References

[1] T. W. Anderson, *An introduction to multivariate statistical analysis*, Wiley, New York 1958.
[2] M. S. Bartlett, *Properties of sufficiency and statistical tests*, Proc. Roy. Soc. London, series A, 160 (1937), pp. 268–282.
[3] L. J. Gleser, *A note on the sphericity test*, Ann. Math. Statist. 37 (1966), pp. 464–467.
[4] E. J. G. Pitman, *Tests of hypothesis concerning location and scale parameters*, Biometrika 31 (1939), pp. 200–215.
[5] E. Spjøtvoll, *Unbiasedness of likelihood ratio confidence sets in cases without nuisance parameters*, J. R. Statist. Soc. B., 34 (1971), pp. 268–273.
[6] N. Sugiura, and H. Nagao, *Unbiasedness of some test criteria for the equality of one or two covariance matrices*, Ann. Math. Statist. 39 (1968), pp. 1686–1692.

Presented at the semester
MATHEMATICAL STATISTICS
September 15–December 18, 1976

MATHEMATICAL STATISTICS
BANACH CENTER PUBLICATIONS, VOLUME 6
PWN—POLISH SCIENTIFIC PUBLISHERS
WARSAW 1980

SOME BASIC RESULTS ON SEARCH LINEAR MODELS WITH NUISANCE PARAMETERS*

J. N. SRIVASTAVA

Colorado State University, Fort Collins, Colo., U.S.A.

Summary

Search linear models, introduced in Srivastava [1] (1975), are now well known. Consider the model

(1) $$E(\mathbf{y}) = A_1\boldsymbol{\xi}_1+A_2\boldsymbol{\xi}_2,\quad V(\mathbf{y}) = \sigma^2 I_N,$$

where $\mathbf{y}(N\times 1)$ is a vector of observations, $\mathbf{A}_1(N\times \nu_1)$, and $\mathbf{A}_2(N\times \nu_2)$, are known matrices, $\boldsymbol{\xi}_1(\nu_1\times 1)$, and $\boldsymbol{\xi}_2(\nu_2\times 1)$ are vectors of unknown parameters, and σ^2 is a known or unknown constant. On $\boldsymbol{\xi}_2$, we have the following partial information. It is known that at most k elements of $\boldsymbol{\xi}_2$ are non-negligible. The problem is to search the non-negligible elements of $\boldsymbol{\xi}_2$ and draw inferences on them as well as on the elements of $\boldsymbol{\xi}_1$. For work on the design or inference aspects of this problem, see the references at the end. In this paper, we consider a variation of the above problem. We assume that interest lies in estimating the fixed set of parameters $\boldsymbol{\xi}_1$ alone. In other words, although some elements of $\boldsymbol{\xi}_2$ are non-negligible, we do not need to search or estimate them. This problem is very important since in many applications a solution to this problem may be considered adequate. We prove fundamental results concerning this new problem under a model which is more general than (1).

1. Introduction

Search linear models of the type (1) are well known. They were introduced in Srivastava [1], where the noiseless case (i.e. when $\sigma^2 = 0$) was specially developed. Of course, in all statistical problems, some noise is present (i.e. $\sigma^2 > 0$). However, it is clear that if we cannot do the search correctly, or estimate the parameters 'precisely' (i.e. with variance zero) in the noiseless case, there is no hope of doing so when noise is present. Indeed, as is elaborated in the papers of the author on the

* This research was supported by the National Science Foundation Grant No. MCS76-23282.

subject, a sample of which is listed at the end for illustration, the design aspect is connected largely with the noiseless case. Broadly speaking, a situation is said to correspond to a *search design* if the structure of the observations (and hence of the matrices A_1 and A_2) is such that, in the noiseless case, all the parameters of interest can be estimated precisely. Thus, a search design corresponds to the case where, essentially, the parameters of interest are estimable (i.e. possess unbiased estimates).

In Srivastava [1], the following fundamental theorem on search designs was proved. (We assume k is known.)

THEOREM 1. *A necessary and sufficient condition that under* (1), *in the noiseless case, the non-negligible parameters in* $\boldsymbol{\xi}_2$ *can be searched correctly, and these and the elements of* $\boldsymbol{\xi}_1$ *can be estimated precisely is that we have*

$$\text{(2)} \qquad \operatorname{Rank}(A_1 \vdots A_{20}) = \nu_1 + 2k,$$

for every $(N\times 2k)$ *submatrix* A_{20} *of* A_2.

Thus, (2) gives a necessary and sufficient condition that the model (1) corresponds to a search design. In Srivastava [1], the subject of search designs was further developed in the direction of application of the model (1) to factorial experiments. Problems concerning the development of search designs from a given design (which is not a search design) by the possible addition of future observations, and certain sequential or rather multi-stage design concepts, were considered in Srivastava [2] (1976). Factorial designs where (1) is applicable with $\boldsymbol{\xi}_1$ corresponding to the general mean, main effects, and two-factor interactions, and $\boldsymbol{\xi}_2$ to the remaining factorial effects, and where $k = 1$, were developed in Srivastava and Ghosh [3] (1977). It has been pointed out in these and other papers that the real life situations where linear models are often used, are usually such that the search linear model actually holds there, and should be employed. This also indicates that the usual theory of optimal designs is really not bias-free, since it is based on ordinary linear models, which correspond to the search linear model with the *assumption* that $k = 0$. It may be argued that usually k will not be known; however, it is intuitively clear that using search designs with $k > 0$, and thus taking into account at least some non-negligible parameters out of $\boldsymbol{\xi}_2$, should be better than assuming $k = 0$. Some new concepts relating to a theory of 'bias-free' optimal designs were introduced and developed in Srivastava [4] (1977).

The inference aspect of search models (for $\sigma^2 > 0$) is studied in Srivastava [1], and Srivastava and Mallenby [5] (1977). In this paper, we consider the situation where, under (1), interest lies only in $\boldsymbol{\xi}_1$. This is studied under a more general model.

2. Search linear model with nuisance parameters

Consider the model

$$\text{(3)} \qquad E(\mathbf{y}) = A_1\boldsymbol{\xi}_1 + A_2\boldsymbol{\xi}_2 + A_3\boldsymbol{\xi}_3, \quad V(\mathbf{y}) = \sigma^2 I_N,$$

where $\mathbf{y}$, A_1, A_2, $\boldsymbol{\xi}_1$, $\boldsymbol{\xi}_2$, and σ^2 are as in (1), $A_3(N\times\nu_3)$ is a known matrix, and $\boldsymbol{\xi}_3(\nu_3\times 1)$ is a vector of unknown parameters. We now assume that interest lies in the estimation of (or, other inference problems concerning) *only* the ν_1 parameters $\boldsymbol{\xi}_1$. Under the above assumptions, the model (3) will be called a *search linear model with nuisance parameters of type* I. There is a variation of the above model, namely *type* II. In a type II model, the elements of $\boldsymbol{\xi}_1$ and the non-negligible elements of $\boldsymbol{\xi}_2$ are of interest. Other seeming variations of the above model are essentially special cases of these two models. It is clear that these nuisance parameter models also are of wide applicability. Block-treatment designs are, for example, a large area of application. We now establish some basic results. We assume k known.

THEOREM 2. *Consider a type* I *model of the form* (3), *and assume* $\sigma^2 = 0$. *A necessary and sufficient condition that the model corresponds to a search design* (*i.e.* $\boldsymbol{\xi}_1$ *is estimable precisely*) *is that both conditions* (a) *and* (b) *below are satisfied.*

$$\text{(4)} \qquad \begin{aligned} &\text{(a) } \operatorname{Rank}(A_1) = \nu_1, \\ &\text{(b) } \operatorname{Rank}(A_1 \vdots A_{20} \vdots A_3) = \operatorname{Rank}(A_1) + \operatorname{Rank}(A_{20} \vdots A_3) \end{aligned}$$

for every $(N\times 2k)$ *submatrix* A_{20} *of* A_2.

Proof. (i) *Necessity.* Condition (4a) is clearly necessary, for otherwise even under an ordinary linear model (where $\nu_2 = \nu_3 = 0$), the parameters $\boldsymbol{\xi}_1$ will not be estimable. Now, suppose (4b) does not hold. Then there exist vectors $\boldsymbol{\theta}_1(\nu_1\times 1)$, $\boldsymbol{\theta}_2(2k\times 1)$, and $\boldsymbol{\theta}_3(\nu_3\times 1)$, and a submatrix $A_{20}(N\times 2k)$ of A_2 such that $\boldsymbol{\theta}_1 \neq \mathbf{0}$, and $(\boldsymbol{\theta}_2, \boldsymbol{\theta}_3) \neq \mathbf{0}$, and

$$\text{(5)} \qquad A_1\boldsymbol{\theta}_1 + A_{20}\boldsymbol{\theta}_2 + A_3\boldsymbol{\theta}_3 = \mathbf{0},$$

where the various zero-vectors (denoted by $\mathbf{0}$) are of appropriate sizes. Let $A_{20} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{2k})$, and let the columns $\boldsymbol{\alpha}_j(N\times 1)$, $(j = 1, \ldots, 2k)$, of A_{20} correspond respectively to the elements $\xi_{2,i_1}, \ldots, \xi_{2,i_{2k}}$ of $\boldsymbol{\xi}_2$. Let $\boldsymbol{\theta}_2' = (\theta_1^2, \ldots, \theta_{2k}^2)$. Then, if $\boldsymbol{\theta}_1^*(\nu_1\times 1)$, $\boldsymbol{\theta}_3^*(\nu_3\times 1)$ are *any* real vectors, it follows from (5) that

$$\text{(6)} \qquad \begin{aligned} &A_1\boldsymbol{\theta}_1^* + A_3\boldsymbol{\theta}_3^* + (-\theta_1^2)\boldsymbol{\alpha}_1 + \ldots + (-\theta_k^2)\boldsymbol{\alpha}_k \\ &\quad = A_1(\boldsymbol{\theta}_1 + \boldsymbol{\theta}_1^*) + A_3(\boldsymbol{\theta}_3 + \boldsymbol{\theta}_3^*) + \theta_{k+1}^2\boldsymbol{\alpha}_{k+1} + \ldots + \theta_{2k}^2\boldsymbol{\alpha}_{2k}. \end{aligned}$$

This shows that the situation where we have $\boldsymbol{\xi}_1 = \boldsymbol{\theta}_1^*$, $\boldsymbol{\xi}_3 = \boldsymbol{\theta}_3^*$ and the elements of $\boldsymbol{\xi}_2$ are all negligible except possibly for $\xi_{2,i_1}, \ldots, \xi_{2,i_k}$ which have values $\theta_1^2, \ldots, \theta_k^2$ respectively, is *not distinguishable* from the one where $\boldsymbol{\xi}_1 = \boldsymbol{\theta}_1 + \boldsymbol{\theta}_1^*$, $\boldsymbol{\xi}_3 = \boldsymbol{\theta}_3 + \boldsymbol{\theta}_3^*$ and the (possibly) non-zero element of $\boldsymbol{\xi}_2$ are $\xi_{2,i_{k+1}}, \ldots, \xi_{2,i_{2k}}$ with values $\theta_{k+1}^2, \ldots, \theta_{2k}^2$, respectively, in the sense that both will give rise to the same value of $\mathbf{y}$ (or of $E(\mathbf{y})$, when $\sigma^2 > 0$). Now, in one of these situations, we have $\boldsymbol{\xi}_1 = \boldsymbol{\theta}_1^*$, and in the other $\boldsymbol{\xi}_1 = \boldsymbol{\theta}_1 + \boldsymbol{\theta}_1^*$. Since $\boldsymbol{\theta}_1 \neq \mathbf{0}$, these two values of $\boldsymbol{\xi}_1$ are different. Hence, if (4b) does not hold, the true value of $\boldsymbol{\xi}_1$ cannot be determined. Thus, conditions (4a, b) are necessary.

(ii) *Sufficiency.* We claim that if (4) holds under the model (3), then $\mathbf{y}$ has a *unique* representation in terms of the columns of A_1, A_3, and a set of *at most* k columns of A_2. For if there are two such representations, then it is clear that a condition like (6) will hold, contradicting (4). On the other hand, the unique representation of $\mathbf{y}$ in terms of the above columns will clearly give the value of ξ_1, whose elements will just be the coefficient of the columns of A_1 in this representation. This completes the proof.

THEOREM 3. *Consider a type* II *model of the form* (3), *with* $\sigma^2 = 0$. *A necessary and sufficient condition that the model corresponds to a search design (i.e. the non-negligible elements of* ξ_2 *can be searched correctly, and these and the elements of* ξ_1 *can be estimated precisely) are that both conditions* (a) *and* (b) *below be satisfied*:

(7) (a) $\operatorname{Rank}(A_1 \vdots A_{20}) = \nu_1 + 2k$,

(b) $\operatorname{Rank}(A_1 \vdots A_{20} \vdots A_3) = \operatorname{Rank}(A_1 \vdots A_{20}) + \operatorname{Rank}(A_3)$

for every $(N\times 2k)$ *submatrix* A_{20} *of* A_2.

Proof. The proof of this result follows on lines similar to that of the last one, and will be omitted for brevity.

3. Search and estimation techniques

We now present some simple techniques for dealing with search and/or estimation problems arising under model (3). First, we introduce some notation. For any matrix M, the matrix M^* will denote *a* conditional inverse of M; note that M^* is not necessarily unique, and has merely to satisfy $MM^*M = M$. Also, $G(M)$ will denote *the* matrix $[I_p - M(M'M)^*M']$, where M is of size $(p\times q)$, and I_p is the $(p\times p)$ identity matrix. Note that, given M, the matrix $G(M)$ is uniquely determined, and is symmetric and idempotent. The following result can easily be established.

THEOREM 4. *Let* $P(m\times p)$ *and* $Q(m\times q)$ *be two matrices over the real field with* $m \geqslant p, q$. *Then the following three conditions are equivalent*:

(8a) $\operatorname{Rank}(Q) = q$, and $\operatorname{Rank}(P \vdots Q) = \operatorname{Rank}(P) + \operatorname{Rank}(Q)$,

(8b) $\operatorname{Rank}\{[G(P)]Q\} = q$,

(8c) *The matrix* $Q'[G(P)]Q$ *is non-singular.*

THEOREM 5. *Consider a type* I *model of the form* (3) *with* $\sigma^2 = 0$. *Suppose* (4) *are satisfied. Let*

$$\mathbf{y}^{13} = [G(A_1 \vdots A_3)]\mathbf{y}. \tag{9}$$

(i) *Then there exists a* $(N\times k_1)$ *submatrix* A_{21} *of* A_2, *and a* $(k_1\times 1)$ *vector* $\mathbf{u}$, *with* $k_1 \leqslant k$, *such that*

$$\mathbf{y}^{13} = [G(A_1 \vdots A_3)]A_{21}\mathbf{u}. \tag{10}$$

(ii) *The matrix* $(A_1'[G(A_{21} \vdots A_3)]A_1)$ *is symmetric and of full rank.*

(iii) *The equation*

$$\xi_1 = (A_1'[G(A_{21} \vdots A_3)]A_1)^{-1}(A_1'[G(A_{21} \vdots A_3)])\mathbf{y}, \tag{11}$$

holds, so that ξ_1 *can be computed using it.*

Proof. It is well known that for any matrix M, we have $[G(M)]M = 0$. Since $\sigma^2 = 0$, we have $E(\mathbf{y}) = \mathbf{y}$. Now, we know that $\mathbf{y}$ has the form as at (3), where ξ_1 is unique, ξ_3 is not necessarily unique, and where for some ξ_3 (say $\xi_3 = \xi_3^*$), there exists a value of ξ_2 (say ξ_2^*) such that *at least* $(N-k)$ elements of ξ_2^* are zero. Let the nonzero elements of ξ_2^* be k_2 in number (where $k_2 \leqslant k$), and let these be denoted by $\theta_1, \theta_2, \ldots, \theta_{k_2}$; let the k_2 columns of A_2 corresponding to these be denoted by $\alpha_1^*, \ldots, \alpha_{k_2}^*$; finally, let $\phi^{*\prime} = (\varphi_1, \ldots, \varphi_{k_2})$, and $A_{21}^* = [\alpha_1^*, \ldots, \alpha_{k_2}^*]$. Then

$$\mathbf{y} = A_1\xi_1 + A_3\xi_3^* + A_{21}^*\phi^*, \tag{12}$$

so that

$$\mathbf{y}^{13} = [G(A_1 \vdots A_3)]\mathbf{y} = [G(A_1 \vdots A_3)]A_{21}^*\phi^*, \tag{13}$$

which satisfies (10) with $A_{21} = A_{21}^*$, and $\mathbf{u} = \phi^*$, and $k_1 = k_2$. This proves (i).

Part (ii) of the theorem follows from part (i), condition (4), and Theorem 4, part (c). To prove (iii), observe that in view of (12), the right-hand side of (9) equals

$$\{A_1'[G(A_{21} \vdots A_3)]A_1\}^{-1}\{A_1'[G(A_{21} \vdots A_3)]\}\{A_1\xi_1 + A_3\xi_3^* + A_{21}^*\phi^*\} = \xi_1 + \zeta, \tag{14}$$

say, where

$$\zeta = LA_1'[G(A_{21} \vdots A_3)]A_{21}^*\phi^*, \tag{15a}$$

$$L = \{A_1'[G(A_{21} \vdots A_3)]A_1\}^{-1}. \tag{15b}$$

Notice that in (14) and (15) we are stressing the fact that ξ_3 and ξ_2 are not necessarily unique, so that the A_{21} and $\mathbf{u}$ occurring in (10) may be distinct from A_{21}^* and ϕ^*. However, in view of part (i), we must have the relationship

$$[G(A_1 \vdots A_3)]A_{21}\mathbf{u} = [G(A_1 \vdots A_3)]A_{21}^*\phi^*, \tag{16}$$

since $G(A_1 \vdots A_3)$, and hence $\mathbf{y}^{13}$, are unique. But, (16) and (4) together imply that $(A_{21}\mathbf{u} - A_{21}^*\phi^*)$ must belong to the column space of A_3. Thus, there exists $\mathbf{w}(\nu_3\times 1)$ such that

$$A_{21}^*\phi^* = A_{21}\mathbf{u} + A_3\mathbf{w}. \tag{17}$$

From (15a) and (17), we find that $\zeta = \mathbf{0}$. This completes the proof of the theorem.

In view of the above, the following procedure for determining ξ_1 may be used in the noiseless case. Obtain $\mathbf{y}^{13}$ and, by some procedure (including trial and error), obtain A_{21} and $\mathbf{u}$ (which may not be unique). Finally, obtain ξ_1 from (11). The same procedure can be used in the noisy case $(\sigma^2 > 0)$, except that to determine A_{21} and $\mathbf{u}$, some kind of least squares may be used; notice that in this case the right-hand side of (11) will not be unique, as expected.

THEOREM 6. *Consider a type* II *model of the form* (3), *with* $\sigma^2 = 0$. *Suppose* (5) *is satisfied. Let*

$$\mathbf{y}^3 = [G(A_3)]\mathbf{y}. \tag{18}$$

Then the following ordinary search linear model of the form (1) *holds*:

(19) $$\mathbf{y}^4 = \{[G(A_3)]\}\xi_1 + \{[G(A_3)]A_2\}\xi_2.$$

Furthermore, (19) *can be used to determine* ξ_1 *and* ξ_2 *as under ordinary search linear models.*

Proof. Equation (19) is obvious in view of (18) and (3). To show that $\mathbf{y}^4$ at (19) has the structure of a search design, we have to show that conditions corresponding to (2) in Theorem 1 hold; thus we need to show that

(20) $$\operatorname{Rank}\{[G(A_3)]A_1 \vdots [G(A_3)]A_{20}\} = \operatorname{Rank}\{[G(A_3)]A_1\} + \operatorname{Rank}\{[G(A_3)]A_{20}\},$$

for all $(N\times 2k)$ submatrices A_{20} contained in A_2. From (7b), (8b), and (7a), taking $Q = A_3$, and $P = [A_1 \vdots A_{20}]$, we find that the l.h.s. of (20) equals $\nu_1 + 2k$. Similarly, the two terms on the r.h.s. of (20) are respectively ν_1 and $2k$. This completes the proof of the theorem.

We close the paper by recalling, for the sake of completeness, one procedure for search and estimation under the ordinary search linear model (1) when conditions (2) hold. We first compute $[G(A_1)]\mathbf{y} = \mathbf{y}^1$, say. Clearly,

(21) $$E(\mathbf{y}^1) = \{[G(A_1)]A_2\}\xi_2,$$

where in view of (2) and (8b), we have Rank $[G(A_1)]A_{20} = 2k$, for all $(N\times 2k)$ submatrices A_{20} of A_2. Then, we project $\mathbf{y}'$ on the sets of k columns of $[G(A_1)]A_2$, until (in the noiseless case) we obtain a set of k columns of A_2 which gives a perfect fit. In the noisy case, ordinary least squares projection may be used. Notice that the technique mentioned in this paragraph is essentially equivalent to method I in Srivastava [1].

References

[1] J. N. Srivastava, *Designs for searching non-negligible effects*, in: *A survey of statistica design and linear models*, ed. by J. N. Srivastava, North-Holland Publ. Company, Inc. New York 1975, pp. 507–519.

[2] —, *Some further theory of search linear models*, in: *Contribution to Applied statistics*, publ. by the Swiss-Australian Region of the Biometry Society, 1976, pp. 249–256.

[3] J. N. Srivastava and S. Ghosh, *Balanced* 2^m *factorial design of resolution V which allow search and estimation of one extra unknown effect* $4 \leqslant m \leqslant 8$, Comm. Statist. — Theor. Meth. A6 (1977), pp. 141–166.

[4] J. N. Srivastava, *Optimal search designs, or designs optimal under bias-free optimality criteria*, in: *Statistical decision theory and related topics, II*, ed. by S. C. Gupta and D. S. Moore, 1977, pp. 375–409.

[5] J. N. Srivastava and D. W. Mallenby, *Some studies on a new method of search in search linear models* (submitted for publication).

Presented to the semester
MATHEMATICAL STATISTICS
September 15–December 18, 1976

MATHEMATICAL STATISTICS
BANACH CENTER PUBLICATIONS, VOLUME 6
PWN—POLISH SCIENTIFIC PUBLISHERS
WARSAW 1980

DEVIATIONS FROM TOTAL INFORMATION AND FROM TOTAL IGNORANCE AS MEASURES OF INFORMATION

ERIK N. TORGERSEN

University of Oslo, Institute of Mathematics, Oslo, Norway

1. Introduction, notations and basic facts

Many interesting possibilities of quantifying the content of information in a statistical experiment have been proposed and studied in the literature. Among the most prominent are Fisher information and Kullback–Leibler information numbers. Several of the principles for comparing designs of experiments are based on ideas of measuring information. Most of the quantifications are designed for particular problems. It is, therefore, not surprising that comparison by different measures may lead to conflicting results. There is, of course, no hope to remedy this and no single real valued quantity is likely to qualify as "the information number". Any measure is bound to be useful within limited scopes only. The particular measures which I shall shortly describe are not exceptions — on the contrary they might even appear quite artificial. I find them more interesting because of their construction than because of their usefulness in concrete applications.

Before proceeding let me at once remark that limitations of time as well as on space, force me to present most of our results without proofs. Anyone interested will find proofs and other information on the subject in [14].

Our point of departure shall be the view of statistical decision theory, i.e. that the performance of a decision procedure is to be judged on the basis of the risk it incures. In order to give precise definitions, let us agree that a statistical experiment $\mathscr{E}$ with parameter set Θ is a family $(P_\theta;\theta\in\Theta)$ of probability measures on a common measurable space, say $(\chi, \mathscr{A})$. We may then write:

$$E = (\chi, A; P_\theta; \theta\in\Theta) = (P_\theta; \theta\in\Theta).$$

It is often convenient to identify experiments with the random variables defining them. Thus, if our observation X is χ-valued and $\mathscr{A}$-measurable and the distribution of X under θ is P_θ, then $\mathscr{E}$ may be considered as the experiment obtained by observing X.

If $\mathscr{E}_i = (P_{\theta,i};\theta\in\Theta)$, $i = 1,\ldots,n$, are experiments, then their product is the experiment $\left(\prod_{i=1}^n P_{\theta,i};\theta\in\Theta\right)$ and we shall use notations as $\mathscr{E}_1\times\ldots\times\mathscr{E}_n$ or $\prod_{i=1}^n \mathscr{E}_i$