BULLETIN OF THE
INSTITUTE OF MATHEMATICS
ACADEMIA SINICA
Volume 10, Number 1, March 1982

CONSISTENCY IN REGRESSION MODELS SPECIFIED BY DIFFERENTIAL EQUATIONS(1)

BY

I. S. CHANG (張憶壽) AND CHAO A. HSIUNG (熊昭)

To the Memory of Jack Kiefer(2)

Abstract. For each $t \in G$, a bounded open subset of R^n, let F_t be a distribution function with mean $m(t)$. Suppose $m(t)$ satisfies an ordinary or a partial differential equation. Let $t_1, t_2, \cdots$ be a sequence of points in G and $X_1, X_2, \cdots$ be an independent sequence of random variables such that the distribution function of X_i is F_{t_i}. We consider estimators $M_n(t) = M_n(X_1, X_2, \cdots, X_n)(t)$ which are classical approximate solutions of the differential equation in question and which assume least-squares from the observation among the class of all classical approximate solutions. We investigate the consistency of these estimators for the cases of a first order ordinary differential equation and a second order elliptic partial differential equation.

1. Introduction. The theory of regression analysis is developed to study the mean behavior of some quantitative variable which is in part randomly determined and in part a function of some set of other variables. The existing theory is often classified in terms of the properties of the non-random function involved. Since differential equations arise frequently when one wants to describe deterministic natural phenomena, it seems desirable to investigate regression models in which the non-random functions satisfy certain differential equations. In these situations, one possible source for random effect is measurement error. In order to illustrate the main ideas and

---

Received by the editors December 8, 1981.

(1) Part of the work was done while both authors were visiting the Department of Mathematical Statistics at Columbia University with fellowships from N. S. C. of R. O. C.

(2) In a letter dated April 29, 1981, Professor J. Kiefer supported the authors in an argument with some people about this work. His sudden death makes the authors, deeply indebted to his precious time and scholarly advice, decide to publish the paper as it is so that his points in that letter are best incorporated.

AMS 1970 subject classifications. Primary 62J05; Secondary 62G20.
Key words and phrases. Strong consistency, Least-squares estimates. Differential equations.

methods of this work, we consider in this paper only two interesting examples, without goint into mathematical generalities.

The first example, discussed in §2, concerns regression model with mean function satisfying a first order ordinary differential equation. We propose for the mean function an estimator based on the method of least squares and the classical method of Cauchy-Euler. (see [4]) The second example, treated in §3, concerns regression model whose mean function satisfies an elliptic partial differential equation. The estimator proposed for its mean function is suggested by the method of least squares and that of finite difference. Finally, in §4, we give remarks about the design of experiment for the ODE case and the optimality of our main result for the PDE case.

To motivate our work, we recapitulate here the main ideas of least squares method in polynomial regression and concave regression. In both cases, certain functions are specified to form the admissible set from which we choose the one assuming the least squares from the observations to be our estimate. The situation is particularly simple for polynomial regression, whose admissible set consists of polynomials of the same degree as its mean functions irrespective of the observing points. For concave regression, however, the admissible set consists of concave functions which are linear except possibly at the observing points. Note that piecewise linearity is imposed to simplify the matters while concavity is desired naturally to fulfill the requirement of the model. ([3])

Since the mean function of our model is the solution of a differential equation, our first choice of admissible set would naturally comprise all solutions of the differential equation in question. On the other hand, we find that, since it is rather difficult and even impossible in some cases to specify explicitly solutions of differential equations, applying least squares method to this admissible set can only make it more unmangeable. Faced with this complication, we compromise and take certain approximate solutions of the differential equation in question to form our admissible set. This conforms to the practice in deterministic mathematics.

Since it has been a good idea to take advantage of some kind of Cauchy-Euler methods to obtain approximate solutions for a first order ODE, we, in section 2, consider the approximate solutions in Cauchy-Euler sense as admissible functions. The rationale behind our choice of admissible set for the PDE case in section 3 can be similarly supplied and is therefore omitted.

It is to be noted that respective consistency theorems are proved for both examples by deliberately utilizing two different approaches. Algorithms and simulation results will be reported elsewhere.

2. Regression model of first order ordinary differential equations.

2.1. The estimator. Let $t_1, t_2, \cdots$ be a sequence of real numbers between 0 and 1. To simplify the presentation, we will assume that all the t_i's are distinct. Let $X_1, X_2, \cdots$ be a sequence of random variables such that $X_i - m(t_i)$ is independent and identically distributed with mean 0 for all $i = 1, 2, \cdots$, where m is a solution of the ordinary differential equation

$$g'(t) = f(t, g(t)) \tag{2.1.1}$$

on $[0, 1]$, for some continuous function f on a neighborhood of $[0, 1] \times R$ satisfying Lipschitz condition in the second variable.

Let $T_n = \{t_1, \cdots, t_n\}$. Let T_n' be a finite subset of $[0, 1]$ containing T_n. Let $S_n = (s_{n1}, \cdots, s_{nb(n)})$ be the ordered set obtained from T_n', where $b(n)$ is the number of elements of T_n'. Assume that the mesh of S_n as a partition of $[0, 1]$ tends to 0 when n gets large. The classical method of Cauchy-Euler adopts the elements of

$$A_n = \{g : g \text{ is a continuous piecewise linear function on } [0, 1] \text{ such that i) } g \text{ is linear at } s_{n1}, \text{ ii) } g'(s) = f(s_{ni}, g(s_{ni})) \text{ if } s_{ni} < s < s_{n(i+1)} \text{ for } i = 1, 2, \cdots, b(n)-1 \text{ or } s_{nb(n)} < s < 1\}$$

as approximate solutions of (2.1.1). Regarding A_n as a subset of R^n by identifying g with $(g(t_1), \cdots, g(t_n))$, we define $M_n = M_n(X_1, \cdots, X_n)$ to be any element of A_n which is closest to the observation $(X_1, \cdots, X_n)$. Because A_n is closed in R^n, there is always an element of A_n which assumes the least distance. M_n is the estimator we propose for the regression model whose mean function satisfies (2.1.1).

2.2. Consistency of the estimator.

THEOREM. *Assume that* $E(|X_1|^2) < \infty$. *Then*

$$\lim_{n\to\infty} \max_{0\le t\le 1} |M_n(t) - m(t)| = 0 \quad a.e.$$

The proof of the theorem is accomplished by establishing the following lemmas.

LEMMA 2.1. *Almost surely there exists a constant M such that for n large enough, $|M_n(t) - m(t)| < M$ for some t in* $[0, 1]$.

Proof. We know, from standard results in the theory of ordinary differential equations (see [4]), that there exists m_n in A_n with $|m_n(t) - m(t)| < 1$ for all t in $[0, 1]$. For this m_n, we have

$$\begin{aligned}
&(1/n)\sum_{k=1}^{n} (M_n(t_k) - m(t_k))^2 \\
&\le (2/n)\sum_{k=1}^{n} ((M_n(t_k) - X_k)^2 + (X_k - m(t_k))^2) \\
&\le (2/n)\sum_{k=1}^{n} ((m_n(t_k) - X_k)^2 + (X_k - m(t_k))^2) \\
&\le (6/n)\sum_{k=1}^{n} ((X_k - m(t_k))^2 + (m(t_k) - m_n(t_k))^2) \\
&\le (6/n)\sum_{k=1}^{n} (X_k - m(t_k))^2 + 6
\end{aligned}$$

Therefore, the lemma follows by applying law of large numbers to the last expression.

LEMMA 2.2. *Almost surely there exists a M such that, for n large enough, $|M_n(t)| \le M$ for all t in* $[0, 1]$.

Proof. This lemma is an immediate consequence of Lemma 2.1 and the claim that for every $M_1 > 0$, there exists $M > 0$ such that, when n is large enough, g in A_n with $\min_{[0,1]}|g(t)| \le M_1$ implies $\max_{[0,1]}|g(t)| \le M$.

To prove this claim, let g in A_n with $|g(t)| \le M_1$ for some t

in [0, 1]. We are going to find a $M > 0$ satisfying the above desired property. Let $a_i = s_{ni}$, $d_i = a_i - a_{i-1}$, $i = 1, 2, \cdots, b(n)$. Assume that $|g(a_j)| \leq M_1$ for some j.

It follows from the definition of A_n that for $k = 2, \cdots, b(n)$,

$$|g(a_k)| \leq |g(a_{k-1})|(1 + cd_k) + cd_k$$

and

$$|g(a_{k-1})| \leq e^{2cd_k}(|g(a_k)| + cd_k)$$

for n so large that $cd_k < \frac{1}{2}$, where c is a constant bigger than the Lipschitz constant of f and all the values of f on $[0, 1] \times \{0\}$.

Therefore, for $k > j$, we have

$$\begin{aligned} |g(a_k)| &\leq M_1 \sum_{j+1}^{k} (1 + cd_v) + c \sum_{j+1}^{k} d_v \prod_{v+1}^{k} (1 + cd_e) \\ &\leq M_1 \prod_{1}^{b(n)} (1 + cd_v) + c \sum_{1}^{b(n)} d_v \prod_{1}^{b(n)} (1 + cd_e) \\ &\leq M_1 e^c + c \sum_{1}^{b(n)} d_v e^c \\ &\leq (M_1 + c)e^c \end{aligned}$$

It follows from here that $|g(t)| \leq M$ for some suitable M, for all t bigger than a_j. A similar argument gives the desired result for t smaller than a_j. Therefore, the proof is completed.

LEMMA 2.3. *Let $0 \leq a < b \leq 1$ be given. Let $\varepsilon > 0$ be given. Almost surely that $|M_n(t) - m(t)| < \varepsilon$ for some t in $[a, b]$ for n large enough.*

Proof. Let $t_0 = (a + b)/2$. Suppose the conclusion is not correct, then there is a subsequence n_k such that

(2.2.1) $$|M_{n_k}(t_0) - m(t_0)| \geq \varepsilon$$

on a set of positive probability. Without loss of generality, we may assume that

(2.2.2) $$M_{n_k}(t_0) - m(t_0) \geq \varepsilon$$

one a set of positive probability.

Let w be the solution of (2.1.1) with $w(t_0) = m(t_0) + \varepsilon$. Let $\delta = (\frac{1}{3}) \min_{0 \leq t \leq 1} (w(t) - m(t))$. It follows from the fundamental inequality of ordinary differential equation (see [4]) and Lemma 2.2 that $M_{n_k}(t) - m(t) > 2\delta$, for all t in [0. 1], for all large n_k.

Define g_n to be the element of A_n with $g_n(s_{n1}) = m(s_{n1})$. We know from the fundamental inequality that $|g_n(t) - m(t)| < \delta$ for all t in [0, 1] when n gets large.

Consider

$$(2.2.3)\quad \begin{aligned} &(1/n)\sum_{k=1}^{n}[(X_k - M_n(t_k))^2 - (X_k - g_n(t_k))^2] \\ &\quad = (1/n)\sum_{k=1}^{n}(M_n(t_k) - g_n(t_k))^2 \\ &\qquad + (2/n)\sum_{k=1}^{n}(g_n(t_k) - m(t_k))(M_n(t_k) - g_n(t_k)) \\ &\qquad + (2/n)\sum_{k=1}^{n}(m(t_k) - X_k)(M_n(t_k) - g_n(t_k)) \end{aligned}$$

Since both M_n and g_n are approximate solutions of (2.1.1), it follows from Lemma 2.2 that all M_n and g_n have their slopes bounded by a common constant for all n. This implies that M_n and g_n are of bounded variation of a common bound for all n; therefore, Lemma 1 of [3] says that

$$(2.2.4)\quad (2/n)\sum_{k=1}^{n}(m(t_k) - X_k)(M_n(t_k) - g_n(t_k)) \longrightarrow 0 \quad \text{a. e.}$$

as n goes to ∞.

The fundamental inequality tells us also that for some constant D,

$$(2.2.5)\quad \begin{aligned} &\left|(2/n)\sum_{k=1}^{n}(g_n(t_k) - m(t_k))(M_n(t_k) - g_n(t_k))\right| \\ &\quad \le (D/n)\sum_{k=1}^{n}|g_n(t_k) - m(t_k)| \longrightarrow 0 \quad \text{a. e.} \end{aligned}$$

as n goes to ∞.

Put (2.2.4) and (2.2.5) in (2.2.3), we know that on a set of positive probability

$$(2.2.6)\quad \overline{\lim_{n\to\infty}}\,(1/n)\sum_{k=1}^{n}((X_k - M_n(t_k))^2 - (X_k - g_n(t_k))^2) \ge \delta^2 .$$

This contradicts the fact that M_n assumes the least square, hence the proof is completed.

LEMMA 2.4. *Let* $\varepsilon > 0$ *be given. Then the event*

$$\{\max_{0\le t\le 1} |M_n(t) - m(t)| > \varepsilon \quad \text{i. o.}\}$$

has probability 0.

Proof. We will prove only that the even

$$\{\max_{0\le t\le 1} (M_n(t) - m(t)) > \varepsilon \quad \text{i. o.}\}$$

has probability 0 and leave out the other part, which can be treated similarly.

Let ω be an element of $\{\max_{0\le t\le 1}(M_n(t) - m(t)) > \varepsilon$ i. o.$\}$. Then for this ω, there are a sequence of positive integers n_k and a sequence of numbers p_k in [0, 1] such that

$$M_{n_k}(p_k) > m(p_k) + \varepsilon \tag{2.2.7}$$

for all $k = 1, 2, \cdots$. Without loss of generality, we can assume p_k has a limit point p. Since both M_n and m have all their slopes bounded by a common constant for all n, there exists $\delta > 0$ such that

$$|u - v| < \delta, \qquad u, v \text{ in } [0, 1] \text{ implies}$$
$$|M_n(u) - M_n(v)| < \varepsilon/3 \qquad \text{for all} \quad n = 1, 2, \cdots$$

and

$$|m(u) - m(v)| < \varepsilon/3 \tag{2.2.8}$$

Consider $I = (p - \delta/2, p + \delta/2) \cap [0, 1]$. Let q be any point of I. We know from (2.2.7) and (2.2.8) that

$$M_{n_k}(q) - m(q) = M_{n_k}(q) - M_{n_k}(p) + M_{n_k}(p)$$
$$- m(p) + m(p) - m(q) > \varepsilon/3$$

Therefore, Lemma 2.3 tells us that this ω is a member of a set of probability 0. This completes the proof.

3. Regression model of an elliptic partial differential equation.

3.1. **Preliminaries.** Let $G \subset R^2$ be a bounded connected region. Let $\bar{G}$ denote the closure of G. For each t in $\bar{G}$, let F_t be a distribution function with mean $m(t)$, which has bounded derivatives of order 4 on G.

Let a_1, a_2 and a be continuous functions defined on G with $a \leq 0$, $|a_1|+|a_2| \leq K$, for some K. Let f be a bounded function on G. Assume that $m(t)$ satisfies the following elliptic partial differential equation

$$\frac{\partial^2 m}{\partial t_1^2} + \frac{\partial^2 m}{\partial t_2^2} + a_1 \frac{\partial m}{\partial t_1} + a_2 \frac{\partial m}{\partial t_2} + am = f \tag{3.1.1}$$

on G.

The regression model we are studying has its mean function a solution of (3.1.1).

We recall that a lattice L of R^2 is a set of points for which there exists a number h, called mesh, such that (p_1, p_2), (q_1, q_2) in L implies that both $p_1 - q_1$ and $p_2 - q_2$ are integral multiples of h. For $p = (p_1, p_2)$ in L, $(p_1 + h, p_2)$, $(p_1 - h, p_2)$, $(p_1, p_2 + h)$ and $(p_1, p_2 - h)$ are in L and constitute the neighborhood of p. Let $D \subset L$. Define the closure of D, denoted by $\bar{D}$, to be

$$\bar{D} = \{p \in L : p \text{ in } D \text{ or } p \text{ is in a neighborhood of some } q \text{ in } D\}.$$

A lattice domain is defined to be a subset D of some lattice such that $\bar{D} \subset \bar{G}$ and $\bar{D} \backslash D$ has all its points closer to ∂G than any points of D. For a lattice domain D we define a finite difference operator by letting

$$\begin{aligned} L_D(u)(t_1, t_2) = {} & (1/h^2)(u(t_1 + h, t_2) + u(t_1 - h, t_2) \\ & + u(t_1, t_2 + h) + u(t_1, t_2 - h) - 4u(t_1, t_2)) \\ & + a_1(t_1, t_2)(1/2h)(u(t_1 + h, t_2) - u(t_1 - h, t_2)) \\ & + a_2(t_1, t_2)(1/2h)(u(t_1, t_2 + h) - u(t_1, t_2 - h)) \\ & + a(t_1, t_2)u(t_1, t_2) \end{aligned} \tag{3.1.2}$$

for all (t_1, t_2) in D.

Identifying functions on $\bar{D}$ as points of R^d, where d is the number of points in $\bar{D}$, we know the set

$$H = \{u : u \text{ is a function on } \bar{D},\ L_D(u) = f \text{ on } D\}$$

is an affine-space in R^d. Let v be the unique function in H which agrees with m on $\bar{D} \backslash D$. Let w denote the restriction of m to $\bar{D}$. Then

$$\max_{t \in \bar{D}} |w(t) - v(t)| = O(h^2) \tag{3.1.3}$$

where h is the mesh of D and the constant involved in (3.1.3) depends only on K, the region G and the fourth derivatives of m. (see [1]) v, or a suitable interpolation of v, is regarded in the theory of differential equations as a respectable approximation of m if the boundary value of m is available.

Let P denote the perpendicular projection from R^d onto H. Using (3.1.3), we know

$$\max_{t\in\bar{D}}|w(t)-P(w(t))|=o(1) \tag{3.1.4}$$

as h goes to 0. It is based on (3.1.4) that we are going to propose an estimator.

3.2. **The estimator.** Let D_n be a sequece of lattice domains with meshes h_n tending to 0. Denote the number of elements in $\bar{D}_n$ by d_n. Then $d_n \leq C(1/h_n^2)$ for some constant C depending only on the diameter of G. Denote the elements of $\bar{D}_n$ by $t_i^{(n)}$, $1 \leq i \leq d_n$. Let $X_{ij}^{(n)}$ be independent random variables with distribution function $F_{t_i^{(n)}}$, where $j=1, 2,\cdots, b_n$; for some positive integer b_n. We think of $X_{ij}^{(n)}$ as the jth observation taken at $t_i^{(n)}$. Let $X_i^{(n)}=(1/b_n)\sum_{j=1}^{b_n}X_{ij}^{(n)}$. Consider $X^{(n)}=(X_1^{(n)},\cdots, X_{d_n}^{(n)})$ as a point in R^{d_n}. Difine $M_n=P(X^{(n)})$, the image of $X^{(n)}$ under the projection P described in §3.1. M_n, or a proper interpolation of M_n, is the estimator we propose for this regression model.

3.3. **Consistency Theorem.** Notations of this section bear the same meanings as those of §3.2.

THEOREM. i) *If* $E|X_{ij}^{(n)}|^r \leq A$ *for some constant* A, *for some* $r \geq 2$, *for all* $1\leq i\leq d_n$, $1\leq j\leq b_n$, $n=1, 2, 3,\cdots$, *then for* b_n, d_n *such that*

$$\sum_{n=1}^{\infty} d_n(1/b_n)^{r/2}<\infty, \tag{3.3.1}$$

we have

$$\lim_{n\to\infty}\max_{t\in\bar{D}_n}|M_n(t)-m(t)|=0 \quad a.e. \tag{3.3.2}$$

ii) *If* $X_{ij}^{(n)}-m(t_i^{(n)})$ *are identically distributed and* $Ee^{t|X_{ij}^{(n)}|}$ *exists for all* $t>0$, *for all* i, j, n, *then for* b_n, d_n *such that for each* $\varepsilon>0$, *there exists* $t>0$,

$$\sum_{1}^{\infty} d_n e^{-t\varepsilon b_n{}^{1/2}} < \infty, \qquad \textit{we then have } (3.3.2).$$

iii) *If* $X_{ij}^{(n)} - m(t_i^{(n)})$ *has normal distribution with mean* 0 *and variance* σ^2 *for all* i, j, n, *then for* b_n, d_n *such that for each* $\varepsilon > 0$, $\sum_1^\infty d_n b_n^{-1/2} e^{-\varepsilon b_n} < \infty$, *we then have* (3.3.2).

Proof. We will prove i) only. The proofs for both ii) and iii) will be omitted because they go the same way as that of i) except applying Lemmas 3.4.2 and 3.4.3 respectively, instead of Lemma 3.4.1, to (3.3.6).

Let N_n be the null space of the finite difference operator L_{D_n} defined by (3.1.2), i. e.

$$N_n = \{u : u \text{ is a function on } \bar{D}_n,\ L_{D_n}(u) = 0 \text{ on } D_n\}.$$

Let

$$H_n = \{u : u \text{ is a function on } \bar{D}_n,\ L_{D_n}(u) = f \text{ on } D_n\}.$$

Let A_n be the matrix, expressed in terms of usual basis of R^{d_n}, of the perpendicular projection from R^{d_n} onto N_n. Let P_n be the perpendicular projection from R^{d_n} onto H_n. Then

$$P_n(u) = P_n(0) + A_n(u) \tag{3.3.3}$$

for all u in R^{d_n}. Note that A_n is an idenpotent, i. e. $A_n^2 = A_n$. Consequently, we have

$$\max_{1\le d\le d_n} \sum_{1=i}^{d_n} (a_{di}^{(n)})^2 \le 1 \tag{3.3.4}$$

where $a_{di}^{(n)}$ is the (d, i)-enty of A_n.

Let w_n denote the restiction of m to $\bar{D}_n$. In view of (3.1.4), we need only to prove

$$\lim_{n\to\infty} \max_{t\in\bar{D}_n} |M_n(t) - P_n(w_n)(t)| = 0 \quad \text{a. e.} \tag{3.3.5}$$

Set $Z_{ij}^{(n)} = X_{ij}^{(n)} - m(t_i^{(n)})$, $Z_i^{(n)} = X_i^{(n)} - m(t_i^{(n)})$. Then

$$\begin{aligned}\max_{t\in\bar{D}_n}|P_n(X^{(n)})-P_n(w_n)| &= \max_{t\in\bar{D}_n}|A_n(X^{(n)})-A_n(w_n)|\\ &= \max_{t\in\bar{D}_n}|A_n(X^{(n)}-w_n)|\\ &= \max_{1\le d\le d_n}\left|\sum_{i=1}^{d_n} a_{di}^{(n)} Z_i^{(n)}\right|\\ &= \max_{1\le d\le d_n}\left|\sum_{i=1}^{d_n}\sum_{j=1}^{b_n} a_{di}^{(n)}(1/b_n)\, Z_{ij}^{(n)}\right|.\end{aligned} \tag{3.3.6}$$

Observe that

$$\begin{aligned}\sum_{n=1}^{\infty}\sum_{d=1}^{d_n}\Big(\sum_{i=1}^{d_n}\sum_{j=1}^{b_n}(a_{di}^{(n)}(1/b_n))^2\Big)^{r/2} &= \sum_{n=1}^{\infty}\sum_{d=1}^{d_n}\Big((1/b_n)\sum_{i=1}^{d_n}(a_{di}^{(n)})^2\Big)^{r/2}\\ &\le \sum_{n=1}^{\infty}\sum_{d=1}^{d_n}(1/b_n)^{r/2}\\ &\le \sum_{n=1}^{\infty} d_n(1/b_n)^{r/2}<\infty.\end{aligned}$$

Therefore, (3.3.6) and hence (3.3.2) goes to 0 almost everywhere as a consequence of Lemma 3.4.1. This completes the proof.

COROLLORY 3.3.1. *Assume the mesh of D_n is $1/n$. Then*

i) *If $E|X_{ij}^{(n)}|^r \le A$ for some A, for some $r\ge 2$, then $b_n=n^p$, $p>6/r$ can quarantee* (3.3.2).

ii) *If $X_{ij}^{(n)}-m(t_i^{(n)})$ is identically distributed and $Ee^{t|X_{ij}^{(n)}|}$ exists for all i, j, n and $t>0$, then $b_n=(\log(n))^2$ can quarantee* (3.3.2).

iii) *If $X_{ij}^{(n)}-m(t_i^{(n)})$ is normally distributed with mean* 0 *and variance σ^2, then b_n with $(\log(n))/b_n=o(1)$ as n goes to ∞ can guarantee* (3.3.2).

3.4. **Some technical lemmas.**

LEMMA 3.4.1. *Let K_n, d_n be two sequences of positive integers. Let $\{a_{dk}^{(n)}: 1\le d\le d_n,\ 1\le k\le K_n,\ n\ge 1\}$ be an array of real numbers with*

$$\sum_{n=1}^{\infty}\sum_{d=1}^{d_n}\Big(\sum_{k=1}^{K_n}(a_{dk}^{(n)})^2\Big)^{r/2}<\infty \tag{3.4.1}$$

for some $r \geq 2$. *Let* $\{Z_k^{(n)}: 1 \leq k \leq K_n,\ n \geq 1\}$ *be identically distributed random variables with mean* 0. *Assume for each* n, $Z_k^{(n)}$*'s*, $1 \leq k \leq K_n$, *are independent. Then*

$$E|Z_k^{(n)}|^r < \infty \tag{3.4.2}$$

implies

$$\sum_{n=1}^{\infty} \mathscr{P}\Big\{\max_{1\leq d\leq d_n} \Big|\sum_{k=1}^{K_n} a_{dk}^{(n)} Z_k^{(n)}\Big| > \varepsilon\Big\} < \infty \tag{3.4.3}$$

for any $\varepsilon > 0$, *for any such array* $a_{dk}^{(n)}$. *Conversely, we have* (3.4.3) *implies* $E|Z_k^{(n)}|^q < \infty$ *for any* $q < r$.

Proof. Assume (3.4.2). It follows from Marcinkiewicz-Zygmund inequality and Hölder inequality that for $r > 2$,

$$\begin{aligned} E\Big|\sum_{k=1}^{K_n} a_{dk}^{(n)} Z_k^{(n)}\Big|^r & \\ &\leq B_r E\Big|\sum_{k=1}^{K_n} (a_{dk}^{(n)} Z_k^{(n)})^2\Big|^{r/2} \\ &= B_r E\Big|\sum_{k=1}^{K_n} (a_{dk}^{(n)})^{(2r-4)/r}(a_{dk}^{(n)})^{4/r}(Z_k^{(n)})^2\Big|^{r/2} \\ &\leq B_r E\bigg(\Big(\sum_{k=1}^{K_n} (|a_{dk}^{(n)}|^{(2r-4)/r})^{r/(r-2)}\Big)^{(r-2)/r} \\ &\qquad \Big(\sum_{k=1}^{K_n} (|a_{dk}^{(n)}|^{4/r}|Z_k^{(n)}|^2)^{r/2}\Big)^{2/r}\bigg)^{r/2} \\ &= B_r E\Big(\sum_{k=1}^{K_n} |a_{dk}^{(n)}|^2\Big)^{(r/2)-1}\Big(\sum_{k=1}^{K_n} |a_{dk}^{(n)}|^2 |Z_k^{(n)}|^r\Big) \\ &= B_r E|Z_k^{(n)}|^r\Big(\sum_{k=1}^{K_n} |a_{dk}^{(n)}|^2\Big)^{r/2} \end{aligned} \tag{3.4.4}$$

for some constant B_r, depending only on r. For $r = 2$, (3.4.4) can be proved in a similar but easier way.

Thus, employing (3.4.4) and (3.4.1), we have

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathscr{P}\left\{\max_{1\le d\le d_n}\left|\sum_{k=1}^{K_n} a_{dk}^{(n)} Z_k^{(n)}\right| > \varepsilon\right\}
&\le \sum_{n=1}^{\infty}\sum_{d=1}^{d_n} \mathscr{P}\left\{\left|\sum_{k=1}^{K_n} a_{dk}^{(n)} Z_k^{(n)}\right| > \varepsilon\right\} \\
&\le \sum_{n=1}^{\infty}\sum_{d=1}^{d_n} (1/\varepsilon^r) E\left|\sum_{k=1}^{K_n} a_{dk}^{(n)} Z_k^{(n)}\right|^r \\
&\le \sum_{n=1}^{\infty}\sum_{d=1}^{d_n} (1/\varepsilon^r) B_r E|Z_k^{(n)}|^r \left(\sum_{k=1}^{K_n} |a_{dk}^{(n)}|^2\right)^{r/2} \\
&= (1/\varepsilon^r) B_r E|Z_k^{(n)}|^r \sum_{n=1}^{\infty}\sum_{d=1}^{d_n} \left(\sum_{k=1}^{K_n} |a_{dk}^{(n)}|^2\right)^{r/2} \\
&< \infty .
\end{aligned}$$

This proves (3.4.3).

Conversely, set $d_n = 1$, $a_{1k}^{(n)} = 0$ for $k \neq K_n$, $a_{1K_n}^{(n)} = n^{-(1/q)}$ for $0 < q < r$. Then (3.4.1) is valid for this choice and we know from (3.4.3) that

$$\sum_{n=1}^{\infty} \mathscr{P}\{|n^{-(1/q)} Z_{K_n}^{(n)}| > \varepsilon\} < \infty,$$

which implies

$$\sum_{n=1}^{\infty} \mathscr{P}\{|Z_1^{(n)}|^q > \varepsilon^q n\} < \infty .$$

Therefore, $E|Z_k^{(n)}|^q < \infty$ for any $q < r$.

COROLLARY 3.4.1. *If, in Lemma* 3.4.1, *we assume a stronger condition that all* $Z_k^{(n)}$*'s are independent, then the conclusion of Lemma* 3.4.1 *remains unchanged except replacing* (3.4.3) *of the converse part by*

$$\lim_{n\to\infty} \max_{1\le d\le d_n} \left|\sum_{k=1}^{K_n} a_{dk}^{(n)} Z_k^{(n)}\right| = 0 \quad a.e. \tag{3.4.5}$$

COROLLARY 3.4.2. *As in Lemma* 3.4.1, *if* $Z_k^{(n)}$*'s are not necessarily identically distributed,* (3.4.3) *is still valid so long as* $E|Z_k^{(n)}|^r < A$, *for some constant* A, *for all* n, k.

LEMMA 3.4.2. *Let* K_n, b_n, d_n *be three sequences of positive integers. Assume that for each* $\varepsilon > 0$, *there exists* $t > 0$, *such that*

$$\sum_{1}^{\infty} d_n e^{-t\varepsilon b_n} < \infty \tag{3.4.6}$$

Let $\{a_{dk}^{(n)} : 1 \le d \le d_n,\ 1 \le k \le K_n,\ n \ge 1\}$ *be an array of real numbers satisfying*

$$\varlimsup_{n\to\infty} \max_{1\le d\le d_n} b_n^2 \sum_{k=1}^{K_n} (a_{dk}^{(n)})^2 < \infty\,. \tag{3.4.7}$$

Let $\{Z_k^{(n)} : 1 \le k \le K_n,\ n\ge 1\}$ *be identically distributed random variables with mean* 0. *Assume for each* n, $Z_k^{(n)}$*'s,* $1 \le k \le K_n$, *are independent. Then the following two statements are equivalent*:

$$\text{i)}\quad E e^{t|Z_k^{(n)}|} < \infty \qquad \text{for all} \quad t > 0, \tag{3.4.8}$$

$$\text{ii)}\quad \sum_{n=1}^{\infty} \mathscr{P}\left\{\max_{1\le d\le d_n} \left|\sum_{k=1}^{K_n} a_{dk}^{(n)} Z_k^{(n)}\right| > \varepsilon\right\} < \infty \tag{3.4.9}$$

for any $\varepsilon > 0$, *for any such array* $a_{dk}^{(n)}$.

Proof. Assume (3.4.8). Without loss of generality, we assume $E(Z_1^{(n)})^2 = 1$. Let $y(s) = Ee^{sZ_1^{(n)}}$. Then y is infinitely differentiable on R with $y(0) = 1$, $y'(0) = 0$, $y''(0) = 1$. Let $s_0 > 0$ be such that $y(s) \le 1 + s^2$ for $|s| \le s_0$.

Given $\varepsilon > 0$, choose $t > 0$ such that (3.4.6) is satisfied. Let $V_d^{(n)} = \{k : 1 \le k \le K_n,\ |tb_n a_{dk}^{(n)}| > s_0\}$. Let $v_d^{(n)}$ be the number of elements in $V_d^{(n)}$. (3.4.7) implies that $v_d^{(n)} < v$ for some v, for all n, d. Keeping these in mind, we have

$$\begin{aligned}
&\mathscr{P}\left\{\sum_{k=1}^{K_n} a_{dk}^{(n)} Z_k^{(n)} > \varepsilon\right\} \\
&\qquad \le e^{-t\varepsilon b_n} E e^{tb_n \Sigma a_{dk}^{(n)} Z_k^{(n)}} \\
&\qquad \le e^{-t\varepsilon b_n} \prod_{k\in V_d^{(n)}} y(tb_n a_{dk}^{(n)}) \prod_{k=1}^{K_n} (1 + (tb_n a_{dk}^{(n)})^2) \\
&\qquad \le e^{-t\varepsilon b_n} B(t)\, e^{t^2 b_n^2 \Sigma (a_{dk}^{(n)})^2} \\
&\qquad \le e^{-t\varepsilon b_n} B(t)\, e^{t^2 C}
\end{aligned} \tag{3.4.10}$$

where both $B(t)$ and C are some constants.

With the help of (3.4.10), we get

$$\sum_{n=1}^{\infty} \mathscr{P}\left\{\max_{1\le d\le d_n}\left|\sum_{k=1}^{K_n} a_{dk}^{(n)} Z_k^{(n)}\right| > \varepsilon\right\}$$

$$\le \sum_{n=1}^{\infty}\sum_{d=1}^{d_n}\left(\mathscr{P}\left\{\sum_{k=1}^{K_n} a_{dk}^{(n)} Z_k^{(n)} > \varepsilon\right\}\right.$$

$$\left. + \mathscr{P}\left\{\sum_{k=1}^{K_n} a_{dk}^{(n)} Z_k^{(n)} < -\varepsilon\right\}\right)$$

$$\le \bar{B}(t)e^{t^2 C}\sum_{n=1}^{\infty} d_n e^{-t\varepsilon b_n}$$

for some constant $\bar{B}(t)$. The proof is therefore completed in view of (3.4.6).

Now let's prove the converse. Take $d_n = 1$, $a_{1k}^{(n)} = 0$ for $k \neq K_n$, $a_{1K_n}^{(n)} = 1/\log(n)$, $b_n = \log(n)$. It is trivial to see that both (3.4.6) and (3.4.7) are satisfied. Let $t > 0$ be given. Choose $\varepsilon > 0$ such that $t\varepsilon \le 1$. Then

$$\sum_{n=1}^{\infty} \mathscr{P}\left\{\max_{1\le d\le d_n}\left|\sum_{k=1}^{K_n} a_{dk}^{(n)} Z_k^{(n)}\right| > \varepsilon\right\}$$

$$\ge \sum_{n=1}^{\infty} \mathscr{P}\{|Z_1^{(n)}| > \varepsilon \log(n)\}$$

$$\ge \sum_{n=1}^{\infty} \mathscr{P}\{e^{t|Z_1^{(n)}|} > n\}\,.$$

Therefore ii) implies i) and the proof is completed.

COROLLARY 3.4.3. *If, in Lemma* 3.4.2, *we assume a stronger condition that all $Z_k^{(n)}$'s are independent, the conclusion of Lemma* 3.4.2 *remains unchanged except replacing* (3.4.9) *of the converse part by*

(3.4.11) $$\lim_{n\to\infty}\max_{1\le d\le d_n}\left|\sum_{k=1}^{K_n} a_{dk}^{(n)} Z_k^{(n)}\right| = 0 \quad a.e.$$

LEMMA 3.4.3. *Let K_n, b_n, d_n be three sequences of positive integers. Assume that for each $\varepsilon > 0$,*

(3.4.12) $$\sum_{n=1}^{\infty} d_n b_n^{-1/2} e^{-\varepsilon b_n} < \infty\,.$$

Let $\{a_{dk}^{(n)} : 1 \le d \le d_n,\ 1 \le k \le K_n,\ n \ge 1\}$ be an array of real numbers satisfying

$$\text{(3.4.13)} \qquad \overline{\lim_{n\to\infty}} \max_{1\le d\le d_n} b_n \sum_{k=1}^{K_n} (a_{dk}^{(n)})^2 < A^2$$

for some constant A. Let $\{Z_k^{(n)} : 1 \le k \le K_n, n \ge 1\}$ *be normally distributed random variables with mean* 0 *and variance* σ^2. *Assume for each n,* $Z_k^{(n)}$*'s,* $1 \le k \le K_n$, *are independent. Then for each* $\varepsilon > 0$,

$$\text{(3.4.14)} \qquad \sum_{n=1}^{\infty} \mathcal{P}\left\{\max_{1\le d\le d_n} \left|\sum_{k=1}^{K_n} a_{dk}^{(n)} Z_k^{(n)}\right| > \varepsilon\right\} < \infty .$$

Proof. Without loss of generality, assume $\sigma = 1$. It is obvious that

$$\text{(3.4.15)} \qquad \begin{aligned} &\mathcal{P}\left\{\sum_{k=1}^{K_n} a_{dk}^{(n)} Z_k^{(n)} > \varepsilon\right\} \\ &\le \mathcal{P}\{N(0, 1) > \varepsilon b_n^{1/2}/A\} \\ &\le (C/b_n^{1/2})\, e^{-1/2(\varepsilon^2 b_n/A^2)} \end{aligned}$$

for some constant C, for n large enough.

Therefore,

$$\begin{aligned} &\sum_{n=1}^{\infty} \mathcal{P}\left\{\max_{1\le d\le d_n} \left|\sum_{k=1}^{K_n} a_{dk}^{(n)} Z_k^{(n)}\right| > \varepsilon\right\} \\ &\le \sum_{n=1}^{\infty} \sum_{d=1}^{d_n} \mathcal{P}\left\{\left|\sum_{k=1}^{K_n} a_{dk}^{(n)} Z_k^{(n)}\right| > \varepsilon\right\} \\ &\le 2\sum_{n=1}^{\infty} \sum_{d=1}^{d_n} (C/b_n^{1/2})\, e^{-1/2(\varepsilon^2 b_n/A^2)} < \infty . \end{aligned}$$

This completes the proof.

4. Concluding remarks.

1. Thanks to the flexibility of Cauchy-Euler method, no condition whatsoever on the design of experiment is imposed for the ODE case in section 2. This becomes quite an advantage especially in the situation that the data is given over, say, time and we are not allowed to choose the observing points t_i freely.

2. It is clear that the convergence rate of our main theorem in §3.3 depends heavily on the matrix A_n. (see (3.3.3)) We conjecture, based on our computer work, that for each n, A_n has at

least one row for which the sum of the squares of its entries is bigger than a positive constant. This conjecture can be best understood probably in the context of finite difference method in elliptic partial differential equations. If this conjecture should be correct, then we would be able to argue, by the following Lemma 4.1, that our main theorem can not be improved and making replicates at each observing point would be necessary as well as desirable.

Notations n the following bear the same meaning as those in §3.

LEMMA 4.1. *Assume* $X_{ij}^{(n)} - m(t_i^{(n)})$ *has normal distribution with mean* 0 *and variance* 1. *Let* $b_n = C(\log(n))$ *for some constant* C *with* $0 < C \leq 2$. *Suppose*

$$\varliminf_{n\to\infty} \sum_{i=1}^{d_n} (a_{di}^{(n)})^2 \geq a^2$$

for some d *and* $a > 0$, *where* $a_{di}^{(n)}$*'s are entries of the matrix* A_n. *Then*

$$\limsup_{n\to\infty} (1/b_n) \sum_{i=1}^{d_n} \sum_{j=1}^{b_n} a_{di}^{(n)}(X_{ij}^{(n)} - m(t_i^{(n)})) \geq a \quad a.e.$$

Proof. Set $Z_{ij}^{(n)} = X_{ij}^{(n)} - m(t_i^{(n)})$. It follows from the normality of $Z_{ij}^{(n)}$ and the assumption of the lemma that

$$\begin{aligned}\sum_{n=1}^{\infty} \mathscr{P}\left\{\sum_{i=1}^{d_n} \sum_{j=1}^{b_n} a_{di}^{(n)} Z_{ij}^{(n)} \geq ab_n\right\} \\ \geq \sum_{n=1}^{\infty} (K/b_n^{1/2})\, e^{-b_n/2} \\ = \infty\end{aligned}$$

where K is a suitable constant.

Therefore, Borel-Cantelli lemma ensures the conclusion of this lemma and the proof is completed.

With this lemma, it can be seen easily that $\max_{t\in \bar{D}_n} |M_n(t) - m(t)|$ would not converge to 0 when n goes to ∞, if the previous conjecture should be correct.

ACKNOWLEDGEMENT. The authors would like to thank Professor Carlos Lima for some helpful conversations. Comments and suggestions from Professors J. Kiefer, T. L. Lai and H. Robbins are much appreciated. Professor Lai, in particular, gave a comment which leads to an improvement of §2.

REFERENCES

1. L. Bers, F. John and M. Schechter, *Partial Differential Equations*, Interscience Publishers, 1964.

2. Y. S. Chow and T. L. Lai, *Limiting Behavior of Weighted Sums of Independent Random Variables*, Ann. Prob. Vol. 1, No. 5, 810–824, 1973.

3. D. L. Hanson and G. Pledger, *Consistency in Concave Regression.* Ann. Statist. Vol. 4, No. 6, 1038–1050, 1976.

4. W. Huriwicz, *Lectures on Ordinary Differential Equations*, MIT Press, 1958.

DEPARTMENT OF MATHEMATICS, NATIONAL CENTRAL UNIVERSITY, CHUNG-LI, TAIWAN, R. O. C.