SOME PROBLEMS IN MULTIVARIATE ANALYSIS
PART I

BY
CHARLES STEIN

TECHNICAL REPORT NO. 6

PREPARED UNDER CONTRACT Nonr-225(21)
(NR-042-993)
FOR
OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

OCTOBER 25, 1956
SOME PROBLEMS IN MULTIVARIATE ANALYSIS

PART I

by

Charles Stein

TECHNICAL REPORT NO. 6

PREPARED UNDER CONTRACT Nonr-225(21)
(NR-042-993)
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

October 25, 1956
1. Introduction. I shall discuss a number of unsolved problems in multivariate analysis together with some results which constitute partial solutions of some of these problems. I shall not attempt to cover the whole field, but instead consider principally the question of whether the commonly used procedures for simple problems such as multiple regression, estimation of means and testing hypotheses concerning means are optimum. The answer seems to be that in general they are not. Nevertheless I think that in most practical applications they will continue to be used. The usual procedures are maximum likelihood estimates and likelihood ratio tests so that for sufficiently large samples they are nearly optimum (see Wald [11], Le Cam [5]). If the dimension of the problem is not large, it seems likely that sample sizes occurring in practice will usually be large enough for these results to be relevant. However, if the dimension $p$ of the basic multivariate distribution is large, it may be that the number $n$ of random vectors observed must be extremely large for the large sample results to apply, for example there are cases where $n/p^3$ must be large.

For all the problems we shall consider, the usual procedures are invariant under all affine transformations (linear transformations together with translations) of the basic $p$-dimensional sample space which leave the problem invariant. In fact, since they are maximum likelihood estimates
and likelihood ratio tests, they must be invariant under all transformations which leave the problem invariant. In addition, they have very strong optimum properties (for example, the tests are uniformly most powerful) in the class of all invariant procedures. Thus our examples will show that in general there does not exist a minimax procedure which is invariant under all affine transformations leaving the problem invariant. The lack of uniqueness of the minimax procedure is disturbing. The first example of the non-existence of an invariant minimax procedure was given by Peisakoff, but never published. It referred to the free group with two generators. Another example has been given, even for the one dimensional translation group (by Girshick and Savage [1], p. 59) but it involves a highly artificial decision problem, and cannot be transferred to non-sequential problems in testing hypotheses. This type of example indicates that some caution must be exercised in formulating theorems asserting the existence of an invariant minimax procedure even in the most favorable cases. Such theorems are given in some work of Hunt and the author reported in Lehmann [6], in an unpublished thesis of Peisakoff [7] and in a paper of Kudo [4]. They assert that, subject to mild restrictions on the decision problem, if the transformation group in question is a locally compact abelian group or a compact group, the result holds. (The topological requirements are relevant because the class of measurable sets is related to the topology). In addition if the result holds for a normal subgroup and for the corresponding quotient group it holds for the full group. Thus it holds for the group of rigid motions in a Euclidean space, and for the common solvable groups, in particular the multiplicative group of non-singular triangular matrices, i.e. those $\alpha$ with
\( a_{ij} = 0 \) for \( j > i \). All of the results in this paper are obtained by considering this latter group, except for the last example in section 3 and a few general theorems in Part II, which has not yet been written.

A general theory of problems of this sort has not yet been developed, so that this paper unavoidably consists of a disconnected treatment of special problems. In addition to being unavoidable at present this may have the advantages of emphasizing the statistical rather than the mathematical aspects of the problem, and making the results accessible to a class of readers (including the author) who would find a general treatment difficult to absorb, at least at the present time.

In section 2, we give a formulation of the notion of invariance of a statistical problem, which has no originality but merely serves to establish a framework within which to discuss the problem. In section 3, we discuss three examples of problems invariant under the real full linear group where there does not exist an invariant minimax test. The first is a simple statistical example, which must have been known for some time, although its relevance to the class of problems of this paper seems to have been overlooked. The second is more artificial, but much more instructive than the first. The third example is, with some simplification, the first one the author obtained, and is mentioned in an abstract [9]. Later sections of Part I discuss some problems which actually occur in practical multivariate analysis, although they are not the most commonly occurring problems. In section 4 we consider the estimation of a completely unknown covariance matrix. Although the choice of a loss function is somewhat artificial, I believe the result that the sample covariance matrix is not a minimax estimate will hold in general,
say for invariant convex loss functions. Also it is of interest that we 
find that in a certain sense (perhaps too demanding) it is impossible to 
estimate the covariance matrix well unless \( n/p \) is large. In section 5 
we consider the problem of finding confidence sets for the mean of a multi-
variate normal distribution, first considered by Hotelling [2]. Again we 
find that the best invariant procedure is not minimax, but again there is 
some question as to the appropriateness of the loss function. In section 
6 we return to the problem of testing for the value of a constant of proportion-
ality between two covariance matrices, considered briefly at the beginning 
of section 3.

I cannot give a detailed table of contents for Part II, but I expect to 
discuss there the following topics:

(i) Multiple regression in the multivariate normal model with the loss 
equal to the ratio of the mean squared error of prediction to that which 
one would incur if one knew the covariance matrix. The usual procedure is 
minimax, but presumably not admissible if the number of predictors is large 
(perhaps \( \geq 3 \), or maybe even 2, especially if the mean is unknown).

(ii) Hotelling's \( T^2 \) test, where I can give only a reduction of the 
problem together with some conjectures and approximate results.

(iii) A method of deriving the probability ratio of a maximal invariant 
based on some simple considerations concerning invariant measures. This is 
used in section 6 of Part I, but the reader should not have great difficulty 
in constructing an alternative derivation.

(iv) A necessary and sufficient condition that a connected simple 
Lie group \( \mathfrak{g} \) should be such that for any testing problem invariant under \( \mathfrak{g}, \)
there exist an invariant minimax test. This condition was privately communicated to the author (in a slightly different context) by H. Rubin.

(v) A theorem to justify our conclusion in section 5 and in places in Part II that there exists an invariant minimax solution (with respect to the group of triangular matrices). A theorem of Kudo covers the application in section 4, and the results of Hunt and Stein given in Lehmann [6] cover that of section 6.

(vi) Some remarks on the author's paper [10] where the treatment is excessively formal.
2. **Invariant decision problems.** The notion of invariance of a statistical decision problem under a transformation is essentially the same as the notion of invariance in any branch of mathematics. It is a general principle that if a problem with a unique solution is invariant under a certain transformation, then the solution will be invariant under that transformation, in an appropriate sense. Perhaps the principal reason for the strong intuitive appeal of invariant decision procedures is the feeling that there should be a unique best (or correct) way to analyze a collection of statistical data. Nevertheless, in cases where the use of an invariant procedure conflicts violently with the desire to make a correct decision with high probability or have a small expected loss, it must be abandoned.

We must first formulate the notion of a statistical decision problem in a degree of generality which will be adequate for our later work, but not excessive. Let \( \mathcal{X} \) be a set, the sample space, \( \mathcal{B} \) a \( \sigma \)-algebra of subsets of \( \mathcal{X} \), \( \Omega \) a set, the parameter space, and \( P \) a function on \( \Omega \) to the set of probability measures on \( \mathcal{B} \). We denote by \( P_\theta \) the probability measure on \( \mathcal{B} \) corresponding to \( \theta \in \Omega \). Let \( A \) be another set, the action space, and \( L \) a real-valued function on \( \Omega \times A \times \mathcal{X} \), the loss function. A pure decision procedure is a function \( d \) on \( \mathcal{X} \) to \( A \). The associated risk \( R(\theta,d) \) for \( \theta \in \Omega \) is defined by

\[
(2.1) \quad R(\theta,d) = E_\theta L(\theta,d(X),X)
\]

where \( E_\theta \) denotes the expectation when \( X \) is distributed according to \( P_\theta \). In order that (2.1) should be meaningful we introduce a \( \sigma \)-algebra \( \mathcal{Q} \) of subsets of \( A \) and require that \( L \) be \( \mathcal{Q} \mathcal{B} \) measurable in its last two
arguments and that $d$ be $(\mathcal{B}, \mathcal{A})$ measurable. More precisely

(1) for each $\theta \in \Theta$ and real $c$,

$$\{(a, x) : L(\theta, a, x) \leq c\} \in \mathcal{A}\mathcal{B}$$

where $\mathcal{A}\mathcal{B}$ is the smallest $\sigma$-algebra of subsets of $A \times X$ containing all Cartesian products $A_1 \times B_1$ with $A_1 \in \mathcal{A}$ and $B_1 \in \mathcal{B}$.

(ii) $\{x : \hat{d}(x) \in A_1\} \in \mathcal{B}$ for all $A_1 \in \mathcal{A}$.

We also require that $L$ be bounded from below. Then $R(\theta, d)$ given by (2.1) is always meaningful as a real number or $+\infty$.

Nor let $g$ be a $1-1$ function on $X$ onto $X$, $(\mathcal{B}, \mathcal{B})$ measurable in both directions and $\hat{g}$ a $1-1$ function on $A$ onto $A$, $(\mathcal{A}, \mathcal{A})$ measurable in both directions. We shall say that the statistical decision problem we have formulated is invariant under $(g, \hat{g})$ if the following conditions hold:

a.) there is a function $\tilde{g}$ on $\Theta$ to $\Theta$ such that for each $\theta$, if $X$ is distributed according to $P_{g\theta}$, then $gX$ is distributed according to $P_{\tilde{g}\theta}$. This can also be expressed by saying that for all $B \in \mathcal{B}$,

(2.2) $P_{g\theta}(B) = P_{\theta}(g^{-1}B)$.

If $P$ is $1-1$, which we shall always assume, it is easy to see that $g$ is $1-1$ onto, and is uniquely determined.

b.) $L(\tilde{g}\theta, \tilde{g}a, gx) = L(\theta, a, x)$

for all $\theta \in \Theta$, $a \in A$ and $x \in X$.

Under these circumstances, the pure decision function $d$ will be said to be invariant under $(g, \hat{g})$ if for all $x$,

(3.3) $\hat{d}(gx) = \hat{g}(dx)$,
i.e.

\[(3.4) \quad \hat{g}^{-1} \circ d \circ g = \hat{g}^{-1}\]

This says that we get the same result whether we use \(d\) directly or first transform the sample point (by \(g\)), then use \(d\) and transform the result back (by \(g^{-1}\)). Roughly speaking, two people using essentially the same decision function \(d\), but different coordinate systems, will get the same result. However, in this formulation it must be understood that the solution is expressed in terms of the numerical coordinates alone without direct reference to the coordinate system used.

Of course there is nothing to prevent the statistician from mixing two decision functions, say by tossing a coin and using the decision function \(d_1\) if it comes up heads but \(d_2\) if it comes up tails. In multivariate analysis, this notion of mixed decision procedures is usually superfluous in principle, since it is ordinarily possible to construct from the sample a random variable which is uniformly distributed on \([0,1]\) independent of a sufficient statistic for all parameter points. Nevertheless, it is convenient to allow randomized decision because this permits us to restrict our attention to procedures based on a minimal sufficient statistic. Also the space of randomized decision functions has a natural structure as a convex set, and thus we can average randomized decision functions in a simple and unambiguous way. We shall use randomization after the sample point has been observed. The equivalence of this with randomization before the experiment is discussed, for example, in [12].

We define a randomized decision function to be a function \(\delta\) on \(X\)
to the set of probability measures on $\mathcal{A}$, measurable in the sense that for every $C \in \mathcal{A}$ and real $r$, \( \{ x : (\mathbb{B} x)A \leq r \} \in \mathcal{B} \). The interpretation is that after observing the sample point $X$, we distribute our action in accordance with the probability measure $\mathbb{B}X$. The randomized decision function $\mathbb{B}$ is said to be invariant under the transformation $(g, \mathbb{B})$ considered earlier if

\[
(2.5) \quad [\mathbb{B}(gx)] (\mathbb{B}C) = (\mathbb{B}x)C
\]

for all $x \in \mathcal{X}$ and $C \in \mathcal{A}$. It is clear that pure decision functions may be considered a special case of randomized decision functions (at least if $\mathcal{A}$ contains all subsets of $A$ consisting of a single point) and that the two notions of invariance coincide in this case.
3. Examples of the non-existence of an invariant minimax procedure for problems in testing hypotheses invariant under the full linear group. The examples given in this section will have the following properties:

(i) The problem is invariant under the multiplicative group \( \mathcal{G} \) of non-singular real \( p \times p \) matrices with \( p \geq 2 \).

(ii) There exists a minimax procedure.

(iii) There does not exist a minimax procedure which is invariant under the group \( \mathcal{G} \).

First consider the case where \( X_1, \ldots, X_m, Y_1, \ldots, Y_n \) are independently normally distributed random \( p \)-dimensional vectors with mean \( 0 \) and covariance matrices

\[
\mathbb{E} X_i X_i^t = \Gamma, \quad \mathbb{E} Y_j Y_j^t = k\Gamma
\]

where \( \Gamma \) and \( k \) are unknown, \( \Gamma \) being a \( p \times p \) positive definite covariance matrix and \( k \) a positive real number. Suppose we want to test

\[ H_0: k = 1 \text{ against } H_1: k = k_1 > 1, \]

with a loss of \( l \) for an incorrect decision. The problem is invariant under the transformations \( X_i \to \alpha X_i, Y_j \to \alpha Y_j \) with \( \alpha \) a non-singular \( p \times p \) matrix. If \( m + n \leq p \), this group is transitive on a set having probability 1 under each hypothesis (the set of \( X_1, \ldots, X_m, Y_1, \ldots, Y_n \) for which rank \( (X_1, \ldots, X_m, Y_1, \ldots, Y_n) = m + n \)). Thus any invariant test must essentially ignore the observations and have power equal to its size. However, if \( X_{11} \) is the first coordinate of \( X_1 \) and \( Y_{1j} \) the first coordinate of \( Y_j \),
then \[ \frac{m \sum Y_{1j}^2}{n \sum X_{1i}^2} \] is distributed as \( k \text{ } F_{n,m} \) so that the test which consists of rejecting \( H_0 \) when \[ \frac{m \sum Y_{1j}^2}{n \sum X_{1i}^2} > c, \] is similar under each hypothesis and has power greater than its size. This shows that no invariant test can be minimax. In a later section we shall show that, for arbitrary \( m \) and \( n \) the best invariant test is not minimax. We shall obtain a minimax test and see that, for fixed \( p \), as \( m,n \to \infty \) the best invariant test becomes more satisfactory.

Next we shall look at a more artificial case which brings out clearly the principle on which nearly all the examples I know are based. Let \( X \) be a random point of the multiplicative group \( \mathcal{G} \) of \( p \times p \) non-singular matrices with \( p \geq 2 \). The reader may find it simpler to consider only the case \( p = 2 \). Under \( H_q \) (for \( q = 0,1 \)), let \( X \) be distributed as \( g H k^{(q)} \) where \( g \) is an unknown element of \( \mathcal{G} \), \( H \) a random orthogonal matrix uniformly distributed over the orthogonal group, and \( k^{(1)}, k^{(2)} \) are given \( p \times p \) triangular matrices (i.e. \( k^{(q)}_{jj} = 0 \) for \( j > i \)). The orthogonal group \( \mathcal{H} \) is of course the multiplicative group of all \( p \times p \) matrices \( h \) for which \( hh' = I \), and the statement that \( H \) is uniformly distributed over \( \mathcal{H} \) means that for any \( h \in \mathcal{H} \) the distribution of \( HH \) is the same as that of \( H \) (and implies that \( HH \) and \( H^{-1} \) also have the same distribution as \( H \)). The problem is clearly invariant under the transformation \( X \to g_1 X \) for any \( g_1 \in \mathcal{G} \). We shall need the well-known

**Lemma:** For any \( g \in \mathcal{G} \) there exist \( h_1, h_2 \in \mathcal{H} \) and \( k_1, k_2 \in \mathcal{H} \) (the group of triangular matrices) such that \( g = h_1 k_1 = k_2 h_2 \).
An analogous factorization for an arbitrary connected simple Lie group (see Iwasawa [11], pp. 525-530) seems likely to be useful in studying the analogous problem for these groups.

Now consider tests of the form: Reject $H_0$ if $X \in S$ where $S$ is a measurable subset of $f$ invariant under transformation on the left by $\mathcal{K}$, i.e. $kS = S$ for all $k \in \mathcal{K}$. It is clear that $P\{gHk(1) \in S\}$ is independent of $g$. For by the above lemma there exist $h \in \mathcal{K}$ and $k \in \mathcal{K}$ such that $g = kh$. Then

$$P\{gHk(1) \in S\} = P\{khHk(1) \in S\} = P\{kHk(1) \in k^{-1}S\} = P\{Hk(1) \in k^{-1}S\}.$$

Thus any such test is similar under $H_0$ and $H_1$. Since any test invariant under $f$ must ignore the observation $X$ and so have power equal to its size, if we can prove that there exist $S$ (invariant under $H$), $k^{(1)}$ and $k^{(2)}$ such that $P\{Hk^{(1)} \in S\} \neq P\{Hk^{(2)} \in S\}$ we shall have shown that no invariant test can be minimax. We shall prove the existence of such objects only in the case $p = 2$, but it is not difficult to extend the argument to the general case.

A maximal invariant for $f$ under multiplication on the left by $\mathcal{K}$ is the ratio $f(g) = e_{11}/e_{12}$ taking on real values or $\infty$. It is easy to see that $f(Hk)$ has a Cauchy distribution with location parameter $k_{21}/k_{22}$ and scale parameter $k_{11}/k_{22}$. We can write $H$ as
\[ H = \begin{pmatrix} \cos T & \sin T \\ -\sin T & \cos T \end{pmatrix} \]

where \( T \) is a real random variable uniformly distributed modulo \( 2\pi \). Then

\[
H_k = \begin{pmatrix} \cos T & \sin T \\ -\sin T & \cos T \end{pmatrix} \begin{pmatrix} k_{11} & 0 \\ k_{21} & k_{22} \end{pmatrix}
\]

\[
= \begin{pmatrix} k_{11} \cos T + k_{21} \sin T, k_{22} \sin T \\ -k_{11} \sin T + k_{21} \cos T, k_{22} \cos T \end{pmatrix}
\]

so that

\[ f(H_k) = \frac{k_{21}}{k_{22}} + \frac{k_{11}}{k_{22}} \cot T. \]

Since \( \cot T \) has a Cauchy distribution with location parameter \( 0 \) and scale parameter \( 1 \), the assertion is proved. Thus we can take for example

\[
k^{(1)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad k^{(2)} = \begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix}
\]

with \( c > 0 \)

and \( S = \{ g : g_{11} / g_{12} > c/2 \} \). By taking \( c \) large we can make the size arbitrarily close to 0 and the power arbitrarily close to 1.

Still another example can be given by showing first that the group \( \mathcal{G} \) of all homeomorphisms of a circle (circumference) \( C \) onto itself does not have the property that for any testing problem invariant under \( \mathcal{G} \) there exists an invariant minimax test. This is essentially the sample described in the abstract [9]. Choose two probability measures \( P_0 \) and \( P_1 \) in \( C^3 \) in the following way. Choose in \( C \), \( n(\geq 3) \) disjoint open intervals. Let both \( P_0 \) and \( P_1 \) be such that if \( (X_1, X_2, X_3) \) is distributed according to \( P_1 \)
then $X_1, X_2, X_3$ lie in the same one of the $n$ selected open intervals, the probability that this is any given one being $1/n$. Under $P_0$ we require that $X_2$ is always between $X_1$ and $X_3$ when considered as points of one of the selected intervals, under $P_1$ that $X_2$ is never between $X_1$ and $X_3$. Now suppose we observe $(Y_1, Y_2, Y_3) = (gX_1, gX_2, gX_3)$ where $g$ is a fixed unknown homeomorphism of $C$ onto itself, and $(Y_1, Y_2, Y_3)$ is distributed according to either $P_0$ or $P_1$. If $(Y_1, Y_2, Y_3)$ lie in some half circle, we shall decide that the distribution was $P_0$ if $Y_2$ is between $Y_1$ and $Y_3$ when considered as points of this half circle, otherwise that the distribution was $P_1$. It is clear that the probability of an incorrect decision is $\leq 1/n$. For $g$ can transform at most one of the $n$ selected intervals into more than a half circle, and if the $(X_1, X_2, X_3)$ are not in this interval, our procedure gives a correct decision. Of course, $g$ is transitive on the subset of $C^3$ where no two coordinates are equal so that any invariant decision procedure must have a maximum probability of error greater than or equal to $\frac{1}{2}$.

Now consider the group of projective transformations on the real projective line $P$. These transformations are homeomorphisms of $P$, which is homeomorphic to the circle. Also the projective group is transitive on the subset of $P^3$ consisting of those points no two of whose coordinates are equal. Thus the counter example given above applies also to the group of projective transformations on the real line.

For the group of sense preserving homeomorphisms of the circle an example can be given in which the group operates on $C^2$. Proceed as before but let $P_0$ be a probability measure in $C^2$ for which $(X_1, X_2)$ always
fall in the same interval with $X_1$ preceding $X_2$ and $P_1$ a distribution under which $X_1$ always follows $X_2$. As before, this yields a counter example for the group of sense preserving projective transformations of the projective real line.

The corresponding problem for the group of all homeomorphisms (or all increasing homeomorphisms) of the real line seems to be unsolved. There are other groups which are not locally compact for which the problem of the existence of an invariant minimax test (or estimate) is interesting.
4. Estimation of a covariance matrix. Suppose \( X_1, \ldots, X_n \) (with \( n \geq p \)) are random \( p \)-dimensional vectors, independently normally distributed with mean 0 and unknown non-singular covariance matrix \( \Gamma \). Then

\[
S = \sum_{i=1}^{n} X_i X_i' \]

is a sufficient statistic and the usual estimate of \( \Gamma \) is \( \Gamma = S/n \). This estimate has many desirable properties.

(i) It is the maximum likelihood estimate.

(ii) It is (to within a constant multiple) the only non-randomized invariant estimate depending only on \( S \). More precisely, if \( \Gamma = \varphi(S) \) is a non-randomized estimate of \( \Gamma \) for which \( \varphi(\alpha S \alpha') = \varphi(S) \alpha' \) for all non-singular \( p \times p \) matrices \( \alpha \), then, for some real constant \( \varphi(S) = \lambda S \).

(iii) If \( \xi \) is any \( p \)-dimensional vector, the estimate \( \xi' S \xi/n \) of the variance \( \xi' \Gamma \xi \) of \( \xi' X_i \) is in many ways a reasonable one. In particular as \( n \to \infty \) it is a consistent estimate (in that the ratio of the estimate to the true value approaches 1 in probability uniformly with respect to \( \xi, \Gamma, \) and \( p \)).

However, the first two properties are not really very strong arguments in favor of \( \hat{\Gamma} \). The principal general optimum properties of the maximum likelihood estimates are asymptotic properties, in the present case, asserting that for fixed \( p \), as \( n \to \infty \) the maximum likelihood estimate becomes good in various ways. It would require a careful study to tell whether, for moderate \( p \) (say 10) the required \( n \) would not be much larger than is ordinarily available. Of course, the second property is hardly more convincing since there are many examples, in the present paper and elsewhere,
which show that the class of invariant procedures does not always contain a good procedure.

The third property, although less pretentious is somewhat more convincing, but there are other ways in which we may want to use the estimate $\hat{\Gamma}$. For example, we may want to estimate $\det \Gamma$, or $\xi' \Gamma^{-1} \xi$ for some $p$-dimensional vector $\xi$. Let us first look at the second problem.

$\xi'(S/n)^{-1} \xi$ is distributed as $\xi' \Gamma^{-1} \xi$ times $n$ times the reciprocal of a $\chi^2$ with $n-p+1$ degrees of freedom. Thus if $p/n$ is not small, $\xi'(S/n)^{-1} \xi$ is a good estimate of $(1 + \frac{p-1}{n-p+1}) \xi' \Gamma^{-1} \xi$, but not of $\xi' \Gamma^{-1} \xi$. Of course we can repair the estimate, by using $\frac{S}{n-p+1}$ rather than $S/n$ but this destroys property (iii) mentioned above. Also $\det S/n$ is a substantial underestimate of $\det \Gamma$. The bias in $\log \det S/n$ as an estimate of $\log \det \Gamma$ only approaches 0 as $\frac{n}{p^2} \to \infty$ and the ratio of the bias to the standard deviation only approaches 0 as $\frac{n}{p^3} \to \infty$. These results are obtained easily from the fact that $\det S$ is distributed as $(\det \Gamma)^{\frac{p}{p-1}} \chi^2_{n-p+1}$ with the $\chi^2$'s independent. These considerations suggest the following apparently unsolved problem: Does there exist a double sequence $\{\varphi_{p,n}\}$ of functions, $\varphi_{p,n}$ being a function on the space $\mathcal{S}$ of positive definite covariance matrices to itself such that for every $c > 1$ and $\epsilon > 0$ there exists $n_0$ such that for all $p$ and $n \geq \max(n_0, cp)$

\[
\text{P} \left\{ \left| \frac{\det \varphi_{p,n}(S)}{\det \Gamma} - 1 \right| < \xi \right\} > 1 - \xi . \tag{4.1}
\]

\[
\text{For all } p\text{-dimensional vectors } \xi, \tag{4.2}
\]
\[ p \left\{ \left| \frac{\mathbf{x}^{\prime} \Phi p, n (S) \mathbf{x}}{\mathbf{x}^{\prime} \Gamma \mathbf{x}} - 1 \right| < \delta \right\} > 1 - \delta. \]

(4.3) For all \( p \)-dimensional vectors \( \mathbf{x} \),

\[ p \left\{ \left| \frac{\mathbf{x}^{\prime} \Phi^{-1} p, n (S) \mathbf{x}}{\mathbf{x}^{\prime} \Gamma^{-1} \mathbf{x}} - 1 \right| < \delta \right\} > 1 - \delta. \]

Of course various modifications of this problem suggest themselves immediately.

Next we shall consider the problem of minimax estimation of \( \Gamma \) with the loss function

(4.4) \[ L(\Gamma, \hat{\Gamma}) = \text{tr} \Gamma^{-1} \hat{\Gamma} - \log \det \Gamma^{-1} \hat{\Gamma} - p. \]

This function is somewhat arbitrary but it is convex in \( \hat{\Gamma} \) and invariant under linear transformations in the sense that

(4.5) \[ L(\alpha \Gamma \alpha', \alpha \hat{\Gamma} \alpha') = L(\Gamma, \hat{\Gamma}). \]

Also \( L(\Gamma, \hat{\Gamma}) \geq 0 \) with equality if and only if \( \Gamma = \hat{\Gamma} \). It would be desirable to study the problem of minimax estimation with an arbitrary loss function satisfying these conditions. Because of the convexity of \( L(\Gamma, \hat{\Gamma}) \) in \( \hat{\Gamma} \) we need look only for a minimax estimate among those of the form \( \hat{\Gamma} = \varphi (S) \), i.e. non-randomized estimates based on the sufficient statistic \( S \). As we have already indicated, the problem is invariant under transformations of the form

\( S \rightarrow \alpha S \alpha' \), \( \Gamma \rightarrow \alpha \Gamma \alpha', \hat{\Gamma} \rightarrow \alpha \hat{\Gamma} \alpha' \).
with $\alpha$ non-singular, in particular for non-singular triangular $\alpha$, i.e. those with $\alpha_{ij} = 0$ for $j > i$. The group $G$ of triangular $\alpha$ is solvable, so that by a theorem of Kudo [4], p. 47, there exists a minimax estimate invariant under $G$ i.e. a minimax $\hat{\varphi} = \varphi(S)$ with

\[ \varphi(\alpha S \alpha') = \alpha \varphi(S) \alpha' \]

for all $\alpha \in G$. Condition (4.6) implies that there exists a positive diagonal matrix $\Delta$ (i.e. $\Delta_{ij} = 0$ for $i \neq j$ and $\Delta_{ii} > 0$) such that

\[ \varphi(S) = k \Delta k' \]

where

\[ S = kk' \text{ with } k \text{ triangular.} \]

For such $\varphi$, $R_\Gamma(\varphi) = \xi L(\Gamma, \varphi(S))$ will be independent of $\Gamma$. We find

\[ L(\Gamma, \hat{\varphi}) = \text{tr} k \Delta k' - \log \det k \Delta k' - p. \]

\[ R_\Gamma(\varphi) = \xi \left\{ \text{tr} k \Delta k' - \log \det \Delta - \log \det kk' - p \right\} \]

\[ = \xi \sum_{i=1}^{p} \left( \Delta_{ii} \chi_{n+p-2i+1}^2 - \log \Delta_{ii} - \log \chi_{n-1+i}^2 \right) - p \]

\[ = \sum_{i=1}^{p} \left( (n+p-2i+1) \Delta_{ii} - \log \Delta_{ii} - \xi \log \chi_{n-1+i}^2 \right) - p \]

since the $i^{th}$ diagonal element of $k'k$ is distributed as $\chi_{n+p-2i+1}^2$. This is minimized by

\[ \Delta_{ii} = \frac{1}{n+p-2i+1}. \]

Thus a minimax estimate is given by (4.7) with $\Delta_{ij} = 0$ for $i \neq j$ and $\Delta_{ii}$ given by (4.11). Since the class (4.7) includes the usual estimate (with $\Delta = \frac{1}{n} I$) and the minimax estimate we have obtained is clearly unique.
in the class \((4.7)\), the usual estimate is not minimax, and so no estimate invariant under the full linear group is minimax.

The minimax estimate we have obtained has several interesting and disturbing features.

(i) It is not unique for \( p \geq 2 \). For any non-singular matrix \( \alpha \), the estimate \( \alpha^{-1} \phi(\alpha \alpha') \alpha^{-1'} \) is also minimax and has the same constant risk. If \( \alpha \) is not triangular this estimate is different from \( \phi \), in general.

(ii) It is not admissible for \( p \geq 2 \). Because of the strict convexity of the loss function and the fact that the minimax estimate \( \phi \) has constant risk, the estimate

\[
(4.12) \quad \hat{\Gamma} = \frac{1}{2} (\phi(S) + \alpha^{-1} \phi(\alpha \alpha') \alpha^{-1'})
\]

for non-triangular \( \alpha \) will have everywhere smaller risk.

(iii) It is unreasonable, in that it does not have property (iii) of the usual estimate mentioned above. In particular the estimate \( \frac{S_{11}}{n + p - 1} \) of the variance of the first coordinate is ridiculously small if \( p/n \) is not small. In principle one might try to reduce the bias by using instead of \( \phi \), the estimate

\[
(4.13) \quad \hat{\Gamma} = \frac{1}{p} \sum_{\alpha} \alpha^{-1} \phi(\alpha \alpha') \alpha^{-1'}
\]

where \( \alpha \) ranges over all matrices having only 0's and 1's with exactly one 1 in each row and each column. However, this is impractical if \( p \) is at all large, and the bias is not completely eliminated anyway. Probably we must conclude that the loss function we have chosen is not realistic despite its superficial plausibility.
(iv) The minimax risk is surprisingly large. For $\varphi$ given by (4.7), (4.8), and (4.11), we have by 4.10

\begin{equation}
R_\varphi(\varphi) = \sum_{i=1}^{p} \left\{ \log (n+p-2i+1) - \xi \log \chi_{n-i+1}^2 \right\}
\end{equation}

But

\begin{equation}
\xi \log \chi_{n-i+1}^2 = \log (n-i+1+O(\sqrt{n-i+1}))
\end{equation}

\begin{equation*}
= \log (n-i+1) + O\left(\frac{1}{\sqrt{n-i+1}}\right)
\end{equation*}

so that

\begin{equation}
R_\varphi(\varphi) = \sum_{i=1}^{p} \log \frac{n+p-2i+1}{n-i+1} + O\left(\frac{p}{\sqrt{n-p}}\right).
\end{equation}

If $n$ is large and $p/n$ is not small this is of the exact order of $p$. But we can write

\begin{equation}
L(\Gamma, \hat{\Gamma}) = \sum_{i=1}^{p} \left( \lambda_i - \log \lambda_i - 1 \right)
\end{equation}

where the $\lambda_i$ are the roots of the equation

\begin{equation}
\det(\hat{\Gamma} - \lambda \Gamma) = 0.
\end{equation}

Thus, for any $c_1 > 0$ there exists $c_2$ such that for all sufficiently large $n$ and $p > c_1 n$,

\begin{equation}
\sum_{i=1}^{p} \xi (\lambda_i - \log \lambda_i - 1) > c_2 p.
\end{equation}

This suggests (but does not quite prove) that with high probability, under the circumstances indicated some of the roots $\lambda_i$ of (4.18) will differ
appreciably from 1. Thus we expect that for any estimator \( \hat{\Gamma} \) there exists \( \varepsilon \) such that either

\[
\sup_{\varepsilon} \frac{\hat{\varepsilon} \hat{\Gamma} \hat{\varepsilon}}{\hat{\varepsilon} \hat{\Gamma} \hat{\varepsilon}} \quad \text{or} \quad \inf_{\varepsilon} \frac{\hat{\varepsilon} \hat{\Gamma} \hat{\varepsilon}}{\hat{\varepsilon} \hat{\Gamma} \hat{\varepsilon}}
\]

will differ appreciably from 1 with high probability, where \( \varepsilon \) ranges over the non-zero \( p \)-dimensional vectors. This is in contrast to the fact that for fixed \( \varepsilon \), \( \frac{\hat{\varepsilon} \hat{\varepsilon}^{\top} (\hat{\varepsilon}/n)^{\varepsilon}}{\hat{\varepsilon} \hat{\varepsilon} \hat{\varepsilon}} \) is close to 1 with high probability as long as \( n \) is large, regardless of the value of \( p \).
5. Confidence sets for the mean of a multivariate normal distribution.

Let \( X_1, \ldots, X_n \) with \( n \geq p + 1 \) be independently normally distributed random \( p \)-dimensional vectors with unknown mean \( \xi \) and unknown covariance matrix \( \Gamma \). Suppose we are interested in finding a function \( f \) on \( \mathcal{N}^n \) (where \( \mathcal{N} \) is the space of \( p \)-dimensional vectors) to the set of subsets of \( \mathcal{N} \) such that

\[
\mathbb{P}_{\xi, \Gamma} \left\{ f(X_1, \ldots, X_n) \geq 1 - \xi \right\}
\]

for all \( \xi, \Gamma \) and in some sense \( f(X_1, \ldots, X_n) \) is as small as possible on the average, subject to (5.1). There is one \( f \) which obviously satisfies (5.1) (with equality rather than inequality) and is commonly used. This is

\[
f_0(x_1, \ldots, x_n) = \left\{ \xi : n(\bar{x} - \xi)'S^{-1}(\bar{x} - \xi) \leq c \right\}
\]

where

\[
\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i
\]

\[
s = \sum_{i=1}^{n} x_i'x_i' - n\bar{x}\bar{x}'
\]

and \( c \) is chosen so that

\[
\frac{\chi^2_p}{\chi^2_{n-p}} \leq c \Rightarrow \mathbb{P} \left\{ \frac{\chi^2_p}{\chi^2_{n-p}} \leq c \right\} = 1 - \xi.
\]

That (5.1) holds for \( f = f_0 \) is equivalent to the fact, proved by Hotelling, that \( n(\bar{x} - \xi)'S^{-1}(\bar{x} - \xi) \), with \( \bar{x}, S \) defined analogously to (5.4), is distributed as \( \chi^2_p/\chi^2_{n-p} \).
We shall search for an \( f_1 \) with the property that (5.1) holds for

\[
\mathcal{R}, \mathcal{F} = f_1 \quad \text{and}
\]

\[
\text{(5.6)} \quad \sup_{\xi, \Gamma} \frac{m[f_1(x_1, \ldots, x_n)]}{(\det \Gamma)^{\frac{1}{2}}} \leq \frac{1}{\xi, \Gamma}
\]

is as small as possible where \( m \) is Lebesgue measure. This is a somewhat arbitrary interpretation of our earlier statement that \( f \) should be as small as possible on the average. Neyman's approach, which is to try to keep the probability of covering incorrect values of the mean small may be preferable but there are some difficulties. One form of this approach is equivalent to the question whether Hotelling's \( \mathcal{T}^2 \) test is minimax for testing \( H_0: \xi = 0 \) against \( H_1: \xi', \Gamma^{-1} \xi = \text{const.} \), and this problem is still unsolved. Also in problems as complicated as the present one it is not clear that this approach is more natural than ours.

In order to find an \( f_1 \) minimizing (5.6) subject to (5.1), we first search for a function \( \varphi \) of \( n + 1 \) arguments, each a \( p \)-dimensional vector, taking on values in \([0,1]\) such that

\[
\text{(5.7)} \quad \xi, \Gamma \varphi(x_1, \ldots, x_n, \xi) \geq 1 - \varepsilon \quad \text{for all} \quad \xi, \Gamma
\]

and, subject to this condition

\[
\text{(5.8)} \quad \sup_{\xi, \Gamma} \frac{1}{\xi, \Gamma} \frac{1}{(\det \Gamma)^{\frac{1}{2}}} \int \varphi(x_1, \ldots, x_n, \xi) \phi \, d\xi
\]

is minimized. The \( \varphi \) we shall obtain will happen to take on only the values 0 and 1 so that if we define
we shall have a solution of the problem mentioned earlier.

Since $(\bar{X}, \bar{S})$ is sufficient and the problem is linear in $\varphi$ so that nothing can be gained by randomization there exists a minimizing $\varphi$ of the form

$$
\varphi(x_1, \ldots, x_n, \xi) = \psi(\bar{X}, \bar{S}, \xi).
$$

The problem is invariant under affine transformations

$$
X_i \rightarrow \alpha X_i + \beta, \quad \xi \rightarrow \alpha \xi + \beta, \quad \Gamma \rightarrow \alpha \Gamma \alpha',
$$

with $\alpha$ a non-singular $p \times p$ matrix and $\beta$ a $p$-dimensional column vector in the sense that if the $X_i$ are independently normally distributed with mean $\bar{x}$ and non-singular covariance matrix $\Gamma$ then the $\alpha X_i + \beta$ are independently normally distributed with mean $\alpha \bar{x} + \beta$ and non-singular covariance matrix $\alpha \Gamma \alpha'$. It follows that

$$
\inf \limits_{\xi, \Gamma} \epsilon_{\xi, \Gamma} \varphi(\alpha x_1 + \beta, \ldots, \alpha x_n + \beta, \alpha \xi + \beta)
= \inf \limits_{\xi, \Gamma} \epsilon_{\xi, \Gamma} \varphi(x_1, \ldots, x_n, \xi)
$$

and

$$
\sup \limits_{\xi, \Gamma} \frac{1}{(\det \Gamma)^{\frac{1}{2}}} \frac{1}{\epsilon_{\xi, \Gamma}} \int \varphi(\alpha x_1 + \beta, \ldots, \alpha x_n + \beta, \xi) d\xi
= \sup \limits_{\xi, \Gamma} \frac{1}{(\det \Gamma)^{\frac{1}{2}}} \frac{1}{\epsilon_{\xi, \Gamma}} \int \varphi(x_1, \ldots, x_n, \xi) d\xi.
$$

The decision function $\varphi$ is said to be invariant under the transformation (5.11) if
(5.14) \[ \varphi(\alpha x_1 + \beta, \ldots, \alpha x_n + \beta, \alpha \xi + \beta) = \varphi(x_1, \ldots, x_n, \xi). \]

For \( \varphi \) of the form (5.10), the condition (5.14) becomes

(5.15) \[ \psi(\alpha x + \beta, \alpha \alpha', \alpha \xi + \beta) = \psi(x, s, \xi). \]

The group of all transformations (5.11) with \( \alpha \) triangular (i.e. \( \alpha_{ij} = 0 \) for \( j > i \)) is solvable, so that by an extension of the arguments of Hunt and the author (see Lehmann [9]) and Kudo [4], there exists a minimizing \( \varphi \) of the form (5.10) satisfying (5.15) for all \( \beta \) and all triangular \( \alpha \). But this is equivalent to the condition that \( \psi \) be of the form

(5.16) \[ \psi(x, s, \xi) = \rho(t_1^2, \ldots, t_p^2) \]

where

(5.17) \[ t_i^2 = n(x - \xi)_i s^{-1}_i (x - \xi)_i \]

where \((x - \xi)_i\) is the i-dimensional vector whose coordinates are the first i coordinates of \(x - \xi\) and \(s_i\) is the upper left i \times i submatrix of s.

Since the group of transformations (5.18)

(5.18) \[(\xi, \gamma) \rightarrow (\alpha \xi + \beta, \alpha \gamma \alpha')\]

of the parameter space with \( \alpha \) triangular is transitive, the condition (5.7) and expression (5.8) become

(5.19) \[ \varepsilon_{0,1} \rho(t_1^2, \ldots, t_p^2) \geq \alpha \]
(5.20) \[ \xi_{0,I} \int \rho(T_{1}^2, \ldots, T_{P}^2) \, d\xi \]

with

(5.21) \[ T_{I}^2 = n(x - \xi)_{(1)} S_{(1)}^{-1} (x - \xi)_{(1)} \]

Thus the problem becomes that of minimizing the linear function (5.20) subject to the single linear condition (5.19), and this is easy.

We make the change of variable

(5.22) \[ \xi^* = \sqrt{n} K^{-1}(x - X) \]

where \( K \) is triangular and

(5.23) \[ S = KK' \]

Then the \( I \)th coordinate \( \xi_{I}^* \) of \( \xi \) satisfies

(5.24) \[ \xi_{I}^{*2} = T_{I}^2 - T_{I-1}^2 \]

and the problem is to minimize

(5.25) \[ \int \rho(\xi_{1}^{*2}, \xi_{2}^{*2} + \xi_{3}^{*2} + \ldots + \xi_{p}^{*2}) \, d\xi^* \]

subject to

(5.26) \[ \varepsilon \rho(\xi_{1}^{*2}, \ldots, \xi_{p}^{*2}) \geq 1 - \varepsilon . \]

If \( \pi \) is the joint density of \( \xi_{1}^{*}, \ldots, \xi_{p}^{*} \), with respect to Lebesgue measure the signs being assigned independently at random with probability \( \frac{1}{2} \), the
minimizing \( \rho \) is given by

\[
(5.27) \quad \rho(\xi_1^*, \ldots, \xi_p^*) = \begin{cases} 
1 & \text{if } \pi(\xi_1^*, \ldots, \xi_p^*) \geq c \\
0 & \text{if } \pi(\xi_1^*, \ldots, \xi_p^*) < c 
\end{cases}
\]

where \( c \) is chosen so that (5.26) is satisfied with equality.

We shall show that the density \( \pi \) is given by

\[
(5.28) \quad \pi(\xi_1^*, \ldots, \xi_p^*) = \frac{\Gamma \left( \frac{n}{2} \right)}{\pi^{p/2} \Gamma \left( \frac{n-p}{2} \right)} \frac{1}{\left( 1 + \sum_{i=1}^{p} \xi_i^* \right)^{n-p+1}} \left( 1 + \sum_{i=1}^{p} \xi_i^* \right)^{p-1} \left( 1 + \sum_{i=1}^{n_p} \xi_i^2 \right)^{1/2} 
\]

It is well known (see for example Rao [12], p. 74) that the conditional distribution of \( \xi_p^* \) given \( \xi_1^*, \ldots, \xi_{p-1}^* \) is that of

\[
\sqrt{\frac{p-1}{\sum_{i=1}^{p-1} \xi_i^2}} \frac{U}{\sqrt{2^{n-p}}} \quad \text{where } U \text{ is a unit normal random variable independent of } \chi^2_{n-p}
\]

of \( \chi^2_{n-p} \). Thus the conditional density of \( \xi_p^* \) given \( \xi_1^*, \ldots, \xi_{p-1}^* \) is

\[
(5.29) \quad \frac{\Gamma \left( \frac{n-p+1}{2} \right)}{\sqrt{\pi} \Gamma \left( \frac{n-p}{2} \right)} \frac{1}{\sqrt{1 + \sum_{i=1}^{p-1} \xi_i^2}} \left( 1 + \sum_{i=1}^{p} \xi_i^2 \right)^{p-1} \left( 1 + \sum_{i=1}^{n_p} \xi_i^2 \right)^{1/2} 
\]

which is the density of an appropriate multiple of Student's \( t \). The joint density \( \pi \) of \( \xi_1^*, \ldots, \xi_p^* \) is obtained by multiplying the expressions obtained from (5.29) by replacing \( p \) by \( 1, 2, \ldots, p \). Thus
(5.30) \( \pi(\xi_1^*, \ldots, \xi_p^*) \)

\[
= \frac{p}{\pi} \frac{\Gamma \left( \frac{n-q+1}{2} \right)}{\sqrt{\pi} \Gamma \left( \frac{n-q}{2} \right)} \left( \frac{1 + q-1 \sum_{k=1}^{\xi_k^*}}{1 + \sum_{k=1}^{\xi_k^*}} \right)^{\frac{n-q}{2}} \left( \frac{1 + \frac{q}{\xi_k^*}}{1 + \xi_k^*} \right)^{\frac{n-q+1}{2}}
\]

\[
= \frac{\Gamma \left( \frac{n}{2} \right)}{\pi^p \Gamma \left( \frac{n-p}{2} \right)} \left( \frac{1}{\prod_{k=1}^{p-1} \left( 1 + \sum_{k=1}^{\xi_k^*} \right)} \right)^{\frac{1}{2}} \left( \frac{1}{1 + \sum_{k=1}^{\xi_k^*} \frac{n-p+1}{2}} \right)
\]

Thus we conclude finally that a minimax solution is obtained by taking as the confidence set for the mean \( \xi \) the set of all \( \xi \) for which

(5.31) \( (1 + T_{p}^2)^{-\frac{n-p+1}{2}} \prod_{k=1}^{p-1} (1 + T_{k}^2) \leq c \)

with the \( T_{k}^2 \) given by (5.21) and \( c \) chosen so that the desired confidence coefficient \( 1 - \varepsilon \) is attained exactly. Unfortunately it does not seem to be easy to compute \( 1 - \varepsilon \) as a function of \( c \).
6. Testing for the value of a constant of proportionality between two covariance matrices. Here we shall give a more complete solution of the problem considered at the beginning of section 3. \( X_1, \ldots, X_m, Y_1, \ldots, Y_n \) are independently normally distributed random \( p \)-dimensional vectors with mean 0 and \( \mathcal{E} X_i X_i' = \Gamma, \mathcal{E} Y_j Y_j' = k \Gamma \) where \( \Gamma \) is an unknown covariance matrix and \( k \) an unknown positive constant. We want to test \( H_0: k=1 \) against \( H_1: k=K \). The problem is invariant under linear transformations \( X_i \rightarrow \alpha X_i, Y_j \rightarrow \alpha Y_j \) with \( \alpha \) a non-singular \( p \times p \) matrix, in particular for triangular \( \alpha \). We know that there exists a minimax test of given size which is invariant under the multiplicative group of all non-singular triangular matrices operating in the above manner. Since the induced group of transformations of the parameter space is transitive on \( H_0 \) and on \( H_1 \), such minimax tests will reject \( H_0 \) if and only if the probability ratio of the maximal invariant (that is the ratio of its density under \( H_1 \) to its density under \( H_0 \)) is greater than a selected constant.

To compute this probability ratio we first suppose \( m+n > p \). The joint density \( p \) of \( X_1, \ldots, X_m, Y_1, \ldots, Y_n \) with respect to Lebesgue measure in the \((m+n)p\) dimensional sample space is

\[
(6.1) \quad p_{\Gamma,k}(x_1, \ldots, x_m, y_1, \ldots, y_n) = \frac{1}{(2\pi)^{m+n} \det \Gamma_0^2 k^2} \exp \left[ -\frac{1}{2} \text{tr} \Gamma_0^{-1} \left( \Sigma x_1 x_1' + \frac{1}{k} \Sigma y_j y_j' \right) \right].
\]

Since the desired probability ratio is independent of \( \Gamma \) we shall need this only for \( \Gamma = I \). A volume element invariant under all linear transformations
\[\pi \frac{dx_i}{\prod_{j=1}^{m+n} \frac{dx_j}{dy_j}} \text{ where } x = (x_1 \ldots x_m), y = (y_1 \ldots y_n). \text{ The density} \]
\[\frac{1}{[\det(xx' + yy')]^{\frac{m+n}{2}}} \exp \left[-\frac{1}{2} \text{tr} \left(xx' + \frac{k}{2} yy'\right)\right].\]

Thus by a lemma to be proved in Part II, the ratio of the density of a maximal invariant under the group of triangular matrices for \(H_1\) to the density for \(H_0\) is
\[\pi_{k,1}(x,y) = \frac{1}{x_1(x,y)} \text{ with} \]
\[(6.3) \quad \pi_k(x,y) = \frac{1}{np} \frac{1}{k^2} \cdot \int \det [(gx)(gx)' + (gy)(gy)']^{\frac{m+n}{2}} e^{-\frac{1}{2} \text{tr}[(gx)(gx)' + \frac{1}{k} (gy)(gy)']} \, \mathbf{d}_\mu(g)\]

where \(\mu\) is a left invariant measure in the multiplicative group of triangular matrices. This measure is given by
\[(6.4) \quad d\mu(g) = \frac{\prod_{i=1}^m \frac{dg_{ii}}{dg_{ii}}}{\prod |g_{ii}|^\frac{1}{2i}}.\]
Thus

\[ (6.5) \quad \pi_k(x,y) = \frac{\det(x^1 + y^1)}{\kappa^2} \quad \frac{m+n}{n_0} \]

\[ \cdot \left( \int \left( \text{det} \ g_{ij} \right)^{m+n/2} e^{-\frac{1}{2} \text{tr} \ g(x^1 + \frac{1}{k} y^1) g_i^j \frac{1}{\Pi} \sum \frac{d_i^j}{d_i^j}} \right) \]

\[ = \frac{\det(x^1 + y^1)}{\kappa^2} \quad \frac{m+n}{n_0} \quad \int_{\mathbb{R}^{m+n-1}} e^{-\frac{1}{2} \text{tr} \ g(x^1 + \frac{1}{k} y^1) g_i^j \frac{1}{\Pi} \sum \frac{d_i^j}{d_i^j}} \]

Let us make the change of variable

\[ (6.6) \quad h = gr \]

where \( r \) is triangular and

\[ (6.7) \quad x^1 + \frac{1}{k} y^1 = r r^t. \]

Then (6.5) becomes

\[ (6.8) \quad \pi_k(x,y) = \frac{\det(x^1 + y^1)}{\kappa^2} \quad \frac{m+n}{n_0} \quad \Pi \left| r_{ii} \right|^{m+n-i} - (m+n-i) - (p-i+1) \]

\[ \cdot \left( \int \Pi \left| h_{ii} \right|^{m+n-1} e^{-\frac{1}{2} \text{tr} \ h h^t \Pi} \right) \]

It follows that the desired probability ratio is
\[
\frac{\pi_{k_l}(x,y)}{\pi_l(x,y)} = \frac{p}{n} \sum_{i=1}^{\frac{m+n+p-2i+1}{2}} \frac{r^{(o)}}{r^{(l)}}
\]

where \( r^{(o)} \) and \( r^{(l)} \) are triangular and

\[
x x' + y y' = r^{(o)} r^{(o)'},
\]

\[
x x' + \frac{1}{k} y y' = r^{(l)} r^{(l)'}
\]
REFERENCES


11 A. Wald, "Tests of statistical hypotheses concerning several parameters when the number of observations is large," *Trans. Amer. Math. Soc.*, Vol. 54 (1943), pp. 426-482.