Strong consistency of least-squares estimates in regression models

T. L. LAI AND HERBERT ROBBINS

Department of Mathematical Statistics, Columbia University, New York, New York 10027

Contributed by Herbert Robbins, April 26, 1977

ABSTRACT A general theorem on the limiting behavior of certain weighted sums of i.i.d. random variables is obtained. This theorem is then applied to prove the strong consistency of least-squares estimates in linear and nonlinear regression models with i.i.d. errors under minimal assumptions on the design and weak moment conditions on the errors. We consider the linear regression model

$$y_i = \alpha + \beta x_i + \epsilon_i \quad (i = 1, 2, \ldots)$$

[1]

where \(\epsilon_1, \epsilon_2, \ldots\) are i.i.d. random variables with \(E\epsilon_1 = 0, E\epsilon_1^2 = \sigma^2, 0 < \sigma^2 < \infty\), and \(x_1, x_2, \ldots\) is an arbitrary sequence of constants, not all equal. The least-squares estimate of \(\beta\) based on \(x_1, \ldots, x_n\) is

$$b_n = \frac{\sum_{i=1}^{n} (x_i - \bar{x}_n) y_i}{\sum_{i=1}^{n} (x_i - \bar{x}_n)^2},$$

[2]

where

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

From [1] we see that

$$b_n - \beta = \frac{\sum_{i=1}^{n} (x_i - \bar{x}_n) \epsilon_i}{\sum_{i=1}^{n} (x_i - \bar{x}_n)^2},$$

[3]

so that

$$E(b_n - \beta)^2 = \sigma^2 / A_n,$$

[4]

where

$$A_n = \frac{\sum_{i=1}^{n} (x_i - \bar{x}_n)^2}{n - 1}.$$

Hence, if the condition

$$\lim_{n \to \infty} A_n = \infty,$$

[5]

holds, then \(b_n\) converges to \(\beta\) in mean square, and hence in probability, as \(n \to \infty\). (In fact, as we shall see, [5] is a necessary as well as sufficient condition for \(b_n\) to converge to \(\beta\) in probability.) We shall be concerned with finding mild conditions on the distribution of \(\epsilon_1\) such that [5] implies the convergence of \(b_n\) to \(\beta\) with probability 1. Theorem 1 below shows that this holds whenever \(E(\epsilon_1^2 (1 + |\epsilon_1|)^r) < \infty\) for some \(r > 1\). A related problem has recently been considered by Anderson and Taylor (refs. 1 and 2).

From [3], \(b_n\) will converge to \(\beta\) if and only if

$$\sum_{i=1}^{n} c_{ni} \epsilon_i \to 0 \text{ as } n \to \infty,$$

[6]

where

$$c_{ni} = (x_i - \bar{x}_n) / \sum_{i=1}^{n} (x_i - \bar{x}_n)^2.$$

It is known (see ref. 3 and its bibliography) that for an arbitrary double array \(c_{ni}\), \(\Sigma_{n=1}^{\infty} c_{ni} \epsilon_i\) will converge to 0 with probability 1 provided that \(\epsilon_1\) is generalized Gaussian and that \(\Sigma_{n=1}^{\infty} c_{ni}^2 = O(1/\log n)\). In terms of the \(x\)-sequence, the latter condition becomes

$$A_n / \log n \to \infty \text{ as } n \to \infty,$$

which is much stronger than [5]. In order to show that [5] suffices, it will be necessary to exploit the special structure of the double array \(c_{ni}\) defined by Eq. 6. It turns out that the sequence \(\{\sum_{i=1}^{n} (x_i - \bar{x}_n) \epsilon_i; n \geq 1\}\) is a wide-sense martingale (ref. 4, p. 164). From this fact and a strong invariance principle we shall be able to prove

**THEOREM 1.** Let \(\epsilon_1, \epsilon_2, \ldots\) be i.i.d. with \(E\epsilon_1 = 0\) and \(E\epsilon_1^2 (\log (1 + |\epsilon_1|)^r) < \infty\) for some \(r > 1\), and let \(A_n\) be defined by [4] for some sequence of constants \(x_1, x_2, \ldots\) not all equal. If [5] holds, then for every \(\delta > 0\),

$$\lim_{n \to \infty} \frac{\sum_{i=1}^{n} (x_i - \bar{x}_n) \epsilon_i}{n^{1/2} (\log A_n)^{1/2}} = 0 \text{ with probability 1.}$$

[7]

It follows that if [5] holds, then \(b_n \to \beta\) with probability 1 as \(n \to \infty\).

We preface the proof of Theorem 1 by two lemmas.

**LEMMA 1.** Let \(\epsilon_1, \epsilon_2, \ldots\) be uncorrelated random variables such that \(E\epsilon_1 = 0, E\epsilon_1^2 = \sigma^2 < \infty\) for all \(n\) and let \(x_1, x_2, \ldots\) be any sequence of constants. Define

$$t_n = \frac{n}{\sum_{i=1}^{n} (x_i - \bar{x}_n) \epsilon_i}, \quad w_n = u_n - u_{n-1}.$$  

[8]

Then

$$E(u_m u_n) = 0 \text{ if } m \neq n,$$

[10]

so that \(\{u_n; n \geq 1\}\) is a wide-sense martingale.

The proof is straightforward. As a consequence we obtain that \(E(u_n^2) = 2 \sigma^2 E(w_n^2)\), which is equivalent to the interesting algebraic identity

$$A_n = \frac{\sum_{i=1}^{n} (x_i - \bar{x}_n)^2}{\frac{n}{\sum_{i=1}^{n} (x_i - \bar{x}_n)^2}} \left(1 - \frac{1}{n}ight) (x_j - \bar{x}_{j-1})^2.$$  

[11]

This shows that the sequence \(A_n\) is nondecreasing in \(n\). Hence, if the \(x's\) are not all equal, either [5] holds or \(A_n \to c\) for some \(0 < c < \infty\). In the latter case, by the convergence theorem for wide-sense martingales (ref. 4, p. 165), \(u_n\) converges in \(L_2\) and hence in probability to some random variable \(u\) with \(E u = 0\) and \(E(u^2) = \sigma^2 c\). It follows that \(b_n - \beta = u_n / A_n\) converges in probability to \(v = u / c\), with \(Ev = 0, Ev^2 = \sigma^2 > 0\). Hence \(b_n\) cannot converge in probability to \(\beta\); that is, the condition [5] is
necessary, as well as sufficient, for \( b_n \) to be a weakly consistent estimator of \( \beta \).

When \( \epsilon_1, \epsilon_2, \ldots \) are i.i.d. \( \mathcal{N}(0, \sigma^2) \), the orthogonality relations [10] are equivalent to independence, so that Lemma 1 implies

**Lemma 2.** Let \( \epsilon_1, \epsilon_2, \ldots \) be i.i.d. \( \mathcal{N}(0, \sigma^2) \). Then the sequence

\[
|u_n; n \geq 1| \text{ has the same joint distribution as } [W(\sigma^2 A_n); n \geq 1], \text{ where } W(t) \text{ for } t \geq 0 \text{ denotes the standard Wiener process.}
\]

**Proof of Theorem 1:** Let

\[
d_j = \frac{1}{\sqrt{j}} (\bar{x}_j - \bar{x}_{j-1}).
\]

By [5] and [11],

\[
\sum_{j=1}^{n} d_j^2 \sim A_n \to \infty \text{ as } n \to \infty,
\]

and by [8] and [9],

\[
\sum_{j=1}^{n} d_j^2 = \sum_{j=1}^{n} d_j (\epsilon_j - \bar{x}_j) = \sum_{j=1}^{n} \left( \frac{n-j}{n} d_j^2 \right) = \sum_{j=1}^{n} d_j (\epsilon_j - \bar{x}_j).
\]

In view of [13],

\[
\sum_{j=1}^{n} d_j^2 / A_n (\log A_n)^{1+b} < \infty
\]

by the integral comparison test, and hence by Kronecker's lemma and a theorem of Kolmogorov

\[
\lim_{n \to \infty} \frac{\sum_{j=1}^{n} d_j^2}{A_n (\log A_n)^{1+b}} = 0 \text{ with probability 1.}
\]

Take \( 1 < p < r \) and let \( \sigma^2 = E(\epsilon_1^2) \). Since \( E(\epsilon_1^2 (\log(1 + |\epsilon_1|))^r) < \infty \), there exist (see Theorem 1 of ref. 5) i.i.d. \( \mathcal{N}(0, \sigma^2) \) random variables \( \epsilon_0, \epsilon_2, \ldots \) such that (by redefining the random variables on a new probability space if necessary)

\[
(\log n)^{p/2} \left( \sum_{i=1}^{n} \epsilon_i - \sum_{i=1}^{n} \epsilon_i^* \right)^2
\]

\[
\to 0 \text{ with probability 1.}
\]

It follows from [16] and the Schwarz inequality that with probability 1

\[
\left( \sum_{j=1}^{n-1} d_j + d_{n+1} (\bar{x}_n - \bar{x}_{n-1}) \right) = 0 \text{ with probability 1.}
\]

Since \( \sum_{j=1}^{n} d_j^2 \sim A_n \), the above implies

\[
\sum_{j=1}^{n} d_j^2 / A_n (\log A_n)^{1+b/2} = 0 \text{ with probability 1.}
\]

Let \( u_{n*} = \epsilon_0 + \epsilon_2 + \cdots + \epsilon_{n-1} \). From Lemma 2, with probability 1.

\[
\limsup_{n \to \infty} |u_{n*}| / \sigma \sqrt{2A_n \log \log A_n} \leq 1
\]

with probability 1. Furthermore, from Eq. 14, \( u_{n*} = \sum_{j=1}^{n} d_j (\epsilon_j - \bar{x}_j) \), and hence, for \( n \geq 2 \),

\[
\limsup_{n \to \infty} \left( \frac{\sum_{j=1}^{n} d_j (\epsilon_j - \bar{x}_j)}{A_n (\log A_n)^{1+b/2}} \right) = 0 \text{ with probability 1.}
\]

From [14], [15], and [21], the desired conclusion [7] follows.

By [17] and [20],

\[
\lim_{n \to \infty} \left( \sum_{j=1}^{n} d_j (\epsilon_j - \bar{x}_j) / \sigma \sqrt{2A_n \log \log A_n} \right) = 0 \text{ with probability 1.}
\]

An examination of the above proof shows that [22] would hold if

\[
\limsup_{n \to \infty} \left( \sum_{j=1}^{n} d_j (\epsilon_j - \bar{x}_j) / \sigma \sqrt{2A_n \log \log A_n} \right) = 1
\]

It is natural to ask under what conditions we can sharpen the conclusion [7] of Theorem 1 to

\[
\limsup_{n \to \infty} \left( \sum_{j=1}^{n} d_j (\epsilon_j - \bar{x}_j) / \sigma \sqrt{2A_n \log \log A_n} \right) = 1
\]

with probability 1. A sufficient condition for the law of the iterated logarithm [23] (see ref. 6, p. 399) is

\[
\sum_{j=1}^{n} d_j^2 / A_n (\log A_n)^{1+b} < \infty
\]

Since \( \sum_{j=1}^{n} d_j^2 / A_n \to \infty \) for every \( r > 1 \) by [13] and the integral comparison test, we therefore obtain the following theorem.

**Theorem 2.** Let \( \epsilon_1, \epsilon_2, \ldots \) be i.i.d. such that \( E(\epsilon_1^2) = 0 \) and \( E(\epsilon_1^2 (\log(1 + |\epsilon_1|))^r) < \infty \) for some \( r > 1 \). Let \( (x_n) \) be a sequence of real constants and let \( A_n \) be defined by [4]. Suppose [5] holds and

\[
(x_n - \bar{x}_{n-1})^2 = 0(A_n^{1-r}) \text{ for some } r > 0.
\]

Then [22] holds.

**Extension to Nonlinear Regression.** Consider the nonlinear regression model

\[
y_i = M(x_i) + \epsilon_i \quad (i = 1, 2, \ldots)
\]

where the errors \( \epsilon_1, \epsilon_2, \ldots \) are i.i.d. with \( E(\epsilon_1^2) = 0 \), and \( M(x) \) is an unknown real-valued function that is twice continuously differentiable in some open neighborhood of a given point \( \theta \). To estimate \( M(\theta) \), we can set the levels \( x_i \) around \( \theta \) so that

\[
\lim_{n \to \infty} x_n = \theta \text{ with probability 1.}
\]

Under this condition on the design, it is easy to see that \( \overline{y}_n \) is a strongly consistent estimator of \( M(\theta) \). The following theorem gives a strongly consistent estimator of \( M'(\theta) \).

**Theorem 3.** For the above nonlinear regression model, suppose the design satisfies conditions [5] and [27], and assume that \( E(\epsilon_1^2 (\log(1 + |\epsilon_1|))^r) < \infty \) for some \( r > 1 \). Then \( b_n \), as defined by [2], is a strongly consistent estimator of \( M'(\theta) \).

To prove Theorem 3, by the assumptions on \( M(x) \) we can write

\[
M(x) = M(\theta) + M'(\theta)(x - \theta) + g(x)(x - \theta),
\]

with probability 1. Moreover, from Eq. 14, \( u_{n*} = \sum_{j=1}^{n} d_j (\epsilon_j - \bar{x}_j) = \sum_{j=1}^{n} d_j (\epsilon_j - \bar{x}_j - \epsilon_j) \), so

\[
\limsup_{n \to \infty} \left( \sum_{j=1}^{n} d_j (\epsilon_j - \bar{x}_j - \epsilon_j) / \sigma \sqrt{2A_n \log \log A_n} \right) \leq 2
\]

with probability 1. By [17] and [20],

\[
\limsup_{n \to \infty} \left( \sum_{j=1}^{n} d_j (\epsilon_j - \bar{x}_j - \epsilon_j) / \sigma \sqrt{2A_n \log \log A_n} \right) \leq 2
\]

with probability 1.

where \( \lim_{x \to y} g(x) = 0 \) and \( g(x) \) is continuously differentiable in some closed interval \( I \) containing \( 0 \) in its interior. Therefore there exists \( C > 0 \) such that
\[
|g(x) - g(y)| \leq C|x - y| \quad \forall \ x, y \in I.
\] [29]

Let \( k_i = g(x_i) \). From [2], [26], and [28],
\[
b_n = M'(\theta) + \frac{\sum_{i=1}^{n} k_i(x_i - \theta)(x_i - \bar{x}_n)}{\sum_{i=1}^{n} (x_i - \bar{x}_n)^2} + \frac{\sum_{i=1}^{n} (x_i - \bar{x}_n)k_i}{\sum_{i=1}^{n} (x_i - \bar{x}_n)^2}. \] [30]

In view of Theorem 1, it therefore suffices to show that as \( n \to \infty \),
\[
\sum_{i=1}^{n} k_i(x_i - \theta)(x_i - \bar{x}_n) = o \left( \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 \right). \] [31]

By [27], \( \lim_{i \to \infty} k_i = 0 \). This, together with [5] and [11], implies that
\[
\sum_{i=1}^{n} k_i(x_i - \bar{x}_n)^2 = o \left( \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 \right). \] [32]

Hence, writing \( x_i - \theta = (x_i - \bar{x}_n) + (\bar{x}_n - \theta) \) in [31], we need only show that
\[
(\bar{x}_n - \theta) \sum_{i=1}^{n} k_i(x_i - \bar{x}_n) = o \left( \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 \right). \] [33]

Let \( \bar{k}_n = n^{-1} \sum_{i=1}^{n} k_i \). In view of [27], [33] would follow if it can be shown that
\[
\sum_{i=1}^{n} (k_i - \bar{k}_n)(x_i - \bar{x}_n) = o \left( \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 \right). \] [34]

Therefore, by the Schwarz inequality, it suffices to show that as \( n \to \infty \),
\[
\sum_{i=1}^{n} (k_i - \bar{k}_n)^2 = o \left( \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 \right). \] [35]

By [29], for all sufficiently large \( i \) and \( j \),
\[
|k_i - k_j| = |g(x_i) - g(x_j)| \leq C|x_i - x_j|.
\]

Hence, [35] follows immediately from the identity
\[
n \sum_{i=1}^{n} (k_i - \bar{k}_n)^2 = \sum_{1 \leq i < j \leq n} (k_i - k_j)^2.
\]

This research was supported by the National Science Foundation and the U.S. Public Health Service under Grants NSF-MCS-75-20905 and R01-GM-16905-08.

The costs of publication of this article were defrayed in part by the payment of page charges from funds made available to support the research which is the subject of the article. This article must therefore be hereby marked "advertisement" in accordance with 18 U. S. C. §1734 solely to indicate this fact.


Statistics: Lai and Robbins