Statistical Science
1999, Vol. 14, No. 2, 201–205

Gustav Elfving's Impact on Experimental Design

Herman Chernoff

1. INTRODUCTION

During my visit at Stanford University in the summer and fall of 1951, some problems proposed by the National Security Agency (NSA) for an Office of Naval Research (ONR) applied research grant led to two of my publications [1, 2] which had a profound effect on my future research. Both papers had relevance to issues in experimental design. One of these concerned optimal design for estimation. Among other results, it demonstrated that, asymptotically, locally optimal designs for estimating one parameter require the use of no more than k of the available experiments, when the distribution of the data from these experiments involves k unknown parameters. A trivial example would be that to estimate the slope of a straight line regression with constant variance, where the explanatory variable x is confined to the interval $[-1, 1]$, an optimal design requires observations concentrated at the two ends, $x = 1$ and $x = -1$.

Shortly after I derived this result, I discovered a related publication by Gustav Elfving [3]. While Elfving's result is restricted to k-dimensional regression experiments, it gives an elegant geometrical representation of the optimal design accompanied by an equally elegant derivation, which I still find pleasure in presenting to audiences who are less acquainted with this paper than they should be.

In some problems, practical considerations make it impossible to apply *optimal* designs. One beauty of the Elfving result is that the graphical representation of his result makes it rather clear how much is lost by applying some restricted suboptimal methods, and gives some guidance to wise compromises between optimality and practicality.

By 1950, experimental design was a well-established field of statistics. Major sources of application were in agriculture and chemistry, and the analysis of variance played an essential role. Combinatorial and number theoretic approaches, including that of finite geometry, tended to be efficient statistically and computationally for estimating many parameters because of the implicit tendencies to have balance, symmetry and orthogonality. One important consequence of the theory that has been slow in penetrating other sciences is that standard approaches of varying one causal variable at a time is inefficient compared to techniques where several variables are manipulated simultaneously. The weighing schemes of Yates [10] and Hotelling [5] made this point very clearly.

In spite of the activity in experimental design, a general theory for optimal design for estimation was lacking. The revolutionary impact of Elfving's contribution was due to the confluence of several factors. The problem he formulated was general in that it applied to much of the known literature, but was not too general. The results had a simple geometric interpretation and were computationally easy before computer technology was highly developed. Finally they were illustrated in terms of two unknown coefficients which made the results easy to comprehend.

2. THE ELFVING PROBLEM

Consider the regression

$$Y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + u_i, \quad i = 1, 2, \ldots, n,$$

where $\mathbf{x}_i = (x_{i1}, x_{i2})^T$ may be selected by the experimenter from a set S, β_1 and β_2 are unknown parameters, the residuals u_i are independent with mean 0 and constant, possibly unknown, variance σ^2 and the Y_i are observed. A particular *level* $\mathbf{x} \in S$ may be selected several times, yielding independent values of Y. It is desired to estimate

$$\theta = a_1\beta_1 + a_2\beta_2$$

for a specified value of $\mathbf{a} = (a_1, a_2)^T$. How should one allocate the n choices of $\mathbf{x}$ so as to yield a most informative estimate $\hat{\theta}$ of θ?

Herman Chernoff is Professor, Department of Statistics, Harvard University, Cambridge, Massachusetts 02138 (*e-mail: chernoff@hustat.harvard.edu*).