Listy Biometryczne – Biometrical Letters
Vol. 28 (1991), No. 2, 89–95

Multivariate analysis under M-estimation theory using a convex discrepancy function

C. Radhakrishna Rao and Z.J. Liu

Pennsylvania State University, University Park, PA 16802
Mississippi State University, Mississippi State, MS 39762

Summary

We consider a general multivariate linear regression model $Y_i = X_i'\beta + \varepsilon_i$, $i = 1,...,n$, where Y_i and ε_i are p-vector random variables, X_i is a $q\times p$ design matrix and β is a q-vector of unknown parameters. We develop a general theory for the estimation of β and tests of hypotheses on β using the concepts of M-estimation. Specifically, we consider the estimation of β by minimizing $\sum_1^n \rho\,(Y_i - X_i'\beta)$, where the discrepancy function, ρ, is convex. The special case of the MANOVA model, where X_i has a simple structure, is considered in some detail.

1. Introduction

We consider a general p-variate regression model

$$\mathbf{Y}_i = \mathbf{X}_i'\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \quad i = 1,...,n\ , \tag{1.1}$$

where $\mathbf{Y}_i$ is a p-vector of observable random variables, $\mathbf{X}_i$ is a $q\times p$ design matrix, $\boldsymbol{\beta}$ is a q-vector of unknown parameters and $\boldsymbol{\varepsilon}_1,...,\boldsymbol{\varepsilon}_n$ are iid p-vector noise random variables. The model (1.1) is more general than the usual MANOVA model

$$\mathbf{Y}_i = \mathbf{B}'\mathbf{x}_i + \boldsymbol{\varepsilon}_i, \quad i = 1,...,n\ , \tag{1.2}$$

where $\mathbf{B}$ is an $m\times p$ matrix of regression parameters and $\mathbf{x}_i$ is an m-vector of concomitant (or design) variables. Note that (1.2) can be written as (1.1) defining $\boldsymbol{\beta} = \text{vec}\,\mathbf{B}$ (i.e., obtained by writing the columns of $\mathbf{B}$ one below the other starting from the first) and $\mathbf{X}_i' = \mathbf{I} \otimes \mathbf{x}_i'$ using the notation of a Kronecker product (Rao, 1973, p.29).

Key words: Fisher-Hsu distribution, least distances estimation, M-estimation, MANOVA, Rao's score test, roots of determinantal equation, Wald test

In this paper, we provide a survey of some recent results on estimation and testing of hypotheses concerning the $\boldsymbol{\beta}$ parameters based on M-estimation theory. It may be recalled that one method of obtaining M-estimates is to minimize

$$\sum_{i=1}^{n} \rho(\mathbf{Y}_i - \mathbf{X}_i' \boldsymbol{\beta}) \tag{1.3}$$

with respect to $\boldsymbol{\beta}$ choosing a suitable discrepancy function ρ of a p-vector variable. Another method is to minimize

$$\frac{n}{2} \log |\boldsymbol{\Sigma}| + \sum_{i=1}^{n} \rho\,(\boldsymbol{\Sigma}^{-1/2}(\mathbf{Y}_i - \mathbf{X}_i' \boldsymbol{\beta})) \tag{1.4}$$

with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$. In this survey we consider the method (1.3) choosing ρ to be a convex function. Although such a choice of ρ covers a wide variety of cases, it may not be suitable when we want to exercise a high degree of control over the influence of large values of the error variables $\varepsilon_1,\ldots,\varepsilon_n$. However, the choice of a convex function enables us to define the estimate without ambiguity and to investigate its properties making a minimal set of assumptions on the model (1.2).

The theory of M-estimation started with the seminal work of Huber (1964) and there is now considerable literature on the subject. For references to some important current contributions on the subject, the reader is referred to the bibliographic sections in the papers by Bai, Rao and Wu (1991), Bai, Rao and Zhao (1990), McKean and Schrader (1987), Rao (1988) and Schrader and Hettmansperger (1980).

We make the following assumptions on the model (1.1) and the discrepancy function ρ in (1.4).

(A_1) $\rho(\mathbf{x})$ is a convex function of a p-vector variable $\mathbf{x}$.

(A_2) Let $\psi(\mathbf{x})$ be any choice of the vector derivative of $\rho(\mathbf{x})$ and denote by D the set of discontinuity points of ψ, which is the same for all choices of ψ. Further let F be the common distribution function of $\varepsilon_1,\ldots,\varepsilon_n$, the error vectors in the model (1.1). Then $F(D) = 0$. [This condition is imposed to provide unique values for certain functionals of ψ which appear in our discussion, and it automatically holds when ρ is differentiable. For instance, if $\rho(\mathbf{x}) = |\mathbf{x}|^r$, $r > 1$, the condition does not impose any restriction on F. We conjecture that this condition is crucial for asymptotic normality but not necessarily for establishing consistency of estimates.]

(A_3) For a p-vector $\mathbf{u}$, there exists a positive definite matrix $\mathbf{A}$ such that

$$\mathrm{E}\,[\psi\,(\varepsilon_i + \mathbf{u})\,] = \boldsymbol{\Lambda}\mathbf{u} + o\,(\|\mathbf{u}\|) \quad \text{as} \quad \|\mathbf{u}\| \to \mathbf{0}.$$

(A_4) $g\,(\mathbf{u}) = \mathrm{E}\,\|\,\psi(\varepsilon_i + \mathbf{u}) - \psi(\varepsilon_i)\,\|^2$ exists for all small $\mathbf{u}$ (i.e., $\|\mathbf{u}\|$ is small) and g is continuous at $\mathbf{u} = \mathbf{0}$.

(A_5) $\mathrm{E}\,[\psi(\varepsilon_i)\psi(\varepsilon_i)'] = \boldsymbol{\Gamma}$ (positive definite).

(A_6) $\mathbf{S}_n = \mathbf{X}_1\mathbf{X}_1' + \ldots + \mathbf{X}_n\mathbf{X}_n'$ is nonsingular for $n \geq n_0$ (some value of n) and
$d_n^2 = \max_{1 \leq i \leq n} \operatorname{tr}(\mathbf{X}_i'\mathbf{S}_n^{-1}\mathbf{X}_i) \to 0$ as $n \to \infty$.

2. Main theorems for the general model

We state the main theorems concerning $\hat{\boldsymbol{\beta}}$, the M-estimate of $\boldsymbol{\beta}$ using (1.3), and tests of hypotheses on $\boldsymbol{\beta}$ for the general model. The proofs are given in Bai, Rao and Wu (1991).

Theorem 2.1. Under the assumptions (A_1)–(A_6)

$$\text{(i) } \hat{\boldsymbol{\beta}} \to \boldsymbol{\beta}_0 \text{ (true value) in pr.} \quad , \tag{2.1}$$

$$\text{(ii) } \mathbf{T}^{-1/2}\mathbf{K}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N_q(\mathbf{0}, \mathbf{I}_q) \quad , \tag{2.2}$$

where

$$\mathbf{T} = \sum_{i=1}^{n} \mathbf{X}_i\boldsymbol{\Gamma}\mathbf{X}_i' \quad \text{and} \quad \mathbf{K} = \sum_{i=1}^{n} \mathbf{X}_i\boldsymbol{\Lambda}\mathbf{X}_i' \quad . \tag{2.3}$$

The result (2.1) enables us to provide standard errors of individual estimates, obtain simultaneous confidence intervals for all or subsets of the parameters and test linear hypotheses on $\boldsymbol{\beta}$ in large samples. The distribution given in Theorem 2.2 can be used to test the hypothesis $\mathbf{H}'\boldsymbol{\beta} = \boldsymbol{\gamma}$, where $\mathbf{H}$ is an $r \times q$ matrix of rank r.

Theorem 2.2. Under the assumptions (A_1)–(A_6) the test statistic

$$U_1 = (\mathbf{H}'\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma})' \, (\mathbf{H}'\mathbf{K}^{-1}\mathbf{T}\mathbf{K}^{-1}\mathbf{H})^{-1} \, (\mathbf{H}'\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}) \tag{2.4}$$

is asymptotically distributed as χ^2 (chi-square) on r degrees of freedom.

We call U_1, the Wald type statistic. There is an alternative statistic of the likelihood ratio type

$$U_2 = \sum_{i=1}^{n} \rho\,(\mathbf{Y}_i - \mathbf{X}_i'\tilde{\boldsymbol{\beta}}) - \sum_{i=1}^{n} \rho\,(\mathbf{Y}_i - \mathbf{X}_i'\hat{\boldsymbol{\beta}}) \quad , \tag{2.5}$$

where $\tilde{\boldsymbol{\beta}}$ is the value of $\boldsymbol{\beta}$ which minimizes (1.3) subject to the null hypothesis $\mathbf{H}'\boldsymbol{\beta} = \boldsymbol{\gamma}$. Theorem 2.3 characterizes the asymptotic distribution of (2.5).

Theorem 2.3. The statistic U_2 is asymptotically distributed as

$$\mathbf{Z}'[\mathbf{T}^{1/2}\mathbf{K}^{-1}\mathbf{H}\,(\mathbf{H}'\mathbf{K}^{-1}\mathbf{H})^{-1}\mathbf{H}'\mathbf{K}^{-1}\mathbf{T}^{1/2}]\mathbf{Z} \quad , \tag{2.6}$$

where $\mathbf{Z}$ is an m-dimensional normal variable with independently distributed components. This is, in general, a mixture of chi-square distributions.

There is a third type of statistic which is Rao's score type. Define

$$\xi(\boldsymbol{\beta}) = \sum_{i=1}^{n} \mathbf{X}_i\, \psi(\mathbf{Y}_i - \mathbf{X}'_i\, \boldsymbol{\beta}) \quad , \tag{2.7}$$

which may be called the score function, and let $\tilde{\boldsymbol{\beta}}$ be an estimate of $\boldsymbol{\beta}$ as defined in (2.5).

Theorem 2.4. The Rao's score statistic for testing $\mathbf{H}'\boldsymbol{\beta} = \boldsymbol{\gamma}$ is

$$U_3 = \xi(\tilde{\boldsymbol{\beta}})' \mathbf{T}^{-1} \xi(\tilde{\boldsymbol{\beta}}) \quad , \tag{2.8}$$

which is asymptotically distributed as χ^2 on r degrees of freedom.

It is seen that while the statistics U_1 and U_2 involve both the matrices $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$, the statistic U_3 involves only $\boldsymbol{\Gamma}$. If these are not known, they can be consistently estimated and substituted for the unknown matrices, and the resulting statistics still have the asymptotic distributions mentioned in Theorem 2.1–2.3. For instance an estimate of $\boldsymbol{\Gamma}$ is

$$\hat{\boldsymbol{\Gamma}} = \frac{1}{n} \sum_{i=1}^{n} [\, \psi(\mathbf{Y}_i - \mathbf{X}'_i\hat{\boldsymbol{\beta}})][\psi(\mathbf{Y}_i - \mathbf{X}'_i\hat{\boldsymbol{\beta}})]' \ . \tag{2.9}$$

If ψ is continuously differentiable and its derivative is denoted by a $p{\times}p$ matrix function η, then an estimate of $\boldsymbol{\Lambda}$ is

$$\hat{\boldsymbol{\Lambda}} = \frac{1}{n} \sum_{i=1}^{n} \eta(\mathbf{Y}_i - \mathbf{X}'_i\hat{\boldsymbol{\beta}}) \ . \tag{2.10}$$

As an example let us choose

$$\rho(\mathbf{x}) = \rho(x_1, \ldots, x_p) = (\Sigma x_i^2)^{1/2} = \|\,\mathbf{x}\,\| \, , \tag{2.11}$$

for which

$$\psi(\mathbf{x}) = \begin{cases} \dfrac{\mathbf{x}}{\|\mathbf{x}\|} & \text{if } \ \mathbf{x} \neq \mathbf{0} \, , \\ 0 & \text{if } \ \mathbf{x} = \mathbf{0} \, , \end{cases} \tag{2.12}$$

$$\psi'(\mathbf{x}) = \begin{cases} \dfrac{1}{\|\mathbf{x}\|}\left(1 - \dfrac{\mathbf{x}\mathbf{x}'}{\|\mathbf{x}\|^2}\right) & \text{if } \ \mathbf{x} \neq \mathbf{0} \, , \\ 0 & \text{if } \ \mathbf{x} = \mathbf{0} \, . \end{cases} \tag{2.13}$$

In this case, the estimates of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Lambda}$ are

$$\hat{\boldsymbol{\Gamma}} = \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{\mathbf{e}}_i \hat{\mathbf{e}}'_i}{\|\,\hat{\mathbf{e}}\,\|^2} \ , \tag{2.14}$$

$$\hat{\Lambda} = \frac{1}{n}\sum_{i=1}^{n} \| \hat{e}_i \|^{-1} (I - \frac{\hat{e}_i \hat{e}_i'}{\| \hat{e}_i \|^2}) \quad , \tag{2.15}$$

where $\hat{e}_i = Y_i - X_i' \hat{\beta}$.

The method of estimation based on the discrepancy function (2.11) is called the least distances (LD) method and the related asymptotic theory was developed in Bai, Chen, Miao and Rao (1990).

3. Main theorems for the MANOVA model

Now, we consider a special case of the general model (1.1), usually known as the MANOVA model,

$$Y_i = B'x_i + \varepsilon_i \ , \quad i = 1,...,n \quad , \tag{3.1}$$

where Y_i is p-vector, B is an $m \times p$ matrix of regression parameters and x_i is an m-vector of independent variable. The discrepancy function ρ is chosen to be convex and the conditions (A_1)–(A_5) imposed on the model (3.1) are the same as those in Section 1. The condition (A_6) is changed to

$(A_6)'$ $S_n = x_1 x_1' + \ldots + x_n x_n'$ is positive definite for $n \geq n_o$ (some value) and

$$d_n^2 = \max_{1 \leq i \leq n} x_i' S_n^{-1} x_i \rightarrow 0 \ \text{ as } \ n \rightarrow \infty .$$

We denote by $\hat{B}$ an estimate of B obtained by minimizing

$$\sum_{i=1}^{n} \rho \, (Y_i - B'x_i) \tag{3.2}$$

and by $\tilde{B}$ an estimate of B obtained by minimizing (3.2) subject to given hypothesis $P'B = C_0$ where P is an $m \times r$ given matrix of rank r and C_o is a given $r \times p$ matrix.

We have the following theorems:

Theorem 3.1. Under the assumptions (A_1)–(A_5) and $(A_6)'$,

(i) $$\text{vec}\, S_n^{1/2} (\hat{B} - B_0) \xrightarrow{\mathcal{D}} N_{mp}(0, \Lambda^{-1}\Gamma\Lambda^{-1} \otimes I) \quad , \tag{3.3}$$

(ii) Under the null hypothesis $P'B = C_0$, the $p \times p$ matrix statistic

$$W_m = (P'\hat{B} - C_0)'(P'S_n^{-1}P)^{-1}(P'\hat{B} - C_0) \tag{3.4}$$

has asymptotically central Wishart distribution

$$W_p(r, \Lambda^{-1}\Gamma\Lambda^{-1}) \tag{3.5}$$

i.e., with r degrees of freedom and covariance matrix $\mathbf{\Lambda}^{-1}\mathbf{\Gamma}\mathbf{\Lambda}^{-1}$.

(iii) Let $\hat{\mathbf{\Lambda}}$ and $\hat{\mathbf{\Gamma}}$ be consistent estimates of $\mathbf{\Lambda}$ and $\mathbf{\Gamma}$ respectively. Then, the test criteria for the null hypothesis $\mathbf{P'B} = \mathbf{C}_0$ can be constructed from the roots of the determinantal equation

$$|\mathbf{W}_n - \theta\hat{\mathbf{\Lambda}}^{-1}\hat{\mathbf{\Gamma}}\hat{\mathbf{\Lambda}}^{-1}| = 0 \tag{3.6}$$

and the asymptotic distribution of the roots is the same as that obtained by Fisher and Hsu in the usual least squares MANOVA theory under the assumption of normality.

The reader is referred to Rao (1973, pp. 556–560) for a discussion of the tests based on the roots of (3.6) and references to original papers by Fisher and Hsu.

An alternative approach based on Rao's score statistic is given in Theorem 3.2.

Theorem 3.2. Let

$$\xi(\mathbf{B}) = \sum_{i=1}^{n} \mathbf{x}_i\,[\psi(\mathbf{Y}_i - \mathbf{B}'\mathbf{x}_i)]' \tag{3.7}$$

and define

$$\mathbf{R}_n = \xi(\tilde{\mathbf{B}}')\mathbf{S}_n^{-1}\xi(\tilde{\mathbf{B}}) \quad , \tag{3.8}$$

where $\tilde{\mathbf{B}}$ is an estimate of $\mathbf{B}$ as defined in (3.2). Then:

(i) The asymptotic distribution of R_n is $\mathbf{W}_p(r, \mathbf{\Gamma})$.

(ii) The asymptotic distribution of the roots of the determinantal equation

$$|\mathbf{R}_n = \theta\hat{\mathbf{\Gamma}}| = 0 \tag{3.9}$$

have the same asymptotic distribution as those of the equation (3.6).

It may be noted that the test criteria based on (3.6) involve both $\hat{\mathbf{\Lambda}}$ and $\hat{\mathbf{\Gamma}}$ whereas those based on (3.9) involve only $\hat{\mathbf{\Gamma}}$.

Tests of the type (3.6) and (3.9) have been introduced earlier in particular cases by Sen (1982), Singer and Sen (1985) in the multivariate case and by Schrader and Hettmansperger (1980) in the univariate case. However, there are some noticeable differences between our approach and theirs. We use fewer assumptions which may be due to our choice of the discrepancy function as a convex function. Further we suggest the use of different consistent estimates of $\mathbf{\Lambda}$ and $\mathbf{\Gamma}$.

To complete the theory of M-estimation and inference based on it for the MANOVA model, we need to provide consistent estimates of $\mathbf{\Lambda}$ and $\mathbf{\Gamma}$ independently of the null hypothesis.

A natural estimate of $\mathbf{\Gamma}$ is

$$\hat{\mathbf{\Gamma}} = n^{-1} \sum_{i=1}^{n} [\psi(\mathbf{Y}_i - \hat{\mathbf{B}}'\mathbf{x}_i)][\psi(\mathbf{Y}_i - \hat{\mathbf{B}}'\mathbf{x}_i)]' \quad , \tag{3.10}$$

where $\hat{\mathbf{B}}$ is the unrestricted estimate of $\mathbf{B}$ as defined in (3.2). To estimate $\mathbf{\Lambda}$, we take a $p \times p$ nonsingular matrix $\mathbf{Z}$ and denote its columns by $\zeta_1,...,\zeta_n$ Take $h = h_n > 0$ such that

$$h_n/d_n \to \infty, \quad h_n \to 0 \quad \text{and} \quad \liminf_{n \to \infty} nh_n^2 > 0 \quad , \tag{3.11}$$

where d_n is as defined in assumption (A_6). Define

$$\eta_{ik} = \psi(\mathbf{Y}_i - \hat{\mathbf{B}}'\mathbf{x}_i + h\,\zeta_k) - \psi(\mathbf{Y}_i - \hat{\mathbf{B}}'\mathbf{x}_i - h\,\zeta_k) \quad , \quad i = 1,...,n; \quad k = 1,..., p \quad , \tag{3.12}$$

and use the $p \times p$ matrix

$$\hat{\mathbf{\Lambda}} = (2nh)^{-1} \sum_{i=1}^{n} (\eta_{i1} : \dots : \eta_{ip}) \mathbf{Z}^{-1} \tag{3.13}$$

as an estimate of $\mathbf{\Lambda}$.

Theorem 3.3. Under the assumptions (A_1)–(A_5) and $(A_6)'$ on the model (3.1) $\hat{\mathbf{\Gamma}} \to \mathbf{\Gamma}$ and $\hat{\mathbf{\Lambda}} \to \mathbf{\Lambda}$ in pr. as $n \to \infty$.

The proofs of the theorems in this section are given in Bai, Rao and Zhao (1990).

REFERENCES

Bai, Z.D., Chen, X.R., Miao, B.Q. and Rao, C.R. (1990). Asymptotic theory of least distance estimates in multivariate analysis. *Statistics* **21**, 503-519.

Bai, Z.D., Rao, C.R. and Wu, Y. (1991). M-estimation of multivariate linear regression parameters under a convex discrepancy function. *Statistica Sinica* (in press).

Bai, Z.D., Rao, C.R. and Zhao, L.C. (1990). MANOVA type tests under M-theory for the standard multivariate linear model. Technical Report No. 90-39, Pennsylvania State University.

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

McKean, J.W. and Schrader, R.M. (1987). Least absolute errors analysis of variance. In *Statistical Data Analysis Based on the L*1*-Norm and Related Methods,* Ed. Y. Dodge. North-Holland, 297-305.

Rao, C.R. (1973). *Linear Statistical Inference and Its Applications.* Wiley, New York.

Rao, C.R. (1988). Methodology based on the L1-norm in statistical inference. *Sankhyā* **50**, 289-313.

Schrader, R.T. and Hettmansperger, T.P. (1980). Robust analysis of variance based upon a likelihood ratio criterion. *Biometrika* **67**, 93-101.

Sen, P.K. (1982). On M tests in linear models. *Biometrika* **69**, 245-248.

Singer, J.M. and Sen, P.K. (1985). M-methods in multivariate linear models. *J. Multivariate Anal.* **17**, 168-184.

Received 28 August 1991