CENTRAL LIMIT THEOREM AND CONSISTENCY IN LINEAR REGRESSION

by

Friedhelm Eicker

University of North Carolina

This research was supported by the Office of Naval Research under Contract No. Nonr-855(06) for research in probability and statistics at Chapel Hill. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Institute of Statistics
Mimeograph Series No. 271
December, 1960

CENTRAL LIMIT THEOREM AND CONSISTENCY IN LINEAR REGRESSION[1]

by

Friedhelm Eicker

Summary.

Several asymptotic properties of the least squares estimators for the parameters in the linear regression model with non-identical, independent errors are derived with regard to infinitely increasing sample size.

The notion of convergence of a sequence of random variables b_ν on a set of F is introduced and applied. That is, the b_ν depend functionally on an arbitrarily chosen sequence of random variables ϵ_μ all of which are elements of the set F, and they converge for each such sequence. If F is any set of random variables with zero means and bounded variances which contains at least one normal variable then there exists a simple necessary and sufficient condition for the regression matrix X such that the estimators are consistent on F (theorem 1). Furthermore, for asymptotic normality of essentially these estimators on any subset $\tilde{F}$ in the set of all zero-mean random variables whose variances exist, necessary and sufficient conditions are given both for X and the set $\tilde{F}$ simultaneously to be fulfilled (theorem 2). This is the central limit theorem for the linear regression model as described above. For the case of unknown error variances a statistic is constructed and sufficient conditions for X and $\tilde{F}$ are given to assure its asymptotic normality on $\tilde{F}$ (theorem 3). One important aspect of the theory developed is its non-parametric character and the width of the range of admitted error distributions. Some examples and illustrations are given.

1. Notations and definitions.

The model used for linear regression is described by the equations

1. This research was supported by the Office of Naval Research under Contract No. Nonr-855(06) for research in probability and statistics at Chapel Hill. Reproduction in whole or in part is permitted for any purpose of the United States Government.

(1) $$y_t = x_{1t}\beta_1 + \ldots + x_{qt}\beta_q + \epsilon_t, \qquad t = 1,2,\ldots,$$

or in vector notation:

(2) $$\underline{y} = X\underline{\beta} + \underline{\epsilon},$$

where the vector of observed values is for a sample of size N $\underline{y} = (y_1, y_2, \ldots, y_N)'$.* The vectors of the known regression constants are $\underline{x}_i = (x_{1t}, \ldots, x_{iN})'$, their matrix is $X = (\underline{x}_1, \ldots, \underline{x}_q)$. Further let $\underline{\beta} = (\beta_1, \ldots, \beta_q)'$ and $\underline{\epsilon} = (\epsilon_1, \ldots, \epsilon_N)'$ where the ϵ's are uncorrelated and have zero mean. The ϵ's need not be distributed identically. The least squares (LS) estimators $\underline{b}$ of $\underline{\beta}$ which are considered here are determined by the normal equations

(3) $$X'\underline{y} = \underline{\ \ b}, \qquad \text{where } P = X'X.$$

By B we denote the q-dimensional space of all possible parameter vectors $\underline{\beta}$(usually the entire q-dimensional real space).

2. <u>A necessary and sufficient condition for consistent estimators</u>.

Let in this paragraph F be any set of random variables (r.v.) with zero means and uniformly bounded variances which contains at least one normal r.v..

The following definition proves to be very appropriate in the scope of this paper. It takes into account that very often the errors ϵ_t are not identically distributed and that the sequence of their distributions $G_1(z)$, $G_2(z)$,... is not known. The definition given here is a special

---

* Prime denotes the transpose of a vector or matrix.

case of the general one stated in 4.

<u>Definition 1</u>: The LS-estimator b is called consistent on BxF if it tends for $N \to \infty$ in probability to the estimated parameter for any particular value of $\beta \in B$ and for each sequence ϵ_t of errors $\epsilon_t \in F$. (B here may be understood as a 1-dimensional subspace of B as defined in 1). B x F is the product space of B and F.

The following simple theorem brings forth a basic property of the regression matrix X resp. of $P = X'X$:

<u>Theorem 1</u>: The LS-estimates $\underline{b}$ of $\underline{\beta}$ are consistent on BxF if and only if $\lambda_{min}(P) \to \infty$. $\lambda_{min}(P)$ is the minimum characteristic value of P.

<u>Remark</u>: The theorem holds also on the product space $B\times\hat{F}$ (in an obvious extension of definition 1).

The space $\hat{F}$ of sequences $\{\epsilon_t\}$ of errors with countably many dimensions may be taken to be more general then the above product space

$$F \times F \times F \times \ldots, \text{ namely as } \hat{F} = F_1 \times F_2 \times F_3 \times \ldots$$

The error ϵ_t shall lie in F_t, and F_t may be taken different for each $t = 1,2,3,\ldots$, Each of the sets F_t is submitted uniformly to the requirements stated above for F.

<u>Proof</u>: From (2) and (3) comes $X'\underline{\epsilon} = P(\underline{b}-\underline{\beta})$. $\underline{b}$ can be uniquely determined if and only if for almost all N, $|P| \neq 0$. Hence for $N \geq N_o$

$$P^{-1}X'\underline{\epsilon} = \underline{b} - \underline{\beta} \,. \tag{4}$$

(I) Sufficiency: Clearly $E(P^{-1}X'\underline{\epsilon}) = \underline{0}$ for all N. The variance of each component of the vector $P^{-1}X'\underline{\epsilon}$ tends to zero if and only if $E(\underline{\epsilon}'XP^{-2}X'\underline{\epsilon}) \to 0$. Because of $E\epsilon_t^2 <$ const· the left side is

$$O(\mathrm{tr}\ XP^{-2}X') = O(\mathrm{tr}\ P^{-1}) = O(1/\lambda_{min}(P)) \to 0.$$

Hence with the Tchebycheff inequality $\underline{b} \xrightarrow[\text{i.p.}]{} \underline{\beta}$ uniformly on B x F.

(II) Necessity: If $\underline{b} \xrightarrow[\text{i.p.}]{} \underline{\beta}$ on B x F then this holds especially for $\epsilon_t \sim N(0,\sigma^2)$, $t = 1,2,\ldots,\sigma > 0$. If $\underline{p}_i'$ denotes the i-th row in P^{-1} then

$$b_i - \beta_i = \underline{p}_i' X' \underline{\epsilon} \sim N(0, \underline{p}_i' X' \Sigma X \, \underline{p}_i),$$

where Σ is a diagonal matrix with $\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2$ in the diagonal. Hence with two positive constants

$$\text{const } (P^{-1})_{ii} < \text{var } (b_i - \beta_i) < \text{const } (P^{-1})_{ii},*$$

$(P^{-1})_{ii}$ being the i-th diagonal element of P^{-1}. This term as well as their sum for $i = 1,2,\ldots,q$ which equals $\sum_i 1/\lambda_i(P)$, must tend to zero because of the consistency of $\underline{b}$. Hence

$$\lambda_{min}(P) \longrightarrow \infty ; \text{ q.e.d.}$$

3. Some applications and remarks.

(I) As $P_{ii} \geq \lambda_{min}(P)$, it follows that $\underline{x}_j^2 \longrightarrow \infty$ for all $j = 1,\ldots,q$, for $N \rightarrow \infty$. $\lambda_{min}(P)$ is a non-negative, non-decreasing sequence in N.

(II) Applying Gershgorin's method we have $\lambda_{min}(P) \longrightarrow \infty$ if for all i

$$\underline{x}_i^2 - \sum_{\substack{j=1 \\ j \neq i}}^{q} |\underline{x}_i' \ \underline{x}_j| \longrightarrow \infty .$$

(III) The definition of the minimum characteristic value for symmetric matrices

$$\lambda_{min}(P) = \min_{\underline{u}} (X\underline{u})^2 \ , \ \underline{u}^2 = 1,$$

*const means any constant, independent of t.

is used in the following example:

If $x_{it} = t^{c_i}$, $c_i > c_{i+1}$, (<u>polynomial</u> <u>regression</u>) then because of $\sum_{t=1}^{N} t^c = O(N^{c+1})$, the order of $X\underline{u}$ is determined by the first non-zero component of $\underline{u}$. The slowest growth of $X\underline{u}$ is therefore of order

$$\min_{\underline{u}} (X\underline{u})^2 = O(N^{2c_q+1}) , \underline{u}^2 = 1.$$

Hence consistent estimates are obtained, also for non-integers c_i, if and only if $\min_i c_i = c_q > -\frac{1}{2}$. Similarly, regression vectors may be treated in which exponentials, or exponentials plus polynominals occur.

(IV) One obtains consistent estimators also in <u>trigonometric</u> <u>regression</u>. Let

$$x_{2i-1,\ t} = \cos \omega_i t \ , \ x_{2i,\ t} = \sin \omega_i t \ , \quad i = 1, 2, \ldots, q,$$

if $\omega_i \neq \omega_j$ and $\omega_i \neq 2\pi - \omega_j$ for $i \neq j$, further $\omega_i < 2\pi$ for all i. P tends in this case to a diagonal matrix whose non-zero elements are O(N).

(V) In the theorem it is required that F contains a normally distributed r.v.. This is certainly not necessary. We do not, however, pursue the question whether simultaneously necessary conditions for X and F can be found for consistency on F as is done in section 4 for convergence to the normal law. One would have to consider the convergence of b_i to a unit step function (compare [5], p. 116). On the other hand, a generalization of the theorem for certain classes of dependent errors seems to be easily possible.

(VI) As the above proof shows, $(P^{-1})_{ii} \longrightarrow \infty$ is a necessary and sufficient condition for $b_i \underset{i.p.}{\longrightarrow} \beta_i$ on BxF.

(VII) It is also possible to give sufficient conditions for consistent LS-estimates in a linear model of time series analysis where regression with regard to lagged variables is included, according to the equation: [2]

$$y_t = \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + \beta_1 x_{1t} + \dots + \beta_q x_{qt} + \epsilon_t .$$

The ϵ's are uncorrelated, but need not be identical. The x_{it} are the exogenous regression constants. The theorem given in 2. is a certain specialization of the result for this general equation.

4. Conditions for asymptotically normal LS-estimators.

To obtain conditions for the errors ϵ_k which are henceforth assumed to be independent, and for the regression vectors in X, assuming the asymptotic normality of the LS-estimates b of β as defined in 1. we use the central limit theorem in either of the two following versions [e.g. 4, 5]:

(A) Lindeberg-Feller version:

In order that the distribution functions of the sums

(5) $$\zeta_N = \xi_{N1} + \xi_{N2} + \dots + \xi_{NN}$$

of independent random variables $\xi_{N1}, \xi_{N2}, \dots, \xi_{NN}$ with

(6) $$E\, \xi_{Nk} = 0 \quad \text{for } k = 1,2,\dots,N, \quad \text{and}$$

(7) $$\sum_{k=1}^{N} \operatorname{var} \xi_{Nk} = 1$$

converge to the normal distribution N(0,1):

$$\Phi(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz$$

it is necessary and sufficient that Lindeberg's condition

$$(8) \qquad \sum_{k=1}^{N} \int_{|z|>\delta} z^2 d\, F_{Nk}(z) \longrightarrow 0 \qquad \text{for } N \longrightarrow \infty$$

be satisfied for every $\delta > 0$. $F_{Nk}(z)$ is the distribution function of ξ_{Nk}.

The second version makes a stronger assumption concerning the moments of the ξ_{Nk}:

(B) Sufficient for the convergence of the ζ_N as given by (5), (6) and (7), to the normal distribution is the condition of Lyapunov that

$$(9) \qquad \sum_{k=1}^{N} \int |z|^{2+\eta} d\, F_{Nk}(z) \longrightarrow 0 \qquad \text{for } N \longrightarrow \infty,$$

η being some positive number.

We apply these theorems to our problem by putting for any fixed i out of 1,2,...,q

$$\zeta_N = B_N^{-1} \underline{p}_i' X' \underline{\epsilon} = B_N^{-1}(b_i - \beta_i)$$

where $\underline{p}_i$ is the i-th row or column in P^{-1}, and

$$(10) \qquad B_N^2 = \text{var } \underline{p}_i' X' \underline{\epsilon}.$$

Putting

$$(11) \qquad \underline{u} = X \underline{p}_i$$

and $\xi_{Nk} = B_N^{-1} u_k \epsilon_k$ we have from $E\, \epsilon_k = 0$, that condition (6) is satisfied, and by definition of B_N and the independence of the ϵ's also condition (7). (To simplify the notation and where confusion is unlikely the subscript N is omitted on various symbols).

If $G_k(\epsilon)$ is the distribution of ϵ_k, we then have in (8)

$$(12) \quad \int_{|z|>\delta} z^2 dF_{Nk}(z) = B_N^{-2} u_k^2 \int_{|\epsilon|>\delta B_N |u_k^{-1}|} \epsilon^2 \, d\, G_k(\epsilon) \quad .$$

We now first proceed in a somewhat heuristic manner to find out what kind of conditions have to be placed upon the regression matrix X and the errors. Preliminary results are given in two corollaries. They prepare the questioning answered by the main result (theorem 2). With (12) we get the inequality for (8)

$$(13) \qquad \sum_{k=1}^{N} \int_{|z|>\delta} z^2 dF_{Nk}(z) \leq B_N^{-2} \left(\sum_{j=1}^{N} u_j^2 \right) \max_{k=1,\dots,N} \left[\int_{|\epsilon|>\delta B_N |u_k^{-1}|} \epsilon^2 \, dG_k(\epsilon) \right] .$$

If $\sigma_k^2 = E\, \epsilon_k^2 \geq m > 0$ for all k then for all N

$$B_N^{-2} \sum_{k=1}^{N} u_k^2 \leq m^{-1} .$$

By definition we have $\underline{u}^2 = \underline{p}_i' \, X' X \, \underline{p}_i = (P^{-1})_{ii} \geq \lambda_1^{-1}(P)$ where $(P^{-1})_{ii}$ is the i-th diagonal element of P^{-1} and λ_k means the k-th largest characteristic root. Furthermore, let $\underline{r}'_j$ be the j-th row in X. Then we have in the expression for the integration interval

$$(14) \qquad B_N^2 \, u_k^{-2} \geq \frac{\underline{u}^2 \quad m}{\underline{r}_k^2 \quad \underline{p}_i^2} \geq \frac{m\, \lambda_q^2(P)}{\underline{r}_k^2 \, \lambda_1(P)} \quad .$$

We want these expressions for all k to tend to $+\infty$ in order to satisfy (13). Hence we may require

$$(15) \qquad \max_{k=1,\dots,N} \underline{r}_k^2 \, \frac{\lambda_1(P)}{\lambda_q^2(P)} \longrightarrow 0 \qquad \text{as } N \longrightarrow \infty .$$

Let us furthermore assume about the error distribution that there exists for $c \geq 0$ a bounded function $g(c)$ with $\lim_{c \to \infty} g(c) = 0$ such that for each error ϵ_k with distribution function G_k holds

$$(16) \qquad \int_{|z| \geq c} z^2 \, dG_k(z) < g(c).$$

We observe that this requirement is stronger than that of uniform boundedness from above of all σ_k. From (15) and (16) we see that (13) tends to zero, thus satisfying the Lindeberg condition. So we have

Corollary: In order that $B_N^{-1}(b_i - \beta_i)$, $i = 1,\ldots,q$, tends to the normal distribution $N(0,1)$ where B_N^2 is given by (10), it is sufficient that

(a1) $\qquad E\,\epsilon_k = 0, \qquad 0 < m \leq E\,\epsilon_k^2$ for all k,

(a2) $\qquad$ there exists a function $g(c)$ with $\lim_{c \to \infty} g(c) = 0$ such that (16) holds for all errors.

(b) $$\frac{\lambda_1(P)}{\lambda_q^2(P)} \max_{k=1,\ldots,N} \underline{r}_k^2 \longrightarrow 0 \qquad \text{as } N \longrightarrow \infty,$$

all quantities as defined above.

Remark: Condition (b) necessarily requires $\lambda_1(P)\lambda_q^{-2}(P) \longrightarrow 0$ and $\lambda_q(P) \longrightarrow \infty$. The first statement follows from the fact that $\max_{k=1,\ldots,N} \underline{r}_k^2 > \text{const} > 0$, the second from the further fact that $\lambda_1(P_N)$ is positive and a non-decreasing function of N (here we have explicitly written P_N instead of P to indicate that X is $(N\text{x}q)$). We have namely with unimodular characteristic vectors $\underline{v}_k$:

$$\lambda_1(P_N) = \sum_{j=1}^{N} (\underline{r}_j' \, \underline{v}_N)^2 \leq \sum_{j=1}^{N+1} (\underline{r}_j' \, \underline{v}_N)^2 \leq \sum_{j=1}^{N+1} (\underline{r}_j' \, \underline{v}_{N+1})^2 = \lambda_1(P_{N+1}).$$

Similarly, $\lambda_q(P_N)$ can be shown to be a non-decreasing function of N for any choice of the $\underline{r}_k$'s.

In going through the proof of this corollary one observes that it holds also given any particular i, for a weaker assumption than (b) which moreover can be shown to be even necessary in assuming asymptotic normality of $B_N^{-1}(b_i - \beta_i)$ on the underlying set F of error random variables defined by conditions (a1) and (a2). To be precise we write down the

Definition 2: $B_N^{-1}(b_i-\beta_i)$ is said to be asymptotically normally distributed on the space F if its distribution converges to the standard normal distribution N(0,1) for each choice of an error sequence $\{\epsilon_k\}$ whose elements ϵ_k lie in F.

This definition can be generalized to

Definition 3: Let $\{b_N\}$, N = 1,2,..., be any sequence of random variables in which each b_N is a function of other random variables ϵ_k, k=1,2,.... All the ϵ_k's are elements of a set F. Then the b_N are said to converge (in some particular sense) on the set F (shortly: on F) if this convergence holds for each possible choice of the sequence ϵ_1, ϵ_2,...

A further generalization can be made as follows:

Definition 3': Let b_N be random variables for N = 1,2,... each of which depends in a certain functional manner on at most countably many other random variables ϵ_1, ϵ_2,.... Let ϵ_i be an element of the set F_i for each i = 1,2,.... All of these sets may be different. Then the b_N are said to converge (in some particular sense) on the product space $\Pi = F_1 \times F_2 \times F_3 \times \ldots$ if this convergence holds for each possible choice of a sequence $\{\epsilon_i\} \in \Pi$, i.e. for each selection $\epsilon_i \in F_i$, i=1,2,...

Corollary: Convergence on F for the r.v.s b_N of the above structure is identical with their convergence on the product space $\Pi = F \times F \times F \times \ldots$

Clearly definitions 1 and 2 are special cases (with regard to the b_N as well as to the kind of convergence) of definition 3.

If the b_N still depend on a set of parameters which lie in a set B, say, as is the case in our application then we will consider convergence on the set BxF resp. $B \times \Pi$ thereby extending the above definitions in a trivial manner.

Corollary: A necessary and sufficient condition that $B_N^{-1}(b_i - \beta_i)$ be asymptotically normally distributed on the set F is (i=1,2,...,q)

$$\text{(22)} \qquad \frac{1}{(P^{-1})_{ii}} \max_{k=1,\ldots,N} (\underline{r}_k' \, \underline{p}_i)^2 \longrightarrow 0 \qquad \text{as } N \longrightarrow \infty .$$

Proof: Sufficiency: from (22) follows that the left side in

$$\frac{u_k^2}{B_N^2} = \frac{(\underline{r}_k' \, \underline{p}_i)^2}{B_N^2} \leq \frac{1}{m(P^{-1})_{ii}} \max_k (\underline{r}_k' \, \underline{p}_i)^2$$

tends to zero as $N \longrightarrow \infty$. Hence, using (16), (13) tends to zero for any $\delta > 0$ wherewith Lindeberg's condition is satisfied.

Necessity: We take a sequence $\{\epsilon_k\}$ of identically distributed errors with variance M and distribution function $G(\epsilon)$. Assume there is a subsequence of (22) whose terms remain $\geq Mc^2 > 0$, say. The corresponding sequence of N's may be denoted by $\{N_j\}$, say. Then (8) becomes for $N=N_j$:

$$\sum_{k=1}^{N_j} \int_{|z|>\delta} z^2 \, dF_{Nk}(z) \geq c^2 \int_{|\epsilon|>\delta/c} \epsilon^2 \, dG(\epsilon) > 0$$

for $\delta > 0$ small enough. As $j \longrightarrow \infty$, the left side should tend to 0 which is impossible. By this contradiction the corollary is proved.

One may ask whether the assumptions about the set F can be relaxed so that still asymptotic normality on the set F holds, especially, whether there are necessary conditions simultaneously for X and F to be fulfilled.

We assume now without loss of generality that in any series of error distributions all their variances are positive. If namely for some k the error ϵ_k in equation (1) would have zero variance, we would not have a stochastic equation. This then gives a linear relation between the coefficients β_i unless $\underline{r}_k = \underline{0}$, and the problem can be reduced to one which contains one zero variance error less (and also one parameter β_i less) than the original problem. As a matter of fact the number of independent linear relations of this kind cannot exceed q, in which case either all the parameters β_i are uniquely determined or the problem is unsolvable.

We notice that this property of F of positive variances for all its r.v.-s, has in the case of convergence already the consequence that for any (fixed) k

$$(30) \qquad q_{kN} \equiv |u_k/B_N| \longrightarrow 0 \qquad \text{as } N \longrightarrow \infty .$$

Otherwise there would exist an infinite sequence of positive integers $\{N_j\}$ such that $q_{kN_j} \geq c > 0$, and (8) had the lower bound $c^2 \int_{|\epsilon| > \delta/c} \epsilon^2 \, d\, G_k(\epsilon)$ which remains positive for all j and a suitably small δ, and because of $\sigma_k^2 > 0$. But this a contradiction.

On the other hand, the assumption about F that all its r.v.-s have finite variances is required because otherwise there may not be a finite variance for the limiting normal distribution, or also there may not be any convergence in distribution at all.

The question about necessary and sufficient properties of a subset of F is answered by the following basic

Theorem 2: Let F be the space of all random variables whose means are zero and whose variances are positive and finite. Let $\tilde{F}$ be any non-empty subser $\subsetneq$ F. Then it is necessary and sufficient for the convergence of $B_N^{-1}(b_i-\beta_i)$ on $\tilde{F}$ to the standard normal law N(0,1) for $N \longrightarrow \infty$ that the following conditions for X and $\tilde{F}$ be simultaneously satisfied:

$$\text{(I)} \qquad \frac{\max\limits_{k=1,\ldots,N} (\underline{r}_k' \, \underline{p}_i)^2}{(P^{-1})_{ii}} \longrightarrow 0 \qquad \text{as } N \longrightarrow \infty$$

(II) There exists a bounded function g(c) for $c \geq 0$ with $\lim\limits_{c \to \infty} g(c) = 0$ such that for each r.v. $\epsilon \in \tilde{F}$ with distribution function $G(\epsilon)$, say,

$$\int_{|\epsilon| \geq c} \epsilon^2 \, dG(\epsilon) < g(c)$$

holds.

(III) There exists a positive constant m such that for each $\epsilon \in \tilde{F}$

$$\operatorname{var} \epsilon > m$$

holds.

As the theorem shows, uniform boundedness of all variances from above is not sufficient.

Proof: 1) Sufficiency: Let $\{\epsilon_k\}$ be any sequence of r. v.-s out of $\tilde{F}$, satisfying (II) and (III). Let var $\epsilon_k = \sigma_k$, $B_N^2 = \sum\limits_{k=1}^{N} u_{kN}^2 \sigma_k^2$,

$u_{kN} = \underline{r}_k' \, \underline{p}_i$ (earlier denoted just by u_k). Then for any $k \leq N$

$$q_{kN}^2 = \frac{u_{kN}^2}{B_N^2} \leq \frac{1}{m(P^{-1})_{ii}} \max_{k=1,\ldots,N} u_{kN}^2 \longrightarrow 0 \text{ as } N \longrightarrow \infty \text{ by (I). Here}$$

$(P^{-1})_{ii} = \sum\limits_{k=1}^{N} u_{kN}^2$ is the i-th diagonal element of $(X'X)^{-1}$, X = (N x q).

We now see that the necessary and sufficient condition ($\mathcal{L}$) of the Lindeberg-Feller central limit theorem is satisfied because for any $\delta > 0$

$$(35) \qquad \sum_{k=1}^{N} \int_{|z|>\delta} z^2 \, dF_{Nk}(z) \leq \frac{(P^{-1})_{ii}}{B_N^2} \max_{k=1,\ldots,N} \int_{|\epsilon|>\delta/q_{kN}} \epsilon^2 \, dG_k(\epsilon)$$

which tends to zero. We have namely $B_N^{-2}(P^{-1})_{ii} < m^{-1}$ and by (II), with G_k the distribution function of ϵ_k, for any $k \leq N$

$$\int_{|\epsilon|>\delta/q_{kN}} \epsilon^2 \, dG_k(\epsilon) < g(\delta/q_{kN}) \longrightarrow 0$$

because $\max\limits_{k=1,\ldots,N} q_{kN} \longrightarrow 0$.

2) The <u>necessity</u> of both of the conditions (II) and (III) is proved in several steps. First we show the necessity of (I) by choosing all errors to be identically distributed, $\epsilon_k \sim G$, G arbitrarily chosen in any $\tilde{F} \subsetneq F$. By assumption on F, $\sigma = \text{var}\, \epsilon_k > 0$. For (8) we have

$$(37) \qquad \sum_{k=1}^{N} q_{kN}^2 \int_{|\epsilon|>\delta/q_{kN}} \epsilon^2 \, dG(\epsilon) \geq \frac{1}{\sigma^2} U_N^2 \int_{|\epsilon|>\delta\, \sigma/|U_N|} \epsilon^2 \, dG(\epsilon) \quad ,$$

where

$$(38) \qquad U_N^2 = \max_{k=1,\ldots,N} u_{kN}^2/(P^{-1})_{ii} \quad .$$

As the left side of the inequality tends for any $\delta > 0$ to zero so must U_N^2. (Paragraphs 1) and 2) are evidently the proof of the last corollary).

3) Given any $\tilde{F} \subseteq F$. Let the distribution function of $\epsilon \in \tilde{F}$ be $G_\epsilon(z)$. Let $f_\epsilon(c) = \int_{|z|>c} z^2 \, dG_\epsilon(z)$ and consider for $c \geq 0$ the non-increasing

function

$$M(c) = \limsup_{\epsilon \in \tilde{F}} f_\epsilon(c).$$

$\tilde{F}$ has property (II) if and only if $\lim_{c \to \infty} M(c) = 0$. We assume henceforth that $\tilde{F}$ does not have property (II). We will then construct a sequence of errors which are elements of $\tilde{F}$ such that (8) does not tend to zero, thus coming to a contradiction. Obviously, $\tilde{F}$ must have infinitely many elements, and for any sequence $c_\nu \to \infty$ we can find a sequence of r.v-s $\epsilon_\nu \in \tilde{F}$ and functions $f_\nu(c)$ such that for all $\nu=1,2,\ldots$

$$f_\nu(c_\nu) = \int_{|z|> c_\nu} z^2 \, dG_\nu(z) > C > 0$$

where $\lim_{c \to \infty} M(c) > C > 0$.

We consider first the case where there exists an infinite subset among the ϵ_ν's in which all r.v.-s have bounded variances $< M$, say. To avoid double subscripts we will assume that we had selected the ϵ_ν right from the beginning in this way; hence for $\nu=1,2,\ldots$

$$C < \sigma_\nu^2 = \text{var}\ \epsilon_\nu < M. \tag{41}$$

By selection of a suitable sequence $\{\epsilon_{\nu_k}\}$ for the errors in (1) from these ϵ_ν we show that at least for a subsequence N_ρ, $\rho=1,2,\ldots$ of the set of positive integers (8) is not true. Because of (41) we have

$$\frac{u_{kN}^2}{M(P^{-1})_{ii}} < q_{kN}^2 < \frac{u_{kN}^2}{C(P^{-1})_{ii}}$$

which tends necessarily to zero because of 2). Let η be a positive constant $< C^{-1}$. Let for each N k_N be the largest integer such that

$$\frac{1}{C(P^{-1})_{ii}} \sum_{k=1}^{k_N} u_{kN}^2 < \eta,$$

yet such that k_N is non-decreasing. Evidently $k_N \longrightarrow \infty$.

Now determine an infinite sequence $N_1 N_2, \ldots \longrightarrow \infty$ such that for $\rho=1,2,\ldots$

$$N_\rho \leq k_{N_{\rho+1}} \quad .$$

Let now ϵ_{ν_k} be the same r.v. for $k=k_{N_\rho}+1, \ldots, N_\rho$, say ϵ_{N_ρ}, with

$$c_{\nu_k} = c_{N_\rho} > \frac{M(P^{-1})_{ii}}{\min_j |u_{jN_\rho}|} \quad .$$

Here j runs from $k_{N_\rho}+1$ to N_ρ except those j for which $u_{jN_\rho} = 0$. Because of $c_\nu \longrightarrow \infty$ a $c_{\nu_k} < \infty$ can always be found.

For each N_ρ can any $\delta > 0$, $\delta < 1$ we now get for (8) a contradiction

$$\sum_{k=1}^{N_\rho} q_{kN_\rho}^2 \int_{|z|> \delta/q_{kN}} z^2 \, dG_k(z) \geq \left(\sum_{k=k_{N_\rho}+1}^{N_\rho} q_{kN_\rho}^2 \right) \int_{|z|> c_{N_\rho}} z^2 \, dG_{N_\rho}(z)$$

$$> \frac{C}{M}\left(1 - \sum_{k=1}^{k_{N_\rho}} \frac{u_{kN_\rho}^2}{(P^{-1})_{ii}}\right) > \frac{C}{M}(1 - C_\eta) > \text{const} > 0.$$

4) We consider now the case that the variances of the r.v.-s in $\tilde{F}$ are not bounded and select a sequence $\{\epsilon_\nu\}$ such that

$$\sigma_\nu^2 \longrightarrow \infty \quad .$$

One can determine these σ_ν^2 in such a way, for the variances of the errors in our model (1), that (8) is bounded away from zero for all N for which $u_{NN} \neq 0$. We write (8) as

$$\sum_{k=1}^{N-1} q_{kN}^2 \int\limits_{|z|>\delta/q_{kN}} z^2\, dG_k(z) + q_{NN}^2 \int\limits_{|z|>\delta/q_{NN}} z^2\, dG_N(z)$$

and, assuming $\delta^2 < \frac{1}{2}$, we determine the last integration interval so that

$$\frac{\delta^2}{u_{NN}^2} B_N^2 < \frac{1}{2} \left(\sum_{k=1}^{N-1} \frac{u_{kN}^2}{u_{NN}^2} \sigma_k^2 + \sigma_N^2 \right) < \kappa^2 \sigma_N^2 ,$$

which is always possible for $u_{NN} \neq 0$ and a constant $\kappa < 1$ and already fixed σ_k's, $k=1,\ldots,N-1$ by making σ_N sufficiently large. Then, taking only the N-th term, (8) is larger than

$$\frac{u_{NN}^2}{B_N^2} \int\limits_{|z|>\kappa\sigma_N} z^2\, dG_N(z) > (1-\kappa^2) \frac{u_{NN}^2 \sigma_N^2}{B_N^2} > \frac{1-\kappa^2}{2\kappa^2} > 0.$$

With this the proof of the necessity of (II) for any $\tilde{F}$ is completed.

5) It remains to prove the necessity of (III). We now assume that an arbitrary subset $\tilde{F} \subseteq F$ does not have property (III) and construct a contradiction by showing that (8) remains above a positive constant for a sequence of positive numbers $N_\rho \longrightarrow \infty$. Any such set $\tilde{F}$ contains necessarily infinitely many r.v.-s and especially sequences $\{\epsilon_\nu\}$ with $\liminf\limits_{\nu\to\infty} \sigma_\nu = 0$.

We first give the proof for a special countable set $\tilde{F}$ of this kind, say $\tilde{S}$, with r.v.-s whose distributions are 2-step functions $S_\nu(z)$ each having steps at + and - σ_ν with weight 1/2 each. Let there be given any non-singular regression matrix $X_{N \times q}$ for $N \longrightarrow \infty$; hence the numbers u_{kN}, $k=1,\ldots,N$ are given. The k-th term in (8):

$$\sum_{k=1}^{N} q_{kN}^2 \int\limits_{|z|>\delta/q_{kN}} z^2\, dS_k(z)$$

gives the contribution

$$\frac{\sigma_k^2 u_{kN}^2}{B_N^2} > \delta^2 \tag{50}$$

if only we have for the integration interval $\sigma_k > \delta/q_{kN}$, which is the same inequality as (50). If we can construct an error sequence $\{\epsilon_\nu\}$ with $\epsilon_\nu \in \tilde{S}$ such that (50) holds for some $\delta_o > 0$ and for infinitely many N_ρ and for at least one k at each N_ρ then we have a contradiction. Clearly, if (50) folds for some $\delta_o > 0$ it holds also for any positive $\delta < \delta_o$. We write (50) in the form, putting $\delta_o^2 = \eta$:

$$(1 - \eta)\sigma_k^2 u_{kN}^2 - \eta \sum_{\substack{j=1 \\ j\neq k}}^{N} \sigma_j^2 u_{jN}^2 > 0. \tag{51}$$

Subcase 5 a): Suppose there are (finite) constants c and k_o such that for an infinite sequence $N_\rho \longrightarrow \infty$ and for at least one k at each N_ρ, say $k_\rho \leq k_o$, holds:

$$u_{k_\rho N_\rho}^2 > c\, U_{N_\rho}^2 \quad , \text{ where } U_N^2 = \max_{k=1,\dots,N} u_{kN}^2 . \tag{52}$$

Then (51) is for $N = N_\rho$, $k = k_\rho$, any ρ, greater than

$$((1 - \eta)\, \sigma_{k_\rho}^2 c - \eta \sum_{j=1}^{\infty} \sigma_j^2)\, U_{N_\rho}^2 > 0$$

if we choose $\sum_{j=1}^{\infty} \sigma_j^2 < \infty$ and $\sigma_1 = \sigma_2 = \dots = \sigma_{k_o}$ = constant; also let η not be greater than a suitably small positive constant. Hence under our assumption about the u_{kN}, condition (III) is necessary for the set $\tilde{S}$.

Subcase 5b): Suppose now that for any pair of constants d and k_o (52) holds only for finitely many N_ρ. With other words, we can find for any k a monotonically decreasing sequence of numbers with $\lim_{N\to\infty} d_N(k) = 0$ such that

(54) $$u_{jN}^2 < k^{-1} d_N(k) U_N^2 \qquad \text{for } j=1,2,\dots,k\ ;\ N=1,2,\dots \ .$$

Furthermore, there exists even a sequence $m_N \longrightarrow \infty$ such that for any given constant D which we assume to be < 1 and for all $N \geq N_o$, say,

$$\sum_{k=1}^{m_N} u_{kN}^2 < DU_N^2 \ .$$

This follows immediately from (54), and we choose for instance

$$m_N \leq m'_N = \max_{k=1,\dots,N} \{k \mid d_N(k) < D\}.$$

Clearly m'_N is non-decreasing because the $d_N(k)$ are decreasing sequences of N, and $\lim_{N\to\infty} m'_N = \infty$ because the $d_N(k)$ tend to zero for all k. Consider now a pair of sequences of integers $m_{N_\rho} \longrightarrow \infty$, $N_\rho \longrightarrow \infty$ for $\rho \longrightarrow \infty$ such that $N_\rho = m_{N_{\rho+1}}$.

Let K_N be defined by $u_{K_N N}^2 = \max_{k=1,\dots,N} u_{kN}^2 = U_N^2$. Then because of $D < 1$ certainly $m_N < K_N$. Assuming now that $\sigma_1, \sigma_2, \dots, \sigma_{m_N}$ for N equal to some N_ρ be any given constants.

We determine inductively $\sigma_{m_N+1}, \sigma_{m_N+2}, \dots, \sigma_N$ such that we have for (51), taken at $N=N_\rho$ and $k=K_{N_\rho}$, the lower bound

$$U_N^2 \ (1-\eta)\ \sigma_{k_N}^2 - \eta\ D \max_{j=1,\dots,m_N} \sigma_j^2 - \eta \sum_{\substack{j=m_N+1 \\ j\neq k_N}}^{N} \sigma_j^2 > 0 \ .$$

This can be done by choosing (i) $\sigma_{k_N} = \text{const} > 0$ for $N=N_1,N_2,\ldots$ (ii) $\sum_{j=m_N+1}^{N} \sigma_j^2 < \text{const} < \infty$ again for $N= N_1,N_2,\ldots$, (iii) $\eta \leq \eta_o$ sufficiently small. Hence also in subcase 5b) condition (III) proves to be necessary for the set $\tilde{S}$.

6) We now turn back to an arbitrary subset $\tilde{F}$ not having property (III). There are sequences of r.v.-s with distributions $G_\nu(z)$ in $\tilde{F}$ such that for their variances $\liminf_{\nu\to\infty} \sigma_\nu = 0$ holds. Consider for each $G_\nu(z)$ a step function $S_\nu(z)$ as defined in 5) which has the same variance σ_ν. We then have for any ν, any $\delta > 0$ and any number q_{kN}

$$\int_{|z| \geq \delta/q_{kN}} z^2 \, dG_\nu(z) \geq \frac{1}{2} \int_{|z| \geq \delta'/q_{kN}} z^2 \, dS_\nu(z)$$

with $\delta' = \sqrt{2}\delta$. It is sufficient to prove this for $\frac{\delta'}{q_{kN}} = \sigma_\nu - 0$. The left side is not smaller than $\sigma_\nu^2(1 - \frac{1}{2})$ which is also the value of the right side. It now follows from 5) that for any set $\{u_{kN}\}$ we can find a sequence of errors with d.f.-s G_k such that for $N=N_\rho$, $\rho\to\infty$

$$\sum_{k=1}^{N_\rho} q_{kN_\rho}^2 \int_{|z|>\delta/q_{kN_\rho}} z^2 \, dG_k(z) \geq \frac{1}{2} \sum_{k=1}^{N_\rho} q_{kN_\rho}^2 \int_{|z|>\sqrt{2}\,\delta/q_{kN_\rho}} z^2 \, dS_k(z) > 2\,\delta^2 > 0.$$

This completes the proof of theorem 2.

As the Lyapunov form of the central limit theorem gives only a sufficient condition for the convergence we cannot expect weaker conditions on X and $\tilde{F}$ than those given in theorem 2. One can easily show e.g. the sufficiency of the conditions on $\tilde{F}$ as stated in the

Corollary to theorem 2: For the convergence in law of $B_N^{-1}(b_i - \beta_i)$ on any $\tilde{F} \subseteq F$ it is sufficient that (I) and (III) holds besides:

(II)' $\qquad E(|\epsilon|^{2+\eta}) <$ const for all $\epsilon \in \tilde{F}$.

Proof: We have in (9)

$$\int |z|^{2+\eta} dF_{Nk}(z) = |B_N^{-1} u_{kN}|^{2+\eta} \int |\epsilon|^{2+\eta} dG_k(\epsilon),$$

and hence there is by (II)' and (III) for the sum (9) the upper bound

$$\text{const } B_N^{-2-\eta} \sum_{k=1}^{N} |u_{kN}|^{2+\eta} \leq \text{const } \frac{\Sigma |u_{kN}|^{2+\eta}}{(\Sigma u_{kN}^2)^{1+\eta/2}} \text{ const } |U_N|^{\eta}$$

which tends to zero by (I) for any $\eta > 0$.- This proof also shows that (II) in theorem 2 must follow from (II)'.

Theorem 2 makes a statement about the quantities $B_N^{-1}(b_i - \beta_i)$. Here B_N is a function of the variances σ_k of the errors ϵ_k. In many cases, however, the σ_k are unknown. If we knew that B_N^2 would converge for any sequence of errors $\{\epsilon_\nu\}$ that might arise we could consider $(b_i - \beta_i)/(P^{-1})_{ii}$. This is a statistic and tends evidently to a normal distribution with zero mean and a variance which is known and equals $Q_N^2 = (B_N^2/(P^{-1})_{ii})^2$ if the limit of this expression is known. (Clearly, it always must lie between m^2 and M^2). But there is no hope for convergence of Q_N (to be equal $|Q_N|$) on any set $\tilde{F}$ which contains at least two r.v.-s with different variances between m and $M > m$, say. In this case namely we can construct a sequence Q_N which has at least two limit values. Because of (I) there is a sequence of integers $m_N \longrightarrow \infty$ with the property

$$\frac{1}{\underline{u}_N^2} \sum_{k=1}^{m_N} u_{kN}^2 < \frac{\eta}{M} \quad , \quad \frac{M-m}{2} > \eta > 0.$$

Hence for any choice of $\{\sigma_k\}$, $\sum_{k=1}^{m_N} \sigma_k^2 \frac{u_{kN}^2}{\underline{u}_N^2} < \eta$. Determine now a sequence $N_\rho \longrightarrow \infty$ such that $m_{N_{\rho+1}} \geq N_\rho$. Then put σ_j, $j = N_\rho + 1,\ldots,N_{\rho+1}$ equal to m resp. M alternating with ρ odd or even. Then either $Q_N < m + \eta$ for $N = N_\rho$, ρ odd, say, or

$$Q_N > M(1 - \frac{1}{\underline{u}_N^2} \sum_{k=1}^{m_N} u_{kN}^2) > M - \eta \quad \text{for } N = N_\rho,\ \rho \text{ even.}$$

In order to have a statistic for the case of an arbitrary set $\tilde{F}$ we can replace B_N by the estimate given in

Theorem 3 : Let

$$C_N^2 = \sum_{k=1}^{N} (\underline{p}_i' \, \underline{r}_k)^2 \, e_k^2$$

where e_k is the k-th component of $\underline{e} = \underline{y} - X\,\underline{b}$. Then it is sufficient for $C_N^{-1}(b_i - \beta_i)$, $i=1,\ldots,q$, to have asymptotically the standard normal distribution for any sequence of errors ϵ_k which are elements of any subset $\tilde{F} \subseteq F$, that (I) and (II) are satisfied:

(I) $$\frac{\max\limits_{k=1,\ldots,N} (\underline{r}_k' \, \underline{p}_i)^2}{(P^{-1})_{ii}} \longrightarrow 0 \qquad \text{as } N \longrightarrow \infty$$

(see also remark (3)),

(II) The fourth moments of all r.v.-s in $\tilde{F}$ exist and lie between two positive constants.

Proof: 1) We show first that $C_N^2 B_N^{-2}$ tends to 1 in probability. With

$\underline{e} = \underline{y} - X\underline{b} = \underline{\epsilon} + X(\underline{\beta} - \underline{b})$, with $\underline{\beta} - \underline{b} = -P^{-1}X'\ \underline{\epsilon}$, we have

$$E(e_k^2) \leq \sigma_k^2 + (M - 2\sigma_k^2)\ \underline{r}_k'\ P^{-1}\ \underline{r}_k$$

where M is an upper bound for all σ_k^2, existing by (II). With

$$\sum_{k=1}^{N} \underline{r}_k'\ P^{-1}\ \underline{r}_k = \operatorname{tr} X\ P^{-1}X' = q,$$ further by (I) and

(II) we have for $N \longrightarrow \infty$

$$\frac{E(C_N^2)}{B_N^2} = 1 + O\Big(\frac{\max_k(\underline{r}_k'\ \underline{p}_i)^2}{B_N^2}\Big) \longrightarrow 1\ .$$

In determining the variance it is convenient to introduce $\underline{v}_k = XP^{-1}\underline{r}_k$,

and with $u_{kN} = \underline{r}_k'\ \underline{p}_i$ we have

$$\operatorname{var}(C_N^2\ B_N^{-2}) \leq B_N^{-4}\ E\left\{C_N^2 - B_N^2\right\}^2 = B_N^{-4}\ E\left\{\sum_{k=1}^{N} u_{kN}^2(\epsilon_k^2 - \sigma_k^2 + (\underline{v}_k'\underline{\epsilon})^2 - 2\ \epsilon_k\underline{v}_k'\underline{\epsilon})\right\}^2 .$$

Because of $E\ \epsilon_k^4 <$ const $< \infty$, the contribution of the pure quadratic terms in the squared sum is $<$ const $B_N^{-4} \sum_{k=1}^{N} u_{kN}^4 <$ const $\dfrac{\max_k u_{kN}^2}{B_N^2} \longrightarrow 0$ by (I).

Use is made of $\sum_{k=1}^{N} \underline{v}_k^2 = q$. Similarly, in the mixed quadratic terms, one may take out of the occurring double sum the factor $B_N^{-2}\max_k u_{kN}^2$ which makes the remaining double sum (which is bounded) tend to zero.

2) Now by theorem 2, $B_N^{-1}(b_i - \beta_i)$ is a sequence of random variables whose distributions tend on $\tilde{F}$ for $N \longrightarrow \infty$ to N(0,1). By 1), $|C_N|B_N^{-1}$ is a sequence of r.v.-s which tend to 1. Hence their quotient $C_N^{-1}(b_i - \beta_i)$ tends to N(0,1), taking for C_N the positive root of C_N^2 [1, p. 254]. Thus theorem 3 is proved.

Remarks:

(1) Condition (I) includes that $N^{-1} \leq \dfrac{\max_{k=1,\ldots,N} u_k^2}{(P^{-1})_{ii}}$. This follows follows from $(P^{-1})_{ii} = \sum_1^N u_k^2 \leq N \max_{k=1,\ldots,N} u_k^2$.

(2) From the earlier corollaries it is seen that condition (I) in theorems 2 and 3 can be replaced by the following stronger assumptions (which of course no longer can be necessary):

$$(\mathrm{I}') \qquad \frac{(P^{-2})_{ii}}{(P^{-1})_{ii}} \max_{k=1,\ldots,N} \underline{r}_k^2 \longrightarrow 0 \quad \text{for } N \longrightarrow \infty ,$$

or even stronger

$$(\mathrm{I}'') \qquad \frac{\lambda_1(P)}{\lambda_q^2(P)} \max_{k=1,\ldots,N} \underline{r}_k^2 \longrightarrow 0 \qquad \text{as } N \longrightarrow \infty$$

which includes $\lambda_1(P)/\lambda_q^2(P) \longrightarrow 0$ and $\lambda_q(P) \longrightarrow \infty$ (remark to the first corollary).

(3) Clearly condition (II)' of the above corollary arising from the Lyapunov theorem does not add anything new to theorem 3 whose condition (II) is much stronger than (II)'.

(4) Theorem 2 suggests itself for a generalization in the sense that the u_{kN} are replaced by more general double sequences (work in this direction is under way). This would mean that limit theorems over certain sets or r.v.-s can be established, utilizing thereby mainly definition 3. Also convergence against any limit law, not only the normal one, can be taken into consideration.

Examples (1): If $q = 1$, there is only one regression vector $\underline{x}$, say. As i assumes only the value 1, we have $P^{-1} = (P^{-1})_{ii} = \underline{p}_i = \dfrac{1}{\underline{x}^2}$, and

and condition (I) reduces to $\frac{\max x_k^2}{\underline{x}^2} \longrightarrow 0$ which is equivalent with the pair of conditions

$$\frac{x_N^2}{\underline{x}^2} \longrightarrow 0 \quad \text{and} \quad \underline{x}_N^2 \longrightarrow \infty$$

for $N \longrightarrow \infty$. A regression vector of this kind is called slowly increasing [6].

(2) In the case where $q = 2$ one may put $\rho_N = \underline{x}_1' \underline{x}_2/(\underline{x}_1^2 \underline{x}_2^2)^{\frac{1}{2}}$, the correlation coefficient between the N-vectors $\underline{x}_1$ and $\underline{x}_2$. Then one easily computes that (I) is equivalent with

$$\frac{1}{\underline{x}_1^2 (1 - \rho_N)} \max_{k=1,\dots,N} \left(x_{1k} - \rho_N \sqrt{\frac{\underline{x}_1^2}{\underline{x}_2^2}} \, x_{2k}\right)^2 \longrightarrow 0$$

plus a similar equation in which subscripts 1 and 2 are inter-changed.

References

(1) H. Cramér, "Mathematical Methods of Statistics," Princeton University Press, 1957.

(2) F. Eicker, "Consistency of parameter - estimates in a linear time series model," Institute of Statistics, Mimeographed Series No. 265, 1960.

(3) F. Eicker, "A necessary and sufficient condition for consistency of the LS-estimates in linear regression," Institute of Statistics, Mimeographed Series No. 266, 1960.

(4) B. W. Gnedenko, "Lehrbuch der Wahrscheinlichkeitsrechnung," (translated from the Russian by Wolfgang Richter and Hans Joachim Rossberg), Akademie - Verlag, Berlin, 1957.

(5) B. V. Gnedenko and A. N. Kolmogorov, "Limit distributions for sums of independent random variables," (translated from the Russian by K. L. Chung and J. L. Doob), Addison-Wesley Publishing Company, Cambridge, Massachusetts, 1954.

(6) Ulf Grenander and Murray Rosenblatt, "Statistical analysis of stationary time series," John Wiley, New York, 1957.