ON CERTAIN TYPES OF BIAS IN CURRENT METHODS OF RESPONSE SURFACE ESTIMATION

by

H. R. van der Vaart

Prepared under Contract No. DA-36-034-ORD-1517 (RD)

(Experimental Designs for Industrial Research)

Institute of Statistics

Mimeo. Series No. 205

July, 1958

ON CERTAIN TYPES OF BIAS IN CURRENT METHODS

OF RESPONSE SURFACE ESTIMATION[1]

by

H. R. van der Vaart[2]

1. Introduction

Let $\xi_1, \ldots, \xi_k$ represent a set of k simultaneous conditions (f.i. temperature, humidity, pressure, etc.) under which some process, biological or chemical, is made to take place. Let η represent the yield or response of the process. Experiments will be made at different sets of levels for each of the k conditions. Interest is in η as a function of $\xi_1, \ldots, \xi_k$, or equivalently in η as a scalar function of the k x 1 matrix (column-vector) $\underline{\xi}$. We shall assume throughout that the response function, or response surface, is quadratic, i.e. that

$$\eta = \eta_0 + \underline{\beta}'\underline{\xi} + \underline{\xi}'\underline{\Phi}\,\underline{\xi}. \qquad (1)$$

Here the k x 1 matrix $\underline{\beta}$ and the k x k matrix $\underline{\Phi}$ represent constants which the experimenter will want to determine; a prime denotes transposition of any matrix, so $\underline{\beta}'$ is a row-vector. Of course there are interesting problems because of the fact that this quadratic model will only rarely be correct (cf. for instance Box and Hunter, 1957, p. 216), but this paper is restricted to certain problems which arise even though this model would be correct.

Evidently, if the different levels of the k components of $\underline{\xi}$ could be experimentally produced with exactitude, and if the corresponding yield η could be observed with exactitude, there would be little need for statistics in the problem.

[1] A report on work sponsored by the Office of Ordnance Research, United States Army, under contract DA-36-034-ORD-1517(RD) with the Institute of Statistics, North Carolina State College of the University of North Carolina, Raleigh, N. C.

[2] Present address: Institute for Theoretical Biology of the University, Leiden, Netherlands.

As is done in most statistical work on the theory of estimation of response surfaces (cf. Box and Wilson, 1951, p. 2; Bloemena, 1956, p. 8; Box and Hunter, 1957, p. 198; Cochran and Cox, 1957, p. 335), we will assume that $\underline{\xi}$ can be experimentally produced with exactitude, but η cannot be so observed: instead of the "true" response η we observe realizations of a random variable y with

$$\mathcal{E}\, y = \eta, \qquad \text{var } y = \sigma^2 , \tag{2}$$

var y = σ^2 being assumed not to depend on the corresponding value of $\underline{\xi}$, and η depending on $\underline{\xi}$ according to (1). Observe that we denote random variables by Latin letters, parameters by Greek letters, square matrices by capital letters, vectors (k x 1 matrices) by underlined small letters.

The general procedure now is to choose wisely a set of N, say, sets of values of the k conditions being studies, i.e., a set of N points in $(\xi_1, \ldots, \xi_k)$-space, and for each of these to observe y, finally to estimate η_0, $\underline{\beta}$ and $\bar{\Phi}$ from these data. This estimation is usually done by the method of least squares (cf. Box and Wilson, 1951, p. 5: Box and Hunter, 1957, p. 198), which yields expectation-unbiased estimators of η_0, $\underline{\beta}$ and $\bar{\Phi}$ provided (a) the random variables y corresponding to the different points in the above-mentioned set of N points in $(\xi_1, \ldots, \xi_k)$-space are uncorrelated and (b) the pattern of these N points in $(\xi_1, \ldots, \xi_k)$-space satisfies a certain condition (mentioned in Section 3, p. 5 of Box and Wilson, 1951) which is easy to comply with. If in addition y is normally distributed, the estimators of η_0 and of the elements of $\underline{\beta}$ and $\bar{\Phi}$ are multinormally distributed, their covariance matrix depending on the pattern (the design) of the N points in $(\xi_1, \ldots, \xi_k)$-space.

The present paper will investigate a few questions arising with respect to current methods of estimating the type of quadratic surface: i.e. estimating whether it has a maximum, or a minimum, or a saddle-point, or a stationary ridge, or a rising ridge, etc.

As is well known, if one wants to investigate the type of quadratic surface defined by equation (1), the thing to do is to reduce this equation to canonical form, i.e., to rotate the $(\xi_1, \ldots, \xi_k)$-axes in such a way that the transform of $\underline{\xi}'\Phi\underline{\xi}$ no longer contains cross-product terms. This means that one seeks to construct an orthogonal matrix Y such that if one effects the transformation

$$\underline{\xi} = Y\underline{\zeta}, \qquad (3a)$$

which throws equation (1) in the form

$$\eta = \eta_0 + \underline{\beta}'Y\underline{\zeta} + \underline{\zeta}'Y'\Phi Y\underline{\zeta}, \qquad (3b)$$

the matrix $Y'\Phi Y$ is diagonal. Denote this diagonal matrix by

$$Y'\Phi Y = \Lambda, \qquad (3c)$$

then (3b) reads

$$\eta = \eta_0 + \underline{\beta}'Y\underline{\zeta} + \underline{\zeta}'\Lambda\underline{\zeta}. \qquad (3d)$$

The elements of Λ, $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_k$, are the latent roots of the matrix Φ. It is well known that such an orthogonal matrix Y can always be constructed. Obviously the type of a quadratic surface can be assessed much easier from an equation like (3d) than from (1). Note that plotting contour lines (which is feasible if $k = 2$) directly from equation (1) (cf. for instance Cochran and Cox, 1957, p. 352) constitutes essentially the same method as reducing equation (1) to canonical form: one may regard reduction to canonical form as a device by which to plot contour lines.

Now, as we set forth in the second and third paragraph of this introduction, η_0, $\underline{\beta}$ and Φ are unknown; only their (expectation-unbiased) estimates, to be denoted by y_0, $\underline{b}$ and F, are available. Yet Box and Wilson, 1951, p. 23-24; Box, 1954, p. 35; Box and Youle, 1955, p. 289; Box and Hunter, 1957, p. 239 ("A fitted second degree equation can be interpreted most readily by writing it in the canonical form") proceed in just the same way: they construct an orthogonal matrix U such that the transformation

$$\underline{\xi} = U \underline{\zeta} \tag{4a}$$

throws the fitted equation

$$\hat{\eta} = y_0 + \underline{b}'\underline{\xi} + \underline{\xi}'F\underline{\xi} \tag{4b}$$

in the form

$$\hat{\eta} = y_0 + \underline{b}'U\underline{\zeta} + \underline{\zeta}'L\,\underline{\zeta}, \tag{4c}$$

where

$$L = U'FU \tag{4d}$$

is the diagonal matrix consisting of the latent roots $\ell_1 \leq \ell_2 \leq \dots \leq \ell_k$ of F. Then they use these latent roots of the estimating matrix F in their inference concerning the type of quadratic response surface.

We shall report on some properties of the distribution of the latent roots of F which seem to explain (at least partly) some difficulties cited in the literature on applications, and which make clear that as long as the present method (of estimating Λ) has not been replaced by a better one, it should at least be applied with caution. Unfortunately this caution calls for more experimental points (i.e., for larger N) than otherwise would be necessary, thus counteracting the third requirement concerning experimental designs for estimating response surfaces which was laid down by Box and Hunter, 1957, on p. 197: "It should not contain an excessively large number of experimental points."

Before carrying out this program, we shall have to make a few general remarks on the effects of scaling.

2. A few remarks on scaling

When writing down an equation like (1) we implicitly assume that the units in which $\xi_1, \xi_2, \dots, \xi_k$ are to be measured have been chosen. A change in the units of some or all ξ' will bring about a change in the values of β and Φ, and by choosing these units judiciously one can make the elements of β and of Φ have almost any value one wants -- except for changes of sign. An analogous conclusion holds for the estimators b and F of β and Φ. Obviously a sphere in $(\xi_1, \ldots, \xi_k)$-space will no longer remain a sphere if the unit(s) of at least one of the variables $\xi_1, \ldots, \xi_k$ are changed. Thus it is understandable that certain aspects of the theory of estimation of response surfaces have been criticized as depending too much on scaling. Specifically, it has been frequently asked whether there is any sense in rotatable designs. In rotatable designs the variance of the estimated response

$$y_0 + b'\xi + \xi' F\xi \qquad (5)$$

in the experimental point represented by the vector ξ depends only on the distance $\xi'\xi$ between this experimental point and the origin (cf. Box and Hunter, 1957, p. 204), hence after any change of scale a rotatable design is no longer rotatable. The values of the latent roots of Φ and F, too, are affected by changes of scale, but fortunately, the numbers of negative, zero, and positive latent roots, respectively, are not affected by these changes.

Because of all this it is important to see what scaling really means. There are some differences in the ways in which Box and his associates have handled this matter in different papers.

Part of the paper by Box and Wilson, 1951, is given without mentioning units at all; on page 8 a kind of normalization is introduced with the purpose of being able to compare two designs: "two designs are regarded as of comparable size when they are measured so that the spread for each of the (k) factors (ξ_i) is the same in the two designs,[3] the spread being some variance-type quantity used to judge if

---

[3] Incidentally, at the Ames regional meeting of the Institute of Mathematical Statistics, April, 1950, Le Roy Folks gave an interesting alternative definition of comparable designs: he wanted to compare all designs which have all their points in the region of immediate experimental interest; hence he considered the range of experimental points rather than their spread.

the N different levels of the factor ξ_i are widely different or not (i = 1, ..., k). In Box and Youle, 1955, p. 292, the variables ξ_i are being coded, apparently only "for convenience." In Box and Hunter, 1957, p. 196, standardized levels

$$\xi^*_{iu} = \frac{\xi_{iu} - \bar{\xi}_i}{\tau_i \sqrt{c}} \quad (u = 1, \ldots, N;\ i = 1, \ldots, k) \qquad (6a)$$

are introduced so that

$$\sum_{u=1}^{N} \xi^*_{iu} = 0, \quad \sum_{u=1}^{N} \xi^{*2}_{iu} = N/c \quad (i = 1, \ldots, k) \quad . \qquad (6b)$$

Here ξ_{iu} is the value which the experimental condition represented by ξ_i takes in the uth experimental point (u = 1, ..., N); $\bar{\xi}_i$ is the mean of ξ_{iu} over u: the $\bar{\xi}_i$ (i = 1, ..., k) determine the center of the design, i.e., the center of the region of immediate experimental interest; $\tau_i^2 = \sum_{u=1}^{N} (\xi_{iu} - \bar{\xi}_i)^2/N$; c is an arbitrary constant, independent of i, hence not playing an essential role in our present problem. Of course, not only the experimental points $(\xi_{1u}, \ldots, \xi_{ku})$ (u = 1, ..., N) are transformed as a consequence of (6a), but the whole of $(\xi_1, \ldots, \xi_k)$-space is transformed into $(\xi^*_1, \ldots, \xi^*_k)$-space according to

$$\xi^*_i = \frac{\xi_i - \bar{\xi}_i}{\tau_i \sqrt{c}} \quad (i = 1, \ldots, k) \qquad (6c)$$

Now, on the one hand, as Box and Hunter, 1957, p. 196, write, equation (6a) allows us to derive designs (i.e., sets of N values of the vector $\underline{\xi} = (\xi_1, \ldots, \xi_k)$) which cover the region of immediate interest in any given experimental problem, from one fixed design[4] described in terms of the variables $\xi^*_i, \ldots, \xi^*_k$. This is done by choosing the quantities τ_i and $\bar{\xi}_i$ so that the points $(\xi_{1u}, \ldots, \xi_{ku})$

---

[4] Of course more than one fixed design in $(\xi^*_1, \ldots, \xi^*_k)$-space exists, according to the properties which one desires the design to have, but (5a) can transform each fixed design in $(\xi^*_1, \ldots, \xi^*_k)$-space in an infinite number of ways so that it may serve on an infinite number of regions to be explored in $(\xi_1, \ldots, \xi_k)$-space.

($u = 1, \ldots, N$) cover the region of immediate interest as fully as possible. On the other hand, however, equation (6c) means that designs which are rotatable with respect to ($\xi_1^*, \ldots, \xi_k^*$)-space have the property that the variance of the estimated response (5) is constant if $\sum_{i=1}^{k} \xi_i^{*2}$ = constant, i.e., on the surface of the ellopsoids of the family

$$\sum_{i=1}^{k} \left(\frac{\xi_i - \overline{\xi}_i}{\tau_i \sqrt{c}} \right)^2 = \text{constant}, \qquad (6d)$$

where τ_i and $\overline{\xi}_i$ reportedly are chosen in such a way that the ellipsoid (cf. (6b))

$$\sum_{i=1}^{k} \left(\frac{\xi_i - \overline{\xi}_i}{\tau_i \sqrt{c}} \right)^2 = \frac{N}{c}$$

follows the boundary of the experimentally interesting region as closely as possible. The only result of scaling according to (6a) and (6c) is to adapt the design to the region the experimenter wants to explore; as for the region to be explored, the experimenter has to make up his mind on this point anyhow. In case of rotatable designs the contours of equally exact prediction are thus made to follow the boundary of this region as closely (in a certain sense) as possible -- which does not seem too unreasonable a procedure in a number of contexts. The importance of the region to be explored (or the region of immediate experimental interest) can easily be seen by considering one response surface (for $k = 2$) and two different regions of interest, one of them happening to be elongated along one principal axis of $\underline{\xi}' \Phi \underline{\xi}$ and short along the other axis, the other region happening to be elongated along the second axis and short along the first. Examples of this kind show that in this problem it is not to be taken for granted that scale-invariant procedures are necessarily preferable.

The results to be discussed in the next section are described for a system of coordinates $\xi_1, \ldots, \xi_k$. One may assume if one wishes that the ξ_i stand for

ξ_i^* in the sense of equation (6c).

3. Some properties of the distribution of the latent roots of symmetric random matrices

Part of the proofs of the results contained in this section can be found in van der Vaart, 1958. The other proofs are yet to be published.

The k x k matrix F introduced in section 1 is real, symmetric, has random variables for its elements. We shall assume throughout that the joint distribution of its elements f_{ij} $(1 \leq i \leq j \leq k)$ is continuous in the usual sense (some results will hold under a somewhat weaker assumption, for instance that the latent roots of F are all different with probability one -- which of course they are if the f_{ij} are continuously distributed). Because of the least squares estimation

$$\mathcal{E} f_{ij} = \phi_{ij} \quad (1 \leq i \leq j \leq k). \tag{7}$$

Under these rather weak conditions ($\ell_1 < \ell_2 < \dots < \ell_k$ being the latent roots of F, $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_k$ the latent roots of Φ):

$$\sum_{g=1}^{k} \mathcal{E} \ell_g = \sum_{g=1}^{k} \lambda_g = \sum_{i=1}^{k} \phi_{ii} , \tag{8}$$

$$\mathcal{E} \ell_1 < \lambda_1, \quad \mathcal{E} \ell_k > \lambda_k, \quad \mathcal{E}(\ell_k - \ell_1) > \lambda_k - \lambda_1 \tag{9}$$

If in addition the $\frac{1}{2}k(k+1)$-variable probability density function of the f_{ij} $(1 \leq i \leq j \leq k)$ is symmetrical with respect to the point with coordinates $\phi_{ij} = \mathcal{E} f_{ij}$ $(1 \leq i \leq j \leq k)$, then

$$\text{Med } \ell_1 \leq \lambda_1, \quad \text{Med } \ell_k \geq \lambda_k, \quad \text{Med } (\ell_k - \ell_1) \geq \lambda_k - \lambda_1 \quad . \tag{10}$$

Under certain weak conditions on the distribution of the f_{ij}, which are satisfied for instance by the multinormal distribution, the inequalities (10) are strict.[5]

---

[5] In the abstract No. 41 in Annals Mathematical Statistics 28, 1957, p. 1069 it is erroneously stated that the first two inequalities in (10) hold under the conditions described in the second paragraph of this section.

Evidently a much larger class of distributions of the f_{ij} will allow just one of the inequalities (10) to hold.

Again if the joint distribution of the f_{ij} is continuous and if (7) holds, the equations (11) and (12) obtain

$$\sum_{g=1}^{k} \operatorname{var} \ell_g = \sum_{i=1}^{k} \sum_{j=1}^{k} \operatorname{var} f_{ij} + \sum_{g=1}^{k} \lambda_g^2 - \sum_{g=1}^{k} (\mathcal{E}\ell_g)^2 \tag{11}$$

If k = 2 and the "amount of expectation-bias" α_β is defined by

$$\alpha_\beta = \mathcal{E}(\ell_2 - \lambda_2) = -\mathcal{E}(\ell_1 - \lambda_1) > 0 \tag{11a}$$

then (11) yields

$$\sum_{g=1}^{2} \operatorname{var} \ell_g = \sum_{i=1}^{2} \sum_{j=1}^{2} \operatorname{var} f_{ij} - 2\alpha_\beta^2 - 2(\lambda_2 - \lambda_1)\alpha_\beta \ . \tag{11b}$$

Concerning covariances

$$\sum_{g=1}^{k} \sum_{h=1}^{k} \operatorname{cov}(\ell_g, \ell_h) = \sum_{i=1}^{k} \sum_{j=1}^{k} \operatorname{cov}(f_{ii}, f_{jj}) \ . \tag{12}$$

As a particular case of a general theorem given in van der Vaart, 1958, we have that for k = 2

$$\operatorname{var} \ell_1 = \operatorname{var} \ell_2 \tag{13a}$$

if f_{11}, f_{12}, f_{22} are trinormally distributed with

$$\operatorname{var} f_{11} = \operatorname{var} f_{22},\ \operatorname{cov}(f_{11}, f_{12}) = 0 = \operatorname{cov}(f_{12}, f_{22}); \tag{13b}$$

under this condition (11b) determines $\operatorname{var} \ell_1 = \operatorname{var} \ell_2$ in terms of α_β.

Equations (11) give some information on the variances of the latent roots of F in terms of the variances of the f_{ij}. The existing literature does not elaborate on this point. Box and Youle, 1955, p. 295 state that "appropriate standard errors for these constants (i.e., the ℓ_g) can be shown to be of the same order of magnitude as those of the original quadratic and interaction terms". Box and Hunter, 1957,

p. 240 state that for any rotatable design the variances of the coefficients are the same in every orientation and since the latent roots "are simply the 'quadratic effects' in the directions of the canonical variables they have the same standard errors as have the quadratic effects (i.e., our f_{ii}) before transformation." Equation (11b) shows that if this were true, then for k = 2

$$\alpha_\beta^2 + (\lambda_2-\lambda_1)\alpha_\beta = \text{var } f_{12} \quad , \tag{14}$$

hence

$$\alpha_\beta = \sqrt{\text{var } f_{12}} \quad \text{if} \quad \lambda_2-\lambda_1 = 0 \quad . \tag{14a}$$

That for any distribution of the f_{ij} $(1 \leq i \leq j \leq 2)$ the amount of expectation-bias should depend solely on the latent roots of Φ and on var f_{12}, not on any higher moment, seems unlikely. In case the f_{ij} $(1 \leq i \leq j \leq 2)$ are multinormally distributed our equations (18a) and (18d) below show that (14) and (14a) do not hold generally. The flaw in the argument of Box and Hunter, l.c., seems to be that though it is derived from a statement (at the top of p. 208 of the same paper) to the effect that in rotatable designs every variance and covariance of the coefficients (our f_{ij})......"must remain constant under rotation" of the design of N experimental points, yet there is a difference between the rotation on p. 208 and the one on p. 240: the first is nonrandom, does not depend on the estimate F, the second is random, does depend on F, reduces the off-diagonal elements of F to zero (L is a diagonal matrix). By their argument var f_{12} would equal var 0 = 0.

If the observational error $(y-\eta)$ is normally distributed, the distribution of the f_{ij} $(1 \leq i \leq j \leq k)$ is multinormal. This will be assumed in the remaining part of section 3. Then we can explicitly compute the probability density function of $\ell_1, \ldots, \ell_k$. As the distribution of the f_{ij} depends on the parameters contained in $\Phi = Y\Lambda Y'$, the probability density of $\ell_1, \ldots, \ell_k$ will in general depend both on Λ and Y. In a sense the elements of Y are nuisance parameters. However, they do not occur in the distribution of $\ell_1, \ldots, \ell_k$ if we restrict attention to those distributions of the f_{ij} for which the covariance matrix can be described in terms of two parameters α and β as follows ($\sigma_{ij,pq} = \mathrm{cov}(f_{ij}, f_{pq})$):

$$\left.\begin{array}{ll}
\sigma_{ii,ii} = \frac{\beta + (k-1)\alpha}{\beta(\beta+k\alpha)} & (i = 1, \ldots, k) \\
\sigma_{ii,pp} = \frac{-\alpha}{\beta(\beta+k\alpha)} & (i, p = 1, \ldots, k;\ i \neq p) \\
\sigma_{ij,ij} = \frac{1}{2\beta} & (i, j = 1, \ldots, k;\ i \neq j) \\
\sigma_{ij,pq} = 0 & (i, j, p, q = 1, \ldots, k;\ i \neq j \text{ or} \\
 & \quad p \neq q, \text{ and } i \neq p \text{ or } j \neq q)
\end{array}\right\} \qquad (15)$$

Here

$$a > -\beta/k, \quad \beta > 0 \qquad (15\underline{a})$$

It is interesting that all second order rotatable designs lead to the set of $\mathrm{cov}(f_{ij}, f_{pq})$ satisfying (15). Observe that

a) In our conditions (15) only the quadratic coefficients f_{ij} of our equation (4b) are involved. In Box and Hunter's (1957, p. 213) conditions for second order rotatability the coefficients of the terms of degree zero and one occur as well.

b) When trying to make the two sets of conditions correspond note that $b_{ii} = f_{ii}$, $b_{ij} = 2f_{ij}$ (the b_{ij} are Box and Hunter's notation, the f_{ij} are ours).

c) For k = 2 the covariance matrix (15) satisfies conditions (13b), hence if (15) holds, var ℓ_1 = var ℓ_2.

d) The 1-1-correspondence between our covariance matrix (15) and Box and Hunter's covariance matrix corresponding to second order rotatable designs is described by

$$\beta = 2\lambda \cdot \frac{N}{\sigma^2} \qquad (16\underline{a})$$

$$\lambda = \frac{\beta}{\beta-2\alpha}, \quad \frac{\alpha}{\beta} = \frac{\lambda-1}{2\lambda} \qquad (16\underline{b})$$

for any $k \geq 2$. Here N is the number of experimental points, σ is defined in (2), λ depends on the design of the N experimental points and influences the function by which the variance of the estimated yield in the point $(\xi_1, \ldots, \xi_k)$ depends on the distance between $(\xi_1, \ldots, \xi_k)$ and the center of the design. According to Box and Hunter, 1957, p. 214, values of λ somewhat less than unity (small negative values of α/β) are satisfactory in various respects. If in our formulae below we let $\beta \dashrightarrow \infty$, we assume λ, hence α/β to remain constant:

$$\alpha/\beta \text{ constant, then } \beta \dashrightarrow \infty \text{ corresponds to } N\,\sigma^{-2} \dashrightarrow \infty \qquad (17)$$

We shall now give some results for $k = 2$, f_{ij} multinormally distributed with covariance matrix (15). Here $\mu\sqrt{2} = \lambda_2-\lambda_1$, ${}_1F_1(a;\, c;\, x)$ is the confluent hypergeometric function (cf. Higher transcendental functions, vol. 1, 1953, p. 248), $I_0(x)$ is the modified Bessel function of order zero (cf. Higher transcendental functions, vol. 2, 1953, p. 5).

$$\mathcal{E}(\ell_2-\lambda_2) = -\mathcal{E}(\ell_1-\lambda_1) = -\frac{\mu}{\sqrt{2}} + \frac{1}{2}\sqrt{\frac{\pi}{\beta}}\cdot{}_1F_1\left(-\frac{1}{2};1;\, -\frac{\beta\mu^2}{2}\right) \qquad (18\underline{a})$$

For μ and/or $\beta \dashrightarrow \infty$ we find:

$$-\mathcal{E}(\ell_1-\lambda_1) \sim + \frac{1}{2\sqrt{2}\,\beta\mu} + \frac{1}{8\sqrt{2}\,\beta^2\mu^3} + \frac{3}{16\sqrt{2}}\frac{1}{\beta^3\mu^5} + O\left(\frac{1}{\beta^4\mu^7}\right) \qquad (18\underline{b})$$

Note that the amount of expectation-bias

$$-\mathcal{E}(\ell_1-\lambda_1) = \mathcal{E}(\ell_2-\lambda_2) \begin{cases} \text{is independent of } \alpha, & (18\underline{c}) \\ = \frac{1}{2}\sqrt{\frac{\pi}{\beta}} \text{ if } \mu = 0, & (18\underline{d}) \\ \text{decreases as } \mu \text{ increases}, & (18\underline{e}) \\ \text{decreases as } \beta \text{ increases}, & (18\underline{f}) \\ \text{is } O\left(\frac{1}{\beta\mu}\right) \text{ as } \mu \dashrightarrow \infty \text{ and/or } \beta \dashrightarrow \infty \ (\mu \neq 0) & (18\underline{g}) \end{cases}$$

Furthermore

$$\text{var } \ell_1 = \text{var } \ell_2 = \frac{\beta + \alpha}{\beta(\beta+2\alpha)} + \frac{1}{2\beta} + \frac{\mu^2}{2} - \frac{\pi}{4\beta}\left[{}_1F_1(-\tfrac{1}{2};1;-\tfrac{\beta\mu^2}{2})\right]^2. \qquad (19\underline{a})$$

For μ and/or $\beta \dashrightarrow \infty$ we find

$$\text{var } \ell_1 = \text{var } \ell_2 \sim \frac{\beta + \alpha}{\beta(\beta+2\alpha)} - \frac{1}{4\beta^2\mu^2} - \frac{1}{2\beta^3\mu^4} - O\left(\frac{1}{\beta^4\mu^6}\right). \qquad (19\underline{b})$$

Note that

$$\text{var } \ell_1 = \text{var } \ell_2 \begin{cases} < \text{var } f_{11} = \text{var } f_{22} = \dfrac{\beta+\alpha}{\beta(\beta+2\alpha)}\,, & (19\underline{c}) \\ \left.\begin{aligned} &= \text{var } f_{11} + \text{var } f_{12} - \frac{\pi}{4\beta} = \\ &= \frac{3\beta+4\alpha}{\beta(\beta+2\alpha)} - \frac{\pi}{4\beta} \end{aligned}\right\} \text{ if } \mu = 0, & (19\underline{d}) \\ \sim \text{var } f_{11} = \dfrac{\beta+\alpha}{\beta(\beta+2\alpha)} \text{ if } \mu \text{ \underline{or} } \beta \dashrightarrow \infty\,, & (19\underline{e}) \\ \text{yet } \not\sim \text{var } f_{11} \text{ for certain combinations of } \beta\text{- and } \mu\text{-values} & (19\underline{f}) \end{cases}$$

We ought to remark that we have as yet no quite rigorous proof that (18e), (18f), (19c) hold good for all values of α, β, μ although the evidence available seems pretty strong.

Finally if ω_a and ω_b are two constants and if

$$\omega_a - \lambda_1 = -\chi_a/\sqrt{2},\ \ \omega_b - \lambda_1 = -\chi_b/\sqrt{2},\ \ \chi_a^0 = \chi_a\sqrt{\beta},\ \chi_b^0 = \chi_b\sqrt{\beta}\ ,$$

$$\mu' = \mu\sqrt{\beta} = (\lambda_2 - \lambda_1)\sqrt{\frac{\beta}{2}}\ ,\ \text{then}$$

$$P(\omega_a < \ell_1 < \omega_b) = P\left(-\frac{\chi_a^0}{\sqrt{2\beta}} < \ell_1 - \lambda_1 < -\frac{\chi_b^0}{\sqrt{2\beta}}\right) = \text{(because of (13}\underline{b}\text{))} =$$

$$= P\left(+\frac{\chi_b^0}{\sqrt{2\beta}} < \ell_2 - \lambda_2 < +\frac{\chi_a^0}{\sqrt{2\beta}}\right) =$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\mu'^2} \int_0^\infty e^{-\frac{1}{2}m^2} \cdot m \cdot I_0(\mu' m)\,dm \int_A^B e^{-\frac{1}{2}y^2}\,dy = \qquad (20\underline{a})$$

$$= \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{1}{2}(m-\mu')^2} \cdot m \cdot e^{-\mu' m} \cdot I_0(\mu' m)\,dm \int_A^B e^{-\frac{1}{2}y^2}\,dy, \qquad (20\underline{b})$$

where $A = (m-\mu'-\chi_a^0)\sqrt{\frac{\beta+2\alpha}{\beta}}$, $B = (m-\mu'-\chi_b^0)\sqrt{\frac{\beta+2\alpha}{\beta}}$, (20c)

Applying Hsu, 1951, starting from (20b) or from the expression which follows from (20b) by the substitution $m = \mu' x$, one can obtain an asymptotic result for μ and/or $\beta \to \infty$: lack of time has prevented me from doing so. If $\mu = 0$, the expressions for A and B in (20c) show that for $\mu = 0$

$$P\left(-\frac{\chi_a^0}{\sqrt{2\beta}} < \ell_1 - \lambda_1 < -\frac{\chi_b^0}{\sqrt{2\beta}}\right) = \text{constant if } \chi_a^0 \text{ and } \chi_b^0 \text{ are constant} \qquad (20\underline{d})$$

If $\chi_a^0 = \infty$, $\chi_b^0 = 0$, the result is

$$P(\ell_1 < \lambda_1) = \frac{1}{2} + \frac{1}{2}\sqrt{\frac{2\alpha + \beta}{2\alpha + 2\beta}} \;. \qquad (20\underline{e})$$

4. Discussion of results

Our investigation has been restricted (a) to $k = 2$, (b) to quadratic response surfaces: possible consequences of third degree terms on the effects described here are not considered, (c) to designs satisfying (15) -- which include second order rotatable designs. Our convention (17) entails that $\beta \to \infty$ corresponds to $N \cdot \sigma^{-2} \to \infty$, i.e., mostly to $N \to \infty$ (σ will not be easily changed in the course of one experiment).

Let us first observe that probably var $\ell_1 <$ var f_{11} (cf. (19c)), certainly var $\ell_1 \neq$ var f_{11} (cf. (19a), (19b)). This contradicts Box and Hunter, 1957, p. 240, cf. our section 3. Yet (19e) and (19f) show that under certain conditions var $\ell_1 \sim$ var f_{11} asymptotically.

As to <u>types of bias</u> in the estimation of the latent roots of Φ (the canonical coefficients of the quadratic), our results are much more complete for $\mu = 0$, i.e. $\lambda_1 = \lambda_2$, than for arbitrary μ.

<u>For $\mu = 0$</u> there is considerable expectation-bias (18<u>d</u>) and median-bias ((20<u>e</u>): if $\alpha = 0$, design both rotatable and orthogonal, $P(\ell_1 < \lambda_1) = .85 = P(\ell_2 > \lambda_2)$; for pentagonal designs $P(\ell_1 < \lambda_1) = .75 = P(\ell_2 > \lambda_2)$; upper limit of this probability is 1.00). If $\beta \dashrightarrow \infty$, median-bias remains constant, expectation-bias $\dashrightarrow 0$. Both $-\mathcal{E}(\ell_1 - \lambda_1) = \mathcal{E}(\ell_2 - \lambda_2)$ and var ℓ_1 = var ℓ_2 being asymptotically $\sim \beta^{-1/2}$ (times some constant) suggests that as β increases the scale of the probability distribution of ℓ_g contracts with λ_g as a center ($g = 1, 2$), at the rate of $\beta^{-1/2}$. This is confirmed by (20<u>d</u>).

<u>For $\mu > 0$</u> there is still expectation-bias and median-bias. The expectation-bias appears to decrease with $\mu \dashrightarrow \infty$ and/or $\beta \dashrightarrow \infty$. As we could not yet explore the expressions (20<u>a</u>) and (20<u>b</u>) we do not know for certain how the median-bias varies with μ and β. As now $-\mathcal{E}(\ell_1 - \lambda_1)/\sqrt{\text{var } \ell_1}$ is not asymptotically independent of β, the contraction of the probability distribution will probably be somewhat more intricate than for $\mu = 0$. Considering $\pm\mathcal{E}(\ell_g - \lambda_g)/\sqrt{\text{var } \ell_g}$ ($g = 1, 2$) suggests that for $\mu > 0$ the median-bias might decrease with $\beta \dashrightarrow \infty$ and/or $\mu \dashrightarrow \infty$ (some part of the probability being pulled over to the other side of the λ_g).

One general conclusion from the fact that the ℓ_g are expectation-biased and median-biased estimators of the λ_g is that a considerable part of the probability distribution of $\ell_1(\ell_2)$ is found on $\ell_1 < \lambda_1$ ($\ell_2 > \lambda_2$) (the marginal distributions of the ℓ_g, not cited in this paper, are quite skew). This suggests that if <u>$\lambda_1 > 0$</u> is small its <u>estimate</u> will be <u>negative</u> "<u>too frequently</u>," if <u>$\lambda_2 < 0$</u> is small its <u>estimate</u> will be <u>positive</u> "<u>too frequently</u>." Of course, without a comparing estimator (cf. <u>van der Vaart</u>, 1957, p. 4) the meaning of "too frequently" cannot be made precise. One conclusion is warranted, though. If more observations are made, the above-mentioned contraction of the probability distribution will eventually pull the large bulk of it over the zero point to λ_1 or λ_2, thus decreasing the frequency of erroneous sign estimates. There is some evidence that for large differences $(\lambda_2-\lambda_1)$ the frequency of erroneous sign estimates will be smaller than for small $(\lambda_2-\lambda_1)$; this seems intuitively plausible. If λ_1 happens to be zero the contraction of the probability distribution to λ_1 will pull nothing over the zero point (as for $\lambda_1 > 0$), but it will eventually concentrate a large portion of it so close to zero that the difference becomes materially unimportant (cf. Hodges and Lehmann's, 1954, concept of material significance as opposite to statistical significance). The fact that the frequency of erroneous estimates of the signs of the latent roots can be decreased by taking more observations is gratifying, but counteracts (as long as no better estimators are known) the third requirement on an experimental design for estimating response surfaces, as was laid down by Box and Hunter, 1957, p. 197: "It should not contain an excessively large number of experimental points."

There is experimental evidence for the theoretical results here presented.

1) On p. 548 of the book on design and analysis of industrial experiments, edited by Davies, 1956, an example is given of an analysis which in the first step yields one negative and one smaller positive latent root (besides a quite small negative one). After the accumulation of a larger body of data the positive root is described as "negligibly small."

2) Mason, 1956, p. 94, Fig. 5.6, presents a diagram of contour lines of a fitted second degree response surface which clearly shows a saddle-point (= col = minimax). (As we said before, plotting contour lines and computing latent roots are equivalent with respect to the present problem.) His cautious remark concerning this occurrence is in direct agreement with our theory: "Such a surface would appear to be difficult to interpret agronomically. One would certainly like some substantiation of this

type of pattern before extending its application too far. A more complete sampling by observation points in the critical region is perhaps in order."

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

I wish to thank Dr. R. J. Hader of the Institute of Statistics, North Carolina State College of the University of North Carolina, Raleigh, for his suggestion that it might be interesting to investigate some questions in the theory of response surface estimation during my stay at this Institute. A seminar I gave on the subject at the Department of Statistics of the University of Chicago during my stay there led to plans for computing on high speed machines certain problems which seemed analytically untractable.

RÉSUMÉ

On étudie la loi de probabilité des valeurs propres d'une matrice réelle, symétrique, à éléments aléatoires. L'application des résultats à la théorie de l'estimation des surfaces de régression quadratiques semble expliquer pourquoi les expérimentateurs ont quelquefois pu trouver un col ou minimax au lieu du maximum auquel ils s'attendaient - et d'autres expériences analogues. De plus, on fait quelques remarques sur l'écart moyen centré quadratique des valeurs propres et sur les dispositions dites "rotatable" des observations.

LITERATURE CITED

Bloemena, A. R. (1956). Experimentele bepaling van optimale condities (Overzichts-rapport). Rapport S 204 (Ov 6) van de Statistische Afdeling van het Mathematisch Centrum, Amsterdam (Netherlands).

Box, G. E. P. (1954). Biometrics 10, 16-60.

Box, G. E. P. and Hunter, J. S. (1957). Ann. Math. Stat. 28, 195-241.

Box, G. E. P. and Wilson, K. B. (1951). Jour. Roy. Stat. Soc., Ser. B, 13, 1-45.

Box, G. E. P. and Youle, P. V. (1955). Biometrics 11, 287-323.

Cochran, W. G. and Cox, G. M. (1957). Experimental Designs, 2nd ed., New York, Wiley, and London, Chapman and Hall, xiv + 611 p.

Davies, O. L. (Editor) (1956). The design and analysis of industrial experiments; published for Imp. Chem. Ind. Lim. by Oliver and Boyd, London and Edinburgh, and Hafner, New York, 2nd ed., xiii + 636 p.

Higher Transcendental Functions, (Bateman Manuscript Project), Vol. 1, 1953, xxvi + 302 p., Vol. 2, 1953, xvii + 396 p., McGraw Hill.

Hodges, J. L. and Lehmann, E. L. (1954). Jour. Roy. Stat. Soc., Ser. B, 16, 261-268.

Hsu, L. C. (1951). Amer. Jour. Math. 73, 625-634.

Mason, D. D. (1956). Functional models and experimental designs for characterizing response curves and surfaces, p. 76-98 in: Methodological procedures in the economic analysis of fertilizer use data (based on a symposium held in 1955 and edited by Baum, E. L., Heady, E. O., Blackmore, J.). Iowa State College Press, xx + 218 p.

Vaart, H. R. van der (1957). Some general remarks on the definition of the concept of biased estimators. Institute of Statistics Mimeo. Series No. 178, 18 p.

Vaart, H. R. van der (1958). Some results on the probability distribution of the latent roots of a symmetric matrix of continuously distributed elements, and some applications to the theory of response surface estimation. Institute of Statistics Mimeo. Series No. 189, 1 + 40 + 2 p.