Journal of Statistical Planning and Inference 144 (2014) 173–187

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Information in a two-stage adaptive optimal design

Adam Lane [a], Ping Yao [b], Nancy Flournoy [c,*]

[a] Cancer and Blood Diseases Institute, 3333 Burnet Avenue, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA
[b] School of Nursing and Health Studies, 253 Wirtz Hall, Northern Illinois University, DeKalb, IL 60115, USA
[c] Department of Statistics, 146 Middlebush Hall, University of Missouri, Columbia, MO 65211, USA

ARTICLE INFO

Available online 27 July 2013

Keywords:
Bioassay
Dose-finding
Experimental design
Fisher's information
Observed information
Optimal design
Phase I
Response-adaptive design

ABSTRACT

In *adaptive optimal designs*, each stage uses a locally optimal design evaluated at the maximum likelihood estimates derived using cumulative data from all prior stages. This dependency on prior stages affects Fisher's information, the asymptotic covariance matrix of the maximum likelihood estimates. Fisher's information is motivated for use in adaptive designs with small samples by deriving the Cramèr–Rao lower bound for such experiments. Then the usefulness of Fisher's information is shown from both a design and an analysis perspective. From a design perspective, the locally optimal stage one sample size is defined in terms of Fisher's information and a procedure to approximate it is suggested. From an analysis perspective, Fisher's information is compared to a commonly used information measure derived by ignoring the stage dependencies and to the observed information. To make the analysis explicit, a two stage design with fixed first stage is examined in the context of a general nonlinear regression model.

1. Introduction

Adaptive designs are popular for phase I dose-finding clinical trials because they may be directed toward an experimental goal. This goal typically lies on a spectrum between ethically treating patients enrolled in the study and maximizing the amount of information collected. This trade off is often referred to as the treatment versus experimentation dilemma; see, for example, Bartroff and Lai (2010), Baldi Antognini and Giovagnoli (2010) and Azriel et al. (2011). Examples of designs that attempt to maximize the ethical treatment of enrolled patients can be found in Li et al. (1995), Whitehead and Williamson (1998) and Thall and Cook (2004). These types of designs remain popular despite examples that lead to inconsistent estimates of the model parameters in Lai and Robbins (1982), Pronzato (2000), Chang and Ying (2009), Oron et al. (2011) and Azriel (2012).

The purpose of this exposition is to examine designs at the other end of the spectrum which use classical methods from the theory of optimal design to produce precise experiments, where precision is defined in terms of a design and the primary experimental concern. A design is denoted $\xi=\{w_i,x_i\}_1^K$, where x_i is the ith treatment, which in general may be a point in multidimensional space, and w_i is the proportion of observations allocated to x_i. We call w_i a design weight and $\sum w_i=1$. The design of a precise experiment minimizes, with respect to ξ, some concave function of the covariance matrix of the model parameter estimates. The concave function is determined by the primary experimental concerns. For example, when efficient parameter estimation is desired, the D-optimality criterion (the inverse of the determinant of Fisher's information) is commonly recommended. For examples of different optimality criteria on and their corresponding concave

* Corresponding author. Tel.: +1 5738826376.
E-mail addresses: adam.lane@cchmc.org (A. Lane), pyao@niu.edu (P. Yao), flournoyn@missouri.edu (N. Flournoy).

0378-3758/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jspi.2013.07.010

functions see Pukelsheim (2006). When the underlying model is linear, precise experiments for a given concave function may be attained because Fisher's information is independent of the model parameters.

There is a wealth of literature on optimal designs for linear models (cf. Fedorov, 1972; Silvey, 1980; Atkinson et al., 2007). However, when the underlying model is nonlinear, Fisher's information matrix, and as a result optimal designs, will depend on the model parameters. This dependence represents a major challenge for the implementation of optimal designs in nonlinear models. The term *locally optimal design* is often used to indicate that optimal designs in nonlinear models are optimal only in the neighborhood of the true parameters. There have been many suggestions on how to deal with the locally optimal design problem. Fisher (1947, Chapter 68) and Chernoff (1953) suggest that optimal designs be approximated by guessing the parameter values; however, this method may be inefficient when the guess is far from the true parameter values. Ford et al. (1992) argue that this procedure of guessing provides an appropriate benchmark by which to gauge the performance of alternate methods. Kitsos et al. (1988) suggest the use of a non-optimal design that has the property of being insensitive to the true parameter values. Dette and Sahm (1998) develop a minimax optimal design for use in nonlinear models. For a review of methods for nonlinear models see Ford et al. (1989) and O'Brien and Funk (2003).

Others have suggested adaptive methods. Atkinson et al. (2007, Chapter 17) suggest a sequential procedure where the model is linearized using an expansion; then the optimal design of the approximate linear model is used in the first stage and updated for consequent stages. Haines et al. (2003) develop a sequential Bayesian optimal design procedure. Many researchers, including Box and Hunter (1965), Fedorov (1972), White (1975) and Silvey (1980) advocate what we refer to as adaptive optimal designs. An *adaptive optimal design* is a procedure in which the first stage is initialized, using expert opinion or prior data. Then each successive stage is allocated according to the estimated optimal design obtained using all the data from the previous stages. Recently Dragalin and Fedorov (2005), Dragalin et al. (2008, 2007) among others have proposed and analyzed such designs. For a comparison and further discussion of designs from both ends of the treatment/experimental spectrum see Fedorov et al. (2012a).

One issue present in the current adaptive optimal design literature is that in place of constructing a likelihood from the joint density of responses and support points, responses have been treated as independent conditional on the design. Silvey (1980) and others point out that the information employed is not by definition Fisher's information. To make this issue explicit in Section 3 we develop Fisher's information for a two-stage adaptive optimal design.

We attempt to clarify the benefit of the unconditional Fisher information from two different perspectives. First, from a design perspective we define the *locally optimal stage one sample size* for a two-stage adaptive experiment in Section 4. Then we propose a method to approximate the locally optimal stage one sample size when the parameter is unknown. Second, in Section 5 from an analysis perspective we examine the effect of using Fisher's information against commonly used alternatives. The alternatives we examine are an approximation based on the information measure derived under conditional independence and the observed information. In Section 6 a simulation is done to compare the performance of the different information measures and their estimates.

Throughout this exposition we assume for simplicity that there are only two stages and that the first stage treatment (support point) is fixed. Responses are assumed to be normal with a nonlinear mean function. It is convenient for our narrative to use such a simple set-up with the understanding that lessons learned apply to more complex scenarios.

2. Information bound in a two-stage experiment

Attainment of the Cramèr–Rao lower bound provides a small sample justification for the use of Fisher's information in the optimum design of experiments in linear models. We now consider this argument from the viewpoint of a two-stage adaptive design for a nonlinear model.

In the first stage, a vector of independent responses, $\boldsymbol{y_1}$, from a distribution containing a single parameter θ is observed from n_1 subjects at a fixed treatment level x_1. To determine the second stage treatment, a deterministic onto function of the first stage data is used, *i.e.*, $x_2 = x_2(x_1, \boldsymbol{y_1})$. Then a vector of responses, $\boldsymbol{y_2}$, is observed from n_2 subjects at the adapted point, x_2. Note the vectors $\boldsymbol{y_1}$ and $\boldsymbol{y_2}$ are composed of independent observations, but are not independent of one another. Rather it is assumed that $\boldsymbol{y_1}$ and $\boldsymbol{y_2}$ are from a joint density

$$f_{\boldsymbol{y_1},\boldsymbol{y_2}|x_1}(\boldsymbol{y_1},\boldsymbol{y_2}|x_1,\theta) = f_{\boldsymbol{y_2}|\boldsymbol{y_1},x_1}(\boldsymbol{y_2}|\boldsymbol{y_1},x_1,\theta) f_{\boldsymbol{y_1}|x_1}(\boldsymbol{y_1}|x_1,\theta), \tag{1}$$

which is bounded and twice differentiable with respect to $\theta \in \Theta$, where θ is an interior point of Θ.

Let $\tilde{\theta}_n$ be an estimator of θ based on the $n = n_1 + n_2$ total subjects from stage one and stage two, with finite expectation $\mathrm{E}[\tilde{\theta}_n] = \theta + b(x_1, \theta)$. The following derivation of the information inequality for adaptive experiments is based on the derivations in Cox and Hinkley (1974, p. 254) and Hogg et al. (2005, p. 322). Let $S = \partial \log f_{\boldsymbol{y_1},\boldsymbol{y_2}|x_1}(\boldsymbol{y_1},\boldsymbol{y_2}|x_1,\theta)/\partial\theta$ denote the score function. Then

$$\mathrm{Cov}\left[\tilde{\theta}_n, S\right] = \mathrm{E}\left[\tilde{\theta}_n S\right] = \mathrm{E}\left[\tilde{\theta}_n \frac{\frac{\partial}{\partial\theta} f_{\boldsymbol{y_1}\boldsymbol{y_2}|x_1}}{f_{\boldsymbol{y_1}\boldsymbol{y_2}|x_1}}\right] = \frac{\partial}{\partial\theta}\,\mathrm{E}\left[\tilde{\theta}_n\right] = 1 + \frac{\partial}{\partial\theta} b(x_1,\theta).$$