Trans. Nat. Acad. Sci. & Tech. (Phils.)
1987.9:97-106

A GENERALIZED ASYMPTOTIC THEORY OF MULTIVARIATE L–ESTIMATES*

Roberto N. Padua
Mathematics Department
De La Salle University

ABSTRACT

Statistics which can be expressed as a linear combinations of order statistics, called L–estimates, are considered in this paper. Much of the current theory on this subject deals with the case of univariate and identical parent populations. The present paper considers the general theory in which the parent populations are multivariate which may or may not be identical. The results of the previous authors are then shown as merely special cases of the present investigation in which the dimension is reduced to $p = 1$.

Introduction

The observation that the sample mean is unduly influenced by extreme observations has prompted present-day Statisticians to develop a class of statistics called robust statistics. This new field of Statistics includes the R, M and L estimates. The R estimates are estimates obtained by using the rank scores of the sample values. The M estimates are estimates obtained by minimizing some functions of $X_i - \theta$ where θ is the unknown parameter. On the other hand, L estimates are estimates of the form:

$$(1)\ldots \qquad \hat{\theta} = \sum_{i=1}^{n} C_i X_{(i)}$$

where $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ are the ordered sample values and the C_i's are weights.

Among the proposed competitors of the sample mean, the L estimates are the easiest to implement computationally. The R estimates may sometimes involved complicated mathematics and their efficiency, in general, is more difficult to assess. On the other hand, no closed forms of the M estimates can be given in general. The determination of the M estimates may, for example, involve the use of Newton-Raphson method.

*With the assistance of: Dr. Khursheed Alam, Clemson University, Clemson, South Carolina, U.S.A.

Because of the simplicity and mathematical tractability of the L estimates, much has been written on its asymptotic behavior in the univariate setting. Lloyd (1952) has derived an optimum L estimate for a fixed sample size. The asymptotic analysis has been linked with asymptotic normality through several approaches by Chernoff, Gastwirth, and Johns (1967), Stigler (1969, 1972), Shorack (1969, 1972), Boos (1977, 1979) and others. The asymptotic normality is derived under various restrictions on the underlying distribution from which the sample is drawn and the weights – generating function of the linear combination of the order statistics, giving the L–estimate.

The standard asymptotic theory of L–estimates deals with sample values which are *univariate* and has a common distribution. A few papers have been written on the case of variable distribution such as those by Shorack (1973) and Stigler (1974).

In the present paper, we develop a general asymptotic theory of L–estimates in the multivariate setting wherein the parent populations may or may not be identical. All the results of the previous authors will then be seen as special cases of the present investigation when the dimension is reduced to $p = 1$. Of particular interest in the case of the asymptotic distribution of the sample median which was derived by Mood (1941) and Lehmann (1984) and again by Padua (1986) under various setting.

Section 2 develops the asymptotic theory, Section 3 considers some applications and finally Section 4 gives some directions for future research.

Multivariate Distribution

Let $X_1, \ldots, X_n$ be n independent p–dimensional random variables with *cdf* $F_1, \ldots, F_n$, respectively. Let X_{ij} denote the jth component of X_i and $X_{(1j)} \leq \ldots \leq X_{(nj)}$ denote the ordered values of $X_{1j}, \ldots, X_{nj}$. Let $L = (L^*_{1n}, \ldots, L^*_{pn})'$, where

$$L^*_{jn} = \frac{1}{n} \sum_{1}^{n} C_i X_{(ij)}$$

$$C_i = n \int_{\frac{i-1}{n}}^{i/n} J(u)\, du,$$

J is a bounded integrable function on $[0, 1]$. For $y = (y_1, \ldots, y_p)$, let

$$H_{ij}(y) = \begin{cases} 0 \text{ for } y_1 < X_{1j} \\ \\ 1 \text{ for } y_1 \geq X_{ij} \end{cases}$$

First we consider the *i.i.d.* case when $F_1 = \ldots = F_n = F^*$, say. Let F_j^* denote the *cdf* of X_{ij}. We shall assume that

$$(2)\ldots \qquad \int_{-\infty}^{\infty} (F_i^*(x)\,(1-F_i^*(x))\,)\ dx < \infty, j = 1, \ldots, p.$$

Let $Z_i^* = (Z_{i1}^*, \ldots, Z_{ip}^*)'$, $\mu^* = (\mu_1^*, \ldots, \mu_p^*)$, and $\Sigma^* = (\sigma_{jk}^*)$, where

$$Z_{ij}^* = \int_{-\infty}^{\infty} (H_{ij}(X) - F_j^*(x))\ J(F_j^*(X))\ dx$$

$$\mu_j^* = \int_{-\infty}^{\infty} xJ(F_j^*(x))\ d\ F_j^*(x) \qquad \text{and}$$

$$^*jk = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} J(F_j^*(u))\ J(F_k^*(v))\ (\min(F_j^*(u), F_k^*(v)) - F_j^*(u)\,F_k^*(v)\ dudv.$$

The proof given in Padua (1986) for the derivation of the asymptotic distribution of L_n'' in the univariate case goes through for each component of L_n^*. Thus we have the asymptotic representation of $\sqrt{n}\ (L_n^* - \mu^*)$ as

$$(3)\ldots \qquad Z^* = \frac{-1}{\sqrt{n}} \sum_{1}^{n} Z_i^*.$$

From (3) and the multivariate central limit theorem we have

Theorem 1. Let J be bounded and continuous *a.e.* F^{*-1} on $[0, 1]$. If (2) is satisfied and Σ^* is positive definite then $\sqrt{n}\ (L_n^* - \mu^*) \xrightarrow{L} N(0, \Sigma^*)$ as $n \longrightarrow \infty$.

Theorem 2. Let J be bounded and continuous *a.e.* F^{*-1} on $[0, 1]$, $j = 1, \ldots, p$, such that $J(u) = 0$ for $0 < u < \alpha$ and $\beta < u < 1$. If the α and β quantiles of F_j^* are uniquely defined for each j, and Σ^* is positive definite then

$$\sqrt{n}\ (L_n^* - \mu^*) \xrightarrow{L} N(0, \Sigma^*) \text{ as } n \longrightarrow \infty.$$

Next we consider the non-*i.i.d.* case. Let F_{ij} denote the *cdf* of x_{ij} and let

$$\hat{F}_j^*(x) = \frac{1}{n} \sum_{i=1}^{n} F_{ij}(x),\ j = 1, \ldots, p.$$

We shall assume that $F_j^*(x)$ tends to a limiting distribution $F_j^*(x)$ for each x, as $n \longrightarrow \infty$.

*Proposition 1**. There exists a positive number N, such that

$$\sqrt{n} \int_{-\infty}^{\infty} |\hat{F}_j^*(x) - F_j^*(x)|\, dx \leq N,\ j = 1, \ldots, p.$$

for sufficiently large n.

*Proposition 11**. There exists a function $Q\,(0 < Q(x) \leq 1)$ and positive numbers a and b $(0 < b < 1)$, such that $Q^b(x)$ is integrable, and for sufficiently large n, $F_{nj}(x) \leq Q^2(x)$ for $x \leq -a$ and $1 - F_{nj}(x) \leq Q^2(x)$ for $x \geq a$, $j = 1, \ldots, p$.

*Proposition 111**. As $n \longrightarrow \infty$

$$\sqrt{n} \int_{-\infty}^{\infty} (F_j^*(x) - \hat{F}_j^*(x))\, J(F_j^*(x))\, dx \longrightarrow c_j\ ,$$

$$j = 1, \ldots, p.$$

where the c_j are constants, such that $-\infty < c_j < \infty$.

Let $\tilde{Z}_i = (\tilde{Z}_{i1}, \ldots, \tilde{Z}_{ip})'$ and

$\tilde{\Sigma}_1 = (\tilde{\sigma}_{ijk})$, given by

$$\tilde{Z}_{ij} = \int_{-\infty}^{\infty} (H_{ij}(x) - F_{ij}(x))\, J(F_j^*(x))\, dx$$

$$\sigma_{ijk}{}^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} J(F_j^*(u))\, J(F_k^*(v))\, (\min(F_{ij}(u), F_{ik}(v) - F_{ij}(u) F_{ik}(v))\, dudv.$$

The proof given in Padua (1986) for the derivation of the asymptotic distribution of L_n'' in the case of variable distributions, goes through for each component of L_n''. Thus we have the asymptotic representation of $\sqrt{n}\,(L_n^*, \mu^*)$ as

(4) . . .
$$-\frac{1}{\sqrt{n}} \sum_{1}^{n} \tilde{Z}_1 + \tilde{c}$$

where $\tilde{c} = (\tilde{c}_1, \ldots, \tilde{c}_p)'$. We let

$$(5)\ldots \qquad \frac{1}{n}\sum_{1}^{n}(\tilde{\Sigma}_i) \longrightarrow \tilde{\Sigma}, \text{ as } n \longrightarrow \infty$$

where $\tilde{\Sigma}$ is a positive definite matrix. It is easy to see that Rao's condition (see Rao (1973), p. 147) for the application of the multivariate central limit theorem to the sum (4) is satisfied. Thus we have

Theorem 3. Let J be bounded and continuous *a.e.* F_j^{*-1} on $[0, 1]$ and $\hat{F}_j^*(x) \longrightarrow F_j^*(x)$ for each x, as $n \longrightarrow \infty$, $j = 1, \ldots, p$. If Propositions 1*, 11* and 111* are satisfied and (5) holds then $\sqrt{n}\,(L_n^* - \mu^*) \xrightarrow{L} N(\tilde{c}, \tilde{\Sigma})$, as $n \longrightarrow \infty$. The analogue of Theorem 2.5 in Padua (1986) for the multivariate distribution is given as follows. We omit the proof.

Theorem 4. Let J be bounded and continuous *a.e.* F_j^{*-1} on $[0, 1]$, $j = 1, \ldots, p$, such that $J(u) = 0$ for $0 < u < \alpha$ and $\beta < u < 1$. If $\hat{F}_j^*(x) \longrightarrow F_j^*(x)$ for each x, the α and β quantiles of F_j^* are uniquely defined, $j = 1, \ldots, p$, Propositions 1* and 111* are satisfied and (5) holds, then $\sqrt{n}\,(L_n^* - \mu^*) \xrightarrow{L} N(\tilde{c}, \tilde{\Sigma})$, as $n \longrightarrow \infty$.

Component-wise Sample Median: The sample median is a special case of a univariate L–estimate, and is treated separately from the general case (see, for example, Lehmann (1983), Theorem 5.3.2) in the literature, for simplicity. We consider similarly the component-wise sample median which is a special case of a multivariate L–estimate. For simplicity we assume that n is an odd integer. Let $\tilde{x} = (\tilde{x}_1, \ldots, \tilde{x}_p)'$, where $\tilde{x}_j$ denotes the median value of $x_{1j}, \ldots, x_{nj}$, $j = 1, \ldots, p$. Let F_{ij} and f_{ij} denote the *cdf* and *pdf*, respectively, of x_{ij} and for $b = (b_1, \ldots, b_p)'$ let $f_i(b) = (f_{i1}(b_i), \ldots, f_{ip}(b_p))'$, $F_1(b) = (F_{i1}(b_i), \ldots, F_{1p}(b_p))'$ and $F_{ijk}(b_j, b_k) = P\{x_{ij} \leq b_j, x_{ik} \leq b_k\}$.

Let $\Omega_i(b) = (v_{ijk})$, given by $V_{ijk} = F_{ij}(b_j)\,(1 - F_{ij}(b_j))$

$$V_{ijk} = F_{ijk}(b_j, b_k) - F_{ij}(b_j)\,F_{ik}(b_k).$$

Let $W_i = (W_{i1}, \ldots, W_{ip})'$ and $S_n = \sum_{1}^{n} W_i$, where

$$W_{ij} = \begin{cases} 1 & \text{if } x_{ij} > b_j / \sqrt{n} \\ 0 & \text{otherwise, } j = 1, \ldots, p. \end{cases}$$

We have $E(W_{ij}) = 1 - F_{ij}(b_j / \sqrt{n})$ and $\operatorname{cov}(W_i) = \Omega(b / \sqrt{n})$.

Clearly

$$(6)\ldots \qquad \sqrt{n}\ \overline{X} \leq b \Leftrightarrow S_n \leq \frac{n-1}{2}\ \underset{\sim}{e}$$

where $\underset{\sim}{e} = (1, \ldots, 1)'$ and $\leq$ means component-wise inequality,

$$E(n\underset{\sim}{e} - S_n) = \sum_1^n F_i(b/\sqrt{n})$$

$$= \sum_1^n F_i(\underset{\sim}{0}) + \frac{1}{\sqrt{n}}(b^* f_i(\underset{\sim}{0})) + 0(\sqrt{n})$$

where $a^*B = (a_1 b_1, \ldots, a_p b_p)'$, $\underset{\sim}{0}$ denotes a p–component null vector and

$$\operatorname{cov}(S_n) = \sum_1^n \Omega_i\ (b/\sqrt{n}$$

$$= \sum_1^n \Omega_i\ (\underset{\sim}{0}) + 0(n).$$

It is assumed that the x_{ij} have a continuous density at the origin. We make the following additional assumptions: As $n \longrightarrow \infty$

Assumption 1. $\frac{1}{n} \sum_1^n f_i(\underset{\sim}{0}) \longrightarrow f$

where f is a bound length vector with positive components.

Assumption 2. $\frac{1}{n} \sum_1^n \Omega_i\ (\underset{\sim}{0}) \longrightarrow \Omega$

where Ω is a non-null matrix, and

Assumption 3. $\frac{1}{n} (\sum_1^n) F_i\ (\underset{\sim}{0}) - \frac{n}{2}\ \underset{\sim}{e}) \longrightarrow (\underset{\sim}{0})$.

By the multivariate central limit Theorem (see Rao (1973), p. 147)

$$(7)\ldots \qquad \frac{S_n - E\ S_n}{\sqrt{n}} \xrightarrow{L} N(0, \Omega)$$

under Assumption 2. If Assumptions 1 and 3 are satisfied then from (2.25) we have for large n

$$(8)\ldots \qquad P(\sqrt{n}\ \bar{X} \leq b) = P\left(S_n \leq \frac{n-1}{2}\ e\right)$$

$$= P\left[\frac{S_n - E\,S_n}{\sqrt{n}} \leq -\frac{n+1}{2\sqrt{n}}\ \underset{\sim}{\theta} + \frac{1}{\sqrt{n}} \sum_{1}^{n} F_i(0) + \frac{1}{n} \sum_{1}^{n} b = f_i\ (\underset{\sim}{0})\right]$$

$$= P\left[\frac{S_n - E\,S_n}{\sqrt{n}} \leq b = f\right] .$$

Combining (7) and (8), we get

Theorem 5. If the x_{ij} have a continuous density at the origin and Assumptions 1, 2 and 3 are satisfied, then

$$\sqrt{n}\ \bar{X}^* = f \xrightarrow{L} N(0, \Omega), \text{ as } n \longrightarrow \infty .$$

We can rephrase Assumptions 1, 2 and 3 with reference to an arbitrary vector d in place of the null vector 0, and rephrase the given theorem accordingly, in an obvious manner.

L–estimate of regression coefficients. The study of robust estimation is particularly important for the general regression problem. In this regard Huber (1973) has noted that "just a single grossly outlying observation may spoil the least squares estimate, and moreover, outliers are much harder to spot in the regression than in the simple location case." Various types of robust estimates of the regression coefficients of a linear model have been considered in the literature. M–estimates of the regression coefficients have been considered by Anscombe (1967), Huber (1973) and Bickel (1975), among others. R–estimates of the regression coefficients have been considered by Adichie (1967), Jureckova (1971) and Maritz (1979). Some other types of robust estimates for the simple linear regression model have been proposed by Mood (1950), Theil (1950), Sen (1968) and Forsythe (1972). A type of M–estimate for the regression coefficients has been proposed by Koenker and Bassett (1978).

Consider the linear model

$$(9)\ldots \qquad Y = X\theta + \epsilon$$

where Y is an n–dimensional vector of response variables, X is an nxp matrix non-stochastic variables, θ is a p–dimensional vector of the regression coefficients and ϵ is an n–dimensional vector of errors. The components of ϵ are *i.i.d.* random

variables. We partition X into m submatrices, according to the rows of X. Let X_i denote the ith submatrix and let Y_i and ϵ_i denote the associated subvector of Y and ϵ, respectively. We assume that each X_i is of rank p. Let

$$\begin{aligned} \tilde{\theta}_i &= (X_i' X_i)^{-1} X_i Y_i \\ &= (X_i' X_i)^{-1} X_j' \epsilon_i + \theta \end{aligned}$$

denote the least squares estimate of θ, as obtained from the ith partition of (Y, X). Let $\tilde{\theta}_{ij}$ denote the jth component of $\tilde{\theta}_i$ and let $\tilde{\theta}_{(1j)} \leq \ldots \leq \tilde{\theta}_{(mj)}$ denote the ordered values of $\tilde{\theta}_{ij}, \ldots, \tilde{\theta}_{mj}$. For a robust estimate of θ, consider a multivariate L–estimate $\hat{\theta}$ whose jth component is given by

$$(10)\ldots \qquad \hat{\theta}_j = c_1 \tilde{\theta}_{(1j)} + \ldots + c_k \tilde{\theta}_{(mj)}$$

where c_1 are suitable constants. Simple estimates such as those for which $\hat{\theta}_j$ is a trimmed mean or a median value of $\tilde{\theta}_{ij}, \ldots, \tilde{\theta}_{mj}$ are particularly interesting. We considered the latter estimate for which

$$(11)\ldots \qquad \hat{\theta}_j = \text{median}\,(\tilde{\theta}_{1j}, \ldots, \tilde{\theta}_{mj}).$$

In many practical situations it is reasonable to assume that the components of ϵ in (9) are symmetrically distributed about the origin. We shall make this assumption here. Therefore, the $\tilde{\theta}_{ij}$ are symmetrically distributed about the origin, $j = 1, \ldots, p$ and $i = 1, \ldots, m$. Denoting by F_{ij} and f_{ij} the *cdf* and density function of $\tilde{\theta}_{ij}$, we get $F_{ij}(0) = \frac{1}{2}$. It can be generally assumed that the matrix X and the error distribution are such that the assumptions of Theorem 4 are satisfied. It follows that

$$\sqrt{m}\ (\hat{\theta} - \theta)^* f \xrightarrow{L} N(0, \Omega) \text{ as } m \longrightarrow \infty \text{ , where}$$

Ω is a positive definite matrix whose diagonal elements are each equal to $\frac{1}{4}$.

We need to find the values of f and Ω. Suppose that the rows of X are independently distributed according to a given distribution. Then given the common distribution of the components of ϵ, we can empirically determine the values of f and Ω by the Monte Carlo method, for example. To compare $\hat{\theta}$ with the least squares estimates we compare the covariances of the asymptotic distribution of

$$\hat{\theta} \text{ and } \bar{\theta} = \frac{1}{m} \sum_{1}^{n} \tilde{\theta}_i ,$$

where $\tilde{\theta}_i$ is the least squares estimate of θ, associated with the *ith* partition of X. In this regard we note that $\tilde{\theta}_j$ is distributed with mean θ and covariance

$$\sigma^2 (X_j' X_j)^{-1},$$

where σ^2 denotes the common variance of the component of ϵ. If the rows of X are generated from the normal distribution $N(\underset{\sim}{0}, V)$, then $(X_i' X_j)^{-1}$ is distributed according to the inverted Wishart distribution. Therefore, for large m

$$\frac{1}{m} \sum_{1}^{m} (X_{ii}' X_i)^{-1} \xrightarrow{p} \frac{1}{m} \sum_{1}^{m} E(X_i' X_i)^{-1}$$

$$= \frac{1}{m} \left(\sum_{1}^{m} \frac{1}{k_i - p - 1} \right) V^{-1}$$

where k_i denotes the number of rows in the ith partition of X. It is assumed that $k_i \geq p + 2$ for each i. If all the k_i are nearly equal to k, say, then by the multivariate central limit theorem

$$\sqrt{m}\,(\tilde{\theta} - \theta \xrightarrow{L} N\left(\underset{\sim}{0}, \frac{\sigma^2}{k-p-1} V^{-1}\right)$$

as $m \longrightarrow \infty$.

Future Research

Although the theory presented here is comprehensive, there are still some avenues for future research in L–estimation. Some of the more important ones are as follows:

1. Establish bounds for the error in normal approximation. Such bounds may be of the Berry-Esseen type which gives the maximum error that one may incur using the normal approximation.
2. Develop software packages which will incorporate L estimates of location parameters and regression coefficients.
3. In the application to robust regression, a theoretical research on the optimal block sizes is needed. Moreover, a theoretical research is also needed to see the effect of multicollinearity on the proposed regression estimates.

These are but a few of the research directions which may interest an applied statistician or a mathematical statistician.

Acknowledgment

The author is greatly indebted to his research advisor and internationally-known statistician, Prof. Khursheed Alam of Clemson University, Clemson, South Carolina, for suggesting the topic to him.

References

1. Anscombe, F.J. 1967. Topics in the investigation of linear relations fitted by the method of least squares. *Jour. Royal. Statist. Soc.* Series B 29: 1-52.
2. Adichle, J.N. 1969. Estimates of regression parameters based on rank tests. *Ann. Math. Statist.* 38: 894-904.
3. Benneth, C.A. 1952. Asymptotic properties of ideal linear estimators. Ph.D. Dissertation, University of Michigan.
4. Bickel, P.J. 1975. One-step Huber estimates in the linear model. *Jour. Amer. Statist. Assoc.* 70: 428-434.
5. Boos, D.D. 1979. The differential approach in statistical theory and robust inference. Ph.D. Dissertation, Florida State University.
6. ———————— . 1979. A differential for L–statistics. *Ann. Statist.* 7: 955-959.
7. Chernoff, H., J.L. Gastwirth and M.V. Johns, Jr. 1967. Asymptotic distribution of linear combination of order statistics, with application to estimations, *Ann. Math. Statist.* 38: 52-57.
8. Feller, W. 1966. An Introduction to Probability and Its Applications, Vol. II, Wiley, New York.
9. Forsythe, A.B. 1972. Robust estimation of straight line regression coefficients by minimizing the pth power deviations. *Technometrics* 14: 159-166.
10. Huber, P.J. 1973. Robust regression: asymptotic conjectures and Monte Carlo. *Ann. Statist.* 1: 799-821.
11. Jureckova, J. 1971. Nonparametric estimate of regression coefficients. *Ann. Math. Statist.* 42: 1328-1338.
12. Koenker, R. and G. Bassett. 1978. Regression quantities. *Econometrica* 46: 33-48.
13. Lehmann, E.L. 1983. Theory of Point Estimation. Wiley Series in Probability and Statistics.
14. Maritz, J.S. 1979. On Theil's method in distribution-free regression. *Austral. J. Statist.* 21: 30-35.
15. Moore, D.S. 1968. An elementary proof of asymptotic normality of linear functions of order statistics. *Ann. Math. Statist.* 39: 263-265.
16. Padua, R.N. 1986. A Simple Proof of the Asymptotic Normality of L–Estimates: IID and Non-IID Cases. *The Philippine Statistician,* October 1986, pp.
17. Rao, C.R. 1973. Linear Statistical Inference and Its Applications, 2nd Ed. Wiley, New York.
18. Serfling, R.J. 1980. Approximation Theorems of Mathematical Statistics, Wiley Series in Probability and Mathematical Statistics.
19. Sen, P.K. 1968. Estimates of regression coefficients based on Kendall's Tau. *Jour. Amer. Statist. Assoc.* 63: 1379-1389.
20. Shorack, G.R. 1969. Asymptotic normality of linear combinations of functions of order statistics. *Ann. Math. Statist.* 40: 2041-2050.
21. 1972. Functions of order statistics. *Ann. Math. Statist.* 43: 412-427.
22. Shorack, G.R. 1973. Convergence of reduced empirical and quantile processes with applications to functions of order statistics in the non-i.i.d. case. *Ann. Statist.* 1: 146.152.
23. Stigler, S.M. 1969. Linear functions or order statistics. *Ann. Math. Statist.* 40: 770-778.
24. 1974. Linear functions of order statistics with smooth weight functions. *Ann. Statist.* 92: 676-693. Correction note (1979). *Ann. Statist.* 7: 466.
25. Theil, H. 1950. A rank-invariant method of linear and polynomial regression analysis, I, II and III. *Nederl. Akad. WetenSch. Frac.* 53: 386-392 and 1397-1412.