The Annals of Statistics
1984, Vol. 12, No. 4, 1298–1309

ASYMPTOTIC BEHAVIOR OF *M*-ESTIMATORS OF p REGRESSION PARAMETERS WHEN p^2/n IS LARGE. I. CONSISTENCY[1]

BY STEPHEN PORTNOY

University of Illinois

Consider the general linear model $Y = x\beta + R$ with Y and R n-dimensional, β p-dimensional, and X an $n \times p$ matrix with rows x_i'. Let ψ be given and let $\hat{\beta}$ be an M-estimator of β satisfying $0 = \sum x_i \psi(Y_i - x_i'\hat{\beta})$. Previous authors have considered consistency and asymptotic normality of $\hat{\beta}$ when p is permitted to grow, but they have required at least $p^2/n \to 0$. Here the following result is presented: in typical regression cases, under reasonable conditions if $p(\log p)/n \to 0$ then $\| \hat{\beta} - \beta \|^2 = \mathscr{O}_p(p/n)$. A subsequent paper will show that $\hat{\beta}$ has a normal approximation in R^p if $(p \log p)^{3/2}/n \to 0$ and that $\max_i | x_i'(\hat{\beta} - \beta) | \to_p 0$ (which would not follow from norm consistency if $p^2/n \to \infty$). In ANOVA cases, $\hat{\beta}$ is not norm consistent, but it is shown here that $\max | x_i'(\hat{\beta} - \beta) | \to_p 0$ if $p \log p/n \to 0$. A normality result for arbitrary linear combinations $a'(\hat{\beta} - \beta)$ is also presented in this case.

1. Introduction. For each n consider a general linear model defined by

$$Y = X\beta + R \tag{1.1}$$

where Y and R are n-dimensional random vectors, X is a $n \times p$ matrix, β is a p-dimensional vector and the coordinates of R are independent and identically distributed. Although X and p depend on n, this dependence will be suppressed in the notation throughout the paper. Let x_i denote the (column) vector in R^p whose coordinates form the ith row of X. The basic questions here concern the asymptotic behavior of M-estimators of β: let $\psi: R \to R$ be given and suppose $\hat{\beta}$ satisfies the vector equation

$$0 = \textstyle\sum_{i=1}^n x_i \psi(Y_i - x_i'\beta). \tag{1.2}$$

Then $\hat{\beta}$ is called an M-estimator.

The asymptotic behavior of $\hat{\beta}$ when p tends to infinity with n has been studied by Huber (1973), Yohai and Maronna (1979) and Ringland (1983) (the last reference giving asymptotic expansions). These papers have provided conditions under which $\hat{\beta} \to \beta$ in an appropriate stochastic manner and $a'(\hat{\beta} - \beta)$ has an asymptotic normal distribution (for appropriate bounded sequences of vectors $a \in R^p$). In appropriate balanced cases Yohai and Maronna (1979) show that $p^2/n \to 0$ is sufficient for consistency and that $p^{5/2}/n \to 0$ is sufficient for

Received October 1981; March 1984.

[1] Research supported in part by National Science Foundation grants MCS-80-02340 and MCS-83-01834.

AMS 1970 *subject classifications*. Primary 62G35; secondary 62E20, 62J05.

Key words and phrases. M-estimators, general linear model, asymptotic normality, consistency, regression, robustness.

normality. Here conditions will be presented with the goal of reducing the growth condition on p as far as possible toward $p/n \to 0$.

Without loss of generality assume $\beta = 0$ throughout the remainder of the paper. Then equation (1.2) becomes

$$0 = \sum_{i=1}^{n} x_i \psi(R_i - x_i' \beta). \tag{1.3}$$

The results here are presented separately for the two basic cases of the general linear model (1.1). The first case (that of linear regression) places rather complicated and seemingly artificial conditions on the design matrix, X. It should be noted that these conditions basically require only that the empirical distribution of the vectors $\{x_i\}$ be near a distribution (in R^p) with an appropriately smooth density. In fact, as shown in Section 4, these conditions will hold in probability whenever $\{x_i\}$ form a sample from a distribution not too concentrated in any fixed direction (equivalently, the directions $\{x_i / \| x_i \|\}$ should be at least somewhat smoothly distributed over the unit sphere). The fundamental result here is the following consistency result: if ψ is increasing, $p(\log p)/n \to 0$, and other relatively mild conditions hold, then $\| \hat{\beta} \|^2 = \mathscr{O}_p(p/n)$. The conditions for the result are stated in Section 2 and the basic results are presented in Section 3.

Asymptotic normality results in the regression case will be presented in a subsequent paper (Portnoy, 1984). Huber (1981) conjectured (on the basis of informal expansions) that additional symmetry conditions may be necessary for asymptotic normality if $p^{3/2}/n$ does not tend to zero. However, there seems to be no way of showing that the error terms in Huber's expansion are small. In the subsequent paper, under stronger conditions the author shows that if $(p \log p)^{3/2}/n \to 0$ then the distribution of $\hat{\beta}$ can be uniformly approximated by a normal distribution in R^p. However, preliminary computations indicate that normality will hold in many nonsymmetric cases, even if $p^{3/2}/n \to +\infty$.

Results for the second case of the general linear model (that of analysis of variance) are also presented in Section 5. In the case of a design with p cells (with cell means, β_j) and n/p observations per cell, classical results (e.g., Huber, 1964) give conditions under which there are consistent, asymptotically normal M-estimators, $\hat{\beta}_j (j = 1, \cdots, p)$. However, clearly $\hat{\beta}_j^2 = \mathscr{O}_p(p/n)$; and, hence, $\| \hat{\beta} \|^2 = \mathscr{O}_p(p^2/n)$ which requires $p^2/n \to 0$ for norm consistency. Here it is shown that if $p(\log p)/n \to 0$ then at least $\max_j | \hat{\beta}_j | = \mathscr{O}_p((p(\log p)/n)^{1/2})$. Also a result giving asymptotic normality for arbitrary linear combinations $a'\hat{\beta}$ is presented. The behavior of $\| \hat{\beta} \|^2$ makes it clear why the two cases of (1.1) should be treated separately, and why conditions in the regression case can not be so simple that they also hold for ANOVA designs.

2. Conditions for consistency in the regression case.

The results of Section 3 require conditions on the rate at which p may go to infinity, on the distribution of R_i and the function, ψ, and conditions on the design matrix.

Condition on p. For the results in Section 3, assume that $p(\log p)/n \to 0$ as $n \to \infty$. For simplicity, the results of Section 3 will actually assume that $p(\log n)/n \to 0$; but this is no stronger: if $p \le \sqrt{n}$, $p(\log n)/n \le (\log n)/\sqrt{n} \to 0$; while if $p \ge \sqrt{n}$, $p(\log n)/n \le 2p(\log p)/n$.

Conditions on R_i and ψ. The following two conditions on ψ and on certain expectations are needed:

P1: ψ is an absolutely continuous function with ψ' bounded satisfying $E\psi(R) = 0$, $E\psi'(R) > 0$, and $E\psi^2(R) \le B < +\infty$. Let c be a constant and define for r real

$$H(c; r) = \inf\{\psi'(r - v): |v| \le c\}. \tag{2.1}$$

P2: There exist constants $b > 0$ and $c > 0$ such that $H(c; \cdot)$ is measurable (hence, $H_i(c) \equiv H(c; R_i)$ is a random variable) and $EH_i(c) \ge b$.

Note that by P1, $H_i(c)$ is bounded (for any c). Also note that P1 and P2 hold for the standard robust ψ functions where ψ' has only finitely many discontinuities.

Conditions on X. The conditions on X are designed to hold in typical regression cases where the rows $\{x_i\}$ of X behave like a sample from a distribution in R^p. In any fixed regression problem, the conditions will hold trivially; but the situation here considers an infinite sequence of regression problems and the conditions restrict the way the sequence, X_n, of design matrices can be constructed. It will be shown in Section 4 that they hold in probability if $\{x_i\}$ are indeed a sample from any of a wide class of distributions for which the distribution of $a'x$ does not depend too strongly on a.

It is also important to note that condition X2 is designed for the situation when $\mathrm{cov}(x_i) = I$. If $\{x_i\}$ form a random sample, the assumption $\mathrm{cov}(x_i) = I$ may be made in reasonable generality since the tranformation $\tilde{x}_i = \sum_n^{-1/2} x_i$, $\theta = \sum_n^{1/2} \beta$ yields an equivalent problem with $\mathrm{cov}(\tilde{x}_i) = I$. If $\{\hat{\theta}_n\}$ is norm consistent and the maximum eigenvalue of $\sum_n$ is bounded, then $\{\hat{\beta}_n\}$ will also be norm consistent (at the same order, $\mathscr{O}_p(p/n)$). Alternatively, for any sequence, $\{\sum_n\}$, if the minimum eigenvalue of $(\tilde{X}'\tilde{X})$ is bounded below by $a \cdot n$ in probability (for some constant a), then (if Theorem 3.2 holds),

$$\left\| \left(\frac{1}{n} X'X\right)^{1/2} (\hat{\beta} - \beta) \right\|^2 = \left\| \left(\frac{1}{n} \tilde{X}'\tilde{X}\right)^{1/2} (\hat{\theta} - \theta) \right\|^2$$

$$= \mathscr{O}_p(\|\hat{\theta} - \theta\|^2) = \mathscr{O}_p\left(\frac{p}{n}\right).$$

The arguments in Section 4 can be used to show that the condition on the minimum eigenvalue will hold in probability if $\{x_i\}$ are a sample from an appropriate distribution in R^p.

For conditions X1 and X2, let

$$I(y, c) = \{i = 1, 2, \cdots, n: |x_i'y| \le c\}, \tag{2.2}$$

and let $\mathscr{S}$ be the ball (in R^p) of radius δ and $\mathscr{S}^*$ be the sphere of radius 1.

X1: for any $c > 0$ there are constants $a > 0$, $\delta > 0$, and $C > 0$ such that for all $\beta \in \mathscr{S}$, $y \in \mathscr{S}^*$, and $n = 1, 2, \cdots$

$$\textstyle\sum_{i \in J} (x_i' y)^2 \geq an \quad \text{where} \quad J = I(\beta, c) \cap I(y, C).$$

X2: for any $c > 0$ and $\varepsilon > 0$ there are constants $\delta' > 0$ and $C > 0$ such that for all $\beta \in \mathscr{S}$, $y \in \mathscr{S}^*$ and $n = 1, 2, \cdots$

$$\textstyle\sum_{i \notin J} (x_i' y)^2 \leq \varepsilon n \quad \text{where} \quad J = I(\beta, c) \cap I(y, C).$$

X3: there is a constant B such that for $n = 1, 2, \cdots$

$$\max\{\| x_i \|^2 : i = 1, 2, \cdots, n\} \leq Bn^2.$$

X4: there is a constant B such that for $n = 1, 2, \cdots$

$$\textstyle\sum_{i=1}^n \sum_{j=1}^p x_{ij}^2 \leq Bpn.$$

3. The basic consistency results.

LEMMA 3.1. *Assume conditions* P1, P2, X1, X2, *and* X3; *and let*

$$p(\log n)/n \to 0.$$

Define $H_i(x_i' \beta) = \inf\{\psi'(R_i - v) : |v| \leq |x_i' \beta|\}$ *as in* P2. *Then there are constants* $a^* > 0$ *and* $\delta > 0$ *such that*

$$P\{\inf\{\textstyle\sum_{i=1}^n (x_i' y)^2 H_i(x_i' \beta) : \|y\| = 1, \|\beta\| \leq \delta\} \geq a^* n\} \to 1 \tag{3.1}$$

as $n \to \infty$.

PROOF. First, let c be as in condition P2, let $I(\beta, c) = \{i : |x_i' \beta| \leq c\}$ and let $J'(\beta, y) = I(\beta, c) \cap I(y, C + 1)$ as in conditions X1 and X2. Then

$$\textstyle\sum_{i=1}^n (x_i' y)^2 H_i(x_i' \beta) \geq \sum_{i \in J'} (x_i' y)^2 H_i(c) - B \sum_{i \notin J'} (x_i' y)^2 \equiv S - T \tag{3.2}$$

where B is a bound on $\psi'(r)$.

First consider S and suppress the argument in $H_i(c)$. By conditions P1 and P2 note that H_i are i.i.d. random variables with $EH_i = d > 0$. Furthermore, since H_i is bounded, $\rho(t) \equiv \log E \exp\{tH_i\}$ has a bounded second derivative ($0 < \rho''(t) < B^2$ where B is a bound on H_i). Therefore,

$$M(t) = E \exp\{-tH_i\} \leq \exp\{-dt + bt^2\} \tag{3.3}$$

for some constant $b > 0$. Now let a, δ and C be given by condition X1 using $c/2$ instead of c, and fix y with $\|y\| = 1$ and β with $\|\beta\| \leq \delta$. Then by the Markov inequality (for any set $J \subset I(y, C)$)

$$\begin{aligned} q(y, \beta) \equiv P\{\textstyle\sum_{i \in J} (x_i' y)^2 H_i \leq \tfrac{1}{2} dan\} &\leq \exp\{\tfrac{1}{2} tadn\} \textstyle\prod_{i \in J} M(-(x_i' y)^2 t) \\ &\leq \exp\{\tfrac{1}{2} tadn - td \textstyle\sum_{i \in J} (x_i' y)^2 + bt^2 \sum_{i \in J} (x_i' y)^4\} \\ &\leq \exp\{\tfrac{1}{2} tadn - td \textstyle\sum_{i \in J} (x_i' y)^2 (1 - (bC^2 t/d))\} \end{aligned}$$

for $t > 0$. So choose t small enough so that $bC^2 t < d/4$. Then, using condition X1, if $J = I(\beta, c/2) \cap I(y, C)$,

$$q(y, \beta) \le \exp\{½\ tdan - ¾\ tdan\} \le \exp\{-a'n\} \tag{3.4}$$

(for $a' = tda/4$).

Now, cover the sphere $\mathscr{S}^* = \{\|y\| = 1\}$ by cubes, $\mathscr{C}(y_k)$ about y_k with sides of length $\varepsilon/n^{5/2}$ (where ε will be chosen shortly). Thus, if $y \in \mathscr{C}(y_k)$, $\|y - y_k\| \le (p\varepsilon^2/n^5)^{1/2}$ and (by X3)

$$\begin{aligned} |(x_i'y)^2 - (x_i'y_k)^2| &\le |x_i'(y - y_k)|\ |x_i'y + x_i'y_k| \\ &\le \|x_i\|\ \|y - y_k\| \cdot 2\ \|x_i\| \cdot 1 \\ &\le 2Bn^2\varepsilon(p/n^5)^{1/2} = B^*\varepsilon\sqrt{p/n} \end{aligned}$$

for some constant B^*. Hence, since $\{H_i\}$ are bounded, ε can be chosen so that for all n and $y \in \mathscr{C}(y_k)$ (for any set J),

$$\left|\sum_{i\in J} (x_i'y)^2 H_i - \sum_{i\in J} (x_i'y_k)^2 H_i\right| \le n \cdot \tilde{B}\varepsilon\sqrt{p/n} \le ¼\ dan. \tag{3.5}$$

Similarly, the ball $\mathscr{S} = \{\|\beta\| \le \delta\}$ can be covered by cubes, $\mathscr{C}'(\beta_k)$ of side ε/n such that for $\beta \in \mathscr{C}'(\beta_k)$

$$|x_i'\beta - x_i'\beta_k| \le \tilde{B}\varepsilon\ \sqrt{p/n}.$$

Thus, ε can be chosen so that for $\beta \in \mathscr{C}'(\beta_k)$ and $y \in \mathscr{C}(y_k)$,

$$J'(\beta, y) \supset I(\beta_k, c/2) \cap I(y_k, C) = J(\beta_k, y_k).$$

Therefore, for $\beta \in \mathscr{C}'(\beta_k)$ and $y \in \mathscr{C}(y_k)$ (using (3.5)),

$$\sum_{i\in J'(\beta,y)} (x_i'y)^2 H_i \ge \sum_{i\in J(\beta_k,y_k)} (x_i'y)^2 H_i \ge \sum_{i\in J} (x_i'y_k)^2 H_i - ¼\ dan.$$

Thus,

$$P\{\textstyle\sum_{i\in J'} (x_i'y)^2 H_i \le ¼\ dan \text{ for some } \beta \in \mathscr{C}'(\beta_k) \text{ and } y \in \mathscr{C}(y_k)\} \le q(y_k, \beta_k).$$

But the number of pairs of cubes needed to cover $\mathscr{S}^* \times \mathscr{S}$ is $N \le (2n^{5/2}/\varepsilon)^p (2n/\varepsilon)^p \le (B_0 n^5)^p$ (for some B_0). Therefore, (by (3.4)),

$$\begin{aligned} &P\{\textstyle\sum_{i\in J'} (x_i'y)^2 H_i \le ¼\ dan \text{ for some } \beta \in \mathscr{S} \text{ and } y \in \mathscr{S}^*\} \\ &\quad\le \textstyle\sum_k q(y_k, \beta_k) \le N \exp\{-a'n\} \\ &\quad= \exp\{-a'n(1 - (5p/n)\log(B_0 n))\} \to 0. \end{aligned}$$

Therefore, the term S in (3.2) satisfies

$$P\{S \ge ¼\ dan \text{ for all } \beta \in \mathscr{S} \text{ and } y \in \mathscr{S}^*\} \to 1.$$

Lastly, for term T in (3.2), from condition X2

$$\sum_{i\notin J'} (x_i'y)^2 \le \sum_{i\notin J} (x_i'y)^2 \le \varepsilon n;$$

and choosing $\varepsilon = ⅛\ da$ yields the desired result (with $a^* = ⅛\ ad$ and δ replaced by $\min(\delta, \delta'))$. □

THEOREM 3.2. *Assume conditions* P1, P2, X1, X2, *and* X3, *and* X4, *and that*

$(p \log n)/n \to 0$. *Let* $F\colon R^p \to R^p$ *be defined by*

$$F_j(\beta) = \sum_{i=1}^n x_{ij}\psi(R_i - x_i'\beta). \tag{3.6}$$

Then there is a root $\hat\beta$ *of the equation* $F(\beta) = 0$ *satisfying*

$$\|\hat\beta\|^2 = \mathscr{O}_p(p/n).$$

PROOF. Result 6.3.4 of Ortega and Rheinboldt (1970, page 163) will be applied; so it suffices to show that $\beta'F(\beta) < 0$ for $\|\beta\|^2 = Bp/n$ in probability. First note that

$$\begin{aligned}\beta'F(\beta) &= \sum_{i=1}^n (x_i'\beta)\psi(R_i - x_i'\beta)\\ &= \sum_{i=1}^n (x_i'\beta)\psi(R_i) - \sum_{i=1}^n (x_i'\beta)\int_0^{x_i\beta} \psi'(R_i - v)\,dv\\ &\equiv A_1 - A_2.\end{aligned}$$

Now $|A_i| \le \|\beta\|\,\|\sum_{i=1}^n x_i\psi(R_i)\|$, and

$$\begin{aligned}E\|\sum_{i=1}^n x_i\psi(R_i)\|^2 &= E\sum_{j=1}^p \sum_{i=1}^n \sum_{\ell=1}^n x_{ij}x_{\ell j}\psi(R_i)\psi(R_\ell)\\ &= \sum_{j=1}^p \sum_{i=1}^n x_{ij}^2 E\psi^2(R_i) \le Bnp\end{aligned}$$

(by conditions P1 and X4). Therefore, using Chebychev's inequality, for any $\varepsilon > 0$ there is a constant B^* such that for all n

$$P\{A_1 \le B^*\sqrt{np}\,\|\beta\| \text{ for all } \beta\} \ge 1 - \varepsilon. \tag{3.7}$$

From the definition of A_2 and Lemma 3.1 (with H_i as defined there),

$$\begin{aligned}A_2 &\ge \sum_{i=1}^n (x_i'\beta)^2 \inf\{\psi'(R_i - v)\colon |v| \le |x_i'\beta|\}\\ &\ge \|\beta\|^2 \inf_{\|y\|=1} \sum_{i=1}^n (x_i'y)^2 H_i(x_i'\beta) \ge a^*n\|\beta\|^2 \qquad (3.8)\\ &\qquad\qquad \text{for all } \beta \text{ with } \|\beta\| \le \delta,\end{aligned}$$

with probability tending to one. Thus, (from (3.7) and (3.8)) there is N such that for $n \ge N$

$$P\{A_1 - A_2 \le B^*\sqrt{np}\,\|\beta\| - a^*n\|\beta\|^2 \text{ for all } \beta \text{ with } \|\beta\| \le \delta\} \ge 1 - 2\varepsilon.$$

Let $\sqrt{B} = 2B^*/a^*$ and choose $N' > N$ so that $B(p/n) \le \delta^2$ for $n \ge N'$. Then

$$\begin{aligned}&P\{\beta'F(\beta) < 0 \text{ for all } \beta \text{ with } \|\beta\|^2 = Bp/n\}\\ &\quad\ge P\{A_1 - A_2 \le -\tfrac{1}{2}Ba^*p \text{ for all } \beta \text{ with } \|\beta\|^2 = Bp/n\} \ge 1 - 2\varepsilon\end{aligned}$$

for $n \ge N'$; and the theorem follows from the result in Ortega and Rheinboldt. □

COROLLARY 3.3. *Under the hypotheses of Theorem* 3.2, $\hat\beta$ *is unique on* $\{\beta\colon \|\beta\| \le \delta\}$ *in probability. If in addition* ψ' *is nonnegative (everywhere), then* $\hat\beta$ *is unique on* R^p *in probability.*

PROOF. For F defined by (3.6), the derivative matrix satisfies

$$(F'(\beta))_{jk} = -\sum_{i=1}^n x_{ij}x_{ik}\psi'(R_i - x_i'\beta).$$

Thus, for any $y \in R^p$, $y'F'(\beta)y = -\sum_{i=1}^n (x_i'y)^2\psi'(R_i - x_i'\beta)$; and Lemma 3.1 immediately implies that $F'(\beta)$ is strictly negative definite on $\{\beta: \|\beta\| \le \delta\}$. Thus, $\hat{\beta}$ is unique on this set. If ψ' is nonnegative, then $F'(\beta)$ is nonpositive definite everywhere and (since it is negative definite on a neighborhood of $\hat{\beta}$) $\hat{\beta}$ is unique on R^p. □

4. Conditions on the design matrix. We now show that the conditions X1, X2, X3, and X4 on the design matrix in Section 2 may be expected to hold when the rows $\{x_i'\}$ of X have an appropriate multivariate distribution. In particular, assume that the row vectors $(x_1', x_2', \cdots, x_n')$ form an i.i.d. sequence with distribution G on R^p. It will be shown that conditions X1, X2, X3, and X4 hold in probability for a wide class of distributions, G. As noted in Section 2, assume that the sequence of covariance matrices, $\{\sum_n = \text{cov}(x_i)\}$ has bounded maximum eigenvalue, so that we may assume $\text{cov}(x_i) = I$. Also assume that $p(\log n)/n \to 0$.

Conditions X3 and X4 can be easily handled by imposing the simple moment condition

$$Ex_{ij}^2 \le B_0 < +\infty \tag{4.1}$$

(for all $i = 1, \cdots, n$ and $j = 1, \cdots, p$). Then $E\|x_i\|^2 \le B_0 p$; and using a (first moment) Chebychev inequality,

$$P\{\max_i \|x_i\|^2 \ge Bn^2\} \le nP\{\|x_i\|^2 \ge Bn^2\} \le \frac{B_0 np}{Bn^2} = \frac{B_0}{B}\frac{p}{n} \to 0.$$

So condition X3 holds in probability. Condition 4.1 and the Weak Law of Large Numbers imply that condition X4 holds in probability.

For conditions X1 and X2 define

$$U_i(c, C) = I(|x_i'\beta| \le c, |x_i'y| \le C) \cdot (x_i'y)^2 \tag{4.2}$$

where the dependence on β and y has been suppressed; and (as before) let $\mathscr{S}$ denote the ball of radius δ and $\mathscr{S}^*$ the sphere of radius 1. Assume that the following condition (on G) holds:

(4.3) for any $c > 0$ and $\varepsilon > 0$, there are constants $\delta > 0$ and $C > 0$ such that for all $\beta \in \mathscr{S}$ and $y \in \mathscr{S}^*$,

$$EU_i(c, C) \ge 1 - \varepsilon.$$

The same argument used in Theorem 3.1 will now be used to prove the following result:

THEOREM 4.1. *Assume $\{x_i\}$ are i.i.d. according to a distribution for which* (4.3) *holds. Then if $(p \log n)/n \to 0$, conditions* X1 *and* X2 *hold in probability.*

PROOF. The argument will be sketched for X1; X2 follows analogously. First, by (4.3) choose δ and C so that with $U_i' = U_i(c - 1/n, C - 1/n)$, $EU_i' \ge b > 0$. As in the proof of Theorem 3.1, there is a constant b' such that for $t > 0$

$$E \exp(-U_i' t) \le \exp\{-tb + b't^2\}.$$

Thus, for fixed (β, y) for any $a < \frac{1}{2} b$,

$$P\{\sum_{i=1}^n U_i' \le 2an\} \le \exp\{-(b - 2a)tn + b'nt^2\} \le \exp\{-a^* n\}$$

for some constant a^*. As in Theorem 3.1, cover $\mathscr{S}$ and $\mathscr{S}^*$ by cubes $\mathscr{C}(\beta_k)$ and $\mathscr{C}(y_k)$ of side $\varepsilon/n^{3/2}$ so that on $\mathscr{C}(\beta_k) \times \mathscr{C}(y_k)$

$$\sum_{i=1}^n U_i \ge \sum_{i=1}^n U_i' - an.$$

Since the number of such pairs of cubes needed for coverage is less than $(Bn)^{3p}$,

$$\begin{aligned} P\{\sum_{i=1}^n U_i \le an \text{ for some } \beta \in \mathscr{S}, y \in \mathscr{S}^*\} &\le (Bn)^{3p} P\{\sum_{i=1}^n U_i' \le 2an \text{ for fixed } (\beta_k, y_k)\} \\ &\le (Bn)^{3p} e^{-a^* n} \to 0 \end{aligned}$$

as in Theorem 3.1; and, hence, X1 holds in probability. A similar proof holds for X2 (where the fact that $E(x_i' y)^2 = \| y \|^2 = 1$ is used). □

REMARK. Condition (4.3) is a condition on the uniformity over β and y of the joint distribution of any pair of linear combinations $(x_i' \beta)$ and $(x_i' y)$. That is, if β and y are fixed, then (by dominated convergence)

$$EU_i \to E(x_i' y)^2 = \| y \|^2 = 1$$

as $\delta \to 0$ and $C \to \infty$. Thus, if U_i tends to be smallest when β and y are orthogonal and the distribution of $(a' x_i)$ does not depend too strongly on the direction, $a/\| a \|$, then (4.3) may be expected to hold. In particular, (4.3) is fairly easy to check if $x_i \sim \mathscr{N}_p(0, I)$; and, in fact, (4.3) will hold if the distribution of x_i is a scale mixture of such normal distributions.

5. Consistency and normality in the ANOVA case. Consider the case of a one-way ANOVA with p cells and n/p observations per cell. Here, the model can be written

$$Y_{ij} = \beta_j + R_{ij}, \quad i = 1, \cdots, n/p, \quad j = 1, \cdots, p, \tag{5.1}$$

where $\{R_{ij}\}$ are i.i.d. Let ψ be given and let $\hat{\beta}_j$ denote the solution (if it exists) of the equation

$$0 = \sum_{i=1}^{n/p} \psi(Y_{ij} - \hat{\beta}_j). \tag{5.2}$$

Elementary extensions of classical methods will be used to obtain the existence of "uniformly" consistent M-estimators, $\{\hat{\beta}_j\}$, under weak conditions. Stronger conditions and more tedious computations will provide asymptotic normality of arbitrary linear combinations $\sum a_j \hat{\beta}_j$ (with $\| a \| = 1$). As before, without loss of generality, assume the true values $\beta_j = 0$.

THEOREM 5.1. *Let ψ be a bounded function such that ψ' is bounded and* *continuous near zero. Suppose* $E\psi(R) = 0$ *and* $E\psi'(R) = d \neq 0$; *and assume* $p(\log p)/n \to 0$. *Then there are solutions* $\{\hat{\beta}_j\}$ *of* (5.2) *and a constant* $B > 0$ *such that for any sequence* $\{\delta_n\}$ *with* $\delta_n \to 0$ *and for* $j = 1, \cdots, p$, *if* $u \leq \delta_n$

$$P\{|\hat{\beta}_j| \geq u\} \leq 2\exp\{-Bu^2 n/p\}. \tag{5.3}$$

As a consequence,

$$P\left\{\max_j |\hat{\beta}_j| \geq \left(\frac{1}{B}\frac{p\log n}{n}\right)^{1/2}\right\} \leq \frac{2p}{n} \to 0. \tag{5.4}$$

PROOF. By the hypotheses on ψ (and the dominated convergence theorem) $(d/du)E\psi(R+u) \to d$ as $u \to 0$. Without loss of generality, suppose $d > 0$. Then there are $\{\delta_n\}$ such that for n large enough and $0 < u \leq \delta_n$,

$$d_n'(u) \equiv E\psi(R-u) \leq -(d/2)u \quad \text{and} \quad d_n''(u) \equiv E\psi(R+u) \geq (d/2)u.$$

Therefore, by the Markov inequality

$$\begin{aligned} &P\left\{\frac{p}{n}\sum_{i=1}^{n/p}\psi(R_{ij}-u) \geq -\frac{d}{4}u\right\} \\ (5.5)\qquad &\leq P\left\{\sum_{i=1}^{n/p}(\psi(R_{ij}-u) - d_n'(u)) \geq \frac{n}{p}\frac{d}{2}u\right\} \\ &\leq \exp\left\{-\frac{nd}{2p}ut\right\}(M(t))^{n/p} \end{aligned}$$

for $t > 0$ where $M(t) = E\{\exp t(\psi(R-u) - d_n'(u))\}$. Since ψ is bounded there is a constant b (independent of u) such that

$$M(t) \leq \exp\{bt^2\}.$$

Thus, the bound in (5.5) becomes

$$\exp\left\{-\frac{nd}{2p}ut + \frac{n}{p}bt^2\right\} \leq \exp\left\{-\frac{d^2u^2}{8b}\frac{n}{p}\right\} \equiv \exp\left\{-Bu^2\frac{n}{p}\right\}$$

for $t = du/(4b)$.

Similarly, for any $j = 1, \cdots, p$ and $|u| \leq \delta_n$,

$$P\left\{\frac{p}{n}\sum_{i=1}^{n/p}\psi(R_{ij}+u) \leq \frac{d}{4}u\right\} \leq \exp\left\{-Bu^2\frac{n}{p}\right\}.$$

Hence, using a standard consistency proof, there is a root $\hat{\beta}_j$ for (5.2) satisfying (5.3).

To obtain (5.4) note that the hypothesis shows that $p(\log n)/n \to 0$. Hence (5.3) holds for $u_n^2 = p(\log n)/(Bn)$ and

$$\begin{aligned} P\{\max_j|\hat{\beta}_j| \geq u_n\} &\leq \sum_{j=1}^{p} P\{|\hat{\beta}_j| \geq u_n\} \\ &\leq 2p\exp\left\{-Bu_n^2\frac{n}{p}\right\} = \frac{2p}{n}. \quad \square \end{aligned}$$

THEOREM 5.2. *Suppose the hypotheses of Theorem* 5.1 *hold, and suppose further that* ψ *is an odd function and is twice continuously differentiable with* $\operatorname{Var} \psi'(R) < +\infty$ *and* $\psi''(u)$ *uniformly bounded. Also let* R_{ij} *have a symmetric distribution and assume* $p(\log p)^2/n \to 0$. *Then for any sequence of vectors* $a = a(n) \in R^p$ *with* $\| a \| = 1$,

$$\sqrt{n/p}\ \textstyle\sum_{j=1}^{p} a_j \hat{\beta}_j \to_{\mathrm{D}} \mathcal{N}(0, \sigma^2)$$

where $\sigma^2 = E\psi^2(R)/(E\psi'(R))^2$ *and where* $\{\hat{\beta}_j\}$ *are given by Theorem* 5.1.

PROOF. First note that there are random variables $\{A_j\}$ such that

(i) $A_j = \sqrt{n/p}\ \hat{\beta}_j$ for $j = 1, \cdots, p$ with probability tending to one,
(ii) A_j is symmetric about zero, and
(iii) $\{A_1, \cdots, A_p\}$ are i.i.d. with EA_j^2 bounded (uniformly in n).

To prove this, let $\varepsilon_n = \sqrt{p(\log n)/(Bn)}$ where B is given in Theorem 5.1, and define

$$A_j = \begin{cases} \sqrt{n/p}\ \hat{\beta}_j & \text{if } \ |\hat{\beta}_j| \le \varepsilon_n \\ 0 & \text{otherwise.} \end{cases}$$

Then $\{A_1, \cdots, A_p\}$ are i.i.d. (since $\{\hat{\beta}_1, \cdots, \hat{\beta}_p\}$ are); and by the symmetry hypotheses and Theorem 5.1, (i) and (ii) follow.

To show (iii), let F denote the c.d.f. of $\hat{\beta}_j$. Then integrating by parts and using Theorem 5.1,

$$\begin{aligned} EA_j^2 &= 2 \int_0^{\varepsilon_n} \frac{n}{p} u^2 dF(u) \\ &= -\frac{2n}{p} u^2(1 - F(u)) \Big|_0^{\varepsilon_n} + \frac{4n}{p} \int_0^{\varepsilon_n} uP\{\hat{\beta}_j \ge u\}\, du \qquad (5.6) \\ &\le \frac{8n}{p} \int_0^{\varepsilon_n} ue^{-Bu^2 n/p}\, du \le 4\, \frac{n}{p} \left(\frac{p}{nB}\right) \int_0^{\infty} ve^{-v^2/2}\, dv = 4/B. \end{aligned}$$

To prove normality let

$$\nu_n = \min\left\{\log \frac{n}{p \log^2 n}, \sqrt{p/2}\right\}. \qquad (5.7)$$

Then $\nu_n \to +\infty$ and $p - \nu_n^2 \to +\infty$. Given $a \in R^p$, let

$$J = \{j = 1, \cdots, p\colon |a_j| \ge 1/\nu_n\}. \qquad (5.8)$$

By (5.7) and (5.8),

$$1 \ge \textstyle\sum_{j \in J} a_j^2 \ge \#J/\nu_n^2 \quad \text{or} \quad \#J \le \nu_n^2. \qquad (5.9)$$

By (i) above, $\sqrt{n/p} \sum a_j \hat{\beta}_j$ has the same asymptotic distribution as the following independent sums;

$$\sqrt{n/p}\ \textstyle\sum_{j \in J} a_j \hat{\beta}_j + \sum_{j \notin J} a_j A_j. \qquad (5.10)$$

By (ii) and (iii) above, $\{A_j: j \notin J\}$ form an i.i.d. sequence with mean zero (symmetry) and finite variance. Since $\max\{|a_j| : j \notin J\} \to 0$ (from (5.7) and (5.8)) and $\# J^c \ge p - \nu_n^2 \to +\infty$ (by (5.9)), the Lindeberg-Feller theorem shows that

$$\sum_{j\notin J} a_j A_j \to_{\mathrm{D}} \mathscr{N}(0, \sigma^2 \textstyle\sum_{j\notin J} a_j^2), \tag{5.11}$$

where the fact that $EA_j^2 \to \sigma^2$ follows from the fact that A_j has the same asymptotic distribution as $\sqrt{n/p}\,\hat{\beta}_j$ which Huber (1964) has shown to be $\mathscr{N}(0, \sigma^2)$.

To handle the first term in (5.10) the standard Taylor's series expansion for $0 = \sum_i \psi(R_{ij} - \hat{\beta}_j)$ yields

$$d\sqrt{n/p} \textstyle\sum_{j\in J} a_j\hat{\beta}_j = \sum_{j\in J} a_j \sqrt{p/n} \sum_{i=1}^{n/p} \psi(R_{ij}) - D_n \tag{5.12}$$

where $d = E\psi'(R)$ and for some $|T_{ij}| \le |\hat{\beta}_j|$,

$$\begin{aligned} D_n &= \sqrt{n/p} \textstyle\sum_{j\in J} a_j\hat{\beta}_j\, (p/n) \sum_{i=1}^{n/p} (\psi'(R_{ij}) - d) \\ &\quad + \tfrac{1}{2}\sqrt{n/p} \textstyle\sum_{j\in J} a_j\hat{\beta}_j^2\, (p/n) \sum_{i=1}^{n/p} \psi''(R_{ij} - T_{ij}) \\ &\equiv D_n' + D_n''. \end{aligned} \tag{5.13}$$

Using Cauchy-Schwarz and (5.4) there is $B' > 0$ such that with probability tending to one,

$$|D_n'| \le B'(\textstyle\sum_{j\in J} a_j^2)^{1/2}\sqrt{\log n}\,(\sum_{j\in J} C_j^2)^{1/2}$$

where C_j is a sample average of n/p sample values $(\psi'(R_{ij}) - d)$. Thus,

$$E\textstyle\sum_{j\in J} C_j^2 = (\# J)(p/n)\mathrm{Var}\,\psi'(R).$$

Therefore, with probability tending to one, for some B^*,

$$|D_n'| \le B^* \sqrt{\# J}\,(p/n \log n)^{1/2}. \tag{5.14}$$

Again, using Cauchy-Schwarz, (5.4) and the bound on $\psi''(u)$, there is B' such that with probability tending to one

$$|D_n''| \le B' \sqrt{n/p}\,(\textstyle\sum_{j\in J} a_j^2)^{1/2}(\sum_{j\in J} \hat{\beta}_j^4)^{1/2} \le B'((p/n)\log^2 n)^{1/2}\sqrt{\# J}. \tag{5.15}$$

Therefore, from (5.7) and (5.9) the bounds in (5.14) and (5.15) tend to zero; and, hence, $D_n \to_P 0$. Thus, it remains to consider the first term on the right-hand side of (5.12):

$$\textstyle\sum_{j\in J} a_j \sqrt{p/n} \sum_{i=1}^{n/p} \psi(R_{ij}) = \sqrt{p/n} \sum_{i=1}^{n/p} U_i,$$

where $U_i = \sum_{j\in J} a_j\psi(R_{ij})$. By hypothesis on ψ, $\{U_1, \cdots, U_{n/p}\}$ are i.i.d. with $EU_i = 0$ and $\mathrm{Var}\,U_i = \sum_{j\in J} a_j^2 E\psi^2(R)$. Thus, the Central Limit Theorem holds and

$$\sqrt{n/p} \textstyle\sum_{j\in J} a_j\hat{\beta}_j \to_D \mathscr{N}(0, \sigma^2 \sum_{j\in J} a_j^2). \tag{5.16}$$

The result now follows from (5.10), (5.11), and (5.16). □

REMARKS. (1) The existence of consistent roots of (5.2) (Theorem 5.1) can be proven under slightly weaker hypotheses. Suppose that ψ is a continuous function with $\psi(R)$ having a finite moment generating function and zero mean, and that ψ satisfies the condition: $E\psi(R+u)$ is a strictly monotonic function for u in a neighborhood of zero. Then the existence of consistent M-estimators $\{\hat{\beta}_j\}$ such that $\max_j|\hat{\beta}_j| \to_P 0$ follows if $(p \log n)/n \to 0$. However, these weaker hypotheses do not appear to be sufficient to obtain asymptotic normality of general linear combinations of $\{\hat{\beta}_j\}$.

(2) The proof of Theorem 5.2 can be made more direct by showing that $E(\hat{\beta}_j)^k < +\infty$ (for fixed even k). This can be proved using large deviation results if ψ is monotonic and bounded and if the c.d.f., F, of R decreases algebraically in the tails. However, if ψ is redescending quickly enough or if $1 - F(u) \approx 1/\log u$ (for u large) then $|\hat{\beta}_j|$ will not have finite moments. In fact, if ψ is monotonic and F smooth enough, large deviation results show that

$$P\{\text{there is a root of (5.2) in } (-u, u)\} \approx (1 - F(u))^{n/2p},$$

and, hence, the existence of moments of $|\hat{\beta}_j|$ depends very strongly on F. The proof of Theorem 5.2 avoids this problem by a truncation argument.

(3) The rate of convergence to normality in Theorem 5.2 is often of order $\sqrt{p/n}$. However, even if p grows rapidly, if $\{a_j\}$ are approximately equal (i.e., $a_j \sim 1/\sqrt{p}$) then the rate of convergence of $a'\beta$ is of order $1/\sqrt{p}$.

REFERENCES

HUBER, P. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101.

HUBER, P. (1973). Robust regression: asymptotics, conjectures, and Monte Carlo. *Ann. Statist.* **1** 799–821.

HUBER, P. (1981). *Robust Statistics.* Page 170. Wiley, New York.

ORTEGA, J. and RHEINBOLDT, W. (1970). *Iterative Solutions of Non-linear Equations in Several Variables.* Page 163. Academic, New York.

PORTNOY, S. (1984). Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. II: Normality. Unpublished manuscript.

RINGLAND, J. (1983). Robust multiple comparisons. *J. Amer. Statist. Assoc.* **78** 145–151.

YOHAI, V. J. and MARONNA R. A. (1979). Asymptotic behavior of M-estimators for the linear model. *Ann. Statist.* **7** 258–268.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ILLINOIS
1409 WEST GREEN STREET
URBANA, ILLINOIS 61801