$$

$$= \lim_n \frac{n^{-1/2} g}{k_n} ,$$

hence for any $M_4 > g/\lambda'(0)$ there exists $n_1 = n_1(M_4)$, $n_1 > n_0$, such that

$$\text{(2.9)} \qquad n > n_1 \to |n^{1/2} k_n| < M_4 .$$

Let $\psi_1(x) = \psi(x) - \psi(F^{-1}(\varepsilon) + \varepsilon)$, note that replacing ψ by ψ_1 in (2.7) does not change the value of the expression and:

$$\int_0^1 \psi^2(F^{-1}(t) - \xi)\, dt < \infty \text{ in a neighborhood of } 0$$

implies:

$$\int_0^1 \psi_1^2 (F^{-1}(t) - \xi)\, dt < \infty \text{ in this same neighborhood,}$$

hence we may suppose: $\psi(x)$ is negative increasing in

$$(- \infty , F^{-1}(\varepsilon) + \varepsilon] .$$

We want to prove:

$$n^{-1/2} \sum_{i=1}^{n} \{J(\tfrac{i}{n+1}) (\psi (F^{-1}(U_{(i)}) - k_n) - \psi(F^{-1}(\tfrac{i}{n+1}) - k_n))$$

$$- (\psi_H(F^{-1}(U_{(i)}) - k_n) - \psi_H(F^{-1}(\tfrac{i}{n+1}) - k_n))\}$$

$$= n^{-1/2} \sum_{i=1}^{n} \{\int_{F^{-1}(\frac{i}{n+1}) - k_n}^{F^{-1}(U_{(i)}) - k_n} (J(\tfrac{i}{n+1}) - J(F(x)))\, d\psi(x)\}$$

is $o_p(1)$.

The strategy of the proof is the following: given $\eta > 0$, we first find

$\delta_j = \delta_j(\eta)$, j=1,2, such that:

$n^{-1/2}\left|\sum_{i=1}^{[n\delta_1]}(\ldots)\right| < \frac{\eta}{3}$ except on a set of probability at most $\frac{\eta}{3}$ for

$n > N_1$,

$n^{-1/2}\left|\sum_{n-[n\delta_2]+1}^{n}(\ldots)\right| < \frac{\eta}{3}$ except on a set of probability at most $\eta/3$

for $n > N_1$.

Once δ_1 and δ_2 are given using an argument similar to the one used in section II.4:

$$L\{n^{-1/2}\sum_{i=[n\delta_1]+1}^{n-[n\delta_2]}(\ldots)\} \xrightarrow[n \to \infty]{} L\{1_{[0,\infty)}\}$$

so that there exists $N_2 > N_1$ such that:

$n > N_2$ implies $n^{-1/2}\left|\sum_{i=[n\delta_1]+1}^{n-[n\delta_2]}(\ldots)\right| < \frac{\eta}{3}$ except on a set of probability at most $\frac{\eta}{3}$.

Choose $\delta_1 = \delta_1(\eta)$ in the following way:

if $\lim_{t\to 0} F^{-1}(t) > -\infty$ take $\delta_1 = 0$

if $\lim_{t\to 0} F^{-1}(t) = -\infty$,

let $M_5 = M_0\{2M_4\ M_2 + M_3\}$ and take $\delta_1 \in (0, 1/2)$ satisfying:

$$\cdot\ \delta_1 < \max\ \{\frac{\varepsilon}{2}\ ,\ \frac{F(-2M_1)}{2}\ ,\ M_1\}\ .$$

note that $\lim_{t\to 0} F^{-1}(t) = -\infty$ implies that $\delta_1 \in (0, 1/2)$ satisfying this condition exists,

- $-\int_0^{\delta_1} \psi(F^{-1}(t) - \varepsilon)\, dt < \frac{\eta}{12M_5}$,

- $F(F^{-1}(2\delta_1) + 2\delta_1) < \varepsilon$

- δ_1 is a continuity point of $\psi(F^{-1}(t))$.

If $\delta_1 = 0$, there is nothing to prove; if $\delta_1 \in (0, {}^1/2)$ we prove that there exists $N_1 \in N$ such that:

$$\left| n^{-1/2} \sum_{i=1}^{[n\delta_1]} (\ldots) \right| < \eta/3 \quad \text{for } N > N_1 \text{ except on a set of}$$

probability $\eta/3$.

Using (2.8) and (2.9), there exists $n_2 = n_2(\varepsilon, \delta_1, \eta, g)$, $n_2 > n_1$ such that

for $n > n_2$ and $i \in \{1,2,\ldots,[\frac{3}{2}\delta_1 n]\}$

$$F^{-1}\left(\frac{i}{n+1}\right) < -2M_1$$

$$\frac{i}{n+1} < 2\delta_1$$

$$|k_n| < \delta_1$$

and
$$U_{(i)} < 2\delta_1$$

$$F^{-1}(U_{(i)}) < -2M_1$$

except on a set of probability at most $\eta/3$. So that $\max\{F(F^{-1}(U_{(i)}) - k_n) , F(F^{-1}(\frac{i}{n+1}) - k_n)\} < \varepsilon$ and using assumption (vi)

$$(2.10) \qquad \sup_{x \,\varepsilon\, [F^{-1}(U_{(i)}) - k_n \,,\, F^{-1}(\frac{i}{n+1}) - k_n]} \left|J(F(x)) - J(\tfrac{i}{n+1})\right|$$

$$\leq M_0 \sup_{x \,\varepsilon\, [F^{-1}(U_{(i)}) - k_n \,,\, F^{-1}(\frac{i}{n+1}) - k_n]} |F(x) - i/(n+1)|$$

$$\leq M_0 \left\{\left|\tfrac{i}{n+1} - F(F^{-1}(U_{(i)}) - k_n)\right| + \left|\tfrac{i}{n+1} - F(F^{-1}(\tfrac{i}{n+1}) - k_n)\right|\right\} .$$

Now, $\max \{F^{-1}(U_{(i)}) - k_n \,,\, F^{-1}(U_{(i)}) \,,\, F^{-1}(\frac{i}{n+1}) - k_n \,,\, F^{-1}(\frac{i}{n+1})\}$

$$< - 2 M_1 + \delta < - M_1 ,$$

so, using the fact that F is differentiable in

$$\{|x| > M_1\} \text{ (assumption vii)} ,$$

$$F(F^{-1}(U_{(i)}) - k_n) - U_i = k_n \; f\,(\theta_{in}) \text{ and}$$

$$F(F^{-1}(\tfrac{i}{n+1}) - k_n) - \tfrac{i}{n+1} = k_n \; f\,(\omega_{in})$$

where θ_{in} and $\omega_{in} < - M_1$ so that:

$$f(\theta_{in}) \text{ and } f(\omega_{in}) < M_2 .$$

Therefore $n^{-1/2}$ (2.10) is less or equal than:

$$n^{-1/2} M_0 \left\{\left|\tfrac{i}{n+1} - U_{(i)}\right| + 2 \,|k_n|\, M_2\right\} .$$

$$\leq \frac{M_0}{n} \{M_3 + 2 \,|k_n| n^{1/2} M_2\}$$

$$\leq \frac{M_0}{n} \{M_3 + 2 M_4 M_2\}$$

for $n > n_2$.

Therefore $n^{-1/2}$ (2.10) is less than $\frac{M_5}{n}$ for $n > n_2$ except on a set of probability at most $\eta/3$.

Hence

$$n^{-1/2} \left| \sum_{i=1}^{[n\delta_1]} (\ldots) \right|$$

$$\leq \frac{M_5}{n} \left| \sum_{i=1}^{[n\delta_1]} \psi(F^{-1}(U_{(i)}) - k_n) - \psi(F^{-1}(\tfrac{i}{n+1}) - k_n) \right|$$

$$\leq -\frac{M_5}{n} \sum_{i=1}^{[n\delta_1]} \psi(F^{-1}(U_{(i)}) - \varepsilon) + \psi(F^{-1}(\tfrac{i}{n+1}) - \varepsilon) \text{ for } n > n_2 ,$$

using the assumption on ψ and the way δ_1 was chosen.

Using lemma II.5 under assumptions (iv) and (v): there exists $n_3 > n_2$ such that

$$n > n_3 \rightarrow \left| n^{-1} \sum_{i=1}^{[n\delta_1]} \psi(F^{-1}(\tfrac{i}{n+1}) - \varepsilon) - \int_0^{\delta_1} \psi(F^{-1}(t) - \varepsilon)\, dt \right| < \frac{\eta}{12M_5}$$

or

$$-n^{-1} \sum_{i=1}^{[n\delta_1]} \psi(F^{-1}(\tfrac{i}{n+1}) - \varepsilon) < \frac{\eta}{6M_5} .$$

To end the proof of this first part, it suffices to show:

$$\{n^{-1} \sum_{i=1}^{[n\delta_1]} \psi(F^{-1}(U_{(i)}) - \varepsilon) - \int_0^{\delta_1} \psi(F^{-1}(t) - \varepsilon)\, dt\}$$

is $o_p(1)$.

Let $\psi^*(x) = \begin{cases} \psi(x - \varepsilon), & x \le F^{-1}(\delta_1) \\ 0 & \text{elsewhere} \end{cases}$

Since $\int_0^1 (\psi^*(F^{-1}(t)))^2 \, dt < \infty$ the weak law of large numbers yields:

$$(n^{-1} \sum_{i=1}^{n} \psi^* (X_i) - \int_0^{\delta_1} \psi(F^{-1}(t) - \varepsilon) \, dt) \text{ is } o_p(1) .$$

It is sufficient to show:

$$n^{-1}\left| \sum_{i=1}^{n} \psi^*(X_i) - \sum_{i=1}^{[n\delta_1]} \psi(F^{-1}(U_{(i)}) - \varepsilon)\right| \text{ is } o_p(1) . \tag{2.11}$$

Choose $\varepsilon_1 \in (0, \delta_1/2)$ such that

$$\frac{3}{2} \, \varepsilon_1 |\psi(F^{-1}(\delta_1 - \varepsilon_1) - \varepsilon)| < \eta/24M_5 \, ,$$

using (2.8) there exists $n_4 = n_4 (M_3 , \eta)$, $n_4 > n_3$ such that for each $n > n_4$ and i satisfying:

$$\frac{i}{n+1} \in (0, \delta_1 - \varepsilon_1/2), \quad U_{(i)} < \delta_1$$

$$\frac{i}{n+1} \in (\delta_1 + \varepsilon_1/2, 1), \quad U_{(i)} > \delta_1 \text{ and}$$

$$U_{([n(\delta_1 - \varepsilon_1/2)] + 1)} > \delta_1 - \varepsilon_1$$

except on this same set of probability at most $\eta/3$.

For $n > n_4$, (2.11) reduces to:

$$n^{-1}\left| \sum_{i=[n(\delta_1-\varepsilon_1/2)] + 1}^{[n\delta_1]} \psi^*(F^{-1}(U_{(i)})) - \psi(F^{-1}(U_{(i)}) - \varepsilon)\right.$$

$$- \sum_{i=[n\delta_1]+1}^{[n(\delta_1+\varepsilon_1/2)]} \psi^* (F^{-1}(U_{(i)}))| \, .$$

Now since ψ is negative in $(-\infty \, , F^{-1}(\varepsilon)+\varepsilon)$ the last expression is bounded by:

$$\frac{3}{2}\,\varepsilon_1|\psi(F^{-1}(U_{([n(\delta_1-\varepsilon_1/2)]+1)}) - \varepsilon)|$$

$$\leq \frac{3}{2}\,\varepsilon_1|\psi(F^{-1}(\delta_1 - \varepsilon_1) - \varepsilon)| \text{ for } n > n_4$$

$$\leq \eta/24M5$$

so, there exists $n_5 > n_4$ such that

$$n > n_5 \rightarrow n^{-1}|\sum_{i=1}^{[n\delta_1]} \psi(F^{-1}(U_{(i)}) - \varepsilon)| < \frac{\eta}{12M_5}$$

and the first part of the proof is completed.

Using a symmetric argument, $\delta_2 = \delta_2(\eta)$ a continuity point of $\psi(F^{-1}(t))$, satisfying:

$n^{-1/2}|\sum_{i=n-[n\delta_2]+1}^{n}(\ldots)| < \eta/3$ except on a set of probability at most $\eta/3$ is easily found.

To prove (2.7), it suffices to show:

$$n^{-1/2}|\sum_{i=[n\delta_1]+1}^{n-[n\delta_2]}(\ldots)| \text{ is } o_p(1) \text{ or}$$

$$n^{-1/2}|\sum_{i=1}^{n} J_1\,(\tfrac{i}{n+1})\;(\psi_H(F^{-1}(U_{(i)}) - k_n) - \psi_H(F^{-1}(\tfrac{i}{n+1}) - k_n))$$

$$- J_2 \left(\frac{i}{n+1}\right) \left(\psi(F^{-1}(U_{(i)}) - k_n) - \psi(F^{-1}(\tfrac{i}{n+1}) - k_n)\right) \Big| \text{ is } o_p(1) ,$$

where $J_1(t) = \begin{cases} 1 , & t \in [\delta_1 , 1-\delta_2] \\ 0 , & \text{elsewhere} \end{cases}$

$$J_2(t) = \begin{cases} J(t) & t \in [\delta_1 , 1-\delta_2] \\ 0 & \text{elsewhere} . \end{cases}$$

Note that J_1 and J_2 fulfill requirements of section II.4 so that

$$L \left(n^{-1/2} \sum_{i=[n\delta_1]+1}^{n-[n\delta_2]} (\ldots) \right) \xrightarrow[n \to \infty]{} L \left(\int_0^1 J_1(t) \; Z(t) d\psi_H(F^{-1}(t)) - \int_0^1 J_2(t) \; Z(t) d\,\psi(F^{-1}(t)) \right) .$$

For any $x(\cdot) \in C[0,1]$ (the set of continuous functions on $[0,1]$), one easily shows:

$$\int_0^1 J_1(t) \; x(t) d\psi_H(F^{-1}(t)) = \int_0^1 J_2(t) \; x(t) d\psi(F^{-1}(t)) .$$

Since $P\{Z(\cdot) \in C[0,1]\} = 1$, $L \left(n^{-1/2} \sum_{i=[n\delta_1]+1}^{n-[n\delta_2]} (\ldots)\right) \xrightarrow[n \to \infty]{} L \{ 1_{[0,\infty]} \}$.

Therefore there exists n_5 , $n_5 > n_4$ such that $n^{1/2} \left| \sum_{i=[n\delta_1]+1}^{n-[n\delta_2]} (\ldots) \right| < \eta/3$ except on a set of probability at most $\eta/3$ and this ends the first part of the lemma.

In order to prove the result we now need to show:

$$n^{-1/2} \sum_{i=1}^{n} (\psi_H(X_i - k_n) - \psi_H (F^{-1}(\tfrac{i}{n+1}) - k_n) - \psi_H(X_i)) \text{ is } o_p(1) .$$

Lemma II.8 insures us that ψ_H fulfills assumptions (iv) and (vi) of lemma II.5 hence:

$$n^{-1/2}\sum_{i=1}^{n}(\psi_H(X_i - k_n) - \lambda_H(k_n)) \text{ and}$$

$$n^{-1/2}\sum_{i=1}^{n}(\psi_H(X_i - k_n) - \psi_H(F^{-1}(\tfrac{i}{n+1}) - k_n))$$

will reach the same limit as $n \to \infty$. It suffices to show:

$$\lim_n P\{n^{-1/2}|\sum_{i=1}^{n}\psi_H(X_i - k_n) - \lambda_H(k_n) - \psi_H(X_i)| > \eta\} = 0$$

for all $\eta > 0$ in order to end the proof of this lemma. That last convergence is an easy consequence of the following condition:

$$\lim_n E(n^{-1/2}\{\sum_{i=1}^{n}\psi_H(X_i - k_n) - \lambda_H(k_n) - \psi_H(X_i)\})^2 = 0 \ .$$

Let $V(x)$ be the variance of x .

Since:

$$E(\psi_H(X_i - k_n)) = \lambda_H(k_n) \quad i \in \{1,2,\ldots,n\}$$

and $\quad E(\psi_H(X_i)) = 0$,

the sum under consideration is a sum of i.i.d. random variables with null expectation. Therefore $E[(\ldots)^2] = V(\ldots)$ and it suffices to show:

$$\lim_n V(\psi_H(X_i - k_n) - \psi_H(X_i) - \lambda_H(k_n)) = 0 \ .$$

The last variance is equal to

$$E(\psi_H(X_1 - k_n) - \psi_H(X_1) - \lambda_H(k_n))^2$$

$$= E(\psi_H(X_1 - k_n) - \psi_H(X_1))^2 + \lambda_H^2(k_n) - 2\lambda_H^2(k_n) \ .$$

Since $\lim_n k_n = 0$ for n big enough, $|k_n| < \varepsilon$ hence

$$\psi_H(x) - \psi_H(x+\varepsilon) \leq \psi_H(x) - \psi_H(x-k_n) \leq \psi_H(x) - \psi_H(x-\varepsilon)$$

and:

$$(\psi_H(x) - \psi_H(x-k_n))^2 \leq (\psi_H(x) - \psi_H(x-\varepsilon))^2 + (\psi_H(x) - \psi_H(x+\varepsilon))^2$$

$$\leq 2(\psi_H^2(x) + \psi_H^2(x-\varepsilon) + \psi_H^2(x) + \psi_H^2(x+\varepsilon)) \quad .$$

The last function is integrable so that applying dominated convergence:

$$\lim_n E\,(\psi_H(X_1-k_n) - \psi_H(X_1))^2 = 0$$

and the lemma is proved.

<u>Proof of theorem II.6</u>:

Let $g \in R$ and $\{k_n\}_{n=1}^{\infty} = \{k_n(g)\}_{n=1}^{\infty}$ be an R sequence such that:

$$\lambda(k_n) = n^{-1/2} g \quad .$$

Now, using lemma II.6,

$$\lim_n P\{n^{-1/2} \sum_{i=1}^{n} (J(\tfrac{i}{n+1})\, \psi(X_{(i)} - k_n) - \lambda(k_n)) < -g\} =$$

$$\lim_n P\{n^{1/2}\, \lambda\,(\hat{T}_n) > g\}$$

provided the asymptotic distribution of the former is continuous. By lemma II.9, the desired limit is equal to:

$$\lim_n P\{n^{-1/2} \sum_{i=1}^{n} \psi_H(X_i-\xi_0) < -g\} \quad .$$

Assumption iii) and lemma II.8 imply

$$V(\psi_H(X_i - \xi_0)) < \infty ,$$

hence applying the Central Limit Theorem,

$$\mathcal{L}(n^{1/2} \lambda (\hat{T}_n)) \xrightarrow[n \to \infty]{} N(0, \sigma_1^2) .$$

Using an argument similar to the one at the end of theorem II.5:

$$\mathcal{L}(n^{1/2}(\hat{T}_n - \xi_0)) \xrightarrow[n \to \infty]{} N(0, \sigma^2) .$$

Lemma II.9 has stronger implications than asymptotic normality.

<u>Theorem II.7</u>

Under the assumptions of theorem II.6, the L-M estimator $\hat{T}_n$ based on ψ and J satisfies:

$$\text{for any} \quad \varepsilon > 0 : \lim_n P\{n^{1/2}|\hat{T}_n - \xi_0 + (\lambda'(\xi_0))^{-1}S_n| > \varepsilon\} = 0$$

where:

$$S_n = n^{-1} \sum_{i=1}^{n} \psi_H(X_i - \xi_0) \quad .$$

<u>Proof</u>: Assume without loss of generality $\xi_0 = 0$, take $\varepsilon > 0$ and define:

$$S_n(g) = n^{-1} \sum_{i=1}^{n} (J(\tfrac{i}{n+1}) \psi(X_{(i)}-k_n(g)) - \lambda_H(k_n(g)))$$

where $\{k_n(g)\}_{n=1}^{\infty}$ has been defined in lemma II.9 .

Since $n^{1/2} \lambda(\hat{T}_n)$ is asymptotically normal, there exist $M=M(\varepsilon)$ and $N_0 = N_0(\varepsilon,M)$ in N such that:

$$n > N_0 \text{ implies } P\{n^{1/2}|\lambda(\hat{T}_n)| \geq M\} \leq \varepsilon/2 .$$

So that:

$$P\{n^{1/2}|\lambda(\hat{T}_n) + S_n| > \varepsilon\}$$

$$\leq \varepsilon/2 + P\{n^{1/2}|\lambda(\hat{T}_n) + S_n| > \varepsilon \text{ and } |\lambda(\hat{T}_n)| < M\} .$$

Choose m in N such that $\varepsilon > 3/m$ and consider:

$$P\{n^{1/2}|\lambda(\hat{T}_n) + S_n| > \varepsilon \text{ and } n^{1/2}|\lambda(\hat{T}_n)| < M\}$$

$$= \sum_{j=-mM}^{mM} P\{\{j/m \leq n^{1/2}\lambda(\hat{T}_n) < (j+1)/m\} \cap$$

$$\{n^{1/2}S_n > \varepsilon - n^{1/2}\lambda(\hat{T}_n) \text{ or } n^{1/2}S_n < -\varepsilon - n^{1/2}\lambda(\hat{T}_n)\}\}$$

$$\leq \sum_{j=-mM}^{mM} P\{j/m \leq n^{1/2}\lambda(\hat{T}_n) < (j+1)/m \text{ and } n^{1/2}S_n < -\varepsilon - j/m\}$$

$$+ P\{j/m \leq n^{1/2}\lambda(\hat{T}_n) < (j+1)/m \text{ and } n^{1/2}S_n > \varepsilon - (j+1)/m\} .$$

Using lemma II.9 and lemma II.6, for each j in $\{-mM, -mM+1, \ldots, mM\}$, there exists:

$$n_j = n_j(m,\varepsilon) \text{ , } n_j > N_0 \text{ such that:}$$

(2.12) $\quad P\{n^{1/2}|S_n - S_n(j/m)| > \frac{1}{m}\} < \varepsilon/4mM$ and

(2.13) $\quad \{n^{1/2}S_n(j/m) < -j/m\} \subset \{n^{1/2}\lambda(\hat{T}_n) \geq j/m\} \subset$

$$\{n^{1/2}S_n(j/m) < -(j-1)/m\}$$

for $n > n_j$.

Take $n > n_j$ and consider:

$$P\,\{j/m \le n^{1/2}\,\lambda(\hat{T}_n) < (j+1)/m \text{ and } n^{1/2}\,S_n > \varepsilon - (j+1)/m\}\ ,$$

by (2.12), this is less or equal than:

$$\frac{\varepsilon}{4Mm} + P\{\{j/m \le n^{1/2}\,\lambda(\hat{T}_n)\} \cap \{n^{1/2}S_n(j/m) > \varepsilon - (j+2)/m\}\}\ ,$$

using (2.13) and the fact that $\varepsilon > 3/m$, the last expression is less or equal than:

$$P\{\{n^{1/2}S_n(j/m) < -\,(j-1)/m\} \cap \{n^{1/2}S_n(j/m) > -\,(j-1)/m\}\}$$

$$+\ \varepsilon/4Mm$$

$$=\ \varepsilon/4Mm\ .$$

Now, take $n > n_{j+1}$, and consider:

$$P\{j/m \le n^{1/2}\,\lambda(\hat{T}_n) < (j+1)/m \text{ and } n^{1/2}S_n < -\,\varepsilon - j/m\}$$

$$\le P\{\{n^{1/2}\,\lambda(\hat{T}_n) < (j+1)/m\} \cap \{n^{1/2}S_n((j+1)/m) < -\,\varepsilon - (j-1)/m\}\}$$

$$+\ \varepsilon/4Mm \quad \text{by (2.12)}$$

$$\le P\{\{n^{1/2}\,\lambda(\hat{T}_n) < (j+1)/m\} \cap \{n^{1/2}\,\lambda(\hat{T}_n) \ge (j+1)/m\}\} + \varepsilon/4Mm$$

$$=\ \varepsilon/4Mm\ .$$

Hence for $n > \max_{j\varepsilon\{-mM,\,-mM+1,\ldots,mM\}} n_j$

$$P\{n^{1/2}|\lambda(\hat{T}_n) + S_n| > \varepsilon\} \le \quad \varepsilon \text{ and}$$

$$n^{1/2}(\lambda(\hat{T}_n) + S_n) \text{ is } o_p(1) .$$

Using an argument similar to the one at the end of theorem II.5 ends the proof.

The next lemma provides an expression of $\lambda'(\xi_0)$ as a function of $\psi' = \frac{d\psi(x)}{dx}$, J and F^{-1} .

Lemma II.10

Assuming $\psi(x)$ is an absolutely continuous function having derivative $\psi'(x) = \frac{d\psi}{dx}$ such that

$$\psi'(x) = h(x)\ g(x) \qquad \text{a.e.dx}$$

where:

a) $h(F^{-1}(t)-x)$ is a positive bounded continuous function of x at ξ_0 a.e. dt

b) there exist $\varepsilon > 0$, and a partition

$$0 < t_1 < s_1 < \ldots < s_{n_0} < 1 \text{ of } [0,1] \text{ such that}$$

i) g is decreasing in $(-\infty, F^{-1}(t_1) + \varepsilon]$ and g is increasing in $[F^{-1}(s_{n_0}) - \varepsilon , \infty)$ and:

$$\int_{[0,t_1) \cup [s_{n_0},1]} g(F^{-1}(t)-\xi)\ dt < \infty , \quad \xi \varepsilon \{\xi : |\xi - \xi_0| < \varepsilon\}$$

ii) g is uniformly continuous in $A = [F^{-1}(t_1), F^{-1}(s_{n_0})] -$

$$\bigcup_{i=1}^{n_0} (F^{-1}(t_i) , F^{-1}(s_i))$$

iii) F has a bounded variation derivative

$$f \text{ in } [F^{-1}(t_1)\ ,\ F^{-1}(s_{n_0})] - A\ ,$$

then:

$$\lambda'(\xi_0) = -\int_0^1 J(t)\ \psi'(F^{-1}(t) - \xi_0)\ dt\ .$$

<u>Proof</u>: Without loosing generality, assume: $\xi_0 = 0$.

Since J is a bounded variation function, there exists M_0 in R such that:

$$J(t) < M_0 \quad t \in [0,1]\ .$$

$$\frac{\lambda(\xi) - \lambda(0)}{\varepsilon} = \int_0^1 J(t)\ \frac{\psi(F^{-1}(t) - \xi) - \psi(F^{-1}(t))}{\xi}\ dt$$

$$= -\int_0^1 \int_0^1 J(t)\ \psi'(F^{-1}(t) - \xi u)\ du\ dt\ .$$

Let $M_1 = \sup_{x\varepsilon R} h(x)$ and consider:

(2.14) $$\int_{[0,t_1] \cup [s_{n_0},1]} \int_0^1 J(t)\ \psi'(F^{-1}(t) - \xi u)\ du\ dt,$$

for $|\xi| < \varepsilon/2$ and $t \in [0,t_1]$,

$$\int_0^1 \psi'(F^{-1}(t) - \xi u)\ du < M_1\ g(F^{-1}(t) - \varepsilon/2)\ ,$$

if $t \in [s_{n_0},1]$

$$\int_0^1 \psi'(F^{-1}(t) - \xi u)\,du < M_1\, g(F^{-1}(t) + \varepsilon/2)\ ;$$

the assumption on h and the dominated convergence theorem imply that (2.14) goes to:

$$\int_{[0,t_1] \cup [s_{n_0},1]} J(t)\ \psi'(F^{-1}(t))\,dt$$

as ξ goes to 0 .

Consider:

$$(2.15) \qquad \int_A \int_0^1 J(t)\ h\ (F^{-1}(t) - \xi u)\ g\ (F^{-1}(t) - \xi u)\,du\,dt\ .$$

Since g is uniformly continuous in A , g is bounded in A and using dominated convergence, as ξ goes to 0 (2.15) tends to:

$$\int_A J(t)\ \psi'(F^{-1}(t))\,dt\ .$$

Using Tonelli theorem:

$$\int_{t_i}^{s_i} \int_0^1 J(t)\ \psi'(F^{-1}(t) - \xi u)\,du\,dt$$

$$= \int_0^1 \int_{t_i}^{s_i} J(t)\ \psi'(F^{-1}(t) - \xi u)\,dt\,du\ .$$

We want to show:

$$\lim_{\xi \to 0} \int_0^1 \int_{t_i}^{s_i} J(t)\ \psi'(F^{-1}(t) - \xi u)\,dt\,du = \int_{t_i}^{s_i} J(t)\ \psi'(F^{-1}(t))\,dt\ .$$

Integrating by part the inner integral of:

$$\int_0^1 \int_{t_i}^{s_i} J(t)\ \left(\frac{\psi'(F^{-1}(t) - \xi u)}{f(F^{-1}(t))} - \frac{\psi'(F^{-1}(t))}{f(F^{-1}(t))}\right) f(F^{-1}(t))\,dt\,du$$

leads to:

$$\int_0^1 \{J(s_i)\ f(F^{-1}(s_i))\ (\psi(F^{-1}(s_i) - \xi u) - \psi(F^{-1}(s_i)))$$

$$J(t_i)\ f\ (F^{-1}(t_i))\ (\psi(F^{-1}(t_i) - \xi u) - \psi(F^{-1}(t_i))) -$$

$$\int_{[t_i,s_i]}(\psi(F^{-1}(t) - \xi u) - \psi(F^{-1}(t)))\,dJ(t)f(F^{-1}(t))\}\ du\ .$$

Using the uniform continuity of ψ and the fact that $J(t)\ f(F^{-1}(t))$ is a bounded variation function in $[t_i,s_i]$, the last expression goes to 0 as $\xi \to 0$ and $\lambda'(0) = -\int_0^1 J(t)\psi'(F^{-1}(t))\ dt$.

Remark: If $\psi(x)$ is differentiable so is $\psi_H(x)$, $\psi_H'(x) = J(F(x+\xi_0))\ \psi'(x)$. $\psi_H(x)$ fulfills the assumptions of lemma II.10 provided $J(F(F^{-1}(t) - x + \xi_0))$ is continuous at ξ_0 a.e. dt; one easily checks that the last condition is verified if F is continuous. Note that the continuity of F implies: $F(F^{-1}(t)) = t$, so that under this additional assumption,

$$\lambda'(\xi_0) = \lambda_H'\ (\xi_0) = \int_0^1 J(t)\ \psi'(F^{-1}(t) - \xi_0)\ dt\ .$$

In the last part of this section, theorem II.7 will be used to prove various results about the asymptotic behaviour of quantiles and L-M estimators.

First, in corollary II.2 a partial generalization of theorem II.7 is proved.

Corollary II.2

Let J(t) be a bounded variation function in [0,1] (note that J(t) is not assumed to be positive) and take $\psi(x) = x$, if the assumptions of theorem II.6 hold with J(t) and $\psi(x)$ then the estimator:

$$\hat{T}_n = \frac{1}{\sum_{i=1}^{n} J(\frac{i}{n+1})} \sum_{i=1}^{n} J(\frac{i}{n+1}) X_{(i)}$$

satisfies:

$$n^{1/2}(\hat{T}_n - \xi_0 - S_n) \text{ is } o_p(1)$$

where: $$S_n = n^{-1} \sum_{i=1}^{n} \psi_H(X_i - \xi_0) / \int_0^1 J(t)\, dt$$

$$\xi_0 = \int_0^1 J(t)\, dF^{-1}(t) / \int_0^1 J(t)\, dt$$

$$\psi_H(y) = \int_0^y J(F(x+\xi_0))\, dx - \int_0^1 \int_0^{F^{-1}(t) - \xi_0} J(F(x+\xi_0))\, dx\, dt .$$

Note that this corollary implies:

$$L\, (n^{1/2}(\hat{T}_n - \xi_0)) \xrightarrow[n \to \infty]{} N(0, V(\psi_H(X_1-\xi_0))/(\int_0^1 J(t)d(t))^2).$$

Proof: Let: $$J_1(t) = \begin{cases} J(t) & \text{if } J(t) > 0 \\ 0 & \text{elsewhere} \end{cases}$$

$$J_2(t) = J_1(t) - J(t) .$$

Note that $J_j(t) \geq 0$ and $J_j(t)$ and $\psi(x)$ fulfill the assumptions of theorem II.6.

Using lemma II.9, with $k_n(g) = 0$,

$\psi(x) = x$ and $\lambda(x) = \int_0^1 J_j(t)\, F^{-1}(t)\, dt - x \int_0^1 J_j(t)\, dt$ for any $\varepsilon > 0$,

$$\lim_n P\{n^{-1/2} \left| \sum_{i=1}^{n} J_j \left(\frac{i}{n+1}\right) (X_{(i)} - \xi_j) - \psi_{jH}(X_i - \xi_j) \right| > \varepsilon\} = 0$$

$j = 1,2,$ where:

$$\xi_j = \int_0^1 J_j(t)\, dF^{-1}(t) / \int_0^1 J_j(t)\, dt$$

and ψ_{jH} corresponds to ψ_H with $J(t)$ and ξ_0 replaced by $J_j(t)$ and ξ_j . In order to prove the result, one needs to show:

$$(2.16) \quad \psi_{1H}(x-\xi_1) - \psi_{2H}(x-\xi_2) = \psi_H(x-\xi_0) .$$

Consider:

$$\int_0^{x-\xi_1} J_1(F(y+\xi_1))\, dy; \quad \text{if} \quad u = y+\xi_1-\xi_0 , \text{ this integral becomes}$$

$$= \int_{\xi_1-\xi_0}^{x-\xi_0} J_1(F(u+\xi_0))\, du ,$$

hence:

$$\psi_{1H}(x-\xi_1) = \int_{\xi_1-\xi_0}^{x-\xi_0} J_1(F(u+\xi_0))\, du - \int_0^1 \int_{\xi_1-\xi_0}^{F^{-1}(t)-\xi_0} J_1(F(u+\xi_0))\, du\, dt$$

and

$$\psi_{1H}(x-\xi_1) = \int_0^{x-\xi_0} J_1(F(u+\xi_0))\, du - \int_0^1 \int_0^{F^{-1}(t)-\xi_0} J_1(F(u+\xi_0))\, du\, dt .$$

Using a similar argument with ψ_{2H} proves (2.16) .

Corollary II.3 Joint asymptotic distribution of L-M estimators.

Let $\{\hat{T}_{jn}\}_{j=1}^{k}$ be a sequence of L-M estimators based on J_j and ψ_j and estimating ξ_j , j=1,2,...,n . Assuming each pair J_j,ψ_j fulfills the assumptions of theorem II.6 ,

$$L \;\{n^{1/2}(\hat{\underset{\sim}{T}}_n - \underset{\sim}{\xi})\} \xrightarrow[n\to\infty]{} N_k(\underset{\sim}{0}\;,\; \Sigma)$$

where $\hat{\underset{\sim}{T}}_n$ is the vector of $\hat{T}_{jn}$'s ,

$\underset{\sim}{\xi}$ is the vector of ξ_j's ,

$\underset{\sim}{0}$ is a k×1 vector of 0's ,

Σ is the k×k matrix of σ_{rs}'s ,

$$\sigma_{rs} = \mathrm{Cov}\left(\frac{\psi_{rH}(X_1-\xi_r)}{\lambda_r'(\xi_r)} \quad \frac{\psi_{sH}(X_1-\xi_s)}{\lambda_s'(\xi_s)}\right) .$$

Proof: This is a straightforward consequence of theorem II.7 .

Using the last corollary, one can find the distribution of the difference between 2 L-M estimators, so that one can prove the following:

Corollary II.4

Let $\hat{T}_{1n}$, $\hat{T}_{2n}$ be two L-M estimators based on J_1 and ψ_1 , J_2 and ψ_2 respectively, assuming:

i) J_j and ψ_j fulfill the assumptions of theorem II.6 j = 1,2 ,

ii) $\hat{T}_{1n}$ and $\hat{T}_{2n}$ are estimating the same parameter ξ_0 ,

then $n^{1/2}(\hat{T}_{1n} - \hat{T}_{2n})$ is $o_p(1)$ if and only if

$$J_1(F(x+\xi_0))\, d\psi_1(x) = J_2(F(x+\xi_0))\, d\psi_2(x) \text{ and}$$

$\lambda_1'(\xi_0) = \lambda_2'(\xi_0)$ (where λ_j corresponds to λ with J and ψ replaced by J_j and ψ_j) .

Such a relation between L and M estimators has been conjectured and proved under very restrictive conditions by Jaeckel (1971) in his theorem 2.

Note that $J_1(F(x+\xi_0))\, d\psi_1(x) = J_2(F(x+\xi_0))\, d\psi_2(x)$ implies:

$$\psi_{1H}(x) = \psi_{2H}(x) \text{ and } \psi_{1H}'(x) = \psi_{2H}'(x) .$$

Hence if ψ_1 and ψ_2 satisfy the assumptions of lemma II.10 and if F is continuous, then:

$$\lambda_1'(\xi_0) = \lambda_{1H}'(\xi_0) = \lambda_{2H}'(\xi_0) = \lambda_2'(\xi_0)$$

so that the assumption $\lambda_1'(\xi_0) = \lambda_2'(\xi_0)$ is fulfilled.

Corollary II.5 Application of theorem II.7 to quantiles

Suppose F, the distribution of the random sample, has a positive derivative, $f(F^{-1}(\alpha))$ at $F^{-1}(\alpha)$. Then the α^{th} quantile $X_{[\alpha n]+1}$ can be seen as an M estimator based on:

$$\psi_\alpha(x) = \begin{cases} \alpha & x > 0 \\ -(1-\alpha) & x \le 0 \end{cases}$$

so that:

$$n^{1/2}(X_{[\alpha n]+1} - F^{-1}(\alpha) - ([\#X_i > F^{-1}(\alpha)] - n(1-\alpha))/f(F^{-1}(\alpha))) \text{ is } o_p(1)$$

where $[\#X_i > a]$ = number of X_i's bigger than a .

Proof: Consider:

$$\psi_\alpha(x) = \begin{cases} \alpha & x > 0 \\ -(1-\alpha), & x \le 0 \end{cases} .$$

The M estimator $\hat{T}_n$ corresponding to ψ_α is a solution of:

$$h(\theta) = \sum_{i=1}^{n} \psi_\alpha(X_i - \theta) = 0 .$$

Note that

$$h(\theta) = [\#X_i > \theta]\alpha - (1-\alpha)\ [\#X_i \le \theta]$$

$$= [\#X_i > \theta]\alpha - (1-\alpha)\ (n - [\#X_i > \theta])$$

$$= [\#X_i > \theta] \ - n(1-\alpha) .$$

So that if $\theta \in [X_{[\alpha n]}, X_{[\alpha n]+1})$,

$$h(\theta) = \alpha n - [\alpha n] \ge 0$$

and if $\theta \in [X_{[\alpha n]+1}, X_{[\alpha n]+2})$,

$$h(\theta) = \alpha n - [\alpha n] - 1 < 0 .$$

Therefore if $\alpha n \ne [\alpha n]$,

$$\sup\{\theta : h(\theta) \ge 0\} = \inf\{\theta : h(\theta) \le 0\} = X_{[\alpha n]+1}$$

and $X_{[n\alpha]+1} = \hat{T}_n$, the M estimator corresponding to $\psi_\alpha(x)$.

Consider:

$$\lambda(x) = \int_0^1 \psi_\alpha(F^{-1}(t) - x)\, dt \;,$$

for x in a neighborhood of $F^{-1}(\alpha)$, F is monotone increasing and: $F^{-1}(F(x)) = x$, therefore:

$$\lambda(x) = -\,(1-\alpha)\,F(x) + \alpha(1-F(x))$$

$$= \alpha - F(x)$$

$$\text{and} \quad \lambda'(F^{-1}(\alpha)) = -\,f(F^{-1}(\alpha)) \quad .$$

Applying theorem II.7 (note that since ψ_α is bounded the regularity assumptions on F are not needed)

$$n^{1/2}\,(X_{[n\alpha]+1} - F^{-1}(\alpha) - n^{-1}\sum_{i=1}^{n} \psi_\alpha(X_i - F^{-1}(\alpha)) \,/\, f(F^{-1}(\alpha))) \text{ is}$$

$o_p(1)$.

Since:

$$\sum_{i=1}^{n} \psi_\alpha(X_i - F^{-1}(\alpha)) = [\#X_i > F^{-1}(\alpha)] - n(1-\alpha)$$

the theorem is proved.

One easily computes:

$$V(\psi_\alpha(X_1 - F^{-1}(\alpha)) = \alpha(1-\alpha)$$

and applying theorem II.6 to the α^{th} quantile:

$$\mathcal{L}\,(n^{1/2}(X_{[\alpha n]+1} - F^{-1}(\alpha))) \xrightarrow[n \to \infty]{} N(0,\, \alpha(1-\alpha)\,/f^2(F^{-1}(\alpha))) \quad .$$

Once more one computes:

$$\text{Cov}\ (\psi_\alpha(X_1-F^{-1}(\alpha)),\ \psi_\beta(X_1-F^{-1}(\beta))) = (1-\alpha)\ \beta$$

if $\alpha > \beta$. Hence applying corollary II.3, with $0 < \alpha_1 < \alpha_2 < \ldots < \alpha_k < 1$, yields the following classical result:

$$L\ (\underset{\sim}{X}_\alpha - \underset{\sim}{\xi}_\alpha) \xrightarrow{n \to \infty} N_k\ (\underset{\sim}{0}\ ,\ \Sigma)$$

where $\underset{\sim}{X}_\alpha$ is the k×1 vector of the $X_{[\alpha_i n]+1}$'s ,

$\underset{\sim}{\xi}_\alpha$ is the k×1 vector of the $F^{-1}(\alpha_i)$'s ,

$\underset{\sim}{0}$ is a k×1 vector of 0's ,

Σ is the k×k matrix of σ_{rs}'s where

$$\sigma_{rs} = (1-\alpha_r)\ \alpha_s / f(F^{-1}(\alpha_r))\ .\ f(F^{-1}(\alpha_s)) \text{ for } r \geq s\ .$$

Remark:

In theorem II.7, writing

$$\lambda'(\xi_0) = -\int_0^1 J(t)\ \psi'(F^{-1}(t) - \xi_0)\ dt$$

yields

$$n^{1/2}\ (\hat{T}_n - \xi_0 - \frac{1}{n}\sum_{i=1}^{n} \frac{\psi_H(X_i-\xi_0)}{\int_0^1 J(t)\psi'(F^{-1}(t)-\xi_0)\ dt}) \text{ is } o_p(1)\ .$$

Note that

$$\frac{\psi_H(x-\xi_0)}{\int_0^1 J(t)\ \psi'(F^{-1}(t)-\xi_0)\ dt} = I\ C\ (F,\mu)\ (x)$$

is Hampel's influence curve (see Hampel (1974)). For the L estimator, this result has been proved by Stigler (1974).

Section II.6 Step estimators

This section is motivated by practical problems. The first one is the computational aspect involved in finding L-M estimates. To solve the equation

$$\sum_{i=1}^{n} J\left(\frac{i}{n+1}\right) \psi\left(X_{(i)} - \theta\right) = 0$$

directly may be time consuming. Using the asymptotic linearity as described in theorem II.7 should lead to very good approximation of that equation's solution.

During recent years, a great deal of interest has been given to M estimators based on non monotone ψ . For example, Hampel's M estimator based on:

$$\psi(x) = -\psi(-x) = \begin{cases} x & 0 \le x < a \\ a & a \le x < b \\ \frac{c-x}{c-b}\, a & b \le x < c \\ 0 & c \le x \end{cases}$$

has shown a highly robust behavior in the Princeton robustness study (Andrews et al. 1972). So that, it should be of some interest to relax the increasingness assumption on ψ and the positiveness assumption on J .

Definition II.8 The step version of an L-M estimator.

Let $\hat{T}_n$ be an L-M estimator based on ψ and J , the j^{th} step estimator $\hat{T}_n^{(j)}$ of $\hat{T}_n$ based on $\hat{T}_n^*$ is defined as:

$$\hat{T}_n^{(j)} = \hat{T}_n^{(j-1)} + \frac{\sum_{i=1}^{n} J\left(\frac{i}{n+1}\right) \psi\left(X_{(i)} - \hat{T}_n^{(j-1)}\right)}{\sum_{i=1}^{n} J\left(\frac{i}{n+1}\right) \psi'\left(X_{(i)} - \hat{T}_n^{(j-1)}\right)}$$

where $\hat{T}_n^{(0)} = \hat{T}_n^*$ satisfies

$$n^{1/2}(\hat{T}_n^* - \theta_0) \text{ is } O_p(1) \text{ for } \theta_0 \in R .$$

Remarks

1) Note that $\hat{T}_n^{(j)}$ is in fact the j^{th} approximation of the solution of

$$\sum_{i=1}^{n} J(\frac{i}{n+1}) \psi (X_{(i)}-\theta) = 0$$

using the well known Newton-Raphson method starting at

$$\hat{T}_n^{(0)} = \hat{T}_n^* .$$

2) If $\psi(x) = x$ then $\hat{T}_n^{(j)} = \hat{T}_n$ $j \in R$.

3) Note that once the observations have been ordered to compute $\hat{T}_n^{(j)}$ does not require more work than to compute a step version of an M estimator.

The next theorem will provide some clues about the asymptotic behaviour of these step-estimators.

Theorem II.8

Let i) $\{\psi_j\}_{j=1}^{n_0}$ be a sequence of left continuous increasing functions and

$$\psi = \sum_{j=1}^{n_0} a_j \psi_j$$

where $a_j \in R$, $j = 1,2,\ldots,n_0$.

ii) $J_1(t)$ and $J_2(t)$ be two positive bounded variation functions and $J(t) = J_1(t) - J_2(t)$,

iii) $\{\hat{T}^*_n\}_{n=1}^{\infty}$ be a sequence of statistics satisfying

$$n^{1/2}(\hat{T}^*_n - \theta_0) \text{ is } O_p(1) \text{ for a } \theta_0 \epsilon R .$$

Assuming each triplet $(\psi_j , J_k , \theta_0)$ $j=1,\ldots,n_0$ $k=1,2$ satisfies the conditions of theorem II.6 and $\lambda_{(j,k)H}(x)$ is differentiable at θ_0 then:

i) $$\lim_n P\{n^{-1/2}\Big|\sum_{i=1}^{n} J(\tfrac{i}{n+1})\ \psi\ (X_{(i)}-\hat{T}^*_n) - \lambda(\hat{T}^*_n) - \psi_H\ (X_i - \theta_0)\Big| > \varepsilon\} = 0 ,$$

ii) if furthermore ψ' exists, $\lambda'(\theta_0) \neq 0$ and:

$$(\sum_{i=1}^{n} \frac{1}{n} J(\tfrac{i}{n+1})\ \psi'\ (X_{(i)} - \hat{T}^*_n) - \lambda'(\theta_0)) \text{ is } o_p(1)$$

then:

$$n^{1/2}\ (\hat{T}^*_n + \frac{\sum_{i=1}^{n} J(\tfrac{i}{n+1})\ \psi\ (X_{(i)} - \hat{T}^*_n) - \lambda(\theta_0)}{\sum_{i=1}^{n} J(\tfrac{i}{n+1})\ \psi'\ (X_{(i)} - \hat{T}^*_n)} - \theta_0 + \frac{1}{n}\sum_{i=1}^{n} \frac{\psi_H(X_i-\theta_0)}{\lambda'(\theta_0)}) \text{ is } o_p(1) .$$

where:

$$\lambda(x) = \int_0^1 J(t)\ \psi\ (F^{-1}(t) - x)\ dt$$

$$\psi_H(x) = \int_0^x J(F(y+\xi_0))\ d\psi(y) - \int_0^1 \int_0^{F^{-1}(t)-\xi_0} J(F(y+\xi_0))\ d\psi\ (y)\ dt$$

$$\lambda_H(x) = \int_0^1 \psi_H(F^{-1}(t) - x)\ dt .$$

Let $\lambda_{(j,k)}(x)$, $\psi_{(j,k)H}(x)$ and $\lambda_{(j,k)H}(x)$ be the λ , ψ_H and λ_H functions corresponding to ψ_j and J_k .

Note that θ_0 is not assumed to be a solution of $\lambda(x) = 0$.

Proof: Without loss of generality assume $\theta_0 = 0$.

Note that:

$$\psi_H(x) = \sum_{k=1}^{2} \sum_{j=1}^{n_0} a_j \, \psi_{(j,k)H}(x)$$

and $$\lambda_H(x) = \sum_{k=1}^{2} \sum_{j=1}^{n_0} a_j \, \lambda_{(j,k)H}(x)$$

Therefore it suffices to prove i) under the assumptions:

a) ψ is an increasing left continuous function in R ,

b) $J(t)$ is a positive increasing bounded function in $[0,1]$.

Using assumption iii), for any $\varepsilon > 0$, there exist $M_1 = M_1(\varepsilon)$ and $n_1 = n_1(\varepsilon, M_1)$ such that:

$$n > n_1 \rightarrow |\hat{T}^*_n| < M_1 \, n^{-1/2}$$

except on a set of probability at most ε . Therefore using corollary II.1, the first half of lemma II.9 is true and:

$$n^{-1/2} \{ \sum_{i=1}^{n} J\left(\frac{i}{n+1}\right) \psi\,(X_{(i)} - \hat{T}^*_n) - \lambda(\hat{T}^*_n) - \psi_H\,(X_{(i)} - \hat{T}^*_n)$$

$$+ \lambda_H\,(\hat{T}^*_n)\} \text{ is } o_p(1) \text{ as } n \rightarrow \infty \, .$$

Hence, to end the proof of i) it suffices to show

(2.17) $$n^{-1/2} \sum_{i=1}^{n} \{\psi_H\,(X_i - \hat{T}^*_n) - \lambda_H(\hat{T}^*_n) - \psi_H\,(X_i)\} \text{ is } o_p(1) \, .$$

Using the differentiability of λ_H at 0 :

$$\frac{\lambda_H(\hat{T}_n^*)}{\lambda_H'(0)\hat{T}_n^*} \longrightarrow 1 \text{ in prob.}$$

as $n \to \infty$.

$$\frac{\lambda_H(-n^{-1/2}M_1)}{-\lambda_H'(0)n^{-1/2}M_1} \longrightarrow 1$$

So that there exist $M_2=M_2(\varepsilon,M_1,\lambda_H)$ and $n_2=n_2(\varepsilon,M_2)$ $n_2>n_1$ in N such that: $n > n_2$ implies

$$|\lambda_H(\hat{T}_n^*)| < n^{-1/2}M_2$$

and

$$|\hat{T}_n^*| < n^{-1/2}M_1$$

except on a set of probability ε and

$$|\lambda_H(-M_1 n^{-1/2})| < n^{-1/2}M_2 \ .$$

Let $\delta_1 = \frac{\varepsilon}{2M_2}$ and consider, for $n>n_2$

$$n^{-1/2}\sum_{i=1}^{[n\delta_1]} \{\psi_H(X_{(i)}-\hat{T}_n^*) - \lambda_H(\hat{T}_n^*) - \psi_H(X_{(i)})\}$$

$$\leq n^{-1/2}\sum_{i=1}^{[n\delta_1]}\{\psi_H(X_{(i)} + M_1 n^{-1/2}) - \lambda_H(-M_1 n^{-1/2}) - \psi_H(X_{(i)})\}$$

$$+ n^{-1/2} [n\delta_1] \{\lambda_H(-M_1 n^{-1/2}) - \lambda_H(\hat{T}_n^*)\}$$

except on a set of probability at most ε . Using the fact that $\delta_1=\frac{\varepsilon}{2M_2}$ and the way M_1 and M_2 were chosen, for $n>n_2$ the last expression is less or equal than:

$$n^{-1/2} \sum_{i=1}^{[n\delta_1]} \{\psi_H(X_{(i)}+M_1 n^{-1/2}) - \lambda_H(-M_1 n^{-1/2}) - \psi_H(X_{(i)})\} + \varepsilon$$

except on a set of probability ε .

One proves the same way, for $n>n_2$:

$$n^{-1/2} \sum_{i=n-[n\delta_1]+1}^{n} \{\psi_H(X_{(i)}-\hat{T}_n^*) - \lambda_H(\hat{T}_n^*) - \psi_H(X_{(i)})\}$$

$$\leq n^{-1/2} \sum_{i=n-[n\delta_1]+1}^{n} \{\psi_H(X_{(i)}+M_1 n^{-1/2}) - \lambda_H(-M_1 n^{1/2}) - \psi_H(X_{(i)})\} + \varepsilon$$

except on this same set of probability at most ε .

So that for $n>n_2$, (2.17) is less or equal than:

$$n^{-1/2} \sum_{i=1}^{n} \{\psi_H(X_i+M_1 n^{-1/2}) - \lambda_H(-Mn^{-1/2}) - \psi_H(X_i)\} +$$

$$n^{-1/2} \sum_{i=[n\delta_1]+1}^{n-[n\delta_1]} \{\psi_H(X_{(i)}-\hat{T}_n^*) - \lambda_H(\hat{T}_n^*) - \psi_H(X_{(i)}+M_1 n^{-1/2})$$

$$+ \lambda_H(-M_1 n^{-1/2})\} + 2\varepsilon$$

except on a set of probability at most ε .

By an argument similar to the one in the second half of lemma II.9 with

$k_n = -M_1 n^{-1/2}$, one proves:

$$n^{-1/2} \sum_{i=1}^{n} \{\psi_H(X_i+M_1 n^{-1/2}) - \lambda_H(-M_1 n^{-1/2}) - \psi_H(X_i)\} \text{ is } o_p(1) .$$

By corollary II.1 and by an argument similar to the one used at the end of the first half of lemma II.9:

$$n^{-1/2} \sum_{i=[n\delta_1]+1}^{n-[n\delta_1]} \{\psi_H(X_{(i)}-\hat{T}^*_n) - \lambda_H(\hat{T}^*_n) - \psi_H(X_{(i)}+M_1 n^{-1/2}) + \lambda_H(-M_1 n^{-1/2})\} \text{ is } o_p(1) .$$

Therefore (2.17) is bounded by $O(\varepsilon)$ except on a set of probability at most ε for big n .

Using a similar argument, it is shown that (2.17) is bigger than $O(-\varepsilon)$ except on a set of probability at most ε for large n and i) is proved.

To prove ii), first note that

$$n^{1/2}\{\frac{\lambda(\hat{T}^*_n)-\lambda(0)}{\lambda'(0)} - \frac{1}{n}\sum_{i=1}^{n}\{\frac{J(\frac{i}{n+1})\psi(X_{(i)}-\hat{T}^*_n)-\lambda(0)}{\lambda'(0)} \quad \frac{-\psi_H(X_i)}{\lambda'(0)}\}\} \text{ is } o_p(1)$$

using part (i) and the fact that $\lambda'(0)$ exists and is non zero.

By an argument similar to the one used at the end of theorem II.5

$$n^{1/2}(\frac{\lambda(\hat{T}^*_n)-\lambda(0)}{\lambda'(0)} - \hat{T}^*_n) \text{ is } o_p(1) .$$

Now consider

$$n^{-1/2}\sum_{i=1}^{n}\{J(\frac{i}{n+1})\psi(X_{(i)}-\hat{T}^*_n)-\lambda(0)\}\ \{\frac{1}{\lambda'(0)} - \frac{n}{\sum_{i=1}^{n} J(\frac{i}{n+1})\psi'(X_{(i)}-\hat{T}^*_n)}\} .$$

To prove that the last expression is $o_p(1)$ it suffices to show:

$$n^{-1/2}\sum_{i=1}^{n}\{J(\tfrac{i}{n+1})\ \psi\ (X_{(i)} - \hat{T}_n^*) - \lambda(0)\}\ \text{is}\ O_p(1)\ .$$

Using the preceding argument:

$n^{1/2}(\lambda(\hat{T}_n^*) - \lambda(0))$ and $n^{1/2}\lambda'(0)\hat{T}_n^*$ have the same asymptotic behaviour, and using part (i):

$$\mathcal{L}\ \{n^{-1/2}\sum_{i=1}^{n}\{J(\tfrac{i}{n+1})\ \psi\ (X_{(i)} - \hat{T}_n^*) - \lambda(\hat{T}_n^*)\}\}\ \xrightarrow[n\to\infty]{}$$

$$N(0,\ E(\psi_H^2(X_1)))\ .$$

So that writing

$$n^{-1/2}\sum_{i=1}^{n}\{J(\tfrac{i}{n+1})\ \psi\ (X_{(i)} - \hat{T}_n^*) - \lambda(0)\} =$$

$$n^{-1/2}\sum_{i=1}^{n}\{J(\tfrac{i}{n+1})\ \psi\ (X_{(i)} - \hat{T}_n^*) - \lambda(T_n^*)\} +$$

$$n^{1/2}\{\lambda(T_n^*) - \lambda(0)\}$$

ends the proof.

Corollary II.6

Under the assumptions of theorem II.8, assuming $\lambda(\theta_0)=0$ then

$$n^{1/2}(\hat{T}_n^{(j)}-\theta_0)\ \text{is}\ O_p(1)\ \ j\varepsilon\{1,2,\ldots\}$$

Proof: Assume without loss of generality $\theta_0=0$.

Using theorem II.8

$$n^{1/2}(\hat{T}_n^{(1)} - \frac{1}{n}\sum_{i=1}^{n}\frac{\psi_H(X_i)}{\lambda'(0)}) \text{ is } o_p(1) .$$

therefore $(\hat{T}_n^{(1)}-\theta_0)$ is $O_p(1)$.

The corollary is proved by iterating this result.

Note that if F is assumed to be symmetric about θ_0 , if an odd $\psi(x)$ and a J(t) symmetric about 1/2 are chosen, then $\lambda(\theta_0) = 0$.

So that if $\hat{T}_n^*$ is any statistic converging to θ_0 (the median or the δ-trimmed mean or a symmetrically weighted sum of selected quantiles) the result of corollary II.6 holds.

The next corollary is dealing with the asymmetric case:

<u>Corollary II.7</u>

Assuming $\lambda'(x)$ and $\lambda''(x)$ exist in a subset of R containing $\{\theta_j\}_{j=0}^{k}$ where θ_j is the j^{th} approximation of the solution of $\lambda(x)=0$ using the Newton Raphson method starting at θ_0 .

Assuming each triplet (ψ,J,θ_j) , (ψ',J,θ_j) fulfil the assumptions of theorem II.8 for $j\epsilon\{0,1,\ldots,k-1\}$

$$n^{1/2}(\hat{T}_n^{(k)}-\theta_k) \text{ is } O_p(1) .$$

<u>Proof</u>: Without loss of generality, assume $\theta_0=0$, using theorem II.8:

$$n^{1/2}(\hat{T}_n^{(1)} - \frac{n\,\lambda(0)}{\sum_{i=1}^{n} J(\frac{i}{n+1})\psi'(X_{(1)}-\hat{T}_n^*)} + \frac{1}{n}\sum_{i=1}^{n}\frac{\psi_H(X_i)}{\lambda'(0)}) \text{ is } o_p(1) .$$

Using the Central Limit theorem:

$$n^{1/2}(\frac{1}{n}\sum_{i=1}^{n} \frac{\psi_H(X_i)}{\lambda'(0)}) \text{ is } 0_p(1) .$$

Note that θ_1 the first approximation of the root of $\lambda(x) = 0$ using the Newton Raphson method starting at 0 is:

$$\frac{-\lambda(0)}{\lambda'(0)} .$$

To prove $n^{1/2}(T_n^{(1)}-\theta_1)$ is $0_p(1)$, it suffices to show:

$$n^{1/2}(\lambda'(0) + n^{-1}\sum_{j=1}^{n} J(\frac{i}{n+1})\ \psi'(X_{(i)} - \hat{T}_n^*)) \text{ is } 0_p(1) .$$

Note that

$$-\lambda'(x) = \lambda^{(1)}(x) = \int_0^1 J(t)\ \psi'(F^{-1}(t)-x)\ dt ;$$

using theorem II.8's part (i) with (J,ψ')

$$n^{-1/2}(\sum_{j=1}^{n} J(\frac{i}{n+1})\ \psi'(X_{(i)} - \hat{T}_n^*) + \lambda'(\hat{T}_n^*)) \text{ is } 0_p(1) .$$

Now using the fact that $\lambda''(x)$ exists,

$$n^{1/2}\{\lambda'(\hat{T}_n^*) - \lambda'(0) - \lambda''(0)\ \hat{T}_n^*\} \text{ is } o_p(1)$$

by an argument similar to the one used at the end of theorem II.5 .

So that, since

$n^{1/2}(\hat{T}_n^*)$ is $0_p(1)$, and

$$n^{1/2}(\lambda'(\hat{T}_n^*) - \lambda'(0)) \text{ is } O_p(1) \text{ , one obtains}$$

$$n^{1/2}(\hat{T}_n^{(1)} - \theta_1) \text{ is } O_p(1) \text{ .}$$

Iterating this result proves the theorem.

In the second part of this section, an extension of the L-M estimator will be presented. In recent literature there is a great deal of interest in M estimators based on non-increasing ψ (Andrews et al. (1972), Collins (1976) (1977)) so that it makes sense to drop the increasingness assumption on ψ and the positiveness assumption on J in order to obtain analogous L-M estimators.

To give a formal definition of those extended L-M estimators presents a problem since the equation:

$$\sum_{i=1}^{n} J(\tfrac{i}{n+1})\ \psi\ (X_{(i)}-\theta) = 0$$

no longer has a unique solution. Furthermore this formal definition should lead to easily computable estimates. The following is suggested:

<u>Definition II.9</u> <u>Extended L-M estimator</u>

Let i) J(t) be a bounded variation function in [0,1] ,

ii) $\psi(x)$ be a left continuous function.

The extended L-M estimator $\hat{T}_n$ based on ψ and J, is defined as the N^{th} step estimator obtained when solving

$$\sum_{i=1}^{n} J(\tfrac{i}{n+1})\ \psi\ (X_{(i)}-\theta) = 0$$

using the Newton Raphson method starting at a given statistic $\hat{T}_n^*$.

Remarks:

1) If the extended L-M estimator is defined as $\hat{T}_n = \lim_k \hat{T}_n^{(k)}$, for each k:

$$n^{1/2}\left(\hat{T}_n^{(k)} - \theta_k + \frac{1}{n}\sum_{i=1}^{n} \frac{\psi_H(X_i - \theta_{k-1})}{\lambda'(\theta_{k-1})} - \frac{\lambda(\theta_{k-1})}{\lambda'(\theta_{k-1})} - \frac{n\lambda(\theta_{k-1})}{\sum_{i=1}^{n} J(i/(n+1))\psi'(X_{(i)} - \hat{T}_n^{(k-1)})}\right) \text{ is } o_p(1)$$

by corollary II.7. To conclude $\hat{T}_n$ is consistent, the last convergence should be uniform in k, this is not obvious (cf Collins (1976)).

2) If $\psi(x) = x$, $\hat{T}_n = \frac{1}{\sum_{i=1}^{n} J(\frac{i}{n+1})} \cdot \sum_{i=1}^{n} J(\frac{i}{n+1})\, X_{(i)}$,

provided $\sum_{i=1}^{n} J(\frac{i}{n+1}) \neq 0$.

Corollary II.8

Suppose ψ, J and $\hat{T}_n^*$ fulfil the assumptions of theorem II.8, the extended L-M estimator $\hat{T}_n$ satisfies:

i) if $\lambda(\theta_0) = 0$,

$$n^{1/2}\left(\hat{T}_n - \xi_0 + \frac{1}{n}\sum_{i=1}^{n} \frac{\psi_H(X_i - \theta_0)}{\lambda'(\theta_0)}\right) \text{ is } o_p(1) ,$$

ii) if furthermore ψ and J satisfy the assumptions of corollary II.7 for k=N and if ξ_0 is the solution obtained when solving $\lambda(x)=0$ using the Newton Raphson method starting at θ_0 after N iterations,

$$n^{1/2}(\hat{T}_n - \xi_0) \text{ is } O_p(1) .$$

Proof: These results are easy consequences of theorem II.8 and corollary II.6 and II.7,

Note that the results of corollary II.3, II.4, II.6 and II.7 are also true for extended L-M estimators in the symmetric case.

Section II.7 Asymptotically efficient and minimax estimators.

In this section it will be shown that, in a parametric context, certain L-M estimators have desirable asymptotic properties.

Excluding superefficiency, see Huber (1969), an estimator $\hat{T}_n$ of a location parameter ξ_0 has asymptotically minimum variance for a distribution F (with density f) if:

$$\mathcal{L}\,(n^{1/2}(\hat{T}_n-\xi_0)) \xrightarrow[n\to\infty]{} N(0,I(f)^{-1})$$

where I(f) is the Fisher information for the location parameter:

$$I(f) = \int_{-\infty}^{\infty} \left(\frac{f'(x+\xi_0)}{f(x+\xi_0)}\right)^2 f(x+\xi_0)\,dx\ .$$

Such an estimator is said to be efficient. The next theorem will provide a characterization of efficient L-M estimators.

Theorem II.9

Let F be a distribution with a twice differentiable density such that $\lim_{n\to\pm\infty} f'(x)=0$ and suppose that the L-M (or extended L-M) estimator $\hat{T}_n$ based on ψ and J has asymptotic variance:

$$\sigma^2 = \frac{V(\psi_H(X-\xi_0))}{[\int_{-\infty}^{\infty}\psi_H'(x-\xi_0)f(x)dx]^2}$$

then $\hat{T}_n$ is an efficient estimator for ξ_0 , the location parameter of F if and only if:

$$\psi_H(x) = a\phi(x) \text{ where } \phi(x) = -\frac{f'(x+\xi_0)}{f(x+\xi_0)} \text{ and } a\varepsilon R\ .$$

Proof: Without loss of generality, assume $\xi_0 = 0$.

Since $E(\psi_H(x)) = 0$,

$$\sigma^2 = \frac{\int_{-\infty}^{\infty} \psi_H^2(x)\ f(x)\ dx}{[\int_{-\infty}^{\infty} f(x)\ \psi_H'(x)\ dx]^2} \quad .$$

Consider:

$$\int_{-\infty}^{\infty} f(x)\ \psi_H'(x)\ dx = f(x)\ \psi_H(x) \Big|_{-\infty}^{\infty}$$

$$- \int_{-\infty}^{\infty} f'(x)\psi_H(x)\ dx ,$$

under the condition:

$$\lim_{x \to \pm\infty} f(x)\psi_H(x) = 0 ,$$

$$\int_{-\infty}^{\infty} f(x)\ \psi_H'(x)\ dx = \int_{-\infty}^{\infty} \phi(x)\psi_H(x)\ f(x)\ dx \ .$$

Using the Cauchy-Schwarz inequality:

$$[\int_{-\infty}^{\infty} \phi(x)f(x)\ \psi_H(x)\ dx]^2 \le \int_{-\infty}^{\infty} \phi(x)^2\ f(x)\ dx \ . \ \int_{-\infty}^{\infty} \psi_H^2(x)\ f(x)\ dx$$

with equality if and only if:

$$\psi_H(x) = a\phi(x) \text{ where } a \varepsilon R \ .$$

So that $\sigma^2 \ge \frac{1}{I(f)}$

with equality if and only if:

$$\psi_H(x) = a\phi(x) \ .$$

Note that $\lim_{x \to \pm\infty} f'(x) = 0$ implies:

$$\lim_{x \to \pm\infty} \phi(x)\, f(x) = 0$$

and the theorem is proved.

Examples

1) Let $X_1,\ldots,X_n$ be a random sample from a distribution $F(x)=\Phi(x-\xi_0)$ where

$$\Phi(x) = \int_{-\infty}^{x} (2\pi)^{-1} e^{-1/2\, y^2}\, dy \; .$$

We want to find an efficient L-M estimator of ξ_0 based on ψ and:

$$J(u) = u(1-u) \quad u\varepsilon[0,1]$$

$$\psi_H(x) = \int_0^x \Phi(y)\,(1-\Phi(y))\,\psi'(y)\,dy ,$$

so that taking

$$\psi(x) = \int_0^x \frac{1}{\Phi(y)\,(1-\Phi(y))}\,dy$$

$\psi_H(x) = \phi(x) = x$ and the L-M estimator $\hat{T}_n$ based on ψ and J is efficient. Note that using corollary II.4

$$n^{1/2}(\hat{T}_n-\bar{X}) \text{ is } o_p(1) \; .$$

2) $X_1,\ldots,X_n$ is a random sample from a distribution $F(x)=G(x-\xi_0)$ where:

$$G(x) = \frac{1}{1+e^{-x}} \quad x\varepsilon R$$

is the logistic distribution

$$\phi(x) = -\frac{d}{dx}\,\ell n\, g(x)$$

$$= -\frac{d}{dx}\,\ell n\, \frac{e^{-x}}{(1+e^{-x})^2}$$

$$= \frac{e^x - 1}{e^x + 1} \ .$$

So that an efficient L-M estimator for ξ_0 with:

$$J(t) = 1 \text{ must have } \psi(x) = \frac{e^x - 1}{e^x + 1} \ .$$

This is the maximum likelihood estimator. Now if $\psi(x) = x$, $J(t)$ must satisfy:

$$\frac{e^x - 1}{e^x + 1} = \int_0^x J\left(\frac{1}{1+e^{-y}}\right) dy \ .$$ Differentiating both sides we have

$$\frac{2e^x}{(1+e^x)^2} = J\left(\frac{1}{1+e^{-x}}\right)$$ so that

if $J(t) = t(1-t)$ and $\psi(x) = x$, we have the efficient L estimator.

If $J(u) = \begin{cases} u^2 , & u<1/2 \\ (1-u)^2, & u \geq 1/2 \end{cases}$

for $x > 0$

$$\frac{e^x - 1}{e^x + 1} = \int_0^x \left(\frac{e^y}{e^y+1}\right)^2 \psi'(y)$$

differentiating:

$$\frac{2e^x}{(e^x+1)^2} = \frac{(e^x)^2}{(e^x+1)^2} \quad \psi'(x)$$

so that if $\psi'(x) = e^{-x}$ for $x > 0$

,

or $\psi(x) = \begin{cases} 1 - e^{-x} & x \geq 0 \\ e^x - 1 & x < 0 \end{cases}$

the L-M estimator based on J and ψ is efficient.

Definition II.10 Strong unimodality

A distribution F is said to be strongly unimodal if its density f exists and satisfies: log f(x) is a convex function within some open interval (a,b) such that $-\infty \leq a < b \leq \infty$

and $\int_a^b f(x)\,dx = 1$.

Note that strongly unimodal distributions are distributions having a monotone likelihood ratio for the location parameter.

Let G_0 be a given strongly unimodal distribution and let:

$$\mathcal{G} = \{G : G = (1-\varepsilon)\, G_o + \varepsilon H \text{ where H is an absolutely continuous distribution function}\} \text{ where } \varepsilon \text{ is fixed in } (0,1) .$$

Given that the X_i's distribution F(x) can be written $G_x(x-\xi_0)$ where $G_x \varepsilon \mathcal{G}$, we want to find an estimator for ξ_0 which is minimax for the family $\mathcal{G}$, i.e. an estimator which minimizes

$$\max_{G \varepsilon \mathcal{G}} \sigma^2(G)$$

where $\sigma^2(G)$ is the asymptotic variance of the estimator if the X_i's distribution is $G(x-\xi_0)$.

The next theorem due to Huber (1969) provides a minimax M estimator.

Theorem II.10

The maximum likelihood estimator for the distribution:

$$f_0(x) = \begin{cases} (1-\varepsilon)\ g_0\ (x_o)\ e^{K(x-x_0)} & x \le x_0 \\ (1-\varepsilon)\ g_0\ (x) & x_0 \le x \le x_1 \\ (1-\varepsilon)\ g_0\ (x_1)\ e^{-K(x-x_1)} & x_1 \le x \end{cases}$$

where •g_0 is the density of G_0

•x_0 and x_1 are the end points of

$$\{x : \frac{\dot{g}_0'(x)}{g_0(x)} < K\}$$

•K satisfies:

$$(1-\varepsilon)\ (\frac{g(x_0)}{K} + \frac{g(x_1)}{K} + \int_{x_0}^{x_1} g(x)\ dx) = 1$$

is minimax for the family $\mathcal{G}$.

The next theorem will provide some minimax L-M estimators.

<u>Theorem II.11</u>

When G_0 and H are symmetric with respect to 0 , any efficient L-M estimator for F_0 satisfying:

i) $\psi(x) = x, x\varepsilon R$,

or ii) J(t) is symmetric with respect to 1/2 , decreasing in [1/2, 1], null in $[F_0(x_1),1]$ and $\psi'(x)$ is an even function decreasing in $[0,\infty)$, is minimax for the family $\mathcal{G}$.

<u>Proof</u>: Without loss of generality assume ξ_0 , the estimated parameter, is 0 . Under condition i), the estimator under consideration is the efficient L estimator for F_0 and the result is proved in Jaeckel (1971) theorem 3.

Under condition ii),

$$\sigma^2(G) = \frac{2\int_0^\infty(\int_0^y J(G(x))\psi'(x)\,dx)^2\,dG(y)}{(2\int_{1/2}^1 J(t)\psi'(G^{-1}(t))dt)^2}$$

using the symmetry assumptions.

For $x\varepsilon[0,x_1]$,

$$G(x) = (1-\varepsilon)\,G_0(x) + \varepsilon\,H(x)$$

$$\geq (1-\varepsilon)\,G_0(x)$$

$$= F_0(x) \ ,$$

so that for $t\varepsilon[1/2,F_0(x_1)]$

$t = G(G^{-1}(t)) \geq F_0(G^{-1}(t))$, ~~hence~~

$$F_0(F_0^{-1}(t)) \geq F_0(G^{-1}(t)) \rightarrow F_0^{-1}(t) \geq G^{-1}(t) \ .$$

Therefore using the fact that J is decreasing in [1/2,1],

$$J(G(x))\,\psi'(x) \leq J(F_0(x))\psi'(x) \quad x\varepsilon[0,x_1]$$

and using the fact that ψ' is decreasing in $[0,\infty]$,

$$J(t)\psi'(G^{-1}(t)) \geq J(t)\psi'(F_0^{-1}(t)) \quad t\varepsilon[1/2,F_0(x_1)] \ .$$

hence:

$$\sigma^2(G) \leq \frac{1}{2}\,\frac{\int_0^\infty(\int_0^y J(F_0(x))\psi'(x)dx)^2\,dG(y)}{[\int_{1/2}^1 J(s)\psi'(F_0^{-1}(s))ds]^2}$$

now since:

$\int_0^y J(F_0(x))\psi'(x)\,dx$ is increasing in $[0,x_1]$, constant $y > x_1$,

$$\sigma^2(G) \le \frac{1}{2}\frac{\int_0^\infty(\int_0^y J(F_0(x))\psi'(x)dx)^2 dF_0(y)}{[\int_{1/2}^1 J(s)\psi'(F_0^{-1}(s))\,ds]^2} = \sigma^2(F_0)$$

and the result is proved.

Chapter III

A formal theory of L-M location parameters

In this chapter, the location parameter will be studied as a functional defined from a set of distribution functions $\mathcal{F}$ into R.

Section III.1 Bickel-Lehmann location parameters

Definition III.1

Let X and Y be random variables distributed according to F_X and F_Y respectively, then

i) X is stochastically greater than Y $\leftrightarrow F_Y(x) \geq F_X(x) \; x \varepsilon R$,

notation $X > Y$ s.,

ii) $Y = a X + b$ s. $\leftrightarrow F_X(\frac{x-b}{a}) = F_Y(x) \; x\varepsilon R$, $a > 0$,

iii) $Y = -X$ s. $\leftrightarrow F_Y(x) = 1 - F_X(-x)$ a.e..

Definition III.2 Bickel-Lehmann location parameter

A functional μ defined from a set of distributions $\mathcal{F}$ (if X is a random variable distributed according to F, $\mu(X)$ or $\mu(F)$ will be used to denote the image of F by μ) into R is said to be a Bickel-Lehmann location parameter (BLLP) if:

i) $X,Y\varepsilon\mathcal{F}$ such that $X > Y$ s. $\rightarrow \mu(X) \geq \mu(Y)$

ii) $X,Y\varepsilon\mathcal{F}$ such that $X = -Y$ s. $\rightarrow \mu(X) = -\mu(Y)$

iii) $X,Y\varepsilon\mathcal{F}$ such that $Y = aX+b$ s. $\rightarrow \mu(Y) = a\mu(X) + b$

where $a\varepsilon R+$, $b\varepsilon R$.

As pointed out in Lehmann and Bickel's (1975) theorem 1:

(3.1) If F is symmetric with respect to θ and μ is a BLLP, $\mu(F) = \theta$.

(3.2) If $P\{X\varepsilon[a,b]\} = 1$ and μ is a BLLP, $b \geq \mu(X) \geq a$.

Definition III.3 L-M location parameter

Let i) $J(t)$ be a positive bounded variation function defined on $[0,1]$,

ii) $\psi(x)$ be an increasing differentiable function defined in R .

The L-M location parameter based on J and ψ for the distribution F, $\mu(F)$ is defined as the solution of:

$$\int_0^1 J(t)\ \psi(F^{-1}(t) - \theta)\ dt = 0 \ .$$

In this section we will find what conditions must J and ψ fulfill in order for μ to be a BLLP.

For convenience, μ is a BLLP on F will really mean μ is a BLLP on the subset of F in which it is defined.

One easily checks that $\mu(X) = E(X)$ is a BLLP on any set F of distributions having finite first moment. According to the last statement we say μ is BLLP on any set F .

Define: $F_\sigma(x) = F(\frac{x}{\sigma})$,

$$F^{(-)}(x) = 1-F(-x) - \lim_{\substack{y\to -x \\ y<-x}} [F(-x) - F(y)] \ .$$

Note that if X has distribution F, σX has distribution F_σ and -X has distribution $F^{(-)}$.

Theorem III.1

Let μ_{J_0} be an L-M location parameter based on J_0 and ψ which is a BLLP on any set of distributions.

Assuming:

i) $J_0(t) > \eta$ for $t \in [1/2 - \eta , 1/2]$ for some $\eta > 0$,

ii) there exists an absolutely continuous, symmetric with respect to 0 distribution F_0 having a density with compact support f_0 which is non null in $\{x : 0 < F_0(x) < 1\}$ satisfying:

for any bounded variation function $J(t)$ defined on [0,1], for any $\sigma > 0$, $x \in R$,

$$\frac{d}{d\varepsilon} \int_0^1 J(t)\ \psi(\sigma\ F_{x\varepsilon}^{-1}(t) - \theta_{\varepsilon\sigma})\ dt$$

$$= \int_0^1 J(t)\ \frac{d}{d\varepsilon}\ \psi(\sigma\ F_{x\varepsilon}^{-1}(t) - \theta_{\varepsilon\sigma})\ dt$$

and the last expression is a continuous function for ε in [0,1/2] where:

$$F_{x\varepsilon} = (1-\varepsilon)\ F_0 + \varepsilon\delta_x$$

$$\delta_x = 1_{[x,\infty)}$$

$$\theta_{\varepsilon\sigma} = \mu_J(F_{x\varepsilon}(y/\sigma)) .$$

Then $J_0(t) = J_0(1-t)$ and

$$\psi(x) = \text{sgn}(x)\ |x|^\alpha \quad \alpha > 0 .$$

Note since f_0 has compact support $\mu_J(F_{x\varepsilon}(\frac{y}{\sigma}))$ is always defined where μ_J is the L-M location parameter based on ψ and J .

<u>Lemma III.1</u>

Let K(t) be an absolutely continuous distribution defined on [0,1] which is symmetric with respect to 1/2 , let $\mathcal{F}$ be a set of distributions and

$$\mathcal{F}_K = \{K(F(x)) : F\varepsilon\mathcal{F}\}$$

if μ is a BLLP on $\mathcal{F}_K$, then

$$\mu_K(F) = \mu(K(F)) \text{ is a BLLP on } \mathcal{F} .$$

<u>Proof</u>: If X is a random variable distributed according to F_X , then X_K is a random variable distributed according to $K(F_X)$.

i) $X > Y \text{ s} \leftrightarrow F_Y(x) > F_X(x)$

$$\rightarrow K(F_Y(x)) \geq K(F_X(x))$$

since K is non decreasing.hence

$X_K > Y_K$ s. so that:

$$\mu_K(X) = \mu(X_K) \geq \mu(Y_K) = \mu_K(Y).$$

ii) $Y = - X \text{ s.} \rightarrow F_Y(x) = 1 - F_X(-x) \text{ a.e.}$

$$\rightarrow K(F_Y(x)) = K(1-F_Y(-x)) \text{ a.e.}$$

$$= 1 - K(F_Y(-x)) \text{ a.e.}$$

since K is absolutely continuous symmetric with respect to 1/2 .

Therefore:

$Y_K = - X_K$ s. and $\mu_K(Y) = - \mu_K(X)$.

iii) $Y = aX + b$ s. $a > 0$ $b \varepsilon R$

$\leftrightarrow F_Y(x) = F_X(\frac{x-b}{a})$

$\leftrightarrow K(F_Y(x)) = K(F_X(\frac{x-b}{a}))$ and

$Y_K = aX_K + b$ s. so that $\mu_K(Y) = a\mu_K(X) + b$.

This ends the proof.

Note that if the distributions in F are absolutely continuous, the absolute continuity of K may be removed and the result is still valid.

Proof of theorem III.1

Consider:

$$F_{x\varepsilon}^{-1}(t) = \inf \{y : (1-\varepsilon) F_0(y) + \varepsilon\delta_x(y) \geq t\}$$

i) if $t < (1-\varepsilon) F_0(x)$

$$F_{x\varepsilon}^{-1}(t) = F_0^{-1}(\frac{t}{1-\varepsilon})$$

ii) if $(1-\varepsilon) F_0(x) \leq t < (1-\varepsilon) F_0(x) + \varepsilon$

$$F_{x\varepsilon}^{-1}(t) = x$$

iii) if $(1-\varepsilon) F_0(x) + \varepsilon \leq t$

$$F_{x\varepsilon}^{-1}(t) = F_0^{-1}(\frac{t-\varepsilon}{1-\varepsilon}) .$$

So that $\lim_{\varepsilon \to 0} F_{x\varepsilon}^{-1}(t) = F_0^{-1}(t)$ hence:

since $\int_0^1 J(t)\, \psi(F_{x\varepsilon}^{-1}(t) - \theta_\varepsilon)\, dt$ is a continuous function of ε ,

$$\int_0^1 \psi(F_0^{-1}(t) - \theta_0)\, dt = 0$$

where $\theta_0 = \lim_{\varepsilon \to 0} \theta_\varepsilon$

and $\theta_0 = \mu_J(F_0) = 0$ if μ_J is a BLLP, by (3.1).

Furthermore $F_{x\varepsilon}^{-1}$ is differentiable and:

$$\frac{d}{d\varepsilon} F_{x\varepsilon}^{-1}(t) = \frac{1}{(1-\varepsilon)^2} \begin{cases} \dfrac{t}{f_0(F_0^{-1}(\frac{t}{1-\varepsilon}))} & t < (1-\varepsilon)\, F_0(x) \\ 0 & (1-\varepsilon)\, F_0(x) < t < (1-\varepsilon)\, F_0(x)+\varepsilon \\ \dfrac{t-1}{f_0(F_0^{-1}(\frac{t-\varepsilon}{1-\varepsilon}))} & (1-\varepsilon)\, F_0(x) + \varepsilon < t\ . \end{cases}$$

Using the differentiability assumption, for ε small enough

$$\int_0^1 J(t)\, \left(\frac{d}{d\varepsilon} F_{x\varepsilon}^{-1}(t) - \frac{d}{d\varepsilon}\theta_\varepsilon\right) \psi'(F_{x\varepsilon}^{-1}(t) - \theta_\varepsilon)\, dt = 0$$

so that if ε goes to 0 ,

$$\left.\frac{d}{d\varepsilon}\theta_\varepsilon\right|_{\varepsilon=0} = \frac{\int_0^1 J(t)\, \dfrac{t-\delta_x(F_0^{-1}(t))}{f_0(F_0^{-1}(t))}\, \psi'(F_0^{-1}(t))\, dt}{\int_0^1 J(t)\, \psi'(F_0^{-1}(t))\, dt}$$

provided μ_J is BLLP . If $t = F_0(y)$

$$\left.\frac{d}{d\varepsilon}\theta_\varepsilon\right|_{\varepsilon=0} = \frac{\int_{-\infty}^{\infty} J(F_0(y))\, (F_0(y) - \delta_x(y))\, \psi'(y)\, dy}{\int_{-\infty}^{\infty} J(F_0(y))\, \psi'(y)\, f_0(y)\, dy}\ .$$

Note that:

$$[(1-\varepsilon)\ F_0 + \varepsilon\delta_x]^{(-)} = (1-\varepsilon)\ F_0^{(-)} + \varepsilon\delta_x^{(-)}$$

$$= (1-\varepsilon)\ F_0 + \varepsilon\delta_{-x}\ .$$

Repeating the preceding argument with

$$\omega_\varepsilon = \mu_J\{[(1-\varepsilon)F_0+\varepsilon\delta_x]^{(-)}\} = -\ \theta_\varepsilon\ ,$$

one obtains

$$-\frac{d}{d\varepsilon}\theta_\varepsilon\bigg|_{\varepsilon=0} = \frac{\int_{-\infty}^{\infty} J(F_0^{(-)}(y))\ (F_0^{(-)}(y) - \delta_x^{(-)}(y))\ \psi'(y)\ dy}{\int_{-\infty}^{\infty} J(F_0^{(-)}(y))\ \psi'(y)\ dF_0^{(-)}(y)}$$

$$= \frac{\int_{-\infty}^{\infty} J(F_0(y))\ (F_0(y) - \delta_{-x}(y))\ \psi'(y)\ dy}{\int_{-\infty}^{\infty} J(F_0(y))\ \psi'(y)\ dF_0(y)}\ .$$

Hence:

$$\int_{-\infty}^{\infty} J(F_0(y))\ (F_0(y) - \delta_{-x}(y))\ \psi'(y)\ dy$$

$$= -\int_{-\infty}^{\infty} J(F_0(y))\ (F_0(y) - \delta_x(y))\ \psi'(y)\ dy$$

differentiating with respect to x:

$$(3.3) \qquad J(F_0(-x))\,\psi'(-x) = J(F_0(x))\ \psi'(x)\ .$$

Note that the fact that f_0 has compact support implies that the preceding integrals are defined.

Now consider:

$$\int_{-\infty}^{\infty} J(F_0(y))\ (F_0(y) - \delta_x(y))\psi'(y)\ dy$$

$$= \int_0^x J(F_0(y))\, \psi'(y)\, dy + \int_{-\infty}^0 J(F_0(y))\, F_0(y)\, \psi'(y)\, dy$$

$$+ \int_0^\infty J(F_0(y))\, [F_0(y)-1]\, \psi'(y)\, dy$$

if $y = -u$ in the second term, one obtains

$$\int_0^\infty J(F_0(-u))\, F_0(-u)\, \psi'(-u)\, du$$

using (3.3) and the fact that

$$F_0(-u) = 1-F_0(u) \ ,$$

$$\int_{-\infty}^\infty J(F_0(y))\, (F_0(y) - \delta_x(y))\, \psi'(y)\, dy =$$

$$\int_0^x J(F(y))\, \psi'(y)\, dy \ .$$

Therefore if μ_j is a BLLP:

$$(3.4) \quad \lim_{\varepsilon \to 0} \frac{\mu_J((1-\varepsilon)\, F_0+\varepsilon\delta_x)}{\varepsilon} = \frac{\int_0^x J(F_0(y))\, \psi'(y)\, dy}{\int_{-\infty}^\infty J(F_0(y))\, \psi'(y)\, dF_0(y)} \ .$$

Note that without loss of generality, one may suppose:

$$0 < \int_{1/2-\eta}^{1/2} J_0(s)\, ds < \frac{1}{2} \ .$$

Let:

$$J_1(t) = \begin{cases} (1-2\eta)^{-1}(1-2\int_{1/2-\eta}^{1/2} J_0(t)\, dt) & t\varepsilon[0,1/2-\eta] \\ J_0(t) & t\varepsilon(1/2-\eta,1/2] \\ J_1(1-t) & t\varepsilon(1/2,1] \end{cases}$$

and $K_1(t) = \int_0^t J_1(s)\, ds$.

Thus K_1 is an absolutely continuous distribution on [0,1] having non null density in [0,1]. Hence $K_1^{-1}(t)$ is an absolutely continuous distribution function in [0,1].

For any set of distributions $\mathcal{F}$, μ_{J_0} is a BLLP in $\mathcal{F}_{K_1^{-1}} = \{K_1^{-1}(F) : F\varepsilon\mathcal{F}\}$. Using lemma III.1 ,

$\mu_{J_2}(F) = \mu_{J_0}(K_1^{-1}(F))$ is a BLLP in $\mathcal{F}$, $\mu_{J_2}(F)$ is solution of:

$$\int_0^1 J_0(t)\ \psi(F^{-1}(K_1(t)) - \theta)\ dt = 0$$

or

$$\int_0^1 \frac{J_0(K_1^{-1}(s))}{J_1(K_1^{-1}(s))}\ \psi(F^{-1}(s) - \theta)\ ds = 0$$

and μ_2 is an L-M location parameter based on

$$J_2(s) = \frac{J_0(K_1^{-1}(s))}{J_1(K_1^{-1}(s))} \quad \text{and } \psi .$$

Note that J_2 is a bounded variation function since

$$J_1(t) \geq \min \left\{\eta\ ,\ \frac{1-2\int_{1/2-\eta}^{1/2} J_0(s)\, ds}{1-2\eta}\right\} \text{ for } t\varepsilon\ [0,1] .$$

Using (3.4):

$$\lim_{\varepsilon \to 0} \frac{\mu_{J_2}((1-\varepsilon)F_0+\varepsilon\delta_x)}{\varepsilon} = \frac{\int_0^x J_2(F_0(y))\ \psi'(y)\ dy}{\int_{-\infty}^{\infty} J_2(F_0(y))\ \psi'(y)\ dF_0(y)} ,$$

repeating a similar argument with:

$[(1-\varepsilon)\ F_0+\varepsilon\delta_x]_\sigma$, one obtains:

$$\lim_{\varepsilon \to 0} \frac{\mu_{J_2}\{[(1-\varepsilon)\ F_0+\varepsilon\delta_x]_\sigma\}}{\varepsilon} = \frac{\int_0^{x\sigma} J_2(F_0(\frac{y}{\sigma}))\ \psi'(y)\ dy}{\int_{-\infty}^{\infty} J_2(F_0(\frac{y}{\sigma}))\ \psi'(y)\ dF_0(\frac{y}{\sigma})}$$

$$= \frac{\sigma\int_0^{x} J_2(F_0(y))\ \psi'(\sigma y)\ dy}{\int_{-\infty}^{\infty} J_2(F_0(\frac{y}{\sigma}))\ \psi'(y)\ dF_0(\frac{y}{\sigma})} .$$

Since μ_{J_2} is a BLLP :

$$\mu_{J_2}\{[(1-\varepsilon)\ F_0+\varepsilon\delta_x]_\sigma\} = \sigma\mu_{J_2}\{(1-\varepsilon)F_0+\varepsilon\delta_x\}$$

and:

$$\frac{\int_0^x J_2(F_0(y))\ \psi'(y)\ dy}{\int_{-\infty}^{\infty} J_2(F_0(y))\ \psi'(y)dF_0(y)} = \frac{\int_0^x J_2(F_0(y))\ \psi'(\sigma y)\ dy}{\int_{-\infty}^{\infty} J_2(F_0(\frac{y}{\sigma}))\ \psi'(y)\ dF_0(\frac{y}{\sigma})}$$

differentiating with respect to x :

$$\frac{J_2(F_0(x))\ \psi'(x)}{\int_{-\infty}^{\infty} J_2(F_0(y))\ \psi'(y)\ dF_0(y)} = \frac{J_2(F_0(x))\ \psi'(\sigma x)}{\int_{-\infty}^{\infty} J_2(F_0(\frac{y}{\sigma}))\ \psi'(y)\ dF_0(\frac{y}{\sigma})} .$$

Therefore:

$$(3.5)\quad \frac{\int_0^x J_2(F_0(y))\ \psi'(\sigma y)\ dy}{\psi'(\sigma x)} = \frac{\int_0^x J_2(F_0(y))\ \psi'(y)\ dy}{\psi'(x)} .$$

Note $J_2(t) = 1 \quad t \in [K_1(\frac{1}{2} - \eta) \, , \, \frac{1}{2}]$ and K_1 is strictly increasing symmetric with respect to 1/2 so that

$$1/2 > K_1(\tfrac{1}{2} - \eta) \; .$$

For $\quad x \in [F_0^{-1}(K_1(\frac{1}{2} - \eta)) \, , \, 0]$

(3.5) is equivalent to:

$$\frac{\psi(\sigma x) - \psi(0)}{\sigma\psi'(\sigma x)} = \frac{\psi(x) - \psi(0)}{\psi'(x)} \; .$$

For a fixed x , integrating with respect to σ leads

$$\log\,(-\psi(\sigma x) + \psi(0)) = \alpha \log \sigma + C \text{ where } \alpha > 0$$

or

$$\psi(\sigma x) = \psi(0) - C_1\sigma^\alpha \text{ where } \alpha > 0 \; , \; C_1 > 0$$

so that

$$\psi(-\sigma) = \psi(0) - C_2\sigma^\alpha \text{ for } \sigma > 0 \; ,$$

repeating a similar argument with $\omega = -\sigma$ leads:

$$\psi(\sigma) = \psi(0) + C_2\sigma^\alpha \text{ for } \sigma > 0 \; ,$$

and

$$\psi(x) = \psi(0) + k\,\mathrm{sgn}(x)\,|x|^\alpha \; .$$

Note that $\psi'(x) = \psi'(-x)$ so that (3.3) implies:

$$J_0(F_0(x)) = J_0(F_0(-x)) , \quad x \in R$$

$$= J_0(1-F_0(x))$$

or: $\quad J_0(t) = J_0(1-t)$ since F_0 is absolutely continuous.

Since $\mu_{J_0}(F_0) = 0$

$$\int_0^1 J_0(t)\ \mathrm{sgn}\ (F_0^{-1}(t))|F_0^{-1}(t)|^{\alpha}\ dt$$

$$+\ \int_0^1 J_0(t)\ \psi(0)\ dt = 0 \quad .$$

Using the symmetry properties the first integral is null so that:

$$\psi(0) \int_0^1 J_0(t)\ dt = 0$$

and $\psi(0) = 0$. Thus the theorem is proved.

This theorem is a generalization of Bickel and Lehmann's (1975) theorem 2.

Theorem III.2

If $\psi(x) = |x|^{\alpha}\ \mathrm{sgn}(x)\ \alpha > 0$ and $J(t) = J(1-t)$ for all t in [0,1], the L-M location parameter μ based on ψ and J defined on any set of distributions $\mathcal{F}$ is a BLLP.

The proof of this theorem is an easy consequence of lemma III.1 and of the following:

Lemma III.2

If $\psi(x) = c\ |x|^{\alpha}\ \mathrm{sgn}(x),\ \alpha > 0\ \mu_H(F)$, the L-M location parameter based on ψ and $J(t) = 1\ t\varepsilon[0,1]$ and defined on any set of distributions is a BLLP .

Proof:

i) Let F_X and $F_Y \varepsilon \mathcal{F}$

$$X > Y \text{ s.} \rightarrow F_Y(x) \geq F_X(x) \quad x \in R .$$

Consider: $$F_X^{-1}(t) = \inf \{x : F_X(x) \geq t\}$$

$$\geq \inf \{x : F_Y(x) \geq t\}$$

$$= F_Y^{-1}(t) \text{ for all } t \in (0,1) .$$

So that:

$$|F_X^{-1}(t) - \theta|^\alpha \geq |F_Y^{-1}(t) - \theta|^\alpha \text{ if } F_Y^{-1}(t) \geq \theta$$

and $$|F_X^{-1}(t) - \theta|^\alpha \leq |F_Y^{-1}(t) - \theta|^\alpha \text{ if } \theta \geq F_X^{-1}(t) .$$

Consider:

$$\int_0^{F_X(\theta)} -|F_Y^{-1}(t) - \theta|^\alpha \, dt - \int_{F_X(\theta)}^{F_Y(\theta)} |F_Y^{-1}(t) - \theta|^\alpha \, dt + \int_{F_Y(\theta)}^1 |F_Y^{-1}(t) - \theta|^\alpha \, dt$$

$$\leq \int_0^{F_X(\theta)} - |F_X^{-1}(t) - \theta|^\alpha \, dt + \int_{F_X(\theta)}^{F_Y(\theta)} |F_X^{-1}(t) - \theta|^\alpha \, dt + \int_{F_Y(\theta)}^1 |F_X^{-1}(t) - \theta|^\alpha \, dt$$

$$= \int_0^1 \psi \, (F_X^{-1}(t) - \theta) \, dt .$$

Therefore:

$$\int_0^1 \psi \, (F_X^{-1}(t) - \mu_H(Y)) \, dt \geq 0$$

which implies:

$\mu_H(X) \geq \mu_H(Y)$ since $\int_0^1 \psi \, (F^{-1}(t) - \theta) \, dt$ is a decreasing function of θ .

ii) if $Y = -X$ s. then $F_Y(x) = 1-F_X(-x)$ a.e.

$\mu(Y)$ is a solution of:

$$\int_{-\infty}^{\infty} \psi(x-\theta)\ dF_Y(x) = 0$$

or
$$\int_{-\infty}^{\infty} \psi(x-\theta)\ d(1-F_X(-x)) = 0$$

since $F_Y(x) = 1-F_X(-x)$ a.e. and ψ is a continuous function.

This implies $\mu_H(Y) = -\ \mu_H(X)$ because ψ is odd.

iii) if $Y = aX + b$ s. $a > 0$ $b \varepsilon R$ $\mu(Y)$ is solution of:

$$\int_{-\infty}^{\infty} \psi\ (x-\theta)\ dF_Y(x) = 0$$

or
$$\int_{-\infty}^{\infty} \psi\ (x-\theta)\ dF_X(\frac{x-a}{b}) = 0$$

or
$$\int_{-\infty}^{\infty} \psi(ay + b-\theta)\ dF_X(y) = 0$$

or
$$\int_{-\infty}^{\infty} \psi(a(y + \frac{b-\theta}{a}))\ dF_X(y) = 0$$

since $\psi(ax) = \psi(a)\ \psi(x)$ if $a > 0$,

$\mu_H(Y) = a\ \mu_H(X) + b$.

This ends the proof.

<u>Remark</u>: 1) The preceding theorem shows that there are two ways to compute an L-M location parameter for a distribution F:

i) as the L-M location parameter based on J and ψ for the distribution F .

ii) as the M location parameter based on ψ for the distribution K(F) where $K(t) = \int_0^t J(s)\ ds$.

2) Note that if $J(t) = 0$, $t \notin [\delta, 1-\delta]$ where $\delta \varepsilon (0,1/2)$, $P\{X_K \varepsilon [F^{-1}(\delta), F^{-1}(1-\delta)]\} = 1$ where X_K is a random variable having distribution $K(F)$. Using (3.2),

$$\mu(F) = \mu_H(K(F)) \varepsilon \ [F^{-1}(\delta) \ , \ F^{-1}(1-\delta)] \ .$$

3) Let $\psi(x) = x$, $x \varepsilon R$,

$$J(t) = \begin{cases} 1/(1-2\delta) & t \varepsilon [\delta, 1-\delta] \\ 0 & \text{elsewhere} \end{cases} .$$

Hence,

$$K(F) = \begin{cases} 0 & x < F^{-1}(\delta) \\ \dfrac{F(x) - \delta}{1-2\delta} & x \varepsilon [F^{-1}(\delta) \ , \ F^{-1}(1-\delta)] \\ 1 & x > F^{-1}(1-\delta) \end{cases}$$

is the δ-trimmed version of F.

The L-M location parameter based on ψ and J is the δ-trimmed mean.

The corresponding M estimator is the mean of the δ-trimmed distribution.

Yet, note that even if estimating the δ-trimmed mean from a random sample of a distribution F and estimating the mean from a random sample of a δ-trimmed version of F , one estimates the same parameter, the two estimators have different asymptotic behaviour (assume F is symmetric):

i) when estimating the trimmed mean of F :

$$\mathcal{L}(n^{1/2}(\overline{X}_\delta - \mu(F))) \xrightarrow[n \to \infty]{} N(0,\sigma_1^2)$$

where $\sigma_1^2 = (1-2\delta)^2 \; [\int_\delta^{1-\delta} F^{-1}(t)^2 dt + \delta(F^{-1}(\delta)^2 + F^{-1}(1-\delta)^2)]$

ii) and when estimating the mean of δ-trimmed F:

$$\mathcal{L}(n^{1/2}(\overline{X}_K - \mu(F)) \xrightarrow[n \to \infty]{} N(0,\sigma_2^2)$$

where $\sigma_2^2 = (1-2\delta)^{-2} \int_\delta^{1-\delta} F^{-1}(t)^2 \, dt$.

Section III.2 Robustness

In this section, two concepts of robustness are investigated.

The first one, absolute robustness, is a property that a functional does or does not possess. Loosely speaking a functional μ is robust if a small variation in F produces a small variation in $\mu(F)$. (See Hampel 1971).

The second one, relative robustness, is a tool to compare two functionals. Loosely speaking μ_1 is more robust than μ_2 if given that a small variation in F produces a small variation in $\mu_2(F)$, the pertubation caused to $\mu_1(F)$ will also be small. This concept has been first discussed by Bickel and Lehmann (1975).

Definition III.4

A functional μ is said to be robust at F_0 if μ is continuous at F_0 with respect to the Lévy (or Prohorov) distance.

A functional μ is said to be robust in a set $\mathcal{F}$ if μ is continuous at every F_0 in $\mathcal{F}$.

Throughout this section, the following characterization of the weak convergence will be used:

Theorem III.3

Let $d(\cdot,\cdot)$ denotes the Lévy distance and $\{F_n\}_{n=1}^{\infty}$ be a sequence of distribution functions:

$$\lim_n d(F_n,F) = 0 \leftrightarrow F_n^{-1}(t) \xrightarrow[n\to\infty]{} F^{-1}(t) \text{ a.e. } .$$

Proof: Let first prove $\rightarrow$.

By definition (see Chung (1974) p. 94):

$$d(F_n,F) = \inf \{\varepsilon : F_n(x-\varepsilon) - \varepsilon \le F(x) \le F_n(x+\varepsilon) + \varepsilon \text{ for all } x\varepsilon R\} .$$

Let $t\varepsilon(0,1)$ be a continuity point of F^{-1} and choose $\delta > 0$ such that $\delta < \frac{1}{2} \min (t,1-t)$. Using the assumption there exists $n_\delta \varepsilon N$ such that:

$$n > n_\delta \rightarrow d(F_n,F) < \delta \quad \text{so that:}$$

$$F(x-\delta) - \delta \le F_n(x) \le F(x+\delta) + \delta \cdot$$

Taking $x=F_n^{-1}(t)$:

$$F(F_n^{-1}(t)+\delta) \ge F_n(F_n^{-1}(t)) - \delta$$

$$\ge t-\delta \text{ since } F_n \text{ is right continuous.}$$

Hence: $F_n^{-1}(t)+\delta \ge F^{-1}(t-\delta)$ since

$$F^{-1}(t-\delta) = \inf \{x : F(x) \ge t-\delta\} .$$

Using the symmetry of the Lévy distance yields:

$$F^{-1}(t+\delta) + \delta \geq F_n^{-1}(t) \geq F^{-1}(t-\delta) - \delta \text{ or}$$

$$F^{-1}(t+\delta) - F^{-1}(t) + \delta \geq F_n^{-1}(t) - F^{-1}(t) \geq F^{-1}(t-\delta) - F^{-1}(t) - \delta .$$

Since F^{-1} is continuous at t:

$$\lim_n F_n^{-1}(t) = F^{-1}(t) ,$$

F^{-1} is continuous a.e. being an increasing function, so that

$$F_n^{-1}(t) \xrightarrow[n \to \infty]{} F^{-1}(t) \text{ a.e.} \quad .$$

To prove the converse, the following criterion for weak convergence will be used (see Chung (1974) p. 87)

$$\lim_n d(F_n,F) = 0 \quad \leftrightarrow$$

(3.6) $$\psi\varepsilon C_K \lim_n \int_R \psi(x)\, dF_n(x) = \int_R \psi(x)\, dF(x) ,$$

where C_K is the set of continuous functions with compact support.

Let X_n be a random variable having distribution F_n , and T be a random variable having the $U[0,1]$ distribution. Using lemma II.1, X_n and $F_n^{-1}(T)$ have the same distribution and:

$$E(\psi(X_n)) = \int_R \psi(x)\, dF_n(x) = \int_{[0,1]} \psi(F_n^{-1}(t))\, dt .$$

Let $\psi\varepsilon C_k$. There exists $M\varepsilon N$ such that $|\psi(x)| < M$, $x\varepsilon R$ and

$$\lim_n \psi(F_n^{-1}(t)) = \psi(F^{-1}(t)) \text{ a.e.}$$

so that the Lebesgue dominated convergence theorem may be used and:

$$\lim_n \int_{[0,1]} \psi(F_n^{-1}(t))\, dt = \int_{[0,1]} \psi(F^{-1}(t))\, dt$$

$$= \int_R \psi(x)\, dF(x) ,$$

hence, using the criterion (3.6), $\lim_n d(F_n,F) = 0$ and the theorem is proved.

In this section, robustness properties of L-M location parameters which are BLLP will be investigated, therefore assume:

- $J(t) = J(1-t) \quad t\varepsilon[0,1]$

- $\psi(x) = c \operatorname{sgn} x\, |x|^\alpha \quad c\varepsilon R^+ \quad \alpha > 0 .$

Theorem III.5

An L-M location parameter is robust in any set $\mathcal{F}$ if and only if:

(3.7) $\qquad J(t) = 0 \quad t\varepsilon[\delta,1-\delta]$ where $\delta\varepsilon(0,1/2)$.

Proof: Suppose that there does not exist any δ such that (3.7) is satisfied, let

$$a_n = \int_{1-1/n}^{1} J(t)\, dt , \quad a_n > 0 \quad n\varepsilon N .$$

Suppose

$$\psi(x) = \operatorname{sgn} x\, |x|^\alpha \quad \alpha > 0 ,$$

let:

$$F(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$F_n(x) = \begin{cases} 0 & x < 0 \\ 1-{}^1/n & 0 \leq x < (\frac{1}{a_n})^{1/\alpha} \\ 1 & x \geq (\frac{1}{a_n})^{1/\alpha} \end{cases} .$$

Note $\mu(F) = 0$,

$$\lim_n d(F_n, F) = 0 \quad \text{and}$$

$$F_n^{-1}(t) = \begin{cases} 0 & 0 < t \leq 1 - \frac{1}{n} \\ (\frac{1}{a_n})^{1/\alpha} & 1 - \frac{1}{n} < t < 1 \end{cases} .$$

$\mu(F_n)$ is solution of:

$$\int_0^1 J(t)\ \psi(F_n^{-1}(t) - \theta)\ dt = 0 .$$

Note that if X_n is a random variable having distribution F_n :

$$P(X_n\ \varepsilon [0, (\frac{1}{a_n})^\alpha]) = 1 \quad \text{so that:}$$

the solution of the last equation is in $[0, (\frac{1}{a_n})^\alpha]$, therefore $\psi(-\theta) = -\theta^\alpha$ and

$$\psi\left(\left(\frac{1}{a_n}\right)^{1/\alpha}-\theta\right) = \left(\left(\frac{1}{a_n}\right)^{1/\alpha}-\theta\right)^{\alpha} .$$

Hence $\mu(F_n)$ is solution of:

$$a_n\left[\left(\frac{1}{a_n}\right)^{1/\alpha}-\theta\right]^{\alpha} - (1-a_n)\theta^{\alpha} = 0$$

$$\leftrightarrow \quad a_n\left(\left(\frac{1}{a_n}\right)^{1/\alpha}\frac{1}{\theta} - 1\right)^{\alpha} - (1-a_n) = 0$$

$$\leftrightarrow \quad \left(\frac{1}{a_n}\right)^{1/\alpha}\frac{1}{\theta} - 1 = \left(\frac{1-a_n}{a_n}\right)^{1/\alpha} \quad \text{and}$$

$$\mu(F_n) = \frac{\left(\frac{1}{a_n}\right)^{1/\alpha}}{\left(\frac{1-a_n}{a_n}\right)^{1/\alpha} + 1}$$

$$= [(1-a_n)^{1/\alpha} + a_n^{1/\alpha}]^{-1} .$$

As $n \to \infty$, $a_n \to 0$ and

$$\lim_n \mu(F_n) = 1 .$$

So that μ is not robust at F .

Conversely, let $\{F_n\}_{n=1}^{\infty}$ be a sequence of distribution functions converging weakly to F .

Without loss of generality assume δ is a continuity point of F .

Using theorem III.4, $F_n^{-1}(1-\delta)$ and $F_n^{-1}(\delta)$

converge to $F^{-1}(1-\delta)$ and $F^{-1}(\delta)$ respectively, so that there exist M and n_1 in N such that

$$n > n_1 \to \max \{|F_n^{-1}(1-\delta)| \ , \ |F_n^{-1}(\delta)|\} < M \ .$$

Using the remark at the end of section III.1 ,

$$n > n_1 \to |\mu(F_n)| < M \ .$$

Therefore there exists a convergent subsequence, say $\{n(k)\}_{k=1}^{\infty}$ satisfying

$$\mu(F_{n(k)}) \xrightarrow[k \to \infty]{} \gamma \ .$$

Consider

$$\int_0^1 J(t) \ \psi(F_{n(k)}^{-1}(t) - \mu(F_{n(k)})) \ dt \ .$$

There exists $n_2 > n_1$ such that for $t \varepsilon [\delta, 1-\delta]$ and $n(k) > n_2$

$$|\psi(F_{n(k)}^{-1}(t) - \mu(F_{n(k)}))| < \psi(M+2|\gamma|)$$

and using dominated convergence

$\int_0^1 J(t) \ \psi(F_{n(k)}^{-1}(t) - \mu(F_{n(k)})) \ dt$ converges to

$\int_0^1 J(t) \ \psi(F^{-1}(t) - \gamma) \ dt$,

and

$$\mu(F) = \gamma \ .$$

Therefore all converging subsequences converge to $\mu(F)$. This implies:

$$\lim_{n} \mu(F_n) = \mu(F)$$

and the theorem is proved.

Intuitively μ_1 is more robust than μ_2 if $\{\mu_2(F_n)\}$ converges to $\mu_2(F)$ implies $\{\mu_1(F_n)\}$ converges to $\mu_1(F)$. As exemplified in the next page $\{\mu(F_n)\}$ may converge to $\mu(F)$ fortuitously, take α in $(0,1)$ and let:

$$P(X_n = x) = \begin{cases} (1/n)^{1+\alpha} & x = -n^2 \\ 1 - (1/n)^{\alpha} - (1/n)^{1+\alpha} & x = 0 \\ (1/n)^{\alpha} & x = n \; . \end{cases}$$

The distribution F_n of X_n tends to:

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

as $n \to \infty$.

$\mu_\alpha(F_n)$, the M location parameter based on $\psi(x) = \operatorname{sgn} x \, |x|^\alpha$, is the solution of

$$\int_{-\infty}^{\infty} \psi(x-\theta) \, dF_n(x) = 0 \quad \text{or}$$

$$-(1/n)^{1+\alpha} (n^2+\theta)^\alpha + (1-(1/n)^{1+\alpha} - (1/n)^\alpha) \, \psi(-\theta) + (1/n)^\alpha (n-\theta)^\alpha = 0 \; .$$

We want to solve the following equation:

$$(3.8) \quad (1-(1/n)^\alpha - (1/n)^{1+\alpha}) \, \psi(\theta) - (1-\theta/n)^\alpha + (1/n)^{1-\alpha}(1-\theta/n^2)^\alpha = 0$$

for $\theta = 0$, the LHS equals:

$$-1 + (^1/n)^{1-\alpha} < 0 \quad \text{for } n > 1 .$$

for $\theta = 2$, the LHS is bigger or equal than:

$$\frac{1}{2^\alpha} \cdot 4^\alpha - (1-2/n)^\alpha + (1/n)^{1-\alpha}(1-2/n^2) \geq 2^\alpha - 1 > 0$$

for n big enough. So that

$$\mu_\alpha(F_n) \in [0,2] \text{ if } n \text{ is big}$$

and as $n \to \infty$, equation (3.8) becomes:

$$\psi(\theta) - 1 = 0 , \text{ i.e. } \quad \theta = 1 ,$$

and $\lim_n \mu_\alpha (F_n) = 1$

while $\lim_n \mu_1(F_n) = 0$.

Hence $\{\mu_\alpha (F_n)\}$ does not converge to $\mu_\alpha(F)$ while $\{\mu_1(F_n)\}$ does.

Therefore a "naive" definition of relative robustness would lead to noncomparability for μ_α and μ_1 when intuitively μ_α is more robust than μ_1 for $\alpha < 1$.

The following approach is suggested. If one is able to find a reasonable condition, $C(\mu)$ say, on a sequence $\{F_n\}_{n=1}^\infty$ of distribution functions such that:

$\{F_n\}$ fulfills $C(\mu)$ implies $\{\mu(F_n)\}$ converges.

Then define: μ_1 is more robust than μ_2 if $\{F_n\}$ fulfills $C(\mu_2)$ implies $\{F_n\}$ fulfills $C(\mu_1)$.

If $\mu(X) = E(X)$, such a condition is known: if the sequence $\{F_n\}$ is uniformly integrable then

$$\{\mu(F_n)\} \text{ converges.}$$

(see Billingsley (1968) p. 32).

<u>Definition III.5</u> <u>Uniform integrability</u>

A sequence $\{F_n\}$ of distribution functions is said to be uniformly integrable if

$$\sup_n \int_{\{|F_n^{-1}(t)| > M\}} |F_n^{-1}(t)| \, dt \xrightarrow[M \to \infty]{} 0 \ .$$

This is the same as:

$$\lim_{M\to\infty} - \mu(\underline{F}_n(\cdot,M)) + \mu(\overline{F}_n(\cdot,M)) = 0$$

uniformly in n , where:

$$\overline{F}(x,M) = \begin{cases} 0 & x < 0 \\ F(M) & 0 \le x \le M \\ F(x) & x > M \end{cases}$$

$$\underline{F}(x,M) = \begin{cases} F(x) & x \le -M \\ F(-M) & -M < x < 0 \\ 1 & x \ge 0 \end{cases} \ .$$

This suggests the following:

Definition III.6 μ uniform integrability

A sequence $\{F_n\}_{n=1}^{\infty}$ of distribution functions is said to be μ uniformly integrable if:

i) $\{F_n\}$ converges weakly

ii) $\lim_{M \to \infty} \mu(\overline{F}_n(\cdot,M)) - \mu(\underline{F}_n(\cdot,M)) = 0$ uniformly in n .

In the general case, the conditions on μ are too weak to prove $\{F_n\}$ is μ uniformly integrable implies $\{\mu(F_n)\}$ converges. We shall therefore restrict ourselves to the L-M case.

Let μ be an L-M location parameter based on J and ψ. If μ is robust, $J(t) = 0 \quad t\varepsilon[\delta,1-\delta]$ for a given $\delta\varepsilon(0,{}^1/2)$ so that if $M > \max\{|F^{-1}(\delta/2)|, F^{-1}(1-\delta/2)\}$ $\mu(\overline{F}_n(\cdot,M)) = \mu(\underline{F}_n,M) = 0$ for n big enough and every sequence $\{F_n\}$ converging weakly is μ uniformly integrable.

Theorem III.4 Let μ be an L-M location parameter based on J and ψ , if $\{F_n\}_{n=1}^{\infty}$ is a μ uniformly integrable sequence converging to F then $\{\mu(F_n)\}_{n=1}^{\infty}$ converges to $\mu(F)$.

Proof: Let $K(t) = \int_0^t J(s)\, ds$.

Notice that:

$$K(\overline{F}(\cdot,M)) = \overline{K(F)}\,(\cdot,M) \text{ and}$$

$$K(\underline{F}(\cdot,M)) = \underline{K(F)}\,(\cdot,M) .$$

Hence a sequence $\{F_n\}$ is μ uniformly integrable if and only if

$\{K(F_n)\}$ is μ_H uniformly integrable where μ_H is the M location parameter based on ψ . Therefore it suffices to prove the theorem with μ replaced by μ_H .

This will be an easy consequence of the following result:

$\{F_n\}$ is μ_H uniformly integrable if and only if:

$$(3.9) \qquad \lim_M \int_{\{|x|>M\}} |\psi(x)| dF_n(x) = 0$$

uniformly in n .

Suppose $\{F_n\}$ is μ_H uniformly integrable, $\mu_H(\overline{F}_n(\cdot,M)) = \theta_n$ satisfies:

$$(3.10) \qquad \int_M^\infty \psi(x-\theta_n)\, d\, F_n(x) = \psi(\theta_n)\, F_n(M) \; .$$

Using the μ_H uniform integrability of $\{F_n\}$ for any $\varepsilon > 0$ there exists $M_0 \varepsilon N$ such that:

$$M > M_0 \rightarrow |\theta_n| < \varepsilon \qquad n\varepsilon N \; .$$

There exists $M_1 = M_1(\psi)$, $M_1 > M_0$, such that:

$$x > M_1 \rightarrow \psi(x-\theta_n) > \psi(x-\varepsilon) > \frac{1}{2}\psi(x)$$

$n\varepsilon N$.

Hence (3.10) leads, for $M > M_1$,

$$\int_M^\infty \psi(x)\, d\, F_n(x) \leq 2\psi(\varepsilon) \qquad n\varepsilon N \; .$$

Using a similar argument with $\mu_H(\underline{F}_n(\cdot,M))$, the first half of statement (3.9) is proved.

To prove the converse, note that (3.9) is equivalent to:

$$\lim_M \int_{\{|x|>M\}} |\psi(x-\theta)| \, d\, F_n(x) = 0 \text{ uniformly in } n \text{ for any } \theta > 0 \text{ .}$$

This implies that for any continuity point M of F , M > 0 ,

$$\lim_n \int_R \psi(x-\theta) \; d\; \bar{F}_n(x,M) = \int_R \psi(x-\theta) \; d\; \bar{F}(x,M) \; .$$

Hence, for such M ,

(3.11) $$\lim_n \mu_H(\bar{F}_n(\cdot,M)) = \mu_H(\bar{F}(\cdot,M)) \; .$$

Using Billingsley's (1968) theorem 5.3 :

$$\liminf_n \int_{\{|x|>M\}} |\psi(x)| d\; F_n(x) \geq \int_{\{|x|>M\}} |\psi(x)| d\; F(x)$$

so that

$$\lim_M \int_{\{|x|>M\}} |\psi(x)| d\; F(x) = 0 \quad .$$

Now, $\theta=\mu_H(\bar{F}(\cdot,M))$ satisfies:

$$\int_M^\infty \psi(x-\theta) \; d\; F(x) = \psi(\theta) \; F(M)$$

or, since $\theta \geq 0$,

$$\frac{\int_M^\infty \psi(x) \; d\; F(x)}{F(M)} > \psi(\theta)$$

and $\lim_M \mu_H(\bar{F}(\cdot,M)) = 0$.

Using that last result and (3.11), there exist $M_2=M_2(\varepsilon)$ and $n_\varepsilon=n_\varepsilon(M_2)$ in N such that:

$n > n_\varepsilon$, $M > M_2$ imply:

$$\mu_H(\overline{F}_n(\cdot,M)) < \varepsilon .$$

And one concludes:

$\lim_M \mu_H(\overline{F}_n(\cdot,M)) = 0$ uniformly for $n\varepsilon N$. Using a similar argument with $\mu_H(\underline{F}(\cdot,M))$ proves statement (3.9).

Note that statement (3.9) may be generalized to the L-M case:

Corollary III.1

$\{F_n\}_{n=1}^{\infty}$ is μ uniformly integrable if and only if:

$$\lim_M \int_{\{t:|F_n^{-1}(t)|>M\}} J(t)|\psi(F_n^{-1}(t))|dt = 0$$

uniformly in n .

Proof: $\{F_n\}$ is μ uniformly integrable if and only if $K(F_n)$ is μ_H uniformly integrable, i.e. if and only if:

$$\lim_M \int_{\{|x|>M\}}|\psi(x)|dK(F_n(x)) = 0$$

uniformly for $n\varepsilon N$.

Let T be a random variable defined on [0,1] having distribution K ; $F_n^{-1}(T)$ has distribution $K(F_n)$ so that

$$\int_{\{|x|>M\}}|\psi(x)|\, dK(F_n(x)) =$$

$$\int_{\{|F_n^{-1}(t)|>M\}} J(t)|\psi(F_n^{-1}(t))|\, dt ,$$

and the corollary is proved.

Theorem III.4 insures us that as far as L-M location parameters are concerned, the following definition is consistent.

Definition III.7

Let μ_1 and μ_2 be two location parameters, μ_1 is said to be more robust than μ_2 if $\{F_n\}_{n=1}^{\infty}$ is μ_2 uniformly integrable implies $\{F_n\}_{n=1}^{\infty}$ is also μ_1 integrable.

Note that if μ is a robust L-M location parameter, μ is more robust than any L-M location parameter using the remark after definition III.6 .

Theorem III.5

Let μ_1 and μ_2 be two L-M location parameters based on J_1 and ψ_1 , J_2 and ψ_2 respectively (note that $\psi_i(x) = \text{sgn } x \; |x|^{\alpha_i}$, $i = 1,2$) . μ_1 is more robust than μ_2 if $\alpha_2 \geq \alpha_1$ and if there exists ε in $(0, {}^1/2)$ such that:

$$K_2(t) \geq K_1(t) \qquad t\varepsilon[0,\varepsilon]$$

$$K_2(t) \leq K_1(t) \qquad t\varepsilon[1-\varepsilon,1]$$

where $K_i(t) = \int_0^t J_i(s)\, ds \quad i=1,2$.

Proof:

Let $\{F_n\}$ be a μ_2 integrable sequence converging weakly to F and consider:

$$\int_{\{|F_n^{-1}(t)| > M\}} |\psi_1(F_n^{-1}(t)|\; dK_1(t) \;.$$

Since $\alpha_2 \geq \alpha_1$ the last expression is less than or equal to:

$$\int_{\{|F_n^{-1}(t)|>M\}} \psi_2(F_n^{-1}(t))\, dK_1(t) \ .$$

Since $\lim_n F_n^{-1}(t) = F^{-1}(t)$ a.e.,

there exists n_ε such that:

$n > n_\varepsilon \rightarrow \{t: F_n^{-1}(t) > |F^{-1}(1-\varepsilon/2)|\}$ is an interval included in $(1-\varepsilon,1]$; hence for $n > n_\varepsilon$

$$\int_{\{F_n^{-1}(t) > |F^{-1}(1-\varepsilon/2)|\}} \psi_2 (F_n^{-1}(t))\, dK_1(t) \leq \int_{\{F_n^{-1}(t) > |F^{-1}(1-\varepsilon/2)|\}} \psi_2(F_n^{-1}(t))dK_2(t)$$

because $K_1(t) \geq K_2(t)$ $t\varepsilon[1-\varepsilon,1]$ and $\psi_2(F_n^{-1}(t))$ is increasing.

Using a similar argument for t near 0, a number M in N can be found such that:

$n\varepsilon N$

$$\int_{\{|F_n^{-1}(t)| > M\}} |\psi_2(F_n^{-1}(t))|dK_1(t) \leq \int_{\{|F_n^{-1}(t)| > M\}} |\psi_2(F_n^{-1}(t))|dK_2(t)$$

so that $\{F_n\}$ μ_2 uniformly integrable implies $\{F_n\}$ μ_1 uniformly integrable.

To end this section, we will prove a partial converse of theorem III.4 .

Theorem III.6

Let $\{F_n\}_{n=1}^{\infty}$ be a sequence of distribution functions converging weakly to F such that:

$$F_n(0^-) = \lim_{\substack{x \to 0 \\ x < 0}} F_n(x) = 0 \quad \text{for any } n \in N$$

$\{\mu(F_n)\}_{n=1}^{\infty}$ converges to $\mu(F)$ if and only if $\{F_n\}_{n=1}^{\infty}$ is μ uniformly integrable.

Proof: The sufficiency of the assumption "$\{F_n\}$ is μ uniformly integrable" is a consequence of theorem III.4 . We only need to show its necessity.

Since $F_n(0^-) = 0$ for any n

$$\int_{\{F_n^{-1}(t) < -M\}} J(t)\, \psi(F_n^{-1}(t))\, dt = 0 \quad \text{for all } n \text{ in } N \text{ and } M > 0$$

and the following has to be proved:

$$\mu(F_n) \xrightarrow[n \to \infty]{} \mu(F) \text{ implies}$$

(3.12) $$\lim_{M} \int_{\{F_n^{-1}(t) > M\}} J(t)\, \psi(F_n^{-1}(t))\, dt = 0 \text{ uniformly}$$

for $n \in N$.

Let $\theta_n = \mu\,(F_n)$, $\theta = \mu(F)$, ε be a fixed positive number and choose

M_0 in the following way:

$$\lim_n \{t : F_n^{-1}(t) \leq M_0\} = \{t : F^{-1}(t) \leq M_0\}$$

- $\displaystyle\int_{\{F^{-1}(t) > M_0\}} J(t)\ \psi(F^{-1}(t) - \theta)\ dt < \varepsilon$ and $M_0 > \theta + \varepsilon$

- $x > M_0$ implies:

$$\frac{\psi(x-\theta-\varepsilon)}{\psi(x)} > \frac{1}{2} .$$

Using dominated convergence:

$$\lim_n \int_{\{F_n^{-1}(t) \leq M_0\}} J(t)\ \psi(F_n^{-1}(t) - \theta_n)\ dt = \int_{\{F^{-1}(t) \leq M_0\}} J(t)\ \psi(F^{-1}(t) - \theta)\ dt .$$

Therefore there exists $n_1 = n_1(\varepsilon, M)$, $n_1 \varepsilon N$ such that $n > n_1$ implies

$$\left| \int_{\{F_n^{-1}(t) \leq M_0\}} J(t)\ \psi(F_n^{-1}(t) - \theta_n)\ dt - \int_{\{F^{-1}(t) \leq M_0\}} J(t)\ \psi(F^{-1}(t) - \theta)\ dt \right| < \varepsilon$$

and $|\theta_n - \theta| < \varepsilon$,

so that for $n > n_1$ and $M > M_0$,

$$\int_{\{F_n^{-1}(t) > M\}} J(t)\ \psi(F_n^{-1}(t))dt < 2 \int_{\{F_n^{-1}(t) > M_0\}} J(t)\ \psi(F_n^{-1}(t) - \theta_n)\ dt$$

$$= 2 \{ \int_{\{F^{-1}(t) > M_0\}} J(t)\, \psi(F^{-1}(t) - \theta)\, dt$$

$$+ \int_{\{F^{-1}(t) \le M_0\}} J(t)\, \psi(F^{-1}(t) - \theta)\, dt \quad - \quad \int_{\{F_n^{-1}(t) \le M_0\}} J(t)\, \psi(F_n^{-1}(t) - \theta_n)\, dt\}$$

$$\le 4\, \varepsilon\, .$$

For each n $\varepsilon\{1,2,\ldots,n_1\}$, one can find M_n such that:

$$M > M_n \rightarrow \int_{\{F_n^{-1}(t) > M_n\}} J(t)\, \psi(F_n^{-1}(t))\, dt < 4\varepsilon\, ,$$

and (3.12) is proved.

Note that this theorem is a generalization of the second part of Billingsley's (1968) theorem 5.4, and theorem III.4 is a generalization of its first part.

Section III.3 Influence Curve

In this section the concept of influence as introduced by Hampel (1974) is discussed.

Let x_1, $x_2,\ldots,x_n$ be a set of observed values with empirical distribution F_n . The L-M estimate $\mu(F_n)$ has been computed where μ is any L-M location parameter. An (n+1)th observation x is added. Let F_{n+1}

be the empirical distribution corresponding to $x, x_1, x_2, \ldots, x_n$; the "influence of x on $\mu(F_{n+1})$" can be defined as:

$$\mu(F_{n+1}) - \mu(F_n) .$$

For instance, if μ is the mean, the influence of x is equal to:

$$(n+1)^{-1}(x-\mu(F_n)) ,$$

if μ is the δ-trimmed mean, the influence of x is equal to:

$$[(n+1)(1-2\delta)]^{-1}\begin{cases} x_{([\delta n]+1)} - \mu(F_n) & \text{if} \quad x \le x_{([\delta n]+1)} \\ x - \mu(F_n) & \text{if} \quad x_{([\delta n]+1)} < x \le x_{(n-[\delta n])} \\ x_{(n-[\delta n])} - \mu(F_n) & \text{if} \quad x_{(n-[\delta n])} < x \end{cases} .$$

The influence curve is the asymptotic counterpart of this notion; it is a function whose value at x is a measure of n times the "asymptotic influence" of x. This leads to the following formal definition:

Definition III.8 Influence Curve

The influence curve of a functional μ at a probability distribution F is defined as:

$$IC(F,\mu)(x) = \lim_{\varepsilon \to 0} \frac{\mu((1-\varepsilon)F+\varepsilon\delta_x) - \mu(F)}{\varepsilon}$$

where $\delta_x(y) = \begin{cases} 0 & \text{if} \quad y < x \\ 1 & \text{if} \quad x \le y \end{cases}$

Note that the IC (influence curve) of a functional μ is nothing more than its first Gateaux or Fréchet "derivative". Von Mises (1947) and Filippova

(1962) have introduced these notions in a probabilistic context.

If μ is the expectation,

$$IC\ (F,\mu)\ (x) = x - \mu(F)\ ,$$

if μ is the δ-trimmed expectation,

$$IC\ (F,\mu)\ (x) = (1-2\delta)^{-1} \begin{cases} F^{-1}(\delta) - \mu(F) & x \leq F^{-1}(\delta) \\ x - \mu(F) & F^{-1}(\delta) < x \leq F^{-1}(1-\delta) \\ F^{-1}(1-\delta) - \mu(F) & F^{-1}(1-\delta) < x \end{cases}$$

These IC are the limits of n times the function defining the influence of the (n+1)th observation for the mean and the δ-trimmed mean respectively.

More generally, if H is any distribution function,

$$\lim_{\varepsilon \to 0} \frac{\mu((1-\varepsilon)\ F+\varepsilon H) - \mu(F)}{\varepsilon} = \int IC\ (F,\mu)\ (x)\ dH(x)\ ;$$

$$\varepsilon \cdot \int IC\ (F,\mu)\ (x)\ dH(x)$$

measures the influence of the contaminant εH on μ (F) since:

$$\mu\ ((1-\varepsilon)\ F+\varepsilon H) \approx \mu(F) + \varepsilon \int IC\ (F,\mu)\ (x)\ dH(x)\ .$$

<u>Theorem III.7</u> <u>IC of M location parameters</u>

If μ is an M location parameter based on ψ assuming $\lambda(x) = \int_0^1 \psi(F^{-1}(t) - x)\ dt$ has a strictly negative derivative at $\mu(F)$,

$$IC\ (F,\mu)\ (x) = -\ \psi(x-\mu(F))/\lambda'(\mu(F))$$

at each point $x - \mu(F)$ where ψ is continuous.

Proof: Let

$$\lambda_\varepsilon(\theta) = (1-\varepsilon) \int_R \psi(y-\theta)\, dF(y) + \varepsilon\psi(x-\theta)$$

and θ_ε be the solution of $\lambda_\varepsilon(\theta) = 0$. Assume without loss of generality, $x > \mu(F)$. Therefore:

$$\lambda_\varepsilon(\mu(F)) > 0 \text{ and } \lambda_\varepsilon(x) < 0 \quad \text{or}$$

$$\theta_\varepsilon \varepsilon [\mu(F), x] \; .$$

So, assuming without loss of generality $\psi(0) = 0$,

$$0 \geq \int_R \psi(y-\theta_\varepsilon)\, dF(y) \geq -\, \varepsilon\psi(x-\mu(F))/(1-\varepsilon)$$

and as ε goes to 0 , $\lambda(\theta_\varepsilon)$ goes to 0 . Since λ has non null derivative at $\mu(F)$, λ^{-1} is continuous at 0 and

$$\lim_{\varepsilon \to 0} \theta_\varepsilon = \mu(F) \; .$$

Now:

$$(1-\varepsilon) \int_R \psi(y-\theta_\varepsilon)\, dF(y) + \varepsilon\psi(x-\theta_\varepsilon) = 0$$

or

$$(1-\varepsilon) \frac{\theta_\varepsilon - \mu(F)}{\varepsilon} \frac{\int_R \psi(y-\theta_\varepsilon)\, dF(y)}{\theta_\varepsilon - \mu(F)} = -\psi(x-\theta_\varepsilon)$$

and as $\varepsilon \to 0$,

$$IC\ (F,\mu)(x) = -\, \psi(x-\mu(F))/\lambda'(\mu(F))$$

assuming ψ is continuous at $x - \mu(F)$.

Remark

This theorem contains a proof of the existence of the IC and provides an algebraic expression for that IC. The previous theorems on the IC for an M location parameter (Andrews et al. (1972) p. 40, Hampel (1974), Huber (1977) p. 14) do not prove the existence.

The proof of the existence of the IC for a general L-M location parameter is not straightforward. The beginning of theorem III.1 contains a proof of the next result under very restrictive regularity conditions:

Theorem III.8 IC for L-M location parameter

Under certain regularity conditions, the IC of μ , an L-M location parameter based on J and ψ is equal to:

$$\text{IC }(F,\mu)\ (x) = \frac{-\int_0^{x-\mu(F)} J(F(y+\mu(F)))\ d\psi(y)}{\lambda'(\mu(F))} + C$$

where C is a constant such that

$$\int \text{IC }(F,\mu)\ (x)\ dF(x) = 0 \ .$$

A formal proof of this theorem would require intricate analysis. Furthermore this result has merely an intuitive interest. Therefore it will not be proved here.

Note that:

$$\text{IC }(F,\mu)\ (x) = -\psi_H(x-\xi_0)/\lambda'(\xi_0)$$

where ψ_H and ξ_0 has been defined in section II.5. So that using this new concept, theorem II.7 may be reformulated in the following way:

Theorem III.9

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution F , let

$$F_n(x) = (n+1)^{-1} \sum_{i=1}^{n} I\,(x-X_i)$$

where $I(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$.

Then for any L-M location parameter μ based on J and ψ satisfying the assumptions of Corollary II.8,

$$n^{1/2}\,(\mu(F_n) - \mu(F) - n^{1/2} \sum_{i=1}^{n} IC\,(F,\mu)\,(X_i)) \text{ is } o_p(1) .$$

So that $n^{1/2}\,(\hat{T}_n - \xi_0) = n^{1/2}\,(\mu(F_n) - \mu(F))$ has the same asymptotic behaviour as the sum of the influences of the X_i's on $\mu(F_n)$.

This result has been proved in a different context under a lot of messy regularity conditions by Filippova (1962). In the location model, it has been conjectured by Huber (1972). It legitimates the use of the IC as a tool in applied robust estimation. Coming back to BLLP,

Theorem III.10

If μ is a BLLP then:

$$IC\,(F_{aX+b}\,,\,\mu)\,(x) = a\;IC\,(F_X,\mu)(\tfrac{x-b}{a})$$

and

$$IC\ (F_{-X},\mu)\ (x) = -IC\ (F_X,\mu)\ (-x)$$

Proof: $\mu((1-\varepsilon)\ F_X(\frac{y-b}{a}) + \varepsilon\delta_x(y)) - a\mu(F_X) - b$

$$= \mu((1-\varepsilon)\ F_X(\frac{y-b}{a}) + \varepsilon\delta_{\frac{x-b}{a}}\ (\frac{y-b}{a})) - a\mu(F_X) - b$$

$$= a\ \mu((1-\varepsilon)\ F_X + \varepsilon\delta_{\frac{x-b}{a}}) - a\mu(F_X)\ ,$$

dividing by ε and taking the limit as ε tends to 0 proves the first statement.

To prove the second, consider:

$$\mu((1-\varepsilon)\ F_{-X} + \varepsilon\delta_x) - \mu(F_{-X})$$

$$= \mu((1-\varepsilon)\ (1-F_X(-y)) + \varepsilon(1-\delta_{-x}(-y))) + \mu(F_X)$$

$$= \mu(1 - (1-\varepsilon)\ F_X(-y) - \varepsilon\delta_{-x}(-y)) + \mu(F_X)$$

$$= -\mu((1-\varepsilon)\ F_X + \varepsilon\delta_{-x}) + \mu(F_X)\ ,$$

dividing by ε and taking the limit as ε goes to 0 ends the proof.

Note that only properties ii) and iii) of a BLLP have been used.

The IC has been useful in the M location parameter context to build robust and highly robust estimators. Hampel ((1974), (1973)).

Huber's minimax M estimator is the prototype of robust M estimators while Hampel's three parts descending M estimator as defined at the beginning of section II.6 is the prototype of highly robust M estimators.

The L counterpart of Huber's estimator, the trimmed mean or in general a trimmed L estimator, has been investigated by Jaeckel (1971), Bickel and Lehmann (1975) and for simulation Andrews et al.(1972). Yet, the L counterpart of Hampel's proposal has never been considered. Figure I A) contains the IC of a three parts descending M estimator up to a constant.

A similar graph would be obtained for the IC of the L estimator based on the following J:

let $0 < t_0 < t_1 < t_2 < {}^1/2$, define for $a > 0$

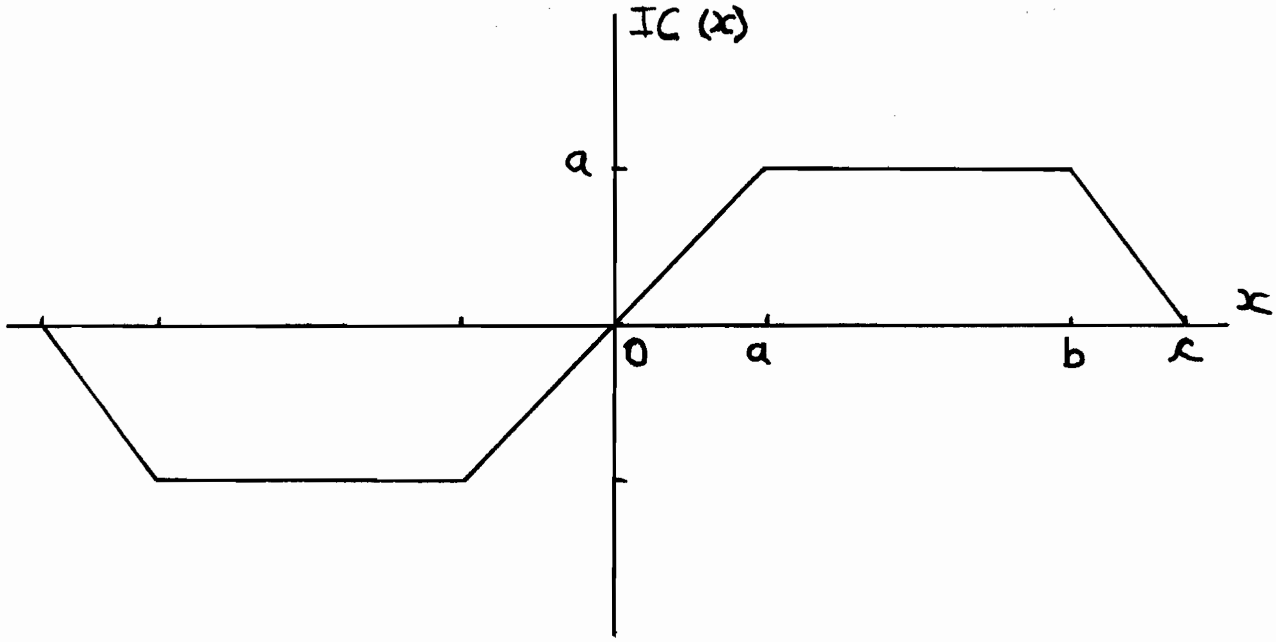
$$J(t) = J(1-t) = \begin{cases} 0 & 0 \leq t < t_0 \\ -{}^1/a & t_0 \leq t < t_1 \\ 0 & t_1 \leq t < t_2 \\ 1 & t_2 < t < {}^1/2 \, . \end{cases}$$

This L estimator is the difference between a trimmed mean and a trimmed outer mean. Assuming F is symmetric with respect to 0 , the corresponding IC is: $IC\ (F,\mu)\ (x) = -\ IC\ (F,\mu)\ (-x) =$

$$\begin{array}{ll} x & x \varepsilon [0, F^{-1}(1-t_2)] \\ F^{-1}(1-t_2) & x \varepsilon [F^{-1}(1-t_2), F^{-1}(1-t_1)] \\ F^{-1}(1-t_2) - a^{-1}(x - F^{-1}(1-t_1)) & x \varepsilon [F^{-1}(1-t_1), F^{-1}(1-t_0)] \\ F^{-1}(1-t_2) - a^{-1}(F^{-1}(1-t_0) - F^{-1}(1-t_1)) & x \varepsilon [F^{-1}(1-t_0), \infty] \end{array}$$

Figure I B) contains the graph of such a function. The IC obtained is a function of the underlying distribution. One should use this "adaptiveness"

A) The IC of the three parts descending M estimator.



B) The IC of the L analogue of the three parts descending M estimator.

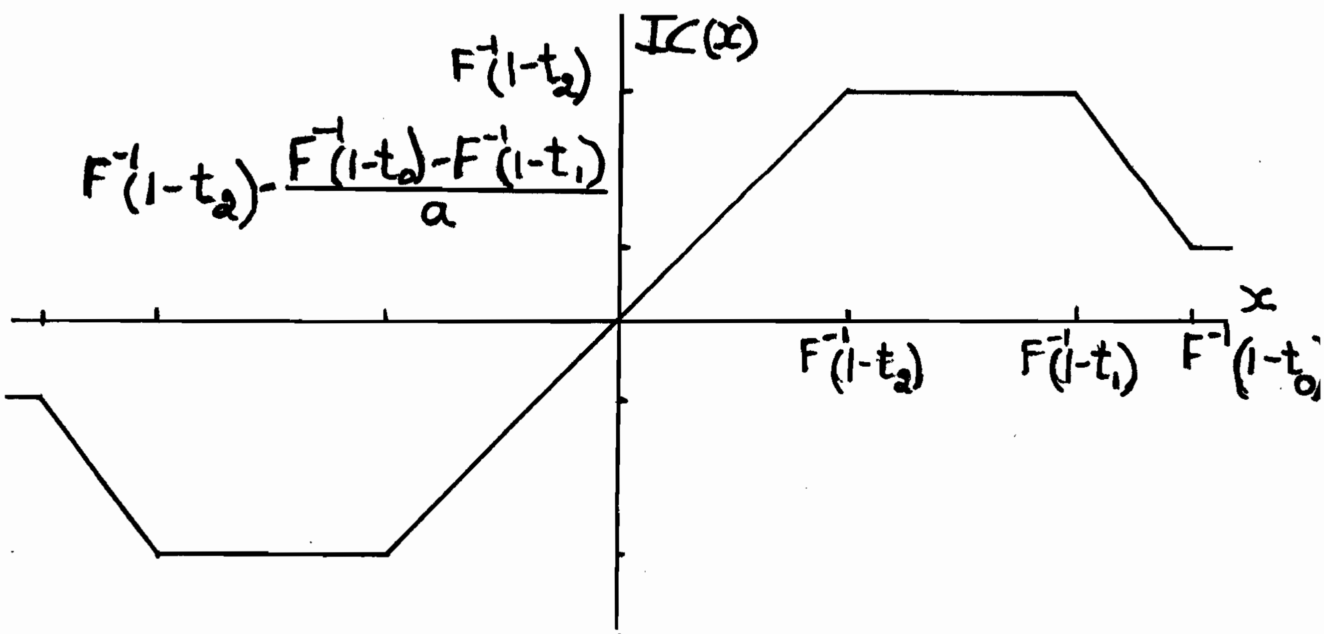


Figure I

of the IC to build "nice" estimators.

The principle underlying the construction of the next estimators is the following: first choose a heavy tailed distribution then design the IC in such a way that the influence of the extreme observations is zero for this special distribution.

The distribution used to build LM_1 , LM_2 , LM_3 and LM_4 , the L estimator based on J_1 , J_2 , J_3 and J_4 is the Cauchy:

$$J_1(t) = J_1(1-t) = {}^1/.36 \begin{cases} 1 & t\varepsilon[.5,.75] \\ -{}^1/2.1 & t\varepsilon[.75,.9] \\ 0 & t\varepsilon[.9,1] \end{cases}$$

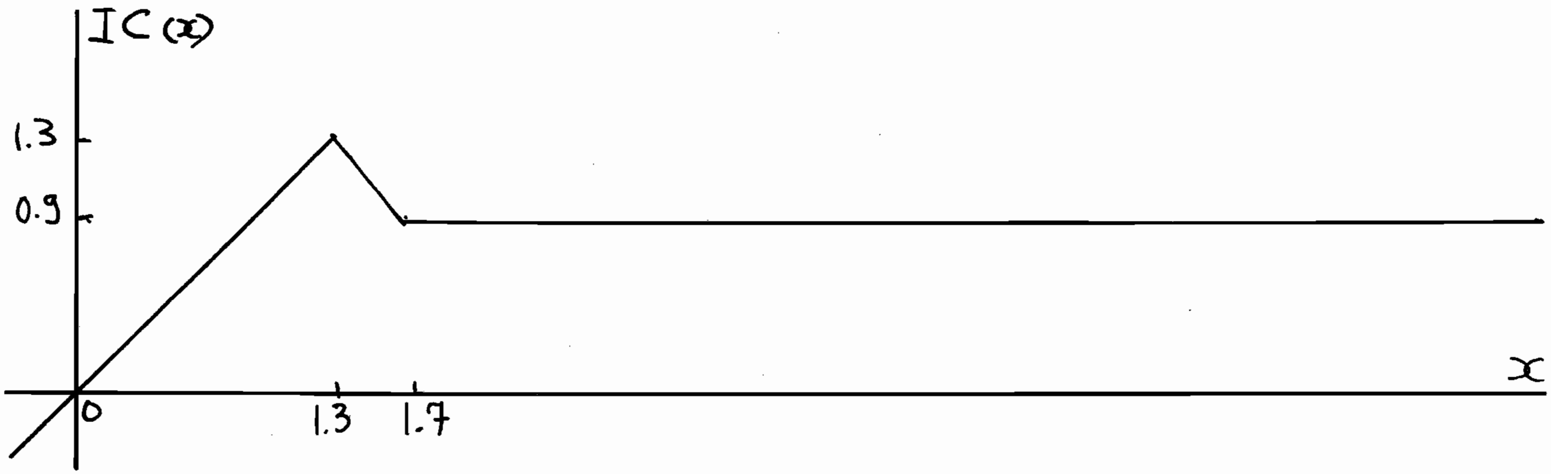
$$J_2(t) = J_2(1-t) = {}^1/.43 \begin{cases} 1 & t\varepsilon[.5,.75] \\ -{}^1/5.4 & t\varepsilon[.75,.95] \\ 0 & t\varepsilon[.95,1] \end{cases}$$

$$J_3(t) = J_3(1-t) = {}^1/.47 \begin{cases} 1 & t\varepsilon[.5,.75] \\ 0 & t\varepsilon[.75,.9] \\ -{}^1/3.3 & t\varepsilon[.9,.95] \\ 0 & t\varepsilon[.95,1] \end{cases}$$

$$J_4(t) = J_4(1-t) = {}^1/.7 \begin{cases} 1 & t\varepsilon[.5,.9] \\ -1 & t\varepsilon[.9,.95] \\ 0 & t\ [.95,1] \end{cases} .$$

Note that seen as functionals, these location parameters satisfy conditions

A) LM_4's IC for the normal



B) LM_4's IC for the t with 3 degrees of freedom

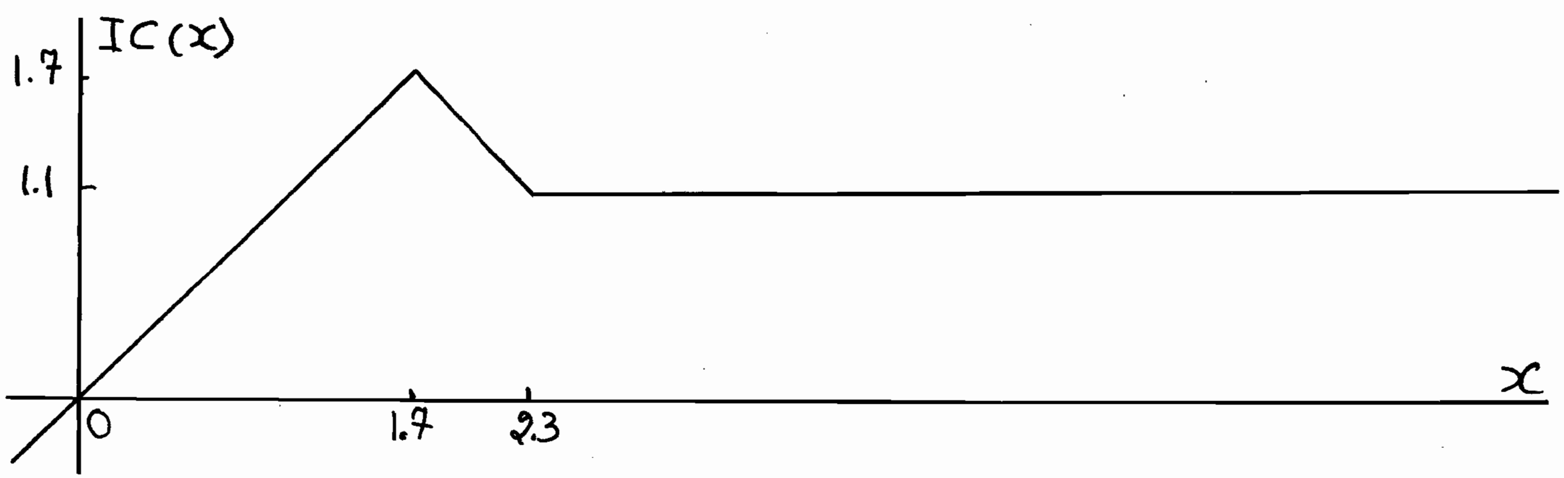
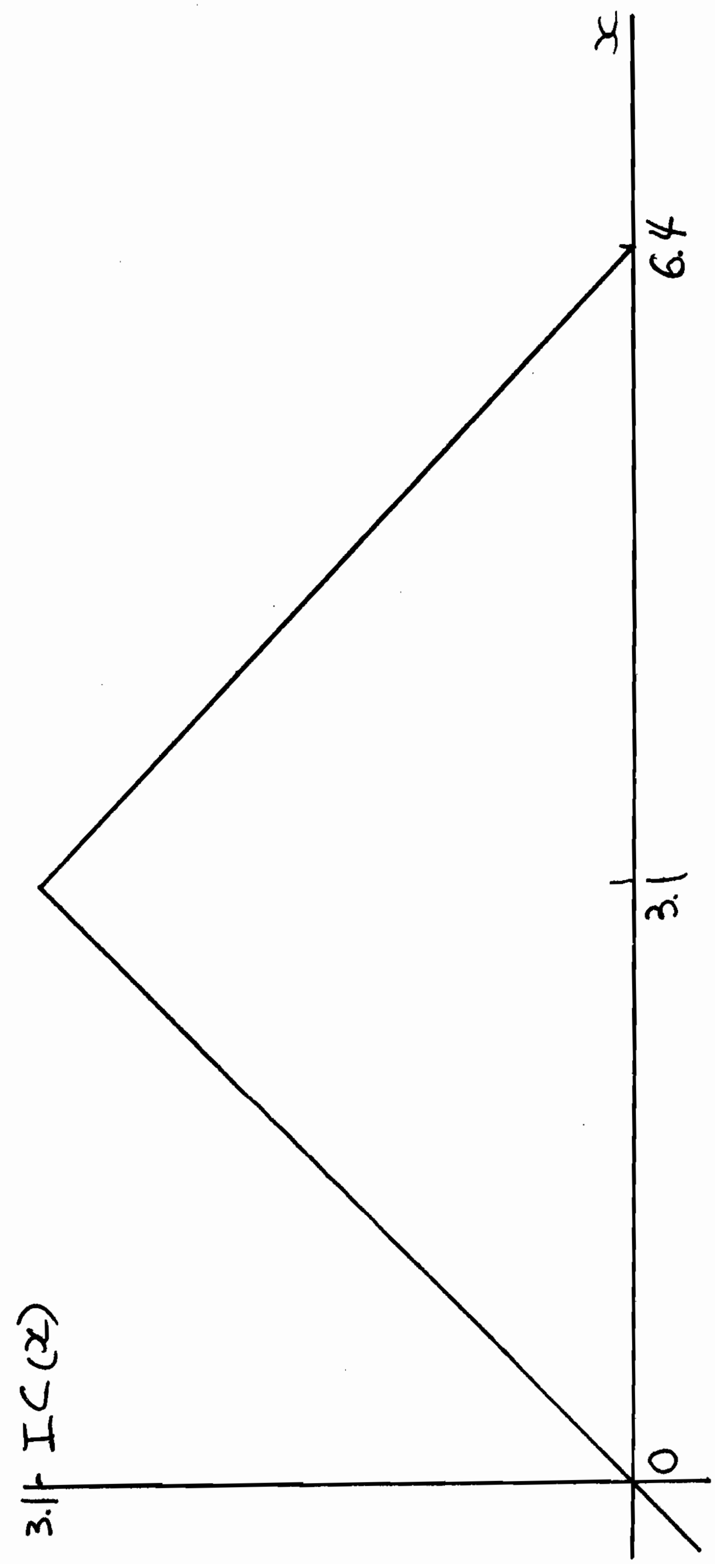


Figure II



LM$_4$'s IC for the Cauchy

Figure III

ii) and iii) of BLLP's definition while highly robust M estimators satisfy only condition ii); they are not scale invariant.

Figure II contains, up to a constant, the graphs of LM_4's IC when the underlying distribution is normal and t with 3 degrees of freedom, Figure III contains a similar graph when the underlying distribution is Cauchy.

(Since the distributions under consideration are symmetric, the IC are odd and IC (F,μ) (x) has been graphed only for $x \geq 0$) .

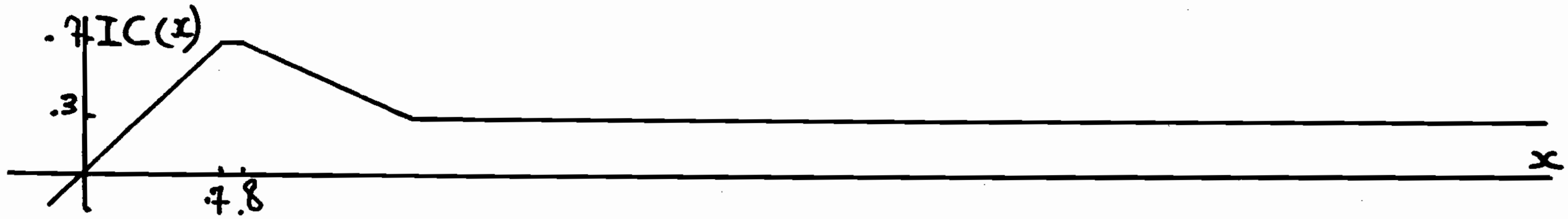
LM_5 and LM_6 , the L estimators based on J_5 and J_6 have been built using the t distribution with 2 degrees of freedom, LM_7 , the L estimator based on J_7 is using a t with 4 degrees of freedom.

$$J_5(t) = J_5(1-t) = {}^1/.37 \begin{cases} 1 & t\varepsilon[.5,.75] \\ 0 & t\varepsilon[.75,.8] \\ -{}^1/2.25 & t\varepsilon[.8,.95] \\ 0 & t\varepsilon[.95,1] \end{cases}$$

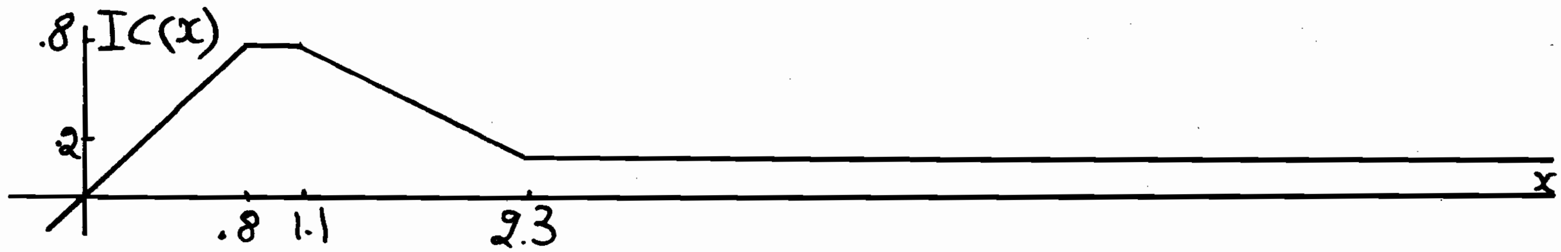
$$J_6(t) = J_6(1-t) = {}^1/.61 \begin{cases} 1 & t\varepsilon[.5,9] \\ -1.9 & t\varepsilon[.9,.95] \\ 0 & t\varepsilon[.95,1] \end{cases}$$

$$J_7(t) = J_7(1-t) = {}^1/.42 \begin{cases} 1 & t\varepsilon[.5,8) \\ 0 & t\varepsilon[.8,9] \\ -{}^1/.8 & t\varepsilon[.9,.95] \\ 0 & t\varepsilon[.95,1] \end{cases}$$

A) LM_5's IC for the normal



B) LM_5's IC for the t with three degrees of freedom



C) LM_5's IC for the Cauchy

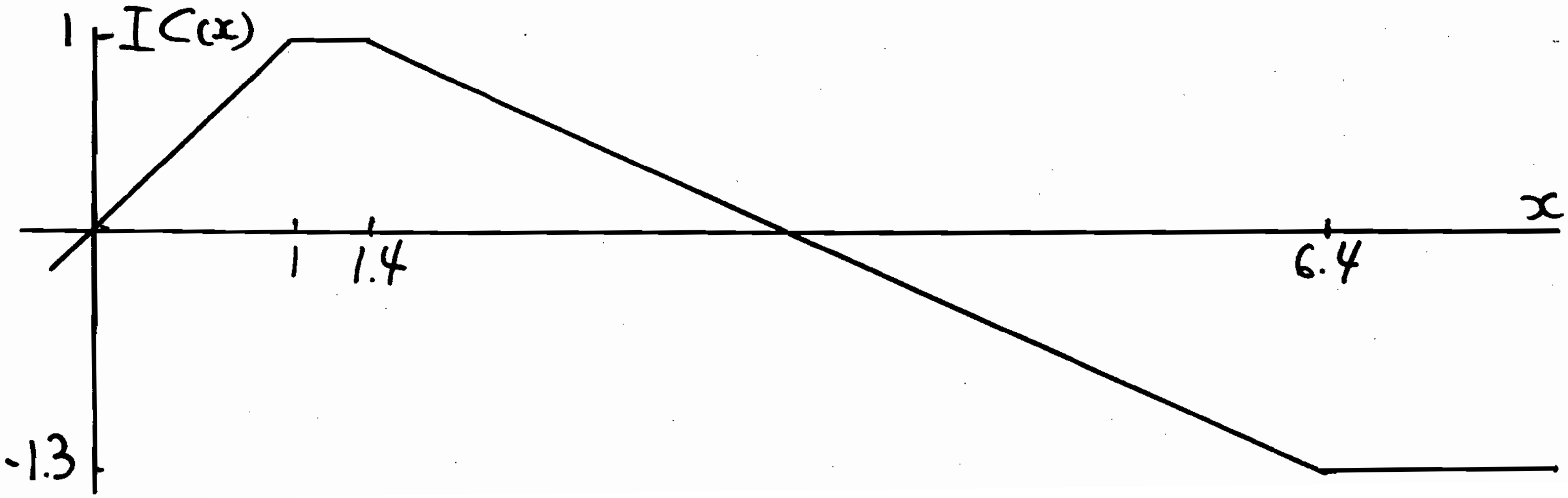


Figure IV

Figure IV contains, up to a constant, the graphs of LM_5's IC when the underlying distribution is normal, t with 3 degrees of freedom and Cauchy ($IC(F,\mu)(x)$ has been graphed for $x \geq 0$) .

The last part of this section contains a Monte-Carlo study of the previously defined estimators. Each estimator was simulated a thousand times from a sample of size 20 for ten sampling distributions. These sampling distributions were distributions of the following type of random variables: X/Y where X is distributed N(0,1) and Y, the contaminating random variable, is positive, distributed independently of X. The distributions of Y and the corresponding sampling distributions are:

Contaminating distribution	Sampling distribution
a) degenerate Y=1 a.e.	$\Phi(x)$
b) $\sqrt{\chi_3^2/3}$	t with 3 d.f.
c) $Y = \begin{cases} c^{-1} & \text{with probability } \alpha \\ 1 & \text{with probability } 1-\alpha \end{cases}$	$\alpha\Phi(x/c) + (1-\alpha)\,\Phi(x)$
d) half normal $2(\Phi(y)-.5)$	Cauchy
e) uniform [0,1]	$\int_0^1 \Phi(xu)\,du$

Remarks

1) The contaminated normal, c) was simulated for 2 values of c, 3 and 10, and for 3 values of α , .05,.1 and .25. The sampling distribution for a given value of α and c is labelled α , cN .

2) e) was added despite its artificial construction to provide a heavy

tailed distribution with an unpeaked center, see Andrews et al. (1972) p. 123.

A variance reduction technique due to Dixon and Tukey (1968), Relles (1970) and apparently based on ideas of Fraser (1968) was used, see Andrews et al (1972). Here is a brief description:

Let: $\tilde{x}$ denote the 20 x 1 vector of normal deviates,

$\tilde{y}$ denote the 20 x 1 vector of contaminating deviates, and

$\tilde{z} = (x_i/y_i)$ is the vector of the observations.

$T(\tilde{z})$, the computed estimate, is location and scale invariant and $T(-\tilde{z}) = T(\tilde{z})$. Hence $E(T(\tilde{z})) = 0$ and $E(T^2(\tilde{z}))$ has to be estimated.

Let: $\hat{z} = (\Sigma\, x_i y_i)/(\Sigma y_i^2)$

given $\tilde{y}$ $\hat{z}$ is distributed $N(0, (\Sigma y_i^2)^{-1})$,

$$s_{\hat{z}}^2 = 19^{-1}\, \Sigma(x_i - y_i \hat{z})^2$$

given $\tilde{y}$ $s_{\hat{z}}^2$ is distributed $\chi_{19}^2/19$.

Define, $\tilde{c}(\tilde{z})$ the configuration in the following way:

$$\tilde{c}(\tilde{z}) = s_{\hat{z}}^{-1}\,(\tilde{z} - \tilde{e}\hat{z})$$

where $\tilde{e}$ is a 20 x 1 vector of 1's . Note that given $\tilde{y}$, $\tilde{c}(\tilde{z})$, $\hat{z}$ and $s_{\hat{z}}^2$ are independent.

Thus:

$$E(T^2(\underset{\sim}{z})) = E(E(T^2(\underset{\sim}{z})|\underset{\sim}{y},\underset{\sim}{c}(\underset{\sim}{z})))$$

and $E(T^2(\underset{\sim}{z})|\underset{\sim}{y},\underset{\sim}{c}(\underset{\sim}{z})) =$

$$E\left(\left(\left(\frac{T(\underset{\sim}{z})-\hat{z}}{s_{\hat{z}}}\right) s_{\hat{z}} + \hat{z}\right)^2 \Big| \ldots\right)$$

$$= E((T(\underset{\sim}{c}(\underset{\sim}{z}))\ s_{\hat{z}} + \hat{z})^2|\ldots)$$

$$= T^2(\underset{\sim}{c}(\underset{\sim}{z}))\ E(s_{\hat{z}}^2|\underset{\sim}{c}(\underset{\sim}{x}),\underset{\sim}{y}) + 2T(\underset{\sim}{c}(\underset{\sim}{z}))\ E(s_{\hat{z}}\hat{z}|\underset{\sim}{c}(\underset{\sim}{x}),\underset{\sim}{y}) + E(\hat{z}^2|\underset{\sim}{c}(\underset{\sim}{x}),\underset{\sim}{y})$$

$$= T^2(\underset{\sim}{c}(\underset{\sim}{z})) + (\Sigma y_i^2)^{-1}$$

using the independence property. To obtain the desired variances, we sum $T^2(\underset{\sim}{c}(\underset{\sim}{z})) + (\Sigma y_i^2)^{-1}$. Table I contains 20 times the sampling variances of the LM estimates for the different sampling distributions. It also contains the corresponding variances for the 5%, 10%, and 25% trimmed means computed by Andrews et al.(1972).

The empirical results are disappointing. LM1, LM2 and LM3 are equal to 25%, the 25% trimmed mean, minus a trimmed outer mean divided by a normalizing constant. These estimates are totally outdone by 25% (except LM1 for Normal/U). LM4 is the most successful; in gentle situation it stands around 5% and 10% but for highly contaminated cases, it breaks down rather surprisingly. LM5, LM6 and LM7 are even poorer than LM1 to LM4. Furthermore LM1 - LM7 are completely outdone by the Hampel estimates (for Hampel's estimate variances, see Andrews et al. (1972)).

Two reasons are set forth to explain these results:

	Sampling distributions									
Estimates	Normal	t_3	.05,3N	.1,3N	.25,3N	.05,10N	.1,10N	.25,10N	Cauchy	Normal/U
LM1	1.48	1.89	1.58	1.68	2.09	1.60	1.82	3.47	5.43	6.48
LM2	1.31	1.70	1.40	1.50	1.89	1.42	1.55	2.22	8.72	8.46
LM3	1.25	1.65	1.34	1.44	1.84	1.36	1.51	2.28	12.05	11.81
LM4	1.13	1.67	1.24	1.36	1.90	1.37	1.84	5.51	35.72	37.67
5%	1.02	1.88	1.19	1.41	2.27	1.23	2.90	14.93	24	35.94
10%	1.05	1.68	1.17	1.33	1.92	1.20	1.46	6.71	7.3	13.60
25%	1.20	1.59	1.29	1.39	1.79	1.29	1.47	2.18	3.1	6.62
LM5	1.55	2.02	1.65	1.76	2.18	1.85	2.70	6.73	56.15	33.94
LM6	1.28	1.94	1.41	1.55	2.16	2.31	4.69	11.18	193.68	161.22
LM7	1.69	2.77	1.88	2.12	3.06	4.20	10.66	30.15	***	***

Table I

20 times the variances of the estimates

1) <u>The normalizing constants</u>

The IC of Hampel's estimates is equal to:

$$\sigma(F)\psi(x/\sigma(F))/E(\psi'(X/\sigma(F)))$$

where X has distribution F and $\sigma(F)=(F^{-1}(3/4)-F^{-1}(1/4))/(\Phi^{-1}(3/4)-\Phi^{-1}(1/4))$,

$$E(\psi'(X/\sigma(F))) = \{P(|X/\sigma(F)|\varepsilon[0,a]) - P(|X/\sigma(F)|\varepsilon[b,c])a/(c-b)\}$$

is the normalizing constant. Here are these constants for Hampel's estimates considered by Andrews et al.(1972) in three sampling situations:

a	b	c	Normal	t_3	Cauchy
2.5	4.5	9.5	.98	.93	.78
2.2	3.7	5.9	.96	.88	.72
2.1	4.	8.2	.95	.89	.76
1.7	3.4	8.5	.89	.84	.72
1.2	3.5	8	.74	.76	.64

For the LM's estimate, the IC is equal to:

$$\int_0^x J(F(x))\,dx/\int_0^1 J(t)\,dt$$

The normalizing constant is $\int_0^1 J(t)\,dt$. For the estimates in Table I these constants are:

Estimate	constant
LM1	.36
LM2	.43
LM3	.47
LM4	.7
5%	.9
10%	.8
25%	.5
LM5	.37
LM6	.61
LM7	.42

The normalizing constants for the L estimates are generally smaller than the ones for the M estimates in all sampling distributions. To robustify L estimates, the price paid on the normalizing constants is bigger than for the M case thereby increasing the effect of the influential observations on the estimates.

2) Non normality

To explain the discrepancy between the LM's variances and the trimmed variances, in vigorous situations, note that the influence theory is asymptotic. The weight functions of the LM estimates are very discontinuous, this may decrease the rate of convergence to the asymptotic situation in highly contaminated cases. The fact that, in the Cauchy and Normal/U cases, the variances increase with the absolute value of the weight given to the extreme observations support this statement:

LM1	gives weight	-1.33	to	x(18)
LM2	" "	- .43	to	x(19)
LM3	" "	- .61	to	x(19)
LM4	" "	-1.44	to	x(19)
5%	" "	1.11	to	x(19)
10%	" "	1.25	to	x(18)
25%	" "	2	to	x(18)
LM5	" "	-2.75	to	x(19)
LM6	" "	-3.11	to	x(19)
LM7	" "	-4.28	to	x(19) .

Appendix I contains a listing of the computer program used in this section.

Chapter IV

Two orderings of distributions

In any estimation procedure, the performance of a certain estimator is mainly a function of the tails of the underlying distribution. In the location model the Princeton Simulation Study, Andrews et al. (1972), for finite samples and Bickel and Lehmann (1975), Rivest (1976) for infinite samples support this statement. Mosteller and Tukey (1977) have pointed out that for real data, the most important deviation from normality is the behaviour in the tails of the underlying distribution.

In this chapter, two methods for the classification of distributions will be discussed.

The first one has been introduced by van Zwet in 1964. van Zwet's ordering is based on F^{-1}, the inverse of the distribution function.

The second ordering has never been defined as such. In this context, G is bigger than F, if G can be regarded as the distribution of XY where X and Y are independent random variables and X has distribution F. This method has been used at the end of Chapter III to generate distributions with tails bigger than the normal ones.

Section IV.1 α unimodality

In this section, a generalization of Khintchine (1938)'s concept of unimodality introduced by Olshen and Savage (1970) is discussed. We will characterize the distributions of the following type of random variables:

$$U_0^{1/\alpha}\, Y$$

where U_0 is a $U[0,1]$ random variable and Y is any random variable distributed independently of U_0 .

<u>Definition IV.1</u> <u>α Unimodality</u>

A distribution function F is said to be α unimodal ($\alpha > 0$) if F has a well defined density f(x), except may be at a point a in R, satisfying

$$f(x)/(x-a)^{\alpha-1} \text{ is decreasing } x > a ,$$
$$f(x)/(a-x)^{\alpha-1} \text{ is increasing } x < a.$$

<u>Remarks</u>

1) For $\alpha=1$, this is analogous to standard unimodality as defined by Gnedenko and Kolmogorov (1954) p. 157.

2) If F is α_0 unimodal, F is α unimodal for any $\alpha \geq \alpha_0$.

3) Without loss of generality, assume that f is right continuous so that $f(x)/(x-a)^{\alpha-1}$ is right continuous except maybe at a.

<u>Example</u>

Let:

$$f(x) = \frac{1}{\Gamma(\alpha)} \beta^{\alpha} x^{\alpha-1} e^{-\beta\alpha} \quad x > 0 ,$$

then $F(x) = \int_0^x f(x)\,dx$, the Gamma with parameter α and β , is α unimodal (taking a=0).

<u>Theorem IV.1</u>

F is α unimodal if and only if:

$$M(x) = F(x) - (x-a)\, f(x)/\alpha$$

is a distribution function.

Proof: Assume without loss of generality a = 0 .

Define for $t < 0$

$$V_1(t) = \int_{-\infty}^{t} (-x)^{\alpha}\, df(x)/(\alpha(-x)^{\alpha-1})$$

and for $t > 0$

$$V_2(t) = F(0) - \int_0^t x^{\alpha}\, df(x)/(\alpha x^{\alpha-1}) ,$$

where $\int_0^t = \int_{(0,t]}$.

Suppose first that f is α unimodal; it suffices to prove

$$M(t) = \begin{cases} V_1(t) & t < 0 \\ V_2(t) & t \geq 0 \end{cases}$$

and $V_1(-\infty) = 0$ and $V_2(\infty) = 1$.

By definition

$$V_1(t) = \lim_{M} \int_{-M}^{t} (-x)^{\alpha}\, df(x)/(\alpha(-x)^{\alpha-1}) .$$

Consider:

$$\int_{-M}^{t} (-x)^{\alpha} df(x)/(\alpha(-x)^{\alpha-1})$$

integrating by parts leads:

$$F(t) - F(-M) - tf(t)/\alpha - Mf(-M)/\alpha .$$

In order to prove $V_1(t) = M(t)$, $t < 0$ and $V_1(-\infty) = 0$, one has to show:

(4.1) $$\lim_{M\to\infty} M\, f(-M)/\alpha = 0 \; .$$

Since f is a density, for any $\varepsilon > 0$, there exists M_ε such that

$$M > M_\varepsilon \text{ implies } \int_{-M}^{-M_\varepsilon} f(x)\, dx < \varepsilon$$

or: $$\varepsilon > \int_{-M}^{-M_\varepsilon} \frac{(-x)^{\alpha-1}\, f(x)\, dx}{(-x)^{\alpha-1}}$$

$$> \quad (f(M)/M^{\alpha-1}) \; . \; (M^\alpha - M_\varepsilon^\alpha)/\alpha \; .$$

Note that $\lim_{M\to\infty} f(-M)/M^{\alpha-1} = 0$ (otherwise, $\int_{-\infty}^{0} f(x)dx = \infty$) and (4.1) is proved. The same way, it is shown that:

(4.2) $$\lim_{M\to\infty} Mf(M) = 0 \; .$$

Consider for $t > 0$:

$$V_2(t) - F(0) = \lim_{\varepsilon\to 0} \int_\varepsilon^t -x^\alpha df(x)/(\alpha x^{\alpha-1}) \quad .$$

Integrating the RHS by parts leads:

$$F(t) - F(\varepsilon) - tf(t)/\alpha + \varepsilon f(\varepsilon)/\alpha \quad .$$

In order to prove $V_2(t) = M(t)$, $t > 0$ and $V_2(\infty) = 1$, one has to show:

(4.3) $$\lim_{\substack{\varepsilon \to 0 \\ \varepsilon > 0}} \varepsilon\, f(\varepsilon) = 0 \; .$$

Since f is integrable, for any $\eta > 0$, there exists ε_0 such that ε in $(0,\varepsilon_0)$ implies:

$$\eta > \int_0^{\varepsilon} f(x)\, dx$$

$$= \int_0^{\varepsilon} x^{\alpha-1} f(x)/(x^{\alpha-1})\, dx$$

$$> (f(\varepsilon)/(\varepsilon^{\alpha-1})\, \varepsilon^{\alpha})/\alpha \ .$$

Therefore (4.3) is proved. A similar argument would show that:

(4.4) $$\lim_{\substack{\varepsilon \to 0 \\ \varepsilon < 0}} \varepsilon f(\varepsilon) = 0 \ .$$

Note that since f is right continuous M(x) is right continuous except maybe at 0 , but (4.3) implies that M(x) is right continuous at 0 .

Since (4.2) implies $V_2(\infty) = 1$ the first part is proved. To prove the converse note that if $x > y > 0$

$$V_2(x) - V_2(y) = M(x) - M(y)$$

therefore if M(x) is increasing, so are $V_2(x)$ and $V_1(x)$, hence

$$f(x)/(-x)^{\alpha-1} \text{ increases in } (-\infty, 0)$$

and $f(x)/x^{\alpha-1}$ decreases in $(0,\infty)$ and f is α unimodal.

<u>Remark</u>

1) Gnedenko and Kolmogorov (1954) p.157 proved the theorem for $\alpha=1$. An alternate proof of theorem IV.1 may be derived using their result and the fact that if X is α unimodal $(X-a)^{\alpha}$ is unimodal.

Theorem IV.2

The following statements are equivalent:

i) F is α unimodal with respect to a,

ii) let U_0 be a random variable uniformly distributed in [0,1], there exists a random variable Y distributed independently of U_0 such that:

$$a + U_0^{1/\alpha}(Y-a) \text{ has distribution } F ,$$

iii) $|F^{-1}(t) - a|^\alpha$ is convex in [0,1] .

Proof: Assume without loss of generality a = 0 .

To prove i) → ii) let Y be a random variable having distribution M(x) . For x > 0

$$P(U_0^{1/\alpha}\, Y \le x) = M(0) + P(U_0^{1/\alpha}\, Y \in (0,x]) .$$

Note that M(0) = F(0) and consider:

$$(4.5) \qquad P(U_0^{1/\alpha}\, Y \in (0,x]) = \int_0^\infty F^{(\alpha)}(x/t)\, dM(t)$$

where
$$F^{(\alpha)}(x) = \begin{cases} 0 & x < 0 \\ x^\alpha & x \in [0,1] . \\ 1 & x > 1 \end{cases}$$

Then (4.5) is equal to:

$$M(x) - M(0) - \int_x^\infty (x^\alpha/t^\alpha)\,(t^\alpha)\, df(t)/(\alpha t^{\alpha-1})$$

$$= M(x) - M(0) + xf(x)/\alpha$$

$$= F(x) - M(0) .$$

Similarly one shows that the result is true for $x > 0$.

To prove ii) → iii), for $x > 0$

$$F(x) = F(0) + \int_0^\infty F^{(\alpha)}(x/t)\, dG(t) ,$$

for a certain distribution G . Let $t_0 = x/2$, for any ε satisfying $|\varepsilon| < t_0$,

$$\varepsilon^{-1}(F(x+\varepsilon) - F(x)) = \int_{t_0}^{\infty} \varepsilon^{-1}(F^{(\alpha)}((x+\varepsilon)/t) - F^{(\alpha)}(x/t))\, dG(t) .$$

Using the mean value theorem and the fact that $f^{(\alpha)}(x) = \frac{d}{dx} F^{(\alpha)}(x)$ is monotone in $[0,1]$ for $t > t_0$

$$\varepsilon^{-1}(F^{(\alpha)}((x+\varepsilon)/t) - F^{(\alpha)}(x/t))$$

$$\leq t^{-1}\alpha\, [|(x+|\varepsilon|)/t|^{\alpha-1} + |(x-|\varepsilon|)/t|^{\alpha-1}]$$

$$\leq \alpha t_0^{-\alpha}\, [|x+|\varepsilon||^{\alpha-1} + |x-|\varepsilon||^{\alpha-1}]$$

which is a G integrable function. Applying dominated convergence:

$$f(x) = \alpha\, x^{\alpha-1} \int_x^\infty t^{-\alpha}\, dG(t) .$$

Therefore for $x > 0$,

$$\frac{\alpha x^{\alpha-1}}{f(x)} \text{ is increasing or:}$$

(4.5) $$\frac{\alpha(F^{-1}(t))^{\alpha-1}}{f(F^{-1}(t))} \text{ is increasing in } (t_1,1)$$

where $t_1 = \inf\{t : F^{-1}(t) > 0\}$. Integrating the last expression we obtain that:

$$(F^{-1}(t))^{\alpha} \text{ is convex in } (t_1,1) \ .$$

For $x < 0$

$$F(x) = \int_{-\infty}^{0} (1-F^{(\alpha)}\ (\tfrac{x}{t}))\ dG(t) \ ,$$

using a similar argument, one shows that:

$$f(x) = \alpha(-x)^{\alpha-1} \int_{-\infty}^{x} (-t)^{-\alpha}\ dG(t)$$

and

(4.6) $$-\frac{\alpha(-F^{-1}(t))^{\alpha-1}}{f(F^{-1}(t))}$$ is increasing so that $(-F^{-1}(t))^{\alpha}$ is convex in $(0,t_0)$ where $t_0 = \sup\{t : F^{-1}(t) < 0\}$.

To prove iii) $\rightarrow$ i) , note that iii) implies that (4.5) and (4.6) hold so i) is true.

Corollary IV.1

F is α unimodal if and only if ϕ the characteristic function of F satisfies:

$$\phi(t) = e^{-iat}\ \alpha|t|^{-\alpha} \int_{0}^{|t|} s^{\alpha-1}\ \nu(\text{sgn}(t)s)\ ds$$

where ν is any characteristic function.

Proof: Without loss of generality assume $a = 0$. Using theorem IV.2, F is α unimodal if and only if $U_0^{1/\alpha}Y$ has distribution F, hence:

$$\phi(t) = E(e^{itU_0^{1/\alpha}Y})$$

$$= E(E(e^{itU_0^{1/\alpha}Y} | U_0))$$

$$= E(\nu(tU_0^{1/\alpha}))$$

where ν is the characteristic function of Y . Hence:

$$\phi(t) = \int_0^1 \nu(tu^{1/\alpha})\, du \, .$$

let $v = |t|\, u^{1/\alpha}$ ie $du = \alpha |t|^{-\alpha} v^{\alpha-1}\, dv$

$$\phi(t) = \alpha |t|^{-\alpha} \int_0^{|t|} v^{\alpha-1}\, \nu(\mathrm{sgn}(t)v)\, dv$$

Remark:

Assume a = 0 , note that $\phi(t)$ is differentiable as a product of differentiable functions. Given $\phi(t)$, $\nu(t)$ can be found in the following way, if $t > 0$,

$$\nu(t) = \alpha^{-1} \quad t^{-\alpha+1} \frac{d}{dt} \; t^{\alpha} \phi(t)$$

$$= \phi(t) + t\phi'(t)/\alpha \, .$$

Using a similar argument for $t < 0$, F is α unimodal if and only if

$$\nu(t) = \phi(t) + t\,\phi'(t)/\alpha$$

is a characteristic function.

Note that this condition on ϕ, the characteristic function, is similar to the condition on F as stated in theorem IV.1 .

In the last part of this section, distributions of

$$U_0^{-1/\alpha}Y$$

will be characterized.

Definition IV.2 α^-unimodality

A distribution F is said to be α^-unimodal ($\alpha > 0$) if F has a well defined density f(x) except maybe at a point a in R satisfying:

$$(x-a)^{\alpha+1} f(x) \text{ increases in } (a,\infty)$$

$$(a-x)^{\alpha+1} f(x) \text{ decreases in } (-\infty,a) .$$

Example

Let

$$f(x) = \Pi^{-1}(1+x^2)^{-1}$$

$F(x) = \int_{-\infty}^{x} f(y)\,dy$ is 1^- unimodal (taking a=0) .

Theorem IV.3

Let X be a random variable having an absolutely continuous distribution F, F is α^-unimodal if and only if the distribution of $(X-a)^{-1}$ is α unimodal.

Proof: Assume a=0 .

$$(4.7) \qquad P(X^{-1} \le x) = \begin{cases} F(0) + 1 - F(x^{-1}) & x > 0 \\ F(0) - F(x^{-1}) & x < 0 . \end{cases}$$

The density of 1/X is:

$$x^{-2} f(x^{-1}) .$$

Therefore, if $x > 0$

$$x^{-(\alpha-1)} (x^{-2} f(x^{-1})) = x^{-(\alpha+1)} f(x^{-1})$$

is a decreasing function of x since F is α^- unimodal. For $x < 0$, a similar argument proves that the distribution of X^{-1} is α unimodal. The converse is proved the same way.

Remark

The assumption F is absolutely continuous implies that $(X-a)^{-1} \neq \infty$ with probability 1 .

Theorem IV.4

Let F be an absolutely continuous distribution. The following statements are equivalent:

i) F is α^- unimodal,

ii) $M(x) = F(x) + \alpha^{-1}(x-a) f(x)$ is a distribution function,

iii) If U_0 is a $\mathcal{U}[0,1]$ distributed random variable, there exists a random variable Y independent of U_0 such that:

$$a + U_0^{-1/\alpha}(Y-a) \text{ has distribution } F,$$

iv) $|F^{-1}(t)-a|^{-\alpha}$ is convex in $(0,t_0)$, $(t_0,1)$ for a certain $t_0 \,\varepsilon[0,1]$ and

$$\lim_{t\to 0} F^{-1}(t) = -\infty \ , \lim_{t\to 1} F^{-1}(t) = \infty \ .$$

Proof: Assume a = 0 . To prove i) $\leftrightarrow$ ii) note that X^{-1} is α unimodal if and only if (using theorem IV.1)

$$G(x) = \begin{cases} F(0) - F(x^{-1}) - (\alpha x)^{-1} f(x^{-1}) & x < 0 \\ \\ F(0) + 1 - F(x^{-1}) - (\alpha x)^{-1} f(x^{-1}) & x > 0 \end{cases}$$

is a distribution. If Y has distribution G, 1/Y has distribution

$$M(x) = F(x) + xf(x)/\alpha$$

using (4.7).

i) $\leftrightarrow$ iii) by theorem IV.2.

To prove i) $\rightarrow$ iv), take $t_0 = F^{-1}(a)$. For $t > t_0$

$$\alpha^{-1} F^{-1}(t)^{\alpha+1} f(F^{-1}(t)) \text{ is increasing}$$

or

$$- \alpha F^{-1}(t)^{-(\alpha+1)} [f(F^{-1}(t))]^{-1} \text{ is increasing.}$$

Integrating we obtain $F^{-1}(t)^{-\alpha}$ is convex in $(t_0,1)$. Similarly, it is shown that $(-F^{-1}(t))^{-\alpha}$ is convex in $(0,t_0)$. To prove iv) $\rightarrow$ i, one uses an argument similar to the one used at the end of theorem IV.2. The condition on $F^{-1}(0)$ and $F^{-1}(1)$ implies $f(x)\neq 0 \ x \,\varepsilon\, R$ which is a necessary condition to have α^-unimodality.

In the last theorem, the absolute continuity of F allows the use of the relationship between α and α^- unimodality to obtain a short proof.

This assumption is not necessary. Arguments similar to the ones used in the proof of theorems IV.1 and IV.2 would prove theorem IV.4 without this assumption.

Finally, we will find an expression for the characteristic function of an α^- distribution.

Corollary IV.2

F is an α^- distribution if and only if ϕ the chracteristic function of F satisfies:

$$\phi(t) = e^{-iat} \; \alpha t^{\alpha} \int_t^{\infty} v^{-(\alpha+1)} \, \nu(v) \, dv \quad \text{for } t > 0$$

where ν is a characteristic function.

Proof: Without loss of generality assume a=0. By theorem IV.5, F is α^- unimodal if and only if

$$\phi(t) = E(e^{itU_0^{-1/\alpha}Y})$$

$$= E(E(e^{itU_0^{-1/\alpha}Y}|U_0))$$

$$= \int_0^1 \nu(t\,u^{-1/\alpha}) \, du$$

where ν is the characteristic function of Y . Now, if $v = tu^{-1/\alpha}$, $-du = \alpha t^{\alpha} v^{-(\alpha+1)} dv$ and

$$\phi(t) = \alpha t^{\alpha} \int_t^{\infty} v^{-(\alpha+1)} \, \nu(v) \, dv \; .$$

Remark:

Assume a=0 . $\phi(t)$ is differentiable as a product of differentiable functions. Given $\phi(t)$, $\nu(t)$ can be found solving:

$$\alpha t^{-(\alpha+1)}\ \nu(t) = -\frac{d}{dt}\ \phi(t)/t^{\alpha}$$

$$\nu(t) = \phi(t) - t\phi'(t)/\alpha \quad \text{for } t > 0 \quad .$$

The conclusion of theorem IV.6 holds if and only if $\nu(t)$ is a characteristic function.

Section IV.2 Applications

In this section, the results of section IV.1 will be used to find the distributions of infinite products of uniformly distributed random variables.

Notation

- $G \varepsilon 0\ (F)$ if given a random variable X having distribution F, a random variable Y, independent of X, can be found such that XY has distribution G .
- $F^{(\alpha)}$ denotes the distribution of $U_0^{1/\alpha}$ where U_0 is a $\mathcal{U}[0,1]$ random variable, F_α denotes the distribution of $U_0^{-1/\alpha}$

$$F^{(\alpha)}(x) = \begin{cases} 0 & x < 0 \\ x^{\alpha} & x \varepsilon [0,1] \\ 1 & x > 1 \end{cases}$$

$$F_\alpha(x) = \begin{cases} 0 & x \leq 1 \\ 1 - x^{-\alpha} & x > 1 \end{cases}$$

- Γ_α denotes the Gamma distribution with parameter α and 1

$$\Gamma_\alpha(x) = \begin{cases} 0 & x \le 0 \\ \int_0^x \Gamma(\alpha)^{-1} y^{\alpha-1} e^{-y}\, dy & x > 0 \end{cases}$$

• $\Phi(x)$ denotes the standard normal distribution

$$\Phi(x) = \int_{-\infty}^x (2\pi)^{-1/2} e^{-1/2\, y^2}\, dy$$

• t_ν denotes the t distribution with ν degrees of freedom

$$t_\nu(x) = \int_{-\infty}^x \frac{\Gamma((\nu+1)/2)}{(\pi\nu)^{1/2}\Gamma(\nu/2)} (1+t^2/\nu)^{-1/2(\nu+1)}\, dt$$

• F_{ν_1,ν_2} denotes the F distribution with ν_1 and ν_2 degrees of freedom

$$F_{\nu_1,\nu_2}(x) = \begin{cases} 0 & x \le 0 \\ \int_0^x \frac{\nu_1^{\nu_1/2} \nu_2^{\nu_2/2} \Gamma((\nu_1+\nu_2)/2) y^{1/2(\nu_1-2)} (\nu_2+\nu_1 y)^{1/2(\nu_1+\nu_2)}}{\Gamma(\nu_{1/2})\Gamma(\nu_{2/2})}\, dy & x > 0 \end{cases}$$

• $B_{p,q}$ denotes the Beta distribution with parameter p and q

$$B_{p,q}(x) = \begin{cases} 0 & x < 0 \\ \int_0^x \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1} (1-y)^{q-1}\, dy & x \in [0,1] \\ 1 & x > 1 \end{cases}$$

• W_b denotes the Weibull distribution with parameter b

$$W_b(x) = \begin{cases} 0 & x \le 0 \\ 1 - e^{-x^b} & x > 0 \end{cases}$$

• $\{U_i\}_{i=0}^\infty$ $\{V_i\}_{i=0}^\infty$ denote two independent sequences of independently distributed $U[0,1]$ random variables while U denotes a $U[-1,1]$ random variable. For positive random variables, the results of section IV.1 can be summarized as follows:

Corollary IV.3

Let F be the distribution function of a strictly positive random variable, then:

i) $F \varepsilon 0(F^{(\alpha)})$ if and only if

$$M(t) = -\int_0^t x^\alpha df(x)/(\alpha x^{\alpha-1})$$

is a distribution where f is the density of F; furthermore if Y has distribution M(t) and is independent of U_0 ,

$$U_0^{1/\alpha} Y \text{ has distribution } F ,$$

ii) $F\varepsilon 0(F_\alpha)$ if and only if

$$M(t) = \alpha^{-1} \int_0^t x^{-\alpha} dx^{\alpha+1} f(x)$$

is a distribution function; furthermore if Y has distribution M(t) and is independent of U_0 ,

$$U_0^{-1/\alpha} Y \text{ has distribution } F .$$

Theorem IV.5 Decomposition of a Gamma distribution

The random variable defined by:

$$\lim_{n \to \infty} (n+\alpha) \prod_{i=0}^{n-1} U_i^{1/(i+\alpha)}$$

has distribution $\Gamma_\alpha(x)$.

Proof: The density of Γ_α is:

$$f(x) = \Gamma^{-1}(\alpha)\, x^{\alpha-1}\, e^{-x} \qquad x > 0 \;;$$

since $f(x)/x^{\alpha-1}$ is decreasing,

$$\Gamma_\alpha \varepsilon 0 \quad (F^{(\alpha)}) \text{ and } U_0^{1/\alpha}\, Y_0 \text{ has distribution } \Gamma_\alpha$$

where U_0 and Y_0 are independent and Y_0 has density:

$$m(t) = [\Gamma(\alpha+1)]^{-1}\, t^\alpha e^{-t} \qquad t > 0$$

by corollary IV.3. Iterating this result:

$$\prod_{i=0}^{n-1} U_i^{1/(i+\alpha)}\, Y_{n-1}$$

has distribution $\Gamma_\alpha(x)$ where Y_{n-1} is independent of the U_i's and has distribution $\Gamma_{\alpha+n}$. Now,

$$E(Y_{n-1}/(n+\alpha)) = 1 \text{ and } V(Y_{n-1}/(n+\alpha)) = (n+\alpha)^{-1} \quad ,$$

hence $(\alpha+n)^{-1} Y_{n-1}$ goes in probability to 1 as n tends to ∞ and the theorem is proved.

Corollary IV.4 Multiplication of Gamma random variables

If $\{X_j\}_{j=0}^{k-1}$ is a sequence of independent random variables distributed $\Gamma_{\alpha+j/k}(x)$ then:

$$\left(\prod_{j=0}^{k-1} X_j\right)^{1/k} \text{ is distributed } \Gamma_{k\alpha}(kx) \quad .$$

<u>Proof</u> Let $\{U_{i(j)}\}_{i=0}^{\infty}{}_{j=0}^{k-1}$ be a sequence of independent $U[0,1]$ random variables. Then

$$\lim_{n} n \prod_{i=0}^{n-1} U_{i(j)}^{1/(i+\alpha+j/k)}$$

has distribution $\Gamma_{\alpha+j/k}$. Since each limit is finite almost surely,

$$\prod_{j=0}^{k-1} (\lim_{n} n \prod_{i=0}^{n-1} U_{i(j)}^{1/(i+\alpha+j/k)})^{1/k}$$

$$= \frac{1}{k} \lim_{n} nk \prod_{i=0}^{n-1} \prod_{j=0}^{k-1} U_{(i)j}^{1/(k\alpha+ki+j)} \quad .$$

Using theorem IV.5 the last random variable has distribution $\Gamma_{k\alpha}(kx)$.

<u>Theorem IV.6</u> <u>Decomposition of a normal distribution</u>

The random variable defined by:

$$\lim_{n\to\infty} \sqrt{2n-1} \ \ U \prod_{i=1}^{n-1} U_i^{1/(2i+1)}$$

is distributed $\Phi(x)$.

<u>Proof:</u> Note that if X has distribution $\Phi(x)$, $|X|$ has density:

$$f(x) = 2(2\Pi)^{-1/2} \exp(-1/2\, x^2) \qquad x > 0$$

and $|X|^2$ is distributed $\Gamma_{1/2}(x/2)$. Hence

$$(\lim_{n\to\infty} 2(n+1/2) \prod_{i=0}^{n-1} U_i^{1/(i+1/2)})^{1/2}$$

$$= \lim_{n\to\infty} (2n+1)^{1/2} \prod_{i=0}^{n-1} U_i^{1/2i+1}$$

has the same distribution as $|X|$. Let Y be a random variable distributed independently of X such that $P(Y=\pm 1)=1/2$, $Y|X|$ has distribution $\Phi(x)$. Now YU_0 is distributed as U and the theorem is proved.

Remarks

1) Note that the double exponential distribution is in $0\ (\Phi)$. Using corollary IV.4 (with $\alpha={}^1/2$, $k = 2$) $\sqrt{2}|X_1|X_2^{\ 1/2}$ is distributed Γ_1 , where X_1 , and X_2 are independent random variables distributed Φ and Γ_1 respectively. Hence $\sqrt{2}\ X_1\ X_2^{\ 1/2}$ is distributed double exponential. This technique has been used in Andrews et al. (1972) to generate the double exponential from the normal distribution.

2) Let X be a Γ_α distributed random variable then $Y(X/\alpha)^\alpha$ has density:

$$h_\alpha(x) = [2\Gamma(\alpha+1)]^{-1} \exp\ (\ x^{\ 1/\alpha}/\alpha) \qquad x\varepsilon R$$

and its representation as an infinite product is:

$$\lim_n \ (n/\alpha+1)^\alpha\ U \prod_{i=1}^{n-1} U_i^{\ 1/(1+i/\alpha)} \quad .$$

The next corollary is an easy consequence of theorem IV.5 and IV.6.

Corollary IV.5 Decomposition of a t , an F and a W distribution

Let $\nu\varepsilon N$, then:

$$U\ \nu^{\ 1/2} \prod_{i=0}^{\infty} U_i^{\ 1/(2i+3)}\ V_i^{-1/(2i+\nu)}$$

is distributed t_ν .

Let ν_1 , $\nu_2 \varepsilon N$, then:

$$(\nu_2/\nu_1) \quad \prod_{i=0}^{\infty} U_i^{\ 1/(\nu_1/2+i)}\ V_i^{-1/(\nu_2/2+i)}$$

is distributed F_{ν_1,ν_2} .

Let b > 0 then

$$\lim_{n} n^{1/b} \prod_{i=1}^{n} U^{1/bi}$$

is distributed W_b .

The following corollary gives a relation between two Weibull distributions when the parameter of one distribution is a multiple of the other. The proof is a direct consequence of corollary IV.4

Corollary IV.6

Let $\{X_i\}_{i=0}^{k-1}$ be a sequence of independent random variables such that X_i is distributed $\Gamma_{1/k+i/k}$ $i=0,1,\ldots,k-2$ and X_{k-1} is distributed W_{kb} then:

$$X_{k-1} \prod_{i=0}^{k-2} X_i^{1/kb} \text{ is distributed } W_b(k^{1/b}x) .$$

Theorem IV.7 Decomposition of the Beta distribution

Let p and q be positive numbers, define:

$$Y_i = \begin{cases} U^{1/(i+q)} & \text{with probability } 1-(i+q)/(i+p+q) \\ 1 & \text{with probability } (i+q)/(i+p+q) \end{cases}$$

then $\prod_{i=0}^{\infty} Y_i$ has distribution $B_{p,q}$.

Proof: Let X_1 be distributed Γ_{p+q} . If X_2 has distribution $B_{p,q}$ and X_2 is independent of X_1 then X_1X_2 has distribution Γ_p . To prove this result, one uses the fact that if Y_1 and Y_2 are distributed Γ_p and Γ_q

respectively, then $Y_1/(Y_1+Y_2)$ is distributed $B_{p,p+q}$ independently of $Y_1 + Y_2$.

Now if $\alpha < \beta$, $F^{(\alpha)} \varepsilon\, O(F^{(\beta)})$ and $U_0^{1/\beta}\, Y$ is distributed $F^{(\alpha)}$ where Y has distribution:

$$M(t) = \begin{cases} -\int_0^t u^\beta \, d\, \alpha u^{\alpha-\beta}/\beta & 0 \le t < 1 \\ 1 & t \ge 1 \end{cases}$$

$$= \begin{cases} (1-\alpha/\beta)\, F^{(\alpha)}(t) & 0 \le t < 1 \\ 1 & t \ge 1 \end{cases} .$$

Therefore if

$$Y_i = \begin{cases} V_i^{1/(i+p)} & \text{with probability } 1 - (i+p)/(i+p+q) \\ 1 & \text{with probability } (i+p)/(i+p+q) \end{cases} ,$$

$Y_i \;\; U_i^{1/(i+p+q)}$ and $V_i^{1/(i+p)}$ have the same distribution, and

$$\lim_n n \prod_{i=0}^{n-1} U_i^{1/(i+p+q)} \; Y_i$$

is distributed Γ_p. Reordering the random variables:

$$(\lim_n n \prod_{i=0}^{n-1} U_i^{1/(i+p+q)}) \prod_{i=0}^{\infty} Y_i$$

has distribution Γ_p .

Now, $X_1\, X_2$ and $X_1 \prod_{i=0}^{\infty} Y_i$ have the same distribution. Since X_2 and

$\prod_{i=0}^{\infty} Y_i$ are bounded random variables, they are determined by their moments, using Carleman criterion, Chung (1974) p. 68. Hence X_2 and $\prod_{i=0}^{\infty} Y_i$ have the same distribution. This ends the proof.

Special cases

1) For q=1 , $B_{p,1} = F^{(p)}$ and if:

$$Y_i = \begin{cases} U_i^{1/(i+p)} & \text{with probability } 1 - (i+p)/(i+p+1) \\ 1 & \text{with probability } (i+p)/(i+p+1) \end{cases}$$

$\prod_{i=0}^{\infty} Y_i$ has the same distribution as $U_0^{1/p}$.

2) $B_{p,1-p}$ is the generalized arc sine distribution (see Feller 1971 p. 470). If:

$$Y_i = \begin{cases} U_i^{1/(i+p)} & \text{with probability } 1 - (i+p)/(i+1) \\ 1 & \text{with probability } (i+p)/(i+1) \end{cases}$$

$\prod_{i=0}^{\infty} Y_i$ has distribution $B_{1-p,p}$. If p = 1/2 , $\Pi\, Y_i$ has the same distribution as:

$$(\text{Sin}\quad U_0/2)^2 .$$

Remarks

1) If X is distributed $B_{p,q}$, for any $\alpha\varepsilon R$, X^α and $\prod_{i=0}^{\infty} Y_i$ have the same distribution where:

$$Y_i = \begin{cases} U_i^{\alpha/(i+p)} & \text{with probability } 1 - (i+p)/(i+p+q) \\ 1 & \text{with probability } (i+p)/(i+p+q) . \end{cases}$$

2) If $\prod_{i=0}^{\infty} Y_i$ has distribution $B_{p,q}$, $\prod_{i=k}^{\infty} Y_i$ has distribution $B_{p+k,q}$.

3) Therefore: if X has distribution $B_{p+k,q}$, $\prod_{i=0}^{k-1} Y_i$ X has distribution B_{pq} , while if X and Y are independent and have distribution $B_{p,q}$ $B_{p+q,r}$ XY has distribution $B_{p,q+r}$ see Rao, (1973) p. 168.

Section IV.3 On a new ordering of distributions

Definition IV.3 Ordering 0

A distribution *G* is said to be bigger than F (notation $G \varepsilon 0(F)$) if G can be viewed as the distribution of the product of two independent random variables, XY where Y is strictly positive and X has distribution F.

Remarks

1) If $G \varepsilon 0(F)$

$$G(x) = \int_0^{\infty} F(x/t)\, dM(t)$$

for a certain distribution M.

2) Note that if $G_\sigma(x) = G(x/\sigma)$ $\sigma > 0$, G_σ ε $0(F)$ for any σ since G_σ is the distribution of σXY . Therefore 0 is scale independent.

3) If $F = \Phi$ and

$$Y = \begin{cases} 1 & \text{with probability } c \\ \alpha & \text{with probability } 1-c , \end{cases}$$

$$G(x) = c\Phi(x) + (1-c)\,\Phi(x/\alpha) .$$

The set of these distributions G is known as the Tukey model. It has been used by astronomer Newcomb to describe distributions with a tail heavier than the normal one in 1886.

4) Let $X_1, X_2, \dots, X_n$ be a random sample from the distribution $F(\frac{x}{\sigma})$ where σ is the scale parameter. Suppose σ is a positive random variable with distribution M , then the distribution of the sample is

$$\int_0^\infty F(xt)\, dM(t)$$

and $O(F)$ can be seen as the set of all possible distributions of a random sample from F with randomized scale.

Let

$$\mathcal{F} = \{F : F \text{ is the distribution of a random variable } X \text{ satisfying } E(|X|^\alpha) < \infty \text{ for an } \alpha \neq 0\}$$

Theorem IV.8

The relation O is a weak ordering in $\mathcal{F}$.

Proof: The antisymmetry property is the only one which needs a proof. Suppose $F \varepsilon O(G)$ and $G \varepsilon O(F)$ where $F, G \varepsilon \mathcal{F}$. If X and Y have distribution F and G respectively, there exist two positive random variables, U and V, independent of X and Y such that:

UX has distribution G

VY has distribution F .

Therefore X and UVX has distribution F . Suppose that for $\alpha > 0$, $E(|X|^\alpha) < \infty$. Then $E(|X|^r) < \infty, r \varepsilon [0,\alpha]$. Hence $E((UV)^r) = 1, r \varepsilon [0,\alpha]$. By Cauchy Schwarz inequality,

$$UV = 1 \quad \text{a.s.}$$

and since U and V are independent,

$$U = C \text{ and } V = C^{-1}$$

where C is a constant. Therefore:

$$O(F) = O(G) .$$

<u>Remark</u>

Note that this ordering is not complete. Consider $F^{(1)}$ and F_1 . $F^{(1)}$ has density

$$f^{(1)}(t) = \begin{cases} 1 & t\varepsilon[0,1] \\ 0 & \text{elsewhere} \end{cases}$$

$t^2 f^{(1)}(t)$ is not increasing in $(0,\infty)$ and $F^{(1)} \notin O(F_1)$. F_1 has density:

$$f_1(t) = \begin{cases} 0 & t < 1 \\ t^{-2} & t \geq 1 \end{cases}$$

and $f_1(t)$ is not decreasing in $(0,\infty)$ and $F_1 \notin O(F^{(1)})$. Using a similar argument, one can show that $F^{(\alpha)}$ and F_β are not comparable.

The following theorem is a direct consequence of section IV.2 results.

<u>Theorem IV.9</u>

i) If $\alpha_1 > \alpha_2 > 0$, $\Gamma_{\alpha_2} \varepsilon O(\Gamma_{\alpha_1})$.

ii) If $v_1 > v_2$ where $v_1\ v_2 \varepsilon N$, $t_{v_2} \varepsilon O(t_{v_1})$.

iii) If $r_1 \geq r_2$, $s_1 \geq s_2$ where $s_i\ r_i \in \mathcal{N}$,

$$F_{r_2\ ,\ s_2} \in 0(F_{r_1,s_1}) \ .$$

iv) If $b_1 > b_2 > 0$, $W_{b_2} \in 0(W_{b_1})$.

v) If $\alpha \geq 1/2$, $H_\alpha \in 0(\Phi)$ (h_α is defined page 157) .

vi) If $p_1 > p_2 > 0$ and $0 < q_1 < q_2$, $p_1 - p_2 \in N$

$$B_{p_2,q_2} \in 0(B_{p_1,q_1}) \ .$$

The fact that if $\alpha > \beta$,

$$F^{(\beta)} \in 0(F^{(\alpha)}) \text{ and}$$

$$F_\beta \in 0(F_\alpha)$$

has an interesting consequence.

Corollary IV.7

Let H_β be the distribution of

$$(\lim_n \ C(n) \prod_{i=0}^{n-1} U_i^{1/\alpha(i)})^\beta$$

where $\alpha(x)$ and $C(n)$ are functions defined from N in R . Then if $\beta_1 > \beta_2$

$$H_{\beta_1} \in 0(H_{\beta_2}) \ .$$

The second ordering has been first studied by van Zwet (1970). It has been used by Barlow and Proschan (1966) in reliability theory and by Bickel and Lehmann (1975) in estimation.

<u>Definition IV.4</u> <u>Ordering 0_v</u>

If F and G are distributions symmetric with respect to 0 ,

$$0_v(F) = \{G : G : G^{-1}(t)/F^{-1}(t) \text{ is increasing in } (1/2,1)\}$$

If F and G satisfy F(x) and G(x) = 0 if x < 0 ,

$$0_v(F) = \{G : G^{-1}(t)/F^{-1}(t) \text{ is increasing in } (0,1)\} .$$

As the other ordering, 0_v is scale invariant: if $F_\sigma(x) = F(x/\sigma), \sigma > 0$,

$$0_v(F) = 0_v(F_\sigma) .$$

van Zwet's definition is more restrictive, he has defined

$$0_v(F) = \{G : G^{-1}(t) = K(F^{-1}(t))\}$$

where K is an even convex function such that K(0) = 0 . In the last part of this section the relation between the two orderings will be investigated. The next lemma will be used to prove theorem IV.10.

<u>Lemma IV.1</u>

(A) Let h_1 and h_2 be two positive increasing functions defined in [a,b] where $-\infty < a < b < \infty$ such that

$$h_i(u) = \int_a^u h_i'(y)\, dy, i = 1,2$$

then if $h_1'(u)/h_2'(u)$ is increasing in (a,b), $h_1(u)/h_2(u)$ is increasing in (a,b) .

(B) Let h_1 and h_2 be two positive decreasing functions defined in [a,b] such that:

$$h_1(u) = -\int_u^b h_1'(y)\,dy + M \ , \ M > 0$$

and

$$h_2(u) = -\int_u^b h_2'(y)\,dy$$

then if $h_1'(u)/h_2'(u)$ is increasing in (a,b), $h_1(u)/h_2(u)$ is increasing in (a,b) .

Proof: See Rivest (1976).

Note that using lemma IV.1 (A), if K is convex, K is increasing and if $K(0) = 0$,

$$K(x)/x \text{ is increasing,}$$

this proves that van Zwet's definition is more stringent than the present one.

Theorem IV.10

For any $\alpha > 0$,

i) $0(F^{(\alpha)}) \subset 0_v(F^{(\alpha)})$ and

ii) $0(F_\alpha) \subset 0_v(F_\alpha)$.

Proof: $(F^{(\alpha)})^{-1}(t) = t^{1/\alpha}$. Applying theorem IV.2 iii) $G \in 0(F^{(\alpha)})$ is equivalent to: $(F^{-1}(t))^\alpha$ is convex in $(0,1)$. Therefore

$$F^{-1}(t)^\alpha/t$$

increases in $(0,1)$ and i) is proved. Now,

$$F_\alpha^{-1}(t) = (1-t)^{-\alpha} \ .$$

Take $G \in O(F_\alpha)$, $G^{-1}(t)^{-\alpha}$ is convex in $(0,1)$, this implies that:

$$-\alpha/(G^{-1}(t)^{\alpha+1} \quad g(G^{-1}(t))$$

is increasing in $(0,1)$ or

$$\frac{-1}{-\alpha/(G^{-1}(t)^{\alpha+1} \quad g\,(G^{-1}(t))}$$

is increasing in $(0,1)$. Applying lemma IV.1 (B) ,

$$(1-t)/(G^{-1}(t)^{-\alpha} - \lim_{t\to 1} G^{-1}(t)^{-\alpha})$$

is increasing or:

$$G^{-1}(t)/F_\alpha^{-1}(t) \text{ is increasing.}$$

This theorem clarifies the relation between O and O_v for an elementary set of distributions. In more complicated cases, the problem is intractable. Bickel and Lehmann (1975) have conjectured that

$$O(F) \subset O_v(F)$$

provided the one parameter family of distributions $F(\alpha x)$ have densities with a monotone likelihood ratio.

The converse of theorem IV.10 is not true.

$$\text{Let } F(x) = \begin{cases} (x/2)^{1/2} & x \in [0,{}^1/2] \\ x & x \in [{}^1/2,1] \,. \end{cases}$$

$$f(x) = \begin{cases} (2\sqrt{2x})^{-1} & x \in [0,{}^1/2] \\ 1 & x \in [{}^1/2,1] \end{cases}$$

hence f(x) is not decreasing. Now

$$F^{-1}(t) = \begin{cases} 2t^2 & t \in [0,{}^1/2] \\ t & t \in [{}^1/2,1] \end{cases}$$

and

$$F^{-1}(t)/t = \begin{cases} 2t & t \in [0,{}^1/2] \\ 1 & t \in [{}^1/2,1] \end{cases} .$$

Therefore $F \in 0_v(F^{(1)})$ while $F \notin 0(F^{(1)})$.

Note that theorem IV.10 holds if $F^{(\alpha)}$ and F_α are replaced by the distribution of $\psi_\alpha(U)$ and $(\psi_\alpha(U))^{-1}$ where U is a random variable distributed uniformly in [-1,1] and

$$\psi_\alpha(x) = \text{sgn}(x)\ |x|^\alpha .$$

Section IV.4 Decomposition of a stable distribution

In this section, the following double significance of the Laplace transform will be exploited: let X be a positive random variable with distribution F ,

$$\int_0^\infty e^{-tx}\, dF(x)$$

can be seen as both: the Laplace transform of X and one minus the distribution of Y/X where Y is independent of X and has distribution Γ_1 .

Definition IV.5 Strictly stable distribution

A distribution R is strictly stable if given two arbitrary constants C_1 , C_2 and two independent random variables X_1 , X_2 having distribution R , there exists a constant C such that $C_1 X_1 + C_2 X_2$ and $C X_1$ have the same distribution.

For the properties of these distributions see Feller (1971) pp. 169-176 and Gnedenko and Kolmogorov (1954) pp. 162-171.

Here only strictly stable distributions will be considered, for convenience, they will be called stable distributions.

Theorem IV.11 Decomposition of a positive stable distribution

Define

$$Y_i = \begin{cases} U_i^{-1/\alpha(i+1)} & \text{with probability } 1-\alpha \\ 1 & \text{with probability } \alpha \end{cases}$$

where $\alpha \, \varepsilon (0,1)$, then

$$\lim_n n^{1-1/\alpha} \prod_{i=0}^{n-1} Y_i$$

has a positive stable distribution with parameter α .

Proof: A positive random variable X_α is stable with parameter α if and only if its Laplace transform is equal to:

$$e^{-t^{\alpha}} \quad \text{for } \alpha \in (0,1) \ . \tag{4.6}$$

see Feller (1971) p. 448. In section IV.3, it has been shown that:

$$W_{\alpha} \in 0(\Gamma_1) \text{ provided } \alpha < 1 \ .$$

Therefore there exists a distribution M such that:

$$1 - e^{-t^{\alpha}} = \int_0^{\infty}(1-e^{-tx}) \, dM(x)$$

or

$$e^{-t^{\alpha}} = \int_0^{\infty} e^{-tx} \, dM(x) \ .$$

Using (4.8) and the unicity of the Laplace transform, X_{α} has distribution M and if Y is distributed Γ_1, independently of X_{α}, Y/X_{α} has distribution W_{α}. Using corollary IV.5

$$\lim_n n \prod_{i=0}^{n} U_i^{1/(i+1)} \text{ has distribution } \Gamma_1$$

and

$$\lim_n n^{1/\alpha} \prod_{i=0}^{n-1} V_i^{1/(\alpha i+\alpha)} \text{ has distribution } W_{\alpha} \ .$$

As shown in theorem IV.7

$$F^{(\alpha+\alpha i)} \in 0(F^{(i+1)})$$

and if

$$Y_i = \begin{cases} U_i^{-1/\alpha(1+i)} & \text{with probability } 1-\alpha \\ 1 & \text{with probability } \alpha \ , \end{cases}$$

$Y_i^{-1} U_i^{1/(1+i)}$ and $V_i^{1/(\alpha+\alpha i)}$ have the same distribution. Hence

$$\lim_n n^{1-1/\alpha} \prod_{i=0}^{n-1} Y_i$$

has distribution M(t) and the theorem is proved.

Special case

If $\alpha = k^{-1}$ where $k \in N$, using corollary IV.6 with $b = 1/k$, and if $\{X_i\}_{i=0}^{k-2}$ are independent random variables distributed $\Gamma_{1/k+i/k}$,

$$Yk(\prod_{i=0}^{k-2} X_i k) \text{ is distributed } W_\alpha(x) ;$$

therefore:

$$k^{-1} \prod_{i=0}^{k-2} (X_i\ k)^{-1} \text{ is positive stable with parameter } \alpha = k^{-1} .$$

For k = 2 , if X is distributed $\Gamma_{1/2}$, $^1/4X$ has Laplace transform

$$e^{-t^{1/2}} .$$

This result is due to Lévy (1940).

Corollary IV.8 Moments of a positive stable distribution

Let X_α be a positive stable random variable with parameter $\alpha < 1$, then

$$E(X_\alpha^r) = \frac{\Gamma(1-r/\alpha)}{\Gamma(1-r)} \qquad r \in (-\infty,\alpha) .$$

Proof: If Y is independent of X_α and has distribution Γ_1 then

$$X_\alpha/Y \text{ and } {}^1/Y_1$$

have the same distribution where Y_1 is distributed W_α . Therefore

$$E(X_\alpha^r) = E(Y_1^{-r})/E(Y^{-r})$$

and the result is proved.

Now using the closed representation for stable distributions with parameter k^{-1}, it is possible to obtain a closed representation for a stable distribution with rational parameter.

Theorem IV.12

Let p, $q \in N$, $q > p > 1$ and $\{Z_i\}_{i=1}^{p-1}$ $\{Y_j\}_{j=1}^{q-1}$ be two independent sequences of independent random variables distributed $B_{i/q},\ i(1/p - 1/q)$ $i=1,\ldots,p-1$ and $\Gamma_{j/q}$ $j=1,\ldots,q-1$ respectively. Then:

$$[q/(p) \prod_{i=1}^{p-1} Z_i\, q/(p)]^{-1/p} \, [\prod_{i=p}^{q-1} q\, Y_i]^{-1/p}$$

has Laplace transform $e^{-t^{p/q}}$.

Proof: According to Feller (1971) p. 176, if X and Y are independent positive stable random variables with exponent α and β respectively, the product $XY^{1/\alpha}$ is stable with exponent $\alpha\beta$.

Let X be a positive stable distribution with parameter p/q, and $\{X_i\}_{i=1}^{p}$ be a sequence of independent random variables distributed $\Gamma_{i/p}$, which is independent of $\{Y_i\}$ and $\{Z_i\}$. Using theorem IV.11 $[p\prod_{i=1}^{p-1} X_i p]^{-1}$ is stable with exponent p^{-1} and $[q\prod_{i=1}^{p-1} q\, Z_i X_i \prod_{i=p}^{q-1} q\, Y_i]^{-1}$ is stable with exponent q^{-1} since $Z_i X_i$ is distributed $\Gamma_{i/q}$ by the third remark following theorem IV.7. Applying Feller's proposition with $p^{-1}=\alpha$ $p/q=\beta$:

$$[p\prod_{i=1}^{p-1} X_i p]^{-1} X^p \text{ and}$$

$$[p\prod_{i=1}^{p-1} X_i p]^{-1}[(q/p)\prod_{i=1}^{p-1} Z_i(q/p)]^{-1}[\prod_{i=p}^{q-1} q\ Y_i]^{-1}$$

have the same distribution which implies that:

$$[\prod_{i=1}^{p} X_i p]^{1/p} X^{-1} \text{ and}$$

(4.7)

$$[\prod_{i=1}^{p} X_i p]^{1/p}\{[(q/p)\prod_{i=1}^{p-1} Z_i(q/p)]^{-1/p}\ [\prod_{i=p}^{q-1} q\ Y_i]^{-1/p}\}^{-1}$$

have the same distribution. Using corollary IV.4, $[\prod_{i=1}^{p} X_i p]^{1/p}$ have distribution Γ_1 . Therefore using (4.7) and the fact that one minus the distribution of $[\prod_{i=1}^{p} X_i p]^{1/p}/X$ is the Laplace transform of X ,

$$X \text{ and } [(q/p)\prod_{i=1}^{p-1} Z_i\ (q/p)]^{-1/p}[\prod_{i=p}^{q-1} q\ Y_i]^{-1/p}$$

have the same Laplace transform. The unicity of the Laplace transform concludes the proof.

To obtain similar results for the symmetric stable distributions, note that if X and X_α are independently distributed $\Phi(x)$ and positive stable with exponent $\alpha(\alpha < 1)$ respectively

$$X\ X_\alpha^{1/2}$$

is symmetric stable with exponent 2α (Feller 1971 p. 176). Hence

Corollary IV.9

1) If $\{Z_i\}$, $\{Y_i\}$ denote the same random variable as in theorem IV.12 and if Z is independent of $\{Z_i\}$, $\{Y_i\}$ with distribution $\Phi(x)$

$$Z_{2p/q} = Z[q/p \prod_{i=1}^{p-1} Z_i(q/p)]^{-1/2p}[\prod_{i=p}^{q-1} q\ Y_i]^{-1/2p}$$

has characteristic function $e^{-|t|^{2p/q}}$

2) if Z_α has characteristic function

$$e^{-|t|^\alpha} \quad \alpha \in (0,2) ,$$

$$E(|Z_\alpha|^r) = \frac{2^{r/2}\Gamma((r+1)/2)}{\Gamma(1/2)} \cdot \frac{\Gamma(1-r/\alpha)}{\Gamma(1-r/2)}$$

for any $r \in (-1,\alpha)$.

Proof: The proof of 1) is a straightforward consequence of the last remark. To prove 2), note that:

$$E(|Z|^r) = \frac{2^{r/2}\,\Gamma(.(r+1)/2)}{\Gamma(1/2)} \quad r > -1 ,$$

and if $X_{\alpha/2}$ is positive stable with exponent $\alpha/2$,

$$E(|X_{\alpha/2}|^{r/2}) = \Gamma(1-r/\alpha)/\Gamma(1-r/2) .$$

Section IV.5 Conclusion and remarks

In the second chapter a unified asymptotic theory of L and M estimators has been presented. An important feature of this theory is the fact that it is still valid if a parameter is replaced by its estimator; this feature has been exploited in section II.6 to derive some asymptotic properties of step estimators. It should also lead to a satisfactory proof of asymptotic normality for the estimators of the scale parameter when the location has to be estimated (see Bickel and Lehmann (1976)).

In the linear regression model both the M and the L estimators have been defined (Relles (1968) and Bickel (1971)) and investigated. To introduce the L-M class in this context is of dubious interest. Nevertheless an application of the Newton Raphson method to the convergence proofs in the regression model should be fruitful (cf Bickel (1975)).

The ordering of distribution functions defined in this chapter contains a very wide range of contamination schemes. Its relation to van Zwet's ordering deserve a special attention because of the numerous properties of the latter ordering (see van Zwet (1970)). Furthermore it should be of some interest to make inferences about these orderings (i.e. to test if the underlying distribution function of a random sample is bigger (or less) than a given distribution function).

BIBLIOGRAPHY

Andrews, D.F. et al. (1972). *Robust Estimates of Location: Survey and Advances*. Princeton University Press, Princeton, N.J.

Barlow, R.E. and Proschan, F. (1966). Tolerance and confidence limits for classes of distributions based on failure rate. *Ann. Math. Statist.*, 37, 1592-1601.

Bennett, C.A. (1952). Asymptotic properties of ideal linear estimators. Unpublished Dissertation, University of Michigan.

Bickel, P.J. (1973). Analogues of linear combinations of order statistics in the linear model. *Ann. Statist.*, 1, 597-616.

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

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

Bickel, P.J. and Lehmann, E.L. (1976). Descriptive statistics for non parametric model III. *Ann. Statist.*, 4, 1139-1158.

Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley & Sons Inc., New York.

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

Chung, K.L. (1974). *A Course in Probability Theory*, Second Edition. Academic Press.

Collins, J.R. (1976). Robust estimation of a location parameter in the presence of asymmetry. *Ann. Statist.*, 4, 68-85.

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

Daniell, P.J. (1920). Observations weighted according to order. *American Journal of Mathematics*, 42, 222-236.

Dixon, W.J. and Tukey, J.W. (1968). Approximate behaviour of the distribution of Winsorized t (Trimming/Winsorization 2), *Technometrics*, 10, 83-98.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications. Volume II.* Second Edition. John Wiley & Sons.

Filippova, A.A. (1962). Mises' theorem on the asymptotic behaviour of functionals of empirical distribution functions and its statistical applications. *Theory of Probability and its Applications*, 7, No. 1, 24-57.

Fraser, D.A.S. (1968). *The Structure of Inference.* John Wiley & Sons Inc.

Gnedenko, B.V. and Kolmogorov, A.N. (1954). *Limit Distributions for Sums of Independent Random Variables.* Addison-Wesley, Cambridge, Mass.

Hájek, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives. *Ann. Math. Statist.*, 39, 325-346.

Hájek, J. and Sidák, Z. (1967). *Theory of rank test.* Academic Press.

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. Wahrscheinlichkeits-theorie und Veriv. Gebeite*, 27, 87-104.

Hampel, F.R. (1974). The Influence Curve and Its Role in Robust Estimation. *J. Amer. Statist. Assoc.*, 69, 383-393.

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

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

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

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

Huber, P.J. (1977). *Robust Statistical Procedures.* Regional Conference Series in Applied Mathematics, 27, Society for Industrial and Applied Mathematics.

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

Jung, J. (1958). On linear estimates defined by a continuous weight function. *Arkiv för Matematik*, Band 3 nr 15, 199-209.

Khintchine, A.Y. (1938). On unimodal distributions. (In Russian). *Izv. Nauchno-Issled. Inst. Mat. Mekh., Tomsk. Cos. Univ.*, 2, 1-7.

Lévy, P. (1940). Sur certains processus stochastiques homogènes, *Composition Mathematica*, 7, 283-339.

Lloyd, E.H. (1952). Least squares estimators of location and scale parameters using order statistics. *Biometrika*, 39, 88-95.

Miller, K.S. (1964). *Multidimensional Gaussian distributions.* John Wiley & Sons, Inc.

Mosteller, F. and Tukey, J.W. (1977). *Data Analysis and Regression, A Second course in statistics*. Addison-Wesley.

Olshen, R.A. and Savage, L.J. (1970). A generalized unimodality. *J. Appl. Prob.*, 7, 21-34.

Rao, C.R. (1973). *Linear Statistical Inference and Its Applications*, Second Edition. John Wiley & Sons.

Relles, D.A. (1968). Robust regression by modified least squares. Ph.D. Thesis, Yale University.

Relles, D.A. (1970). Variance reduction techniques for Monte Carlo sampling from Student distributions. *Technometrics*, 12, 499-515.

Rivest, L.P. (1976). Sur la Théorie Asymptotique des Estimateurs d'un Paramètre de Location. Unpublished M.Sc. Thesis, Université de Montréal.

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.

Stigler, S.M. (1969). Linear functions of order statistics. *Ann. Math. Statist.*, 42, 412-427.

Stigler, S.M. (1973). Simon Newcomb, Percy Daniell, and the History of robust estimation 1885-1920. *Journal of the American Statistical Association*, 68, 872-879.

Stigler, S.M. (1974). Linear functions of order statistics with smooth weight functions. *Ann. Statist.*, 4, 676-693.

Tukey, J.W. (1960). A survey of sampling from contaminated distributions. *Contributions to Probability and Statistics*, I. Olkin, ed., Stanford University Press, Stanford, California.

van Zwet, W.R. (1970). *Convex transformations of random variables*. Second Edition. Mathematish Centrum Amsterdam.

Von Mises, Richard (1947). On the asymptotic distributions of differentiable statistical functions. *Ann. Math. Statist.*, 18, 309-348.

APPENDIX I

This appendix contains a listing of the computer program used in section III.3.

```
      CALL RSTART(1294,26876)
      DO 5 III=1,1000
C GENERATION OF THE SAMPLE
      SX=0
      MO=0
      MX=0
      DO 10 I=1,20
      X(I)=RNOR(0)
      Y(I)=1
      IF (UNI(0).LT.0.25) Y(I)=1./10.
      MO=MO+Y(I)**2
      MX=MX+X(I)*Y(I)
      SX=SX+X(I)**2
   10 X(I)=X(I)/Y(I)
      MX=MX/MO
      SX=(SX-((MX**2)*MO))/(19)
      SX=DSQRT(SX)
C   COMPUTATION OF THE CONFIGURATION
      DO 20 I=1,20
   20 X(I)=(X(I)-MX)/SX
C    ORDERING OF THE CONFIGURATION
      DO 30 I=1,20
      M=21-I
      Y(I)=X(1)
      K=1
      DO 40 II=1,M
      IF (X(II).GE.Y(I)) GO TO 40
```

```
      Y(I)=X(II)
      K=II
   40 CONTINUE
   30 X(K)=X(M)
C COMPUTATION OF THE LMI ESTIMATORS
      S(1)=0
      S(2)=0
      S(3)=0
      DO 50 I=6,15
   50 S(1)=S(1)+Y(I)
      DO 60 I=3,5
      J=20-I+1
   60 S(2)=S(2)+Y(I)+Y(J)
      S(3)=Y(2)+Y(19)
      LM(1)=(-S(2)/2.1+S(1))/(20*T(1))
      LM(2)=(-(S(2)+S(3))/5.4+S(1))/(20*T(2))
      LM(3)=(-(S(3)/3.3)+S(1))/(20*T(3))
      LM(4)=(-S(3)+S(2)+S(1))/(20*T(4))
      LM(5)=(-(S(2)+S(3)-Y(16)-Y(5))/(2.25)+S(1))/(20*T(5))
      LM(6)=(-(S(3)*1.9)+S(1)+S(2))/(T(6)*20)
      LM(7)=(-(S(3)*1.8)+S(1)+Y(16)+Y(5))/(T(7)*20)
C   CALCULATION OF THE VARIANCES
      DO 300 I=1,7
  300 VLM(I)=VLM(I)+LM(I)**2+1/MC
    5 CONTINUE
```

RNOR(0) denotes a normal deviate, UNI(0) denotes a random number chosen in [0,1]. These random numbers were generated by the McGill random number generator package Super Duper.

The T(I) involved in the computation of the LM(I) estimates are the normalizing constants (T(1) = .36 etc.).

The variances contained in Table I page 135 are obtained by dividing the VLM(I)'s obtained after 1000 iterations by 50.

Computations were done in double precision on an IBM 370/158.