The Annals of Statistics
1984, Vol. 12, No. 2, 416–423

JACK KIEFER'S CONTRIBUTIONS TO EXPERIMENTAL DESIGN

BY HENRY P. WYNN
Imperial College, London

1. History. Careful experimentation is part and parcel of the scientific method developed in the eighteenth and nineteenth century. John Stuart Mill was probably the first to give clear prescriptions on how to carry out experiments. He separated experiments into "spontaneous" experiments, what we would now call observational studies, and "artificial" experiments, namely controlled experiments. Mill and others were firmly of the belief that controlled experimentation was better, if the subject matter allowed it. This was carried through into this century with the "crucial experiment" becoming the cornerstone of the falsification ideas of Karl Popper and his followers. The details of experimental strategy, however, were neglected by the philosophers, except that it was recognized that careful variation in the levels of "agents" A, B, C, $\cdots$ would yield an analysis of their effects a, b, c, $\cdots$.

The breakthrough into a more versatile approach to experimental design came with the work of Ronald A. Fisher and his followers, notably Frank Yates, at the Rothamstead Experimental Station in England. A number of useful concepts were introduced such as balance, orthogonality, blocking and aliasing. This led to an explosion of work on combinatorial design which took seed in the USA through the work of Raj Chandra Bose and collaborators.

Here and there in the combinatorial literature the idea of efficiency—usually relative to some standard design—had been discussed. However, at the end of the second world war the theory of optimum design was almost nonexistent except for a remarkable early paper by Smith (1918) and the important paper of Wald (1943). It is no accident that the modern theory of optimum design has its roots in the decision theory school of U.S. statistics founded by Abraham Wald. The idea of "risk," developed formally by Wald and arising out of the earlier work of Neyman and Pearson, was the most important innovation of that school. There were parallel developments in utility theory, mathematical programming and mathematical economics so that the early history of the subjects were interwoven. Together with Wald, Jacob Wolfowitz and Jack Kiefer were leading members of this school. They started the second great advance in the science of experimentation in this century by applying decision-theoretic ideas, and over the subsequent twenty years Jack Kiefer himself nurtured this science to maturity.

2. Continuous theory. Since Wald's paper, a number of papers had appeared by Elfving (1952), Hoel (1958), Guest (1958) and important work by Box and Draper and co-workers on response surface design. Thus, the literature

Received December 1983.

leading up to the first Kiefer-Wolfowitz paper [23] had gradually liberated the allowable region of experimentation to grow from one-at-a-time methods through combinatorial design to multifactor experimentation and response surface design. It had reached the point where Kiefer and Wolfowitz could allow an almost arbitrary design region $\mathscr{X}$ in the same way that decision theory had earlier allowed an almost arbitrary action space. The other brave step, technical rather than conceptual, was to abolish, in a stroke, computational difficulties involved in changing the sample size. Thus, a design became a probability measure ξ over the design space $\mathscr{X}$, and a "continuous theory" was born.

It is worth restating briefly the basic Kiefer-Wolfowitz setup. At each point in $\mathscr{X}$ there is a potential observation Y_x whose expectation is

$$E(Y_x) = \sum_{i=1}^{k} \theta_i f_i(x) = \theta^T f(x),$$

where $\theta = (\theta_1, \cdots, \theta_k)^T$ are unknown parameters and the f_i's are continuous functions on $\mathscr{X}$, which is taken to be compact. For an (exact) experiment, observations $Y_{x_1}, \cdots, Y_{x_N}$ are taken and assumed to be uncorrelated with equal variance σ^2. The $k \times k$ information matrix is $X^T X$, where $X^T = [f(x_1), \cdots, f(x_N)]$ and $\text{cov}(\hat{\theta}) = \sigma^2 (X^T X)^{-1}$ is the covariance matrix of the least squares estimate of θ based on the observations.

The normalized version of $X^T X$ namely, $(1/N) X^T X$, generalizes to the moment matrix

$$M(\xi) = \int_{\mathscr{X}} f(x) f(x)^T \xi(dx).$$

A key mathematical benefit of this approach is that the set of all $M(\xi)$, the moment space, is closed and convex. In addition, many of the optimality criteria which had been introduced in earlier work, when extended to $M(\xi)$, involved minimization of a convex functional. The most important of these was D-optimality (introduced by Wald):

$$\min_\xi \{-\log \det M(\xi)\}.$$

This is one of a wider class of Φ_p criteria introduced by Kiefer in later work:

$$\min_\xi [\text{trace}(M(\xi)^{-p})]^{1/p}.$$

$-1 \leq p \leq \infty$. The case $p = 0$ gives D-optimality, $p = \infty$, the so-called E-optimality (minimizing the maximum eigenvalue of $M(\xi)^{-1}$), and $p = 1$, A-optimality which had been studied mostly in block design settings. The power of the extension to measures is demonstrated by the proof of the beautiful General Equivalence Theorem (GET) [29]. This showed that D-optimality was equivalent to G-optimality which achieves

$$\min_\xi \max_{x \in \mathscr{X}} f(x)^T M(\xi)^{-1} f(x).$$

The quantity $f(x)^T M(\xi)^{-1} f(x)$ is the generalization of the (normalized) variance function, the variance of the estimated response $\hat{Y}_x$.

Casting the problem as a convex program brought it into the arena with

Lagrangian theory, game theory and minimax/saddle point theory. It was clear then that more general results than the GET could be established. Jack Kiefer tied up much of this in [58] for a general smooth class of optimality criteria. Earlier, Karlin and Studden had given a game theoretical proof of the GET. More recently, Pukelsheim (1980) has used the Fenchel duality theorem to achieve duality theorems for Φ_p optimality following an emphasis on the duality approach by Silvey and Titterington (1973). Thus, the core of continuous optimum design can now be seen to have an extra existence as a fascinating subculture of optimization theory.

The ability of the equivalence theorem to throw up rich examples led to pioneering work by Jack Kiefer and his co-workers. The connection with orthogonal polynomials proved vital to this analysis. This arose because of the simple observation that the (generalized) variance function satisfied, for polynomial models,

$$d(x, \xi) = \sum_{i=1}^{k} \phi_i^2(x),$$

where $\phi_1, \cdots, \phi_k$ are orthonormal polynomials with respect to the design. The technique was to guess at a nice class of symmetric-looking designs possibly using the invariance of D-optimality under linear transformations of the parameters. The class would be defined up to unknown weights $\alpha, \beta, \gamma, \cdots$ at certain support points. A general expression for the orthonormal polynomials would be found and the optimal α, β, γ calculated using two additional features of the GET: (i) the minmax value of $d(x, \xi)$ is k (the number of parameters) and (ii) $\max_x d(x, \xi)$ is achieved at a support point for optimal ξ. This technique proved very successful for quadratic regression on a simplex and hypercube ([31], [46]). It transpired that the classical orthogonal functions were not necessarily optimum but, in a mysterious way, provided the support for the optimum designs. A beautiful paper by Kiefer and Studden [64] exploited the classical theory of orthogonal polynomials to find limiting designs for large k. After such analytic methods became difficult (as for example when $\mathscr{X}$ was a sphere), Kiefer and Galil used a mixture of analysis, search and direct computation to find solutions, as in [81].

One particularly difficult area computationally is the so-called singular case which arises in the extension of D-optimality (and Φ_p-optimality) to subsets of parameters. The difficulty arises basically from the fact that, although a chosen vector of parameters $\alpha = B\theta$ may be estimable under a particular design (and indeed at the optimum), so that $BM(\xi)B^T$ is nonsingular, $M(\xi)$ itself may be singular. This leads to technical problems in the specification of a solution. A discussion of this case with some history appears in the recent book by Silvey (1980).

3. Exact theory. Despite the success of the continuous theory, Jack Kiefer always had in mind that it was important for the development of the subject to solve outstanding problems in the exact theory: that is, to find optimum designs for fixed sample size or under more rigid combinatorial restrictions. Indeed, his very first paper on design [22] established under mild conditions the optimality in a wide sense of incomplete block designs and Latin squares. He made use of a delightfully simple but very useful lemma (Proposition 1, [61]) for symmetric matrices.

In block design settings, or more general m-way layouts, one often works with the so-called C-matrix, written C_d, where d refers to the design. This is the information matrix for estimating treatment contrasts. If there are v treatments (v treatment parameters) C_d will be a $v \times v$ matrix but have maximal rank $v - 1$. The lemma says that C_d, and hence d, is optimal if, within the class of all C_d, (a) it is completely symmetric (that is, all diagonal elements equal and all off diagonal elements equal), (b) it has maximum trace. Optimal here means "universally optimal": that is, it maximizes any convex, nonincreasing, permutation invariant function on C_d. It therefore includes A-, E- and D-optimality. This was the starting point for two main lines of research.

The first of the problems on which Kiefer spent a considerable effort was on what happens in the Latin square type situations when the combinatorial restrictions are relaxed to allow v treatments in a row and column design with b_1 rows and b_2 columns ($b_1 = b_2 = v$ is the Latin square case). He defined a generalized Youden square (GYS). This is based on a balanced block design (BBD) that is, a block design with v treatments in b blocks of size k. Treatment i appears n_{ij} times in block j and (1) all the $r_i = \sum_j n_{ij}$ are equal, (2) all the $\lambda_{ih} = \sum_j n_{ij}n_{hj}$ are equal, and (3) $| n_{ij} - k/v | < 1$ for all i and j. A GYD is a design in which the treatment/row design and treatment/column design are each separately BBD's. He had proved [22] the universal optimality of a GYD in the so-called regular case when v divides b_1 (or b_2). It was the failure of D-optimality in some nonregular settings that led Jack Kiefer on a long and difficult quest. First A- and E-optimality could be proved. He then produced a remarkable theorem ([61], [83]) which says that D-optimality holds when $v \neq 4$ and clarified the case when $v = 4$. In this work we see him at his most ingenious, using completely original analytic arguments to solve an extremely hard combinatorial problem. There was an immediate spinoff in the study of GYD's both by himself [60] and others (Ash (1981), Ruiz and Seiden (1974), Seiden and Wu (1978)). C.-S. Cheng (1981) extended some of the work to Youden hyper-rectangles.

The other problem was what happens when no BIBD exists so that universal optimality is not so apparent. The natural place to look for optimal designs is among partially balanced block designs with two associate classes or, more particularly, group divisible designs. In a series of papers C.-S. Cheng (for example, Cheng (1978)) proved optimality for certain members of a class of designs called regular graph designs. He exploited important links between optimum combinatorial design and certain problems in graph theory concerned with finding graphs with a maximal number of spanning trees. It is a tribute to the work of Kiefer and Cheng that upon translation into the language of graph theory their theorems gave new results.

The blending of optimality and combinatorics continues to be a very live research area. For example, there has been a burst of recent work using Schur-convexity, a slightly weaker notion than universal optimality; see Giovagnoli and Wynn (1981), unpublished work by Gregory Constantine and the paper of Cheng (1979).

Jack Kiefer had a most fruitful collaboration on experimental design in the last ten years with Zvi Galil. The second part of this consisted of an in-depth study of the maximization of $\det(X^T X)$ when X only has elements with values ± 1. This is a linear regression problem in which each variable is allowed to take values ± 1. It is usually called a weighing design problem because it derives from an experiment to estimate weights when objects are weighed, possibly together, on a chemical balance. When X is square then the solution, when 4 divides k, is to take $X = H$, a Hadamard matrix if such a matrix exists—a well known but unsolved conjecture (now established for all $k \leq 200$). They address a far less studied case than the Hadamard conjecture when $k = 1, 2, 3 \bmod 4$, concentrating mostly on the most interesting case $k = 3 \pmod 4$. Ehlich (1964a, 1964b) had made some original contributions to this problem, which Kiefer and Galil extended to an almost complete solution (modulo the truth of the Hadamard conjecture). Taking X to be an $m \times k$ matrix ($m = 4\ell - 1$), they showed that, for $m \geq 2k - 5$, the following "easy" solution X_e is D-optimum. Take an $(m + 1) \times (m + 1)$ Hadamard matrix, delete a row and select any k columns. They also filled in many cases between the bound and the saturated case $m = k$, [85], [92], [95].

Jack Kiefer had a barely disguised love for combinatorics although his work was never really in the main stream of combinatorial theory. His profound originality proves that he could have turned his hand to algebra or number theory as easily as to mathematical statistics and probability theory.

4. Algorithms. An outgrowth from the continuous theory was the introduction and development of algorithms of the steepest ascent type to generate optimum designs computationally. They consisted of augmenting design measures

$$\xi_{n+1} = (1 - \alpha_n)\xi_n + \alpha_n \xi'_n \quad (0 \leq \alpha \leq 1),$$

where ξ_n was the previous design measure and ξ'_n a measure giving the "direction" of movement, often a single point mass. Wynn (1970) dealt with the natural case $\alpha_n = 1/(n + 1)$, work which was simultaneous with that of V. V. Fedorov (1972). The work took firm hold in the United Kingdom notably with papers by Silvey and Titterington (1973), and in the U.S.S.R. and Eastern Europe (Fedorov, (1972), Pazman (1974)). The main problems were with the unboundedness of the functionals ($-\log \det M(\xi)$ is infinite when $M(\xi)$ is singular) and with the "singular case" referred to above. These were essentially infinite dimensional algorithms since they dealt with measures (see Wu and Wynn (1978)). C.-F. Wu (1978) did important work bringing to bear the extensive literature in optimization algorithms from the optimization literature. By keeping the support of the design finite, finite dimensional algorithms could be developed.

Mitchell (1974) had developed special algorithms involving "excursions" in which new points would be added and old points thrown out (the original Wynn algorithm could be considered as a single infinite excursion). Kiefer and Galil [86] improved on the Mitchell algorithm using a variety of sophisticated computing techniques. They were, as they claimed, "15 to 50 times faster". The algorithm was used to back up and explore the more theoretical work on weighing designs. It was an excellent example of the computer being the slave of mathematicians and a fine blend of sophisticated mathematics and computation. There remain many interesting problems concerned with the speed of these "optimum subset" algorithms. More recently, Welch (1982) has developed a branch and bound integer program leading to exact optima.

5. Non-standard models. A criticism of the classical optimum design theory was that, whereas the mathematics was nice, problems of investigating different models had been ignored. Naturally Kiefer was aware of this. From his files and research it is clear that he was an authority on the variance-bias approach introduced by Box and Draper (1959) in their pioneering paper following earlier work on response surfaces. He was particularly familiar with the extensions introduced by Karson, Manson and Hader (1969). Roughly, the idea is to guard against the possible presence of a more complex model—for example, a quadratic model when a linear model is fitted. In [57] Kiefer gave an illuminating and thorough discussion of the Box-Draper and Karson-Manson-Hader approaches, pointing out some defects and giving methodology to get around some of them. He explored the issues further in [63] and [75] and showed in particular that, for dimensions greater than two, the D-optimum design for quadratic regression on a simplex is more protective against a cubic departure than the Box-Draper approach. Results of Draper and Herzberg based on the same approach are discussed in [84]. The problem still seems largely unresolved and it is possible that D-optimum designs will continue to be robust in higher dimensions. Other authors have taken different approaches, for example, Marcus and Sacks (1977).

Models in which errors are correlated have a small but long history going back to a paper on analysis of variance by Papadakis (1937). Jack Kiefer, in his Berkeley symposium paper [28], tackled head-on a problem posed by Williams to give optimum one-dimensional exact designs to estimate treatment differences in the presence of autocorrelated errors in discrete time. At each point in time one observes

$$Y_t = \alpha_{[t]} + \varepsilon_t,$$

where ε_t is the process and $\alpha_{[t]}$ one of k treatment parameters allocated at time t. He showed for an autoregressive process of order two that, asymptotically, the best designs lay among one of the following patterns (letters are treatments): (1) $AA \cdots BB \cdots$, (2) $ABAB \cdots$, (3) $AABBAABB \cdots$ or (with three or more treatments) (4) $ABCABC \cdots$. Over the two years up to 1981 he and Wynn [98] have given a fairly complete theoretical solution to the problem for a pth order process and a complete combinatorial solution in the case $p = 3$. They were also able to push forward the combinatorial theory into higher dimensions and develop a close connection with the theory of stationary discrete state processes [96]. As in Jack Kiefer's previous work on exact design the door was opened on a new class of combinatorial objects and it was gratifying to find connections in other fields, in this case communication theory. They had also investigated the robustness of classical designs, Latin squares and BIBD's, for simple autocorrelated models [88].

There has been parallel work in continuous time and the seminal work of Jerome Sacks and Donald Ylvisaker ((1969) and other papers) should be mentioned. Like them, Kiefer had seen the importance of studying time dependent and spatial processes and the possibilities this gives for extending the scope of optimum data collection into less controlled environments.

6. Personal note. To work with Jack Kiefer was a privilege and a joy. His untimely death was a dreadful loss to his family, friends and colleagues and his work remains a monument to a great scholar and a delightful human being.

REFERENCES

ASH, A. S. (1981). Generalized Youden designs: construction and tables. *J. Statist. Plann. Inference* **5** 1–25.

BOX, G. E. P. and DRAPER, N. R. (1959). A basis for the selection of a response surface design. *J. Amer. Statist. Assoc.* **54** 622–654.

CHENG, C.-S. (1978). Optimality of certain asymmetrical experimental designs. *Ann. Statist.* **6** 1239–1261.

CHENG, C.-S. (1979). Optimal incomplete block designs with four varieties. *Sankhyā B* **41** 1–14.

CHENG, C.-S. (1981). Optimality and construction of pseudo-Youden designs. *Ann. Statist.* **9** 201–205.

EHLICH, H. (1964a). Determinantenabschatzungen für binare matrizen. *Z. Wahrsch. verw. Gebiete* **83** 123–132.

EHLICH, H. (1964b). Determinantenabschatzungen für binare matrizen mit $n \equiv 3 \bmod 4$. *Z. Wahrsch. verw. Gebiete* **84** 438–447.

ELFVING, G. (1952). Optimum allocation in linear regression theory. *Ann. Math. Statist.* **23** 255–262.

FEDOROV, V. V. (1972). *Theory of Optimal Experiments*. Academic Press, New York.

GIOVAGNOLI, A. and WYNN, H.P. (1981). Optimum continuous block designs. *Proc. Roy. Soc., London Ser. A* **377** 405–416.

GUEST, P. G. (1958). The spacing of observations in polynomial regression. *Ann. Math. Statist.* **29** 214–299.

HOEL, P. G. (1958). Efficiency problems in polynomial estimation. *Ann. Math. Statist.* **29** 1134–1146.

KARLIN, S. and STUDDEN, W. J. (1966). Optimal experimental designs. *Ann. Math. Statist.* **37** 783–815.

KARSON, M. J., MANSON, A. R. and HADER, R. J. (1969). Minimum bias estimation and experimental designs for response surfaces. *Technometrics* **11** 461–476.

MARCUS, M. B. and SACKS, J. (1977). Robust designs for regression problems. *Statistical Decision Theory and Related Topics II*. (ed. S. S. Gupta and D. S. Moore). Academic Press, New York.

MITCHELL, T. J. (1974). An algorithm for the construction of "D-optimal" experimental designs. *Technometrics* **16** 203–210.

PAPADAKIS, J. S. (1937). Méthode statistique pour des experiences sur champ. *Institut D'Amélioration des Plantes à Salonique, Bullitin Scientifique* **23** 1–30.

PAZMAN, A. (1974). A convergence theorem in the theory of D-optimum experimental designs. *Ann. Math. Statist.* **2** 216–218.

PUKELSHEIM, F. (1980). On linear regression designs. *J. Statist. Plann. Inference* **4** 339–364.

RUIZ, F. and SEIDEN, E. (1974). On construction of some families of generalized Youden designs. *Ann. Statist.* **2** 503–519.

SACKS, J. and YLVISAKER, D. (1969). Designs for regression problems with correlated errors. III. *Ann. Math. Statist.* **41** 2057–2074.

SIEDEN, E. and WU, C.-J. (1978). A geometric construction of some families of generalized Youden designs for a prime power. *Ann. Statist.* **6** 451–460.

SILVEY, S. D. and TITTERINGTON, D. M. (1973). A geometric approach to optimal design theory. *Biometrika* **60** 21–32.

SILVEY, S. D. (1980). *Optimum Design.* Chapman Hall, London.

SMITH, K. (1918). On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations. *Biometrika* **12** 1–85.

WALD, A. (1943). On the efficient design of statistical investigations. *Ann. Math. Statist.* **14** 134–140.

WELCH, W. J. (1982). Branch-and-bound search for experimental designs based on D-optimality and other criteria. *Technometrics* **24** 41–48.

WU, C.-F. (1978). Some algorithmic aspects of the theory of optimal design. *Ann. Statist.* **6** 1286–1301.

WU, C.-F. and WYNN, H. P. (1978). The convergence of general step length algorithms for regular optimum design criteria. *Ann. Statist.* **6** 1273–1285.

WYNN, H. P. (1970). The sequential generation of D-optimal experimental designs. *Ann. Math. Statist.* **41** 1655–1664.

IMPERIAL COLLEGE OF SCIENCE
AND TECHNOLOGY
DEPARTMENT OF MATHEMATICS
HUXLEY BUILDING
180 QUEEN'S GATE
LONDON SW7 2BZ
ENGLAND