DESIGN AND INFERENCE IN NONLINEAR PROBLEMS

by

Christos Par. Kitsos

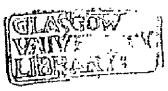
A dissertation submitted to the

UNIVERSITY OF GLASGOW

for the degree of

Doctor of Philosophy

1986



ACKNOWLEDGEMENTS

I am deeply indebted to my supervisors Professor Michael Titterington and Dr Ian Ford for introducing me to the problem and for their keen involvement and interest in the project that led to the final version of this thesis. Their invaluable and generous assistance was of vital importance.

I would like to express my graditute to the Ministry of National Economy of Greece, Technical Assistance Division for their financial assistance through the NATO scholarships. My thanks are also due to the Ministry of National Defence of Greece, where I am employed, for their permission to make use of the NATO scholarship.

Last, but not least. I would like to thank my wife and our three children for their understanding and support. Helen, I want you to know that I am grateful.

Ελένη,

Δήμητρα, Παρασκευάς, Μαρία

SUMMARY

The aim of this thesis is to review and augment the theory and methods of optimal nonlinear experimental design. It represents a continuation of the work on experimental design in the Department of Statistics in Glasgow University (Silvey and Titterington (1973), Ford (1976), Silvey (1980), Titterington (1980a,b), Ford and Silvey (1980), Torsney (1981) among others).

Chapter 1 serves as an introduction to the nonlinear problem. In Chapter 2 we formulate the appropriate notation needed for the development of this thesis. The main assumptions, which we will recall if needed, and the necessary theory is discussed.

In Chapter 3 the idea of the optimal nonlinear experimental design is formulated for any convex criterion function. This leads to traditional definitions as special cases. We also focus on the canonical form of a design under $c(\theta)$-optimality. Partially nonlinear models are reviewed and the design for a subset of parameters is discussed in the context of the general optimality criterion function. The geometrical aspects of the nonlinear case are compared and contrasted with the linear case.

Chapters 4 and 5 are devoted to strategies for the construction of the nonlinear optimal designs. Alternative approaches for the static design problem are discussed in Chapter 4. Emphasis is given to the sequential approach to design in Chapter 5. There, binary response problems are also tackled and the stochastic

approximation method is reviewed and discussed.

Chapter 6 is devoted to confidence intervals. The problem of constructing confidence intervals if the sequential principle of design is adopted is discussed and a suggestion is given. As a result a simulation study is presented.

In Chapter 7 two more simulation studies are analysed, the first for a one parameter binary problem and the second for a two parameter regression problem. Different designing procedures are applied and more emphasis is given to sequential methods. The stochastic approximation method is discussed as a fully sequential method. The performance of approximate confidence intervals is investigated.

Chapter 8 considers a compromise between the static and fully sequential design. The calibration problem is used as an example and investigated in a (yet another) simulation study. The maxi-min efficiency design is derived and investigated.

In Chapter 9 we examine a design problem in rhythmometry involving the cosinor function. Different design criteria are introduced for the full sample space as well as a truncated form. Geometrical ideas provide a solution to solve this problem. An analytical approach is also offered as a method of solution of this practical design problem.

CONTENTS

TABLES

FIGURES

CHAPTER 1

INTRODUCTION

One of the most famous data sets is the one collected in 1793 for defining a new unit: the metre (Stigler, 1981). It was that data set which Gauss used to claim the method of least squares, so named in 1805 by Legendre, who produced the first publication about this invaluable tool of Science, as his "own Method", ("meine Method", in Gauss's words); see Plackett (1972).

What is also very important and we would like to put special emphasis on this, is that for this particular data set:

- In statistical terms it was a nonlinear function that had to be finally estimated.
- A linear approximation had been used in 1755 for this problem and a second order expansion was applied in 1832.

The metre was defined as equal to one 10000000th part of the meridian quadrant i.e the distance from the north pole to the equator along a parallel of longitude passing through Paris. Moreover there is a relationship between the modulus of arc length (s), latitude (d) and meridian quadrant (L) and the linear approximation. is of the form $\eta(L)$ with

$$\eta(L) = \frac{s}{d} = \theta_1+\theta_2\sin^2 L \qquad (1.1)$$

with θ_1 being the "length of a degree at the equator"

θ_2 "the excess of a degree at the pole over one at the equator", see Stigler (1981).

Then the ellipticity (e) is estimated through θ_1 and θ_2 by the nonlinear relationship

$$\frac{1}{e} = 3\,\frac{\theta_1}{\theta_2} + \frac{3}{2} \quad . \qquad (1.2)$$

Both Gauss and Legendre were mathematicians and they treated the data in their own remarkable mathematical way. The first statistical treatment of a nonlinear function comes from the pioneer of modern statistics, R.A. Fisher. In Rothamsted Experimental Station, around 1922, he came across what is known as the dilution series problem. A brief description is as follows. For a small volume u taken from the volume V of a liquid containing n tiny organisms (such as bacteria) the probability P that u contains no organism is

$$P = \left[\, 1 - \frac{u}{V} \,\right]^n \cong \exp(-nu/V) = \exp(-\theta u), \qquad (1.3)$$

say. The parameter θ, the density per unit volume, has to be estimated. The question is how should we perform the experiment to get the best possible estimate. Relation (1.3) is nonlinear in the sense that the parameter θ does not appear as a linear term in the model. Fisher solved this **nonlinear** problem in 1922, using a concept of his own : his information.

Since Fisher's pioneering work in experimental design, Statistics has become involved in all the experimental sciences : chemistry, biology, pharmacology, psychology, and so on. Of course statisticians do not provide methods for designing experiments in isolation. However, in cooperation with the experimenter, who makes clear the ojective of an experiment, the statistician provides the

most informative pattern of the experiment in order that the required objective can be achieved.

The analysis of an experiment can be summarized in the flow-chart of Table 1.1.

The objectives of the experimenter can be :

(i) To obtain an estimate for a response η say in some particular region using variables $u=(u_1, u_2, \ldots, u_k)$. This is the response surface problem introduced by G.E.P. Box and Draper (1959).

(ii) To determine the best mathematical model which describes most precisely the investigated phenomenon. This is the discrimination problem between rival models and it has been reviewed by Hill (1976).

(iii) To estimate optimally, in some sense, all or a subset of the parameters of a model that is assumed correct.

The above mentioned objectives are common to linear and nonlinear experimental designs (LED and NLED) i.e when the assumed suitable model is linear or nonlinear with respect to its parameters. The terms linear and nonlinear are explained clearly with examples in Chapter 2.

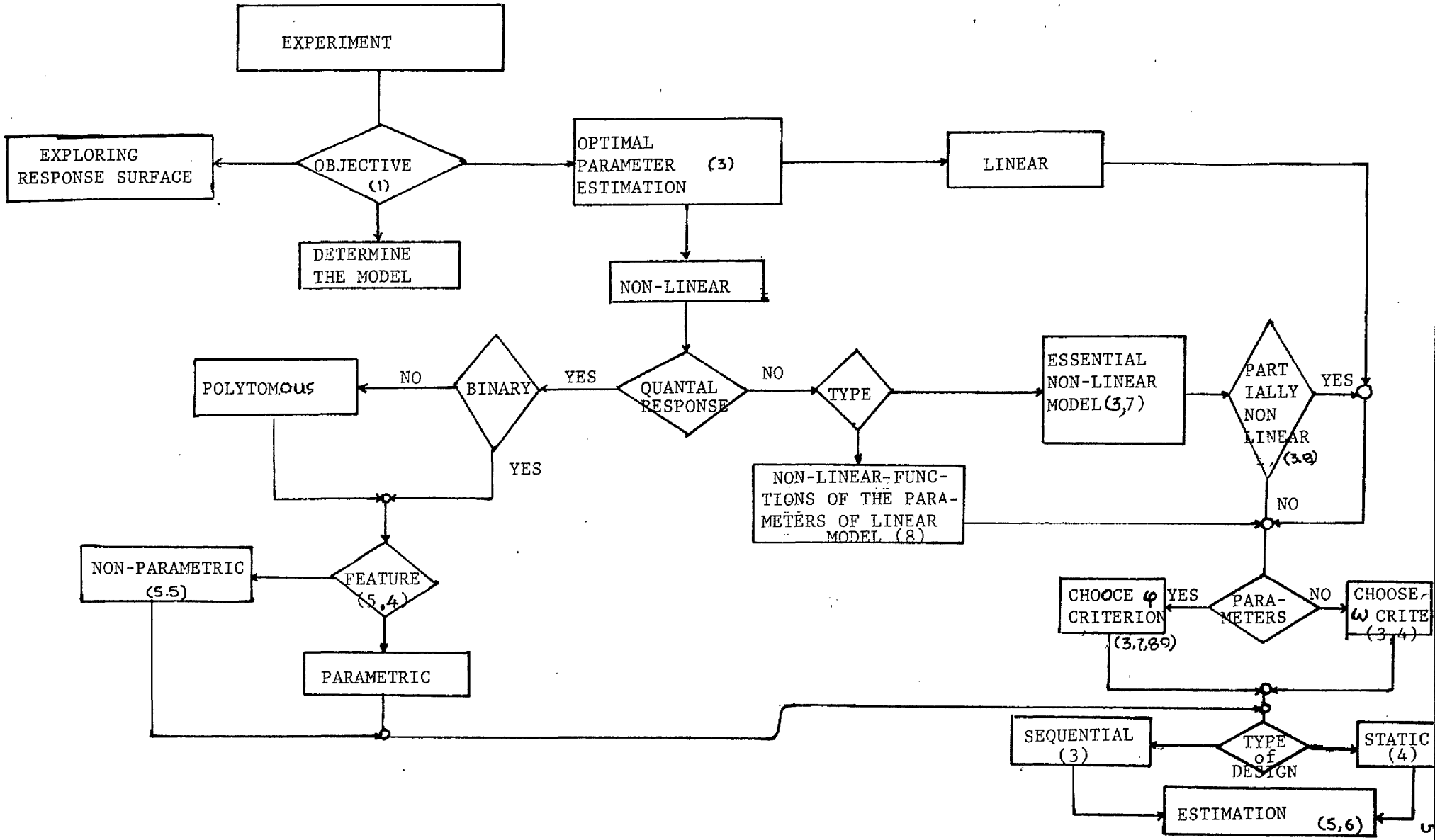
The fact is that more attention has been paid to LED than to NLED. In their recent review work Steinberg and Hunter (1985) devoted only one paragraph to nonlinear models, as did St. John and Draper (1975) ten years earlier. Other review works on LED are those of Ford (1976) and Atkinson (1982). Ash and Hedayat (1978) provide an extensive bibliography covering the work of Eastern countries in this field, Titterington (1980b) reviewed the geometric approach to LED, Pazman (1980) contributed on a theoretical level, Pukelseim and Titterington (1983) offered a general approach to optimal LED and

Torsney (1981) followed the optimum linear design problem through the general optimization problem. Fedorov (1972) and Silvey (1980) contributed excellent monographs on LED.

There is no such volume of review work in NLED, although work on experimental design started with a nonlinear problem, as Cohran (1973) pointed out in his review paper. Davis (1971) compared some sequential procedures in bioassay, Abdelbasit and Plackett (1981,1983) review the nonlinear case for certain types of problems, but give no attention to regression type experiments, and Wu (1985) has worked recently on binary response problems. However, the nonlinear experimental design problem finds applications in many fields : as a regression problem in kinetics (chemistry, biology), as a binary model in testing explosives, biological assays, fatigue experiments, educational studies and life testing.

The target of this thesis is to review and augment the theory of NLED; to compare LED and NLED; to provide, for both, general optimality criteria; to provide methods and discuss problems associated with nonlinear problems. The emphasis will be on the target (iii) described above and its related difficulties.

Table 1.1: The flow chart of an approach to an experiment (In parenthesis the Chapter or Section where it is discussed)



CHAPTER 2

MOTIVATION

2.1 Introduction

To perform any experiment the following are needed :

- The experimental unit.
- The range of experimental conditions.
- The measurements or responses (y, say) obtained at certain values of the explanatory variables (u, say)
- The computer

Usually, in chemical reactions involving kinetics, the experimenter designs in blocks, and the experimental unit is the apparatus which provides the measurements. In psychology or medicine the experimental unit is the individual under investigation, through a test or a medicine, and thus the experiment is performed through single observations.

The Book of Science has been written in the mathematical alphabet. Thus in this chapter we introduce the notation, the necessary assumptions and definitions for the mathematical formulation of the nonlinear experimental design problem.

2.2 Notation

The Euclidean space $U \subseteq R^k$ in which the predictor variables or covariates or explanatory variables or independent variables $u=(u_1,u_2,\ldots,u_k)$ take their values is known as the **experimental region** or **design space**. A typical example from kinetics is "time".

The parameter space $\Theta \subseteq R^p$ is the set where the parameters $\theta=(\theta_1,\theta_2,\ldots,\theta_p)$ takes their values. Let Ξ be the family of measures ξ such that

$$\xi(u) \geq 0 \ , \quad u \in U \quad \text{and} \quad \int_U \xi(du) = 1. \tag{2.2.1}$$

Definition 2.2.1 We shall refer to such a measure ξ as a **design measure**.

Definition 2.2.2 The pair (U,ξ) will be called the **design**.

Definition 2.2.3 The **support** of the design (U,ξ), $Supp(\xi)$, say, is the set of points u for which $\xi(u)>0$.

Denoting by $Mat(\nu,m)$ the set of $\nu x m$ size matrices we let

$\theta \in Mat(p,1)$ be the vector of parameters,

$u_i \in Mat(k,1)$ the predictor variable,

$i=1,2,\ldots,n$, n the sample size.

For the response y we assume that either $y \in \Psi \subseteq R$ or $y \in \{0,1,\ldots,\lambda\}$, $\lambda \geq 1$. When the response y is supposed to take any value in Ψ we also suppose that a **regression model** (in general nonlinear) exists consisting of the deterministic portion $f(u,\theta)$ and the stochastic portion, e, known as error, linked through the formulation

$$y_i = f(u_i,\theta) + e_i. \tag{2.2.2}$$

If we assume that f(.) is of the form $\theta^T g(u)$ with g being a vector of continuous function of u then the **nonlinear** problem is reduced to the so called **linear** problem, as the function f(.) is a linear function of θ and $g(u)=(g_1(u),g_2(u),\ldots,g_p(u))$ is known.

When $\Psi=\{0,1,2,\ldots\lambda\}$ we have the **multiresponse** problem. The most common one is $\lambda=1$, i.e a **binary response**. In this case the outcome $Y_i=1$ or 0 is linked with the covariates and the parameters through a probablity model "T" with

$$P(Y_i=1) = T(u_i,\theta),\ P(Y_i=0) = 1-T(u_i,\theta), \qquad (2.2.3)$$

where u_i is the value of u going with observation Y_i

Examples 2.2.1 (i) For a chemical irreversible reaction with reaction rate mechanism q, where q=1,2,3 or 4 the deterministic function is (see Hill et al, 1968)

$$f(u_i,\theta_q) = \begin{cases} \exp(-\theta_{q1}u_{i1}\exp(-\theta_{q2}/u_{i2})) & \text{if } q=1 \\ \{1+(q-1)\theta_{q1}u_{i1}\exp(-\theta_{q2}/u_{i2})\}^{(1-q)^{-1}} & \text{if } q>1 \end{cases} \qquad (2.2.4)$$

i.e the vector of parameters is $\theta_q=(\theta_{q1},\theta_{q2})$, q=1,2,3,4 and presents the rate constants. The covariates are $u_i=(u_{i1},u_{i2})$, i=1,2,...,n.

(ii) A typical situation in bioassay is to consider the **logistic** model for T in which, if u is a scalar,

$$T(u,\theta) = \{1+\exp(-\theta_1(u-\theta_2))\}^{-1} \qquad (2.2.5)$$

or the probit model in which

$$T(u,\theta) = (\sqrt{2\pi\theta_2^2})^{-1}\int_{-\infty}^{u}\exp(-(v-\theta_1)^2/(2\theta_2^2))dv. \qquad (2.2.6)$$

In both cases $\theta=(\theta_1,\theta_2)$ with θ_1 the location parameter and θ_2 the scale parameter.

2.3 Assumptions

The following main assumptions will be considered throughout this thesis. We shall refer to them as Assumption 1 or 2 etc when we recall them.

When limiting results for the sequence of estimators θ_n are considered the parameter space Θ is assumed to be compact. For the errors the main assumption which is imposed is:

Assumption 1: The errors e_i are independent and identically distributed with

$$E(e_i)=0 \text{ and } V(e_i)=\sigma^2>0, \quad i=1,2,\ldots n. \qquad (2.3.1)$$

Under Assumption 1 for model (2.2.2) we have that

$$\eta = \eta(u,\theta) = E(Y) = f(u,\theta). \qquad (2.3.2a)$$

For the model (2.2.3) we assume that

$$\eta = \eta(u,\theta) = E(Y) = T(u,\theta). \qquad (2.3.2b)$$

(For brevity the dependence of T on θ will sometimes be omitted.)
When inference is to be made, for the regression models, Assumption 1 is changed to the typical normal error assumption i.e

Assumption 2: The independent identically distributed errors are normally distributed with mean 0 and variance $\sigma^2>0$.

It is usually desirable for the design to be insensitive to any violation of Assumption 2. For both f and T functions in (2.3.2) some assumptions are considered. As far as f is concerned we basically want it to be smooth in the neighborhood of θ_t, the true value of θ. That is why we assume

Assumption 3: The function $f(u,\theta)$ is continuous in θ at θ_t and the second order derivatives of f with respect to θ exist at and near θ_t.

Function T plays an important role in binary response problems and it can be known, as in Example 2.2.1(ii), or unknown. In the latter case it is suggested by Wu (1985) that it be approximated locally by the logistic. For T we assume :

Assumption 4: T(u) is a monotonic differentiable function.

Recall (2.3.2b). The function T(u) can be considered as a cumulative distribution function of the random variable Z defined through the random variable Y as follows

$$Y = Y(u) = \begin{cases} 1 & \text{if } Z \leqslant u \\ 0 & \text{if } Z > u \ . \end{cases} \qquad (2.3.6)$$

Then :

$P[Y=1] = P[Z \leq u] = T(u)$, say

$P[Y=0] = P[Z > u] = 1-T(u)$.

Therefore :

$E(Y) = 1xT(u)+0x(1-T(u)) = T(u) = \pi$

$Var(Y) = T(u)x(1-T(u))$.

In many practical cases T is related to the explanatory variable u and the parameter θ, through a linear function of u and θ, thus $T=T(\theta^T u)$ and $f(u,\theta)=f(\theta^T u)$. As discrimination between rival models is not a target of this thesis we will assume throughout this thesis that

Assumption 5: The model used to plan the design is correct.

The existence of the estimators is discussed in the next section.

2.4 On the existence of estimtors

After collection of the data the question arises as to whether it is possible to get estimates in all problems, that is those of binary response and regression.

For the model (2.2.2) we introduce the quantity

$$S_n(\theta) = \Sigma (y_i - f(u_i,\theta))^2 = ||y-f(u,\theta)||_2 \qquad (2.4.1)$$

where $||.||_2$ is the l_2-norm. An estimate $\hat{\theta}$ will be called the **least squares estimate** (LSE) iff

$$S_n(\hat{\theta}) = \min \{ S_n(\theta);\ \theta \epsilon \Theta \} \qquad (2.4.2)$$

Jennrich (1969), imposing some assumptions, proved that the model (2.2.2) has a LSE, $\hat{\theta}$, as a measurable function $\Psi \rightarrow \Theta$, where Ψ is the space of values of Y's. Under Assumption 2 it is known that the LSE coincides with the **maximum likelihood estimators** (MLE).

For the binary response problem Silvapulle (1981) provided conditions under which the **likelihood function** L,

$$L \propto \Pi\{T(u_i,\theta)\}^{y_i}\{1-T(u_i,\theta)\}^{1-y_i} \qquad (2.4.3)$$

can provide maximum likelihood estimators (Appendix A3.II). Roughly speaking that occurs when the intersection of the sets of values taken by the explanatory variables corresponding to 1's and to 0's is not the null set. This happens to be a necessary and sufficient condition for the logit and probit models.

Now, having ensured that the likelihood equation can provide MLE and denoting by ℓ the log-likelihood we define the matrix

$$S(\theta,\xi_n,y) = -\left(\frac{\partial^2 \ell}{\partial\theta_i \partial\theta_j}\right) \qquad (2.4.4)$$

where ξ_n is the design measure on n observations. The matrix

$$S(\hat{\theta},\xi_n,y) = -\left(\frac{\partial^2 \ell}{\partial\theta_i \partial\theta_j}\ \Big|_{\theta=\hat{\theta}}\right) \qquad (2.4.5)$$

will be called the **sample information matrix.**

Example 2.4.1 Maximum likelihood estimates for the logistic can be obtained through the **"normal equations"**

$$\Sigma T_i = \Sigma y_i$$

$$\Sigma u_i T_i = \Sigma y_i u_i$$

with $T_i = T(u_i;\theta)$ as in (2.2.5).

2.5 Fisher's Information matrix

In the regression problems the variance σ^2 is sometimes of the form $\sigma^2(u,\theta)$. That is σ^2 depends on the design point and the parameter vector. In the linear case it is often assumed independent of the parameter θ. In practice it may or may not be possible to assume that is known.

Let $\nabla\eta$ denote the vector of partial derivatives

$$\nabla\eta = \left(\frac{\partial\eta}{\partial\theta_1}, \frac{\partial\eta}{\partial\theta_2}, \ldots, \frac{\partial\eta}{\partial\theta_p} \right)^T. \tag{2.5.1}$$

For the exponential family of models **Fisher's information** matrix is defined to be

$$I(\theta,u) = \sigma^{-2} (\nabla\eta)(\nabla\eta)^T. \tag{2.5.2}$$

It is easy to see that there exists a vector v such that

$$I(\theta,u) = vv^T. \tag{2.5.3}$$

Moreover in many of the nonlinear problems the covariate u and the parameter θ appear together linearly in the form $\theta^T u$. Thus, if

$$\eta = \eta(\theta^T u) \tag{2.5.4}$$

Then

$$\nabla\eta = [w(\theta^T u)]^{1/2} u \tag{2.5.5}$$

with

$$w(z) = [\partial\eta/\partial z]^2 , \quad z=\theta^T u. \tag{2.5.6}$$

Therefore

$$I(\theta,u) = \sigma^{-2}w(\theta^T u)uu^T. \qquad (2.5.7)$$

The concept of the **average-per-observation information matrix** will play an important role in our scenario for the nonlinear experiment design problem. It is defined for ξ_n the n-point design measure, to be

$$M(\theta,\xi_n) = n^{-1} \Sigma\ I(\theta,u_i). \qquad (2.5.8a)$$

For the continous case

$$M(\theta,\xi) = \int_U I(\theta,u)\xi(du). \qquad (2.5.8b)$$

The idea of Caratheodory's Theorem (Appendix A1.II), so essential for the linear experiment design, Silvey (1980), can be used for the average information matrix in nonlinear problems, as has been noted by Titterington (1980a) in the following theorem.

Theorem 2.5.1 (Titterington, 1980)

For any $\theta\epsilon\Theta$

(i) $M(\theta,\xi)$ is symmetric nonegative definite for any $\xi\epsilon\Xi$.

(ii) $M(\theta) = \{M(\theta,\xi);\ \xi\epsilon\Xi\}$ is convex and compact

(iii) The extreme points of $M(\theta)$ are each of the form $I(\theta,u)$ for some u. Further for any $\xi\epsilon\Xi$ there exists $\zeta\epsilon\Xi$ assigning positive weight to at most $2^{-1}k(k+1)+1$ points in U and such that $M(\theta,\xi)=M(\theta,\zeta)$.

Suppose the matrix $M=M(\theta,\xi)$ is partitioned in the form

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{12}^{\tau} & M_{22} \end{pmatrix} \qquad (2.5.9)$$

with $M_{11}\epsilon Mat(s,s)$, $M_{12}\epsilon Mat(s,p-s)$, $M_{22}\epsilon Mat(p-s,p-s)$, $1\leq s<p$.

We define the matrix

$$M_s = M_s(\theta,\xi) = M_{11}-M_{12}M_{22}^{-}M_{12}^{\tau} \qquad (2.5.10)$$

with M_{22}^{-} being the generalised inverse of M_{22} (Appendix A4.I). The information matrix I(.) will be considered partitioned in the same fashion. These partitions are helpful when our interest lies in estimating the leading $s<p$ parameters in the vector θ as it will be explained in section 3.4.

2.6 Linearization of the model

The idea of the (desig)n matrix X being known is essential in dealing with linear models (Graybill, 1976). In nonlinear models we can not define a matrix X in the same fashion. This can be done only approximately through the partial derivatives of θ, with θ taking its "true" value, θ_t. We define the nxp matrix

$$X = (x_{ij}) = \left.\frac{\partial f(u_i,\theta)}{\partial \theta_j}\right|_{\theta=\theta_t} . \qquad (2.6.1a)$$

Then the matrix $X=X(\theta)$ is formed as a function of θ. Function $f(u,\theta)$ can be linearized through a Taylor series expansion in the neighborhood of θ_t as

$$f(u,\theta) = f(u,\theta_t)+\Sigma(\theta_j-\theta_{tj})(\partial f(u,\theta)/\partial\theta_j)|_{\theta=\theta_t} \qquad (2.6.1b)$$

Following the pattern of linear regression models in the nonlinear regression case, an approximation to the covariance matrix, of the estimates of the parameters, can be defined as

$$C \cong [X^T(\theta_t)X(\theta_t)]^{-1}\sigma^2. \qquad (2.6.2)$$

Moreover for all nonlinear problems a useful approximation to the covariance matrix is

$$C^{-1} \cong n\, M(\theta_t,\xi). \qquad (2.6.3)$$

M.J.Box (1971b) "linearized" the nonlinear function through a Taylor series expansion of second order and, through this, he evaluated the bias of θ and $f(u,\theta)$, when the model was fitted. For this one could seek minimum bias experiments. This has had little application in practice for nonlinear models in contrast with the linear case where the idea was introduced by G.E.P. Box and Draper (1959). M.J. Box (1970) suggested also that cost optimal designs could be constructed for the classical nonlinear regression models. The assumption in this case is that the cost of the experimentis represented by its duration. In the sequel we will assume that experiments are equally costly.

The linearization idea can be applied to the logit model (Cox, 1970) in the following example.

Example 2.5.1 Given that

$$[1+\exp(-\theta_1(u-\theta_2))]^{-1} \cong 1/2 + 1/6\ \theta_1(u-\theta_2)$$

when $|\theta_1(u-\theta_2)| \leqslant 3$, then the normal equations of Example 2.4.1 are

approximately

$$n/2 + (\theta_1/6)\Sigma(u_i-\theta_2) = \Sigma y_i$$

$$(1/2)\Sigma u_i + (\theta_1/6)\Sigma u_i(u_i-\theta_2) = \Sigma u_i y_i$$

In the next section we present some examples to clarify the idea of Fisher's information matrix. These examples will be reconsidered in the sequel.

2.7 Examples

(i) Consider the model, in which

$$P(Y=1) = T(\theta^T u).$$

Let: $\theta_1+\theta_2 u_1=z=\theta^T u$ and $T'(z)>0$, $\theta=(\theta_1,\theta_2)$, $u=(1,u_1)$. Then the log-likelihood ℓ, will be

$$\ell = \log\{T(z)^Y[1-T(z)]^{1-Y}\} + \text{const.} \qquad (2.7.1)$$

Therefore we evaluate $I(\theta,\xi)$ as

$$E\{(\nabla\ell)(\nabla\ell)^T\} = \alpha(\theta)uu^T$$

with $$\alpha(\theta)=T'^2[T(1-T)]. \qquad (2.7.2)$$

Application: T might be either the logit or probit function.

(ii) For the nonlinear regression model

$$\eta = \theta_1 - \exp(-\theta_2 u), \quad u \in [-1,1]$$

we have that

$$I(\theta,\xi) = (\nabla\eta)(\nabla\eta)^T = \begin{pmatrix} 1 & u\exp(-\theta_2 u) \\ u\exp(-\theta_2 u) & u^2\exp(-2\theta_2 u) \end{pmatrix}. \tag{2.7.3}$$

We note that $I(\theta,\xi)$ is a function only of θ_2.

(iii) Consider the nonlinear regression model (Box and Draper, 1959)

$$\eta = \theta_1 \exp(\theta_2 u), \quad u \in [a,b]$$

This is a model used to describe growth phenomena. The x_{ij} vectors $i=1,2,\ldots,n$, $j=1,2$ can be formed according to (2.6.1) as

$$x_{i1} = \partial\eta/\partial\theta_1 = \exp(\theta_2 u_i), \quad x_{i2} = \partial\eta/\partial\theta_2 = \theta_1 u_i \exp(\theta_2 u_i) \tag{2.7.4a}$$

We form $S_n(\theta)$ as

$$S_n(\theta) = \Sigma(y_i - \theta_1 \exp(\theta_2 u_i))^2. \tag{2.7.4b}$$

To find the estimator $\hat{\theta}$, which minimizes $S_n(\theta)$ the partial derivatives of $S_n(\theta)$ are needed to obtain the "normal equations"

$$\Sigma(y_i - \theta_1 \exp(\theta_2 u_i))\exp(\theta_2 u_i) = 0 \tag{2.7.5a}$$

$$\Sigma(y_i - \theta_1 \exp(\theta_2 u_i))\theta_1 u_i \exp(\theta_2 u_i) = 0. \tag{2.7.5b}$$

The existence of the least square estimates is shown by Jennrich (1969). Moreover, in this case, on evaluating the Hessian and its' expected value we have

$$\sigma^{-2}nM(\theta,\xi_n) = \begin{pmatrix} \Sigma\exp(2\theta_2 u_i) & \theta_1\Sigma u_i\exp(2\theta_2 u_i) \\ \theta_1\Sigma u_i\exp(2\theta_2 u_i) & \theta^2\Sigma u^2\exp(2\theta_2 u_i) \end{pmatrix}. \qquad (2.8.6a)$$

and

$$\sigma^{-2}S(\hat{\theta},\xi_n,y) = \begin{pmatrix} \Sigma\exp(2\hat{\theta}_2 u_i) & \hat{\theta}_1\Sigma u_i\exp(2\hat{\theta}_2 u_i) \\ \hat{\theta}_1\Sigma u_i\exp(2\theta_2 u_i) & A(y,u,\hat{\theta}) \end{pmatrix} \qquad (2.8.6b)$$

where $\quad - A(y,u,\hat{\theta}) = \hat{\theta}_1\Sigma u_i^2\exp(\hat{\theta}_2 u_i)\ \{y_i - 2\hat{\theta}_1\exp(\hat{\theta}_2 u_i)\}.$

We shall work on this example in chapter 7, which reports a simulation study dealing with different procedures for tackling the design problem.

Having introduced most of the notation needed in the sequel and the appropriate definitions concerning nonlinear models, we are ready to formulate the design problem. In the next chapter the idea of a locally optimal design is introduced and discussed.

CHAPTER 3

LOCALLY OPTIMAL DESIGNS

3.1 Introduction

We recall that the aim of the experiment we shall consider is to estimate as well as possible the parameters included in the model. It might be all p parameters, or a set of $s<p$ linear combinations of the p parameters. In the sequel when only s of the parameters are to be estimated we will assume that they are the first s components of the vector $\theta=(\theta_1,\theta_2,\ldots,\theta_p)$, and we shall use the notation $\theta_{(s)}=(\theta_1,\ldots,\theta_s)$. The average information matrix, related approximately to the covariance matrix, as in (2.5.3), is a natural starting point for the establishment of an optimality criterion. That is, some real-valued function of $M(\theta,\xi)$ can serve as a criterion on the basis of which to answer the question of why one design is better than the other. The sense of optimality is then dictated by the criterion that has been chosen and offers a way of comparing one design - the optimal - with others, which might be optimal under another criterion.

Now, why might the design be called **locally optimal**?
It is the θ-dependence which leads to the term "locally optimal": the optimal design **depends** on the true value of θ.
This θ-dependence is the main point of difference between linear experimental design, which originated in Smith (1918) and the

nonlinear case, originating in Fisher (1922).

Thus in this chapter we provide criteria describing what we mean by suggesting one design take best in comparison with others. In addition, the geometrical interpretation of these criteria is discussed.

3.2 Formulation

Suppose we wish to estimate a set of linear combinations of the parameter vector $\theta=(\theta_1,\ldots,\theta_p)$. This might lead to an estimation of the vector θ itself, some linear combinations of the p components of θ or to $s\leq p$ components. Let $Q\in Mat(s,p)$, $1\leq s\leq p$ be the matrix of the known coefficients defining the above mentioned linear transformation: that is,the quantities of interest are $Q\theta$. If rank(Q)=p, when s=p, the matrix Q is nonsingular. If $s<p$ we suppose that rank(Q)=s.

On the basis of the experiment the average information matrix $M=M(\theta,\xi)$ is obtained. In the sequel we regard θ as taking its true value. Then we can define the following operator J_Q applied to M, through the above matrix Q :

$$J_Q[M] = QM^{-}(\theta,\xi)Q^T \qquad (3.2.1)$$

with M^{-} a generalized inverse of M and $Q^T\in Mat(p,s)$.
It is easy to verify that

$$Q_1=Q_2 \text{ implies } J_{Q1}[M] = J_{Q2}[M]. \qquad (3.2.2)$$

The converse is not true. The matrix $QM^{-}Q^T\in Mat(s,s)$ is

assumed nonsingular. Indeed

$$\text{rank}\{QM^{-}Q^{T}\} \leq \min\{\text{rank}(Q), \text{rank}(M^{-})\} = \text{rank}[Q] = s \qquad (3.2.3)$$

Given the above notation we need a real valued function, ω say, applied to J_Q to be used as an optimality criterion. We choose ω to be a convex decreasing function on the set of nonnegative definite matrices ,NMat(s,s) say, i.e $\omega(A) \leq \omega(B)$ if the matrix $A-B \in NMat(s,s)$ and $A,B \in NMat(s,s)$.

Definition 3.2.1 The design measure ξ^* is called ω-optimal iff

$$\omega\{J_Q[M(\theta,\xi^*)]\} := \min\{\omega\{QM^{-}(\theta,\xi)Q^{T}\}, \xi \in \Xi\}. \qquad (3.2.4)$$

Having introduced the above definition we now examine special cases of ω and Q which will lead us to familiar definitions. The cases we shall consider for ω are

$$\omega(.) = \begin{cases} \omega_1(.)=\text{logdet}(QM^{-}Q^{T}) \\ \omega_2(.)=\text{tr}(QM^{-}Q^{T}) \\ \omega_3(.)=\text{max eigenvalue of } (QM^{-}Q^{T}) \\ \omega_4(.)=\sup \text{tr}\{I(\theta,u)QM^{-}Q^{T}\} \end{cases} \qquad (3.2.5)$$

and for Q are

$$Q = \begin{cases} A \in Mat(p,p),\ \text{rank}(A)=p \\ I \in Mat(p,p) \text{ - the identity matrix} \\ A \in Mat(s,p),\ \text{rank}(A)=s \\ [I_s : \dot{0}],\ I_s \in Mat(s,s),\ \dot{0} \in Mat(s,p-s) \\ c \in Mat(p,1) \end{cases} \qquad (3.2.6)$$

Any $\omega(.)$ combined with some Q will lead to some criterion. In the next paragraph we consider these special cases under the light of definition 3.2.1 along with (3.2.5) and (3.2.6).

We note that, if the whole discussion is expressed in terms of $[J_Q(M^-)]^{-1}$, use, as optimaslity criteria, maximize concave ω functions. Following Silvey (1980) for technical reasons when $J_Q(M)$ is singular $\omega(.)$ is defined to be $+\infty$ $(-\infty)$ for convex (concave) ω functions.

3.3 Special cases

The following definitions obtain, for $\omega_i(.)$ i=1,2,3,4 as in (3.2.5) and Q=A as in (3.2.6)

Definition 3.3.1 The design measure ξ^* is called

$D_A(\theta)$-optimal iff $\omega_1(.)$ is considered
$A_A(\theta)$-optimal iff $\omega_2(.)$ is considered
$E_A(\theta)$-optimal iff $\omega_3(.)$ is considered
$G_A(\theta)$-optimal iff $\omega_4(.)$ is considered

When $Q=I\in Mat(p,p)$ the operator $J_I(.)$ provides just the generalized inverse of M. In this case we obtain the Φ criterion, introduced by Titterington (1980a). Actually under our notation the Φ criterion is

$$\Phi = \omega \circ J_I \qquad (3.3.1)$$

where o denotes the composition of two functions. Corresponding to ω_i, i=1,2,3,4 of (3.2.5) we obtain $\Phi_i=\omega_i o J_I$, i=1,2,3,4.

Traditional definitions of these are the following.

Definition 3.3.2 An optimal design measure ξ^* is called

D(θ)-optimal iff $\Phi_1(.)$ is considered
A(θ)-optimal iff $\Phi_2(.)$ is considered
E(θ)-optimal iff $\Phi_3(.)$ is considered
G(θ)-optimal iff $\Phi_4(.)$ is considered.

The optimality criteria D(θ) and G(θ) were introduced by White (1973) who also extended Kiefer and Wolfowitz's (1960) theorem as follows.

Theorem 3.3.1 (White, 1973)

For the optimal design measure ξ^* the following are equivalent.

(i) ξ^* is D(θ)-optimal.

(ii) ξ^* is G(θ)-optimal.

(iii) $\sup d(u,\xi,\theta)=p=\dim\theta$.

where $\dim\theta$ is the dimension of $\theta\in\Theta\subseteq R^p$ and

$$d(u,\xi,\theta)=\mathrm{tr}\{I(\theta,u)M(\theta,\xi)^{-1}\}$$

(in this theorem the optimal M will be nonsigular.

Following Whittle (1973), Titterington (1980a) stated the extension of the above theorem to any criterion Φ. The idea of a directional derivative (Appendix A1.I) Φ for Φ, when Φ is differentiable, is used as a generalization of derivative, giving a useful tool for hill-climbing optimization techniques. This theorem plays, obviously, an important role in nonlinear theory and it is as follows.

Theorem 3.3.2 (Titterington, 1980a)

For any $\theta\in\Theta$ the following are equivalent.

(i) $\Phi[M(\theta,\xi)]$ is minimized at $\Phi[M(\theta,\xi^*)]$.

(ii) $\Phi[M(\theta,\xi^*),M(\theta,\zeta)]\geq 0$ for all $\zeta\in\Xi$.

If Φ is differentiable at $M(\theta,\xi^*)$ we also have the equivalents

(iii) $\Phi[M(\theta,\xi^*),I(\theta,u)]\geq 0$ for all $u\in U$.

(iv) $\Phi[M(\theta,\xi^*),I(\theta,u)]=0$ for any u weighted positively in ξ^*; that is, for any u in the $\mathrm{Supp}(\xi^*)$. Note that Φ is not differentiable when $M(\theta,\xi^*)$ is singular.

In terms of our notation, with $A \in Mat(p,p)$ nonsingular, there is no difference between $D_A(\theta)$ and $D(\theta)$-optimality as

$$\det[AM^{-1}(\theta,\xi)A^T] = \det[M^{-1}(\theta,\xi)][\det(A)]^2 \qquad (3.3.2)$$

but there is a difference with the other criteria.

$D(\theta)$-optimality minimizes the volume of approximate confidence ellipsoids for θ, centered at $\hat{\theta}$. Moreover, the information matrix $M(\theta,\xi^*)$ corresponding to ξ^* is unique - when θ takes its true value - since ω in this case is a strictly concave function. This is related to the duality theory for the linear case first tackled by Silvey (1972) and established by Sibson (1972). The linear result can be applied in the nonlinear case when $\theta=\theta_t$, for the other criteria as well. Thus:

$G(\theta)$-optimality minimizes the maximum approximate variance of the estimated future response. The interpretation for both $D(\theta)$ and $G(\theta)$ optimality has been made, under Assumption 2, that of normal errors.

$A(\theta)$-optimality minimizes the sum of approximate variances of the parameter estimates, as in the linear case (Titterington (1980a)).

$E(\theta)$-optimality seeks to minimize the variance of the worst-estimated linear combination $c^T\theta$, with $c^Tc=1$.

Ford (1976) describes in detail the above properties of the Φ criterion and its concavity. Titterington (1980b) reviews the geometrical aspects of the linear case.

The geometry of these criteria will be discussed in section 3.7. Torsney (1980) works with generalizations of the above criteria in linear case criteria and Silvey (1980) reviews the criteria in his excellent monograph.

The cases of $Q=A \in Mat(s,p)$, $rank(A)=s$ is similar to that of $Q=A \in Mat(p,p)$. Of course relation (3.3.2) no longer holds.

A particular case is that of $[I_s;\acute{O}]$, $I_s \epsilon Mat(s,s)$ the unit matrix and $\acute{O}\epsilon Mat(s,p-s)$ the zero matrix. This case is discussed in section 3.4.

Now, consider $Q=c\epsilon Mat(p,1)$, i.e Q is a vector. Assuming that $M^{-1}(\theta,\xi)$ exists and recalling definition 3.2.1, we have

$$\omega\{J_c[M(\theta,\xi^*)]\} = \min \{c^T M^{-1}(\theta,\xi)c,\ \xi\epsilon\Xi\} \qquad (3.3.3)$$

with ω the identity function, id say. This criterion minimizes the approximate variance of a linear combination of $\hat{\theta}$ and it is known as $c(\theta)$-optimality. We shall refer to it as Φ_s, i.e

$$\Phi_s = id \circ J_c. \qquad (3.3.4)$$

3.4 Applications

In the linear case the above criteria are independent of θ and thus we refere to them as D,G,A,E-optimality. We can treat the nonlinear case as linear by supposing θ to be known.

The $D(\theta)$-optimality criterion has been the most commonly used in practice, ever since the pioneer work of Box and Lucas (1959) who obtained locally optimal designs when n=p for a number of nonlinear models. For the exact locally optimal design when n=p maximisation of $\det(X^TX)$ (see 2.6.1 for definition of X) is equivalent to maximisation of $\det(X)$ because,

$$\Delta = \det(X^TX) = [\det(X)]^2. \qquad (3.4.1)$$

Atkinson and Hunter (1968) suggested that when n=rp one should perform the experiment of Box and Lucas for the p-point p-term model and replicate the experiment r times. If the design is restricted to these p points replication minimizes the generalized variance of the p-term model. This may not, however, necessarily give the optimum among all experiments with rp-observations. M.J. Box (1968a) gave a generalization to $n=rp+k$, $r>1$ $0 \leq k \leq p-1$.

Optimal or at least near-optimal designs can be produced by replicating, as nearly as possible (depending on the conditions) the experiment of the optimal design using the Box and Lucas idea.

In chemical experimentation the $D(\theta)$-optimal criterion has been very popular since Behnken (1964) obtained optimum designs for determining reactivity ratios. Chemical work was also the motivation for Bard and Lapidus (1968), Hunter and Atkinson (1966), Henson and Hunter (1969), Hill and Hunter (1974), Hunter, Hill and Henson (1974) and M.J. Box (1968a, 1970).

In Table 3.1 we summarize all the work dealing with kinetic models mostly from chemistry and in biology. Given the "true" θ the optimal design points for the covariates involved has been listed.

More work on $D(\theta)$-optimality has been developed in various fields. Currie (1982) estimated the Michaelis-Menten enzyme kinetic function. Hohman and Jung (1975) obtained $D(\theta)$-optimal designs for a special regression set up. We shall refer to this again in Section 3.7.

The aesthetic appeal of $D(\theta)$-optimality extends also to binary response problems.

Begg and Kalish (1984) applied different optimality criteria to logistic model for obtaining the optimal proportion of cases for treatment allocation. Maxim, Hendrickson and Cullen (1977) considered binary response for bivariate models such as bivariate exponential and Weibull distributions. Their result generalizes the following result of Chernoff (1953) for the exponential model under A(θ)-optimality (which coincides for one parameter with D(θ)-optimality).

A(θ)-optimality was tackled by Chernoff (1953) in his early work. He suggested that the maximum number of optimal points needed for $A_s(\theta)$-optimality is $s(2p-s+1)/2$, $1 \leq s \leq p$. It is interesting that when s=1 or p Chernoff's theorem leads to Elfving's theorem (Appendix 3) for c-optimality. A(θ)-optimality has been suggested by Titterington (1980a) for dynamic systems.

Little attention has been paid to E(θ)-optimality in applications.

We shall use D(θ), A(θ) and c(θ)-optimality in later chapters. It is of interest to comment that criteria Φ_i, i=1,2,3,4 coincide for the one parameter model.

We now pay some attention to logit and probit models under D(θ)-optimality, because of their use in applications. Consider the quantal response model of the form $T=T(\theta^T u)$ (recall example 2.7). In this case, (Ford (1976)), the D(θ)-optimal design is concentrated at two points, namely

$$u_1=(u_0-\theta_1)/\theta_2 \quad , \quad u_2=(-u_0-\theta_1)/\theta_2 \tag{3.4.2a}$$

with

$$\xi_1=\xi_2=0.5 \tag{3.4.2b}$$

and

$$D=\det M(\theta,u_0)=\{u_0\alpha(u_0)\}^2/\theta_2 \tag{3.4.2c}$$

with $\alpha(u_0)=\alpha(u_0,\theta)$ as in (2.7.2).

The function $D=D(u_0)$ has a unique maximum at $\hat{u}_0$. Then the optimal points turn to be of the form $(\pm\hat{u}_0-\theta_1)/\theta_2 \epsilon U$.

For the logistic case : $\hat{u}_0 = 1.54$

For the probit case : $\hat{u}_0 = 1.14$

If U is symmetric about $-\theta_1/\theta_2$ and $(\pm\hat{u}_0-\theta_1)/\theta_2 \epsilon U=[\kappa,\lambda]$ then the D(θ)-optimal design is

$$u_1=\kappa \quad , \quad u_2=\lambda. \qquad (3.4.2d)$$

Table 3.1

Local Optimal Design points for nonlinear models (see Appendix 2)

No.	Model, n = (1)	Initial θ_0 - Restrictions (2)	Optimal settings of covariates (3) u_1	u_2	u_3
1	$\theta_1 + \theta_2 \exp(\theta_3 u)$	$\theta_0 = (\theta_{01}, \theta_{02}, \theta_{03})$	κ	$-\frac{1}{\theta_{03}} + \frac{\kappa e^{\theta_{03}} - \lambda e^{\theta_{03}\lambda}}{\exp(\theta_{03}\kappa) - \exp(\theta_{03}\lambda)}$	λ
2	$\theta_1\{1-\exp(\theta_2 u)\}$	$\theta_0 = (\theta_{01}, \theta_{02})$	λ	$-\frac{1}{\theta_{02}} - \frac{\lambda\exp(\theta_{02}\lambda)}{1 - \exp(\theta_{02}\lambda)}$	
3	$\theta_1 - \exp(-\theta_2 u)$	$\theta_0 = (\theta_{01}, \theta_{02})$	(i) κ		
		(i) $\frac{1}{\theta_{02}} \notin [\kappa, \lambda]$	λ		
		(ii) $\frac{1}{\theta_{02}} \in (\kappa, \lambda)$	(ii) $\frac{1}{\theta_{02}}$		
			$\min\left(\frac{\partial n}{\partial \theta_2}\Big\|_{\theta_{02}}^{u=\kappa}, \frac{\partial n}{\partial \theta_2}\Big\|_{\theta_{02}}^{u=\lambda}\right)$		

Table 3.1 (cont.)

No.	(1)	(2)	(3) u_1	u_2	u_3
4	$\theta_1 \exp(\theta_2 u)$	$\theta_0 = (\theta_{01}, \theta_{02})$ (i) $\theta_{02} < 0$ (ii) $\theta_{02} > 0$	(i) κ $\kappa - \frac{1}{\theta_{02}}$ (ii) $\lambda - \frac{1}{\theta_{02}}$ λ		
5	$\exp\left\{-\theta_1 t \exp(-\theta_2(\frac{1}{T}-\frac{1}{T_0}))\right\}$ $u_1 = t$, $u_2 = T$	$\theta_0 = (1.0, 16.000)$ $T_0 = 400$	0.419 8.209	420 380	
6	$\exp(-\theta_1 t_1 \exp(-\theta_2 t_2)$ $u_1 = t_1$, $u_2 = t_2$	$\theta_0 = (\theta_{01}, \theta_{02})$	$\frac{1}{\theta_{01}} \exp(\theta_{02}\kappa)$	$\frac{1}{\theta_{01}} \exp(\theta_{02}\lambda)$	
7	$\frac{\theta_1}{\theta_1-\theta_2}\{\exp(-\theta_1 t)-\exp(-\theta_2 t)\}$ $u_1 = t$	$\theta_0 = (0.7, 0.2)$ (i) estimate both (ii) estimate θ_1 (iii) estimate θ_2	(i) 6.858 1.229 (ii) 1.17 7.74 (iii) 3.52 5.64		

Table 3.1 (cont.)

No.	(1)	(2)	(3) u_1	u_2	u_3
8	As 7 with $\theta_i = \kappa_i \exp(-E_i(\frac{1}{T}-\frac{1}{T_0}))$ (see Appendix 2)	(i) $\theta_0 = (0.7,0.2)$ $\kappa = (0.015,0.003)$ $E = (5000,10000)$	(i) 48.864 21.988 8304.5 319.42	600 600 450 450	
		(ii) $\theta_i \sim N(\theta_{i0}, w_i^2)$ $\delta_i = \sigma/w_i$ i=1,2	(ii) 1.15 1.15		
9	$\frac{\theta_1\theta_3 p_1}{1 + \theta_1 p_1 + \theta_2 p_2}$ $u_1 = p_1$, $u_2 = p_2$	$\theta_0 = (2.9,12.2,0.69)$ (i) estimate θ	(i) 0.2 3.0 3.0	0.0 0.0 1.0	
		(ii) estimate (θ_1,θ_2) when $\theta_3 = 0.69$	(ii) 0.345 3.0	0 0.795	
10	$\frac{\theta_1}{\theta_1+\theta_2-\theta_3}\{e^{-\theta_3 t} - e^{-(\theta_1+\theta_2)t}\}$ $u_1 = t$	$\theta_0 = (0.25,0.5,0.5)$ (i) estimate θ (iiα) estimate (θ_1,θ_2) (iiβ) estimate (θ_2,θ_3) (iiγ) estimate (θ_2,θ_3) (iiiα) estimate θ_1 (iiiβ) estimate θ_2 (iiiγ) estimate θ_3	(i) 0.4 , 1.3 , 2.0 (iia) 0.4 , 1.2 , 2.0 (iiβ) 0.4 , 1.2 , 2.0 (iiγ) 0.4 , 1.25 , 2.0 (iiiα) 0.4 , 1.4 , 2.0 (iiiβ) 0.3 , 1.3 , 2.0 (iiiγ) 0.3 , 1.3 , 2.0		

Table 3.1 (cont.)

No.	(1)	(2)	(3) u_1	u_2	u_3
11	$\frac{\theta_2\theta_3 p_1}{(1-p_1)\{1+(\theta_2-1)p_1\}}$ $u_1 = p_1$	(i) $\theta_2 = 2$ (ii) $\theta_2 = 10,40,80,100,1000$	(i) 0.13 0.30 (ii) 0.05 0.30		
12	$\frac{\theta_1\theta_3(u_2 - u_1/1.632)}{1+\theta_2 u_1 + \theta_3 u_2 + \theta_4 u_3}$	$\theta_0 = (386,044,.022,.107)$ $\beta = .72$ (exponent of power transformation)	471 107 107 107	294 294 69 294	11 (*) 121 (*) 11 (**) 11 (·)
13	$\frac{\theta_1^2 e^{-\theta_3^2 u_3} + \theta_2^2 e^{-\theta_4^2 u_3} u_1 u_2}{\theta_1^2 e^{-\theta_3^2 u_3} u_1 + u_4 \theta_2^2 e^{-\theta_4^2 u_3} u_2}$ (..)	$\theta_0 = (0.001123, 0.00527, 110, 107)$ $\beta = -0.218$	1. 4. 16. 1. 16. 7. 1. 4.	673 673 623 623	5.75 (*) 5.75 (*) 5.75 (**) 5.75 (·)

(*) : no. of runs : 6

(**) : no. of runs : 7

(·) : no. of runs : 8

(..) : all values of u_1, u_2, u_3, u_4 are multiplied by 10^3.

3.5 Canonical form of a design (c(θ)-optimality)

It is known, Federov (1972, p 81), that D-optimal designs are invariant with respect to any non-degenerate linear transformation of the parameters.

It is in fact convenient if we have a design criterion which will remain invariant under certain transformations of the design space. We would then be able to have a canonical form of the design, which, when transformed, would produce other "daughter designs". The transformation we introduce is of the form:

$$h : U \subseteq R^k \longrightarrow Z \subseteq R^k : u \longmapsto h(u)=z=Bu \qquad (3.5.1)$$

with $B \in Mat(p,p)$, nonsingular.

Consider nonlinear models in which the parameters appear in the linear combination $\theta^T u=\theta_0+\theta_1 u_1$ (recall example 2.7 (i)). Take as criterion Φ that of c(θ)-optimality i.e

$$\Phi(M_u) = c^T M_u^{-1} c \qquad (3.5.2)$$

with $M_u=M_u(\theta,\xi)$ the average information matrix in U-space (recall example 2.7 (i)). Thus

$$M_u = \Sigma\, \alpha(\theta_0+\theta_1 u_{1i}) c c^T \qquad (3.5.3)$$

with $\alpha(.)$ as in example 2.7, and $c_i=(1,u_{1i})^T$, $c=(1,u_1)^T$.

Let

$$B = \begin{pmatrix} 1 & 0 \\ \theta_0 & \theta_1 \end{pmatrix} \qquad (3.5.4)$$

and

$$z = Bu = \begin{pmatrix} 1 & 0 \\ \theta_0 & \theta_1 \end{pmatrix}\begin{pmatrix} 1 \\ u_1 \end{pmatrix} = \begin{pmatrix} 1 \\ \theta_0+\theta_1 u_1 \end{pmatrix} = \begin{pmatrix} 1 \\ z_1 \end{pmatrix} . \qquad (3.5.5)$$

Criterion (3.5.3) can be written

$$c^T M_u^{-1} c = c^T [\ \Sigma\ \alpha(.) c_i c_i^T\ ]^{-1} c =$$

$$c^T B^T (B^T)^{-1} [\ \Sigma\ \alpha(.) c_i c_i^T\ ]^{-1} B^{-1} B c =$$

$$(Bc)^T (\Sigma\ \alpha(.)[Bc_i][Bc_i]^T)^{-1}(Bc) = c_Z^T M_Z^{-1} c_Z$$

with $c_Z=(1,\theta_0+\theta_1 u_1)^T=(1,z_1)^T$, $M_Z=\Sigma\alpha(.) z_i z_i^T$, $z_i=(1,z_{1i})^T$.

Thus the equivalence of c(θ)-optimality in U space and Z space has been proved. This is of practical use as a design can be constructed on a "suitable" design space say with θ_0, θ_1 fixed and then transformed back to the design for θ_0, θ_1 of interest.

The c(θ)-optimality criterion can be used when the percentile of a logistic curve is to be estimated. The 100p percentile, L_p, of the response curve T(u) is defined as the solution of

$$T(L_p) = p \qquad (3.5.6)$$

when T(.) is the logistic, Λ(.) say, we have

$$\Lambda(L_p;\theta) = \{1+\exp(-(\theta_0+\theta_1 L_p))\} = p \qquad (3.5.7)$$

Therefore

$$L_p = -\theta_1^{-1}[\theta_0+\ln(p^{-1} - 1)] = \tilde{L}_p(\theta_0,\theta_1) \qquad (3.5.8)$$

Thus L_p has been expressed as a non linear function of θ_0, θ_1. In bioassays the median is the most common percentile of interest. It is easy to see that for p=0.5 (3.5.8) is reduced to

$$L_{.5} = - (\theta_0/\theta_1) \qquad (3.5.8a)$$

for the logit case. Clearly, designing as well as possible to get the best estimate of L_p, has a practical use. We proceed to evaluate the vector $\nabla\tilde{L}_p$,

$$\nabla\tilde{L}_p = -\theta_1^{-1}(1,L_p)^T. \qquad (3.5.9)$$

Let n_i be the number of observations at u_i for i=1,2,...,k. Then, for the MLE $\hat{\theta}=(\hat{\theta}_0,\hat{\theta}_1)$ (recall examples 2.4.1, 2.5.1), it is known that for $k\geq 2$ and $\Sigma n_i=n$ large

$$\mathrm{Var}(\hat{\theta}_0,\hat{\theta}_1) \cong \{\Sigma\Lambda(u_i)(1-\Lambda(u_i))u_iu^Tn_i\}^{-1}, \qquad (3.5.10)$$

in which $u_i=(1,u_{i1})^T$. From (3.5.9) and (3.5.10) we have that

$$\mathrm{Var}(\hat{L}_p) \cong (\nabla\tilde{L}_p)^T\mathrm{Var}(\hat{\theta}_0,\hat{\theta}_1)\nabla\tilde{L}_p \qquad (3.5.11)$$

Therefore minimization of $\mathrm{Var}(\hat{L}_p)$ is approximately equivalent to minimization of

$$c^TM^{-1}(\xi)c \qquad (3.5.12)$$

with $c=(1,L_p)^T$ and $M^{-1}(\xi)$ given by the right had side of (3.5.10).

We thus require to find a $c(\theta)$-optimal design, for the desired $\theta=(\theta_0,\theta_1)$.

We can then use the "canonical form" idea described above to obtain the design in the Z space.

For the logistic of the form (3.5.8) Meeker and Hahn (1977), obtained the two point optimum design for $p\epsilon[0.083,0.917]$, working on survival probability at design test.

The logistic model has also received extensive attention in bioassays, from, among others, Mc Leish and Tosh (1983), Tsutakawa (1980), Cox(1970) and Finney (1978). Recently Wu (1985) suggested that a binary response model of unknown type can be locally approximated by the logistic. The method he proposed will be discussed in chapter 7.

3.6 Designs for subsets of parameters

The asymptotic generalized variance of the estimators of $\theta_{(s)}$ is defined to be the quantity

$$V(\theta,\xi) = n^{-1} \det[M_s]. \qquad (3.6.1)$$

where M_s is as in (2.5.10). Note that M_s has to be nonsingular for $\theta_{(s)}$ to be estimable. With the operator notation it is easy to see that when $A=[I_s:\dot{0}]$, $J_A(M)=M_s$. Under the Φ notation of (3.3.1) we shall use the notation Φ_s for

$$\Phi_s = \omega \circ J_{[I_s:\dot{0}]}. \qquad (3.6.2)$$

For the cases Φ_i, i=1,2,3,4 we have the following definition

Definition 3.4.1 The optimal design measure ξ^* is called

$D_S(\theta)$-optimal iff $\Phi_{1S}(.)$ is considered

$A_S(\theta)$-optimal iff $\Phi_{2S}(.)$ is considered

$E_S(\theta)$-optimal iff $\Phi_{3S}(.)$ is considered

$G_S(\theta)$-optimal iff $\Phi_{4S}(.)$ is considered.

White (1973) stated an equivalence theorem for $D_S(\theta)$ and $G_S(\theta)$ optimality, similar to that for the linear case. As with $D(\theta)$-optimality, $D_S(\theta)$-optimality is relevant and appealing in applications. Begg and Kalish (1984) apply $D_S(\theta)$-optimality to the logistic problem. The M_S matrix and D_S-optimality arise in the chemical kinetics literature; see Hunter, Hill and Henson (1974) for application to the first order chemical reaction (Appendix 2).

3.7 Geometrical aspects

The geometrical approach to the linear optimal design of experiments has been extensively covered by Titterington (1975,1980b). The geometrical insight into the linear problem is based on the equivalence theorems of Kiefer and Wolfwitz (1960), for D-optimality, and Karlin and Studden (1966), for D_s-optimality. Geometric interpretation of these problems has been achieved by Silvey and Titterington (1973) for D_s-optimality. For the linear model (2.3.3) the geometry was built up not on the design space U, but on its image through g, $U_0=g(U)$ say, as described by Silvey and Titterington (1973). Furthermore, for this transformation they proved that the information matrix $M(\xi)$ is preserved expressed in terms of the family of design measures

$$\Xi_0 = \{ \xi_0=\xi g^{-1}, \ \xi\in\Xi \} \qquad (3.7.1)$$

Under the above discussion the following two "duality" theorems play an important role in the realm of linear experimental design.

Theorem 3.7.1 (Sibson, 1972)

Let U_0 be a compact set which spans R^p. The D-optimal design problem for U_0 is the dual of the minimal ellipsoid problem for U_0 and the two problems share a common extreme value.

Theorem 3.7.2 (Silvey and Titterington, 1973)

Let U_0 be any compact subset of R^p which spans the leading s-dimensional co-ordinate subspace. Then for U_0, the D_s-optimal design problem is the dual of the thinnest cylinder problem and the two problems share a common extreme value.

We note that the minimal ellipsoid problem is that of finding an ellipsoid of minimal content centred at the origin, containing U_0. The kernel matrix of the ellipsoid is $M^{-1}(\xi)$. It was Silvey (1972) who pointed out that the only possible effective support points of D-optimal designs are points in U which, through g, correspond to the points where the minimal ellipsoid hits U_0.

The thinnest cylinder problem is that of finding a cylinder of minimum cross-sectional content - with cylinder axis required to pass through the origin - which spans R^s and which contains U_0. The only possible effective support points, according to Silvey and Titterington (1973), are points in U whose images in U_0 are points where the thinnest covering cylinder hits U_0.

We discuss these ideas in the linear case to clarify where this ellipsoid is, and how it is influenced by the design measure and the information matrix. Now we proceed to the nonlinear case.

The fact that the information matrix depends on the unknown parameter θ, and since f is not, linear as is g, prompts us to approach the problem slightly differently.

Recall the definition of the matrix X in (2.6.1). In the linear case in which $Ey=\Sigma g_i(x)\theta_i=f(u,\theta)$, it is easy to see that $x_{ij}=g_j(x_i)$. This is not the case in the nonlinear situation. Expanding $f(u,\theta)$ in a Taylor series, we can write the linearized model as

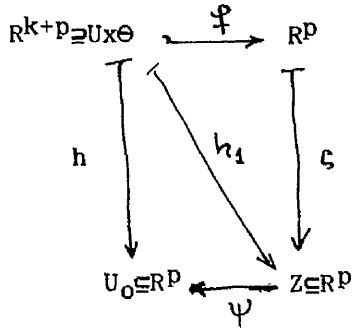
$$W \simeq Xb \qquad (3.7.2)$$

with: X as in (2.6.1a), $b=\theta-\theta_t$, and $W=f(u,\theta)-f(u,\theta_t)$. Thus from the space R^p, through (3.7.2) we obtain the space Z, say, through a transformation, ζ say, corresponding to matrix X. Following M.J. Box (1968a), another transformation ψ say, can be defined from Z to $U_0 \subseteq R^p$

through the relation

$$\psi(z) = u_0 = (X_1 X_2 \ldots X_p)^{-1} z \qquad (3.7.3)$$

with "X_i being the experiments which constitute an optimal design of p experiments to estimate p parameters". The following diagram describes the above discussion.



$$h_1 = \zeta \circ f \qquad (3.7.4)$$

$$h = \psi \circ \zeta \circ f$$

Then following Halmos (1958, p 164), the families of measures

$$\Xi_0' = \{\xi_0' = \xi h_1^{-1},\ \xi \in \Xi\}\ , \quad \Xi_0'' = \{\xi_0'' = \xi h^{-1},\ \xi \in \Xi\} \qquad (3.7.5)$$

"preserve the measure". That is, the optimal design constructed on Z or U_0 is transferred to U. Thus the induced design space U_0 in which the geometry can be built up has been defined.

When θ takes its true value all the geometric aspects covered by Titterington (1980b) can be applied to the nonlinear case. When we end up the experiment with an estimate of θ, this estimate forms the geometry in a "local" sense, as the estimate might any θ in the neighborhood of θ_t. Thus D(θ)-optimality corresponds to the minimal local ellipsoid.

In the case n=p, recall section 4.2. Box and Lucas (1951) gave the geometrical interpretation that the optimal design points should be a vertex of the p+1-hedron, a simplex defined by the p-optimal points and the origin. M.J. Box (1968a) states that the optimal experimental sites must lie on the boundary of U_0, where U_0 is obtained using (3.7.4). Atkinson and Hunter (1968) gave two theorems which ensure that the p+1-hedron must lie wholly within Z in the case of n=rp, i.e when we replicate the p-term optimum design r times. Hill and Hunter (1974) gave a geometrical interpretation for $D_S(\theta)$-optimality.

The geometry of c-optimality is covered by Elfving (1952; see Appendix A3.I). We shall use Elfving's geometrical argument extensively in chapters 8 and 9.

The dependence on θ influences the geometry in the nonlinear case, as the geometric ideas are based on the approximation of the covariance matrix by $[nM(\theta,\xi)]^{-1}$.

We now discuss the particular special case of the partially nonlinear models.

3.8 Partially nonlinear models

Recall the Gaussian regression model (2.2.2). If $f(u,\theta)$ is of a special form some useful results can be obtained.

Hohman and Jung (1975) consider the case where

$$f(u,\theta) = \theta_1+g(u,\theta_2),\ u\in U \tag{3.8.1}$$

a two parameter model linear in one parameter and nonlinear in the other. The D-optimal design provides a two point design with design measure $\xi=0.5$ at each point. The two design points are either the end points of U, or points depending only on the true θ_2, θ_{2t}, say. The fact that $\xi=0.5$ is in accordance with the general result for D-optimality which allocates weight 1/p for the regression model $f(u,\theta)=\theta^Tu$ when the optimal design is supported on p points.

Hill (1980) defined a regression model to be **partially nonlinear** for the k parameters, $k<p$, iff

$$\nabla f(u,\theta) = B(\theta)h(u,\theta_{(k)}) \tag{3.8.2}$$

where $B(\theta)$ is a matrix not depending on u but just on $\theta=(\theta_1,\theta_2,\ldots,\theta_p)$, $\theta_{(k)}$ is the vector of the k parameters which appear in a nonlinear way and h is a vector of functions depending on $\theta_{(k)}$. For the model (3.8.2) the $D(\theta)$-optimal design will depend (1980) only on $\theta_{(k)}$; see Hill (1980)

Example 3.8.1 Consider the model describing decay (or growth) phenomena

$$f(u,\theta) = \theta_1\exp(\theta_2u).$$

Then

$$\nabla f(u,\theta) = \begin{pmatrix} 1 & 0 \\ 0 & \theta_1 \end{pmatrix} \begin{pmatrix} \exp(\theta_2 u) \\ u\exp(\theta_2 u) \end{pmatrix}$$

Therefore the D(θ)-optimal design will depend only on θ_2. This is in accordance with Box and Lucas (1959). We shall reconsider this example in chapter 7.

Khuri (1984) provided, for the model (3.8.2), a sufficient condition of the following form for $D_s(\theta)$-optimality.

Let $\theta_{(s)}$ be the "linear terms", $\theta_{(p-s)}$ be the "nonlinear terms" and consider $B(\theta)$ partitioned as $B(\theta) = [B_1^T(\theta), B_2^T(\theta)]$, $B_1 \in Mat(s,p)$, $B_2 \in Mat(p-s,p)$. Suppose the corresponding $M(\theta,\xi)$ is as in (2.5.10). A sufficient condition for a locally $D_s(\theta)$ optimal design (for $\theta_{(s)}$) to be dependent only on $\theta_{(p-s)}$ is that matrix $B_2(\theta)$ should be expressible in the form

$$B_2(\theta) = P(\theta)[I_{p-s} : K] \qquad (3.8.3)$$

where $P(\theta) \in Mat(p-s,p-s)$ and is nonsingular, I_{p-s} is the identity matrix and $K \in Mat(p-s,s)$ does not involve θ.

Example 3.7.2 Consider the Michaelis-Menten model (Currie, 1982) with

$$f(u,\theta) = \theta_1 u/(\theta_2+u).$$

For this model

$$\nabla f(u,\theta) = (u(\theta_2+u)^{-1} , \theta_1 u(\theta_2+u)^{-2})^T$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & \theta_1 \end{pmatrix} \begin{pmatrix} u(\theta_2+u)^{-1} \\ u(\theta_2+u)^{-2} \end{pmatrix}.$$

Thus (3.8.2) can be applied and therefore the D-optimal design for

estimating $\theta=(\theta_1,\theta_2)$ depends only on θ_2. This is a theoretical justification for Currie (1982, section 3.1) who confines himself "primarily to K" (K is our θ_2). As far as D_1-optimality concerned it can be easily verified that (3.8.3) does not hold.

We now consider a more general model than (3.8.1), which we shall call proper partially nonlinear, as follows :

$$f(u,\theta) = l(u,\beta_1) + m(u,\beta_2) \qquad (3.8.4)$$

where $\beta_1=(\theta_1,\theta_2,\ldots,\theta_s)$, $\beta_2=(\theta_{s+1},\ldots,\theta_p)$ i.e $\theta=(\beta_1,\beta_2)$

$l(u,\beta_1)=\theta_0+\theta_1 u+\ldots+\theta_s u_s$

$m(u,\beta_2)$ any nonlinear function of β_2.

We state the following proposition.

Proposition 3.8.1 For the model (3.8.4) the $D(\theta)$-optimal design depends on β_2. Moreover the $D_s(\theta)$-optimal design for estimating β_1 also depends on β_2.

Proof :

$M=M(\theta,\xi)=M(\beta_2,\xi)$ as, trivially, ∇f does not depent on β_1. Therefore the $D(\theta)$-optimal design depends on β_2.

Therefore $M_{22} = M_{22}(\beta_2,\xi)$ and thus the ratio $\det(M)/\det(M_{22})$ is a function of β_2 only. Therefore the $D_s(\theta)$-optimal design depends on β_2.

This proposition can be viewed as a generalization of the work of Hohman and Jung (1975).

Examples 3.8.2 (i) Let (Hohman and Jung (1975))

$$f(u,\theta)=\theta_1-\exp(-\theta_2 u), \quad \theta_2>0$$

The optimal design depends only on θ_2.

(ii) Let

$$f(u,\theta)=\theta_1+\theta_2 u+\sin(\theta_3 u)+\cos(\theta_4 u)$$

It is easy to see that M_{22} is a function of (θ_3,θ_4) . Therefore $D(\theta)$ and $D_2(\theta)$-optimality depends only on $\beta_2=(\theta_3,\theta_4)$.

3.9 Discussion

Experimental design in the linear case started as an optimum allocation of the observations at the treatment points; see Smith (1918), de la Garza (1954). Fedorov (1972) summarized all the linear work, but he does not supply the general formulation of the problem which is extensively covered by Ford (1976) who proves the concavity of the Φ_i i=1,2,3,4 in the linear case. Silvey (1980) contributes with an excellent monograph. The main target through this theoretical framework is to obtain methods - possibly based on algorithms - to get the optimum design measure for estimating θ. Fedorov provided the first algorithm, but it was only in Wu and Wynn (1978) that a general dichotomous convergence theorem was obtained, concerning the convergence of the sequence of design measures. The theoretical framework in the linear case is completed by the duality theory which first came to light in Lagrangian theory (Silvey 1972, Sibson 1972, Silvey and Titterington 1973). Pukelsheim and Titterington (1983) placed the general optimal experiment problem for parameter estimation in linear regression models under a general framework. Thus the linear theory not only has a well defined theoretical background but this theoretical frame turns out to be helpful from the point of view of applications.

The nonlinear theory suffers from the dependence on the parameters which we want to estimate! Recall (2.6.3) for the average information matrix. Thus any function of $M(\theta,\xi)$ has to be based on the knowledge of the parameters we want to estimate. That made Cochran (1973) comment that the statistician could ask the experimenter "You tell me the value of θ and I promise to design the best experiment for estimating θ"! Silvey (1980) emphasizes this dependence by using the notation Φ_θ.

This θ-dependence also occurs when, in linear regression models, interest lies in a nonlinear function of its parameters. An example is the curvature of the second degree linear model (Ford and Silvey, 1980). We shall come to this work in chapter 8.

Certain problems of course exist when we have fitted a nonlinear regression model. Draper and Smith (1981), Bard and Lapidus (1968), Bard (1974), among others, discuss these problems, which are usually overcome through linearization.

As the nonlinear theoretical framework is based on the true θ, it is easy to see how the so-called equivalence theorems were extended to the nonlinear case from the linear case. We can certainly say that any experiment E_1 with design measure ξ_1 is preferable to experiment E_2 with design measure ξ_2, performed on the same set-up with given θ iff

$$\Phi(M(\theta,\xi_1)) \prec \Phi(M(\theta,\xi_2)) \qquad (3.9.1)$$

However, how to get the minimum $\Phi(.)$ is another story.

So far as (3.2.4) is concerned, the formulation in sections 3.2, 3.3 and 3.4 concerning the general local optimum nonlinear experimental design problem can be stated as follows.

Consider a convex and decreasing ω, a known matrix Q and the operator J_Q. Then :

$$\begin{aligned} \text{minimize} \quad & : \omega \circ J_Q \\ \text{subject to} \quad & : \xi \in \Xi,\ \theta \text{ is the true value} \\ & \quad M(\theta,\xi) \text{ is positive definite.} \end{aligned}$$

CHAPTER 4

STATIC DESIGN

4.1 Introduction

In the linear case, of which an example is that of p-term polynomial regression the D-optimal design has a tendency to use as optimal design points the "end" points of the design space, among others, when $p \geq 2$. This is true even when non classical lines of thought are applied; see Kitsos (1976). Moreover under some considerations, (Fedorov (1972, Th.2.2.3)) the design points for D-optimality can be defined as roots of a particular polynomial (Legendre, Jacobi, Laguerre, Hermite). The design then allocates measure 1/p at these points.

The situation is different in nonlinear problems. The design points can only defined under the "true" θ. Therefore a "guess" about θ has to be supplied. The aim is then to gain knowledge about θ with an efficient estimate, $\hat{\theta}$, say, so that the covariance matrix $\hat{C}=C(\hat{\theta})$ will approximate $n^{-1}M^{-1}(\theta,\xi)$, θ the true value, as well as possible.

This θ-dependence requires the development of alternative strategies for the construction of experimental designs in practice. We shall investigate two procedures that the experimenter might use.

- Procedure 1: Choose design points. Perform the experiment at these points. We shall refer to this procedure as **static design**.
- Procedure 2: Choose initial design points. Perform the experiment at these points and estimate the parameters. Re-assess the design points. Perform the experiment at these new points and get new estimates. Continue the procedure until a predefined stopping rule is satisfied. We shall refer to this procedure as **sequential design**.

In this chapter the static designs will be tackled. Sequential designs will be discussed in chapter 5.

4.2 Locally optimal approach

In practice the true θ is unknown. Thus a guess for θ might be submitted either from previous experimental work or from thepretical considerations. The local optimality criteria Φ_i, i=1,2,3,4 discussed in chapter 3, therefore give a line of thought for applications of using that guess instead of the true θ, which hopefully eventually be approached by its estimator θ. Therefore static designs can obtained by using a guess for θ instead the true θ. Table 3.1 summarizes published work on Gaussian regression models and provides the locally optimum settings of the covariates which might be used when a guess for θ is provided.

4.3 Lauter's approach

An attempt to avoid θ-dependence has been made through S-optimality. Lauter (1974) defined S-optimality as follows.

Definition 4.2.1 The design measure ξ^* is called S-optimal iff

$$S(\xi^*) = \max\{ S(\xi),\ \xi\in\Xi \} \tag{4.3.1a}$$

where

$$S(\xi) = \int_{\Theta} \ln|M(\theta,\xi)|\nu(d\theta) \tag{4.3.1b}$$

and ν is a given measure defined on some σ-algebra of subsets of $\Theta\subseteq R^p$. For S-optimality she proved an equivalence theorem like the one of Kiefer and Wolfowitz (1960) and the one stated by White (1973).

Theorem 4.3.1 (Lauter, 1974)

Let $d_1(\theta,\xi,u) = (\nabla f)^T M^{-1}(\theta,\xi)(\nabla f)$

$$\Xi_o = \{\xi:\ \int_{\Theta} d_1(\theta,\xi,u)\nu(d\theta)<\infty\}.$$

Then for $\Xi_o \neq \emptyset$ and $|S(\xi)|<\infty$ for every $\xi\in\Xi_o$ the following three conditions are equivalent.

(i) $\xi\in\Xi o$ is S-optimal.

(ii) ξ^* minimizes $\max\{ \int_{\Theta} d_1(\theta,\xi,u)\nu(d\theta),\ u\in U\}=\Gamma$.

(ii) $\Gamma=p\int_{\Theta}\nu(d\theta)$.

To avoid θ-dependence a prior distribution can be assumed for θ and then we work using an average information matrix, independent of θ, of the form

$$M(\xi) = E_{\theta}[M(\theta,\xi)]. \tag{4.3.2}$$

any weighting function w(.) on the parameter space Θ, which may or may not be a formal prior density. Then we can use

$$M(\xi)= \int_{\Theta} M(\theta,\xi)w(d\theta) \qquad (4.3.3)$$

or construct a new criterion

$$\Phi_w(\xi) = \int_{\Theta} \Phi[M(\theta,\xi)]w(d\theta) \qquad (4.3.4)$$

with Φ as in (3.3.1). In both cases equivalence theorems like that of Lauter (1974) can be written down.

4.4 Stone and Morris approach

Stone and Morris (1985), in a recent paper, which in their own words, "raises more questions both of theoretical and of practical nature, than it resolves", propose two alternative criteria. These criteria are for non-sequential non-local, non-linear design ie, for the static problem. One of their criteria is based on log-likelihood and the other on sum of squares and both require knowledge of **two** values θ', θ'' of the parameter of interest θ, and include the possibility of a nuisance parameter δ. Their first criterion function, which must be maximized, is

$$C_L = E(LR|\theta',\delta) - E(LR|\theta'',\delta) \qquad (4.4.1a)$$

where

$$LR = \log[p(y|\theta',\delta')/p(y|\theta'',\delta'')] \qquad (4.4.1b)$$

and LR is the logarithm of the likelihood ratio for θ' and θ'', with ζ evaluated at ζ' and ζ'', the conditional maximum likelihood estimates for ζ in each contet. The "design for discrimination" character of this criterion is obvious. Moreover the assumption of a common ζ in (4.4.1a), which must be prespecified, must reduce the practical utility of this criterion. Their second criterion also requires prespecification of two θ's , θ', θ''. This criterion is

$$C_S = \inf\{ \Sigma[\eta_i(\theta',\zeta') - \eta_i(\theta'',\zeta'')]^2,\ \zeta',\zeta''\epsilon A\}, \quad (4.4.2)$$

where $\eta_i(\theta,\zeta)$ denotes the expectation of the i-th observation and A is a prespecified set. For this criterion, if there is no nuisance parameter, the result may be a singular design, from which θ will be unestimable. Note that the specification of the set A may present practical difficulties as well.

The fact that both C_L and C_S are based on two specified values for θ, which we aim to estimate eventually, makes these discrimination criteria rather weak, as far a inference is concerned. Both the above criteria require prespecification of a number of quanties. Stone and Morris (1985) do not investigate problems of misspecification. Therefore if inference about the parameters is of interest rather than discrimination, we reserve judgement on the practical usefulness of this particular approach.

4.5 Maxi-min criterion

Another alternative method of avoiding the θ-dependence problem is the maximin design approach. That is we solve

$$\max_{\xi \in \Xi} \min_{\theta \in \Theta} [h_{\Phi}(M[\theta,\xi])] \qquad (4.5.1)$$

where by $h_{\Phi}(.)$ we mean a function h of the criterion Φ. The maximin design from (4.5.1) will provide that design whose minimum value of $h_{\Phi}(.)$ is greater than that of any other. Even if the locally optimum criterion Φ is invariant under transformations of the parameter space is not necessary that the maximin criterion be invariant. The locally optimal values of Φ, the criterion function, may vary considerably with θ, indicating that some θ values may dominate the construction of the good design. A function h_{Φ} which is of great use in this respect (Silvey, 1980 p 58) is the efficiency measure defined as

$$h_{\Phi}(M[\theta,\xi]) := Eff(\theta,\xi) = \Phi[M(\theta,\xi)] \; \Delta \; \Phi[M(\theta,\xi^{*})] \qquad (4.5.2)$$

with Δ = / or - and $\xi^{*}=\xi^{*}(\theta)$ the locally optimum design for θ.

Silvey (1980) has applied the criterion in two examples and casts some doubt on how useful this approach might be. The maximin efficiency criterion is applied to a particular problem in chapter 8.

4.6 Constant information designs

A constant information design is one where the information of the associated with a nonlinear design, $M(\theta,\xi)$ say, is at least approximately independent of θ.

Fisher (1922,1966) came across this property in the dilution series experiment, which will be discussed extensively in chapter 7. Abdelbasit and Plackett (1981,1983) discuss and extend Fisher's work. In their 1983 work they state that "constant information is a desirable property because the asymptotic dispersion matrix of the estimators is then the same, whatever the values of the parameters". While there is no doubt that this is an interesting property we doupt that this might be considered as the only goal for a design of experiment.

A design with a constant information structure does not remain invariant under nonlinear transformations. Moreover in theory we can obtain, under maximin criteria, designs which are at least as good as equal information designs.

It is interesting to note that Fisher applied this criterion to the dilution series experiment but he never came back to this approach again. Abdelbasit and Plackett generalize this concept to any other problem.

An alternative to the static designs is the sequential way of designing, which is extensively discussed in the next chapter.

CHAPTER 5

SEQUENTIAL DESIGNS

5.1 Introduction

Our objective is to construct a design that eventually estimates the unknown parameter vector θ as well as possible. Adopting the sequential procedure we choose an initial design using prior knowledge on θ and get an estimate of the parameters. This estimate is useful as an initial guess to redesign, reestimate and so on. Some important questions are as follows :

- How do we choose the initial design?
- What measures of optimality can we use?
- How do we revise or continue the design?
- How will inference be made?

In the rest of this chapter we will try to answer these questions. We mention here that we can proceed by either designing in batches of observations or adding a single observation at a time into the design. The latter procedure will be called **fully-sequential** design, adopting Ford's (1976) terminology.

5.2 Background

Let us assume that the initial design has been constructed and an estimate $\hat{\theta}$ has been obtained. When a new design point is added (in terms of a batch of observations or a single observation) a new Fisher information matrix is obtained and a new estimator is evaluated - through least squares for instance. Thus a sequence of least squares estimates $\hat{\theta}_n$ is obtained. Jennrich (1969) proved the existence of these estimators when the design is developed sequentially but in a manner not dependent on $\hat{\theta}_n$. Moreover, he established the strong consistency of the sequence of estimators. provided Θ is compact. i.e

$$\hat{\theta}_n \xrightarrow{a.s} \theta, \text{ as } n \to \infty \qquad (5.2.1)$$

The sequence of average information matrices obtained in this way is also a strongly consistent sequence i.e, as $n \to \infty$

$$M(\hat{\theta}_n, \xi) \xrightarrow{a.s} M(\theta, \xi). \qquad (5.2.2)$$

Finally, he showed that. as $n \to \infty$

$$\sqrt{n}\ (\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} N(0, \sigma^2 M^{-1}(\theta, \xi)) \qquad (5.2.3)$$

where by $\mathcal{L}$ we mean convergence in distribution.

Wu (1981) relaxed Jennrich's assumption and proved the same results.

It is suggested that the initial design should be built up at the optimum points of the corresponding locally optimal design, on the basis of an initial guess for θ. Table 3.1 will be of use in

this context. Fedorov (1972) suggested that the next design point should be that which minimizes the estimator's generalized variance. That is a D(θ)-optimality criterion is used for choosing the next design point. This defines an algorithm with the following steps :

1.- Define initial values θ_0 for θ and perform the experiment.

2.- From the initial design obtain an estimate, $\hat{\theta}_1$, of θ.

3.- Choose as the next design point u_{n+1}, n=1,2,.. that which minimizes

$$d(\hat{\theta}_n, \xi_n, u_{n+1}) = n^{-1}[\nabla f(\hat{\theta}_n, u_{n+1})]^T M^{-1}(\hat{\theta}_n, \xi_n)[\nabla f(\hat{\theta}_n, u_{n+1})]$$

4.- Perform the experiment at u_{n+1} and get $\hat{\theta}_{n+1}$

5.- Perform steps 3. and 4. the required number of times.

The above algorithm is based on the results of Jennrich (1969) and Wu (1981) mentioned earlier. White (1975) considered the sequence $\{M(\theta, \xi_n)\}$ rather than $\{M(\hat{\theta}_n, \xi)\}$. Under very strong assumptions she proved that, an $n \to \infty$,

$$\det M(\theta, \xi_n) \xrightarrow{a.s.} \det M(\theta, \xi^*) \qquad (5.2.4)$$

where ξ^* is the optimal measure. But this limiting result has only been shown to hold for the particular case of D(θ)-optimality and one of the strong assumptions which it is based is that $\hat{\theta}_n \to \theta$.

5.3 Extensions

Establishment of convergence of $M(\hat{\theta}_n, \xi_n)$ to $M(\theta, \xi^*)$ under some function Φ has many technical difficulties. One main virtue of the linear theory is the dichotomous convergence theorem of Wu and Wynn (1978) for any function Φ, and for the sequence $M(\xi_n)$.

Titterington (1980a) generalized, for the regression model, the relation proposed (5.2.4) for D(θ)-optimality. For any criterion Φ (recall Definition 3.3.2) and its corresponding directional derivative Φ (appendix A1.I) he suggested choosing as the next design point u_{n+1} that which minimizes the quantity

$$d_\Phi(\hat{\theta}_n, \xi_n, u_{n+1}) = \Phi[M(\hat{\theta}_n, \xi_n),\ I(\hat{\theta}_n, u_{n+1})]. \qquad (5.3.1)$$

A somewhat similar iterative structure appears in the extension of the Wynn type algorithm for linear designs to nonlinear problems in the following sense.

Algorithm : (Titterington, 1980a)

Consider a sequence α_n, n=1,2,... such that $\lim\alpha_n=0$ as $n\to\infty$ and $\Sigma\alpha_n=\infty$, $0<\alpha_n<1$. Let u_{n+1} minimize

$$\Phi[M(\theta, \xi_n),\ I(\theta, u_{n+1})]. \qquad (5.3.2)$$

Given an initial $\xi_0\in\Xi$ and subject to certain conditions, the sequence of designs generated by the convex iteration scheme

$$\xi_{n+1} = (1-\alpha_n)\xi_n + \alpha_n\xi(u_{n+1}) \qquad (5.3.3)$$

converges to a Φ-optimal design which puts measure 1 on u_n.

There are two difficult features of sequential design.

Firstly, as any design point comes into the design on the basis of a previous estimate of the parameter, the design points are not statistically independent. Thus the "information" matrix, as defined earlier, is not Fisher's information matrix in the sense that it does not necessarily provide an approximate covariance matrix.

Secondly, at the s-th stage, say, the estimator θ_s has to be evaluated. This can be done through the Newton-Raphson method (Stoer and Bulirsh (1979)) which of course might diverge if a poor initial guess is made.

These two problems are related to inference about θ and will be faced in practice in chapter 7.

On the first point Ford (1976) discussed the confidence interval problem for sequential designs under the repeated sampling and the strong likelihood principle of inference. Ford and Silvey (1980), on the basis of a simulation study, made the suggestion of ignoring the fact that the design is sequential and instead, of using the values $u_1, u_2, \ldots, u_n$ of the design as if they were prespecified. Ford, Titterington and Wu (1985) discuss procedures for obtaining valid inferences when the sequential nature of design is adopted.

On the second point, the Newton-Raphson iteration scheme is the numerical method which supplies the estimate at stage s, say, through the iteration

$$\hat{\theta}_{s,k+1} = \hat{\theta}_{s,k} - S^{-1} q_s \qquad k=1,2.... \qquad (5.3.4)$$

where $\hat{\theta}_{s,k}$ is an estimate of the k-th iteration at the s-stage and S is the appropriately-evaluated Hessian of the log-likelihood, which has to be inverted, and q_s the vector of first partial derivatives. For discussion on the Newton-Raphson method in nonlinear problem see

Bard (1974). A statistical version of the Newton-Raphson method, known as stochastic approximation (or the Robbins-Monro scheme) will be discussed in paragraph 5.5.

The sequential idea of designing has also been faced from a Bayesian point of view. The criterion for parameter estimation is to choose that value θ_m of θ for which the posterior density is a maximum. Moreover, in principle for large sample situations the posterior distribution for θ, $\pi(\theta|y_n,u_n)$, given a prior distribution should be approximately normal with

$$\pi(\theta|y_n,u_n) \cong N(\theta_m,(B+S(\theta_m,\xi_n,y_n)^{-1}) \qquad (5.3.5)$$

where the matrix B reflects the prior information. The interesting point is that the Bayesian approach leads to $D(\theta)$-optimality, see Ford (1976) among others. Box and Hunter (1967) and Draper and Hunter (1967a,1967b) obtained sequential designs within a Bayesian framework. When interest is based on a subset s of parameters Henson and Hunter (1969) and Hill and Hunter (1974) used a criterion, based on the corresponding marginal distribution, which leads to $D_s(\theta)$-optimality.

Finally, as it is noted by Silvey (1980 p 66), in sequential design it is only inference obtained from the likelihood function, which remains the same whether the design points u_i, $i=1,2,\ldots,n$ have been predefined or evaluated sequentially. If the repeated sampling approach is adopted the situation is not that clear.

5.4 Binary response problems

Experiments with dichotomous outcomes can be faced in a variety of practical situations. In these cases the "response" and the "non-response" outcome can be presented in different ways. Some practical situations are as follows :

- In testing explosives:

Usually a weight is dropped on the explosive mixture from a certain height. The dichotomous variable takes value "explode" or "not explode".

- In entomological assays

A critical dose level is associated with the insect of interest. The response is "killed" or "not killed".

- In fatigue experiments.

The strength of a certain material is tested. This response is "strong" or "not strong".

- In educational studies.

The tutor might have questions of the form: "right" or "wrong".

- In life testing.

Experiment on the life of a photographic film or safety equipment such as fire extinguishers.

In this kind of problem the main interest is usually devoted to the estimation of a percentile L_p of the response curve. Usually this percentile is the median $L_{0.5}$. The commonly used sequential methods are the following:

- Spearman - Karber's method (Finney, 1978,section 18.7)
- Up-and-Down method (UD) of Dixon and Mood (1948)
- Stochastic Approximation (SA) of Robbins and Monro(1951)

The Spearman-Karber estimator was used in the early 1940's. Finney (1978) formulates this method, which estimates the mean of the response with no computational difficulties but with assumptions which are very unlikely to be attained in practice.

The Up-and-Down sequential scheme can be described by the model:

$$u_{n+1} = \begin{cases} u_n - \delta & \text{if } Y_n=1 \\ u_n + \delta & \text{if } Y_n=0 \end{cases} \qquad \delta>0 \qquad (5.4.1)$$

Dixon and Mood (1948) assumed a probit model with parameters μ an σ^2 for the detonation level, when they applied the method on testing explosives. The choise of the "step size" δ is a problem. One suggestion is that it should be a rough estimate of σ.
Wetherill (1963,1975), Wetherill et al (1966) and Choi (1971) used a logit analysis. Brownlee et al (1953) discuss the method as applied to small samples and Kershaw (1983) provides an extensive simulation study on the method. Wasan (1969, chapter 8) and Tsutakawa (1967) discussed the method as an example of a Markov process : from the definition of the method the choise of each run depends only on the current situation.

As far as applications are concerned in binary response problems we mention that McLeish and Tosh (1983) estimate extreme quantiles of the logistic response. Wu (1985) suggested a local approximation, to any unknown response, by the logistic, when the

quantile L_p, $p\epsilon[0.1,0.9]$ is to be estimated. Wu (1985) obtained fully efficient estimates using the Stochastic Approximation scheme which we discuss in the next section.

As far as life testing is concerned Izeman and Rinoff (1977) worked on the exponential distribution to provide a sequential design. Bergman and Turnbull (1983) placed the life testing problem with the framework of sequential design and the method was applied to data from animal experiments.

5.5 Stochastic Approximation

The Stochastic Approximation (SA) method can be applied to an experiment that is fully sequential in Ford's (1976) terms, i.e when we build up knowledge about θ by adding one experimental unit at each stage of the experiment. It is a stochastic version of the Newton-Raphson (NR) iteration : this numerical method motivated Robbins and Monro (1951) in their pioneer work on SA. Since then the method has attracted much attention in the literature more because of its theoretical framework than because of its potential in applications. The practical aspects of the method are discussed in Chapter 7. Here we present a critical review of the theoretical foundation of the method, while avoiding most of the technicalities.

The SA method deals with relations (2.2.3) and (2.3.2). Namely: evaluate the root θ of the equation

$$E\{Y(u)\} := T(u) = p, \quad p \in R, \tag{5.5.1}$$

where θ is unique and T, p provided. Robbins and Monro (1951) imposed the following two main assumptions on the Borel measurable function T

$$\text{(A1)} \quad (u-\theta)[T(u)-p] > 0 \tag{5.5.2a}$$

$$\text{(A2)} \quad \Pr[|Y(u)-p| \leq K_1] = 1 \text{ for every } u \tag{5.5.2b}$$

with K_1 being a constant. Using the conditions

$$\text{(C1)} \quad \inf|T(u)-p| \geq \delta > 0 \tag{5.5.3a}$$

$$\text{(C2)} \quad T(u) \text{ nondecreasing and } T'(\theta)=b>0 \tag{5.5.3b}$$

they proved that then there exists a sequence α_n, n=1,2,... with

$$\alpha_n>0, \quad \Sigma\alpha_n=\infty, \quad \Sigma\alpha_n^2<\infty \tag{5.5.4}$$

such that the sequence of stimuli

$$u_{n+1} = u_n - \alpha_n(y_n-p) \quad n=2,3,.., \quad u_1 \text{ arbitrary} \tag{5.5.5}$$

converges to θ in mean square, i.e, as $n\to\infty$

$$\lim E(u_n-\theta)^2 = 0. \tag{5.5.6}$$

Recall that (C2) is actually Assumption 4 mentioned in section 2.2. The physical meaning of the sequence α_n can be thought of as the "weight" associated with trial n, as discussed in an application by Guttman and Guttman (1959). A typical α_n might be $\alpha_n=n^{-1}$ or more generally any sequence satisfying the relation

$$c'/n \leq \alpha_n \leq c''/n, \quad c',c'' \text{ constants} \tag{5.5.7}$$

Wolfowitz (1952) weakened (A2) and proved that, as $n\to\infty$

$$u_n \longrightarrow \infty, \quad \text{wp.1.} \tag{5.5.8}$$

Kiefer and Wolfowitz (1952) modified the method to that of evaluating the extremes of a function rather than the roots of an equation. Different SA schemes were developed and these are reviewed by Wasan (1969). Letting

$$\alpha_n = c/(nb),\ b=T'(\theta)=T'(u)|_{u=\theta} \qquad (5.5.9)$$

and assuming that

$$(A3)\quad Y_i = a+b(u_i-\theta)+e_i,\quad a=T(\theta) \qquad (5.5.10)$$

with e_i under Assumption 1 of section 2.2, then we have that, as $n\to\infty$,

$$\lim E(u_n-\theta) = 0 \qquad (5.5.11a)$$

$$\lim Var(u_n-\theta) = \sigma^2c^2/[nb^2(2c-1)],\ \text{with } c>.5. \qquad (5.5.11b)$$

It was Chung (1954) and later Sacks (1958) who looked at asymptotic normality. More assumption were needed of course and were imposed by Chung. In addition to (A1),(A2) he required a lower bound for Var{Y(u)} and that (C1) holds for every $\varepsilon>0$ when $|u-\theta|<\varepsilon$. Then with any α_n of the form

$$\alpha_n = n^{-(1-\epsilon)},\ \epsilon<1/2 \qquad (5.5.12)$$

relations (5.5.11) still hold. With one more assumption

$$(A4)\quad E\{Y(u)-T(u)\}^2 = \sigma^2>0,\quad \text{for every } u, \qquad (5.5.13)$$

and for the same sequence α_n with ϵ bounded below it can be proved that, as $n\to\infty$

$$n^{(1-\epsilon)/2}\ (u_n-\theta) \xrightarrow{\mathcal{L}} N(0,\sigma^2/2b). \qquad (5.5.14)$$

Moreover for the sequence

$$\alpha_n = C/n \ , \ n=1,2,.. \qquad (5.5.15)$$

more assumptions were needed. Chung (1954) imposed them and Hodges and Lehmann (1956) made them slightly weaker. However, the main result is that, for α_n, n=1,2,... as in (5.5.15) with

$$C>1/2K \quad \text{where} \quad K\in \inf[(T(u)-p)/(u-\theta)] \qquad (5.5.16)$$

then, as $n\to\infty$,

$$\sqrt{n}(u_n-\theta) \xrightarrow{\mathcal{L}} N(0, \sigma^2C^2/(2bC-1)). \quad bC>1/2. \qquad (5.5.17)$$

From (5.5.17) it is obvious that the asymptotic variance is minimized with an optimal choice of C, C_{opt} say, namely

$$C_{opt} = b^{-1} = [T'(\theta)]^{-1}. \qquad (5.5.18)$$

Recall that p is given but the c.d.f T will usually not be known, which happens in real life problems. So again we face the common problem of nonlinear situations. Without the knowledge of some quantities (here T,θ) we can not obtain the optimal procedure (here C_{opt} and consequently u_{n+1}).

Thus the problem of creating the sequence (5.5.5) contains the intirnsic problem of creating an approximation for C_{opt}. Approximations need iterations, iterations are sequences, so we need a "daughter" sequence of u_n, β_n say, which might converge, hopefully, to b. Simultaneously the "parent" sequence, will converge to θ.

Sakrison (1965) and Venter (1967) tried to overcome the

difficulty of the unknown T. Anbar (1978) considered a method which is the simplest, from the point of view of applications and well behaved from a theoretical point of view. His idea is based on the main virtue of SA, that SA is a kind of regression of Y on u, and at the same time, C_{opt} is a kind of slope for the unknown $T(\theta)$. Thus at stage n+1 he suggested as C the slope coefficient of the regression line formed from the data u_i and $Y(u_i)$ i=1,2,... namely

$$C \simeq \hat{\beta}_n = \Sigma(u_i-\bar{u}_n)Y_i \;/\; \Sigma(u_i-\bar{u}_n)^2 \;,\; n\geq 2 \qquad (5.5.19)$$

where $\bar{u}_n = n^{-1}\Sigma u_i$.

Anbar (1978) imposed different assumptions the main one being that, for all u, and such that $T(\theta)=p$,

$$K|u-\theta| < |T(u)| < K_1|u-\theta| \qquad (5.5.20)$$

with K, K_1 being constants. Relation (5.5.20) simply means that the derivative of T lies between K and K_1. Moreover he restricted the interval $[K,K_1]$ to a subinterval $[K^*,K_1^*]$ in which he assumed $K_1^*<2K$. Although he made heavy use of this assumption, Lai and Robbins (1981) proved that the following results are still valid without the assumption. Suppose as $n\to\infty$

$$\lim \hat{\beta}_n = b \;,\quad \hat{\beta}_n \text{ as in (5.5.19).} \qquad (5.5.21)$$

Then

$$\text{(i)}\quad \sqrt{n}(u_n-\theta) \xrightarrow{\mathcal{L}} N(0,\sigma^2/b^2),\ \text{as } n\to\infty \qquad (5.5.22)$$

with b as in (5.5.4) and σ^2 as in (5.5.14), and

(ii) $\lim u_n = \theta$ (5.5.23)

moreover

$$\lim \{\Sigma(u_i-\theta)^2/\log n\} = \sigma^2/b^2. \quad (5.5.24)$$

The quantity $\Sigma(u_i-\theta)^2$ has been named the **cost of the experiment** by Lai and Robbins (1979). For the sequence $\hat{\beta}_n$, a truncation has been suggested by both Anbar (1978) and Lai and Robbins (1981) rather than using at each stage $1/\hat{\beta}_n$. Under this truncation idea the sequence $1/\hat{\beta}_n$ is restricted to a prespecified interval. Therefore when a value outside this interval is obtained, the sequence is truncated to the interval limits. Wu (1985) applied the truncation idea and we used the truncation of the SA scheme when obtaining estimates for the dilution series problem which we discuss in chapter 7.

Example 5.5.1. Let $\ell_n(\theta)$ be the log-likelihood of the n observations for a model with c.d.f $p(y_i|u_i,\theta)$. Then for the n+1 observations the log-likelihood will be ℓ_{n+1} and equals

$$\ell_{n+1}(\theta)=\Sigma\log p(y_i|u_i,\theta)=\ell_n(\theta)+p(y_{n+1}|u_{n+1},\theta). \quad (5.5.25)$$

Let $\hat{\theta}_n$, $\hat{\theta}_{n+1}$ be the MLE obtained from $\ell_{n+1}(\theta)$ and $\ell_n(\theta)$ respectively. Taking the derivatives of the two sides of (5.5.25) we have

$$\frac{\partial \ell_{n+1}(\theta)}{\partial \theta} = \frac{\partial \ell_n(\theta)}{\partial \theta} + \frac{\partial \log p(y_{n+1}|u_{n+1},\theta)}{\partial \theta}. \quad (5.5.26)$$

From the definition of MLE

$$0 = \left.\frac{\partial \ell_n(\theta)}{\partial \theta}\right|_{\hat{\theta}_{n+1}} + \left.\frac{\partial \log p(y_{n+1}|u_{n+1},\theta)}{\partial \theta}\right|_{\hat{\theta}_{n+1}}$$

$$= \frac{\partial \ell_{n+1}(\hat{\theta}_n+\hat{\theta}_{n+1}-\hat{\theta}_n)}{\partial \theta} + S_c(y_{n+1}|u_{n+1},\hat{\theta}_{n+1})$$

$$\simeq 0 + (\hat{\theta}_{n+1}-\hat{\theta}_n)\frac{\partial^2 \ell_n(\hat{\theta}_n)}{\partial \theta} + S_c(y_{n+1}|u_{n+1},\hat{\theta}_n) \quad (5.5.27)$$

based on a first order Taylor expansion about θ_n, where by $S_c(.|.)$ we denote the score function for a single observation.

From (5.5.27) we obtain the appropriate recursion

$$\hat{\theta}_{n+1} = \hat{\theta}_n - \frac{1}{\frac{\partial^2 \ell_n(\hat{\theta}_n)}{\partial \theta^2}} S_c(y_{n+1}|u_{n+1},\hat{\theta}_n) \quad (5.5.28)$$

Approximating the Hessian of ℓ by Fisher's information

$$\hat{\theta}_{n+1} = \hat{\theta}_n + I^{-1}(\hat{\theta}_n)S_c(y_{n+1}|u_{n+1},\hat{\theta}_n) \quad n=0.1,2... \quad (5.5.29)$$

Example 5.5.2. Consider the regression model

$$y = \exp(-\theta u)+e$$

For the error term e we assume that it is under Assumption 2. The notation of example 5.5.1 is used. Namely

$$\ell_n = \text{const.} -(1/(2\sigma^2))\Sigma(y_i-\exp(-\theta u_i))^2 = \Sigma \log p(y_i|u_i,\theta)$$

$$\partial \ell_n/\partial \theta = -\sigma^{-2}\Sigma u_i \exp(-\theta u_i)(y_i-\exp(-\theta u_i)) = \Sigma S_c(y_i|u_i,\theta)$$

$$\partial^2 \ell_n/\partial \theta^2 = -\sigma^{-2}\Sigma[(-u_i^2\exp(-\theta u_i))(y_i-\exp(-\theta u_i))+u_i^2\exp(-2\theta u_i)]$$

$$I(\theta) = \sigma^{-2}\Sigma u_i^2 \exp(-2\theta u_i)$$

Applying the recursion formula (5.5.29) we get

$$\hat{\theta}_{n+1} = \hat{\theta}_n - I^{-1}(\hat{\theta}_n)[u_{n+1}\exp(-\hat{\theta}_n u_{n+1})(y_{n+1}-\exp(-\hat{\theta}_n u_{n+1})] \quad (5.5.30)$$

The information $u_i^2\exp(-2\theta u_i)$ is asked to be minimized in each stage, as an optimum design rule. Therefor taking the logarithm of the information and evaluating the root of the derivative it can be shown that the optimum design rule occurs when

$$u_{i+1} = 1/\hat{\theta}_i \qquad i=0,1,.. \quad (5.5.31)$$

Substituting (5.5.31) in (5.5.30) we get

$$\hat{\theta}_{n+1}=\hat{\theta}_n-[(1/\hat{\theta}_n)e^{-1}(y_{i+1}-e^{-1})]/[\overset{n+1}{\Sigma}(1/\hat{\theta}_{i-1}^2)\exp(-2\hat{\theta}_n/\hat{\theta}_{i-1})], \quad (5.5.32)$$

as one point recursion. Assuming that $(\hat{\theta}_n/\hat{\theta}_{i-1})\cong 1$, the scheme (5.5.32) is approximated by

$$\hat{\theta}_{n+1} = \hat{\theta}_n - [(1/\hat{\theta}_n)e(y_{n+1}-e^{-1})]/[\overset{n+1}{\Sigma}(1/e_{i-1}^2)]. \quad (5.5.33)$$

In a long run the summation $\overset{n+1}{\Sigma}(1/e_{i-1}^2)$ will be approximately equal to $(n+1)\hat{e}_n^{-2}$. Therefore (5.5.33) can be approximated by

$$\hat{\theta}_{n+1} = \hat{\theta}_n - [\hat{\theta}_n e/(n+1)](y_{n+1}-e^{-1}). \quad (5.5.34)$$

Scheme (5.5.34) is a stochastic approximation scheme in which the sequence α_n is

$$\alpha_n = \hat{\theta}_n e/(n+1). \qquad (5.5.35)$$

Moreover if we assume that the sequence of estimates lie in between θ_l and θ_u then

$$\Sigma\alpha_n \geq \theta_l e \Sigma 1/(n+1) \text{ and } \Sigma\alpha_n \leq \theta_u^2 e^2 \Sigma 1/(n+1)^2 \qquad (5.5.36)$$

Thus for the truncated sequence in which $\hat{\theta}_{n+1}$ is defined by (5.5.34) unless the right-hand side is less than θ_l (in which case $\hat{\theta}_{n+1}=\theta_l$) or is greater than θ_u (in which case $\hat{\theta}_{n+1}=\theta_u$), condition (5.5.4) holds (because of (5.5.30), and therefore the SA scheme converges to the root of the equation $\partial\ell/\partial\theta=0$. i.e to the MLE θ.

5.6 Discussion

The methods UD and SA tackled in sections 5.4 and 5.5 have the following common characteristics.

(i) They deal with the fully sequential way of designing.

(ii) There is a non-parametric flavour to them.

(iii) They are developed for estimating the parameter of interest, usually a single one and not a subset of several parameters.

The fully sequential nature is obvious as one observation comes at each stage. Moreover both UD and SA are "Markovian" in the sense that the choice of each run depends only on the current situation. The martingale (see Appendix 10) structure of SA has been considered by Lai and Robbins (1979).

Recall (5.5.5). When the "regression" equation (5.5.5) is of the form

$$T(L_p) = p,\ p\epsilon(0,1) \tag{5.6.1}$$

i.e the 100p percentile of the response is to be evaluated, (5.5.5) reduces to

$$L^{(n+1)} = L^{(n)} - \alpha_n(y_n - p). \tag{5.6.2}$$

When p=0.5 then the median, $m=L_{.5}$, is to be estimated through

$$m_{n+1} = m_n - \alpha_n(y_n - 1/2) = \tag{5.6.3}$$

$$= m_n - (\alpha_n/2)(2y_n-1)$$

(If we take $\alpha_n=2\delta$ to be a constant sequence, then)

$$= m_n - \delta(2y_n-1) \qquad (5.6.4)$$

and relation (5.6.4) is equivalent to (5.4.1). Thus the SA scheme has been reduced to the UD method, which is used for the median of the response. In other words the UD method is a special case of the SA scheme.

The nonparametric feature of the methods is based on the fact that we try to estimate a functional of the unknown response T(.), usually the p-th percentile L_p. The assumption which is made about the cdf T(.) is usually one of two, either normal or logistic, which leads to probit or logit analysis. Indeed both methods have been used. The virtue of the logit model is that it is both simple and approximates the normal very well in the range $p\epsilon[0.2,0.8]$ (Cox,1968, Table 2.1, Finney, 1978, section 17.4).

The SA scheme is a fully sequential procedure for solving an equation : recall (5.6.1) as a special case of (5.5.5). For the logistic the percentile L_p is as in (4.3.8a); that is, a ratio has to be estimated through (5.6.1).

The superiority of SA in comparison to UD is that the Up AND Down method has only two positions to which to move. The SA method on the other hand is more flexible in terms of the step length taken : SA moves the sequence according to the gradiant of the tangent to T(.) as the scheme is of the form

$$u_{n+1} = u_n - (1/(n\hat{\beta}_n))(y_n-p),\ \hat{\beta}_n \text{ as in (5.5.19)} \qquad (5.6.4)$$

where $\hat{\beta}_n$ is the slope of a linearization of T(.) i.e $\hat{\beta}_n \cong T'(\theta)$. Thus even from a geometrical argument UD can be regarded as a restricted

direction SA.

The analogy between SA and NR is obvious. Thus, the use of a good initial guess in SA is based on the fact that NR converges for an initial value in the neighborhood of the solution (see : Appendix A6.I). The idea that T is a Borel measurable function is as essential as is the assumption in Numerical Analysis of a continues differentiable function. Moreover the derivative sequence in NR has been replaced by a sequence of real numbers in SA.

Now, let us try to compare the form of SA and the generalization of Wynn's algorithm from Titterington (1980b) which is presented in section 5.3.

For the sequence α_n, common to both, we have (we shall refer to Titterington's algorithm as (A))

$$\text{(SA)} \quad \alpha_n>0,\ \Sigma\alpha_n = \infty,\ \Sigma\alpha_n^2 < \infty \tag{5.7.4a}$$

$$\text{(A)} \quad 0<\alpha_n\leq 1,\ \Sigma\alpha_n = \infty,\ \lim \alpha_n = 0. \tag{5.7.4b}$$

We are applying a convex iteration scheme in both situations. Indeed for SA the design points are produced through the iteration

$$\text{(SA)} \quad u_{n+1} = u_n - \alpha_n(y_n-p) \qquad n=1,2,\ldots \tag{5.8.4a}$$

For algorithm (A) a convex iteration of the design measures is used, namely

$$\text{(A)} \quad \xi_{n+1} = \xi_n - \alpha_n(\xi_n-\xi(u_n)) \quad n=1,2,\ldots \tag{5.8.4b}$$

The optimal choice of the sequence $\alpha_n=C/n$ in SA (recall that in (A) a typical α_n, is $\alpha_n=1/n$) depends on the optimal choice of C.

This has been chosen from (5.5.18) as that which minimizes the asymptotic variance of the quantile estimator, i.e

$$1/C_{opt} = T'(\theta). \qquad (5.8.5)$$

Under Assumption 4 and using the directional derivative idea (Appendix A1.I)

$$T'(\theta) = \Phi_T(\theta,\theta+1) \qquad (5.8.6)$$

with Φ_T being the Frechet directional derivative of T. Algorithm (A) suggests that the next design point u shall be that which minimizes the quantity

$$\Phi\{M(\theta,\xi_n),I(\theta,u)\} \qquad (5.8.7)$$

with Φ the Frechet directional derivative of the criterion Φ which is being considered. Thus beyond the point that algorithm (A) is a "steepest ascent" method and SA is also searching for some optimal direction through C and its "regression" approximation (recall β_n in (5.5.19) the two schemes share a number of interesting points.

Chapters 4 and 5 were devoted to design for point estimation under different procedures. The next chapter tackles the interval estimation problem .

CHAPTER 6

CONFIDENCE INTERVALS

6.1 Introduction

After performing the experiment and the data have been collected, statistical inference needs point estimation and the construction of the appropriate, possibly approximate, confidence intervals.

In nonlinear problems we try to apply the linear theory in constructing confidence intervals. The cost we have to pay is the approximation involved and its lack of accuracy. The accuracy depends on how nonlinear the function of interest is and thus the idea of a measure of nonlinearity, which is discussed in the rest of this chapter, was introduced.

When the design is constructed sequentially the question must be asked of how we should obtain the confidence intervals. Ford (1976) and Ford and Silvey (1980) studied this problem. An extension of their approach is tried in chapter 7. Here we apply their argument in section 6.5.

6.2 Background

Recall model (2.2.3), i.e a classical nonlinear regression model, under Assumptions 2,3 and 5. As the function f(.), the deterministic part of model (2.2.3), is nonlinear it is of interest to see how nonlinear it is, by a Taylor expansion.

The tangent hyper-plane to the solution locus (the surface in the sample space generated by the points $\eta_i = f(\theta, u_i)$, $i=1,2,\ldots,n$ with θ regarded as a variable) at point $f(\hat{\theta})$ is given by

$$q(\theta) = f(\hat{\theta}) + \hat{X}'(\theta - \hat{\theta}) \qquad (6.2.1)$$

where $f(\theta) = (f(\theta,u_1), f(\theta,u_2), \ldots, f(\theta,u_n))^T$, $\hat{X} = X(\hat{\theta})$ and X as in (2.6.1). In principle θ_t (the true θ) is needed, instead of $\hat{\theta}$, but $\hat{\theta}$ is used in practice.

Beale (1960) suggested a dimensionless **empirical measure of nonlinearity** Λ^*, defined by

$$\Lambda^* = ps^2(d_2/d_4) \qquad (6.2.2)$$

with s^2 as in (2.5.1), p the number of parameters involved and

$$d_i = \Sigma \, ||\eta(\theta_k) - q(\theta_k)||^i \quad i=2,4 \qquad (6.2.3)$$

where $\{\theta_k\}$ is a set of m points in the neighborhood of θ.
The **theoretical measure of nonlinearity** according to Beale, Λ say, is the same as Λ^* but with σ^2 in place of s^2 and with $m \to \infty$. The minimum value of the theoretical measure of nonlinearity, Λ_0 say, was named

the intrinsic nonlinearity of the assumed correct model.

Guttman and Meter (1965) criticized Beale's measures of nonlinearity on the basis of some real life models. Later M.J. Box (1971b) provided a lower bound for nonlinearity depending on the bias of the estimator θ : thus bias and nonlinearity were connected. Measure Λ_0 is a sort of curvature of the solution locus. Bates and Watts (1980), using ideas from differential geometry, proved that Λ_0 is one quarter of the mean square intrinsic curvature. Moreover they proved that by replicating the design r times the curvature at any point in any direction is reduced by a factor $1/\sqrt{r}$. Thus replication obtained its own geometrical interpretation.

Although the measures of nonlinearity have a strong theoretical background thanks to the work of Hamilton, Watts and Bates (1982) and Bates and Watts (1981) the linear approximation is what is applied in practice. One reason is that departure from linearity needs at least the evaluation of the Hessian which might prove computationally inefficient, even these days. More recent work appears in Hamilton (1986).

6.3 Confidence regions

In constructing confidence regions the target is always to minimize their volume/area/length. Hence optimal design might lead to minimum approximate confidence regions.

If (6.2.1) is true, i.e the model is linear, then a $100(1-\alpha)\%$ confidence region corresponds to

$$S_n(\theta)-S_n(\hat{\theta}) = (\theta-\hat{\theta})^T(X^TX)(\theta-\hat{\theta}) \leq ps^2F(\alpha;p,n-p) \qquad (6.3.1)$$

with $S_n(\theta)$ as in (2.4.1), $S_n(\theta)$ being the residual sum of squares, s^2 an estimate of σ^2, X as in (2.6.1a) and $F(\alpha;p,n-p)$ as usual the $100(1-\alpha)\%$ of the F distribution. We note that in nonlinear problems the estimator of σ^2

$$s^2 = S_n(\hat{\theta})/(n-p) \qquad (6.3.2)$$

is not an unbiased estimator of σ^2 (Draper and Smith, 1981 ,p 504). Recall that matrix X depends on an estimate for θ. Thus the approximation is based both on the linearity and the dependence of X on θ.

Beale (1960) treated confidence regions for the classical regression problems assuming a uniform prior distribution over the solution locus. He adjusted the confidence region to be of the usual form, using

$$S_n(\theta) - S_n(\hat{\theta}) \leq \lambda p s^2 F(\alpha;p,n-p), \qquad (6.3.3)$$

in which

$$\lambda = \begin{cases} 1, \text{ linearization without Beale's assumption} \\ 1+(n/n-1)\Lambda_0 \quad \text{if } p=1 \\ 1+[n(p+2)/(n-p)p]\Lambda_0 \quad \text{if } p\geq 2 \end{cases} \qquad (6.3.4)$$

Thus Beale (1960) used the measure of nonlinearity he developed to adjust an approximation from linearity. when a confidence region is constructed.

In the nonlinear case confidence regions sometimes appear to have "banana-shapes". Under suitable transformation they can sometimes be made ellipsoidal and thus easier to deal with.

Hamilton and Watts (1985) argue that elliptical confidence regions (under D(θ)-optimality) that are suitable for large samples are not appropriate for small samples. Thus they try to construct a quadratic approximation to the volume of small sample confidence regions. They propose a design criterion of the form

$$G = a\cdot\log\det\{M(\theta,\xi)\}-\log\{1+b\cdot trQ(\theta,\xi)\} \qquad (6.3.5)$$

where a.b are constants. $M(\theta,\xi)$ is the average information matrix and $Q(\theta,\xi)$ is a matrix describing parameter effects. Because of the presence of the second term in (6.3.5) the criterion G is not invariant under (nonlinear) transformations of θ. Moreover their criterion requires an estimate of σ^2 which is not always available.

In section 6.5 we tackle the small sample problem in constructing confidence intervals, adopting a sequential design procedure. In principle, for large samples, the covariance matrix is approximated by the inverse of Fisher's information matrix (recall (2.6.2)) and thus approximate confidence intervals can be obtained. However, account has to be taken of the nature of the experiment, i.e if it is a sequential or static one.

It was Ford (1976) who stated that the sequential nature of the design is irrelevant to any method of inference based on the strong likelihood principle. Therefore maximum likelihood estimates can be calculated as if the design points $u_1,\ldots,u_n$ were predetermined. Moreover in practice Fisher's information matrix can be approximated by the sample information (recall (2.4.5)) when the likelihood function has been evaluated.

Unfortunately the situation is not that clear when the repeated sampling approach is adopted, although Ford and Silvey

(1980) suggest that in setting up confidence intervals, even in the singular case "we may effectively ignore the fact that the design is sequential". Ford (1976) in Section 4.4 provides an illustration. Thus the inference made takes into consideration the design procedure used in experiment an experiment E is well defined by the following four elements

$$E = (U,\mathcal{E},F_i,P_j) \quad i=1,2,\ j=1,2 \qquad (6.3.6)$$

where $(U,\mathcal{E})$ form the design, (recall definition 2.2.2), F_i, $i=1,2$ indicates whether the problem is "quantal" or not and P_j, $j=1,2$ describes the design procedure : sequential or static. After E has been performed, inference is made from the results.

6.4 Simulation study

In sequentially constructed designs, the work of Ford and Silvey (1980) plays an important role. Moreover, there are cases in which the design must, by necessity, to be built up sequentially. Ford, Titterington and Wu (1985) discuss various procedures for obtaining valid inferences in sequential design.

The features of a fully sequential design appear in the autoregressive model.

$$y_{i+1} = \theta y_i + \epsilon_{i+1}\ ,\ i=1,2,\ldots,n. \qquad (6.4.1)$$

Note that this is of the form $y_{i+1}=\theta x_{i+1}+\epsilon_{i+1}$, where $x_{i+1}=y_i$. The value y_i is given and the errors ϵ_{i+1} satisfy Assumption 2. Then an estimate of θ, θ_n, is given by

$$\hat{\theta}_{n+1} = \Sigma y_i y_{i+1} \; / \; \Sigma y_i^2 \qquad (6.4.2)$$

where the summation runs from 1 to n. The sample information I_n/σ^2 can be evaluated from

$$I_n = \Sigma y_i^2 \qquad (6.4.3a)$$

and Lai and Siegmund (1983) point out that the asymptotic result

$$I_n^{1/2} (\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} N(0,\sigma^2) \qquad (6.4.3b)$$

holds for $|\theta|<1$.

The result of Lai and Siegmund (1983) can be put into context by noting that if the design points were fixed in the model $y_i=\theta x_i+\epsilon_i$, i=1,2,..n, and $\hat{\theta}_n=\Sigma y_i x_i/\Sigma x_i$, then it holds exactly for any n and θ that $I_n^{1/2}(\hat{\theta}_n-\theta) \xrightarrow{\mathcal{L}} N(0,\sigma^2)$ where $I_n=\Sigma x_i^2$. Ford, Titterington and Wu (1985) also discuss this model. In the simulation study we use small sample sizes of n=10,5 observations, for different nominal levels α=0.05, 0.20. We took y_1=0.0 and used error variance σ^2=1. Different values of θ were taken from the range $-1.5 \leq \theta \leq 1.5$. Confidence limits were evaluated according to the formula

$$\hat{\theta} \mp t(n-1;1-\alpha/2)\sqrt{[(RSS/n-1)I_n^{-1}]} \qquad (6.4.4a)$$

with RSS as the residual sum of squares, namely

$$RSS = \Sigma y_i^2 - \hat{\theta}\Sigma y_{i+1} y_i. \qquad (6.4.4b)$$

Of course (6.4.4) will only give exact confidence intervals in the case where $y_i=\theta x_i+\epsilon_i$ with all the x_i's fixed in advance or

selected independently of the other y_i's.

To test the normality of the sequence θ_n the skewness and the kurtosis were evaluated. The results are presented in Tables 6.1, 6.2, for 1000 simulated experiments.

From Tables 6.1, 6.2 it is easy to see that so far as the normality is concerned the results are unsatisfactory when $\theta \notin (-1,1)$. The mean squared error (MSE) is , of course, larger when the sample size is reduced from n=10 to n=5 observations. i.e the accuracy of the estimators has been reduced.

However, so far as the coverage probabilities are concerned, the study provides evidence that, even with small sample sizes, the approximation is with some exceptions reasonably valid. In general from the results of Tables 6 the nominal level did not influence the study.

Although model (6.4.1) is linear, the estimate θ is a ratio of the data obtained sequentially, i.e not independently. The work of Ford and Silvey (1980) and this simulation study encourages us to ignore the sequential nature of the design, when inference has to be carried out, though we note the increased non normality of $\hat{\theta}$ in extreme cases.

We indeed present three more simulation studies to support this. The binary response problem and a two variable regression problem are discussed in the next chapter.

Table 6.1

Simulation study on autoregressive model (6.4.1)

Nominal level α=0.05. Number of simulations N=1000

Sample size n=10,5 , y_1=0.0, σ=1

n	θ	P	MSE	S	K	$\bar{\theta}$
10	-1.5	.919	0.03	3.88	21.07	-1.44
	-1.0	.947	0.09	1.39	5.47	-.84
	-0.5	.970	0.09	0.63	3.27	-0.42
	0.0	.967	0.10	-0.06	2.56	0.00
	0.5	.965	0.10	-0.58	3.01	0.41
	1.0	.955	0.09	-1.29	5.45	0.85
	1.5	.928	0.03	-3.96	22.21	1.45
5	-1.5	.968	0.27	1.13	5.79	-1.29
	-1.0	.969	0.27	0.08	10.00	-0.82
	-0.5	.979	0.25	-0.09	6.05	-0.39
	0.0	.970	0.25	0.22	4.92	0.02
	0.5	.973	0.33	-0.58	4.95	0.79
	1.0	.975	0.61	-0.14	4.98	0.38
	1.5	.957	0.29	-1.23	6.02	1.29

P : estimated coverage, S : Average skewness of estimates

K : Average kurtosis of estimates, $\bar{\theta}$ = Average$(\hat{\theta}_i)$ $i=1,\ldots,N$

Table 6.2

Simulation Study on autoregressive model (6.4.1)

Nominal level α=.20. Number of simulations N=1000

Sample size n=10,5 , y_1=0.0, σ=1

n	θ	P	MSE	S	K	$\bar{\theta}$
10	-1.5	.773	0.04	4.78	31.35	-1.36
	-1.0	.778	0.09	1.12	4.64	-0.85
	-0.5	.868	0.08	0.61	3.43	-0.42
	0.0	.856	0.09	-0.11	2.73	0.00
	0.5	.839	0.09	-0.65	3.37	0.41
	1.0	.827	0.07	-1.38	5.84	0.87
	1.5	.770	0.03	-3.28	19.39	1.44
5	-1.5	.764	0.25	1.56	6.98	-0.88
	-1.0	.851	0.27	0.65	4.83	-0.80
	-0.5	.862	0.27	-0.48	7.83	-0.39
	0.0	.904	0.24	0.09	5.50	0.00
	0.5	.866	0.25	-0.28	4.83	0.38
	1.0	.838	0.27	-0.51	4.84	0.81

$\bar{\theta}$, P,S,K as in Table 6.1

CHAPTER 7

SIMULATION STUDIES

7.1 Introduction

We have already reviewed and augmented the theory of the nonlinear design problem. This chapter and, the following two, are devoted to applications. We try "to put theory to work" on particular problems. We discuss the difficulties which arise and the results obtained. Both the binary and continuous cases are tackled for one parameter and two parameters respectively. The simulation studies were carried out on the ICL computer of Glasgow University.

In the sequel we describe the problems simulated, and given interpretations of the results. These simulations are

- Simulation I

  The dilution series problem (sections 7.2 to 7.6).

- Simulation II

  The first order growth law (sections 7.7 to 7.9).

7.2 The dilution series assesment

Experimenters and statisticians are indebted to Rothamsted Experimental Station as it offered a job to the jobless Fisher! Since then Fisher developed the theory of experimental design and

tackled the first nonlinear design problem in 1922. This problem is that of dilution series which we now describe.

It is desired to determine the concentration of migro-organisms in a solution. In this case various dilutions are sampled. For every dilution we record whether or not there is a sterility. We use the following notation.

u : a small volume that is taken out of a volume, say V, of liquid which contains N tiny organisms. Let $U=[U_l, U_u]$ be the design space.

θ : The density per unit volume, i.e $\theta=N/V$. The probability that the volume contains no organisms is

$$p = (1-u/V)^N \simeq \exp(-Nu/V) = \exp(-\theta u).$$

y : The binary response describing the phenomenon is

y=1 : no organism in u (sterile)

y=0 ; organisms in u (fertile).

The probability model describing the experiment is therefore

$$p(y|u,\theta) = \begin{cases} \exp(-\theta u) & y=1 \\ 1 - \exp(-\theta u) & y=0 \end{cases} \qquad \theta>0 \qquad (7.2.1)$$

The aim is to estimate θ as well as possible. Model (7.2.1) might describe also the probability that an insect survives a dose of u units of a certain insecticide.

Fisher's information $I(u,\theta)$ for model (7.2.1) can be calculated as (Appendix A7.I)

$$I(\theta,u) = u^2/(\exp(\theta u)-1). \qquad (7.2.2)$$

On maximizing $I(\theta,u)$ we get (Appendix A7.I)

$$\theta u^* = 1.59. \qquad (7.2.3a)$$

Thus the optimum design point, the one which minimizes the variance, i.e, that which corresponds to D(θ)-optimality, depends on θ according to

$$u^* = \begin{cases} 1.59/\theta & \text{if } 1.59/\theta \in U \\ U_l \text{ or } U_u & \text{otherwise.} \end{cases} \qquad (7.2.3b)$$

The form of the probability model (recall (2.4.3) with $T(u,\theta)$ as in (7.2.1)) will be binomial with success probability $p=p(1|u,\theta)$ and number of successes the number of sterile samples. In terms of probability, the value $u^*=1.59/\theta$ corresponds to $p=0.2$. We use values of p when θ is equal to its true value to define the space U. It seems reasonable to keep the probability levels between [0.025,0.975]. Therefore, throughout the simulations we use the bounds U_l, U_u, which can be evaluated from the relations

$$\exp(-\theta_t U_u)=.025 \ , \ \exp(-\theta_t U_l)=.975 \qquad (7.2.4)$$

where θ_t is the "true" value of θ. For this value we choose $\theta_t=3.18$ which corresponds to $u^*=0.5$ from (7.2.3a). Thus from (7.2.4) we get

$$U_l = 0.00796 \qquad U_u = 1.160. \qquad (7.2.5a)$$

Note that U_l would represent the optimal design point for θ=199.70 and U_u similarly for θ=1.370. With this in mind we restrict the parameter space to $\Theta=[\theta_l,\theta_u]$, with

$$\theta_l = 1.370 \qquad \theta_u = 199.70. \tag{7.2.5b}$$

Thus a bounded parameter space is obtained, useful in the simulation study.

7.3 The strategy of Simulation I

Adopting a sequential procedure for designing we choose the design points entering the design to be these ones which minimize Fisher's information. At each stage , s say. s≥1 the data are generated in a batch of r replications of the form

$$y_i = \begin{cases} 1 & \text{if } \exp(-\theta_t u_i) > U(0,1) \\ & \qquad i=1,2,\ldots,r \\ 0 & \text{otherwise} \end{cases} \tag{7.3.1}$$

where U(0,1) is a uniform (0.1) random number generated by the NAG subroutine G05CAF.

In the first stage. s=1. the estimator of θ. $\hat{\theta}_1$ say, can be evaluated explicity from the data (proof Appendix A7.II (ii)) as

$$\hat{\theta}_1 = -u_1 \ln(r^{-1}\Sigma y_i). \tag{7.3.2}$$

We use the bounds of Θ as estimates of θ, in the extreme

cases. It can be proved that if all y_i's are 1 then $\hat{\theta}=0$ and when all y_i's are 0 then $\hat{\theta}=\infty$. Therefore we avoid situations, especially with small batches, when the MLE could not be evaluated (Appendix A3.II). The MLE exists and was evaluated iteratively when $0<\Sigma y_i<n$. The numerical method of Newton-Raphson (NR) (Appendix A6.I) was used to solve, at each stage, the likelihood equation (Appendix A7.II) for evaluating $\hat{\theta}$. Newton-Raphson converges when the initial value lies in the neighbourhood of the solution. Therefore we had to overcome this difficulty, which happened often when small batches were used in early stages of the sequential design. We used the Bisection method (Appendix A6.II) using a rather "large initial interval", [0.01,100.3] to obtain a "good" initial value and then the Newton-Raphson method was started off.

As far as the design points are concerned the procedure can therefore be described by

$$u_{s+1} = \begin{cases} 1.59/\theta_s & \text{if } u_{s+1}\in U \\ U_l & \text{if } \hat{\theta}_s<\theta_l \\ U_u & \text{if } \hat{\theta}_s>\theta_u \end{cases} \quad s=1,2\ldots,s_{max} \qquad (7.3.3)$$

The maximum number of stages, s_{max}, say, depends on the number of replications r chosen. Simulations were carried out for r=5,25.50. We kept n=100 and therefore s_{max}=20,4,2, respectively.
The estimates corresponding to (7.3.3) were obtained through

$$\hat{\theta}_s = \begin{cases} -u_0 \ln\bar{y} & \text{if } s=1 \\ \theta_u & \text{if all } y_i's=0 \quad ,s>1 \\ \theta_l & \text{if all } y_i's=1 \\ \hat{\theta} \text{ evaluated through} \begin{cases} \text{NR if r=50} \\ \text{Bisection and NR if r=5,25.} \end{cases} \end{cases} \qquad (7.3.4)$$

Therefore the design points and the estimates at each point are well defined through the truncation we have introduced and the numerical techniques we used.

To investigate the dependence of the design procedure on the initial design points we choose the values $\theta_1=2.196$ and $\theta_2=7.15$ as the starting values. These values of θ lead to corresponding design points u=.72 and u=.22 in (7.3.3). These design points correspond to probability levels p=0.1 and p=0.5 respectively when $\theta_t=3.18$ i.e values to the right and to the left of the local optimum design point $u^*=0.5$ corresponding to $\theta_t=3.18$ with probability level p=0.2. For the final estimate $\hat{\theta}$, i.e when n=100 observations were used, an approximate confidence interval was evaluated for $\hat{\theta}$ by using the formula

$$\hat{\theta} \mp 1.96\sqrt{(1/S(\hat{\theta},\xi_n,y))} \qquad (7.3.5)$$

where $S(\hat{\theta},\xi_n,y)$ is the sample information. recall (2.4.5) (see also Appendix A7.II). That is, although in the sequential design the design points are not entering the design independently of the response we follow Ford and Silvey (1980) who constructed the confidence interval by "pretending" that the design points were independent of the response. We shall come to this point in the discussion of the results of this simulation study. The experiment was repeated 1000 times. The "confidence intervals" were constructed and it was checked whether the "true" value of θ was captured. The estimated confidence probabilities are reported.

7.4 Simulation procedures

Different procedures were applied to tackle the dilution problem under the strategy described above. We will refer to these as P1,P2 etc. In all the cases n=50 or 100.

P1. Static design.

Chernoff (1953), in his early work on A(θ)-optimality, suggested that the optimal static design (recall chapter 4) will be that one which takes all the observations at the locally optimal point for the true θ, as in (7.2.3b). Therefore the n observations were taken at $u^*=u^*(\theta_{st})$ where θ_{st} is the "starting value" for θ. Data were generated and the MLE was calculated. For P1 the case n=1000 was also investigated.

Results in Tables 7.1.

P2. Sequential design. equal batches.

The batch sequential method of designing was adopted. Equal batches were used to reach the total sample size n.

Results in Tables 7.2

P3. Sequential design, unequal batches.

We start off the design with a batch of 25 or 50 observations. The MLE was evaluated explicitly at the first stage. Thereafter, i.e when $s=2,3,..s_{max}$, the number of replications, r' say, was taken to be 5. The values of s_{max} are 15 and 10 and

correspond to the initial batches 25 and 50 observations.

Results in Tables 7.3.

P4. Fully sequential design (Stochastic Approximation).

Batches of 5,25,50 observations were used to start off the design. One observation was then added, i.e $r'=1$, to the design and only one step of the Newton-Raphson "iteration" was used to produce the estimate of θ. This is the Stochastic Approximation scheme discussed in section 5.5.

Results in Tables 7.4.

P5. Fully sequential design (Full Maximum Likelihood at the end)

Here we use the data generated by P4 and obtain the exact MLE at the end of the experiment.

Results in Tables 7.5.

We comment that a fully sequential design with a "fully" evaluated MLE in all the steps, i.e with a "full" Newton-Raphson iteration at each point, was not investigated. The reason was not only that the computational time was very large, but we have evidence to believe that little was to be gained by using at each stage a "full" Newton-Raphson iteration. We shall come to this point in section 7.5.

Below we describe the output presented in Tables 7.1-7.5 for the procedures described above.

θ_{st} : Starting value for θ.

r : Number of observation per batch, r=5,25,50.

ECP : Estimated Coverage Probability i.e the proportion of times out of the 1000 simulations the true value of θ was captured in the confidence interval.

$\bar{\theta}$: The average value of the estimates, $\hat{\theta}_i$, produced in N=1000 simulations.

S : Estimated skewness of $\hat{\theta}$.

K : Estimated kurtosis of $\hat{\theta}$.

EMSE : Estimated Mean Square Error of the 1000 evaluated $\hat{\theta}_i$'s through the relation

$$EMSE = est.Var(\hat{\theta}) + [est.Bias(\hat{\theta})]^2 \qquad (7.4.1)$$

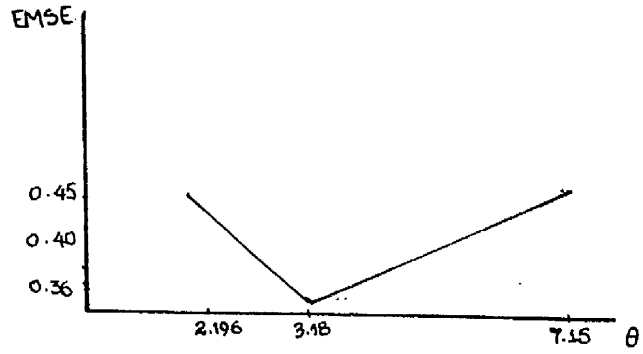
The "true" value of θ in all the cases was 3.18, and the sample size n was either 50 or 100. In Fig.7.1-7.5 lines —— represent r=50, ••• represent r=25, – – – represent r=5.

Table 7.1 : Simulation Study I, Procedure P1.

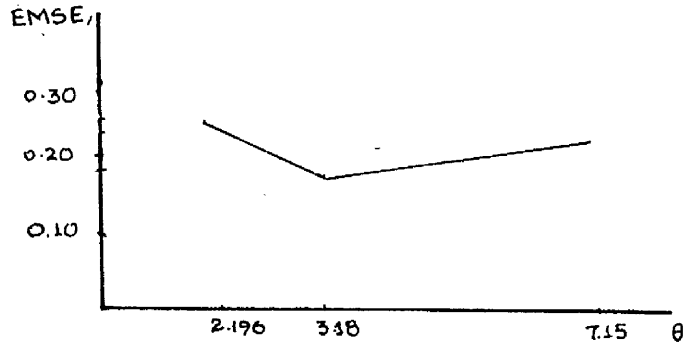
n	θ_{st}	ECP	$\bar{\theta}$	S	K	EMSE
1000	2.196	.953	3.18	.23	2.97	.018
	3.18	.954	3.19	.02	2.97	.015
	7.15	.958	3.18	.05	2.93	.019
100	2.196	.950	3.24	.59	3.98	.204
	3.18	.952	3.2	.53	3.47	.172
	7.15	.952	3.0	.17	3.29	.201
50	2.196	.977	3.31	1.52	5.09	.454*
	3.18	.958	3.28	.56	3.43	.365
	7.15	.935	3.22	.38	3.27	.462

* : Two "outliers" were not considered, therefore N=998.

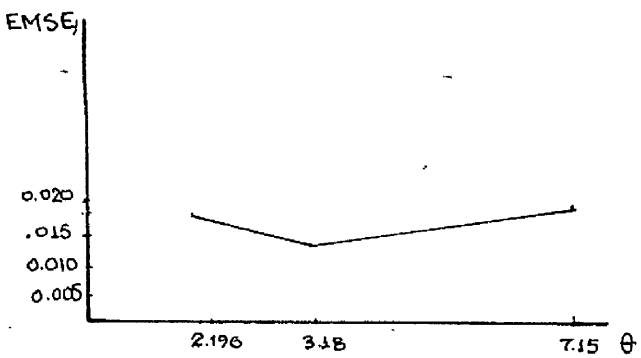
Fig. 7.I : Simulation Study I. Graphs of EMSE for PI.



n = 50



n = 100



n = 1000

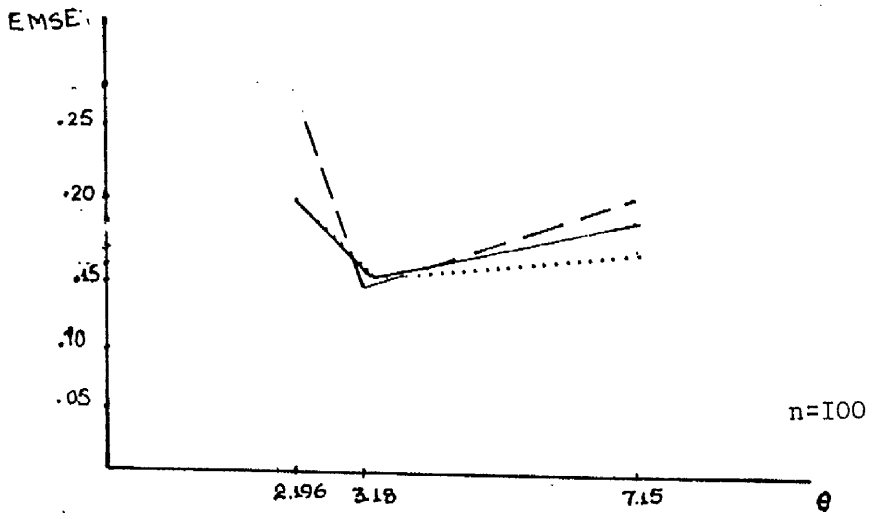
Table 7.2 : Simulation Study I, Procedure P2.

n	e_{st}	r	ECP	$\bar{e}$	S	K	EMSE
100	2.196	5	.937	3.23	.55	3.86	.20
		25	.946	3.24	.66	3.85	.20
		50	.945	3.25	1.64	8.46	.22*
	3.18	5	.946	3.26	.51	3.63	.16
		25	.950	3.19	.45	3.72	.16
		50	.950	3.20	.42	3.35	.15
	7.15	5	.955	3.23	.36	2.83	.17
		25	.946	3.23	.36	3.09	.19
		50	.955	3.24	.36	2.96	.20
50	2.196	5	.954	3.28	.83	4.32	.43
		25	.962	3.46	3.12	12.16	1.14#
	3.18	5	.952	3.26	.77	3.78	.39
		25	.948	3.24	.99	4.32	.35#
	7.15	5	.947	3.29	.68	3.64	.42
		25	.940	3.28	.57	3.46	.43

* : Two "outliers" were not considered, therefore N=998.

: Three "outliers" were not considered, therefore N=997.

Fig. 7.2 : Simulation Study I. Graphs of EMSE for P2.



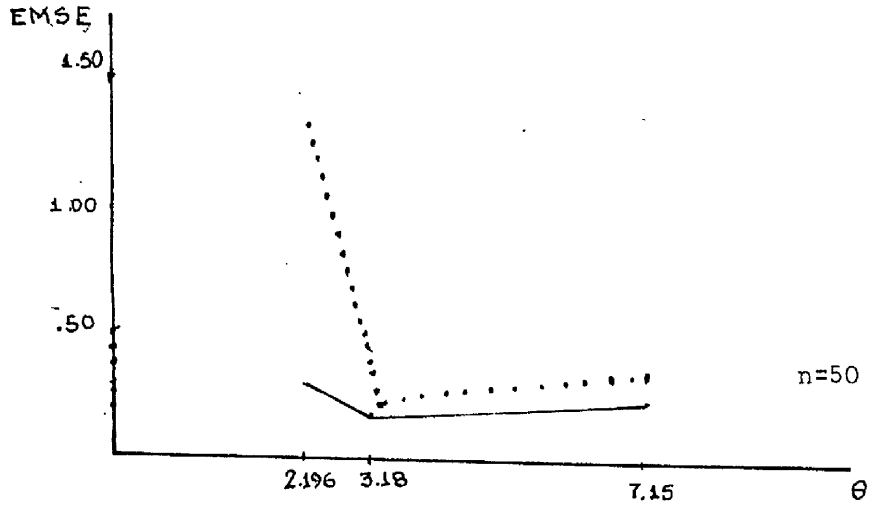
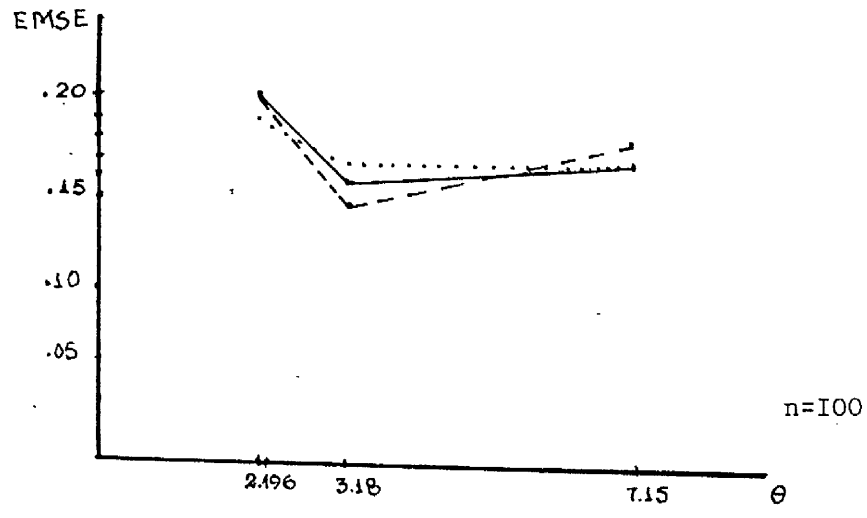


Table 7.3 : Simulation Study I, Procedure P3.

n	θ_{st}	r	ECP	$\bar{\theta}$	S	K	EMSE
100	2.196	5*	.937	3.23	.55	3.86	.20
		25	.944	3.21	.51	3.43	.19
		50	.948	3.23	.75	3.97	.20
	3.18	5*	.946	3.20	.51	3.63	.16
		25	.947	3.20	.51	3.28	.17
		50	.961	3.20	.46	3.31	.15
	7.15	5*	.955	3.23	.36	2.83	.17
		25	.947	3.20	.44	3.36	.17
		50	.949	3.19	.36	3.21	.18
50	2.196	5*	.954	3.28	.83	4.32	.43
		25	.942	3.21	1.09	1.28	.51
	3.18	5*	.952	3.26	.77	3.78	.39
		25	.953	3.26	.88	5.43	.38
	7.15	5*	.947	3.29	.68	3.64	.42
		25	.948	3.26	.43	2.98	.39

* : From Table 7.2 (r=r'=5).

Fig. 7.3 : Simulation Study I. Graphs of EMSE for P3.



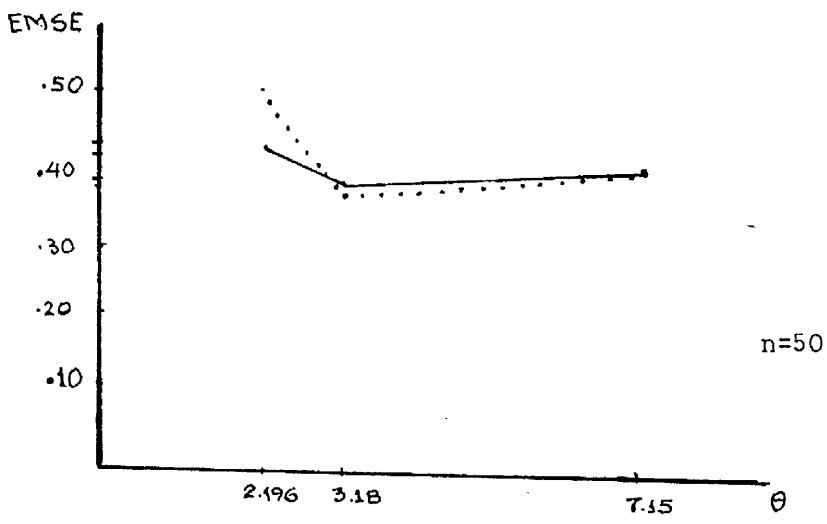


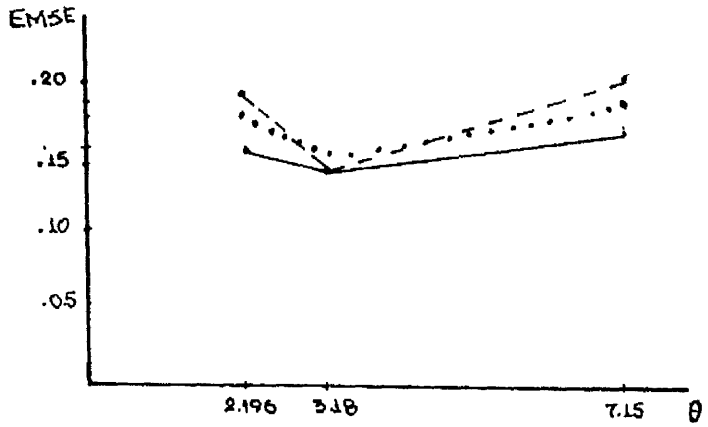
Table 7.4 : Simulation Study I, Procedure P4.

n	θ_{st}	r	ECP	$\bar{\theta}$	S	K	EMSE
100	2.196	5	.955	3.21	.45	3.53	.16
		25	.959	3.23	.48	3.21	.18
		50	.957	3.23	.60	3.83	.19
	3.18	5	.960	3.21	.46	3.33	.15
		25	.952	3.21	.75	4.37	.16
		50	.962	3.20	.59	3.84	.15
	7.15	5	.953	3.24	.32	2.77	.17
		25	.956	3.24	.41	3.23	.19
		50	.946	3.22	.37	3.25	.20
75	2.196	5	.943	3.22	.64	3.74	.24
		25	.951	3.24	.57	3.22	.25
		50	.954	3.25	.96	5.51	.28
	3.18	5	.948	3.23	.60	3.65	.23
		25	.945	3.23	.82	4.45	.23
		50	.955	3.23	.68	3.80	.22
	7.15	5	.958	3.25	.38	3.05	.23
		25	.953	3.26	.49	3.38	.26
		50	.941	3.21	.43	3.43	.28
50	2.196	5	.956	3.26	.67	3.53	.37
		25	.948	3.30	.92	4.28	.46
	3.18	5	.949	3.24	.89	3.89	.33$^{\natural}$
		25	.946	3.24	.98	4.04	.34*
	7.15	5	.948	3.26	.69	4.02	.38
		25	.948	3.29	.46	3.14	.44

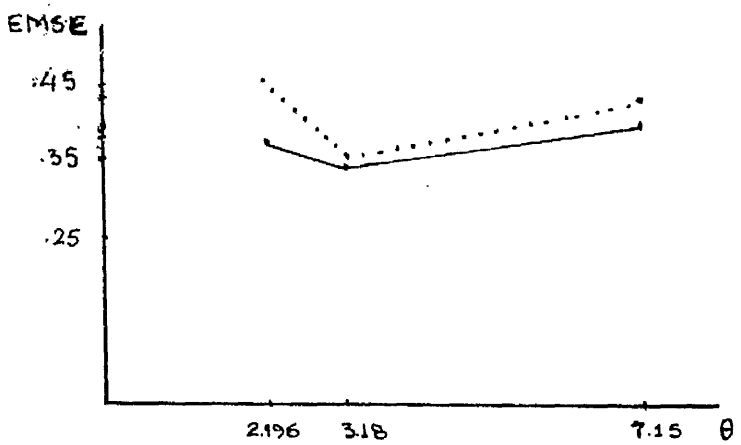
$\natural$:N=997

*:N=998

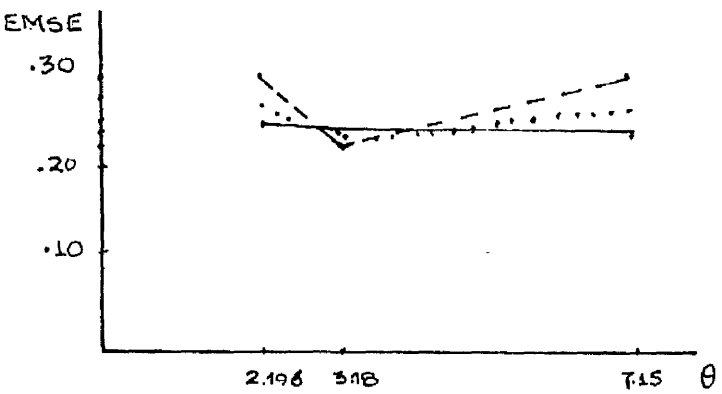
Fig. 7.4 : Simulation Study I. Graphs of EMSE for P4.



n=100



n=50



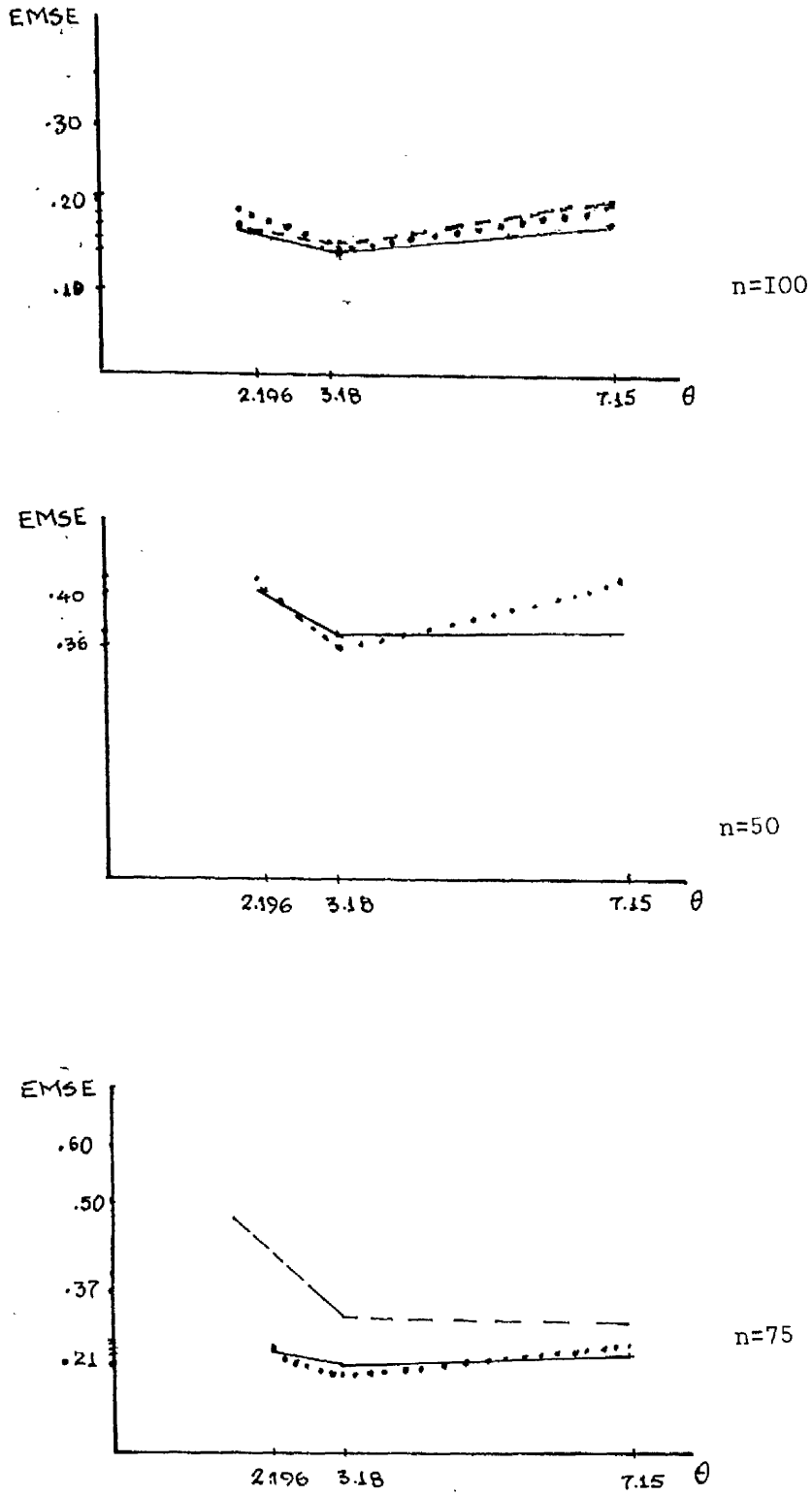
n=75

Table 7.5 : Simulation Study I, Procedure P5 .

n	θ_{st}	r	ECP	$\bar{\theta}$	S	K	EMSE
100	2.196	5	.961	3.24	.58	3.90	.17
		25	.946	3.23	.44	3.17	.18
		50	.956	3.23	.41	3.43	.17
	3.18	5	.967	3.23	.37	3.17	.15
		25	.958	3.21	.32	3.17	.15
		50	.954	3.21	.44	3.55	.16
	7.15	5	.958	3.28	.58	3.47	.17
		25	.952	3.23	.57	4.05	.18
		50	.954	3.21	.20	3.06	.18
75	2.196	5	.955	3.27	.64	3.78	.25
		25	.947	3.24	.55	3.46	.24
		50	.976	3.33	.86	3.90	.46*
	3.18	5	.961	3.24	.59	3.64	.22
		25	.955	3.22	.54	3.66	.21
		50	.955	3.25	.82	4.43	.37
	7.15	5	.951	3.24	.65	3.80	.23
		25	.951	3.25	.60	3.86	.25
		50	.951	3.22	.24	3.15	.23
50	2.196	5	.959	3.31	.70	3.65	.40
		25	.957	3.28	.82	4.61	.41
	3.18	5	.950	3.26	.78	4.11	.37
		25	.947	3.25	.73	4.24	.36
	7.15	5	.955	3.28	.58	3.47	.37
		25	.958	3.28	.82	4.52	.41

*:N=999

Fig. 7.5 : Simulation Study I. Graphs of EMSE for P5.



7.5 Discussion I

We now discuss the results described in section 7.4. Firstly we shall discuss each procedure and then we compare the procedures.

We comment that the total information for θ, and therefore the variance of θ can be evaluated (asymptotically) explicitly as

$$(nI(\theta,u))^{-1} = [nu^2/(\exp(\theta_t u)-1)]^{-1} \qquad (7.5.1)$$

when the design takes the n observations at the point u. Table 7.6 provides the values of $n^{-1}I^{-1}(\theta,u)$ for the design points we select to start the design under different sample sizes. Therefore, a guide for the evaluated mean squares is provided when a static design is performed so that it can be compared with the sequential procedures.

Table 7.6 : Evaluating the asymptotic variance.

u / n	.72	.50	.22
1000*	.0171	.0156	.0208
100	.171	.156	.208
75#	.231	.208	.278
50	.342	.312	.416

* : only for P1, # : only for P4,P5

We now consider each procedure separately.

- P1

We not only tried sample sizes n=100 or 50 but also a sample size of n=1000 (!) to study the asymptotic behaviour of the one point, one stage design. For n=1000, the EMSE is not too far from the expected value (see Table 7.6). As the sample size gets smaller, EMSE of course increases. The normality of the vector of estimates behaves quite well when n=1000. but it gets worse when n drops to 50.

Thus the sample size is very critical especially when we are near to the end points of the probability levels in (7.2.4). This happens when u_1=.72 which corresponds to probability level p=0.1, because of this truncation may take place, corresponding to a batch of y_i's all equal to zero. Thus the one-stage design does not have the opportunity to improve the estimate, when n=50, from this pathological situation.

- P2

When n=100 and r=50 (or n=50 and r=25) we have the so called two stages design (which we will discuss extensively in chapter 8). In the two stages design if, in the first stage the estimator is a "bad" one, the design does not have the opportunity to recover in the next stage. But when the initial estimate is "reasonable" it is improved in the second stage. This is less likely when the initial batch is 5 or 25 observations. The design behaves similarly with r=5,25 when n=100.

- P3

There is no two stage design in P3. When 5 observations are used in the first batch, P3 coincides with P2. When 25 observations were used as the first batch the EMSE obtained was slightly better than the equal batch procedure P2, when we start off from a point far from the true value. When r=50 the design had "enough time to recover" from a possible bad estimate at first stage. For this particular procedure there is not too much effect from the initial batch size and the e_{st} value chosen.

- P4

As only one observation was added at each stage, we also used the sample size n=75 as an intermediate stage between n=50 and n=100. Our aim was to check how far we can improve matters by adding only one observation. The performance of the procedure is largely independend of the initial batch size and the value e_{st}, although there is a little more variability when n=50.

- P5

There is little difference between P4 and P5. Under different sample sizes the EMSE are close to that of P4. The comments for P4 are similar to those of P5.

We now try to present a general view of this simulation study.

The procedures mentioned above can be divided into two categories

- One stage design (Procedure P1)
- Sequential design :

  Block design (Procedures P2,P3)

  Fully sequential design (Procedures P4,P5)

The results of this simulation study support the work of Ford and Silvey (1980). All the sequential procedures provided satisfactory coverage probabilities, i.e around .95, on the average, when 95% confidence intervals were to be constructed with performance getting better from P2 to P3 to P4. This encouraging result leads us to use the same approach for a two parameter model, presented later.

Wu (1985) applied the truncation idea in fully sequential design in a different way. Although he carried out only 500 simulations per case there are cases where, in 114 or 56 of them the initial estimator could not be evaluated and he choose to ommit these runs. Ofcourse he had a two parameter model. He faced the problem of existence the MLE and thus things were worse than in our case, in which

- With small batch size, i.e for r=5 or sometimes r=25, but never with r=50, it was quite likely for us to obtain either all successes (y_i's=1) or all failures (y_i's=0), in which case the MLE at the first stage could not be evaluated.

- At some stage of the sequential design, before the design had reached 50 observations the Newton-Raphson (Appendix A6.I) diverged.

We overcame these main difficulties, and we are consequently able to report 1000 simulations, using

- The truncation of the design.
- The Bisection method (Appendix A6.II).

Truncation helped us to "bring the design back" to the sample space we had defined as it is - needless to say - a waste of time to look outside of Θ for estimates of θ.

The Bisection method is a rather slow numerical method with convergence rate (Appendix A6.II) 1/2, whilst Newton-Raphson has at least second order convergence rate. The Bisection method has the advantage that no derivatives are needed. Thus with Bisection the initial value θ_0 to feed Newton-Raphson was in the neighbourhood of the solution.

Abdelbasit and Plackett (1981,1983) discuss the dilution series model in the context of constant information design. Indeed it was proved by Fisher (1922) that if $u=c\alpha^{-y}$, $c>0$, $\alpha>1$, $y=0,1,2,..$ with α and c constants (α is usually 2 or 10) then the information for $\log\theta$, $I(\log\theta)$ is approximately $\pi^2/6\log\alpha$, provided that α does not greatly exceed 1. Thus Abdelbasit and Plackett (1981) suggest using the constant information criterion for designing; see section 4.6.

We also compared the stochastic approximation scheme with a "full likelihood sequential procedure" in a small number of simulated experiments. In the latter maximum likelihood estimate was evaluated at each stage through Newton-Raphson iteration. Whereas in the former only "one iteration" was used. An initial batch of ten observations was chosen and various starting values for θ were suggested. We report in Fig.7.1 the study for the "true" $\theta=3.18$ and the "far" $\theta=7.18, 7.29$. At every stage up to 100 observations the estimates were obtained using the two techniques. In Fig.7.1 we denote by * those estimates which were obtained by stochastic approximation and by . those estimates which were obtained by Newton-Raphson. Both figures 7.1 provide evidence that the estimates do not differ. Therefore a gain in computing effort exists when, applying a "quick" calculation, the estimate is obtained by

stochastic approximation and is used to redesign. The speed of calculation in SA might lead to important improvements over a full "maximisation at each stage" procedure in control engineering problems.

As far as normality of the estimators is concerned acceptable behaviour is observed with n=100. It is less good when the sample size is reduced to n=50. We recall the saying of R.A. Fisher (1935) "Nothing that we say shall be true, except in the limit when the sample is indefinitely increased; a limit obviously never obtained in practice". This situation is described in P1 where with n=1000 everything seems to be acceptable, except the sample size. That idea of "practice" was behind this simulation study and it seems to us that n=100 is quite "large" and n=50 "reasonable".

Normality of the estimates was also investigated by using the NSCORES function of the MINITAB package (Ryan et al (1981)). The outcome is not presented in this thesis but gives similar evidence to the kurtosis and skewness measures. It improved from P2 to P3 to P4. Thinking in terms of probability levels p the values of θ_{st}=2.196 and 3.18 correspond to starting design points with values of p=0.1 and 0.2 respectively. That is low probability of successes compared with the p=0.5, the probability level of the "far" θ_{st} value 7.15. Therefore there is an "unbalanced" (as one of np or n(1-p) is close to n) set of 0's responses which influence the normality of the estimates corresponding to starting values 2.196 and 3.18. This seems to be more frequent in two stages design (Table 7.2 : θ_{st}=2.196, n=100, r=50 and r=25 or θ_{st}=3.18, 2.196, n=50, r=25) or in cases where the first batch produces such an "unbalanced" set of 0's that the design has no time to recover (Table 7.5 : θ_{st}=2.196, n=75, r=50).

Note that with respect to this problem the optimal design, that is optimal on an asymptotic criterion, has not got particularly

good properties for finite sample sizes. At the optimal design point p=0.2 and hence the probability of a run of 0's is increased. This point can be illustrated in the static design for θ_{st}=2.196 when n=50. In two of the 1000 simulations all 50 y's where equal to zero and hence θ was recorded 199.70, the upper bound of our allowed range for Θ. This means that the corresponding estimates of skewness and kurtosis and mean squared error will be greatly effected by such outliers. Note that in the tables when outliers occured they were ommited from the calculation of the quantitiesmentioned above, but not from the calculation of the estimated confidence probability.

Intuition suggests that if you have 50 or 100 observations to make and you do not know the optimal point, do not waste all your observations at one point but design sequentially. If you can design in blocks, as might be the case in chemical design, choose P2 or P3; if you design per observation, as in some psychological or engineering work, choose P4 or P5.

With the experience of this Simulation Study I behind us, we move on to Simulation Study II.

7.6 Two Graphs

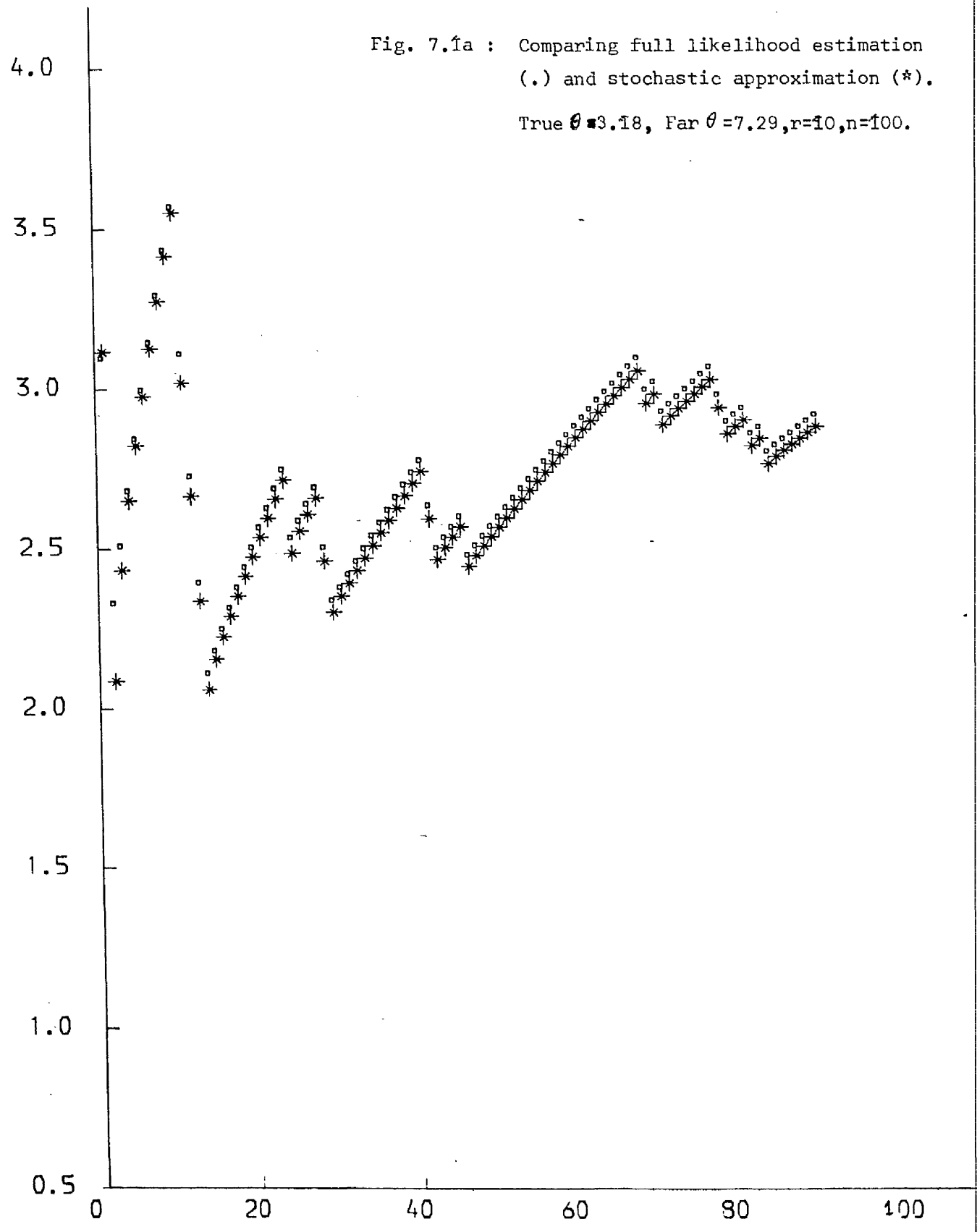
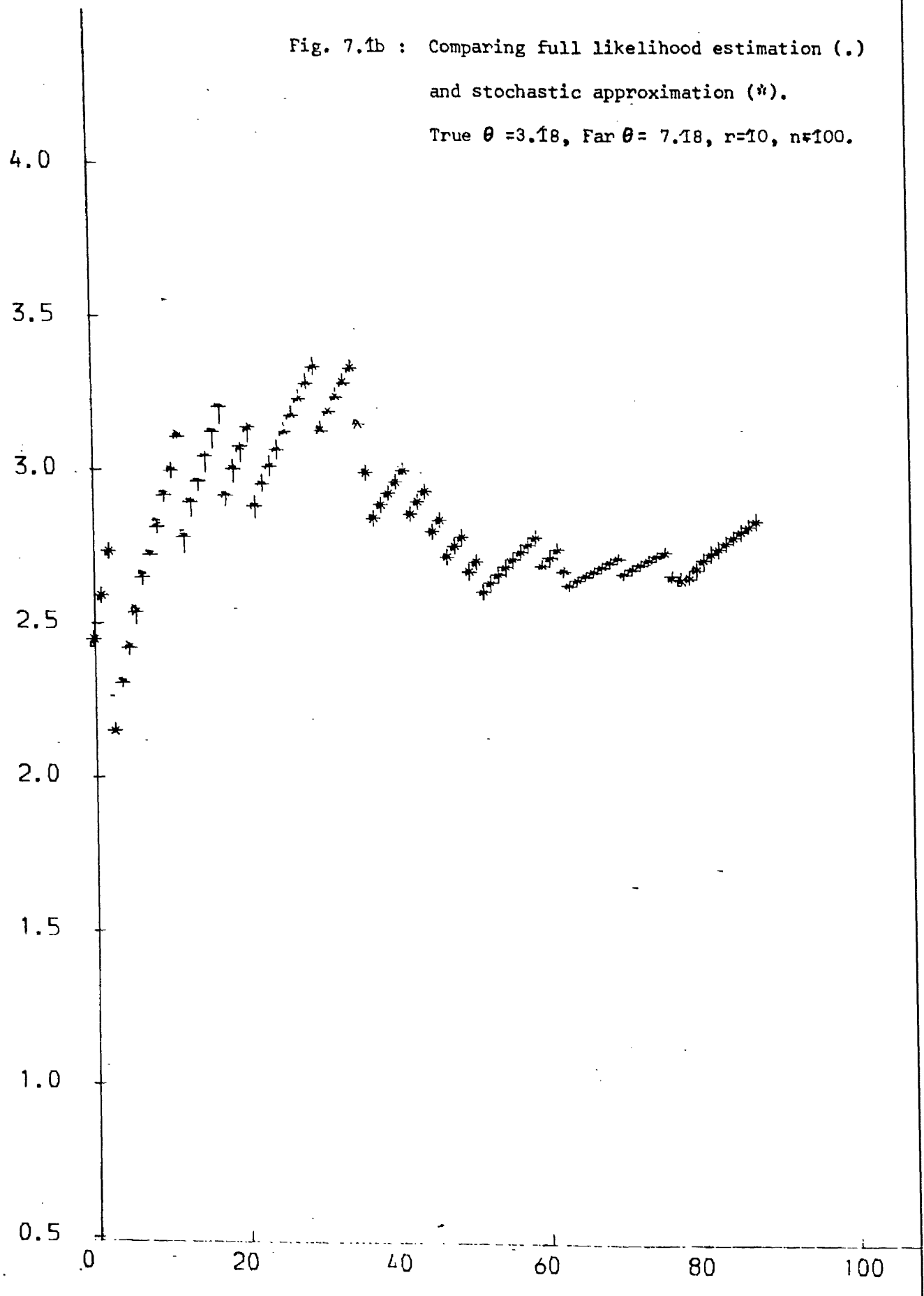


Fig. 7.1a : Comparing full likelihood estimation (.) and stochastic approximation (*). True θ =3.18, Far θ =7.29, r=10, n=100.

Fig. 7.1b : Comparing full likelihood estimation (.) and stochastic approximation (*).
True θ =3.18, Far θ = 7.18, r=10, n=100.



7.7 The first order growth law

Biological processes concerning a measure of growth, y say, of plants or animals can be expressed through a regression set-up, known as **growth law**.

Consider the period ranging from U_l up to U_u with u denoting time. The expected initial value of y (i.e when u=0) is denoted by $\theta_1>0$. The rate of increase of the biological process is denoted by $\theta_2>0$. The phenomenon can be described by the nonlinear regression model

$$y_i = \theta_1 \exp(\theta_2 u_i) + e_i, \quad i=1,2,\ldots,n \quad u\in U=[U_l,U_u]. \qquad (7.7.1)$$

For the error term e_i we use Assumption 2, when inference is made. Under the criterion of D(θ)-optimality Box and Draper (1959) considered this model and produced the locally optimal two point design which allocates half observation at the points

$$u_1 = U_u - 1/\theta_2 \quad , \quad u_2 = U_u \qquad (7.7.2)$$

where $\theta_2>0$. Therefore supplying a "guess" θ_{20}, say, for θ_2 a static design can be produced. For the so called decay model, i.e when $\theta_2<0$, the design points are

$$u_1 = U_l \quad , \quad u_2 = U_l - 1/\theta_2. \qquad (7.7.3)$$

The designs (7.7.2) or (7.7.3) depend only on θ_2, as it is partially-nonlinear (recall example 3.7.1).

Jennrich (1969) proved for this model the existence of the least squares estimates which, under Assumption 2, coincide with the maximum likelihood estimators. Our aim is to use this model under different sequential design procedures and to investigate the distribution of the parameter estimators as well as the construction of confidence intervals.

7.8 Strategy and procedures of Simulation II

The optimum design measure is $\xi^*=1/2$ for the two points (7.7.2). We kept the sample size n=40 and the following procedures (we refer to them as $\Pi 1$, $\Pi 2$,...) are discussed.

$\Pi 1$. Static design.

One stage design. Allocate half observations as in (7.7.2) providing θ_{20}, a guess for θ_{2t}.

$\Pi 2$. Two-stage design.

Use half the observations in the first stage, that is allocate one fourth at each of the optimal points. Obtain the estimates $\hat{\theta}$. Use them to redesign as in (7.7.2) using $\hat{\theta}_2$ instead of θ_2.

The other three procedures are sequential ones and only the number of stages is changed.

Π3. Five-stage design.

Use 8 observations in each stage.

Π4. Ten-stage design.

Use 4 observations in each stage.

Π5. "Fully-sequential" design.

Two observations at each stage, i.e one observation at each "optimal" point at each stage.

Any design space which is an interval on R can be transformed to U=[0,1] and local D(θ)-optimality remains invariant under linear transformations. Therefore we consider U=[0,1] through out the simulation. For the procedures we denote the maximum number of stages by S_{max}, S_{max}=1,2,5,10,20. The corresponding number of replications, at each optimal point, is r=20,10,4,2,1. We speed up the simulation by generating at each stage, only two normal deviates $\bar{y}_{j..}$, j=1,2 using the NAG subroutine G05DDF (Appendix A7.IV) where

$$\bar{y}_{j..} \sim N(\theta_{1t}\exp(\theta_{2t}u_j^*),\ \sigma^2/r)\ ,\ j=1,2 \qquad (7.8.1)$$

with u_j^* j=1.2 taken to be the optimal points, y_{jik} the k-th, k=1,...,r observation at stage i=1,...,S_{max} for the two points u_j^* j=1,2, as in (7.7.2). We start off the design with different θ_2 values, of the form θ_{2f}, $\theta_{2f}=\theta_{2t},\theta_{2t}\pm 2$, and θ_{2t}=1,2,3,4 the "true" value of θ_2. The value of θ_1 was kept constant, θ_1=10.0. The design points were evaluated according to

$$u^*_{1s+1} = 1 - 1/\hat{\theta}_{2s} \quad , \quad u^*_{2s+1} = 1 \qquad (7.8.2)$$

with $\hat{\theta}_{2s}$ being the estimate of θ_2 at stage s. For the first stage (s=1), in sequential procedures, or for the static design the estimate for θ can be evaluated explicitly (Appendix A7.V) as

$$\hat{\theta}_1 = \bar{y}_{11.}\exp(-\hat{\theta}_2 u_{11}), \quad \hat{\theta}_2 = [\ln\bar{y}_{11.} - \ln\bar{y}_{12.}]/u_{12}. \qquad (7.8.3)$$

In other stages, when s>1, the estimates were obtained through the modified Newton-Raphson scheme (Appendix A6.I). We settled on λ=0.5 as the modification parameter, i.e a "half-step" of the Newton-Raphson iteration was used to approach the solution, i.e to solve equations (2.7.9) (see also Appendix A7.VI).

At each stage the estimate $\hat{\theta}_{2s}$ was substituted into the Hessian (recall 2.7.10 or Appendix A7.VI) when $\hat{\theta}_{s+1} = (\hat{\theta}_{1s+1}, \hat{\theta}_{2s+1})$ was to be evaluated. The information matrix $M=M(\theta,\xi)$ (recall (2.7.9)) was evaluated at $\theta=\hat{\theta}(S_{max})$, i.e the estimate at the last stage. Simultaneous and individual approximate confidence intervals were produced through

$$(\theta-\hat{\theta})^T M(\theta,\xi)(\theta-\hat{\theta}) \leq 2s^2 F(2.38;.95) \qquad (7.8.4a)$$

$$\hat{\theta}_i \mp 1.96\sqrt{(M^{-1}_{ii}(\theta,\xi)s^2)}, \quad i=1,2 \qquad (7.8.4b)$$

respectively, with s^2 a suitable estimate of σ^2, i.e residual sum of squares divided by 38 df, F(.) as usual denoting the F distribution and $M^{-1}_{ii}(.)$ the diagonal elements of $M^{-1}(.)$.

Approximate confidence intervals are obtained through (7.8.4) when the design points are predetermined and not obtained sequentially. We followed the work of Ford and Silvey (1980) as in section 7.3 and applied this approach to sequential designs. The

coverage probabilities for both $\hat{\theta}_1$ and $\hat{\theta}_2$ individually and jointly were evaluated through the procedures mentioned above.

Moreover, the EMSE's (Estimated Mean Square Error) for θ_1 and θ_2 were evaluated as in (7.4.1), as well as $\log\det M_0$, with M_0 as the right hand side of (2.7.6a) with $\theta=\hat{\theta}(S_{max})$. The results are in Tables 7.9-7.13. Table 7.12 summarizes the estimated measures of skewness and kurtosis for the 1000 evaluated $\hat{\theta}_1$ and $\hat{\theta}_2$ under the different procedures. Table 7.12 provides measures of efficiency evaluated for θ_2 individually and for the design, i.e for θ. As such measures we used

$$\text{Eff}(\theta_i) = \frac{\text{EMSE(for } \theta_i \text{ in static design)}}{\text{EMSE(for } \theta_i \text{ in design under study)}}, \quad i=1,2 \qquad (7.8.5a)$$

$$\text{Eff}(\theta) = \frac{\log\det M_0 \text{ of design under study}}{\log\det M_0 \text{ for static design}} \qquad (7.8.5b)$$

Where the locally optimal design for θ_{2t} was used in the static design.

Table 7.7: Simulation Study II, Procedure III

θ_{2T}	θ_{2F}	Coverage probabilities			EMSE		ln det M_ω
		θ_1 and θ_2	θ_1	θ_2	θ_1	θ_2	
1.0	1.0	.943	.961	.954	50(.)	57(*)	5.470
	3.0	.954	.955	.953	138	168	5.096
2.0	2.0	.942	.955	.954	26	28	6.606
	4.0	.945	.952	.946	50	56	6.437
3.0	1.0	.947	.958	.958	46	46	7.201
	3.0	.961	.956	.954	8.4	9	7.991
	5.0	.943	.942	.941	12	13	7.895
4.0	2.0	.957	.968	.967	3.4	3	9.211
	4.0	.947	.940	.940	2.2	2.3	9.478
	6.0	.953	.953	.954	2.7	2.3	9.415

(.) All values x 10^{-3}
(*) All values x 10^{-5}

Table 7.8 : Simulation Study II, Procedure II2

θ_{2T}	θ_{2F}	Coverage probabilities			EMSE		Ln det M_σ
		θ_1 and θ_2	θ_1	θ_2	θ_1	θ_2	
1.0	1.0	.955	.962	.959	47(•)	53(*)	5.470
	3.0	.945	.952	.954	73	88	5.355
2.0	2.0	.943	.939	.942	30	32	6.606
	4.0	.945	.950	.950	35	40	6.541
3.0	1.0	.953	.949	.949	14	14	7.758
	3.0	.946	.954	.947	9	9.5	7.990
	5.0	.954	.959	.960	10	10	7.950
4.0	2.0	.957	.968	.971	2.4	2	9.367
	4.0	.952	.943	.942	2.2	2.3	9.478
	6.0	.949	.968	.969	2.2	2	9.451

(.) All values x 10^{-3}
(*) All values x 10^{-5}

Table 7.9 : Simulation Study II , Procedure II3

θ_{2T}	θ_{2F}	Coverage probabilities			Σ MSE		ln det M_o
		θ_1 and θ_2	θ_1	θ_2	θ_1	θ_2	
1.0	1.0	.945	.949	.952	50 (·)	55 (*)	5.470
	3.0	.944	.949	.950	56	65	5.435
2.0	2.0	.953	.956	.956	27	29	6.605
	4.0	.948	.944	.946	31	34	6.584
3.0	1.0	.941	.945	.943	10	10	7.913
	3.0	.950	.938	.942	9	9.5	7.991
	5.0	.946	.945	.944	9.4	10	7.977
4.0	2.0	.947	.943	.945	2.3	2.3	9.438
	4.0	.957	.960	.960	2	2.1	9.478
	6.0	.951	.953	.954	2.1	2.4	9.468

(.) All values $\times 10^{-3}$
(*) All values $\times 10^{-5}$

Table 7.10: Simulation Study II, Procedure II4

θ_{2T}	θ_{2F}	Coverage probabilities			E MSE		ln det M_θ
		θ_1 and θ_2	θ_1	θ_2	θ_1	θ_2	
1.0	1.0	.950	.941	.945	52 (·)	60 (*)	5.470
	3.0	.948	.938	.939	56	63	5.453
2.0	2.0	.955	.962	.960	26	28	6.606
	4.0	.942	.943	.943	30	32	6.595
3.0	1.0	.948	.941	.946	10	10	7.953
	3.0	.942	.940	.938	10	10	7.990
	5.0	.957	.964	.965	8	8	7.984
4.0	2.0	.953	.957	.957	2	2	9.458
	4.0	.950	.951	.950	2	2.2	9.478
	6.0	.955	.958	.960	2	2	9.473

(·) All values x 10^{-3}
(*) All values x 10^{-5}

Table 7.11: Simulation Study II, Procedure II5

θ_{2T}	θ_{2F}	Coverage probabilities			EMSE		ln det M_o
		θ_1 and θ_2	θ_1	θ_2	θ_1	θ_2	
1.0	1.0	.944	.955	.953	46 (-)	53 (*)	5.471
	3.0	.954	.950	.955	52	59	5.463
2.0	2.0	.936	.931	.935	30	32	6.605
	4.0	.946	.954	.951	27	30	6.601
3.0	1.0	.946	.944	.946	9.5	10	7.973
	3.0	.956	.951	.950	8.7	9.4	7.990
	5.0	.946	.944	.946	9	9	7.987
4.0	2.0	.957	.947	.945	2.2	2	9.468
	4.0	.941	.944	.940	2.4	2.5	9.478
	6.0	.958	.970	.971	1.9	2	9.476

(.) All values x IO^{-3}

(*) All values x IO^{-5}

Table 7.12: Simulation Study II, Skewness and Kurtosis for different procedures for the parameter θ_2

Procedure		Π1		Π2		Π3		Π4		Π5	
θ_{2T}	θ_{2F}	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
1.0	1.0	-0.001	2.91	0.015	2.67	0.074	2.99	0.042	2.88	0.020	3.16
	3.0	-0.002	3.01	0.092	3.05	0.125	3.29	-0.066	3.00	0.092	2.87
2.0	2.0	-0.012	2.98	-0.023	3.17	0.005	2.60	-0.008	2.83	0.083	3.24
	4.0	0.001	3.16	0.014	2.83	-0.101	2.85	0.019	2.58	0.020	3.17
3.0	1.0	0.097	3,40	0.020	2.96	-0.135	2.91	0.101	2.85	0.020	2.86
	3.0	-0.052	2.97	-0.039	2.94	-0.045	2.84	-0.043	3.09	0.019	2.69
	5.0	0.044	2.74	0.014	3.04	0.067	2.86	0.009	2.90	0.035	3.06
4.0	2.0	-0.034	2.88	0.034	2.76	0.038	2.94	-0.195	2.99	0.089	2.88
	4.0	-0.013	2.90	-0.001	3.20	0.064	2.66	0.054	3.13	0.010	2.97
	6.0	0.019	3.09	-0.064	2.99	-0.005	2.89	-0.194	2.98	-0.109	2.94

(1) Skewness θ_2

(2) Kurtosis θ_2

Table 7.13: Simulation study II, Evaluating efficiencies for θ_1.

Procedure		Π_1	Π_2	Π_3	Π_4	Π_5
θ_{2T}	θ_{2F}					
1.0	1.0	1.0	1.06	1.0	.96	1.08
	3.0	.36	.68	.89	.89	.96
2.0	2.0	1.0	.86	.96	1.0	.81
	4.0	0.5	.74	.83	.86	.96
3.0	1.0	.20	.60	.84	.84	.88
	3.0	1.0	.93	.93	.84	.96
	5.0	.75	.84	.89	1.05	.93
4.0	2.0	.64	.91	.95	1.10	1.0
	4.0	1.0	1.0	1.10	1.10	.91
	6.0	.81	1.0	1.04	1.10	1.15

Table 7.14: Simulation Study II, Evaluating efficiencies for θ_2 and $\theta=(\theta_1,\theta_2)$ under different procedures

Procedure		$\Pi 1$		$\Pi 2$		$\Pi 3$		$\Pi 4$		$\Pi 5$	
θ_{2T}	θ_{2F}	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
1.0	1.0	1.0	1.0	1.0	1.0	1.03	1.0	.95	1.0	1.07	1.0
	3.0	.34	.93	.64	.98	.88	.99	.90	.99	.97	.99
2.0	2.0	1.0	1.0	.87	1.0	.96	1.0	1.0	1.0	.87	1.0
	4.0	.50	.97	.70	.99	.82	.99	.88	.98	.93	.99
3.0	1.0	.20	.90	.64	.97	.90	.99	.90	.99	.90	.99
	3.0	1.0	1.0	.95	1.0	.95	1.0	.90	1.0	.96	1.0
	5.0	.69	.99	.90	.99	.90	1.0	1.12	.99	1.0	.99
4.0	2.0	.76	.97	1.15	.99	1.0	.99	1.15	.99	1.15	.99
	4.0	1.0	1.0	1.0	1.0	1.09	1.0	1.04	1.0	.92	1.0
	6.0	1.0	.99	1.15	1.0	.96	.99	1.15	.99	1.15	1.0

(1) Efficiency for θ_2 recall (7.8.5a)

(2) Efficiency for θ recall (7.8.5b)

7.9 Discussion II

The $\log\det M_0$, of course, achieves its maximum value when $\theta_{2f}=\theta_{2t}$ i.e $M(\theta,\xi)$ becomes large at the "true" values, that is the "local" ellipsoid is minimum at that point, recall section 3.7. The EMSE are, as expected, smaller when $\theta_{2f}=\theta_{2t}$. There is not much difference when we approach the true value either from smaller or larger values.

As far as the coverage probabilities are concerned on the average they are close to .95 and all the methods perform well. Among the sequential procedures ($\Pi 3, \Pi 4, \Pi 5$), the fully sequential procedure, $\Pi 5$, leads to better EMSE. The normality of the vectors of estimates, obtained from 1000 simulations, seems to behave very well. Table 7.14 provides evidence for this as all kurtosis values are very close to 3 and the skewneses very close to zero.

The efficiency of the static design for θ_2, when we start off from values far from the true value, is rather poor, since in one-stage design θ_1 is treated as known and there is no chance - as there is other stage - for the estimator to deviate much from its true value, this supports the adoption of the sequential design procedure. Table 7.14 supports the comment that the efficiency in the procedures is getting on the average better in the order $\Pi 1 < \Pi 2 < \Pi 3 < \Pi 4 < \Pi 5$.

Simulation Study II extends the results of Ford and Silvey (1980) for a two parameter nonlinear model.

Thus the sequential nature of the design for nonlinear models

- May often be irrelevant to the manner of obtaining estimators and constructing confidence intervals based on familiar sampling theory methods.

- There are cases in which sequential design procedures can result in "tighter" inferences, i.e shorter confidence intervals. Among them the fully sequential design might provide the tightest inference.

Thus, although the static design for the true θ might be experimentally economical, the absence of knowledge about θ suggests that a sequential procedure should be adopted. The inference can be obtained as in the static case.

CHAPTER 8

TWO-STAGE DESIGN

8.1 Introduction

In chapter 7 the sequential method of design was adopted and different procedures were applied. The idea of the two-stage design was discussed. Now we apply this method of design in the calibration problem. Here, the nonlinear feature of the problem under investigation will evolve from an underlying linear model.

The geometry of the problem and the application of Elfving's theorem will also be discussed. The maxi-min efficiency design is discussed along with the results obtained from a further simulation study. We shall refer to this simulation study as Simulation Study III. The procedures followed are discussed in section 8.6.

8.2 The problem

Consider the regression model with

$$\eta = E(y|u) = \theta_0+\theta_1 u_1, \quad u_1 \epsilon U=[-1,1], \qquad (8.2.1)$$

where U is the design space. Our target is to estimate the value of $u_1=u_0$ given $\eta=C$ i.e,

$$u_0 = (C-\theta_0)/\theta_1. \qquad (8.2.2)$$

We consider two design procedures

- One-stage
- Two-stage.

For the one-stage design, we might use as criterion function Φ, (recall section 3.3), either D-optimality for (θ_0,θ_1) or c-optimality for estimating u_0. The D-optimal design is of interest because it will be intepended of θ since we have a linear model. It is of interest to investigate the effectiveness of the D-optimal design as measured by the c-optimality criterion. Under c-optimality, thanks to Elfving's theorem, we can construct a locally optimal two point design geometrically. The criterion we would like to use is

$$\min \mathrm{Var}(\hat{u}_0) \qquad (8.2.3)$$

with $\hat{u}_0$ given by (8.2.5b) below. Strictly speaking $\mathrm{Var}(\hat{u}_0)$ does not exist, however the asymptotic formula for $\mathrm{Var}(\hat{u}_0)$ is still useful for the construction of confidence intervals as long as θ_1/σ is not small. In the sequel $\mathrm{Var}(\hat{u}_0)$ will refer to the asymptotic formula.

We can proceed by considering u_0 as a function of θ_0 and θ_1. That is, with $u_0=u_0(\theta_0,\theta_1)$.

$$\nabla u_0 = (-1/\theta_1)u, \text{ where } u=(1,u_0)^T. \qquad (8.2.4)$$

Then

$$\mathrm{Var}(\hat{u}_0) = \theta_1^{-2}\, u^T \mathrm{Var}(\hat{\theta}) u \qquad (8.2.5a)$$

with

$$\hat{u}_0 = (C-\hat{\theta}_0)/\hat{\theta}_1. \qquad (8.2.5b)$$

In the sequel one stage and two-stage designs will be discussed, with an aim of estimating u_0 as well as possible.

8.3 One-stage design

From geometrical considerations it is clear that the optimal design measure ξ either under D-optimality or c-optimality can be defined by p and 1-p, the proportions of observations at the end points, of the design space, that is,

$$\xi = \xi(p) = \begin{cases} p & \text{at } +1 \\ 1-p & \text{at } -1. \end{cases} \qquad (8.3.1)$$

The corresponding design matrix M is of the form

$$M = M(p) = \begin{pmatrix} 1 & 2p-1 \\ 2p-1 & 1 \end{pmatrix}. \qquad (8.3.2)$$

We are interested in minimizing, from (8.2.5), the quantity v(p) where,

$$v(p) = (n\theta_1^2)^{-1} u^T M^{-1} u, \qquad (8.3.3a)$$

i.e

$$v(p) = (4n\theta_1^2 p(1-p))^{-1}(u_0^2+2u_0(1-2p)+1). \qquad (8.3.3b)$$

Under D-optimality we allocate half observation at the end points +1, -1. That is

$$p = 1-p = 1/2. \tag{8.3.4}$$

The value of (8.3.3b) under this design, V_D say, is

$$V_D = (n\theta_1^2)^{-1}(u_0^2+1). \tag{8.3.5}$$

The locally c-optimal design can be obtained using Elfving's theorem (appendix A3.I). The percentage of the observations p allocated at +1 will depend on u_0, i.e $p=p(u_0)$. The induced design space U_0 (recall section 3.7) has to be formed for the model (8.2.1). Therefore as $U_0=f(U)$, with $f(u_1)=\theta_0+\theta_1 u_1$, U_0 will be

$$U_0 = \{ v: (v_1,v_2),\ v_1=1,\ v_2=u_1 \in U \}. \tag{8.3.6a}$$

The induced design space U_0 and its reflexion $-U_0$ form the two line segments as in Fig. 8.1. The geometry of the design space gives evidence of the symmetry in the problem, as a square is formed centered at the origin. From Elfving's theorem, we may take the support points (recall definition 2.2.3) to be the end points, i.e

$$\text{Supp}(\xi^*) = \{-1,\ 1\}. \tag{8.3.6b}$$

The weight of observations at each point can also be evaluated according to Elfving's theorem.

Consider the following two cases :

(i) $|u_0|<1$

Take $u_0=u_1$ as in Fig. 8.1. Then it is easy to see that

$$p = (u_1+1)/2, \qquad 1-p=(1-u_1)/2. \tag{8.3.7}$$

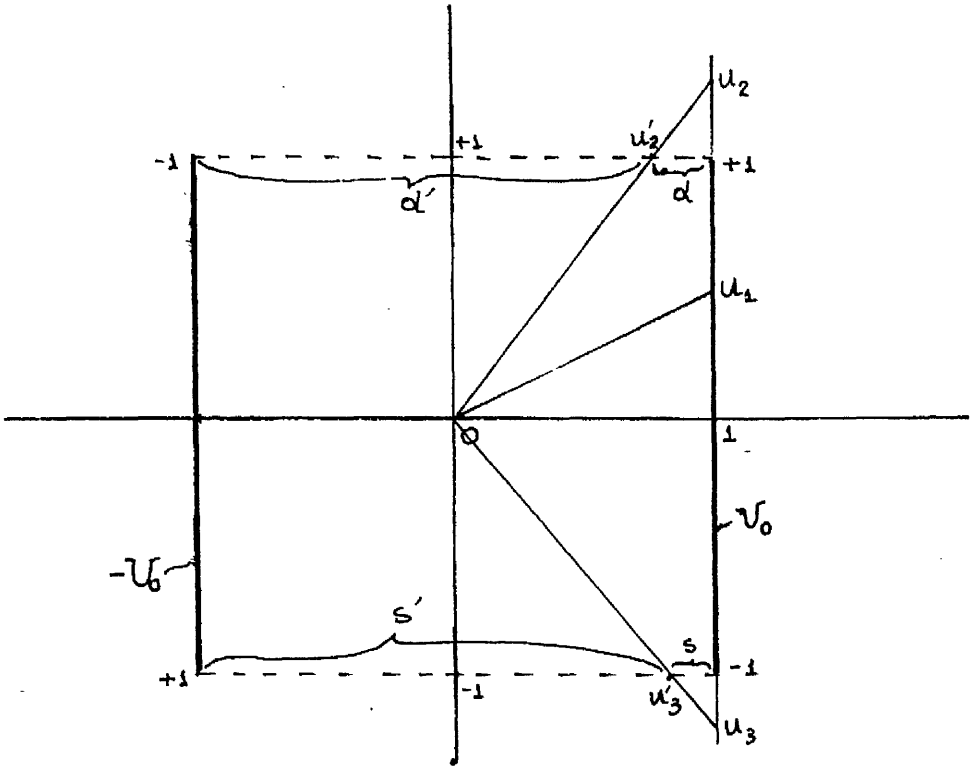


Fig. 8.1 The induced design space U_0 and its reflexion $-U_0$ for the model $\eta=E(y|u)=\theta_0+\theta_1 u$. $u\in[-1,1]$.

(ii) $|u_0|>1$

Take u_0 equal to u_2 and u_3 as in Fig. 8.1. Then:

- For u_2, from the similar triangles, it is easy to see that

$$d=(u_2-1)/u_2, \quad d'=2-d=(u_2+1)/u_2.$$

Hence,

$$p = (u_2+1)/(2u_2) \quad , \quad 1-p=(u_2-1)/(2u_2).$$

- For u_3 it is easy to see that

$$s=(u_3+1)/u_3 \quad , \quad s'=2-s=(u_3-1)/u_3.$$

Hence, we obtain in a symmetrical fashion,

$$1-p=(u_3+1)/u_3, \quad p=(u_3-1)/(2u_3).$$

To summarize the above discussion

$$p = \begin{cases} (u_0+1)/2 & \text{if } |u_0|<1 \\ (u_0+1)/(2u_0) & \text{otherwise} \end{cases} \qquad (8.3.8a)$$

$$1-p = \begin{cases} (1-u_0)/2 & \text{if } |u_0|<1 \\ (u_0-1)/(2u_0) & \text{otherwise.} \end{cases} \qquad (8.3.8b)$$

Therefore under this criterion the optimal v(p) value, V_C say, is (appendix A8.I)

$$V_C = \begin{cases} (ne_1^2)^{-1} & \text{if } |u_0|<1 \\ (ne_1^2)^{-1}u_0 & \text{otherwise.} \end{cases} \qquad (8.3.9)$$

There is interest in comparing (8.3.5) and (8.3.9). We use the efficiency measure, Eff(C,D) say, to assess the efficiency of the

D-optimal design relative to the local c-optimal design.

$$\text{Eff}(C,D) = U_C/U_D = \begin{cases} (u_0^2+1)^{-1} & |u_0|<1 \\ u_0^2/(u_0^2+1) & \text{otherwise.} \end{cases} \qquad (8.3.10)$$

A plot of Eff(C,D) vs u_0 is shown in Fig. 8.2. Fig. 8.2 indicates that the efficiency at the end points ∓ 1 is 50%. Moreover the D-optimal design at any other points in the interval (-1, 1) has a greater efficiency. The efficiency also increases outside the end points.

It is of interest to investigate the maxi-min efficiency design for this particular problem.

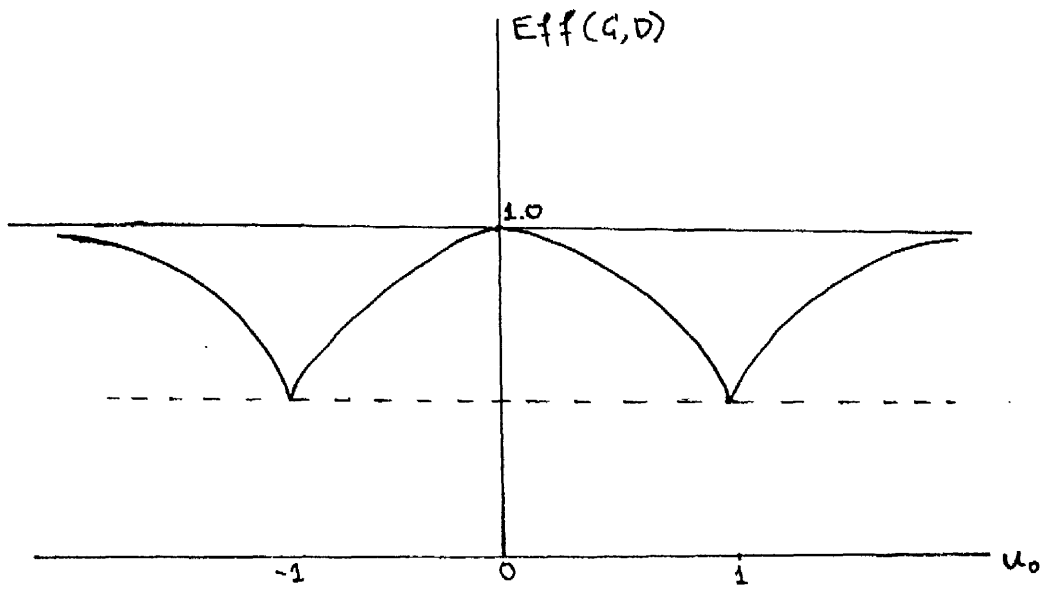


Fig. 8.2 The efficiency of the calibration design for different values of u_0.

8.4 Maxi-min efficiency design

Recall the maxi-min efficiency design measure introduced as another approach to static design in section 4.5. That is, we choose that design of the form $\xi(p)$ whose minimum efficiency $E(\xi)$ is greater than that of any other design i.e

$$\max\{ \min [Eff(\xi(p), u\epsilon U], \xi\epsilon\Xi \}, \tag{8.4.1}$$

recall (4.5.1) and (4.5.3) with $\Delta = /$.

From the discussion in section 8.3, interest in maximum efficiency design for the special problem means interest in

$$\max_{p} \min_{u_o} [V_i(u_o)/v(p)] \tag{8.4.2}$$

where $v(p)$ is as in (8.3.3b) and V_i is either V_D as in (8.3.5) or V_C as in (8.3.9). Then comparing the approximate variances from the general design $\xi(p)$ with the D-optimal design and the locally c-optimal $v(p)$ we can evaluate the following efficiencies.

- For D-optimality

$$Eff[\xi(1/2)] = V_D/v(p) = (u_o^2+1)/v(p). \tag{8.4.3}$$

- For $\xi(p)$

$$Eff[\xi(p)] = \begin{cases} [4p(1-p)/(u_o^2+2u_o(1-2p)+1)] & |u_o|<1 \\ [4u_o^2p(1-p)/(u_o^2+2u_o(1-2p)+1)] & \text{otherwise.} \end{cases} \tag{8.4.4}$$

The minimum values of (8.4.4) are (Appendix, A8.II) p, 1-p and 4p(1-p). Therefore, according to (8.4.1) we are looking for a design which satisfies

$$\max_{p} \min \{p, 1-p, 4p(1-p)\}. \qquad (8.4.5)$$

A heuristic proof that p=1/2 is provided in Fig. 8.3. The simultaneous graph of the function in (8.4.5) illustrates that the maxi-min value occurs when p=1/2.

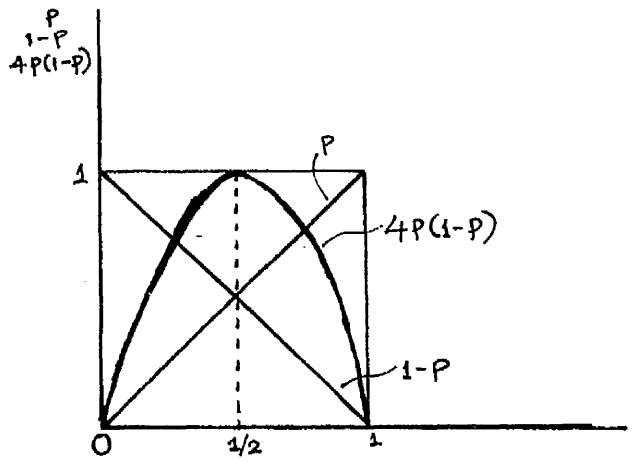


Fig. 8.3 A simultaneous sketch of the functions p, 1-p, 4p(1-p).

Therefore for the calibration design problem curently under discussion the maxi-min efficiency design turns out to be the D-optimal design.

8.5 Two-stage design

It is of interest to compare the maxi-min design with the two stage design which makes of use of our knowledge of the locally c-optimal design.

As the true values of θ_0, θ_1 are not known the experiment must obtain estimates of them. This might be done by using a portion of observations p_0 say to obtain these estimators, $\hat{\theta}_0$, $\hat{\theta}_1$, say, under D-optimality i.e allocating $p_0/2$ observations at each of the end design points. Then we estimate

$$\hat{u}_0 = (C-\hat{\theta}_0)/\hat{\theta}_1 \qquad (8.5.1)$$

with C as in (8.2.2). Therefore this value can be used as if it were the true one and the remaining proportion $1-p_0$ of the observations can be used to construct a locally c-optimal design. That is, depending on $\hat{u}_0$ we allocate the remaining observations at the end points according to Elfving's theorem.

Thus a two-stage, two-point design can be constructed to tackle this calibration problem. The advantage over the static design is certainly the use of an objective estimate of u_0, rather than a subjective guess. We study this method in a simulation study which we discuss in the next section.

8.6 Simulation Study

The ideas discussed above are applied in a simulation study. We let the true $\theta=(\theta_0,\theta_1)$ be $\theta_t=(0,1)$. Therefore $u_0=C$.

As different values of C we considered C=.1, .3, .5, .7, .9, 1.1, 1.3. The corresponding values of p, when a local c-optimal design is adopted, can be evaluated through (8.3.8a). In fact p=p(C) is calculated as p=.55, .65, .75, .85, .95, .954545, .884615 for the above values of C. When the D-optimal design is adopted, p=1/2. We let n=10,20,50. Problems can arise if σ, the standard deviation of the response is large. When σ is large the distribution of u_0 is unstable since $\hat{\theta}_1$ is likely to be occasionaly close to 0.0. In the results presented we shall assume σ=.25. Tables 8.1 and 8.2 summarize the outcomes obtained for the one stage design under c and D-optimality. For the one-stage design, under different values of C approximate confidence intervals were evaluated for $\hat{u}_0$. As stansdard error of $\hat{u}_0$, $se(\hat{u}_0)$ say, we used (recall (8.3.3b))

$$se(\hat{u}_0)=(v(p)s^2)^{0.5}, \tag{8.6.1}$$

with s^2 a suitable estimator of σ^2, namely s^2=RSS/(n-2), where RSS is the residual sum of squares. The values ECP and EMSE presented in Tables 8.1 and 8.2 are the estimated confidence probabilities and the estimated mean square error of $\hat{\theta}_1$, $EMSE(\hat{\theta}_1)$ say, evaluated as

$$EMSE(\hat{\theta}_1) = 10^{-3}\sum_{i}^{1000}(\theta_{1t}-\hat{\theta}_{1i})^2. \tag{8.6.2}$$

A two stage design simulation study was also carried out. For sample sizes n=10,20,50 we took p_0= .2,.4,.6,.8 as the proportion of observations in the first stage. Then we allocate the rest of the observations according to (8.3.8a). For the two-stage design the estimated value of u_0, $\hat{\hat{u}}_0$ say, was obtained after the second stage and an approximate confidence interval was constructed for $\hat{\hat{u}}_0$. As standard error for $\hat{\hat{u}}_0$, $se(\hat{\hat{u}}_0)$ say, we used (8.6.1) with the

appropriate p.

Tables 8.3 provide the average confidence limits for $\hat{\hat{u}}_0$. under different p_0 values for the specified u_0. Table 8.4 provides a comparison of the one stage D-optimal and local c-optimal designs with the two stage design. These designs were compared by standardising with respect to the mse for the c-optimal design. The efficiencies were evaluated and listed in Table 8.5 for the different designs constructed. The procedures described above will be denoted by π_1, π_2, π_3, namely,

- π_1 : One-stage - using D-optimality
- π_2 : One-stage - using local c-optimality
- π_3 : Two-stage design.

The simulation was carried out in a similar manner to simulation study II (see : Appendix A8.III).

Table 8.1

Simulation Study III - Procedure π_1

Values of ECP and √EMSE for different u_o values and different sample sizes

n	10		20		50	
u_o	ECP	√EMSE	ECP	√EMSE	ECP	√EMSE
0.1	.929	0.075	.945	0.057	.936	0.037
0.3	.926	0.082	.934	0.058	.935	0.038
0.5	.917	0.088	.936	0.064	.933	0.040
0.7	.933	0.097	.942	0.068	.941	0.044
0.9	.955	0.047	.935	0.077	.938	0.048
1.1	.912	0.119	.932	0.085	.942	0.052
1.3	.921	0.129	.945	0.090	.936	0.058

Table 8.2

Simulation Study III - Procedure π_2

Values of ECP and √EMSE for different u_0 values and different sample sizes

n	10		20		50	
u_0	ECP	√EMSE	ECP	√EMSE	ECP	√EMSE
0.1	.908	0.082	.942	0.056	.948	0.034
0.3	.919	0.080	.943	0.056	.939	0.035
0.5	.915	0.080	.935	0.055	.951	0.034
0.7	.911	0.081	.926	0.059	.946	0.036
0.9	.945	0.036	.940	0.056	.933	0.038
1.1	*	*	.934	0.066	.946	0.040
1.3	.929	0.106	.939	0.074	.946	0.047

* The corresponding p=.954545 so that no two point design can obtained with n=10.

Table 8.3a

Simulation Study III - Procedure π_3.

Values of ECP and √EMSE for different u_0 values.

Sample size : n=10.

p_0	0.2		0.4		0.6		0.8	
u_0	ECP	√EMSE	ECP	√EMSE	ECP	√EMSE	ECP	√EMSE
0.1	.907	0.082	.921	0.079	.921	0.080	.924	0.080
0.3	.906	0.081	.917	0.083	.903	0.082	.918	0.081
0.5	.895	0.082	.920	0.079	.904	0.084	.917	0.084
0.7	.908	0.081	.905	0.087	.898	0.087	.900	0.091
0.9	.935	0.038	.949	0.038	.944	0.041	.949	0.043
1.1	.885	0.096	.900	0.100	.893	0.107	.879	0.115
1.3	.915	0.105	.922	0.107	.913	0.115	.903	0.117

Table 8.3b

Simulation Study III - Procedure π_3.

Values of ECP and √EMSE for different u_o values.

Sample size : n=20.

p_o	0.2		0.4		0.6		0.8	
u_o	ECP	√EMSE	ECP	√EMSE	ECP	√EMSE	ECP	√EMSE
0.1	.937	0.056	.927	0.056	.930	0.056	.942	0.056
0.3	.932	0.057	.934	0.055	.933	0.058	.921	0.059
0.5	.916	0.057	.923	0.059	.930	0.059	.946	0.057
0.7	.931	0.057	.936	0.057	.927	0.061	.939	0.061
0.9	.928	0.600	.933	0.062	.927	0.063	.933	0.067
1.1	.934	0.064	.928	0.066	.921	0.071	.937	0.077
1.3	.930	0.077	.945	0.078	.947	0.077	.946	0.085

Table 8.3c

Simulation Study III - procedure π_3.

Values of ECP and √EMSE for different u_0 values.

Sample Size : n=50.

p_0	0.2		0.4		0.6		0.8	
u_0	ECP	√EMSE	ECP	√EMSE	ECP	√EMSE	ECP	√EMSE
0.1	.946	0.035	.956	0.034	.921	0.080	.948	0.035
0.3	.947	0.036	.940	0.036	.945	0.034	.943	0.037
0.5	.948	0.035	.946	0.037	.945	0.037	.953	0.038
0.7	.944	0.037	.949	0.037	.951	0.037	.948	0.041
0.9	.958	0.035	.938	0.039	.948	0.039	.938	0.044
1.1	.948	0.040	.950	0.042	.952	0.040	.943	0.048
1.3	.952	0.046	.952	0.048	.952	0.050	.938	0.054

Table 8.4

Simulation Study III

Summary of one stage and two-stage design comparing the efficiencies (*). Sample size n=10,20,50

u_o	D-optimal	Two stages design			
		0.2	0.4	0.6	0.8
0.1	1.0	0.99	1.0	1.0	1.0
	0.99	1.0	1.0	1.0	1.0
	0.92	1.0	0.99	1.0	0.98
0.3	0.98	0.98	0.97	0.98	0.99
	0.95	0.97	1.0	0.96	0.93
	0.92	0.96	0.97	0.94	0.92
0.5	0.91	0.98	1.0	0.95	0.95
	0.87	0.96	0.94	0.94	0.97
	0.84	0.97	0.94	0.93	0.90
0.7	0.83	1.0	0.92	0.93	0.89
	0.87	1.0	1.0	0.96	0.96
	0.83	0.99	0.99	0.93	0.88
0.9	0.77	0.96	0.96	0.88	0.84
	0.73	0.94	0.91	0.88	0.83
	0.78	1.0	0.97	0.95	0.86
1.1	-	-	-	-	-
	0.78	1.0	1.0	0.93	0.86
	0.76	0.99	0.95	0.86	0.83
1.3	0.82	1.0	0.99	0.92	0.90
	0.82	0.96	0.95	0.95	0.87
	0.81	1.0	0.97	0.92	0.86

8.7 Discussion

In this application we are estimating a nonlinear function of the parameters of a linear model and we are using an approximation for obtaining the variance of the estimator (8.2.5b).

As would be expected, under procedure π_2, the evaluated $\sqrt{EMSE}$ for each value of u_o is smaller than the $\sqrt{EMSE}$ under π_1. This holds for the different sample sizes n=10,20,50.

Overall, Tables 8.3 indicate that the policy of allocating p_o=0.8 i.e 80% of the observations at the first stage might provide the worst EMSE's. For n=50 the D-optimal design and the two-stage design with p_o=0.4 provide very similar EMSE's. This is not so when p_o=0.8. The situation is much better with p_o=0.2 with the smaller sample size n=10,20 in that the two-stage design, even with p_o=0.2, provides smaller EMSE's than the D-optimal static design. With small size n=10 the confidence intervals do not have the expected nominal level. There is a tendency to be lower, close to .90. With n=20 the results are improved, although still low. For n=50 the results are more reliable.

From Table 8.4, comparing the efficiencies of the designs in general, the two-stage design with p_o=0.2,0.4,0.6 is better than the D-optimal static design.

This simulation study, dealing with estimation in the calibration problem provides empirical evidence that efficient procedures can be achieved by the two-stage designs. It is interesting that the D-optimal design is also the maxi-min and, over the contexts covered in the simulations, performs fairly well but not as well as the two-stage design.

Abdelbasit and Plackett (1983) consider two-stage designs for the dilution series problem and for the two aprameter logistic problem. They examine efficiency based on approximation to the total information for the two stages. In their problems the first stage requires an initial estimate for the unknown parameters. Their results suggest that a two stage design given a particular starting estimate can be quite efficient relative to a design with more stages. Our results suggest that even with n not very large the two-stage design might prove useful, even when we use only 20% of our observations at first stage.

CHAPTER 9

OPTIMAL DESIGN IN RHYTHMOMETRY

9.1 Introduction

In this chapter we present a particular illustration of a nonlinear design problem, based on an application from physiology. The so-called **cosinor** model (Nelson et al. (1979)) has been proposed as a model for biological time series. An example of such a time series is that of circadian rhythms in airway calibre in normal and asthmatic patients. Normal subjects were recruited for this study and agreed to record their peak expiratory flow rate at different times of the day for a certain number of days. A form of the cosinor model, depending on clock or calendar time, has been applied by Hetzel and Clark (1980). Confidence intervals and related statistical analysis on the proposed model have been developed by Nelson et al. (1979).

We study the problem from the point of view of experimental design, that is what are the optimum times during the day that the measurements have to be recorded, how many times per day should the measurement take place, and how should these times be weighted optimally.

In the practical problem considered a nonlinear function of the parameters of the linearized cosinor model is to be estimated.

Various optimal design procedures are discussed from a geometrical and analytical point of view, and their efficiencies are compared with the locally optimum design. As the design depends on time we will replace u by t in the sequel. The unit for t is time in days.

9.2 Background

Some diurnal rhythms can be described by the following cosine model, known as the **cosinor model** in the biomedical literature.

$$y(t) = \eta(t,\theta) + \epsilon \qquad (9.2.1a)$$

with

$$\eta(t,\theta)=\theta_0+\theta_1\cos(\omega t+\theta_2) \qquad (9.2.1b)$$

where :

$y(t)$: is the response at time t, i.e the biological variable (rhythm) we want to study.

θ_0 : the **mesor**, i.e the "mean" value about which oscillation occurs.

θ_1 : the amplitude, i.e the half difference between the highest and lowest value during the osciilation in a complete cycle (360^0 or 24 hours).

θ_2 : acrophase, i.e timing of high point in degrees.

ω : angular frequency=degrees/unit time ($2\pi=360^0$ corresponds to a complete cycle). We consider $\omega=2\pi$ to correspond to a daily cycle.

ϵ : the error term under (recall chapter 2)

Assumption 2, when inference is made

Assumption 1, when the design is discussed.

$t \in [0,1]$

The model is illustrated in Fig. 9.1 below.

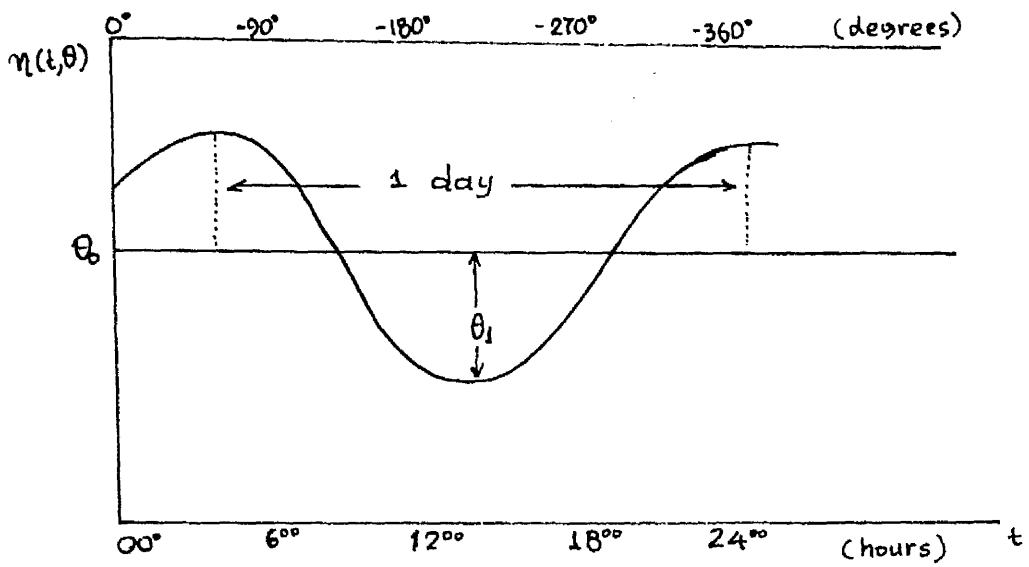


Fig 9.1 : A typical cosine function of the form

$$\eta(t,\theta)=\theta_0+\theta_1\cos(2\pi t+\theta_2)$$

From a clinical point of view the ratio θ_1/θ_0 is the parameter of interest. This represents the ratio of the amplitude of the cyclic variation to the overall mean.

We are assuming a period of one day, i.e we consider $\omega=2\pi$. The reference point for phase is 0^0 or 00.00 hours since $\cos 0^0=1$. Zero time is taken as 00.00 hours on the first day the study is started. It is easy to see, from Fig. 9.1, that the case $\theta_1/\theta_0<1$ is the only practical one in real life situations. When $\theta_1/\theta_0>1$ we would have negative values for η which has no physical meaning in the problem.

Expanding the cosine term, $\eta(t,\theta)$ we have

$$\eta(x,\theta) = \theta_0 x_0+\beta_1 x_1+\beta_2 x_2 \qquad (9.2.2)$$

with

$$\beta_1=\theta_1\cos\theta_2,\ \beta_2=-\theta_1\sin\theta_2,\ x_0=1,\ x_1=\cos 2\pi t,\ x_2=\sin 2\pi t \quad (9.2.2c)$$

Therefore model (9.2.1) can be written as

$$y(t) = W^T(t)\beta+\epsilon \ ,\beta=(\theta_0,\beta_1,\beta_2),\ W^T(t)=(x_0,x_1,x_2). \quad (9.2.3)$$

When the model (9.2.2) is fitted estimates for θ_1 and θ_2 can be obtained through the relations

$$\hat{\theta}_1 = \sqrt{(\hat{\beta}_1^2+\hat{\beta}_2^2)},\quad \hat{\theta}_2 = \hat{\omega}+\kappa \quad (9.2.4)$$

say, where $\hat{\omega}=\arctan|\hat{\beta}_2/\hat{\beta}_1|$ and κ is an appropriate constant. The value $2\theta_1$ is the peak to trough estimate and ω is the estimate of the phase of the rhythm i.e the time of the computed acrophase. For different values of $\hat{\beta}_1, \hat{\beta}_2$ we have (see Nelson et al (1979)).

$\hat{\beta}_1,\ \hat{\beta}_2 > 0$ then $\hat{\theta}_2 = -\hat{\omega}$ $(\kappa=0)$

$\hat{\beta}_1<0,\ \hat{\beta}_2>0$ then $\hat{\theta}_2 = -\pi+\hat{\omega}$ $(\kappa=-\pi)$

$\hat{\beta}_2<0,\ \hat{\beta}_1<0$ then $\hat{\theta}_2 = -\pi-\hat{\omega}$ $(\kappa=-\pi)$

$\hat{\beta}_1>0,\ \hat{\beta}_2<0$ then $\hat{\theta}_2 = -2\pi+\hat{\omega}$ $(\kappa=-2\pi)$

In the sequel the design problem will be discussed for a nonlinear function of the parameters of the linearized model (9.2.1), namely (9.2.3).

9.3 D-optimal design

For the model (9.2.1) clinicians are interested in efficient estimation of the relative stability of breathing i.e the ratio

$$g=g(\theta_0,\theta_1)=\theta_1/\theta_0 \qquad (9.3.1)$$

is to be estimated as well as possible.

Therefore we consider optimum designs for estimation of g. For the model (9.2.2) the design space, $\mathbf{X}$ say, is a circle, defined by

$$x_0 = 1, \quad x_1^2+x_2^2 = 1 . \qquad (9.3.2)$$

The centre of the circle is on the x_0 axis at point (1,0,0); see Fig. 9.2. It follows then (Fedorov. 1972 p.75) that the points of the D-optimal design must lie on the given circle. Moreover any equally-weighted design whose support coincides with the vertices of any regular polygon inscribed in the circle is D-optimal. For instance a four point, equally spaced and equally weighted design will be a D-optimal design. We comment that, in contrast, under c-optimality a two point unequal weighted design will be produced in section 9.4.

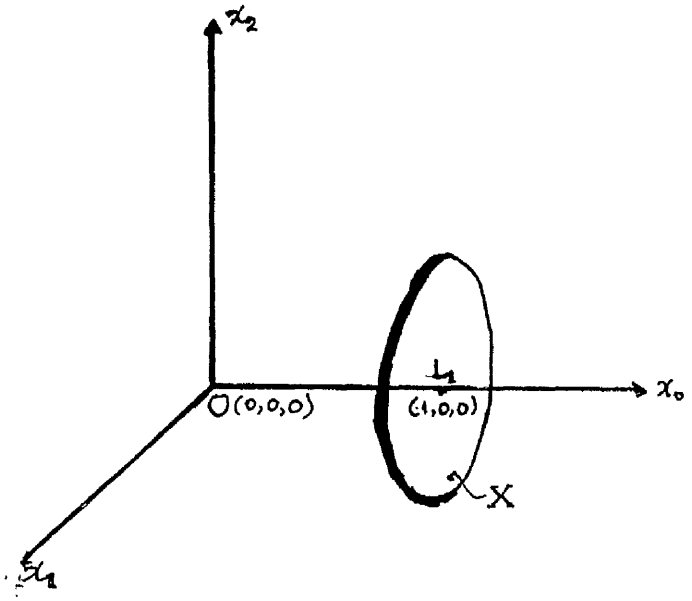


Fig. 9.2 : Position of the design space X

For the model (9.2.2) we evaluate the average information per observation matrix for this four point design, mentioned above. It equals (recall (2.5.8c))

$$nM(\xi) = \begin{pmatrix} n & \Sigma\cos(2\pi t_i) & \Sigma\sin(2\pi t_i) \\ \Sigma\cos(2\pi t_i) & \Sigma\cos^2(2\pi t_i) & \Sigma\cos(2\pi t_i)\sin(2\pi t_i) \\ \Sigma\sin(2\pi t_i) & \Sigma\cos(2\pi t_i)\sin(2\pi t_i) & \Sigma\sin^2(2\pi t_i) \end{pmatrix} \quad (9.3.3)$$

Take the 4 points to be

$$t,\ t+1/4,\ t+1/2,\ t+3/4$$

i.e, in angles,

$$2\pi t,\ 2\pi t+\pi/2,\ 2\pi t+\pi,\ 2\pi t+3\pi/2.$$

Let $\theta_i=2\pi t+\tau_i$, $\tau_i=0,\pi/2,\pi,3\pi/2$ it is easy to see that

$$\Sigma\cos\theta_i=\Sigma\sin\theta_i=0,\quad \Sigma\cos^2\theta_i=\Sigma\sin^2\theta_i=2. \quad (9.3.4)$$

Thus for n observations obtained in n/4 days (9.3.3) is reduced to

$$nM = \begin{pmatrix} n & 0 & 0 \\ 0 & 2(n/4) & 0 \\ 0 & 0 & 2(n/4) \end{pmatrix} = ndiag(1,1/2,1/2). \qquad (9.3.5)$$

Interest is in estimating (9.3.1) written as

$$g = (\surd(\beta_1^2+\beta_2^2))/\theta_0. \qquad (9.3.6)$$

Thus the approximate variance of g is

$$nVar(\hat{g}) \cong \sigma^2(\nabla g)^T M^{-1}(\nabla g) \qquad (9.3.7)$$

where ∇g is the vector of partial derivatives of g(.) and equals

$$(\nabla g)^T = (-(\surd(\beta_1^2+\beta_2^2))/\theta_0^2 \ , \ \beta_1/(\theta_0\surd(\beta_1^2+\beta_2^2)), \ \beta_2/\theta_0\surd(\beta_1^2+\beta_2^2))$$
$$= \theta_0^{-1} \ (-\theta_1/\theta_0, \ \beta_1/\theta_1, \ \beta_2/\theta_1) \qquad (9.3.8)$$

Substituting (9.3.8), (9.3.5) in (9.3.7) we obtain the approximate variance, V_4 say, of an equally-spaced equally-weighted 4-point design,

$$V_4 = (\sigma^2/(n\theta_0^2)) \ [(\theta_1/\theta_0)^2 + 2]. \qquad (9.3.9)$$

Note, from (9.3.7) that our problem is approximately equivalent to a locally c-optimal design, where "c" is given by (9.3.8).

9.4 C-optimal design

For given θ_0, θ_1, θ_2 and therefore β_1, β_2 the locally c-optimal design problem is to

$$\min \{c^T M^-(\xi)c,\ \xi\in\Xi\} \qquad (9.4.1)$$

with $c=(c_0,c_1,c_2)^T=\nabla g$ as in (9.3.7) and $nM(\xi)$ as in (9.3.3) i.e ξ imposes a measure which puts weight $1/n$ at $t_1,t_2,\ldots,t_n\in[0,1)$. In principle we require an optimal measure ξ^* on $[0,1)$ to solve (9.4.1). Elfving (1952) developed a geometrical approach to finding c-optimal designs. With this in mind, and considering the reflection, -X, of the design space X, a cylinder is formed, connecting X and -X with the x_0-axis as axis and directrix the circle X, see Fig. 9.3.

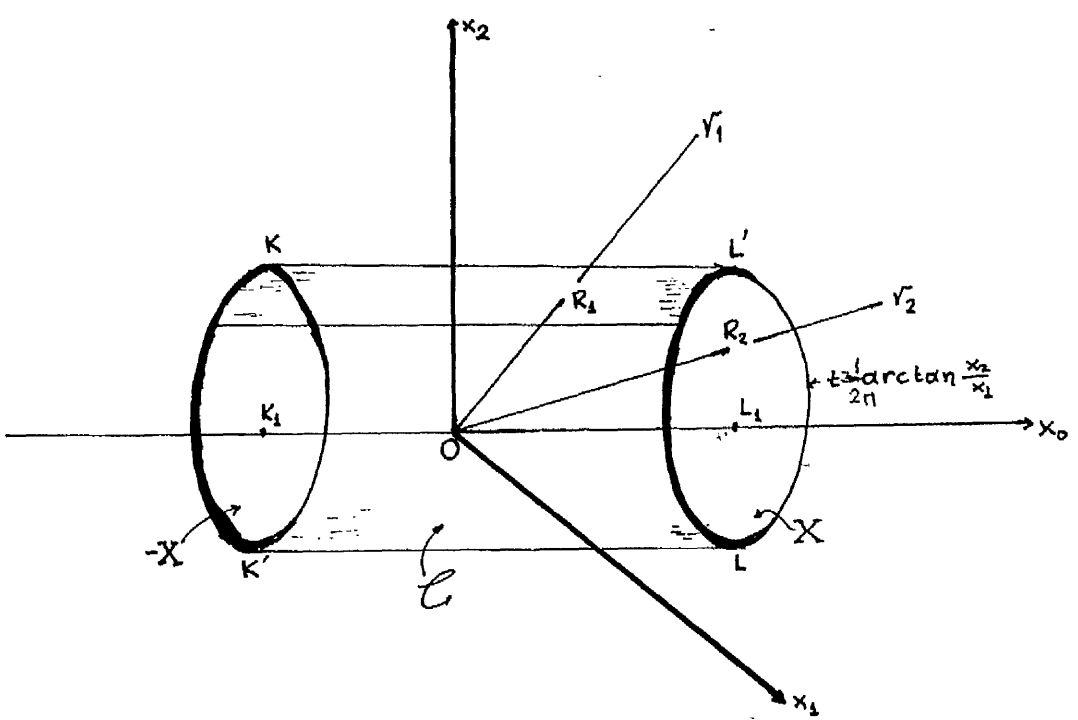


Fig. 9.3 The design space X and its reflection -X form the cylinder $\mathcal{C}$ for the model (9.2.2)

The equation of the cylinder is

$$\mathcal{C} = \{ (x_0, x_1, x_2) : -1 \leq x_0 \leq 1, \; x_1^2 + x_2^2 = 1 \}. \qquad (9.4.2)$$

Moreover any point on the cylinder $\mathcal{C}$ is either

(i) On the curved surface (ray r_1, point R_1 in Fig. 9.3) or

(ii) On one of the ends (ray r_2, point R_2 in Fig. 9.3).

Any ray, R say, can be written

$$R = \{ (\Delta c_0, \Delta c_1, \Delta c_2) \;, \; \Delta > 0 \} \qquad (9.4.3)$$

for some c_0, c_1, c_2. In particular, we consider the case where c_0, c_1, c_2 is as in (9.3.7). The ray hits $x_0=1$ at $\Delta = 1/c_0 = -\theta_0/\theta_1$ and therefore the point of intersection is $(1, c_1/c_0, c_2/c_0)$. We distinguish cases (i) and (ii) as follows :

$$\text{if} \quad (c_1^2+c_2^2)^2/c_0^2 \; \begin{cases} > 1 & \text{then } R_1 \text{ is considered} \\ < 1 & \text{then } R_2 \text{ is considered} \end{cases} \qquad (9.4.4)$$

It is easy to verify that

$$(c_1^2+c_2^2)^2/c_0^2 = (\theta_0/\theta_1)^2. \qquad (9.4.5)$$

The geometry of the problem suggests the use of Elfving's theorem (Appendix 3) to tackle the two cases described above.

- Case I : Consider points such as R_1, ie $\theta_1/\theta_0 < 1$. The side elevation for point R_1 is presented in Fig. 9.4. The ray hits the cylinder at $(\Delta c_0, \Delta c_1, \Delta c_2)$. Thus

$$(\Delta c_1)^2+(\Delta c_2)^2=1, \quad \Delta^2(c_1^2+c_2^2)=1, \quad \Delta=1/\sqrt{(c_1+c_2^2)}.$$

So the point R_1 is

$$R_1 = (\sqrt{(c_1^2+c_2^2)})^{-1} \ (c_0, c_1, c_2). \tag{9.4.6}$$

Applying Elfving's theorem and using the notation of Fig. 9.4 we get

$$KR_1/R_1L' = \xi/(1-\xi) \text{ i.e } \xi/(1-\xi) = (1+c_0/\sqrt{(c_1^2+c_2^2)})/(1-c_0\sqrt{(c_1^2+c_2^2)})$$
$$= (1-\theta_1/\theta_0)/(1+\theta_1/\theta_0) \tag{9.4.7}$$

where ξ, $1-\xi$ will be the weights assigned under the c-optimal design.

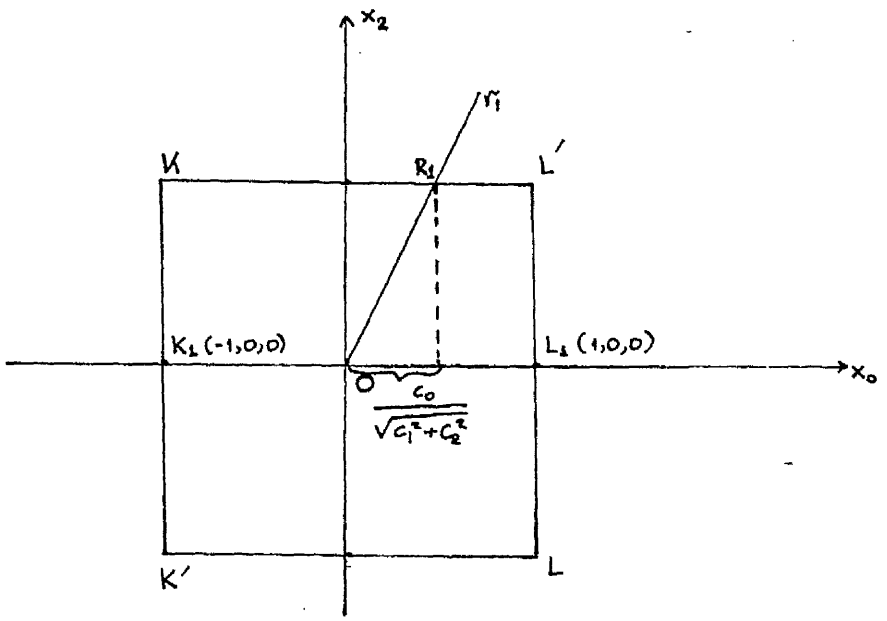


Fig. 9.4 : Side elevation for point R_1 from the cylinder $\mathcal{C}$ in Fig. 9.3.

Therefore from (9.4.7) we get

$$\xi = 0.5(1 - \theta_1/\theta_0), \quad \theta_1/\theta_0 < 1 \ . \tag{9.4.8}$$

The corresponding $2\pi t^*$ value will be (with t^* being the optimal value)

$$2\pi t^* = \arctan c_2/c_1 = \arctan((\beta_2/(\theta_1\theta_0))/(\beta_1/(\theta_1\theta_0))) =$$

$$\arctan\beta_2/\beta_1 = \tan^{-1}(-\tan\theta_2) = \tan^{-1}(\tan(-\theta_2)) = -\theta_2$$

Thus for $\theta_1/\theta_0 < 1$ allocate

$$\xi = 0.5(1-\theta_1/\theta_0) \text{ obs at } -\theta_2/2\pi$$

$$1-\xi = 0.5(1+\theta_1/\theta_0) \text{ obs at } \pi-\theta_2/2\pi. \tag{9.4.9}$$

For the two point design the corresponding 3x3 matrix $M=M(\theta,\xi)$ is singular with rank(M)=2. Considering the general form of $M(\theta,\xi)$ in (9.3.3) for this particular case it is easy to verify that under (9.4.9)

$$M(\theta,\xi) = \begin{pmatrix} 1 & (2\xi-1)\cos\theta_2 & -(2\xi-1)\sin\theta_2 \\ (2\xi-1)\cos\theta_2 & \cos^2\theta_2 & -\cos\theta_2\sin\theta_2 \\ -(2\xi-1)\sin\theta_2 & -\cos\theta_2\sin\theta_2 & \sin^2\theta_2 \end{pmatrix} \tag{9.4.10}$$

Substituting ξ in (9.4.10) from (9.4.9) we get, for the optimal design measure $\xi=\xi_2$ say

$$M(\theta,\xi_2) = \begin{pmatrix} 1 & -(\theta_1/\theta_0)\cos\theta_2 & (\theta_1/\theta_0)\sin\theta_2 \\ -(\theta_1/\theta_0)\cos\theta_2 & \cos^2\theta_2 & -\cos\theta_2\sin\theta_2 \\ (\theta_1/\theta_0)\sin\theta_2 & -\cos\theta_2\sin\theta_2 & \sin^2\theta_2 \end{pmatrix}.$$

To solve (9.4.1) the generalized inverse $M^-(\theta,\xi_2)$ is needed. Using a

matrix result (Rao (1965) p.26, Appendix A4.II (v)) we have $\left(\theta_2 \neq \frac{\pi}{2}\right)$

$$M^C(\theta,\xi_2) = \begin{pmatrix} \cos^2\theta_2 & (\theta_1/\theta_0)\cos\theta_2 & 0 \\ (\theta_1/\theta_0)\cos\theta_2 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \frac{1}{\cos^2\theta_2\left(1-\frac{\theta_1^2}{\theta_0^2}\right)} \quad (9.4.11)$$

Hence for $\xi=\xi_2$ and

$$c^T=(\nabla g)^T=(1/\theta_0)(-\theta_1/\theta_0,\ \beta_1/\theta_1,\ \beta_2/\theta_1)=(1/\theta_0)(-\theta_1/\theta_0, \cos\theta_2, -\sin\theta_2)$$

we obtain

$$c^T M^-(\theta,\xi)c = (1/\theta_0)^2 \quad (9.4.12)$$

and therefore the approximate variance V_2 for the two point design is

$$V_2 = \mathrm{Var}(c^T\hat{\theta}) = (1/\theta_0)^2\,\sigma^2/n, \quad \theta_1/\theta_0<1. \quad (9.4.13)$$

-Case II : Consider points such as R_2, i.e $\theta_1/\theta_0>1$. The sectional diagram showing point R_2 is presented in Fig. 9.5.

In section 9.2 we noted that for that particular problem our interest is restricted to $\theta_1/\theta_0<1$. For completeness however, we also discuss the other case now. Interest is focused on a result, following the same procedure as in Case I, which is complementary to that in (9.4.9).

Consider Fig. 9.5. Any chord of the type $\Gamma\Gamma'$, i.e going through R_2, corresponds to a design resulting in the point R_2. We choose the diameter D_1D_2 of the circle with centre $L_1(1,0,0)$. The point R_2 can be obtained by allocating weight ξ on D_2 and $1-\xi$ to D_1.

point R_2 can be obtained by allocating weight ξ on D_2 and $1-\xi$ to D_1. Therefore

$$\xi/(1-\xi)=D_2R_2/R_2D_1=[1+\Delta\sqrt{(c_1^2+c_2^2)}]/[1-\Delta\sqrt{(c_1^2+c_2^2)}]=$$

$$=[1+(\sqrt{(c_1^2+c_2^2)})/c_0]/[1-(\sqrt{(c_1^2+c_2^2)})/c_0]=(1-\theta_0/\theta_1)/(1+\theta_0/\theta_1).$$

Thus for $\theta_1/\theta_0>1$ the weights are

$$\xi=0.5(1-\theta_0/\theta_1)\ ,\quad 1-\xi=0.5(1+\theta_0/\theta_1). \tag{9.4.14}$$

Therefore there is a symmetry in the result obtained for both cases, I and II, as can be seen by comparing (9.4.14) and (9.4.9).

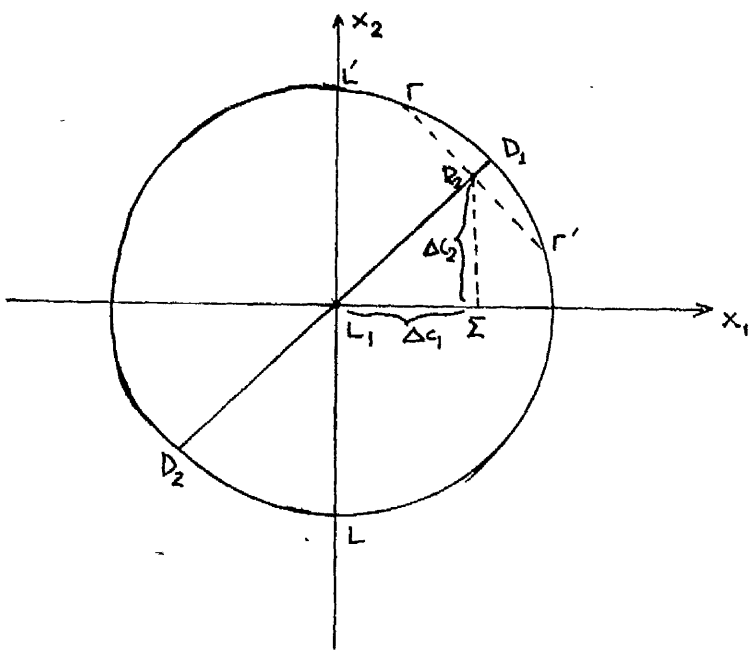


Fig. 9.5 : Sectional diagram for point R_2 from the cylinder $\mathcal{C}$ in Fig. 9.2.

9.5 Restricted Design space

An important practical difficulty with the optimal designs of section 9.4 is that they require measurements to be made when the response function is maximum and minimum. The latter typically occurs in the early hours of the morning.

It might be desirable for the design to be restricted to more social hours, i.e avoid taking measurements during the night. We restrict the design to a portion 1-T, say of the day, where T is the length of the night-time period, e.g 11pm till 7am, see Fig. 9.6. Moreover we assume that the minimum of the response function occurs at the middle of T and the maximum, in 1-T, occurs at the middle of this interval. Any design depends on θ_1/θ_0 and θ_2. Moreover the restriction on time means that the new design space, X_N, say, is no longer a circle and hence the idea of a full cylinder is no any longer applicable. The cylinder will be "truncated". say, as is shown in Fig. 9.7. A side elevation is presented in Fig. 9.8. Thus we have to reevaluate the equations of interest.

Firstly we evaluate the equation of the ray r_1. This is the equation of the line through the points $O(0,0)$ and $R_1(-\theta_1/\theta_0,1)$; see Fig 9.8. This is

$$y = (-\theta_0/\theta_1)x. \qquad (9.5.1)$$

(The point R_1 corresponds the case discussed in Fig. 9.4). The equation of the line through the points $L'(1,1)$ and $W(-1,\cos\Psi)$, with Ψ being the angle corresponding to the portion of 2π which is equivalent to T/2, is

$$(y-1)/(\cos\Psi-1) = (x-1)/(-1-1). \qquad (9.5.2)$$

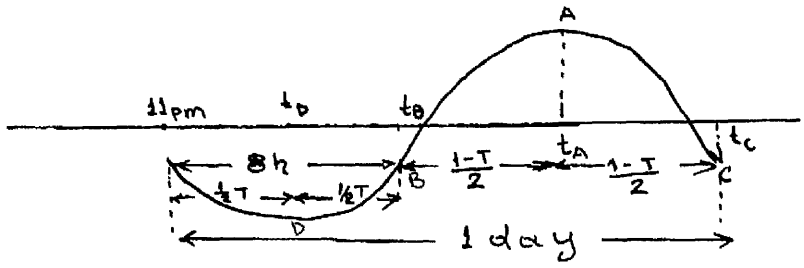


Fig 9.6 A typical situation when the rythm is "going down" during the night.

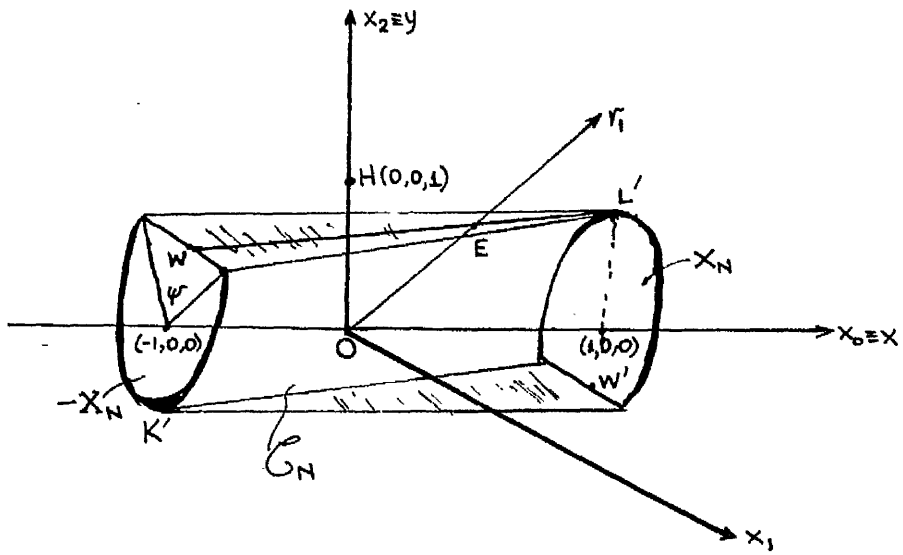


Fig. 9.7 The truncated space X_N and its corresponded truncated cylinder.

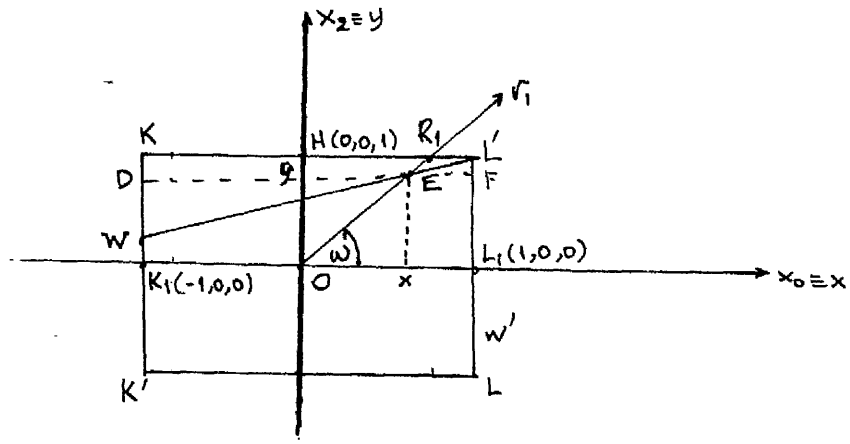


Fig. 9.8 Side elevation of Fig. 9.7

As T is the portion of the day which defines the design space, $\mathbf{X}$, we have $2\psi=2\pi T$. Therefore from (9.5.2) we get

$$y = 0.5[\cos(\pi T)+1] + 0.5[1-\cos(\pi T)]x. \qquad (9.5.3)$$

The coordinates of E, Fig. 9.8, are the solutions of the simultaneous equations (9.5.3) and (9.5.1) and therefore

$$x = 0.5[\cos(\pi T)+1]/(0.5\cos(\pi T)-0.5-\theta_0/\theta_1), \qquad (9.5.4)$$

giving a corresponding y.

It is easy to see that, with T=0, $x=-\theta_1/\theta_0$ and therefore y=1 i.e the point R_1 in the untruncated case. Now using Elfving's theorem (recall Fig.9.8)

$$(1-\xi)/\xi = DE/EF = (1+x)/(1-x) \qquad \xi = 0.5(1-x).$$

Therefore

$$\xi = [-0.5(\theta_0/\theta_1+1)]/[0.5\cos(\pi T)-0.5-\theta_0/\theta_1]. \qquad (9.5.5)$$

Therefore design weight $1-\xi$ is applied at A and 0.5ξ at each of B and C, under the symmetry assumption which we have imposed; recall Fig. 9.6. For definiteness we consider T=1/3, that is a restriction to a 16 hour period. The information matrix can be written in the form

$$M = (1-\xi)M_A + 0.5\xi M_B + 0.5\xi M_C; \qquad (9.5.6)$$

recall Fig. 9.6 and the regression (9.2.2c). It is easy to see that, for any T,

$$2\pi t_A+\theta_2=0 \text{ i.e } t_A=-\theta_2/2\pi, \quad t_D=t_A-1/2$$

$$t_B=t_D+(1/2)T=1/2[-\theta_2/\pi -1+T]$$

$$t_C=t_D-(1/2)T+1=(1/2)[-\theta_2/\pi +1-T].$$

If we set T=1/3 in the above relations, the vectors corresponding to $W(t_i)$ i=A,B,C are (recall (9.2.2c)) :

$$W(t_A)=(1,\ \cos\theta_2,\ -\sin\theta_2)^T$$

$$W(t_B)=(1,\ \cos(\theta_2+2\pi/3),\ -\sin(\theta_2+2\pi/3))^T$$

$$W(t_C)=(1,\ \cos(\theta_2-2\pi/3),\ -\sin(\theta_2-2\pi/3))^T \ .$$

We can therefore write $M_i=W(t_i)W^T(t_i)$, i=A,B,C. Moreover it is easy to prove the following Lemma.

Lemma 9.5.1 : The average per observation information matrix $M(\xi)$ as in (9.5.6) can be written as

$$M(\xi)=(1-\xi)A_1A_1^T+\xi A_2A_2^T+\xi A_3A_3^T$$

with

$$A_1=(1,\cos\theta_2,-\sin\theta_2)$$

$$A_2=(1,-1/2\cos\theta_2,1/2\sin\theta_2)$$

$$A_3=(0,\sqrt{3}/2\sin\theta_2,\sqrt{3}/2\cos\theta_2).$$

Our target is to evaluate c^TM^-c. Therefore the following Lemma provides the evaluation of the desired quanity.

Lemma 9.5.2 : The matrix $M(\xi)$ can be written as

$$M(\xi) = \lambda_c cc^T+\lambda_d dd^T+\lambda_e ee^T \qquad (9.5.7)$$

where the vectors c,d,e are orthogonal. Moreover

$$\rho = c^T M^{-}(\xi) c = \lambda_c^{-1} / [1 - \lambda_{ce}^2 \lambda_c^{-1} \lambda_e] \qquad (9.5.8)$$

For details of the proof see Appendix A5.III. The quantities in (9.57) and (9.5.8) can be proved equal to :

$$c^T = (1/\theta_0)(-\theta_1/\theta_0,\ \cos\theta_2,\ -\sin\theta_2)$$

$$d^T = (0,\ \sin\theta_2,\ \cos\theta_2), \quad e^T = (\theta_0/\theta_1,\ \cos\theta_2,\ -\sin\theta_2)$$

$$\lambda_c = (1-\xi)\lambda_1 + \xi\lambda_3, \quad \lambda_d = 3/4\ \xi,$$

$$\lambda_e = (1-\xi)\lambda_2 + \xi\lambda_4, \quad \lambda_{ce} = (1-\xi)\lambda_1\lambda_2 + \xi\lambda_3\lambda_4$$

with

$$\lambda_1 = [(1-\kappa)\theta_0]/[\kappa^2+1], \qquad \lambda_2 = [(\kappa+1)\kappa]/[\kappa^2+1]$$

$$\lambda_3 = [-\theta_0(2\kappa+1)]/[2(\kappa^2+1)], \quad \lambda_4 = [2\kappa-\kappa^2]/[2(\kappa^2+1)], \quad \kappa = \theta_1/\theta_0.$$

After some algebra we find that ρ is given by

$$\rho = (1/3\theta_0)^2[(4\kappa+1)^2(1-\xi) + (2-\kappa)^2\xi]/[\xi(1-\xi)], \quad \kappa = \theta_0/\theta_1. \qquad (9.5.9)$$

Therefore the approximate variance for the three point optimally-weighted truncated design is

$$V_3 = \mathrm{Var}(c^T\hat{\theta}) = \rho\sigma^2/n. \qquad (9.5.10)$$

The corresponding design weight ξ_3 can be evaluated from (9.5.5) with T=1/3 as

$$\xi_3 = [0.5(\kappa+1)]/[0.25+\kappa], \qquad (9.5.11)$$

Thus the design measure still depends on the fraction θ_1/θ_0 which we are trying to estimate. If we wanted to construct an

equally-weighted 3 point design in this truncated case, the design measure would be defined by $\xi=2/3$, with corresponding approximate variance

$$V_3^* = (2/3\theta_0^2)[2(\kappa+1)^2,(2-\kappa)^2]\sigma^2/n. \qquad (9.5.12)$$

In the next paragraph a synopsis is presented of the above results and the efficiencies of the designs are evaluated.

9.6 Synopsis

We summarize the results obtained in previous sections for the optimal design for estimating the function $g=\theta_0/\theta_1$ of the parameters θ_0, θ_1 of the model $\eta(t,\theta)=\theta_0+\theta_1\cos(2\pi t+\theta_2)$. The results are tabulated in Table 9.1.

For the results evaluated in section 9.3, 9.4, 9.5 efficiencies can be obtained. For the untruncated case, Table 9.1 compares the four point and the two point designs.

$$E_{2,4} = V_2/V_4 = 1/[(\theta_1/\theta_0)^2+2]. \qquad (9.6.1)$$

The truncated design compared with the untruncated unequally weighted design gives efficiencies

$$E_{2,3} = V_2/V_3 = [9\xi(1-\xi)]/[\kappa_1(1-\xi)+\kappa_2\xi] \qquad (9.6.2)$$

with

$\kappa_1=(\theta_1/\theta_0+1)^2$, $\kappa_2=(2-\theta_1/\theta_0)^2$, ξ as in (9.5.11).

For the equally weighted truncated design compared with the two point design

$$E^{*}_{2,3} = V_2/V^{*}_3 = 3[2(\theta_1/\theta_0)^2+(2-\theta_1/\theta_0)^2]^{-1}. \qquad (9.6.3)$$

For different values of $\theta_1/\theta_0<1$ the efficiencies and the design measures have been evaluated for the different designs in Tables 9.2 and 9.3. It is interesting that truncation does not greatly influence the nature of the design. Thus an equally weighted truncated design can be recommended. In practice, from previous experience the ratio θ_1/θ_0 does not exceed 0.3. In principle any optimal design, in the nonlinear case, depends on the parameters it is planning to estimate. In this particular case Table 9.2 reflects this dependence. However for small values of θ_1/θ_0 there is little difference between the optimal designs. Thus we choose the design procedure we adopt, we provide a guess for θ_1/θ_0 and the appropriate design measure can be evaluated, as in Chapter 4. That is we know what percentage of the omeasurements will be allocated to each point.

Table 9.1: Synopsis of the proposed designs for estimating $g = \frac{\theta}{\theta_o}$, $\frac{\theta_1}{\theta_o} < 1$. from the model $\eta(t,\theta) = \theta_o + \theta_1 \cos(2\pi t + \theta_2)$

Designs	Approximate Variance	Design Measures	Optimal design points
1. 4 equal spaced equally weighted points	$V_4 = \frac{\sigma^2}{n\,\theta_o^2}\left[\left(\frac{\theta_1}{\theta_o}\right)^2 + 2\right]$ (for every $\frac{\theta_1}{\theta_o}$)	$\xi_L^{(4)} = \frac{1}{4}$ L = 1,2,3,4	$\theta_1^* = 2\pi t$, $\theta_2^* = 2\pi t + \frac{\pi}{2}$ $\theta_3^* = 2\pi t + \pi$, $\theta_4^* = 2\pi t + \frac{3\pi}{2}$
2. 2 point unequally weighted c-optimal	$V_2 = \frac{\sigma^2}{n\,\theta_o^2}$	$\xi_1^{(2)} = \frac{1}{2}\left(1 - \frac{\theta_1}{\theta_o}\right)$ $\xi_2^{(2)} = 1 - \xi_1^2$	$\theta_1^* = -\theta_2 = 2\pi t$ $\theta_2^* = \pi - \theta_2 = 2\pi t + \pi$
3. 3 point, T period truncated unequally weighted c-optimal	$V_3 = \frac{\sigma^2}{9\theta_o^2 n} \frac{4\left(\frac{\theta_1}{\theta_o} + 1\right)^2(1-\xi_1) + \left(2 - \frac{\theta_1}{\theta_o}\right)^2 \xi}{\xi\,(1-\xi)}$	$\xi^* = \frac{-\frac{1}{2}\left(\frac{\theta_o}{\theta_1} + 1\right)}{\frac{1}{2}\cos(\pi T) - \frac{1}{2} - \frac{\theta_o}{\theta_1}}$ $\xi_1^{*(3)} = 1 - \xi_1^*$, $\xi_2^{*(3)} = \xi_3^{*(3)} = \frac{1}{2}\xi^*$	$\theta_1^* = -\theta_2$ $\theta_2^* = \frac{1}{2}\left[-\theta_2 + T - 1\right]$ $\theta_3^* = \frac{1}{2}\left[-\theta_2 + 1 - T\right]$
4a. When T = 1/3	V_3 as above in 3.	$\xi^* = \frac{\frac{1}{2}\left(\frac{\theta_o}{\theta_1} + 1\right)}{\frac{1}{2} + \theta_o/\theta_1}$ $\xi_1^{(3)} = 1 - \xi$, $\xi_2^{(2)} = \xi_3^{(3)} = \frac{1}{2}\xi$	$\theta_1^* = -\theta_2$ $\theta_2^* = -\frac{\theta_2}{2} - \frac{1}{3}$ $\theta_3^* = -\frac{\theta_2}{2} + 1$
4b. when T = 1/3 and $\xi_o = 2/3$	$V_3^* = \frac{2}{3\theta_o^2}\left[2\left(\frac{\theta_1}{\theta_o} + 1\right)^2 + \left(2 - \frac{\theta_1}{\theta_o}\right)^2\right]\frac{\sigma^2}{n}$	as above with $\xi_o^* = 2/3$	as 4a $t^* = 2\pi\theta^*$

Table 9.2 : Efficiencies when $e_1/e_0<1$ for the designs described in Table 9.1 and relations refered in sections 9.6.

e_1/e_0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$E_{2,4}$	0.50	0.49	0.48	0.46	0.44	0.42	0.40	0.37	0.35
$E_{2,3}$	0.53	0.51	0.48	0.46	0.44	0.42	0.40	0.39	0.37
$E_{2,3}$	0.49	0.49	0.48	0.46	0.44	0.42	0.40	0.37	0.35

Table 9.3 : Evaluating design measures when $e_1/e_0<1$ for the designs described in Table 9.1

Design	e_1/e_0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1.	ε_i	0.25 for all i=1,2,3,4 and for every e_1/e_0								
2.	ε_1	0.45	0.40	0.35	0.30	0.25	0.20	0.15	0.10	0.05
	ε_2	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
4b.	ε_1	0.46	0.42	0.40	0.36	0.33	0.30	0.28	0.25	0.22
	$\varepsilon_{2=3}$	0.27	0.29	0.30	0.32	0.33	0.35	0.36	0.37	0.39

9.7 Analytic approach

In chapter 3 we presented the geometric aspect of the design, after producing the analytic forms of the confidence intervals. In this chapter we have proceeded by using firstly geometrical arguments and now we describe the analytic approach of the problem. Both approaches are of course the two sides of the same coin.

In our practical context interest is focused only on Case I, i.e $\theta_1/\theta_0<1$. We now tackle this case using a different approach.

For the two point design the matrix $M(\xi)$ is as in (9.4.10) and the design allocates proportion ξ at $-\theta_2$ and $1-\xi$ at $\pi-\theta_2$. Consider the two vectors b_1, b_2 as follows

$$\sqrt{2}b_1 = (1,\cos\phi,-\sin\phi)^T, \quad \sqrt{2}b_2 = (1,-\cos\phi,\sin\phi)^T \qquad (9.7.1)$$

which are orthogonal with norm 1. Suppose a vector h can be written as

$$h = \nu b_1+(1-\nu)b_2 \quad , \nu\in R. \qquad (9.7.2)$$

Moreover we can write $M(\theta,\xi)$, recall (9.4.10), as

$$M(\theta,\xi) = 2\xi b_1 b_1^T + 2(1-\xi)b_2 b_2^T \qquad (9.7.3a)$$

and therefore using a matrix result (Appendix A4.I, (vii, viii))

$$M^-(\theta,\xi) = \tfrac{1}{2}\xi^{-1}b_1 b_1^T + \tfrac{1}{2}(1-\xi)^{-1}b_2 b_2^T. \qquad (9.7.3b)$$

For any h as in (9.7.2), considering $M^- = M^-(\varepsilon)$ as in (9.7.3) we find that

$$2\, h^T M^-(\varepsilon) h = \nu^2/\varepsilon + (1-\nu)^2/(1-\varepsilon) \qquad (9.7.4)$$

Relation (9.7.4) gives the value of the criterion function for any ε and it can easily be shown to be minimized when

$$|\nu/(1-\nu)| = |\varepsilon/(1-\varepsilon)| \text{ i.e } \pm\nu/(1-\nu) = \varepsilon/(1-\varepsilon). \qquad (9.7.5)$$

Thus for the particular h in which we are interested namely, $(-\theta_1/\theta_0)h=(1,(-\theta_0/\theta_1)\cos\theta_2, (\theta_0/\theta_1)\sin\theta_2)$ we can obtain

$$\nu = 0.5(1-\theta_0/\theta_1). \qquad (9.7.6)$$

This is of course the value of the optimum ε evaluated with the geometrical argument in (9.4.14). From (9.7.5) we have that either $\varepsilon=\nu$ or $\varepsilon=-\nu/(1-2\nu)$. Thus, taking into consideration (9.7.6) we obtain relations (9.4.14) and (9.4.10) respectively, namely

$$\varepsilon = 0.5(1-\theta_0/\theta_1) \quad \text{if} \quad \theta_1/\theta_0>1 \qquad (9.7.7a)$$

$$\varepsilon = 0.5(1-\theta_1/\theta_0) \quad \text{if} \quad \theta_1/\theta_0<1. \qquad (9.7.7b)$$

Consider now the general problem of truncated daily time interval T and the allocation of the observations as follows (recall Fig. 9.6 allocating observations at the end points and the middle of 1-T) :

$1-\varepsilon$ at $t_A = -\theta_2$

0.5ε at $t_B = -\theta_2-2\pi*1/2(1-T) = -\phi-\pi(1-T) = -\phi-\delta$

0.5ε at $t_C = -\theta_2+2\pi*1/2(1-T) = -\phi+\pi(1-T) = -\phi+\delta,$

with $\delta=\pi(1-T)$. The corresponding vectors $W(t)$ (recall (9.2.2c)) will then be

$$W(t_A)^T = (1,\ \cos\theta_2,\ -\sin\theta_2)$$
$$W(t_B)^T = (1,\ \cos(\theta_2+\delta),\ -\sin(\theta_2+\delta))$$
$$W(t_C)^T = (1,\ \cos(\theta_2-\delta),\ -\sin(\theta_2-\delta)).$$

Therefore the matrix $M=M(\theta,\xi)$ can be evaluated explicitly as

$$M = (1-\xi)W(t_A)W(t_A)^T + 0.5\xi[W(t_B)W(t_B)^T + W(t_C)W(t_C)^T]. \quad (9.7.8)$$

For the matrix M as in (9.7.8) the following Lemma holds (for proof see Appendix A5.I).

Lemma 9.7.1 : The matrix M can be written as

$$M = M_0 + \alpha_{12}[b_1 b_2^T + b_2 b_1^T] \quad (9.7.9)$$

with M_0 matrix defined in Appendix A5.I and $\alpha_{12}=2\alpha(1-\alpha)\xi$, with $\alpha=0.5(\cos\delta+1)$.

The inverse of M is needed as the general problem can be formulated as

$$\min\ h^T M^{-1} h\ ,\ \xi\in\Xi,\quad h=(-\theta_1/\theta_0)(1,(-\theta_0/\theta_1)\cos\theta_2, \theta_0/\theta_1\sin\theta_2)^T. \quad (9.7.10)$$

Therefore the following Lemma can be proved (Appendix A5.I).

Lemma 9.7.2 : The quantity $h^T M^{-1} h$ is evaluated to be

$$h^T M^{-1} h=(\theta_1^2/2\theta_0^4)(1-\alpha)^{-2}\{(1-\nu)^2/\xi + (\nu-\alpha)^2/(1-\xi)\}, \quad (9.7.11)$$

with $\alpha=0.5(\cos\xi+1)$

Minimizing (9.7.11), we rediscover (9.7.5) and obtain

$$\xi = [0.5(1+\theta_0/\theta_1)]/[0.5-0.5\cos(\pi T)+\theta_0/\theta_1]. \qquad (9.7.12)$$

With T=1/3, the particular case discussed earlier, (9.7.12) corresponds to (9.5.5).

CHAPTER 10

EPILOGUE

The cardinal target of this thesis was to gain knowledge, by means of theoretical and applied work, on optimal nonlinear experimental design.

The two different routes of inquiry, theoretical argument and laboratory experimentation -through the simulations - were followed. We only wanted to serve the sense of "measure", an ancient Greek principle, which in recent terminology says: put equal weight on both theory and practice.

The statistician moves sequentially from the Scylla of theory to the Charybdis of application. That is, he has to find the optimal route between theory and application. Throughout our applications we have assumed that the model was known. More work is needed to study the robustness of optimal nonlinear designs. On the theoretical side there is still no general theory concerning the convergence of the sequence $M(\theta_n, \xi_n)$ in sequentially designed experiments. We have commented on the link between fully sequential designs and the Wynn type algorithm. Possibly links of this type might aid development of a general theory of convergence.

The linear optimal design problem has flourished in the work of, for example, Wu and Wynn (1978) and Pukelsheim and Titterington (1983). The former offer a dichotomous theorem to check whether an appropriately created sequence of design measures

converges to the optimal one. The latter place optimal linear experimental design under a general convex analysis setting. It would be nice if nonlinear design could be taken closer to these targets.

Ironically optimal designs for nonlinear problems require knowledge of the unknown parameters. Some static designs are based on initial point estimates for the unknown parameters while others require specification of a range of plausible values or a prior distribution for the unknown parameters. More experience is needed on the application of these approaches to practical problems.

Extending the work of Ford and Silvey (1980) we performed a number of simulation studies for a variety of design strategies and problems. It appears that approximate inference can be carried out ignoring the sequential nature of the design assuming that the sample sizes are reasonably large. The most complicated model considered had two parameters. It would be of interest to study more complex problems with more parameters.

Our results suggest that two stage designs might provide a useful compromise for an experimenter unable to carry out a fully sequential design. More studies might be needed to clarify the situation as it seemed that the nature of the problem influenced the performance of the two-stage design.

Geometrical aspects play an active role in this thesis. It is not just because we are pure lovers of Plato, who did not permit anybody to enter his Academy without the knowledge of Geometry. It is because we also believe that the beautiful is equally as useful as the necessary. In chapter 9 we have another practical example where geometrical interpetation is in accordance with an analytic approach, to prove that theory serves the Nature and nature can be ruled from Theory.

APPENDIX 1

A1.I Differentiability

Let g be a function from R^n to $[-\infty,\infty]$ and $x=(x_1,x_2,\ldots,x_n)$ a point where g is finite. Then we define

Definition : The (Gateaux) **directional derivative** at x in the direction of y is defined to be

$$G(x,y) = \lim_{\epsilon \to 0^+} \epsilon^{-1}\{g(x+\epsilon y)-g(x)\} \qquad (A1.1)$$

If g is differentiable

$$G(x,y) = \Sigma y_i(\partial g(x)/\partial x_i)$$

Note that

$$G(x,e_j) = (\partial^+ g/\partial x_j)$$

i.e the right hand partial derivative of g with respect to the j-th component of x, with e_j being the j-th unit vector.

The Frechet direction derivative is defined as

$$F(x,y) = G(x,y-x) \qquad (A1.2)$$

The following theorem can be proved.

Theorem : For a concave function g and x a point where g is finite then G(x,y) exists for all y ; this is so wehether or not g is differentiable at x.

Comment : The entries x, y might be matrices. In the design contest with criterion Φ and derectional derivative Φ we have corresponds to $g=\Phi$ and $G=\Phi$ and x, y are matrices.

References : Silvey (1980), Torsney (1981).

A1.II Convex sets

Definition 1 : The set S is called **convex** if all points $s \in S$ the form

$$s=\alpha s_1+(1-\alpha)s_2 \ , \quad s_1, s_2 \in S, \ \alpha \in [0,1]$$

are elements of S.

Definition 2 : The set of points, S^* say, with elements

$$s^* = \Sigma \alpha_i s_i, \quad \Sigma \alpha_i = 1, \quad \alpha_i \in [0,1], \ s_i \in S$$

is a convex set. S^* is called the convex hull of the set S.

Theorem (Caratheodory) : Each point s^* in the convex hull S^* of any subset S, of the n-dimensional space, can be represented in the form

$$s^* = \sum_{i}^{n+1} a_i s_i \ , \quad a_i \geq 0, \quad s_i \in S, \quad \sum_{i}^{n+1} a_i = 1$$

If s^* is a boundary point of the set S^*, then a_{n+1} can be set equal to zero.

References : Fedorov (1972), Silvey (1972).

APPENDIX 2

Many phenomena from Biology and Chemistry involve nonlinear functions, especially kinetic data and growth curve phenomena. We present a survey of the literature in which nonlinear experimental design ideas have been used. Methods for fitting nonlinear models are presented in Ratkowsky (1983). Table 3.1 summarizes the optimum design points suggested for various nonlinear models, as defined by the response function $\eta=f(u,\theta)$. The design space U is the "operating region" for the experimenter.

A2.I Mitscherlish equation of diminishing returns.

Model : $\eta=\theta_1+\theta_2\exp(\theta_3 u)$

$U=[\kappa,\lambda]\subseteq R,\ \theta\in R^{+}xR^{-}$

where :

η : the expected amount of growth

θ_1: hypothetical growth from an infinite amount of fertilizer

$\theta_1+\theta_2$: measures the rate at which additional increment of fertilizer decreases.

u : the amount of added fertilizer

(see : Box and Lucas (1959))

A2.I.1 If it is assumed in A2.I that $f(0,\theta)=0$ then $\theta_2=-\theta_1$

Let $\theta_3=\theta_2$. Then the monomolecular model is

Model : $\eta=\theta_1\{1-\exp(\theta_2 u)\}$

$U=[\kappa,\lambda]\subseteq R,\ \Theta\subseteq R^+xR^-$

where :

η : the amount of product formed at time u from a simple decay law.

(see : Box and Lucas (1959), Hohmann and Jung (1975), Katz et al (1981))

A2.II The growth (or decay) law.

Model : $\eta=\theta_1\exp(\theta_2 u),\ \theta_2>0\ \ (\theta_2<0)$

$U=[\kappa,\lambda]\subseteq R,\ \Theta\subseteq R^+xR$

where :

η : the amount of substance growth (or decay)

(see : Box and Lucas (1959). Jennrich (1969))

A2.III Irreversible reaction $B \longrightarrow C$.

Model : $\eta = \exp\{-\theta_1 t_1 \exp[-\theta_2(1/T - 1/T_0)]\}$

$$u=(t_1,T)\epsilon U\subseteq R^{+}x\Delta,\ \Delta=[380,450],\ \Theta\subseteq R^{2}$$

(see : Box and Lucas (1959), M.J. Box (1968a))

If it is assumed that $f(\alpha,\theta)-f(0,\theta)=1$ then the model is reduced to

$$\eta=\theta_1-\exp(-\theta_2 u)$$

In order to estimate the reaction rate constant k and its variation with temperature (T) the Arrhenieus law gives

$$k=A\exp(-E/RT),$$

were : A is the frequency factor, E the activity energy. R the gas constant and T the temperature (^{o}K). If η is the fraction of the original material remaining after the reaction has continued for a time t at temperature T then $\eta=\exp(-A\exp(-E/RT))$. Rewriting this in terms of the rate θ_1 at specific temperature T_0 and letting $\theta_2=E/R$ be the proportion of the activation energy we obtain model A3.III.

If $t_2=1/T - 1/T_0$ then

$$\eta=\exp\{-\theta_1 t_1\exp(-\theta_2 t_2)\}\ ,\ u=(t_1,t_2)\subseteq R^{2}$$

This is the first order decay law with rate a function of temperature

(see : Hunter and Atkinson (1966)).

A2.IV Chemical reactions $A \longrightarrow B \longrightarrow C$,

Model : $\eta=[\theta_1/(\theta_1-\theta_2)]\{\exp(-\theta_2 t)-\exp(-\theta_1 t)\}$

$u=t\in U=R^+,\ \theta\subseteq R^2$

where:

η: the amount of B present after time t, expressed as a function of the total material present when initially (t=0) only material A is present ($\eta=[B]$)

θ_1: rate constant $A \longrightarrow B$

θ_2: rate constant $B \longrightarrow C$

We can also assume that the θ's are functions of the temperature according to the Arrhenius law, i.e

$$\theta_i=k_i\exp(-E_i(1/T - 1/T_o))$$

with k_i, E_i parameters. The model is considered with

$$u=(t,T)\in U\subseteq R^+ x\Delta,\quad \Delta=[380,450].$$

(see : Fedorov (1972) p 227, Katz et al (1981), Draper and Hunter (1967), Hunter, Hill and Henson (1969))

A2.V Chemical reaction $R \longrightarrow P_1+P$

The catalytic dehydration of Hexyl Alcohol Reaction.

Model : $\eta=(\theta_1\theta_3P_1)/(1+\theta_1P_1+\theta_2P_2)$

$u=(P_1,P_2)\in\Delta x\Delta$, $\Delta=[0,3]$, $\Theta\subseteq R^3$

where :

η : speed of chemical reaction

P_1: partial pressure of product P_1

P_2: partial pressure of product P

θ_1: the absorption equilibrium constant for the reactant R

θ_2: the absorption equilibrium for the product P_1

θ_3: effective reaction rate constant.

(see: Fedorov (1972))

A2.VI BET (Brunauer-Ermet-Teller) equation.

Model : $\eta=(\theta_2\theta_1P_1)/[(1-P_1)(1+(\theta_2-1)P_1)]$

$u=P_1\in\Delta\subseteq R$. $\Delta=[0.05,030]$, $\Theta\subseteq R^2$

where :

η : Volume of gas absorbed on the solid P_1

P_1: P/P_o relative pressure

θ_1: the monolayer capacity (U_m in Chemistry)

θ_2: constant characterestic of the gas-solid.

(see : Hill and Hunter (1974), Henson and Hunter (1969), Khuri(1984))

A2.VII Catalyst and reaction study

Model : $\eta=(\theta_1/(\theta_2+\theta_1-\theta_3))\{\exp(-\theta_3 t)-\exp(-(\theta_1+\theta_2)t)\}$

$u=t\in U\subseteq R, \quad \theta\subseteq R^3$

where :

A : reactant reacting to form

B : the desired product

C : undesired by-products

Also B further reacts to form C

$\theta_1,\theta_2,\theta_3$ rate constants

(see : Hill and Hunter (1974))

A2.VIII Isomerization of n-pentane to i-pentane in the presense of hydrogen.

Model : $$\eta = \frac{\theta_1\theta_3(u_2-u_1/1.632)}{1+\theta_2u_1+\theta_3u_2+\theta_4u_3}$$

$u=(u_1,u_2,u_3)\in U=\Delta_1 x\Delta_2 x\Delta_3\subseteq R^3$

$\Delta_1=[107,471],\ \Delta_2=[69,294],\ \Delta_3=[11,121]$

$\theta\subseteq R^4$

where :

η : the rate of disappearence on n-pentane

θ_1: reaction parameter

$\theta_2,\theta_3,\theta_4$ equilibrium constants

u_1, u_2, u_3 partial pressures of hydrogen, n-pentane, i-pentane respectively.

(see : Prichard and Bacon (1977)).

A2.IX Oxidation of benzene.

$$\text{Model : } \eta = \frac{\theta_1 \exp(-\theta_3 u_3)\theta_2 \exp(-\theta_4 u_3) u_1 u_2}{\theta_1 \exp(-\theta_3 u_3) u_1 + u_4 \theta_2 \exp(-\theta_4 u_3) u_2}$$

$$u=(u_1,u_2,u_3,u_4) \in U=\Delta_1 x \Delta_2 x \Delta_3 x \{5.75\} \subseteq R^4 \ , \quad \theta \subseteq R^4$$

$$\Delta_1=[10^{-3}, 16*10^{-3}], \ \Delta_2=[10^{-3}, 4*10^{-3}], \ \Delta_3=[623, 673]$$

where :

η : the initial reaction rate

u_1 : the concentration of oxygen

u_2 : the concentration of benzene

u_3 = 1/T - 0.0015428, wehere T is the absolute temperature of the reaction

u_4 : the observed stoichiometric number

$\theta_1, \theta_2, \theta_3, \theta_4$: model parameters arrising in Arrhenius's law.

(see : Prichard and Bacon (1977), Prichard et al (1977))

A2.X Michaelis - Menten model.

Model : $\eta=\theta_1 u/(\theta_2+u)$

$u \in U \subseteq R$, $\theta \subseteq R^2$

where :

η : the reaction velocity

θ_1 : maximum velocity of the reaction

θ_2 : the half saturation constant (i.e maxη)

u : the concentration of substrate.

(see : Currie (1982))

APPENDIX 3

A3.I Elfving's theorem

Elfving (1952) stated the geometrical characterization of c-optimality as follows :

Consider the model $\eta=E(y|u)=f^T(u)\theta$, $u\in U$. As far as c-optimality is concerned, the optimal design weights of the observations can be obtained through the following geometrical argument:

Draw the convex hull $\mathcal{C}$ of the space $U_0=f(U)$ and its reflection in the origin, $-U_0$. Draw the vector c and let T_1 be the point of intersection of c with $\mathcal{C}$. Then T_1 lies on the line which joins $A\in U_0$ and $B'\in(-U_0)$ and which forms part of the surface $\mathcal{C}$. Note that B' is the reflection of B. Then the c-optimal design is located at A and B with weights ξ, $1-\xi$ respectively where

$$\xi/(1-\xi) = T_1B'/AT_1.$$

Moreover the optimum minimum variance is $[OT/OT_1]^2$. The cases in which U_0 is a line segment or a circle are discussed in chapters 8 and 9 respectively.

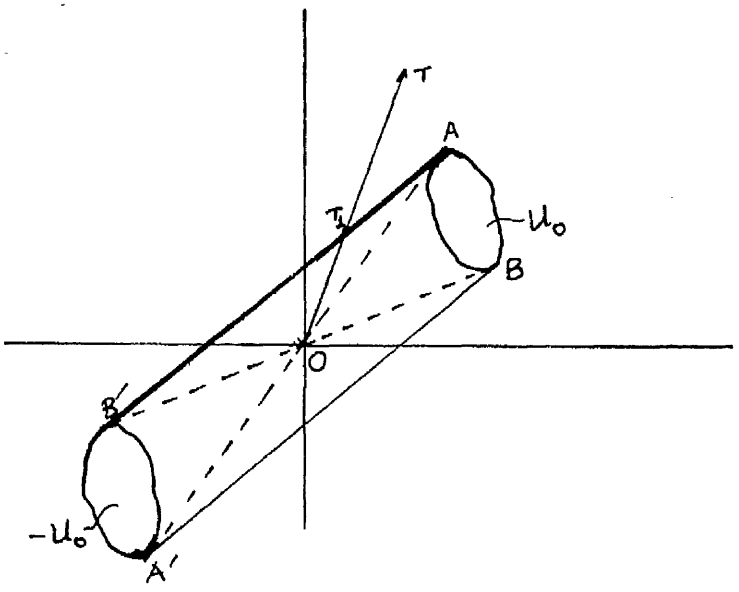


Fig. A3.1 Elfving's theorem.

A3.II Silvapulle's theorem

Silvapulle (1981) stated and proved the conditions under which the existence of the MLE in binary problems is guaranteed.

Let $u_1, u_2, \ldots, u_r$ be the design points corresponding to responses $y_i=1$, $i=1,\ldots,r$ and $u_{r+1}, \ldots, u_n$ corresponding to responses $y_i=1$, $i=r+1,\ldots,n$. Consider the convex cones $S = \{\Sigma\kappa_i u_i, \kappa_i \geq 0, \forall_{i=1,\ldots r}\}$ $F = \{\Sigma k_j u_j, \kappa_j \geq 0, \forall j=r+1,\ldots,n\}$. Then the following theorem holds.

Theorem : Let the condition (L) be defined by

(L) $S \cap F \neq \emptyset$ or one of S or F is $R^p \supseteq \Theta$

Then for the binomial response model

$$\text{Prob}(y_i=1)=T(u_i^T\theta)$$

(i) The MLE $\hat{\theta}$ of θ exists and the minimum set $\{\theta\}$ is

bounded only when (L) is satisfied.

(ii) Suppose that

$$\ell(\theta)=-\Sigma\log T(u_i'\theta) - \Sigma\log(1-T(u_i'\theta))$$

is a proper closed convex function on R^p. Then the MLE $\hat{\theta}$ exists and the minimum set $\{\hat{\theta}\}$ is bounded if and only if (L) is satisfied.

(iii) Suppose that $-\log T$ and $\log(1-T)$ are convex and $u_{ii}=1$ for every i. Then $\hat{\theta}$ exists and the minimum set $\{\hat{\theta}\}$ is bounded if and only if $S\cap F\neq\emptyset$. Let us further assume that T is strictly increasing at every t satisfying $0<T(t)<1$. Then θ is uniquely defined if and only if $S\cap F=0$.

As an example where the MLE does not exist consider Fig.A3.2 where there is no "interblocking" condition between S and F.

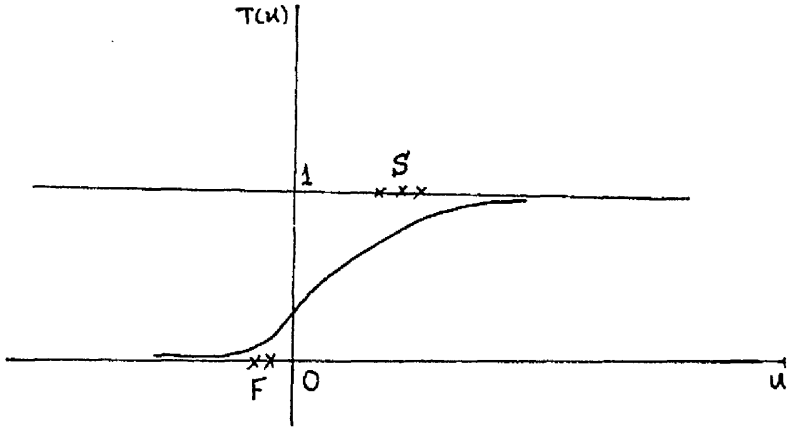


Fig. A3.2 : No intersection between S and F

APPENDIX 4

A4.I Generalized inverse

Definition : Let $A \in M(m,n)$, with $M(m,n)$ the set of mxn matrices then A^- is the Moore-Penrose generalized inverse iff :

(i) AA^- and A^-A are symmetric

(ii) $A^-AA^- = A^-$ and $AA^-A = A^-$

Properties :

(i) When A^- exists it is of size nxm and it is unique.

(ii) $(A^-)^- = A$

(iii) $(A^T)^- = (A^-)^T$

(iv) $\text{rank}(A) = \text{rank}(A^-) = \text{rank}(AA^-) = \text{rank}(A^-A) = \text{rank}(AA^-A) = \text{rank}(A\ AA^-)$

(v) If $A = A^T$ then $A^- = (A^-)^T$

(vi) If $A=A_1+A_2+\ldots+A_k$ and $A_iA_j^T=0$ for all $i,j=1,2,\ldots k$ $i\neq j$ then $A^- = A_1^-+A_2^-+\ldots+A_k^-$

(vii) If a is a nonzero vector then $a^-=(a^Ta)^-a^T=||a||^{-2}a^T$

A4.II Conditional inverse

Definition : Let $A \in Mat(m,n)$. A^C is a conditional inverse iff

$$AA^CA = A$$

- The generalized inverse is also a conditional inverse. The opposite not necessarily true.
- For the linear model $Y=X\beta+\epsilon$ the normal equations are $X^TX\beta=X^TY$. Moreover it is known $v^T(X^TX)^Cv=v^T(X^TX)^-v$, v belongs to the column space of X^T.

Properties :

(i) $rank(A^C) \geq rank(A)$

(ii) $rank(A^CA)=rank(AA^C)=rank(A)=tr(AA^C)=tr(A^CA)$

(iii) AA^C, A^CA are idempotent matrices

(iv) $A^CA=I \Rightarrow rank(A)=n$

$AA^C=I \Rightarrow rank(A)=m$

(v) If A of rank r is partitioned as

$$A = \begin{pmatrix} B & C \\ D & E \end{pmatrix} \quad \text{then} \quad A^C = \begin{pmatrix} B^{-1} & 0 \\ 0 & 0 \end{pmatrix}$$

with B of rank r.

References : Graybill (1969), Rao (1965)

APPENDIX 5

A5.I Evaluation of $\min\{ h^T M^{-1}(\theta,\xi)h,\ \xi\in\Xi\}$

(i) Proof of Lemma 9.7.1

(See also A5.II for details about trigonometric equations)

$$M = M(\theta,\xi) = (1-\xi)\begin{pmatrix} 1 & \cos\theta_2 & -\sin\theta_2 \\ \cos\theta_2 & \cos^2\theta_2 & -\cos\theta_2\sin\theta_2 \\ -\sin\theta_2 & -\cos\theta_2\sin\theta_2 & \sin^2\theta_2 \end{pmatrix}$$

$$+\tfrac{1}{2}\xi\begin{pmatrix} 1 & \cos(\theta_2+\delta) & -\sin(\theta_2+\delta) \\ \cos(\theta_2+\delta) & \cos^2(\theta_2+\delta) & -\sin(\theta_2+\delta)\cos(\theta_2+\delta) \\ -\sin(\theta_2+\delta) & -\sin(\theta_2+\delta)\cos(\theta_2+\delta) & \sin^2(\theta_2+\delta) \end{pmatrix}$$

$$+\tfrac{1}{2}\xi\begin{pmatrix} 1 & \cos(\theta_2-\delta) & -\sin(\theta_2-\delta) \\ \cos(\theta_2-\delta) & \cos^2(\theta_2-\delta) & -\sin(\theta_2-\delta)\cos(\theta_2-\delta) \\ -\sin(\theta_2-\delta) & -\sin(\theta_2-\delta)\cos(\theta_2-\delta) & \sin^2(\theta_2-\delta) \end{pmatrix}$$

$$= (1-\xi)\begin{pmatrix} 1 & \cos\vartheta_2 & -\sin\vartheta_2 \\ \cos\vartheta_2 & \cos^2\vartheta_2 & -\cos\vartheta_2\sin\vartheta_2 \\ -\sin\vartheta_2 & -\cos\vartheta_2\sin\vartheta_2 & \sin^2\vartheta_2 \end{pmatrix}$$

$$+\tfrac{1}{2}\xi\begin{pmatrix} 2 & 2\cos\delta\cos\vartheta_2 & -2\cos\delta\sin\vartheta_2 \\ 2\cos\delta\cos\vartheta_2 & 2[\cos^2\delta\cos^2\vartheta_2+\sin^2\delta\sin^2\vartheta_2] & -\cos 2\delta\sin 2\vartheta_2 \\ -2\cos\delta\sin\vartheta_2 & -\cos 2\delta\sin\vartheta_2 & 2[\cos^2\delta\cos^2\vartheta_2+\sin^2\delta\sin^2\vartheta_2 \end{pmatrix}$$

(A5.1

$$\therefore\ M(\theta,\xi) = (1-\xi)\begin{pmatrix}1\\ \cos\vartheta_2\\ -\sin\vartheta_2\end{pmatrix}(1\ \cos\vartheta_2\ -\sin\vartheta_2)$$

$$+\xi\begin{pmatrix}1\\ \cos\delta\cos\vartheta_2\\ -\cos\delta\sin\vartheta_2\end{pmatrix}(1\ \cos\delta\cos\vartheta_2\ -\cos\delta\sin\vartheta_2)$$

$$+\xi\begin{pmatrix}0\\ \sin\delta\sin\vartheta_2\\ \sin\delta\cos\vartheta_2\end{pmatrix}(0\ \sin\delta\sin\vartheta_2\ \sin\delta\cos\vartheta_2). \qquad (A5.2)$$

We define

$$b_1 = \frac{1}{\sqrt{2}}(1, \cos\vartheta_2, -\sin\vartheta_2)^T$$

$$b_2 = \frac{1}{\sqrt{2}}(1, -\cos\vartheta_2, \sin\vartheta_2)^T$$

$$d = (0, \sin\vartheta_2, \cos\vartheta_2)^T \quad , b_1 \perp b_2,\ b_2 \perp d.$$

we can write

$$\begin{pmatrix}1\\ \cos\delta\cos\vartheta_2\\ -\cos\delta\sin\vartheta_2\end{pmatrix} = \alpha\begin{pmatrix}1\\ \cos\vartheta_2\\ -\sin\vartheta_2\end{pmatrix} + (1-\alpha)\begin{pmatrix}1\\ -\cos\vartheta_2\\ \sin\vartheta_2\end{pmatrix} \qquad (A5.2a)$$

where $\alpha = \frac{1}{2}(\cos\delta + 1)$.

$$\therefore\ M(\theta,\xi) = (1-\xi)2b_1b_1^T + \xi\left[\sqrt{2}\alpha b_1 + \sqrt{2}(1-\alpha)b_2\right]\left[\sqrt{2}\alpha b_1 + \sqrt{2}(1-\alpha)b_2\right]^T$$

$$+\xi\sin^2\delta\, dd^T =$$

$$= 2(1-\xi)b_1b_1^T + 2\xi\{\alpha^2 b_1b_1^T + \alpha(1-\alpha)b_1b_2^T + (1-\alpha)\alpha b_2b_1^T +$$

$$(1-\alpha)^2 b_2b_2^T\} + \xi\sin^2\delta\, dd^T = \alpha_1 b_1b_1^T + \alpha_2 b_2b_2^T + \alpha_3 dd^T +$$

$$\alpha_{12}\left[b_1b_2^T + b_2b_1^T\right] := M_0 + \alpha_{12}b_1b_2^T + \alpha_{12}b_2b_1^T := M_1 + \alpha_{12}b_2b_1^T, \qquad (A5.3)$$

with

$$\alpha_1 = 2(1-\xi) + 2\xi\alpha^2$$

$$\alpha_2 = 2\xi(1-\alpha)^2$$

$$\alpha_3 = \xi\sin^2\delta$$

$$\alpha_{12} = 2\alpha(1-\alpha)\xi.$$

Our target is to evaluate $M^{-1} = M^{-1}(\theta, \xi)$.

Firstly $M_0^{-1} = \alpha_1^{-1} b_1 b_1^T + \alpha_2^{-1} b_2 b_2^T + \alpha_3^{-1} d d^T$ (A5.4)

Secondly $M_1^{-1} = (M_0 + \alpha_{12} b_1 b_2^T)^{-1} = M_0^{-1} - \dfrac{\alpha_{12} M_0^{-1} b_1 b_2^T M_0^{-1}}{1 + \alpha_{12} b_2^T M_0^{-1} b_1}$

$$= M_0^{-1} - \alpha_{12} M_0^{-1} b_1 b_2^T M_0^{-1} \qquad (A5.5)$$

Because $b_2^T M_0^{-1} b_1 = 0$,

Thus

$$M^{-1} = (M_1 + \alpha_{12} b_2 b_1^T)^{-1} = M_1^{-1} - \frac{\alpha_{12} M_1^{-1} b_2 b_1^T M_1^{-1}}{1 + \alpha_{12} b_1^T M_1^{-1} b_2} \qquad (A5.6)$$

We then have

(i) $b_1^T M_1^{-1} b_2 = b_1^T (M_0^{-1} - \alpha_{12} M_0^{-1} b_1 b_2^T M_0^{-1}) b_2 =$

$$= b_1^T M_0^{-1} b_2 - \alpha_{12} b_1^T M_0^{-1} b_1 b_2^T M_0^{-1} b_2 = -\alpha_{12} \alpha_1^{-1} \alpha_2^{-1}, \qquad (A5.7)$$

Because

$b_1^T M_0^{-1} b_2 = 0$

$b_1^T M_0^{-1} b_1 = b_1^T (\alpha_1^{-1} b_1 b_1^T + \alpha_2^{-1} b_2 b_2^T + \alpha_3^{-1} d d^T) b_1 = \alpha_1^{-1}$

$b_2^T M_0^{-1} b_2 = \alpha_2^{-1}$;

(ii) $M_1^{-1} b_2 b_1^T M_1^{-1} = (M_0^{-1} - \alpha_{12} M_0^{-1} b_1 b_2^T M_0^{-1}) b_2 b_1^T (M_0^{-1} - \alpha_{12} M_0^{-1} b_1 b_2^T M_0^{-1})$

$= M_0^{-1} b_2 b_1^T M_0^{-1} - \alpha_{12} M_0^{-1} b_1 b_2^T M_0^{-1} b_2 b_1^T M_0^{-1} - M_0^{-1} b_2 b_1^T \alpha_{12} M_0^{-1} b_1 b_2^T M_0^{-1}$

$+ \alpha_{12}^2 M_0^{-1} b_1 b_2^T M_0^{-1} b_2 b_1^T M_0^{-1} b_1 b_2^T M_0^{-1}$;

(iii) $b_i^T M_0^{-1} b_i = \alpha_i^{-1} \quad i = 1, 2$.

$$\therefore M = M_0^{-1} - \alpha_{12} M_0^{-1} b_1 b_2^T M_0^{-1}$$

$$- \frac{\alpha_{12} \{ M_0^{-1} b_2 b_1^T M_0^{-1} - \alpha_{12} M_0^{-1} b_1 b_1^T M_0^{-1} / \alpha_2 - \alpha_{12} M_0^{-1} b_2 b_2^T M_0^{-1} / \alpha_1 + \alpha_{12}^2 M_0^{-1} b_1 b_2^T / \alpha_2 \alpha_1 \}}{1 - \alpha_{12}^2 / \alpha_1 \alpha_2} \qquad (A5.8)$$

(ii) On proof of Lemma 9.7.2

We want to minimize the quantity $h^T M^{-1} h$, with

$$h = -\frac{\vartheta_1}{\vartheta_0^2}\left(1, -\frac{\vartheta_0}{\vartheta_1}\cos\vartheta_2, \frac{\vartheta_0}{\vartheta_1}\sin\vartheta_2\right).$$

The vector h can be written as

$$h = -\frac{\vartheta_1}{\vartheta_0^2}\{\nu b_1 + (1-\nu)b_2\}, \quad \nu = \frac{1}{2}\left(1 - \frac{\vartheta_0}{\vartheta_1}\right).$$

Thus

$$h^T M^{-1} h = \frac{\vartheta_1^2}{\vartheta_0^4}\{\nu b_1 + (1-\nu)b_2\}^T M^{-1}\{\nu b_1 + (1-\nu)b_2\}$$

$$= \frac{\vartheta_1^2}{\vartheta_0^4}\{\nu^2 b_1^T M^{-1} b_1 + \nu(1-\nu) b_1^T M^{-1} b_2 + (1-\nu)\nu b_2^T M^{-1} b_1$$

$$+ (1-\nu)^2 b_2^T M^{-1} b_2\} =$$

$$\frac{\vartheta_1^2}{\vartheta_0^4}\{\nu^2 b_1^T M^{-1} b_1 + \nu(1-\nu)[b_1^T M^{-1} b_2 + b_2^T M^{-1} b_1] + (1-\nu)^2 b_2^T M^{-1} b_2\}. \quad (A5.9)$$

We have

(i) $b_1^T M^{-1} b_1 = b_1^T M_0^{-1} b_1 - \alpha_{12} b_1^T M_0^{-1} b_1 b_2^T M_0^{-1} b_1 - \frac{\alpha_{12}}{1-\alpha_{12}^2/\alpha_1\alpha_2} b_1^T (A5.8) b_1$

$$= \alpha_1^{-1} - \frac{\alpha_{12}}{1-\alpha_{12}^2\alpha_1^{-1}\alpha_2^{-1}}\{\underbrace{b_1^T M_0^{-1} b_2}_{0} b_1^T M_0^{-1} b_1 - \frac{\alpha_{12}}{\alpha_2}\underbrace{b_1 M_0^{-1} b_1}_{\alpha_1}\underbrace{b_1^T M_0^{-1} b_1}_{\alpha_1}$$

$$- \frac{\alpha_{12}}{\alpha_1}\underbrace{b_1^T M_0^{-1} b_2}_{0} b_2^T M_0^{-1} b_1 + \frac{\alpha_{12}^2}{\alpha_2\alpha_1} b_1 M_0^{-1} b_1 \underbrace{b_2^T M_0^{-1} b_1}_{0}\} =$$

$$= \alpha_1^{-1} - \frac{\alpha_{12}}{1-\alpha_{12}^2/\alpha_1\alpha_2}\left(-\frac{1}{\alpha_2}\frac{\alpha_{12}}{\alpha_1\alpha_2}\right) = \frac{1}{\alpha_1} + \frac{\alpha_{12}^2/\alpha_1^2\alpha_2}{1-\alpha_{12}^2/\alpha_1\alpha_2} \quad (A5.10)$$

By symmetry

$$\text{(ii)} \quad b_2^T M^{-1} b_2 = \frac{1}{\alpha_2} + \frac{\alpha_{12}^2/\alpha_1\alpha_2}{1-\alpha_{12}^2/\alpha_1\alpha_2}. \quad (A5.11)$$

Moreover

$$\text{(iii)} \quad b_1^T M^{-1} b_2 = \underbrace{b_1^T M_0^{-1} b_2}_{0} - \alpha_{12}\underbrace{b_1^T M_0^{-1} b_1}_{\alpha_1^{-1}}\underbrace{b_2^T M_0^{-1} b_2}_{\alpha_2^{-1}}$$

$$- \frac{\alpha_{12}}{1-\alpha_{12}^2/\alpha_1\alpha_2} b_1^T (A5.8) b_2 =$$

$$= -\frac{\alpha_{12}}{\alpha_1\alpha_2} - \frac{\alpha_{12}}{1-\alpha_{12}^2/\alpha_1\alpha_2}\left\{b_1^T M_0^{-1} b_2\, b_1^T M_0^{-1} b_2 - \frac{\alpha_{12}}{\alpha_2} b_1^T M_0^{-1} b_1 \underbrace{b_1^T M_0^{-1} b_2}_{0}\right.$$

$$\left. - \frac{\alpha_{12}}{\alpha_1}\underbrace{b_1^T M_0^{-1} b_2}_{0}\, b_2^T M_0^{-1} b_2 + \frac{\alpha_{12}^2}{\alpha_1\alpha_2}\underbrace{b_1^T M_0^{-1} b_1}_{\alpha_1^{-1}}\,\underbrace{b_2^T M_0^{-1} b_2}_{\alpha_2^{-1}}\right\}$$

$$= -\frac{\alpha_{12}}{\alpha_1\alpha_2} - \frac{\alpha_{12}}{1-\alpha_{12}^2/\alpha_1\alpha_2}\left(\frac{\alpha_{12}^2}{\alpha_1^2\alpha_2^2}\right) =$$

$$= -\frac{\alpha_{12}}{\alpha_1\alpha_2} - \frac{\alpha_{12}^3/\alpha_2^2\alpha_1^2}{1-\alpha_{12}^2/\alpha_1\alpha_2}\,. \qquad \text{(A5.12)}$$

(v) $b_2^T M^{-1} b_1 = b_2^T\Big(M_0^{-1} - \alpha_{12} M_0^{-1} b_1 b_2^T M_0^{-1} -$

$$- \frac{\alpha_{12}}{1-\alpha_{12}^2/\alpha_1\alpha_2}\Big(M_0^{-1} b_2 b_1^T M_0^{-1} - \alpha_{12} M_0^{-1} b_1 b_1^T M_0^{-1}\frac{1}{\alpha_2} -$$

$$\left. - \alpha_{12} M_0^{-1} b_2 b_2^T M_0^{-1}\frac{1}{\alpha_1} + \frac{\alpha_{12}^2}{\alpha_1\alpha_2} M_0^{-1} b_1 b_2^T M_0^{-1}\Big)\right) b_1 =$$

$$= b_2^T M_0^{-1} b_1 - \alpha_{12}\underbrace{b_2^T M_0^{-1} b_1}_{0}\, b_2^T M_0^{-1} b_1 - \frac{\alpha_{12}}{1-\alpha_{12}^2/\alpha_1\alpha_2}\left\{\underbrace{b_2^T M_0^{-1} b_2}_{\alpha_2^{-1}}\,\underbrace{b_1 M_0^{-1} b_1}_{\alpha_1^{-1}}\right.$$

$$- \frac{\alpha_{12}}{\alpha_2}\underbrace{b_2^T M_0^{-1} b_2}_{0}\, b_1^T M_0^{-1} b_1 - \frac{\alpha_{12}}{\alpha_1} b_2^T M_0^{-1} b_2\, b_2^T M_0^{-1} b_1$$

$$\left. + \frac{\alpha_{12}^2}{\alpha_1\alpha_2} b_2^T M_0^{-1} b_1\,\underbrace{b_2^T M_0^{-1} b_1}_{0}\right)$$

$$= \frac{\alpha_{12}}{1-\alpha_{12}^2/\alpha_1\alpha_2}\left(\alpha_2^{-1}\alpha_1^{-1}\right). \qquad \text{(A5.13)}$$

Therefore substituting (A5.10), (A5.11), (A5.12), (A5.13) to (A5.9) we have

$$h^T M^{-1} h = \frac{\vartheta_1^2}{\vartheta_0^4}\left\{ \frac{\nu^2/\alpha_1 + (1-\nu)^2/\alpha_2 - 2\nu(1-\nu)\alpha_{12}/\alpha_1\alpha_2}{1 - \alpha_{12}^2/\alpha_1\alpha_2} \right\}$$

$$= \frac{\vartheta_1^2}{\vartheta_0^2} \cdot \frac{\nu^2\alpha_2 + (1-\nu)^2\alpha_1 - 2\nu(1-\nu)\alpha_{12}}{\alpha_1\alpha_2 - \alpha_{12}^2}$$

$$= \frac{\vartheta_1^2}{\vartheta_0^4} \cdot \frac{2(1-\nu)^2(1-\xi) + 2\xi\{(1-\alpha)\nu - \alpha(1-\nu)\}^2}{4\xi(1-\xi)(1-\alpha)^2} =$$

$$\frac{1}{2}\,\frac{\vartheta_1^2}{\vartheta_0^4}\,\frac{1}{(1-\alpha)^2}\left\{ \frac{(1-\nu)^2}{\xi} + \frac{(\nu-\alpha)^2}{1-\xi} \right\}. \qquad \text{(A5.14)}$$

The minimum of (A5.14) with respect to ξ is obtained when

$$\left| \frac{\xi}{1-\xi} \right| = \left| \frac{1-\nu}{\nu-\alpha} \right| \qquad \text{(A5.15)}$$

$$\nu = \tfrac{1}{2}\left(1 - \frac{\vartheta_0}{\vartheta_1}\right) \quad \therefore \quad \xi = \frac{1-\nu}{\alpha+1-2\nu}$$

$$= \frac{1/2\,(1 + \vartheta_0/\vartheta_1)}{1/2 - 1/2\cos(\pi T) + \vartheta_0/\vartheta_1}$$

$$= \frac{1/2\,(1 + \vartheta_0/\vartheta_1)}{1/2 - 1/2\cos(\pi T) + \vartheta_0/\vartheta_1}. \qquad \text{(A5.16)}$$

A5.II Some elementary trigonometric results

(i) $\cos^2(-\theta_2-\delta) + \cos^2(-\theta_2+\delta) = \cos^2(\theta_2+\delta) + \cos^2(\delta-\theta_2)$

$= 2\left[\cos^2\theta_2\cos^2\delta + \sin^2\theta_2\sin^2\delta\right]$

(ii) $\sin^2(-\theta_2-\delta) + \sin^2(-\theta_2+\delta) = \sin^2(\theta_2+\delta) + \sin^2(\theta_2-\delta)$

$= 2\left[\cos^2\delta\sin^2\theta_2 + \sin^2\delta\cos^2\theta_2\right]$

(iii) $\cos(-\theta_2-\delta) + \cos(-\theta_2+\delta) = \cos(\theta_2+\delta) + \cos(\delta-\theta_2)$

$= 2\cos\delta\cos\theta_2$

(iv) $-\left[\sin(\theta_2+\delta) + \sin(\theta_2-\delta)\right] = -2\cos\delta\sin\theta_2$

(v) $-\sin(\theta_2+\delta)\cos(\delta+\theta_2) - \sin(\theta_2-\delta)\cos(\theta_2-\delta) =$

$= -(\sin\theta_2\cos\delta + \cos\theta_2\sin\delta)(\cos\theta_2\cos\delta - \sin\theta_2\sin\delta)$

$-(\sin\theta_2\cos\delta - \cos\theta_2\sin\delta)(\cos\theta_2\cos\delta + \sin\theta_2\sin\delta)$

$= -2\sin\theta_2\cos\theta_2\left[\cos^2\delta - \sin^2\delta\right] = -\sin 2\theta_2\cos 2\delta.$

A5.III Special case T=1/3

The proof of Lemma 9.5.2 is based on Appendix A5.I when cosθ=1/2. From Lemma 9.5.1 we express vectors A_1, A_2, A_3 as linear combinations of the vectors c, d, e as follows

$$A_1 = \lambda_1 c + \lambda_2 e$$
$$A_2 = \lambda_3 c + \lambda_4 e$$
$$A_3 = \frac{\sqrt{3}}{2} d .$$

The analysis of (9.5.7) can be evaluated through an analysis similar to that of A5.I or can be considered as special case of α_1, α_2, α_3, α_{12} in (A5.3a). For (9.5.8) consider A5.I with M^{-1} evaluated as

$$M^{-1} = M_0^{-1} - \lambda_{ce} M_0^{-1} c e^T M_0^{-1} -$$
$$- \frac{\lambda_{ce}\{M_0^{-1} - \lambda_{ce}(M_0^{-1}c)(e^T M_0^{-1}) e c^T \{M_0^{-1} - \lambda_{ce} M_0^{-1} c e^T M_0^{-1}\}}{1 - \lambda_{ce}^2 (c^T M_0^{-1} c)(e^T M_0^{-1} e)} .$$

$$\therefore c^T M^{-1} c = c^T M_0^{-1} c + \frac{\lambda_{ce}^2 (c^T M_0^{-1} c)^2 (e^T M_0^{-1} e)}{1 - \lambda_{ce}^2 (c^T M_0^{-1} c)(e^T M_0^{-1} e)}$$

$$= \frac{c^T M_0^{-1} c}{1 - \lambda_{ce}^2 (c^T M_0^{-1} c)(e^T M_0^{-1} e)} = (9.5.8).$$

$$M_0 = \lambda_c c c^T + \lambda_d d d^T + \lambda_e e e^T .$$

APPENDIX 6

A6.I Newton-Raphson scheme.

Let f be a function $f:R^n \longrightarrow R^n$ with a root ξ i.e $f(\xi)=0$. Iterative methods are considered to evaluate such as ξ. When n=1 the iteration. known as Newton-Raphson, is

$$x_{i+1} = x_i - f(x_i)/f'(x_i),\ i=0,1,2... \qquad (A6.1)$$

When n>1 scheme (A6.1) is generalized to

$$x_{i+1} = x_i - (Df(x_i))^{-1}f(x_i) \quad i=0,1,2,... \qquad (A6.2)$$

where $Df(x_i)$ is the nxn Jacobian matrix with elements $\partial f_k/\partial x_{ij}$, k,j=1,2.....n. We assume thet $Df(x_i)$ is nonsigular. Iteration schemme (A6.2) can be modified in the form

$$x_{i-1} = x_i - \lambda_i(Df(x_i))^{-1}f(x_i),\ \lambda_i \epsilon (0.1),\ i=0,1,... \qquad (A6.3)$$

Newton and Kantorovitch's theorem (see Ortega and Rheinfold. 1970. p.421) provides conditions under which the Newton-Raphson scheme (A6.2) converges, provided that the initial guess lies in the neighborhood of the solution ξ.

The method is a very rapidly converging scheme. Its convergence is of quadratic order i.e

$$||x_{j+1}-\xi|| \leq k||x_j-\xi||^p, \quad p \geq 2$$

with k being a constant and $||.||$ the ℓ_2-norm.

The method was proposed by Sir Isaak Newton in Analysis per Aequationes Numero Terminorum Infinitas (1666) and by J.R Raphson, Esq. in Analysis Aequationum Universalis (1690). It was only in 1818 that Fourier was able to use the iteration (A6.1).

A6.II Bisection method

A slow (convergence rate 0.5) but very reliable method is the Bisection rule (Stoer and Bulirsh. 1980, p.285) for solving the equation $f(x)=0$.

We define a sequence of subintervals, starting from an initial interval $[a_0,b_0]$ which contains ξ, the root of f. We define

$$\mu_j = (a_j+b_j)/2 \qquad j=0,1,2,\ldots.$$

and let

$$b_{j+1} = \mu_j, \quad a_{j+1} = a_j \quad \text{if} \quad f(\mu_j)f(b_j)>0$$

or

$$a_{j+1} = \mu_j, \quad b_{j+1} = b_j \quad \text{if} \quad f(\mu_j)f(b_j)<0$$

Then the sequence $[a_{j+1},b_{j+1}] \subseteq [a_j,b_j]$ approaches ξ. Obviously

$$|a_{j+1} - b_{j+1}| = 0.5|a_j - b_j|.$$

APPENDIX 7

1. On Simulation Study I.

A7.I Evaluation of $I(\theta,u)$, for a single observation

From (7.2.1) $\ell(y|u,\theta)=\log p(y|u,\theta)$ will be

$$\ell=\ell(y|u,\theta) = \begin{cases} -\theta u & y=1 \\ \log(1-\exp(-\theta u)) & y=0 \end{cases}$$

Therefore

$$\partial^2\ell/\partial\theta^2 = \begin{cases} 0 & y=1 \\ -[u^2\exp(-\theta u)]/[1-\exp(-\theta u)]^2 & y=0 \end{cases}$$

Thus

$$I(\theta,u) = E(-\partial^2\ell/\partial\theta^2) =$$

$$0p(1|u,\theta)+[u^2\exp(-\theta u)]/[1-\exp(-\theta u)]^2 p(0|u,\theta)=$$

$$u^2\exp(-\theta u)/(1-\exp(-\theta u)).$$

Moreover

$$\partial I(\theta,u)/\partial u = 0 \qquad 2=2\exp(-\theta u)+\theta u.$$

The equaqtion $\Phi(\Psi)=2\exp(-\Psi)+\Psi-2=0$, $\Psi=\theta u$ can be solved numerically by the Newton-Raphson scheme.

A7.II Calculation of the MLE for the model (7.2.1)

(i) The likelihood is

$$L(\theta;u,y) = \Pi \exp(-\theta u_i)^{y_i}(1-\exp(-\theta u_i))^{1-y_i}$$

$$= \exp(-\theta\Sigma u_i y_i)\ \Pi(1-\exp(-\theta u_i))^{1-y_i} \Rightarrow$$

$$\ell(\theta) = \log(L(.)) = -\theta\Sigma u_i y_i + \Sigma(1-y_i)\ln(1-\exp(-\theta u_i)) \Rightarrow$$

$$\ell'(\theta) = -\Sigma u_i y_i + \Sigma(1-y_i)[u_i\exp(-\theta u_i)]/[1-\exp(-\theta u_i)] \quad (*)$$

$$\ell''(\theta) = -\Sigma(1-y_i)[u_i\exp(-\theta u_i)]/[1-\exp(-\theta u_i)]^2.$$

To get the MLE $\ell'(\theta)=0$ has to be solved. The Newton-Raphson iteration is applied to get

$$\theta_{\nu+1} = \theta_\nu - \ell'(\theta_\nu)/\ell''(\theta_\nu)\ ,\ \theta_0 \text{ given. } \nu=1,2....$$

(ii) Evaluation of $\hat{\theta}_i$.

For the first batch, $u_i=u_1$, $i=1,2...,r$. $r=5,25,50$. Thus from (*)

$$\ell'(\theta)=-u_1\overset{r}{\Sigma}y_i + u_1\overset{r}{\Sigma}(1-y_i)\exp(-\theta u_1)/(1-\exp(-\theta u_1)) = 0 \Rightarrow$$

$$\overset{r}{\Sigma}y_i=[\exp(-\theta u_1)/(1-\exp(-\theta u_1)]\overset{r}{\Sigma}(1-y_i) \Rightarrow$$

$$\overset{r}{\Sigma}y_i - \exp(-\theta u_1)\overset{r}{\Sigma}y_i = \exp(-\theta u_1)(r-\overset{r}{\Sigma}y_i) \Rightarrow$$

$$r\exp(-\theta u_1)=\overset{r}{\Sigma}y_i \Rightarrow (7.3.1).$$

2. On Simulation Study II.

A7.IV Sum of Squares for Simulation Study II.

Let j=1,2 denote the two points where the observations are obtained, let I denote the number of stages and n_i the number of observations at each point in each stage.

Then the sum of squares SS equals

$$SS = \sum_{J=1}^{2}\left\{\sum_{L=1}^{I}\left[\sum_{K=1}^{n_i}(y_{jlk} - \theta_1 e^{\theta_2 u_{jL}})^2\right]\right\} =$$

$$= \sum_{J=1}^{2}\left[\sum_{L=1}^{I}\left[\sum_{K=1}^{n_i}(y_{jlK} - \bar{y}_{lj.})^2 + n_i(\bar{y}_{jl.} - \theta_1 e^{\theta_2 u_{jL}})^2\right]\right] =$$

$$= \sum_{J=1}^{2}\sum_{L=1}^{I}\sum_{K=1}^{n_i}(y_{jlK} - \bar{y}_{lj.})^2 + \sum_{J=1}^{2}\sum_{i=1}^{I} n_i(\bar{y}_{jl.} - \theta_1 e^{\theta_2 u_{jL}})^2 =$$

$$= SS1 + SS2.$$

Note that

$$SS1 \sim \chi^2\left(2\sum_{i=1}^{I}(n_i - 1)\right)\sigma^2 \equiv \chi^2(n-2I)\sigma^2$$

$$\bar{y}_{jl.} \sim N(\theta_1 e^{\theta_2 u_{jl}}, \frac{\sigma^2}{n_i}).$$

We therefore need to generate normal and Chi-square random numbers. We used the ICL 2988 computr of Glasgow University the NAG routines G05DDF for $N(\mu,\sigma)$ and G05DHF for χ^2_ν.

We kept n_i constant through the experimentation, i.e. equal replications, r say, at each point each time. Thus

$$SS2 = r\sum_{J=1}^{2}\sum_{i=1}^{I}(\bar{y}_{jl.} - \theta_1 e^{\theta_2 u_{jl}})^2.$$

A7.V Estimates at first stage.

At first stage I=1 from (*) we want to minimize SS2, namely

$$(y_{11.}-\theta_1\exp(\theta_2 u_{11})^2+(y_{21.}-\theta_1\exp(\theta_2 u_{21}))^2.$$

That is, both the terms must be equal to zero ie

$$\hat{\theta}_1\exp(\hat{\theta}_2 u_{11})=\bar{y}_{11.} \quad , \quad \hat{\theta}_1\exp(\hat{\theta}_2 u_{21})=\bar{y}_{21.}$$

$$\bar{y}_{11.}/\bar{y}_{21.} = \exp(\hat{\theta}_2(u_{11}-u_{21})).$$

Hence we obtain relation (7.8.3).

A7.VI Minimizing SS2

$$F_1 = \frac{\partial SS2}{\partial \theta_1} = -2\sum\sum n_i(\bar{y}_{ji.}-\theta_1 e^{\theta_2 u_{ji}})e^{\theta_2 u_{ji}}$$

$$F_2 = \frac{\partial SS2}{\partial \theta_2} = -2\sum\sum n_i(\bar{y}_{ji.}-\theta_1 e^{\theta_2 u_{ji}})\theta_1 u_{ji} e^{\theta_2 u_{ji}}$$

$n_i = r$

$$\therefore \quad F_1 = -r\sum\sum \bar{y}_{ji.} e^{\theta_2 u_{ji}} + \theta_1 r\sum\sum (e^{\theta_2 u_{ji}})^2$$

$$F_2 = -r\theta_1\sum\sum \bar{y}_{ji.} u_{ji} e^{\theta_2 u_{ji}} + \theta_1^2 r\sum\sum u_{ji}(e^{\theta_2 u_{ji}})^2.$$

The equations $F_1=0$ and $F_2=0$ have to be solved using the Newton-Raphson method. The Hessian H(i,j) i=1,2, j=1,2 has terms.

$$H(1,1) = r\sum\sum (e^{\theta_2 u_{j1}})^2$$

$$H(1,2) = -r\sum\sum \bar{y}_{j1}.u_{j1}e^{\theta_2 u_{j1}} + 2\theta_1 r\sum\sum u_{j1}(e^{\theta_2 u_{j1}})^2$$

$$H(2,1) = -r\sum\sum \bar{y}_{j1}.u_{j1}e^{\theta_2 u_{j1}} + 2\theta_1 r\sum\sum u_{j1}(e^{\theta_2 u_{j1}})^2$$

$$H(2,2) = -r\theta_1\sum\sum \bar{y}_{j1}.u_{j1}^2 e^{\theta_2 u_{j1}} + 2\theta_1^2 r\sum\sum (u_{j1}e^{\theta_2 u_{j1}})^2.$$

The residual sum of squares has been evaluated as

$$RSS = \chi^2_{(n-2)}\sigma^2 + SS2(\hat{\theta}_1, \hat{\theta}_2).$$

APPENDIX 8

A8.I Evaluating V_C in (8.3.9).

(i) $|u_0|<1$

$$p = p(u_0) = \frac{1+u_0}{2}$$

$$\therefore \quad M = \begin{pmatrix} 1 & 2p-1 \\ 2p-1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & u_0 \\ u_0 & 1 \end{pmatrix}.$$

From (8.3.3a) we have

$$n\theta_1^2 V(p) = \frac{1}{1-u_0^2} (1, u_0) \begin{pmatrix} 1 & -u_0 \\ -u_0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ u_0 \end{pmatrix}$$

$$= \cdots = 1.$$

(ii) $|u_0|>1$

$$p = p(u_0) = \frac{1+u_0}{2u_0}$$

$$\therefore \quad M = \begin{pmatrix} 1 & 1/u_0 \\ 1/u_0 & 1 \end{pmatrix}$$

by (8.3.3a) we have

$$n\theta_1^2 V(p) = \frac{u_0^2}{u_0^2-1} (1, u_0) \begin{pmatrix} 1 & -1/u_0 \\ -1/u_0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ u_0 \end{pmatrix} =$$

$$= \cdots = u_0^2.$$

A8.II The study of (8.4.4)

(i) $|u_0|<1$.

$$\frac{d}{du} Eff(\xi(p)) = 4p(1-p)\frac{d}{du_0}\left[u_0^2+2u_0(1-2p)+1\right]^{-1}$$

$$= 4p(1-p)\left\{-\frac{2u_0+2(1-2p)}{(u_0^2+2u_0(1-2p)+1)^2}\right\} = 0$$

$\therefore \quad u_0 = 2p-1$

$$\frac{d^2}{du_0^2}\Big|_{u_0=2p-1} Eff(\xi(p)) = -4p(1-p)\frac{d}{du_0} Eff(\xi(p)) = -\frac{1}{4p(1-p)} < 0 .$$

Therefore $u_0=2p-1$ gives the maximum.

The corresponding value of efficiency is 1. We check at the end points.

- At u=1 $\quad Eff(\xi(p)) = \dfrac{4p(1-p)}{1+2(1-2p)+1} = p$

- At u=-1 $\quad Eff(\xi(p)) = \dfrac{4p(1-p)}{4p} = 1-p .$

(ii) $|u_0|>1$.

$$\frac{d}{du_0} Eff(\xi(p)) = 4p(1-p)\frac{d}{du_0}\frac{u_0^2}{u_0^2+2u_0(1-2p)+1} = 0$$

Therefore $u_0=(2p-1)^{-1}$ or $u_0=0$ the latter of which is not accepted.

$$\frac{d^2}{du_0^2} Eff(\xi(p)) = 4p(1-p)\frac{d}{du_0} Eff(\xi(p))$$

$$= -4p(1-p)\frac{3u_0^2+2u_0(1-2p)+1}{(u_0^2+2u_0(1-2p)+1)^3} =$$

$$\left(u_0 = \frac{1}{2p-1} \right)$$

$$= -4p(1-p)\,\frac{3\left[\left(\frac{1}{2p-1}\right)-1\right]}{\left[\left(\frac{1}{2p-1}\right)^2+1\right]^3} < 0 \quad .$$

Therefore $u_0=(2p-1)^{-1}$ is a maximum and the corresponding value of Eff(1/(2p-1))=1.

A8.III Simulation strategy

Let y_{ijk} be the k-th observation at $i=\pm1$ point for the $j=1,..r$ replication. We know that

$$\bar{y}_{-1..} \sim N\left(\theta_0 - \theta_1, \frac{\sigma^2}{n_{-1}}\right) \qquad \text{(A8.1a)}$$

$$\bar{y}_{+1..} \sim N\left(\theta_0 + \theta_1, \frac{\sigma^2}{n_{+1}}\right) \qquad \text{(A8.1b)}$$

with n_{-1}, n_{+1} the number of observations at ±1, and

$$\hat{\theta}_0 = \frac{1}{2}\left(\bar{y}_{-1} + \bar{y}_{+1}\right) \qquad \text{(A8.2a)}$$

$$\hat{\theta}_1 = \frac{1}{2}\left(\bar{y}_{+1} - \bar{y}_{-1}\right). \qquad \text{(A8.2b)}$$

We took advantage of (A8.1) and (A8.2) as the sufficient statistics concentrate the information in two normally distributed quantities. Thus only two normal numbers need to be generated for obtaining estimates for θ_0, θ_1. Moreover the residual sum of squares

r_{ij} will be

$$r_{ij} = \sum_{k} (y_{ijk} - \bar{y}_{ij.})^2 \qquad \text{(A8.3)}$$

and the residual sum of squares r is

$$\begin{aligned} r &= \sum_{jk} (y_{+1jk} - \bar{y}_{+1..})^2 + \sum_{jk} (y_{-1jk} - \bar{y}_{-1..})^2 \\ &= \sum_{jk} (y_{+1jk} - \bar{y}_{+1j.})^2 + \sum_{jk} n_{+1j} (\bar{y}_{+1j.} - \bar{y}_{+1..})^2 + \\ &\sum_{jk} (y_{-1jk} - \bar{y}_{-1j.})^2 + \sum_{jk} n_{-1j} (\bar{y}_{-1j} - \bar{y}_{-1..})^2 \\ &= r_{11} + r_{12} + n_{-11} (\bar{y}_{-11.} - \bar{y}_{-1..})^2 + n_{-12} (\bar{y}_{-12.} - \bar{y}_{-1..})^2 \\ &+ r_{21} + r_{22} + n_{+11} (\bar{y}_{+11.} - \bar{y}_{+1..})^2 + n_{+12} (\bar{y}_{+12.} - \bar{y}_{+1..})^2. \end{aligned}$$

The r_{ij}'s are independent $\sigma^2 \chi^2_\nu$ with $\nu = \sum(n_{ij} - 1) = n - 4$.

Therefore a single $\sigma^2 \chi^2_{(n-4)}$ random number was generated for r.

The SUBROUTINES used from the NAG routines of the ICL computer of the University of Glasgow were

- G05DDF for Normal
- G05DHF for Chi-Square.

APPENDIX 9

Let X_n, n=1,2,3,... be a stochastic process such that the joint distribution of $(X_1,X_2,\ldots,X_n)$ has a strictly positive continuous density p_n. In the sequel the variables X_n and Y_n are supposed to have well-defined expectations.

Definition 1: The sequence X_n, n=1,2,3,... is called **absolutely fair** if

$$E(X_1)=0$$

$$E(X_{n+1}|X_1,\ldots,X_n)=0.$$

Definition 2: A sequence Y_n is a **martingale** if

$$E(Y_{n+1}|Y_1,\ldots,Y_n)=Y_n$$

with

$$Y_n=X_1+X_2+\ldots+X_n+c$$

X_n being as in Definition 1, and c a constant.

The following theorem is known as the martingale convergence theorem.

Theorem : Let S_n be an infinite martingale with $E(S_n^2)<c<\infty$ for all n. There exists a random variable S such that

$$S_n \xrightarrow{w.p.1} S, \quad \text{as } n\to\infty$$

Furthermore

$$E(S_n)=E(S) \quad \text{for all n.}$$

Note : A sequence of random variables Z_n converges with probability one (w.p.1) or almost everywhere (a.e) to the random variable Z iff

$$P[\lim Z_n=Z] = 1.$$

REFERENCES

1. Abdelbasit, K.M. and Plackett, R.L. (1981). Experimental Design for Categorized Data. Int. Statist. Review, Vol.49, 111-126.

2. Abdelbasit, K.M. and Plackett, R.L. (1983). Experimental Design for Binary Data. JASA, Vol.78, 90-98.

3. Anbar, D. (1978). A Stochastic Newton-Raphson Method. J. of Stat. Planning and Inference, Vol.2, 153-163.

4. Anderson, T.W. and Taylor, J.B. (1979). Strong consistency of least squares estimates in dynamic models. Ann. Stat., Vol.7, 484-489.

5. Apostol, T. (1961). Calculus, Vol.I Blaisdell Publishing Co., London.

6. Atkinson, A.C. and Hunter, W.G. (1968). The Design of Experiments for parameter Estimation. Technometrics, Vol.10, 271-289.

7. Atkinson, A.C. (1982). Developments in the design of Experiments. Inst. Stat. Rev., Vol.50, 161-177.

8. Ash, A. and Hedayat, A. (1978). An introduction to design optimality with an overview of the literature. Commun. Stat., Series A, Vol.7, 1295-1325.

9. Bard, Y. and Lapidus, L. (1968). Kinetics Analysis by Digital Parameter Estimation. Catalysis Review, Vol.2(1), 67-112.

10. Bard, Y. (1974). Nonlinear Parameter Estimation. Academic Press, N.York.

11. Barlow, R.E. and Hsing, J.W. (1983). Expected Information from a life Test Experiment. The Statistician, Vol.32, 35-45.

12. Basu, A.P. and Ghosh, J.K. (1980). Asymptotic properties Of a solution to the Likelihood Equation with Life Testing Applications. JASA, Vol.75, 410-414.

13. Bates, D.M. and Watts, D.G. (1980). Relative Curvature Measures of Nonlinearity. J.R.S.S.B., Vol.42, 1-25.

14. Bates, D.M. and Watts, D.G. (1981). Parameter transformations for improved approximate confidence regions in nonlinear least squares. Ann. Stat., Vol.9., 1152-1167.

15. Bates, D.M., Hamilton, D.C., Watts, D.G. (1983). Calculation of intrinsic and parameter-effects curvatures for nonlinear regression moidels. Commun. Stat., Series B, Vol.12, 469-477.

16. Beale, E.M.L. (1960). Confidence Regions in Non-linear Estimation. J.R.S.S.B., Vol.22, 41-88.

17. Begg, C.B. and Kalish, L.A. (1984). Treatment allocation for Nonlinear Models in Clinical Trials: The Logistic Model. Biometrics. Vol.40, 409-420.

18. Behnken, D.W. (1964). Estimation of Copolymer Reactivity Ratios. An Example of Nonlinear Estimation. J. of Polymer Science, Part A., Vol.2, 645-668.

19. Bergman, S.W. and Turnbull, B.W. (1983). Efficient sequential designs for destructive life testing with application to animal serial sacrifice experiments. Biometrika., Vol.70, 305-314.

20. Berkson, J. (1944). Application to the logistic function to bio-assay. JASA, Vol.39, 357-365.

21. Berkson, J. (1953). A statistically precise and relatively simple method of estimating the bioassay with quantal response based on the logistic function. JASA, Vol.48., 565-599.

22. Berkson, J. (1955). Maximum Likelihood and minimum X^2 Estimates of the logistic function. JASA, Vol.50, 130-162.

23. Box, G.E.P. and Draper, N.R. (1959). A Basis for the selection of a response surface design. JASA, Vol.54, 622-654.

24. Box, G.E.P. and Lucas, H.L. (1959). Design of experiments in Nonlinear Situation. Biometrika, Vol.49, 77-90.

25. Box, G.E.P. and Hunter, W.G. (1962). A useful Method for Model-Building. Technometrics, Vol.4, 301-318.

26. Box, G.E.P. and Hunter, W.C. (1963). Sequential design of Experiments for Nonlinear models. Proceedings of the IBM Scientific Computing Symposium on Statistics, 113-137.

27. Box, G.E.P. and Tiao, G.C. (1972). Bayesian Inference in Statistical Analysis. Addison-Wesley Pub. Co.

28. Box, G.E.P. and Hill, W.J. (1974). Correcting Inhomogeneity of variance with Power Transformation Weighting. Technometrics, Vol.19, 385-389.

29. Box, M.J. (1968a). The Occurrence of Replications in Optimal Designs of Experiments to Estimate Parameters in Non-linear Models. J.R.S.S.B., Vol.30, 290-302.

30. Box, M.J. and Draper, N.R. (1968b). Non-linear Model Building under Non-homogeneous Variance Assumptions. Technical Report, Imperial Chemical Industries Limited.

31. Box, M.J. (1970a). Improved Parameter Estimation. Technometrics, Vol.12, 219-229.

32 . Box, M.J. (1970b). Some Experiences with a Nonlinear Experimental Design Criterion. Technometrics, Vol.12, 569-589.

33. Box, M.J., Draper, N.R. and Hunter, W.G. (1970c). Missing values in Multiresponse nonlinear model fitting. Technometrics, Vol.12, 613-620.

34. Box, M.J. (1971a). A parameter estimation criterion for Multiresponse Models Applicable when some observations are missing. Appl. Stat., Vol.20, 1-7.

35. Box, M.J. (1971b). Bias in Nonlinear Estimation. J.R.S.S.B., Vol.32, 171-201.

36. Brownlee, K.A., Hodges, J.L. and Rosenbatt, M. (1953). The Up-and-Down method with small samples. JASA, Vol48, 262-277.

37. Brunauer, S., Emmett, P.H. and Teller, E.J. (1938). Adsorption on Gases in Multimolecular Layers. J. Am. Chem. Soc., Vol.60, 309-315.

38. Chernoff, H. (1953). Locally optimal designs for estimating parameters. Ann. Math. Stat., Vol.30, 586-602.

39. Chernoff, H. (1975). Approaches in Sequential Design of Experiments. In " A Survey of Statistical Design and Linear Models", 67-90, editor J.N. Srivastava, N. York.

40. Choi, S.C. (1971). An investigation of Wetherill's method of estimation for the Up-and-Down experiment. Biometrics, Vol.27, 961-970.

41. Chung, K.L. (1954). On a Stochastic Approximation method. Ann. Math. Stat., Vol.25, 463-483.

42. Cochran, W.G. (1973). Experiments for nonlinear functions. JASA, Vol.68, 771-781.

43. Cox, D.R. (1970). Analysis of Binary Data. Methvens Monographs.

44. Currie, D.J. (1982). Estimating Michaelis-Menten Parameters Bias, Variance and Experimental Design. Biometrics, Vol.38, 907-919.

45. Davis, M. (1971). Comparison of Sequential Bioassays in Small Samples. J.R.S.S.B, Vol.33,78-87.

46. Davis, A.W. and Veitch, G.L. (1978). Approximate Confidence Intervals for Nonlinear Functions with Applications to a Cloud Seeding Experiment. Technometrics, Vol.20, 227-230.

47. Dixon, W.J. and Mood, A.M. (1948). A method for obtaining and analyzing sensitivity data. JASA, Vol.43, 109-126.

48. Draper, N.R. and Hunter, W.G. (1967a). The use of prior distributions in the design of experiments for parameter estimation in nonlinear situations. Biometrika, Vol.54, 147-153.

49. Draper, N.R. and Hunter, W.G. (1967b). The use of prior distributions in the design of experiments for parameter estimation in nonlinear situations: multi-response case. Biometrika, Vol.54, 662-665.

50. Draper, N.R. and Hunter, W.G. (1966). Design of experiments for parameter estimation in multiresponse situations. Biometrika, Vol.53, 525-533.

51. Draper, N.R. and Smith, H. (1981). Applied Regression Analysis, 2d edition. J. Wiley, N.York.

52. Dupac, V. and Kral, F. (1970). Robbins-Monro procedure with both variables subject to experimental error. Ann. Math. Stat., Vol.43, 1089-1095.

53. Elfving, G. (1952). Optimum Allocation in Linear Regression Theory. Ann. Math. Stat., Vol.23, 255-262.

54. Fabian, V. (1967). Stochastic Approximation of minima with improved asymptotic speed. Ann. Math. Stat., Vol.38 191-200.

55. Fedorov, V.V. (1972). Theory of Optimal Experiments. Academic Press.

56. Fertig, K.W. and Mann, N.R. (1980). Life-Testing Sampling Plans of Two-Parameter Weibull Populations. Technometrics, Vol.22, 165-177.

57. Finney, D.J. (1978). Statistical Methods in Biological Assay. C. Griffin and Co. LTD.

58. Fisher, R.A. (1922). On the Mathematical Foundation of Theoretical Statistics. Phil. Trans. Roy. Soc., London, ser. A., Vol.22, 309-368.

59. Fisher, R.A. (1935). The logic of Inductive Inference. J.R.S.S., Vol.XCVIII, 39-54.

60. Fisher, R.A. (1947). The Design of Experiments. Oliver and Boyd, London.

61. Ford, I. (1976). Optimal Static and Sequential design: A critical review. Ph.D Thesis, U. of Glasgow.

62. Ford, I. and Silvey, S.D. (1980). A sequentially constructed design for estimating a nonlinear parameter function. Biometrika, Vol.67, 381-388.

63. Ford, I., Titterington, D.M. and Wu, C.F.J. (1985). Inference and sequential design. Biometrika, Vol.72, 545-551.

64. Fryer, J.G. and Pethybridge, R.J. (1975). Maximum Likelihood Estimation for a special type of grouped data with an application to a dose response problem. Biometrics, Vol.31, 633-642.

64a. de la Garza, A. (1954). Spacing of information in polynomial regression. Ann. Math. Stat., Vol. 25, 123-130.

65. Gallant, A.R. (1975). Nonlinear Regression. Amer. Statist., Vol.29, 73-81.

66. Cramer, H. (1946). Mathematical methods in Statistics. Princeton University Press.

67. Graybill, F.A. (1969). Introduction to Matrices with Application to Statistics. Wadsworth Pu. Co.

68. Graybill, F.A. (1976). Theory and Application of Linear Models. Duxbury Press, Massachusetts.

69. Gross, J.M. and Wiseall, B. (1972). Principles of physical Chemistry (Chapter XXI). McDonald and Evans LTD.

70. Guttman, I. and Meeter, D.A. (1965). On Beale's Measure of Nonlinearity. Technometrics, Vol.7, 623-637.

71. Guttman, L. and Guttman, R. (1959). An Illustration of the use of Stochastic Approximation. Biometrics, Vol.15, 551-559.

72. Halmos, P.R. (1958). Measure theory.
Van Nostrad, Princeton, N.J.

73. Hamilton, D.C., Watts, D.G. and Bates, D.M. (1982). Accounting for intrinsic nonlinearity in nonlinear regression parameter inference regions.
Ann. Stat., Vol.10, 386-393.

74. Hamilton, D.C. and Watts, D.G. (1985). A quadratic Design Criterion for Precise Estimation in Nonlinear Regression Models. Technometrics, Vol.27, 241-250.

75. Hamilton, D.C. (1986). Confidence regions for parameter subsets in nonlinear regression.
Biometrika, Vol.73, 57-64.

76. Henson, T.L. and Hunter, W.G. (1969). Estimation of Monolayer Capacity for BET Equation.
J. of Catalysis, Vol.11(2), 197-200.

77. Hetzel, M.R. and Clark, T.J.H. (1980). Comparison of normal and asthmatic circadian rhythms in peak expiratory flow rate. Thorax, Vol.35, 732-738.

78. Hill, P.D.H. (1976). Optimal Experimental Design for Model Discrimination. Ph.D. thesis, U. Of Glasgow.

79. Hill, P.D.H. (1980). D-optimal Designs for partially Nonlinear Regression Models. Technometrics, Vol.22, 275-276.

80. Hill, W.J., Hunter, W.G. and Wichern, D.W. (1968). A joint design criterion for the dual problem of Model Discrimination and Parameter estimation.
Technometrics, Vol.10, 145-160.

81. Hill, W.J. and Hunter, W.G. (1974). Design of Experiments for subsets of Parameters. Technometrics, Vol.16, 425-434.

82. Hodges, J.L. and Lehmann, E.C. (1956). Two approximations to the Robbins Monro Process. Proceedings of the Third Berkley Symposium on Mathematical Statistics and Probability, U. of California Press, Berkley, 95-104.

83. Hohman, G. and Jung, W. (1975). On sequential and non-sequential D-optimal experiment design. Biom. Z. Bd., Vol.5, 329-336.

84. Izenman, A.J. and Rinoft, Y. (1977). Sequential Designs in Exponential Life Tests. Technometrics, Vol.19, 455-459.

85. Jennrich, R.J. (1969). Asymptotic properties of non-linear Least-Squares estimators. Ann. Math. Stat., Vol.40, 633-643.

86. Karlin, S. and Studden, W.J. (1966). Optimal experimental designs. Ann. Math. Stat., Vol.37, 783-815.

87. Kati, S.K. (1962). Some methods of Estimation for Poisson Bionomial distribution. Biometrics, Vol.18, 42-51.

88. Katz, D., Azen, S.P. and Schumitzky, A. (1981). Bayesian Approach to the Analysis of Nonlinear Models: Implementation and Evaluation. Biometrics, Vol.37, 137-142.

89. Kershaw, C.D. (1983). Sequential estimation for binary response. Ph.D thesis, Edinburgh University.

90. Kesten, H. (1958). Accelerated stochastic approximation. Ann. Math. Stat., Vol.29, 41-59.

91. Khuri, A.I. (1984). A Note on D-optimal Designs for Partially Nonlinear Regression Models. Technometrics, Vol.26, 59-61.

92. Kiefer, J. (1957). Optimum sequential search and Approximation Methods undes Minimum Regularity Assumptions. SIAM, Vol.5, 105-136.

93. Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). Ann. Stat., Vol.2, 849-879.

94. Kiefer, J. and Wolfowitz, J. (1952). Stochastic estimation of the maximum of a regression function. Ann. Math. Stat., Vol.23, 462-466.

95. Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems. Canadian J. Math., Vol.12, 363-366.

96. Kitsos, C.P. (1978). Choosing the order of a response polynomial - Structural approach with minimax criterion. Master's thesis, U. of New Brunswick, Canada.

97. Lai, T.L. and Robbins,H. (1981). Adaptive design and Stochastic Approximation. Ann. Stat., Vol.7, 1196-1221.

98. Lai, T.L. and Robbins, H. (1979). Consistency and Asymptotic Efficiency of Slope Estimates in Stochastic Approximation Schemes. Z. Wahrscheinlichkeitstheorie verw. Gebite, Vol.56, 329-360.

99. Lai, T.L. and Wei, C.Z. (1982). Least Squares estimates in stochastic regression models with applications to identification and control of dynamic systems. Ann. Stat., Vol.10, 154-166.

100. Lai, T.L. and Siegmund, D. (1983). Fixed accuracy estimation of an autoregressive parameter. Ann. Stat., Vol.11, 478-485.

101. Lauter, E. (1974). A method of designing experiments for nonlinear models. Math. Operationsforsch. U. Stat., Vol.5, 697-708.

102. Lord, M.F. (1971). Tailored Testing, An application of Stochastic Approximation. JASA. Vol.66, 707-711.

103. Little, R.E. (1974). A mean square error comparison of certain median response estimates for the Up-and-Down method with small samples. JASA, Vol.69, 202-210.

104. Maistrov, L.E. (1974). Probability theory, a Historical sketch. Academic Press, N. York.

105. Marshall, A.W. and Olkin, I. (1967). A multivariate exponential distribution. JASA, Vol.62, 30-44.

106. Maxim. L.D., Hendrickson, A.D. and Cullen, D.E. (1977). Experimental design for sensitivity Testing: The Weibull Model. Technometrics, Vol.19, 405-412.

107. McCullagh, P. and Nelder, J.A. (1983). Generalized Linear Models. Chapman and Hall.

108. McLeish, D. and Tosh, D. (1983). The estimation of extreme quantiles in logistic bioassay. Biometrika, Vol.70, 625-632.

109. Meeker, W.Q. and Hahn, G.J. (1978). A comparison of Accelerated Test Plans to Estimate the Survival Probability at a Design Stress. Technometrics, Vol.20, 245-247.

110. Meeker, W.Q. and Hahn, G.J. (1977). Asymptotically Optimum Over-Stress Tests to Estimate the Survival Probability at a Condition with a Low Expected Failure Probability. Technometrics, Vol.19, 381-399.

111. Nelson, W., Tong, Y.L., Lee, J.K. and Halberg F. (1979). Methods for Cosinor - Rhythmometry. Chronobiologia, Vol.6, 305-323.

112. Ortega, J.M. and Rheinbolt. W.C. (1970). Iterative Solution of Nonlinear equation in several variables. Academic Press.

113. Owen, R.J. (1975). A Bayesian Sequential Procedure for Quantal Response in Context of Adaptive Mendal Testing. JASA, Vol.70, 351-356.

114. Papakyriazis, P.A. (1978). Optimal Experimental Design in Econometrics, The Time Series Problem. J. Econometrics, Vol.7, 351-372.

115. Pazman, A. (1980). Some features of the optimal Design Theory - A survey. Math. Operationsforsch. Statist. Ser. Statistics, Vol.11, 415-446.

116. Pazman, A. (1981). Probability distribution of the multivariate nonlinear least squares estimators. Kybernetika, Vol.20, 209-230.

117. Plackett, R.L. (1972). The discovery of the method of least squares. Biometrika, Vol.59, 239-251.

118. Pritchard, D.J. and Bacon, D.W. (1977). Accounting for Heteroscedasticity in Experiment Design. Technometrics, Vol.19, 109-115.

110. Pritchard, D.J., Downie, J. and Bacon, D.W. (1977). Further consideration of Heteroscedasticity in Fitting kinetic Models. Technometrics, Vol.19, 227-236.

111. Pukelsheim, F. and Titterington, D.M. (1983). General Differential and Lagrangian theory for Optimal Experiment Design. Ann. Stat., Vol.11, 1060-1068.

112. Ratkowsky, D.A. (1983). Nonlinear Regression Modeling. A Unified Practical Approach. Mercel Dekker, Inc.

113. Rao, C.R. (1965). Linear Statistical Inference and its Application. J. Willey.

114. Robbins, H. and Monro, S. (1951). A stochastic Approximation Method. Ann. Math. Stat., Vol.22, 400-407.

115. Ruppert, D. (1979). A new dynamic stochastic approximation procedure. Ann. Stat., Vol.7, 1179-1195.

116. Ruppert, D. (1981). Stochastic approximation of an implicity defined function. Ann. Stat., Vol.9, 555-566.

117. Ryan, T.A., Joiner, B.L and Ryan, B.F. (1981). MINITAB Reference Manual.

118. Sacks, J. (1958). Asymptotic distribution of stochastic approximation procedure. Ann. Math. Statist., Vol.29, 373-405.

119. Sakrison, D.J. (1965). Efficient recursive estimation; application to estimating the parameters of a covariance function. Int. J. Eng. Science, Vol.3, 461-483.

120. Seber, G.A.F. (1977). Linear Regression Analysis. J. Wiley.

121. Sibson, R. (1972). Discussion of paper by H. P. Wynn. J.R.S.S.B., Vol.34, 181-183.

122. Silvapulle, M.J. (1981). On the existence of the Maximum Likelihood Estimators for the Bionomial Response Models. J.R.S.S.B., Vol.43, 310-313.

123. Silvey, S.D. (1972). Discussion of parer by H.P. Wynn. J.R.S.S.B., Vol.34, 174-175.

124. Silvey, S.D. (1980). Optimal Design. Chapman and Hall

125. Silvey, S.D. and Tittertington, D.M. (1973). A g eometric approach to optimal design theory. Biometrika, Vol.60, 21-32.

126. Smith, K. (1918). On the standard deviation 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, Vol.12, 1-85.

127. Steinberg, D.M. and Hunter, W.G. (1984). Experiment Design: Review and Comment. Technometrics, Vol.26, 71-97.

128. St John, R.C. and Draper, N.R. (1975). D-optimality for Regression Designs: A Review. Technometrics, Vol.17, 15-23.

129. Stigler, S.M. (1981). Gauss and the invention of Least Squares. Ann. Stat., Vol.9, 465-474.

130. Stoer, J. and Bulirsch, R. (1979). Introduction to Numerical Analysis. Springer-Verlag.

131. Stone, M. and Morris, J.A. (1985). Explanatory studies in non-sequential non-local non-linear design. J. of Statstical Planning and Inference, Vol.12, 1-9.

132. Titterington, D.M. (1975). Optimal design: some geometrical aspects of D-optimality. Biometrika, Vol.62, 313-320.

133. Titterington, D.M. (1980a). Aspects of Optimal Design in Dynamic Systems. Technometrics, Vol.22, 287-299.

134. Titterington, D.M. (1980b). Geometric Approaches to Design of Experiment. Math. Oper. Stat., Vol.11, 151-163.

135. Torsney, B. (1981). Algorithms for a constrained optimization problem with applications in Statistics and optimum Design. Ph.D thesis, U. of Glasgow.

136. Torsney, B. (1985). Personal communication.

137. Tsutakawa, R.K. (1967). Random walk design in Bio-assay. JASA, Vol.62, 842-856.

138. Tsutakawa, R.K. (1972). Design of Experiment in Bioassay JASA, Vol.67, 584-590.

139. Tsutakawa, R.K. (1980). Design of Dose Levels for estimating a Percentage Point of a Logistic Quantal Response Curve. Appl. Statist., Vol.29, 25-33.

140. Venter, J.H. (1967). An extension of the Robbins-Monro procedure. Ann. Math. Stat., Vol.38, 181-190.

141. Volund, A. (1978). Application of the four-parameter Logistic Model to Bioassay. Comparison with slope Ratio and Parallel Line Methods. Biometrics, Vol.34, 357-365.

142. Wang, S.G. and Wu, C.F.J. (1983). Further results on the consistent direction of least squares estimators. Ann. Stat., Vol.11, 1257-1262.

143. Wasan, M.T. (1969). Stochastic Approximation. Cambridge, At the University Press.

144. Watts, D.G. (1986). Influential Data Analysis Fourth Sheffield Conference, Inluence Data Analysis, 8-11 April 1986.

145. Wetherill, G.B. (1963). Sequential estimation of quantal response curves. J.R.S.S.B., Vol.25, 1-38.

146. Wetherill, G.B., Chen, H. and Vasudeva, R.B. (1966). Sequential estimation of quantal response curves : A new method of estimation. Biometrika, Vol.53, 439-454.

147. Wetherill, G.B. (1975). Sequential methods in Statistics. Chapman and Hall.

148. White, L.V. (1973). An extension to the General Equivalence Theorem for nonlinear models. Biometrika, Vol.60, 345-348.

149. White, L.V. (1975). The optimal design for the experiments for estimation in nonlinear models. Ph.D thesis, Imperial College, London.

150. Whittle, P. (1973). Some general points in the theory of optimal experiment design. J.R.S.S.B., Vol.35,123-130.

151. Wolfowitz, J. (1973). On the stochastic approximation method of Robbins and Monro. Ann. Math. Stat., Vol.23, 457-461.

152. Wu, C.F.J. and Wynn, H.P. (1978). The convergence of general step-length algorithms for regular optimum design criteria. Ann. Stat., Vol.6, 1273-1285.

153. Wu, C.F.J. (1981). Asymptotic Theory of Nonlinear Least Squares estimates. Ann. Stat., Vol.9, 501-513.

154. Wu, C.F.J. (1985a). Efficient Sequential Designs With Binary Data. JASA, Vol.80, 974-984.

155. Wu, C.F.J. (1985b). Asymptotic inference from sequential design in a nonlinear situation. Biometrika, Vol.72, 553-558.

156. Wu, C.F.J. (1985c). Personal Communication.