AD-A098 169   FLORIDA STATE UNIV TALLAHASSEE DEPT OF STATISTICS   F/G 12/1
              LINEAR LEAST SQUARES ESTIMATES AND NONLINEAR MEANS.(U)
              MAR 81   R L BERGER, N A LANGBERG   DAAG29-79-C-0158
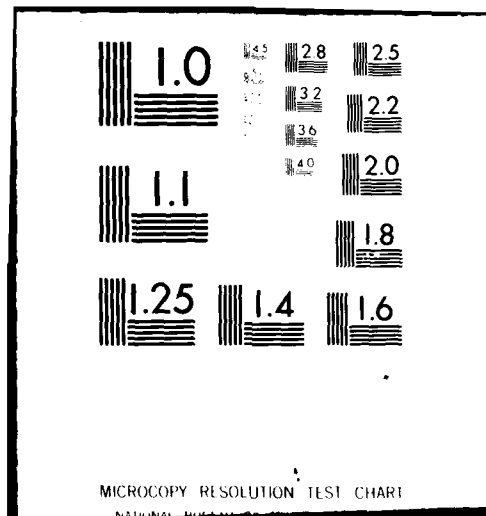UNCLASSIFIED FSU-STATISTICS-M-573   ARO-16713.15-M   NL

END
DATE
FILMED
5-81
DTIC

1.0

4.5

8.

2.8   2.5

3.2   2.2

3.6

1.1

4.0   2.0

1.8

1.25   1.4   1.6

MICROCOPY RESOLUTION TEST CHART

LEVEL

LINEAR LEAST SQUARES ESTIMATES AND NONLINEAR MEANS.

by

Roger L. Berger[1] and Naftali A. Langberg[2]

March 1981

FSU Statistics Report M-573
USARO Technical Report D-50
AFOSR Technical Report No. 78-120

APR 24 1981

Department of Statistics
The Florida State University
Tallahassee, FL. 32306

AMS 1970 subject classification: 62F12, 62F35, 62J05

DTIC FILE COPY

1

ABSTRACT

The consistency and asymptotic normality of a linear least squares estimate of the form $(X'X)^- X'Y$ when the mean is not $X\beta$ is investigated in this paper. The least squares estimate is a consistent estimate of the best linear approximation of the true mean function for the design chosen. The asymptotic normality of the least squares estimate depends on the design and the asymptotic mean may not be the best linear approximation of the true mean function.

Key words: Consistency, asymptotic normality, best linear approximation, model robustness.

Linear Least Squares Estimates
and Nonlinear Means

1. Introduction

In the standard linear models theory, the mean of an observation vector
Y is given $X\beta$ where X is a known matrix of constants and β is an unknown
vector of parameters. The estimate $\hat{\beta} = (X'X)^- X'Y$, where $(X'X)^-$ is a generalized inverse of $X'X$, has many well-known and desirable properties. Considering that both regression and analysis of variance problems are part
of the linear models theory, estimates like $\hat{\beta}$ are among the most widely used
of all parameter estimates.

If the mean of Y is not $X\beta$, the properties of $\hat{\beta}$ are not described by
the standard linear models theory. Indeed, it may not be clear what (if
anything)$\hat{\beta}$ is estimating. The behavior of $\hat{\beta}$ for large samples when the mean
is not $X\beta$ is investigated in this paper. This behavior is found to depend
on the design used in the experiment. But, for a given design, $\hat{\beta}$ is shown
to estimate the best linear approximation of the true mean function in a sense
to be defined.

In Section 2, the model and notation are defined. In Section 3,$\hat{\beta}$ is
shown to be a consistent estimate of the best linear approximation. The
asymptotic normality of $\hat{\beta}$ is investigated in Section 4. Section 5 contains
some examples.

Other authors have considered questions related to those addressed herein.
The question of model robustness", what $\hat{\beta}$ is estimating if the mean is not
$X\beta$, has been investigated by authors such as Box and Draper (1959), Atwood
(1971), Stigler (1971) and Huber (1975). These authors considered specific
alternative mean functions to $X\beta$. Royall and Herson (1973) have described
the mean of $\hat{\beta}$ for an arbitrary mean function. Surprisingly recently other

authors such as Drygas (1976), Lai, Robbins and Wei (1978), Anderson and Taylor (1979) and Wu (1980) have considered the consistency of $\hat{\beta}$ under various conditions, but always under the assumption that the mean is $X\beta$.

2. Model and Notation

Let X denote a subset of an r-dimensional Euclidean space ($r \geq 1$). X is the set of possible values of the independent variable x. A __design__ for a sample of size n is a specification of n points, $x_{1,n}, \ldots, x_{n,n}$, from X where the n points specify the values of the independent variable at which observations are to be taken. The $x_{i,n}$ need not be all distinct. If a point x is repeated p times then p observations are taken at x. A design can be completely described by a discrete probability measure ξ_n on X where the probability ξ_n assigns to a point x is the proportion of the n observations to be taken at x. ξ_n will also be called the design.

The observation vector is $Y_n = (Y_{1,n}, \ldots, Y_{n,n})'$. It is assumed that for each n

$$Y_{i,n} = m(x_{i,n}) + \epsilon_{i,n} \qquad i = 1, \ldots, n,$$

where m, the unknown mean function, is a real-valued bounded and continuous function on X and $\epsilon_{i,n}$, $n = 1, 2, \ldots$; $i = 1, \ldots, n$, are identically distributed random variables with mean zero and variance σ^2. It is also assumed that $\epsilon_{1,n}, \ldots, \epsilon_{n,n}$ are independent for each $n = 1, 2, \ldots$. Thus $Y_{1,n}, \ldots, Y_{n,n}$ are independent random variables with $E \, Y_{i,n} = m(x_{i,n})$ and variance σ^2.

Let $f(x) = (f_1(x), \ldots, f_n(x))'$ denote a p × 1 vector of real-valued, bounded and continuous functions on X. For a given design, $x_{1,n}, \ldots, x_{n,n}$, let X_n be the n × p matrix with (i, j)th element $f_j(x_{i,n})$. The asymptotic behavior of the estimate $\hat{\beta}_n = (X_n' X_n)^{-} X_n' Y_n$ will be investigated in this paper.

$\hat{\beta}_n$ will be called a <u>linear least squares estimate</u> since, for a given observation $y_n = (y_{1,n}, \ldots, y_{n,n})'$, $\hat{\beta}_n$ is the vector β which minimizes $\Sigma_{i=1}^n (y_{i,n} - \beta' f(x_{i,n}))^2$. Conditions under which $X_n' X_n$ is non-singular and $\hat{\beta}_n$ is uniquely defined are given in Lemmas 3.1 and 3.4.

Let ξ be a probability measure on X. A $p \times 1$ vector $\beta(m,\xi)$ will be called a <u>best linear approximation of</u> <u>m(x)</u> if

$$\int (m(x) - \beta'(m,\xi)f(x))^2 d\xi(x) = \inf_\beta \int (m(x) - \beta' f(x))^2 d\xi(x). \qquad (2.1)$$

Since $\beta(m,\xi)$ depends on the unknown mean function m, $\beta(m,\xi)$ is a parameter. Let $M(\xi)$ denote the $p \times p$ matrix with (i,j)th element $\int f_i(x) f_j(x) \, d\xi(x)$ and let $c(\xi)$ denote the $p \times 1$ vector with (i)th coordinate $\int f_i(x) m(x) \, d\xi(x)$. In optimal design theory (see, e.g., Kiefer (1962)) a multiple of $M(\xi)$ is called the information matrix of ξ. If $M(\xi_n)$ is non-singular, $\sigma^2 n^{-1} M^{-1}(\xi)$ is the covariance matrix of $\hat{\beta}_n$. By equating the partial derivatives $\partial \int (m(x) - \beta' f(x))^2 d\xi(x)/\partial \beta_j$, $j = 1, \ldots, p$, to zero it is easily verified that if $M(\xi)$ is non-singular then $\beta(m,\xi)$ is unique and equals $M^{-1}(\xi)c(\xi)$. Under conditions relating a sequence of designs ξ_n to ξ, it will be shown that $\hat{\beta}_n$ is a consistent estimate of $\beta(m,\xi)$ and $\hat{\beta}_n$ is asymptotically normal with mean $\beta(m,\xi)$. In this sense, the best linear approximation $\beta(m,\xi)$ is the parameter being estimated by the linear least squares estimate $\hat{\beta}_n$ when the mean $m(x)$ is not necessarily linear.

3. Consistency

Let ξ_n be a fixed sequence of designs for sample sizes $n = 1,2,\ldots$. Throughout it will be assumed that ξ_n converges weakly to a probability measure

ξ where weak convergence is defined in Billingsley (1968). The main result in this section, the proof of which is deferred until the end of the section, is the following theorem.

Theorem 3.1: If $M(\xi)$ is non-singular, then $\hat{\beta}_n$ is a consistent estimate of $\beta(m,\xi)$ in that the random vectors $\hat{\beta}_n$ converge in probability to the real-valued vector $\beta(m,\xi) = M^{-1}(\xi) \, c(\xi)$.

It is important to note that the parameter $\beta(m,\xi)$ which $\hat{\beta}_n$ estimates depends on ξ. The experimenter chooses ξ when the sequence of designs $\xi_1, \xi_2, \ldots$ is chosen. ξ should be chosen so that the definition of best linear approximation in (2.1) accurately reflects how the experimenter wishes to measure the closeness of $\beta'f(x)$ to $m(x)$. If the support of ξ consists only of a finite number of points, $\beta'f(x)$ will be compared to $m(x)$ only at these points in determining $\beta(m,\xi)$. If the aim is to estimate a β so that $\beta'f(x)$ is close to $m(x)$ for all x in X, a choice of a ξ whose support is all of X, e.g., uniform on X, seems more appropriate. In optimal design theory (see, e.g., Kiefer (1959) or Karlin and Studden (1966)) the optimal design often has a support with a finite number of points. These designs may not be very appropriate if the mean $m(x)$ is not of the form $\beta'f(x)$. The fact that what $\hat{\beta}_n$ is estimating depends on the design if $m(x) \neq \beta'f(x)$ has been recognized previously. See, for example, Draper and Smith (1966, Chapter 2, Section 12) or Royall and Herson (1973).

The following lemma gives a condition under which $M(\xi)$ will be non-singular. The functions $f_1(..),\ldots,f_p(x)$ are called <u>linearly independent</u> - ξ if, for a $p \times 1$ real-valued vector $\alpha, \alpha'f(x) = 0$ a.s. ξ implies $\alpha = 0$.

Lemma 3.1: If $f_1(x),\ldots,f_p(x)$ are linearly independent -ξ then $M(\xi)$ is non-singular.

Proof: Let α be a $p \times 1$ real-valued vector. Assume $M(\xi)\alpha = 0$. To show $M(\xi)$ is non-singular it suffices to show $\alpha = 0$. $M(\xi)\alpha = 0$ implies $0 = \alpha' M(\xi)\alpha = \int (\alpha' f(x))^2 d\xi(x)$. This implies $\alpha' f(x) = 0$ a.s. ξ. Since $f_1(x), \ldots, f_p(x)$ are linearly independent $-\xi$, $\alpha = 0$. ||

To study the asymptotic behavior of $\hat{\beta}_n$, it is useful to note that

$$\hat{\beta}_n = (X_n' X_n)^{-} X_n' Y_n \tag{3.1}$$

$$= (n^{-1} X_n' X_n)^{-}(n^{-1} X_n' \epsilon_n) + (n^{-1} X_n' X_n)^{-}(n^{-1} X_n' m_n)$$

where $\epsilon_n = (\epsilon_{1,n}, \ldots, \epsilon_{n,n})'$ and $m_n = (m(x_{1,n}), \ldots, m(x_{n,n}))'$. The asymptotic behavior of $n^{-1} X_n' \epsilon_n$, $n^{-1} X_n' m_n$ and $n^{-1} X_n' X_n$ is described in the following three lemmas.

Lemma 3.2: Under the model the random vectors $n^{-1} X_n' \epsilon_n$ converge in probability to 0.

Proof: It suffices to show that for any $\delta > 0$,

$$\lim_{n \to \infty} P(|\Sigma_{i=1}^n f_r(x_{i,n}) \epsilon_{i,n}| > n\delta) = 0 \text{ for } r = 1, \ldots, p. \tag{3.2}$$

Let $a = \max_{1 \le r \le p} \sup_x f_r^2(x)$. Since the f_r are bounded, $a < \infty$. Fix $\delta > 0$. By Chebyshev's Inequality and the fact that $\epsilon_{1,n}, \ldots, \epsilon_{n,n}$ are uncorrelated with mean zero,

$$P(|\Sigma_{i=1}^n f_r(x_{i,n}) \epsilon_{i,n}| > n\delta) \le E(\Sigma_{i=1}^n f_r(x_{i,n}) \epsilon_{i,n})^2 / n^2 \delta^2$$

$$= \sigma^2 \Sigma_{i=1}^n f_r^2(x_{i,n}) / n^2 \delta^2 \le \sigma^2 a / n\delta^2. \text{ Thus Equation (3.2) is true.} ||$$

For the sake of completeness it should be noted that the result of Lemma 3.2 and hence the consistency result of Theorem 3.1 holds under the condition that for each n, $\epsilon_{1,n}, \ldots, \epsilon_{n,n}$ are uncorrelated with means all zero and variances all bounded above by σ^2. The stronger condition that $\epsilon_{i,n}, \ldots, \epsilon_{n,n}$ are i.i.d. is not necessary.

Lemma 3.3: Under the assumptions of the model, $\lim\limits_{n\to\infty} n^{-1}X_n'm_n = c(\xi)$.

Proof: Coordinatewise, this result is

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} f_r(x_{i,n})m(x_{i,n}) = \int f_r(x)m(x)d\xi(x), \quad r = 1,\ldots,p. \qquad (3.3)$$

Note that $n^{-1} \sum_{i=1}^{n} f_r(x_{i,n})m(x_{i,n}) = \int f_r(x)m(x)d\xi_n(x)$.

Since f_r and m are bounded and continuous and ξ_n converges weakly to ξ, Equation

(3.3) is true.||

Lemma 3.4: If $M(\xi)$ is non-singular then (i) $\lim\limits_{n\to\infty} n^{-1}X_n'X_n = M(\xi)$,

(ii) $n^{-1}X_n'X_n$ is non-singular for all sufficiently large n and

(iii) $\lim\limits_{n\to\infty} (n^{-1}X_n'X_n)^- = M^{-1}(\xi)$.

Proof: Elementwise, statement (i) is

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} f_r(x_{i,n}) f_s(x_{i,n}) = \int f_r(x)f_s(x)d\xi(x), \quad r, s = 1,\ldots,p. \qquad (3.4)$$

Note that $n^{-1} \sum_{i=1}^{n} f_r(x_{i,n})f_s(x_{i,n}) = \int f_r(x)f_s(x)d\xi_n(x)$.

Since f_r and f_s are bounded and continuous and ξ_n converges weakly to ξ,

Equation (3.4) is true.

Since the determinant is a continuous function, by (i),

$\lim\limits_{n\to\infty} |n^{-1}X_n'X_n| = |M(\xi)| \neq 0$. Thus (ii) is true.

If A_n and A are non-singular matrices and $\lim\limits_{n\to\infty} A_n = A$, then $\lim\limits_{n\to\infty} A_n^{-1} = A^{-1}$.

By (i) and (ii), $n^{-1}X_n'X_n$ is non-singular for all large n and (iii) is true.||

Proof of Theorem 3.1: This theorem follows from Equation (3.1) and Lemmas

3.2, 3.3 and 3.4.||

4. Asymptotic Normality

In this section, the asymptotic normality of $n^{1/2}(\hat{\beta}_n - \beta(m,\xi))$ is investi-

gated. As in Section 3, it is assumed that ξ_n is a fixed sequence of designs

for sample sizes n = 1, 2,... which converges weakly to a probability measure

ξ. The asymptotic normality result is given in Theorem 4.1.

__Theorem 4.1:__ Assume $M(\xi)$ is non-singular. Let $m_n = (m(x_{1,n}),\ldots, m(x_{n,n}))'$ Assume that $\lim_{n\to\infty} n^{1/2}((X_n'X_n)^- X_n' m_n - \beta(m,\xi))$ exists and equals the $p \times 1$ vector b. Then $n^{1/2}(\hat{\beta}_n - \beta(m,\xi))$ converges weakly to a multivariate normal random vector with mean b and covariance matrix $\sigma^2 M^{-1}(\xi)$.

__Proof:__ Let $T_n = n^{-1/2} X_n' \epsilon_n$. Note that

$$n^{1/2}(\hat{\beta}_n - \beta(m,\xi)) = (n^{-1} X_n' X_n)^- T_n + n^{1/2}((X_n' X_n)^- X_n' m_n - \beta(m,\xi)). \tag{4.1}$$

By assumption, the last term converges to b. By Lemma 3.4, $\lim_{n\to\infty} (n^{-1} X_n' X_n)^- = M^{-1}(\xi)$. So to prove the desired result it suffices, by an extension of Slutzky's Theorem (Billingsley (1968), Theorem 5.5), to show that the random __vectors__ T_n converge weakly to U where U is a p-dimensional multivariate normal random __vector__ with mean 0 and covariance matrix $\sigma^2 M(\xi)$.

To prove the convergence of T_n to U, it suffices, by the Cramer-Wold device (Billingsley (1968), p. 48), to show that $\alpha' T_n$ converges weakly to $\alpha' U$ for all p-dimensional vectors α. Fix $\alpha \neq 0$. Let $W_{j,n} = n^{-1/2}(\alpha' f(x_{j,n})) \epsilon_{j,n}$. Then $\alpha' T_n = \sum_{j=1}^n W_{j,n}$ so it suffices to show that $\sum_{j=1}^n W_{j,n}$ converges weakly to $\alpha' U$. By the definition of $\epsilon_{j,n}$, $W_{1,n},\ldots, W_{n,n}$ are independent random variables with zero means and variances respectively equal to $n^{-1}(\alpha' f(x_{j,n}))^2 \sigma^2$. Thus it suffices, by the Normal Central Limit Theorem (Loéve (1963), p. 288), to show that

$$\lim_{n\to\infty} \sum_{j=1}^n E W_{j,n}^2 = \sigma^2 \alpha' M(\xi)\alpha \tag{4.2}$$

and that

$$\lim_{n\to\infty} \sum_{j=1}^n E(W_{j,n} I(|W_{j,n}| > \delta))^2 = 0 \text{ for all } \delta > 0 \tag{4.3}$$

where I is the indicator function.

To verify Equation (4.2) note that $\sum_{j=1}^n E W_{j,n}^2 = \sigma^2 \alpha'(n^{-1} X_n' X_n)\alpha$. By Lemma 3.4, $\lim_{n\to\infty} (n^{-1} X_n' X_n) = M(\xi)$ so Equation 4.2 is true.

Finally to verify Equation 4.3, let $B = \sup_x |\alpha' f(x)|$. Since the f_i are bounded, $B < \infty$. Fix $\delta > 0$. Since $\epsilon_{1,n},\ldots,\epsilon_{n,n}$ are i.i.d. random variables,

$$\overline{\lim} \, \Sigma_{j=1}^n EW_{j,n}^2 \, I(|W_{j,n}| > \delta)$$

$$\leq B^2 \, \overline{\lim} \, n^{-1} \Sigma_{j=1}^n E\epsilon_{j,n}^2 I(|\epsilon_{j,n}| > \delta n^{1/2}B^{-1})$$

$$= B^2 \, \overline{\lim} \, E\epsilon_{1,1}^2 \, I(|\epsilon_{1,1}| > \delta n^{1/2}B^{-1})$$

$$= 0.\|$$

In theorem 4.1, the asymptotic mean b depends on the unknown function m(x). In order for Theorem 4.1 to be useful for making large sample inferences about $\beta(m,\xi)$, say to construct confidence sets for $\beta(m,\xi)$, b must be zero for all bounded, continuous functions m(x). Lemmas 3.3 and 3.4 show that $(X_n'X_n)^- X_n'm_n$ converges to $\beta(m,\xi)$ for all bounded, continuous m(x). It is reasonable to expect that b = 0 if ξ_n converges to ξ fast enough in some sense. Conditions under which this is true are examined in the remainder of this section. A useful result in this regard is given in Corrollary 4.1. It should be noted that one important situation in which b = 0 is if $m(x) = \beta'f(x)$ for some β. In this case $(X_n'X_n)^- X_n'm_n = \beta = \beta(m,\xi)$ for all large n.

To simplify notation, for the remainder of this section assume that $X = (u,v)$ an interval on the real line. Similar results hold if x is a r-dimensional vector. Let $\xi_n(x)$ and $\xi(x)$ also denote the distribution functions of the probability measures ξ_n and ξ. Assume there is a function h(x) such that $\lim_{n \to \infty} n^{1/2}(\xi_n(x) - \xi(x)) = h(x)$ almost everywhere with respect to Lebesque measure on (u,v). Assume that $\int_u^v \sup_n n^{1/2}|\xi_n(x) - \xi(x)|dx < \infty$. Assume the functions $m, f_1,\ldots, f_p$ are differentiable on (u,v). Let g' denote the derivative of a function g. The mean vector b in Theorem 3.1 can be written in terms of the function h. To obtain this result, the following lemma will be used.

Lemma 4.1: Let $\phi(x)$ be a real valued function defined on (u, v) with a derivative ϕ' and let G(x) be a distribution function with support contained in

(u, v). Then

$$\int_u^V \phi(x)dG(x) = \int_u^V \phi'(x)(1 - G(x))dx + \phi(u). \tag{4.4}$$

Proof: $\int_u^V \phi(x)dG(x) = \int_u^V \int_u^x \phi'(u)du\, dG(x) + \phi(u).$

Equation 4.4 follows from Fubini's Theorem (Loéve (1963), p. 135) by inter-changing the integration order. ||

The next theorem gives an expression for the mean b of Theorem 4.1 in terms of matrices which depend on the function h. Let d denote the $p \times 1$ vector with (i)th coordinate $-\int_u^V (f_i(x)m(x))'h(x)dx.$ Let D denote the $p \times p$ matrix with (i, j)th element $\int_u^V (f_i(x)f_j(x))'h(x)dx.$

Theorem 4.2: Assume $M'(\xi)$ is non-singular. Then $b = M^{-1}(\xi)(d + DM^{-1}(\xi)c(\xi)).$

Proof: Let $A_n = n^{-1}X_n'X_n.$ By Lemma 3.4, A_n is non-singular for all large n so we shall write A_n^{-1} for A_n^-

$$b = \lim_{n\to\infty} n^{1/2}(A_n^{-1}(n^{-1}X_n'm_n) - M^{-1}(\xi)c(\xi))$$

$$= \lim_{n\to\infty} n^{1/2}(A_n^{-1}(n^{-1}X_n'm_n - c(\xi)) + (A_n^{-1} - M^{-1}(\xi))c(\xi)).$$

To prove the desired result it suffices to prove that

$$\lim_{n\to\infty} n^{1/2}A_n^{-1}(n^{-1}X_n'm_n - c(\xi)) = M^{-1}(\xi)d. \tag{4.5}$$

and

$$\lim_{n\to\infty} n^{1/2}(A_n^{-1} - M^{-1}(\xi))c(\xi) = M^{-1}(\xi)DM^{-1}(\xi)c(\xi). \tag{4.6}$$

By Lemma 3.4, $\lim_{n\to\infty} A_n^{-1} = M^{-1}(\xi).$ Thus to prove Equation 4.5 it suffices to show that

$$\lim_{n\to\infty} n^{1/2}(\int_u^V f_r(x)m(x)d\xi_n(x) - \int_u^V f_r(x)m(x)d\xi(x))$$

$$= -\lim_{n\to\infty} \int_u^V (f_r(x)m(x))'n^{1/2}(\xi_n(x) - \xi(x))dx \tag{4.7}$$

$$= -\int_u^V (f_r(x)m(x))'h(x)dx, \quad r = 1,\ldots,p.$$

The first equality is true by Lemma 4.1 and the second equality is true by the assumptions made about $n^{1/2}(\xi_n(x) - \xi(x))$ and the Dominated Convergence Theorem

(Loéve (1963), p. 125).

Note that $A_n^{-1} - M_n^{-1}(\xi) = A_n^{-1} (M(\xi) - A_n)M^{-1}(\xi)$. By Lemma 3.4,
$\lim_{n\to\infty} A_n^{-1} = M^{-1}(\xi)$. Thus to prove Equation 4.6 it suffices to show that

$$\lim_{n\to\infty} n^{1/2} (\int_u^v f_r(x)f_s(x)\, d\xi(x) - \int_u^v f_r(x)f_s(x)\, d\xi_n(x))$$

$$= \lim_{n\to\infty} \int_u^v (f_r(x)f_s(x))^{\wedge} n^{1/2} (\xi_n(x) - \xi(x))dx \qquad (4.8)$$

$$= \int_u^v (f_r(x)f_s(x))h(x)dx, \quad r = 1,\ldots,p;\ s = 1,\ldots, p.$$

The first equality is true by Lemma 4.1 and the second equality is true by the assumptions made about $n^{1/2}(\xi_n(x) - \xi(x))$ and the Dominated Convergence Theorem.||

Corrollary 4.1: Assume $M(\xi)$ is non-singular. If $\lim_{n\to\infty} n^{1/2}(\xi_n(x) - \xi(x)) = 0$ almost everywhere with respect to Lebesque measure on (u, v) then $n^{1/2}(\hat\beta_n - \beta(m,\xi))$ converges weakly to a multivariate normal random vector with mean 0 and covariance matrix $\sigma^2 M^{-1}(\xi)$.

Proof: $h(x) = \lim_{n\to\infty} n^{1/2}(\xi_n(x) - \xi(x)) = 0$ a.e. implies $d = 0$ and $D = 0$. Thus by Theorem 4.2, the mean b in Theorem 4.1 is 0.||

5. Examples

In this section two examples are considered. In the first example a fairly general method of constructing designs which satisfy the conditions of Corollary 4.1 is given. The second example gives a sequence of designs for which the function $h(x)$ is non-zero.

Example 1: Let $\xi(x)$ be a fixed continuous and strictly increasing distribution function on the real line. Let $\xi^{-1}(x)$ denote the inverse of $\xi(x)$. Let $n_1, n_2,\ldots$ and $m_1, m_2,\ldots$ be sequences of positive integers such that n_i/m_i is an integer for each i and $\lim_{i\to\infty} m_i/n_i^{1/2} = 0$. For $k = 1,\ldots, n_i/m_i$, let $I_{k,i}$ denote the interval $(\xi^{-1}((k-1)m_i/n_i), \xi^{-1}(km_i/n_i))$. Let ξ_{ni} denote any design with m_i observations in each of the intervals $I_{k,i}$. Then ξ_{ni} converges weakly to ξ. Furthermore, $\sup_X |\xi_{n_i}(x) - \xi(x)| \le m_i/n_i$. So since $\lim_{i\to\infty} m_i/n_i^{1/2} = 0$,

$\lim_{i \to \infty} n_i^{1/2}(\xi_{n_i}(x) - \xi(x)) = 0$. By Corollary 4.1, the mean vector in Theorem 4.1 is zero for a sequence of designs constructed in this way. A special case of interest is the case of $m_i = N$ for some fixed N and $n_i = Ni$.

Example 2: Here is an example of a sequence of designs for which the $h(x)$ function is non-zero and, by Theorem 4.2, the mean vector b in Theorem 4.1 is not zero for all mean functions $m(x)$. Let $[u, v] = [0, 1]$ and let ξ be the uniform distribution on $[0, 1]$. Let $n_i = i^2$ for $i = 1, 2, \ldots$. Let ξ_{n_i} be the design which takes one observation at each of the $i^2 - i$ points $1/i + k/i^2$, $k = 0, 1, \ldots, i^2 - i - 1$, and i observations at 1. For each i, $x - 1/i < \xi_{n_i}(x) \le x$ for all $x \in [0, 1]$ so $\xi_{n_i}(x)$ converges to $\xi(x)$. But for any i and any $x \in (1/i, 1)$, $-1/i < \xi_{n_i}(x) - \xi(x) \le -1/i + 1/i^2$. Thus for all $x \in (0, 1)$, $h(x) = \lim_{i \to \infty} n_i^{1/2}(\xi_{n_i}(x) - \xi(x)) = \lim_{i \to \infty} i(\xi_{n_i}(x) - \xi(x)) = -1$. Suppose $f_1(x) = 1$ and $f_2(x) = x$. Suppose $m(x) = x^2$. Then using the result of Theorem 4.2, the mean vector b can be calculated to be $b = (-1, 2)$.

References

Anderson, T.W. and Taylor, J.B. (1979) Strong consistency of least squares estimates in dynamic models, Ann. Statist., 7, 484-489.

Atwood, C.E. (1971) Robust procedures for estimating polynomial regression, JASA, 66, 855-860.

Billingsley, P. (1968) Convergence of Probability Measures, John Wiley & Sons, Inc., New York.

Box, G.E.P. and Draper, N.R. (1959) A basis for the selection of response curve design, JASA, 54, 622-654.

Draper, N. and Smith, H. (1966) Applied Regression Analysis, John Wiley & Sons, Inc., New York.

Drygas, H. (1976) Weak and strong consistency of least squares estimators in regression models, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 34, 119-127.

Huber, P.J. (1975) Robustness and designs, A Survey of Statistical Design and Linear Models, J.N. Srivastava, ed., North Holland Pub Co., Amsterdam, 287-301.

Karlin, S. and Studden, W.J. (1966) Optimal experimental designs, Ann. Math. Statist., 37, 783-815.

Kiefer, J. (1962) Two more criteria equivalent to D-optimality of designs, Ann. Math. Statist., 33, 792-796.

Kiefer, J. and Wolfowitz, J. (1959) Optimum designs in regression problems, Ann. Math. Statist., 30, 271-294.

Lai, T.L., Robbins, H. and Wei, C.Z. (1978) Strong consistency of least squares estimates in multiple regression, Proc. Natl. Acad. Sci., USA, 75, 3034-3036.

Loeve, M. (1963) Probability Theory, 3rd Edition, D. Van Nostrand Co., Inc., Princeton.

Royall, R.M. and Herson, J. (1973) Robust estimation in finite populations I, JASA, 68, 880-889.

Stigler, S.M. (1971) Optimum experimental design for polynomial regression, JASA, 66, 311-318.

Wu, C.F. (1980) Characterizing the consistent directions of least squares estimates, Ann. Statist., 8, 739-801.

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1. REPORT NUMBER FSU No. M573 USARO No: D-50 AFOSR No: 78-120	2. GOVT ACCESSION NO. AD-A098 169	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and subtitle) Linear Least Squares Estimates and Nonlinear Means		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER FSU Statistics Report M 573
7. AUTHOR(s) Roger L. Berger , Naftali A. Langberg		8. CONTRACT OR GRANT NUMBER(s) USARO DAAG 29-79-C-0158 AFOSR 78-3678
9. PERFORMING ORGANIZATION NAME AND ADDRESS The Florida State University Department of Statistics Tallahassee, FL. 32306		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Office-Durham, P.O. Box 12211 Research Triangle Park, NC 27709 U.S. Air Force, Air Force Office of Scientific Research, Bolling Air Force Base, D.C. 20332		12. REPORT DATE March, 1981
		13. NUMBER OF PAGES 13
		15. SECURITY CLASS. (of this report) Unclassified
14. Monitoring Agency Name & Address (if different from Controlling Office)		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS

Consistency, asymptotic normality, best linear approximation, model robustness.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The consistency and asymptotic normality of a linear least squares estimate of the form $(X'X)^{-}X'Y$ when the mean is not $X\beta$ is investigated in this paper. The least squares estimate is a consistent estimate of the best linear approximation of the true mean function for the design chosen. The asymptotic normality of the least squares estimate depends on the design and the asymptotic mean may not be the best linear approximation of the true mean function.

DAT
ILM