INTERSECTION REGION CONFIDENCE PROCEDURES WITH AN APPLICATION TO THE LOCATION OF THE MAXIMUM IN QUADRATIC REGRESSION

By David L. Wallace

University of Chicago

1. Summary. Confidence region procedures for multidimensional quantities sometimes require prohibitive amounts of computation and the regions are difficult to represent in a useful way. Some approximate procedures are constructed by using regions obtained as the intersection of several regions, each much easier to construct. The procedures are applicable to the solution of simultaneous equations, whose coefficients are subject to random error. Approximations by convex polyhedra and by parallelepipeds are proposed. The procedures are illustrated for setting a confidence region for the location of the vertex of a quadratic regression surface.

2. Confidence regions. In this section, I give a subjective evaluation of the requirements for a useful confidence region procedure.

Suppose that λ is a (multidimensional) quantity defined as a function of the parameters of the distribution sampled. The problem of constructing confidence regions for the true value(s) of λ will be considered.

A confidence region and a point estimate for λ are often used to summarize the information about λ in the observed sample. Their use is an attempt to convey in a comprehensible way some idea of the extent and character of the determination of λ, taking account of the inaccuracies of measurement. Any use of the confidence region in making decisions about further experimentation, process operations, etc., will be informal. The exact confidence level is not important and even the frequency interpretation of the procedure is not essential, both serving principally as "benchmarks" for purposes of comparison and familiarity. What is important is that the region be represented geometrically or analytically so that the user can comprehend its size, shape and location. Approximations to the region which simplify this representation will be valuable as long as they do not greatly change the confidence level.

The theoretical specification of confidence procedures and the investigation of their statistical properties (level, power, etc.) are usually accomplished through an associated family of tests of hypotheses. The condition that the true quantity have the particular value λ is a condition on the parameters and hence a statistical hypothesis. Denote it by H_λ. Then, given a level α test of H_λ for each value of λ, the confidence procedure defined by $R = \{\lambda: H_\lambda \text{ not rejected}\}$ is an error level α confidence procedure. The "error level" ($= 1 -$ "confidence level") of a confidence procedure is usually more convenient than the confidence level

Received November 5, 1956.

and coincides with usage in the related testing procedures. (Strictly, there should be a notational distinction between a confidence procedure R as a set-valued random variable and a confidence region R as the realization for a particular sample. However, the meaning of the symbol R should be clear from the context or the verbal distinction between confidence procedure and confidence region.)

This method does not necessarily give usable confidence regions, even when the test of each H_λ separately is satisfactory. In order that the tests need not actually be carried out for each λ, some continuity in λ must be required of the family of tests. Any (non-randomized) test of H_λ can be represented (not uniquely) by a statistic $h(\lambda)$ where, for any sample z, H_λ is rejected if and only if $h(\lambda, z) > 0$. If there is a choice of h which is, for a fixed sample z, continuous in λ, then the confidence region R is a closed set with boundary equation $h(\lambda, z) = 0$.

However, continuity of $h(\lambda)$ is not generally enough. If λ is one dimensional; a useful confidence region is usually an interval. A solution providing the limits of the interval is satisfactory, but one providing only a complex equation $h(\lambda) = 0$ for the limits may not be.

When λ is multidimensional, the problems of computation and representation are greatly magnified. The boundary equation will likely increase in complexity rapidly with increasing dimension. But more serious is the difficulty of representing the region even when $h(\lambda)$ is given explicitly in terms of simple functions. The boundary can be plotted in two dimensions, as can cross sections in more than two dimensions, though with effectiveness decreasing with increasing dimension. A principal difficulty is that few shapes are readily visualized in more than two dimensions, or, what is more essential, that comprehension of a region from the equation of its boundary is restricted to very simple surfaces.

The simplest regions are the parallelepipeds which can be completely described by giving limits on each coordinate of a coordinate system related by an affine transformation to the original coordinate system, or equivalently, by giving p linear double inequalities on the coordinates of λ.

The next simplest regions would seem to be the convex polyhedra. When the number of faces is small, the region is simply described by giving the linear inequalities corresponding to each face and is only slightly more complex than the parallelepiped. As an approximate representation of a region with corners, the number of faces is likely too large to permit use of the inequalities and the region must be thought of, with greater difficulty and less adequately, in terms of the corners (vertices).

Ellipsoidal confidence regions are important, largely because they occur naturally in the classical normal theory of means and regression coefficients and also in the general large sample confidence theory. They are probably visualized as rounded boxes and their description by a center and lengths and directions of principal axes corresponds closely to the parallelepiped description.

3. Geometrical idea of intersection confidence regions. In many multidimensional confidence problems, interest centers more on the separate coordinates

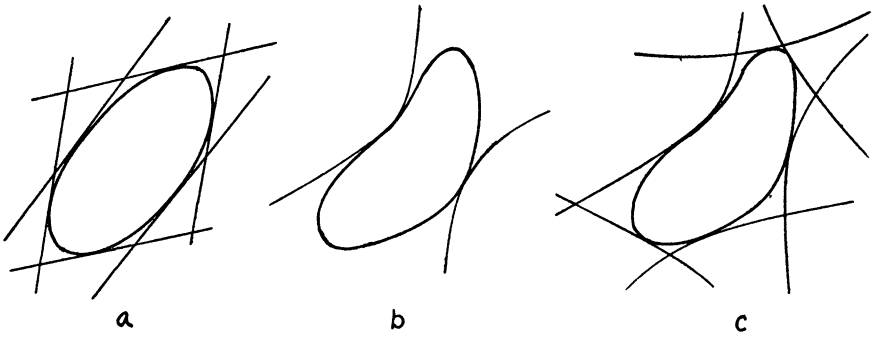

FIG. 1. (a) An ellipse as intersection of straight strips. (b) A standard region and a curved strip. (c) A standard region as intersection of curved strips.

or on linear combinations of them than on the multidimensional quantity (e.g., the means in an analysis of variance). The usual ellipsoidal confidence region is of little value. Scheffé's [12] multiple comparisons procedure amounts to representing the ellipsoid as the intersection of all the slabs between parallel pairs of tangent hyperplanes (see figure 1a). Each slab gives a confidence interval for a single linear combination of coordinates. The totality of such intervals has the same joint error level as does the ellipsoidal region. The procedure permits making as many confidence statements on linear combinations as desired and permits the posterior selection of "most interesting" statements.

The same representation by slabs is valid and useful for any convex region and this is the basis for the multiple comparisons procedure given by Tukey [13], Roy and Bose [11] and others.

Even in problems in which the multidimensional quantity is of principal interest, the multiple comparison methods provide a means for approximating convex confidence regions by convex polyhedra, regions more easily described and visualized. Often, the linear inequalities defining the polyhedron are much more easily obtained (computationally) than is the boundary equation of the exact region.

In the small sample theory of more complicated problems (such as the location of a regression surface maximum), the standard confidence regions are not ellipsoids and may not even be convex, connected, or bounded. There is no practical way to determine from a particular boundary equation if the region is convex. The intersection region procedures developed here are an attempt to construct some usable approximate representations for some of these problems.

The idea is to approximate a standard region as the intersection of several regions each of which is fairly easy to represent and to compute. They are typically (in two dimensions) curved strips rather than straight strips (Figures 1b and 1c). The regions are determined essentially by applying the multiple comparisons theory at an earlier stage in the confidence region construction. The approximation is carried one stage further in which the curved strips are approximated by straight strips and their intersections by convex polyhedra.

4. Intersection region procedures for families of general linear hypotheses. The most important class of quantities amenable to intersection region procedures arise from general linear hypotheses in general linear models (cf. Wilks [14], Chap. 8). An n dimensional vector $\mathbf{z} = \mathbf{X}\beta + \varepsilon$ is observed in which $\mathbf{X}$ is a known $n \times m$ matrix of rank m, β the unknown m-dimensional vector of 'regression' coefficients and the n components of ε are independently and normally distributed each with zero mean and variance σ^2. Least squares estimates of β and σ^2 are

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{z})$$

$$s^2 = \frac{1}{n - m}[\mathbf{z}'\mathbf{z} - (\mathbf{z}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{z})],$$

distributed independently as a normal with mean β and covariance matrix $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ and as $\sigma^2\chi^2/(n - m)$ on $\nu = n - m$ degrees of freedom.

Many quantities of interest can be represented as the roots of sets of simultaneous linear equations in the regression coefficients; as the root in λ of the equations

$$\sum_{j=1}^{m} \beta_j \delta_{ij}(\lambda) = \delta_{i0}(\lambda); \qquad (i = 1, \cdots, p)$$

linear in the regression coefficients, but of arbitrary though specified form in λ.

Linear combinations of means or regression coefficients are included by choosing all $\delta_{ij}(\lambda)$ to be constants and taking $\delta_{i0}(\lambda) = \lambda_i$. Two one-dimensional quantities typical of the more complicated problems motivating the intersection procedures are:

(i) β_1/β_2: $\delta_{11} = 1$, $\delta_{12} = -\lambda_1$, other δ_{ij} and δ_{i0} zero.

(ii) location of vertex of regression curve $\beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$: $\delta_{11} = 1$ $\delta_{12} = 2\lambda_1$, $\delta_{13} = 3\lambda_1^2$, other δ_{ij} and δ_{i0} zero.

If λ is the true value of the quantity, it satisfies the equations:

$$H_\lambda : \sum \beta_j \delta_{ij}(\lambda) = \delta_{i0}(\lambda); \qquad (i = 1, \cdots, p)$$

or written in vector form (with natural definitions):

$$H_\lambda : \Delta_\lambda \beta = \delta_{0\lambda}$$

and is a "general linear hypothesis." Any procedure for testing H_λ for every λ leads to a confidence region procedure for λ. Several procedures will be used.

Let $\delta_i(\lambda) = \sum \beta_j \delta_{ij}(\lambda) - \delta_{i0}(\lambda)$ and $\delta_\lambda = [\delta_1(\lambda), \cdots, \delta_p(\lambda)]' = \Delta_\lambda \beta - \delta_{0\lambda}$. For each λ, the least squares estimate of δ_λ is $\mathbf{d}_\lambda = \Delta_\lambda \mathbf{b} - \delta_{0\lambda}$. $\mathbf{d}_\lambda$ is normally distributed with $E(\mathbf{d}_\lambda) = \delta_\lambda$ and Cov $(\mathbf{d}_\lambda) = \sigma^2 \Delta_\lambda (\mathbf{X}'\mathbf{X})^{-1} \Delta_\lambda' = \sigma^2 \mathbf{V}_\lambda$. (Note that $\mathbf{d}_\lambda$ and $\mathbf{V}_\lambda$ will generally depend on λ except when Δ_λ is a matrix of constants—as it is for the usual simple problems).

In all that follows, the observations are used only to compute $\mathbf{d}_\lambda$ and $\mathbf{V}_\lambda$ and the sample variance s^2. $\mathbf{V}_\lambda$ *is assumed nonsingular (and hence positive definite) for every* λ.

The likelihood ratio test of the hypothesis $H_\lambda : \delta_\lambda = 0$ is: reject H_λ if $T_\lambda \geqq A$ with test statistic

$$T_\lambda = \frac{\mathbf{d}_\lambda' \mathbf{V}_\lambda^{-1} \mathbf{d}_\lambda}{ps^2}.$$

When H_λ is true, T_λ has an F distribution with p and ν degrees of freedom. In general, T_λ has a noncentral F distribution with noncentrality parameter $\delta_\lambda' \mathbf{V}_\lambda^{-1} \delta_\lambda / \sigma^2$. The test with critical value $F_{p,\nu;\alpha}$ (the upper 100 α percent point of the F distribution) is a similar level α test and is the uniformly most powerful invariant level α test of H_λ. Throughout the paper, this test will be called the "standard" test of the general linear hypothesis.

The corresponding confidence procedure for λ is the "standard" level α procedure: $R_S = \{\lambda: T_\lambda \leqq F_{p,\nu;\alpha}\}$ which can be written as

$$R_S = \left\{\lambda: \begin{vmatrix} ps^2 F_{p;\nu,\alpha} & \mathbf{d}_\lambda' \\ \mathbf{d}_\lambda & \mathbf{V}_\lambda \end{vmatrix} \geqq 0\right\}$$

or

$$R_S = \left\{\lambda: \sum_{i,j} d_{\lambda i}\, d_{\lambda j}\, V_{\lambda ij} - ps^2 F_{p;\nu,\alpha} |\mathbf{V}_\lambda| \leqq 0\right\} \tag{4.1}$$

in which $V_{\lambda ij}$ is the cofactor of the element $v_{\lambda ij}$ in $\mathbf{V}_\lambda$. This confidence procedure has been constructed and used by Box and Hunter [2].

The confidence procedure can be difficult to use especially when the elements of $\mathbf{V}_\lambda$ depend on λ. For then the boundary equation and sometimes the region itself can be very complicated and the necessary computation messy.

The intersection region procedure is based on working separately with the p single equations $\delta_i(\lambda) = 0$ composing H_λ or, more conveniently, with linear combinations of these equations. When $\delta_i(\lambda) = \beta_i - \lambda_i$ so that the quantity of interest is the vector of regression coefficients, the procedure reduces to the multiple comparisons procedure of setting confidence limits on some or all linear combinations of the $\{\beta_i\}$.

Let $\mathbf{k}_1, \cdots, \mathbf{k}_r$ be any r prescribed p-dimensional vectors, and let $H_{\lambda i}$ denote the hypothesis $\mathbf{k}_i' \delta_\lambda = 0$. Every $H_{\lambda i}$ is true when H_λ is, and if the vectors $\{\mathbf{k}_i\}$ span p-space, the truth of all r "component" hypotheses implies that H_λ is true.

Suppose each hypothesis $H_{\lambda i}$ were tested according to: reject $H_{\lambda i}$ if $T_{\lambda i} > A$. A natural joint test of H_λ is to reject H_λ if any $H_{\lambda i}$ is rejected, i.e. if

$$U_\lambda = \max_{1 \leqq i \leqq r} T_{\lambda i} > A. \tag{4.2}$$

Corresponding to each component set of tests is a confidence region R_i such that

$$R_i = \{\lambda: T_{\lambda i} < A\}$$

and the intersection region R_I is defined as the intersection of the r regions $\{R_i\}$, or equivalently as the region defined by the joint test:

$$R_I = \bigcap_{i=1}^{r} R_i = \{\lambda: U_\lambda < A\}.$$

Each $H_{\lambda i}$ is a linear hypothesis with standard test statistic (which will be used)

$$T_{\lambda i} = \frac{(\mathbf{k}_i' \mathbf{d}_\lambda)^2}{s^2(\mathbf{k}_i' \mathbf{V}_\lambda \mathbf{k}_i)}.$$

The component region R_i is given by

$$R_i = \{\lambda: (\mathbf{k}_i'\mathbf{d}_\lambda)^2 - As^2(\mathbf{k}_i'\mathbf{V}_\lambda\mathbf{k}_i) \leqq 0\}$$

and is usually much simpler, computationally and geometrically, than the standard region R_S. When H_λ is true, each $T_{\lambda i}$ is distributed as $F_{1,\nu}$. The joint distribution of the $T_{\lambda i}$ follows, in principle, from the fact that

$$x_i = \frac{\mathbf{k}_i' \mathbf{d}_\lambda}{(\mathbf{k}_i' \mathbf{V}_\lambda \mathbf{k}_i)^{1/2}}; \qquad (i = 1, \cdots, r.) \tag{4.3}$$

are distributed as a multivariate normal with zero means, (when H_λ is true), variances σ^2, and correlations depending on the $\{\mathbf{k}_i\}$ and on λ.

The choice of the critical value A must be a compromise between control of the error level of procedure, ease of computation, and simplicity of the resulting boundary equation. In order that the intersection region procedure have a constant error level α, A must be the 100 α percent point of the distribution of U_λ, the studentized maximum of the squares of correlated normal deviates. But except for a few special cases, these percent points cannot now be obtained without major computation. And since they would likely depend on λ through the correlations, the boundary equations of the intersection region would be complicated by the presence of the function $A(\lambda)$. The use of the exact percent point, if obtainable, for some single "compromise" value of λ might be an excellent choice. I assume throughout that a constant (in λ) critical value is used. Attention here will mainly be restricted to two approximate choices, each "conservative" in the sense that the error level of the intersection region procedure does not exceed the nominal level α.

THEOREM 4.1. *For any set of prescribed* $\mathbf{k}_1, \cdots, \mathbf{k}_r$, *the confidence region procedure* R_I *using critical value* $A = F_{1,\nu;\alpha/r}$ *has error level not exceeding* α.

Since U_λ exceeds A if and only if at least one $T_{\lambda i}$ exceeds A,

$$P(U_\lambda > A) \leqq \sum_{i=1}^{r} P(T_{\lambda i} > A).$$

(This holds generally, without regard for the meaning of $T_{\lambda i}$ and leads to an immediate generalization of the theorem for joint tests and intersection procedures based on any separate tests of any set of component hypotheses.) When H_λ is true, every $T_{\lambda i}$ has an $F_{1,\nu}$ distribution so with $A = F_{1,\nu;\alpha/r}$, the right hand

side is exactly α and the joint test of H_λ has error level not exceeding α. This holds for every λ, so the error level of the intersection region procedure is also so bounded.

The actual error level using $A = F_{1,\nu;\alpha/r}$ will depend on λ through the correlations of the $\{\mathbf{k}_i'\mathbf{d}_\lambda\}$. When these correlations are small, the error level will be quite close to α. Some results on the closeness of the bound α are given in section eight.

As the number r of linear combinations is increased, the correlations increase and the bound gets worse. The behavior of the intersection region is best studied in the limiting case where all linear combinations are used. The distribution theory is exactly that used by Scheffé [12] and is based on an algebraic lemma.

LEMMA (Scheffé). *If* $\mathbf{d}$ *is any* p*-vector,* $\mathbf{V}$ *any symmetric positive definite* $p \times p$ *matrix, then*

$$\sup_{\text{all } \mathbf{k}} \frac{(\mathbf{k}'\mathbf{d})^2}{(\mathbf{k}'\mathbf{V}\mathbf{k})} = \mathbf{d}'\mathbf{V}^{-1}\mathbf{d}.$$

THEOREM 4.2. *The intersection region procedure using all linear combinations, each with critical value* $A = pF_{p,\nu;\alpha}$, *is identical to the level* α *standard region procedure.*

COROLLARY. *Any intersection region based on* r *prescribed combinations* $\{\mathbf{k}_i\}$ *and critical value* A_0 *always contains the standard region with error level*

$$P\{F_{p,\nu} > A_0/p\}.$$

Applying the lemma to $\mathbf{d}_\lambda$ and its covariance matrix $\sigma^2\mathbf{V}_\lambda$ and studentizing with s^2,

$$\sup_{\text{all } \mathbf{k}} \frac{(\mathbf{k}'\mathbf{d}_\lambda)^2}{s^2(\mathbf{k}'\mathbf{V}_\lambda\,\mathbf{k})} = \frac{\mathbf{d}_\lambda'\,\mathbf{V}_\lambda^{-1}\mathbf{d}_\lambda}{s^2}.$$

But the left-hand side is the test statistic associated with the intersection procedure (over all $\mathbf{k}$) and the right-hand side is p times the standard test statistic.

Any intersection procedure can be treated as an approximation to some standard procedure. The intersection region will always contain the standard region and will converge to it as more linear combinations are used. The gain in simplicity of the component regions may more than compensate for the large number of regions and the imperfect approximation.

The use of $A = pF_{p,\nu;\alpha}$ has one advantage over all other choices, approximate or exact, for a finite r. The distribution theory of the $\{T_{\lambda i}\}$ and related statistics is valid only if the vectors $\{\mathbf{k}_i\}$ are chosen independently of $\mathbf{d}_\lambda$ and s^2. But Theorem 4.2 is based on all linear combinations and thus can be used for $\{\mathbf{k}_i\}$ selected after studying the data. A useful a posteriori choice of linear combinations will be illustrated in the application in section seven. (This advantage is a primary motivation of Scheffé's [12] multiple comparison procedure.)

When the equations defining λ are homogeneous linear functions of the regression coefficients, confidence regions for λ can have shapes and behavior not

occurring in classical confidence regions. Several such properties follow from Theorem 4.3.

THEOREM 4.3. *If for all* λ, *each component of* δ_λ *is a homogeneous linear function of the regression coefficients of the linear model, then for any sample there is a nonzero* α^* *which will depend on the sample but not on* λ, *such that the standard confidence region for* λ *is the entire space for any error level less than* α^*.

COROLLARY. *The theorem holds, with a possibly different* α^*, *for the intersection confidence region.*

The theorem follows easily from the well-known interpretation of the test statistic T_λ of the general linear hypothesis, that

$$T_\lambda = \frac{S_{1\lambda} - S_0}{S_0} \cdot \frac{\nu}{p}$$

where $S_0 = \nu s^2$ is the sum of squares of residuals after the least squares fit of the linear model to the data, while $S_{1\lambda}$ is the sum of squares of residuals after the best least squares fit subject to the restriction of the hypothesis $H_\lambda: \delta_\lambda = 0$ (cf. Wilks [14]). One of the possible fits satisfying the homogeneous restrictions $\delta_\lambda = 0$ is that with all regression coefficients estimated to be zero, leaving a residual sum of squares $\sum z^2$, the original sum of squares. Consequently, $S_{1\lambda} \leqq \sum z^2$ for all λ and for any sample, T_λ is bounded by a constant depending on the sample but not on λ. Since $F_{p,\nu;\alpha}$ approaches infinity as α approaches zero, for the sample z there is a nonzero $\alpha^*(z)$ for which H_λ will be accepted for all λ at any significance level $\alpha < \alpha^*(z)$, and the corresponding confidence region will be the entire space. The corollary follows using the corollary to Theorem 4.2.

The theorem shows that the confidence regions need not be bounded. Since the value(s) of λ for which T_λ is maximum will generally be finite, the confidence region as a function of the error level will close in around the maximizing point(s), and the resulting region will be neither convex nor simply connected and perhaps not even connected. (In the usual problems with means and regression coefficients, the hypotheses are not homogeneous and, what is essential, the constant term depends on λ.)

5. Geometry of intersection regions for a class of equations linear in λ. Further study of intersection regions requires specifying the form in λ of the defining equations. An interesting class of equations is suggested by the problem of locating the maximum of a quadratic regression surface (section seven). Suppose that λ is a p-dimensional vector $\mathfrak{\lambda}$ and that the equation $\delta_\lambda = 0$ is linear and *homogeneous* in the regression coefficients and *linear* in $\mathfrak{\lambda}$. Introduce the notation $\delta_\lambda = \gamma + \Gamma\mathfrak{\lambda}$ in which the elements of the p-vector γ and the $p \times p$ matrix Γ are homogeneous linear functions of the regression coefficients. Let $\mathbf{c}$ and $\mathbf{C}$ be the corresponding least squares estimates of γ and Γ and let $\mathbf{d}_\lambda = \mathbf{c} + \mathbf{C}\mathfrak{\lambda}$. The covariance matrix $\sigma^2\mathbf{V}_\lambda$ of $\mathbf{d}_\lambda$ is an inhomogeneous quadratic function of $\mathfrak{\lambda}$. The particular forms of γ and Γ are of no interest except for the evaluation of $\mathbf{d}_\lambda$ and $\mathbf{V}_\lambda$ (a tedious but straightforward task) and to verify the two

assumptions: *Assume that* $\mathbf{V}_\lambda$ *is nonsingular for all* λ. *Assume that the (random) matrix* $\mathbf{C}$ *is nonsingular with probability one.*

Then a unique solution $\hat{\lambda}$ (the maximum likelihood estimate of λ) of $\mathbf{d}_\lambda = 0$ will exist. $\boldsymbol{\Gamma}$ may be singular for a particular set of population regression coefficients, so that the "population value of λ" need not be unique or even exist. (If no solution of $\boldsymbol{\delta}_\lambda = 0$ exists, the confidence problem is vacuous.)

The development of this section concerns geometric properties of intersection regions. Throughout, the observed sample will be held fixed, arbitrarily, except for the above mentioned set of probability zero.

The component region R_i based on the combination $\mathbf{k}_i'\boldsymbol{\delta}_\lambda$ is $\{\lambda: h_i(\lambda) \leqq 0\}$ with quadratic boundary equation $h_i(\lambda) = (\mathbf{k}_i'\mathbf{d}_\lambda)^2 - As^2(\mathbf{k}_i'\mathbf{V}_\lambda\mathbf{k}_i) = 0$. Since $\mathbf{V}_\lambda$ is positive definite for all λ, all points on the hyperplane $\mathbf{k}_i'\mathbf{d}_\lambda = 0$ (call it M_i) lie in the interior of R_i. The two parts (if they exist) of the exterior (complement) of R_i on either side of M_i are each convex (Theorem 9.1).

Thus, R_i is the region between the sheets of a two-sheeted hyperboloid, the exterior of an ellipsoid, or limiting and transitional forms of these. The boundary (call it F_i) can never be one of the one-sheeted hyperboloids.

Considered as a function of the critical value A (or of the significance level), R_i is the hyperplane M_i when $A = 0$ and expands monotonically with A, first as a "curvilinear slab" between the two sheets of a hyperboloid of small curvature, then widening and curving until it eventually becomes the outside of an ellipsoid and finally fills the entire space. By Theorem 4.3, this last will occur for a finite A.

The component region will be most easily described and comprehended when it is a "curvilinear slab," being almost a confidence interval for the linear combination $\mathbf{k}_i'\mathbf{d}_\lambda$. In any case, the computations involved in using R_i are relatively simple.

The component regions are studied as a preliminary to forming intersection regions. In this section, restrict attention to p linearly independent combinations. The intersection of p slabs is a parallelepiped. At best, the component regions are bounded by hyperboloids of small curvature, and the intersection of p such regions is a "curvilinear parallelepiped." The $2p$ "faces" become concave and the 2^p corners are moved at least enough that the 2^{p-1} corners on any "face" are not coplanar.

Let $R_I = \bigcap R_i$ be the intersection region determined by $\mathbf{k}_1, \cdots, \mathbf{k}_p$ and let $J = \bigcap F_i$ be the intersection of the boundaries. The points of J—the "corners" of R_I—are the solutions of the p simultaneous quadratic equations

$$\{h_i(\lambda) = 0, i = 1, \cdots, p\}.$$

Except with probability zero, J will contain not more than 2^p points. Let R^1 be the convex closure of J. R^1 is a convex polyhedron, all of whose vertices are points of J (not necessarily conversely).

In the intuitive discussion above, R_I would be contained in R^1 so that R^1 would be a conservative and perhaps close approximation to R_I. But the ap-

proximation depends critically on R_I being a "curvilinear parallelepiped." If R_I is unbounded, disconnected, or otherwise misshaped, the formal calculation of R^1 will usually lead to a clearly bad approximation, but it can lead to an apparently good but erroneous approximation. Complex shaped intersection regions may arise because of poor choices of components (e.g., too highly correlated) or because of the inadequacy of the sample data in accurately determining λ. The R^1 approximation can be so bad that it has no points in common with R_I except the corners. This will occur if all the corners lie on the part of one boundary (say F_i) on one side of its hyperplane M_i, as will necessarily occur if any R_i is an ellipsoid (Theorem 9.2).

A positive result, representing the behavior understood in the intuitive discussion is that R^1 contains R_I if there are exactly 2^p corners, one in each of the 2^p parts of p-space formed by the p hyperplanes $\{M_i\}$ (Theorem 9.3).

To use R^1, one must find all corners of R_I. There is need for some simpler approximations, to provide a first approximation for the complicated calculations of the corners of R_I, and also to be easier to describe and use than is R_I. Two simple polyhedral approximations R^2 and R^3 are suggested, both based on the idea that the boundaries $\{F_i\}$ are hyperboloids of small curvature, at least in the region of interest.

R^2 is obtained by replacing each hyperboloid F_i by a pair of tangent hyperplanes and taking the convex polyhedron formed by these $2p$ hyperplanes. R^3 is the parallelepiped formed by approximating each F_i by a pair of hyperplanes, both parallel to the corresponding M_i. These approximations are easily determined once the point of tangency or intersection on each boundary is chosen. The suggested choice is the pair of points lying also on every M_j, $j \neq i$. If, for each i, the two points lie on opposite sides of M_i (the reverse is a sure indication that R_I is not a "curvilinear parallelepiped"), R^2 is always contained in R_I (Theorem 9.4). If the hyperboloids are nearly flat, the approximation is good. By dilating the region until all corners are outside R_I, a "conservative" approximation of the same convenient shape could be obtained. *R^2 or a dilation is probably the most useful of all the suggested regions.* R^3 is a much rougher approximation whose virtue is simplicity of shape and computation.

6. On the computation of intersection regions and approximations for equations linear in λ. The computations needed in the approximations of section five are simplified by a change of coordinates. For a particular choice of $\mathbf{k}_1, \cdots, \mathbf{k}_p$, and a particular sample, define new (oblique) coordinates for the space of λ by

$$\xi_i = \mathbf{k}_i' \mathbf{d}_\lambda ; \qquad (i = 1, \cdots, p.)$$

The inverse transformation must be obtained, requiring the solution of p simultaneous linear equations in $\{\lambda_i\}$. (The unique solution is guaranteed by the assumed linear independence of the $\{\mathbf{k}_i\}$ and nonsingularity of $\mathbf{C}$.)

The coordinate hyperplane $\xi_i = 0$ is M_i and the maximum likelihood estimate

$\hat{\lambda}$ is at the origin $\hat{\xi} = 0$. Each $h_i(\lambda)$ can be written in ξ coordinates as

$$h_i(\lambda) = g_i(\xi) = \xi_i^2 - As^2(\mathbf{k}_i'\mathbf{V}_\lambda\mathbf{k}_i)$$

and is a quadratic function of the $\{\xi_i\}$.

The computation of R^2 and R^3 is immensely simplified. The pair of points lying on F_i and every M_j, $j \neq i$ have all ξ coordinates but the ith zero, and that given by the roots $(e_i' < e_i'')$ of the quadratic $g_i(0, \cdots, 0, \xi_i, 0, \cdots, 0) = 0$. If they have the same sign, the intersection region, at least for the particular choice of components, is dangerously complicated. If $e_i' < 0 < e_i''$, then the equation of the tangent hyperplane approximation to the part of F_i with $\xi_i > 0$ (say) is

$$0 = \sum_{\substack{j=1 \\ j \neq i}}^{p} \xi_j \left[\frac{\partial g_i(\xi)}{\partial \xi_j}\right]_{\xi=(0,\cdots,e_i'',\cdots,0)} + (\xi_i - e_i'') \cdot \left[\frac{\partial g_i(\xi)}{\partial \xi_i}\right]_{\xi=(0,\cdots,e_i'',\cdots,0)}.$$

and is easily written down when $g_i(\xi)$ is given explicitly. R^2 is defined by the $2p$ corresponding inequalities, each chosen so that $\hat{\xi} = 0$ satisfies it. R^3 is given as $\{\xi: e_i' < \xi_i < e_i''; i = 1, \cdots, p\}$. The inequalities for R^2 or R^3 are easily changed back to linear inequalities in λ if desired.

The R^1, R^2 and R^3 constructions have no unique or natural extensions to more than p component regions (in p dimensions). One possible procedure would be to repeat the R^2 construction for another set of p components (perhaps with some overlaps), using a new set of coordinates, then converting both sets of inequalities to some one convenient coordinate system. The approximate region would then be the convex polyhedron defined by all of the inequalities. Some dilation of the R^2 region as discussed in section five would be desirable to prevent serious underapproximation of the confidence region.

7. An application to the location of the maximum of a quadratic regression surface. Many aspects of the problem of determining the values of the input variables of a process to yield a maximum response have been studied by Box and colleagues ([1], [2], [3], [4], [5]). Here we use the simple model in which each observed response z is distributed normally and independently with variance σ^2 and mean

$$E(z) = \gamma_0 + \boldsymbol{\gamma}'\mathbf{y} + \tfrac{1}{2}\mathbf{y}'\boldsymbol{\Gamma}\mathbf{y} \tag{7.1}$$

with input variables $\mathbf{y}' = (y_1, \cdots, y_p)$ and regression coefficients

$$\gamma_0, \boldsymbol{\gamma}' = (\gamma_1, \cdots, \gamma_p), \qquad \boldsymbol{\Gamma} = [\gamma_{ij}]$$

with $\gamma_{ij} = \gamma_{ji}$. n sets of $(z, \mathbf{y})$ comprise the observed data and the model is a "general linear model" of section four with $m = 1 + p + p(p+1)/2$ regression coefficients. Denote by c_0, $\mathbf{c}$, $\mathbf{C}$ and s^2 the least squares estimates of γ_0, $\boldsymbol{\gamma}$, $\boldsymbol{\Gamma}$ and σ^2. The fitted surface is

$$\hat{z} = c_0 + \mathbf{c}'\mathbf{y} + \tfrac{1}{2}\mathbf{y}'\mathbf{C}\mathbf{y}. \tag{7.2}$$

The estimates and their variances and covariances can be computed from the formulas of section four or from those given by Box and Wilson [4].

If the surface (7.1) has a maximum at $\mathbf{y} = \boldsymbol{\lambda}$, $\boldsymbol{\lambda}$ will satisfy the stationarity equation $H_\lambda : \boldsymbol{\gamma} + \boldsymbol{\Gamma}\boldsymbol{\lambda} = 0$. This equation will be satisfied by any vertex—maximum, minimum or saddle point—of the surface and all confidence regions are for the location of a vertex, type unspecified. Box and Hunter [2] construct the standard confidence region and show that if the region is bounded, then the region can be said to represent one particular type of vertex in the sense that for every $\boldsymbol{\lambda}$ in the region and for each fitted surface with vertex $\boldsymbol{\lambda}$ that does not give a "significantly poor fit," the vertex is of the same type. Their argument extends to intersection regions based on joint tests of the form of equation (4.2.).

From the structure of $\mathbf{d}_\lambda = \mathbf{c} + \mathbf{C}\boldsymbol{\lambda}$ it follows that even with the maximum of balance and symmetry in the design, each diagonal element of the covariance matrix $\mathbf{V}_\lambda$ of $\mathbf{d}_\lambda$ has at least a constant term and all p square terms. Each off-diagonal element has at least a term in $\lambda_i\lambda_j$. Consequently, in the equation (4.1) for the standard confidence region R_s, each term will generally be a polynomial in the $\{\lambda_i\}$ of degree $2p$. Even for $p = 2$, the equation is already unwieldly. Box and Hunter [3] show that with a rotatable design, the equation can be reduced to a quartic for any p.

H_λ is exactly of the form studied in section five and the nonsingularity assumptions for $\mathbf{V}_\lambda$ and $\mathbf{C}$ are met. Intersection region procedures are applicable. The simplest choice of linear combinations is the direct use of the p components of $\mathbf{d}_\lambda$. With this choice, the critical value $A = F_{1,\nu;\alpha/p}$ could be used to give an error level bounded by α.

A better choice is suggested by a canonical analysis of the fitted surface such as is obtained by introducing new coordinates $\{x_i\}$ with origin at the center and with axes the principal axes of the fitted surface (7.2). If $\{\mathbf{m}_i\}$ and $\{b_{ii}\}$ are the eigenvectors and corresponding eigenvalues of $\mathbf{C}$ and if $b_i = \mathbf{m}_i'\mathbf{c}$, then $x_i = \mathbf{m}_i'\mathbf{y} + b_i/b_{ii}$ and the fitted surface becomes $\hat{z} = \text{constant} + \frac{1}{2}\sum_{i=1}^{p} b_{ii}x_i^2$. (Box and Wilson [4] and Box [1] give more details and interpretations.)

The suggested choice for $\mathbf{k}_i$ is $\mathbf{m}_i/b_{ii}$ which corresponds to using the separate stationarity equations found by equating to zero the partial derivatives of the true surface in the directions of the principal axes of the fitted surface. This set being dependent on the data, the only simple valid critical value is $A = pF_{p,\nu;\alpha}$ (with which arbitrarily many more combinations can be used without increasing the error level of the intersection region). With $\mathbf{k}_i = \mathbf{m}_i/b_{ii}$, the linear form defining ξ_i and used for the component region R_i is

$$\begin{aligned} \xi_i &= \mathbf{k}_i'\mathbf{d}_\lambda = \mathbf{m}_i'\mathbf{d}_\lambda/b_{ii} = \mathbf{m}_i'\mathbf{c}/b_{ii} + \mathbf{m}_i'\mathbf{C}\lambda/b_{ii} \\ &= b_i/b_{ii} + \mathbf{m}_i'\boldsymbol{\lambda}. \end{aligned} \tag{7.3}$$

Thus, the transformation to the ξ coordinate system, useful in the computation of intersection region approximations, is here identical to the transformation to the principal coordinates $\{x_i\}$ useful in understanding the shape of the regression surface.

The intersection region R_I and the approximation R^1, R^2 and R^3 will be illustrated on the numerical example given by Box and Hunter [2] to illustrate R_S. For the fitted surface $c_0 = 77.95$, $s^2 = 1.07$, $\nu = 9$,

$$\mathbf{c} = \begin{bmatrix} 3.76 \\ -1.57 \end{bmatrix} \qquad \mathbf{C} = \begin{bmatrix} -5.74 & 3.84 \\ 3.84 & -5.28 \end{bmatrix}$$

with covariance matrix given by equation (27) of [2]. The fitted surface has a maximum $\hat{\lambda} = (.889, .349)$. The surface in the principal coordinates x is: $\hat{z} = 79.35 - \frac{1}{2}(9.357\, x_1^2 + 1.663\, x_2^2)$ and the transformation of coordinates is

$$\begin{aligned} x_1 &= .7279\, y_1 - .6857\, y_2 - .4076 \\ x_2 &= .6857\, y_1 + .7279\, y_2 - .8631. \end{aligned} \tag{7.4}$$

The transformation consists of a translation of the point $\hat{\lambda}$ to the origin and a rotation through $-43°17.2'$.

Using the second set of recommended $\{\mathbf{k}_i\}$ and the critical value $A = 2F_{2,9;.05} = 8.52$, the transformation (7.3) from the λ to ξ coordinates is identical with the transformation (7.4) from $\mathbf{y}$ to $\mathbf{x}$ coordinates. The boundary equations in the ξ coordinates of the two component regions R_1 and R_2 are:

$$g_1(\xi) = -.035 - .014\, \xi_1 - .054\, \xi_2 + .929\, \xi_1^2 - .045\, \xi_2^2 + .035\, \xi_1\xi_2$$

$$g_2(\xi) = -.433 + .048\, \xi_1 - .468\, \xi_2 - 1.414\, \xi_1^2 + .514\, \xi_2^2 + .709\, \xi_1\xi_2$$

The corners of the intersection region R_I are $(-.88, 3.10)$, $(.43, 1.31)$, $(-.12, -.55)$, $(.16, -.66)$ of which one lies in each quadrant, satisfying the hypothesis of Theorem 9.3, so that R_I is contained in the quadrilateral R^1 formed as convex closure of these points. The parallelepiped approximation R^3 is given by the inequalities $(-.19 \leqq \xi_1 \leqq .20)$, $(-.57 \leqq \xi_2 \leqq 1.48)$ and the polyhedral approximation R^2 is the intersection of the four tangent half spaces:

$$T_1^- = \{\xi:\ .067 + .360\, \xi_1 + .061\, \xi_2 \geqq 0\}$$

$$T_1^+ = \{\xi:\ .073 - .360\, \xi_1 + .047\, \xi_2 \geqq 0\}$$

$$T_2^- = \{\xi:\ .599 + .355\, \xi_1 + 1.053\, \xi_2 \geqq 0\}$$

$$T_2^+ = \{\xi:\ 1.558 - 1.097\, \xi_1 - 1.053\, \xi_2 \geqq 0\}$$

The regions R_1, R_2, R_I, R^1, R^2, R^3, and R_S (the latter taken from [2], p. 198) are illustrated in Figure 2. (In two dimensions, the approximations to R_I are unnecessary except for simple analytic description and are shown principally to illustrate the different approximations.)

8. Bounds on the error level of intersection procedures. A lower bound on the error-level can be obtained that gives some indication of the closeness of the bound in Theorem 4.1.

Fix λ, let $u_i = x_i/s$ with x_i defined by equation (4.3) and let ρ_{ij} = correlation (x_i, x_j). The joint distribution under H_λ of the $\{u_i\}$ is an r-variate generalization

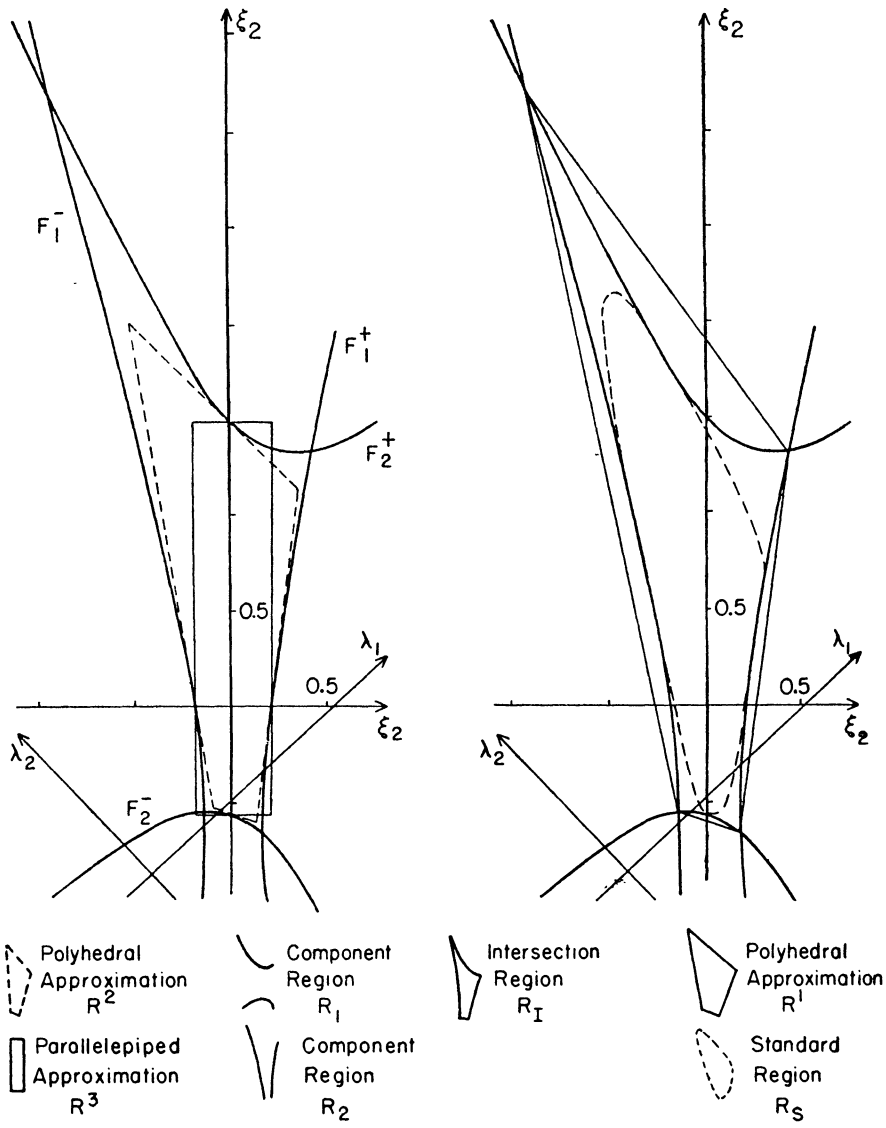

FIG. 2. Various approximate confidence regions for the location of the maximum of the quadratic regression example of section seven.

of the t distribution (cf. Dunnett and Sobel [6]) with ν degrees of freedom and correlation matrix $[\rho_{ij}]$ of the associated r-variate normal distribution. Denote the bivariate distribution integrals by

$$d_\nu(a, b, \rho_{ij}) = P(u_i > a, u_j > b)$$

$$f_\nu(a, b, \rho_{ij}) = P(|u_i| > a, |u_j| > b)$$

K. Pearson [10, Table VIII and IX] gives $d_\infty(a, b, \rho)$ for a selection of a, b, ρ and Dunnett and Sobel [6] give something simply related to $d_\nu(a, a, \pm 0.5)$ for a selection of a and ν, and formulas for computing other values. The marginal $f_\nu(a, 0, \rho)$ is the double-tail probability of Student's t distribution and is independent of ρ.

The error level of the joint test is

$$P(U_\lambda > A) = P\left(\max_{1 \leqq i \leqq r} |u_i| \geqq \sqrt{A}\right).$$

Bounds on this probability are given in Theorem 8.1.

THEOREM 8.1.

$$P\left(\max_{1 \leqq i \leqq r} |u_i| \geqq a\right) \leqq r f_\nu(a, 0, -) \tag{1}$$

$$P\left(\max_{1 \leqq i \leqq r} |u_i| \geqq a\right) \geqq r f_\nu(a, 0, -) - \sum_{i<j} f_\nu(a, a, \rho_{ij}) \tag{2}$$

$$P\left(\max_{1 \leqq i \leqq r} |u_i| \geqq a\right) \geqq r f_\nu(a, 0, -) - \binom{r}{2}\left[\left(1 - \frac{\rho_1}{\rho_0}\right) f_\nu(a, a, 0) + \frac{\rho_1}{\rho_0} f_\nu(a, a, \rho_0)\right] \tag{3}$$

in which $\rho_0 = \max |\rho_{ij}|$ *and* $\rho_1 = \sum_{i<j} |\rho_{ij}| \Big/ \binom{r}{2}$.

Equality occurs in (1) *if* $r = 1$ *and in* (2) *and* (3) *if* $r = 2$.

Inequalities (1) and (2) are direct applications of Bonferroni's inequalities (cf. Feller [7]) to the events $\{|u| \geqq a\}$. The inequality (3) follows on combining with inequality (2) the symmetry and convexity of $f_\nu(a, a, \rho)$ proved in the lemma below. For since $0 \leqq |\rho_{ij}| \leqq \rho_0$,

$$f_\nu(a, a, \rho_{ij}) = f_\nu(a, a, |\rho_{ij}|) \leqq \left(1 - \frac{|\rho_{ij}|}{\rho_0}\right) f_\nu(a, a, 0) + \frac{|\rho_{ij}|}{\rho_0} f_\nu(a, a, \rho_0)$$

and

$$\sum_{i<j} f_\nu(a, a, \rho_{ij}) \leqq \binom{r}{2}\left[\left(1 - \frac{\rho_1}{\rho_0}\right) f_\nu(a, a, 0) + \left(\frac{\rho_1}{\rho_0}\right) f_\nu(a, a, \rho_0)\right].$$

LEMMA. $f_\nu(a, a, \rho)$ *is a symmetric, convex function of* ρ.

Use $f_\nu(a, a, \rho) = 2d_\nu(a, a, \rho) + 2d_\nu(a, a, -\rho)$. Write each d_ν as the double integral of the bivariate t-density, change to "elliptical polar coordinates" ([6], p. 154), and integrate out the angle. Straightforward calculation of $\partial^2 f_\nu / \partial \rho^2$ shows convexity.

Table 8.1 gives the upper bound (1) and lower bound (3) for a selection of values of a, r, ν, ρ_0, ρ_1 chosen to give upper bound near .05 or .01. The upper bound would seem to be sufficiently accurate for most uses provided the correlations and r are not very large.

TABLE 8.1

Bounds on $P\left(\max_{1\leq i\leq r} |u_i| > a\right)$ *from Theorem* 8.1

r	ν	a	Upper Bound (1)	Lower Bound (3)					
				$\rho_0 = 0$ $\rho_1 = 0$	$\rho_0 = .5$		$\rho_0 = .8$		
					$\rho_1 = .2$	$\rho_1 = .5$	$\rho_1 = .2$	$\rho_1 = .5$	$\rho_1 = .8$
2	∞	2.2	.056	.055*	—	.052*	—	—	.045*
3	∞	2.4	.049	.048*	.046	.043	—	.038	.031
5	∞	2.6	.047	.046*	.043	.038	.038	.027	.016
8	∞	2.7	.055	.054*	.048	.039	.039	.016	<0
10	∞	2.8	.051	.050*	.043	.033	.033	.007	<0
2	∞	2.8	.0102	.0102*	—	.0098*	—	—	.0080*
3	∞	2.9	.0112	.0112*	.0109	.0104	—	.0092	.0080
5	∞	3.1	.0097	.0096*	.0092	.0086	.0084	.0065	.0046
8	∞	3.2	.0110	.0109*	.0102	.0092	.0086	.0050	.0014
10	∞	3.3	.0097	.0096*	.0089	.0079	.0071	.0033	<0
2	14	2.5	.051	.049*	—	.046*			
5	14	3.0	.048	.044	.039	.033			
3	8	3.0	.051	.047	.044	.040			
2	6	3.0	.048	.045*	—	.042*			
5	7	3.5	.050	.041	.036	.029			

* Exact value.

—Impossible ρ_0, ρ_1 combination.

9. Mathematical results for the geometry of section five. The notation of sections five and six will be used, and all results are for a fixed sample. For terminological convenience, all work will be done in terms of a Euclidean p-space E_p with *rectangular* coordinates ξ. The affine transformation of the space does not affect the properties of interest.

Let Var $\xi_i = \sigma^2(\mathbf{k}_i'\mathbf{V}_\lambda\mathbf{k}_i)$ be the variance of the linear form defining ξ_i, transformed for the particular sample to a function of ξ. If S_i is any set in E_p indexed by i, let

$$S_i^+ = S_i \cap \{\xi\colon \xi_i > 0\}, \qquad S_i^- = S_i \cap \{\xi\colon \xi_i < 0\}, \qquad S_i^0 = S_i \cap \{\xi\colon \xi_i = 0\}.$$

Let $S^* = E_p - S$, and $\bar{S}$ = closure of S.

THEOREM 9.1. *R_i^{*+} and R_i^{*-} are convex.*

To prove R_i^{*+} convex, it is sufficient to show that for any two points ξ_1 and ξ_2 in R_i^{*+} and any constant θ such that $0 < \theta < 1$, $\xi_0 = \theta\xi_1 + (1 - \theta)\xi_2$ is in R_i^{*+}. But

$$\xi \in R_i^{*+} \Leftrightarrow \{g_i(\xi) > 0 \quad \text{and} \quad \xi_i > 0\}.$$

$\xi_{0i} > 0$ is immediate. Expanding $g_i(\xi_0)$, using the Cauchy inequality for covariance, $g_i(\xi_j) > 0$ for $j = 1, 2$ and $\xi_{1i}\xi_{2i} > 0$,

$$g_i(\xi_0) = \theta^2[\xi_{1i} - (As^2/\sigma^2) \text{ Var } \xi_{1i}]$$
$$+ (1 - \theta)^2[\xi_{2i} - (As^2/\sigma^2) \text{ Var } \xi_{2i}]$$
$$+ 2\theta(1 - \theta)[\xi_{1i}\xi_{2i} - (As^2/\sigma^2) \text{ Cov } (\xi_{1i}, \xi_{2i})]$$
$$\geqq \theta^2 g_i(\xi_1) + (1 - \theta)^2 g_i(\xi_2)$$
$$+ 2\theta(1 - \theta)[\xi_{1i}\xi_{2i} - (As^2/\sigma^2)\sqrt{(\text{Var } \xi_{1i})(\text{Var } \xi_{2i})}]$$
$$> 2\theta(1 - \theta)[\xi_{1i}\xi_{2i} - (\xi_{1i}^2\xi_{2i}^2)^{1/2}] = 0.$$

COROLLARY. *$\overline{R_i^{*+}}$ and $\overline{R_i^{*-}}$ are convex.*

THEOREM 9.2. *If for some i and some sign (say $+$), $J \cap F_i^+ = J$, then $R_I \cap R^1 \subset F_i^+$.*

By hypothesis, $J \subset F_i^+ \subset \overline{R_i^{*+}}$. By the corollary to Theorem 9.1, $\overline{R_i^{*+}}$ is convex and closed so that $R^1 \subset \overline{R_i^{*+}}$.

$$R_I \cap R^1 \subset R_I \cap \overline{R_i^{*+}} \subset R_I \cap (R_i^{*+} \cup F_i^+) \subset R_I \cap F_i^+ \subset F_i^+.$$

COROLLARY. *If any boundary F_i is an ellipsoid, $R_I \cap R^1 \subset F_i$.*

The entire ellipsoid must be on one side of the plane $\xi_i = 0$.

THEOREM 9.3. *If J contains exactly 2^p points, with one point in each of the 2^p open orthants formed by the p coordinate hyperplanes $M_1, \cdots, M_p$, then $R_I \subset R^1$.*

Three lemmas will be proved, from which the theorem follows immediately.

By hypothesis, J contains 2^p points, one in each open orthant defined by the p coordinate planes $\{\xi_i = 0\}$. Denote the 2^p points ("corners") by

$$\{\mathbf{e}_u' = (e_{u1}, \cdots, e_{up}); u = 0, \cdots, 2^p - 1\}$$

with the subscript u assigned so that if $[u_1, \cdots, u_p]$ is the binary expansion of u, then $e_{uj} > 0$ ($\mathbf{e}_u \varepsilon M_j^{*+}$) if $u_j = 1$ and $e_{uj} < 0$($\mathbf{e}_u \varepsilon M_j^{*-}$) if $u_j = 0$. The point $\mathbf{e}_u$ is in the diagonally opposite orthant from $\mathbf{e}_{2^p-1-u}$. The "diagonal line" $D(u, 2^p - 1 - u)$ through $\mathbf{e}_u$ and $\mathbf{e}_{2^p-1-u}$ can be parametrized as

$$\{\xi: \xi = \theta\mathbf{e}_u + (1 - \theta)\mathbf{e}_{2^p-1-u}\}$$

and is the union of three disjoint parts: the "diagonal segment" $D^0(u, 2^p - 1 - u)$ with $0 \leqq \theta \leqq 1$, and two "outer diagonals" $D(u)$ with $\theta > 1$ and $D(2^p - 1 - u)$ with $\theta < 0$. $D(u)$ is contained in the same open orthant as $\mathbf{e}_u$. Define

$$Q_i^+ = \text{convex hull of } \bigcup_{\{u: u_i = 1\}} D(u)$$

$$Q_i^- = \text{convex hull of } \bigcup_{\{u: u_i = 0\}} D(u)$$

$$Q = \bigcup_{i=1}^{p} (Q_i^+ \cup Q_i^-).$$

LEMMA 1. *Under the tonditions of Theorem 9.3, $R_I \subset Q^*$.*

For each value of u and i, the diagonal $D(u, 2^p - 1 - u)$ intersects the boundary F_i in the two points $\mathbf{e}_u$ and $\mathbf{e}_{2^p-1-u}$. The diagonal segment crosses each

coordinate hyperplane M_i. Since the boundary equation of R_i is quadratic, the segment $D^0(u, 2^p - 1 - u)$ is in R_i and the outer diagonal D_u is in R_i^*. Then

$$\bigcup_{\{u:u_i=1\}} D_u \subset R_i^{*+} \quad \text{and} \quad \bigcup_{\{u:u_i=0\}} D_u \subset R_i^{*-}.$$

By Theorem 9.1, R_i^{*+} and R_i^{*-} are convex, so $Q_i^+ \subset R_i^{*+} \subset R_i^*$ and $Q_i^- \subset R_i^{*-} \subset R_i^*$ for every i. Then $Q \subset \cup_{i=1}^p R_i^* = R_I^*$ and $R_I \subset Q^*$.

LEMMA 2. *Under the conditions of Theorem* 9.3, R^1 *contains a cube with the origin in the interior.*

By induction on the dimension p, the cube with faces $|\xi_i| = a = \min_{u,j}|e_{uj}|$ will be shown to be in R^1. Let ξ^0 be any point with $|\xi_i^0| \leqq a$ for all i. If $p = 1$, then $e_{01} \leqq -a \leqq \xi_1^0 \leqq a \leqq e_{11}$ for the one coordinate and ξ^0 is in the convex closure of $\mathbf{e}_0$ and $\mathbf{e}_1$.

For arbitrary p, let $\mathbf{e}_u$ be an arbitrary corner. Suppose that $e_{up} > 0$. Let $\mathbf{e}_{u'}$ be the corner with sign $e_{u'j} = \text{sign } e_{uj}$ for $j < p$ and $e_{u'p} < 0$. Then $e_{u'p} \leqq -a \leqq \xi_p^0 \leqq a \leqq e_{up}$ and there is a convex linear combination $\mathbf{d}_u = \theta_u \mathbf{e}_u + (1 - \theta_u)\mathbf{e}_{u'}$ with $0 \leqq \theta_u \leqq 1$ such that $d_{up} = \xi_p^0$. For the other coordinates, $|d_{uj}| \geqq \min(|e_{uj}|, |e_{u'j}|) \geqq a$ and sign $d_{uj} = \text{sign } e_{uj}$. There are 2^{p-1} points $\mathbf{d}_u$ with $u = (u_1, \cdots, u_{p-1}, 1)$, satisfying the conditions of the theorem on the $p - 1$ dimensional hyperplane $\xi_p = \xi_p^0$ with restricted $\min_{u,j}|d_{uj}| \geqq a$. By the induction hypothesis, the point ξ^0 lies in the convex closure of the $\{\mathbf{d}_u\}$ but since each of these $\mathbf{d}_u$ is in the convex closure of R^1, so also is ξ^0, completing the proof of the lemma.

LEMMA 3. *Under the conditions of Theorem* 9.3, $R^1 \supset Q^*$.

The lemma will be proved by defining an expansion of R^1 to the entire space using only points of Q.

By Lemma 2, R^1 contains a cube P with corners $(\pm a, \cdots, \pm a)$. Denote the corners of P by $\{\mathbf{c}_u\}$ with $\mathbf{c}_u$ in the same orthant as $\mathbf{e}_u$. The cube P can be decomposed into disjoint open simplexes of dimension p and less, all vertices being corners of P and every face of every simplex in the collection. A simplex S of dimension q and with vertices $\mathbf{c}_{u(1)}, \cdots, \mathbf{c}_{u(q+1)}$ is defined as

$$S = \left\{\xi: \xi = \sum_{j=1}^{q+1} \theta_{u(j)}\mathbf{c}_{u(j)}; \sum_{j=1}^{q+1} \theta_{u(j)} = 1, \text{ all } \theta_{u(j)} > 0\right\}.$$

Taking $\theta_u = 0$ for all $u \neq u(j)$ for any j, the $\{\theta_u, u = 0, \cdots, 2^p - 1\}$ are the barycentric coordinates of the point ξ with respect to the simplicial decomposition. The barycentric coordinates are continuous functions over each closed simplex $\bar{S}$ and uniquely defined over P and so are continuous over P. (Cf. Lefschetz [9], p. 97.)

One such decomposition of the cube consists of the $p!$ p-dimensional simplexes $S_i = \{\xi: -a < \xi_{i_1} < \cdots < \xi_{i_p} < a\}$ in which $i = (i_1, \cdots, i_p)$ is any permutation of the integers $(1, \cdots, p)$, and of all faces of the $\{S_i\}$. S_i can be written as

$$S_i = \left\{\xi: \xi = \sum_{j=1}^{p+1} \theta_{u(i,j)}\mathbf{c}_{u(i,j)}; \sum_{j=1}^{p+1} \theta_{u(i,j)} = 1, \theta_{u(i,j)} > 0\right\}$$

in which $c_{u(i,j)}$ is that corner of P whose i_kth coordinate is $-a$ for $k < j$ and $+a$ for $k \geqq j$. The correspondence between the two representations of S_i is given by:

$$\theta_{u(i,j)} = \frac{1}{2a}(\xi_{i_j} - \xi_{i_{j-1}}), \qquad (j = 1, \cdots, p+1):$$

$$(\xi_{i_0} \equiv -a, \xi_{i_{p+1}} \equiv +a).$$

Any simplex lies entirely in a face, say $\{\xi_i = +a\}$, of P or else does not intersect a face. For if all vertices of S lie in $\{\xi_i = a\}$ then so does S, and if one or more vertices do not lie in $\{\xi_i = a\}$ then the ith coordinate of each point in S is less than $+a$. Let $\mathcal{S}$ be the collection of all simplexes lying in the face F of P. $\mathcal{S}$ is a disjoint simplicial decomposition of F and the barycentric coordinates are continuous over F. Each simplex in $\mathcal{S}$ lies in some single face of the cube. Let S with corners $\mathbf{c}_{u(1)}, \cdots, \mathbf{c}_{u(k)}$ be an arbitrary member of $\mathcal{S}$. Suppose that S lies in the face $\{\xi_i = a\}$ and hence in M_i^{*+}.

Define a deformation of F as follows. For $\xi \varepsilon S$ with $\xi = \sum \theta_{u(j)}\mathbf{c}_{u(j)}$; $\sum \theta_{u(j)} = 1$; $\theta_{u(j)} > 0$; define

$$f_t(\xi) = t\sum_{j=1}^{k} \theta_{u(j)}\mathbf{e}_{u(j)} + (1-t)\sum_{j=1}^{k} \theta_{u(j)}\mathbf{c}_{u(j)}, \qquad 0 \leqq t \leqq 1$$

$$\left.\begin{aligned} f_t(\xi) &= t\sum_{j=1}^{k} \theta_{u(j)}\mathbf{e}_{u(j)} + (1-t)\sum_{j=1}^{k} \theta_{u(j)}\mathbf{e}_{2p-1-u(j)} \\ &= \sum_{j=1}^{k} \theta_{u(j)}(t\mathbf{e}_{u(j)} + (1-t)\mathbf{e}_{2p-1-u(j)}). \end{aligned}\right\} \quad t > 1$$

Let $F(t)$ be the image of F under f_t. The deformation consists in moving each corner of the cube to the corresponding corner of R_I, then out the diagonals.

For each t, f_t is a continuous mapping of F into E_p since the barycentric coordinates are continuous over F. Further, the family of mappings is jointly continuous in t and ξ. f_0 is the identity mapping on F and f_t is homotopic to f_0 for all t. Write $f_t/F \sim f_0/F$; $/F$ to indicate the domain of the mapping and $\sim$ for the homotopy equivalence relation (cf. Lefschetz [9], p. 42).

For $0 \leqq t \leqq 1$, $F(t) \subset R^1$ by Lemma 2 and the convexity of R^1. For $t > 1$, $t\mathbf{e}_{u(j)} + (1-t)\mathbf{e}_{2p-1-u(j)}$ lies on the outer diagonal $D(u(j))$. Since the corners of S all lie in M_i^{*+}, so do the $\{\mathbf{e}_{u(j)}\}$. Therefore $f_t(S) \subset Q_i^+$, the convex hull of all outer diagonals in M_i^{*+}. Finally then, $F(t) \subset Q$ for all $t > 1$ and

$$P \cup (\bigcup_{t>0} F(t)) \subset R^1 \cup Q.$$

The proof of the lemma will be completed if any point not in P can be shown to lie in $F(t)$ for some $t > 0$.

The distance from the origin to the image $S(t)$ of any simplex S of the face of the cube is never less than a and increases to infinity. For, if S lies in M_i^{*+}, then $c_{u(j),i} = a$, $e_{u(j),i} \geqq a$ and $e_{2p-1-u(j),i} \leqq -a$. Then for any point ξ in $S(t)$, $\xi_i \geqq a$ if $t \leqq 1$ and $\xi_i \geqq (2t-1)a$ if $t > 1$.

If $\mathbf{x}$ is any point in E_p, denote by π_x the mapping of $E_p - \mathbf{x}$ onto the unit sphere centered at the origin, which maps ξ into the projection from the origin of $\xi - \mathbf{x}$ (vector subtraction). Denote by π_x/B the mapping π_x with domain restricted to the set B. The following topological theorem is needed:[1]

THEOREM (Hurewicz and Wallman [8], Theorem VI-10). *Let B be a closed bounded subset of E_p. Two points $\mathbf{x}$, $\mathbf{y}$ neither contained in B are separated by C, if and only if the mapping π_x/B and π_y/B are not homotopic: $\pi_x/B \not\sim \pi_y/B$.*

Assume there is a point $\mathbf{x}$ not contained in $P \cup (\cup_{t>0} F(t))$. Since the distance $(\mathbf{0}, F(t)) \to \infty$, choose t_1, such that dist. $(\mathbf{0}, F(t)) >$ dist. $(\mathbf{0}, \mathbf{x})$. Since $\mathbf{x}$ does not lie in the cube P, the points $\mathbf{0}$ and $\mathbf{x}$ are separated by the cube boundary F according to the Jordan separation theorem. Applying the topological theorem with $B = F$, $\pi_0/F \not\sim \pi_x/F$. Then for any t

$$\pi_0/F(t) \equiv \pi_0 f_t/F \sim \pi_0 f_0/F \equiv \pi_0/F \not\sim \pi_x/F \equiv \pi_x f_0/F \sim \pi_x f_t/F \equiv \pi_x/F(t),$$

since (a) f_0/F is the identity mapping, (b) $f_0/F \sim f_t/F$ by construction, (c) $\phi f_0/F \sim \phi f_t/F$ by composition for any mapping ϕ with correct domain ([9], p. 42). Since homotopy is an equivalence relation, $\pi_0/F(t) \not\sim \pi_x/F(t)$. Applying the topological theorem again with $B = F(t_1)$, it follows that the points $\mathbf{0}$ and $\mathbf{x}$ are separated by $F(t_1)$. The line segment $[\mathbf{0}, \mathbf{x}]$ being connected, must intersect $F(t_1)$ which is impossible since dist. $(\mathbf{0}, \mathbf{x}) <$ dist. $(\mathbf{0}, F(t_1))$. Therefore $P \cup (\cup_{t>0} F(t)) = E_p$, completing the proof of the lemma and Theorem 9.3.

THEOREM 9.4. *If for each i, $e_i' < 0 < e_i''$ then $R^2 \subset R_I$.*

The line $L_i = \cap_{j \neq i} M_j$ intersects F_i^+ at $\xi_i = e_i'' > 0$ and F_i^- at $\xi_i = e_i' < \mathbf{0}$. Approximate F_i^+ and F_i^- by their tangent hyperplanes at these points. If T_i^+ and T_i^- denote the closed half-spaces bounded by these tangent hyperplanes with halves chosen to contain the origin, approximate R_i by $T_i^+ \cap T_i^-$ and R_I by $R^2 = \cap_{i=1}^p (T_i^- \cap T_i^-)$. Since R_i^{*+} is open and convex and contains the line L_i for $\xi_i > e_i''$, then R_i^{*+} does not intersect the tangent hyperplane or the half-space T_i^+. Similarly, $R_i^{*+} \cap T_i^- = 0$ so that $T_i^+ \cap T_i^- \subset R_i$ and $R^2 \subset R_I$.

10. Acknowledgment. The author is indebted to Professor John W. Tukey for suggesting the subject of this paper, and for his advice and encouragement during its preparation.

REFERENCES

[1] G. E. P. BOX, "The exploration and exploitation of response surfaces: some general considerations and examples," *Biometrics*, Vol. 10 (1954), pp. 16–60.

[2] G. E. P. BOX AND J. S. HUNTER, "A confidence region for the solution of a set of simultaneous equations with an application to experimental design," *Biometrika*, Vol. 41 (1954), pp. 190–198.

[3] G. E. P. BOX AND J. S. HUNTER, "Multi-factor experimental designs for exploring response surfaces," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 195–241.

[4] G. E. P. BOX AND K. B. WILSON, "On the experimental attainment of optimum conditions," *J. Roy. Stat. Soc., Series B*, Vol. 13 (1951), pp. 1–45.

[1] The use of this theorem was indicated to the author by Professor E. H. Spanier.

[5] O. L. DAVIES (Editor), *Design and Analysis of Industrial Experiments*, London, Oliver and Boyd, 1954, Chap. 11.
[6] C. W. DUNNETT AND M. SOBEL, "A bivariate generalization of Student's t-distribution, with tables for certain special cases," *Biometrika*, Vol. 41 (1954), pp. 153–169.
[7] W. FELLER, *Probability Theory and Its Applications*, New York, John Wiley and Sons, 1950.
[8] W. HUREWICZ AND H. WALLMAN, *Dimension Theory*, Princeton, Princeton University Press, 1941.
[9] S. LEFSCHETZ, *Introduction to Topology*, Princeton, Princeton University Press, 1949.
[10] K. PEARSON, *Tables for Statisticians and Biometricians, Part II*, London, Cambridge University Press for the Biometrika Trustees, 1931.
[11] S. N. ROY AND R. C. BOSE, "Simultaneous confidence interval estimation," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 513–536.
[12] H. SCHEFFÉ, "A method for judging all contrasts in the analysis of variance," *Biometrika*, Vol. 40 (1953), pp. 87–104.
[13] J. W. TUKEY, "The problem of multiple comparisons," duplicated manuscript, Princeton University (1953).
[14] S. S. WILKS, *Mathematical Statistics*, Princeton, Princeton University Press, 1943.