A Robust Bayesian Design Criterion for Nonlinear Models

Sydney Akapame[*], **John J. Borkowski**

Department of Mathematical Sciences

Montana State University, Bozeman, MT 59715, USA

[*] corresponding author; e-mail: kwasi.akapame@gmail.com

Abstract

Nonlinear models pervade the statistical literature on drug development, and specifically in pharmacokinetics (PK), pharmacodynamics (PD), and the biological and physical sciences in general. Obtaining efficient experimental designs for such models is non-trivial due to the well-documented parameter-sensitivity problem. Bayesian methods, which integrate prior information about the model parameters into the design process, have been proposed as a solution to the problem. In implementing such methods, the assumption is made that a single prior distribution exists for the parameters which may not be the case. In this research, we discuss situations in which there may be multiple (or competing) prior distributions and propose a robust design criterion for obtaining efficient designs in such cases.

1. Introduction

Experimentation, critical to the scientific method and manufacturing, is also an efficient way of learning about the world. While most of the pioneering work in experimental design started in the field of agriculture, notably by R. A. Fisher, experiments are performed in the educational, biological, chemical and physical sciences. These and many other fields have benefited immensely over the past century from the extensive research in the design of experiments. Fisher is largely credited with the role of statistics in experimental design and perhaps more importantly, the role of the statistician in experimentation: from being an after-the-fact technician to being an active collaborator at all stages of an investigation [1].

The importance of a statistically sound experimental design cannot be over-emphasized. Response surface designs including central composite designs, Plackett-Burman designs, as well as full and fractional factorial designs [2] have been used to model relationships such as those between a chemical endpoint or response and a set of input variables. Response surface designs are widely used in industrial experiments to determine optimal operating conditions for a system. Optimal two-stage studies are used in epidemiological studies to maximize information gain [3]. Fedorov and Leonov [4] discuss optimal experimental designs in the drug development literature, especially in the study of pharmacokinetics (PK) and pharmacodynamics (PD).

The procedure of designing an experiment, according to Atkinson and Bailey [5] in their review of design of experiments, consists of three important phases:

   i. Choice of treatments,

  ii. Choice of experimental units and,

 iii. Deciding which treatment to apply to which experimental unit.

The authors mention that the relative importance of each phase depends on the application. in many clinical trials, patients enter the trial sequentially in which case phase (ii) is very important, and where there is structure, for example sex or age, in the experimental units, phase (iii) becomes important. In this paper, we will focus on phase (i), that is, the choice of treatments. Following [5], in an experimental setting, the choice of treatments may be qualitative or quantitative (or both in some cases, like in the chemical industry). In a clinical trial,

for example, the efficacy of two drugs can be compared to a placebo in which case the treatments are qualitative. In the foregoing case, the choice of treatment is important but not mathematically sophisticated. For example, there are no intermediate levels for the treatments. As Chaloner and Verdinelli [6] point out, the main idea behind experimental design is that statistical inference about quantities of interest can be improved remarkably by appropriately selecting the values of the treatment or control variables. This makes the choice of treatments in an experiment quite critical if we hope to learn anything about a particular system (like in PK studies) or a process. For example in Atkinson, Chaloner, Herzberg, and Juritz [7], the question is asked, "When a horse is injected with a drug, at what times must blood be drawn from the animal in order to model the passage of the drug through the bloodstream?". In dose-response studies where quantal models are used, the question is asked: "What quantity of a dose will result in, for example, a 50% success rate?". Questions of this sort are common in practice and have led to so-called optimal experimental designs [8]. Optimal designs are experimental designs that are based on one or more design optimality criteria. For a complete review of optimal designs in terms of applications, see Berger and Wong [9].

In this paper, we focus on situations when treatments are quantitative, that is, having potential levels throughout an interval I of $\mathcal{R}$. Also, we assume that a model is known *a priori*. Section 2 discusses the theory of optimal designs and more importantly, design optimality criteria. In Section 3, we focus on the optimal design problem for nonlinear models and discuss some potential solutions. Section 4 specifically addresses a Bayesian solution to the nonlinear design problem. We introduce our solution to the design problem in Section 5 and apply it to the four-parameter logistic (4PL) model in section 6, and conclude with a discussion in Section 7.

2. Overview of Optimality Theory and Design Optimality Criteria

Although not often mentioned in the optimal design literature, the work of Smith [10] was a precursor to Kiefer [8] and Kiefer and Wolfowitz [11] as far as the theory of optimal design is concerned. For example, in her seminal work, [10] obtained optimal designs for a polynomial in one factor over a $[-1, 1]$ design region. However, the concept of optimal designs is largely credited to Kiefer and his colleagues. Expositions on the subject of optimality theory have

been given in Fedorov [12], Pukelsheim [13] and Atkinson et al. [14]. Optimal designs were originally for linear models, in particular, response surface models over regular regions [5]. More recently, optimal designs have been widely studied in the context of nonlinear models.

To motivate the theory of optimal design, consider a univariate linear model

$$y_i = \eta(x_i, \theta) + \epsilon_i \tag{1}$$

where y_i is the response for the ith experiment, the functional form of $\eta(x_i, \theta)$ is assumed to be known, θ is a $p \times 1$ vector of unknown parameters, $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ is the ith random error, and x_i is the value of the explanatory variable x for the ith experiment. The optimal design goal is to choose r distinct design points in the design space $\mathcal{X}$ such that the optimal design

$$\xi = \left\{ \begin{array}{ccccc} x_1, & x_2, & . & . & . & , x_r \\ w_1, & w_2, & . & . & . & , w_r \end{array} \right\} \tag{2}$$

produces experimental data for estimating some function of the p-dimensional parameter vector θ with high efficiency. Here w_k, $k = 1, ..., r$, is the amount of experimental effort at the kth design point. Thus, the w_i weights are positive and sum to one. These weights are often irrational in which case the optimal design is said to be *continuous*. Otherwise, if n_k out of of a total of n observations are made at the kth support point, then $w_k = n_k/n$ and the design is said to be *exact*. Although continuous designs are not implementable in practice, the mathematical problem of finding optimal designs is simplified by treating ξ as a design measure [15] which satisfies:

$$\xi(x) \geq 0, \; x \in \mathcal{X} \text{ and } \int_{\mathcal{X}} \xi dx = 1. \tag{3}$$

Pukelsheim and Reider [16] show how optimum continuous designs can be converted into near-optimum exact designs, with minimum loss in efficiency, so that they can be used in practice.

The information matrix, apart from a multiplicative constant, for a continuous design is given by

$$M(\xi, \theta) = \int_{\mathcal{X}} f(x, \theta) f^T(x, \theta) \xi(\mathrm{d}x) \tag{4}$$

where $f(x, \theta)$ is the vector of partial derivatives of $\eta(x, \theta)$ with respect to θ. For the linear model, the dependence of M on θ can be dropped in general unless

interest is in estimating a nonlinear function of the parameters. Consequently, it is important to note that for linear models, the information matrix M depends only on the support points, x. Optimal design theory is concerned with maximization (or minimization) of a concave (or convex) function of the information matrix M. In general, the optimal design problem is a constrained nonlinear mathematical programming problem. The experimental objective in a particular setting, determines the suitable convex function of M to be optimized. Functions of this sort are known as design optimality criteria.

In most applications the experimental objective is efficient estimation of model parameters. For example, for certain fixed-effects analysis of variance models, Wald [17] proposed maximizing the determinant $|M|$, or equivalently minimizing $|M^{-1}|$, in an attempt to maximize the power of the F-ratio for testing a linear hypothesis on the parameters. This criterion, called the *D-optimality* criterion in Kiefer and Wolfowitz [18], is the most widely-used optimality criterion. An equivalent criterion is $\log |M(\xi)|$. Maximizing the D criterion is equivalent to maximizing the product of the eigenvalues of M. Often, in analysis of variance models, estimating a set of contrasts with minimum variance is of interest. In this case, the *C-optimality* criterion is used, and the function to be minimized is $tr\{C^T M^{-1}(\xi)C\}$, where the rows of the $s \times p$ matrix C^T are the vectors of constants corresponding to the contrasts. In response surface methodology, minimizing the maximum prediction variance in the design space is often of interest. In this regard, [10] proposed what is now known as the *G-optimality* criterion. In a $G-$optimal design, the maximum value of the variance of prediction

$$d(x,\xi) = f^T(x)M^{-1}(\xi)f(x) \tag{5}$$

over the design space is minimized. Other optimality criteria exist for different experimental objectives. When more than one objective is necessary, combinations of optimality criteria can be used. See Chapter 10 of [14] for other optimality criteria.

Kiefer [8,11, 18-21] and other authors developed the general theory of optimal designs. The most important theoretical result is the *general equivalence theorem* (GET). The GET is practically useful in that it led to the development of algorithmic methods for optimal design construction and methods for checking the optimality of a proposed design. It is must be noted that, in general, the GET applies to continuous designs but not exact designs. However, a family

of algorithms have been derived based on the theorem [22]. Another important result of the theorem is that $G-$ and $D-$optimal continuous designs are equivalent. Although the work of Kiefer and his colleagues is focused on linear models, White [23] extended it to nonlinear models which is discussed in Section 3. To conclude our overview of optimality theory, we point out that, in general, the optimal design for a $k-$parameter model has k distinct support points. An advantage over classical designs is that the number of support points is minimized. However, this also poses a problem because the inadequacy (or lack-of-fit) of the model cannot be checked and is the main criticism of optimal designs. O'Brien [24] and others suggest ways this can be overcome. An important consequence of this problem is the published research on *robust* designs.

3. Optimal Design Problem for Nonlinear Models

There are many cases in practice where a linear regression model is not appropriate for data analysis. For example, the true relationship between the response and the predictor is a differential equation or a solution to a differential equation. In nonlinear models, the expectation function is not linear in the model parameters. Nonlinear situations can also arise in linear models when interest is in estimating a nonlinear function of the parameters. The importance of nonlinear models cannot be overstated given their use in the biological, chemical and physical sciences as well as in industry. Bates and Watts [25] and Seber and Wild [26] provide a thorough discussion of nonlinear regression analysis. Traditionally, researchers have resorted to transformations that linearize the nonlinear model in order to use estimation methods that are applicable to linear models. Common among the transformations used is the natural logarithmic transformation, which has worked quite well in many cases. However, it is important to point out here that in cases where the errors are additive, rather than multiplicative, the use of natural logarithmic transformation may not be advisable [27].

The Michaelis-Menten model [25] for enzyme kinetics relates the initial velocity of an enzymatic reaction to the substrate concentration. To illustrate the optimal design problem for nonlinear models, consider the expectation function

for the ith observation of the Michaelis-Menten model

$$\eta(x_i, \theta) = \frac{\theta_1 x_i}{\theta_2 + x_i} \tag{6}$$

where $\eta(x_i, \theta)$ is the mean velocity of the reaction given a substrate concentration x_i. The vector of partial derivatives is

$$f(x_i, \theta) = \begin{bmatrix} \dfrac{\partial \eta(x_i, \theta)}{\partial \theta_1} & \dfrac{\partial \eta(x_i, \theta)}{\partial \theta_2} \end{bmatrix}^T = \begin{bmatrix} \dfrac{x_i}{\theta_2 + x_i} & \dfrac{-\theta_1 x_i}{(\theta_2 + x_i)^2} \end{bmatrix}^T. \tag{7}$$

Then for a continuous design with r support points and weights $w_1, w_2, \ldots, w_r$, the information matrix is given by

$$M(\xi, \theta) = \sum_{i=1}^{r} w_i f(x_i, \theta) f(x_i, \theta)^T \tag{8}$$

which depends on $\theta = (\theta_1, \theta_2)$. In particular, the ith information matrix

$$M_i(\xi, \theta) = \begin{bmatrix} \dfrac{x_i^2}{(\theta_2 + x_i)^2} & \dfrac{-\theta_1 x_i^2}{(\theta_2 + x_i)^3} \\ \dfrac{-\theta_1 x_i^2}{(\theta_2 + x_i)^3} & \dfrac{\theta_1^2 x_i^2}{(\theta_2 + x_i)^4} \end{bmatrix} \tag{9}$$

clearly depends on θ. The optimal design problem for nonlinear models is a result of the dependence of the information matrix on the unknown parameters. To put things in context, experiments are designed to learn about unknown model parameters. The dependence of the information matrix on these unknown parameters means that to obtain, for example, a $D-$optimal design for the Michaelis-Menten model, the experimenter needs to have prior knowledge of the model parameters. This paradox was noted as an absurd bargain between experimenter and statistician by Cochran [28]: "You tell me the value of θ and I promise to design the best experiment for estimating θ."

Various approaches have been proposed in the literature for addressing this issue. The earliest and perhaps simplest approach is the idea of *local optimality* by Chernoff [29] and Box and Lucas [30]. If a guess of the parameter vector can be made, then an optimal design can be obtained at this local value of the parameter vector, and the resulting design is said to be locally optimal. Obviously, locally optimal designs seem reasonable in single parameter models or where most of the parameters are conditionally linear, like θ_2 in the Michaelis-Menten model. In fact, locally optimal designs will approximate the *true* optimal designs quite closely if reasonable guesses of the parameter vector can be made. These designs, however, can be very inefficient if the *true* parameter

values are not close to the guesses. Although they may not be practical in many situations, Ford, Titterington, and Kitsos [31] identify some very important uses of locally optimal designs. Box and Hunter [32], Chernoff [33], and others have proposed sequential experimental designs in situations where initial experiments have been done to obtain prior parameter estimates. Although conceptually ideal, sequential designs are not desirable in problems where the cost of data collection is large. Quite popular in the literature and more suited to our purposes in this paper is the subject of Bayesian optimal designs which we discuss in the next section.

4. Bayesian Optimal Designs

Often prior information will exist before an experiment is performed. This information may be in the form of prior point estimates and/or distributions of model parameters, or observations from a previous experiment. Another important source of prior information is expert opinion. It is the opinion of the authors that expert opinion constitutes the most important form of prior information especially in the case of nonlinear design problems where, paradoxically, knowledge of the unknown parameters is required in order to obtain an optimal design. This is because prior observations or point estimates are frequently unavailable before an experiment is conducted while expert opinion will be available. The availability of prior information makes the Bayesian paradigm useful in experimental design as it provides the framework to incorporate such information. A detailed review of the Bayesian approach to experimental design is given by [6] and DasGupta [34].

In the Bayesian paradigm, the design problem for both linear and nonlinear models is treated as a decision problem [35] and prior information is quantified by a prior probability distribution, $p(\theta)$. Following [6], the experimental design ξ is chosen from a set of possible designs Ξ, and data y is then observed, and followed by a terminal decision $d \in \mathcal{D}$. The decision is in two parts:

1. The selection of the design $\xi \in \Xi$ and,

2. The terminal decision (or goal) of the experiment.

The terminal decision reflects the objective of the experiment and a utility function $U(d, \xi, \theta, y)$ is specified accordingly. The Bayesian solution to the

design problem is then to find the design ξ^* that maximizes expected utility

$$U(\xi) = \int_y \max_{d \in \mathcal{D}} \int U(d, \xi, \theta, y) p(\theta|y, \xi) \cdot p(y|\xi) \mathrm{d}\theta \mathrm{d}y. \qquad (10)$$

Intuitively, the Bayesian decision-theoretic approach to design, according to [6], is to

1. Specify a utility function that reflects the experimental objective,

2. Regard the design choice as a decision problem and,

3. Select a design that maximizes expected utility.

Bayesian equivalents of several optimality criteria can be obtained through the use of appropriate utility (or loss) functions. For example, for the model in (1), use of Shannon information [36] results in the Bayesian D-optimality criterion. For the normal linear model in (1), if the prior distribution of θ is such that $\theta|\sigma^2 \sim N(\theta_0, \sigma^2 R^{-1})$, then the Bayesian D-optimality criterion is [6]

$$\phi(\xi) = \log \det\{M(\xi) + R\}, \qquad (11)$$

where R is the $p \times p$ prior precision matrix with known elements. Thus, (11) is maximized to obtain the Bayesian D-optimal design for (1). It is important to note that the design does not depend on the prior location of θ but on the precision matrix R. This makes intuitive sense since the optimal design, in the case of a linear model, does not depend on the model parameters. The prior information matrix, $\sigma^2 R^{-1}$ is in a sense equivalent to some (not necessarily an integer) N_0 prior observations. Also notable is the fact that if prior information about θ is weak, that is, $R \to 0$, then $\phi(\xi) \to \log \det M(\xi)$ which is the classical D-optimality criterion.

However, in the case of nonlinear models, the exact expected utility is often a complicated integral [6], and so a normal approximation is used. Following [14], by ignoring the prior information matrix, the design maximizing expected Shannon information will maximize

$$\Phi(\xi) = \int \log |M(\xi, \theta)| p(\theta) \mathrm{d}\theta. \qquad (12)$$

Unlike in a linear model, the optimal design depends both on the prior location and the precision matrix. In Section 2, we mentioned the fact that (a non-Bayesian) optimal design for a k-parameter model typically has k unique support points. Bayesian designs generally differ from classical optimal designs

because the number of support points increases as prior information becomes more disperse or less informative. This also is intuitive in the sense that more support points are needed to estimate θ if the amount of information about θ prior to the experiment is uninformative. Similarly, where substantial prior information is available, fewer support points are needed to estimate θ. Chaloner and Larntz [37] give the Bayesian version of the Kiefer-Wolfowitz equivalence theory results for Bayesian designs.

Bayesian experimental designs are based on a single prior distribution and are quite sensitive to its choice [38]. In most practical situations, the statistician will elicit prior information from more than one subject matter expert. This inevitably results in multiple prior distributions, and hence, increases the variability of the model parameters. For example, Tsai and Chaloner [39] describe a design problem where prior distributions are elicited from over 50 clinical experts. This paper focuses on a method of designing experiments efficiently in situations where there are multiple prior distributions.

5. Robust Design Criterion for Nonlinear Models

Authors have looked into the issue of robustness in Bayesian experimental designs. The earliest work we could find as far as robustness of the optimal design to the particular prior distribution is [40] who obtained optimal robust designs that minimize Bayes risk for a fixed prior distribution subject to being robust to misspecification of the prior. Also, [41] and [42] also examine the situation where there is a class of plausible prior distributions in the context of analysis of variance (ANOVA). Work by Dasgupta, Mukhopadhyay, and Studden [43] is also relevant. Common among the foregoing is the fact that they examined the problem in the context of linear models only.

For nonlinear models, we propose the robust design criterion

$$\Psi(\xi) = \left(\prod_{i=1}^{k} \int \Gamma(\xi, \theta) p_i(\theta) \mathrm{d}\theta \right)^{\frac{1}{k}} \tag{13}$$

where k is the number of plausible priors, $p_i(\theta)$ is the ith prior distribution and $\Gamma(\xi, \theta)$ is any of the alphabetic optimality criteria. The optimal robust or compromise design is then

$$\xi_C = \arg\max_{\xi} \Psi(\xi). \tag{14}$$

The criterion introduced here is really an extension of the existing Bayesian design criterion. It is clear that if there is no variability in the prior information, i.e, for $k = 1$, then the criterion reduces to (10). Essentially, the criterion is a geometric mean of the Bayesian optimality criterion values for k prior distributions, and it has at least one important advantage. A design found using this criterion is guaranteed to perform efficiently for a wide range of values of θ than one which is based on an arithmetic mean. This property is particularly important given that the objective of this criterion is to arrive at compromise designs in the face of varying prior distributional assumptions. Weights that suggest the likelihood of the priors are ignored in this definition of the criterion.

6. Application to 4PL Model

Biological assays are methods that investigate the biological properties of a compound (e.g., a drug) by the analysis of its effects on living matter. In a typical bioassay, a stimulus (e.g., a dose of drug) is applied to a subject yielding a change in a measurable characteristic (or response) of the subject. In drug development research, the relationship between the dose of a drug and a clinical endpoint (response) is of paramount interest. Consequently, estimating the parameters of the model describing the dose-response relationship is critical. In most pharmacological studies, the four-parameter logistic (4PL) model has been found to adequately model this relationship. The mean number of cells at log concentration x is given by

$$\eta(x, \theta) = \theta_3 + \frac{\theta_4 - \theta_3}{1 + (x/\theta_1)^{\theta_2}} \tag{15}$$

where θ_1 is often denoted as ED_{50}; θ_2 is a slope parameter; θ_3 and θ_4 are lower and upper asymptotes [4]. In designing an experiment that will optimally estimate the model parameters, suppose that prior elicitation results in two multivariate Normal distributions $p_1(\theta)$ and $p_2(\theta)$ with means and covariance matrices $\mu_1, \mathbf{V}_1$ and $\mu_2, \mathbf{V}_2$ respectively, where

$$
\begin{aligned}
\mu_1 &= (15.03, 1.31, 530, 1587), & \mathbf{V}_1 &= diag(1.00, 0.01, 1, 0.50) & \text{and} \\
\mu_2 &= (5.01, 0.44, 177, 529), & \mathbf{V}_2 &= diag(2.00, 0.02, 2.00, 1.00).
\end{aligned}
$$

It is insightful to look at the distribution of logistic curves under these two prior distributions. Figure 1 contains logistic curves based on a random sample of

200 sets of parameter values from each of the two prior distributions. The plots show that there is a large number of different profiles (or shapes) that the 4PL curve can assume. Our aim is then to find a design that performs sufficiently well, for estimation purposes, for example, across these different profiles.
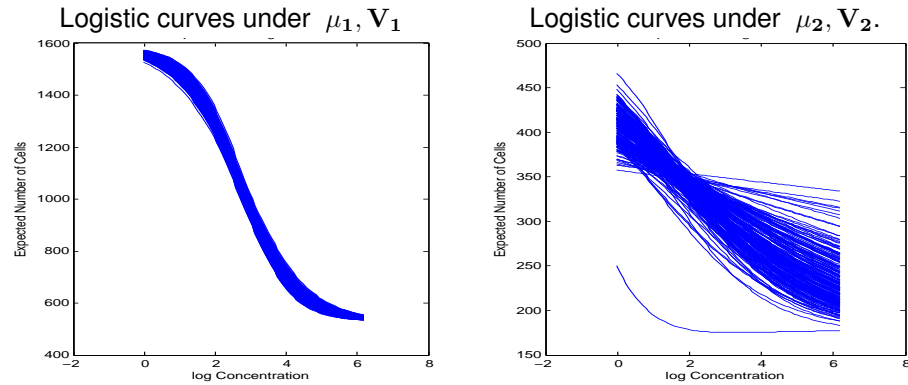

Figure 1: Distribution of a random sample of logistic curves under two prior parameter distributions.

The distribution of curves under $\mu_2, \mathbf{V_2}$ are more variable compared to those under $\mu_1, \mathbf{V_1}$ due to the relatively larger variability in the second prior. The objective of this section is to show that a design that is a function of the two information (or precision) matrices is more desirable than one that is based on exactly one of the prior distributions.

The following Bayesian D-optimal designs are obtained by maximizing the criterion in (12) using priors $p_1(\theta)$ and $p_2(\theta)$:

$$\xi_{D1} = \left\{ \begin{array}{cccc} 0.0338 & 1.9794 & 3.5168 & 6.1215 \\ 0.2726 & 0.2611 & 0.1943 & 0.2719 \end{array} \right\} \tag{16}$$

and

$$\xi_{D2} = \left\{ \begin{array}{cccc} 0.0379 & 1.5555 & 3.7501 & 6.1409 \\ 0.2152 & 0.1721 & 0.3496 & 0.2631 \end{array} \right\} \tag{17}$$

respectively. The proposed criterion in (13) is also maximized to obtain the robust design

$$\xi_C = \left\{ \begin{array}{cccc} -0.0138 & 1.8378 & 3.5650 & 6.1625 \\ 0.2416 & 0.2431 & 0.2801 & 0.2352 \end{array} \right\}. \tag{18}$$

To evaluate the performance of the robust design, we use relative efficiency. For two design measures ξ_1 and ξ_2, the relative D-efficiency of ξ_1 compared to ξ_2 is

$$D_{rel-eff}(\theta) = \left\{ \frac{|M(\xi_1,\theta)|}{|M(\xi_2,\theta)|} \right\}^{1/p}, \tag{19}$$

where p is the number of model parameters. Thus, $D_{rel-eff} > 1$ implies ξ_1 is more efficient than ξ_2 for estimating θ. In particular, we obtain the efficiency of ξ_C relative to locally optimal designs based on each of the prior distributions. Thus, we compute

$$D_{rel-eff}(\theta^1) = \left\{ \frac{|M(\xi_C,\theta^1)|}{|M(\xi_{\theta^1},\theta^1)|} \right\}^{1/p}, \tag{20}$$

where $\theta^1 \in p_1(\theta)$, and

$$D_{rel-eff}(\theta^2) = \left\{ \frac{|M(\xi_C,\theta^2)|}{|M(\xi_{\theta^2},\theta^2)|} \right\}^{1/p}, \tag{21}$$

where $\theta^2 \in p_2(\theta)$, and ξ_{θ^1} and ξ_{θ^2} are locally optimal designs at θ^1 and θ^2, respectively.
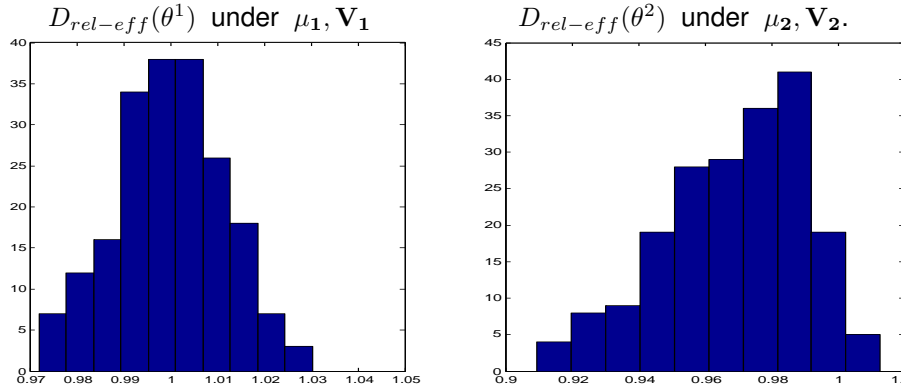

Figure 2: Empirical distribution of relative efficiencies of the robust design, ξ_C across two prior distributions.

In Figure 2, we provide plots of the distribution of relative efficiencies of ξ_C (based on a random sample of 200 sets of parameter values from each of the two prior distributions) across the two prior distributions. Numerical summaries of the these plots are also given in Table 1. The relative efficiencies are generally greater than 1 or within a small neighborhood of it, as shown in Table 1. This suggests the robustness of ξ_C to the two prior distributions.

Table 1: Numerical summaries of the empirical distribution of the relative frequency of the robust (or composite) design across the two priors.

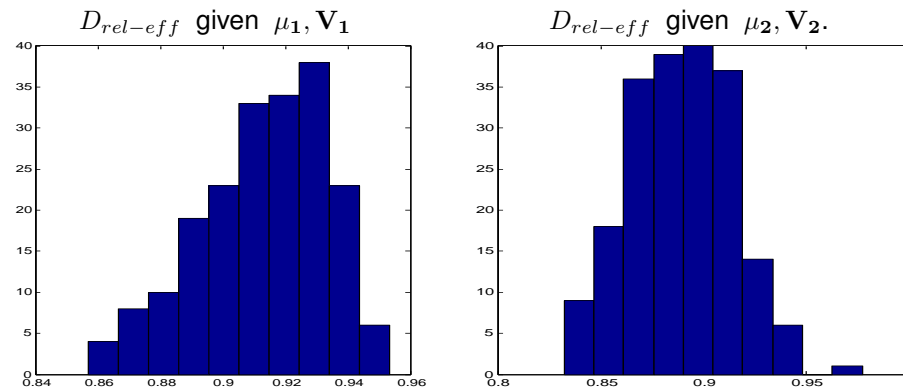| Summary	Relative D-efficiencies	
Statistic	$\mu_1, \mathbf{V}_1$	$\mu_2, \mathbf{V}_2$
Minimum	0.972	0.910
Maximum	1.030	1.013
Median	1.000	0.972
Mean	1.000	0.970

Figure 3: Empirical distribution of relative efficiency of the robust design relative to locally optimal designs across two prior distributions.

The essential features of the distributions in Figure 3 are summarized in Table 2 below where it is worth noting that the relative efficiencies are all less than 1. This is indicative of the sub-optimality of the ξ_{D1} for $p_2(\theta)$ and also ξ_{D2} for $p_1(\theta)$.

7. Discussion

Multiple prior distributions can occur when multiple experts are consulted or when different amounts of information are available to these experts. Variation in the training of experts (in a group) can also result in variable prior distributional assumptions of model parameters. We have shown in the previous section that when there are two prior distributions (or assumptions) about the parameters of a model, both priors should be accounted for in the design stage. A design that is optimal with respect to a particular prior is not necessarily optimal for another prior. As a result, it makes sense to design an experiment such that the resulting experimental design is a function of both information

Table 2: Numerical summaries of the distribution of relative efficiencies of the Bayesian optimal designs ξ_{D1} and ξ_{D2} on $p_2(\theta)$ and $p_1(\theta)$, respectively.

Summary	Relative D-efficiencies	
Statistic	ξ_{D1} on μ_2, V_2	ξ_{D2} on μ_1, V_1
Minimum	0.857	0.832
Maximum	0.954	0.977
Median	0.916	0.890
Mean	0.913	0.889

matrices. Ignoring one of the priors can lead to sub-optimal designs.

The criterion proposed here is an extension of the Bayesian optimality criterion when there is more than one prior. Essentially, it is a product of expectations. An intuition for the criterion can be gained by assuming, for example, that for some one-parameter nonlinear model, θ^* and θ^{**} are assumed to be the true parameter values by two different experts with similar training. In this case, we can think of two different locally optimal designs ξ^* and ξ^{**}, respectively, corresponding to the two expert guesses. If θ^* is actually the truth but θ^{**} is used in the design stage, then the resulting experimental design will be sub-optimal for θ^*, and vice versa. A reasonable approach in this situation is to obtain a design ξ that optimizes, for example, the product of information matrices corresponding to the two guesses. This idea is similar to that of product optimality suggested by Atkinson and Cox [44]. In the present context, instead of considering degenerate distributions like in the case of local optimality, we restrict our attention to the more practical situation of prior distributions and propose a criterion which is essentially the product of expected information matrices.

In many cases, expert opinions, usually in the form of probability distributions, are pooled or aggregated [45] with the view to obtaining a consensus distribution. Typically, non-negative weights w_i such that $\sum w_i = 1$ are used in the pooling process to reflect confidence in an expert's opinion. Weights can be seamlessly incorporated into the proposed criterion so that the resulting design has better efficiency over the prior with larger weight. In fact, the criterion in (13) belongs to a class of possible criteria for nonlinear models that will be discussed in a subsequent paper.

References

[1] Steinberg, David M. and Hunter, William G., Experimental Design: Review and Comment, *Technometrics*, 1984; 26(2): 71-97.

[2] Myers, R. H. and Montgomery, D. C. and Anderson-Cook, C.M., *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, J. Wiley, 2009.

[3] Reilly, M. and Salim, A., *Designing Optimal Two-Stage Epidemiological Studies*, John Wiley & Sons, Ltd, 2005.

[4] Fedorov, V. V. and Leonov, S. L., *Response-Driven Designs in Drug Development*, John Wiley & Sons, Ltd, 2005, 103-136.

[5] Atkinson, A. C. and Bailey, R. A., One Hundred Years of the Design of Experiments on and off the Pages of Biometrika, *Biometrika*, 2001; 88(1): 53-97.

[6] Chaloner, K. and Verdinelli, I., Bayesian Experimental Design: A Review, *Statistical Science*, 1995; 10(3): 273-304.

[7] Atkinson, A. C. and Chaloner, K. and Herzberg, A. M. and Juritz, J., Optimum Experimental Designs for Properties of a Compartmental Model, *Biometrics*, 1993; 49(2): 325-337.

[8] Kiefer, J., Optimum experimental designs (with discussion), *Journal of the Royal Statistical Society, Series B*, 1959; 21: 272-319.

[9] Berger, M.P.F. and Wong, W.K., *Applied Optimal Designs*, Wiley, New York, 2005.

[10] Smith, K., On the Standard Deviations of Adjusted and Interpolated Values of an Observed Polynomial Function and its Constants and the Guidance they give Towards a Proper Choice of the Distribution of Observations, *Biometrika*, 1918; 12: 1-85.

[11] Kiefer, J., and Wolfowitz, J., The equivalence of two extremum problems, *Canadian Journal of Mathematics*, 1960; 12: 363-366.

[12] Fedorov, V.V., Probability and mathematical statistics, *Theory of optimal experiments*, Academic Press, 1972.

[13] Pukelsheim, F., *Optimal Design of Experiments* , Wiley, 1993.

[14] Atkinson A. C., Donev A. N., and Tobias R., *Optimum Experimental Designs, With SAS*, Oxford University Press, 2007.

[15] St. John, R. C. and Draper, N. R., D-Optimality for Regression Designs: A Review, *Technometrics*, 1975: 17(1); 15-23.

[16] Pukelsheim, F., and Reider, S., Efficient rounding of approximate designs, *Biometrika*, 1992; 79: 763-770.

[17] Wald, A., On the Efficient Design of Statistical Investigations, *The Annals of Mathematical Statistics*, 1943; 14(2): 134-140.

[18] Kiefer, J. and Wolfowitz, J., Optimum Designs in Regression Problems, *The Annals of Mathematical Statistics*, 1959; 30(2): 271-294.

[19] Kiefer, J., Optimum Designs in Regression Problems, II, *The Annals of Mathematical Statistics*, 1961; 32(1): 298-325.

[20] Kiefer, J., *Optimum experimental designs V, with applications to rotatable designs*, Proceedings of the Fourth Berkeley Symposium on Mathematical Statist. and Probability, 1961; 381-405.

[21] Kiefer, J., General Equivalence Theory for Optimum Designs (Approximate Theory, *The Annals of Statistics*, 1974; 2(5): 849-879.

[22] Wu C., Some Algorithmic Aspects of the Theory of Optimal Designs, *The Annals of Statistics*, 1978; 6(6): 1286-1301.

[23] White, L. V., An Extension of the General Equivalence Theorem to Nonlinear Models, *Biometrika*, 1973; 60(2): 345-348.

[24] O'Brien, T. E., Optimal Design and Lack of Fit in Nonlinear Regression Models, *Statistical Modelling*, Springer New York, 201-206.

[25] Bates, D. M. and Watts, D. G., *Nonlinear regression analysis and its application*, John Wiley & Sons, 1988.

[26] Seber, G. A. F., and Wild, C. J., *Nonlinear regression*, Wiley, New York, 1989.

[27] Myers, R. H. and Montgomery, D. C. and Vining, G. G., *Generalized linear models: with applications in engineering and the sciences*, J. Wiley, 2002.

[28] Cochran, W. G., Experiments for Nonlinear Functions, *Journal of the American Statistical Association*, 1973; 68(344) : 771-781.

[29] Chernoff, H., Locally Optimal Designs for Estimating Parameters, *The Annals of Mathematical Statistics*, 1953; 24(4): 586-602.

[30] Box, G. E. P. and Lucas, H. L., Design of experiments in nonlinear situations, *Biometrika*, 1959; 46: 77-90.

[31] Ford, I. and Titterington, D. M. and Kitsos, C. P., Recent Advances in Nonlinear Experimental Design, *Technometrics*, 1989; 31(1): 49-60.

[32] Box, G .E. P. and Hunter, W. G., Sequential Design of Experiments for Nonlinear Models, *IBM Scientific Computing Symposium: Statistics*, 1965: 113-137.

[33] Chernoff, H., Sequential Design of Experiments, *The Annals of Mathematical Statistics*, 1959; 30(3): 755-770.

[34] DasGupta, A., *Review of optimal Bayes designs*, Purdue Univ, 1995.

[35] Lindley, D.V., *Bayesian Statistics, A Review*, Society for Industrial and Applied Mathematics, 1972.

[36] Shannon, C. E., A mathematical theory of communication, *Bell System Technical Journal*, 1948; 27: 379-243, 623-656.

[37] Chaloner K. and Larntz K., Optimal Bayesian design applied to logistic regression experiments, *J. Statist. Plann. Inference*, 1989; 21(2): 191-208.

[38] Toman, B. and Notz, W. I., Bayesian optimal experimental design for treatment - control comparisons in the presence of two-way heterogeneity, *Journal of Statistical Planning and Inference*, 1991; 27(1): 51 - 63.

[39] Tsai, C. and Chaloner, K., Using Prior Opinions to Examine Sample Size in Two Clinical Trials, *Case Studies in Bayesian Statistics*, 2002; 162: 407-421.

[40] DasGupta, A. and Studden, W. J., Robust Bayesian Experimental Designs in Normal Linear Models, *The Annals of Statistics*, 1991; 19(3): 1244-1256.

[41] Toman, B. Bayesian robust experimental designs for the one-way analysis of variance, *Statistics & Probability Letters*, 1992; 15(5): 395 - 400.

[42] Toman B. and Gastwirth J. L., Robust Bayesian experimental design and estimation for analysis of variance models using a class of normal mixtures, *Journal of Statistical Planning and Inference*, 1993; 35(3): 383 - 398.

[43] Dasgupta A. and Mukhopadhyay S. and Studden W. J., Compromise designs in heteroscedastic linear models, *Journal of Statistical Planning and Inference*, 1992; 32(3): 363 - 384.

[44] Atkinson, A. C. and Cox, D. R., Planning Experiments for Discriminating Between Models, *Journal of the Royal Statistical Society. Series B*, 1974; 36(3): 321-348.

[45] Genest, C. and Zidek, J. V., Combining Probability Distributions: A Critique and an Annotated Bibliography, *Statistical Science*, 1986; 1(1): 114-135.