A MULTIVARIATE CENTRAL LIMIT THEOREM FOR RANDOM LINEAR VECTOR FORMS[1]

By F. Eicker

University of Freiburg (Breisgau) and Columbia University

0. Summary. In [2] a central limit theorem [CLT] for sequences of univariate random linear forms was proved. That result is extended in this note to a multivariate CLT for q-dimensional linear forms of constant q-vectors with real-valued random weight coefficients. Some applications are indicated in Section 3.

1. Notation. We use the following notation which differs slightly from that used in [2]. Let $\mathfrak{F}$ be a (non-empty) set of distribution functions (d.f.'s) of random variables (r.v.'s) with zero means and positive, finite variances. Let $\mathcal{E}(\mathfrak{F})$ be the set of all sequences of independent random variables (independent within each sequence) whose d.f.'s belong to $\mathfrak{F}$, but are not necessarily the same from term to term of the sequence. A generic member of $\mathcal{E}(\mathfrak{F})$ will be denoted by $\epsilon = \{\epsilon_k ; k = 1, 2, \cdots\}$ or, when we discuss sequences of members of $\mathcal{E}(\mathfrak{F})$, by $\epsilon(n) = \{\epsilon_{nk} ; k = 1, 2, \cdots\}, n = 1, 2, \cdots$.

Let

$$A_n = (a_1(n), a_2(n), \cdots, a_{k_n}(n)) = (a_{jk}(n)), n = 1, 2, \cdots, j = 1, \cdots, q,$$

be a sequence of $q \times k_n$ matrices with column vectors $a_k(n) \varepsilon R_q$ (real q-dimensional Euclidean space with zero element ϕ) and elements $a_{jk}(n)$. Let $\min_n k_n \geqq q$, $\min_n \operatorname{rank} A_n = q$, $a_{k_n}(n) \neq \phi$ for all n.

For $\epsilon(n) \varepsilon \mathcal{E}(\mathfrak{F})$ we consider the random linear form $\sum_{k=1}^{k_n} a_k(n)\epsilon_{nk}$ (a random weighting of the vectors $a_1(n), a_2(n), \cdots$ with the elements $\epsilon_{n1}, \epsilon_{n2}, \cdots$ of $\epsilon(n)$). A short notation for this form is the matrix product $A_n\epsilon(n)$ where in this combination we interpret the symbol $\epsilon(n)$ as the vector $(\epsilon_{n1}, \cdots, \epsilon_{nk_n})'$ [$'$ denotes the transpose].

The covariance matrix of the vector $A_n\epsilon(n)$ is

$$B_n^2 = A_n\Sigma_nA_n' \tag{1}$$

where $\Sigma_n \equiv \operatorname{diag}(\sigma_{n1}^2, \cdots, \sigma_{nk_n}^2)$ is the covariance matrix of the vector $\epsilon(n)$. B_n is the unique positive definite square root of B_n^2. Thus, the random q-vectors

$$\zeta(n) = B_n^{-1}A_n\epsilon(n) \tag{2}$$

have mean ϕ and covariance matrix I_q (the q-dimensional identity matrix).

2. A CLT for random vector forms. We shall prove

$$\zeta(n) \rightarrow_L N(0, I_q) \tag{3}$$

Received 6 April 1965; revised 27 May 1966.

[1] Research supported in part by NSF-Grant no. NSF-GP-3694 at Columbia University.

($\rightarrow_L$ means convergence in distribution, $N(0, I_q)$ is the q-dimensional standard normal d.f.) for any sequence of sequences $\{\epsilon(n)\}$, $n = 1, 2, \cdots, \epsilon(n) \varepsilon \mathcal{E}(\mathfrak{F})$, if the following three conditions are simultaneously satisfied:

$$\text{(I}^*\text{)} \qquad \max\nolimits_{k=1,\cdots,k_n} a_k'(n)(A_nA_n')^{-1}a_k(n) \rightarrow 0$$

$$(4) \qquad \text{(II)} \qquad \sup\nolimits_{G\varepsilon\mathfrak{F}} \int_{|x|>c} x^2\, dG(x) \rightarrow 0 \quad \text{as} \quad c \rightarrow \infty$$

$$\text{(III)} \qquad \inf\nolimits_{G\varepsilon\mathfrak{F}} \int x^2\, dG(x) > 0.$$

In this paper all limits hold for $n \rightarrow \infty$ unless otherwise stated.

If the $\zeta(n)$ for all n are formed with one and the same sequence $\epsilon \varepsilon \mathcal{E}(\mathfrak{F})$ we write $\zeta(n; \epsilon)$. We say that *the summands of* $\zeta(n; \epsilon)$ *are infinitesimal* if

$$(5) \qquad \max\nolimits_{k=1,\cdots,k_n} P(\|B_n^{-1}a_k(n)\|\, |\epsilon_k| > \delta) \rightarrow 0 \quad \textit{for all} \quad \delta > 0$$

($\|\cdot\|$ means the Euclidean norm in R_q). We now have

THEOREM. $\xi(n; \epsilon) \rightarrow_L N(0, I_q)$ *and* (5), *both uniformly in* $\epsilon \varepsilon \mathcal{E}(\mathfrak{F})$, $\Leftrightarrow$ (I*), (II), (III).

The previous univariate result [2] is completely contained in the above theorem for $q = 1$. We remark that conditions (II) and (III) concern only the set $\mathfrak{F}$ and that (I*) requires no knowledge of the particular sequence ϵ occurring in a given situation. For $q = 1$, (I*) reduces to

$$\text{(I)} \qquad \max\nolimits_{k=1,\cdots,k_n} |a_{1k}(n)|\left(\sum\nolimits_{k=1}^{k_n} a_{1k}^2(n)\right)^{-\frac{1}{2}} \rightarrow 0.$$

For the proof of the theorem we need a lemma on the convergence in distribution of a sequence of random q-vectors $\{\xi(n)\}$, $n = 1, 2, \cdots$. We first derive Lemma 1 where F denotes a q-dimensional d.f., X a random vector $\sim F$, and S_q the unit sphere $\subset R_q$.

LEMMA 1. $\xi(n) \rightarrow_L F \Leftrightarrow \delta_n \equiv E \exp(ib_n'\xi(n)) - E \exp(ib_n'X) \rightarrow 0$ *for all sequences* $\{b_n\}$, $b_n \varepsilon S_q$.

PROOF. ($\Leftarrow$) Choose all $b_n \equiv \beta \varepsilon S_q$ and apply the multidimensional continuity theorem for characteristic functions ([1], p. 102).

($\Rightarrow$) If $b_n \rightarrow \beta$ in Euclidean norm, then the assertion follows from the Helly-Bray theorem and the continuity on R_q of a q-dimensional characteristic function. If b_n does not converge suppose $\lim_n \sup |\delta_n| = \delta > 0$. Then $\delta_{n'} \rightarrow \delta$ for a suitable subsequence $n' \uparrow \infty$, and $b_{n''} \rightarrow \beta^*$ for $\{n''\} \subset \{n'\}$, $n'' \uparrow \infty$, some $\beta^* \varepsilon R_q$, which implies $\delta_{n''} \rightarrow 0$ and thus yields a contradiction.

An immediate consequence is

LEMMA 2. $\xi(n) \rightarrow_L N(0, I_q)$ *if and only if* $b'(n)\xi(n) \rightarrow_L N(0, 1)$ *for all sequences of constant vectors* $\{b(n), n = 1, 2, \cdots\}$ *with* $b(n) \varepsilon S_q$.

Another implication of Lemma 1 is the well known equivalence: $\zeta(n) \rightarrow_L F \Leftrightarrow \lambda'\zeta(n) \rightarrow_L \lambda'X$ for all $\lambda \varepsilon S_q$ (comp. [4], p. 103).

We remark, although this observation is not needed subsequently, that also the following is true: $\xi(n) \rightarrow_L N(0, I_q)$ *if and only if* $v'(n)\xi(n) \rightarrow_L N(0, 1)$ *for all sequences of random q-vectors* $v(n)$ *for which there exists a sequence of constant vectors* $b(n) \varepsilon S_q$ *such that* $v(n) - b(n) \rightarrow 0$ i.p..

PROOF OF THE THEOREM. (3) and (5) hold uniformly in ϵ if and only if both statements hold for every sequence of sequences $\epsilon(1), \epsilon(2), \cdots$. To prove these modified statements we use Lemma 2 with

$$b(n) = \|B_n c(n)\|^{-1} B_n c(n), \tag{6}$$

$\{c(n)\}$ being any sequence of vectors $c(n) \varepsilon S_q$. Then

$$b'(n)\zeta(n) = (c'(n)B_n^2 c(n))^{-\frac{1}{2}} c'(n) A_n \epsilon(n). \tag{7}$$

Putting $c'(n)a_k(n) = \alpha_{nk}$ (3) is seen to be equivalent with the normal convergence of

$$\left(\sum_{k=1}^{k_n} \alpha_{nk}^2 \sigma_{nk}^2\right)^{-\frac{1}{2}} \sum_{k=1}^{k_n} \alpha_{nk} \epsilon_{nk} \tag{8}$$

for all $\{c(n)\}$ and (5) is seen to be equivalent with

$$\sup_{c(n)\varepsilon S_q} \max_{k=1,\cdots,k_n} P\left(\left|\left(\sum_{j=1}^{k_n} \alpha_{nj}^2 \sigma_{nj}^2\right)^{-\frac{1}{2}} \alpha_{nk}\epsilon_{nk}\right| > \delta\right) \to 0 \tag{9}$$

for all $\delta > 0$. To see the latter we write $\max_{k=1,\cdots,k_n} \sup_{b(n)\varepsilon S_q}$ instead of $\sup_{c(n)\varepsilon S_q} \max_{k=1,\cdots,k_n}$ and note that

$$\left(\sum_{j=1}^{k_n} \alpha_{nj}^2 \sigma_{nj}^2\right)^{-\frac{1}{2}} \alpha_{nk} = b'(n) B_n^{-1} a_k(n),$$

$$\sup_{b(n)\varepsilon S_q} |b'(n) B_n^{-1} a_k(n)| = \|B_n^{-1} a_k(n)\|.$$

Hence the left hand sides of (9) and (5) [replacing here ϵ_k by ϵ_{nk}] are equal.

Now by Theorem 1 of [2], (8) and (9) are jointly equivalent to the three statements (II), (III), and

$$\sup_{k;c(n)\varepsilon S_q} \left(\alpha_{nk}^2 \Big/ \sum_{j=1}^{k_n} \alpha_{nj}^2\right) \to 0. \tag{10}$$

The left hand side of (10) equals

$$\begin{aligned}\sup_{k;c\varepsilon S_q} \{(c'a_k(n))^2/(c'A_nA_n'c)\} &= \sup_{k;c\varepsilon S_q} \{(c'(A_nA_n')^{-\frac{1}{2}} a_k(n))^2\} \\ &= \max_k a_k'(n)(A_nA_n')^{-1} a_k(n).\end{aligned}$$

Thus (10) is equivalent to (I*), and the theorem is proved.

3. Remarks. The above theorem can be applied, e.g., to determine the joint asymptotic distribution of the least squares parameter estimates in a multiple linear regression system with not necessarily identically distributed error random variables [3].

In the applications of the theorem the d.f.'s of the random variable of the sequence ϵ are frequently unknown and consequently also their variances σ_k^2 that occur in the matrices B_n. It can be shown, however, that without further assumptions these σ_k^2 can be replaced by ϵ_k^2 and that the statements (3) and (5) of the theorem remain valid [3].

Conditions (I*), (II), (III) presumably remain necessary, if instead of the whole class $\mathcal{E}(\mathfrak{F})$ only a subset is admitted. However, we do not pursue this question further.

REFERENCES

[1] CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton Univ. Press.

[2] EICKER, F. (1963). Central limit theorems for families of sequences of random variables. *Ann. Math. Statist.* **34** 439–446.

[3] EICKER, F. (1965). Limit theorems for regressions with unequal and dependent errors. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* Univ. of California Press.

[4] RAO, C. R. (1965). *Linear Statistical Inference and Its Applications*. Wiley, New York.