EXISTENCE OF CONSISTENT ESTIMATES OF THE DIRECTIONAL PARAMETER IN A LINEAR STRUCTURAL RELATION BETWEEN TWO VARIABLES[1]

By Jerzy Neyman

University of California, Berkeley, California

Summary. Let Z_n denote the system of $8n$ independent pairs of measurements (X_{ik}, Y_{ik}), for $i = 1, 2, \cdots, n$ and $k = 1, 2, \cdots, 8$, of two nonobservable random variables ξ_{ik} and η_{ik}, known to satisfy a linear relation of the form $\xi_{ik} \cos \theta^* + \eta_{ik} \sin \theta^* - p = 0$, where p is an arbitrary real number and θ^* may have any value between the limits

$$-\tfrac{1}{2}\pi < \theta^* \leqq \tfrac{1}{2}\pi.$$

The purpose of the paper is to construct a class of estimates $T_n(Z_n)$ of the parameter θ defined as follows: when $\theta^* = \frac{1}{2}\pi$ then $\theta = 0$; otherwise $\theta \equiv \theta^*$. Each estimate $T_n(Z_n)$ of the class considered converges in probability to θ as $n \to \infty$ under the following conditions: (i) except when $\theta = 0$, the variables ξ_{ik} are nonnormal; (ii) any nonnormal components of the errors of measurements, $X_{ik} - \xi_{ik}$ and $Y_{ik} - \eta_{ik}$, are mutually independent, independent of ξ_{ik} and of the normal components of these errors; (iii) the normal components of the errors may be correlated but as a pair are independent of ξ_{ik}.

1. Introduction. Let ξ and η be two random variables known to be linearly connected, so that there exist two numbers, θ^* and p,

$$(1) \qquad -\tfrac{1}{2}\pi < \theta^* \leqq \tfrac{1}{2}\pi, \qquad -\infty < p < +\infty,$$

such that the simultaneous values of ξ and η satisfy the condition

$$(2) \qquad \xi \cos \theta^* + \eta \sin \theta^* - p = 0.$$

We consider the case where ξ and η are not directly observable but where the observations yield the simultaneous values of two other random variables X and Y, connected with ξ and η by the equations

$$(3) \qquad X = \xi + U, \qquad Y = \eta + V.$$

Here U and V are unobservable random variables interpreted as errors in measuring ξ and η, respectively. Equation (2) is described as the linear structural relation between the variables X and Y. Throughout the paper it is assumed that the errors U and V may be correlated or not but, as a pair, are independent

[1] This paper was prepared with the partial support of the Office of Naval Research. It presents an extension of the contents of the Second Rietz Memorial Lecture delivered by the author at the Summer Meeting of the Institute of Mathematical Statistics at Boulder, Colorado, September 1st, 1949.

of the variables ξ and η. The problem considered is that of using a sequence $\{X_m, Y_m\}$ of completely independent pairs of observations on X and Y to construct a consistent estimate of θ^*. This is an old problem and a number of the earlier attempts to solve it are described by Wald in an important paper [1].

Early attempts to obtain a consistent estimate of θ^* were based exclusively on the sample variances and covariance of X and Y. However, as early as 1916, Godfrey Thomson showed [2] that the same first and second moments of the simultaneous distribution of X and Y are compatible with an infinity of different values of θ^* and that, therefore, attempts to estimate this parameter using only second order sample moments are doomed to failure. The writings of Thomson appear to have been overlooked and more and more studies were published using sample moments of the first and second orders as basic functions on which the estimates of θ^* were built. In 1936 [3] it was pointed out that, should it happen that the unobservable random variables ξ and η and also the errors U and V are normally distributed, then no consistent estimate of θ^* is possible because, in this event, the joint distribution of X and Y is also normal, and is determined by moments of the first two orders. Since these moments are consistent with an infinity of different values of θ^*, the latter is nonidentifiable. Between 1936 and the appearance of the paper by Wald in 1940 several studies were published, of which we will mention one by R. G. D. Allen [4], adding more precision to the facts just described.

Wald's paper brought a new idea into the situation. Namely, in certain cases something may be known about the particular values assumed by the unobservable random variable ξ. When this condition obtains, a method due to Wald gives a consistent estimate of θ^*. This estimate is again based on the arithmetic means of the observations on X and Y, appropriately grouped. Wald's idea took root and led to the paper by Housner and Brennan [5]. The same idea, a little more developed, is at the base of papers by Berkson [6] and by Hemelrijk [7]. However, important as these developments may be in various fields of application, it is obvious that they do not constitute a solution of the original problem of estimating θ^* when no knowledge of the particular values assumed by the unobservable random variables is postulated [8].

A new era in the study of the problem began following the result of Reiersøl [9][2] who proved that the case of nonidentifiability of θ^* noted in 1936 is an exception rather than a rule. This discovery stimulated the paper by Scott [10] giving a consistent estimate of θ^* applicable in a new category of cases, when no information on the particular values of ξ is postulated. However, the consistency of the estimate of Scott depends on the existence of a certain number of moments of the variable ξ.

The present paper is concerned with the case where the errors of measurement may be split into two components

[2] Although this paper appeared in print in 1950, the author became acquainted with it in the spring of 1948 from a lecture delivered by Reiersøl in a seminar meeting at the Statistical Laboratory, University of California, Berkeley.

$$(4) \qquad \begin{cases} U = U_1 + U_2, \\ V = V_1 + V_2, \end{cases}$$

where U_1 and V_1 are mutually independent and, as a pair, are independent of (U_2, V_2), and where U_2 and V_2 follow an arbitrary normal distribution. With the exception of the above independence, no restriction is placed on the distributions of U_1 and V_1. The purpose of the paper is to give an explicit construction of an estimate of a parameter θ (closely allied to but not identical with θ^*) which remains consistent in the most general case of identifiability, that is when ξ and η follow an arbitrary nonnormal distribution. No knowledge of particular values of ξ is postulated.

Since the above hypotheses admit the possibility that X and Y have no moments at all, the conventional methods of constructing the estimate have to be abandoned. Essentially, the estimate is defined as the abscissa which corresponds to the minimum ordinate of a point on a random curve. A search for this minimum among the roots of the derivative may be embarrassing. In fact, the derivative need not exist at all points. Therefore, the estimate is defined as the outcome of a specially devised interpolation procedure. The proof is based on a lemma which seems to have an interest of its own and may be applicable in other cases.

2. Concepts of identifiability and of consistent estimability. In order to define the concepts of identifiability[3] and of consistent estimability, we shall consider a variable point ϑ (parameter) capable of assuming any one of a set s of positions ϑ'. Every $\vartheta' \varepsilon s$ will be described as a possible value of ϑ. For every $\vartheta' \varepsilon s$ consider a specified set $\omega(\vartheta')$ of distribution functions and let ω stand for the union of all $\omega(\vartheta')$ for $\vartheta' \varepsilon s$.

DEFINITION 1. *We shall say that the parameter ϑ is identifiable in ω if, whatever $\vartheta' \varepsilon s$ and whatever $\vartheta'' \varepsilon s$, $\vartheta' \neq \vartheta''$, the corresponding sets $\omega(\vartheta')$ and $\omega(\vartheta'')$ have no elements in common.*

If ϑ is identifiable in ω, then to every distribution function $F \varepsilon \omega$ there corresponds a uniquely defined value of ϑ, say $\vartheta(F) \varepsilon s$.

From now on we shall restrict ourselves to sets ω of distribution functions F defined in the same Euclidean space of a fixed number m of dimensions. For every $F \varepsilon \omega$ we shall consider an m-dimensional random variable $X(F)$ whose distribution function is F. For $n = 1, 2, \cdots$ the symbol $Y_n(F)$ will denote the set of n completely independent observations made on $X(F)$. Thus, $Y_n(F)$ may be considered as a random variable of dimensionality mn. Let y_n denote a point in the mn-dimensioned Euclidean space R_{mn}. Consider a sequence of Borel measurable functions $\{T_n(y_n)\}$, each from R_{mn} to s. Obviously, the result $T_n(Y_n(F))$ of substituting $Y_n(F)$ for y_n in $T_n(y_n)$ is a random variable.

DEFINITION 2. *If the parameter ϑ is identifiable in ω and if, whatever be $F \varepsilon \omega$,*

[3] Important discussion of this concept, in a slightly different form, is due to Koopmans and Reiersøl [11]. This paper contains a substantial bibliography.

the sequence $\{T_n(Y_n(F))\}$ *converges in probability to* $\vartheta(F)$ *as* $n \to \infty$, *then this sequence is called a consistent estimate of* ϑ *in* ω.

DEFINITION 3. *If the parameter* ϑ *is identifiable in* ω *and if there exists a consistent estimate of* ϑ *in* ω, *then we shall say that* ϑ *is consistently estimable in* ω.

3. Identifiability of the directional parameter in the linear structural relation of two random variables. Returning to the general situation described in Section 1, denote by θ the parameter defined as follows:

$$\text{if} \quad -\tfrac{1}{2}\pi < \theta^* < \tfrac{1}{2}\pi, \quad \text{then} \quad \theta = \theta^*,$$

$$\text{if} \quad \theta^* = \tfrac{1}{2}\pi, \qquad \text{then} \quad \theta = 0.$$

The parameter θ thus defined will be called the directional parameter of the structural relation (2).

Denote by S the set of possible values of θ, $-\frac{1}{2}\pi < \theta < \frac{1}{2}\pi$. For every value ϑ of this set we shall now define a set $\Omega(\vartheta)$ of joint distributions of the variables X and Y of formulae (3). We begin by defining $\Omega(0)$.

If $\theta = 0$ then either $\theta^* = 0$ or $\theta^* = \frac{1}{2}\pi$. Accordingly, $\Omega(0)$ is defined as the union of the two sets of distributions, $\Omega^*(0)$ and $\Omega^*(\frac{1}{2}\pi)$, each corresponding to a particular value of θ^*. If $\theta^* = 0$ then formula (2) implies that ξ is degenerate and $\xi = p$. Assume the following hypotheses:

(a) The variable η has an arbitrary distribution.

(b) $U = U_1 + U_2$ and $V = V_1 + V_2$, where U_1 and V_1 are mutually independent, as a pair are independent of ξ and η but otherwise arbitrarily distributed, and where (U_2, V_2) represent a pair of arbitrary normal variables, independent of the triplet η, U_1, V_1. In particular, U_2 and V_2 may be correlated.

(c) $-\infty < p < +\infty$.

Obviously, every specific set of hypotheses regarding η, U_1, V_1, U_2, V_2 and p implies a specific distribution of the pair X, Y. Then $\Omega^*(0)$ denotes the set of all such distributions.

In order to define $\Omega^*(\frac{1}{2}\pi)$ we notice that, if $\theta^* = \frac{1}{2}\pi$ then (2) implies that η is degenerate and $\eta = p$. $\Omega^*(\frac{1}{2}\pi)$ is defined to contain every joint distribution of X and Y implied by an arbitrary assumption regarding the distribution of ξ and by hypotheses (b) and (c).

As mentioned $\Omega(0)$ is the union of $\Omega^*(0)$ and $\Omega^*(\frac{1}{2}\pi)$. However, the reader will verify easily that the sets $\Omega^*(0)$ and $\Omega^*(\frac{1}{2}\pi)$ coincide. Therefore θ^* is not identifiable in $\Omega(0)$.

For every possible value ϑ of θ, other than $\vartheta = 0$, the set $\Omega(\vartheta)$ is defined to contain every joint distribution of X and Y defined by formulae (3), implied by the assumption that ξ follows an arbitrary nondegenerate, nonnormal distribution, that η is connected with ξ by equation (2) with p having an arbitrary real value, and that the errors U and V are arbitrarily distributed, subject to condition (b). It will be seen that the equality $\theta = 0$ characterizes the case where at least one of the variables ξ and η is degenerate so that, instead of being linearly connected, these variables may be considered as mutually independent.

Reiersøl proved [9] that the parameter θ is identifiable in the set Ω of distributions of X and Y defined as the union of all sets $\Omega(\vartheta)$ for $-\frac{1}{2}\pi < \vartheta < \frac{1}{2}\pi$. Since it is known that the restriction of nonnormality imposed on ξ and η when $\vartheta \neq 0$ cannot be relaxed without destroying the identifiability of θ, it follows that Ω is the broadest set of joint distributions of X and Y within which θ is identifiable, consistent with the assumption that the errors of measurement U and V satisfy assumption (b). The purpose of the present paper is to provide an explicit construction of an estimate of θ consistent in Ω.

4. A few preliminaries. It will be convenient to use the concept of uniform convergence in probability. Let $G(x)$ denote a function defined over a non-degenerate closed interval $x \varepsilon [a, b]$. Further, let $\{Z_n\}$ be a infinite sequence of random variables and $\{G_n(Z_n, x)\}$ a sequence of functions of two arguments Z_n and x. Each $G_n(Z_n, x)$ is assumed to be defined for every $x \varepsilon [a, b]$ and for every possible value of the random variable Z_n. Furthermore, when x is fixed, $G_n(Z_n, x)$ is a Borel measurable function of Z_n. Thus, it is a random variable.

DEFINITION 4. *We shall say that the sequence $\{G_n(Z_n, x)\}$ of random functions converges in probability to $G(x)$ uniformly in $[a, b]$, if there exists a function $m(n)$ defined for all $n = 1, 2, \cdots$ such that*

$$\lim_{n\to\infty} m(n) = \infty \tag{5}$$

and such that, whatever $\epsilon > 0$,

$$\lim_{n\to\infty} \left(m(n) \sup_{x \varepsilon [a,b]} P\{ | G_n(Z_n, x) - G(x) | > \epsilon \} \right) = 0. \tag{6}$$

Every function $m(n)$ satisfying the above conditions will be described as the norm of uniform convergence of $\{G_n(Z_n, x)\}$. Obviously, it may always be assumed that the norm $m(n)$ assumes only positive integer values.

In order to illustrate this concept, assume that for every $x \varepsilon [a, b]$ and for every $n = 1, 2, \cdots$ we have

$$E[G_n(Z_n, x)] = G(x) \tag{7}$$

and that the variance $\sigma_n^2(x)$ of $G_n(Z_n, x)$ is bounded by

$$\sigma_n^2(x) \leqq \frac{1}{n} \sigma_0^2, \tag{8}$$

where $\sigma_0 > 0$ is a constant. Using the inequality of Bienaymé-Tchebycheff we may write

$$P\{ | G_n(Z_n, x) - G(x) | > \epsilon \} < \frac{\sigma_n^2(x)}{\epsilon^2} \leqq \frac{\sigma_0^2}{n\epsilon^2} \tag{9}$$

for every $x \varepsilon [a, b]$. Thus

$$\sup_{x \varepsilon [a,b]} P\{ |G_n(Z_n, x) - G(x) | > \epsilon \} < \frac{\sigma_0^2}{n\epsilon^2}, \tag{10}$$

and it is seen that, under conditions (7) and (8), the sequence $\{G_n(Z_n, x)\}$ converges in probability to $G(x)$ uniformly in $[a, b]$. For example, the norm of uniform convergence may be defined as the greatest integer not exceeding the square root of n,

$$m(n) = [\sqrt{n}]. \tag{11}$$

Another convenient concept will be described as the m-lattice minimal point of a function f. This is defined as follows. Let $[a, b]$ denote a nondegenerate closed interval and $f(x)$ a real function defined on $x \varepsilon [a, b]$. Let m be an arbitrary integer $m > 1$ and

$$a_{mk} = a + k\frac{b-a}{m-1} \quad \text{for} \quad k = 0, 1, \cdots, m-1. \tag{12}$$

We shall say that the m points a_{mk} form the m-lattice on $[a, b]$. Now consider the values $f(a_{mk})$ of $f(x)$ corresponding to the points of the m-lattice and use the symbol f_m to denote the smallest of these. In general, there will be r points of the lattice, say

$$a_{mk_1} < a_{mk_2} < \cdots < a_{mk_r} \tag{13}$$

such that $f(a_{mk_i}) = f_m$. Let $\mu = [(r + 1)/2]$. The point $a_{m\mu}$ will be described as the m-lattice minimal point of the function $f(x)$. It will be denoted by $M_m(f(x))$.

FUNDAMENTAL LEMMA. *If the real function $G(x)$ is defined and continuous on a nondegenerate closed interval $[a, b]$ in which it has an absolute minimum $G(x_0)$ attained at a single point x_0, if $\{Z_n\}$ is a sequence of random variables and if the sequence of real random functions $\{G_n(Z_n, x)\}$ converges to $G(x)$ uniformly in $[a, b]$ with an integer valued norm $m(n)$, then the sequence $\{M_{m(n)}[G_n(Z_n, x)]\}$ of $m(n)$-lattice minimal points of $G_n(Z_n, x)$ converges in probability to x_0.*

PROOF. Assume that the conditions of the lemma are satisfied. The proof consists in showing that, whatever $\epsilon > 0$ and $\eta > 0$, a number $N(\epsilon, \eta)$ can be found such that the inequality $n > N(\epsilon, \eta)$ implies

$$P\{| M_{m(n)}[G_n(Z_n, x)] - x_0 | > \epsilon\} < \eta. \tag{14}$$

Fix ϵ and η and denote by g the minimum value of $G(x)$ attained in the part of $[a, b]$ outside of the open interval $| x - x_0 | < \epsilon$. Obviously $g > G(x_0)$. Let $\delta < \epsilon$ be a sufficiently small positive number such that $| x - x_0 | < \delta$ implies

$$G(x_0) \leqq G(x) < G(x_0) + \tfrac{1}{3}(g - G(x_0)). \tag{15}$$

Denote by N_1 the smallest integer such that $n > N_1$ implies

$$\frac{b-a}{m(n)-1} < \delta \tag{16}$$

and by N_2 the smallest integer such that $n > N_2$ implies

$$m(n) \sup_{x\varepsilon[a,b]} P\{| G_n(Z_n, x) - G(x) | > \tfrac{1}{3}(g - G(x_0))\} < \eta. \tag{17}$$

Finally, let $N(\epsilon, \eta) = \max(N_1, N_2)$. It is easy to see that for $n > N(\epsilon, \eta)$, the inequality (14) is satisfied. We notice first that with $n > N(\epsilon, \eta) \geqq N_1$ the interval $(x_0 - \delta, x_0 + \delta)$ will include some points of the $m(n)$-lattice. Further, in order that $| M_{m(n)}[G_n(Z_n, x)] - x_0 | > \epsilon$ it is necessary that at least one of the values of $G_n(Z_n, x)$ assumed at points of the $m(n)$-lattice outside of the interval $(x_0 - \epsilon, x_0 + \epsilon)$ not exceeded any of the values assumed by this function on the $m(n)$-lattice within $(x_0 - \delta, x_0 + \delta)$. But outside of $(x_0 - \epsilon, x_0 + \epsilon)$ we have

$$G(x_0) < g \leqq G(x) \tag{18}$$

and inside of $(x_0 - \delta, x_0 + \delta)$

$$G(x) < G(x_0) + \tfrac{1}{3}(g - G(x_0)). \tag{19}$$

It follows that, if at each point of the $m(n)$-lattice the random function $G_n(Z_n, a_{mk})$ differs from $G(a_{mk})$ by at most $\frac{1}{3}(g - G(x_0))$, then

$$| M_{m(n)}[G_n(Z_n, x)] - x_0 | \leqq \epsilon.$$

Thus, the probability that $| M_{m(n)}[G_n(Z_n, x)] - x_0 | > \epsilon$ is at most equal to the probability, say π, that for at least one point a_{mk} of the $m(n)$-lattice $| G_n(Z_n, a_{mk}) - G(a_{mk}) | > \frac{1}{3}(g - G(x_0))$. However,

$$\begin{aligned} \pi &\leqq \sum_{k=0}^{m(n)-1} P\{| G_n(Z_n, a_{mk}) - G(a_{mk}) | > \tfrac{1}{3}(g - G(x_0))\} \\ &\leqq m(n) \sup_{x \varepsilon [a,b]} P\{| G_n(Z_n, x) - G(x) | > \tfrac{1}{3}(g - G(x_0))\} < \eta \end{aligned} \tag{20}$$

because of (17), and the proof of the lemma is completed.

5. Consistent estimates of the directional parameter of a linear structural relation between two variables. We return to the problem of the consistent estimation of the directional parameter θ of the structural relation (2). The parameter θ was defined in Section 3. Also it will be assumed that the joint distribution function F of the variables X and Y belongs to the set Ω defined in Section 3. Consider a set of $N(n) = 8n$ independent observations to be made on the pair of variables X and Y. These observations will be divided into n eight-tuples and denoted by (X_{ij}, Y_{ij}) for $i = 1, 2, \cdots, n$ and $j = 1, 2, \cdots, 8$. The ith eight-tuple will be denoted by Z_i^*. The totality of n eight-tuples will be denoted by Z_n.

In defining the estimate of θ we shall need three (identical or different) probability density functions $w_1(x)$, $w_2(x)$, and $w_3(x)$, and their characteristic functions, say $\Phi_1(t)$, $\Phi_2(t)$, and $\Phi_3(t)$, respectively. These probability density functions can be selected arbitrarily out of a class Γ which we shall define by the following conditions: every $w(x) \varepsilon \Gamma$ is symmetric about zero, $w(-x) = w(x)$, and there exists a positive number a such that $w(x) > 0$ for every $| x | < a$.

It will be observed that the symmetry of $w_k(x)$ implies that the corresponding characteristic function $\Phi_k(t)$ is real.

Speaking in terms of the characteristic functions Φ_1, Φ_2, Φ_3, we shall define a class C of consistent estimates of θ. Any particular choice of the functions Φ_1, Φ_2 and Φ_3 will determine a particular estimate of the class C. For example, we may choose to consider the following probability densities of class Γ: (1) the normal probability density with zero mean and unit variance, (2) the Cauchy probability density with unit scale and zero location parameter, and (3) the rectangular probability density between $-a$ and $+a$. Each of the corresponding characteristic functions, $\exp\{-\frac{1}{2}t^2\}$, $\exp\{-|t|\}$, and $\sin at/at$, respectively, may be taken to represent either Φ_1 or Φ_2 or Φ_3, or any two, or all three $\Phi_1 = \Phi_2 = \Phi_3$.

Assume that the choice of the functions $\Phi_k(t)$ is made. Denote by ϑ an arbitrary number between the limits $-\frac{1}{2}\pi \leqq \vartheta \leqq +\frac{1}{2}\pi$. For the kth eight-tuple of observations define the following symbols

$$(21)\quad \begin{cases} A(Z_k^*, \vartheta) = \Phi_1[(X_{k1} - X_{k2} + X_{k3} - X_{k4}) \cos \vartheta \\ \qquad + (Y_{k1} - Y_{k2} + Y_{k3} - Y_{k4}) \sin \vartheta]\Phi_2(X_{k1} - X_{k2} + X_{k5} - X_{k6}), \\ B(Z_k^*) = \Phi_3(Y_{k1} - Y_{k2} + Y_{k7} - X_{k8}), \\ C(Z_k^*) = \Phi_3(Y_{k1} - Y_{k4} - Y_{k6} + Y_{k7}), \\ D(Z_k^*) = \Phi_3(Y_{k3} - Y_{k4} + Y_{k5} - Y_{k6}), \end{cases}$$

$$(22)\quad H(Z_k^*, \vartheta) = A(Z_k^*, \vartheta)\{B(Z_k^*) - 2C(Z_k^*) + D(Z_k^*)\}.$$

Finally, let

$$(23)\quad G_n(Z_n, \vartheta) = \frac{1}{n}\sum_{k=1}^{n} H(Z_k^*, \vartheta).$$

Put $m(n) = [\sqrt{n}]$ and consider the $m(n)$-lattice on the closed interval $[-\frac{1}{2}\pi, +\frac{1}{2}\pi]$. For every fixed value Z_n' of Z_n we consider $G_n(Z_n', \vartheta)$ as a function of $\vartheta \in [-\frac{1}{2}\pi, +\frac{1}{2}\pi]$ and then $M_{m(n)}(G_n(Z_n', \vartheta))$ will denote its $m(n)$-lattice minimal point. After these preliminaries we define the estimate $T_n(Z_n)$ of θ as follows.

$$(24)\quad \begin{aligned} &\text{(i) If } G_n(Z_n, 0) \leqq \frac{1}{\sqrt[4]{n}}, \quad \text{then} \quad T_n(Z_n) = 0. \\ &\text{(ii) If } G_n(Z_n, 0) > \frac{1}{\sqrt[4]{n}}, \quad \text{then} \quad T_n(Z_n) = M_{m(n)}[G_n(Z_n, \vartheta)]. \end{aligned}$$

THEOREM. *The sequence $\{T_n(Z_n)\}$ represents an estimate of θ consistent in Ω.*

PROOF. We begin by noticing that, since the symbols in (21) are defined in terms of characteristic functions, their absolute values cannot exceed unity. Therefore,

$$(25)\quad |H(Z_k^*, \vartheta)| \leqq 4,$$

and thus all moments of $H(Z_k^*, \vartheta)$ exist. In particular, we shall be interested in the first moment, say

$$E\{H(Z_k^*, \vartheta)\} = E\{G_n(Z_n, \vartheta)\} = G(\vartheta), \tag{26}$$

and in the variance, say $\sigma^2(\vartheta)$, of $H(Z_k^*, \vartheta)$. Obviously, $\sigma^2(\vartheta) \leqq 16$. Since the successive variables $H(Z_k^*, \vartheta)$ are completely independent, the variance of $G_n(Z_n, \vartheta)$, say $\sigma_G^2(\vartheta)$, is

$$\sigma_G^2(\vartheta) = \frac{\sigma^2(\vartheta)}{n} \leqq \frac{16}{n}, \tag{27}$$

and it follows that the sequence $\{G_n(Z_n, \vartheta)\}$ converges in probability to $G(\vartheta)$ uniformly in $[-\frac{1}{2}\pi, \frac{1}{2}\pi]$. As we have seen before (see Section 4) the function $m(n) = [\sqrt{n}]$ may be taken as the norm of the uniform convergence.

Our next step in the proof consists in showing that the function $G(\vartheta)$ has the following properties.

(A) If the random variables X and Y follow a distribution $F \varepsilon \Omega$ such that $\theta(F) = 0$, then $G(\vartheta) \equiv 0$ for all $\vartheta \varepsilon [-\frac{1}{2}\pi, \frac{1}{2}\pi]$ including $\vartheta = 0$.

(B) If $\theta(F) \neq 0$, then $G(\vartheta) > 0$ for all $\vartheta \varepsilon [-\frac{1}{2}\pi, \frac{1}{2}\pi]$ with the exception of $\vartheta = \theta(F)$ where $G[\theta(F)] = 0$.

(C) $G(\vartheta)$ is continuous for $\vartheta \varepsilon [-\frac{1}{2}\pi, \frac{1}{2}\pi]$.

Once these three properties of $G(\vartheta)$ are established, the proof of the theorem is completed as follows. Assume first that $\theta(F) = 0$. Then, by the theorem of Bienaymé-Tchebycheff,

$$\begin{aligned} P\left\{G_n(Z_n, 0) \leqq \frac{1}{\sqrt[4]{n}}\right\} &\geqq P\left\{ |G_n(Z_n, 0)| \leqq \frac{1}{\sqrt[4]{n}}\right\} \\ &> 1 - \sigma_G^2(0)\sqrt{n} \geqq 1 - \frac{16}{\sqrt{n}}. \end{aligned} \tag{28}$$

The definition of $T_n(Z_n)$ implies that it is equal to zero whenever

$$G_n(Z_n, 0) \leqq \frac{1}{\sqrt[4]{n}}, \quad \text{unconditionally,} \tag{29}$$

and also whenever

$$G_n(Z_n, 0) > \frac{1}{\sqrt[4]{n}} \quad \text{and} \quad M_{m(n)}[G_n(Z_n, \vartheta)] = 0. \tag{30}$$

Consequently, the probability

$$P\{T_n(Z_n) = 0\} \geqq P\left\{G_n(Z_n, 0) \leqq \frac{1}{\sqrt[4]{n}}\right\} \geqq 1 - \frac{16}{\sqrt{n}} \tag{31}$$

and tends to unity as $n \to \infty$.

Assume now that $\theta(F) \neq 0$. According to the fundamental lemma, in this case $M_{m(n)}(G_n(Z_n, \vartheta))$ converges in probability to $\theta(F)$. To prove that the same is true for $T_n(Z_n)$ it is sufficient to show that the probability $P\{T_n(Z_n) \neq M_{m(n)}[G_n(Z_n, \vartheta)]\}$ tends to zero as $n \to \infty$. Obviously this last probability does not exceed the probability that $G_n(Z_n, 0) \leqq n^{-\frac{1}{4}}$. According to property (B) we have $G(0) > 0$ in the case considered. When $n > G(0)^{-4}$, we have

$$P\left\{G_n(Z_n, 0) \leqq \frac{1}{\sqrt[4]{n}}\right\} \leqq P\left\{ \mid G_n(Z_n, 0) - G(0) \mid > G(0) - \frac{1}{\sqrt[4]{n}}\right\} < \frac{16}{n\left(G(0) - \frac{1}{\sqrt[4]{n}}\right)^2}, \tag{32}$$

and it follows that, as $n \to \infty$, the probability that $T_n(Z_n)$ will coincide with $M_{m(n)}(G_n(Z_n, \vartheta))$ tends to unity. It is seen that the properties (A), (B), and (C) of the function $G(\vartheta)$ combined with (26) imply that, whatever $F \varepsilon \Omega$, the estimate $\{T_n(Z_n)\}$ converges in probability to $\theta(F)$ or, in other words, that $T_n(Z_n)$ is an estimate of θ consistent in Ω. Therefore, in order to prove the theorem, we shall establish that the expectation (26) has the properties (A), (B), and (C). This will be done in Section 6 in the following order. First we shall use the postulated properties of the observable random variables X and Y and define a function $G(\vartheta)$ having the properties (A), (B), and (C). Next we shall show that the function $G(\vartheta)$ so defined coincides with the expectation (26).

6. Structural definition of $G(\vartheta)$. The structural definition of $G(\vartheta)$ is based on the properties of the characteristic function, say $\phi(t_1, t_2)$, of the joint distribution of X and Y. According to the usual definition

$$\phi(t_1, t_2) = E(e^{it_1X+it_2Y}), \tag{33}$$

where

$$X = \xi + U_1 + U_2, \quad Y = \eta + V_1 + V_2. \tag{34}$$

Assume first that $\theta = 0$. In this case the components $\xi + U_1$ and $\eta + V_1$ are mutually independent and the possible dependence of X and Y will be due to the correlation that may exist between the normal components of errors U_2 and V_2. Since the logarithm of the characteristic function of two normal variables is a polynomial of the second order, when $\theta = 0$ the characteristic function of X and Y has the form, say

$$\phi(t_1, t_2 \mid \theta = 0) = e^{\psi_1(t_1)+\psi_2(t_2)+\gamma t_1 t_2}, \tag{35}$$

where $\psi_i(t_i)$ is a function of t_i alone, $i = 1, 2$. We note this form of $\phi(t_1, t_2 \mid \theta = 0)$ for future reference and proceed to the next case, where $\theta \neq 0$.

In this case $\theta = \theta^*$ and the structural relation (2) may be solved with respect to

$$\eta = \frac{p}{\sin \eta} - \xi \cot \theta. \tag{36}$$

Substituting this expression into (34) and denoting the logarithm of the characteristic function of ξ by $\chi(t)$, we have

$$\phi(t_1, t_2) = e^{\chi(t_1 - t_2 \cot\theta) + \psi_1(t_1) + \psi_2(t_2) + \gamma t_1 t_2}, \tag{37}$$

where the symbols ψ_1 and ψ_2 are again used to denote functions of one argument only, either t_1 or t_2. These functions in (37) have a meaning different from that in (35). However, this difference is of no importance because in both cases the essential point is that ψ_1 depends on t_1 but not on t_2 and that ψ_2 depends on t_2 but not on t_1. It will be convenient to consider that $\phi(t_1, t_2)$ always has the form (37) with the understanding that, when $\theta = 0$, then $\chi(t) \equiv 0$.

Since $\psi_1(t)$, $\psi_2(t)$ and $\chi(t)$ are defined in terms of logarithms of characteristic functions, they vanish at $t = 0$ and are continuous at this point. In addition, we shall use the following important property of $\chi(t)$. This is that, whenever $\theta(F) \neq 0$, then however small $\delta > 0$, the function $\chi(t)$ cannot coincide with a polynomial of second order on the whole of the interval $(-\delta, \delta)$. This property is implied by the hypothesis that, whenever $\theta \neq 0$ and therefore $\chi(t) \not\equiv 0$, then ξ is not normally distributed. In fact, assume that there exists a positive number δ^* such that $\chi(t) = a + bt + ct^2$ for all $|t| < \delta^*$. It is easy to see that in this case all the derivatives of the characteristic function of ξ would exist at $t = 0$ and would determine all the moments of ξ. Furthermore, these moments would coincide with the moments of a normal distribution, from which it would follow that ξ itself is normally distributed, contrary to the hypothesis. Thus it follows that, if $\chi(t)$ coincides with a quadratic in t over an interval, this interval cannot include $t = 0$.

Select a number $\vartheta \, \varepsilon \, [-\frac{1}{2}\pi, \frac{1}{2}\pi]$ and three arbitrary real numbers t, τ_1, τ_2. We shall consider $\phi(t_1, t_2)$ at the following eight points which, to abbreviate the formulae, will be denoted by lower case Roman numerals. Thus, for example, $\phi(\text{i})$ will denote the value of $\phi(t_1, t_2)$ evaluated at the first of the eight points. The coordinates of the first four points are

(i) $\quad t_1 = t\cos\vartheta + \tau_1, \quad t_2 = t\sin\vartheta + \tau_2,$

(ii) $\quad t_1 = t\cos\vartheta + \tau_1, \quad t_2 = t\sin\vartheta,$

(iii) $\quad t_1 = t\cos\vartheta, \quad t_2 = t\sin\vartheta + \tau_2,$

(iv) $\quad t_1 = t\cos\vartheta, \quad t_2 = t\sin\vartheta.$

The coordinates of points (v) through (viii) are obtained from those of (i) to (iv), respectively, by substituting $t = 0$. Thus

(v) $\quad t_1 = \tau_1, \quad t_2 = \tau_2,$

(vi) $\quad t_1 = \tau_1, \quad t_2 = 0,$

(vii) $\quad t_1 = 0, \quad t_2 = \tau_2,$

(viii) $\quad t_1 = 0, \quad t_2 = 0.$

Obviously $\phi(\text{viii}) = 1$. Now we form the function

$$h(\vartheta, t, \tau_1, \tau_2) = \phi(\text{i})\phi(\text{iv})\phi(\text{vi})\phi(\text{vii}) - \phi(\text{ii})\phi(\text{iii})\phi(\text{v})\phi(\text{viii}). \tag{38}$$

Easy algebra gives

$$h(\vartheta, t, \tau_1, \tau_2) = \Psi_1\Psi_2 - \Psi_1\Psi_3, \tag{39}$$

where

$$\begin{cases} \Psi_1 = \exp\{\psi_1(t\cos\vartheta + \tau_1) + \psi_1(t\cos\vartheta) + \psi_1(\tau_1) \\ \qquad + \psi_2(t\sin\vartheta + \tau_2) + \psi_2(t\sin\vartheta) + \psi_2(\tau_2) \\ \qquad + \gamma[(t\cos\vartheta + \tau_1)(t\sin\vartheta + \tau_2) + t^2\cos\vartheta\sin\vartheta]\}, \\ \Psi_2 = \exp\{\chi(At + \tau_1 - \tau_2\cot\theta) + \chi(At) + \chi(\tau_1) + \chi(-\tau_2\cot\theta)\}, \\ \Psi_3 = \exp\{\chi(At + \tau_1) + \chi(0) + \chi(At - \tau_2\cot\theta) + \chi(\tau_1 - \tau_2\cot\theta)\}, \end{cases} \tag{40}$$

with

$$A = \frac{\sin(\theta - \vartheta)}{\sin\theta}. \tag{41}$$

For any $x > 0$ we shall use the symbol $\sigma(x)$ to denote the set of triplets (t, τ_1, τ_2) such that $|t| < x$, $|\tau_1| < x$ and $|\tau_2| < x$. Because of the properties of the functions ψ_1, ψ_2, and χ there exists a positive number δ such that within $\sigma(\delta)$ the functions Ψ_1 and Ψ_3 do not vanish. Consequently, for $(t, \tau_1, \tau_2)\ \varepsilon\ \sigma(\delta)$ we may write

$$\begin{aligned} h(\vartheta, t, \tau_1, \tau_2) &= \Psi_1\Psi_3\left(\frac{\Psi_2}{\Psi_3} - 1\right) \\ &= \Psi_1\Psi_3\Big(\exp\{[\chi(At + \tau_1 - \tau_2\cot\theta) - \chi(At + \tau_1) \\ &\qquad - \chi(At - \tau_2\cot\theta) + \chi(At)] \\ &\qquad - [\chi(\tau_1 - \tau_2\cot\theta) - \chi(\tau_1) - \chi(-\tau_2\cot\theta) + \chi(0)]\} - 1\Big). \end{aligned} \tag{42}$$

The idea of the function $h(\vartheta, t, \tau_1, \tau_2)$ originated from the paper of Reiersøl and this function is the key to the whole construction of the estimate $T_n(Z_n)$. The function $h(\vartheta, t, \tau_1, \tau_2)$ is defined as a combination of values of the characteristic function of the observable random variables X and Y at eight arbitrarily selected points. Consequently, the definition of $h(\vartheta, t, \tau_1, \tau_2)$ is independent of the value of $\theta(F)$. However, the properties of $h(\vartheta, t, \tau_1, \tau_2)$ do depend on $\theta(F)$, as follows.

(a) If $\theta(F) = 0$, then $h(\vartheta, t, \tau_1, \tau_2) = 0$ for all values of the four arguments $\vartheta\ \varepsilon\ [-\frac{1}{2}\pi, \frac{1}{2}\pi]$ and $-\infty < t, \tau_1, \tau_2 < +\infty$.

(b) If $\theta(F) \neq 0$ and $\vartheta = \theta(F)$, then $h(\vartheta, t, \tau_1, \tau_2) = 0$ for all combinations of values of t, τ_1, τ_2, $-\infty < t, \tau_1, \tau_2 < +\infty$.

(c) If $\theta(F) \neq 0$ and $\vartheta \neq \theta(F)$ then, whatever $\delta_1 > 0$, the cube $\sigma(\delta_1)$ contains a subset of points (t, τ_1, τ_2) of positive three-dimensional measure within which $h(\vartheta, t, \tau_1, \tau_2) \neq 0$.

In order to prove (a) we notice that the case $\theta(F) = 0$ is characterized by the identity $\chi(t) \equiv 0$. Making this substitution in (42) it is immediately seen that, in this particular case, $h(\vartheta, t, \tau_1, \tau_2) \equiv 0$ for all combinations of values of the four arguments.

In order to prove (b) we notice that $\vartheta = \theta(F)$ implies

$$A = \frac{\sin (\vartheta - \theta)}{\sin \theta} = 0. \tag{43}$$

Then (42) implies that $h(\theta, t, \tau_1, \tau_2) \equiv 0$ for all combinations of values of the three arguments t, τ_1, τ_2.

In proving (c) we shall use the hypothesis that ξ is not a normal variable and that, therefore, however small $\delta > 0$, the function $\chi(t)$ cannot coincide with a polynomial of second order on the whole of the interval $|t| < \delta$. Assume that the assertion (c) is not true and that, with $\vartheta \neq \theta(F) \neq 0$, there exists a positive number δ^* such that, for $(t, \tau_1, \tau_2) \,\varepsilon\, \sigma(\delta^*)$ we have identically $h(\vartheta, t, \tau_1, \tau_2) \equiv 0$. Then this identity will also hold for all sufficiently small $|t|$ and $|\tau_1|$ and

$$\tau_2 = -\tau_1 \tan \theta. \tag{44}$$

Within the common part of $\sigma(\delta)$ and $\sigma(\delta^*)$ the functions Ψ_1 and Ψ_3 do not vanish. Therefore, we must conclude that the result of substituting (44) into Ψ_2 and Ψ_3 must give $\Psi_2/\Psi_3 \equiv 1$ for all sufficiently small $|t|$ and $|\tau_1|$. This however, implies that

$$\chi(At + 2\tau_1) - 2\chi(At + \tau_1) + \chi(At) \equiv \chi(2\tau_1) - 2\chi(\tau_1) + \chi(0). \tag{45}$$

It will be seen that the expressions on both sides of this identity represent second differences of $\chi(t)$ at steps τ_1 evaluated at points At and zero, respectively. Thus, the assumption $h(\vartheta, t, \tau_1, \tau_2) \equiv 0$ in $(t, \tau_1, \tau_2) \,\varepsilon\, \sigma(\delta^*)$ leads to the conclusion that there must exist a certain vicinity W of the point $t = 0$ where, however small $|\tau_1|$, the second difference of the function $\chi(t)$ computed at steps τ_1 has a value possibly depending on τ_1 but not on the point at which it is evaluated. Since $\chi(t)$ is continuous, it must then coincide with a polynomial of second order in t over the whole interval W. This, however, is contrary to the hypothesis. Therefore, if $\vartheta \neq \theta(F) \neq 0$, whatever be $\delta > 0$ the cube $\sigma(\delta)$ must contain at least one point t', τ_1', τ_2' such that $h(\vartheta, t', \tau_1', \tau_2') \neq 0$. Since h is continuous in (t, τ_1, τ_2) it then follows that $\sigma(\delta)$ must contain a set of three-dimensional positive measure where $h(\vartheta, t, \tau_1, \tau_2) \neq 0$. This establishes (c).

When $h(\vartheta, t, \tau_1, \tau_2) \neq 0$, it may be represented by a real or by a complex number. It is known that by changing the signs of the arguments of any characteristic function one obtains a value which is conjugate to the original value of this characteristic function. It is easy to see that the same applies to $h(\vartheta, t, \tau_1, \tau_2)$. Therefore, the product

$$h(\vartheta, t, \tau_1, \tau_2)h(\vartheta, -t, -\tau_1, -\tau_2) = |h(\vartheta, t, \tau_1, \tau_2)|^2 = g(\vartheta, t, \tau_1, \tau_2), \tag{46}$$

say, is equal to the square of the modulus of $h(\vartheta, t, \tau_1, \tau_2)$. It follows from the preceding that the function $g(\vartheta, t, \tau_1, \tau_2)$ is real valued, nonnegative and continuous in (t, τ_1, τ_2). Also, it is easy to see that $g(\vartheta, t, \tau_1, \tau_2)$ cannot be greater than 4. Furthermore, if $\theta(F) = 0$ then g is identically zero. Also, it is zero identically in t, τ_1, τ_2 if $\theta(F) \neq 0$ but $\vartheta = \theta(F)$.

On the other hand, if $\theta(F) \neq 0$ and $\vartheta \neq \theta(F)$, then in every vicinity of $t = \tau_1 = \tau_2 = 0$ there is a set of positive three-dimensional measure where $g(\vartheta, t, \tau_1, \tau_2) > 0$. Now, let $w_1(x)$, $w_2(x)$ and $w_3(x)$ be three (identical or different) probability density functions of class Γ (that is, each symmetric about $x = 0$ and nonvanishing in a nondegenerate interval $|x| < a$). Also, let

$$(47)\quad G(\vartheta) = \int_{-\infty}^{+\infty} w_1(t) \int_{-\infty}^{+\infty} w_2(\tau_1) \int_{-\infty}^{+\infty} w_3(\tau_2) g(\vartheta, t, \tau_1, \tau_2)\, dt\, d\tau_1\, d\tau_2 .$$

It is obvious that, whatever the chosen probability density functions w_1, w_2, and w_3,

$$\text{if } \theta(F) = 0, \text{ then } G(\vartheta) = 0 \text{ for every } \vartheta \,\varepsilon\, [-\tfrac{1}{2}\pi, \tfrac{1}{2}\pi],$$

$$\text{if } \theta(F) \neq 0 \text{ and } \vartheta = \theta(F), \text{ then } G(\vartheta) = 0,$$

$$\text{if } \theta(F) \neq 0 \text{ and } \vartheta \neq \theta(F), \text{ then } G(\vartheta) > 0.$$

Also, because of the definition of $g(\vartheta, t, \tau_1, \tau_2)$ in terms of the characteristic function of X and Y, $G(\vartheta)$ is a continuous function of ϑ. It follows that the function G defined in formula (47) possesses the properties (A), (B), and (C) mentioned at an earlier stage of the proof of the theorem (Section 5). In order to complete this proof, we now show that, if $\Phi_k(t)$ denotes the characteristic function of $w_k(x)$, $k = 1, 2, 3$, then the expectation of $H(Z_m^*, \vartheta)$, defined by (22) and (21), is equal to $G(\vartheta)$ of formula (47), identically in $\vartheta \,\varepsilon\, [-\tfrac{1}{2}\pi, \tfrac{1}{2}\pi]$.

For this purpose we return to the function $g(\vartheta, t, \tau_1, \tau_2)$ and reexamine its definition (46) in terms of $h(\vartheta, t, \tau_1, \tau_2)$ and ultimately in terms of the characteristic function $\phi(t_1, t_2)$ as in (38). It is seen that $g(\vartheta, t, \tau_1, \tau_2)$ and $G(\vartheta)$ may be written conveniently as linear combinations of four terms each, say

$$(48)\qquad G(\vartheta) = G_1(\vartheta) - G_2(\vartheta) - G_3(\vartheta) + G_4(\vartheta),$$

$$(49)\qquad g(\vartheta, t, \tau_1, \tau_2) = g_1 - g_2 - g_3 + g_4,$$

where, for $k = 1, 2, 3, 4$,

$$(50)\qquad G_k(\vartheta) = \iiint g_k w_1(t) w_2(\tau_1) w_3(\tau_2)\, dt\, d\tau_1\, d\tau_2,$$

and g_k stands for the product of from six to eight factors, each factor representing the characteristic function of X and Y evaluated at specified values of the two arguments. Upon inspecting (46) and (38) the reader will have no difficulty in writing down the expressions of the four components g_k. To save space we shall reproduce only the expression of g_1 represented by the product of eight factors, as follows:

$$g_1 = \phi(t \cos \vartheta + \tau_1, t \sin \vartheta + \tau_2)\phi(-t \cos \vartheta - \tau_1, -t \sin \vartheta - \tau_2) \cdot \phi(t \cos \vartheta, t \sin \vartheta)\phi(-t \cos \vartheta, -t \sin \vartheta) \cdot \phi(\tau_1, 0)\phi(-\tau_1, 0)\phi(0, \tau_2)\phi(0, -\tau_2). \tag{51}$$

Consider the kth eight-tuple of independent observations on X and Y and let (X_{kj}, Y_{kj}) represent the jth pair of this eight-tuple. Obviously, we may write

$$\phi(t \cos \vartheta + \tau_1, t \sin \vartheta + \tau_2) = E[\exp\{it(X_{k1} \cos \vartheta + Y_{k1} \sin \vartheta) + i\tau_1 X_{k1} + i\tau_2 Y_{k1}\}], \tag{52}$$

$$\phi(-t \cos \vartheta - \tau_1, -t \sin \vartheta - \tau_2) = E[\exp\{-it(X_{k2} \cos \vartheta + Y_{k2} \sin \vartheta) - i\tau_1 X_{k2} - i\tau_2 Y_{k2}\}], \tag{53}$$

etc. Because of the complete independence of all the eight pairs (X_{kj}, Y_{kj}), the expression of g_1 may be written as the expectation of a single exponential,

$$g_1 = E[\exp\{it((X_{k1} - X_{k2} + X_{k3} - X_{k4}) \cos \vartheta + (Y_{k1} - Y_{k2} + Y_{k3} - Y_{k4}) \sin \vartheta) + i\tau_1(X_{k1} - X_{k2} + X_{k5} - X_{k6}) + i\tau_2(Y_{k1} - Y_{k2} + Y_{k7} - Y_{k8})\}]. \tag{54}$$

This expectation is just a convenient symbol for an eightfold Stieltjes integral with respect to the distribution function $F(x_j, y_j)$ of each pair (X_{kj}, Y_{kj}). Thus the component $G_1(\vartheta)$ of $G(\vartheta)$ is an elevenfold integral. Since this integral is absolutely convergent, we may invert the order of integration and write

$$G_1(\vartheta) = E\left(\int_{-\infty}^{+\infty} e^{it((X_1-X_2+X_3-X_4)\cos\vartheta+(Y_1-Y_2+Y_3-Y_4)\sin\vartheta)} w_1(t)\, dt \cdot \int_{-\infty}^{+\infty} e^{i\tau_1(X_1-X_2+X_5-X_6)} w_2(\tau_1)\, d\tau_1 \cdot \int_{-\infty}^{+\infty} e^{i\tau_2(Y_1-Y_2+Y_7-Y_8)} w_3(\tau_2)\, d\tau_2\right), \tag{55}$$

or, remembering the definition of Φ_1, Φ_2, and Φ_3,

$$G_1(\vartheta) = E\{\Phi_1[(X_1 - X_2 + X_3 - X_4) \cos \vartheta + (Y_1 - Y_2 + Y_3 - Y_4) \sin \vartheta] \cdot \Phi_2(X_1 - X_2 + X_5 - X_6)\Phi_3(Y_1 - Y_2 + Y_7 - Y_8)\}, \tag{56}$$

or, finally

$$G_1(\vartheta) = E[A(Z_k^*, \vartheta)B(Z_k^*)], \tag{57}$$

with the symbols $A(Z_k^*, \vartheta)$ and $B(Z_k^*)$ defined for every eight-tuple of completely independent observations as in formulae (21). Similarly it is easy to show that

$$\begin{cases} G_2(\vartheta) = G_3(\vartheta) = E[A(Z_k^*, \vartheta)C(Z_k^*)], \\ G_4(\vartheta) = E[A^*(Z_k^*, \vartheta)D(Z_k^*)]. \end{cases} \tag{58}$$

This, however, implies that

$$G(\vartheta) = E[H(Z_k^*, \vartheta)], \tag{59}$$

and the proof of the theorem is completed.

7. Acknowledgment. The results presented in this paper differ in several respects from the contents of the Second Rietz Memorial Lecture of 1949. Among other things it was possible to remove a certain restrictiveness of the original estimate of θ. The parameter considered in 1949 was not θ itself but rather $\beta = \cot \theta$. In order to construct the original estimate of β, it was necessary to use a number B known to exceed $|\beta|$. It is a pleasure to acknowledge the author's indebtedness to Professor Charles M. Stein for a useful suggestion which led to the present construction of the estimate of θ, independent of any advance knowledge of the value of this parameter.

REFERENCES

[1] Abraham Wald, "The fitting of straight lines if both variables are subject to error," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 284–300.

[2] Godfrey H. Thomson, "A hierarchy without a general factor," *British Jour. Psych.*, Vol. 8 (1916), pp. 271–281.

[3] J. Neyman, "Remarks on a paper by E. C. Rhodes," *Jour. Roy. Stat. Soc.*, Vol. 100 (1937), pp. 50–57.

[4] R. G. D. Allen, "The assumptions of linear regression," *Economica*, N. S., Vol. 6 (1939), pp. 191–204.

[5] W. G. Housner and J. F. Brennan, "The estimation of linear trends," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 380–388.

[6] Joseph Berkson, "Are there two regressions?" *Jour. Am. Stat. Assn.*, Vol. 45 (1950), pp. 164–180.

[7] J. Hemelrijk, "Construction of a confidence region for a line," *Nederl. Akad. Wetensch., Proc.*, Vol. 52 (1949), pp. 995–1005.

[8] J. Neyman and Elizabeth L. Scott, "On certain methods of estimating the linear structural relation," *Annals of Math. Stat.*, Vol. 22 (1951), pp. 352–361.

[9] Olav Reiersøl, "Identifiability of a linear relation between variables which are subject to error," *Econometrica*, Vol. 18 (1950), pp. 375–389.

[10] Elizabeth L. Scott, "Note on consistent estimates of the linear structural relation between two variables," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 284–288.

[11] T. C. Koopmans and O. Reiersøl, "The identification of structural characteristics," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 165–181.