THE MINIMUM DISTANCE METHOD[1]

By J. Wolfowitz

Cornell University

1. Summary and Introduction. The present paper gives the formal statements and proofs of the results illustrated in [1]. In a series of papers ([2], [3], [4]) the present author has been developing the minimum distance method for obtaining strongly consistent estimators (i.e., estimators which converge with probability one). The method of the present paper is much superior, in simplicity and generality of application, to the methods used in the papers [2] and [4] cited above. Roughly speaking, the present paper can be summarized by saying that, in many stochastic structures where the distribution function (d.f.) depends continuously upon the parameters and d.f.'s of the chance variables in the structure, those parameters and d.f.'s which are identified (uniquely determined by the d.f. of the structure) can be strongly consistently estimated by the minimum distance method of the present paper. Since identification is obviously a necessary condition for estimation by *any* method, it follows that, in many actual statistical problems, identification implies estimatability by the method of the present paper.

Thus problems of long standing like that of Section 5 below are easily solved. For this problem the whole canonical complex (Section 6 below; see [1]) has never, to the author's knowledge, been estimated by any other method. The directional parameter of the structure of Section 4 seems to be here estimated for the first time.

As the identification problem is solved for additional structures it will be possible to apply the minimum distance method. The proofs in the present paper are of the simplest and most elementary sort.

In Section 8 we treat a problem in estimation for nonparametric stochastic difference equations. Here the observed chance variables are not independent, but the minimum distance method is still applicable. The treatment is incomparably simpler than that of [4], where this and several other such problems are treated. The present method can be applied to the other problems as well.

Application of the present method is routine in each problem as soon as the identification question is disposed of. In this respect it compares favorably with the method of [4], whose application was far from routine.

As we have emphasized in [1], the present method can be applied with very many definitions of distance (this is also true of the earlier versions of the minimum distance method). The definition used in the present paper has the convenience of making a certain space conditionally compact and thus eliminating the need for certain circumlocutions. Since no reason is known at present for

Received January 31, 1956.

[1] Research under contract with the Office of Naval Research.

preferring one definition of distance to another we have adopted a convenient definition. It is a problem of great interest to decide which, if any, definition of distance yields estimators preferable in some sense. The definition of distance used in this paper was employed in [9].

As the problem is formulated in Section 2 below (see especially equation (2.1), the "observed" chance variables $\{X_i\}$ are known functions (right members of (2.1)) of the "unobservable" chance variables $\{Y_i\}$ and of the unknown constants $\{\theta_i\}$. In the problems treated in [3], [9], and [11], it is the distribution of the observed chance variables which is a known function of unobservable chance variables and of unknown constants, and not the observed chance variables themselves. However, the latter problems can easily be put in the same form as the former problems. Moreover, in the method described below the values of the observed chance variables are used only to estimate the distribution function of the observed chance variables (by means of the empiric distribution function). Consequently there is no difference whatever in the treatment of the problems by the minimum distance method, no matter how the problems are formulated.

The unobservable chance variables $\{Y_i\}$ correspond to what in [11] are called "incidental parameters"; the unknown constants $\{\theta_i\}$ are called in [11] "structural parameters". In [9] there is a discussion of the fact that in some problems treated in the literature the incidental parameters are considered as constants and in other problems as chance variables. In contradistinction to the present paper [3] (in particular its Section 5) treats the incidental parameters as unknown constants. The fundamental idea of both papers is the same: The estimator is chosen to be such a function of the observed chance variables that the d.f. of the observed chance variables (when the estimator is put in place of the parameters and distributions being estimated) is "closest" to the empiric d.f. of the observed chance variables. The details of application are perhaps easier in the present paper; the problems are different and of interest per se.

2. The minimum distance method. Let m, m', k, k', be integers such that $0 \leqq m \leqq m'$, $0 \leqq k \leqq k'$. For $j = 1, 2, \cdots$ ad inf. let $(Y_{j1}, \cdots, Y_{jk'})$ be independent, identically distributed vector chance variables with the common d.f. G_0 which is unknown to the statistician. The constants $\theta_1, \cdots, \theta_{m'}$, are also unknown to the statistician. It is known that, for $j = 1, 2, \cdots$ ad inf.,

$$X_{ji} = t_i(Y_{j1}, \cdots, Y_{jk'}, \theta_1, \cdots, \theta_{m'}) \qquad i = 1, \cdots, h \tag{2.1}$$

where the t_i, for $i = 1, \cdots, h$, are known Borel-measurable functions of the arguments exhibited. Define the common d.f. of $(Y_{j1}, \cdots, Y_{jk})$, $j = 1, 2, \cdots$ ad. inf., by

$$G(y_1, \cdots, y_k) = G_0(y_1, \cdots, y_k, +\infty, \cdots, +\infty).$$

Let $\theta = (\theta_1, \cdots, \theta_{m'})$. Let $A = \{(\bar{\alpha}, g)\}$ be a space of couples $(\bar{\alpha}, g)$, the first member of which is a real m'-dimensional vector $(\alpha_1, \cdots, \alpha_{m'})$, and the second member of which is a k'-dimensional d.f. It is known that (θ, G_0) is in A. On A we define a metric δ as follows:

$$\delta([\bar{\alpha}_1, g_1], [\bar{\alpha}_2, g_2]) = \sum_{j=1}^{m'} | \text{arc tan } \alpha_{1j} - \text{arc tan } \alpha_{2j} | + \int | g_1(z) - g_2(z) | e^{-|z|} dz_1 \cdots dz_{k'} \tag{2.2}$$

where

$$\bar{\alpha}_i = \alpha_{i1}, \cdots, \alpha_{im'} \qquad i = 1, 2$$

$$z = z_1, \cdots, z_{k'}$$

We shall also use δ to denote a metric on any Euclidean space or on any space of d.f.'s of the same dimensionality. In that case δ is to be understood as the expression corresponding, respectively, to the first or second term of the right member of (2.2).

Our problem is to give (strongly) consistent estimators of G and $(\theta_1, \cdots, \theta_m)$, i.e., for $n = 1, 2, \cdots$ ad inf., to construct measurable functions (Q_{n1}, Q_{n2}) from hn-dimensional Euclidean space (of $X_{11}, \cdots, X_{1h}, \cdots, X_{n1}, \cdots, X_{nh}$) to A such that, *whatever be* (θ, G_0) (in A), we have, with probability one (w.p. 1), both

$$Q_{n1}^{(j)} \to \theta_j, \qquad j = 1, \cdots, m$$

where $Q_{n1}^{(j)}$ is the jth component of Q_{n1}, and

$$Q_{n2}(y_1, \cdots, y_k, +\infty, \cdots, +\infty) \to G(y_1, \cdots, y_k)$$

at every point of continuity of the latter.

Let $J(\bar{\alpha}, g)$ be the (h-dimensional) d.f. of $(X_{j1}, \cdots, X_{jh})$ when $\theta = \bar{\alpha}$ and $G_0 = g$. In this notation the generic point in h-space is suppressed since it will rarely come explicitly into play, and the emphasis is on the fact that this is a transformation from A into the space of h-dimensional d.f.'s. We shall make the following *Identification and Continuity (I.C.) Assumption*:

Let $\{\bar{\alpha}_i, g_i\}$ *be any Cauchy sequence (i.e., as* $i \to \infty$, $\delta[\bar{\alpha}_i, g_i], [\bar{\alpha}_{i+n}, g_{i+n}]) \to 0$ *uniformly in* n*) such that*

$$\delta(J(\bar{\alpha}_i, g_i), J(\theta, G_0)) \to 0 \tag{2.3}$$

as $i \to \infty$. *Then, as* $i \to \infty$,

$$\alpha_{ij} \to \theta_j, \qquad j = 1, \cdots, m \tag{2.4}$$

(a_{ij} is the jth component of $\bar{\alpha}_i$) and

$$g_i(y_1, \cdots, y_k, +\infty, \cdots, +\infty) \to G(y_1, \cdots, y_k) \tag{2.5}$$

at every point of continuity of the latter.

Let

$$C_n = \{(X_{j1}, \cdots, X_{jh}), j = 1, \cdots, n\}$$

and $F_n(C_n)$ be the empiric d.f. of C_n, i.e., an h-dimensional d.f. such that its value at $(y_1, \cdots, y_h)$ is $1/n$ times the number of elements in C_n for which $X_{ji} < y_i$, $i = 1, \cdots, h$. Let $\gamma(n)$ be any positive function of n which approaches zero as $n \to \infty$. Let $S_n(C_n) = (\theta_n^*, G_{on}^*)$ be any function from the hn-dimensional Euclidean space of C_n to the space A which is measurable and such that

$$\delta(J(\theta_n^*, G_{0n}^*), F_n(C_n)) < \inf_{(\bar{\alpha}, g)\varepsilon A} \delta(J(\bar{\alpha}, g), F_n(C_n)) + \gamma(n) \tag{2.6}$$

$S_n(C_n)$ is a minimum distance estimator, for which the following holds:

THEOREM. *If the I.C. Assumption holds, then, with probability one (w.p. 1),*

$$\theta_{nj}^* \to \theta_j, \qquad j = 1, \cdots, m \tag{2.7}$$

(θ_{nj}^* is the jth *component of θ_n^*) and*

$$G_{0n}^*(y_1, \cdots, y_k, +\infty, \cdots, +\infty) \to G(y_1, \cdots, y_k) \tag{2.8}$$

at every continuity point of the latter.

(In view of (2.7) and (2.8) it is actually the appropriate components of $S_n(C_n)$ which could be called minimum distance estimators of the θ_j and G).

PROOF: By the Glivenko-Cantelli theorem we have that, w.p.1,

$$\delta(F_n(C_n), J(\theta, G_0)) \to 0 \tag{2.9}$$

as $n \to \infty$. Hence, w.p.1,

$$\inf_{(\bar{\alpha}, g)\varepsilon A} \delta(F_n(C_n), J(\bar{\alpha}, g)) \to 0 \tag{2.10}$$

Hence, w.p.1,

$$\delta(J(\theta_n^*, G_{on}^*), J(\theta, G_0)) \to 0. \tag{2.11}$$

Since the space A is (sequentially) conditionally compact (with respect to the metric δ) it follows that, at every sample point and from every subsequence of $S_n(C_n)$, $n = 1, 2, \cdots$, we can select a Cauchy subsequence. For every such sequence the relation corresponding to (2.11) holds, except on a set of sample points of probability zero. When the relation corresponding to (2.11) holds, then, by the I.C. Assumption, relations corresponding to (2.7) and (2.8) hold. Thus, we have proved that, except on a set of sample points of probability zero, *every* subsequence of $S_n(C_n)$ contains a subsequence for which the equations corresponding to (2.7) and (2.8) hold. But this easily implies the theorem.

3. Discussion of the Identification and Continuity Assumption. We have seen that the proof of the strong consistency of the minimum distance estimator follows almost trivially from the I.C. Assumption. Let us now examine this assumption more carefully.

The constants $(\theta_1, \cdots, \theta_m)$ and the d.f. G, which belong to the "structure" (system) (2.1), are said to be "identified in A" if, when $(\bar{\alpha}, g)$ is in A, and

$$J(\theta, G_0) = J(\bar{\alpha}, g) \tag{3.1}$$

identically in the h arguments, then

$$\theta_i = \alpha_i \qquad i = 1, \cdots, m \tag{3.2}$$

$$G(y_1, \cdots, y_k) = g(y_1, \cdots, y_k, +\infty, \cdots, +\infty) \tag{3.3}$$

identically in $y_1, \cdots, y_k$ (of course (θ, G_0) is in A). It is obvious that identification in A is an indispensable condition for our problem of estimating consistently the constants $\theta_1, \cdots, \theta_m$ and the d.f. G, for no particular value of the sequence

$$\{(X_{j1}, \cdots, X_{jh}), j = 1, 2, \cdots\}$$

can furnish more information than the function $J(\theta, G_0)$ itself.

In most, if not all, actual statistical problems, J will be a continuous function on A, i.e., whenever

$$(\bar{\alpha}_i, g_i) \to (\bar{\alpha}, g) \text{ in } A,$$

then

$$\delta(J(\bar{\alpha}_i, g_i), J(\bar{\alpha}, g)) \to 0. \tag{3.4}$$

We shall assume that this is so in the remainder of this section. Then the following considerations will help to understand the I.C. Assumption and to furnish a convenient way of proving that it is satisfied.

Let $\bar{C}_1$ be the map of A under J, i.e.,

$$\bar{C}_1 = \{J(\bar{\alpha}, g), (\bar{\alpha}, g) \varepsilon A\}.$$

Let $\{\bar{\alpha}'_i, g'_i\}$ be *any* Cauchy sequence in A which does *not* have a limit in A, and for which

$$J(\{\bar{\alpha}'_i, g'_i\}) = \lim_{i \to \infty} J(\bar{\alpha}'_i, g'_i)$$

exists. Let $\bar{C}_2$ be the totality of all such $J(\{\bar{\alpha}'_i, g'_i\})$. ($\bar{C}_1$ and $\bar{C}_2$ need not be disjoint).

The *indispensable* condition of identification[2] in A may be stated as follows: If

$$(\bar{\alpha}_i, g_i) \to (\bar{\alpha}, g) \text{ in } A \tag{3.5}$$

and

$$J(\bar{\alpha}_i, g_i) \to J(\theta, G_0) \tag{3.6}$$

then

$$\alpha_j = \theta_j, \qquad j = 1, \cdots, m \tag{3.7}$$

$$g(y_1, \cdots, y_k, +\infty, \cdots, +\infty) = G(y_1, \cdots, y_k) \tag{3.8}$$

identically in $y_1, \cdots, y_k$. If *either* of the two following conditions is *also* met the I.C. Assumption is fulfilled:

---

[2] We remind the reader that J is assumed to be continuous on A.

A) $J(\theta, G_0)$ is not a member of $\bar{C}_2$

B) If $J(\theta, G_0) = J(\{\bar{\alpha}_i', g_i'\})$ is[3] in $\bar{C}_2$, then $\alpha_{ij}' \to \theta_j$, $j = 1, \cdots, m$, and $g_i'(y_1, \cdots, y_k, +\infty, \cdots, +\infty) \to G(y_1, \cdots, y_k)$ at every point of continuity of the latter.

Thus the I.C. Assumption is, for most A to be encountered in actual problems where J is continuous, not much more onerous than the indispensable identification condition. In the important examples to be discussed below condition A, and hence the I.C. Assumption, will hold.

4. A linear relationship between two chance variables subject to independent errors. We illustrate the last two sections by application to the following very important structure: Suppose it is known to the statistician that, for $j = 1, 2, \cdots$, ad inf.,

$$X_{j1} = \xi_j + v_{j1} \tag{4.1}$$

$$X_{j2} = \alpha + \beta\xi_j + v_{j2} \tag{4.2}$$

where α and β are constants[4] unknown to the statistician, and $\{v_{j1}\}$, $\{v_{j2}\}$, and $\{\xi_j\}$ are sequences of independent, identically distributed chance variables, with respective d.f.'s L_1, L_2, L_3, say, which are unknown to the statistician. The different sequences are known to be independent of each other. We shall consider first the problem of estimating β.

Let d be the generic designation of a complex

$$\{a, b, p_1, p_2, p_3\}$$

whose first two elements are real numbers, and whose last three elements are one-dimensional d.f.'s. Let

$$d_0 = \{\alpha, \beta, L_1, L_2, L_3\}$$

The symbol $J(d)$ will have the same meaning as in Section 2.

We shall assume that A is the totality of all d's such that g_3 is not a normal d.f.; for the purposes of this definition and elsewhere in this paper a d.f. which assigns probability one to a single point is to be considered normal (with variance zero). It was proved by Reiersol [5] that β is identified in A; actually an examination of his proof (especially equation (4.5)) shows that Reiersol proved somewhat more, namely that, if

$$J(d_0) = J([a^0, b^0, p_1^0, p_2^0, p_3^0]) \tag{4.3}$$

---

[3] The preceding symbol has been defined in the first displayed equation which precedes (3.5).

[4] This formulation does not include the case when the regression line is parallel to the axis of X_2, i.e., when $\beta = \infty$; in that case $X_1 = \text{constant} + v_1$, $X_2 = \xi + v_2$. This omission is made only in the interest of simplicity. We invite the reader to verify that the formulation where this case is also a possibility can be treated by the methods of the present paper in exactly the same way as this is done in Sections 4, 5, and 6.

where a^0 and b^0 are finite, p_1^0 and p_2^0 are d.f.'s, but the d.f. p_3^0 is not *required* to be not normal, then $b^0 = \beta$ and p_3^0 *must* be not normal. Thus $(a^0, b^0, p_1^0, p_2^0, p_3^0)$ *is* in A.

It is obvious that $J(d)$ is a continuous function of d (on A).

Let $\{d_i = (a_i, b_i, p_{i1}, p_{i2}, p_{i3}), i = 1, 2, \cdots\}$ be *any* Cauchy sequence in A which does *not* have a limit in A and which is such that

$$J(\{d_i\}) = \lim_{i\to\infty} J(d_i) \tag{4.4}$$

exists. Let $d^* = (a^*, b^*, p_1^*, p_2^*, p_3^*)$ be such that $\delta(d_i, d^*) \to 0$. Then at least one of the following four properties must hold:

1) p_3^* is a normal d.f., a^* and b^* are both finite, and p_1^* and p_2^* have variation one.

2) p_3^* is a normal d.f., either $a^* = \pm\infty$ or $b^* = \pm\infty$ or both, and p_1^* and p_2^* have variation one.

3) p_3^* is a non-normal d.f. (therefore has variation one), either $a^* = \pm\infty$ or $b^* = \pm\infty$ or both, and both p_1^* and p_2^* have variation one.

4) the variation of at least one of p_1^*, p_2^*, p_3^* is less than one.

We shall now show that the I.C. Assumption is satisfied in the present probent problem. We shall try first to show that $J(d_0)$ is not in $\bar{C}_2$; we will be able to to achieve this except for one obstacle which we will treat somewhat differently. Suppose then that $J(d_0) = J(\{d_i\})$ were[5] in $\bar{C}_2$. Then d^* could not have the first of the above properties, because of Reiersol's result cited above. If d^* had property 2 above then either the variation of $J(\{d_i\})$ would be less than one (which cannot be true of $J(d_0)$) or else $J(\{d_i\})$ is the same as $J(a'', b'', p_1'', p_2'', p_3''])$, where a'' and b'' are finite, p_3'' assigns probability one to a single point and is therefore normal, and p_1'' and p_2'' have variation one (which cannot be true of $J(d_0)$ because of Reiersol's result cited above). If d^* had property 3 above then $J(\{d_i\})$ would be of variation less than one, which cannot be true of $J(d_0)$. If d^* had property 4 above then either $J(\{d_i\})$ has variation less than one (in which case $J(d_0)$ is not in $\bar{C}_2$) or else $J(\{d_i\}) = J(d_0)$ is in $\bar{C}_2$! To see how this can happen we note that, if z and z' are any real numbers,

$$\xi_j + v_{j1} = (\xi_j + z) + (v_{j1} - z)$$

$$\alpha + \beta\xi_j + v_{j2} = (\alpha + z' - \beta z) + \beta(\xi_j + z) + (v_{j2} - z').$$

If either z or z' or both approach $\pm\infty$ the variations of some or all of p_1^*, p_2^*, p_3^* will be zero. This difficulty is easily overcome. One can show that in this case condition B of Section 3 holds. A method which is essentially the same but formally simpler is the following: Without changing the problem or any loss of generality we can reduce the set A so as to prevent the occurrence of this case. We simply define A as the totality of all d's which, in addition to the conditions previously imposed, satisfy the requirement that the smallest median of both p_1

[5] The preceding symbol was defined in (4.4).

and p_2 is zero. It is clear that the estimation of β is not affected by this additional restriction, and that, under this restriction, $J(d_0)$ cannot be in $\bar{C}_2$. (The definition of d_0, but not the value of β, may be affected by this restriction).

It is obvious that, unless the space A is suitably reduced, the parameter α cannot be identified. In [5] Reiersol states the result that, if the space A is that subset of the originally defined $A = \{d\}$ where the d's are subject to the further restriction

$$\text{median of } p_1 = \text{median of } p_2 = 0, \tag{4.5}$$

then α is also identified on (the new) A. It seems to the writer that one must make precise which median is meant in order to make the proof of [5] go through. Either of the following conditions, for example, will permit the proof of [5] to go through:

$$p_1 \text{ and } p_2 \text{ each have zero as the unique median} \tag{4.6}$$

$$p_1 \text{ and } p_2 \text{ have zero as the smallest (largest) median.} \tag{4.7}$$

What will suffice is a condition such that, if $p_1(x)$, $p_2(x)$ are the third and fourth components of a point in A, $p_1(x + c_1)$, $p_2(x + c_2)$ cannot be the third and fourth elements of any point in A unless $c_1 = c_2 = 0$.

Suppose, for example, one adopts the restriction (4.6) above. Then α is identified on (the new) A, by the result of [5]. Let $\bar{A}$ be the totality of limit points of A which are not in A. $\bar{A}$ will include points whose third and fourth elements will not have zero as a *unique* median. In order to show, *just as before*, that $J(d_0)$ is not in $\bar{C}_2$, we need the additional result analogous to the one implicit in [5] about β and cited above, namely that, if $J(d_0) = J(\{d_i\})$, then the first element of d^* is α. However, this result does not seem to be implicit in [5] under the condition (4.6), and a stronger condition may be needed.

5. A linear relationship between two chance variables whose errors are jointly normally distributed. The following structure is a very famous one with a long history of study (see [2] and [5], for example). Let it be known to the statistician that X_{j1} and X_{j2} satisfy (4.1) and (4.2) respectively, that α and β are unknown constants,[6] that the two sequences $\{\xi_j\}$ and $\{v_{j1}, v_{j2})\}$, $j = 1, 2, \cdots$, ad inf., of independent and identically distributed chance variables are distributed independently of each other, and that the common distribution of $\{(v_{j1}, v_{j2})\}$ is normal, with zero means and covariances $\sigma_{11}(= E(v_{j1})^2)$, $\sigma_{12}(= E(v_{j1}v_{j2}))$, and $\sigma_{22}(= E(v_{j2}^2))$, unknown to the statistician. Designate the common unknown d.f. of $\{\xi_j\}$ by L.

Let m be the generic designation of a complex

$$(a, b, c_{11}, c_{12}, c_{22}, l)$$

[6] See footnote 4.

such that a, b are real numbers, c_{11}, c_{12}, and c_{22} are real numbers such that the matrix

$$\begin{Bmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{Bmatrix}$$

is non-negative definite, and l is a one-dimensional non-normal d.f. Let A be the totality of all m. It follows from the results of Reiersol ([5]) and the Cramér-Lévy theorem ([7], p. 52, Th. 19) that, if

$$J(\mu) = J(m^0)$$

where

$$\mu = (\alpha, \beta, \sigma_{11}, \sigma_{12}, \sigma_{22}, L),$$
$$m^0 = (a^0, b^0, c_{11}^0, c_{12}^0, c_{22}^0, l^0),$$

μ is of course in A, and m^0 satisfies all the requirements imposed on the elements of A except that l^0 is not *required* to be not normal, then l^0 *must* be not normal and $\alpha = a^0$, $\beta = b^0$.

Obviously $J(m)$ is a continuous function of m on A. We will show that condition A of Section 3 is satisfied, so that the I.C. Assumption is fulfilled, and the minimum distance estimator of α and β is strongly consistent.

Let $\bar{A}$ for the present problem be defined as in Section 4. If a point $m^{00} = (a^{00}, b^{00}, c_{11}^{00}, c_{12}^{0}, c_{22}^{00}, l^{00})$ is in $\bar{A}$, and, for a sequence $\{m_i\}$ in A, $m_i \to m^{00}$ and $J(m_i) \to J(\{m_i\})$ in $\bar{C}_2$, at least one of the following must be true:

1) l^{00} is a normal d.f., and c_{11}^{00} and c_{22}^{00} are finite
2) l^{00} is a d.f. which is not normal, and either $a^{00} = \pm\infty$ or $b^{00} = \pm\infty$ or both
3) either $c_{11}^{00} = \infty$ or $c_{22}^{00} = \infty$ or both
4) the variation of l^{00} is less than one

Suppose $J(\mu)$ were in $\bar{C}_2$ and $= J(\{m_i\})$. If the first of the conditions above held then either $J(\{m_i\})$ would be of variation less than one, or $J(\{m_i\})$ would be normal, neither of which can be true of $J(\mu)$. If one of conditions 2, 3, and 4 held, then the variation of $J(\{m_i\})$ would be less than one, which of course cannot be true of $J(\mu)$. This completes the proof that $J(\mu)$ is not in $\bar{C}_2$.

6. Estimation of the remainder of the structure of Section 5. Let $H(y)$ be any one-dimensional d.f. The Gaussian component of H is the largest value of λ for which H can be expressed as the convolution of a normal d.f. with variance λ, and another d.f. H is said to have no Gaussian component if its Gaussian component is zero.

The elements σ_{11}, σ_{12}, σ_{22}, L of μ are not, in general, uniquely determined by $J(\mu)$. Among the, in general, infinitely many m such that $J(m) = J(\mu)$, there is exactly one, say

$$\mu_0 = (\alpha, \beta, \sigma_{11}^0, \sigma_{12}^0, \sigma_{22}^0, L_0),$$

which is such that L_0 has no Gaussian component. We have $\sigma_{11}^0 + \sigma_{22}^0 > c_{11} + c_{22}$ for any other complex m such that $J(m) = J(\mu_0) = J(\mu)$; all such complexes are readily determinable from μ_0. These remarks follow from (4.1), (4.2), and the results of Reiersol [5]. We shall call the complex μ_0 "canonical" and estimate *all* its components in a strongly consistent manner. Of course α and β have already been estimated in Section 5; the present method will also estimate them, inter alia.

Let $Z_1, \cdots, Z_n$ be any independent chance variables with the common df. $H(z)$ and the empiric d.f. $H_n(z)$. Let $d(n)$ be any positive function defined on the positive integers such that $d(n) \to 0$ as $n \to \infty$ and

$$P\{\delta(H(z), H_n(z)) > d(n) \text{ for infinitely many } n\} = 0.$$

There are many such functions; it is easy to verify that $n^{-1/10}$ is such a function, but this is a crude result. For $H(z)$ continuous and one-dimensional and δ the Fréchet distance between two d.f.'s there is available the sharp result of Chung [6], according to which the function

$$\left(\frac{\log \log n}{cn}\right)^{1/2}$$

is a function $d(n)$ if $0 < c < 2$.

Let

$$U(C_n) = (a^*(n), b^*(n), c_{11}^*(n), c_{12}^*(n), c_{22}^*(n), L_n^*) \tag{6.1}$$

be any function from the $2n$-dimensional space of C_n to the space $A = \{m\}$ which is measurable and such that

$$\delta(J(U(C_n)), F_n(C_n)) < d(n) \tag{6.2}$$

and

$$c_{11}^*(n) + c_{22}^*(n) + \gamma(n) > \sup (c_{11} + c_{22}) \tag{6.3}$$

where the supremum operation in the right member is performed over all m such that

$$\delta(J(m), F_n(C_n)) < d(n). \tag{6.4}$$

When there is no m which satisfies (6.4) let $U(C_n)$ be defined in any manner provided it is measurable. It will follow from the general considerations of the next section that

$$\delta(U(C_n), \mu_0) \to 0 \tag{6.5}$$

w.p.1. Thus the elements of $U(C_n)$ are strongly consistent estimators of the elements of the canonical complex.

7. The method of the maximum index. We shall now generalize the considerations of the preceding section.

Consider the structure (2.1), and the totality of $(\bar{\alpha}, g)$ in A such that $J(\bar{\alpha}, g) =$

$J(\bar{\alpha}^*, G^*)$; call this totality $T(\bar{\alpha}^*, G^*)$. In every $T(\bar{\alpha}, g)$ let there be defined a unique member called the canonical complex of $T(\bar{\alpha}, g)$; we may denote this element by $D(\bar{\alpha}, g)$. If $(\bar{\alpha}_1, g_1)$ and $(\bar{\alpha}_2, g_2)$ are such that $J(\bar{\alpha}_1, g_1) = J(\bar{\alpha}_2, g_2)$, then we must have $D(\bar{\alpha}_1, g_1) = D(\bar{\alpha}_2, g_2)$. Suppose that there is defined on A a real-valued function $\psi(\bar{\alpha}, g)$ such that, whenever

$$(\bar{\alpha}_i, g_i) \rightarrow (\bar{\alpha}, g) \text{ in } A, \tag{7.1}$$

then

$$\liminf_{i \rightarrow \infty} \psi(\bar{\alpha}_i, g_i) \leqq \psi(\bar{\alpha}, g) \tag{7.2}$$

and such that, whenever $(\bar{\alpha}^*, g^*)$ is a canonical complex,

$$\psi(\bar{\alpha}^*, g^*) > \psi(\bar{\alpha}, g) \tag{7.3}$$

for every other $(\bar{\alpha}, g)$ in $T(\bar{\alpha}^*, g^*)$.

Let $d(n)$ be any function defined on the positive integers such that $d(n) \rightarrow 0$ as $n \rightarrow \infty$ and

$$P\{\delta(J(\theta, G_0), F_n(C_n)) > d(n) \text{ for infinitely many } n\} = 0 \tag{7.4}$$

Let $U(C_n) = (\theta_n^{**}, G_{0n}^{**})$ be any function from the hn-dimensional Euclidean space of C_n to the space A, which is measurable and such that

$$\delta(J(\theta_n^{**}, G_{0n}^{**}), F_n(C_n)) < d(n) \tag{7.5}$$

and

$$\psi(\theta_n^{**}, G_{0n}^{**}) + \gamma(n) > \sup \psi(\bar{\alpha}, g) \tag{7.6}$$

where the supremum in the right member is over all $(\bar{\alpha}, g)$ such that

$$\delta(J(\bar{\alpha}, g), F_n(C_n)) < d(n). \tag{7.7}$$

When there is no $(\bar{\alpha}, g)$ which satisfies (7.7) let $(\theta_n^{**}, G_{0n}^{**})$ be defined in any manner provided it is measurable. We will call $U(C_n)$ a maximum index estimator (of $D(\theta, G_0)$) and prove the following

THEOREM. *If J is a continuous function on A, i.e., whenever $(\bar{\alpha}_i, g_i) \rightarrow (\bar{\alpha}, g)$ in A, $J(\bar{\alpha}_i, g_i) \rightarrow J(\bar{\alpha}, g)$, and if $J(\theta, G_0)$ is not in $\bar{C}_2$, then, w.p.1,*

$$\delta(U(C_n), D(\theta, G_0)) \rightarrow 0, \tag{7.8}$$

so that $U(C_n)$ is a strongly consistent estimator of $D(\theta, G_0)$.

PROOF: Obviously $\delta(J(\theta_n^{**}, G_{0n}^{**}), F_n(C_n)) \rightarrow 0$, w.p. 1. Hence

$$\delta(J(\theta_n^{**}, G_{0n}^{**}), J(D[\theta, G_0])) \rightarrow 0, \qquad \text{w.p.1.}$$

If (7.8) were not true w.p.1, then, with positive probability, we may choose a Cauchy subsequence (A is conditionally compact; the particular sequence may depend upon the sample point in the probability space) which converges to a point $(\bar{\alpha}', g')$ in A (since $J(\theta, G_0)$ is not in $\bar{C}_2$) and $(\bar{\alpha}', g')$ is not $D[\theta, G_0]$.

It is impossible that, with positive probability,

$$\psi(D[\theta, G_0]) > \psi(\bar{\alpha}', g'), \tag{7.9}$$

because of (7.2) and the fact that, w.p.1, $\delta(J(D[\theta, G_0]), F_n(C_n))$ is eventually less than $d(n)$.

Suppose that, with positive probability,

$$\psi(D[\theta, G_0]) < \psi(\bar{\alpha}', g'). \tag{7.10}$$

Then $(\bar{\alpha}', g')$ would not be in $T(\theta, G_0)$ and $J(\bar{\alpha}', g')$ and $J(\theta, G_0)$ would not be identical. Since J is continuous we must have that, for the Cauchy subsequence, $\lim_{i\to\infty} J(\theta^{**}_{n_i}, G^{**}_{0n_i}) = J(\bar{\alpha}', g')$. Since $\delta(J(\theta^{**}_n, G^{**}_{0n}), J(\theta, G_0)) \to 0$ w.p.1, it follows that $J(\bar{\alpha}', g')$ and $J(\theta, G_0)$ are identical, contradicting the above. Hence (7.10) cannot occur.

Suppose then that, with positive probability,

$$\psi(D[\theta, G_0]) = \psi(\bar{\alpha}', g') \tag{7.11}$$

but $(\bar{\alpha}', g')$ were not $D(\theta, G_0)$. Because of the maximizing property of ψ (on each T) it would follow that $(\bar{\alpha}', g')$ is not in $T(\theta, G_0)$. But then $J(\bar{\alpha}', g')$ and $J(\theta, G_0)$ could not be identical. We have already seen that this cannot be. This leaves, as the only remaining possibility, that $(\bar{\alpha}', g')$ is $D(\theta, G_0)$, a contradiction which proves the theorem.

It is easy to verify that the postulated conditions are verified in the problem of Section 6. We have already seen that there $J(\theta, G_0)$ is not in $\bar{C}_2$. Let

$$\psi(a, b, c_{11}, c_{12}, c_{22}, l) = c_{11} + c_{22}.$$

Then, in any $T(m)$, ψ attains its unique maximum on the canonical comple: The function ψ is obviously continuous on A. Thus it satisfies the requirements of the theorem of the present section.

8. Application to stochastic difference equations. Let it be known to the statistician that $u_0, u_1, u_2, \cdots$ are independent chance variables with the common one-dimensional d.f. G, which is unknown to the statistician. Also it is known that, for $j = 1, 2, \cdots$

$$X_j = u_j + \alpha u_{j-1} \tag{8.1}$$

where α is a constant less than one in absolute value but otherwise unknown to the statistician. The problem is to estimate α consistently, under minimal assumptions on G.

Let q be the generic designation of a couple (a, L), with a real and less than one in absolute value, and L a one-dimensional d.f. which does not assign probability one to a single point. Let $A = \{q\}$. Let $J(q)$ be the d.f. of (X_1, X_2) when $\alpha = a$ and $G = L$. Let F_n be the two-dimensional empiric d.f. of

$$\{(X_{2i-1}, X_{2i}), i = 1, 2, \cdots n\}. \tag{8.2}$$

Finally let $q_0 = (\alpha, G)$; of course, q_0 is in A.

If G were to assign probability one to a single point then it is obvious that α would not be identified. The necessary condition, that G not assign probability one to a single point, is also sufficient, and the d.f. of (X_1, X_2) then determines $\alpha(|\alpha| < 1)$ uniquely. Even more: Let q' be a couple (a', L'), where $|a'| \leqq 1$, and L' is a d.f. which does not assign probability one to a single point. Suppose that $J(q_0) = J(q')$. Then $a' = \alpha$, hence is less than one in absolute value, and q' must be in A. For it follows from Theorem 1 of [10] that, if α were not uniquely determined, G would have to be normal. The possibility that G is normal and $\alpha \neq a'$ is then easily eliminated. (In [1] through an oversight it is erroneously stated that the d.f. of X_1 already determines α uniquely. Attention has been called to this error in, e.g., [8], page 211, footnote 6.)

Although the members of the sequence (8.2) are not independent, the two sequences made up of alternate members of this sequence are sequences of independent chance variables, and it is easy to show, as was done in [4], that

$$\delta(J(q_0), F_n) \rightarrow 0. \tag{8.3}$$

The minimum distance estimator of α is obtained in the usual manner. Let $S_n = (\alpha_n^*, G_n^*)$ be any function from the $2n$-Euclidean space of $(X_1, \cdots, X_{2n})$ to the space A which is measurable and such that

$$\delta(J(\alpha_n^*, G_n^*), F_n) < \inf_{q \varepsilon A} \delta(J(q), F_n) + \gamma(n). \tag{8.4}$$

Then α_n^* is a minimum distance estimator of α, to which it converges w.p.1.

To prove the latter we have only to show that $J(\alpha, G) = J(q_0)$ is not in $\bar{C}_2$. Let $\bar{A}$ be as defined in Section 4. Any member $\bar{q} = (\bar{a}, \bar{L})$ of $\bar{A}$ has one of the following properties:

1) $\bar{L}$ assigns probability one to a single point.

2) $\bar{L}$ is a d.f. which does not assign probability one to a single point, and $\bar{a} = \pm 1$.

3) $\bar{L}$ has variation less than one.

Suppose $J(q_0) = J(\bar{q})$. Then $\bar{q}$ cannot have the first of these properties, because then X_i = constant w.p.1. Also $\bar{q}$ cannot have the second of these properties, by the result described in the third paragraph of this section. If $\bar{q}$ had the third of these properties then either $J(\bar{q})$ would have variation less than one or $J(\bar{q})$ would assign probability one to a single point, neither of which can be true of $J(q_0)$. Hence $J(q_0)$ is not in $\bar{C}_2$.

The author is grateful to Professors Henry Teicher and Lionel Weiss for reading the manuscript.

REFERENCES

[1] J. Wolfowitz, "Estimation of the components of stochastic structures," *Proc. Nat. Acad. Sci.*, Vol. 40 (1954), pp. 602–606.

[2] J. Wolfowitz, "Consistent estimators of the parameters of a linear structural relation," *Skand. Aktuarietids.* 1952, pp. 132–151.

[3] J. Wolfowitz, "Estimation by the minimum distance method," *Ann. Inst. Stat. Math.* (Japan), Vol. 5 (1953), pp. 9–23.

[4] J. Wolfowitz, "Estimation by the minimum distance method in non-parametric stochastic difference equations," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 203–217.

[5] O. Reiersol, "Identifiability of a linear relation between variables which are subject to error," *Econometrica*, Vol. 18 (1950), pp. 375–89.

[6] K. L. Chung, "An estimate concerning the Kolmogoroff distribution," *Trans. Amer. Math. Soc.*, Vol. 67 (1949), pp. 36–50.

[7] H. Cramér, *Random Variables and Probability Distributions*, Cambridge University Press, 1937.

[8] M. Kac, J. Kiefer, and J. Wolfowitz, "On tests of normality and other tests of goodness of fit based on distance methods," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 189–211.

[9] J. Kiefer and J. Wolfowitz, "Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters" *Ann. Math. Stat.*, Vol. 27 (1956), pp. 887–906.

[10] H. Teicher, "Identification of a certain stochastic structure," *Econometrica*, Vol. 24 (1956), pp. 172–177.

[11] J. Neyman and E. L. Scott, "Consistent estimators based on partially consistent observations," *Econometrica*, Vol. 16 (1948), pp. 1–32.