SOME RESULTS ON THE PROBABILITY DISTRIBUTION OF THE
LATENT ROOTS OF A SYMMETRIC MATRIX OF CONTINUOUSLY
DISTRIBUTED ELEMENTS, AND SOME APPLICATIONS
TO THE THEORY OF RESPONSE SURFACE ESTIMATION

Prepared under Contract No. DA-36-034-GRD-1517 (RD)
(Experimental Designs for Industrial Research)

by
H. R. van der Vaart

Institute of Statistics
Mimeo. Series No. 189
January, 1958
CONTENTS

1 Introduction .......................... page 1
   1.1 Preliminaries, notations, statement of the problem 1
   1.2 Definition of certain key matrices 5

2 Distribution-free results on the latent roots distribution 7
   2.1 Some theorems connected with the bias of \( \Lambda \) as an estimator of \( \Lambda \) 9
   2.2 Some theorems on the variances and covariances of the elements of \( L \) and \( \hat{L} \) 17
   2.3 A condition for symmetry in the distribution of the latent roots 21

3 Application of the preceding theory to some aspects of the estimation of quadratic response surfaces 26
   3.1 A few general considerations 26
   3.2 Bias in the estimation of type of quadratic surface 31
   3.3 The variance of the coefficients of the canonical form 36
   3.4 Proposal of a new approach to some problems in the evaluation of response surfaces 38

Literature cited 39
SOME RESULTS ON THE PROBABILITY DISTRIBUTION OF THE
LATENT ROOTS OF A SYMMETRIC MATRIX OF CONTINUOUSLY
DISTRIBUTED ELEMENTS, AND SOME APPLICATIONS
TO THE THEORY OF RESPONSE SURFACE ESTIMATION

by
H. R. van der Vaart
Institute of Statistics, University of North Carolina
Raleigh, North Carolina, U. S. A.
and
Leiden University, Netherlands

1. Introduction

1.1 Preliminaries, notations, statement of the problem.

Consider a $k \times k$ real matrix $M$, each of its elements being allowed to take
values in a certain space. If a probability measure is defined over the product
$\pi_M$ of these $k^2$ spaces, $M$ will be called a random matrix. This probability measure
will be called singular (relative to the $k^2$-dimensional Lebesgue measure to be
defined on $\pi_M$) if a subset of $\pi_M$ exists which has Lebesgue measure zero and
probability measure one (because of probabilities being non-negative the more
sophisticated definitions of Saks, 1937, p. 30, or Doob, 1953, p. 611 can be
simplified in this context). This subset (defined up to a set of Lebesgue measure
zero) is called the singular set of the probability measure in question (relative
to the afore-said Lebesgue measure) or alternatively its set of increase (cf. Doob,
l.c.). We shall denote this singular set of the probability measure over $\pi_M$
by the term $N$-space and the probability distribution (= probability measure) over
$N$-space by the term (probability) distribution of $M$. We may need a term to denote
$\pi_M$ and use enlarged $N$-space to this end.

Examples of singular probability measures defined on $\pi_M$: the elements of $M$ may
be restricted to integer values, $M$ may be prescribed to be orthogonal ($N$-space
\( \frac{1}{2} k(k-1) \)-dimensional) or \( M \) may be symmetric \((M\text{-space } \frac{1}{2} k(k+1) \text{-dimensional})\). In the second and third example \( M \)-space is a hypersurface in \( \pi_M \), and may be defined for instance by equations regarding the elements of \( M \), or by parametrization of these elements.

We shall briefly discuss the case of a symmetric random matrix \( M \). Here \( M \)-space may be defined as a subset of \( \pi_M \) by the parametrization

\[
\begin{align*}
    m_{ij} &= t_{ij} \ (j \geq i; \ i = 1, \ldots, k) \\
    m_{ji} &= t_{ji} \ (j < i; \ i = 1, \ldots, k).
\end{align*}
\]

(1.1.1)

However, it is slightly more convenient to identify the \( t_{ij} \) and the \( m_{ij} \) (\( j \geq i \)) and so to use \( (m_{11}, \ldots, m_{1k}, m_{21}, \ldots, m_{2k}, \ldots, m_{k-1,k-1}, m_{k-1,k}^*, m_{kk}^*) \)-space for \( M \)-space and to use the projection of the probability distribution over the set defined by (1.1.1) on this \( m_{ij} \) (\( 1 \leq i \leq j \leq k \))-space for the distribution of \( M \). We shall call this distribution of a symmetric random matrix \( M \) continuous if it is absolutely continuous with respect to \( \frac{1}{2} k(k+1) \)-dimensional Lebesgue measure on \( m_{ij} \) (\( 1 \leq i \leq j \leq k \))-space.

Let again \( 1^M \) and \( 2^M \) be two \( k \times k \) real (otherwise arbitrary) matrices. As is well known, they are said to be equal,

\[
1^M = 2^M,
\]

if each of the \( k^2 \) elements of \( 1^M \) equals the corresponding element of \( 2^M \). If now \( 1^M \) and \( 2^M \) are in addition random, they will be called equivalent (relative to the probability measure considered) if \( 1^M = 2^M \) with probability one (cf. Kolmogorov, 1933 (1950), p. 33 for \( M \) a scalar; if one needs to distinguish between this type of equivalence and the one described by MacDuffee, 1946, p. 40 seqq., one may add the qualification stochastical to the above-defined type). This entails that the probabilities of any subsets of the enlarged \( 1^M \)-space and of the enlarged
\( M \)-space are equal as soon as these subsets can be transformed into each other by
the equation \( 1^M = 2^M \), hence the distributions of \( 1^M \) and of \( 2^M \) are the same: if \( 1^M \)
and \( 2^M \) are equivalent random matrices, the expectation of \( f(1^M) \) equals the expecta-
tion of \( f(2^M) \) (provided these expectations exist).

Before stating the problem to be investigated, let us add some more notation
conventions. We shall write the expectation of a matrix for the matrix of the
expectations

\[
\mathcal{E}(M) = \mathcal{E}(\begin{bmatrix} m_{ij} \\ \| m_{ij} \| \end{bmatrix} _{i=1}^{1} \cdots _{k=1}^{k}) = \mathcal{E}(m_{ij}) _{i=1}^{i=1} \cdots _{k=1}^{k} = \mu_{ij} _{i=1}^{i=1} \cdots _{k=1}^{k};
\]

(1.1.2)

if the Greek capital \( \mu \) were different from the Latin capital \( m \), we would have
replaced the fourth member of (1.1.2) by one single Greek capital in accordance
with the next convention: a Greek small letter will denote a parameter, and the
\corresponding Latin small letter a random variable closely related to it (Greek
letter denotes expectation of Latin letter, or Latin letter denotes an estimator
of the Greek one for instance): a Greek capital will denote a square matrix of
parameters, a Latin capital a square matrix of random variables. "Estimating \( \Phi \"
will stand for "jointly estimating the \( \varphi_{ij} \)" (\( \Phi = \begin{bmatrix} \varphi_{ij} \\ \| \varphi_{ij} \| \end{bmatrix} _{i=1}^{i=1} \cdots _{k=1}^{k} \)). A symbol
like \( m_{ij} \) will denote the \( 1 \times k \) matrix or row vector consisting of the elements in
the \( i \) th row of \( M \), \( m_{j} \) will denote the \( k \times 1 \) matrix or column vector consisting of the
elements in the \( j \) th column of \( M \). Sometimes we shall denote vectors by under-
lining (printing in bold type) small letters without any subscript.

Now we shall state our problem. Let \( \Phi = \begin{bmatrix} \varphi_{ij} \\ \| \varphi_{ij} \| \end{bmatrix} _{i=1}^{i=1} \cdots _{k=1}^{k} \) be a real sym-
m\etic \( k \times k \) matrix: \( \varphi_{ij} = \varphi_{ji} \). Let \( F = \begin{bmatrix} f_{ij} \\ \| f_{ij} \| \end{bmatrix} _{i=1}^{i=1} \cdots _{k=1}^{k} \) be a real symmetric
\( k \times k \) random matrix which is continuously distributed and satisfies

\[
\mathcal{E}(F) = \Phi
\]

(1.1.3)
As both $\Phi$ and $F$ are real and symmetric, their latent roots, $k$ in number, will be real. Denote the latent roots of $\Phi$ by $\lambda_h$ and those of $F$ by $\xi_h$ ($h = 1, \ldots, k$), and assign the subscripts in such a way that

\[
\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_k \tag{1.1.4}
\]

\[
\xi_1 \leq \xi_2 \leq \cdots \leq \xi_k \tag{1.1.5}
\]

The object of this paper is to investigate the joint distribution of the $\xi_h$, where the $\xi_h$ will be allowed to be negative - contrary to a number of previous investigations on the distribution of latent roots. Though two or more roots $\lambda_h$ may be equal, the probability that two or more roots $\xi_h$ be strictly equal is always zero because of $F$ being continuously distributed. The following notations will be used:

\[
\Lambda = \| \delta_{gh} \lambda_h \|_{h=1}^{g=1} \cdots _{h=k}^{g=k}, \tag{1.1.6}
\]

\[
L = \| \delta_{gh} \xi_h \|_{h=1}^{g=1} \cdots _{h=k}^{g=k}, \tag{1.1.7}
\]

where $\delta_{gh}$ is the well known Kronecker symbol. The distribution of $L$ (note that $L$-space will be "$k$-dimensional" in the sense that its points are defined by $k$ coordinates) has an interest of its own as well as in connection with certain specific problems. Chapter 3 will discuss one of these specific problems, viz. the estimation of the type of response surfaces (cf. for instance Box and Youla, 1955), whereas Chapter 2 will give a number of distribution-free results on the distribution of $L$ (i.e. results which are to a large extent independent of the distribution of $F$).

Remark. The above-mentioned condition, that $F$ be continuously distributed will be maintained throughout the paper. Yet it may be weakened on some occasions, which are left to the reader to point out.
One example is the zero probability of two or more roots $\ell_h$ being strictly equal. Certain discrete distributions of $F$ will share this property with the continuous distributions.

1.2 Definition of certain key matrices.

As is well known a real orthogonal matrix

\[
\begin{pmatrix}
Y \\
U
\end{pmatrix}
\]

can always be found such that

\[
\Phi Y = YA \quad \text{and} \quad FU = UL. \quad (1.2,1a)
\]

Here the columns of

\[
\begin{pmatrix}
Y \\
U
\end{pmatrix}
\]

constitute the mutually orthogonal latent vectors of

$\Phi$, while $v_h$ corresponds to $\lambda_h$; \quad $F$, while $u_h$ corresponds to $\ell_h$.

It is well known that if

$v_h$ is a latent vector, so is $-v_h$; \quad $u_h$ is a latent vector, so is $-u_h$;

besides, if two or more latent roots of $\Phi$ or $F$ are equal:

$\lambda_{h_1} = \lambda_{h_2} = \ldots = \lambda_{h_d}$, say, \quad $\ell_{h_1} = \ell_{h_2} = \ldots = \ell_{h_d}$, say,

then there is an infinite set of systems of corresponding mutually orthogonal latent vectors:

\[
\begin{pmatrix}
v_{h_1} \\
v_{h_2} \\
\vdots \\
v_{h_d}
\end{pmatrix}, \quad \begin{pmatrix}
u_{h_1} \\
u_{h_2} \\
\vdots \\
u_{h_d}
\end{pmatrix}
\]

Neither of these possibilities will have consequences in the course of our investigations, \quad The probability of two or more latent roots being equal is zero

(cf. the second paragraph of
because the elements of $Y$ will be seen to drop out of all interesting developments at an early stage.

section 1.1). Hence the columns $u_h$ of $U$ are uniquely defined with probability one, except for their sign; as for this point see the last paragraph but one in this section.

For easy reference the following formulas, equivalent to (1.2,1a), are listed:

$$\Phi = Y \Lambda Y'$$
$$Y' \Phi = \Lambda Y'$$
$$Y' \Phi Y = \Lambda$$

$$F = U L U' \quad (1.2,1b)$$
$$U' F = L U' \quad (1.2,1c)$$
$$U' F U = L \quad (1.2,1d)$$

Another real matrix to play an important role in this investigation is $V$, defined by one of the equivalent relations:

$$V = Y' U, \quad Y V = U, \quad Y = U V'$$
$$V^2 = U V', \quad V V' = U', \quad Y' = V U' \quad (1.2,2)$$

Note that as $U$ is a random matrix so is $V$, and that if $Y$ and $U$ are restricted so as to be proper orthogonal, so is $V$.

As for the signs of the column vectors of $U$, it is evident that the equations (1.2,2), once $Y$ has been selected, establish a one-one correspondence between the signs of the columns of $U$ and $V$. Now in deriving the distribution-free results of this paper these signs will be seen to be of no consequence. However, they will become important when the distribution of $L$ is studied under more precise assumptions, e.g., that $F$ be multinormally distributed. Then the Haar measure on the orthogonal group of $U$ and $V$ matrices will come into play (cf. Anderson, 1951, and James, 1954), and the sets of admitted matrices $U$ and $V$ will then have to be chosen in harmony with the normalization of this measure.
Both \( Y \) and \( U \), hence \( V \), will be restricted throughout to be proper orthogonal, i.e., their determinants are equal to \( +1 \).

2. Distribution-free results on the latent roots distribution

Under the assumptions of section 1.1, viz., if the \( k \times k \) real symmetric random matrix \( F \) is distributed continuously and if

\[
\mathcal{E}(F) = \Phi, \tag{2.1}
\]

one finds from equations (1.2, l4), (2.1), (1.1,2), and (1.2,2) that

\[
\Lambda = Y^\prime \Phi Y = Y^\prime \mathcal{E}(F)\cdot Y = \mathcal{E}(Y^\prime FY) = \mathcal{E}(VU^\prime FUV) = \mathcal{E}(VLV^\prime) \tag{2.2}
\]

Thus, though the off-diagonal elements of the real symmetric random matrix

\[
Y^\prime FY = VLV^\prime = \hat{\Lambda}, \text{ say} \tag{2.3}
\]

are different from zero with probability one, and though \( \hat{\Lambda}_{11} \) may exceed \( \hat{\Lambda}_{22} \) with positive probability (see Remark at the end of section 2), yet this matrix \( \hat{\Lambda} \) would be an expectation-unbiased estimator of \( \Lambda \) (as for this concept, see \textit{van der Vaart}, 1957a, section 2). Unfortunately, in the usual cases one will not know \( \Phi \) or \( Y \), but only be able to observe a value, or rather a realization, of the random matrix \( F \). Hence among the three matrices \( Y, V, U \) occurring in equations (1.2,2), (2,2) and (2,3) \( Y \) and \( V \) will be essentially unknown, so that generally speaking \( \hat{\Lambda} \) cannot serve as an estimator of \( \Lambda \). Note that if \( Y \) should happen to be the identity matrix \( I \) (which would entail that \( \Phi \) is the diagonal matrix \( \Lambda \)), then equations (2,2) and (2,3) show that \( F = \hat{\Lambda} \) rather than the commonly used \( U^\prime FU = \Lambda \) ought to serve as an estimator of \( \Lambda \). This seems to be a first corroboration of a conjecture made by R. L. Anderson (and privately communicated to the present author during his stay at the Institute of Statistics, Raleigh) to the effect that if some off-diagonal element of \( F \) happens to be smaller than its standard error, it might be a good idea to estimate the response surface as if this off-diagonal
element were zero. This leads to altering $U$ and to using this altered $U$ in the estimate $U^T F U$ of $\Lambda$ (if all off-diagonal elements of $F$ are small — which is a fairly probable occurrence in the above-mentioned case where $Y = I$, $\Phi = \Lambda$ — then this conjecture correctly leads to $U = I$, hence to $F$ as an estimator of $\Lambda$). It is intended to give this matter further consideration in another paper.

It is clear from equation (2.2) that though the off-diagonal elements of $L$ are identically zero, yet this matrix $L$ may leave much to be desired when used as an estimator of $\Lambda$. As this use is quite common in the evaluation of response surfaces (see for instance Box and Youle, 1955, p. 295, and Box and Hunter, 1957, p. 239, and section 3 of the present paper), it is worth while to investigate the properties of $L$ as an estimator of $\Lambda$. In section 2.1 some bias properties of $L$ will be investigated, section 2.2 will be concerned with variances and covariances of the elements of $L$ and $\hat{\Lambda}$, whereas section 2.3 will investigate a sufficient condition to be imposed on the distribution of $F$ in order that the distribution of $L$ will be symmetrical in a certain sense.

Remark. Note that because the off-diagonal elements of $\Lambda$ and $L$ equal zero by definition, no confusion can arise from using a single subscript in $\lambda_h$, $\ell_h$, the diagonal elements of $\Lambda$ and $L$, respectively. However, the diagonal elements of $\hat{L}$ ought to be denoted by $\hat{\ell}_{hh}$ or $\hat{\ell}_{gg}$ as opposed to $\hat{\ell}_{gh}$ ($g \neq h$) for the off-diagonal elements. Yet we shall frequently write $\hat{\ell}_h$ or $\hat{\ell}_g$ instead of $\hat{\ell}_{hh}$ or $\hat{\ell}_{gg}$: if one subscript is attached to $\hat{\ell}$, this symbol will represent a diagonal element of $\hat{L}$, if two subscripts are attached, the symbol will have its usual meaning.

The proof that the off-diagonal elements of $\hat{L}$ are different from zero with probability one, is similar to the proof of Lemma 2.1.1 in the next section: the probability measure (absolutely continuous with
respect to Lebesgue measure) which is defined on F-space is \( \frac{1}{2} k(k+1) \) -
dimensional, the dimension of the set of points with \( \hat{\nu}_{ij} f_{ij} \hat{\nu}_{ij} = 0 \) \( (g \neq h) \) is less than \( \frac{1}{2} k(k+1) \), hence the probability of this point set is zero.

An example that \( P(\hat{\nu}_1 > \hat{\nu}_2) \) may be positive is given by: be \( k = 2 \),
\( Y = I, \Phi = I = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \) with \( \lambda_1 < \lambda_2 \); \( F = \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix} \) \( \sim \) L. Be the distribution of \( F \) multinormal, then by section 1.1 \( \mathcal{E}(f_{11}) = \lambda_1 < \lambda_2 \)
\( = \mathcal{E}(f_{22}) \), yet \( P(f_{11} > f_{22}) \) is known to be positive.

2.1 Some theorems connected with the bias of \( L \) as an estimator of \( \Lambda \).

The first theorem is on absence of bias rather than on its presence.

Theorem 2.1.1.

If \( F \) is continuously distributed and \( \mathcal{E}(F) = \Phi \), then
\[
\mathcal{E}(\text{tr } \hat{L}) = \mathcal{E}(\text{tr } L) = \text{tr } \Lambda = \text{tr } \Phi = \mathcal{E}(\text{tr } F) \tag{2.1.1}
\]

Remark. As usual, \( \text{tr } \Phi \) stands for \( \sum_{i=1}^{k} \phi_{ii} \), etc.

Proof.

Proof is immediate from equation (2.2) and the following properties of traces:
\[
\text{tr } \mathcal{E}(F) = \mathcal{E}(\text{tr } F) \text{ if } F \text{ is any random matrix (cf. eq. 1.1.2)} \tag{2.1.2}
\]
\[
\text{tr } (AB) = \text{tr } (BA) \text{ for any square matrices } A \text{ and } B; \tag{2.1.3}
\]
for instance, \( \text{tr } \Lambda = (\text{by eq. (2.2)}) = \text{tr } [\mathcal{E}(V L V')] = \mathcal{E}[\text{tr}(V L V')] =
\]
\( = \mathcal{E}[\text{tr}(L V V')] = (\text{since } V \text{ is orthogonal}) = \mathcal{E}(\text{tr } L).

Now, in order to prove the next two theorems on bias we need

Lemma 2.1.2.

If the probability distribution of \( F \) is continuous and \( \mathcal{E}(F) = \Phi \); then
Proof.

As \( V \) is an orthogonal matrix \( \sum_{h=1}^{k} v_{1h}^2 = 1, \sum_{h=1}^{k} v_{kh}^2 = 1 \).

Equation (2.3) yields \( \hat{\ell}_1 = \sum_{h=1}^{k} v_{1h}^2 \ell_h, \hat{\ell}_k = \sum_{h=1}^{k} v_{kh}^2 \ell_h \).

Hence

\[
\hat{\ell}_1 - \ell_1 = \sum_{h=1}^{k} v_{1h}^2 (\ell_h - \ell_1) \tag{2.1.5a}
\]

\[
\ell_k - \hat{\ell}_k = \sum_{h=1}^{k} v_{kh}^2 (\ell_h - \ell_k) \tag{2.1.5b}
\]

As \( v_{gh}^2 \geq 0 \) and \( \ell_h - \ell_g \geq 0 \) if \( h > g \) (cf. inequalities (1.1.5)), it is trivial that \( \hat{\ell}_1 - \ell_1 \) and \( \ell_k - \hat{\ell}_k \) are essentially non-negative. Hence all that needs proof is

\[
P(\hat{\ell}_1 - \ell_1 = 0) = 0 \tag{2.1.6a}
\]

\[
P(\ell_k - \hat{\ell}_k = 0) = 0 \tag{2.1.6b}
\]

We shall write down the proof for equation (2.1.6a). The equation

\[
F = U\Lambda U' = V\Gamma V'
\]

which ensues from equations (1.2.1b) and (1.2.2) parametrizes (for this term see e.g. Husem Berg, 1947, p. 245) the \( \frac{1}{2} k(k+1) \)-dimensional \( F \)-space (\( F \) is symmetric) in terms of the Cartesian product of \( \frac{1}{2} k(k-1) \)-dimensional \( V \)-space (\( V \) is orthogonal) and \( k \)-dimensional \( L \)-space. Now equation (2.1.5a) shows that the condition \( \hat{\ell}_1 - \ell_1 = 0 \) can be satisfied only if \( \ell_1 = \ldots = \ell_g < \ell_{g+1} \) and \( 0 = v_{1g} \ell_{g+1} = \ldots = v_{1k} \) for \( g = 1, \ldots, k \). Hence \( \hat{\ell}_1 - \ell_1 = 0 \) defines a union of \( k \) subsets of the Cartesian product of \( V \)-space and \( L \)-space of dimension lower than \( \frac{1}{2} k(k+1) \). In consequence
the image of \( \hat{\ell}_1 - \ell_1 = 0 \) in \( F \)-space has \( \frac{1}{2} k(k+1) \)-dimensional Lebesgue measure zero, and as the probability distribution over \( F \)-space is assumed to be absolutely continuous with respect to this Lebesgue measure over \( F \)-space, it follows that \( P(\hat{\ell}_1 - \ell_1 = 0) = 0 \) (cf. equation (3.5) in Busemann, l.c.).

**Corollary of Lemma 2.1.1.**

\[
P \left[ \sum_{i=1}^{k} \sum_{j=1}^{k} (u_{ij} u_{j}^{l} - u_{ij} u_{j}^{l}) f_{ij} > 0 \right] = 1 \tag{2.1.7a}
\]

\[
P \left[ \sum_{i=1}^{k} \sum_{j=1}^{k} (u_{ij} u_{j}^{k} - u_{ij} u_{j}^{k}) f_{ij} > 0 \right] = 1 \tag{2.1.7b}
\]

**Proof.**

It is easily seen from equations (2.3) and (1.2,1d) that \( \hat{\ell}_1 - \ell_1 \) is the element in the first row and first column of the first member of the matrix equality (which ensues from equations (1.2,1d) and (1.2,2)):

\[
V L V' - L = Y^* F Y - U^* F U \quad . \tag{2.1.8}
\]

Writing down the element in the first row and first column of the second member of (2.1,8) and applying Lemma 2.1,1 yields the desired result. A similar argument pertains to \( \hat{\ell}_k - \ell_k \).

Now we are ready to prove that the smallest latent root of \( F \) is a negatively expectation-biased estimator of the smallest latent root of \( \Phi \), that the largest latent root of \( F \) is a positively expectation-biased estimator of the largest latent root of \( \Phi \), and that their difference is a positively expectation-biased estimator of the corresponding difference.

**Theorem 2.1.2.**

If \( F \) is continuously distributed and \( E(F) = \Phi \), then
\[\mathcal{E}(\ell_1) < \lambda_1, \; \mathcal{E}(\ell_k) > \lambda_k, \; \mathcal{E}(\ell_k - \ell_1) > \lambda_k - \lambda_1\]  
(2.1,9)

**Proof.**

As by equations (2.3) and (2.2) \(\lambda_1 = \mathcal{E}(\hat{\ell}_1), \lambda_k = \mathcal{E}(\hat{\ell}_k)\), one finds readily

\[\lambda_1 - \mathcal{E}(\hat{\ell}_1) = \mathcal{E}(\hat{\ell}_1 - \ell_1) = \int (\hat{\ell}_1 - \ell_1) dP > 0;\]

the last inequality sign holding true because of Lemma 2.1.1, see for instance Halmos, 1950, p. 104. Similarly

\[\mathcal{E}(\ell_k) - \lambda_k = \int (\ell_k - \hat{\ell}_k) dP > 0\]

\[\mathcal{E}(\ell_k - \ell_1) - (\lambda_k - \lambda_1) = \int (\ell_k - \hat{\ell}_k) dP + \int (\hat{\ell}_1 - \ell_1) dP > 0\]

The next theorem contains a result closely related to distribution-bias:

**Theorem 2.1.3.**

If \(F\) is continuously distributed and \(\mathcal{E}(F) = \Phi\), then

\[P(\ell_1 \leq \xi_1) \geq P(\hat{\ell}_1 \leq \xi_1) \text{ for any } \xi_1\]  
(2.1,10a)

\[P(\ell_k \geq \xi_k) \geq P(\hat{\ell}_k \geq \xi_k) \text{ for any } \xi_k\]  
(2.1,10b)

\[P(\ell_k - \ell_1 \geq \xi) \geq P(\hat{\ell}_k - \ell_1 \geq \xi) \text{ for any } \xi.\]  
(2.1,10c)

**Remark.** If \(\hat{\ell}_1\) and \(\hat{\ell}_k\) could be regarded as estimators (which, unfortunately, they cannot in general, cf. section 2) and if each of the three inequalities in this theorem were strict for at least one value of \(\xi, \xi_1, \xi_k\), then these inequalities would mean that \(\hat{\ell}_1\) would be negatively, \(\hat{\ell}_k\) and \(\ell_k - \ell_1\) positively distribution-biased estimators of \(\lambda_1, \lambda_k, \text{ and } \lambda_k - \lambda_1\) with respect to the comparing estimators \(\hat{\ell}_1, \hat{\ell}_k, \text{ and } \hat{\ell}_k - \hat{\ell}_1\) (cf. van der Vaart, 1957a, definition 3). As it is, this theorem may provide us with limits for certain probabilities and in a special case with a proof of median-bias (see corollary b of this theorem).
Proof.

The proof ensues immediately from the simple lemmata 1 and 1' in section 4 of van der Vaart (1957a). In lemma 1, contending that

if $P(z \geq \xi) = 1$, then $P(m \leq \xi) \geq P(m+z \leq \xi + \zeta)$, whether the random variables $m$ and $z$ are independent or not,

substitute $\zeta$ by 0, $z$ by $(\ell_1 - \ell_1)$, $m$ by $\ell_1$ to find (2.1,10a). In lemma 1', i.e., contending that

if $P(z \leq \xi) = 1$, then $P(m \geq \xi) \geq P(m+z \geq \xi + \zeta)$, whether $m$ and $z$ are independent or not,

substitute $\zeta$ by 0, $z$ by $(\hat{\ell}_k - \ell_k)$, $m$ by $\ell_k$ to find (2.1,10b), and $\xi$ by 0, $z$ by $(\hat{\ell}_k - \ell_k - \ell_1 + \ell_1)$, $m$ by $\ell_k - \ell_1$ to find (2.1,10c).

Corollary of Theorem 2.1.3.

If each interval in $f_{ij} (1 \leq i \leq j \leq k)$-space has positive probability, then inequalities (2.1,10a, b, and c) are strict.

Proof.

We may remark that by interval we mean a set of points in $F$-space such that

$\alpha_{ij} < f_{ij} < \beta_{ij} (1 \leq i \leq j \leq k),$

where $\alpha_{ij}$ as well as $\beta_{ij}$ are allowed to be either finite or infinite (in the latter case: $\alpha_{ij} = -\infty$, $\beta_{ij} = +\infty$). Now if we want to prove the three inequalities (2.1,10a, b and c) to be strict, it will be sufficient to show that

$P[(\hat{\ell}_1 > \xi_1) \cap (\ell_1 < \xi_1)] > 0,$ \hspace{1cm} (2.1,11a)

$P[(\hat{\ell}_k < \xi_k) \cap (\ell_k > \xi_k)] > 0,$ \hspace{1cm} (2.1,11b)

$P[(\hat{\ell}_k - \ell_k < \xi) \cap (\ell_k - \ell_1 > \xi)] > 0,$ \hspace{1cm} (2.1,11c)
respectively (cf. equation (4.2) in Lemma 1 in van der Vaart, 1957a). As by the condition of the corollary each interval containing any point in F-space, hence each corresponding neighbourhood of any point in the Cartesian product of V-space and L-space, has positive probability, all we have to show is that there are points in (V-space) x (L-space) in which both
\[ \hat{\ell}_1 - \xi_1 > 0 \quad \hat{\ell}_k - \xi_k < 0 \quad \hat{\ell}_k - \hat{\ell}_1 - \xi < 0 \]
and
\[ \ell_1 - \xi_1 < 0 \quad \ell_k - \xi_k > 0 \quad \ell_k - \ell_1 - \xi > 0 \]
For, we may then surround such points by a neighbourhood (corresponding to an interval in F-space) all points of which satisfy the same inequalities, whence the proof will be complete.

Now given any
\[ \xi_1 \quad \xi_k \quad \xi > 0 \]
choose
\[ \ell_1 < \xi_1, \quad \ell_k > \xi_k \]
\[ \ell_2 \text{ with } \ell_2 - \xi_2 > \xi_1 - \ell_1, \quad \ell_{k-1} \text{ with } \xi_k - \xi_{k-1} > \ell_k - \xi_k \]
\[ \ell_h \geq \ell_{h-1} \text{ (h=3, \ldots, k),} \quad \ell_h \leq \ell_{h+1} \text{ (h=1, \ldots, k-2),} \]
\[ v_{11}^2 < \frac{1}{2}, \quad v_{kk}^2 < \frac{1}{2} \]
\[ \xi_1 \text{ and } \xi_k \text{ satisfying } \]
\[ \xi = \xi_k - \xi_1; \ell_1, \ell_k, \ell_2 \]
\[ \ell_{k-1}, v_{11}^2, \text{ and } v_{kk}^2, \text{ as in the first two columns,} \]
though in such a way that
\[ \ell_{k-1} > \ell_2 \text{ and } \ell_2 \leq \ell_3 \leq \cdots \leq \ell_{k-1} \]
Then
\[ \hat{\ell}_k = \sum_{h=1}^{k} v_{kh} \ell_h < \ell_k < \xi_k < \hat{\xi}_k \]

\[ \hat{\ell}_1 = \sum_{h=1}^{k} v_{lh} \ell_h = \sum_{h=2}^{k} v_{lh} \ell_h < \ell_1 < \xi_1 < \hat{\xi}_1 \]

\[ \xi_1 \]

\[ \ell_1 - \ell_1 < \xi_1 - \xi_1 = \xi \]

If in (2.1,11g) \( \xi \leq 0 \) the proof in the third column does not apply. Because of \( P(\ell_k - \ell_1 > 0) = 1 \), all we have to show in this case is that there are points in the Cartesian product of V-space and L-space in which

\[ \hat{\ell}_k - \hat{\ell}_1 = \sum_{h=1}^{k} (v_{kh}^2 - v_{lh}^2) \ell_h < \xi \leq 0. \]

Choose \( \ell_1 \) and \( \ell_2 \) such that \( \ell_k - \ell_1 > -\xi \geq 0 \), hence \( \ell_1 - \ell_k < \xi \leq 0 \); choose \( v_{kl}^2 = 1 = v_{lk}^2 \). Then \( \hat{\ell}_k - \hat{\ell}_1 = \ell_1 - \ell_k < \xi \).
Corollary b of Theorem 2.1.3

In order that \( \ell_1, \ell_k \) and \( \ell_k - \ell_1 \) be negatively, positively and positively median-biased estimators of \( \lambda_1, \lambda_k \), and \( \lambda_k - \lambda_1 \) respectively, it is sufficient that the \( \frac{1}{2} k(k+1) \)-variate probability density function of the \( f_{ij} \) \( (1 \leq i \leq j \leq k) \) be symmetrical with respect to the point with coordinates \( \varphi_{ij} = \mathcal{E}(f_{ij})(1 \leq i \leq j \leq k) \).

Proof.

As the probability density function is determined only up to its values on a set of Lebesgue measure (on \( F \)-space) zero, the statement of this corollary ought to be read with this qualification in mind. Now put

\[
\xi_1 = \lambda_1, \quad \xi_k = \lambda_k, \quad \xi = \lambda_k - \lambda_1
\]

in the inequalities

\[
(2.1.10a), (2.1.10b), (2.1.10c),
\]

respectively of Theorem 2.1.3. Then the proof will be complete if

\[
P(\hat{\ell}_1 \leq \lambda_1) = \frac{1}{2}, \quad P(\hat{\ell}_k \geq \lambda_k) = \frac{1}{2}, \quad P(\hat{\ell}_k - \hat{\ell}_1 \geq \lambda_k - \lambda_1) = \frac{1}{2},
\]

respectively. Because of equations (2,3) and (1.2,1d).

\[
Y^{\dagger}F_Y = \Lambda \quad \text{and} \quad Y^{\dagger}\phi Y = \Lambda.
\]

Thus the medians of

\[
\sum_{i=1}^{k} \sum_{j=1}^{k} u_{ij} f_{ij} - \varphi_{ij}, \quad \sum_{i=1}^{k} \sum_{j=1}^{k} u_{ik} u_{jk} (f_{ij} - \varphi_{ij}), \quad \sum_{i=1}^{k} \sum_{j=1}^{k} (u_{ik} u_{jk} - u_{il} u_{jl}) (f_{ij} - \varphi_{ij})
\]

are to be shown to be zero. As these expressions when equated to zero represent hyperplanes in \( F \)-space which contain the center of symmetry, the proof is evident.

(It was originally thought that \( \ell_k \) and \( \ell_1 \) are median-biased for any continuous distribution of \( F \), cf. van der Vaart, 1957b; for other than symmetrical
distributions we have not been able to prove or disprove this idea, however; for symmetrical distributions the median-bias of $\hat{\lambda}_k - \lambda_1$ has been added to the original idea.)

Finally we want to remark that though the existence of negative and positive expectation-bias of the latent roots is to a large extent distribution-free, and the existence of median-bias is so to a lesser extent, the amounts

$$\mathcal{E}(\hat{\lambda}_1) - \lambda_1, \quad \mathcal{E}(\hat{\epsilon}_k) - \lambda_k$$

of the expectation-bias, and the amounts

$$\text{Med}(\hat{\lambda}_1) - \lambda_1, \quad \text{Med}(\hat{\epsilon}_k) - \lambda_k, \quad \text{Med}(\hat{\epsilon}_k - \hat{\lambda}_1) - \lambda_k + \lambda_1,$$

say,

of the median-bias, depend on the distribution of $F$. In case $k = 2$, the formula

$$\mathcal{E}(\lambda_1 - \hat{\lambda}_1) = \mathcal{E}(\hat{\epsilon}_1 - \hat{\lambda}_1) = \mathcal{E}(\frac{\lambda_2}{\lambda_1}(\hat{\epsilon}_2 - \hat{\lambda}_1)),$$

(2.1,12)

a special case of (2.1,5a), will turn out to be useful in computations of this amount under more precise assumptions concerning the distribution of $F$.

2.2 Some theorems on the variances and covariances of the elements of $L$ and $\hat{L}$.

As we are going to use many summation signs in the next two sections, we want to simplify their notation: we shall replace

$$\sum_{g=1}^{k} \text{by } \sum_g, \quad \sum_{i=1}^{k} \text{ by } \sum_i, \quad \sum_{i=1}^{k} \sum_{j=1}^{k} \text{ by } \sum_{i,j}, \quad \sum_{i=1}^{k-1} \sum_{j=1}^{k} \text{ by } \sum_{i<j}.$$

To begin with, we shall prove a theorem on the sum of the variances of the latent roots of $F$, which will be very useful as a check on computations concerning these variances:

Theorem 2.2.1

If $F$ is continuously distributed and $\mathcal{E}(F) = \Phi$, then
\[
\sum_g \text{var} \, \hat{\epsilon}_g = \sum_{i,j} \text{var} \, f_{ij} + \sum_{i,j} \chi^2_{ij} - \sum_g (\epsilon_g)^2 \\
= \sum_{i,j} \text{var} \, f_{ij} + \sum_g \chi^2_g - \sum_g (\epsilon_g)^2 \tag{2.2.1}
\]

**Proof.**

The equality of the last two members of (2.2.1) follows from

\[
\sum_g \chi^2_g = \text{tr}(\Lambda \Lambda) = \text{tr}(Y'\Psi Y'\Psi Y) = \text{tr}(Y'\Phi Y) = \text{tr}(\Phi \Phi) = \sum_{i,j} \chi^2_{ij} \tag{2.2.2}
\]

Besides the definition of trace and the orthogonality of \( Y \) the equations (1.2.1d) and (2.1.3) have been applied here. In quite the same way one obtains

\[
\sum_g \epsilon^2_g = \sum_{i,j} f^2_{ij} \tag{2.2.3}
\]

The equality of the first two members of (2.2.1) then results if

\[
\sum_g (\epsilon_g)^2 + \sum_{i,j} \chi^2_{ij}
\]

is subtracted from both sides of the equation

\[
\sum_g \epsilon_g^2 = \sum_{i,j} f_{ij}^2,
\]

which ensues from taking expectations in (2.2.3).

If \( k = 2 \), Theorem 2.2.1 can be reduced in an interesting way:

**Corollary of Theorem 2.2.1**

If \( k = 2 \), then under the conditions of Theorem 2.2.1

\[
\sum_g \text{var} \, \hat{\epsilon}_g = \sum_{i,j} \text{var} \, f_{ij} - 2\lambda_2 \chi_\beta - 2(\lambda_2 - \lambda_1) \chi_\beta', \tag{2.2.4a}
\]

here \( \chi_\beta \) is the amount of expectation-bias of the latent roots,

\[
\chi_\beta = \epsilon(\lambda_2 - \lambda_2) = -\epsilon(\lambda_1 - \lambda_1) > 0 \tag{2.2.4b}
\]
Proof.

As for (2.2,1b) compare Theorems 2.1.1 and 2.1.2. Equation (2.2,1a) ensues from equation (2.2,1) in Theorem 2.2.1 since for k = 2

\[ \sum_g \left[ (E \ell g^2 - \lambda g^2) \right] = \sum_g (E \ell g + \lambda g)(E \ell g - \lambda g) = \sum_g (E \ell g - \lambda g + 2\lambda g)(E \ell g - \lambda g) = \]

\[ = (\alpha \beta + 2\lambda_1)(\alpha \beta - \alpha \beta) + (\alpha \beta + 2\lambda_2)\alpha \beta = 2\lambda \alpha^2 + 2(\lambda_2 - \lambda_1)\alpha \beta . \]

In those cases in which var \( \ell_1 \) = var \( \ell_2 \), this corollary enables us to compute the variances of both latent roots as soon as \( \alpha \) is known. Section 2.3 will derive a condition which is sufficient for these variances to be equal.

A result on \( \hat{L} \) which is analogous to Theorem 2.2.1 on \( L \), is contained in

**Theorem 2.2.2.**

If \( F \) is continuously distributed and \( E(F) = \Phi \), then

\[ \sum_{g,h} \text{var} \hat{\ell} gh = \sum_{g} E(\ell^2 g - \lambda g^2) = \sum_{i,j} \text{var} f_{ij} \] (2.2,5)

Proof.

The main difference with Theorem 2.2.1 is that \( L \) is a diagonal matrix while \( \hat{L} \) is not, and that \( \hat{L} \) is an expectation-unbiased estimator of \( L \), while \( L \) is not. Using equations (2.3) and (1.2,1d) one finds by the same kind of argument that was used in equation (2.2,2):

\[ \text{tr}(\hat{L}^2) = \text{tr}(LL) = \text{tr}(FF), \]

i.e.

\[ \sum_{g,h} \hat{\ell}^2 gh = \sum_{g} \ell^2 g = \sum_{i,j} f^2_{ij} \]

After taking expectations and subtracting

\[ \sum_{g,h} \hat{\ell}^2 gh \lambda h = \sum_{g} \lambda^2 g = \sum_{i,j} \varphi^2_{ij} \] (cf. eq. (2.2,2) )
- note that $\mathcal{E}_h^L = \Lambda$, hence $\mathcal{E}_{gh} = \delta_{gh} \lambda_h$, cf. eq. (1.1.6) - one finds equation (2.2.5).

Finally we list the following result on the covariances of certain elements of $L$ and $\hat{L}$

**Theorem 2.2.3**

If $F$ is continuously distributed and $\mathcal{E}(F) = \Phi$, then

$$\sum_{g,h} \text{cov} (\ell_g^g, \ell_h^h) = \sum_{i,j} \text{cov} (\ell_{ii}^i, \ell_{jj}^j) = \sum_{g,h} \text{cov} (\hat{\ell}_g^g, \hat{\ell}_h^h). \quad (2.2.6)$$

**Proof.**

It is easy to prove from

$$\text{cov} (\ell_g^g, \ell_h^h) = \mathcal{E}(\ell_g^g \ell_h^h) - \mathcal{E}\ell_g^g \mathcal{E}\ell_h^h$$

that

$$\sum_{g,h} \text{cov} (\ell_g^g, \ell_h^h) = \mathcal{E} (\text{tr } L)^2 - (\mathcal{E} \text{tr } L)^2.$$

Likewise

$$\sum_{i,j} \text{cov} (\ell_{ii}^i, \ell_{jj}^j) = \mathcal{E} (\text{tr } F)^2 - (\mathcal{E} \text{tr } F)^2$$

and

$$\sum_{g,h} \text{cov} (\hat{\ell}_g^g, \hat{\ell}_h^h) = \mathcal{E} (\text{tr } \hat{L})^2 - (\mathcal{E} \text{tr } \hat{L})^2.$$

Now as a consequence of equation (2.1.3) and the orthogonality of $U$ and $V$ we find, using equations (1.2.1d) and (2,3):

$$\text{tr } L = \text{tr } (U^t FU) = \text{tr } F,$$

$$\text{tr } L = \text{tr } (VLV^t) = \text{tr } \hat{L}.$$

This proves the theorem.

Note that if $k = 2$ this theorem together with equations (2.2.4) of the Corollary of Theorem 2.2.1 enables us to compute $\text{cov} (\ell_1, \ell_2)$ solely from the amount of expectation-bias $\alpha_\beta$. 
2.3 A condition for symmetry in the distribution of the latent roots.

In connection with the Corollary concerning the case \( k = 2 \) of Theorem 2.2,1 we observed that it would simplify certain computations if we knew that \( \text{var} \, \ell_1 = \text{var} \, \ell_2 \); in fact the knowledge of \( \lambda_2 = \mathcal{E}(\ell_2 | \lambda_2) \) would then be sufficient to compute \( \text{var} \, \ell_1 \) and \( \text{var} \, \ell_2 \). Hence it would be interesting to know conditions which, if satisfied by the distribution of \( F \), would guarantee that \( \text{var} \, \ell_1 = \text{var} \, \ell_2 \).

As this paper is restricted to distribution-free results, we shall replace this equality of variances by a condition which can be handled more readily by distribution-free methods: that of some kind of symmetry in the joint distribution of \( \ell_1 \) and \( \ell_2 \). If \( q \) denotes the joint probability density of \( \ell_1 \) and \( \ell_2 \), then because of definition (1.1,5) of \( \ell_1 \) and \( \ell_2 \)

\[
q(\ell_1, \ell_2) = 0 \text{ if } \ell_1 > \ell_2 .
\]

Hence a symmetry like

\[
q(\ell_1, \ell_2) = q(\ell_2, \ell_1)
\]

would lead to \( q \) differing from zero only if \( \ell_1 = \ell_2 \). A far more interesting case is afforded by the symmetry defined by

\[
q(\ell_1, \ell_2) = q(\gamma - \ell_2, \gamma - \ell_1) .
\]

We shall consider the problem for arbitrary \( k \) instead of \( k = 2 \). We shall first list a few properties of such distributions of \( L \) as satisfy

\[
q(\ell_1, \ell_2, \ldots, \ell_k) = q(\gamma - \ell_k, \gamma - \ell_{k-1}, \ldots, \gamma - \ell_1) \quad (2.3,1a)
\]

and then find sufficient conditions on the distribution of \( F \) in order that the distribution of the latent roots of \( F \) satisfied (2.3,1a).

Condition (2.3,1a) is not the only possible generalization of the case of only two latent roots, but it is well suited to our purpose.
Writing condition (2.3,1a) as a condition of equivalence of two random matrices (see section 1.1) makes the development transparent. If we use the symbol \( \sim \) for equivalence, (2.3,1a) assumes the form:

\[
L \sim \gamma I - ^*L.
\]  

(2.3,1b)

Here \( I \) is the identity matrix, while

\[
L = \left\| \delta_{gh} \right\|_{h=1}^{k} \quad \text{and} \quad ^*L = \left\| \delta_{gh} \right\|_{h=1}^{k-1},
\]

(2.3,1c)

cf. definition (1.1,7).

As the expectations of functions of equivalent matrices are equal (cf. section 1.1), we have

\[
\mathcal{E}L + \mathcal{E}^*L = \gamma I,
\]

(2.3,2a)

whence

\[
\mathcal{E} \ell_g + \mathcal{E} \ell_{k+1-g} = \gamma \quad (g = 1, \ldots, k)
\]

(2.3,2b)

and

\[
2 \mathcal{E} \text{tr} ^*L = 2 \mathcal{E} \text{tr} L = 2 \text{tr} \Lambda = 2 \text{tr} \Phi = k \gamma
\]

(2.3,2c)

Considering the direct product of \( L \) and \( L \) (cf. MacDuffee, 1946, p. 81; as \( L = L \), we need not distinguish between left and right direct product), we find

\[
L \times L \sim (\gamma I - ^*L) \times (\gamma I - ^*L),
\]

\[
\mathcal{E} [L \times L] = \mathcal{E} [(\gamma I - ^*L) \times (\gamma I - ^*L)]
\]

\[
= \gamma^2 (I \times I) - \mathcal{E} (\gamma I - ^*L) \times \mathcal{E} (\gamma I - ^*L)
\]

Subtracting from this equality the equality

\[
(\mathcal{E}L) \times (\mathcal{E}L) = [ \mathcal{E} (\gamma I - ^*L)] \times [ \mathcal{E} (\gamma I - ^*L)]
\]

\[
= \gamma^2 (I \times I) - (\mathcal{E}^*L) \times \gamma I - \gamma I \times (\mathcal{E}^*L) + (\mathcal{E}^*L) \times (\mathcal{E}^*L)
\]

(\mathcal{E}L) \times (\mathcal{E}L) = [ \mathcal{E} (\gamma I - ^*L)] \times [ \mathcal{E} (\gamma I - ^*L)]
we find
\[ \mathcal{E} [L \times L - (\mathcal{E} L) \times (\mathcal{E} L)] = \mathcal{E} [^*L \times ^*L - (\mathcal{E}^*L) \times (\mathcal{E}^*L)] \]
whence from the definition of direct product
\[ \text{cov} (\ell_{g_1}, \ell_{g_2}) = \text{cov} (\ell_{k+1-g_1}, \ell_{k+1-g_2}), \quad (2.3,2a) \]
and for \( g_1 = g_2 = g \)
\[ \text{var} \ell_g = \text{var} \ell_{k+1-g}, \quad (2.3,2a) \]
whence it appears that \( (2.3,1b) \) is sufficient to guarantee the equality of certain variances. Hence we want now to find a condition for the distribution of \( F \) which is sufficient in order that \( L \sim \gamma I^{-^*L} \). We are going to prove

**Theorem 2.3.1.**

Denote by \( \Omega_1 \) and \( \Omega_2 \) two arbitrary real, non-singular, \( k \times k \) matrices, which may be non-random or random, and let \( L \) and \( ^*L \) be defined as in \((2.3,1c)\). Let \( F \) be a real symmetric, random \( k \times k \) matrix. Then in order that the random (diagonal)

matrices
\[ L \quad \text{and} \quad \gamma I^{-^*L} \]

be equivalent, it is sufficient that some pair of matrices \( \Omega_1 \) and \( \Omega_2 \) exists such that the random matrices

\[ \Omega_1^{-1} F \Omega_1 \quad \text{and} \quad \gamma I^{-\Omega_2^{-1}} F \Omega_2 \]

are equivalent. This condition entails that
\[ k \gamma = 2 \text{ tr } \mathcal{E} (F) = 2 \text{ tr } \Phi, \quad (2.3,1n) \]
and in case \( \Omega_1 \) and \( \Omega_2 \) are not random,
\[
\sum_{i,j} \varphi_{ij} \left[ \{ \Omega_1^{-1} \}_i \{ \Omega_1 \}_j + \{ \Omega_2^{-1} \}_i \{ \Omega_2 \}_j \right] = \gamma \cdot \delta_{gh} \quad (g,h=1, \ldots, k)
\]
(2.3,4b)

Here \( \delta_{gh} \) is the Kronecker symbol and the symbol \( \frac{M}{j_{i,ij}} \) (\( M \) any matrix) stands for \( m_{ij} \).

**Remark.** As \( \Omega_1 \) and \( \Omega_2 \) may be random as well as non-random, we could not conform to our general notation system of Latin capitals for random matrices, Greek capitals for non-random matrices.

**Proof.**

The \( k \) latent roots \( \ell_1 \leq \ell_2 \leq \cdots \leq \ell_k \) of any matrix \( F \) define the matrix \( L \) as a function of \( F \):
\[
L = \psi(F),
\]
(2.3,5a)

Though \( F \) and \( L \) may be random, the functional relation \( \psi \) is not. It is well known that the latent roots of the matrices \( \Omega_1^{-1} F \Omega_1 \) and \( F \) are the same, hence
\[
L = \psi(F) = \psi(\Omega_1^{-1} F \Omega_1).
\]
(2.3,5b)

Likewise it is well known that if \( \ell_1 \leq \ldots \leq \ell_k \) are the latent roots of \( F \), then
\[
\gamma^{-\ell_k} \leq \ldots \leq \gamma^{-\ell_1}
\]
are the latent roots of \( (\gamma I - F) \), so
\[
\psi(\gamma I - F) = \gamma I - \gamma L.
\]

Hence as the latent roots of \( (\gamma I - F) \) and of \( \Omega_2^{-1} (\gamma I - F) \Omega_2 = \gamma I - \Omega_2^{-1} F \Omega_2 \) are the same, we find,

**cf.** Lappo-Danilevsky (1953), vol. 1, p. 1.
\[ \gamma I - \gamma L = \gamma I - \Omega_{2}^{-1} F \Omega_{2}. \]  

Thus the matrix \((\gamma I - \gamma L)\) is connected with the matrix \((\gamma I - \Omega_{2}^{-1} F \Omega_{2})\) by exactly the same functional relationship by which \(L\) is connected with \((\Omega_{1}^{-1} F \Omega_{1})\). Hence if two matrices \(\Omega_{1}\) and \(\Omega_{2}\) can be found such that \((\Omega_{1}^{-1} F \Omega_{1})\) and \((\gamma I - \Omega_{2}^{-1} F \Omega_{2})\) are equivalent, then \(L\) and \((\gamma I - \gamma L)\) are equivalent. Finally if \((\Omega_{1}^{-1} F \Omega_{1})\) and \((\gamma I - \Omega_{2}^{-1} F \Omega_{2})\) are equivalent, then

\[ E \text{tr}(\Omega_{1}^{-1} F \Omega_{1}) = E \text{tr}(\gamma I - \Omega_{2}^{-1} F \Omega_{2}) = k \gamma = E \text{tr}(\Omega_{2}^{-1} F \Omega_{2}), \]

or

\[ \text{tr} \Phi = k \gamma - \text{tr} \Phi, \]

which proves equation (2.3,2a), which is as it should be because of (2.3,2a).

Further the expectations of corresponding elements of the two equivalent matrices \((\Omega_{1}^{-1} F \Omega_{1})\) and \((\gamma I - \Omega_{2}^{-1} F \Omega_{2})\) are equal, which proves equation (2.3,4b).

Example.

Take \(k = 2, \Omega_{1} = I, \Omega_{2} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \), write

\[ \Omega_{2}^{-1} F \Omega_{2} = \begin{pmatrix} f_{22} & -f_{21} \\ -f_{12} & f_{11} \end{pmatrix}, \]

Then Theorem 2.3,1 shows that if

\[ F \sim \gamma I - \gamma F, \]

then

\[ L \sim \gamma I - \gamma L. \]

It is interesting to see which consequences condition (2.3,6a) has for the first and second order moments of the distribution of \(F\). We find from

...
\[ E_F = \gamma I - E_F^* \]

that
\[ E_{f_{11}} + E_{f_{22}} = \gamma, \quad \text{cf. (2.3,4b)} \]

in this case equation (2.3,4b) yields just this. Furthermore, from
\[ F \times F \sim (\gamma I - E^*) x (\gamma I - E^*) \]

we derive by the same argument which has yielded equations (2.3,2a) and (2.3,2b):
\[ E [F \times F - (E^* F) x (E^* F)] = E [E^* F x E^* F - (E^* F) x (E^* F)] \]

whence from the definition of direct product:
\[ \text{var} f_{11} = \text{var} f_{22}, \quad \text{cov} (f_{11}, f_{12}) = -\text{cov} (f_{12}, f_{22}) \]  

(2.3,7b)

There are no consequences for \( E(f_{12}), \text{var} f_{12}, \text{cov} (f_{11}, f_{22}) \).

Finally one might note that with the simple \( \Omega_1 \) and \( \Omega_2 \) applied in this example the resulting condition (2.3,6a) is very closely related to the type of symmetry in the distribution of \( F \) one would intuitively expect to guarantee the desired symmetry in the distribution of \( L \). Yet with more general \( \Omega_1 \) and \( \Omega_2 \) the resulting symmetry appears to be somewhat more intricate.

3. Application of the preceding theory to some aspects of the estimation of quadratic response surfaces

3.1 A few general considerations.

Be \( \xi \) and \( \beta \) real \( k \times 1 \) matrices (column vectors), \( \Phi \) a \( k \times k \) real symmetric matrix, \( \eta \) a real scalar variable, a function of \( \xi, \beta \) and \( \Phi \), which has the value \( \eta_0 \) if \( \xi \) is a vector of zeros; here \( \xi, \beta, \Phi \) and \( \eta \) are non-random quantities, \( \beta \) and \( \Phi \) being thought of as unknown constants (parameters) in a given experimental problem, \( \xi \) and \( \eta \) being supposed to be given (and to assume, respectively) different values in different experiments connected with the same given problem...
(f.i., \( \xi_1, \ldots, \xi_k \) may represent a set of \( k \) "conditions" like temperature, humidity, etc., each of which may be made to vary a great deal over different experiments concerning one problem, \( \eta \) may represent the "yield" or "response" of some process which is made to take place at each of a number of different levels of this set of \( k \) conditions). The equation of a quadratic response surface has the form

\[
\eta = \eta_0 + \beta' \xi + \xi' \Phi \xi,
\]

which determines the yield \( \eta \) as a function of the \( k \) conditions represented by the vector \( \xi \). If \( \xi \) and \( \eta \) could be experimentally realized and observed with exactitude, there would be little need for statistics in this problem. However, as soon as one of them cannot, statistics has to come in. As in most, if not all, statistical work on the estimation theory of response surfaces, we are going to assume that \( \xi \) can indeed be experimentally realized with exactitude, but \( \eta \) cannot: instead of the "true" yield \( \eta \) we observe realizations of a stochastic variable \( y \) with

\[
\xi y = \eta, \quad \text{var} \ y = \sigma^2,
\]

(cf. Box and Wilson, 1951, p. 2, Box and Hunter, 1957, p. 198, Bloemen, 1956, p. 8, Cochran and Cox, 1957, p. 335. Note that in equations (3.1,2) \( \eta \) is a function of \( \xi \) according to equation (3.1,1) whereas \( \sigma^2 \) is not: \( \text{var} \ y \) is assumed to have the same value no matter what values the conditions \( \xi \) have. It should be emphasized that we will investigate only quadratic response surfaces, hence will not consider possible consequences of the model being at fault, i.e. of specification errors.

The general procedure now is to choose wisely a set of \( N \), say, values of \( \xi \), i.e. a set of \( N \) combinations of experimental conditions, and for each of these to observe \( y \), finally to estimate \( \eta_0, \beta \) and \( \Phi \) from these data. This estimation is usually done by the method of least squares (cf. Box and Wilson, 1951, p. 5, Box
and Hunter, 1957, p. 198), which under the assumptions stated yields expectation-unbiased estimators of $\eta_0$, $\beta$ and $\Phi$. If one wants to find out the type of quadratic surface one is dealing with, one may reduce its equation to canonical form if $\eta_0$, $\beta$ and $\Phi$ are known exactly. Box and Wilson, 1951, pp. 23-24, Box, 1954, p. 35, Box and Youle, 1955, p. 289, Box and Hunter, 1957, p. 239 ("A fitted second degree equation can be interpreted most readily by writing it in the canonical form") apply this device also if only estimates of $\eta_0$, $\beta$ and $\Phi$ are known. The next section will discuss this practice.

Before we start this discussion, however, it will be appropriate to devote some thought to the question of scaling, i.e. to the question of the choice of the units in which $x_1$, $x_2$, ..., $x_k$ are to be measured. Before an equation like (3.1,1) can be written down, one must have chosen a set of units. A change in the units of some or all $x$'s will bring about a change in the value of $\beta$ and $\Phi$ and by choosing these units one can make the elements of $\beta$ and of $\Phi$ have any value one wants. On top of this, the papers by Box and his associates use a kind of normalization of the units, i.e. of coding of the $N$ values of $\xi$ for which $y$ is going to be observed, which is handled slightly differently in different papers. Box and Wilson, 1951, give a certain amount of theory on an equation like (3.1,1) without mentioning units at all (not restricting themselves to second degree polynomials), and then on p. 8 introduce a kind of normalization with the sole purpose of being able to compare two designs, i.e. two different sets of values of the vector $\xi$ for which $y$ is to be observed: in each of the two designs they change the units in which $\xi_i$ will be measured in such a way that

$$s^2_i = \sum_{u=1}^{N} \frac{(\xi_iu - \bar{\xi}_i)^2}{N}$$

(3.1,3)

will have the same value in both designs ($i=1$, ..., $k$). Here $\xi_iu$ is the value of
the condition represented by $\xi_i$ in the $u^{th}$ set of experimental conditions 
($u=1, \ldots, N$), and $\bar{\xi}_i$ is the mean of $\xi_{iu}$ over $u$. Evidently they allow $s_i^2$ to 
have different values for different $i$-values. In Box and Youle, too, 1955, p. 292, 
the $\xi$'s are coded, though obviously with the aim of getting simple $\xi$-values 
("for convenience"), and no mention is made of any quantity like $s_i^2$. In Box and 
Hunter, 1957, p. 196 the original idea is used again and even to a greater extent: 
standardized levels

$$x_{iu} = \frac{\xi_{iu} - \bar{\xi}_i}{s_i \sqrt{c}} \quad (u=1, \ldots, N; i=1, \ldots, k), \quad (3.1,4)$$

cf. (3.1,3), are introduced (c an arbitrary constant, independent of $i$) so that

$$\sum_{u=1}^{N} x_{iu} = 0, \quad \sum_{u=1}^{N} x_{iu}^2 = N/c \quad (i=1, \ldots, k). \quad (3.1,4a)$$

So in this set-up, if one would calculate a quantity similar to $s_i^2$ (cf. equation 
(3.1,3)) in terms of these new variables $x_{iu}$, then $s_i^2$ would have the same value 
for all $i$, that is for each experimental condition. As Box and Hunter, 1957, 
p. 196, write, equation (3.1,4) allows us to derive designs (i.e. sets of $N$ values 
of the vector $\xi$) which cover the region of immediate interest in any given 
experimental problem, from one fixed design in terms of the variables $x$. (Of 
course more than one fixed design exists, according to the properties which one 
desires for the design one is going to apply, but by means of (3.1,4) each fixed 
design in $(x_1, \ldots, x_k)$-space can be transformed to serve on an infinite range of 
regions to be explored in $(\xi_1, \ldots, \xi_k)$-space).

All this has an interesting consequence for the concept of rotatable designs. 
A rotatable design (defined on p. 205 of Box and Hunter, 1957) leads to the variance 
of the prediction $y$ of the true yield $\eta$ being a function only of
\[ \sum_{i=1}^{k} x_i^2 \]  

(3.1.4b)

the distance between the point in \((x_1, ..., x_k)\)-space where one wants to predict and the center of the design. However, as the whole of their theory is based upon the standardized variables \(x\) rather than the original variables \(\xi\) (cf. equation (4) on page 196 and the consistent use of \(x\)-symbols rather than \(\xi\)-symbols throughout Box and Hunter, 1.c.), this definition of rotatability needs to be restated in terms of the original variables \(\xi\) if one wants to see what it really amounts to.

Now of course the transformation (3.1.4) does not hold just for the points in which \(y\) will be observed. In general it runs

\[ x_i = \frac{\xi_i - \bar{\xi}}{s_i \sqrt{c}} \quad (i=1, ..., k) \]  

(3.1.4c)

(where the \(\bar{\xi}_i\)\((i=1, ..., k)\)) determine the center of the design, i.e. the center of the region of immediate experimental interest), which means that the variance of \(y\) in "rotatable" designs is a function of

\[ \sum_{i=1}^{k} \frac{(\xi_i - \bar{\xi})^2}{c s_i^2} \]  

(3.1.4d)

In view of the criticism which is often voiced against rotatable designs, and which consists in questioning the justification of the definition of distance involved, it seems in order to state exactly what (3.1.4d) reveals as the origin of this*) distance definition: if the values of \(s_i^2\) for different \(i\) are different, this is caused by the fact that the range of interest in the variables \(\xi_1\) was thought by the experimenter to differ from one \(i\)-value to another. Now any experimenter who wants to explore yield as a function of experimental conditions has to make up his mind about this point whether he uses rotatable designs or *) rather than Euclidean distance or yet another type.
does not use them. Hence this seems hardly a cause for criticism of rotatable designs. All they do is roughly this: they make the contour lines of equally exact prediction follow the boundary of the experimentally interesting region as closely as possible.

In the subsequent section we are not going to use the transformation (3.1,10) and we are not going to replace \( \xi \) by \( x \). Once the units have been chosen our theory will apply, whether or not these units have been chosen so as to standardize the variables.

3.2 Bias in the estimation of type of quadratic surface.

If one wants to know the type of quadratic surface represented by equation (3.1,1), then - as we reminded the reader in the previous section - one may reduce this equation to canonical form i.e. rotate the coordinate-axes in such a way that the transform of \( \xi' \Phi \xi \) no longer contains cross-product terms like \( \xi_i \xi_j \). Referring to section 1.2, one would effect the transformation

\[
\xi = Y \zeta
\]

(3.2,1)

by which according to (1.2,1d) the equation (3.1,1) of the quadratic response surface acquires the form:

\[
\eta = \eta_0 + \beta'Y \zeta + \zeta' \Lambda \zeta,
\]

(3.2,2)

where \( Y \) is orthogonal, \( \Lambda \) is diagonal.

The general shape of a surface represented by (3.2,2) can be assessed much easier than that of a surface represented by (3.1,1). Of course if the surface is truly quadratic the matrix \( \Lambda \) decides on the shape everywhere on the surface. If it is not truly quadratic, this will be possible only in the neighbourhood of \( \xi_0 = 0 \), or \( \zeta_0 = 0 \), respectively (equations (3.1,1) and (3.2,2) being regarded as the first terms of a Taylor series).
Assume for a moment that \( k = 2 \). It is clear that in order to assess the shape of the surface one may just as well calculate contour lines from equation (3.1,1), and plot them, inferring the general shape of the surface from them, cf. Cochran and Cox, 1957, p. 352. Note that reducing equation (3.1,1) to canonical form and plotting contour lines are essentially identical methods: one may regard reduction to canonical form as a time-saving device by which to plot contour lines, or in other words, the contour lines are not changed by the transformation (3.2,1).

Now in the papers cited in the second paragraph of section 3.1 \( \eta_0, \beta \) and \( \Phi \) are unknown but (expectation-unbiased) estimates \( y_o, \beta, \Phi \) are available. The authors of these papers then reduce to canonical form the second degree polynomial

\[
y_o + b^t \xi + \xi^t F \xi, \quad \text{cf. (3.1,1)}
\]

i.e. - if we use the notation and the matrices introduced in section 1.2, then they effect the transformation

\[
\tilde{\xi} = U \xi
\]

(3.2,3)

to obtain

\[
y_o + b^t U \xi + \xi^t L \xi, \quad \text{(3.2,4)}
\]

then base their inference about the general shape of the response surface (3.1,1) on the diagonal matrix \( L \). Others, like Mason, 1956, did not compute \( L \), but plotted

\[
y = y_o + b^t \xi + \xi^t F \xi,
\]

which, as was shown above, amounts to the same thing.

Unfortunately, however, Theorem 2.1.2 and Theorem 2.1.3 together with its corollaries \( a \) and \( b \) show that \( L \) is a biased estimator of \( \Lambda \) according to different definitions of bias: the smallest latent root \( \lambda_1 \) of \( F \) is too small, the largest
latent root \( \ell_k \) of \( \Phi \) is too large, and their difference is too large (a) in the sense of expectation-bias (on the average) under quite weak conditions, (b) in the sense of median-bias under stronger conditions which, for instance, are satisfied by \( \Phi \) being multinormally distributed (c) again under quite weak conditions in the sense of distribution-bias with respect to \( \hat{\ell}_1, \hat{\ell}_k, \hat{\ell}_k - \hat{\ell}_1 \), where \( \hat{\ell} \) has been defined by (2,3) and is no true estimator in that it contains elements (\( Y \) or \( V \)) which are unknown and cannot be estimated. Yet, (c) completes the picture set by (a) and (b) which shows that using \( I \) as an estimator of \( \Lambda \) (or drawing contour lines of

\[
y = y_0 + b' \xi + \xi' \Phi \xi
\]

leads to underestimating \( \lambda_1 \), the smallest root of \( \Lambda \), and to overestimating \( \lambda_k \), the largest root of \( \Lambda \), as well as \( \lambda_k - \lambda_1 \), their difference, both on the average and too frequently.

Remark. There is a difference, of course, between underestimation "on the average" and "too frequently" (cf. van der Vaart, 1957a): the frequency (probability) of winding up with an estimate smaller than the true value may be less than one half, yet the mean of these estimates may be smaller than the true value.

Hence, if either the smallest latent root of \( \Phi \) is positive and "close" to zero or the largest latent root is negative and "close" to zero, this method will tend to lead unduly often to the conclusion that the roots have different signs whereas in reality they have not (or that one root is zero whereas either all are positive or all are negative): in other words the conclusion of this method of analysis will be, more frequently than it should, that there is a minimax, (col, or saddlepoint) if there is a minimum or a maximum or a stationary ridge in reality, or that there is a stationary ridge if there is a minimum or a maximum
in reality.

And, as a matter of fact, there seems to be experimental evidence for this conclusion from the theory.

(1) On p. 548 of the book on design and analysis of industrial experiments edited by Davies (1956) an example is given of an analysis which in the first step yields one negative and one smaller positive latent root (besides a negative one that is very close to zero). After a series of experiments with the explicit aim of checking the increase of the yield in the direction indicated by this positive root (whereby the yield turned out to be "remarkably constant" in this direction) another estimate was made possible by the accumulation of a larger body of data; this time the positive root was negligibly small.

Remark. The behaviour of this positive root is in accordance with certain provisional results we obtained if F is multinormally distributed. More data result in a decrease of the overall variance; the amount of bias is a monotonous increasing function of this variance, hence decreases with it. Another interesting point here is that under certain conditions the variance of the estimators \( \hat{\lambda}_g \) of the latent roots \( \lambda_g \) also decreases with the overall variance, and in such a way that the statistical significance of the difference between zero and the latent root in question remains unaltered. Yet it seems justified to say (as is done on p. 548 of Davies, l.c.) that the estimate of such a latent root turns out to be negligibly small - though the authors do not give an explicit statement as to what they mean by this term: this seems to be another example of the concept of material significance as opposite to the concept of statistical significance introduced by Hodges and Lehmann (1954). Finally all this means that \( \lambda \), though it is a biased estimator of \( \lambda \), in the application to
response surface estimation is a consistent estimator of \( \Lambda \) in the sense that the estimates are closer as the design contains more points - whether one makes a number of replicates of a simple design or uses more elaborate designs which contain a large number of points. From the point of view of estimation of \( \Lambda \) this result counteracts (at least as long as no better estimator of \( \Lambda \) is constructed) the third requirement on an experimental design for estimating response surfaces, which was laid down by Box and Hunter (1957), p. 197: "It should not contain an excessively large number of experimental points".

(2) Mason, 1956, p. 94, Fig. 5.6, presents a diagram of contour lines of a fitted second degree response surface which clearly shows a saddlepoint (\( \text{col} = \text{minimax} \)). What he writes concerning this occurrence is very interesting and in direct agreement with the conclusion of our preceding Remark: "Such a surface would appear to be difficult to interpret agronomically. One would certainly like some substantiation of this type of pattern before extending its application too far. A more complete sampling by observation points in the critical region is perhaps in order."

(3) On the session on applications of response surface designs at the Atlantic City Meeting of the American Statistical Association (September 1957) there were reports of similar experiences.

As a conclusion to this section on bias in the estimation of the type of quadratic surface we want to make two statements. Firstly, so far I have never found a report of the estimation of \( \Lambda \) by \( L \) resulting in a minimum or a maximum in cases where a saddlepoint was to be expected. Theoretically this must happen, though "too infrequently", and it would be interesting to hear of any such case. Secondly, we have refrained entirely (cf. the end of the first paragraph of section 3.1) from considering the possible effects of specification errors. It would seem
necessary, however, to investigate the effects of fitting a second degree surface in a third degree situation, both on the relevance of the matrix $\Lambda$ (may be the problem to be considered for general surfaces, even non-polynomial ones, would be the estimation of curvature rather than the estimation of $\Lambda$; the problem thus stated seems to include even the problem which the method of steepest ascent sets out to solve), and on bias in the estimation of $\Lambda$.

3.3 The variance of the coefficients of the canonical form.

If the canonical form (3.2.4) of the estimator $y = b'x + \xi'F\xi$ of the second member of the equation (3.1,1) of a quadratic response surface is used as an estimator of the canonical form (3.2.2) of this response surface, i.e. if $L$ is used as an estimator of $\Lambda$, it is certainly worth while to know something about the distribution of $L$. We have already found a number of results on the expectation and median of the elements of $L$. We shall now make a few observations on their variance, based on Theorem 2.2.1 and its corollary.

The existing literature does not elaborate on this point. Box and Youle (1955), p. 295 state that "appropriate standard errors for these constants can be shown to be of the same order of magnitude as those of the original quadratic and interaction terms". Box and Hunter (1957), p. 240, state that "for any rotatable design the variances of the coefficients are the same in every orientation and since the $B_{ii}$ (our $\ell_i$) are simply the 'quadratic effects' in the directions of the canonical variables they have the same standard errors as have the quadratic effects $b_{ii}$ (our $f_{ii}$) before transformation".

Let us consider the latter, more precise, statement from the standpoint of equality (2.2.1) in Corollary of Theorem 2.2.1. If $\text{var} \ell_1 = \text{var} f_{11}$ and $\text{var} \ell_2 = \text{var} f_{22}$, then this equality would yield

$$\text{var} f_{12} = \beta^2 + (\lambda_2 - \lambda_1)\beta$$

(3.2.5)
This would mean that once \( \Phi = \mathcal{E}(F) \) (determining \( \lambda_1 \) and \( \lambda_2 \), hence \( \lambda_2 - \lambda_1 \)) and \( \text{var} \ f_{12} \) would be fixed, no other feature of the distribution of \( F \) could have any influence on the amount of expectation-bias \( \chi_\beta \): neither \( \text{var} \ f_{11} \) or \( \text{var} \ f_{22} \), nor any higher moments; \( \chi_\beta \) (being positive by definition) would be fixed at

\[
\chi_\beta = \frac{(\lambda_2 - \lambda_1)}{2} + \sqrt{\frac{(\lambda_2 - \lambda_1)^2}{2} + 4 \text{var} \ f_{12}}
\]  

(3.2, 5a)

This seems very unlikely, and provisional calculations based on \( F \) being multi-normally distributed seem to indicate that it is not true.

Where, then, would be the error in the argument on p. 240 of Box and Hunter (1957)? It is based on the top of p. 208 of their paper which states that "every variance and covariance of the b's (f's in our notation) .... must remain constant under rotation", i.e. constant if the design, or the set of \( N \) different values of the vector \( \mathbf{\xi} \) for which \( y \) is to be observed, is rotated. Applying this to the rotation which reduces (using our notation)

\[
y_0 + b' \mathbf{\xi} + \mathbf{\xi}' F \mathbf{\xi}
\]

to canonical form, they conclude that for instance var \( \ell_1 = \text{var} \ f_{11} \). This must be wrong, for this would mean that var \( \ell_{12} = \text{var} \ f_{12} \), whereas \( \ell_{12} = 0 \) per definition, hence with probability one, so that \( \text{var} \ f_{12} = 0 \). The reason why their statement on p. 208 cannot be applied to the rotation which reduces

\[
y_0 + b' \mathbf{\xi} + \mathbf{\xi}' F \mathbf{\xi}
\]

to canonical form, seems to be that this rotation is a random one (characterized by the matrix \( U \)), depending on \( F \), whereas the rotations discussed on p. 208 do not depend on \( F \), and may not be random at all.
Proposal of a new approach to some problems in the evaluation of response surfaces.

Suppose the surface has really an optimum. Then the combination of the methods of steepest ascent and of higher order designs in order to determine this optimum might be proved to be consistent under rather weak conditions by using results and methods in multidimensional stochastic approximation. See for instance Blum (1954) and the basic idea behind the paper by Dvoretzky (1956) on stochastic approximation = numerical iteration + a stochastic element. In this train of thought the results of Crockett and Chernoff (1955) might be applied.
LITERATURE CITED


