Adaptive Procedures for Optimum Observed Information

Adam Lane*
Cincinnati Children's Hospital Medical Center, Cincinnati, USA - Adam.Lane@cchmc.org

Abstract

Fisher expected information can be found *a priori* and as a result its inverse is the primary variance approximation used in the design of experiments. However, it is more common to use the inverse of Fisher observed information for the analysis of experiments. The full Fisher observed information cannot be known *a priori*, however, if an experiment is conducted sequentially (in a series of runs) then the Fisher observed information from previous runs is available. In the current work an adaptive procedure is proposed that uses the Fisher observed information from previous runs to design current and future runs. The objective of this procedure is to optimize the observed efficiency of Fisher observed information from the entire experiment. By optimizing Fisher observed information instead of Fisher expected information it provides a better connection between the design and the analysis of the experiment. This procedure is examined in the context of a Gamma Hyperbola regression model. In this simulated example the OM following the adaptive procedure had greater efficiency compared to well established designs.

Keywords: experimental design; Fisher information; efficiency.

1 Introduction

Fisher observed information (OM) and Fisher expected information (FIM) are measures of the accuracy of estimates of the unknown model parameters. Large values indicate more information corresponding to smaller variability associated with these estimates. In most experimental settings FIM can be found *a priori* and as a result in the optimal design of experiments, in maximum likelihood estimation, an intuitive goal is to "maximize" this quantity. When the maximum likelihood estimate (MLE) is a minimal sufficient statistic FIM contains all the information in the data and is suitable for use in the analysis of the experiment. However, when the MLE is not minimally sufficient there is information loss in FIM.

Information loss can be recovered, at least in part, by considering the distribution of the MLE, $\hat{\mu}$, conditional on an ancillary random variable, a. An ancillary random variable is any variable with a distribution that does not depend on the model parameters. In order for information to be recovered it is required that a be a function of the minimal sufficient statistic. The objective of the current work is to develop an adaptive procedure that uses the OM from previous runs to inform the designs of the current run, in cases where $(\hat{\mu}, a)$ is the minimal sufficient statistic. The use of OM to define designs is motivated by Efron and Hinkley (1978) where it is shown that the inverse of OM is more closely related to $\text{Var}[\hat{\mu}|a]$ than the inverse of FIM.

Adaptive designs have been proposed and examined in optimal designs where FIM depends on the model parameters, see Dragalin and Fedorov (2005), Lane, Yao, and Flounoy (2014), among others. In such cases the dependence of FIM on the model parameters results in optimal designs that are locally optimal, i.e., designs are only optimal neighborhood of the true parameters. The majority of the literature with adaptive procedures for optimal designs use the MLE based on previous stages to determine locally optimal designs for the current stage. While we develop our procedure for a general model that includes cases where FIM depends on the model parameters it is not required for our procedure to be used.

2 Model

An exact design, ξ_n, is represented by a set of design points $\{x_1, x_2, \ldots, x_d\}$ and corresponding weights $\{n_1/n, n_2/n, \ldots, n_d/n\}$, where each x_i is a point in the design region, $\mathscr{X}$; n_i is the sample size at each design point and; the total sample size $n = \sum n_i$.

Consider a set of independent responses $\boldsymbol{y} = (y_1, \ldots, y_n)$ with density $f(y|\mu_0)$, where $\mu_0 = \mu(x, \boldsymbol{\theta_0})$ is a scaler, $\boldsymbol{\theta_0}$ is a p dimensional vector of the unknown but true parameters, $x \in \mathscr{X}$ is a q dimensional vector within the design space $\mathscr{X}$ with $q \leq p$. Let $\mathcal{D}_n = \mathcal{D}_n(\xi_n)$ be all available data collected from an experiment of size n and note that it is a function of the exact design.

Let $\boldsymbol{y_i} = (y_{i1}, \ldots, y_{in_i})$ be the responses obtained from observations with covariate equal to x_i,

$$u(\boldsymbol{y_i}, \mu_i) = -\sum_{k=1}^{n_i} \frac{\partial^2 \log f(y_{ik}|\mu_i)}{\partial \mu_i^2} \quad \text{and} \quad v(\mu_i) = E[u(\boldsymbol{y_i}, \mu_i)].$$

Note $u(\boldsymbol{y_i}, \mu_i)$ and $v(\mu_i)$ are scalars. The OM for $\boldsymbol{\mu} = \{\mu_1, \ldots, \mu_d\}$, $I_{\hat{\boldsymbol{\mu}}}(\mathcal{D}_n)$, is a diagonal matrix with entry $I_{\hat{\mu}_i}(\mathcal{D}_n) = u(\boldsymbol{y_i}, \hat{\mu}_i)$ on the ith diagonal. The FIM is $\mathscr{F}_{\boldsymbol{\mu}}$ is a diagonal matrix with $\mathscr{F}_{\mu_i} = v(\mu_i)$ on the ith diagonal.

Now write $u(\boldsymbol{y_i}, \mu_i) = u(\boldsymbol{y_i}, x_i, \boldsymbol{\theta})$ and $v(\mu_i) = v(x_i, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is an arbitrary point in the parameter space, $\boldsymbol{\Theta}$. Further let $V(x_i, \boldsymbol{\theta})$ be a matrix with jkth element equal to

$$V_{jk}(x_i, \boldsymbol{\theta}) = E\left[\frac{\partial^2 \log f\{y|\mu(x_i, \boldsymbol{\theta})\}}{\partial \theta_j \theta_k}\right] = v(x_i, \boldsymbol{\theta}) \frac{\partial \mu(x_i, \boldsymbol{\theta})}{\partial \theta_j} \frac{\partial \mu(x_i, \boldsymbol{\theta})}{\partial \theta_k}.$$

and $U(\boldsymbol{y_i}, x_i, \boldsymbol{\theta})$ be a matrix with jkth element equal to

$$U_{jk}(\boldsymbol{y_i}, x_i, \boldsymbol{\theta}) = u(\boldsymbol{y_i}, x_i, \boldsymbol{\theta}) \frac{\partial \mu(x_i, \boldsymbol{\theta})}{\partial \theta_j} \frac{\partial \mu(x_i, \boldsymbol{\theta})}{\partial \theta_k}.$$

Further let

$$I_\theta(\mathcal{D}_n) = \sum_{i=1}^{d} U(\boldsymbol{y_i}, x_i, \boldsymbol{\theta}) \quad \text{and} \quad \mathscr{F}_{\boldsymbol{\theta}}(\xi_n) = \sum_{i=1}^{d} n_i V(x_i, \boldsymbol{\theta}).$$

Then OM and FIM for $\boldsymbol{\theta}$ are $I_{\hat{\boldsymbol{\theta}}}(\mathcal{D}_n)$ and $\mathscr{F}_{\boldsymbol{\theta}}(\xi_n)$, respectively. In most cases OM is thought of as a post experiment measure, i.e., $\hat{\boldsymbol{\theta}}$ has been obtained following an experiment and $I_{\boldsymbol{\theta}}(\mathcal{D}_n)$ can be directly evaluated at the observed $\hat{\boldsymbol{\theta}}$. However, for our purposes we will at times consider a pre-experiment version of OM, $I_{\boldsymbol{\theta}}(\mathcal{D}_n)$, where $\boldsymbol{\theta}$ is some pre-experimental guess of the true parameter. This is of interest since it is analogous to what is done in traditional optimal design, details given in Section 2.1, where some initial guess of $\boldsymbol{\theta}$ is used to determine locally optimal designs.

2.1 Traditional Optimal Design

The requirement that all n_i are integers is restrictive in the search for optimal designs. The concept of continuous designs relaxes the restriction such that design points have corresponding weight w_i where $0 \leq w_i \leq 1$ and $\sum w_i = 1$. A continuous design, ξ, is a probability measure on $\mathcal{X}$ and is a member of the set, Ξ, which represents all measures defined on the Borel Field generated by the open sets of $\mathcal{X}$ such that $\int_{\mathcal{X}} \xi(dx) = 1$.

For a continuous design, ξ, the per-subject information matrix is

$$M(\xi, \theta) = \int_{\mathcal{X}} V(x, \boldsymbol{\theta}) \xi(dx) = \sum_{i=1}^{d} w_i V(x_i, \boldsymbol{\theta}).$$

The exact analog of the per-subject information is $M(\xi_n, \boldsymbol{\theta}) = \frac{1}{n} \mathscr{F}_{\boldsymbol{\theta}}(\xi_n)$.

We use the case of D-optimal design to describe traditional optimal design and the adaptive procedures. The traditional optimal design as well as the adaptive procedures do not depend on the D-optimality criterion and can readily be developed for other traditional criteria. A locally D-optimal design $\xi^*_{\boldsymbol{\theta}}$ is defined as

$$\xi^*_{\boldsymbol{\theta}} = \arg\max_{\xi \in \Xi} \log |M(\xi, \boldsymbol{\theta})|. \quad (1)$$

When $M(\xi, \boldsymbol{\theta})$ depends on the parameter θ we refer to this as a locally D-optimal design since, $\xi^*_{\boldsymbol{\theta}}$, is a function of $\boldsymbol{\theta}$. If $M(\xi, \boldsymbol{\theta})$ does not depend on $\boldsymbol{\theta}$ it is a globally optimal design.

From the General Equivalence Theorem for nonlinear models in White (1973) it is known that a design, $\xi^*_{\boldsymbol{\theta}}$, is D-optimal if and only if

$$d(x, \xi^*_\theta, \boldsymbol{\theta}) = \text{tr}\left\{V(x, \boldsymbol{\theta}) M^{-1}(\xi^*_{\boldsymbol{\theta}}, \boldsymbol{\theta})\right\} \leq p$$

for all $x \in \mathcal{X}$ with equality if and only if x is a support of the optimal design. The function $d(x, \xi^*, \boldsymbol{\theta})$ is referred to as the sensitivity function and can be used to verify if a design is optimal. The procedures we propose do not attempt to find this optimal design. It is expected that these designs can be found using existing methods and are therefore known. Throughout we refer to $\xi^*_{\boldsymbol{\theta}}$ as the locally optimal design satisfying (1) for a fixed $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. The the design points and corresponding weights in $\xi^*_{\boldsymbol{\theta}}$ will be denoted x_i^* and w_i^*, respectively, for $i = 1, \ldots, d$.

2.2 Benchmark for Observed Information

In traditional optimal design exact design, ξ_n, are often assessed using D-efficiency, defined as

$$D_{\text{eff}} = \left\{ \frac{|M(\xi, \boldsymbol{\theta})|}{|M(\xi^*, \boldsymbol{\theta})|} \right\}^{1/p}.$$

In the current setting the same design can lead to different realized OM. If the OM from one experiment is greater than the OM from another experiment it does not necessarily indicate a superior design but instead may indicate more fortunate observed data. Instead we introduce an observed version of D-efficiency defined below.

Definition 2.1. *For any observed data $\mathcal{D}_n$, the observed D_c-efficiency is*

$$D_{\textit{c-obs-eff}}\{\mathcal{D}_n\} = \left\{ \frac{|I(\mathcal{D}_n)|}{|cM(\xi^*, \boldsymbol{\theta})|} \right\}^{1/p}.$$

For observed D_c-efficiency to be useful c must be an informative benchmark or lower bound on the OM. Consider the following:

Theorem 2.1. *For any observed data $\mathcal{D}_n$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ let*

$$q_{\boldsymbol{\theta}}(\mathcal{D}_n) = \sum_{i=1}^{d} \frac{u(\boldsymbol{y_i}, x_i, \boldsymbol{\theta})}{v(x_i, \boldsymbol{\theta})}, \quad \omega_i = \frac{1}{q_{\boldsymbol{\theta}}(\mathcal{D}_n)} \frac{u(\boldsymbol{y_i}, x_i, \boldsymbol{\theta})}{v(x_i, \boldsymbol{\theta})}$$

for $i = 1, \ldots, d$ and

$$\tau = \left\{ \begin{array}{cccc} x_1 & x_2 & \ldots & x_d \\ \omega_1 & \omega_2 & \ldots & \omega_d \end{array} \right\}.$$

If $\tau \in \Xi$ then

$$\log |I_{\boldsymbol{\theta}}(\mathcal{D}_n)| \le \log |q_{\boldsymbol{\theta}}(\mathcal{D}_n) M_{\boldsymbol{\theta}}(\xi^*_{\boldsymbol{\theta}}, \boldsymbol{\theta})|.$$

Proof. We can write

$$\begin{aligned} \log \left| \frac{1}{q_{\boldsymbol{\theta}}(\mathcal{D}_n)} I_{\boldsymbol{\theta}}(\mathcal{D}_n) \right| &= \log \left| \frac{1}{q_{\boldsymbol{\theta}}(\mathcal{D}_n)} \sum_{i=1}^{d} \left[\sum_{k=1}^{n_i} u(y_k, x_i, \boldsymbol{\theta}) \right] \left[\frac{\partial^2 \mu(x_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} \right] \left[\frac{\partial^2 \mu(x_i, \boldsymbol{\theta})}{\partial \theta^2} \right]^T \right| \\ &= \log \left| \sum_{i=1}^{d} a_i v(x_i, \boldsymbol{\theta}) \left[\frac{\partial^2 \mu(x_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} \right] \left[\frac{\partial^2 \mu(x_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} \right]^T \right| \\ &= \log |M(\tau, \boldsymbol{\theta})| \le \log |M(\xi^*_{\boldsymbol{\theta}}, \boldsymbol{\theta})| \end{aligned}$$

where the inequality is a direct application of the General Theorem in White (1973) since $\tau \in \Xi$. □

Theorem 2.1 shows that using $c = q_{\boldsymbol{\theta}}(\mathcal{D}_n)$ in Definition 2.1 compares the efficiency of $I(\mathcal{D}_n)$ to a lower bound $q_{\boldsymbol{\theta}}(\mathcal{D}_n) M(\xi^*, \boldsymbol{\theta})$.

A useful assessment of two OM from two different experiments is what we will refer to as observed relative efficiency

Definition 2.2. *For any two sets of observed data, $\mathcal{D}_{1n}$ and $\mathcal{D}_{2n}$, the observed relative efficiency is*

$$D_{\textit{obs-rel-eff}}(\mathcal{D}_{1n}, \mathcal{D}_{2n}) = \left\{ \frac{|q_\theta(\mathcal{D}_{2n}) I_\theta(\mathcal{D}_{1n})|}{|q_\theta(\mathcal{D}_{1n}) I_\theta(\mathcal{D}_{2n})|} \right\}^{1/p}.$$

3 Adaptive Procedure in D-Optimal Designs

Consider an experiment with J sequential runs. Each run has m_j independent observations $j = 1, \ldots, J$. Denote the total sample size up to and including the jth run as $M_j = \sum_{k=1}^{j} m_k$ and note $n = M_J$. To frame the problem in a similar fashion as traditional optimal design denote the exact and continuous design of the jth run as

$$\lambda_{jm_j} = \left\{ \begin{array}{cccc} x_{1j} & x_{2j} & \ldots & x_{d_jj} \\ m_{1j}/m_j & m_{2j}/m_j & \ldots & m_{d_jj}/m_j \end{array} \right\} \quad \text{and} \quad \lambda_j = \left\{ \begin{array}{cccc} x_{1j} & x_{2j} & \ldots & x_{d_jj} \\ w_{1j} & w_{2j} & \ldots & w_{d_jj} \end{array} \right\},$$

respectively, where x_{ij}, $i = 1, \ldots, d_j$ are the design points at run j, $m_j = \sum_{i=1}^{d_j} m_{ij}$ and $\sum_{i=1}^{d_j} w_{ij} = 1$. Here λ_j is a probability measure on $\mathcal{X}$ and is a member of the set, Λ_j, which represents all measures defined on the Borel Field generated by the open sets of $\mathcal{X}$ such that $\int_{\mathcal{X}} \lambda_j(dx) = 1$.

In the context of an adaptive procedure $\xi_{M_j} = \sum_{k=1}^{j} \lambda_{km_k}$ represents the exact design from the first j runs. Let $\mathcal{D}_{M_j} = \mathcal{D}(\xi_{M_j})$ be all available data collected from the first j runs and note that it is a function of the exact design ξ_{M_j}.

A meaningful quantity, around which the proposed adaptive procedure will be based, given the observed data from the previous $j-1$, $\mathcal{D}_{M_{j-1}}$, is

$$K_{\boldsymbol{\theta}}(\lambda_j, \mathcal{D}_{M_{j-1}}) = m_2 \sum_{i=1}^{d_j} w_{ij} V(x_{ij}, \boldsymbol{\theta}) + I_{\boldsymbol{\theta}}(\mathcal{D}_{M_{j-1}}) = m_2 M(\lambda_j, \boldsymbol{\theta}) + I_{\boldsymbol{\theta}}(\mathcal{D}_{M_{j-1}}). \tag{2}$$

We define a continuous design, λ_j^*, to be $D(\mathcal{D}_{M_{j-1}})$-optimal if

$$\lambda_j^*(\theta) = \arg \max_{\lambda_j \in \Lambda_j} \log |K_{\boldsymbol{\theta}}(\lambda_j, \mathcal{D}_{M_{j-1}})|$$

We write this $D(\mathcal{D}_{M_{j-1}})$-optimal since the optimal design for run j will depend on the observed data from the previous $j-1$ stages. For this set-up the following general equivalence theorem can be shown to hold.

Theorem 3.1. *For all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ the following are equivalent*

1. λ_j^* *is* $D(\mathcal{D}_{M_{j-1}})$*-optimal*
2. $tr\left[V(x, \boldsymbol{\theta})\{M_{\alpha_j}(\lambda_j^*, \boldsymbol{\theta})\}^{-1}\right] \leq tr\left[M(\lambda_j^*, \boldsymbol{\theta})\{M_{\alpha_j}(\lambda_j^*, \boldsymbol{\theta})\}^{-1}\right]$
3. $d_K(x, \lambda_j^*, \boldsymbol{\theta}) \leq p$

for all $x \in \mathcal{X}$, with equality if and only if x is a support point of λ_j^*, where $\alpha_j = m_j/M_j$,

$$d_K(x, \lambda_j, \boldsymbol{\theta}) = \alpha_j tr\left[V(x, \boldsymbol{\theta})\{M_{\alpha_j}(\lambda_j, \boldsymbol{\theta})\}^{-1}\right] + \frac{1}{M_j} tr\left[I(\mathcal{D}_{M_{j-1}})\{M_{\alpha_j}(\lambda_j, \boldsymbol{\theta})\}^{-1}\right]$$

and

$$M_{\alpha_j}(\lambda_j, \boldsymbol{\theta}) = \alpha_j M(\lambda_j, \boldsymbol{\theta}) + \frac{1}{M_j} I(\mathcal{D}_{M_{j-1}}),$$

for $j = 2, \ldots, J$.

The proof of Theorem 3.1 is a direct application of Theorem 11.6 and Lemma 6.16 from Pukelsheim (1993). The results from Pukelsheim (1993) do not require D-optimality and thus Theorem 3.1 can easily be adapted for other common optimality criteria.

From Theorem 3.1 we define an adaptive procedure for the OM quantity in (2) as

1. Initiate the design by placing the first run at a design that ensures $I(\mathcal{D}_{M_1})$ is nonsingular. Ideally 1 observation at each of the d optimal design points in ξ^*.
2. For runs $j = 2, \ldots, J$ allocate patients according to design λ_j^* or $\lambda_{jm_j}^*$.

4 Example: Gamma Hyperbola Regression (GHR)

Here we develop a regression model based on the bivariate Gamma Hyperbola model. This model is characterized by two independent random variables $(s,t) = (z_1 e^{\mu}, z_2 e^{-\mu})$ observed simultaneously, where $z_k \sim Gamma[\beta, 1]$, $k = 1, 2$, where β is known.

For the development of the model and relative quantities we assume we have some fixed design ξ_n. In a regression setting, with $\mu(x, \boldsymbol{\theta}) = \boldsymbol{\theta}^T \boldsymbol{f_x}(x)$, two independent random variables

$$(s_j, t_j) = (z_{1j} e^{\mu(x_{ij}, \boldsymbol{\theta})}, z_{2j} e^{-\mu(x_{ij}, \boldsymbol{\theta})}),$$

$j = 1, \ldots, n$ are observed simultaneously. We refer to (s_j, t_j) as a pair since they are observed simultaneously, however, we stress that they are uncorrelated. Let $n_i = \sum_{k=1}^{n} I(x_{ij} = x_i)$, $S_i = \sum_{k=1}^{n} I(x_{ij} = x_i) s_k$, $T_i = \sum_{k=1}^{n} I(x_{ij} = x_i) t_k$ and $r_i = \sqrt{S_i T_i}$ then

$$g(r_i) = \frac{2 r_i^{2 n_i \beta - 1}}{\Gamma[n_i \beta]^2} \int_{-\infty}^{\infty} e^{-2 r_i \cosh(u_i)} du_i.$$

for $i = 1, \ldots, d$. Note $\hat{\mu}_i = \frac{1}{2} Log[S_i / T_i]$ is the MLE of $\mu(x_i, \boldsymbol{\theta})$. Further let $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, \ldots, \hat{\mu}_d)$ and $\boldsymbol{r} = (r_1, \ldots, r_d)$ then joint distribution of $(\hat{\boldsymbol{\mu}}, \boldsymbol{r})$

$$f(\hat{\boldsymbol{\mu}}, \boldsymbol{r}) = e^{-\sum_{i=1}^{d} 2 r_i \cosh\{\hat{\mu}_i - \mu(x_{ik}, \boldsymbol{\theta})\}} \prod_{i=1}^{d} \frac{2 r_i^{2 n_i \beta - 1}}{\Gamma[n_i \beta]^2}$$

From the factorization theorem of sufficiency we can see that the statistic $(\hat{\boldsymbol{\mu}}, \boldsymbol{r})$ is a sufficient statistic for $\boldsymbol{\mu}$. Further the conditional distribution of $\hat{\boldsymbol{\mu}}$ given $\boldsymbol{r}$ is

$$f(\hat{\boldsymbol{\mu}} | \boldsymbol{r}) = \frac{e^{-2 \sum_{i=1}^{d} r_i \cosh\{\hat{\mu}_i - \mu(x_i, \boldsymbol{\theta})\}}}{\prod_{i=1}^{d} \int_{-\infty}^{\infty} e^{-2 r_i \cosh(u_i)} du_i}.$$

For the GHR model in the sequential setting suppose there are J total runs. Within each run there are $2m_j$ observations. For $k = 1, \ldots, m_j$ two observations are observed simultaneously at the same x_{jk}, where the value of x_{jk} is at the experimenters discretion. Denote the kth set of observations within the jth run as (s_{jk}, t_{jk}) for $k = 1, \ldots, m_j$ and $j = 1, \ldots, J$. Let $S_{ij} = \sum_{l=1}^{j} \sum_{k=1}^{m_l} I(x_{lk} = x_i) s_{lk}$, $T_{ij} = \sum_{l=1}^{j} \sum_{k=1}^{m_l} I(x_{lk} = x_i) t_{lk}$ be the sums of the observations from all runs up to and including run j. Further let $r_{ij} = \sqrt{S_{ij} T_{ij}}$. The OM for the sequential experiment is

$$I(\mathscr{D}_j) = 2 \sum_{i=1}^{d} \boldsymbol{f_x}(x_i) \boldsymbol{f_x}(x_i)^T r_{ij}$$

for $j = 1, \ldots, J$.

Fisher expected information can be easily derived as $M(\xi) = 2\beta \sum_{i=1}^{d} w_i \boldsymbol{f_x}(x_i) \boldsymbol{f_x}^T(x_i)$. This is convenient since the optimal design is equivalent to the optimal design in the linear model case.

4.1 Simulation

The simulation consisted of 10,000 iterations from a GHR model with $\boldsymbol{\theta} = (1, 1, 1)$, $\boldsymbol{f_x}(x) = (1, x, x^2)$, $\mathcal{X} = [-1, 1]$ and $\beta = 1$. For this model the continuous optimal design does not depend on $\boldsymbol{\theta}$, i.e., $\xi^*_{\boldsymbol{\theta}} = \xi^*$, and consists of three points $\{-1, 0, 1\}$ each with a weight of 1/3. Let $\mathscr{D}_n(K)$ and $\mathscr{D}_n(\xi^*)$ be the observed data from experiments conducted using the adaptive procedure with and the optimal design ξ^*, respectively. Let $\hat{\boldsymbol{\theta}}_K$ and $\hat{\boldsymbol{\theta}}_{\xi^*}$ be defined similarly.

Different values of β and n were used with $m_j = 1$ for all j. In the simulation all sample sizes n are multiples of p. This ensures that the exact optimal design is equivalent to the continuous optimal design. Results are presented in Table 1. Consider first rows corresponding to the observed efficiency, $D_{q\{\mathscr{D}_n\}-obs-eff}$. Data are presented as 10th, 25th, 50th, 75th and 90th percentiles. In every case and for every percentile the observed efficiency of $\mathscr{D}_n(K)$ was superior to the observed efficiency of $\mathscr{D}_n(\xi^*)$. The results are the most significant for smaller β or a smaller sample size. For example when $\beta = 0.1$ and $n = 60$ approximately 25% of simulation conducted using the fixed optimal design had an observed efficiency less than 0.65. Compare this to the results for the adaptive procedure where the 25th percentile had an observed efficiency of 0.92. In this same case the 25th percentile of the observed efficiency for the adaptive procedure is greater than the 75th percentile of the observed efficiency of $\mathscr{D}_n(\xi^*)$. Considering all simulations for observed efficiency the adaptive procedures improve the observed efficiency for almost every experiment.

$\beta = 0.1$	n	36	60
	$D_{q\{\mathscr{D}_n(K)\}-obs-eff}$	0.74, 0.86, 0.94, 0.98, 0.99	0.84, 0.92, 0.97, 0.99, 1.00
	$D_{q\{\mathscr{D}_n(\xi^*)\}-obs-eff}$	0.36, 0.54, 0.72, 0.87, 0.95	0.51, 0.65, 0.80, 0.91, 0.97
	$D_{obs-rel-eff}\{\mathscr{D}_n(K), \mathscr{D}_n(\xi^*)\}$	1.55	1.62
	$\lvert Var[\hat{\boldsymbol{\theta}}_K]\rvert / \lvert Var[\hat{\boldsymbol{\theta}}_{\xi^*}]\rvert^{-1/p}$	1.29	1.16
$\beta = 0.5$	n	12	24
	$D_{q\{\mathscr{D}_n(K)\}-obs-eff}$	0.77, 0.87, 0.94, 0.98, 0.99	0.94, 0.97, 0.99, 1.00, 1.00
	$D_{q\{\mathscr{D}_n(\xi^*)\}-obs-eff}$	0.59, 0.72, 0.85, 0.94, 0.97	0.80, 0.87, 0.93, 0.97, 0.99
	$D_{obs-rel-eff}\{\mathscr{D}_n(K), \mathscr{D}_n(\xi^*)\}$	1.13	1.06
	$\lvert Var[\hat{\boldsymbol{\theta}}_K]\rvert / \lvert Var[\hat{\boldsymbol{\theta}}_{\xi^*}]\rvert^{-1/p}$	1.13	1.11
$\beta = 1.0$	n	12	24
	$D_{q\{\mathscr{D}_n(K)\}-obs-eff}$	0.90, 0.94, 0.97, 0.99, 1.00	0.97, 0.99, 0.99, 1.00, 1.00
	$D_{q\{\mathscr{D}_n(\xi^*)\}-obs-eff}$	0.80, 0.87, 0.93, 0.97, 0.99	0.90, 0.94, 0.97, 0.99, 1.00
	$D_{obs-rel-eff}\{\mathscr{D}_n(K), \mathscr{D}_n(\xi^*)\}$	1.05	1.03
	$\lvert Var[\hat{\boldsymbol{\theta}}_K]\rvert / \lvert Var[\hat{\boldsymbol{\theta}}_{\xi^*}]\rvert^{-1/p}$	1.07	1.05

Table 1: Results from 10,000 iterations from the GHR model of the observed efficiency, $D_{q\{\mathscr{D}_n\}-obs-eff}$, are summarized as the 10th, 25th, 50th, 75th and 90th percentiles. $D_{obs-rel-eff}$ and the efficiency of the variance of the MLEs are the mean from the simulations. Values $\boldsymbol{\theta} = (1, 1, 1)$, $f_x(x) = (1, x, x^2)$, $m_j = 1$, $\mathscr{X} = [-1, 1]$ and $\beta = 1$ were used.

5 Conclusion

The majority of existing optimum design literature uses the inverse of FIM as the *a priori* variance approximation. This represents a disconnect with the more common recommendation to use OM in the analysis of experiments. This work proposes an adaptive procedure that uses the well developed tools of optimum design to incorporate the OM from previous runs in the design of current and future runs with the objective of increasing the efficiency of the OM from the entire experiment. A simulated example in the context of a GHR model showed that the OM following the adaptive procedure had greater efficiency when compared to a fixed design carried out at the traditional optimal design.

References

V. Dragalin and V. Fedorov. Adaptive designs for dose-finding based on efficacy–toxicity response. *Journal of Statistical Planning and Inference*, 136:1800–1823, 2005.

Bradley Efron and David V. Hinkley. Assesing the accuracy of the maximum likelihood estimate: Observed versus fisher information (with discussion). *Biometrika*, 65:457–483, 1978.

Adam Lane, Ping Yao, and Nancy Flounoy. Information in a two-stage adaptive optimal design. *Journal of Statistical Planning and Inference*, 144:173–187, 2014.

Friedrich Pukelsheim. *Optimal Design of Exeriments.* Wiley, Ney York, 1993.

Lynda V. White. An extension of the general equivalence theorem. *Biometrika*, 60:345–348, 1973.