MAXIMUM LIKELIHOOD ESTIMATION IN REGRESSION WITH UNIFORM ERRORS

Herbert Robbins

Brookhaven National Laboratory

and

Cun-Hui Zhang

State University of New York at Stony Brook

The simple linear regression model $y = \alpha + \beta x + \varepsilon$ with i.i.d. uniform errors is considered, and some properties of the maximum likelihood estimators (MLE's) of α and β are derived. In particular, the asymptotic mean square error of the MLE of β when α is known to be zero is proportional to $(\Sigma_1^n |x_i|)^{-2}$ instead of to $(\Sigma_1^n x_i^2)^{-1}$ as it is for the usual least squares estimator (LSE). The MLE's are also superefficient compared with the LSE's when both α and β are unknown.

1. Introduction.

Consider the simple linear regression model with i.i.d. errors

$$y_i = \alpha + \beta x_i + \varepsilon_i, \quad i=1,2,..., \tag{1.1}$$

where we are interested in estimating the parameters α and β. The usual LSE's of α and β are MLE's when the ε_i are normal, but not when the normality assumption fails to hold. We shall obtain some properties of MLE's when

*Research supported under contract DE-AC02-76CH00016 with the U.S. Department of Energy, and by the National Science Foundation.

AMS 1980 subject classifications. Primary 62J05, 62F12, 62F11; Secondary 62E20.

Key words and phrases. Linear regression, maximum likelihood estimator, superefficiency, rate of convergence.

the errors are uniform $(-\Theta/2, \Theta/2)$.

There are three cases of interest: the one-parameter model with known α and Θ, the two-parameter model with known α, and the three-parameter model with unknown α, β and Θ. We shall always assume $\alpha = 0$ in the one and two-parameter models without loss of generality, so that the regression line $y = \beta x$ passes through the origin.

Let

$$(1.2) \qquad b_-(t) = b_{-,n}(t) = \max_{\substack{1 \le i \le n \\ x_i \ne 0}} [y_i/x_i - t/|x_i|], \quad t \ge 0,$$

and

$$(1.3) \qquad b_+(t) = b_{+,n}(t) = \min_{\substack{1 \le i \le n \\ x_i \ne 0}} [y_i/x_i + t/|x_i|], \quad t \ge 0.$$

In the one-parameter model, a statistic b_n is an MLE of β if and only if

$$(1.4) \qquad b_-(\Theta/2) \le b_n \le b_+(\Theta/2) \quad \text{a.s.} \quad .$$

Since $b_+(\Theta/2) - \beta$ and $\beta - b_-(\Theta/2)$ are two identically distributed nonnegative random variables and the observations y_i with $x_i \ne 0$ are sufficient for β, we shall estimate β by

$$(1.5) \qquad b_n' = b_n(\Theta/2), \text{ where } b_n(t) = (b_+(t) + b_-(t))/2.$$

It will be shown in Theorem 1 below that the estimator b_n' possesses certain optimality properties.

In the two-parameter model, a statistic b_n is an MLE of β if and only if

$$(1.6) \qquad b_-(w_n) \le b_n \le b_+(w_n), \quad \text{a.s.},$$

where w_n is the MLE for $\Theta/2$ given by

$$w_n = \min_t [\max_{1 \leqslant i \leqslant n} |y_i - tx_i|]. \tag{1.7}$$

When the x_i are all non-zero,

$$b_+(w_n) = b_-(w_n), \quad \text{a.s.,} \tag{1.8}$$

and the unique MLE b_n is also given by

$$\max_{1 \leqslant i \leqslant n} |y_i - b_n x_i| = w_n. \tag{1.9}$$

In the three-parameter model, statistics a_n and b_n are MLE's of α and β if and only if

$$\max_{1 \leqslant i \leqslant n} |y_i - a_n - b_n x_i| = \min_{s,t} [\max_{1 \leqslant i \leqslant n} |y_i - s - tx_i|]. \tag{1.10}$$

When $\alpha = 0$ and $x_i = 1$ for all i, the models reduce to the classical location-scale case in which

$$\begin{aligned} b_n' = b_n(w_n) &= [(\max_{1 \leqslant i \leqslant n} y_i) + (\min_{1 \leqslant i \leqslant n} y_i)]/2 \\ &= \text{midrange of the } y_i\text{'s} \end{aligned} \tag{1.11}$$

and

$$\begin{aligned} 2w_n &= (\max_{1 \leqslant i \leqslant n} y_i) - (\min_{1 \leqslant i \leqslant n} y_i) \\ &= \text{range of the } y_i\text{'s}. \end{aligned} \tag{1.12}$$

Again, we do not have a unique MLE in the one-parameter case. A statistic b_n is an MLE if and only if it lies between $b_-(\Theta/2)$ and $b_+(\Theta/2)$, and it turns out that

$$b_-(\Theta/2) = (\max_{1 \leqslant i \leqslant n} y_i) - \Theta/2 \text{ and} \tag{1.13}$$

(1.14) $$b_+(\Theta/2) = (\min_{1 \le i \le n} y_i) + \Theta/2.$$

It is well known that

(1.15) $$n(b_n' - \beta)/\Theta \text{ has the limiting density } \exp[-2|t|]$$

and

(1.16) $$\lim n^2 E(b_n' - \beta)^2 = \Theta^2/2.$$

The results in this paper may be regarded as an extension of these facts.

We summarize the properties of MLE's for the one, two, and three-parameter models in Theorems 1, 2 (and 2'), and 3, which are proved in Sections 2, 3 and 4 respectively. In Section 5 we consider the case when the empirical distribution of x_1 to x_n converges, and give a number of examples.

THEOREM 1. Let $y_1, y_2, \ldots$ be given by (1.1) with $\alpha = 0$ and known $\Theta > 0$. Let b_n be any MLE for β given by (1.4) and let b_n' be given by (1.5).

(i)

(1.17) $$E[(b_n - \beta)/\Theta]^2 \le 4/(\sum_{i=1}^{n} |x_i|)^2.$$

(ii) The statistic b_n' is an MLE for β based on $y_1, \ldots, y_n$, $(b_n' - \beta)/\Theta$ has a symmetric distribution which does not depend on the parameter β and the value of Θ, and

(1.18) $$E[(b_n' - \beta)/\Theta]^2 \le 1/(\sum_{i=1}^{n} |x_i|)^2.$$

(iii) The following two statements are equivalent:

(1.19) $$\sum_{i=1}^{\infty} |x_i| = \infty$$

(1.20) There exist a sequence of statistics $\delta_n = \delta_n(y_1,\ldots,y_n,\Theta)$ and two numbers β_1 and β_2 such that $\delta_n \to \beta_i$ in probability when $\beta = \beta_i$, $i=1,2$, and $0 < |\beta_1 - \beta_2| < \Theta/\max_n|x_n|$ when $\max_n|x_n| < \infty$.

(iv) Let $\delta_n = \delta_n(y_1,\ldots,y_n,\Theta)$ be a sequence of statistics. If

(1.21) $$\max_{1\leq i\leq n}|x_i|/\Sigma_{i=1}^n|x_i| \to 0 \text{ as } n \to \infty ,$$

then the set

(1.22) $$B = \{ \beta: \ \limsup_n E_\beta(\delta_n - \beta)^2/E_\beta(b_n' - \beta)^2 < 1\}$$

has Lebesgue measure zero.

(v) If (1.21) holds, then

(1.23) $$\lim_n \{ (d/dt)P\{ (\Sigma_{i=1}^n|x_i|)(b_n' - \beta)/\theta \leq t\} \} = e^{-2|t|},$$

and

(1.24) $$\lim_n (\Sigma_{i=1}^n|x_i|)^2 E[(b_n' - \beta)/\theta]^2 = 1/2 .$$

Remarks: (i) and (ii) give bounds for the mean square errors of MLE's for β. It follows from (iii) that (1.19) is a minimal condition for the existence of a consistent estimator for β whether θ is known or unknown. Actually, if (1.19) fails to hold, it is impossible to have an estimator that is consistent at even two points with big enough difference. It is shown by (iv) that b_n' is asymptotically optimal and asymptotically locally minimax when θ is known. (v) is the extension of (1.15) and (1.16) of the classical location-scale model.

THEOREM 2. Let $y_1, y_2, \ldots$ be given by (1.1) with $\alpha = 0$ and unknown β and θ.

(i) Assume that b_n is an MLE for β. Then (1.17) holds.

(ii) Let the MLE of $\theta/2$, w_n, be given by (1.7). Then $1/2 - w_n/\theta$ has a nonnegative distribution that does not depend on the parameters β and θ, and

(1.25) $$P\{1/2 - w_n/\theta > t\} < 2\exp[-nt] \text{ for any } t > 0.$$

THEOREM 2'. Let $y_1, y_2, \ldots$ be given by (1.1) with $\alpha = 0$ and unknown β and θ. Suppose that $x_i \neq 0$ for every i. Let b_n be the unique MLE for β given by (1.9).

(i) The statistic b_n is almost surely uniquely defined by (1.9) for each n, $(b_n - \beta)/\theta$ has a distribution symmetric about zero that does not depend on the parameters β and θ, and (1.17) holds.

(ii) The following two statements are equivalent to (1.19):

(1.26) $$\lim_n (b_n - \beta) = 0 \text{ in probability}$$

(1.27) $$\limsup_n \{ (\Sigma_{i=1}^n |x_i|)|b_n - \beta|/\log(\Sigma_{i=1}^n |x_i|) \} < e\,\theta, \text{ a.s.}$$.

(iii) Suppose that (1.21) holds. Then as $n \to \infty$,

(1.28) $$(d/dt)P\{ (\Sigma_{i=1}^n |x_i|)(b_n - \beta)/\theta < t \}$$

$$= (1 + o(1)) \int_1^\infty (y/2)e^{-y|t|} dG_n(y) ,$$

where the distribution function G_n assigns probability $2n(n + i)^{-1}(n + i - 1)^{-1}$ to $(n+i)z_i + 1 - s_i$, and z_i, s_i are defined for each n as follows:

(1.29) $$s_i = \Sigma_{j=1}^i z_j, \quad i = 1, \ldots, n$$

(1.30) $(z_1, \ldots, z_n)$ is the permutation of $\{|x_i|/\Sigma_{j=1}^n |x_j|,\ i=1,\ldots n\}$

for which $z_1 \leq z_2 \leq \ldots \leq z_n$.

G_n is such that

(1.31) $\quad G_n(1) = 0,\ G_n(c) \geq (c - 3)/(c - 2)$ for $c \geq 3$.

COROLLARY 1. Suppose that $x_i \neq 0$ for every i and (1.21) holds. Let b_n be defined by (1.9) and G_n be the same as in (iii) of Theorem 2'. Then

(1.32) $$\liminf_n (\Sigma_{i=1}^n |x_i|)^2 E[(b_n - \beta)/\theta]^2 \geq 1/2 ,$$

(1.33) $$\limsup_n (\Sigma_{i=1}^n |x_i|)^2 E[(b_n - \beta)/\theta]^2 \leq 2 ,$$ and

(1.34) $$1/2 \leq \int_1^\infty 2y^{-2} dG_n(y) \leq 2.$$

COROLLARY 2. Let $\alpha = 0$. Then $(b_n(w_n) - \beta)/w_n$ has a distribution symmetric about zero that does not depend on β and θ, where $b_n(t)$ and w_n are defined by (1.5) and (1.7) respectively. Furthermore, under the conditions of (iii) of Theorem 2',

(1.35) $$P\{ (\Sigma_{i=1}^n |x_i|)|b_n - \beta|/w_n \geq t \} = (1 + o(1))\int_1^\infty \exp[-yt]\, dG_n(y) \ , \text{ for any } t > 0,$$

where $G_n(\cdot)$ is defined by (1.28) through (1.30).

Remarks: (iii) of Theorem 2' is again an extension of (1.15) and (1.16) of the classical location-scale model. When $x_i = 1$ for every i, the distribution function G_n is degenerate at 2. Corollary 2 can be used to construct an asymptotic confidence interval for the unknown parameter β. In Section 5, we study the case when the empirical distribution of the x_i converges.

The LSE of β based on $y_1, \ldots, y_n$ is $\beta_n = \Sigma_1^n y_i x_i / \Sigma_1^n x_i^2$, and

$E(\beta_n - \beta)^2 = \mathrm{Var}(\varepsilon)/\ \Sigma_1^n x_i^2$. When $(\Sigma_1^n |x_i|)^2$ tends to infinity at a faster rate than $\Sigma_1^n x_i^2$, by (i) of Theorems 1, 2, and 2', the MLE's b_n are superefficient compared with the LSE β_n for the uniform error case. Huber (1973), Bickel (1973), and others have considered the so-called M, R, and L-estimators in linear regression. These robust estimators are asymptotically normal, with asymptotic variances proportional to $(\Sigma_1^n x_i^2)^{-1}$, and hence b_n is again superefficient compared with them. This phenomenon is not surprising if we regard estimating β as a generalization of the problem of estimating a location parameter from i.i.d. uniform observations. In fact, if $x_1 = \dots = x_n = 1$, then b_n is just the midrange of the observations, which estimates the center of the uniform distribution with variance proportional to n^{-2}. When a family of distributions does not have a common support the estimation problem is often said to be non-regular. Usually, varying support enables one to find estimators with a superior rate of convergence. The non-regular case for a location parameter has been studied by Kempthorne (1966), Polfeldt (1970), Woodroofe (1972), Giesbrecht-Kempthorne (1976), and Hall (1982). There are possibilities to generalize some of their results to the linear model by combining the methods of the present paper with those of Bickel (1973). Part (ii) of Theorem 2 is analogous to results of Lai-Robbins-Wei (1979) and Wu (1981). Most results of Theorems 1, 2, and 2' can be generalized to the three-parameter model, and some of them can be generalized to the multi-linear regression model under appropriate regularity conditions of the design matrix. An extension of part (i) of Theorem 2 to the case $\alpha \neq 0$ is provided as follows.

THEOREM 3. Let y_1, y_2, ... be given by (1.1) and $n > 3$. Let a_n and b_n be any MLE's of α and β given by (1.10). Then

$$E[(a_n - \alpha)/\theta]^2 \leq 64[n^{-2} + m_n^2\ (\Sigma_{i=1}^n |x_i - \bar{x}_n|)^{-2}] \tag{1.36}$$

$$E[(b_n - \beta)/\theta]^2 \leq 32(\Sigma_{i=1}^n |x_i - \bar{x}_n|)^{-2}, \tag{1.37}$$

where $\bar{x}_n$ is the average of $x_i, \ldots, x_n$ and m_n is the median of $x_1, \ldots, x_n$.

Remark. Since

$$(1.38)\qquad (\Sigma_1^n |x_i - \bar{x}_n|)^2 \geq n(m_n - \bar{x}_n)^2 + \Sigma_1^n (x_i - \bar{x}_n)^2,$$

the estimators a_n and b_n are again superefficient compared with the LSE's for α and β.

2. Proof of Theorem 1.

We assume without loss of generality that $\beta = 0$, $\theta = 1$, and $x_i > 0$ for the proofs of (ii) and (v), which will be given first. Let

$$(2.1)\qquad b_+ = \min_{1 \leq i \leq n}[(y_i + 1/2)/x_i] = \min_{1 \leq i \leq n}[(\varepsilon_i + 1/2)/x_i]$$

$$(2.2)\qquad b_- = \max_{1 \leq i \leq n}[(y_i - 1/2)/x_i] = \max_{1 \leq i \leq n}[(\varepsilon_i - 1/2)/x_i]$$

$$(2.3)\qquad x_n^* = \max_{1 \leq i \leq n} x_i$$

Then $P\{b_+ \geq 0\} = P\{b_- \leq 0\} = 1$, and for any $t \geq 0$, $s \geq 0$, and $1/2 - sx^*_n \geq tx^*_n - 1/2$

$$(2.4)\qquad \begin{aligned} &P\{b_+ \geq t \text{ and } b_- \leq -s\} \\ &= P\{1/2 - sx_i \geq \varepsilon_i \geq tx_i - 1/2 \text{ for every } i=1, \ldots, n \\ &= \exp[\Sigma_{i=1}^n \log(1 - tx_i - sx_i)] \end{aligned}$$

Therefore

$$(2.5)\qquad P\{b_+ \geq t\} = P\{b_- \leq -t\} \leq \exp[-t(\Sigma_{i=1}^n x_i)],$$

(2.6) $Eb_+^2 = Eb_-^2 = 2\int_0^\infty P\{b_+ \geq t\}\, tdt \leq 2/(\Sigma_1^n x_i)^2$, and

(2.7) $E(b_n')^2 \leq E(b_+^2 + b_-^2)/4 \leq 1/(\Sigma_1^n x_i)^2$, since $b_n' = (b_+ + b_-)/2$,

which proves (1.18).

Let $z_1,\ldots,z_n$ be given by (1.30). Taking derivatives on both sides of (2.4),

(2.8) $$(d/dt)\,(d/ds)P\,\{\,(\Sigma_1^n x_i)b_+ \geq t \text{ and } (\Sigma_1^n x_i)b_- \leq -s\}$$

$$= \Sigma_{i=1}^n \Sigma_{\substack{j=1\\ j\neq i}}^n z_i z_j \exp[\,\Sigma_{\substack{k=1\\ k\neq i,j}}^n \log(1 - tx_k - sx_k)] \leq \exp\,[-t-s].$$

Under the condition (1.21), for any $t \geq 0$ and $s \geq 0$

(2.9) $$(d/dt)\,(d/ds)\,P\,\{\,(\Sigma_1^n x_i)b_+ \geq t \text{ and } (\Sigma_1^n x_i)b_- \leq -s\}$$

$$= (1 + o(1))\exp[-t-s] \text{ as } n \text{ tends to infinity.}$$

Integrating over the line $(t-s)/2 = u$, we have by (2.8) and (2.9)

(2.10) $$(d/du)\,P\{(\Sigma_1^n x_i)\,b_n \leq u\} = (d/du)\,P\{(\Sigma_1^n x_i)\,(b_+ + b_-)/2 \leq u\}$$

$$= (1 + o(1))e^{-2|u|}.$$

To prove (iii) and (iv) we shall still assume that $\theta = 1$ and $x_i > 0$ for every i. Let $f_\beta(y_1,\ldots,y_n)$ be the density of $y_1,\ldots,y_n$ and define

(2.11) $$A_n(s,t) = \{\,(y_1,\ldots,y_n):\ f_s(y_1,\ldots,y_n) = f_t(y_1,\ldots,y_n) = 1\}$$

(2.12) $$A(s,t) = \bigcap_{n=1}^\infty A_n(s,t).$$

By the definitions,

$$(2.13) \qquad P_\beta\{A(s,t)\} = \exp[\ \Sigma_1^\infty \log(1-|s-t|x_i)^+] \text{ for } \beta = s,t.$$

It follows from (1.20) that for any $0 < \delta < |\beta_1 - \beta_2|/2$

$$(2.14) \qquad \lim_n P_{\beta_i} \{A_n(\beta_1, \beta_2) \cap [|\delta_n - \beta_i| > \delta\} = 0 \text{ for } i = 1,2.$$

Since the likelihood ratio is unity on $A_n(\beta_1,\beta_2)$, (2.14) implies that

$$\lim_n P_{\beta_1}\{A_n(\beta_1, \beta_2)\} = \lim_n \Sigma_{i=1}^2 P_{\beta_1}\{A_n(\beta_1,\beta_2) \cap [|\delta_n - \beta_i| > \delta]\}$$

$$= \lim_n \Sigma_{i=1}^2 P_{\beta_i} \{A_n(\beta_1,\beta_2) \cap [|\delta_n - \beta_i| > \delta\ ]\} = 0.$$

Hence, by (2.12) and (2.13), (1.20) implies that

$$\exp[\ \Sigma_1^\infty \log(1 - |\beta_1 - \beta_2|x_i)^+\ ] = 0 \text{ for some } 0 < |\beta_1 - \beta_2| < 1/\max_n x_n,$$

which implies (1.19). That (1.19) implies (1.20) is clear by (ii).

We shall assume that the set B defined by (1.22) has a positive Lebesgue measure and prove (iv) by contradiction. Let $\delta > 0$ be small enough that $\mu(B(\delta)) > 0$, where μ is Lebesgue measure and $B(\delta) = \{\beta:\ \limsup_n E_\beta(\delta_n - \beta)^2/E_\beta(b_n - \beta)^2 \leq 1 - \delta\}$. Since $B(\delta)$ can be covered by an open set A with arbitrarily small $\mu(A - B(\delta))$, there exists a finite open interval $B^* = (\beta_1, \beta_2)$ such that $\mu(B^* \cap B(\delta)) \geq (1 - \delta/16)(\beta_2 - \beta_1) > 0$. Let b_+ and b_- be given by the first equations of (2.1) and (2.2). Since $P_\beta\{b_- \leq \beta \leq b_+\} = 1$, we may assume that $b_- \leq \delta_n \leq b_+$ a.s. so that

$$(2.15) \qquad E_\beta(\delta_n - \beta)^2 \leq E_\beta[(b_+ - \beta)^2 + (b_- - \beta)^2] \leq 4/(\Sigma_1^n x_i)^2, \text{ by (2.6)}.$$

It follows from (1.24), (2.15), and the definition of B^* that

(2.16) $$\limsup_n \int_{\beta_1}^{\beta_2} (\Sigma_1^n x_i)^2 E_\beta(\delta_n - \beta)^2 d\beta$$

$$\le 4\mu(B^* - B(\delta)) + (1/2 - \delta/2)\mu(B^* \cap B(\delta)) \le (1/2 - \delta/4)(\beta_2 - \beta_1).$$

On the other hand, the Bayes estimator for the uniform (β_1, β_2) prior is $b_n^* = [\min(b_+, \beta_2) + \max(b_-, \beta_1)]/2$, and by (2.5) and (1.24),

(2.17) $$\lim_n E_\beta (b_n^* - \beta)^2 (\Sigma_1^n x_i)^2 = 1/2 \quad \text{for any } \beta_1 < \beta < \beta_2.$$

Hence

$$\liminf_n \int_{\beta_1}^{\beta_2} (\Sigma_1^n x_i)^2 E_\beta(\delta_n - \beta)^2 d\beta$$

$$\ge \lim_n \int_{\beta_1}^{\beta_2} (\Sigma_1^n x_i)^2 E_\beta(b_n^* - \beta)^2 d\beta = (\beta_2 - \beta_1)/2,$$

which contradicts (2.16).

Finally, let us prove (i). It follows from (1.4) and (2.4) that

$$E(b_n - \beta)^2 \le E[(b_- - \beta)^2 + (b_+ - \beta)^2].$$

Hence, (1.17) follows from (2.6). The proof of Theorem 1 is complete.

3. Proofs of Theorems 2 and 2'.

We shall first prove Theorem 2'. Set

(3.1) $$\varepsilon_i' = \varepsilon_i \quad \text{if } x_i > 0, \text{ and } -\varepsilon_i \text{ if } x_i < 0.$$

By the definition (1.9) of b_n,

$$(3.2)\qquad \max\ \{\ |\varepsilon_i'/\theta - (b_n - \beta)|x_i|/\theta|\ :\ 1 \le i \le n,\ |x_i| > 0\}$$

$$= \min_b \max\ \{\ |\varepsilon_i'/\theta - b|x_i||\ :\ 1 \le i \le n,\ |x_i| > 0\ \}.$$

It is clear that the minimum of the right side of (3.2) is almost surely uniquely reached at $b = (b_n - \beta)/\theta$. Since $\{\ \varepsilon_i'/\theta,\ i \ge 1\ \}$ is a sequence of i.i.d. uniform $(-1/2,1/2)$ random variables, the joint distribution of the sequence $\{\ (b_n - \beta)/\theta\ \}$ does not depend on β, θ, and the signs of x_i. We shall therefore assume throughout this section that $\theta = 1$, $\beta = 0$, and $x_i > 0$ for all i, so that (1.9) becomes

$$(3.3)\qquad \max_{1\le i\le n} |\varepsilon_i - b_n x_i| = \min_b \max_{1\le i\le n} |\varepsilon_i - bx_i|$$

$$\le \max_{1\le i\le n}|\varepsilon_i| \le 1/2.$$

Since by (3.3) $b_n \ge t \ge 0$ implies that $\varepsilon_i \ge tx_i - 1/2$ for every $i = 1,\ldots,n$,

$$(3.4)\qquad P\ \{b_m \ge t \text{ for some } m \ge n\} \le P\{\varepsilon_i \ge tx_i - 1/2 \text{ for every } i=1,\ldots,n\}$$

$$\le \exp[-t\ \Sigma_{i=1}^n x_i] \qquad\qquad \text{for any } t \ge 0.$$

It follows that

$$Eb_n^2 = \int_0^\infty P\ \{b_n^{\ 2} \ge t^2\ \}\ dt^2$$

$$= 4 \int_0^\infty P\ \{b_n \ge t\}\ tdt \le 4 \int_0^\infty \exp[-t\ \Sigma_1^n\ x_i]tdt$$

$$= 4/[\ \Sigma_{i=1}^n x_i)^2\ ,$$

and the proof of (i) is complete.

It is clear that (1.27) implies (1.26), and the equivalence of (1.19) and

(1.26) is implied by (iii) of Theorem 1 and (i). Therefore, for (ii) we need only prove that (1.19) implies (1.27). Define the integers n_k by

$$\Sigma_{i=1}^{n_k-1} x_i < e^k \leq \Sigma_{i=1}^{n_k} x_i .$$

Then for any $t > 0$,

$$P\{ (\log \Sigma_1^n x_i)^{-1} (\Sigma_1^n x_i) b_n \geq t \text{ for some } n_k \leq n < n_{k+1}\}$$

$$\leq P\{ k^{-1} e^{k+1} b_n \geq t \quad \text{for some } n \geq n_k\}$$

$$\leq \exp[- kte^{-k-1} \Sigma_{i=1}^{n_k} x_i] \qquad \text{by (3.4)}$$

$$\leq \exp[- kt/e] .$$

Therefore

$$\Sigma_{k=1}^{\infty} P\{ (\log \Sigma_1^n x_i)^{-1} (\Sigma_1^n x_i) b_n \geq t \text{ for some } n_k \leq n < n_{k+1}\}$$
$$< \infty \qquad \text{for any } t > e,$$

provided that (1.19) holds, and the proof of (1.27) is complete.

To begin the proof of (iii), define

$$(3.5) \qquad d_{ij} = (\varepsilon_i + \varepsilon_j)/(x_i + x_j) , \qquad i,j \geq 1$$

$$(3.6) \qquad w_{ij} = \varepsilon_i - d_{ij} x_i , \qquad i,j \geq 1$$

and for any fixed $n \geq 2$ let

$$(3.7) \qquad I = \text{smallest } i=1,\ldots,n \text{ for which } |\varepsilon_i - b_n x_i| = \max_{1\leq j\leq n} |\varepsilon_j - b_n x_j|$$

$$(3.8) \qquad J = \text{largest } j=1,\ldots,n \text{ for which } |\varepsilon_j - b_n x_j| = \max_{1\leq i\leq n} |\varepsilon_i - b_n x_i| .$$

Then I and J are uniquely defined with probability one, and

$$(3.9) \qquad b_n - \beta = b_n = d_{IJ}, \; \max_{1 \le i \le n} |\varepsilon_i - b_n x_i| = |w_{IJ}| ,$$

so that

$$(3.10) \qquad P\{ b_n \le t \}$$

$$= \Sigma_{i=1}^n \Sigma_{j=i+1}^n P\{ d_{ij} \le t \text{ and } |\varepsilon_k - d_{ij} x_k| \le |w_{ij}| \text{ for } 1 \le k \ne i,j \le n\}$$

$$= \Sigma_{i=1}^n \Sigma_{j=i+1}^n \int_{-\infty}^{t} \int P\{|\varepsilon_k - sx_k| \le |w| \text{ for } 1 \le k \ne i,j \le n\}$$
$$dP\{ w_{ij} \le w \,|d_{ij} = s\} \, dP\{d_{ij} \le s \},$$

where the element of measure is

$$(3.11) \qquad dP\{|w_{ij}| \le w \mid d_{ij} = s\} \, dP\{ d_{ij} \le s\}$$

$$= (I\{0 \le w \le 1/2 - |sx_i|\} + I\{0 \le w \le 1/2 - |sx_j|\}) \cdot$$
$$I\{|sx_i| + |sx_j| \le 1\} \, (x_i + x_j) dwds \quad \text{if } |sx_i|, \; |sx_j| \le 1/2.$$

Let $z_1, \ldots, z_n$ be given by (1.30). It follows from (1.21), (3.10), and (3.11) that for large n

$$(3.12) \qquad (d/dt) \, P\{ (\Sigma_1^n x_i) b_n \le t \}$$

$$= \Sigma_{i=1}^n \Sigma_{j=i+1}^n \int_0^\infty P\{|\varepsilon_k - tz_k| \le w \text{ for } 1 \le k \ne i, \; j \le n \} \, (z_i + z_j) \cdot$$

$$(I\{0 \le w \le 1/2 - |t| z_i\} + I\{ 0 \le w \le 1/2 - |t| z_j\}) \cdot$$

$$I\{|t|(z_i + z_j) \le 1 \} \, dw$$

$$= \Sigma_{i=1}^n \Sigma_{j=i+1}^n \int_0^\infty \exp [- (1 + o(1)) \, \Sigma_{k=1}^n (u + \max(u, |t| z_k)] (z_i + z_j) \cdot$$
$$(I\{ |t| z \le u_i \le 1/2\} + I\{|t| z \le u_j \le 1/2 \}) \cdot$$

$$I\{|t|\,(z_i + z_j) < 1\}\,du \quad ,\ u = 1/2 - w$$

$$= (1+o(1)) \int_0^\infty \exp\,[\,-\,v - \Sigma_{k=1}^n \max(v/n, |t|z_k)]\ \cdot$$

$$\Sigma_{i=1}^n I\{\,z_i < v/(n|t|)\}\,(z_i + 1/n)dv \quad ,\ v = nu$$

$$= (1+\,o(1))[\,\int_{n|t|z_1}^{n|t|z_2} + \int_{n|t|z_2}^{n|t|z_3} + \dots + \int_{n|t|z_n}^{\infty}\ ]$$

$$= (1 + o(1))\ \Sigma_{i=1}^n\ (z_1 + \dots + z_i + i/n)(1+i/n)^{-1}.$$
$$\Big[\exp[-\,(1 + i/n)v - |t|(z_{i+1} + \dots + z_n)]\Big]_{n|t|z_{i+1}}^{n|t|z_i},$$

where z_{n+1} is defined to be infinity.

Let $s_1, \dots, s_n$ and G_n be given by (1.28) and (1.29). Then

$$\Sigma_1^n \exp\,(s_i + i/n)\,(1 + i/n)^{-1} \Big[\exp[\,-\,(1 + i/n)v - |t|(1 - s_i)]\Big]_{n|t|z_{i+1}}^{n|t|z_i}$$

$$= \Sigma_1^n \exp\,[\,-\,|t|\,((n + i)z_i + 1 - s_i)].$$

$$[(s_i + i/n)(1 + i/n)^{-1} - (s_{i-1} + (i - 1)/n)\,(1 + (i - 1)/n)^{-1}]$$

(3.13)

$$= \Sigma_1^n \exp[-\,|t|((n + i)z_i + 1 - s_i)].$$

$$n(n + i)^{-1}(n + i - 1)^{-1}[(n + i)z_i + 1 - s_i]$$

$$= \int_1^\infty (y/2)\exp[-\,|t|y]dG_n(y)\ .$$

Hence

$$(d/dt)P\ \{\ (\ \Sigma_{i=1}^n x_i)b_n < t\ \}$$

$$= (1 + o(1)) \int_1^\infty (y/2)e^{-y|t|}dG_n(y) \quad ,$$

which proves (iii).

To prove Theorem 2, (i) follows from (i) of Theorem 1. For (ii),

$$P\{1/2 - w_n > t\} = 2\,P\{1/2 - w_n > t \text{ and } b_n > 0\}$$

$$< 2\,P\{\varepsilon_i > -1/2 + t \quad \text{for every } i = 1, \dots, n\}$$

$$< 2e^{-nt} \qquad \text{for any } t > 0.$$

Since $P\{w_n < 1/2\} = 1$, the proof is complete.

Finally, we prove Corollary 1. Since (1.32) and (1.33) follow from Theorem 2' and (1.34), and the second inequality of (1.34) is trivial, we need only prove the first inequality of (1.34), which is purely analytic. Consider the design in which $x_n = x_i$ if $n = km + i$ for some integers k and $i = 1, \dots, m$ where m is fixed. By the definitions, G_n converges to G_m weakly. It follows from (iv) and (v) of Theorem 1 that

$$\int_1^\infty 2y^{-2}dG_m(y) = \lim_n \int_1^\infty 2y^{-2}dG_n(y)$$

$$= \lim_n \int_0^1 (\Sigma_1^n x_i)^2 E_\beta(b_n - \beta)^2 d\beta > 1/2.$$

4. Proof of Theorem 3.

By definitions (1.1) and (1.10),

$$(4.1)\quad \max_{1<i<n} \left| \frac{\varepsilon_i}{\theta} - \frac{(a_n - \alpha)}{\theta} - \frac{(b_n - \beta)}{\theta} x_i \right| = \min_{a,b} \max_{1<i<n} \left| \frac{\varepsilon_i}{\theta} - a - bx_i \right| .$$

Therefore, we can assume without loss of generality that $\alpha = \beta = 0$ and $\theta = 1$.

First, let us prove (1.37). For any $t > 0$, (4.2) $$\{ b_n > t \} = \{ b_n > t, a_n + b_n\bar{x}_n < 0 \} \cup \{b_n > t, a_n + b_n\bar{x}_n > 0\}$$

$$\subset \{ \varepsilon_i > t \mid x_i - \bar{x}_n \mid - 1/2 \text{ for all } i \ni x_i > \bar{x}_n \}$$
$$\cup \{ \varepsilon_i < 1/2 - t|x_i - \bar{x}_n \mid \text{ for all } i \ni x_i < \bar{x}_n \} \quad .$$

Hence

$$P \{ b_n > t \} < 2 \exp [- (t/2) \Sigma_{i=1}^n |x_i - \bar{x}_n|] \text{ , and}$$

$$Eb_n^2 = 2 \int_0^\infty P \{ |b_n| > t \} tdt = 4 \int_0^\infty P \{ b_n > t \} tdt$$

$$< 8 \int_0^\infty \exp [- (t/2) \Sigma_1^n |x_i - \bar{x}_n|]tdt = 32/[\Sigma_{i=1}^n |x_i - \bar{x}_n|)^2 ,$$

which proves (1.37). To prove (1.36) we first consider $E(a_n + b_n m_n)^2$.

$$P \{a_n + b_n m_n > t \} = P \{ a_n + b_n m_n > t, b_n > 0 \} + P \{ a_n + b_n m_n > t, b_n < 0 \}$$

$$< P \{ \varepsilon_i > t - 1/2 \text{ for } \forall i \ni x_i > m_n\} + P \{ \varepsilon_i > t - 1/2 \text{ for } \forall i \ni x_i < m_n \}$$

$$< 2e^{-nt/2} \qquad \text{for any } t > 0.$$

Hence

$$E(a_n + b_n m_n)^2 = 2 \int_0^\infty P \{ |a_n + b_n m_n | > t \} tdt$$
$$= 4 \int_0^\infty P \{a_n + b_n m_n > t\} tdt < 8 \int_0^\infty te^{-nt/2}dt = 32/n^2 .$$

It follows from (1.37) that

$$Ea_n^2 < 2 [E(a_n + b_n m_n)^2 + m_n^2 Eb_n^2] < 2 [32/n^2 + 32m_n^2 / (\Sigma_1^n |x_i - \bar{x}_n|)^2],$$

and the proof of (1.36) is complete.

5. Limit of G_n and examples.

We assume that the conditions of Theorem 2' (iii) hold in this section. Let $G_n(y)$ be given by (1.28) through (1.30). Set

$$h_n(x) = [(1 + H_n(x))x + \int_{x+}^{\infty} tdH_n(t)]/\int_0^{\infty} tdH_n(t) \ , \tag{5.1}$$

where $H_n(x) = n^{-1}\Sigma_1^n I\{|x_i| < x\}$. By the definitions, $dG_n(h_n(x))/dH_n(x) = 2(1 + H_n(x))^{-1}(1+H_n(x) - 1/n)^{-1}$. Suppose that $\lim H_n = H$ weakly and $\lim \int_0^{\infty} tdH_n(t) = \int_0^{\infty} tdH(t) = \mu > 0$. Then

$$\lim h_n(x) = h(x) = \begin{cases} 1 & \text{if } \mu = \infty \\ [(1 + H(x))x + \int_{x+}^{\infty} tdH(T)]/\mu & \text{otherwise} \end{cases} \tag{5.2}$$

(5.3) $\lim G_n = G$ weakly such that $G(\{1\}) = 1$ if $\mu = \infty$, and

$$dG(h(x))/dH(x) = 2(1 + H(x))^{-1}(1 + H(x-))^{-1} \text{ otherwise}$$

(5.4) the density of $\Sigma_{i=1}^n|x_i|(b_n - \beta)/\theta$ at t converges to

$$f(t) = f(t;G) = \int_1^{\infty} (y/2)\exp[-y|t|]dG(y)$$

$$\lim (\Sigma_1^n|x_i|)^2E[(b_n - \beta)/\theta]^2 = \int_1^{\infty} 2y^{-2}dG(y) \ , \tag{5.5}$$

where b_n is defined by (1.9).

Example 1. $H(\{1\}) = H(\{2\}) = H(\{3\}) = 1/3$. Then $G(\{3/2\}) = 1/2$, $G(\{13/6\}) = 3/10$, $G(\{3\}) = 1/5$, and $\int_1^{\infty} 2y^{-2}dG(y) = 0.6167$.

Example 2. $H(\{1\}) = H(\{10\}) = H(\{50\}) = 1/3$. Then $G(\{64/61\}) = 1/2$, $G(\{100/61\}) = 3/10$, $G(\{300/61\}) = 1/5$, and $\int_1^{\infty} 2y^{-2}dG(y) = 1.1482$.

Example 3. $dH(x)/dx = M^{-1}I\{0 < x < M\}$. Then $dG(y)/dy = y^{-1.5}$ on $1 < y < 4$, and $\int_1^{\infty} 2y^{-2} dG(y) = 0.775$.

Example 4. $H(\{x\}) = 1$ for some $x > 0$. Then $G(\{2\}) = 1$.

Example 5. $H(x) = x/(1 + x)$. Then $G(\{1\}) = 1$ and $\int_1^\infty 2y^{-2}dG(y) = 2$.

Remarks: As shown by Examples 4 and 5, the inequalities in Corollary 1 are sharp. The above results remain valid if we replace the definition of H_n by $H_n(x) = \Sigma_1^n I\{|x_i|/m_n < x\}/n$, since G_n only depends on $z_1, \ldots, z_n$ given by (1.30). For example, if $x_n = n$ for every n, then we have the same results as in Example 3 and $\int_1^\infty 2y^{-2}\, dG_n(y)$ tends to 0.775.

REFERENCES

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

Bickel, P. J. (1973). On some analogues to linear combinations of order statistics in the linear model. Ann. Statist. 1, 597-616.

Giesbrecht, F. and Kempthorne, O. (1976). Maximum likelihood estimation in the three-parameter lognormal distribution. J. Roy. Statist. Soc. (B) 38, 257-264.

Hall, P. (1982). On estimating the endpoint of a distribution. Ann. Statist. 2, 556-568.

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

Kempthorne, O. (1966). Some aspects of experimental inference. J. Amer. Statist. Assoc. 61, 11-34.

Lai, T. L., Robbins, H. and Wei, C. Z. (1979). Strong consistency of least squares estimates in multiple regression. J. Multivariate Anal. 9, 343-361.

Polfeldt, T. (1970). Asymptotic results in non-regular estimation. Skand. Akt. Tidskr. Suppl. 1-2.

Woodroofe, M. (1972). Maximum likelihood estimation of a translation parameter of a truncated distribution. _Ann. Math. Statist._ 43, 113-122.

Wu, C. F. (1981). Asymptotic theory of nonlinear least squares estimation. _Ann. Statist._ 9, 501-513.