Bayesian Experimental Design: A Review

Kathryn Chaloner[1] and Isabella Verdinelli[2]

Abstract

This paper reviews the literature on Bayesian experimental design, both for linear and nonlinear models. A unified view of the topic is presented by putting experimental design in a decision theoretic framework. This framework justifies many optimality criteria, and opens new possibilities. Various design criteria become part of a single, coherent approach.

Key-words and Phrases: Decision Theory. Hierarchical Linear Models. Logistic Regression. Nonlinear Design. Nonlinear Models. Optimal Design. Optimality Criteria. Utility Functions.

1 Introduction

Non-Bayesian experimental design for linear models has been reviewed by Steinberg and Hunter (1984) and in the recent book by Pukelsheim (1993). Ford, Kitsos and Titterington (1989) reviewed non-Bayesian design for nonlinear models. This paper considers the theory of non-Bayesian design only as needed for the development. DasGupta (1995) presents a complementary review of Bayesian and non-Bayesian optimal design.

1.1 Experimental Design

The design of experiments is an important part of scientific research. Design involves specifying all aspects of an experiment and choosing the values of variables that can be controlled before the experiment starts. Control variables might include: choosing which treatments to study, defining the treatments precisely, choosing blocking factors, choosing how to randomize, specifying the experimental units to be used, specifying a length of time for the experiment to be performed, choosing a sample size and choosing the proportion of observations to allocate to each treatment. These are all relevant aspects in design.

[1]Kathryn Chaloner is Associate Professor, Department of Statistics, University of Minnesota, Minneapolis, Minnesota

[2]Isabella Verdinelli is Associate Professor, Dipartimento di Statistica, Probabilità e Statistiche Applicate, Università di Roma "La Sapienza" and Visiting Associate Professor, Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

Common sense, available resources, and knowledge of the motivation for carrying out the experiment often help in selecting important features that depend on the specific problem. In designing a clinical trial, where experiments are to be performed on people, the choice of the treatments to be compared is a central ethical issue. Whether or not a placebo is ethical and, if not, what to use as a control treatment, is a problem that does not arise in designing a field experiment to compare yields from different fertilizer combinations. Not all aspects of experimental design are susceptible to abstract mathematical treatment. Choosing values for the control variables however can be simply expressed in a mathematical framework. This problem has been considered at length in the scientific literature and is focused on in this paper.

When designing an experiment, decisions must be made before data collection, and data collection is restricted by limited resources. Because information is usually available prior to experimentation and, indeed, often motivates doing the experiment, Bayesian methods are ideally suited to contribute to experimental design. Bayesian decision theory also motivates precise specification of the reason the experiment is being conducted. Like most areas of Bayesian statistics, Bayesian experimental design has gained popularity in the past two decades. But also like many areas of Bayesian statistics, applications to actual experiments still lag behind the theory. Apart from Flournoy (1993) there are no true "case studies" that we know of, where Bayesian ideas have been formally applied to the design of an actual scientific experiment before it is done. This is a very important area for future work. There are however several examples of examining an experiment in a Bayesian design framework after it has been done: for example Clyde, Müller and Parmigiani (1994) and some examples presented in this paper.

The basic idea in experimental design is that statistical inference about the quantities of interest can be improved by appropriately selecting the values of the control variables. In estimation problems, estimators with small variance are usually desirable. Control variables should therefore be selected to achieve small variability for the estimator chosen. Much depends however on what is to be estimated, and how it will be estimated. Specifying the purpose of the experiment generates various criteria for the choice of a design.

We address the fundamental principles of design by providing a general Bayesian decision

theoretic framework for a coherent approach. Most of the work in Bayesian design can be included as special cases in this general structure. The usefulness of this approach and the improvement that can be obtained over designing within the non-Bayesian theory is shown with some examples. Three examples are presented in Section 1.2. They will be examined again in Sections 3 and 6 to illustrate the type of improvement that can be obtained over non-Bayesian designs and that, sometimes, the experiments used in practice are approximately Bayes optimal.

1.2 Examples

Example 1. Consider the one-way analysis of variance model where a given total of n observations must be divided among t groups in an optimal way. The observations are measurements on the experimental units and the groups correspond to t treatments whose effects are of interest. Assume also that the variance of the observations is known. Assigning the same number of observations to each group is a possibility, but different choices of these proportions might be more appropriate, depending on the type of experiment. The same one-way analysis of variance model can be used either in a trial to study the effect of t different treatments, or when the effects of $t-1$ similar treatments are to be compared with a standard control, or, perhaps, when a given drug is to be tested at t different dose levels. Intuitively, it may be sensible to have different designs for each of these situations.

To give an idea of the difference a Bayesian approach can make, consider the following two experiments. One experiment consists in comparing $t-1$ similar treatments with a control. A second one takes observations in t different groups to study the effect of increasing levels $z_1, z_2, \ldots z_t$ of a certain drug. Both experiments are modeled through a one way analysis of variance but they are essentially different. Section 3 presents a way to use prior knowledge in these examples. A prior distribution used in the first experiment will not necessarily be appropriate in the second. A non-Bayesian approach to design would typically consider the two experiments exactly the same and the same design would be chosen for both of them. The Bayesian approach has instead more flexibility as is shown in Section 3.

Example 2. At a University of Minnesota laboratory large numbers of animal experiments were done to assess the potency of individual batches of drug. The laboratory performed

logistic regressions on many different drugs and biologic material.

For one particular drug under study, 54 similar experiments were performed and the same type of design was used for each of the experiments. The design usually consisted of six equally spaced doses with ten mice exposed to each dose. Sixty animals were required for each experiment. Occasionally less than sixty animals were available in which case less than ten animals were exposed to the highest dose. The responses measured were the number of surviving mice one week after being given the dose. Different numbers of mice died depending on the potency of the batch of drug and by chance. Typically a high proportion (80% or 90%) of the mice died at the high levels and a lower proportion (20% or 10%) died at the low levels. After each experiment, the potency of the batch was calculated using maximum likelihood to estimate the LD50, the dose at which the probability of a mice dying was estimated to be 0.50. Two typical data sets are given in table 1 from a experiments that looked at the potency of different concentrations of albumen.

TABLE 1.

Dose	Number Exposed	Number Dead	Dose	Number Exposed	Number Dead
2.5	10	0	2.5	10	0
3.0	10	1	3.0	10	2
3.5	10	1	3.5	10	1
4.0	10	3	4.0	10	4
4.5	10	5	4.5	10	7
5.0	10	6	5.0	10	8
BATCH 1			BATCH 2		

This example will be discussed further in Section 6.2. The design used, six equally spaced doses with ten animals at each dose, was chosen for convenience. It is straightforward to do the experiment, and with large numbers of experiments, it is simpler to use the same design each time. It is natural to ask whether the experiments could have been designed differently. It is also natural to use results from this set of experiments for constructing a prior distribution to use for subsequent experiments. The 54 estimates of LD50 can be thought of a sample from a distribution of possible values. We therefore consider a prior distribution for the LD50 that reasonably reflects the observed sample and has approximately the same first two moments as the sample.

Example 3. Atkinson, Chaloner, Herzberg and Juritz (1993) examined a designed experiment to investigate bioavailability. The experiment, described in a 1979 unpublished PhD thesis by Button at Texas A&M University, consisted of giving 15mg/kg of theophylline as aminophylline to a number of horses by intra-gastric administration. Blood samples were then drawn at different times, t, after injection and the concentration of drug, y, measured. The value of y was modeled to be related to t through an open one-compartment model with first-order absorption input:

$$y = \theta_3(e^{-\theta_1 t} - e^{-\theta_2 t}) + \epsilon.$$

The observation errors ϵ are independent and normally distributed with mean zero. The unknown parameters $(\theta_1, \theta_2, \theta_3)$ are such that $\theta_2 > \theta_1$. At time $t = 0$ the expected response is zero and, as t increases, it increases up to a maximum and then decreases to zero as t gets larger. Several quantities are of interest including the area under the expected response curve, the time at which the maximum is reached, and the value at the maximum. The design problem is choosing the times at which to take blood samples. The design used in Button's thesis is fairly typical of these experiments and is an 18 point design with one measurement at each time and the times are approximately uniform on a log scale.

As in example 2, this is another case of a nonlinear design problem. Atkinson, Chaloner, Herzberg and Juritz (1993) looked at the efficiency of the 18 point design used by the experimenter and constructed Bayesian optimal designs under several prior distributions suggested by the data. This is further discussed in Section 6.4.

1.3 Overview of the Bayesian Approach

Experimental design is the only situation where it is meaningful within the Bayesian theory to average over the sample space. As the sample has not yet been observed, the general principle of averaging over what is unknown applies. Following Raiffa and Schlaifer (1961), Lindley (1972, page 19 and 20) presented a decision theory approach to experimental design. Lindley's argument is essentially the following.

A design η must be chosen from some set $\mathcal{H}$, and data $\mathbf{y}$ from a sample space $\mathcal{Y}$ will be observed. Based on $\mathbf{y}$ a decision d will be chosen from some set $\mathcal{D}$. The decision is in two parts: first the selection of η, and then the choice of a terminal decision d. The unknown

parameters are θ, and the parameter space is Θ. A general utility function is of the form $U(d, \theta, \eta, \mathbf{y})$.

For any design η, the expected utility of the best decision is given by

$$U(\eta) = \int_{\mathcal{Y}} \max_{d \in \mathcal{D}} \int_{\Theta} U(d, \theta, \eta, \mathbf{y})\, p(\theta|\mathbf{y}, \eta)\, p(\mathbf{y}|\eta) d\theta d\mathbf{y}, \tag{1}$$

where $p(\cdot)$ denotes a probability density function with respect to an appropriate measure. The Bayesian solution to the experimental design problem is provided by the design η^* maximizing: (1)

$$U(\eta^*) = \max_{\eta \in \mathcal{H}} \int_{\mathcal{Y}} \max_{d \in \mathcal{D}} \int_{\Theta} U(d, \theta, \eta, \mathbf{y})\, p(\theta|\mathbf{y}, \eta)\, p(\mathbf{y}|\eta) d\theta d\mathbf{y}. \tag{2}$$

In other words, Lindley's argument suggests that a good way for designing experiments is to specify a utility function reflecting the purpose of the experiment, regard the design choice as a decision problem, and select a design that maximizes the expected utility.

The present paper pursues Lindley's approach as a unifying formulation for the theory of Bayesian experimental design. Selecting a utility function that appropriately describes the goals of a given experiment is very important. A design that is optimal for estimation is not necessarily also optimal for prediction. Even restricting attention to optimal designs for estimation, there are a variety of criteria that lead to different designs, depending on what is to be estimated and what utility function is used. The choice of a utility (or loss) function expresses various reasons for carrying out an experiment.

In the linear model, the analogs of widely known non-Bayesian alphabetical design criteria (Box, 1982) such as A-optimality, D-optimality and others can be given decision theoretic justification. When inference about the parameters is the main goal of the analysis, for example, a utility function based on Shannon information leads to Bayesian D-optimality in the normal linear model (see, Bernardo, 1979). In addition, Shannon information can be used for prediction and in mixed utility functions that describe several simultaneous goals for an experiment. Bayesian equivalents of some other popular optimality criteria can also be derived by choosing appropriate utility functions. Some, but not all of the alphabetical optimality criteria, have a utility-based Bayesian version.

There are cases where prediction might be considered more important than inference when designing an experiment. This might be the case, for example, in settings like reliability and quality control where the future level of output has to be kept on target, or in clinical trials when it is important to obtain information on how patients will respond to some treatment. For these types of problems the predictive Bayesian approach is appropriate for both design and analysis. For a detailed treatment of this topic, see Geisser (1993).

Other utility functions can be devised for designing experiments that take into account more specific issues. For example as argued by Lindley and Novick (1981) randomization is unnecessary for inference in a Bayesian experiment: it is "merely useful". Randomization is an important practical aspect of design, especially in clinical trials. Verdinelli (1990) and Ball, Smith and Verdinelli (1993) considered this problem for the linear model within the theory of Bayesian optimal experimental design.

1.4 Notation

In the linear model with n independent observations, X stands for a $n \times k$ design matrix. The rows of X, $\mathbf{x}_j^T, j = 1 \ldots n$ are elements of a compact set $\mathcal{X}$ of design points available to the experimenter. The matrix $X^T X$ is denoted by nM and it is often referred to as the information matrix, since the Fisher information matrix is equal to $\sigma^{-2} nM$. If n_i observations are taken at the point $\mathbf{x}_i \in \mathcal{X}$, then the information matrix can be written as $n \sum (n_i/n)\ \mathbf{x}_i \mathbf{x}_i^T$ with $\sum n_i/n = 1$. Following Fedorov (1972, page 62) and many other authors, define $\eta_i = n_i/n$ so $nM = n \sum \eta_i \mathbf{x}_i \mathbf{x}_i^T$. A design can now be seen as a probability measure η on the region $\mathcal{X}$ of design points. It is usually convenient to relax the requirement for the n_i's to be integers so that the design problem becomes that of finding an optimal design measure η from the set of all probability measures on $\mathcal{X}$; this set is denoted $\mathcal{H}$. We will use both notations nM and $nM(\eta)$ for the information matrix. In some situations, it may be of interest to find exact optimal designs where the probability measure η is such that, for a specified n, the values $n\eta_i$ are all integers.

In some cases, using a linear model, exact calculations for expected utility, $U(\eta)$ as given by (1) and (2) in Section 1.3, are possible. For nonlinear models, expected utilities do not have a closed form representation. Approximations are therefore required. It is often still

possible, however, to formulate the problem in a similar way. The design problem is still to choose values of the control variables $\mathbf{x}_j,\ j = 1, \ldots, n$ from a compact set $\mathcal{X}$. If, just as in the linear case, we denote η_i to be the proportion of observations at a point $\mathbf{x}_i$ then in both linear and nonlinear models the design problem can be thought of as choosing a probability measure η over $\mathcal{X}$ from $\mathcal{H}$. We will see in Sections 4, 5 and 6 that design for nonlinear models presents some challenges. A Bayesian approach can provide practical insight and lead to useful solutions.

Relaxing the requirement for $n\eta_i$ to be integer values makes the problem more tractable. Designs where the proportions are not constrained to correspond to integers for some n are referred to as *approximate* or *continuous* designs. An approximate design can be rounded to an exact design without losing too much efficiency (see for example Pukelsheim, 1993, Chapter 12, for some rounding algorithms and discussion). Without the relaxation to non-integer designs the design problem is that of a hard integer programming problem. Majumdar (1988, 1992) derived Bayesian exact designs for a two way analysis of variance model considering a special subclass of prior distributions. This is a particularly useful approach when dealing with the constraints of incomplete blocks. Toman (1994) derived Bayes optimal exact designs for two- and three-level factorial experiments, with and without blocking. One of the important problems she solved is that of choosing a fraction of the full factorial design.

Most approaches to design assume that there is a fixed number n of observations to be taken. Subject to this constraint, a probability measure on $\mathcal{X}$ should be chosen to maximize the expected utility. This formulation has led to a research area known as "Optimal design" or "Optimal Bayesian design". One of the most powerful tools for finding designs is the General Equivalence Theorem (Kiefer, 1959, Whittle, 1973). Of course there may be other constraints such as a fixed total cost, C, and each observation may cost a different amount c_i. The problem then becomes to maximize utility subject to a fixed cost C. The equivalence theorem can easily be adapted to deal with this extension. See for example Chernoff (1972, p. 16) who showed that a simple linear transformation can modify the problem to the more familiar one with a fixed sample size. This is applied to Bayesian linear design problems in Chaloner (1982). Tuchscherer (1983) finds Bayesian linear optimal designs for particular cost functions.

1.5 Structure of the Paper

Sections 2 and 3 of this paper deal with designs for linear models. Bayesian analogs of alphabetical design criteria are introduced in Section 2.2 and are examined in 2.3. Other design criteria within the Bayesian decision theory approach are discussed in Section 2.4. The case of unknown error variance is considered in 2.5. Section 3 is devoted to the simple but important case of analysis of variance models. The examples considered illustrate the effect of incorporating prior information in the linear model.

Nonlinear models are examined in Sections 4 and 5. Various possible approximations to expected utility are investigated in 4.2. Section 4.3 deals with some of the different Bayesian approaches. Local optimality is considered in 4.4. The approximations are compared in 4.5. Properties of optimal nonlinear Bayesian design are discussed in Section 5. For example it is shown that the number of support points in an optimal design may depend on the prior distribution. Some exact results are given and the available software is reviewed. Section 6 considers a few other specific problems in nonlinear design such as sample size in clinical trials and design for reliability and quality control.

Nonlinear problems generated from a linear model are considered in Section 7. Additional problems, such as design for variance components, for a mixture of linear models and for model discrimination, are discussed in Section 8. Section 9 contains concluding remarks.

2 Bayesian designs for the normal linear model

2.1 Introduction

Consider the problem of choosing a design η for a normal linear regression model. The data $\mathbf{y}$ is a vector of n observations where $\mathbf{y}|\theta, \sigma^2 \sim N(X\theta, \sigma^2 I)$, θ is a vector of k unknown parameters, σ^2 is known and I is the $n \times n$ identity matrix. Suppose that the prior information is such that $\theta|\sigma^2$ is normally distributed with mean θ_0 and variance-covariance matrix $\sigma^2 R^{-1}$, where the $k \times k$ matrix R is known. Recall, from Section 1.4, that the matrix $X^T X$ is denoted by nM or, equivalently, $nM(\eta)$. The posterior distribution for θ is also normal with mean vector $\theta^* = (nM(\eta) + R)^{-1}(X^T\mathbf{y} + R\theta_0)$ and covariance matrix $\sigma^2 D(\eta) = \sigma^2(nM(\eta) + R)^{-1}$; $D(\eta)$ is a function of the design η and of the prior precision matrix $\sigma^{-2}R$.

2.2 Bayesian Alphabetical Optimality: Overview

Following Lindley's (1956) suggestion, several authors (Stone, 1959 a, b; DeGroot, 1962, 1986; Bernardo, 1979) considered the expected gain in Shannon information given by an experiment as a utility function (Shannon, 1948). These authors proposed choosing a design that maximizes the expected gain in Shannon information or, equivalently, maximizes the expected Kullback-Leibler distance between the posterior and the prior distributions:

$$\int \log \frac{p(\theta|\mathbf{y},\eta)}{p(\theta)}\, p(\mathbf{y},\theta|\eta)\; d\theta d\mathbf{y}. \tag{3}$$

The prior distribution does not depend on the design η, so the design η maximizing the expected gain in Shannon information is the one that maximizes:

$$U_1(\eta) = \int \log\{p(\theta|\mathbf{y},\eta)\}\, p(\mathbf{y},\theta|\eta)\; d\theta d\mathbf{y}, \tag{4}$$

This is the expected Shannon information of the posterior distribution. This expected utility $U_1(\eta)$ might be appropriate when the experiment is conducted for inference on the vector θ. In the normal linear regression model

$$U_1(\eta) = -\frac{k}{2}\log(2\pi) - \frac{k}{2} + \frac{1}{2}\log\det\{\sigma^{-2}(nM(\eta)+R)\}.$$

This utility therefore reduces to maximizing the function $\phi_1(\eta) = \det\{nM(\eta)+R\}$ and it is known as Bayes D-optimality. Non-Bayesian D-optimality maximizes the determinant of $M(\eta)$. Note the symbol $\phi(\cdot)$ is used to denote a design criterion function and $U(\cdot)$ is used to denote an expected utility function.

In the non-Bayesian design literature, there are papers discussing the augmentation of a previous design. That is, for D-optimality choosing η to maximize the determinant of $(nM + X_0^T X_0)$ where $X_0^T X_0$ is fixed, typically from a design obtained previously. This is clearly algebraically identical to Bayesian D-optimality and is discussed in Covey-Crump and Silvey (1970), Dykstra (1971), Evans (1979), Mayer and Hendrickson (1973), Johnson and Nachtsheim (1983) and Heiberger, Bhaumik and Holland (1993).

A variation of non-Bayesian D-optimality is D_S-optimality, see, for example, Silvey (1980

p. 10-11). This criterion maximizes the inverse determinant of the covariance matrix for the least squares estimator of a linear function $\psi = \mathbf{s}^T\theta$ of the parameters. The equivalent Bayesian criterion is obtained considering the posterior distribution of ψ in (4). Not much attention has been given to this criterion in the Bayesian literature, but its use is straightforward.

Bayesian D-optimality can be derived from other utility functions as well. Assume that interest is in inference for θ and that $p(\cdot)$ is chosen to represent its probability density function. The following utility function is associated with the true value of the parameter θ and with the function $p(\cdot)$ selected as probability density function for θ:

$$U(\theta, p(\cdot), \eta) = 2p(\theta) - \int p^2(\tilde{\theta})d\tilde{\theta} \ . \tag{5}$$

This utility function is a proper scoring rule, first introduced by de Finetti (1962) for discrete θ. Buehler (1971) proposed its use for eliciting beliefs about θ, both in the discrete and in the continuous case. Spezzaferri (1988) adopted (5) for designing experiments for model discrimination and parameter estimation. He also showed that in the normal linear model, when interest is in estimation of θ, (5) reduces to

$$\left(2\sigma\sqrt{\pi}\right)^{-k} \left\{\det\left[nM(\eta) + R\right]\right\}^{1/2},$$

thus obtaining the D-optimality criterion. Eaton, Giovagnoli and Sebastiani (1994) also use utility functions based on proper scoring rules for prediction and also derive D-optimality as a special case.

Another justification of Bayesian D-optimality was derived by Tiao and Afonja (1976) through the following two valued utility function:

$$U(\hat{\theta}, \theta, \eta) = \begin{cases} 0 & |\hat{\theta} - \theta| < a \\ -1 & |\hat{\theta} - \theta| > a \end{cases} , \tag{6}$$

where $\hat{\theta}$ denotes an estimator for θ and a is an arbitrarily small positive constant.

When the specific reason for conducting an experiment is to obtain a point estimate of the parameters, or of linear combinations of them, a quadratic loss function might be appropriate. In this case a design can be chosen to maximize the following expected utility:

$$U_2(\eta) = -\int (\theta - \hat{\theta})^T A(\theta - \hat{\theta})\; p(\mathbf{y}, \theta|\eta)\; d\theta d\mathbf{y}, \tag{7}$$

where A is a symmetric non negative definite matrix. The Bayes procedure yields as expected utility $U_2(\eta) = -\sigma^2 tr\{AD(\eta)\}$ and a corresponding criterion $\phi_2(\eta) = -tr\{AD(\eta)\} = tr\{A(nM(\eta) + R)^{-1}\}$. A design that maximizes $\phi_2(\eta)$ is called Bayes A-optimal, a generalization of the non-Bayesian A-optimality criterion, that minimizes $tr\{AM(\eta)^{-1}\}$. This criterion also arises when minimizing the expected squared error loss for estimating $\mathbf{c}^T\theta$ or when minimizing the squared error of prediction at $\mathbf{c}$, where $\mathbf{c}$ is not necessarily fixed and a distribution is specified on it. See Owen (1970), Brooks (1972, 1974, 1976, 1977), and Duncan and DeGroot (1976). Chaloner (1984) showed how an equivalence theorem can be used for this criterion, derived a bound on the number of support points in an optimal design and presented some examples. Toman and Notz (1991) considered this criterion for analysis of variance models with two-way heterogeneity. Toman (1992a) and Toman and Gastwirth (1993) dealt with A-optimality in a robustness context and Toman (1994) examined A-optimality for factorial experiments.

A special case of A-optimality is when $\text{rank}(A) = 1$, that is $A = \mathbf{c}\mathbf{c}^T$ and $U_2(\eta) = -\sigma^2 \mathbf{c}^T D(\eta)\mathbf{c}$; this variation of A-optimality is called Bayes c-optimality and it parallels the non-Bayesian c-optimality. This optimality criterion is also obtained when the expected squared loss is used for estimating a given linear combination of the parameters: $\psi = \mathbf{c}^T\theta$ where $\mathbf{c}$ is fixed. A Bayesian modification of the geometric argument in Elfving's (1952) theorem for c-optimality was given in Chaloner (1984) and extended in El-Krunz and Studden (1991) and Dette (1993a, b).

An extension of the notion of the c-optimality criterion is E-optimality, for which the maximum posterior variance of all possible normalized linear combinations of parameter estimates is minimized. As a heuristic argument to motivate E-optimality, consider an experiment to estimate the linear function $\psi = \mathbf{c}^T\theta$, for unspecified $\mathbf{c}$, with the normalizing

constraint $\|\mathbf{c}\| = w$. A minimax approach leads to searching for a design that is good for different choices of $\mathbf{c}$. Denoting the maximum eigenvalue of a matrix H by $\lambda_{max}[H]$, an E-optimal design minimizes

$$\sup_{\|\mathbf{c}\|=w} \mathbf{c}^T D(\eta)\mathbf{c} = w^2 \lambda_{max}[D(\eta)]. \tag{8}$$

This criterion appears not to correspond to any utility function and so, although it is referred to as Bayes E-optimality, its Bayesian justification, in a decision theoretic context, is unclear.

Closely related to Bayesian E-optimality is Bayesian G-optimality. A G-optimal design is chosen to minimize $\sup_{\mathbf{x}\in\mathcal{X}} \mathbf{x}^T D(\eta)\mathbf{x}$. Similarly to E-optimality, this does not correspond to maximizing a utility function (although there is an equivalence theorem, see Pukelsheim, 1993, sect. 11.6, that states that continuous G-optimal designs are numerically identical to a corresponding continuous D-optimal design).

Tiao and Afonja (1976) presented other utility functions aimed at the problems of selecting the best of k parameters and of ranking the parameters. They also proposed, in addition to the utility (6), the quadratic utility in (7) and the following exponential utility:

$$U(\eta) = 1 - \exp\left\{-\frac{\alpha}{2}(\hat{\theta} - \theta)^T(\hat{\theta} - \theta)\right\}.$$

They considered the problem of choosing among a class of balanced designs to illustrate the use of the above utilities and to show that a design often has to be selected from a limited range of available ones.

It is important to recall briefly the main relations between Bayesian and non-Bayesian design criteria. A characteristic of optimal Bayesian design measures is the dependence on the sample size n, since $D(\eta) = n^{-1}(M(\eta) + n^{-1}R)^{-1}$. This identity shows that any differences between a Bayesian design and its corresponding non-Bayesian one are unimportant if n is large, since, in this case, $(M(\eta) + n^{-1}R)$ is approximately equal to $M(\eta)$. This is intuitively reasonable: in experiments where the sample size is large the posterior distribution will be driven by the data and will not be sensitive to the prior distribution. In contrast, if n is small the prior distribution will have more of an effect on the posterior distribution and on

the design.

Letting $n \to \infty$ is equivalent to $R \to 0$ and a similar limiting result is seen. When there is little prior information available, optimal Bayesian designs are close to the corresponding non-Bayesian ones. Hence, when a noninformative prior distribution is used for inference, as may often be the case, there is no advantage to using the Bayesian approach for design.

This limiting behavior is not seen in design for nonlinear models where usual non-Bayesian optimal designs are again special cases of Bayesian design but correspond to a point mass prior distribution rather than non informativeness. This is discussed further in section 4.

Note also that non-Bayesian design criteria, such as c-optimality and D_S-optimality must be adapted to allow for designs where the optimal choice of $nM(\eta)$ may be singular. For Bayesian design criteria, no such adaptation is required. The matrix R is non-singular for a proper informative prior distribution, so the matrix $nM(\eta) + R$ is always non-singular irrespective of whether $nM(\eta)$ alone is.

2.3 Bayesian Alphabetical Optimality: Related Work.

In the 1970's Lindley's work had a profound influence on many aspects of Bayesian statistics. In the area of experimental design, a set of papers by Brooks (Brooks 1972, 1974, 1976, 1977) were inspired by work of Lindley's on the choice of variables in multiple regression (Lindley 1968). Brooks followed Lindley's approach to motivate the problem of choosing the best subset of regressors and the design points in a linear regression model. Predicting the future value of the dependent variable is the goal of the experiment and the predictor is obtained substituting the Bayesian estimator in the regression function, rather than considering the predictive distribution for the future observation. A quadratic loss function, plus costs, is used to evaluate the difference of the future value of y and its predictor. Bayes A-optimality with added costs is the design criterion derived. In his 1974 paper, Brooks also looked for optimal sample size using the same loss function and in his 1977 paper he dealt with design problems when controlling for the future value of y to be at a preassigned value y_0. The setting considered in Brooks' early papers is too general to allow for many explicit solutions and few special cases are explored. Straight line regression is examined in his 1976 paper. Brooks' work can be seen as a statement of the general

principle that the Bayesian method has a way for dealing with the design problem. Bayes optimality criteria are considered as elements of a class of linear criteria. This last feature shows the influence of Fedorov's 1972 book. It is also found in Pukelsheim (1980) and in Pilz's work (for example Pilz,1991) where Bayesian design criteria are seen mostly as extensions of the corresponding non-Bayesian criteria, the focus often being placed in showing that non-Bayesian criteria are limiting cases when diffuse prior information is considered. See also Fedorov (1980, 1981).

Brooks also examined the case of σ^2 being unknown and used the simple solution to the problem that substitutes the value of σ^2 with its prior mean wherever it appears in the final expression of the criterion. This approach was also used by other authors, for example Sinha (1970), Guttman (1971) and more recently Pukelsheim (1993, chapter 11). They defined optimality criteria without a decision theory based framework and so have no clear extension to the case where σ^2 is unknown. In contrast, with a decision theory based framework, the extension to the case where σ^2 is unknown is conceptually easy but, as is shown in Section 2.5, algebraically hard.

Pilz dealt with Bayes experimental designs for a linear model in a series of papers (Pilz 1979a, b, c, d, 1981a, b, c Näther and Pilz 1980, Gladitz and Pilz 1982a, b, Bandemer, Näther and Pilz 1987). See also the monograph, Pilz (1983) and the revised reprint of the monograph, Pilz (1991). His approach is very general, with no distributional assumptions for the model or for the prior distribution. Pilz defined Bayes alphabetical optimality criteria as an extension of the corresponding non-Bayesian criteria and looked at them as special cases of a general "Linear optimality criterion". D-optimality and E-optimality do not fall into this setting, so Pilz often derived separate results for these criteria. The methodology used throughout Pilz' work has the flavor of classical decision theory. For example, he considered admissible and complete classes of designs to find conditions for the existence of Bayes designs in an admissible class. Pilz also adapted much of the existing theory on optimal design to the Bayesian case. He used Whittle's (1973) general version of the equivalence theorem to find relations among the different design criteria and to find bounds for the designs. Pilz also showed that under certain conditions, Bayes alphabetical designs can be constructed as A-optimal designs for a transformed model. In some cases, A-optimality coincides with

D- and E- optimality, but the conditions under which the above holds do not seem easy to satisfy. Pilz did not give explicit designs and examine their practical implications and his work is somewhat abstract.

2.4 Other Utility Functions

As noted in section 1, in certain experiments, prediction can be more important than estimation. In quality control and in clinical trials prediction of future observations can be of special interest. In these cases the Bayesian approach uses predictive analysis which can also be helpful in designing the experiment. The expected gain in Shannon information on a future observation y_{n+1} is used rather than the expected gain in information on the vector of parameters. The expected Kullback-Leibler distance between the predictive distribution $p(y_{n+1}|\mathbf{y},\eta) = \int p(y_{n+1}|\theta)p(\theta|\mathbf{y},\eta)\ d\theta$ (posterior predictive) and the marginal distribution $p(y_{n+1})$ (prior predictive) on y_{n+1} is the equivalent of the quantity (3) in section 2.2. The prior predictive distribution does not depend on the design and the design that maximizes the expected gain in Shannon information on y_{n+1} is equivalent to the design that maximizes the expected utility:

$$U_3(\eta) = \int \log p(y_{n+1}|\mathbf{y},\eta)\ p(\mathbf{y},y_{n+1}|\eta)\ d\mathbf{y}dy_{n+1}. \tag{9}$$

This utility function has been used by San Martini and Spezzaferri (1984) for a model selection problem and by Verdinelli, Polson and Singpurwalla (1993) for accelerated life test experiments. In the normal linear model, maximizing $U_3(\eta)$ with respect to η corresponds to maximizing

$$-\frac{1}{2}\left\{\log(2\pi) + 1 + \log\left[\sigma^2\mathbf{x}_{n+1}^T D(\eta)\mathbf{x}_{n+1} + \sigma^2\right]\right\},$$

where the next observation is going to be taken at the point $\mathbf{x}_{n+1} \in \mathcal{X}$. This is equivalent to minimizing the predictive variance

$$\sigma_{n+1}^2 = \sigma^2[\mathbf{x}_{n+1}^T D(\eta)\mathbf{x}_{n+1} + 1].$$

In the special case of prediction of y_{n+1} at a fixed point $\mathbf{c} = \mathbf{x}_{n+1}$, the design maximizing

$U_3(X)$ corresponds to the Bayes c-optimal design presented in section 2.2.

Yet another situation is where the experimenter is concerned with the value of the response variable y. In these cases, one might be interested not only with inference on the parameters, but also with obtaining a large value of the outcome. Experimentation might be considered only if the design proposed is expected to produce a large value of outcome as well as a large value of information. In such cases, one possibility is to look for a design that maximizes a combination of the expected total output and the expected Shannon information for the posterior distribution. Verdinelli and Kadane (1992) proposed the following expected utility:

$$U_4(\eta) = \int \left[\rho \mathbf{y}^T \mathbf{1} + \beta \ \log p(\theta|\mathbf{y},\eta)\right] p(\mathbf{y},\theta|\eta) d\mathbf{y} d\theta. \tag{10}$$

The non-negative weights ρ and β express the relative contribution that the experimenter is willing to attach to the two components of $U_4(\eta)$. In the normal linear model, these weights affect the choice of the design through the ratio β/ρ. A design maximizing $U_4(\eta)$ is equivalent to a design maximizing

$$\int \mathbf{y}^T \mathbf{1} p(\mathbf{y}) d\mathbf{y} \quad + \quad \frac{\beta}{2\rho} \ \log\det\{D(\eta)\}.$$

Verdinelli (1992) suggested the use of another expected utility function when the goal of the experiment is both inference about the parameters and prediction about the future observation. It is given by a combination of $U_1(\eta)$ and $U_3(\eta)$, namely:

$$U_5(\eta) = \gamma \int \log p(y_{n+1}|\mathbf{y},\eta) p(\mathbf{y}, y_{n+1}|\eta) \ d\mathbf{y} dy_{n+1} + \omega \int \log p(\theta|\mathbf{y},\eta) \ p(\mathbf{y},\theta|\eta) \ d\mathbf{y} d\theta. \tag{11}$$

As in $U_4(\eta)$, the weights γ and ω express the relative contribution of the predictive and the inferential components of the utility. In this case, the two components are expressed in the same units. In the linear model the expected utility U_5 is maximized by a design that maximizes

$$-\frac{\gamma}{2}\left\{\log(2\pi) + 1 + \log\left[\sigma^2 \mathbf{x}_{n+1}^T D(\eta) \mathbf{x}_{n+1} + \sigma^2\right]\right\} - \frac{\omega}{2}\left\{k\log(2\pi) + k - \log\det(\sigma^{-2} D^{-1}(\eta))\right\}.$$

This is equivalent to minimizing $\sigma^2_{n+1}\det\{\sigma^2 D(\eta)\}$, where σ^2_{n+1} is the predictive variance, defined earlier. It turns out that the weights γ and ω do not affect the choice of the design.

Yet another formulation of the design problem as a decision problem is given in Toman (1995). She examined design when the purpose of the experiment is hypothesis testing.

2.5 Unknown Variance

If the variance σ^2 in the linear model of section 2.1 is unknown then the optimality criteria induced by the utility functions of the earlier sections may need to be modified, although conceptually the goal of maximizing a utility remains the same. Let the prior distribution for (θ, σ^2) be conjugate in the normal-inverted gamma family: $\theta|\sigma^2 \sim N(\theta_0, \sigma^2 R^{-1})$ and $\sigma^{-2}|\alpha, \beta \sim \text{Ga}(\alpha, \beta)$, so that $p(\sigma^2|\alpha,\beta) \propto (\sigma^2)^{-(\alpha+1)}\exp\{-\beta\sigma^{-2}\}$. This implies that both the prior and the posterior marginal distributions for θ are multivariate t distributions. Denote by $t_\delta[m, \mu, \Sigma]$ the probability distribution of an m-variate t random variable with δ degrees of freedom, mean vector μ and scale matrix Σ (see for example DeGroot 1970, sec 5.6 or Box and Tiao 1973 page 117). Recall that $\theta^* = (nM(\eta) + R)^{-1}(X^T\mathbf{y} + R\theta_0)$. Let $h(\eta, \mathbf{y})$ denote the quantity $(2\alpha+n)^{-1}\left\{(\mathbf{y} - X\theta_0)^T\left[I - X(nM(\eta) + R)^{-1}X^T\right](\mathbf{y} - X\theta_0) + 2\beta\right\}$ and let $a = \beta/\alpha$. The prior and posterior marginal distributions for θ are:

$$\theta \sim t_{2\alpha}\left[k, \theta_0, aR^{-1}\right] \qquad \text{and} \qquad \theta|\mathbf{y}, \eta \sim t_{2\alpha+n}\left[k, \theta^*, h(\eta, \mathbf{y})(nM(\eta) + R)^{-1}\right].$$

The distribution of $\mathbf{y}$ conditional on θ alone is multivariate t: $\mathbf{y}|\theta \sim t_{2\alpha}[n, X\theta, aI]$. In addition, the marginal distribution of the data $\mathbf{y}$ is multivariate t:

$$\mathbf{y}|\eta \sim t_{2\alpha}\left[n, X\theta_0, a[I - X(nM(\eta) + R)^{-1}X^T]^{-1}\right],$$

and the posterior predictive distribution for y_{n+1}, a new observation at $\mathbf{x}_{n+1}$, is univariate t: $y_{n+1}|\mathbf{y}, \eta \sim t_{2\alpha+n}\left[1, \mathbf{x}_{n+1}\theta^*, h(\eta, \mathbf{y})\{\mathbf{x}_{n+1}(nM(\eta) + R)^{-1}\mathbf{x}_{n+1} + 1\}\right]$.

Evaluating the expected utilities presented in sections 2.2 and 2.4 is now a more complicated task. The integrals that define U_1, U_3, U_4 and U_5 are now intractable since no closed form expression can be derived. Numerical approaches or approximations, such as the normal approximations (12) or (13) described later, in section 4.2, are needed to find Bayesian

designs.

Things are somewhat simpler for A-optimality and U_2. In the expression for $U_2(\eta)$, letting $A = I$, the integral in (7) reduces to $\int \text{tr Var}(\theta|\mathbf{y})p(\mathbf{y})d\mathbf{y}$ where $\text{Var}(\theta|\mathbf{y})$ denotes the posterior covariance matrix and $p(\mathbf{y})$ is the marginal distribution of $\mathbf{y}$. The A-optimality criterion reduces to finding a design η that minimizes

$$\frac{2\alpha + n}{2\alpha + n - 2} \text{ tr } (nM(\eta) + R)^{-1} \times \left\{ \int h(\eta, \mathbf{y})p(\mathbf{y})d\mathbf{y} \right\}.$$

The integral in the above formula is equal to $[2\beta n(2\alpha - 1)^{-1} + 2\beta]/(2\alpha + n)$, which does not depend on $\mathbf{y}$. Hence Bayes A-optimality is insensitive to the knowledge of σ^2 and in this sense it is a robust criterion for choosing a design. See also Chaloner (1984). This feature of A-optimality makes it appealing to use. It remains to be seen how design developed from distributional distances are influenced by the prior distribution on σ^2.

3 Design for analysis of variance models

3.1 Introduction

In section 2, we showed how a decision theoretic setting for experimental design leads to well defined optimality criteria for the linear model. This section deals with the important special case of models for the analysis of variance. In these cases criteria from Section 2 sometimes allow the derivation of explicit forms for optimal designs. Two different ways of building normal prior distributions for the vector θ are examined. Bayesian optimal designs are considered when θ has prior mean θ_0 and covariance matrix $\sigma^2 R^{-1}$, as in section 2.1. In addition, Bayesian optimal designs under a hierarchical prior distribution, as in Lindley and Smith (1972), are also derived. The hierarchical normal linear model can be used to represent different experimental settings. A given criterion like, say, Bayes D-optimality yields different designs for various choices of the hierarchical structure that describes the experiment.

3.2 Analysis of Variance Models

In the one way analysis of variance model, when the effects of t treatments are of interest, the matrix nM is simply $\text{diag}\{n_1, n_2, \ldots n_t\}$, where n_i is the number of observations in the

i-th group. Choosing an optimal design for this model consists in choosing the number of observations n_i or the proportions of observations $\eta_i = n_i/n$ on each treatment.

Duncan and DeGroot (1976) considered the problem of Bayesian optimal design for the one way analysis of variance model using the A-optimality criterion, defined in section 2.2. In one of the cases they examined, one of the t treatments is a control and the contrasts of interest compare the $t-1$ treatments to the control.

In the two-way case, with the second factor being a blocking variable, there might be t treatments and b blocks. The choice of a design for this model is equivalent to the choice of n_{ij}, the number of observations taken on the i-th treatment in the j-th block. If the block sizes k_j are fixed, this is the same as choosing the proportions $\eta_{ij} = n_{ij}/k_j$ of units to assign to the treatments in each of the blocks. Owen (1970) and Giovagnoli and Verdinelli (1983) considered Bayesian designs for the two-way model with treatments and blocks. One of the treatments is a control and the parameters of interest are the contrasts of the treatments with the control. Owen dealt with A-optimality while Giovagnoli and Verdinelli examined a class of criteria proposed, in a non-Bayesian context, by Kiefer (1975). The class is defined for a parameter $p \geq 0$ as $\Phi_p = \{k^{-1}\mathrm{tr}[D(\eta)]^p\}^{1/p}$. Bayesian A-optimality is a special case when $p = 1$, Bayesian D-optimality results when $p \to 0$ and Bayesian E-optimality when $p \to \infty$. Having defined this class, Giovagnoli and Verdinelli then focused on D-optimal designs. Simeone and Verdinelli (1989) used nonlinear programming techniques to derive E-optimal Bayes designs for the same model.

Bayesian designs for analysis of variance models were derived in Toman (1992a, 1994, 1995). Designs for models with two blocking factors were examined by Toman and Notz (1991), who mainly considered A-optimality criterion, but also presented solutions for D- and E-optimality.

3.3 Example 1 Continued

Following Duncan and DeGroot (1976) let us now consider the A-optimality criterion in the one way analysis of variance model. Let $\theta = (\alpha_1, \alpha_2, \ldots \alpha_t)^T$ represent the treatment effects and suppose the experiment is designed to study the contrasts $\alpha_i - \alpha_1$ of the effects of $t-1$ new treatments compared with a control for $i = 2, \ldots, t$. Assume that the treatment

effects α_i are independent and normally distributed with prior means μ_i and variance τ_i^2. The use of utility function (7) for Bayesian A-optimality, leads to an optimal proportion of observations on the control

$$\eta_1 = \max\left\{0\ , \frac{1+\sum_{j=1}^{t} 1/n\tau_j^2}{1+\sqrt{t-1}} - \frac{1}{n\tau_1^2}\right\}$$

and on the ith new treatment

$$\eta_i = \max\left\{0\ , \frac{\sqrt{t-1}\left(1+\sum_{j=1}^{t} 1/n\tau_j^2\right)}{1+\sqrt{t-1}} - \frac{1}{n\tau_i^2}\right\}, \text{for } i = 2, \ldots, t;$$

with the constraint $\sum_{j=1}^{t} \eta_j = 1$. If the same prior mean and variance μ_2 and τ_2^2 say, are assigned to all the new treatments, to represent that they are thought to be independent and have the same prior distribution, then the A-optimal proportions of observations can be written as:

$$\begin{aligned} \eta_1 &= \max\left\{0\ , \sigma^{-2}\frac{\sigma^2+\delta_1+(t-1)\delta_2}{1+\sqrt{t-1}}\right\} \\ \eta_2 &= (t-1)^{-1}(1-\eta_1), \end{aligned}$$

where $\delta_1 = \sigma^2/n\tau_1^2$ and $\delta_2 = \sigma^2/n\tau_2^2$. From these expressions we can see the limiting behavior of η_1 and η_2. As the value of the prior variances gets large with respect to σ^2/n, that is for δ_1 and δ_2 both small, the result approaches the non-Bayesian A-optimal proportion. A proportion $\{\sqrt{t-1}+1\}^{-1}$ of the observations are on the control and the rest are equally divided among the other $t-1$ treatments. This design is sometimes called "the square root rule", since it places the same number of observations on all the new treatments and $\sqrt{t-1}$ as many on the control. When, instead, δ_1 is large compared with δ_2 – meaning that prior information is less precise on the new treatments that it is on the control – then the A-optimal design puts no observations on the control. Similarly, if δ_2 is large compared to δ_1 it may be optimal to put all the observations on the control.

Assume now that the utility function chosen is $U_1(\eta)$ in (4) and the experiment is designed to be Bayesian D-optimal. Suppose that the new treatment effects are still assumed to be

independent and identically distributed. The optimal proportion of observations on the control, η_1, depends again on the ratios $\delta_1 = \sigma^2/n\tau_1^2$ and $\delta_2 = \sigma^2/n\tau_2^2$. When δ_1 and δ_2 are both small, the non-Bayesian D-optimal design is obtained, that places the same proportions of observations $1/t$ on the new treatments and on the control. In contrast, if δ_1 is large, there is precise knowledge of the effect of the control, and it may be optimal to take no observations on the control, just as in Bayesian A-optimality. Similarly if δ_2 is large the prior information about the new treatments is precise and no observations need to be taken on them.

When the optimal design takes no observations on a treatment, then the only information on that treatment in the posterior distribution will be from the prior information. Some experimenters might well find this feature unappealing: some might argue that this is not even an experiment. In implementing such a design the assumption is clearly critical that the prior information really does represent accurate information about the experimental units in this particular experiment. This is always an important assumption to examine, especially when the optimal Bayesian design is so different from the corresponding optimal non-Bayesian design. But of course it is in exactly these cases of precise prior information or, equivalently, of small planned sample size, that Bayesian optimal designs can improve over non-Bayesian designs if the critical assumption holds.

Similar results are obtained when the utility function chosen is (10) and concern is both on inference and on yielding a large value of the total output. In this case, the optimal proportion η_1 on the control depends both on the prior means and on the prior variances. It can be shown that there are two threshold values, F and G, functions of δ_1 and δ_2 only, such that if $\mu_2 - \mu_1 \leq F$ then $\eta_1 = 1$ and if $\mu_2 - \mu_1 \geq G$ then $\eta_1 = 0$. Hence the optimal design does not take observations on the new treatments if the prior mean μ_2 of the new treatment effect is small compared with the prior mean of the control effect μ_1. Similarly no observations are taken on the control if μ_2 is large compared to μ_1.

3.4 Hierarchical form for the prior distribution

The use of a hierarchical normal linear model is motivated by Lindley and Smith (1972). The basic model consists of three stages. The first stage is the sampling distribution and it is just the usual normal linear model with a vector of parameters θ, say, as described in

section 2.1. The second and third stages together are used to model the prior distribution for θ. The linear models of earlier sections are obtained when the prior distribution is expressed through one stage only. We now consider prior distributions specified in two stages. The distribution of θ at the first stage is expressed through a vector of hyperparameters and a second stage is added to specify the distribution of the hyperparameters.

For example in the one way analysis of variance model, let the sampling distribution be such that the y_{ij} are independent and

$$y_{ij}|\theta_i \sim N(\theta_i, \sigma^2),$$

with σ^2 known. To represent the information that all the group effects θ_i are similar, then the first stage of the prior distribution is that, conditional on some value μ, the θ_i are independent with mean μ, and with the same known variance τ^2. That is the θ_i are a sample from the same distribution $\theta_i|\mu \sim N(\mu, \tau^2)$. The second stage of the prior distribution represents the uncertainty in μ: for example $\mu \sim N(0, \omega^2)$. Then the marginal distribution of the parameters θ_i is such that the θ_i are exchangeable, but not independent. The θ_i's are positively correlated, representing that they are believed to be similar. Even if $\omega^2 \to \infty$, representing vague prior knowledge, the distribution still retains a correlation structure.

As Lindley and Smith (1972, page 7) remarked, it is the type of experiment that often suggests the specifications for the first stage of the prior, that describe the relationship existing among the elements of θ. At the second stage, knowledge is likely to be weak, so it is natural to express this by assuming a distribution for the hyperparameters that is dispersed. Under this type of prior distribution, the marginal distribution of the data, $\mathbf{y}$ is formally that of a random effects model rather than a fixed effects model.

Under this hierarchical structure, the Bayesian optimal design criteria derived from D-optimality (4) and A-optimality (7) are different than under a prior distribution set only in one stage. Relatively little research has been done on design with a hierarchical prior distribution and more work is needed in this area.

3.5 Example 1 Continued

Assume again there is a control group and $t-1$ treatment groups. The first stage of the prior distribution is such that, conditional on μ_1 and μ_2, the control effect is normally distributed with mean μ_1 and variance τ_1^2, and the $t-1$ treatment effects $(\alpha_2, \alpha_3, \ldots, \alpha_t)$ are normal with a mean vector $\mathbf{1}\mu_2$, where $\mathbf{1}$ is a $(t-1)$ vector of one's, and a variance matrix $\tau_2^2 I$. The variances τ_1^2 and τ_2^2 are assumed to be known. Let the prior distribution of μ_1 and μ_2 be flat and improper to represent that not much is known about treatments and control, apart from the fact that μ_1 and μ_2 are thought to be different. Collapsing the two-stages gives a singular prior precision matrix $\sigma^{-2}R$:

$$R = \frac{\sigma^2}{(t-1)\, n\tau_2^2} \begin{bmatrix} 0 & : & \mathbf{0}^T \\ \cdot & \cdot & \cdot \\ \mathbf{0} & : & (t-1)I - J \end{bmatrix},$$

where $J = \mathbf{1}\mathbf{1}^T$. The matrix R is such that the mean of the control effect is independent of the mean effects of the new treatments, that the new treatments means are exchangeable, but not independent of each other, and that the prior distribution is non-informative with respect to the control. The symmetry built into this model is such that for any of the utility functions of section 2, the optimal proportions of observations on each new treatment is the same, η_2 say, with $0 \leq \eta_2 \leq (t-1)^{-1}$, and the proportion of observations on the control is $\eta_1 = 1 - (t-1)\eta_2$.

In this case too, different designs are generated from different utilities. When (4) is chosen, for Bayesian D-optimality, and interest is on inference on either the vector $\theta = (\alpha_1, \alpha_2, \ldots, \alpha_t)$ or on the contrasts $\alpha_i - \alpha_1$; for $i = 2, \ldots, t$, the optimal proportion of observations on the control η_1, can be expressed as a function of the ratio $\delta = \sigma^2/n\tau_2^2$. When $\delta \to \infty$ the prior variance for the new treatments is small compared with the error variance σ^2/n. This implies that the new treatments are believed to be very similar to each other (which might often be the case in practice) and $\lim_{\delta\to\infty} \eta_1(\delta) = 1/2$. Hence half of the observations are on the control, and the rest are equally divided among the new treatments. Intuitively, to compare two independent treatments, the observations would

be divided equally between the two and this is, essentially, the situation when $\delta \to \infty$. In contrast, if the prior variance for the new treatments is large compared with σ^2/n, that is $\delta \to 0$, than $\lim_{\delta \to 0} \eta_1(\delta) = 1/t$. Exchangeability in this case reduces to the new treatments being independent, and all treatments, control and new, get an equal allocation of observations.

The design derived from (7), is such that the A-optimal proportion η_1 also depends only on δ. When $\delta \to \infty$ and the new treatment means are believed to be very similar then, $\lim_{\delta \to \infty} \eta_1(\delta) = 1/2$. This solution is the same found for D-optimality when $\delta \to \infty$. In fact, for this limiting case, all alphabetically optimal criteria coincide. In contrast, when the new treatment means are independent, represented by $\delta \to 0$, we get again the square root rule given in Section 3.2.

As mentioned in Section 1.2, let us consider now a different experiment (Smith and Verdinelli, 1980) where the prior means of each group are formed to study the effects of increasing levels $z_1, z_2, \ldots z_k$ of a given drug. This experiment too can be modeled specifying the prior distribution in two stages. At the first stage the prior means of each group are on a response curve described by a low degree polynomial. Hence

$$\alpha_i \sim N(\gamma_0 + \gamma_1 z_i + \ldots + \gamma_r z_i^r, \tau^2)$$

where $i = 1, 2, \ldots, t$; $r < t$; a straight line corresponds to choosing $r = 1$ and a quadratic to $r = 2$. At the second stage, the prior distribution for the hyperparameters $\gamma_0, \gamma_1, \ldots, \gamma_r$ is chosen to be non informative, thus representing that the only knowledge available is about the type of response surface, not about its actual form. Deriving optimal designs for this type of assumptions requires numerical implementation.

Figure 1 shows how the D-optimal proportions η_i vary when $\delta = \sigma^2/n\tau^2$ increases, for equally spaced z_i and an orthogonal polynomial representation. The left hand side of the figure shows D-optimal proportions for seven groups ($t = 7$) and a straight line at the first stage of the prior distribution ($r = 1$). The right hand side of the figure is essentially the same, but the D-optimal proportions are for nine groups ($t = 9$) and for a second degree polynomial at the first stage of the prior distribution ($r = 2$). Note how the D-optimal proportions in the groups behave consistently with the strength of prior beliefs in

the polynomial at the first stage, as represented by the ratio δ between sample and prior

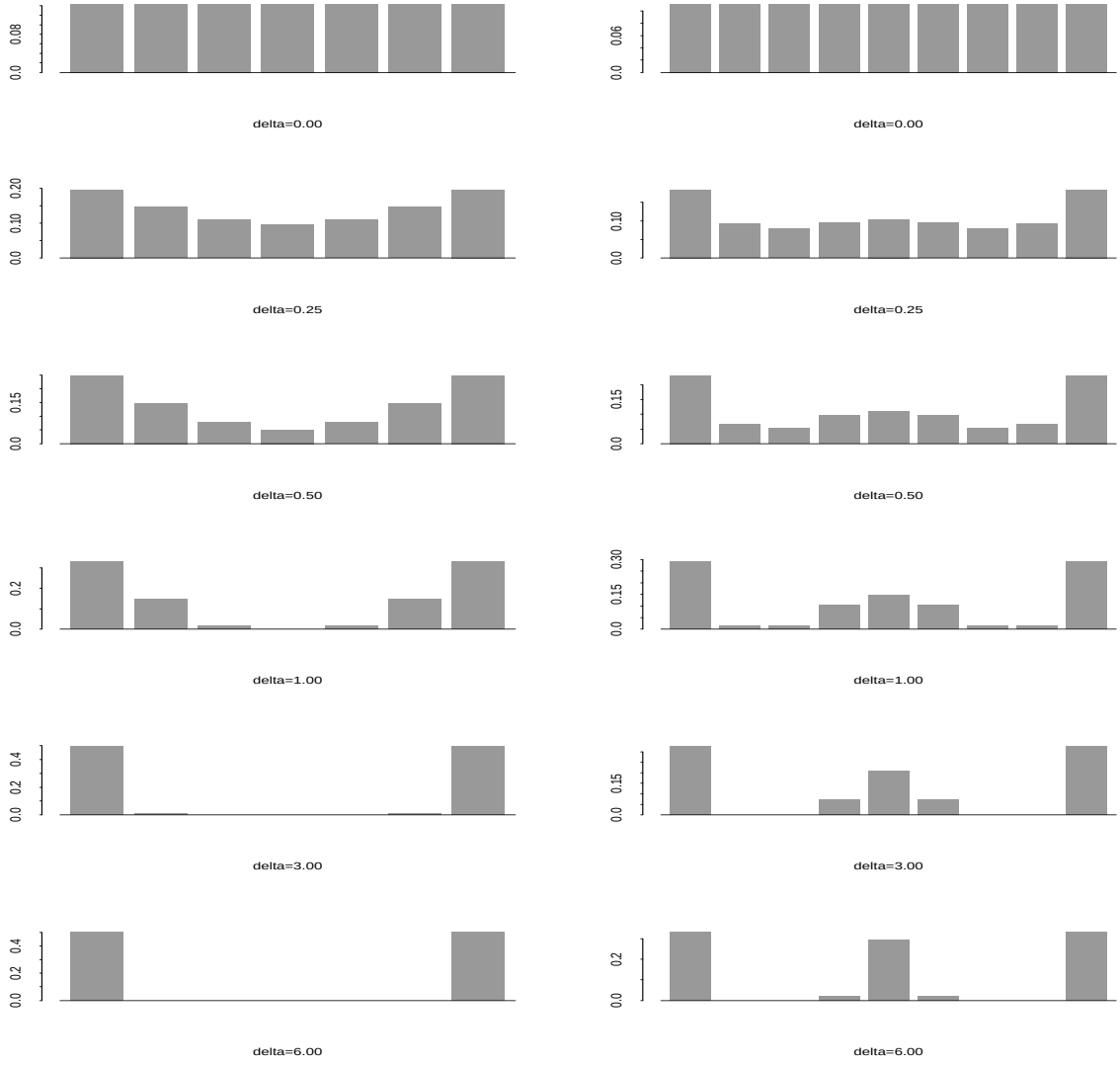

Figure 1

Left D-optimal proportions for $k = 7$ groups and a straight line at the first stage of the prior

Right D-optimal proportions for $k = 9$ groups and a parabola at the first stage of the prior

variances. The optimal proportions of observations on the t groups vary from the non-Bayesian D-optimal design for the one way model $\eta_i = 1/t$, when δ is small, to the non-Bayesian D-optimal designs for the polynomial chosen when δ is large. This last case corresponds to assuming a strong prior knowledge about the polynomial relationship for the t groups, while not considering the one-way structure particularly relevant.

4 Nonlinear design problems

4.1 Introduction

Design is more difficult when the model is not linear or when a nonlinear function of the coefficients of a linear model is of interest. Such problems are referred to as "nonlinear design problems". It will be shown that the design problem can be formulated as maximizing expected utility but approximations must typically be used as the exact expected utility is often a complicated integral. Designs can still be denoted by a probability measure η over the design space $\mathcal{X}$ and the set of all such measures be denoted $\mathcal{H}$. The measures may be arbitrary probability measures representing approximate, or continuous, designs, or measures corresponding to exact designs which have mass $1/n$ on n, not necessarily distinct, points.

4.2 Approximations to expected utility

Most approximations to expected utility involve using a normal approximation to the posterior distribution. Several normal approximations are possible, see for example Berger (1985, p. 224), and involve either the expected Fisher information matrix or the matrix of second derivatives of the logarithm of either the likelihood or the posterior density. The expected Fisher information matrix for a model with unknown parameters θ, a design η and a sample size of n is denoted by $n\mathcal{I}(\theta, \eta)$. Note that the matrix of moments, M, used in the previous sections on linear design, is a very special case of $\mathcal{I}(\theta, \eta)$, where $\mathcal{I}(\theta, \eta)$ does not depend on θ. For consistency with the literature, and to emphasize that $\mathcal{I}(\theta, \eta)$ is not necessarily a moment matrix, this separate notation is used for linear and nonlinear problems.

Let $\hat{\theta}$ denote the maximum likelihood estimate of θ. One normal approximation might be:

$$\theta | y, \eta \sim N(\hat{\theta}, [n\mathcal{I}(\hat{\theta}, \eta)]^{-1}). \tag{12}$$

In (12) the posterior normal approximation only depends on the data through $\hat{\theta}$. An alternative approximation is:

$$\theta | y, \eta \sim N(\hat{\theta}, [R + n\mathcal{I}(\hat{\theta}, \eta)]^{-1}) \tag{13}$$

where $\hat{\theta}$ now denotes the mode of the joint posterior distribution of θ (also called the gener-

alized maximum likelihood estimate of θ as in Berger, 1985, p.133), and R is the matrix of second derivatives of the logarithm of the prior density function, or the precision matrix of the prior distribution.

Several other approximations are possible, for example using the exact posterior mean and variance as the mean and variance of the approximating normal distribution, or using the observed rather than expected Fisher information matrix. Although in specific problems there may be reasons to prefer one approximation to another, and the observed, rather than the expected information matrix, almost always gives a better normal approximation to the posterior distribution, in general there is no obviously best one to use. For design purposes the expected Fisher information matrix is usually algebraically much more tractable. Using approximations other than (12) and (13) is an area for future research.

If, for illustration, Shannon information is the choice of utility then the expected utility $U_1(\eta)$ is given by equation (4), as in the linear model. $U_1(\eta)$ is the exact expected utility, which involves $p(\mathbf{y}|\eta)$, the marginal distribution of the data for a design η. As in the linear model

$$p(\mathbf{y}|\eta) = \int p(\mathbf{y}|\theta,\eta)p(\theta)d\theta.$$

In most cases this marginal distribution of $\mathbf{y}$ must also be approximated. When the posterior utility only depends on $\mathbf{y}$ through some consistent estimate $\hat{\theta}$ a further approximation, of the same order as (12) and (13), is to take the predictive distribution of $\hat{\theta}$ to be the prior distribution. Using this approximation together with (12) gives an approximate value of $U_1(\eta)$:

$$-\frac{k}{2}\log(2\pi) - \frac{k}{2} + \frac{1}{2}\int \log\det\{n\mathcal{I}(\theta,\eta)\}\ p(\theta)\ d\theta. \tag{14}$$

As in earlier sections $U(\cdot)$ will be used to denote exact expected utility and $\phi(\cdot)$ a design criterion. The constant terms and multiplier in (14) can be dropped to give

$$\phi_1(\eta) = \int \log\det\{n\mathcal{I}(\theta,\eta)\}p(\theta)\ d\theta \tag{15}$$

as a design criterion. Similarly the design criterion derived using (13) gives:

$$\phi_{1R}(\eta) = \int \log\det\{n\mathcal{I}(\theta,\eta) + R\}p(\theta)\ d\theta. \tag{16}$$

Suppose now that the only quantity to be estimated is a function of the coefficients $g(\theta)$ and squared error loss is appropriate, so that the utility is $U_2(\cdot)$ in (7). Define the k vector $\mathbf{c}(\theta)$ to be the gradient vector of $g(\theta)$. That is, the ith entry of $\mathbf{c}(\theta)$ is:

$$c_i(\theta) = \frac{\partial g(\theta)}{\partial \theta_i}. \tag{17}$$

Then, using (12), the approximate expected utility is

$$\phi_2(\eta) = -\int \mathbf{c}(\theta)^T\{n\mathcal{I}(\theta,\eta)\}^{-1}\mathbf{c}(\theta)\ p(\theta)\ d\theta. \tag{18}$$

A slightly different approximation involving R is given when (13) is used:

$$\phi_{2R}(\eta) = -\int \mathbf{c}(\theta)^T\{R + n\mathcal{I}(\theta,\eta)\}^{-1}\mathbf{c}(\theta)\ p(\theta)\ d\theta. \tag{19}$$

Should more than one function of θ be of interest, the total expected loss is the sum of the expected losses for all the nonlinear functions. This sum could be a weighted sum to represent some functions being of more interest than others. If the matrix $A(\theta)$ is the sum, or corresponding weighted sum, of the individual matrices $\mathbf{c}(\theta)\mathbf{c}(\theta)^T$ then the approximate expected utility is

$$\phi_2(\eta) = -\int \mathrm{tr}\{A(\theta)[n\mathcal{I}(\theta,\eta)]^{-1}\}\ p(\theta)\ d\theta \tag{20}$$

with a similar expression involving the matrix R if (13) is used. Criteria (15), (18) and (20) will be referred to as Bayesian D-optimality, Bayesian c-optimality and Bayesian A-optimality respectively.

Clyde (1993a) suggested that as these Bayesian design criteria are based on approximate normality it is appropriate to design to ensure that, with high probability, the posterior distribution is, approximately, normal. She suggested several approaches, including maximizing the criteria discussed above subject to some constraints that help ensure normality.

The constraints she used are developed from the ideas of Slate (1991) and Kass and Slate (1994) who gave diagnostics for posterior normality. For small sample sizes the constraints are active but for large sample sizes posterior normality is more likely and so the constraints are typically satisfied by the design maximizing the Bayesian criterion. She also looked at other ways of combining the two objectives of maximizing approximate utility and attaining approximate normality. Hamilton and Watts (1985) and Pázman and Pronzato (1992a, b), took a related non-Bayesian approach.

Müller and Parmigiani (1995) suggested estimating the exact expected utility using Markov Chain Monte Carlo methods but the effectiveness of this suggestion in specific problems has yet to be demonstrated.

4.3 Bayesian criteria

Some of the earliest papers putting design for nonlinear models in a Bayesian perspective are Tsutakawa (1972) and Zacks (1977). They both used the matrix $\mathcal{I}(\theta, \eta)$ in their design criteria. Tsutakawa considered a one parameter logistic regression with known slope coefficient and unknown LD50, denoted by θ. The criterion he maximized was the univariate case of (19), that is:

$$\phi(\eta) = -\int \{R + n\mathcal{I}(\theta, \eta)\}^{-1} p(\theta)\, d\theta. \tag{21}$$

where the integrand is a scalar. Tsutakawa numerically found designs maximizing (21), restricting the designs to equally spaced design points with equal numbers of Bernoulli observations at each design point. He gave the arguments of Section 4.2 to justify (19). In a later paper (Tsutakawa 1980), he extended similar ideas to design for the estimation of other percentile responses.

Zacks (1977) considered problems where the data are to be sampled from an exponential family with known scale parameter and where some function of the mean is linear in an explanatory variable. This class of generalized linear models includes quantal response models and models for exponential lifetimes. The Fisher information matrix has a common form for these models and Zacks considered designs that maximize the expected value of the determinant of $\mathcal{I}(\theta, \eta)$, that is:

$$\phi(\eta) = \int \det\{n\mathcal{I}(\theta,\eta)\}\ p(\theta)\ d\theta. \tag{22}$$

Zacks examined several examples and also found optimal multistage designs for quantal response experiments. Note that the criterion in (22), unlike (15), is not readily interpretable as an approximation to the expected utility (4).

A similar approach to that of Tsutakawa and Zacks to design for nonlinear models, called "robust experimental design", was developed in the field of pharmacokinetics and biological modeling. This was initially developed without mention of any Bayesian motivation. These procedures are described in Walter and Pronzato (1985) and Pronzato and Walter (1985, 1987, 1988), see also Landaw (1982, 1984). This work also relates to work on dynamic systems as in Mehra (1974) and Goodwin and Payne (1977).

In addition to using the criterion (22) Pronzato and Walter also used several other criteria such as:

$$\phi(\eta) = -\int [\det \mathcal{I}(\theta,\eta)]^{-1}\ p(\theta)\ d\theta. \tag{23}$$

$$\phi(\eta) = \det\left\{\mathcal{I}\left[\int \theta p(\theta)\ d\theta, \eta\right]\right\}. \tag{24}$$

They also discussed and derived designs based on minimax criteria.

There is a rich related literature, mostly non-Bayesian, on design, for complex pharmacokinetic and biological models. A feature which makes these methods different is that often allowances are made for inter- and intra-subject variability. Another feature of such models that is often used is non-constant error variance. Further references can be found in Launay and Iliadis (1988), Mallet and Mentre (1988), D'Argenio and Van Guilder (1988), Thomaseth and Cobelli (1988) and D'Argenio (1990). With a few exceptions, such as Katz and D'Argenio (1983), this important work is not in the mainstream statistics literature but in the scientific literature of pharmacokinetics and mathematical biology.

4.4 Local optimality

A crude approximation to expected utility would be to approximate the marginal distribution of $\hat{\theta}$ by a one point distribution. The one point would represent a "best guess". This

approach, known as local optimality, has been used extensively in nonlinear design and is due to Chernoff (1953, 1962). It is also used in the pioneering paper of Box and Lucas (1959) where the important issues in design for nonlinear regression were identified. Although they used local optimality, Box and Lucas suggested extending this by taking into account a prior distribution on the parameter values. Draper and Hunter (1967a) extend the work of Box and Lucas. White (1973, 1975) showed how results from linear design theory can be adapted to apply to local optimality in nonlinear models and she also derived locally optimal designs for binary regression experiments.

As local optimality is a very crude approximation to expected utility, it can be considered as being approximately Bayesian although it is typically not justified in this way and is usually used in a non-Bayesian framework.

The experimenter is required to specify a best guess, θ_0 for the unknown parameters θ. Local D-optimality involves choosing the design η maximizing

$$\phi_{1\theta_0}(\eta) = \det\{\mathcal{I}(\theta_0, \eta)\}. \tag{25}$$

for a fixed value θ_0. Similarly, local c-optimality is to choose η to maximize:

$$\phi_{2\theta_0}(\eta) = -\mathbf{c}^T(\theta_0)\mathcal{I}(\theta_0, \eta)^{-1}\mathbf{c}(\theta_0) = -\mathrm{tr}A(\theta_0)\mathcal{I}(\theta_0, \eta)^{-1} \tag{26}$$

which can clearly be generalized to local A-optimality. As in (18) and (19) the vector $\mathbf{c}(\theta_0)$ is the gradient vector of the function of interest, evaluated at θ_0. Typically $\mathbf{c}(\theta_0)$ depends on θ_0 as does the matrix $A(\theta_0) = \mathbf{c}(\theta_0)\mathbf{c}(\theta_0)^T$. If more than one function of the parameters is of interest then the matrix $A(\theta_0)$ is the, possibly weighted, sum of matrices corresponding to the individual functions. The weights are the relative importance of each nonlinear function. To our knowledge versions of (25) and (26) involving the matrix R have not been used.

4.5 Comparison of the approximations

The various ways to approximate (1) presented earlier and their implications will now be compared. For Bayesian D-optimality and maximizing Shannon information we compare criteria (15) and (16) and, for Bayesian c-optimality and minimizing squared error loss, (18)

and (19). These criteria are asymptotic approximations of the same order. Several aspects do distinguish them. The criteria (16) and (19)

- require the specification of R.
- give optimal designs which depend on the sample size
- avoid technical problems using prior distributions with unbounded support where, for a design with bounded support, $\mathcal{I}(\theta, \eta)$ may be arbitrarily close to being singular (as discussed in Tsutakawa, 1972).

The criteria (15) and (18) alternatively,

- can be interpreted as a procedure where a different prior distribution will be used for the analysis than was used in the design stage. A noninformative prior distribution will be used in the analysis, hence giving R identically zero, but all available prior information will be used in the design process and an informative $p(\theta)$ will be used to average over in the integral. This echoes the idea given in Tsutakawa (1972) of using different prior distributions for design and for analysis. (See also Etzione and Kadane, 1993).
- for similar reasons these criteria are appealing in a non-Bayesian framework where it is accepted that prior information must be used in design but should not be used in the analysis. Indeed this is the motivation of Pronzato and Walter (1985, 1987).

For these reasons we prefer (15) and (18) over (16) and (19). But note that for large sample sizes, or for cases where the matrix R corresponds to imprecise information, there will be very little difference between the two sets of criteria.

Versions of these criteria using the observed rather than expected information matrix, or the second derivative of the logarithm of the posterior do not appear to have been investigated and might give better designs, especially for small samples. Similarly not much is known in general about how well these criteria, which approximate expected utility, perform empirically. For a special case of example 3, Atkinson, Chaloner, Herzberg and Juritz

(1993) showed by simulation that the Bayesian criteria do well empirically. Clyde (1993) also presented some simulations. A recent paper by Sun, Tsutakawa and Lu (1995) showed by simulation that the numerical approximation of Tsutakawa (1972) for design in the one parameter logistic regression example is remarkably accurate.

4.6 Discussion

Apart from the ideal approach of maximizing exact expected utility precisely as in say (1), no single approach can comfortably be labeled as the definitive "Bayesian nonlinear design criterion". The criteria derived in this section are all approximations to the ideal. This has not always been fully understood. For example Atkinson and Donev (1992) present "Five versions of Bayesian D-optimality" in Table 19.1. They explain that the (15) corresponds to "pre-posterior expected loss" but do not explain that it is Shannon information as utility rather than loss, and it is an approximation.

5 Optimal nonlinear Bayesian design

5.1 Introduction

Chaloner (1987) and Chaloner and Larntz (1986, 1988, 1989) developed the use of criteria such as those given by (15) which are the expectation, over a prior distribution of a local optimality criterion. We refer to such criteria as "Bayesian design criteria". These design criteria are concave on $\mathcal{H}$, the space of all probability measures on $\mathcal{X}$. Subject to some regularity conditions, an equivalence theorem can be derived. The equivalence theorem was stated by Whittle (1973) in the context of linear design problems, but its application to nonlinear problems was not then apparent and the regularity conditions required for its use in the nonlinear case not stated. See also Läuter (1974, 1976) and Dubov (1977). The theorem states that, in order to verify that a design measure is optimal, it is necessary only to check that the appropriate directional derivative at that design measure, in the direction of all one point design measures is everywhere non positive. A candidate optimal approximate design can be found using numerical optimization and the theorem makes it easy to check whether the candidate design is indeed globally optimal over $\mathcal{H}$.

The theorem applies to any criterion that is an average, over a prior distribution, of a

local optimality criterion concave on $\mathcal{H}$. Most of the criteria in common usage, including those given by (15), (16), (18), (19) and (20) satisfy this condition.

For a criterion $\phi(\cdot)$ the derivative at a design measure η in the direction of another measure η_* is, when the limit exists:

$$D(\eta, \eta_*) = \lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} [\phi\{(1-\epsilon)\eta - \epsilon\eta_*\} - \phi(\eta)] \ .$$

The extreme points of $\mathcal{H}$ are the measures putting point mass at a single x in $\mathcal{X}$ and are denoted η_x. The directional derivative of $\phi(\eta)$ in the direction η_x is $D(\eta, \eta_x)$ and is denoted $d(\eta, x)$.

For example $\phi(\cdot)$ defined by (15), Bayesian D-optimality, the derivative is:

$$d(\eta, x) = \int \mathrm{tr}\mathcal{I}(\theta, \eta_x)\mathcal{I}(\theta, \eta)^{-1} p(\theta) d\theta - k \ ,$$

where k is the dimension of θ.

Regularity conditions that are sufficient for the equivalence theorem to hold are that there is at least one design η such that $\phi(\eta)$ is finite, that $\phi(\eta)$ is continuous on $\mathcal{H}$ in some topology such as weak convergence, and that the derivatives $d(\eta, x)$ of $\phi(\eta)$ exist and are continuous in x.

The extension of Bayesian criteria to situations involving nuisance parameters is straightforward under the general approach of maximizing expected utility. For Bayesian c-optimality and A-optimality no extension is required as nuisance parameters are inherent in their definition. A Bayesian D_s-criterion and its corresponding equivalence theorem can also be easily derived.

Unlike in linear problems the criterion function $\phi(\cdot)$ is not necessarily a concave function over a *finite* dimensional space and so the equivalence theorem does not provide any bound for the minimum number of points in an optimal design. This is discussed in the following section.

5.2 Number of support points

In most non-Bayesian linear problems an upper bound on the number of support points

in an optimal design is available, see Pukelsheim (1993, p. 188-9). For linear models deriving the bound relies on the fact that the matrix M depends only on the first few moments of the design measure η and Carathéodory's theorem is used. The D-optimality criterion in linear models typically leads to an optimal number of support points that is the same as the number of unknown parameters and the design takes an equal number of observations at each point (Silvey, 1980, p. 42, and Pukelsheim, 1993, section 9.5 for polynomial models).

Designs on a small number of support points are easy to find and their theoretical properties are readily examined. They are not very appealing in practice, however, as they do not allow for checking of the model after the experiment is performed.

The bound also applies to most local optimality criteria and Bayesian criteria for linear models (see, for example Chernoff, 1972, p. 27 and Chaloner, 1984). In contrast for nonlinear models there is no such bound available on the number of support points. Although the criteria are concave on $\mathcal{H}$, the space of probability measures, they are not concave functions on a finite dimensional moment space and so Carathéodory's theorem cannot be invoked.

Chaloner and Larntz (1986, 1989) gave the first examples of how the number of support points in an optimal Bayesian design increases as the prior distribution becomes more dispersed. They found that for prior distributions that have support over a very small region the Bayesian optimal designs are almost the same as the locally optimal design and they have the same number of support points as the number of unknown parameters. For more dispersed prior distributions there are more support points. This is a useful feature for a design as, if there are more support points than unknown parameters, the model assumptions can be checked with data from the experiment. This is discussed further in Section 8.5.

Other examples of Bayesian nonlinear designs where the number of support points is not fixed can be found in Atkinson and Donev (1992), O'Brien and Rawlings (1994 a, b, c), Ridout (1994), Chaloner (1993) and Atkinson, Chaloner, Juritz and Herzberg (1993).

5.3 Exact Results

For local optimality there are several papers deriving closed form expressions for designs: for example White (1975), Kitsos, Titterington and Torsney (1988), Ford, Torsney and Wu (1992) and Wu (1988). For a particular value of the unknown parameters the problem often

reduces to an equivalent linear problem.

Finding optimal Bayesian designs algebraically is much harder and thus implementing Bayesian design criteria requires that designs be found by numerical optimization. Exceptions to this are simple special cases: these cases are not very useful in practice, but they give insight into properties of the optimal designs for more realistic and practical situations. Exact, algebraic results are quite difficult to derive as none of the tools from local optimality are very helpful.

In Chaloner (1993) for example, in a one parameter problem, with prior distributions with only two support points, it is possible to examine exactly how the transition from a one point optimal design to a two point optimal design occurs as the prior distribution is changed. Mukhopadhyay and Haines (1995), Dette and Neugebauer (1995a, b), Dette and Sperlich (1994b) and Haines (1995) all considered some nonlinear regression problems involving an exponential mean function, and gave conditions under which the optimal design is of a particular form. Loosely speaking these results can be generalized to say that if the prior distribution is not too dispersed and does not have heavy tails then an optimal Bayesian design has the same number of support points as there are unknown parameters. Haines (1995) gave an insightful geometric interpretation of this and demonstrated how, for a prior distribution with finite support, the problem reduces to a particular convex programming problem.

5.4 Design Software

It is clear that if Bayesian designs for nonlinear problems are to be used in practice then software must be readily available. Chaloner and Larntz (1988) describe such software for logistic regression. These are menu driven FORTRAN programs that are easy to use and compile and are available from the authors by email. A more powerful and flexible Bayesian design system is the object-oriented environment of Clyde (1993b), developed within XLISP-STAT (Tierney, 1990). This system enables both exact designs and approximate design measures to be easily found for both linear and nonlinear problems. Locally optimal designs and non-Bayesian linear designs can also be found as a special case of Bayesian designs. The system also allows for constraints in the optimization process as suggested in Clyde (1993a).

This powerful software environment is a little difficult to use initially but can be easily adapted to solve a multitude of design problems and makes Bayesian design a very practical reality. The availability of this software makes it straightforward to derive designs for a variety of prior distributions, model assumptions and criteria and so examine robustness. The software is available from the author (by email) and requires the NPSOL FORTRAN library of Gill et al (1986) to be loaded. Documentation, installation and availability by ftp is described in Clyde (1993b).

Warner (1993) describes some other software, which we have not examined, using the Gibbs sampler.

When software provides a continuous (approximate) design and an exact design is required then Pukelsheim and Rieder (1992) and Pukelsheim (1993, p. 424) can be consulted for procedures rounding continuous design measures to exact designs.

5.5 Sequential design

In any design problem an optimal sequential design procedure must be at least as good as a fixed design procedure. In most linear design problems, however, both Bayesian and non-Bayesian, the optimal sequential procedure is the fixed, non-sequential procedure. There is nothing to be gained by designing sequentially. This is easily seen when the error variance σ^2 is known: the posterior utility depends on the design η, but does not depend on the data $\mathbf{y}$. For the case when σ^2 is unknown it is not so clear. For A-optimality and σ^2 unknown with a conjugate prior distribution, the analysis of section 2.5 shows that there is nothing to be gained by sequential design in this case. For other linear problems it is unclear whether sequential design is better. For nonlinear problems the posterior utility clearly depends on the data $\mathbf{y}$, or a function of $\mathbf{y}$ such as $\hat{\theta}$, and there should be a gain from choosing design points sequentially.

Sequential design, however, may be unrealistic in practice. Consider for illustration the experiments of example 2 done in the University of Minnesota laboratory. Theoretically the dose for each one of the 60 animals could be decided upon one at a time and the extensive statistical literature on sequential design of binary response experiments consulted (see for

example Wu, 1985). But

1. as death over the 7 days following injection of the drug is the response, the experiment would be prolonged from a total of 7 days to many months.

2. time trends or seasonal effects may be introduced if the experimental conditions change over time. Similar animals might not always be available and the drugs deteriorate over time.

3. the probability of error in doses and calculations is increased when 60 calculations are done to determine the next dose. A non-sequential procedure is easily implemented and requires less training of laboratory staff.

Several powerful sequential Bayesian design procedures have been developed: see, for example Berry and Fristedt (1985) who reviewed the extensive work in bandit problems and Kuo (1983) who develops procedures for nonparametric binary regression. Freeman (1970) solved the Bayesian sequential design problem exactly for a very small and simple binary regression experiment. We do not attempt to review this work here.

Batch sequential procedures rather than fully sequential procedures (as in Zacks, 1977, and Ridout, 1995) might prove to be more practical. There is a practical concern, however, that the experimental conditions from one batch to the next might be different.

5.6 Discussion

Whatever criterion is used, Bayesian or non-Bayesian, prior information must be considered for nonlinear design as, unlike in a linear model, the posterior utility of a design depends on the data. An experimenter may be willing to specify an informative prior distribution in designing the experiment but may prefer to use a noninformative prior distribution for inference.

6 Specific nonlinear design problems

6.1 Binary response models

Tsutakawa (1972, 1980), Owen (1975), Zacks (1977), Chaloner and Larntz (1989), Flournoy (1993), and Clyde, Müller and Parmigiani (1994) all use Bayesian design ideas

in binary regression models. These models are important and have many applications in toxicology and reliability studies. They are also interesting from a design perspective because they are so very different from linear regression models.

Consider, for example, a simple linear regression model and a closed design interval $\mathcal{X}$. It is straightforward to show that the linear D-optimal design, under a vague prior distribution, is to take half the observations at one extreme of the interval and the other half at the other extreme. In contrast consider a binary regression with a binary response variable which is, "success" or "failure". Suppose that the probability of success near one extreme of the design interval $\mathcal{X}$ is close to 0 and at the other extreme it is close to 1. A design that puts all observations at the two extremes of $\mathcal{X}$ would be very inefficient. There would be a good chance that the experiment will yield no useful information: all the responses at the high value of x might be successes and all the responses at the other value of x might be failures. In this case the likelihood has no well defined mode and the experiment is not very informative.

It can be shown that design points of the Bayesian D-optimality criterion for a binary regression are spread throughout over the interval $\mathcal{X}$ and, as the support of the prior distribution gets wider, the number of support points of the optimal design increase. Good designs for binary regression problems have, therefore, quite different properties than good designs for linear regression problems.

6.2 Example 2 continued

Recall example 2 where the University of Minnesota laboratory performed many logistic regression experiments on several different drugs and biologic material.

For one particular drug under study, 54 similar experiments were performed. The drug was at one of several concentrations: 120, 121, 122 or 124 mg/ml and a similar design was used for each of the 54 experiments. The design was a design of 6 equally spaced doses of 2.5, 3.0, 3.5, 4.0, 4.5 and 5.0, with 10 mice exposed to each dose. 60 animals were used in each experiment. Sometimes less than 60 animals were available in which case less than 10 animals were exposed to the highest dose. The responses measured were the number of surviving mice, usually 7 days after being given the dose. Estimates of the LD50 were

calculated for each experiment and these estimates range from 3.2 to 4.2 and the slopes range from about -4.0 to -1.5. The LD50's were used to estimate the potency of each batch of drug.

The design does not correspond to any locally optimal design, as a locally optimal design has two dose levels only, with half of the animals at each dose.

Could Bayesian design ideas have been useful in this example? To examine this question the set of 54 estimates can be used to construct a prior distribution to design future experiments. The 54 estimates can be thought of a sample from a distribution of possible values that might be encountered in future experiments. A prior distribution was therefore constructed where the LD50 and the slope both have independent Beta distributions, Beta(4,4), and the LD50 lies between 3.2 and 4.2 and the slope between -4.0 and 1.5. This prior distribution reasonably reflects the sample and has, approximately, the same first two moments as the sample. It reflects the actual values obtained in the experiments performed. Before the experiments were performed a more realistic prior distribution might be one representing more uncertainty and so a second prior distribution was constructed which is uniform on the same interval.

For the independent Beta(4,4) distributions the Bayesian ϕ_2-optimal design for minimizing the posterior variance of the LD50 is easily found using the software of Chaloner and Larntz (1989) or Clyde (1993b). It is a 4 point design, symmetric around the prior mean for the LD50 of 3.6, and it takes observations at 3.07, 3.47, 3.73, 4.13 with weights 0.30, 0.20, 0.20 and 0.30 respectively. The design points are not equally spaced. Under this prior distribution, the design actually used in the lab, with an equal number of animals at each of 6 equally spaced doses between 2.5 and 5.0, has a ϕ_2-criterion value 1.52 times of that of the ϕ_2-optimal design. A 5 point design obtained by omitting the highest dose of 5.0 from the 6 point design and dividing the 60 animals equally between the remaining 5 doses has a criterion value of 1.29 times that of the optimal value. If the lowest dose in the 5 point design is also omitted and the animals equally divided between the remaining 4 doses of 3.0, 3.5, 4.0, and 4.5 the criterion value is 1.13 times that of the optimal value. Thus, if the Beta(4,4) distributions reflected the experimenters' beliefs well, if they were willing to use the optimal Bayesian design they could have reduced the variability of their estimates

considerably. If they wanted to use equally spaced doses at convenient values spaced 0.5 units apart and include integer values they could have got very close to an optimal Bayesian design using a 4 point design by omitting the two extreme design points of their design.

The Beta(4,4) prior distributions correspond to quite accurate knowledge of values to expect and so, for further illustration, consider the prior distribution that is uniform over the same interval. This prior distribution might have represented beliefs before the experiments were done. In this case the optimal design is a 5 point design, again centered at 3.6 units, taking observations at 2.78, 3.21, 3.6, 3.99, 4.42 with weights 0.28, 0.15, 0.14, 0.15 and 0.28 respectively. Although the points are almost equally spaced there is more mass at the extremes than at the center points. The equally spaced, equal weight, designs considered earlier are amazingly efficient for this prior distribution. The 6 point design used by the experimenters with equal weight at 2.5, 3.0, 3.5, 4.0, 4.5, 5.0 has a criterion value 1.13 times that of the optimal value, the 5 point design with equal weight at 2.5, 3.0, 3.5, 4.0, 4.5 has a criterion value 1.02 times that of the optimal one and the 4 point design with equal weight at 3.0, 3.5, 4.0, 4.5 has a criterion value 1.08 of the optimal value.

If the expectations of the experimenters could be reasonably represented by the uniform distribution then the design they used is close to the Bayesian optimal design. This example has, therefore, not illustrated that Bayesian design could have greatly improved efficiency of estimation in this laboratory, but rather illustrated that what they were doing may well have been close to being optimal in a Bayesian sense.

6.3 Nonlinear regression models

In a nonlinear regression model, the mean of a normally distributed response variable y is related to explanatory variables x by a nonlinear function $f(x, \theta)$. That is for $i = 1, \ldots, n$, we have $y_i = f(x_i, \theta) + e_i$. The errors e_i are independent and normally distributed with mean zero and variance σ^2. The expected Fisher information matrix for these models depends on the gradient vector $g(x, \theta)$ and is

$$n\mathcal{I}(\theta, \eta) = \sum_{i=1}^{n} g(x_i, \theta) g^T(x_i, \theta).$$

Design for nonlinear regression models has recently received considerable attention and

Bayesian criteria, such as ϕ_1-optimality and ϕ_2-optimality, have been influential. Pronzato, Huang, Walter, Le Roux and Frydman (1989), Huang, Walter and Pronzato (1991) and Atkinson, Chaloner, Herzberg and Juritz (1993) focused on compartmental models and found designs numerically. As did Chaloner and Larntz (1986, 1989), these authors all noted that as the prior distribution becomes more dispersed the number of support points typically increases. Chaloner (1993), Mukhopadhyay and Haines (1994), He, Studden and Sun (1995), Dette and Neugebauer (1995a), Dette and Neugebauer (1995b) and Dette and Sperlich (1994b) all examine simple special cases and prove optimality analytically.

The important paper by Haines (1995) is quite different and introduces some novel geometric interpretations of Bayesian optimal designs and also identifies several parallels between optimal Bayesian design and other areas. The paper by Dette and Sperlich (1994a) is also noteworthy as it uses an expansion of the Stieltjes transform of the design measure. The result provides a different perspective on the numerical optimization problem and gives valuable examples.

6.4 Example 3 continued

Example 3 is a case of design for nonlinear regression. The design problem is to choose times at which to take blood samples to measure the level of a drug. The experimenter used an 18 point design with the observations approximately equally spaced in the logarithm of time. The 18 point design takes one observation at times (in hours) 0.166, 0.333, .5, .666, 1, 1.5, 2, 2.5, 3, 4, 5, 6, 8, 10, 12, 24, 30, and 48. Atkinson, Chaloner, Herzberg and Juritz constructed Bayesian optimal designs under two prior distributions suggested by the data. They also constructed locally optimal designs. Under each prior distribution separate ϕ_2-optimal designs were constructed for estimating each of the three functions of interest (the area under the expected response curve or AUC, the time to maximum concentration or t_{max} and the maximum concentration c_{max}). One of the two prior distributions is such that θ_1 has a uniform distribution on $.05884 \pm .04$ and, independently, θ_2 has a uniform distribution on 4.298 ± 4.0: the parameter θ_3 is taken to be point mass at 21.80. For this prior distribution the 18 point design used by the experimenter is actually fairly efficient for estimating these three quantities.

Specifically denote η_{18} to be the 18 point design and η_t to be the ϕ_2-optimal design for estimating t_{max} under this prior distribution. The ϕ_2-optimal design η_t is a five point design with mass .29, .29, .15, .22, .06 at times .25, .94, 2.8, 8.8 and 24.7. The ratio $\phi_2(\eta_{18})/\phi_2(\eta_t)$ is only 1.3 which means that the 18 point design is fairly efficient for estimating t_{max}, with an expected posterior variance of t_{max} of only 1.3 times the best possible value. For estimating c_{max} the optimal design minimizes an appropriate ϕ_2-criterion and is denoted η_c. This is also a five point design with mass .10, .36, .32, .16 and .06 at times .37, 1.1, 2.4, 6.1 and 24.1. The ratio $\phi_2(\eta_{18})/\phi_2(\eta_c)$ is 1.4. Again the 18 point design is fairly efficient. For AUC the corresponding ϕ_2-optimal design is a 4 point design putting mass .01, .03, .26 and .70 at times .29, 1.7, 13.1 and 39.6 and the corresponding ratio of criteria is 3.2, and so the 18 point design is not as efficient for estimating the AUC as it is t_{max} and c_{max}. Atkinson, Chaloner, Herzberg and Juritz showed that under this prior distribution it is possible to improve on the 18 point design, but not by much. The AUC was found to be not well estimated under any design except one specifically designed to estimate it. Designs efficient for the AUC are very inefficient for estimating t_{max} and c_{max}. If there is very precise prior information, however, or if the area under the curve is of primary importance, the 18 point design can be improved upon considerably using Bayesian design.

So, interestingly, this is a similar situation as example 2, in that it may well be that what the experimenters were doing in practice was close to a Bayesian optimal design.

6.5 Sample size for clinical trials

Sample size calculations are especially important in the design and planning of clinical trials to compare two or more different treatments. The primary non-Bayesian approach is to specify the magnitude of the effect that the trial should be able to detect and choose the sample size to give a required power for a hypothesis test at that alternative. It is usually recommended to make allowances for patients who do not take their assigned medication (noncompliance) and patients who take a different medication than assigned (switchover). This is described for example in Lakatos (1988), Dupont and Plummer (1990) and Wu, Fisher and DeMets (1980). Several computer programs are available to implement variations on these methods.

A partially Bayesian approach to this problem is given in Spiegelhalter and Freedman (1986) who used a prior distribution, as an approximation to a predictive distribution, to average over the power. Berry (1991) described a Bayesian approach for a very simple situation using dynamic programming and sequential updating. Achcar (1984) looked at Bayesian calculations of sample size when sampling from a single Weibull distribution. A completely Bayesian approach is advocated in Brooks (1987) who considered the expected gain in information from a two group experiment with Weibull lifetimes. He dealt with the sample size, the proportion of observations in each group, the length of time to accrue patients and how long to follow them. He obtained some closed form expressions for the gain in Shannon information under normal prior distributions for the unknown parameters and also made some approximations. Some of these calculations are similar to Brooks (1982) where he discussed the information lost, for exponential lifetimes, when censoring is present. Sylvester (1988) examines the sample size for a Phase II clinical trial using Bayesian decision theory when the responses are Bernoulli.

These Bayesian approaches appear not to have been used much in practice. Perhaps this is because they fail to account explicitly for noncompliance and switchover. Or, perhaps this is because there are no freely available computer programs to make these methods accessible. For the non-Bayesian solutions, Shih (1995) described a SAS macro computer program that implements the method of Lakatos (1988).

6.6 Other sample size problems

Deciding on the sample size n in an experiment is always part of design. DasGupta and Mukhopadhyay (1994) take a Bayesian approach choice of sample size for a sample from a single normal distribution with a conjugate normal prior distribution. They define criteria which make the sample size robust to the future data. DasGupta and Vidakovic (1994) take a Bayesian approach to sample size choice for hypothesis testing in a one way analysis of variance model of example 1 where hypothesis testing is the purpose of the experiment. They also give Mathematica code for their method.

There are opportunities for further research in this area for more general, non-normal models. See DasGupta (1995).

6.7 Design problems in reliability and quality control

DeGroot and Goel (1979) considered a Bayesian approach to designing studies of exponential lifetimes where experimental units may, or may not, be subject to an increased stress and where units may be subject to a high stress, if they do not fail in a specified period of time under a low stress. DeGroot and Goel call this "tampering". They derived exact Bayesian optimal designs under particular loss functions and costs. DeGroot and Goel (1988) is a review of this work and appeared in a volume, Clarotti and Lindley (1988), devoted to Bayesian analysis and design in reliability. This volume also contains other relevant papers: for example the chapter by Barlow, Mensing and Smiriga (1988) discusses influence diagrams and their use in optimal design.

Chaloner and Larntz (1992) took the approach described in Section 4.2 to derive Bayesian optimal designs for accelerated life testing where the lifetimes have either Weibull or lognormal distributions and the length of time available for the experiment is fixed. Their methods are extended in Naylor (1994). Verdinelli, Polson and Singpurwalla (1993) discussed Bayesian design for accelerated life testing experiments where prediction is the goal. They used Shannon information in (9) as utility and considered the case where the lifetimes have a lognormal distribution. Mitchell and Scott (1987) also designed to maximize Shannon information in a group testing experiment: they provided free software for their method.

Verdinelli and Wynn (1988) examined some aspects of keeping an expected response on target which is an important problem in the Taguchi approach to design. They proposed, as a Bayesian alternative to non-Bayesian methods, to set the predictive mean at the target value and to minimize the predictive variance.

6.8 Large computer experiments

Some exciting recent developments have occurred in applying ideas from optimal design to the problem of choosing the values at which to run a large deterministic computer model. The situation can be thought of as having a response surface which is known to be smooth but its general form is unknown and the values of the response surface can be determined without error. Sacks, Welch, Mitchell and Wynn (1989) review this work. Recent advances are described in Welch et al (1992), Morris, Mitchell and Ylvisaker (1993) and Bates, Buck,

Riccomagno and Wynn (1995). Much of this work involves sequential design but non-sequential design has also been found helpful as in Currin, Mitchell, Morris and Ylvisaker (1991). A Bayesian formulation of the problem has proved fruitful. Rather than attempt to review this work here the reader is referred to the above references. This important problem has unique aspects.

6.8 Other nonlinear design problems

Ridout (1994) applied Bayesian design ideas to a seed testing experiment similar to the dilution assay problem. Parmigiani (1993) and Parmigiani and Kamlet (1993) used Bayesian decision theory to study the design problem of when to screen for disease and applied this to breast and cervical cancer screening. They presented a powerful case for the use of Bayesian methods in these types of designs and decision making.

Parmigiani and Berry (1994) examined several problems using the exact expected utility, as calculated by (1), for clinical design problems. They mainly consider exponential or binomial responses with conjugate prior distributions. Lad and Deely (1995) also do exact calculations for a simple decision problem and elicit prior probabilities and utilities directly.

Apart from Draper and Hunter (1966, 1967b) little research has been done in using Bayesian design for multivariate response models, either linear or nonlinear. Draper and Hunter developed and used a criterion similar to (16) where a prior precision is incorporated into the criterion. in their examples they either used a prior estimate for the nonlinear parameters, similar to local optimality, or they used sequential design. It is a potential area of research to use criteria which more closely approximate expected utility in the multivariate response case.

7 Nonlinear estimation within a linear model

7.1 General problem

When a nonlinear function of the regression coefficients in a linear model is of primary interest then the expected utility cannot be calculated exactly and the problem has more in common with nonlinear design than with linear design. Asymptotic approximations similar to those in section 4.2 can be used to give design criteria.

Assume that the model is as in section 2 and that a nonlinear function of the parameters

$g(\theta)$ is of interest. Define the k vector $\mathbf{c}(\theta)$ to be the gradient vector of $g(\theta)$ as in (17). Approximations similar to those in section 4.2 give a squared error loss of either

$$\sigma^2 c(\hat{\theta})^T (nM)^{-1} c(\hat{\theta}) \qquad \text{or,} \qquad \sigma^2 c(\hat{\theta})^T (nM + R)^{-1} c(\hat{\theta})$$

where R is either the prior precision matrix or the matrix of second derivatives of the prior distribution. As in section 4.2, the criteria

$$\phi_2 = \int \sigma^2 \mathbf{c}(\theta)^T (nM)^{-1} \mathbf{c}(\theta)\ p(\theta, \sigma) d\theta d\sigma$$

and

$$\phi_{2R} = \int \sigma^2 \mathbf{c}(\theta)^T (R + nM)^{-1} \mathbf{c}(\theta)\ p(\theta, \sigma) d\theta d\sigma$$

can be expressed as a form of A-optimality. That is the design, η, should be chosen to minimize either $trAM^{-1}$ or $trA(R + M)^{-1}$ with $A = E[\sigma^2 \mathbf{c}(\theta)\mathbf{c}(\theta)^T]$, the expectation being over the prior distribution of θ. If more than one nonlinear function of θ is of interest, say $g_i(\theta)$ for $i = 1, \ldots, m$, then the matrix A is the sum, or possibly the weighted sum, of individual matrices $E[\sigma^2 \mathbf{c}_i(\theta)\mathbf{c}_i(\theta)^T]$. Note however that, unlike the case for the usual linear A-optimality, it should be possible to get a better design by choosing the design points sequentially.

One such design problem is that of estimating the turning point in a quadratic regression. This problem is discussed in Mandal (1978), Buonaccorsi and Iyer (1984, 1985), Buonaccorsi (1985), and Chaloner (1989). Buonaccorsi and Iyer (1986) also examined several other problems involving design for the ratio of the coefficients in linear model. A special case of estimating such a ratio is the calibration problem where n independent observations y_i are taken from a simple linear regression model. That is

$$y_i = \theta_0 + \theta_1 x_i + e_i$$

where $e_i, i = 1, \ldots, n$ are normally distributed with mean zero and variance σ^2. There are n observations $\mathbf{y}$ and an $(n + 1)$st observation y_{n+1} for which it is required to estimate

the corresponding value of x_{n+1}. One solution is to estimate the nonlinear function $g(\theta) = (y_{n+1} - \theta_0)/\theta_1$. Buonaccorsi and Iyer (1986) discussed design for this problem using both local optimality and Bayesian A-optimality. A different but related Bayesian approach was taken in Barlow, Mensing and Smiriga (1991) who put a prior distribution on x_{n+1}.

7.2 Turning point example

As in other nonlinear design problems the usual non-Bayesian approach to these problems is to use the local optimality approach of Chernoff (1953). The problem of estimating the turning point in a quadratic regression will be used to illustrate an important limitation of local optimality. This example is used to illustrate non-Bayesian nonlinear design in Ford and Silvey (1980) and Ford, Titterington and Wu (1985) and Bayesian nonlinear design in Chaloner (1989).

Suppose that the expectation of the response y at x is $\theta_0 + \theta_1 x + \theta_2 x^2$. Then the turning point is $g(\theta) = -\theta_1/(2\theta_2)$. Define $\mathbf{c}(\theta)$ to be the gradient vector $(0, 1/(2\theta_2), \theta_1/(2\theta_2^2))^T$. The asymptotic variance of the maximum likelihood estimator of $g(\theta)$ is then:

$$\sigma^2 c(\theta)^T M^{-1} c(\theta) \tag{27}$$

with $nM = n\sum_{i=1}^{k} \eta_i \mathbf{x}_i \mathbf{x}_i^T$ defined in Section 1. Local optimality requires a best guess for θ, θ_0 say. The value of θ_0 is substituted into (27) and the design η is chosen to minimize (27). Suppose now that observations x_i can be taken anywhere in the interval $[-1, 1]$ and that $g(\theta_0) = \frac{1}{2}$ is the best guess value to be used for local optimality. It is straightforward to show that the locally optimal design takes half the observations at $x = 1$ and half at $x = 0$, giving nM as a singular 3×3 matrix of rank 2. The two design points $x = 1$ and $x = 0$ are two points where the expected value of y is equal and the turning point $x = g(\theta)$ is half way between these two points. If this experiment were to be carried out, using no prior information in the estimation process, then it is clearly impossible to fit a quadratic regression to two data points and estimate the turning point. The locally optimal design is therefore useless for practical purposes.

The above illustrates the general point that the locally optimal design for cases of nonlinear estimation within a linear model, can lead to a matrix nM which minimizes (27) but is

singular and in this case the quantity of interest $g(\theta)$ may not be estimable. This is different from the linear c-optimality case where, although the optimal design may give a singular matrix nM, the contrast of interest, $\mathbf{c}^T\theta$, is always estimable.

8 Other design problems

8.1 Variance components models

Designs for the estimation of variances are important in quality control research. Anderson (1975), for example, reviewed this topic. More recently Mukerjee and Huda (1988) examined optimality and Giovagnoli and Sebastiani (1989) considered the design problem when both the variance components and the fixed effects are of interest. The approach has always been to use local optimality until the recent paper of Lohr (1995), who looked at a Bayesian approach to design. She used Bayesian D-optimality and A-optimality for the estimation of the variance components or their sum or their ratio. She gave conditions under which a balanced design is optimal and showed the optimality of a balanced design under a large class of prior distributions. In the context of hierarchical models with unknown variance components for multi-center clinical trials Stangl and Mukhopadhyay (1993) also used Bayesian methods for design.

8.2 Mixtures of linear models

Läuter (1974, 1976) proposed a design criterion that is an average of design criteria, the average being over a number m of models. She used a criterion $\phi(\eta) = \sum_{i=1}^{m} w_i\phi_i(\eta)$, where, for example, $\phi_i(\eta)$ is the D-optimality criterion under the ith of m candidate models. The weight, w_i on the ith model is the prior probability on that model. Cook and Nachtsheim (1982) applied such a criterion to design for polynomial regression when the degree of the polynomial is unknown. The criterion, $\phi(\eta)$, they used was based on A-optimality for predicting the response over the design interval. For the ith model and a design η the variance of the predicted mean response over the design interval is proportional to trA_iM^{-1} where A_i is a specified matrix. This criterion is sometimes referred to as Q- or L-optimality. But rather than average the A-optimality criteria directly Cook and Nachtsheim averaged efficiency criteria. Specifically let η_i be the A-optimal design for the ith model, $i = 1, \ldots, m$, for minimizing the variance of prediction over the design region, and let M_i be the information

matrix for the ith model. Then they maximized

$$\phi(\eta) = -\sum_{i=1}^{m} w_i \frac{\mathrm{tr} A_i M_i(\eta)^{-1}}{\mathrm{tr} A_i M_i(\eta_i)^{-1}}.$$

They gave a number of numerical examples using this criterion to predict the uranium content of a log.

These ideas are similar to those of Bayesian non-linear design although the motivation of Cook and Nachtsheim is not Bayesian. This is apparent through their use of average efficiency: it is unclear how this corresponds to maximizing expected utility. A Bayesian approach, with squared error loss, would argue for an averaging of the A-optimality criteria directly rather than their efficiencies. In other words a Bayesian approach would use the criterion

$$\phi(\eta) = -\sum_{i=1}^{m} w_i tr A_i M_i(\eta)^{-1}.$$

Similarly in using D-optimality averaged over a collection of models, it is unclear, unless utility is considered, whether to average $\phi_i(\eta) = \log\det(M_i)$, or $\phi_i(\eta) = \det(M_i)$, or $\phi_i(\eta) = \sqrt{\det}(M_i)$, or perhaps an efficiency measure, like that of Cook and Nachtsheim, such as $\phi(\eta) = \det[M_i(\eta)]/\det[M_i(\eta_i)]$, where η_i is the D-optimal design for the ith model. From a Bayesian perspective of maximizing expected utility, however, the answer is clear: expected utility should be maximized, not expected efficiency.

An excellent summary of the mathematics of such criteria and how the general equivalence theorem can be applied is in Pukelsheim (1993, p. 286-296). Dette (1990) gave some general results for D-optimality and polynomial regression. Dette (1991, 1993a, 1993b) used mixtures of Bayesian linear model criteria involving the prior precision matrix. He also derived a version of Elfving's (1952) theorem for this case. Dette and Studden (1994) provided further results characterizing the optimal design in terms of its canonical moments. Haines (1995) gave further geometric insight into such criteria.

8.3 Design for model discrimination

In an experiment where several models are compared, in order to select one of them, a number of non-Bayesian approaches to design have been suggested. Usually a method

for discriminating between models and a method for estimating the parameters within each model are combined. These procedures are reviewed in Pukelsheim and Rosenberger (1993) who also provide valuable insight into the mixture criteria of the previous section and suggest a number of ways of designing for a number of simultaneous objectives. See also Ponce De Leon and Atkinson (1991).

Spezzaferri (1988) presented a Bayesian approach to design for choosing between two linear models and to design with the dual goal of model selection and parameter estimation. He used the utility function in (5) of section 2.2 for both problems. For discriminating between two models, the design criterion he derived leads to minimizing the expectation of the posterior probability of one model, when the other is assumed to be true. In the case of multivariate normal nested models, when using diffuse prior information, this criterion is the same as non-Bayesian D-optimality for testing the hypothesis $\theta_0 = 0$, where θ_0 is the subvector of extra parameters in the larger model (see, for example, Atkinson 1972).

For the dual purpose of model discrimination and parameter estimation for two nested normal linear models, Spezzaferri showed that the optimality criterion using utility (5) is given by the product of two factors. One is the determinant of the information matrix of the smaller model. The other factor is the expectation of the posterior probability of the smaller model, when it is assumed to be true. The optimal design for discrimination and estimation maximizes the product of these factors.

8.4 Robustness

It is important to check the sensitivity of the design to prior distribution. DasGupta and Studden (1991) constructed a framework for robust Bayesian experimental design for linear models. They found designs that maximize expected utility for a fixed prior distribution subject to being robust for a class of prior distributions. DasGupta, Mukhopadhyay and Studden (1992) gave a detailed approach to design in a linear model when the variance of the response is proportional to an exponential or power function of the mean response. They developed examples of "compromise designs" where the experimenter wants to find a design that is highly efficient for several design problems. They considered both Bayesian and non-Bayesian formulations of the design criteria.

Seo and Larntz (1992) suggested some criteria for nonlinear design that make the design robust to specification of the prior distribution. They used the design problem of estimating the turning point in a quadratic regression as their motivating example. They suggested a criterion of designing for a "major" prior distribution subject to a constraint of attaining a certain efficiency over a class of closely related prior distributions.

Toman (1992a, b) and Toman and Gastwirth (1993, 1994) also considered robustness of Bayesian design in the normal linear model with respect to the prior distribution. These papers dealt mainly with the one way analysis of variance model. To allow for possible misspecification of prior variances, Toman (1992a, b) proposed using a class of normal prior distributions where the variances take values in specified intervals. The criteria she suggested for choosing designs are maximizing the average, over the class of posterior distributions, of either the determinant or the trace of the posterior precision matrix. Averages are taken with respect to a distribution on the prior precision parameters.

Toman and Gastwirth (1993) examined both robust estimation and robust design for analysis of variance models when the prior distribution is in a class of finite mixtures of normals. They used a squared loss function and considered an average of the posterior risk over the class of corresponding posterior distributions.

Toman and Gastwirth (1994) suggested specifying the prior distribution on treatment means using results from a pilot study. They assumed that the error variances of the pilot and of the follow up studies to be unknown, but that the intervals in which they vary can be specified. They adopted a squared loss function and proposed to use, for the design and the estimator, a minimax criterion, over the class of posterior distributions.

8.5 Model unknown

A major criticism of traditional optimal design for linear models is that the number of support points in an optimal design is often the same as the number of parameters – in which case no model checking can be done. In addition, under the assumption that the model is known, the design points are usually at the boundary of the design region – but if the linear response surface is, as is quite usual, a linear approximation to some smooth but unknown surface, then it is at the boundary of this region that the approximation is most inaccurate.

These criticisms are not new (see for example Box and Draper, 1959, Sacks and Ylvisaker, 1984, 1985) and apply to both Bayesian and non-Bayesian optimal design for linear models.

As discussed in Section 5.2 these criticisms *sometimes* do not apply to Bayesian optimal designs for nonlinear problems. In these cases there is no bound on the number of support points in an optimal design and the support points may be spread throughout the experimental region. It is unclear, however, under what circumstances this is so.

Among attempts at incorporating model uncertainty into the design problem is the mixture approach as described in section 8.2. More recent work by DuMouchel and Jones (1994) introduced a modified Bayesian D-optimal approach for the special case of factorial models. They constructed a prior distribution with a structure recognizing "primary" and "potential" terms. The resulting Bayesian D-optimal designs have very desirable properties. Indeed they provided a Bayesian justification for resolution IV designs. DuMouchel and Jones showed several compelling examples of the use of their methods. This work recognizes model uncertainty, which is almost always present in a practical setting. It specifically accounts for the belief, that has long been held by practitioners, that when certain interactions or effects are assumed to be zero to derive a fractional design the experimenter does not believe that such effects are exactly zero but rather that they are small compared to other effects. DuMouchel and Jones have succeeded in formalizing the otherwise heuristic justification for resolution IV designs over other designs which have the same value of the D-optimality criterion.

Steinberg (1985) considered two-level factorial experiments to represent a response surface problem and also used a Bayesian formulation to introduce uncertainty about the adequacy of the proposed model. He derived a method for choosing the scale of the two factor experiment: that is he chose the "high" and the "low" levels for each factor conditional on a particular fractional factorial design being used. In this way the trade off is recognized between choosing design points on the boundary of the design regions to maximize information and choosing them towards the center of the region where the model is believed to hold to better approximation. Steinberg's approach is reminiscent of earlier work by O'Hagan (1978). O'Hagan considered a Bayesian approach to design for curve fitting where the curve to be fit is a smooth function and design points are chosen based on the predictive distribution.

These approaches all use Bayesian ideas to solve the very practical aspect of real design problems. In real problems the model is almost never known exactly. There is clearly a need for further research here.

9 Concluding remarks

Bayesian design is an exciting and fast developing area of research. The Bayesian methodology has much to offer in experimental design, where prior information has always been used for the choice of experiment, explanatory factors, sample size, and model. A Bayesian approach to design gives a mechanism for formally incorporating such information into the design process. The decision theoretic formulation presented in this paper shows that utility functions can clarify the approach to design.

The examples presented, especially examples 2 and 3 of nonlinear problems, illustrate that some experimenters may already be actually using designs which can be justified as approximately optimal under a Bayesian formulation. A formal Bayesian approach to experimental design may well lead to substantial improvements. It does remain regrettable, however, that so few real case studies appear in the statistical literature of Bayesian optimal design. The same can be said of non-Bayesian nonlinear design where there is considerable theoretical research but few real case studies.

There are many specific design problems that remain to be investigated by a Bayesian approach. In particular, within the linear model context, there is a need for methods incorporating hierarchical linear models and hierarchical prior distributions and unknown variance components. The simple examples presented in section 3 illustrate that more sensible designs can be obtained when the prior distribution is specified within the hierarchical linear model. But, as remarked by Goldstein (1992), there is also the need for these ideas to be applied to actual experiments.

In both linear and nonlinear problems there is the need for methods which reflect the reality that the model for analysis is almost never known with certainty before the experiment is done. The experimental design process should incorporate model uncertainty into the design process.

There is also a parallel need for methods to be developed for the specification and quantifi-

cation of prior beliefs. Prior beliefs maybe entirely subjective, based on personal experience, or may be based on previous experiments and past data. Whatever the source of prior information very little guidance is available on how to collect and quantify such information. A notable exception to this is the important work of Garthwaite and Dickey, for example Garthwaite and Dickey (1988), who have developed useful methods for elicitation for the linear model. It remains a challenge to develop methods for prior elicitation for distributions to be used in design for nonlinear models. A welcome beginning is the study of Flournoy (1993) who gives a nice example of the entire design process, including expert elicitation.

Bayesian design also requires a specification of a utility function. It is clearly helpful, in the design process, to carefully consider the reason the experiment is being done and to consider what utility should be used. Although Shannon information and squared error have been widely used in the statistical literature, it would also be interesting to see alternatives constructed and explored in future research.

As most Bayesian methods for design require numerical optimization and integration there is a need for software to find such designs, both exact and continuous. Without available and user friendly software these methods will not be used in real problems. The software of Clyde (1993b) has the potential to make Bayesian designs accessible to the scientist.

Acknowledgements

We are grateful to Rob Kass, the referees and Larry Wasserman for carefully reading the manuscript and for providing many helpful suggestions.

REFERENCES

ACHCAR, J.A. (1984). Use of Bayesian analysis to design of clinical trials with one treatment. *Comm. Statist., Theory Methods* **13(14)** 1693-1707.

ANDERSON, R.L. (1975). Designs and estimators for variance components. In *A Survey of Statistical Designs and Linear Models*, J.N. Srivastava ed. 1-29. North-Holland, Amsterdam.

ATKINSON, A.C. (1972). Planning Experiments to Detect Inadequate Regression Models. *Biometrika* **59** 275-293.

ATKINSON, A.C., CHALONER, K., JURITZ, J. and HERZBERG, A.M. (1993). Optimum experimental designs for properties of a compartmental model. *Biometrics* **49** 325-337.

ATKINSON, A.C. and DONEV, A.N. (1992). *Optimum Experimental Designs.* Clarendon Press, Oxford.

BALL, F.G., SMITH, A.F.M. and VERDINELLI, I. (1993). Biased Coin Designs with a Bayesian Bias. *Journal of Statistical Planning and Inference* **34** 403-421.

BANDEMER, NÄTHER and PILZ (1987). Once more: optimal experimental designs for regression models. (with discussion) *Statistics* **18** 171-217.

BARLOW, R.E., MENSING, R.W. and SMIRIGA, N.G. (1988). An application of influence diagrams. *Accel. Life Testing and Experts' Opinions in Reliability*, C. A. Clarotti, D. V. Lindley eds. 158-172. Elsevier, North-Holland.

BARLOW, R.E., MENSING, R.W. and SMIRIGA, N.G. (1991). Computing the optimal design for a calibration experiment. *Journal of Statistical Planning and Inference* **29** 5-19.

BATES, R.A., BUCK, R.J., RICCOMAGNO, E. and WYNN, H.P. (1995). Experimental design and observation for large systems. *J. Roy. Statist. Soc. Ser. B* **57** to appear.

BERGER, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag, New York.

BERNARDO, J.M. (1979). Expected Information as Expected Utility. *Ann. Statist.* **7** 686-690.

BERRY, D.A. (1991). Experimental design for drug development: A Bayesian Approach. *J. of Biopharmaceutical Statistics* **1** 81-101.

BERRY, D.A. and FRISTEDT, B. (1985). *Bandit problems: sequential allocation of experiments.* Chapman and Hall, London.

BOX, G.E.P. (1982). Choice of Response Surface Design and Alphabetic Optimality. *Utilitas Mathematica* B **21** 11-55.

BOX, G.E.P. and DRAPER, N.R. (1959). A basis for the Selection of a Response Surface Design. *J. Amer. Statist. Assoc.* **54** 622-654

BOX, G.E.P. and LUCAS, H.L. (1959). Design of experiments in non-linear situations. *Biometrika* **46** 77-90.

BOX, G.E.P. and TIAO, G.C. (1973). *Bayesian Inference in Statistical Analysis.* Addison Wesley.

BROOKS, R.J. (1972). A decision theory approach to optimal regression designs. *Biometrika* **59** 563-571.

BROOKS, R.J. (1974). On the choice of an experiment for prediction in linear regression. *Biometrika* **61** 303-311.

BROOKS, R.J. (1976). Optimal regression designs for prediction when prior knowledge is available. *Metrika* **23** 217-221.

BROOKS, R.J. (1977). Optimal regression design for control in linear regression. *Biometrika* **64** 319-325.

BROOKS, R.J. (1982). On the loss of information through censoring. *Biometrika* **69** 137-144.

BROOKS, R.J. (1987). On the design of comparative lifetime studies. *Comm. Statist., Theory Methods* **16** 1221-1240.

BUEHLER, R.J. (1971). Measuring information and uncertainty. *Foundation of Statistical Inference.* V.P. Godambe, and D.A. Sprott eds. 330-341 Holt, Rinehart and Winston, Canada.

BUONACCORSI, J.P. (1985). Ratios of linear combinations in the general linear model. *Comm. Statist., Theory Methods* **14(3)** 635-650.

BUONACCORSI, J.P. and IYER, H.K. (1984). A comparison of confidence regions and designs in estimation of a ratio. *Comm. Statist., Simul. Comput.* **13(6)** 723-741.

BUONACCORSI, J.P. and IYER, H.K. (1986). Optimal designs for the ratios of linear combinations in the general linear model. *Journal of Statistical Planning and Inference* **13** 345-356.

CHALONER, K. (1982). Optimal Bayesian experimental design for linear models. PhD Dissertation. Dept. Statistics, Carnegie Mellon University.

CHALONER, K. (1984). Optimal Bayesian Experimental Designs for Linear Models. *Ann. Statist.* **12** 283-300.

CHALONER, K. (1987). An Approach to design for generalized linear models. In *Proceedings of the workshop on Model-oriented Data Analysis, Wartburg* – Lecture Notes in Economics and Mathematical Systems, #297, 3-12. Springer-Verlag, Berlin.

CHALONER K. (1989). Bayesian design for estimating the turning point of a quadratic regression. *Comm. Statist., Theory Methods*, **18(4)** 1385-1400.

CHALONER, K. (1993). A note on optimal Bayesian design for nonlinear problems. *Journal of Statistical Planning and Inference* **37** 229-235.

CHALONER, K. and LARNTZ, K. (1986). Optimal Bayesian Designs applied to Logistic Regression Experiments. *University of Minnesota Technical Report.*

CHALONER, K. and LARNTZ, K. (1988). Sofware for logistic regression experiment design. In *Optimal Design and Analysis of Experiments*, Y. Dodge, V. Fedorov and H. Wynn eds. North-Holland, Amsterdam.

CHALONER, K. and LARNTZ, K. (1989). Optimal Bayesian Designs applied to Logistic Regression Experiments. *Journal of Statistical Planning and Inference* **21** 191-208.

CHALONER, K. and LARNTZ, K. (1992). Bayesian Design for accelerated life testing. *Journal of Statistical Planning and Inference* **33** 245-259.

CHERNOFF, H. (1953). Locally optimal designs for estimating parameters. *Ann. Statist.* **24** 586-602.

CHERNOFF, H. (1962). Optimal accelerated life designs for estimation. *Technometrics* **4** 381-408.

CHERNOFF, H. (1972). *Sequential analysis and optimal design.* Society for Industrial and Applied Mathematics, Philadelphia PA.

CLAROTTI, C. A. and LINDLEY, D. V. eds. (1988). *Accelerated life testing and experts' opinions in reliability.* North-Holland, Amsterdam.

CLYDE, M.A. (1993a). *Bayesian optimal designs for approximate normality.* Ph. D. Thesis, University of Minnesota.

CLYDE, M.A. (1993b). An object-oriented system for Bayesian nonlinear design using xlisp-stat. University of Minnesota, School of Statistics, Technical Report 587.

CLYDE, M., MÜLLER, P. and PARMIGIANI, G. (1994). Optimal design for heart defibrillators. To appear in: *Case Studies in Bayesian Statistics II.* C. Gatsonis, J. Hodges, R. E. Kass, N. Singpurwalla eds. Springer-Verlag, New York.

COOK, R.D. and NACHTSHEIM, C.J. (1982). Model robust, linear-optimal designs. *Technometrics* **24** 49-54.

COVEY-CRUMP, P.A.K. and SILVEY, S.D. (1970). Optimal regression designs with previous observations. *Biometrika* **57** 551-556.

CURRIN, C., MITCHELL, T.J., MORRIS, M.D. and YLVISAKER, D. (1991). Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *J. Amer. Statist. Association* **86** 953-963.

D'ARGENIO, D.Z. (1990). Incorporating prior parameter uncertainty in the design of sampling schedules for pharmacokinetic parameter estimation experiments. *Mathematical Biosciences* **99** 105-118.

D'ARGENIO, D.Z. and VAN GUILDER, M. (1988). Design of experiments for parameter estimation involving uncertain systems, with applications to Pharmacokinetics. In *Proc. 12th IMACS World Congress on Scientific Computation, 1988, Paris*, 511-513.

DASGUPTA, A. (1995). Review of optimal Bayes designs. Technical report, Purdue University.

DASGUPTA, A., MUKHOPADHYAY, S. and STUDDEN, W.J. (1992). Compromise designs in heteroscedastic linear models. *Journal of Statistical Planning and Inference* **32** 363-384.

DASGUPTA, A. and MUKHOPADHYAY, S. (1994). Uniform and subuniform posterior robustness: the sample size problem. *Journal of Statistical Planning and Inference* **40** 189-204.

DasGupta, A. and Studden, W.J. (1991). Robust Bayes Designs in Normal Linear Models. *Ann. Statist.* 19 1244-1256.

DasGupta, A. and Vidakovic, B. (1994). Sample sizes in ANOVA: The Bayesian point of view. Technical report, Purdue University.

de Finetti, B. (1962). Does it make sense to speak of "Good Probabiity Appraisers?" *The Scientist Speculates.* I. J. Good ed. 357-364. Basic Books, New York.

DeGroot, M.H. (1962). Uncertainty, Information and Sequential Experiments. *Ann. Statist.* **33** 404-419.

DeGroot, M.H. (1970). *Optimal Statistical Decisions.* McGraw-Hill: New York.

DeGroot, M.H. (1986). Concepts of Information based on Utility. In *Recent Developments in the Foundations of Utility and Risk Theory.* L. Daboni, A. Montesano, and M. Lines eds. 265-275. Reidel, Dordrecht, Holland.

DeGroot, M.H. (1970). *Optimal Statistical Decisions.* McGraw-Hill: New York.

DeGroot, M.H. and Goel, P.K. (1979). Bayesian estimation and optimal design in partially accelerated life testing. *Naval Res. Log. Quart.* **26** 223-235.

DeGroot, M.H. and Goel, P.K. (1988). Bayesian design and analysis of accelerated life testing with step stress. *Accelerated Life Testing and Experts' Opinions in Reliability.* C. A. Clarotti, D. V. Lindley eds. 193-202. Elsevier, North Holland.

Dette, H. (1990). A generalization of D$-$ and D_1-optimal designs in polynomial regression. *Ann. Statist.* **18** 1784-1804.

Dette, H. (1991). A note on robust designs for polynomial regression. *Journal of Statistical Planning and Inference* **28** 223-232.

Dette, H. (1993a). Elfving's theorem for D-optimality. *Ann. Statist.* **21** 753-766.

Dette, H. (1993b). A note on Bayesian c- and D-optimal designs in nonlinear regression models. *Manuscript.*

Dette, H. and Neugebauer, H.M. (1995a). On the existence of Bayesian optimal one point designs for one parametric nonlinear regression models. To appear in *Journal of Statistical Planning and Inference.*

Dette, H. and Neugebauer, H.M. (1995b). Bayesian D-optimal designs for exponential regression models. To appear in *Journal of Statistical Planning and Inference.*

Dette, H. and Sperlich, S. (1994a). Some applications of continued fractions in the construction of optimal designs for nonlinear regression models. *Manuscript.*

Dette, H. and Sperlich, S. (1994b). A note on Bayesian D-optimal designs for general exponential growth models. *Manuscript.*

Dette, H. and Studden, W.J. (1994). Optimal designs for polynomial regression when the degree is not known. *Manuscript.*

Draper, N.R. and Hunter, W.G. (1966). Design of experiments for parameter estimation in multiresponse situations. *Biometrika* **53** 525-533.

DRAPER, N.R. and HUNTER, W.G. (1967a). The use of prior distributions in the design of experiments in non-linear situations. *Biometrika* **54** 147-153.

DRAPER, N.R. and HUNTER, W.G. (1967b). The use of prior distributions in the design of experiments for parameter estimation in non-linear situations: multiresponse case. *Biometrika* **54** 662-665.

DUBOV, E.L. (1977). D-optimal designs for non-linear models under the Bayesian approach. In: *Regression Experiments*, Moscow University Press, Moscow (in Russian).

DUMOUCHEL, W. and JONES, B. (1994). A Simple Bayesian Modification of D-optimal Designs to Reduce Dependence on an Assumed Model. *Technometrics* **36** 37-47.

DUNCAN, G. and DEGROOT, M.H. (1976). A mean squared error approach to Optimal Design Theory. *Proceedings of the 1976 Conference on Information: Science and systems*. 217-221. The Johns Hopkins University.

DUPONT, W.D. and PLUMMER, W.D. JR. (1990). Power and sample size calculations. *Controlled Clinical Trials* **11** 116-128.

DYKSTRA, O. JR. (1971). The augmentation of experimental data to maximize $|X^TX|$. *Technometrics* **13** 682-688.

EATON, M.L., GIOVAGNOLI, A. and SEBASTIANI, P. (1994). A predictive approach to the Bayesian design problem with application to normal regression models. *Technical Report 598*, School of Statistics, University of Minnesota.

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

EL-KRUNZ, S.M. and STUDDEN, W.J. (1991). Bayesian optimal designs for linear regression models. *Ann. Statist* **19** 2183-2208.

ETZIONE, R. and KADANE, J.B. (1993). Optimal experimental design for another's analysis. *J. Amer. Statist. Association* **88** 1404-1411.

EVANS, J.W. (1979). Computer augmentation of experimental designs to maximize $|X^TX|$. *Technometrics* **21** 321-330.

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

FEDOROV, V.V. (1980). Convex design theory. *Math. Operationsforsch. und Statist., Ser. Statistics* **11** 403-413.

FEDOROV, V.V. (1981). Active regression experiments. in *Mathematical Methods of Experimental Design*, Nauka Press, Siberian Section 19-73. (In Russian).

FLOURNOY, N. (1993). A clinical experiment in bone marrow transplantation: estimating a percentage point of a quantal response curve. In *Case Studies in Bayesian Statistics*, C. Gatsonis, J. Hodges, R.E. Kass and N. Singpurwalla eds. 324-336. Springer-Verlag, New York.

FORD, I., KITSOS, C.P. and TITTERINGTON, D.M. (1989). Recent advances in nonlinear experimental designs. *Technometrics* **31** 49-60.

FORD, I. and SILVEY, S.D. (1980). A sequentially constructed design for estimating a nonlinear parametric function. *Biometrika* **67** 381-388.

FORD, I., TITTERINGTON, D.M. and WU, C.F. (1985). Inference and sequential design. *Biometrika* **72** 545-551.

FORD, I., TORSNEY, B. and WU, C.F. (1992). The use of a canonical form in the construction of locally optimal designs for nonlinear problems. *J. Roy. Statist. Soc. Ser. B* **54** 569-583.

FREEMAN, P.R. (1970). Optimal Bayesian sequential estimation of the median effective dose. *Biometrika* **57** 79-89.

GARTHWAITE, P.H. and DICKEY, J.M. (1988). Quantifying Expert Opinion in Linear Regression Problems. *J. Roy. Statist. Soc. Ser. B* **50** 462-474.

GEISSER, S. (1993). *Predictive Inference: An Introduction.* Chapman and Hall, London

GILL, P.E., MURRAY, W., SAUNDERS, M.A. and WRIGHT, M.H. (1986). *User's Guide for NPSOL (Versison 4.0): A Fortran Package for Nonnlinear Programming.* Technical Report SOL 86-2, Stanford University, Department of Operations Research.

GIOVAGNOLI, A. and SEBASTIANI, P. (1989). The optimum design of experiments in the presence of uncontrolled variability and prior information. In *Optimal Design and Analysis of Experiments*, Y. Dodge, V.V. Fedorov and H.P. Wynn eds. 327-344. North-Holland, Amsterdam.

GIOVAGNOLI, A. and VERDINELLI, I. (1983). Bayes D-Optimal and E-Optimal Block Designs. *Biometrika* **70** 695-706.

GIOVAGNOLI, A. and VERDINELLI, I. (1985). Optimal Block Designs under a Hierarchical Linear Model, in *Bayesian Statistics 2.* J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith eds. 655-661. North Holland, Amsterdam.

GLADITZ, J. and PILZ, J. (1982a). Construction of optimal designs in random coefficient regression models. *Math. Operationsforsch. Stat.*, Ser. Statistics **13** 371-385.

GLADITZ, J. and PILZ, J. (1982b). Bayes designs for multiple linear regression on the unit sphere. *Math. Operationsforsch. Stat.*, Ser. Statistics **13** 491-506.

GOLDSTEIN, M. (1992). Comment on Verdinelli. In *Bayesian Statistics 4*, 477. Oxford University Press.

GOODWIN, G.C. and PAYNE, R.L. (1977). *Dynamic System Identification: Experiment Design and Data Analysis*, Academic Press, New York.

GUTTMAN, I. (1971). A remark on the optimal regression designs with previous observations of Covey-Crump and Silvey. *Biometrika* **58** 683-685.

HAINES, L. (1995). A geometric approach to optimal design for one-parameter non-linear models. *J. Roy. Statist. Soc. Ser. B* **57** 575-598.

HE, Z., STUDDEN, W.J. and SUN, D. (1995). Optimal design for rational models. *Manuscript.*

HAMILTON, D.C. and WATTS, D.G. (1985). A quadratic design criterion for precise estimation in nonlinear regression models. *Technometrics* **27** 241-250.

HEIBERGER, R.M., BHAUMIK, D.K. and HOLLAND, B. (1993). Optimal data augmentation strategies for additive models. *J. Amer. Statist. Association* **88** 926-938.

HUANG, C-Y., WALTER, E. and PRONZATO, L. (1991). Experiment Design Policies for Pharmokinetic Models. In *IMACS 91: 13th World Congress on Computation and Applied Mathematics Proceedings* **3** 1484-1487.

JOHNSON, M. E. and NACHTSHEIM, C. J. (1983). Some guidelines for constructing exact D-optimal designs on convex design spaces. *Technometrics* **25** 271-277.

KASS, R.E. and SLATE, E.H. (1994). Some diagnostics of maximum likelihood and posterior non-Normality. *Ann. Statist.* **22** 668-695.

KATZ, D. and D'ARGENIO, D.Z. (1983). Experimental design for estimating integrals by numerical quadrature, with applications to Pharmakokinetic studies. *Biometrics* **39** 621-628.

KIEFER, J. (1959). Optimum experimental designs (with discussion). *J. Roy. Statist. Soc. Ser. B* **21** 272-319.

KIEFER, J. (1975). Construction and optimality of generalized Youden designs. In *A Survey of Statistical Design and Linear Models.* J.N. Srivastava, ed. 333-353. North Holland, Amsterdam.

KITSOS, C.P., TITTERINGTON, D.M. and TORSNEY, B. (1988). An optimal design problem in rhythometry. *Biometrics* **44** 647-671.

KUO, L. (1983). Bayesian bioassay design. *Ann. Statist.* 11 886-895.

LAD, F. and DEELY (1994). Experimental design from a subjective utilitarian viewpoint. In *Aspects of Uncertainty*, P.R. Freeman and A.F.M. Smith eds. 267-282. Wiley, Chichester.

LAKATOS, E. (1988). Sample sizes based on the log-rank statistic in complex clinical trials. *Biometrics* **44** 229-241.

LANDAW, E.M. (1982). Optimal multicompartmental sampling designs for parameter estimation: practical aspects of the identification problem. *Math. and Computers in Simulation* **24** 525-530.

LANDAW, E.M. (1984). Robust sampling designs for compartmental models under large prior eigenvalue uncertainties. In *Mathematics and Computers in Biomedical Applications*, J. Eisenfeld and C. DeLisi eds. 181-187. Elsevier North-Holland, Amsterdam.

LAUNAY, M-C. and ILIADIS, A. (1988). A Feasibility study of optimal sampling schedules in clinical Pharmacokinetics. In *Proc. 12th IMACS World Congress on Scientific Computation, 1988, Paris*, 131-133.

LÄUTER, E. (1974). Experimental design in a class of models. *Math. Operationsforsch. Statist.* **5** 379-396.

LÄUTER, E. (1976). Optimal multipurpose designs for regression models. *Math. Operationsforsch. Statist., Ser. Statist.* **7** 51-68.

LINDLEY, D.V. (1956). On the Measure of Information provided by an Experiment. *Ann. Statist.* **27** 986-1005.

LINDLEY, D.V. (1968). The Choice of Variables in Multiple Regression. *J. Roy. Statist. Soc. Ser. B* **30** 31-53.

LINDLEY, D.V. (1972). *Bayesian Statistics – A Review.* SIAM, Philadelphia.

LINDLEY, D.V. and NOVICK, M.R. (1981). The role of exchangeability in inference. *Ann. Statist.* **9** 45-58.

LINDLEY, D.V. and SMITH, A.F.M. (1972). Bayes Estimates for the Linear Model (with Discussion). *J. Roy. Statist. Soc. Ser. B* **34** 1-41.

LOHR, S.L. (1995). Optimal Bayesian design of experiments for the one-way random effects model. *Biometrika* **82** 175-186.

MANDAL, N.K. (1978). On estimation of the maximal point of a single quadratic regression. *Calcutta Statistical Association Bull.* **27** 119-125.

MAJUMDAR, D. (1988). Optimal Block Designs for Comparing new Treatments with a Standard Treatment. In *Optimal Design and Analysis of Experiments*, Y. Dodge, V. Fedorov and H. Wynn eds. 15-27. North-Holland, Amsterdam.

MAJUMDAR, D. (1992). Optimal Designs for Comparing Test Treatments with a Control Utilizing Prior Information. *Ann. Statist.* **20** 216-237.

MALLET, A. and MENTRE, F. (1988). An approach to the design of experiments for estimating the distribution of parameters in random models. In *Proc. 12th IMACS World Congress on Scientific Computation, 1988, Paris*, 134-137.

MAYER, L.S. and HENDRICKSON, A.D. (1973). A method for constructing an optimal regression design after an initial set of input values has been selected. *Comm. Statist., Theory Methods* **2(5)** 465-477.

MEHRA, R.K. (1974). Optimal Input Signals for Parameter Estimation in Dynamic Systems – Survey and New Results. *IEEE Trans. Auto. Control* **19** 753-768.

MITCHELL, T.J. and SCOTT, D.S. (1987). A computer program for the design of group testing experiments. *Comm. Statist., Theory Methods* **16(10)** 2943-2955.

MORRIS, M.D., MITCHELL, T.J. AND YLVISAKER, D. (1993). Bayesian design and analysis of computer experiments: use of derivatives in surface prediction. *Technometrics* **35** 243-255.

MUKHOPADHYAY, S. and HAINES, L. (1995). Bayesian D-optimal designs for the exponential growth model. To appear *Journal of Statistical Planning and Inference.*

MUKERJEE, R. and HUDA, S. (1988). Optimal design for the estimation of variance components. *Biometrika* **75** 75-80.

MÜLLER, P. and PARMIGIANI, G. (1995). Numerical evaluation of information theoretic measures. To appear in *Bayesian Statistics and Econometrics: Essays in Honor of A. Zellner*, D.A. Berry. K.M. Chaloner, J.F. Geweke eds. Wiley, New York.

NÄTHER, W. and PILZ, J. (1980). Estimation and experimental design in a linear regression model using prior information. *Zastosowania Matematyki* **16** 565-577.

NAYLOR, J.D.H. (1994). *Optimal design for accelerated life tests with restricted resources.* Ph. D. Thesis. School of Statistics, University of Minnesota.

O'BRIEN, T.E. and RAWLINGS, J.O. (1994a). A new subset design strategy for nonlinear regression models. *Manuscript.*

O'Brien, T.E. and Rawlings, J.O. (1994b). A non-sequential design procedure for parameter estimation and model discrimination in nonlinear regression models. *Manuscript.*

O'Brien, T.E. and Rawlings, J.O. (1994c). A practical design methodology for nonlinear regression models. *Manuscript.*

O'Hagan, A. (1978). Curve fitting and optimal design for prediction (with discussion). *Jour. Roy. Statist. Soc., Ser. B* **40** 1-41.

Owen, R.J. (1970). The optimum Design of a two-factor experiment using prior information. *Ann. Statist.* **41** 1917-34.

Owen, R.J. (1975). A Bayesian sequential procedure for quantal response in the context of adaptive mental testing. *J. Amer. Statist. Assoc.* **70** 351-356.

Parmigiani, G. (1993). On optimal screening ages. *J. Amer. Statist. Association* **88** 622-628.

Parmigiani, G. and Berry, D.A (1994). Applications of Lindley information measure to the design of clinical experiments. In *Aspects of Uncertainty.* P.R. Freeman and A.F.M. Smith eds. 329-348. Wiley, Chichester.

Parmigiani, G. and Kamlet, M.S. (1993). A cost-utility analysis of alternative strategies in screening for breast cancer. In *Bayesian Statistics in Science and Technology: Case Studies.* C. Gatsonis, J. Hodges, R.E. Kass, and N. Singpurwalla, eds. Springer-Verlag, New York.

Pázman, A. and Pronzato, L. (1992a). Nonlinear experimental design based on the distribution of estimators. *Journal of Statistical Planning and Inference* **33** 385-402.

Pázman, A. and Pronzato, L. (1992b). Nonlinear experimental design for constrained LS estimation. In *Model Oriented Data Analysis*, V.V. Fedorov, W.G. Müller and I.N. Vuchkov eds. 67-76. Springer-Verlag, Heidelberg.

Pilz, J. (1979a). Konstruktion von optimalen diskreten Versuchsplänen für eine Bayes-Schätzung im linearen Regressionsmodell. *Freiberger Forschungshefte* **117** 67-94.

Pilz, J. (1979b). Optimalitätskriterien, Zulässigkeit und Vollständigkeit im Planungsproblem für eine bayessche Schätzung im linearen Regressionsmodell. *Freiberger Forschungshefte* **117** 67-94.

Pilz, J. (1979c). Das bayessche Schätzproblem im linearen Regressionsmodell. *Freiberger Forschungshefte* **117** 21-55.

Pilz, J. (1979d). Entscheidungstheoretische Darstellung des Problems der bayesschen Schätzung und Versuchsplanung im linearen Regressionsmodell. *Freiberger Forschungshefte* **117** 7-20.

Pilz, J. (1981a). Robust Bayes and minimax-Bayes estimation and design in linear regression. *Math. Operationsforsch., Stat. Ser. Statistics* **12** 163-177.

Pilz, J. (1981b). Bayesian one-point designs in linear regression. Paper presented at the *Fifth International Summer School on Problems of Model Choice and Parameter Estimation in Regression Analysis*, March 1981, Sellin, D.D.R.

PILZ, J. (1981c). Bayessche Schätzung und Versuchsplanung für die multiple lineare Regression. IX. Internat. Kongreß über Anwendungen der Mathematik in den Ingenieurwissenschafen. Weimar 1981 **4** 54-57.

PILZ, J. (1983). *Bayesian Estimation and Experimental Design in Linear Regression Models.* Teubner, Leipzig.

PILZ, J. (1991). *Bayesian Estimation and Experimental Design in Linear Regression Models.* Second Edition. Wiley, New York.

PONCE DE LEON, A.C. and ATKINSON, A.C. (1993). Optimum experimental design for discriminating between two rival models in the presence of prior information. *Biometrika* **78** 601-608.

PRONZATO, L., HUANG, C.Y., WALTER, E., LE ROUX, Y. and FRYDMAN, A. (1989). Planification d'expériences pour l'estimation de paramètres pharmacocinetiques. In *Informatique et Santé 2*, A. Venst and P. Degoulet eds. 3-13. Springer-Verlag, Paris.

PRONZATO, L. and WALTER, E. (1985). Robust experiment design via stochastic approximation. *Mathematical Biosciences* **75** 103-120.

PRONZATO, L. and WALTER, E. (1987). Robust experiment designs for nonlinear regression models. In *Model-Oriented Data Analysis*, V.V. Fedorov and H. Lauter, eds. 77-86. Springer-Verlag, Berlin.

PRONZATO, L. and WALTER, E. (1988). Qualitative and Quantitative experiment design for nonlinear models. *Proc. of the IFAC Symposium, Venice, Italy, 1988.*

PUKELSHEIM, F. (1980). On linear regression designs which maximize information. *Journal of Statistical Planning and Inference* **4** 339-364.

PUKELSHEIM, F. (1993). *Optimal Design of Experiments.* Wiley, New York.

PUKELSHEIM, F. and RIEDER, S. (1992). Efficient rounding of approximate designs. *Biometrika* **79** 763-770.

PUKELSHEIM, F. and ROSENBERGER, J.L. (1993) Experimental designs for model discrimination. *J. Amer. Statist. Association* **89** 642-649.

RAIFFA, H. and SCHLAIFER, R. (1961). *Applied Statistical Decision Theory.* Division of Research, Harvard Business School, Boston.

RIDOUT, M. (1995). Three stage designs for seed testing experiments. *Applied Statistics* **44** 153-162.

SACKS, J., WELCH, W.J., MITCHELL, T.J. and WYNN, H.P. (1989). Design and analysis of computer experiments. *Statistical Science* **4** 409-423.

SACKS, J. and YLVISAKER, D. (1984). Some Model Robust Designs in Regression. *Ann. Statist.*, **12** 1324-1348.

SACKS, J. and YLVISAKER, D. (1985). Model Robust Designs in Regression. In *Proc. of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer.* L.M. LeCam and R.A. Olshen, eds. **2** 667-679. Wadsworth, Monterey, Calif.

SAN MARTINI, A. and SPEZZAFERRI, F. (1984). A Predictive Model Selection Criterion. *J. Roy. Statist. Soc. Ser. B* **46** 296-303.

SEO, H.S. and LARNTZ, K. (1992). Restricted Bayesian optimal design. *Technical Report 574*, School of Statistics, University of Minnesota.

SHANNON, C.E. (1948) A Mathematical theory of Communication. *Bell System Tech. J.*, **27**, 379-423 and 623-656.

SHIH, J. (1995). Sample size calculation for complex clinical trials with survival endpoints. To appear in *Controlled Clinical Trials.*

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

SIMEONE, B. and VERDINELLI, I. (1989). A Feasible directions Method for computing Bayes E-optimal Block Design. *Comp. Statist. & Data Analysis* **7** 23-38.

SINHA, B.K. (1970). A Bayesian approach to optimum allocation in regression problems. *Calcutta Statistical Association Bulletin* **19** 45-52.

SLATE, E.H. (1991). Reparametrizing Statistical Models. Ph.D. Dissertation, Dept. Statistics, Carnegie Mellon University.

SMITH, A.F.M. and VERDINELLI, I. (1980). A Note on Bayesian Design for Inference using a Hierarchical Linear model. *Biometrika* **67** 613-619.

SPEZZAFERRI, F. (1988). Nonsequential designs for model discrimination and parameter estimation. *Bayesian Statistics 3*, J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith eds. 777-783. Oxford University Press, Oxford.

SPIEGELHALTER, D.J. and FREEDMAN, L.S. (1986). A predictive approach to selecting the size of a clinical trial, based on subjective clinical opinion. *Statistics in Medecine* **5** 1-13.

STANGL, D. and MUKHOPADHYAY, S. (1993). Balancing Centers and observations in multicenter clinical trials. Technical Report, Institute of Statistics and Decision Sciences, Duke University.

STEINBERG, D.M. and HUNTER, W.G., (1984). Experimental design: review and comment. *Technometrics* **26** 71-97.

STEINBERG, D.M., (1985). Model robust response surface designs: scaling two-level factorials. *Biometrika* **72** 513-26.

STONE, M. (1959a). Application of a measure of Information to the Design and comparison of Regression Experiment. *Ann. Stat.*, **30**, 55-70.

STONE, M. (1959b). Discussion of Kiefer, *J. Roy. Statist. Soc. Ser. B* **21** 313-315.

SUN, D., TSUTAKAWA, R.K. and LU, W-S. (1995). Bayesian design of experiment for quantal responses: what's promised versus what's delivered. *Manuscript.*

SYLVESTER, R.J. (1988). A Bayesian approach to the design of Phase II clinical trials. *Biometrics* **44** 823-836.

TIAO, G. C. and AFONJA, B. (1976). Some Bayesian Considerations of the Choice of Design for Ranking, Selection and Estimation. *Ann. Instit. Statist. Mathem.* **28** 167-186.

TIERNEY, L. (1990). *LISP-STAT: An Object Oriented Environment for Statistical Computing and Dynamic Graphics*. Wiley, New York.

TOMAN, B. (1992a). Bayesian Robust Experimental Designs for the one-way analysis of Variance. *Statistics and Probability Letters* **15** 395-400.

TOMAN, B. (1992b). Discussion of Verdinelli, *Bayesian Statistics 4*. J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith eds. Oxford University Press. 477-478.

TOMAN, B. (1995). Bayesian experimental design for multiple hypothesis testing. To appear in *J. Amer. Statist. Assoc.*

TOMAN, B. (1994). Bayes Optimal Designs for Two and Three Level Factorial Experiment. *J. Amer. Statist. Assoc.* **89** 937-946.

TOMAN, B. and GASTWIRTH J. L. (1993). Robust Bayesian Experimental design and Estimation for analysis of variance models using a class of Normal Mixtures. *Journal of Statistical Planning and Inference* **35** 383-398.

TOMAN, B. and GASTWIRTH J. L. (1994). Efficiency Robust Experimental Design and Estimation Using a Data-Based Prior. Technical Report # 93-19, Department of Statistics/Computer and Information Systems. The George Washington University. To appear in Statistica Sinica.

TOMAN, B. and NOTZ, W. (1991). Bayesian optimal experimental design for treatment-control comparisons in the presence of two-way heterogeneity. *Journal of Statistical Planning and Inference* **27** 51-63.

TOMASETH, K. and COBELLI, C. (1988). Optimal input design for physiological systems identification: theory and practice. In *Proc. 12th IMACS World Congress on Scientific Computation, 1988, Paris*, 514-516.

TSUTAKAWA, R.K. (1972). Design of an experiment for bioassay. *J. Amer. Statist. Association* **67** 584-590.

TSUTAKAWA, R.K. (1980). Selection of dose levels for estimating a percentage point of a logistic quantal response curve. *Applied Statistics* **29** 25-33.

TUCHSCHERER, A. (1983). Experimental design for Bayesian estimations in the linear regression model taking costs into account. *Biometrical Journal.* **25** 515-525.

VERDINELLI, I. (1983). Computing Bayes D- and E-Optimal Designs for a two-way Model. *The Statistician* **32** 161-167.

VERDINELLI, I. (1990). Procedure di Randomizzazione nella Statistica Bayesiana. *Atti della XXV Riunione Scientifica della SIS.* 35-41. Cedam, Padova.

VERDINELLI, I. (1992). Advances in Bayesian Experimental Design. *Bayesian Statistics 4*. J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith eds. 467-481. Oxford University Press, Oxford.

VERDINELLI, I. and KADANE, J. B. (1992). Bayesian Designs for maximizing information and outcome. *J. Amer. Statist. Assoc.* **87** 510-515.

VERDINELLI, I., POLSON, N. and SINGPURWALLA, N. (1993). Shannon Information and Bayesian Design for Prediction in Accelerated life testing. *Reliability and Decision Making*. R.E. Barlow, C.A. Clarotti, and F. Spizzichino eds. 247-256. Chapman and Hall, London.

VERDINELLI, I. and WYNN, H. P. (1988). Target attainment and Experimental Design, a Bayesian Approach. In *Optimal Design and Analysis of Experiments*, Y. Dodge, V. Fedorov and H. Wynn eds. 319-324. North-Holland, Amsterdam.

WALTER, E. and PRONZATO, L. (1985). How to design experiments that are robust to parameter uncertainty. *Preprints 7th IFACS/IFORS Symposium Identification and System Parameter Estimation*, York, vol. 1, 921-926.

WARNER, J. (1993). Optimal Bayesian designs for nonlinear problems via the Gibbs sampler. To appear in *The Proceedings of the Section on Bayesian Satistics*, ASA, San Francisco 1993.

WELCH, W.J., BUCK, R.J., SACKS, J., WYNN, H.P., MITCHELL, T.J. and MORRIS, M.D. (1992). Screening, predicting, and computer experiment. *Technometrics* 15-25.

WHITE, L. V. (1973). An extension of the general equivalence theorem to nonlinear models. *Biometrika* **60** 345-348.

WHITE, L. V. (1975). *The optimal design of experiments for estimation in nonlinear models*. Ph.D. Dissertation. University of London.

WHITTLE, P. (1973), Some general points in the theory and construction of D-optimum experimental designs. *J. Roy. Statist. Soc. Ser. B* **35** 123-130.

WU, C.F. (1985). Efficient sequential designs with binary data. *J. Amer. Statist. Association* **80** 974-984.

WU, C.F. (1988). Optimal design for percentile estimation of a quantal response curve. In *Optimal Design and Analysis of Experiments*, Y. Dodge, V.V. Fedorov, and H.P. Wynn eds. 213-223. North-Holland, Amsterdam.

WU, M., FISHER, M. and DEMETS, D. (1980). Sample sizes for long-term medical trial with time-dependent dropout and event rates. *Controlled Clinical Trials* **1** 109-121.

ZACKS, S. (1977). Problems and approaches in design of experiments for estimation and testing in non-linear models. In *Multivariate Analysis IV*, 209-223, P.R. Krishnaiah ed. 209-223. North-Holland, Amsterdam.