SERIATION OF MULTIVARIATE OBSERVATIONS
THROUGH SIMILARITIES

BY

ALAN E. GELFAND

TECHNICAL REPORT NO. 146
AUGUST 1, 1969

THIS RESEARCH WAS SPONSORED BY THE ARMY RESEARCH OFFICE,
OFFICE OF NAVAL RESEARCH, AND AIR FORCE OFFICE OF
SCIENTIFIC RESEARCH BY CONTRACT NO.
Nonr-225(52) (NR-342-022)

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
SERIATION OF MULTIVARIATE OBSERVATIONS THROUGH SIMILARITIES

by

Alan E. Gelfand

TECHNICAL REPORT NO. 146

August 1, 1969

PREPARED UNDER CONTRACT Nonr-225(52)
(NR-342-022) FOR
OFFICE OF NAVAL RESEARCH

Reproduction in Whole or in Part is Permitted for any Purpose of the United States Government

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section III:</th>
<th>Methods</th>
<th>33</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.0</td>
<td>Remarks</td>
<td>33</td>
</tr>
<tr>
<td>3.1</td>
<td>Method I</td>
<td>34</td>
</tr>
<tr>
<td>3.2</td>
<td>Method II</td>
<td>39</td>
</tr>
<tr>
<td>3.3</td>
<td>Method III</td>
<td>44</td>
</tr>
<tr>
<td>3.4</td>
<td>Method IV</td>
<td>46</td>
</tr>
<tr>
<td>3.5</td>
<td>Method V</td>
<td>49</td>
</tr>
<tr>
<td>3.6</td>
<td>Several Comments and a Useful Criterion</td>
<td>51</td>
</tr>
<tr>
<td>Section IV:</td>
<td>Simulation Models and Applications of Procedures</td>
<td>53</td>
</tr>
<tr>
<td>4.1</td>
<td>Several Criteria</td>
<td>53</td>
</tr>
<tr>
<td>4.2</td>
<td>Simulation Models</td>
<td>60</td>
</tr>
<tr>
<td>4.3</td>
<td>Tables and Results</td>
<td>63</td>
</tr>
<tr>
<td>4.4</td>
<td>Applications to Archaeological Data</td>
<td>69</td>
</tr>
<tr>
<td>Section V:</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>Estimation</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>5.1 Introduction and Approach</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>5.2 Two Special Cases</td>
<td>78</td>
<td></td>
</tr>
<tr>
<td>5.3 Time Series Approaches</td>
<td>81</td>
<td></td>
</tr>
<tr>
<td>5.4 Generalized Least Squares and Maximum Likelihood Technique</td>
<td>82</td>
<td></td>
</tr>
<tr>
<td>References</td>
<td>94</td>
<td></td>
</tr>
</tbody>
</table>
SERIATION OF MULTIVARIATE OBSERVATIONS
THROUGH SIMILARITIES

by

Alan E. Gelfand

SECTION I
INTRODUCTION AND SUMMARY

For certain types of problems in multivariate data reduction, seriation and scaling may be reasonable approaches. An informal definition of these terms can be attempted in the following way. Assume we are given a collection of \( n \) objects. Seriation techniques try to order these objects on a one-dimensional scale in the sense of assigning a rank from one to \( n \) to each object. Scaling techniques attempt to do more by assigning a numerical value to each object so that not only is order achieved but also some quantitative measure of relative closeness is computed. These two approaches are not mutually exclusive. Quite frequently we can apply a seriation technique and obtain a reasonably well ordered configuration upon which we can try a scaling technique. In this paper our major concern will be to find a "good" method of assigning ranks be it a seriation or a scaling approach. Shortly we shall define the above concepts more rigorously but first let us consider some applications that motivate this kind of study.

An important problem in archaeology is the sequencing in temporal order of collections of deposits or graves in which archaeological artifacts have been located. An issue in political science is the ordering
of a group of individuals on a scale from "Liberal" to "Conservative" on the basis of their responses to political questionnaires. There are many psychological applications. One example is attempting to order a group of children on an intelligence scale through IQ tests. In these any in many other situations scaling may be desirable as well.

The original motivation for this research arises out of several discussions with Professor Herbert Solomon who suggested the study through his interest in previous work in this subject. Previous published investigations on this topic are few. D.G. Kendall [6] offered a statistical approach to the archaeological seriation problem in particular with respect to certain principles of archaeological sequence-dating as first discussed by Sir W.M. Flinders Petrie [11], [12] in a very nontechnical way in the 1890's.

Kendall's method treats the problem entirely as one of estimation. He assumes that there are independent vectors, $Y_i$, with independent components, each component, $Y_{ij}$, indicating the number of occurrences of a particular (j-th) variety of pottery in that particular (i-th) grave, and that each $Y_{ij}$ has a Poisson distribution with mean $\mu_{ij}$ where $\mu_{ij}$ is a function of several other parameters indicating abundance, centrality, and dispersion. He also lets $P$, the true temporal permutation, be a parameter. Kendall employs a maximum likelihood approach and is able to maximize over the $\mu_{ij}$'s independently of $P$ and thus obtain a scoring function $S(P)$ which he then must maximize over all permutations of $P$. However, for any reasonable $n$ the number of permutations becomes prohibitive. We remark that this same difficulty will obviously occur in a Bayesian approach to the problem. However,
Kendall suggests a two stage reduction procedure through the use of similarity functions. In an earlier less technical paper, Robinson [14] also suggested the use of similarity functions to handle the problem. His technique has several weaknesses and still requires the checking of a large number of permutations. Sternin [17] points this out in his paper and offers an interesting although not precisely defined solution using properties of the first and second eigenvectors of certain types of similarity matrices. Sternin's techniques lack general applicability and there is insufficient verification of their effectiveness when applicable. However, all investigators suggest the use of similarities which we will also in this paper. This will be discussed in detail in the next section.

Several psychologists such as Guttman [5] and Coombs [3] have mentioned particular problems in psychological seriation.

The problem of scaling (or multidimensional scaling as it is more generally labeled) once again considers the similarities between the $n$ k-dimensional members of our collection. Taking these as given, there is an attempt to reconstruct $n$ points in a lower dimensional ($< k$) space whose distances in some sense match the observed experimental similarities. In our case the lower dimensional space is one-dimensional. Indeed since this is the case we will not need to worry too much about the theory of multidimensional scaling since we can view the problem in a much simpler way as will be seen in subsequent sections. Unfortunately even the best scaling techniques have not been too successful in projecting multi-dimensional data into one-dimensional representations.
With regard to the scaling problem psychologists have made notable contributions. A discussion of scaling theory and methods up through 1958 is contained in the book by Torgerson [18]. However, as J. B. Kruskal [8], [9] indicates, their rationale is not fully satisfactory and statistical questions of the "goodness" of the solution are not adequately treated. A major advance was made by Shepard [15], in 1962 which enables one to find solutions through an iterative procedure requiring a computer. Subsequently Kruskal [8], [9] and Shepard and Carroll [16] have proposed appealing solutions. Modifications of these techniques to deal with our problem will be discussed later.

Let us examine the situation more formally. Our problem is the following:

We observe n k-dimensional vectors $Y_1, Y_2, \ldots, Y_n$. Let us assume some flexibility in the choice of the size of $k$ and in the interpretation of the components. In other words, $Y_i$ is a suitably chosen attribute vector for the $i$-th person or object.

The $Y$'s will ordinarily be assumed dependent although it may be reasonable to assume that they are conditionally independent given the parameters of their distributions. This assumption is employed by other investigators. See, for example, Kendall's work [6].

For instance in the archaeological problem the $Y$'s refer to the different graves or deposits and the components of the $Y$'s might
be the occurrence of or the number of a particular type of variety of artifact.

On the basis of comparison of the components of the Y's within and between vectors we seek "good" methods for associating with each $Y_i$ a rank $r(Y_i)$, $1, 2, \ldots, n$ and thus order these vectors on a one-dimensional scale. We must assume that the scale we are interested in (time scale, for example) is related to this scale via a continuous and strictly monotone transformation. Since rank orders are invariant under this group of functions we can consider the two scales as equivalent.

Lastly we assume that there is a true unknown order for the Y's on this scale which we designate as $X_1, \ldots, X_n$ and which we are trying to achieve via a "good" method.

These considerations bring us to a fundamental concept, which we mention informally now, that underlies our whole problem. This is the assumption that the components of the Y's are in some sense related to or give information about the order scale for the Y's. In other words we need some sort of "continuity" between the observed vectors and the $r(Y)$ which says that $Y_i$ is "close" to $Y_j$ if and only if $r(Y_i)$ is "close" to $r(Y_j)$.

Considering the importance of similarities to our problem we will devote section 2 to definition and discussion of similarity functions and similarity matrices. We will consider a large class of similarity functions. In addition we will look at the "continuity" problem in two different ways and discuss some related theory.
In section 3 we will look further at the previously mentioned concept of "continuity" and offer several methods of solving our problem from each point of view. These methods include scaling and seriation techniques. In addition, we offer a useful criterion for comparing different estimated orders.

In section 4 we present the results of a large scale simulation for comparison of our methods according to several criteria. In addition we apply our methods to some "real" data to compare them with previous results.

In the last section, section 5, we offer some estimation theory relating to some of the models discussed in section 2, and the techniques of section 3.
SECTION II
BACKGROUND AND THEORY

2.1 A Large Class of Similarity Functions.

For a collection \( Y_1, \ldots, Y_n \) each a \( k \)-dimensional vector, a measure of similarity between a pair \( Y_i \) and \( Y_j \) should reflect the closeness of the components of \( Y_i \) and \( Y_j \) and in particular the similarity between \( Y_i \) and \( Y_j \) should decrease as the discrepancy between their corresponding components increases and vice versa.

Thus we shall consider a similarity function in our case as a mapping \( F \) where

\[
(1) \quad F: \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R} \text{ (usually } \mathbb{R}^+). \]

Given \( X = (X_1, \ldots, X_k) \), \( Y = (Y_1, \ldots, Y_k) \) we want to compute a "discrepancy" for each pair \( X_\ell \) and \( Y_\ell \), \( \ell = 1, \ldots, k \) and then combine these discrepancies suitably in accordance with the above remarks. A very general class of measures, \( F \), of this kind may be written in the form

\[
(2) \quad F(X,Y) = g(\sum_{\ell=1}^{k} f_\ell(X_\ell, Y_\ell))
\]

where we assume that for the \( f_\ell \)

(a) \( f_\ell(x, y) = f_\ell(y, x) \) (symmetry property)

(b) \( f_\ell(y, y) = 0 \) (not essential)

(c) \( f_\ell(x, y) \geq f_\ell(x, z) \) if \( x \leq z \leq y \) or \( y \leq z \leq x \).

As for \( g \) we assume

(d) \( g(x) \leq g(x') \) if \( x' < x \) (monotone decreasing property)
(e) \( g(x) \) is bounded from above on \( \mathbb{R}^+ \) i.e., \( g(0) = c \)

(f) \( g(x) \geq 0 \) (not essential).

As a result of the above assumptions we have

\[
(3) \quad c = g(0) = F(X,X) \geq F(X,Y) \geq 0 .
\]

Defining \( G(X,Y) = c - F(X,Y) \), \( G \) is called the dissimilarity between \( X \) and \( Y \). We note that while

\[
(4) \quad G(X,X) = 0 \leq G(X,Y)
\]

and \( G(X,Y) = G(Y,X) \)

\( G \) is not necessarily a metric. That is, the triangle inequality

\[
(5) \quad G(X,Y) + G(Y,Z) \geq G(X,Z)
\]

need not hold. For example, let \( k = 2 \)

\[
(6) \quad f_g(x,y) = (x - y)^2 \quad \ell = 1, 2
\]

\[
g(x) = \frac{1}{1+x}
\]

\( X = (.1, .1) \)

\( Y = (.4, .4) \)

\( Z = (.7, .7) \)

then \( G(X,Y) = G(Y,Z) = \frac{.18}{1.18} \), \( G(X,Y) = \frac{.72}{1.72} \) so that

\[
G(X,Y) + G(Y,Z).
\]
Nonetheless $G(X,Y) \leq G(X,Z)$ means $Y$ is less dissimilar or more similar ("closer") to $X$ than $Z$ is, so that $G$ retains an intuitive interpretation as a distance.

$F$ will usually be scaled down to the interval $[0,1)$ where $F(X,Y)=1$ implies $X = Y$, that is $X_\ell = Y_\ell$ for all $\ell$. $F(X,Y) = 0$ means $X$ and $Y$ are as dissimilar as possible. In particular some examples that satisfy assumptions (a)-(f) are the following: For vectors whose components exist on $[-\infty, \infty]$ we can use

\[(8)\]
\[F(X,Y) = \frac{1}{1 + \sum_{\ell=1}^{k} w_\ell |X_\ell - Y_\ell|^p} \]

where the $w_\ell$ are non-negative weighting constants and $p > 0$.

Then $0 \leq F \leq 1$.

For vectors whose components sum to one (percentages) we can use

\[(9)\]
\[F(X,Y) = \frac{2 - \sum_{\ell=1}^{k} |X_\ell - Y_\ell|^p}{2}, \quad p > 0.\]

Then again $0 \leq F \leq 1$. This measure was suggested for the archaeological problem by Robinson [14].

For dichotomized data, vectors with 0's or 1's as components, we can use

\[(10)\]
\[F(X,Y) = \frac{1}{k} \sum_{\ell=1}^{k} \max(Y_\ell X_\ell, (1-X_\ell)(1-Y_\ell)) \]

\[= \frac{1}{k} \sum_{\ell=1}^{k} (1 - |X_\ell - Y_\ell|) \]

\[= 1 - \frac{1}{k} \sum_{\ell=1}^{k} |X_\ell - Y_\ell| \]

which again means $0 \leq F \leq 1$. 

9
2.2 The "Continuity" Problem.

Now returning to our problem, we have \( n \) observation vectors and if we calculate \( F(Y_i, Y_j) \) for all \( i \) and \( j \) we can arrange these numbers in a symmetric \( n \times n \) matrix with a constant diagonal which we define as a similarity matrix

\[
F = \begin{pmatrix}
F(Y_1, Y_1) & F(Y_1, Y_2) & F(Y_1, Y_3) & \cdots & F(Y_1, Y_n) \\
F(Y_2, Y_1) & F(Y_2, Y_2) & F(Y_2, Y_3) & \cdots & F(Y_2, Y_n) \\
F(Y_3, Y_1) & F(Y_3, Y_2) & F(Y_3, Y_3) & \cdots & F(Y_3, Y_n) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
F(Y_n, Y_1) & F(Y_n, Y_2) & F(Y_n, Y_3) & \cdots & F(Y_n, Y_n)
\end{pmatrix}
\]  

(11)

Now if we think of the \( Y \)'s as deterministic then the \( F \)'s are non-random and if \( X_1, X_2, \ldots, X_n \) is the true underlying order of the \( Y \)'s then we would like to assume a basic continuity condition such as

\[
F(X_i, X_j) \text{ increases in } j \text{ for } j < i, \ i=1,2,\ldots,n
\]

(12)

\[
F(X_i, X_j) \text{ decreases in } j \text{ for } j > i, \ i=1,2,\ldots,n
\]

as well as the already assumed

(15) \[ F(X_i, X_j) = F(X_j, X_i) \]

(14) \[ F(X_i, X_i) = k. \]

We might even reasonably assume a condition such as stationarity, that is

(15) \[ F(X_i, X_j) = f(|i-j|) = b|_{i-j}|. \]
In trying to solve our original problem a most reasonable approach is to consider the observed Y's and hence the ordered X's and the F's as random variables so that conditions (12) need not hold for our particular observations. However, as indicated earlier we must have some "continuity" relationship between the components of the Y's and the true Y ranks if there is to be any hope of solving the problem in a meaningful way.

Even with some sort of continuity relationship there may be no "reasonable" technique that can reconstruct the exact true order from a particular set of observed similarities since (12) need not hold. For example if \( n = 3 \) and \( X_1, X_2, X_3 \) is the true order, and if \( F(X_1, X_2) = .7, F(X_2, X_3) = .5 \) and \( F(X_1, X_3) = .6 \) then the best any rational technique can hope to conclude is an order \( X_2 X_1 X_3 \) or \( X_3 X_1 X_2 \).

Nonetheless, in all our methods for solution we will begin with a given set of similarities and attempt to extract a "best" estimated order from this set of ordinal numbers. From this perspective it is reasonable to consider "continuity" criteria and assumptions with respect to only the observed similarities. Thus we must offer the warning that the choice of components of our Y's to be considered for a particular problem should be made such that we really believe that they reflect or contain information regarding the true order.

Furthermore, the choice of an appropriate F from our class may be important as well.

All these considerations are important in order that the observed \( F(Y_i, Y_j) \) are a satisfactory measure of whatever proximity exists between the Y's.
In this paper we consider two approaches to the continuity problem. Each approach leads to different types of solutions to our basic problem. The solutions will be discussed in detail in the next section. At this point, the two approaches are outlined.

Approach I is based on the following: Consider the Y's as sets of numbers. Select some "reasonable" mathematical expression to measure the "continuity" between the corresponding observed F's and the sought-after scale order. Then select as an estimate of the true order the permutation of the Y's that "best" satisfies our continuity measure.

Approach II is based on the following: Consider the Y's as random variables and thus so are the F's. Under some reasonable assumptions about their underlying joint mean and covariance structure "good" procedures can be suggested to obtain an estimate of the true order.

We remark that the basis of Approach I is somewhat similar to that for scaling techniques while Approach II is geared strictly for seriation problems. In fact we can use Approach II to obtain a good estimate to incorporate in an Approach I-type scaling procedure. Furthermore Approach I is non-probabilistic while Approach II has random features.

The remainder of this section is devoted to the consideration of various assumptions on the underlying mean and covariance structure of our observed similarities and their ramifications. We shall be concerned only with the correctly ordered set which again we label as \( X_1, \ldots, X_n \).

2.3 Some Assumptions on the Mean Structure.

We define

\[
E[F(X_i, X_j)] = a_{ij}
\]
\[ \text{(17)} \quad \text{cov}[F(X_i, X_j), F(X_k, X_l)] = \sigma_{ij} \cdot k \cdot l. \]

Then in place of (12) we can assume that the mean structure of \( F \) has the monotonicity property.

\[ \text{a}_{ij} \text{ increases in } j \text{ for } j < i, \ i=1,2,\ldots, n \]

\[ \text{(18)} \quad \text{a}_{ij} \text{ decreases in } j \text{ for } j > i, \ i=1,2,\ldots, n. \]

Additionally

\[ \text{(19)} \quad \text{a}_{ij} = a_{ji} \]

\[ \text{(20)} \quad a_{ii} = c \text{ and thus } c \geq a_{ij}. \]

We might also assume stationarity,

\[ \text{(21)} \quad a_{ij} = b |i-j| \]

and obtain constants

\[ \text{(22)} \quad c = b_o \geq b_1 \geq \cdots \geq b_{n-1}. \]

Assumption (18) is essentially the type of assumption that motivates our Approach II-type techniques. It is a continuity premise in the sense that it asserts that we expect \( F(X_i, X_j) > F(X_i, X_k) \) if \( X_j \) is closer to \( X_i \) than \( X_k \) is to \( X_i \) in the true ordering.

Without getting deeply into distribution theory we may ask what sort of assumptions on the components of the \( Y \)'s (hence \( X \)'s) are sufficient for (18) to hold for all similarity functions belonging to the class defined by (2).
The following two theorems are examples. We assume that the \( X \) vectors (which are, as usual, the correctly ordered \( Y \)'s) are conditionally independent given the parameters of their distributions. When we write probability statements for the \( X \)'s we will assume the parameters are included.

Theorem 2.3.1. If the components of the \( X_i \) are independent of each other and if for each \( X_i, X_j, X_k \) such that \(|i-j| < |i-k|\)

\[
P(|X_{i,k} - X_{j,k}| < c) \geq P(|X_{i,k} - X_{k,k}| < c) \quad c \geq 0
\]

and if \( F \) belongs to the class defined by (2) then (18) holds.

Proof. We need only show that \( a_{ij} \geq a_{ik} \).

Since the \( f_{\ell}(X, Y) \) are increasing functions of \(|X-Y|\), (23) implies for each \( \ell \)

\[
P\left( \sum_{\ell=1}^{k} f_{\ell}(X_{i,\ell}, X_{j,\ell}) < c \right) \geq P\left( \sum_{\ell=1}^{k} f_{\ell}(X_{i,\ell}, X_{k,\ell}) < c \right) \quad c \geq 0.
\]

Our independence assumptions imply \( \sum_{\ell=1}^{k} f_{\ell}(X, X') \) is a sum of independent random variables. Thus (24) implies

\[
P\left( \sum_{\ell=1}^{k} f_{\ell}(X_{i,\ell}, X_{j,\ell}) < c \right) \geq P\left( \sum_{\ell=1}^{k} f_{\ell}(X_{i,\ell}, X_{k,\ell}) < c \right) \quad c \geq 0
\]

via an induction on \( k \) through a convolution argument.

Since \( g \) is monotone decreasing,

\[
P\left( g\left( \sum_{\ell=1}^{k} f_{\ell}(X_{i,\ell}, X_{j,\ell}) \right) > c \right) \geq P\left( g\left( \sum_{\ell=1}^{k} f_{\ell}(X_{i,\ell}, X_{k,\ell}) \right) > c \right) \quad c \geq 0
\]
\[ a_{ij} = E F(X_i, X_j) \geq E F(X_i, X_k) = a_{ik}. \]

Remark: Without independent components the conclusion need not hold.

**Theorem 2.3.2.** If each \( X_1 \sim \pi(\mu_1, \Sigma) \) and if for each \( X_i, X_j, X_k \) such that \( |i-j| < |i-k| \)

\[(27) \quad |\mu_i - \mu_j| < |\mu_i - \mu_k|\]

where \( |y| = (|v_1|, |v_2|, \ldots, |v_k|) \) (that is, \( |y| \) has as components the absolute value of the corresponding components of \( y \)) and if \( F \) belongs to the subclass of (2) such that for each \( \ell \)

\[(28) \quad f_\ell(x, y) = w_\ell |x - y|^{p_\ell} \quad w_\ell \geq 0, \quad p_\ell \geq 1, \]

then (18) holds.

**Proof.** Again we need only show that \( a_{ij} \geq a_{ik} \). It is sufficient that (25) holds. From (28) \( \sum_{\ell=1}^{k} f_\ell(X_\ell, Y_\ell) \) is symmetric in \( X - Y \) and

\[(29) \quad (X - Y; \sum_{\ell=1}^{k} f_\ell(X_\ell, Y_\ell) < c) \text{ is convex for all } c.\]

Furthermore,

\[(30) \quad X_i - X_j \sim \pi(\mu_i - \mu_j, 2\Sigma)\]
\[(30) \quad X_i - X_k \sim \pi(\mu_i - \mu_k, 2\Sigma)\]

From a geometric argument, (27), (29) and (30) imply that (25) holds.

For a more general result with analytic proof see Anderson [1], Corollary 2, page 172, with particular reference to its possible extension to general symmetric unimodal distributions.
Returning to our $a_{ij}$ we define the matrix

$$(31) \quad A = (a_{ij})$$

where we always assume (18), (19), and (20) holds, that is $A$ is a

Robinson matrix in the terminology of Kendall [7]. If in particular

(22) holds with $c = 1$ and $b_{n-1} \geq 0$, $A$ is defined to be Toeplitz
decreasing.

If there exist constants $0 < c_1 < \cdots < c_n$ and we define

$$(32) \quad a_{ij} = \begin{cases} 
\frac{c_i}{c_j} & i \leq j \\
\frac{c_j}{c_i} & i \geq j 
\end{cases}$$

then $A$ is defined to be a Green's matrix.

If (22) holds with $a_{ij} = r^{1-i-j}$, $r < 1$ then $A$ is defined to be

an exponential matrix.

"Arithmetic decay" is an example of a Toeplitz decreasing matrix.

That is let

$$(33) \quad b_i = 1-\frac{k}{n} \quad (k < 1/n).$$

"Exponential decay" is an example of a Green's matrix, i.e., let

$$(34) \quad c_i = e^{g(i)}$$

where $g(x)$ is a suitably chosen monotonically increasing function of

$x$. In fact we have the following simple theorem.
Theorem 2.3.3. If \( A \) is both a Green's matrix and a Toeplitz decreasing matrix then \( A \) is exponential; on the other hand if \( A \) is exponential, \( A \) is both a Toeplitz decreasing matrix and a Green's matrix.

Proof. The second statement is obvious. We just verify the first.

Since \( A \) is a Green's matrix

\[
a_{i,i+1} = \frac{c_i}{c_{i+1}}.
\]

But if \( A \) is Toeplitz decreasing

\[
b_{i} = a_{i,i+1} \quad i=1,\ldots,n-1
\]

so

\[
c_{1} = b_{1} c_{2} = b_{1}^{2} c_{3} = \cdots = b_{1}^{n-1} c_{n}
\]

and

\[
b_{i} = (b_{1})^{i}.
\]

Sternin [17], Kendall [6] and others have considered \( F \) matrices that are of one of the above special forms and propose solutions based on these assumptions. In Section V we shall consider estimation and testing for these special structures.

We shall also be concerned with finding conditions such that the "second order" similarities, \( \sum_{j=1}^{n} a_{ij}, \ i=1,2,\ldots,n \), have the property that
\[
\min \left( \sum_{j=1}^{n} a_{1j}, \sum_{j=1}^{n} a_{nj} \right) = \min_{1 \leq i \leq n} \left( \sum_{i=1}^{n} a_{ij} \right).
\]

Intuitively we would guess that an extreme element should have the smallest "second order" similarity. However, this need not be the case as the following \( A \) matrix which satisfies (18), (19), and (20) shows

\[
A = \begin{pmatrix}
6 & 5.9 & 4 & 3.9 & 2.7 & 2.6 \\
5.9 & 6 & 4.1 & 4 & 2.8 & 2.7 \\
4 & 4.1 & 6 & 4.1 & 2.9 & 2.8 \\
3.9 & 4 & 4.1 & 6 & 5 & 4.9 \\
2.7 & 2.8 & 2.9 & 5 & 6 & 5.9 \\
2.6 & 2.7 & 2.8 & 4.9 & 5.9 & 5
\end{pmatrix}
\]

\[
\Sigma a_{1j} = 25.1, \quad \Sigma a_{2j} = 25.5, \quad \Sigma a_{3j} = 23.9, \quad \Sigma a_{4j} = 27.9, \quad \Sigma a_{5j} = 25.3, \quad \Sigma a_{6j} = 24.9
\]

Thus, \( \Sigma a_{3j} < \Sigma a_{1j}, \Sigma a_{3j} < \Sigma a_{6j} \).

We also note that \( a_{14} = 3.9 > 2.9 = a_{35} \). Thus, it need not be true that

\[
\min_{|i-j|=k} a_{ij} > \max_{|i-j| \leq k+1} a_{ij}.
\]

What conditions on \( A \) apart from (18), (19), and (20) will insure that (35) holds.

Remark: It is trivial to show that (18), (19), and (20) imply

\[
\Sigma a_{1j} < \Sigma a_{2j}, \quad \Sigma a_{nj} < \Sigma a_{n-1,j}
\]
Theorem 2.3.4. If \( A \) is Toeplitz decreasing then (35) holds.

Proof.

\[
\sum_{j=1}^{n} a_{1j} = \sum_{j=1}^{n} a_{nj} = \sum_{i=0}^{n-1} b_{i}
\]

\[
\sum_{j=1}^{n} a_{ij} = b_{0} b_{1} + \cdots + b_{n-1} + b_{1} + \cdots + b_{n-1} > b_{0} b_{1} + \cdots + b_{n-1} + b_{n-1} + \cdots + b_{n-1}
\]

\[
= \sum_{j=1}^{n} a_{1j} = \sum_{j=1}^{n} a_{nj}.
\]

We remark that condition (35) need not hold for all Green's matrices. For example if \( n = 6 \) and

\[
\begin{align*}
c_{1} &= 1, & c_{2} &= 1.1, & c_{3} &= 12 \\
c_{4} &= 120, & c_{5} &= 130, & c_{6} &= 140
\end{align*}
\]

it is clear that

\[
\sum_{j=1}^{3} a_{3j} < 1.6, \quad \sum_{j=1}^{6} a_{1j} > \sum_{j=1}^{1} a_{1j} > 1.9.
\]

Nonetheless, we have the following

Theorem 2.3.5. Let \( A \) be a Green's matrix.

Suppose

\[
\begin{align*}
c_{1} &= \min_{j < i} \frac{c_{j} c_{i-j+1}}{c_{i}}
\end{align*}
\]

(that is, minimum is achieved when \( j = 1 \))
(39) \[ c_n = \max_{j < 1} \frac{c_i c_{n-(i-j)}}{c_j} \]

(that is, maximum is achieved when \( i = n \)).

Then if either (38) or (39) holds then (35) holds.

**Proof.** We just show that (38) implies (35)

\[
\sum_{j=1}^{n} a_{i,j} - \sum_{i=1}^{n} a_{i,j} = \sum_{i=1}^{n} \left[ c_i - \frac{c_i}{c_{i+1}} \right] \]

\[
= \sum_{j=2}^{i-1} \left( \frac{c_i}{c_j} - \frac{c_i}{c_{j+1}} \right) + (c_i - c_1) \sum_{j=i+1}^{n} \frac{1}{c_j} \]

\[
> \sum_{j=2}^{i-1} \left( \frac{c_i}{c_j} - \frac{c_i}{c_{j+1}} \right) \geq 0 \text{ by (38).} \]

Remark: In this case we have \( \sum a_{i,j} > \sum a_{n,j} \).

Since (38) and (39) are not always easy to check we have the following

**Theorem 2.3.6.** Let \( A \) be a Green's matrix and \( b_i = c_i - c_{i-1} \) with \( c_0 = 0 \). Then

(i) If \( b_1 \geq b_2 \geq \cdots \geq b_n \) then (38) holds

(ii) If \( b_1 \leq b_2 \leq \cdots \leq b_n \) then (39) holds.

**Proof.** We will prove (i). The proof of (ii) is similar.

Consider a pair \( i, j \), \( i < j \)
\[
\frac{c_i}{c_i} - \frac{c_j}{c_i-j+1} = \frac{c_{i-j+1} - c_i}{c_{i-j} + c_i} - \frac{c_i - c_j}{c_i} \\
\geq \frac{1}{c_i} [(c_{i-j+1} - c_i) - (c_i - c_j)] .
\]

Now for any \( k \), \( c_k = \sum_{k=1}^{k} b_k \). So the above becomes

\[
= \frac{1}{c_i} \left[ \sum_{k=2}^{i-j+1} b_k - \sum_{k=j+1}^{i} b_k \right] \geq 0 \text{ since } b_1 \geq b_2 \geq \cdots \geq b_n .
\]

From Theorems 2.3.5 or 2.3.6 we have the following examples where (35) holds.

I. \( c_i = c+ki \quad c, k \geq 0 \)

II. \( c_i = c+ri \quad c, r \geq 0 \)

III. \( c_i = e^{g(i)} \) where \( g(x) \) is increasing in \( x \), and \( g \) is either convex or concave.

Remark: \( g \) convex gives \( \Sigma a_{ij} \) as minimum (easy to verify from (38))
\( g \) concave gives \( \Sigma a_{ij} \) as minimum (easy to verify from (39))

Now let us consider the case for general \( A \).

Let

\[
(40) \quad c_k = \min_{|i-j|=k} a_{ij} , \quad b_k = \max_{|i-j|=k} a_{ij} \\
r_k = b_k - c_k , \quad t_k = b_k - b_{k+1} .
\]

Consider the following conditions
(A) \( c_i \geq b_{i+1} \)

(B) \( r_i = r \)

(C) \( t_i = t \)

(D) \( r_o = \max_i r_i < \min_{j<i} b_j - c_i \)

**Theorem 2.3.7.** Let \( A \) be a Robinson matrix. Then

(i) (A) and (B) imply (35) holds

(ii) (A) and (C) imply (35) holds

(iii) (A) and (D) imply (35) holds.

**Proof.** We just show (i) and (iii). The proof of (ii) is similar to that of (i).

(i) \( \Sigma a_{i,j} \geq 1 + a_{11} + a_{12} + \cdots + a_{1n-1} + a_{21} + \cdots + a_{n-1,1} \)

\[ \Sigma a_{i,j} \leq 1 + a_{11} + b_1 + b_2 + \cdots + b_{i-1} + b_i + \cdots + b_{n-1} \cdot \]

Hence

\[ \Sigma a_{i,j} - \Sigma a_{i,j} \]

\[ \geq (a_{11} + \cdots + a_{1n-2}) - (b_{11} + \cdots + b_{i-2}) \]

\[ + (a_{11} + \cdots + a_{n-1}) - (b_{11} + \cdots + b_{n-1}) \]

\[ \geq -(i-2)r + (n-i)(i-2)r \geq 0 \text{ since } 1 < i < n \]

(iii) For \( 1 \leq n/2 \) we show \( \Sigma a_{i,j} \leq \Sigma a_{i,j} \) with a similar proof for the other side. By earlier remarks we can assume \( i \geq 3 \). As in (i) we have
\[
\sum_{i,j} \Sigma_{ij} - \Sigma_{1j} \geq - \sum_{j=1}^{i-2} (b_j - c_j) + \sum_{j=1}^{n-i} (c_j - b_{j+i-1}) \\
\geq -(i-2)r_o + (n-i)r_o \\
\geq 0 \text{ since } i \leq n/2. 
\]

We now return to the covariance matrix defined by (17). We have an 
\[
\frac{n(n-1)}{2} \times \frac{n(n-1)}{2}
\]
matrix \( \Sigma \) with entries \( \sigma_{ij}^2 \) and \( \sigma_{ij \cdot k\ell} \).

2.4 Some Assumptions on the Covariance Structure.

Consider the following specializations of \( \Sigma_i \).

I. \( \sigma_{ij}^2 = \sigma^2 \)

II. \( \sigma_{ij \cdot k\ell} = m \sigma^2, \ -1 < m < 1 \) when there is exactly one subscript in common. (It seems reasonable to think of \( m > 0 \).)

III. \( \sigma_{ij \cdot k\ell} = km \sigma^2, \ 0 \leq k \leq 1 \) when there are no subscripts in common. (Perhaps \( k = 0 \) may be reasonable in some cases since if \( X_i, X_j, X_k, X_{\ell} \) are assumed conditionally independent given their distributions, \( F(X_i, X_j) \) and \( F(X_k, X_{\ell}) \) must be.)

We use the notation

(41) \( \Sigma = \Sigma(\sigma^2, m, k) \)

to refer to this matrix for a particular \( \sigma^2, m, \) and \( k. \)

It is clear that \( \Sigma \) is not necessarily a covariance matrix for all \( k, m \) on their ranges. In fact we have the following theorems.

(We assume \( \Sigma \) is \( \frac{n(n-1)}{2} \times \frac{n(n-1)}{2} ) \).
Theorem 2.4.1. For $\Sigma(\sigma^2, m, l)$ there exists a probability space with random variables defined on it if and only if

$$(42) \quad -\frac{2}{n(n-1)^2} < m < l.$$ 

Proof. We need only verify that $\Sigma$ is positive-definite for $m$ on this range. Without loss of generality we assume $\sigma^2 = 1$. Now it is easy to show that for each $r$ the $r \times r$ matrix, $\Sigma_r$ with ones on the diagonal and $m$'s off the diagonal has determinant

$$(43) \quad \text{Det}(\Sigma_r) = (1-m)^{r-1}(1+(r-1)m).$$

Thus by the well known necessary and sufficient condition for positive definiteness, $\Sigma$ is positive definite if and only if

$\text{Det}(\Sigma_r) > 0$ for $1 \leq r \leq n$, that is,

$$(1-m)^{r-1}(1+(r-1)m) > 0 \quad r = 1, \ldots, \frac{n(n-1)}{2}$$

or equivalently

$$(1+(r-1)m) > 0 \quad r = 1, \ldots, \frac{n(n-1)}{2}.$$ 

Therefore we must have

$$1+(\frac{n(n-1)}{2} - 1)m > 0.$$ 

Whence the lower bound in (42) follows.

Note: In the following two theorems we consider $\alpha_{ij}$, for $1 \leq i < j \leq n$ and if we write $\alpha_{ij}$, $i > j$ we interpret it as $\alpha_{ji}$.
Theorem 2.4.2. For $\Sigma(\sigma^2, m, 0)$ there exists a probability space with random variables defined on it if and only if

$$-\frac{1}{2(n-2)} < m < \frac{1}{2}.$$  \hfill (44)

(Remark: Since we are usually concerned with $m > 0$, we have an upper bound independent of $n$.)

Proof. Again we need only verify that $\Sigma$ is positive definite if and only if $m$ belongs to this range. Without loss of generality we assume $\sigma^2 = 1$. We must have $\text{var}(\beta_{i}X_{F_{i,j}}) > 0$ for all sets of constants $\alpha_{ij}$ such that $\alpha_{ij}$ is not identically zero for all $i$ and $j$, i.e.,

$$\sum_{i<j} \alpha_{ij}^2 + \left[ \sum_{i<j} \sum_{k \neq i,j} \alpha_{ij}(\alpha_{kj} + \alpha_{ik}) \right]m > 0.$$  \hfill (45)

Let

$$T_1 = \sum_{i<j} \alpha_{ij}^2$$

$$T_2 = \sum_{i<j} \sum_{k \neq i,j} \alpha_{ij}(\alpha_{kj} + \alpha_{ik}).$$

Then we obtain

$$m > -\frac{T_1}{T_2} \text{ if } T_2 > 0$$  \hfill (46)

$$m < -\frac{T_1}{T_2} \text{ if } T_2 < 0.$$  

We thus seek
(47) \[ \sup_{\{\alpha_{ij}\}} \left( -\frac{T_1}{T_2} \right), \text{ subject to } T_2 > 0 \]

\[ = \sup_{\{\alpha_{ij}\}} \left( \frac{T_2}{T_1} \right), \text{ subject to } T_2 > 0 \]

\[ = \sup_{\{\alpha_{ij}\}} \left( \frac{T_2}{T_1} \right), \text{ (since } T_1 > 0) \]

and

(48) \[ \inf_{\{\alpha_{ij}\}} \left( -\frac{T_1}{T_2} \right), \text{ subject to } T_2 < 0 \]

\[ = \inf_{\{\alpha_{ij}\}} \left( \frac{T_2}{T_1} \right), \text{ subject to } T_2 < 0 \]

\[ = \inf_{\{\alpha_{ij}\}} \left( \frac{T_2}{T_1} \right), \text{ (since } T_1 > 0) \]

Claim 1: \[ \sup_{\{\alpha_{ij}\}} \left( \frac{T_2}{T_1} \right) = 2(n-2) \]

\[ \text{so that } m > - \frac{1}{2(n-2)} \]

Claim 2: \[ \inf_{\{\alpha_{ij}\}} \left( \frac{T_2}{T_1} \right) - 2 \]

\[ \text{so that } m < \frac{1}{2} \]

For Claim 1 the Cauchy-Schwarz Inequality implies
\[(49) \quad (\sum_{j=1}^{n} \alpha_{ij})^2 \leq (n-1) \sum_{j=1}^{n} \alpha_{ij}^2 \]

so that

\[(50) \quad \sum_{i=1}^{n-1} (\sum_{j=1}^{n} \alpha_{ij})^2 \leq (n-1) \sum_{i=1}^{n-1} \sum_{j=1}^{n} \alpha_{ij}^2 = 2(n-1)T_1 \]

Clearly

\[(51) \quad 2T_1 + T_2 = \sum_{i=1}^{n-1} (\sum_{j=1}^{n} \alpha_{ij})^2 \]

so

\[2T_1 + T_2 \leq 2(n-1)T_1 \]

whence

\[\frac{T_2}{T_1} \leq 2(n-2) \]

Since the equalities can be achieved Claim 1 follows. Obviously from

\[(51) \quad 2T_1 + T_2 > 0 \quad i.e., \quad \frac{T_2}{T_1} > -2 \]

so that Claim 2 follows.

The following theorem generalizes both Theorems 2.4.1 and 2.4.2.
Theorem 2.4.3. For $\Sigma(\sigma^2, m, k)$ there exists a probability space with random variables defined on it if and only if

\[ \frac{-2}{(n-2)(4+(n-3)k)} < m < \frac{1}{2-k}. \]

(Remark: Again we have an upper bound independent of $n$. Furthermore $k = 1$ yields (42), $k = 0$ yields (44).)

Proof. Again we need to verify that $\Sigma$ is positive definite if and only if $m$ belongs to this range. Without loss of generality we assume $\sigma^2 = 1$.

Again we must have $\operatorname{var}(\Sigma \alpha_{i,j}^T F_{i,j}) > 0$ for all sets of constants $\alpha_{i,j}$ such that $\alpha_{i,j}$ is not identically zero for all $i$ and $j$, i.e.,

\[ \sum_{i < j} \alpha_{i,j}^2 + \left[ \sum_{i < j} \sum_{k \neq i,j} \alpha_{i,j}(\alpha_{i,k} + \alpha_{k,j}) \right] m + \left[ \sum_{i < j} \sum_{k \neq j, k < l} \alpha_{i,j} \alpha_{k,l} \right] km > 0. \]

Again let

\[ T_1 = \sum_{i < j} \alpha_{i,j}^2 \]

\[ T_2 = \sum_{i < j} \sum_{k \neq i, k \neq j, k < l} \alpha_{i,j}(\alpha_{i,k} + \alpha_{k,j}) \]

\[ T_3 = \sum_{i < j} \sum_{k \neq j, k < l} \alpha_{i,j} \alpha_{k,l} \]

Thus we obtain
\[0 < (1 - \frac{t}{x})^2 + z + \frac{t}{x}\]

To verify Claim 2, we need only show that

\[\sum_{x \neq y \in \mathbb{P}} \frac{z}{(z-u)(z-u)} = \sum_{x \neq y \in \mathbb{P}} \frac{z}{(z-u)(z-u)} \geq \sum_{x \neq y \in \mathbb{P}} \frac{z}{(z-u)(z-u)} \]

and thus

\[\sum_{x \neq y \in \mathbb{P}} \frac{z}{(z-u)(z-u)} + \sum_{x \neq y \in \mathbb{P}} \frac{z}{(z-u)(z-u)} \geq \sum_{x \neq y \in \mathbb{P}} \frac{z}{(z-u)(z-u)} \]

so that

\[\sum_{x \neq y \in \mathbb{P}} \frac{z}{(z-u)(z-u)} = \frac{z}{(z-u)(z-u)} \]

Theorem (28) follows from the fact that for each \( z \), there are terms in the summation over \( k \) and \( \ell \) such that

\[\sum_{x \neq y \in \mathbb{P}} \frac{z}{(z-u)(z-u)} \geq \frac{z}{(x-u)(x-u)} \]

as in the previous theorem.

For Claim 1, the Cauchy-Schwarz inequality implies

\[\sum_{x \neq y \in \mathbb{P}} z < \sum_{x \neq y \in \mathbb{P}} z \]

so that

\[z = \left( \frac{z}{x-z} \right) \sum_{x \neq y \in \mathbb{P}} \frac{z}{x-z} \]

Claim 2:
\[
\begin{align*}
\text{Claim 1:} & \quad \frac{c}{\xi(\xi - u) + \xi + u} + (\xi - u)\xi = \left(\frac{I}{\xi + \xi + u}\right)^{1 \over \eta} \\
& \quad \text{(since)} \\
& \quad \frac{c}{\xi(\xi - u) + \xi + u} = \left(\frac{I}{\xi + \xi + u}\right)^{1 \over \eta} \\
& \quad (0 < \frac{I}{\xi + \xi + u}) \\
& \quad \text{subject to} \\
& \quad 0 > \xi + \xi + u \\
& \quad \text{(55)}
\end{align*}
\]
From before we know that \( \frac{T_2}{T_1} > -2 \) so that if \( T_2 > T_1 \) we are finished.

If \( T_2 < T_1 \) then the worst case is when \( k = 1 \) in which case we must show

\[
\frac{T_2}{T_1} + 2 + \frac{T_3}{T_1} - 1 > 0
\]

or

\[
T_1 + T_2 + T_3 > 0
\]

but

\[
T_1 + T_2 + T_3 = (\sum_{i<j} \alpha_{ij})^2 > 0
\]

so we are done.

Again the supremum and infimum can be attained. As a matter of fact whenever all \( \alpha_{ij} \) are equal we attain the supremum. The infimum can be attained if for example \( \alpha_{12} = 1, \alpha_{13} = -1, \alpha_{24} = -1, \alpha_{34} = 0 \), all other \( \alpha_{ij} = 0 \).

In the estimation section we will be interested in the marginal distribution of the set of \( n-1 \) variables, \( F(x_1, x_2), F(x_2, x_3), \ldots, F(x_{1},x_{1+1}) \ldots F(x_{n-1},x_{n}) \) which we label as \( W_1, W_2, \ldots, W_{n-1} \). We shall be interested in particular in assuming that

\[
E(W_1) = E(W_2) = \cdots = E(W_{n-1})
\]

and

\[
\text{cov}(W_i, W_j) = m_{|i-j|} \sigma^2
\]

where \( m_\alpha \) decreases as \( \alpha \) increases and \( m_0 = 1 \).
Two special cases of (66) are

I. \( m_1 = m, \ m_i = km, \ i > 1, \ 0 \leq k \leq 1, \) in which case we obtain a positive-definite covariance matrix for \( m \) on the range given by Theorem 2.4.3. If \( k = 0 \) we have the situation of first order dependence.

II. \( m_1 = m^1, \) in which case we obtain a positive definite covariance matrix for \( -1 < m < 1 \) since for each \( r \) the determinant of the \( r \times r \) matrix having this form is \( (1-m^2)^{r-1} > 0. \)
SECTION III

METHODS

3.0 Remarks.

From this point on we shall assume that we are given the set of
\[ \frac{n(n-1)}{2} \]
observed similarities \( F(Y_i, Y_j), 1 \leq i < j \leq k \), and from them
we seek a "good" estimated serial order. The question of handling "ties"
in the F's and the possibility of dealing with "missing" observed
similarities will be discussed separately for each method.

At this point we must mention that through similarities the best
we can hope to do is obtain an estimated order up to reversibility.
This is clear from the fact that an estimated order and its reverse
have the objects in the same relative order and it is up to us to inter-
pret or determine the direction of the underlying scale for each parti-
cular problem. Realistically this should present no difficulty since
expertise should enable us to distinguish oldest from youngest, most
liberal from most conservative, etc. once we have a "good" estimated
order. Of course in situations where all our objects are located within
a small segment of the ordering scale such as the case of excavating a
collection of graves all say only a hundred years apart this "reversi-
bility" difficulty may not be so easily handled. Having mentioned this
problem we will no longer consider it. When we write down an estimated
order we will assume that we may want its reverse and in our simulation
we assumed that from expertise one was always able to select the
appropriate one.
We now introduce and discuss the various techniques we have looked at. As mentioned previously we have two approaches to the "continuity" problem. We first consider methods of the Approach I - type.

3.1 Method I.

The continuity measure we construct in this method is similar to the one considered by Shepard and Carroll in [16]. Associate with each $Y_i$ a number $t_i$ and we will let $r(Y_i) = r(t_i)$ so that our estimated order is produced as soon as we specify the $t_i$. How does one choose the $t_i$? The $t$'s can be considered as "underlying independent" variables.

Suppose for the moment that the observed "dependent" $Y$'s are one-dimensional as well. Then for a particular set of $t$'s arranged in increasing order, to say that the corresponding $Y$'s change in a "continuous" manner as we move along the $t$ axis means essentially that the changes in $Y$ as we move from one $t$ to the next tend to be small compared to the changes in $Y$ generally associated with larger changes in $t$. Thus we can define a measure of "continuity" between $Y$ and $t$ as follows:

$$C(t_1, \ldots, t_n) = \sum_{i=1}^{n-1} \left[ \frac{Y_{i+1} - Y_i}{t_{i+1} - t_i} \right]^2.$$  \hspace{1cm} (1)

This measure has been considered by Von Neumann [19] and others. However, it seems rather arbitrary to restrict consideration to adjacent points only. In other words we should include a $\left( \frac{\Delta Y}{\Delta t} \right)^2$ term for every pair of points. However, in our measure we would want to weigh the $\left( \frac{\Delta Y}{\Delta t} \right)^2$ terms by a factor which decreases as $\Delta t$ decreases. Hence we generalize (1) to
\( C(t_1, \ldots, t_n) = \sum_{i<j} \left( \frac{t_i - t_j}{t_i - t_j} \right)^2 \omega_{ij} \)

where \( \omega_{ij} \) is a monotone decreasing function of \( |t_j - t_i| \).

How do we generalize (2) to the case of \( k \)-dimensional \( Y \) vectors?

Recalling the definition of the dissimilarity between \( Y_i \) and \( Y_j \), \( G(Y_i, Y_j) \), given in Section II it seems reasonable to use these quantities as the "distances" between the \( Y \)'s. If we let \( d_{ij} = |t_i - t_j| \) then in the denominator of (2) we should put \( d_{ij}^p \) where \( p > 0 \) is chosen appropriately for the particular \( G(F) \) being used.

For example if

\[
G(Y_i, Y_j) = \frac{1}{k} \sum_{\ell=1}^{k} \omega_{\ell} (Y_{i\ell} - Y_{j\ell})^2 + 1
\]

\( p \) can be chosen to be two.

With these modifications (2) becomes

\( G(F; t_1, \ldots, t_n) = \sum_{i<j} \frac{G_{ij}}{d_{ij}^p} \omega_{ij} \)

where again \( \omega_{ij} \) is a monotone decreasing function of \( |t_j - t_i| \). Now let us say more about \( \omega_{ij} \). Besides being monotone decreasing in \( |t_j - t_i| \) one could also want the ratio \( \frac{\omega_{ij}}{\omega_{k\ell}} \) to be invariant under scale and location changes in the \( t \)'s. These requirements force \( \omega_{ij} \) to have the form

\[
\omega_{ij} = c |t_i - t_j|^{-r} \quad r > 0 .
\]
To see this consider \( \omega_{ij} = f(|t_j-t_i|) \) where we assume \( f \) (Lebesgue) measurable.

The following conditions have been imposed on \( f \).

1. \( f(x) > 0 \) defined for \( x > 0 \).
2. \( f(x) \) decreases as \( x \) increases.
3. Invariance under scale change of ratios i.e., for \( a > 0 \)

\[
\frac{f(x)}{f(y)} = \frac{f(ax)}{f(ay)} \quad x,y > 0 .
\]

Remark: Be definition the \( \omega_{ij} \) are invariant under location changes in the \( t's \) since they are a function of the difference of two \( t's \).

The result (4) will follow if we can show that the only measurable \( f \) satisfying the three conditions is

\[
f(x) = cx^k \quad k < 0 .
\]

However, let \( g(x) = \frac{f(x)}{f(1)} = \frac{f(x)}{c} \). Then \( g \) satisfies (5) and \( g(1) = 1 \). Also (5) implies, in terms of \( g \), that

\[
g(xy) = g(x)g(y) \quad x,y > 0 .
\]

Letting \( x = e^u \), \( y = e^v \) (7) becomes

\[
g(e^{u+v}) = g(e^u)g(e^v) \quad \text{for all } u,v
\]

which is the famous Cauchy functional equation and implies that

\[
g(e^u) = e^{ku} = (e^u)^k
\]
or,

(10) \[ g(x) = x^k \]

and thus

(11) \[ f(x) = cx^k. \]

In order that \( f \) be monotone decreasing \( k \) must be less than zero. Thus (3) becomes

(12) \[ C(F; t_1, \ldots, t_n; r) = \sum_{i < j} \frac{G_{i,j}}{(d_{i,j})^{p+r}} \quad p, r > 0. \]

By the preceding remarks the best selection of a set of \( t \)'s would be one minimizing (12). However (12) is an unconstrained minimum for we need only let the distances between the \( t \)'s become large and (12) can be made arbitrarily small. This fact requires that we adjust (12) by a normalizing factor. It seems reasonable to suggest that the minimum of (12) occurs when the \( t \)'s are such that the \( d_{i,j}^p \) are proportional to the \( G_{i,j} \). In addition, in order that the minimum be constrained, the exponent of \( d_{i,j} \) in the normalizing factor must be \( p+r \). Thus we are led to the normalization by

(13) \[ \left[ \sum_{i < j} \frac{1}{d_{i,j}^r} \right]^{\frac{p+r}{r}}. \]

Hence we obtain our final continuity measure
(14) \[ C(F; t_1, \ldots, t_n: r) = \frac{\sum_{1 \leq j} \frac{G_{i,j}}{(d_{i,j})^{p+r}}}{\left[ \sum_{1 \leq j} \frac{1}{(d_{i,j})^{r}} \right]^{p+r}} \quad p, r > 0. \]

To see that (13) achieves the desired minimum in (14) let us suppose that \((d_{i,j})^k = G_{i,j}\) instead of \((d_{i,j})^p = G_{i,j}\).

Then (14) becomes

\[ \frac{\sum_{1 \leq j} \frac{1}{G_{i,j}}}{\left[ \sum_{1 \leq j} \frac{1}{(G_{i,j})^{r/k}} \right]^{p+r}} \]

while with \((d_{i,j})^p = G_{i,j}\) (14) becomes

\[ \frac{\sum_{1 \leq j} \frac{1}{G_{i,j}}^{r/p}}{\left[ \sum_{1 \leq j} \frac{1}{(G_{i,j})^{r/p}} \right]^{p+r}} = \frac{1}{\left[ \sum_{1 \leq j} \frac{1}{(G_{i,j})^{r/p}} \right]^{p/r}} \]

and (16) is less than or equal to (15) by Hölder's Inequality, that is,

\[ \left[ \sum_{1 \leq j} \frac{1}{(G_{i,j})^{r/k}} \right]^{p+r} \leq \left[ \sum_{1 \leq j} \frac{1}{(G_{i,j})^{r/k}} \right] \left[ \sum_{1 \leq j} \frac{1}{(G_{i,j})^{r/p}} \right]^{p/r} \]

since by writing
\[
\left(\frac{1}{c_{ij}}\right)^{r/k} = \left(\frac{1}{d_{ij}}\right)^{r/k-r/p+r} \left(\frac{1}{g_{ij}}\right)^{r/p+r}
\]

(18)

\[
\sum_{i<j} \left(\frac{1}{d_{ij}}\right)^{r/k} \leq \left[ \sum_{i<j} \left(\frac{1}{d_{ij}}\right)^{p+r-1} \right]^{r/p+r} \left[ \sum_{i<j} \left(\frac{1}{g_{ij}}\right)^{r/p} \right]^{p+r/p}
\]

and (17) follows.

To reiterate, this index measures continuity inversely i.e., a "better" (meaning smoother) fit of \( t \)'s will yield a smaller \( C \). We thus seek

\[
\min_{(t_1, \ldots, t_n)} C(F; t_1, \ldots, t_n; r).
\]

(19)

Several remarks are in order at this point. (i) This minimization can be performed by a method of steepest descent-type program which is available. (ii) The choice of \( r \) is a pragmatic one, namely select the value that seems to work best. (iii) The problem of ties in the \( F_{ij} \) is clearly nonexistent. (iv) In the case of a few missing similarities we can delete them and their corresponding \( d_{ij} \)'s from both the numerator and denominator in (14). (v) This technique has scaled the \( Y \)'s as well as ordered them.

3.2 Method II.

Again let us associate with each \( Y_i \) a number \( t_i \) and we will let \( r(Y_i) = r(t_i) \). In choosing the \( t \)'s we follow the non-metric multidimensional scaling technique of Kruskal [8], [9] applied to a one-dimensional scaling space.
Again we consider the dissimilarities \( G_{ij} \) as our "distances" between Y's and again let \( d_{ij} = |t_i - t_j| \).

In this approach we calculate for each \( (t_1, \ldots, t_n) \) a non-negative number \( S(F; t_1, \ldots, t_n) \) called the "stress". We then seek

\[
(20) \quad \min_{(t_1, \ldots, t_n)} S(F; t_1, \ldots, t_n)
\]

and the corresponding set of \( t \)'s we refer to as the configuration of minimum stress.

The stress is a measure of "continuity" but it is arrived at through quite different considerations than the measure developed in Method I. We define \( S(F; t_1, \ldots, t_n) \) for a particular set of \( t \)'s in the following way.

Let us arrange the \( G_{ij} \) (\( i < j \)) in increasing order and relabel the \( \frac{n(n-1)}{2} \) numbers as

\[
(21) \quad G_1 < G_2 < \cdots < G_{\frac{n(n-1)}{2}}.
\]

For the moment we assume no ties. We treat this situation later.

Denote the corresponding \( d_{ij} \) as \( d_1, \ldots, d_{\frac{n(n-1)}{2}} \).

What would a perfect fit be? Surely it should mean that if \( G_{ij} < G_{k'l'} \) then \( d_{ij} < d_{k'l'} \). In other words, a perfectly fitting set of \( d \)'s would have the property that

\[
(22) \quad d_1 \leq d_2 \leq \cdots \leq d_{\frac{n(n-1)}{2}}.
\]

For such a solution we would want \( S(F; t_1, \ldots, t_n) = 0 \). To measure how
far we are from "perfect fit" or monotonicity we look at all sets of numbers \( \delta_1, \ldots, \delta_{\frac{n(n-1)}{2}} \) satisfying

\[
\delta_1 \leq \delta_2 \leq \cdots \leq \delta_{\frac{n(n-1)}{2}},
\]

and define \( S(F; t_1, \ldots, t_n) \) as

\[
(24) \quad S(F; t_1, \ldots, t_n) = \min_{\delta} \sum_{i=1}^{\frac{n(n-1)}{2}} |d_i - \delta|^{2k}
\]

for some \( k > 0 \), where the minimum is taken over all sets of \( \delta \)'s satisfying (23).

Remark: If (22) holds the stress is zero.

It should be pointed out that there need not be and probably will not be a set of \( t \)'s that will yield a particular set of \( \delta \)'s satisfying (23). In our attempt we seek only a set of \( \delta \)'s that are as close to the \( d \)'s as possible.

As a measure of continuity (24) is invariant under translations but not under scaling changes in the \( t \)'s. We would want in addition that

\[
(25) \quad S(F; t_1, \ldots, t_n) = S(F; at_1, \ldots, at_n) \quad a > 0.
\]

In order that (25) hold a normalizing factor is required. One choice is easily seen to be
\[
\frac{n(n-1)}{2} \sum_{i=1}^{k} d_i^k
\]

and thus as our final measure of stress we employ

\[
S(P; t_1, \ldots, t_n) = \min_{\delta} \frac{n(n-1)}{2} \sum_{i=1}^{k} \frac{|d_i - \delta_i|^k}{d_i^k}
\]

and then we use the criterion (20) to determine the best set of \( t \)'s.

Again we offer several remarks. (i) The double minimization can be easily performed by a method of steepest descent-type program which is available, and is discussed in detail in Kruskal [9]. (ii) The choice of \( k \) does not seem important and we usually set \( k = 2 \) to correspond to the usual "residual sum of squares". (iii) To handle ties we follow Kruskal's suggestion. If

\[
G_\alpha = G_{ij} = G_{k\ell} = G_\beta
\]

then we really do not care which of \( \delta_\alpha \) and \( \delta_\beta \) is the smaller or larger and hence the constraint (23) should be modified to

\[
G_\alpha < G_\beta \implies \delta_\alpha \leq \delta_\beta
\]

This modification is included in the program mentioned above. (iv) Once again to handle missing similarities we just delete them from (26) as before. (v) Obviously this method scales the \( Y \)'s as well as orders them.
These first two techniques have several drawbacks when one is interested in the seriation problem alone.

I. Each requires specification of a somewhat arbitrary "continuity" measure which may not be desirable for particular problems. Furthermore, in obtaining a "best" solution the corresponding minimum value $C(F; t_1, \ldots, t_n)$ or $S(F; t_1, \ldots, t_n)$ is an absolute number and we cannot interpret it in terms of how well we have done with our solution.

II. Since we really only want the ranks of the $t$'s and not the $t$'s themselves we have the feeling that we are asking for more than we need and perhaps not doing as well as we might.

III. The major difficulty is that both methods require a "good" starting configuration for the $t$'s in order to work well. This is due to the fact that using the method of steepest descent programs available it is quite easy to be trapped in a local minimum of $C$ or $S$. For example consider the following three points $t_i, t_j, t_k$ as in Figure 1.

```
\begin{center}
  \begin{tikzpicture}
    \draw (0,0) -- (1,0);
    \draw (1,0) -- (2,0);
    \draw (0,0) -- (0,1);
    \draw (2,0) -- (2,1);
    \draw (1,0) -- (1,1);
    \filldraw (0,0) circle (2pt) node[above] {$t_i$};
    \filldraw (2,0) circle (2pt) node[above] {$t_k$};
    \filldraw (1,0) circle (2pt) node[above] {$t_j$};
  \end{tikzpicture}
\end{center}
```

Figure 1

Suppose $t_k$ "belongs" closer to $t_i$ than to $t_j$. Then $t_k$ is being pulled toward $t_i$ and being pushed away from $t_j$ and hence may be "trapped" into staying relatively where it is. In higher dimensions the $t$'s have more directions to move in and this problem occurs less frequently, hence the success of these methods in higher dimensional scaling.
IV. We frequently have a collection of a size \( n \) which is fairly large and is not feasible to handle by these methods.

In light of the previous remarks our next set of approaches (Type II) will seek a "good" serial estimate possibly with the intent of using it as a suggestion for a starting configuration in Method I or II. These techniques will be motivated by assumptions discussed in section II.

3.3 Method III.

Method III is a so-called "quick and dirty" technique. It is a crude initial approach which may work well in certain situations. Although applicable to any observed \( F \) matrix, the motivation for and confidence in the estimated order it produces will be a function of whether one believes it to be approximately the case that (i) \( A \) is a Robinson matrix, (ii) Condition (35), section II, holds for \( A \).

Suppose \( Y_\alpha \) is the \( Y \) such that

\[
(28) \quad \sum_{j=1}^{n} F(Y_\alpha, Y_j) = \min_{1 \leq i \leq n} \sum_{j=1}^{n} F(Y_i, Y_j).
\]

Then by consulting \( F \), arrange the row of similarities for \( Y_\alpha \) in decreasing order, assuming no ties and labeling the successive \( Y \)'s as \( Y_\alpha_2, \ldots, Y_\alpha_n \) i.e.,

\[
(29) \quad F(Y_\alpha_1, Y_\alpha_i) > F(Y_\alpha_1, Y_\alpha_i) > \cdots > F(Y_\alpha_1, Y_\alpha_n)
\]

and thus we obtain the order.
(30) \[ Y_{\alpha_1}, \ldots, Y_{\alpha_n} \]

To handle ties use a comparison with \( Y_{\alpha_2} \). If \( F(Y_{\alpha_1}, Y') = F(Y_{\alpha_1}, Y'') \) compare \( F(Y_{\alpha_2}, Y') \) and \( F(Y_{\alpha_2}, Y'') \). If there is a tie between the two \( Y \)'s most similar to \( Y_{\alpha_1} \), then we can resolve the tie by comparison with the third most similar \( Y \) to \( Y_{\alpha_1} \).

The order thus obtained may suffer from the fact that the similarity measure \( F \) often tends to be a better "large similarity" discriminator than a small one. For example with respect to the true order \( X_1, \ldots, X_n \), \( F(X_1, X_2) \) is more likely to be bigger than \( F(X_1, X_j) \) than \( F(X_1, X_{n-1}) \) is to be bigger than \( F(X_1, X_n) \). In other words it often tends to be the case that

\[
(31) \quad a_{i,i+1} - a_{i,i+2} > a_{i,i+j} - a_{i,i+j+1}
\]

for all \( i \) and for \( j \) much bigger than \( 1 \). Implicitly we will be aware of this tendency in subsequent methods as well.

Thus the order (30) as written may have many errors on the right end and this suggests we consider applying the following modifying technique to (30).

Beginning with \( i=1 \) and continuing to \( i=n-2 \), look at \( Y_{\alpha_1}, Y_{\alpha_{i+1}} \), \( Y_{\alpha_{i+2}} \) at each stage (possibly modified from previous stages) and calculate

\[
(32) \quad \min(F(Y_{\alpha_1}, Y_{\alpha_{i+1}}), F(Y_{\alpha_1}, Y_{\alpha_{i+2}}), F(Y_{\alpha_{i+1}}, Y_{\alpha_{i+2}}))
\]

If the first term is the minimum interchange \( Y_{\alpha_{i+2}} \) and \( Y_{\alpha_{i+1}} \). If the second term is the minimum make no change. If the third term is
the minimum interchange $Y_{\alpha_i}$ and $Y_{\alpha_{i+1}}$. In the case of a tie for the minimum make no changes.

Shortly, (in 3.6) we shall offer a criterion for determining whether at each stage, this modifying technique has made any overall improvement.

Three comments are in order.

(i) This method yields an estimate which is invariant under the observed order of the $Y$'s.

(ii) Since the method does not consider all the data in $F$ we should not expect it to be as strong as one that does.

(iii) For small $n$ an order can be gotten quickly by hand and in general this method is quickly applied.

Generally this method will be quickest but weakest.

3.4 Method IV.

If the true order is $X_1, \ldots, X_n$ this technique attempts to pick out $F(X_1, X_2)$, $F(X_2, X_3)$, $\ldots$, $F(X_{n-1}, X_n)$.

In other words from the $F$ matrix arrange the $\frac{n(n-1)}{2}$ terms, $F(Y_i, Y_j)$, in decreasing order (as in the Kruskal technique), again assuming no ties. We then select the $n-l$ largest $F$'s sequentially subject to the following conditions.

I. $n-2$ subscripts appear exactly twice and $2$ subscripts appear exactly once among the $2(n-1)$ subscripts of the $(n-1)$ $F$'s picked.

II. All subscripts "communicate" in the sense that given $i$ and $j$ there exists $k_1, \ldots, k_p$ for some $p$ such that among the $(n-1)$ $F$'s are included
\( F(Y_{1,k_1}), F(Y_{k_2}, Y_{k_2}), \ldots, F(Y_{k_p}, Y_j) \).

These two conditions insure that we can construct an estimated order. We label one of the subscripts that appears once as \( \alpha_1 \), the other as \( \alpha_n \). Since they communicate and appear only once and all other pairs communicate, the set \( k_1, \ldots, k_p \) must have \( p = n-2 \) and labeling \( k_1 \) as \( \alpha_2 \), \( k_2 \) as \( \alpha_3 \) etc. we obtain the order

\[ Y_{\alpha_1}, Y_{\alpha_2}, \ldots, Y_{\alpha_n} \cdot \]

Let us illustrate the method for a case with \( n=5 \). Suppose

\[ F_{14} > F_{35} > F_{25} > F_{14} > F_{15} > F_{24} > F_{34} > F_{12} \cdot \]

Then we pick \( F_{14}, F_{25}, F_{25} \). If we now pick \( F_{45} \) condition I would be violated. To pick \( F_{23} \) would violate condition II since the subscripts 2, 3, and 5 would not communicate with subscripts 1 and 4. \( F_{13} \) is acceptable. We thus obtain the order \( Y_4 Y_1 Y_2 Y_5 Y_2 \).

Suppose instead

\[ F_{15} > F_{13} > F_{14} > F_{25} > F_{45} > F_{23} > F_{35} > F_{24} > F_{34} > F_{12} \cdot \]

Then we pick \( F_{15}, F_{13}, F_{14} \) violates condition I. \( F_{25} \) is acceptable. \( F_{45} \) violates condition I. \( F_{23} \) violates condition I since then \( n-1 = 4 \) subscripts have appeared twice and none once. \( F_{35} \) fails but \( F_{24} \) is acceptable and we obtain the order \( Y_3 Y_1 Y_5 Y_2 Y_4 \).

In the case of ties we suggest the following. First of all they may not matter, that is both may be included or both may be excluded. If it is the case that including one excludes the other than we suggest
calculating both resulting estimated orders and comparing them via the criterion to be introduced in 3.6.

Confidence in this method will come partially from the belief that A is approximately a Robinson matrix, but what is more important is that it be reasonable that for each i

\[ \min(a_{i-1,i}, a_{i,i+1}) > a_{i,i+j} \quad j = \pm 2, \pm 3, \ldots \]

or even better that for each i

\[ \min(P(F(x_i, x_{i+1}) > c), P(F(x_{i-1}, x_i) > c)) > \max_j P(F(x_i, x_{i+j}) > c) \]

for \( j = \pm 2, \pm 3, \ldots \) and for all \( c > 0 \).

Assumptions (36) or (37) may be unacceptable especially when the \( X_i \) are not approximately equally spaced on the one-dimensional scale. If the \( X_i \) tend to form clusters on the scale then for some \( X_i \)'s the similarities on one side will tend to be larger than those on the other side. Hence the order (33) may not be quite so good and the technique should not be employed.

Once again we must remark that (i) The estimated order is invariant under the observed order of the \( Y \)'s. (ii) Again, this method does not necessarily consider all the entries in the \( F \) matrix. Furthermore, in just being concerned with the largest of the \( F_{ij} \)'s, this method may lose some of the overall perspective in the relationships between the \( Y \)'s which is contained in the whole \( F \) matrix. Of course, if we feel that \( A \) has approximately the additional features mentioned above then ignoring certain observations should not matter. (iii) This technique has been easily programmed and can be done by hand for smaller \( n \).
(iv) Lastly, one might try the modifying procedure introduced in the previous part on (33) to see if any improvement results in terms of the discussion in 3.6.

3.5 Method V.

This method requires only that A be approximately a Robinson matrix in order to be reasonable. It will be seen to be our best technique.

For each \(Y_i\) we construct an estimated order as follows.

Consult the \(F\) matrix and arrange the similarities containing \(Y_i\) in decreasing order, assuming no ties. Without loss of generality we put the \(Y\) most similar to \(Y_i\) to the right and order each subsequent \(Y\) in the following way. Suppose that when we get to the \(k\)-th \(Y\) the order of the previous \(k-1\) is

\[
Y_{\beta_1}, \ldots, Y_{\beta_{k-1}}.
\]

Then if, for this \(Y\)

\[
F(Y,Y_{\beta_1}) > F(Y,Y_{\beta_{k-1}})
\]

put \(Y\) on the left in (38)

while if

\[
F(Y,Y_{\beta_1}) < F(Y,Y_{\beta_{k-1}})
\]

put \(Y\) on the right in (38).

If

\[
F(Y,Y_{\beta_1}) = F(Y,Y_{\beta_{k-1}})
\]

compare \(Y\) with \(Y_{\beta_2}\) and \(Y_{\beta_{k-2}}\).
If at stage $k$ there are two $Y's$, $Y'$ and $Y''$ such that $F(Y_i, Y') = F(Y_i, Y'')$ then if they are assigned to different sides of (38) there is no problem. If they are assigned to the same side (say the left) then a comparison of $Y'$ and $Y''$ with $Y_{\beta_1}$ (or even with $Y_{\beta_2}$ if necessary) will resolve the tie, with a similar resolution on the right.

Thus, in this way we obtain an estimated order for each $i$, $i=1, \ldots, n$

\[(39) \quad Y^{(i)}_{\alpha_1}, Y^{(i)}_{\alpha_2}, \ldots, Y^{(i)}_{\alpha_n} \]

We assume that for each $i$ the orders constructed in (39) are such that $\alpha_1$ indicates the left end of the scale and $\alpha_n$ the right end.

Now let $R^{(i)}(Y_j)$ be the rank of $Y_j$ in the $i$-th estimated order. Calculate

\[(40) \quad T_j = \sum_{i=1}^{n} R^{(i)}(Y_j), \quad j=1, \ldots, n. \]

Note: $\sum_{j=1}^{n} T_j = \frac{n^2(n+1)}{2}$ (as a check).

Arrange the $T_j$ in increasing order and thus obtain the following final estimated order.

\[(41) \quad Y_{\alpha_1}, \ldots, Y_{\alpha_n} \]

where in (41) $Y_{\alpha_j}$ is that $Y$ which corresponds to the $j$-th smallest $T$. In the case of ties amongst the $T's$ each order should be compared by the criterion introduced in 3.6 to determine the "best".

This method has several good features.
(i) Again it is invariant with respect to the order of the observed Y's. (ii) Since we use all the information in the F matrix and since we "average over i" we would expect this technique to "smooth out" possibly spurious similarities and produce a reasonably good estimate. (iii) For anything but small n hand computation is rather laborious. However, the technique is readily programmed.

3.6 Several Comments and a Useful Criterion.

The estimates offered in 3.3, 3.4, and 3.5 proceed from different assumptions on A. In the absence of further information which estimate should one choose?

For a particular permuted order of the Y's one should permute the F matrix accordingly to see if it assumes approximately a Robinson structure. Looking at this "adjusted" F matrix might suggest desirable modifications in the estimate. In fact, with respect to the archaeological problem, this sort of general goal was the aim of Robinson's initial approaches. (See section 4.4 for examples.)

However, we will want to be more precise about the question of improvement and thus in the view of some of our earlier discussion we would like to define a measure of continuity between a particular permutation of the Y_i's i=1,...,n and the observed set of F_{ij}'s. Such a criterion provides a link-up between Approach I and Approach II type techniques.

Using the ideas of Von Neumann and Shepard and Carroll as mentioned in 3.1, given the "distances" (dissimilarities), G_{ij}, between the Y's
and letting \( Y_{\alpha_1}, Y_{\alpha_2}, \ldots, Y_{\alpha_n} \) be an estimated order, then as in expression (12), we define the continuity index of the particular permutation \( \alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n) \) as

\[
C(F; \alpha) = \sum_{i<j} \frac{G_{i,j}}{|\alpha^{-1}(i) - \alpha^{-1}(j)|^{p+r}} = \sum_{i<j} \frac{G_{\alpha_i, \alpha_j}}{|i-j|^{p+r}}
\]

for appropriate \( p \) and \( r \) as discussed earlier.

The permutation that minimizes (42) could be considered as being most continuous with the data. We need not worry about a normalizing factor since \( \alpha \) ranges over a finite set for a particular \( n \).

However, this criterion can not be converted into a technique since we run into the old problem of having too many permutations to consider.

Given several "good" estimated orders or given an order which we might want to try to modify slightly, the criterion (42) provides a way of comparing or verifying improvement in order to select from among those under consideration the estimate which is most continuous with the data.
SECTION IV

SIMULATION MODELS AND APPLICATIONS OF PROCEDURES

4.1 Several Criteria.

Given a collection of random variables $F(Y_i, Y_j)$, $1 \leq i < j \leq n$ along with their joint distribution one might try to compare methods discussed in Section 3 probabilistically according to certain criteria.

We confess at the outset that exact probabilistic evaluations will not be feasible for any of the methods and any assumed distributions except in the case of $n=2$ (which is not meaningful) and $n=3$ which will be discussed shortly. For general $n$ even the simplest of the methods is too complex and the difficulty in handling $\frac{n(n-1)}{2}$ dimensional distributions is such that exact evaluation is not possible. Thus we were led to the simulation which is discussed in the next two sections as the only reasonable way to compare the various techniques.

Let us introduce the criteria of comparison. Suppose $X_1, X_2, \ldots, X_n$ is the true order and in terms of the $X$'s let $X_{\pi_1}, X_{\pi_2}, \ldots, X_{\pi_n}$ be the estimated order. Consider the following criteria.

(i) Probability of correct ordering for a particular method $m_i$.

Probability of correct ordering, designated by $P_{m_i}^{(C.O.)}$, is given by

\[
P_{m_i}^{(C.O.)} = P_{m_i}^{(\pi_j=j, j=1, \ldots, n)}.
\]
(ii) Expected Number of Errors.

For a particular method \( m_i \) let us look at the random variable

\[
d_j = \begin{cases} 
1 & \text{if } \pi_j \neq j \\
0 & \text{if } \pi_j = j.
\end{cases}
\]

(2)

Then the expected number of errors, \( E_{m_i}(d) \), is given by

\[
E_{m_i}(d) = E\left(\sum_{j=1}^{n} d_j\right).
\]

(iii) Expected Sum of Discrepancy.

Again for a particular method \( m_i \) let

\[
S_j = |\pi_j - j|.
\]

(4)

Then the expected sum of discrepancy, \( E_{m_i}(S) \) is given by

\[
E_{m_i}(S) = E\left(\sum_{j=1}^{n} S_j\right).
\]

(5)

More or less equivalently one might be interested in

\[
E_{m_i}(S^2) = E\left(\sum_{j=1}^{n} S_j^2\right).
\]

(6)

(iv) Expected Index of Continuity.

In light of the continuity criterion introduced in 3.6 (42) for a particular method we might be interested in the expected value of this index. We denote this criterion by \( E_{m_i}(c) \).
Of course this index is non-probabilistic and was really designed to look for the best fitting order for a particular set of \( F \)'s and thus it is meaningless to average such numbers. In fact considering the deterministic \( A \) matrix instead of \( F \), it is not even necessarily true that for an arbitrary Robinson \( A \) matrix,

\[
\min_{\pi} \sum_{1 < j} \frac{1-a_{ij}}{(\pi_i - \pi_j)^{p+r}} = \sum_{1 < j} \frac{1-a_{ij}}{(i-j)^{p+r}}
\]

for fixed \( p \) and \( r \) where the minimum is taken over all permutations \( \pi \) of \((1,2,\ldots,n)\).

Thus it will be more valuable to use this criterion in our "real data" applications in 4.4 where it is our only method of comparison among several orders. Therefore this index was not considered as part of the simulation.

In choosing among methods we would like to maximize (i) and minimize (ii) and (iii). We remark that (i) is not too meaningful in a realistic sense. In addition for larger \( n \) it will be a very small number making it difficult to simulate accurately. Thus it was left out of the simulation. Let us look at (ii) and (iii) briefly.

With respect to (ii) for a particular method the resulting permutation \( \pi = (\pi_1, \pi_2, \ldots, \pi_n) \) of the integers \( 1,2,\ldots,n \) may be thought of as being placed in order on squares marked \((1,2,\ldots,n)\) in which case (dropping the \( m_i \) subscript)
(7) \[ E(d) = n - \sum_{i=1}^{n} P(\text{exactly } i \text{ matches}) \]  

The quantity \( P_{i} = P(\text{exactly } i \text{ matches}) \) is the object of the well known "Problème Des Rencontres" except that for this problem \( \pi \) is usually assumed random. Nonetheless by the law of inclusion and exclusion (see Feller, Vol. I, Chapter IV, p. 96) we get for general \( P_{i} \)

(8) \[ P_{i} = S_{1}- (i+1)S_{i+1} + \frac{(i+1)}{2} S_{i+2}, \ldots, + (-1)^{n-i} S_{n} \]

where

\[ S_{1} = \sum_{j=1}^{n} P(\pi_{j}=j) \]

\[ S_{2} = \sum_{i<j} P(\pi_{i}=i, \pi_{j}=j) \]

(9) \[ S_{3} = \sum_{i<j<k} P(\pi_{i}=i, \pi_{j}=j, \pi_{k}=k) \]

\[ \vdots \]

\[ \vdots \]

etc.

Unfortunately the expressions in (9) are untractable to calculate for any of our methods.

With respect to (iii) what are the ranges of \( \Sigma S_{j} \) and \( \Sigma S_{j}^{2} \)? It is easy to see that the largest value of each sum will occur when \( \pi_{1} = n+1-i \) in which case
\[ \Sigma S_j = 2 \sum_{i=1}^{\left\lfloor \frac{n}{2} \right\rfloor} (n-2i+1) \]

so that

\[
\Sigma S_j \begin{cases} 
\frac{n^2}{2} & \text{if } n \text{ is even} \\
\frac{n^2-1}{2} & \text{if } n \text{ is odd}
\end{cases}
\]

and

\[ \Sigma S_j^2 = 2 \sum_{i=1}^{\left\lfloor \frac{n}{2} \right\rfloor} (n-2j+1)^2 \]

\[ = \frac{n^3-n}{3} \text{ for all } n. \]

It is fairly easy to see that \( \Sigma S_j \) must be even in which case the expectation (5) can be written as (dropping the method subscript)

\[ E(S) = 2 \sum_{k=0}^{\ast} k P(\Sigma S_j=2k) \]

where from (10), \(* = \frac{n^2}{4}, \text{ for } n \text{ even; and } \frac{n^2-1}{4}, \text{ for } n \text{ odd.} \)

But

\[ P(\Sigma S_j=2k) = \sum_{\ell=1}^{n} P(\Sigma S_j=2k|\text{exactly } \ell \text{ errors}) P_{n-\ell} \]

when the \( P_{n-\ell} \) are given by (8).

In calculating the conditional probabilities in (13) we can use the notion of partitions. We seek an enumeration of the partitions of \( 2k \) having exactly \( \ell \) parts. Of course barring further assumptions
the different possibilities are not equiprobable so that the problem
does not reduce simply to just calculating \( P_{n-\beta} \). Thus the difficulties
are even worse than before.

Lastly let us show that Methods III, IV and V are equivalent
(ignoring ties) in terms of the previous criteria for \( n=3 \) and for
any distribution on the \( F' \)s. For Method III:

\[
P_{m_3} (C.O.) = P \left[ \begin{align*}
& F(X_1, X_2) + F(X_1, X_3) < F(X_1, X_2) + F(X_2, X_3), \\
& F(X_1, X_2) + F(X_1, X_3) < F(X_1, X_2) + F(X_2, X_3), \\
& \text{and } F(X_1, X_3) < F(X_1, X_2)
\end{align*} \right]
\]

(14)

\[
= P[F(X_1, X_3) < F(X_1, X_2) < F(X_1, X_3)]
\]

\[
+ P[F(X_1, X_3) < F(X_2, X_3) < F(X_1, X_3)]
\]

\[
= P[F(X_1, X_3) < F(X_1, X_2), F(X_1, X_3) < F(X_2, X_3)]
\]

Similarly the probability of obtaining the order \( X_1, X_2, X_3 \) is
given by

(15) \[ P[F(X_1, X_2) < F(X_1, X_3), F(X_1, X_2) < F(X_2, X_3)] \]

and the probability of obtaining the order \( X_2, X_1, X_3 \) is given by
\begin{equation}
P[F(X_2, X_3) < F(X_1, X_2), F(X_2, X_3) < F(X_1, X_2)]
\end{equation}

For Method IV:

\begin{align*}
P_{m_4}(C.0.) &= P[F(X_1, X_2) > F(X_2, X_3) > F(X_1, X_3)] \\
&\quad + P[F(X_2, X_3) > F(X_1, X_2) > F(X_1, X_3)] \\
&= P[F(X_1, X_2) < F(X_1, X_3), F(X_2, X_3) < F(X_1, X_2)]
\end{align*}

as in (14).

Similarly the probability of obtaining the order $X_1, X_3, X_2$ is given by (15) and the probability of obtaining the order $X_2, X_1, X_3$ by (16).

For Method V:

If $F(X_1, X_3) < F(X_1, X_2)$ and $F(X_1, X_3) < F(X_2, X_3)$ then it is clear that we obtain the order $X_1, X_2, X_3$.

If $F(X_2, X_3) < F(X_1, X_2)$ and $F(X_2, X_3) < F(X_1, X_2)$ we obtain the order $X_2, X_1, X_3$ and

If $F(X_1, X_2) < F(X_2, X_3)$ and $F(X_1, X_2) < F(X_1, X_3)$ we obtain the order $X_1, X_2, X_3$. Thus the probabilities are the same as (14), (15) and (16) respectively and thus all three methods are equivalent with respect to the criteria (i), (ii), and (iii).

Methods I and II were not considered in the above due to their non-probabilistic nature. They were also excluded from the simulation for several reasons.

(i) They are more important as scaling techniques and in fact as mentioned earlier, methods III, IV and V will often be used as a good
starting permutation for I and II.

(ii) They are time consuming on the computer for even one iteration, let alone many iterations of many cases in which we were interested for the simulation.

(iii) Again they are non-probabilistic so that interpretation as well as accuracy with respect to the criteria seemed inadequate.

Sternin's [17] techniques were not considered as well. His method would require too much computer time to iterate. In addition the choice of cutoff points for his variants A and B changes from one set of data to the next. These considerations would not concern us in actual applications. In fact in 4.4 we compare for several real data problems his solutions with ours and others.

4.2 Simulation Models.

We wish to construct random \( F \) matrices having the various mean and covariance structures discussed in sections 2.3 and 2.4.

We desire \( 0 \leq F(Y_i, Y_j) \leq 1 \) for all \( i \) and \( j \) and thus we chose the \( a_{ij} \), \( \sigma^2 \), \( m \) and \( k \) accordingly. For simplicity and without much loss we assume the \( F \)'s follow a multivariate normal distribution.

We thus inserted the modification that if \( F(Y_i, Y_j) > 1 \) set \( F(Y_i, Y_j) = 1^- \) and if \( F(Y_i, Y_j) < 0 \) set \( F(Y_i, Y_j) = 0^+ \). For our choice of parameters we almost never had to apply this rule.

Our approach follows:

For each \( n \) and corresponding choice of \( A \) and \( \Sigma \) we performed \( T \)
iterations. For each iteration we constructed $\frac{n(n-1)}{2}$ normal $F(X_i, X_j)$'s having the above $A$ and $\Sigma$ (we show shortly how this was done) and arranged them in a matrix form. Thus having the true order we applied a known random permutation to the subscripts of the $F$'s and reconstructed a new $F$ matrix as our observed set of $F(Y_i, Y_j)$'s to which we applied each of the methods III, IV, and V, in attempting to regain approximately the true order. It was assumed that the question of reversibility was resolved.

At each iteration for each method we calculated criteria (ii) and (iii). In addition we calculated one of the estimates considered in section V, $b = \bar{W}$ defined by

$$
 b = \frac{1}{n-1} \sum_{i=1}^{n-1} \frac{F(Y_{\alpha_i}, Y_{\alpha_{i+1}})}{F(Y_i, Y_j)}.
$$

(17)

We then averaged the above three estimates over the $T$ iterations, calculated their respective standard errors and tabulated the results for the various cases. An enumeration of some of the cases considered along with the corresponding outputs appears in 4.3. The size of $T$ was chosen such that the standard error of the averaged estimates was small as was the ratio of the standard error of the estimate to the estimate itself.

We generated $F$'s having a particular mean matrix $A = \{a_{ij}\}$ and each of the following $\Sigma$'s

(i) $\Sigma = \Sigma(\sigma^2, m, l)$
\[(i)\] \[\Sigma = \Sigma(\sigma^2, m, 0)\]

\[(ii)\] \[\Sigma = \Sigma(\sigma^2, m, k)\]

Given \(\frac{n(n-1)}{2} + n\) independent normal random variables with mean zero and variance one, labeled \(U_{ij}, 1 \leq i < j \leq n\) and \(V_i, 1 \leq i \leq n\), respectively, if

\[(18)\] \[F(X_i, X_j) = \sigma \sqrt{1-m} U_{ij} + \sigma \sqrt{m} V_i + \alpha_{ij}\]

then \(F\) will have covariance structure given by (i). If

\[(19)\] \[F(X_i, X_j) = \sigma \sqrt{1-2m} U_{ij} + \sigma \sqrt{m} (V_i + V_j) + \alpha_{ij}\]

then \(F\) will have covariance structure given by (ii) and if

\[(20)\] \[F(X_i, X_j) = \frac{\sigma U_{ij} + \sigma \alpha (V_i + V_j) + \sigma \beta \sum_{k=1, j}^{n} V_k}{1 + 2\alpha^2 + (n-2)\beta^2} + \alpha_{ij}\]

where \(\alpha\) and \(\beta\) are chosen appropriately for \(m\) and \(k\), then \(F\) will have covariance structure given by (iii). In fact

\[(21)\] \[m = \frac{\alpha^2 + (n-3)\beta^2 + 2\alpha\beta}{1 + 2\alpha^2 + (n-2)\beta^2}\]

and

\[(22)\] \[k = \frac{(n-4)\beta^2 + 4\alpha\beta}{\alpha^2 + (n-3)\beta^2 + 2\alpha\beta}\]
4.3 Tables and Results.

In the light of previous discussion one would expect Method V to be the best among Methods III, IV, and V. Our simulation verified this expectation with respect to the previous criteria for various choices of n, A and Σ with only one noteworthy exception. In the case that A is a Green's matrix with \( c_1 = e^i \) (see Table II) Method IV worked best. This is not surprising in view of the fact that for this case \( a_{1,1+1} = \frac{1}{e} \approx .37 \) and \( a_{1,1+2} = \frac{1}{2} \approx .14 \) which is a very favorable situation for Method IV.

The number of replications, T, was set at 100.

A sample of some typical results for several different cases is given. For Tables I, II and III n is set equal to 10. For Table IV n = 20. In Tables I and IV A was chosen to be Toeplitz decreasing. In Table II A was taken to be a Green's matrix. In Table III A has a systematically defined structure which is not stationary. In each table at least two different covariance matrices are considered.

We must remark that in some cases there was difficulty in achieving convergence of the simulation estimates with respect to the criteria due to the small number of mistakes made by a particular method but in all cases the differences between the method producing the best orders and the others were significant.

Since, in Tables I, II, and IV, A is stationary it makes sense to estimate \( b_1 \). We note that in Table I, Methods IV and V are about the same. Method III is unacceptable and all seem to underestimate
slightly. In Table II only the estimate of Method IV is acceptable.
In Table IV once again Methods IV and V are about the same and all
seem to underestimate slightly. See Section V for further discussion.
**TABLE I**

$n = 10$, $A$ is Toeplitz Decreasing with $b_1 = 1 - (.1)i$ ($b_1 = .9$)

Case I: $\Sigma = \Sigma(0.10, 0.50, 1.00)$

<table>
<thead>
<tr>
<th></th>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E(d)$</td>
<td>2.980</td>
<td>2.190</td>
<td>0.420</td>
</tr>
<tr>
<td>S.E.$E(d)$</td>
<td>0.208</td>
<td>0.206</td>
<td>0.086</td>
</tr>
<tr>
<td>$E(s)$</td>
<td>3.300</td>
<td>2.700</td>
<td>0.420</td>
</tr>
<tr>
<td>S.E.$E(s)$</td>
<td>0.245</td>
<td>0.283</td>
<td>0.086</td>
</tr>
<tr>
<td>$\hat{b}_1$</td>
<td>0.878</td>
<td>0.891</td>
<td>0.888</td>
</tr>
<tr>
<td>S.E.$\hat{b}_1$</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
</tr>
</tbody>
</table>

Case II: $\Sigma = \Sigma(0.05, 0.40, 0)$

<table>
<thead>
<tr>
<th></th>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E(d)$</td>
<td>1.960</td>
<td>1.550</td>
<td>0.240</td>
</tr>
<tr>
<td>S.E.$E(d)$</td>
<td>0.185</td>
<td>0.181</td>
<td>0.065</td>
</tr>
<tr>
<td>$E(s)$</td>
<td>2.060</td>
<td>1.960</td>
<td>0.240</td>
</tr>
<tr>
<td>S.E.$E(s)$</td>
<td>0.204</td>
<td>0.257</td>
<td>0.065</td>
</tr>
<tr>
<td>$\hat{b}_1$</td>
<td>0.889</td>
<td>0.896</td>
<td>0.899</td>
</tr>
<tr>
<td>S.E.$\hat{b}_1$</td>
<td>0.003</td>
<td>0.002</td>
<td>0.002</td>
</tr>
</tbody>
</table>
TABLE II

\( n = 10, \ A \) is Green's Matrix with \( c_1 = e^i, \ (b_1 = .37) \)

Case I: \( \Sigma = \Sigma(0.05, 0.50, 1.00) \)

<table>
<thead>
<tr>
<th></th>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>E(a)</td>
<td>5.650</td>
<td>0</td>
<td>3.450</td>
</tr>
<tr>
<td>S.E.(E(d))</td>
<td>0.148</td>
<td>0</td>
<td>0.224</td>
</tr>
<tr>
<td>E(s)</td>
<td>13.660</td>
<td>0</td>
<td>4.660</td>
</tr>
<tr>
<td>S.E.(E(s))</td>
<td>0.479</td>
<td>0</td>
<td>0.391</td>
</tr>
<tr>
<td>( \hat{b}_1 )</td>
<td>0.227</td>
<td>0.372</td>
<td>0.259</td>
</tr>
<tr>
<td>S.E.(( \hat{b}_1 ))</td>
<td>0.009</td>
<td>0.008</td>
<td>0.009</td>
</tr>
</tbody>
</table>

Case II: \( \Sigma = \Sigma(0.05, 0.40, 0) \)

<table>
<thead>
<tr>
<th></th>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>E(a)</td>
<td>6.400</td>
<td>0</td>
<td>4.620</td>
</tr>
<tr>
<td>S.E.(E(d))</td>
<td>0.162</td>
<td>0</td>
<td>0.218</td>
</tr>
<tr>
<td>E(s)</td>
<td>15.720</td>
<td>0</td>
<td>6.600</td>
</tr>
<tr>
<td>S.E.(E(s))</td>
<td>0.589</td>
<td>0</td>
<td>0.429</td>
</tr>
<tr>
<td>( \hat{b}_1 )</td>
<td>0.223</td>
<td>0.370</td>
<td>0.275</td>
</tr>
<tr>
<td>S.E.(( \hat{b}_1 ))</td>
<td>0.006</td>
<td>0.003</td>
<td>0.006</td>
</tr>
</tbody>
</table>

Case III: \( \Sigma = \Sigma(0.10, 0.70, 0.80) \)

<table>
<thead>
<tr>
<th></th>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>E(a)</td>
<td>5.250</td>
<td>0</td>
<td>3.020</td>
</tr>
<tr>
<td>S.E.(E(d))</td>
<td>0.173</td>
<td>0</td>
<td>0.225</td>
</tr>
<tr>
<td>E(s)</td>
<td>11.920</td>
<td>0</td>
<td>3.860</td>
</tr>
<tr>
<td>S.E.(E(s))</td>
<td>0.508</td>
<td>0</td>
<td>0.346</td>
</tr>
<tr>
<td>( \hat{b}_1 )</td>
<td>0.246</td>
<td>0.371</td>
<td>0.311</td>
</tr>
<tr>
<td>S.E.(( \hat{b}_1 ))</td>
<td>0.006</td>
<td>0.004</td>
<td>0.007</td>
</tr>
</tbody>
</table>
TABLE III

\( n = 10, \ \text{A is Non-Stationary with} \ a_{ij} = 1 - \frac{|j-i|}{10} + (0.02)(-1)^i \)

Case I: \( \Sigma = \Sigma(0.10, 0.50, 1.00) \)

<table>
<thead>
<tr>
<th></th>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>E(d)</td>
<td>2.670</td>
<td>2.270</td>
<td>0.280</td>
</tr>
<tr>
<td>S.E.(E(d))</td>
<td>0.182</td>
<td>0.170</td>
<td>0.070</td>
</tr>
<tr>
<td>E(s)</td>
<td>3.000</td>
<td>2.700</td>
<td>0.280</td>
</tr>
<tr>
<td>S.E.(E(s))</td>
<td>0.217</td>
<td>0.235</td>
<td>0.070</td>
</tr>
</tbody>
</table>

Case II: \( \Sigma = \Sigma(0.05, 0.40, 0) \)

<table>
<thead>
<tr>
<th></th>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>E(d)</td>
<td>2.210</td>
<td>1.360</td>
<td>0.260</td>
</tr>
<tr>
<td>S.E.(E(d))</td>
<td>0.174</td>
<td>0.172</td>
<td>0.067</td>
</tr>
<tr>
<td>E(s)</td>
<td>2.320</td>
<td>1.620</td>
<td>0.260</td>
</tr>
<tr>
<td>S.E.(E(s))</td>
<td>0.184</td>
<td>0.228</td>
<td>0.067</td>
</tr>
</tbody>
</table>
TABLE IV

\( n = 20, \ A \text{ is Toeplitz Decreasing with } b_1 = 1-(.05)i \) \quad (b_1 = .95)

Case I: \( \Sigma = \Sigma(0.05, 0.50, 1.00) \)

<table>
<thead>
<tr>
<th></th>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E(d) )</td>
<td>5.780</td>
<td>5.000</td>
<td>0.020</td>
</tr>
<tr>
<td>S.E.((E(d)))</td>
<td>0.307</td>
<td>0.337</td>
<td>-</td>
</tr>
<tr>
<td>( E(s) )</td>
<td>6.640</td>
<td>7.040</td>
<td>0.020</td>
</tr>
<tr>
<td>S.E.((E(s)))</td>
<td>0.378</td>
<td>0.587</td>
<td>-</td>
</tr>
<tr>
<td>( \hat{b}_1 )</td>
<td>0.938</td>
<td>0.945</td>
<td>0.947</td>
</tr>
<tr>
<td>S.E.((\hat{b}_1))</td>
<td>0.004</td>
<td>0.003</td>
<td>0.003</td>
</tr>
</tbody>
</table>

Case II: \( \Sigma = \Sigma(0.05, 0.40, 0) \)

<table>
<thead>
<tr>
<th></th>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E(d) )</td>
<td>9.540</td>
<td>8.530</td>
<td>2.620</td>
</tr>
<tr>
<td>S.E.((E(d)))</td>
<td>0.307</td>
<td>0.362</td>
<td>0.173</td>
</tr>
<tr>
<td>( E(s) )</td>
<td>12.600</td>
<td>15.220</td>
<td>2.680</td>
</tr>
<tr>
<td>S.E.((E(s)))</td>
<td>0.463</td>
<td>1.053</td>
<td>0.182</td>
</tr>
<tr>
<td>( \hat{b}_1 )</td>
<td>0.923</td>
<td>0.939</td>
<td>0.940</td>
</tr>
<tr>
<td>S.E.((\hat{b}_1))</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
</tr>
</tbody>
</table>

68
4.4 Applications to Archaeological Data.

In Robinson's [14] paper he discusses two examples in which archaeological deposits are to be ordered chronologically. One case produced the $8 \times 8$ similarity matrix given in Table V in initial form and in Table VI according to Robinson's estimated final order, that is, Robinson's $\alpha$ vector for $Y_{\alpha_1}, Y_{\alpha_2}, \ldots, Y_{\alpha_n}$ would be

$$\alpha = (1, 6, 7, 4, 8, 5, 2, 3).$$

The other case produced a $17 \times 17$ similarity matrix which is given in Table VII and indicates a late, perhaps final stage of rearrangement according to Robinson.

In [10], Meggers and Evans produced an $8 \times 8$ matrix as a result of findings at 8 sites (cemeteries) in the Amazon region. The matrix is given in Table VIII reflecting the final estimated order achieved by Meggers and Evans.

In all the above cases the choice of similarity measure $F$ was twice the expression given in 2.1 (9) with $p=1$.

In addition since Stetterin [17] applied his procedure to the above three cases it seems reasonable to compare our best method (Method V) with Stetterin's and with the above estimates in these cases. We consider several other reasonable orders in each case. We assume that Table VI, Table VII and Table VIII respectively represent the observed order of the $Y$'s. We let $\alpha_R$ be Robinson's estimate, $\alpha_M$ be Meggers and Evans', $\alpha_S$ be Stetterin's and $\alpha_G$ be ours.

69
For each estimated order we calculate \( C(F; \alpha) \) given by 3.6(42) with \( p = 1 \) and \( r = 2 \) (although the choice of \( r \) did not seem to affect the conclusions).

For Table VI.

\[
\begin{align*}
\alpha_R &= (1, 2, 3, 4, 5, 6, 7, 8), \quad C(F; \alpha_R) = 2.592 \\
\alpha_S &= (1, 2, 3, 4, 5, 6, 7, 8), \quad C(F; \alpha_S) = 2.592 \\
\alpha_G &= (1, 2, 3, 4, 5, 6, 7, 8), \quad C(F; \alpha_G) = 2.592 \\
\alpha_1 &= (1, 2, 3, 4, 5, 6, 8, 7), \quad C(F; \alpha_1) = 2.587 \\
\alpha_2 &= (1, 3, 2, 4, 5, 6, 7, 8), \quad C(F; \alpha_2) = 2.739 \\
\alpha_3 &= (1, 3, 2, 4, 5, 6, 8, 7), \quad C(F; \alpha_3) = 2.734
\end{align*}
\]

Thus \( \alpha_R = \alpha_S = \alpha_3 \) is possibly slightly better. Interestingly enough our Method IV produced \( \alpha_1 \) and our Method V is virtually a toss up between \( \alpha_G \) and \( \alpha_1 \), that is \( T_7 = 59 \) and \( T_8 = 60 \).

For Table VII.

\[
\begin{align*}
\alpha_R &= (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17), \quad C(F; \alpha_R) = 4.146 \\
\alpha_S &= (1, 2, 3, 4, 9, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17), \quad C(F; \alpha_S) = 4.092 \\
\alpha_G &= (1, 2, 4, 3, 5, 9, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17), \quad C(F; \alpha_G) = 3.933 \\
\alpha_1 &= (1, 2, 4, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17), \quad C(F; \alpha_1) = 4.061 \\
\alpha_2 &= (1, 2, 4, 5, 9, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17), \quad C(F; \alpha_2) = 4.045 \\
\alpha_3 &= (1, 2, 3, 4, 5, 9, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17), \quad C(F; \alpha_3) = 4.018
\end{align*}
\]

Clearly \( \alpha_G \) seems the best as is further indicated by the fact that \( \alpha_1 \) is better than \( \alpha_R \), \( \alpha_2 \) is better than \( \alpha_S \) and \( \alpha_G \) is better than \( \alpha_3 \).
For Table VIII.

\[ \alpha_M = (1, 2, 3, 4, 5, 6, 7, 8), \quad C(F; \alpha_M) = 2.712 \]

\[ \alpha_S = (2, 3, 1, 4, 5, 6, 7, 8), \quad C(F; \alpha_S) = 2.705 \]

\[ \alpha_G = (2, 3, 1, 4, 5, 6, 7, 8), \quad C(F; \alpha_G) = 2.653 \]

\[ \alpha_1 = (1, 2, 3, 4, 6, 5, 7, 8), \quad C(F; \alpha_1) = 2.663 \]

\[ \alpha_2 = (1, 2, 3, 6, 4, 5, 7, 8), \quad C(F; \alpha_2) = 2.671 \]

\[ \alpha_3 = (2, 3, 1, 6, 4, 5, 7, 8), \quad C(F; \alpha_3) = 2.769 \]

Once again \( \alpha_G \) seems best, in particular when compared with \( \alpha_S \) and \( \alpha_1 \).

Finally we remark that Method V worked remarkably well especially in the absence of any knowledge as to the feasibility of the assumption of an underlying true Robinson A matrix in the above cases.
TABLE V

(8 x 8 Matrix - Robinson [14])

<table>
<thead>
<tr>
<th></th>
<th>Y_1</th>
<th>Y_2</th>
<th>Y_3</th>
<th>Y_4</th>
<th>Y_5</th>
<th>Y_6</th>
<th>Y_7</th>
<th>Y_8</th>
<th>Σ F(Y_i, Y_j)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y_1</td>
<td>2.00</td>
<td>0.05</td>
<td>0.01</td>
<td>0.39</td>
<td>0.04</td>
<td>0.66</td>
<td>0.67</td>
<td>0.11</td>
<td>3.95</td>
</tr>
<tr>
<td>Y_2</td>
<td>0.05</td>
<td>2.00</td>
<td>1.96</td>
<td>1.08</td>
<td>1.95</td>
<td>0.03</td>
<td>0.29</td>
<td>1.14</td>
<td>8.50</td>
</tr>
<tr>
<td>Y_3</td>
<td>0.01</td>
<td>1.96</td>
<td>2.00</td>
<td>1.07</td>
<td>1.96</td>
<td>0.01</td>
<td>0.26</td>
<td>1.15</td>
<td>8.42</td>
</tr>
<tr>
<td>Y_4</td>
<td>0.39</td>
<td>1.08</td>
<td>1.07</td>
<td>2.00</td>
<td>1.10</td>
<td>0.50</td>
<td>0.82</td>
<td>1.72</td>
<td>8.68</td>
</tr>
<tr>
<td>Y_5</td>
<td>0.04</td>
<td>1.95</td>
<td>1.96</td>
<td>1.10</td>
<td>2.00</td>
<td>0.04</td>
<td>0.30</td>
<td>1.19</td>
<td>8.50</td>
</tr>
<tr>
<td>Y_6</td>
<td>0.66</td>
<td>0.03</td>
<td>0.01</td>
<td>0.50</td>
<td>0.04</td>
<td>2.00</td>
<td>1.01</td>
<td>0.27</td>
<td>4.52</td>
</tr>
<tr>
<td>Y_7</td>
<td>0.69</td>
<td>0.29</td>
<td>0.26</td>
<td>0.82</td>
<td>0.30</td>
<td>1.01</td>
<td>2.00</td>
<td>0.66</td>
<td>6.03</td>
</tr>
<tr>
<td>Y_8</td>
<td>0.11</td>
<td>1.14</td>
<td>1.15</td>
<td>1.72</td>
<td>1.19</td>
<td>0.27</td>
<td>0.66</td>
<td>2.00</td>
<td>8.24</td>
</tr>
</tbody>
</table>

TABLE VI

(8 x 8 Matrix - Robinson [14])

<table>
<thead>
<tr>
<th></th>
<th>Y_1</th>
<th>Y_2</th>
<th>Y_3</th>
<th>Y_4</th>
<th>Y_5</th>
<th>Y_6</th>
<th>Y_7</th>
<th>Y_8</th>
<th>Σ F(Y_i, Y_j)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y_1</td>
<td>2.00</td>
<td>0.66</td>
<td>0.69</td>
<td>0.39</td>
<td>0.11</td>
<td>0.04</td>
<td>0.05</td>
<td>0.01</td>
<td>3.95</td>
</tr>
<tr>
<td>Y_2</td>
<td>0.66</td>
<td>2.00</td>
<td>1.01</td>
<td>0.50</td>
<td>0.27</td>
<td>0.04</td>
<td>0.03</td>
<td>0.01</td>
<td>4.52</td>
</tr>
<tr>
<td>Y_3</td>
<td>0.69</td>
<td>1.01</td>
<td>2.00</td>
<td>0.82</td>
<td>0.66</td>
<td>0.30</td>
<td>0.29</td>
<td>0.26</td>
<td>6.03</td>
</tr>
<tr>
<td>Y_4</td>
<td>0.39</td>
<td>0.50</td>
<td>0.82</td>
<td>2.00</td>
<td>1.72</td>
<td>1.10</td>
<td>1.08</td>
<td>1.07</td>
<td>8.68</td>
</tr>
<tr>
<td>Y_5</td>
<td>0.11</td>
<td>0.27</td>
<td>0.66</td>
<td>1.72</td>
<td>2.00</td>
<td>1.19</td>
<td>1.14</td>
<td>1.15</td>
<td>8.24</td>
</tr>
<tr>
<td>Y_6</td>
<td>0.04</td>
<td>0.04</td>
<td>0.30</td>
<td>1.10</td>
<td>1.19</td>
<td>2.00</td>
<td>1.95</td>
<td>1.96</td>
<td>8.58</td>
</tr>
<tr>
<td>Y_7</td>
<td>0.05</td>
<td>0.03</td>
<td>0.29</td>
<td>1.08</td>
<td>1.14</td>
<td>1.95</td>
<td>2.00</td>
<td>1.96</td>
<td>8.50</td>
</tr>
<tr>
<td>Y_8</td>
<td>0.01</td>
<td>0.01</td>
<td>0.26</td>
<td>1.07</td>
<td>1.15</td>
<td>1.96</td>
<td>1.96</td>
<td>2.00</td>
<td>8.42</td>
</tr>
</tbody>
</table>
TABLE VII

(17 x 17 Matrix - Robinson [14])

<table>
<thead>
<tr>
<th>Y_1</th>
<th>Y_2</th>
<th>Y_3</th>
<th>Y_4</th>
<th>Y_5</th>
<th>Y_6</th>
<th>Y_7</th>
<th>Y_8</th>
<th>Y_9</th>
<th>Y_10</th>
<th>Y_11</th>
<th>Y_12</th>
<th>Y_13</th>
<th>Y_14</th>
<th>Y_15</th>
<th>Y_16</th>
<th>Y_17</th>
<th>( \sum_{j=1}^{17} F(Y_{ij} Y_{ij}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y_1</td>
<td>2.00</td>
<td>0.92</td>
<td>1.32</td>
<td>1.05</td>
<td>1.24</td>
<td>1.07</td>
<td>1.16</td>
<td>1.08</td>
<td>1.13</td>
<td>1.06</td>
<td>0.91</td>
<td>1.07</td>
<td>0.92</td>
<td>0.84</td>
<td>0.82</td>
<td>0.60</td>
<td>0.78</td>
</tr>
<tr>
<td>Y_2</td>
<td>0.92</td>
<td>2.00</td>
<td>1.05</td>
<td>1.24</td>
<td>1.07</td>
<td>1.16</td>
<td>1.08</td>
<td>1.13</td>
<td>1.06</td>
<td>0.91</td>
<td>1.07</td>
<td>0.92</td>
<td>0.84</td>
<td>0.82</td>
<td>0.60</td>
<td>0.78</td>
<td>0.64</td>
</tr>
<tr>
<td>Y_3</td>
<td>1.32</td>
<td>1.05</td>
<td>2.00</td>
<td>1.53</td>
<td>1.45</td>
<td>1.35</td>
<td>1.46</td>
<td>1.38</td>
<td>1.50</td>
<td>1.28</td>
<td>1.35</td>
<td>1.26</td>
<td>1.10</td>
<td>1.00</td>
<td>0.91</td>
<td>1.02</td>
<td>1.05</td>
</tr>
<tr>
<td>Y_4</td>
<td>1.04</td>
<td>1.24</td>
<td>1.53</td>
<td>2.00</td>
<td>1.44</td>
<td>1.42</td>
<td>1.50</td>
<td>1.51</td>
<td>1.68</td>
<td>1.40</td>
<td>1.34</td>
<td>1.21</td>
<td>1.07</td>
<td>1.13</td>
<td>1.00</td>
<td>0.76</td>
<td>0.99</td>
</tr>
<tr>
<td>Y_5</td>
<td>1.01</td>
<td>1.07</td>
<td>1.45</td>
<td>1.44</td>
<td>2.00</td>
<td>1.47</td>
<td>1.66</td>
<td>1.67</td>
<td>1.54</td>
<td>1.59</td>
<td>1.49</td>
<td>1.50</td>
<td>1.39</td>
<td>1.47</td>
<td>1.33</td>
<td>1.31</td>
<td>1.14</td>
</tr>
<tr>
<td>Y_6</td>
<td>0.84</td>
<td>1.16</td>
<td>1.35</td>
<td>1.42</td>
<td>1.47</td>
<td>2.00</td>
<td>1.64</td>
<td>1.69</td>
<td>1.66</td>
<td>1.57</td>
<td>1.55</td>
<td>1.54</td>
<td>1.42</td>
<td>1.32</td>
<td>1.29</td>
<td>1.19</td>
<td>1.31</td>
</tr>
<tr>
<td>Y_7</td>
<td>0.95</td>
<td>1.08</td>
<td>1.46</td>
<td>1.50</td>
<td>1.66</td>
<td>1.64</td>
<td>2.00</td>
<td>1.81</td>
<td>1.70</td>
<td>1.68</td>
<td>1.67</td>
<td>1.64</td>
<td>1.48</td>
<td>1.40</td>
<td>1.38</td>
<td>1.29</td>
<td>1.37</td>
</tr>
<tr>
<td>Y_8</td>
<td>0.88</td>
<td>1.13</td>
<td>1.38</td>
<td>1.51</td>
<td>1.67</td>
<td>1.69</td>
<td>1.81</td>
<td>2.00</td>
<td>1.68</td>
<td>1.58</td>
<td>1.67</td>
<td>1.60</td>
<td>1.51</td>
<td>1.43</td>
<td>1.40</td>
<td>1.43</td>
<td>1.33</td>
</tr>
<tr>
<td>Y_9</td>
<td>0.94</td>
<td>1.06</td>
<td>1.50</td>
<td>1.68</td>
<td>1.54</td>
<td>1.66</td>
<td>1.70</td>
<td>1.68</td>
<td>2.00</td>
<td>1.53</td>
<td>1.46</td>
<td>1.49</td>
<td>1.29</td>
<td>1.28</td>
<td>1.20</td>
<td>1.20</td>
<td>1.21</td>
</tr>
<tr>
<td>Y_{10}</td>
<td>0.92</td>
<td>0.91</td>
<td>1.28</td>
<td>1.40</td>
<td>1.59</td>
<td>1.57</td>
<td>1.68</td>
<td>1.58</td>
<td>1.53</td>
<td>2.00</td>
<td>1.47</td>
<td>1.59</td>
<td>1.49</td>
<td>1.48</td>
<td>1.43</td>
<td>1.34</td>
<td>1.39</td>
</tr>
<tr>
<td>Y_{11}</td>
<td>0.81</td>
<td>1.07</td>
<td>1.35</td>
<td>1.34</td>
<td>1.49</td>
<td>1.55</td>
<td>1.67</td>
<td>1.67</td>
<td>1.46</td>
<td>1.47</td>
<td>2.00</td>
<td>1.53</td>
<td>1.52</td>
<td>1.43</td>
<td>1.27</td>
<td>1.52</td>
<td>1.57</td>
</tr>
<tr>
<td>Y_{12}</td>
<td>0.72</td>
<td>0.92</td>
<td>1.26</td>
<td>1.21</td>
<td>1.50</td>
<td>1.54</td>
<td>1.64</td>
<td>1.60</td>
<td>1.49</td>
<td>1.59</td>
<td>1.53</td>
<td>2.00</td>
<td>1.78</td>
<td>1.54</td>
<td>1.64</td>
<td>1.49</td>
<td>1.52</td>
</tr>
<tr>
<td>Y_{13}</td>
<td>0.55</td>
<td>0.84</td>
<td>1.10</td>
<td>1.07</td>
<td>1.39</td>
<td>1.42</td>
<td>1.48</td>
<td>1.51</td>
<td>1.29</td>
<td>1.49</td>
<td>1.52</td>
<td>1.78</td>
<td>2.00</td>
<td>1.71</td>
<td>1.72</td>
<td>1.66</td>
<td>1.61</td>
</tr>
<tr>
<td>Y_{14}</td>
<td>0.46</td>
<td>0.82</td>
<td>1.00</td>
<td>1.13</td>
<td>1.47</td>
<td>1.32</td>
<td>1.40</td>
<td>1.29</td>
<td>1.48</td>
<td>1.43</td>
<td>1.54</td>
<td>1.71</td>
<td>2.00</td>
<td>1.75</td>
<td>1.52</td>
<td>1.45</td>
<td>23.20</td>
</tr>
<tr>
<td>Y_{15}</td>
<td>0.44</td>
<td>0.60</td>
<td>0.91</td>
<td>1.00</td>
<td>1.33</td>
<td>1.29</td>
<td>1.38</td>
<td>1.40</td>
<td>1.28</td>
<td>1.43</td>
<td>1.27</td>
<td>1.64</td>
<td>1.72</td>
<td>1.75</td>
<td>2.00</td>
<td>1.45</td>
<td>1.39</td>
</tr>
<tr>
<td>Y_{16}</td>
<td>0.51</td>
<td>0.78</td>
<td>1.02</td>
<td>0.76</td>
<td>1.11</td>
<td>1.19</td>
<td>1.39</td>
<td>1.45</td>
<td>1.20</td>
<td>1.34</td>
<td>1.52</td>
<td>1.49</td>
<td>1.66</td>
<td>1.52</td>
<td>1.45</td>
<td>2.00</td>
<td>1.54</td>
</tr>
<tr>
<td>Y_{17}</td>
<td>0.49</td>
<td>0.64</td>
<td>1.05</td>
<td>0.99</td>
<td>1.14</td>
<td>1.31</td>
<td>1.37</td>
<td>1.33</td>
<td>1.21</td>
<td>1.39</td>
<td>1.57</td>
<td>1.52</td>
<td>1.61</td>
<td>1.45</td>
<td>1.39</td>
<td>1.54</td>
<td>2.00</td>
</tr>
<tr>
<td></td>
<td>$Y_1$</td>
<td>$Y_2$</td>
<td>$Y_3$</td>
<td>$Y_4$</td>
<td>$Y_5$</td>
<td>$Y_6$</td>
<td>$Y_7$</td>
<td>$Y_8$</td>
<td>$\sum_{j=1}^{8} F(Y_i,Y_j)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_1$</td>
<td>2.00</td>
<td>1.29</td>
<td>1.11</td>
<td>1.21</td>
<td>1.11</td>
<td>1.26</td>
<td>1.13</td>
<td>0.95</td>
<td>10.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_2$</td>
<td>1.29</td>
<td>2.00</td>
<td>1.15</td>
<td>0.95</td>
<td>0.72</td>
<td>0.79</td>
<td>0.84</td>
<td>0.60</td>
<td>8.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_3$</td>
<td>1.11</td>
<td>1.15</td>
<td>2.00</td>
<td>0.90</td>
<td>0.95</td>
<td>0.95</td>
<td>0.91</td>
<td>0.78</td>
<td>8.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_4$</td>
<td>1.21</td>
<td>0.95</td>
<td>0.90</td>
<td>2.00</td>
<td>1.10</td>
<td>1.49</td>
<td>1.30</td>
<td>1.14</td>
<td>10.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_5$</td>
<td>1.11</td>
<td>0.72</td>
<td>0.95</td>
<td>1.10</td>
<td>2.00</td>
<td>1.44</td>
<td>1.36</td>
<td>1.02</td>
<td>9.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_6$</td>
<td>1.26</td>
<td>0.79</td>
<td>0.95</td>
<td>1.49</td>
<td>1.44</td>
<td>2.00</td>
<td>1.63</td>
<td>1.32</td>
<td>10.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_7$</td>
<td>1.13</td>
<td>0.84</td>
<td>0.91</td>
<td>1.30</td>
<td>1.36</td>
<td>1.63</td>
<td>2.00</td>
<td>1.45</td>
<td>10.62</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Y_8$</td>
<td>0.95</td>
<td>0.60</td>
<td>0.78</td>
<td>1.14</td>
<td>1.02</td>
<td>1.32</td>
<td>1.45</td>
<td>2.00</td>
<td>9.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
SECTION V

ESTIMATION

5.1 Introduction and Approach.

Throughout this section let us define the following

1. $Y_1, Y_2, \ldots, Y_n$ is the observed order of the $Y$'s

2. $\tilde{Y}_{\alpha_1}, \tilde{Y}_{\alpha_2}, \ldots, \tilde{Y}_{\alpha_n}$ is an estimate of the true order of the $Y$'s

3. $X_1, X_2, \ldots, X_n$ is the true order.

Suppose we are interested in problems of estimation with respect to given $A$ and $\Sigma$ matrices for our observed $F$'s. In particular we shall be interested in some of the special forms of $A$ and $\Sigma$ discussed in Section II.

The obvious major difficulty is that we do not observe $F(X_i, X_j)$ but $F(Y_i, Y_j)$ for all $i$ and $j$, and of course do not know the permutation relation between the two orders.

The above, however, suggests a two-stage approach to the problem.

(I). Under the assumption that the $F(X_i, X_j)$ have a certain distributional structure, construct "best" estimates given the $F(X_i, X_j)$.

(II). Then using a "best" permutation obtained say by a method in section III we can obtain the estimates in terms of the $F(Y_i, Y_j)$ by inversion.
In 5.2 we will exhibit two special cases where we can avoid this two stage approach.

Thus from here on we need only deal with stage I, that is, considering only the $F(X_i, X_j)$ except for the following comments.

With the assumption of an arbitrary mean structure $A$ we can not hope to do much. On the other hand it may be unreasonable to assume, for example, that $A$ has a Toeplitz-decreasing or Green's structure. We adopt a middle course and assume only that

\[(4) \quad E[F(X_i, X_{i+1})] = a_{i,i+1} = b \quad \text{for all } i\]

and the usual assumption

\[(5) \quad b \geq a_{i,j} \quad \text{for all } i,j, i \neq j.\]

Thus we will restrict ourselves to looking at the marginal distribution of $F(X_1, X_2), \ldots, F(X_{n-1}, X_n)$.

In this way we obtain the largest possible sample size $(n-1)$ using an assumption like (4) and take advantage of the fact that $F$ tends to be a better "large similarity" discriminator than a small one (as discussed in Section III, 3.3). Obviously, the feasibility of (4) must be considered in any application.

Now with respect to stage II, each estimated order (2) gives as our $(n-1)$ observations $F(Y_{\alpha_i}, Y_{\alpha_{i+1}})$, $i=1, \ldots, n-1$. Recalling our techniques discussed in Section III we note that Method IV was designed specifically to try to pick out $F(X_i, X_{i+1})$, $i=1, \ldots, n-1$ and works well when (4) and (5) hold. Thus we might consider its estimated order
in this context more carefully even if it is not necessarily the "best"
permutation estimate in accordance with the criterion in 3.6. In fact
since Method IV should work well when (4) and (5) hold, the case that
it does not do so will might lead us to question the feasibility of these
two assumptions.

Therefore from now on (with the exception of 5.2) we look at the
random vector

(6) \[ W = (W_1, W_2, \ldots, W_{n-1}) \]

where

(i) \( W \) is such that \( E(W_i) = b \).

(ii) \( W \) may or may not be assumed to have a multivariate normal
distribution.

(iii) \( W \) has a stationary covariance structure, that is,

(7) \[ \text{Cov}(W_i, W_j) = m_{|i-j|} \sigma^2 \]

where \( m_0 = 1, m_{\alpha} \) decreases as \( \alpha \) increases and \( m_{\alpha} \geq 0 \) for all \( \alpha \).

A special case of this structure is \( \Sigma = \Sigma(\sigma^2, m, k) \), where we use
the notation of Section II applied to this \((n-1) \times (n-1)\) covariance
matrix.

Another special case is that of \( r \)-th order dependence, that is,
\( m_1 = 0, i > r \). In particular, \( \Sigma = \Sigma(\sigma^2, m, 0) \) is the case of first
order dependence with \( m_1 = m \).

Another special case is \( m_{\alpha} = m^\alpha \) \((0 < m < 1)\) in which case we
note the resulting covariance matrix by \( \Sigma(\sigma^2, m) \).
First let us dispose of the two special cases.

5.2 Two Special Cases.

Suppose it is reasonable to assume that $A$ is completely determined by one unknown parameter $\rho$. Then it may be possible to estimate $\rho$ directly from the observations. For example suppose $A$ is Toeplitz decreasing and the $b_i$ have one of the following two forms,

(8) \[ b_i = 1 - \rho^i \quad 0 < \rho < 1/n \]

(9) \[ b_i = \rho^i \quad 0 < \rho < 1. \]

Then defining

(10) \[ T_n = \sum_{i<j} \frac{F(Y_i, Y_j)}{\frac{n(n-1)}{2}} \]

we have

(11) \[ E(T_n) = E \left[ \frac{\sum_{i<j} F(Y_i, Y_j)}{\frac{n(n-1)}{2}} \right] = E \left[ \frac{\sum_{i<j} F(X_i, X_j)}{\frac{n(n-1)}{2}} \right] = E \left[ \frac{\sum_{i=1}^{n} (n-1)b_i}{\frac{n(n-1)}{2}} \right]. \]

Using (8) this becomes
\begin{equation}
E(T_n) = \frac{\sum_{i=1}^{n-1} (n-i)(1-\rho)}{\frac{n(n-1)}{2}}
\end{equation}

\begin{equation}
= 1 - \rho \left[ \frac{n \sum_{i=1}^{n-1} i - \sum_{i=1}^{n-1} i^2}{\frac{n(n-1)}{2}} \right]
\end{equation}

\begin{equation}
= 1 - \rho [n - \frac{2n-1}{2}] = 1 - \frac{\rho}{2} (n+1).
\end{equation}

Thus to estimate \( \rho \) we can let

\begin{equation}
\hat{\rho}_n = \frac{3(1-T_n)}{n+1}
\end{equation}

and use

\begin{equation}
\hat{\rho}_n' = \min(\rho_n, 1/n).
\end{equation}

Now

\begin{equation}
\text{Var}(\hat{\rho}_n) = \frac{9}{(n+1)^2} \text{Var}(T_n)
\end{equation}

\begin{equation}
= \frac{9}{(n+1)^2} \frac{4}{n^2(n-1)^2} \epsilon^T \Sigma \epsilon
\end{equation}

where \( \epsilon^T \) is a 1 x \( \frac{n(n-1)}{2} \) dimensional vector of 1's, that is, \( \epsilon^T \epsilon \) is the sum of the entries in \( \Sigma \). Thus

\begin{equation}
\text{var} \hat{\rho}_n \leq \frac{9}{(n+1)^2} \frac{4}{n^2(n-1)^2} \left( \frac{n^2(n-1)^2}{4} \right) \left( \max_{i,j} \sigma_{ij} \right)
\end{equation}

and we see that \( \hat{\rho}_n \) is consistent in all the cases of interest.
Now let

\[ U_n = \frac{\sum_{i<j} f(Y_i, Y_j)}{n}. \]

Then using (9) as in (11)

\[ E(U_n) = \frac{\sum_{i=1}^{n-1} (n-1) \rho^i}{n}. \]

For large \( n \) we have

\[ \sum_{i=1}^{n-1} i \rho^i \approx \frac{n}{(1-\rho)^2}. \]

Thus

\[ E(U_n) \approx \frac{\rho}{1-\rho} - \frac{1}{n} \frac{\rho-n+1}{(1-\rho)^2}. \]

It is easy to show that \( E(U_n) \) is exactly \( \frac{\rho}{1-\rho} - \frac{1}{n} \frac{\rho-n+1}{(1-\rho)^2} \). If we use \( U_n \) as an estimate of \( \frac{\rho}{1-\rho} \) then we can use as our estimate of \( \rho \)

\[ \hat{\rho}_n = \frac{U_n}{U_n + 1} \]

\( \hat{\rho}_n \) will estimate \( \rho \) consistently if \( U_n \) estimates \( \frac{\rho}{1-\rho} \) consistently.
(22) \[ \text{var}(U_n) = \frac{1}{n^2} e^T \Sigma e \quad (e^T \text{ as before}) . \]

Thus for example if \( \Sigma \) is r-th order dependent, \( \hat{\rho}_n \) will be consistent.

Using a first order approximation we have

(23) \[ \text{var}(\hat{\rho}_n) \approx (1-\rho)^2 \text{var}(U_n) = \frac{(1-\rho)^2}{n^2} e^T \Sigma e . \]

5.3 Time Series Approaches.

Given a random vector \( W \) satisfying the assumptions at the end of 5.1 we can look at \( W \) as a set of \( n-1 \) observations and try to find "best" estimates of \( b \) and \( \Sigma \) for specialized \( \Sigma \) such as \( \Sigma(\sigma^2, m, k) \) under normal and non-normal assumptions. We do this in 5.4.

Alternatively, it may be reasonable in certain situations to think of \( W \) as \( n-1 \) observations from a stationary time series. Using this approach we need not make any further assumptions on \( \Sigma \). In addition using the well-known theory of applied time series we readily obtain simple estimates of our unknown parameters. These are the following:

(24) \[ \hat{b} = \frac{1}{n-1} \sum_{i=1}^{n-1} W_i = \bar{W} \]

(25) \[ \hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n-1} W_i^2 \frac{(n-1)\hat{b}^2}{b} \]

(26) \[ \hat{m}_1 = \frac{1}{(n-1)\hat{\sigma}^2} \sum_{j=1}^{n-1-i} W_j W_{j+i} \frac{(n-1)\hat{b}^2}{b} \]

81
Clearly for larger $i$, the numerator of (26) will not be a very good estimator of $m_1 \sigma^2$ and hence $\hat{m}_1$ will not be too robust. Thus, beyond a certain $i^*$, it might be more reasonable to assume $m_i = 0$ and therefore that $\Sigma$ is $i^*$-dependent.

In a time series framework, under an assumption such as $\sum_{i=1}^{\infty} |m_i| < \infty$, these estimators will have certain good asymptotic properties such as unbiasedness, consistency, and normality. Proof of and confidence in these properties obviously depends on this framework and thus the choice of whether or not to use these estimators must really depend on how well they seem to work.

For further details one should refer to most any text that deals with applied time series.

5.4 Generalized Least Squares and Maximum Likelihood Techniques.

Suppose again $W^T = (W_1, W_2, \ldots, W_{n-1})$ has mean $b e^T$ where $e^T$ is a $1 \times (n-1)$ dimensional vector of ones, and covariance $\Sigma$.

We shall be interested in the following cases

(i) $\Sigma = \Sigma(\sigma^2, m, 1)$

(ii) $\Sigma = \Sigma(\sigma^2, m, 0)$ and extensions to $r$-th order dependence for $r \ll n-1$

(iii) $\Sigma = \Sigma(\sigma^2, m, k)$

First suppose $\Sigma$ is known. Then the best linear unbiased estimator (in terms of minimum variance) of $b$ is given by

$$b = e^T \Sigma^{-1} W.$$
This is easily seen in the following way.

\begin{equation}
\text{var}(\hat{b}) = \frac{1}{e^{T^{-1}e}}. 
\end{equation}

For any other unbiased estimator, \(a^TW\) we have

\begin{equation}
a^Te = 1
\end{equation}

\begin{equation}
\text{var}(a^TW) = a^T\Sigma a.
\end{equation}

Then

\[
\frac{\text{var}(a^TW)}{\text{var}(\hat{b})} = (a^T\Sigma a)(e^T\Sigma^{-1}e) \geq (a^Te)^2 = 1.
\]

We remark that in the case that \(W\) is normal, \(\hat{b}\) given by (27) is clearly the maximum likelihood estimate.

Let us look at \(\hat{b}\) for the cases of \(\Sigma\) above.

Letting \(\ell^T = e^T\Sigma^{-1}\) (27) becomes

\[
\hat{b} = \frac{\ell^TW}{\ell^Te} = \frac{\sum_{i=1}^{n-1} \ell_i W_i}{\sum_{i=1}^{n-1} \ell_i}
\]

where

\[\ell^T = (\ell_1, \ell_2, \ldots, \ell_{n-1})\]

such that

\[\ell_i = \sum_{j=1}^{n-1} \sigma^{-1}_{i,j}\] (that is, the sum of the \(i\)-th row or column of \(\Sigma^{-1}\))

83
and
\[ e^T e = \sum_{j=1}^{n-1} \sum_{i=1}^{n-1} \sigma_{ij}^{-1} \] (that is, the sum of all the entries in \( \Sigma^{-1} \)).

From (31) we see that we really do not need \( \Sigma^{-1} \) but just \( e \). In addition (28) becomes

(32) \[ \text{var}(\hat{b}) = \frac{1}{\hat{e}^T \hat{e}}. \]

We note the following:

I. \( \ell_i \geq 0 \) for all \( i \)

II. For our stationary \( \Sigma \) given by (7), \( 0 \leq \ell_i \leq \frac{1}{\sigma^2} \),

III. \( \ell_i = \ell_0 \) for all \( i \) implies \( \hat{b} = \bar{W} = \frac{\sum_{i=1}^{n-1} \hat{W}_i}{n-1} \).

In addition from (32) \( \text{var}(\hat{b}) = \frac{1}{\ell_0^2} (e^T \Sigma e)^{-1} \).

IV. \( \Sigma \) has constant row (= column) sums if and only if \( \Sigma^{-1} \) has constant row (= column) sums. In fact if \( d_i \) are the row (= column) sums of \( \Sigma \), then \( d_i = d_0 \) for all \( i \) implies \( \ell_i = \frac{1}{d_0} \) for all \( i \).

V. \( d_i = d_0 \) for all \( i \) implies \( b = \bar{W} \)

VI. If \( \Sigma = \Sigma(\sigma^2, m, 1) \), \( \hat{b} = \bar{W} \)

VII. If \( \Sigma = \Sigma(\sigma^2, m, k) \) and if we change \( \sigma_{1,n-1} \) from \( km \) to \( m \) then \( \hat{b} = \bar{W} \).
Thus it is quite often the case that \( \hat{b} \) will be independent of \( \Sigma \).
In any case we see that for \( \Sigma \) such that \( \Sigma = \sigma^2 V \), (\( V \) positive definite) and in particular for \( \Sigma \) given by (7) \( \hat{b} \) given by (27) is always independent of \( \sigma^2 \). Thus \( \hat{b} \) is at most a function of \( m_1, m_2, \ldots, m_{n-1} \).

As an example let us solve for \( \hat{b} \) in the case that \( m_1 = m \), \( m_i = 0, i > 1 \). (From earlier considerations it is necessary that \( 0 < m < \frac{1}{2} \)). In view of the above we set \( \sigma^2 = 1 \) and obtain the following system of equations for \( \ell \):

\[
(33) \quad E^T \Sigma(1,m,0) = e^T
\]

or equivalently

\[
\begin{align*}
\ell_1 + m\ell_2 &= 1 \\
m\ell_1 + \ell_2 + m\ell_3 &= 1 \\
m\ell_2 + \ell_3 + m\ell_4 &= 1 \\
\ell_4 &= 1 \\
m\ell_{n-3} + \ell_{n-2} + m\ell_{n-1} &= 1 \\
m\ell_{n-2} + \ell_{n-1} &= 1
\end{align*}
\]

(34)

From (34) it is clear that \( \ell_1 = \ell_{n-1} \) and thus we need only find \( \ell_1, \ell_2, \ldots, \ell_{\left[ \frac{n}{2} \right]} \). (34) also shows that we can solve recursively for \( \ell_i \) given \( \ell_1 \).

Denoting the determinant of the \( r \times r \) matrix \( \Sigma(1,m,0) \) by \( D_r \) and defining \( D_0 = 1 \) one can easily show that
\[ \sigma_{1r} = \frac{(-1)^{r-1} m^{-1} D_{n-r-1}}{D_{n-1}} \]

and thus

\[ \ell_1 = \frac{1}{D_{n-1}} \sum_{j=0}^{n-2} (-1)^j m^j D_{n-j-2} . \]

Thus we only need \( D_r \). Once again it is easy to see that

\[ D_{r+2} = D_{r+1} - m^2 D_r . \]

Thus we need only solve the second-order difference equation.

\[ X_n - X_{n-1} + m^2 X_{n-2} = 0 \]

with

\[ X_1 = 1 \]
\[ X_2 = 1 - m^2 . \]

The general solution is given by

\[ X_k = c_1 (a_1)^k + c_2 (a_2)^k \]

where \( a_1, a_2 \) are the roots of \( a^2 - a + m = 0 \), i.e.,

\[ a_1 = \frac{1 + \sqrt{1 - 4m}}{2} \quad \text{and} \quad \frac{1}{2} < a_1 < 1 \]
\[ a_2 = \frac{1 - \sqrt{1 - 4m}}{2} \quad \text{and} \quad 0 < a_2 < \frac{1}{2} . \]

From (39), (40) and (41) we obtain
\[ (42) \quad c_1 = \frac{1-m^2-a_2}{a_1(a_1-a_2)} \]

\[ (43) \quad c_2 = \frac{1-m^2-a_1}{a_2(a_2-a_1)} \]

and thus

\[ (43) \quad D_r = c_1 \left( \frac{1+\sqrt{1-h_0^2}}{2} \right)^r + c_2 \left( \frac{1-\sqrt{1-h_0^2}}{2} \right)^r. \]

One can easily show, in addition, that

\[ (44) \quad D_r > 0 \]

\[ D_r > D_{r+1} \]

\[ \lim_{r \to \infty} D_r = 0 \]

so that \( D_r \) decreases to zero.

We remark that the above solution clearly has applications in other settings.

Remark: This approach fails for other than a first-order dependent \( \Sigma \) and one must invert \( \Sigma \) to find \( \hat{b} \) or just use \( \overline{W} \).

In the case that \( \Sigma \) is \( r \)-th-order dependent both \( \hat{b} \) and \( \overline{W} \) are consistent. This is clear since
\begin{equation}
\text{var}(\hat{b}) \leq \text{var}(\tilde{w}) = \frac{1}{(n-1)^2} e^T \Sigma e
\end{equation}
\begin{equation}
= \frac{\sigma^2}{(n-1)^2} ((n-1) + 2(n-2)m_1 + \cdots + 2[n-(r+1)]m_r)
\end{equation}
\begin{equation}
\leq \frac{2\sigma^2 (r+1)}{(n-1)} \to 0 \text{ as } n \to \infty.
\end{equation}

In fact, by the Central Limit Theorem for \( r \)-dependent random variables since the \( W_i \) are usually assumed bounded or normal, we can show that \( \tilde{w} \) is asymptotically normal with variance \( \sigma^2 (1 + 2 \sum_{i=1}^{r} m_i) \). (See, for example, Fraser [4] Chapter 6, Section 4, p. 219.)

In the case that \( \Sigma = \sigma^2 V \), \( V \) known, the least squares estimate of \( \sigma^2 \) is given by
\begin{equation}
\hat{\sigma}^2 = \frac{(W - \hat{w}e)^T V^{-1} (W - \hat{w}e)}{n-1}.
\end{equation}

To see this we note that \( Z = V^{-1/2} \tilde{w} \) is a vector of uncorrelated random variables with mean vector \( bV^{-1/2} e \) and common variance \( \sigma^2 \). Thus the least squares estimate of \( \sigma^2 \) is
\begin{equation}
\hat{\sigma}^2 = \frac{(Z - \hat{\Sigma}_{-1/2} e)^T (Z - \hat{\Sigma}_{-1/2} e)}{n-1}
\end{equation}
which is equivalent to (46).

Furthermore if \( \hat{b} \) is consistent then \( \hat{\sigma}^2 \) is, and in fact from (47) we see that \( (n-1) \frac{\hat{\sigma}^2}{\sigma^2} \) is tending to a \( \chi^2 \) distribution with \( n-2 \) degrees of freedom.
Suppose now that $\Sigma$ is given by (7) and $\sigma^2$ is known but the $m_i$ are not. If $b$ is a function of the $m_i$ it seems reasonable to look for consistent estimators $\hat{m}_i$ of the $m_i$ (of course, we can not allow the $\hat{m}_i$'s to be functions of $b$). Replacing $m_i$ by $\hat{m}_i$, $b$ will now be asymptotically unbiased and still consistent. (Asymptotic unbiasedness is clear since the $W_i$ are ordinarily assumed bounded or normal.)

Suppose $\Sigma = \Sigma(\sigma^2, m, 1)$ and that $W$ is normally distributed. Then as mentioned earlier $W_1, \ldots, W_{n-1}$ can be thought of as having come from $n$ independent, identically distributed normal random variables, $U_1, U_2, \ldots, U_n$ having mean 0 and variance 1 in the following way

$$W_i = \sigma \sqrt{1-m} U_i + \sigma \sqrt{m} U_n + b \quad i = 1, \ldots, n-1.$$  

Consider

$$Q_{n-1} = \frac{\sum_{\alpha_i, \alpha_j, \alpha_k} (W_{\alpha_i} - W_{\alpha_j})(W_{\alpha_j} - W_{\alpha_k})}{(n-1)(n-2)(n-3)}$$  

so that

$$Q_{n-1} = \frac{\sigma^2(1-m) \sum_{\alpha_i, \alpha_j, \alpha_k} (U_{\alpha_i} - U_{\alpha_j})(U_{\alpha_j} - U_{\alpha_k})}{(n-1)(n-2)(n-3)}$$  

where the summation is taken over all permutations of size 3 from the integers 1, 2, ..., $n-1$. From (50) we see that $Q_{n-1}$ is a $U$ statistic with kernel of order 3. From a well known theorem of Hoeffding (see,
for example, Fraser [4], Chapter 6, Section 5, p. 225) we have $Q_{n-1}$ asymptotically normal with asymptotic variance $2\sigma^2(1-m)^2$, that is

$$\sqrt{n-1} \frac{(Q_{n-1}-(1-m)\sigma^2)}{\sqrt{2} \sigma^2(1-m)} \xrightarrow{d} Z \sim \mathcal{N}(0,1).$$

Thus

$$\hat{m} = 1 - \frac{Q_{n-1}}{\sigma^2}$$

will be an unbiased, consistent, asymptotically normal estimate of $m$, with asymptotic variance $2(1-m)^2$.

If $W$ is not normally distributed a transformation like (48) is not possible. If the $W_i$ are bounded as is usually the case then all moments exist and we can show (via an arduous variance calculation) that $Q_{n-1}$ and hence $\hat{m}$ are consistent.

Suppose now that $\Sigma = \Sigma(\sigma^2,m,0)$ and that the $W_i$ are either normal or bounded so that higher moments exist.

Let us look at the following two estimates

$$R_{n-1} = \frac{\sum_{i=1}^{n-2} (W_{i+1} - W_i)^2}{n-2}$$

and

$$S_{n-1} = \frac{\left\lceil \frac{n-1}{2} \right\rceil}{\left\lfloor \frac{n-1}{2} \right\rfloor} \sum_{i=1}^{\left\lceil \frac{n-1}{2} \right\rceil} (W_{2i-1} - W_{2i-1})^2.$$

Now $E(R_{n-1}) = E(S_{n-1}) = 2\sigma^2(1-m)$. 

90
Let us compare their variances.

Since $S_{n-1}$ is a sum of independent random variables, it is clearly consistent and in fact has variance,

\[ \text{var}(S_{n-1}) = \frac{1}{\binom{n-1}{2}} \text{var}((W_2-W_1)^2). \]  

(55)

For $R_{n-1}$

\[ \text{var}(R_{n-1}) = \frac{1}{n-2} \text{var}((W_2-W_1)^2) + \frac{2(n-3)}{(n-2)^2} \text{cov}((W_2-W_1)^2,(W_3-W_2)^2) \]

(56)

so that $R_{n-1}$ is also consistent.

Both $S_{n-1}$ and $R_{n-1}$ are asymptotically normal, $S_{n-1}$ via the Central Limit Theorem and $R_{n-1}$ by the $r$-dependent Central Limit Theorem mentioned earlier with $r=1$.

For large $N$ we see that

\[ \text{var}(S_{n-1}) - \text{var}(R_{n-1}) \]

(57)

\[ \approx \frac{1}{n-2} [\text{var}((W_2-W_1)^2) - 2 \text{cov}((W_2-W_1)^2,(W_3-W_2)^2)]. \]

(58)

In the case that the $W_i$ are normal a calculation shows that (58) is positive for $0 < m < 1/2$. The calculation is most easily done by letting $T_1 = W_2-W_1$, $T_2 = W_3-W_2$ and noting that $E(T_1^2 T_2^2) = E_{t_1}^h - \frac{1}{2} E(T_1 + T_2)^2 E(T_1 - T_2)^2$. Thus $R_{n-1}$ is preferred. Therefore as an estimate of $m$ we would use
\[
\hat{m} = 1 - \frac{R_{n-1}}{2\sigma^2}.
\]

The above remarks suggest that when \( \Sigma \) is \( r^{th} \)-order dependent with \( \sigma^2 \) known and \( r \ll n-1 \) we use as an estimate of \( m_i \), \( i=1,\ldots,r \).

\[
\hat{m}_i = 1 - \frac{R_{n-1}}{2\sigma^2}^{(i)}
\]

where

\[
R_{n-1}^{(i)} = \frac{\sum_{j=1}^{n-1-i} (W_{j+1} - W_j)^2}{n-1-i}
\]

As before \( \hat{m}_i \) will be unbiased, consistent and asymptotically normal.

Lastly if \( \sigma^2 \) is unknown we need a good estimate of \( \sigma^2 \) that is independent of the \( m_i \) as well as \( \hat{b} \) if \( \hat{b} \) depends on \( m_i \). Consider the usual estimate

\[
\hat{\sigma}^2 = \frac{\sum_{i=1}^{n-1} W_i^2 - (n-1)\bar{W}^2}{n-2}
\]

This estimate is unbiased and is easily shown to be consistent in the case that \( \Sigma \) is \( r^{th} \)-order dependent. Thus replacing \( \sigma^2 \) by \( \hat{\sigma}^2 \) our estimates \( \hat{m}_i \) and \( \hat{b} \) will be consistent and now are asymptotically unbiased if the \( W_i \) are assumed bounded or normal.

In conclusion our estimation procedure assuming all parameters are unknown will be the following:
(i) Estimate $\sigma^2$ via $\hat{\sigma}^2$

(ii) Estimate the $m_1$ using $\hat{\sigma}^2$

(iii) Estimate $b$ via $\hat{b}$ (using the $\hat{m}_1$ if necessary).
REFERENCES


REFERENCES (Cont.)


SERIATION OF MULTIVARIATE OBSERVATIONS THROUGH SIMILARITIES

For certain types of problems in multivariate data reduction, seriation and scaling may be reasonable approaches.

Given a collection of $n$ objects, seriation techniques try to order these objects on a one-dimensional scale in the sense of assigning a rank from one to $n$ to each object. Scaling techniques attempt to do more by assigning a numerical value to each object so that not only is order achieved but also some quantitative measure of relative closeness is computed.

Similarity functions are employed to measure the "closeness" between pairs of vectors. Two general approaches are considered encompassing five methods.

Lastly a section is devoted to several estimation problems that arise from considering the similarities between pairs of vectors as random variables having certain underlying mean and covariance structures.
### Security Classification

**INSTRUCTIONS**

1. **ORIGINATING ACTIVITY:** Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization (corporate author) issuing the report.

2a. **REPORT SECURITY CLASSIFICATION:** Enter the overall security classification of the report. Indicate whether "Restricted Date" is included. Marking is to be in accordance with appropriate security regulations.

2b. **GROUP:** Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. **REPORT TITLE:** Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. **DESCRIPTIVE NOTES:** If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. **AUTHOR(S):** Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. **REPORT DATE:** Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. **TOTAL NUMBER OF PAGES:** Enter total number of pages. The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. **NUMBER OF REFERENCES:** Enter the total number of references cited in the report.

8a. **CONTRACT OR GRANT NUMBER:** Enter the applicable number of the contracting or grant under which the report was written.

8b, 8c, & 8d. **PROJECT NUMBER:** Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. **ORIGINATOR’S REPORT NUMBER(S):** Enter the official number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. **OTHER REPORT NUMBER(S):** If the report has been assigned any other report numbers (either by the originator or by the sponsor), also enter this number(s).

10. **AVAILABILITY/LIMITATION NOTICES:** Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

   (1) "Qualified requesters may obtain copies of this report from DDC."

   (2) "Foreign announcement and dissemination of this report by DDC is not authorized."

   (3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through ."

   (4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through ."

   (5) "All distribution of this report is controlled. Qualified DDC users shall request through ."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. **SUPPLEMENTARY NOTES:** Use for additional explanatory notes.

12. **SPONSORING MILITARY ACTIVITY:** Enter the name of the departmental project office or laboratory sponsoring (paying for) the research and development. Include address.

13. **ABSTRACT:** Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as [(F), (C), (G), or (U)]. There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. **KEY WORDS:** Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.