IDENTIFICATION AND ESTIMATION OF LINEAR MANIFOLDS IN n-DIMENSIONS[1]

By T. A. Jeeves

University of California, Berkeley

1. Summary. This paper investigates the problem of identifiability and estimability of linear structures in n dimensions. The concept of identifiability is examined to elucidate the senses in which it may be interpreted in the present problem. Particular attention is given to the question of treating linear subspaces rather than specific coordinate systems. Necessary and sufficient conditions for identifiability are obtained under the assumption that the "errors" follow a multinormal distribution.

2. Introduction. In many fields of statistical application it is not possible to observe directly the variables of interest but only to observe related random variables. Let $X = (X_1, X_2, \cdots, X_n)$ be a random (row) vector which is "unobservable" and $Y = (Y_1, Y_2, \cdots, Y_n)$ be a random (row) vector which is "observable." Assume that $Y = X\mathbf{B} + U$, where $\mathbf{B}$ is a parameter having $n \times n$ matrices of sure numbers for values and $U = (U_1, U_2, \cdots, U_n)$ is a random (row) vector which is stochastically independent of X.

In this paper particular attention will be given to the case in which U has a multinormal distribution, and it is desired to determine the row space S of the value of $\mathbf{B}$. Two problems are considered: (a) identifiability, whether S is determined if the distribution of Y is known [1], [2], and (b) estimability, whether S can be estimated consistently [1] from an infinite sequence of observations on Y.

Similar problems were considered in 1901 by Pearson [3]. As early as 1916, Thomson [4] showed that estimates based on moments no higher than the second would not be consistent. In 1936, Neyman [5] indicated a set of conditions in which, because of nonidentifiability, no consistent estimates existed. A summary of the state of the problem in 1940 was given by Wald [6], who brought an entirely new approach. An answer for the case of two dimensions was supplied by Reiersøl [7] in 1948.

3. Identifiability. The problem of identification in n dimensions introduces features not present in the two dimensional problem. In particular, just what is to be identified and hence estimated must be clarified. In n dimensions a greater variety of possible interpretations is available. To elucidate the sense in which the problem is treated here, and to bring out the relationships to other work, it seems necessary and profitable to begin with some general remarks on identification, culminating in the definition of identifiability (Definition 3) utilized in the remainder of the paper.

Received 12/15/52, revised 2/12/54.

[1] This paper was prepared with the partial support of the Office of Naval Research.

Following Neyman [1], the concept of identifiability is defined in the following way. Let L be a relation $L(\vartheta, F)$ between the elements ϑ of a space Θ and the elements F of a set Ω of distribution functions. Let $\omega(\vartheta) = \{F \mid L(\vartheta, F)\}$. Let θ be a parameter with range $\Theta_s \subset \Theta$.

DEFINITION 1. θ is *identifiable* (L) if the sets $\omega(\vartheta)$ are disjoint for every $\vartheta \varepsilon \Theta_s$.

This definition generalizes that of Neyman, in that Θ and Θ_s are not necessarily identical. The definition emphasizes the relation L between elements of Θ and elements of Ω. Essentially the definition states that identifiability obtains if no two distinct elements of Θ_s are related to the same distribution function. However, for the succeeding discussion, it is important to notice that this definition implies the existence of a relation among the elements of the larger space Θ, and that this relation characterizes the nature of the identification. The following theorem, which follows easily from the definition, brings out this point. The following notation is introduced: for any $F \varepsilon \Omega$,

$$\gamma(F) = \{\vartheta \varepsilon \Theta \mid F \varepsilon \omega(\vartheta)\}, \qquad \Gamma(\vartheta) = \bigcup_{F \varepsilon \omega(\vartheta)} \gamma(F), \qquad \Omega_s = \bigcup_{\vartheta \varepsilon \Theta_s} \omega(\vartheta).$$

THEOREM 1. *θ is identifiable (L) if and only if there is a relation R between the elements of Θ such that for every $\vartheta \varepsilon \Theta_s$ and every $\vartheta^* \varepsilon \Gamma(\vartheta)$,*

(i) *$R(\vartheta, \vartheta^*)$ holds, and*

(ii) *$R(\vartheta^*, \vartheta)$ holds only if $\vartheta^* = \vartheta$.*

The relation R is uniquely defined by the relation L for any $\vartheta \epsilon \Theta_s$ and $\vartheta^* \varepsilon \Gamma(\vartheta)$; conversely, specifying R implies restrictions on L.

From this it is seen that if θ is identifiable (L) then there is a one-to-one correspondence for $\vartheta \varepsilon \Theta_s$ between ϑ and $\omega(\vartheta)$, and also between ϑ and $\Gamma(\vartheta)$. Further, every $F \varepsilon \Omega_s$ determines a unique value of ϑ and there exists a function I with the domain Ω_s and range Θ_s such that if $F \varepsilon \Omega_s$, then $F \varepsilon \omega(I(F))$. If $\Theta = \Theta_s$, then the relation R is equality. In the following, particular attention will be given to the case in which the ϑ are linear spaces and R is the relation of inclusion.

To apply the definition to the problem considered, it is necessary to exhibit the relation L. Let M be the set of $n \times n$ matrices and Θ a family of subsets of M. Let Ω be the set of n-dimensional distribution functions and $\mathfrak{X}$, $\mathfrak{U}$, and $\mathfrak{Y}$ be nonempty subsets of Ω associated with the random variables X, U, and Y, respectively. Let F_X, F_U, and F_Y be the distribution functions of the random variables X, U, and Y, respectively.

DEFINITION 2. For any sets $\mathfrak{X}$, $\mathfrak{U}$, and Θ, the relation $L(\vartheta, F_Y)$ holds if $B \varepsilon \vartheta$ and if $Y = XB + U$ for some X and U such that $F_X \varepsilon \mathfrak{X}$ and $F_U \varepsilon \mathfrak{U}$.

It has been shown [8] that conditions must be imposed on both $\mathfrak{X}$ and $\mathfrak{U}$ if θ is to be identifiable (L).

Further analysis of the problem requires consideration of the effect on the matrix B of a nonsingular transformation P. Identification problems may be proposed in which the space Θ_s is so specialized that PB no longer belongs to an element of Θ_s, or may not for certain P. Such problems will not be considered in this paper. It will be assumed that Θ_s has the following property: for any nonsingular $n \times n$ matrix P, if $B \varepsilon \vartheta \varepsilon \Theta_s$, then $PB \varepsilon \vartheta^*$ for some $\vartheta^* \varepsilon \Theta_s$. Considering the definition of L above it follows that:

THEOREM 2. *In any sets* Θ, Θ_s, $\mathfrak{X}$, *and* $\mathfrak{U}$ *such that* $\Theta_s \varepsilon \Theta$ *and* Θ_s *has the above property, if* θ *is identifiable* (L) *then for each nonsingular matrix* P *either*

(a) $F_X \varepsilon \mathfrak{X}$ *implies* F_{XP} *not a member of* $\mathfrak{X}$ *for every* X, *or*

(b) $B \varepsilon \vartheta$ *implies* $P^{-1}B \varepsilon \vartheta$ *for every* $\vartheta \varepsilon \Theta_s$.

The content of this theorem indicates two broad categories of problems, those in which condition (a) is satisfied by all nonsingular matrices and those in which condition (b) is satisfied by all nonsingular matrices. Mixed problems in which some matrices P satisfy (a) and some (b) might also be considered. The assumption that condition (a) is satisfied for every P leads to the consideration developed by Koopmans [2], [9].

This paper explores the implications of assuming that condition (b) is satisfied for all P, that is, that the matrices belonging to ϑ are all row equivalent or have the same row space. It will thus be convenient to think of ϑ as a row space. With this interpretation the problem being considered below is that of identifying the row space of the matrix B. The row space is a natural parameter in the problem of general linear structures. As such problems frequently arise, the components of X are presumed to lie in a linear subspace of Euclidean n-space; the determination of this linear subspace is desired. The specification of a particular set of coordinates on this subspace (that is, the determination of B) is frequently not required.

Throughout the remainder of this paper it will be assumed that the elements of Θ_s are the sets of row-equivalent matrices corresponding to the various row spaces of dimension s and that $\Theta = \bigcup_1^n \Theta_s$. It will also be assumed that $\mathfrak{U}$ is the set of multinormal distributions. Since $\mathfrak{X}$ is not specified, the relation L is not completely determined. Instead of specifying the set $\mathfrak{X}$, it will suffice to select the relation R (see Theorem 1) and investigate what conditions on $\mathfrak{X}$ are necessary and sufficient for identifiability. Two natural relations among linear spaces are the relation of equality and the relation of inclusion. The treatment here will be confined to the relation of inclusion. Similar results for the relation of equality have been obtained [8].

In view of the preceding considerations, the definition of identifiability may be particularized for the relation of inclusion as follows:

DEFINITION 3. θ is *identifiable* (L^*) if $\mathcal{S}(\vartheta) \subset \mathcal{S}(\vartheta^*)$ for every $\vartheta \varepsilon \Theta_s$ and $\vartheta^* \varepsilon \Gamma(\vartheta)$.

Here $\mathcal{S}(\vartheta)$ denotes the row space of ϑ, that is, the vector space spanned by the row vectors of any element of ϑ, while L^* denotes a relation L which gives rise to the relation R of inclusion (see Theorem 1). Here $R(\vartheta, \vartheta^*)$ means $\mathcal{S}(\vartheta) \subset \mathcal{S}(\vartheta^*)$.

4. Necessary and sufficient conditions. As in the case of two dimensions [7], identifiability is related to a lack of normality in the random variable X. This concept of the amount of nonnormality of a random variable is defined below.

DEFINITION 4. The *dimension* of a random variable U is the smallest dimension of all linear subspaces which contain U with probability one.

DEFINITION 5. Let $nn(Y)$ be the least value of d such that $Y = U + V$ with U and V independent, V having a multinormal distribution and U having dimension d. This value $nn(Y)$ will be called the *nonnormality* of Y.

DEFINITION 6. The nonnormality of $\mathcal{Y}$ is s (i.e., $nn(\mathcal{Y}) = s$) if $nn(Y) = s$ for every Y such that $F_Y \varepsilon \mathcal{Y}$.

In terms of the definition of nonnormality, the main result on identifiability of linear manifolds in n-dimensions can be stated as follows.

THEOREM 3. *θ is identifiable (L^*) if and only if $nn(\mathcal{Y}) = s$.*

The proof of the theorem depends on the following lemmas, the proofs of which are straightforward [8] and will not be given.

LEMMA 1. *If M is an $n \times n$ matrix with rank s and if $(n - s)$ columns of M are identically zero, then every row either belongs to some $s \times s$ submatrix with rank s, or else is identically zero.*

LEMMA 2. *If A is a symmetric matrix, E is a diagonal matrix with ones and zeros on the main diagonal, and EAE is positive semidefinite, then there exist matrices C, G, and H such that*

$$CAC' = DGD + EHE,$$

where $D + E = I$ (the identity) and H is a diagonal matrix with ones and zeros on the main diagonal, C is nonsingular, and $CD = D$.

The following choice of notation has been made. The symbol $f(t)$ will be used to denote some polynomial of the second degree in t, but not necessarily the same polynomial at each usage. Distinct polynomials will not generally be distinguished. The characteristic function of a random variable X will be denoted by $\varphi_X(t) = \int e^{itx'}\, dF_X(x)$, where t and x are row vectors. Further, $\psi_X(t) = -\log \varphi_X(t)$.

LEMMA 3. *If $\psi_Y(t) = \psi_X(tB') + \psi_U(t)$, and if U has a multinormal distribution, then for any matrix C which is idempotent and row-equivalent to B,*

$$\psi_Y(t) = \psi_Y(tC') + f(t).$$

In particular C may be the canonical form of B.

DEFINITION 7. The *canonical form* of the matrix B is a matrix C which is row equivalent to B, with elements satisfying, for each $i = 1, \cdots, n$,

(a) $c_{ii} = 0$ or $c_{ii} = 1$;

(b) if $c_{ii} = 0$, then $c_{ij} = 0$ for all j and $c_{ji} = 0$ for $j \geqq i$;

(c) if $c_{ii} = 1$, then $c_{ij} = 0$ for $j < i$ and $c_{ji} = 0$ for $j \neq i$.

Lemma 2 can be used to prove

LEMMA 4. *If $\varphi_Y(t) = \psi_Y(tB') + f(t)$, then $\psi_Y(t) = \psi_Y(tF') + \psi_Y(t(I - F)')$, where*

(i) *F is idempotent and row-equivalent to B*

(ii) $\exp\{-\psi_Y[t(I - F)']\}$ *is the characteristic function of a multinormal random variable.*

From the preceding lemma and the definition of nonnormality one easily obtains

LEMMA 5. *$nn(Y) = s$ if and only if s is the minimum rank of all matrices A such that*

$$\psi_Y(t) = \psi_Y(tA') + f(t).$$

Lemma 5 hence furnishes an alternate definition of nonnormality.

PROOF OF THEOREM 3. If $B \varepsilon \vartheta$, then $r(B) = s$. Let Y be any random variable such that $L^*(\vartheta, F_Y)$, and let $t = nn(Y)$. Then

$$Y = XB + U. \tag{1}$$

(a) From the relation L^*, it follows that

$$\psi_Y(t) = \psi_X(tB') + \psi_U(t). \tag{2}$$

Hence by Lemma 3, $\psi_Y(t) = \psi_Y(tC') + f(t)$, where C is idempotent and row-equivalent to B. Therefore, by Lemma 5, $t \leqq s$.

(b) Assume θ is identifiable (L^*). By Lemma 5 there exists a matrix A of rank t such that $\psi_Y(t) = \psi_Y(tA') + f(t)$. Hence there exists a matrix F having the properties enumerated in Lemma 4. But from (2) above, it follows that

$$\psi_Y(tF') = \psi_X(t(BF)') + \psi_U(tF').$$

Therefore, letting $B^* = BF$ and $U^* = UF + Y(I - F)$, one obtains that X and U^* are independent and U^* has a multinormal distribution. Further,

$$Y = XB^* + U^*,$$

so that $L^*(\vartheta^*, F_Y)$ holds, where ϑ^* is the set having B^* as an element. Now $r(B^*) \leqq r(F) = t$. But since θ is identifiable, $\mathcal{S}(\vartheta) \subset \mathcal{S}(\vartheta^*)$ so that $r(B) \leqq r(B^*)$. Whence $s \leqq t$, and part (a) then implies $s = t$, that is, the condition of the theorem is necessary.

(c) Assume θ is not identifiable. In view of part (a) it is required to show that $t < s$. By hypothesis, there exist random variables X^* and U^* and a matrix $B^* \varepsilon \vartheta^*$, such that $L^*(\vartheta^*, F_Y)$,

$$Y = X^*B^* + U^* \tag{3}$$

and $\mathcal{S}(\vartheta) \not\subset \mathcal{S}(\vartheta^*)$. Equations (1) and (3) and Lemma 3 imply

$$\psi_Y(t) = \psi_Y(tC') + f_1(t) = \psi_Y(tC^{*\prime}) + f_2(t), \tag{4}$$

where C and C^* are idempotent and respectively row-equivalent to B and B^* and have s and s^* rows which are not identically zero. There exists a nonsingular transformation P which reduces C^* to a diagonal matrix $D^* = C^*P$ having only ones and zeros on its main diagonal. Let $A = CP$, then (4) yields

$$\psi_Y(tA') = \psi_Y(tD^*) + f(t), \qquad \psi_Y(tD^*A') = \psi_Y(tD^*) + f(t),$$

$$\psi_Y(tA') = \psi_Y(tD^*A') + f(t). \tag{5}$$

Equation (5) will be analyzed in three cases.

CASE I, $r(D^*A') < s$. Let $G' = P'^{-1}D^*A'$. Then $r(G) < s$, and

$$\psi_Y(tC') = \psi_Y(tG') + f(t).$$

This, together with equation (4) and Lemma 5, implies $t \leqq r(G)$, so that $t < s$.

CASE II, $r(D^*A') = s$, and there exists a diagonal matrix D having only ones and zeros on the main diagonal such that

$$r(D) = s, \qquad r(DA') = s, \qquad r(DD^*) < s.$$

Substitution of $\sigma = tD$ for t in (5) gives

$$\psi_Y(\sigma DA') = \psi_Y(\sigma DD^*A') + f(\sigma).$$

The vector $\tau = \sigma DA'$ has exactly s components which are not identically zero. Since $r(DA') = s$, there exists a matrix α such that $r(\alpha) = s$ and $\sigma = \tau\alpha$. Hence

$$\psi_Y(\tau) = \psi_Y(\tau\alpha DD^*A') + f(\tau).$$

Since tA' has the same nonvanishing components as τ,

$$\psi_Y(tC') = \psi_Y(tH') + f(t)$$

where $H' = P'^{-1}A'\alpha DD^*A'$. This, together with equation (4) and Lemma 5, implies $t \leqq r(H)$, and since $r(G) \leqq r(DD^*) < s$, it follows $t < s$.

CASE III, $r(D^*A') = s$, and for every diagonal matrix D having ones and zeros on the main diagonal, $r(DD^*) = s$ whenever $r(D) = s$ and $r(DA') = s$. Let a_{i_j} for $j = 1, \cdots, m$ be the row vectors of A' which are not identically zero. Then by Lemma 1 each row a_{i_j} is included in some s-rowed minor of A' of rank s. That is, there exists a diagonal matrix D_j such that $r(D_jA') = s$ and $r(D_j) = s$ with elements $d_{i_ji_j} = 1$ for $j = 1, \cdots, m$. Since $r(D_jD^*) = s$, by hypothesis, then $d^*_{i_ji_j} = 1$ for $j = 1, \cdots, m$. Hence, it follows that $AD^* = A$. Since D^* is idempotent, then $\mathcal{S}^*(A) \subset \mathcal{S}^*(D^*)$. Here the notation $\mathcal{S}^*(A)$ denotes the space spanned by the row vectors of A. It then follows that $\mathcal{S}^*(C) \subset \mathcal{S}^*(C^*)$, and hence $\mathcal{S}(\vartheta) \subset \mathcal{S}(\vartheta^*)$, contradicting the hypothesis that θ is not identifiable (L^*). Case III is therefore impossible.

This completes the proof of sufficiency for Theorem 3, as these three cases exhaust the possible situations arising from (5). The corollaries below are easy consequences of Theorem 3.

COROLLARY 1. *Denoting XB by S, θ is not identifiable (L^*) if and only if $S = ZG + V$, where Z and V are independent, V has a multinormal distribution, and $r(G) < r(B)$.*

COROLLARY 2. *If $X = ZG + V$ and $r(G) < r(B)$, then θ is not identifiable (L^*).*

COROLLARY 3. *If θ is identifiable (L^*), then the nonnormality of X is not less than s.*

The expression of Theorem 3 in terms of the random variables X is more natural if the problem is reformulated in an equivalent way [8]. Let Y_n denote a vector with n components and B_{sn} a matrix with s rows and n columns, $s \leqq n$. Let $\mathfrak{X}_s$ denote a set of s-dimensional distribution functions.

THEOREM 4. *When the relation* L^{**} *is characterized by the equation* $Y_n = X_s B_{sn} + U_n$ *with* $r(B_{sn}) = s$, *then* θ *is identifiable* (L^{**}) *if and only if* $nn(\mathfrak{X}_s) = s$.

Taking $n = 2$ and $s = 1$, one obtains the result of Reiersøl [7]. Exactly similar results are obtained if the relation R is taken to be equality rather than inclusion. Again, if Θ_s were chosen as the set where elements are all the sets of row-equivalent matrices, so that $\Theta_s = \Theta$, then one would have:

THEOREM 5. θ *is identifiable* (L^*) *if no linear combination of the components of* X *is normally distributed.*

5. Estimation of linear structures. In this section, an estimate is constructed which converges with probability one to the linear structure. An infinite sequence of vector random variables $(Y_1, Y_2, \cdots)$ is considered. No assumption whatever is made concerning the existence of moments of Y_i. Each Y_i satisfies definitions 2 and 3; furthermore, Y_i and Y_j are independent if $i \neq j$. It is assumed that s is known. For every N, let $Z_N = (Y_1, \cdots, Y_N)$. A function $T_N(Z_N)$ will be constructed such that $P\{T_N(Z_N) \to \mathcal{S}(B) \text{ as } N \to \infty\} = 1$, where T_N is a linear vector space and the convergence $T_N(Z_N) \to \mathcal{S}(B)$ is defined in

DEFINITION 8. If C_N and C are linear vector spaces, then $C_N \to C$ as $N \to \infty$, provided $\Delta_N \to 0$ as $N \to \infty$, where $\Delta_N = \max_k \min_r |k - r|$ for all unit vectors k in C_N, and all unit vectors r in C. The quantity Δ_N will be called the *distance* between the sets C_N and C.

Hence if C_N and C are linear vector spaces and C_N a random variable, then C_N converges almost surely to C if $P\{\Delta_N \to 0\} = 1$. A unit vector is here a vector of length one.

DEFINITION 9. A matrix B is *related* to a random variable Y if $B \,\varepsilon\, \vartheta$ for some ϑ such that $\vartheta \,\varepsilon\, \Theta_s$ and $\vartheta \,\varepsilon\, \gamma(F)$ (cf. Definition 1).

From part (c) of Theorem 3 one obtains:

LEMMA 6. *If* $\psi_Y(t) = \psi_Y(tC_i') + f_i(t)$ *for* $i = 1, 2$, *where* C_i *is idempotent with rank* S_i, *then*

(i) *either* $nn(Y) < s$, *or else* $\mathcal{S}(C_1) \subset \mathcal{S}(C_2)$, *and*

(ii) *either* $nn(Y) < s_2$ *or else* $\mathcal{S}(C_2) \subset \mathcal{S}(C_1)$.

LEMMA 7. *If* θ *is identifiable* (L^*), *if* B *is related to* Y, *and if* G *is idempotent, then* $\psi_Y(t) = \psi_Y(tG') + f(t)$ *if and only if* $\mathcal{S}(B) \subset \mathcal{S}(C)$.

PROOF. Suppose $\mathcal{S}(B) \subset \mathcal{S}(G)$ and C is an idempotent matrix row-equivalent to B. Then $\psi_Y(t) = \psi_Y(tC') + f(t)$, and $r(C) = nn(Y)$. Since $CG = C$, then $\psi_Y(tG') = \psi_Y(tC') + f(t)$ and $\psi_Y(t) = \psi_Y(tG') + f(t)$.

Conversely, suppose $\psi_Y(t) = \psi_Y(tG') + f(t)$. Then, since $nn(Y) = r(C)$, Lemma 6 implies $\mathcal{S}(C) \subset \mathcal{S}(G)$.

Lemmas 4 and 7 imply

LEMMA 8. *If* θ *is identifiable* (L^*) *and* B *is related to* Y, *then, for any idempotent* G *such that* $\mathcal{S}(B) \subset \mathcal{S}(G)$, *there exists* F *idempotent and row-equivalent to* G *such that* $\psi_Y(t) = \psi_Y(tF') + \psi_Y(t(I - F)')$, *and* $Y(I - F)$ *has a multinomial distribution.*

LEMMA 9. *If F is idempotent with rank $n - 1$, then $F = I - r'a/ra'$ for unique row vectors r and a.*

From Lemmas 8 and 9, it follows that $\psi_Y(t) = \psi_Y(tF') + \frac{1}{2}\sigma^2(ta')^2$ if and only if $\mathcal{S}(B) \subset \mathcal{S}(F)$, where F is chosen as in Lemma 9. This property is made the basis of a criterion to determine $\mathcal{S}(B)$. Letting

$$L(t) = L(t; \quad a, r, \alpha) = \varphi_Y(t) = \varphi_Y(tF')\alpha^{-(ta')^2} \qquad \text{where } \alpha = \exp[-\tfrac{1}{2}\sigma^2],$$

it follows that $L(t) =_t 0$ if and only if $\mathcal{S}(B) \subset \mathcal{S}(F)$ and α is suitably chosen. Define $G(r) = \min_{a,\alpha} \int L(t)L(-t)\, d\lambda(t)$, where $\lambda(t)$ is a strictly increasing bounded function and the integration is taken over the entire space. Then $G(r) = 0$ if and only if r is orthogonal to $\mathcal{S}(B)$. Thus if F_Y were known, an investigation of the zeros of $G(r)$ would yield explicit knowledge of $\mathcal{S}(B)$.

Determining a random variable which converges almost surely to G will enable the desired estimate to be constructed. To this end the sample characteristic function is defined by

$$\varphi_N(t; Z_N) = \frac{1}{N} \sum_{k=1}^{N} e^{[-itY_k']}.$$

Then $G_N(r; \quad Z_N)$ is defined by replacing $\varphi_Y(t)$ by $\varphi_N(t; \quad Z_N)$ in the definition of $G(r)$. The space C is complementary to $S = \mathcal{S}(B)$, that is the space spanned by the unit vectors r for which $G(r) = 0$.

The estimate $T_N(Z_N)$ is defined to be the linear space orthogonal to the linear vector space C_N spanned by the vectors $k_1, k_2, \cdots, k_{n-s}$. The vectors k_j are defined by the following construction.

(i) k_1 is any unit vector for which $G_N(k_1; \quad Z_N) = \min_r G_N(r; \quad Z_N)$.

(ii) for $j = 2, \cdots, n - s$, k_j is any unit vector such that $G_N(k_j; \quad Z_N) = \min_{r \varepsilon O_j} G_N(r; \quad Z_N)$, where O_j is the linear space orthogonal to $k_1, \cdots, k_{j-1}$.

The proof that the estimate converges almost surely is based on the following lemma which is a corollary of a theorem of Rubin [10].

LEMMA 10. *For any finite cell T of Euclidean n-dimensional space,*

$$P\{\lim_{N\to\infty} \varphi_N(t; Z_N) = \varphi_Y(t) \quad \textit{uniformly for} \quad t \,\varepsilon\, T\} = 1.$$

Taking F as in Lemma 9 and since α^{-u^2} is bounded for $\alpha \,\varepsilon\, [0, 1]$ and u real, then

$$P\{\lim_{N\to\infty} L_N(t) = L(t) \quad \text{uniformly for} \quad t \,\varepsilon\, T \quad \text{and} \quad r, a, \alpha\} = 1.$$

Here $L_N(t)$ is defined by replacing $\varphi_Y(t)$ by $\varphi_N(t; \quad Z_N)$ in the definition of $L(t)$. From this, since r, a, α are on compact sets, it follows that

LEMMA 11. $P\{\lim_{N\to\infty} G_N(r; \quad Z_N) = G(r)$ *uniformly in* $r\} = 1$.

LEMMA 12. *If Δ_N is the distance between C_N and C, then $P\{\Delta_N \to 0$ as $N \to \infty\} = 1$, provided θ is identifiable.*

PROOF. For any $\tau > 0$, let C_τ be the set of unit vectors k such that $\min_{r \varepsilon C} |k - r| < \tau$, where r is a unit vector. For any $\eta > 0$, take

$$\tau = \eta n^{-1/2}, \qquad \xi = \min G(r) \text{ with } r \text{ not a member of } C_\tau \qquad \epsilon < \xi/2.$$

Then, there exists N_η such that for every r and all $N > N_\eta$, $| G_N(r; \quad Z_N) - G(r) | < \epsilon$ with probability one and, hence, both

$$\min G_N(r; \quad Z_N) \text{ with } r \text{ not a member of } C_\tau \geqq \xi - \epsilon > \xi/2$$

$$\min G_N(r; Z_N) \text{ with } r \text{ not a member of } C_\tau < \epsilon < \xi/2.$$

Therefore, if k satisfies $\min_r G_N(r; \quad Z_N) = G_N(k; \quad Z_N)$, it follows that $k \varepsilon C_\tau$, and hence $k_1 \varepsilon C_\tau$.

It can be shown similarly that, if $n - s \geqq 2$, then $k_2 \varepsilon C_\tau$, since in this case there must be a unit vector r such that $r \varepsilon k$ and $r \varepsilon C$. Likewise it can be shown by induction that $k_j \varepsilon C$ for $j = 1, \cdots, n - s$.

Let k be any unit vector in C_N. Then $k = \sum_{j=1}^{n-s} d_j k_j$ and $kk' = 1$ implies $\sum_{j=1}^{n-s} d_j^2 = 1$. Since $k_j \varepsilon C_\tau$ there are vectors $r_j \varepsilon C$ such that $|k_j - r_j| < \tau$ for $j = 1, \cdots, n - s$. Then

$$| k - 4 | < \tau \sum_{j=1}^{n-s} | d_j | \leqq \tau \sqrt{n - s} < \eta.$$

Hence, for any η there exists N_η such that $C_N \subset C_\eta$, provided $N > N_\eta$, and therefore $\Delta_N < \eta$ with probability one.

The following lemma is straightforward.

LEMMA 13. *C_N converges to C if and only if S_N converges to S, where C_N and C are the complements of S_N and S, respectively.*

Lemmas 12 and 13 then imply

THEOREM 6. *If θ is identifiable (L^*), then the estimate $T_N(Z_N)$ converges almost surely to $\mathcal{S}(B)$.*

ACKNOWLEDGMENT. I wish to acknowledge my debt to Professor J. Neyman for suggesting this problem and for his constant encouragement and stimulation during the development of this paper.

REFERENCES

[1] J. NEYMAN, "Existence of consistent estimates of the directional parameter in a linear structural relation between two variables," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 497–512.

[2] TJALLING C. KOOPMANS, "Identification problems in economic model construction," *Econometrica*, Vol. 17 (1949), pp. 125–144.

[3] KARL PEARSON, "On lines and planes of closest fit to systems of points in space," *Philos. Mag.* Series 6, Vol. 2 (1901), pp. 559–572.

[4] GODFREY H. THOMSON, "A hierarchy without a general factor," *British Journal of Psych.*, Vol. 8 (1916), pp. 271–281.

[5] J. NEYMAN, "Remarks on a paper by E. C. Rhodes," *Jour. Roy. Stat. Soc.*, Vol. 100 (1939), pp. 50–57.

[6] ABRAHAM WALD, "The fitting of straight lines if both variables are subject to error," *Ann. Math. Stat.*, Vol. 11 (1940), pp. 284–300.

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

[8] T. A. JEEVES, "Identifiability and almost sure estimability of linear structures in n-dimensions," University of California, 1952, unpublished thesis.

[9] T. C. KOOPMANS AND O. REIERSØL, "The identification of structural characteristics," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 165–181.

[10] HERMAN RUBIN, "Uniform convergence of random functions in the applications to statistics," Technical Report No. 7, Contract No. N6-onr-251 Task III (1951), Department of Statistics, Stanford University.