Published in final edited form as:

J Stat Plan Inference. 2016 November ; 178: 128–139. doi:10.1016/j.jspi.2016.07.002.

Optimal designs based on the maximum quasi-likelihood estimator

Gang Shen^a, Seung Won Hyun^{a,*}, and Weng Kee Wong^b

^aDepartment of Statistics, North Dakota State University, Fargo, ND 58102, USA

^bDepartment of Biostatistics, University of California, Los Angeles, CA 90095, USA

Abstract

We use optimal design theory and construct locally optimal designs based on the maximum quasi-likelihood estimator (MqLE), which is derived under less stringent conditions than those required for the MLE method. We show that the proposed locally optimal designs are asymptotically as efficient as those based on the MLE when the error distribution is from an exponential family, and they perform just as well or better than optimal designs based on any other asymptotically linear unbiased estimators such as the least square estimator (LSE). In addition, we show current algorithms for finding optimal designs can be directly used to find optimal designs based on the MqLE. As an illustrative application, we construct a variety of locally optimal designs based on the MqLE for the 4-parameter logistic (4PL) model and study their robustness properties to misspecifications in the model using asymptotic relative efficiency. The results suggest that optimal designs based on the MqLE can be easily generated and they are quite robust to misspecification in the probability distribution of the responses.

Keywords

Approximate Design; Design Efficiency; Dose-finding Study; Equivalence Theorem; Heteroscedasticity; Maximum likelihood Estimator

1. Introduction

Model assumptions are frequently made for drawing statistical inference but they may not be tenable in practice. The optimal design constructed under wrong model assumptions can be very inefficient. When the model is nonlinear, a further complication is that nominal values for the model parameters are required before the optimal designs can be implemented. It is thus important to find a design that is robust to various forms of mis-specification in the statistical model.

*Corresponding author: Seung.W.Hyun@NDSU.edu (Seung Won Hyun).

The contents in this paper are solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Different approaches have been proposed to find designs robust to various departures from the model assumptions. They typically concern mis-specifications in the error distribution, the response mean function, and the form of heteroscedasticity in the model. For instance, Burridge and Sebastiani (1994) and Dette and Wong (1999) proposed optimal designs when the variance is a function of the mean response, and Chen et al. (2008) proposed optimal minimax designs for a heteroscedastic polynomial model. However the bulk of the work focused on a single violation from the model assumptions. Some examples are Läuter (1974); Stigler (1971); Lee (1987, 1988); Studden (1982); Song and Wong (1998a,b); Dette and Wong (1996, 1999); the first six considered model mis-specification in mean function and the last three concerned mis-specification in the heteroscedasticity structure.

Much of the research in optimal design assumes that the design criterion is formulated in terms of the Fisher Information matrix. This matrix can only be computed under the full specification of the probability distribution for the response. In contrast, the quasi-likelihood method introduced by Wedderburn (1974) assumes only a functional relationship between the mean and the variance of the response. This is a tenable assumption in practical situations, see for example, Box and Hill (1974), Bickel (1978) and, Jobson and Fuller (1980). This suggests that designs based on the quasi-likelihood method are likely to be more robust to model assumptions than methods that require a full parametric probability model.

There is very little work on optimal designs based on the MqLE for nonlinear regression models despite numerous successful applications of the quasi-likelihood method for generalized linear models (McCullagh and Nelder, 1989). A clear exception is work by Niaparast (2009), and Niaparast and Schwabe (2013), who derived optimal designs based on the MqLE for Poisson regression models. The first paper focused on the theoretical construction of D-optimal designs for the Poisson models with random intercepts, and the second paper extended the work to a general mixed effects Poisson model with random coefficients and so the paper deals with random slopes rather than random intercepts. However, their work concerns only finding locally D-optimal designs for the case when the response follows a Poisson distribution.

We consider a more general situation and develop theory to find various types of optimal designs based on the MqLE where the response may be a discrete or a continuous random variable. We also show such optimal designs can be generated from algorithms currently used to find optimal designs based on the Fisher Information matrix. We list advantages of such designs compared to other designs based on other efficient estimators such as the MLE and the LSE when the sample size is large and study the robustness properties of D-optimal designs to mis-specified nominal parameter values. We focus on locally optimal designs (Chernoff, 1953) where nominal values for the model parameters are required before they can be implemented. Bayesian optimal designs can also be directly constructed from our methodology but will not be discussed here. One of the key results is that we show here that asymptotically, locally optimal designs based on the MqLE also have the same property as those constructed based on the MLEs but without the normality distributional assumption. When responses are from a member of the class of exponential family distributions, we prove that the optimal designs based on the MqLE is asymptotically just as efficient as the

optimal designs based on the MLE. Such designs are also asymptotically more efficient than the designs based on asymptotic best linear unbiased estimators, such as LSE. When responses are not from a member of the exponential family of distributions, we show optimal designs based on the MqLE also perform well based on an asymptotic relative efficiency measure.

Section 2 reviews preliminaries and discusses the asymptotic optimality of MqLEs. In Section 3, we review approximate designs and equivalence theorems for checking optimality of an approximate design. We then develop the theory and an algorithm to find a variety of optimal designs based on the MqLE. As an illustrative application, we construct locally optimal designs for dose response studies using the widely used 4-parameter logistic model (4PL-model). We construct optimal designs for estimating models and meaningful function of the model parameters in the 4PL-model and compare its performance with those based on MLEs and a uniform design with 10 points when there is mis-specification in (i) the nominal values of the model parameters, (ii) the heteroscedasticity structure, and (iii) the error distributions. Section 5 concludes with a summary of our work and recommendations.

2. Preliminaries

Suppose we have a dose response study and we have resources to take n observations at d

distinct doses $x_1, \dots, x_d$ with n_i replicates at x_i , $i = 1, \dots, d$ and $\sum_{i=1}^d n_i = n$. Let y be the response at x with $E(y) = \mu$ and $Var(y) = v(\mu)$ where $\mu = \mu(\theta, x)$ and μ is a known function of an unknown r -dimensional parameter vector θ and dose x . For simplicity, we write $\mu(\theta, x)$ as $\mu(\theta)$ or simply μ if there is no confusion. Our stipulation $Var(y) = v(\mu)$ covers the situation of Niaparast (2009) but not the situation of Niaparast and Schwabe (2013) since the latter considered $Var(y) = \mu + \mu^2 \alpha(x)$ and $\alpha(x)$ is not a function of μ . Another difference is that the two papers assumed the n_i replications at x_i are not independent and we do here as in most dose response studies. Assuming that $\inf_{\mu} v(\mu) > 0$, the log quasi-likelihood of y (with dispersion ϕ) is defined as

$$q(\mu; y) \triangleq \int_{-\infty}^{\mu} \frac{y-t}{\phi v(t)} dt.$$

Let $\dot{q} \triangleq \frac{\partial}{\partial \theta} q(\mu; y)$ (quasi-score) and $\ddot{q} \triangleq \frac{\partial^2}{\partial \theta \partial \theta'} q(\mu; y)$. Wedderburn (1974) showed that $q(\mu; y)$ behaves like a likelihood function in the sense that

$$E(\dot{q}) = 0, \quad E(\dot{q} \dot{q}') = -E(\ddot{q})$$

and $-E(\ddot{q}) = \frac{1}{\phi v(\mu)} \dot{\mu}(\theta) \dot{\mu}'(\theta)$ (quasi-information), where $\dot{\mu}(\theta) \triangleq \frac{\partial \mu}{\partial \theta}$. Given $y_1, y_2, \dots, y_n$ the maximum quasi-likelihood estimate (MqLE) of θ maximizes the joint quasi-likelihood

$$\sum_{i=1}^n q(\mu_i; y_i).$$

More generally, let $\mathbf{y} = (y_1, \dots, y_n)'$ be the vector of observations at the n doses and let $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)'$ be its mean vector. Without loss of generality, let the first d elements of $\mathbf{y}$ correspond to observations from the d distinct doses $x_1, \dots, x_d$ and let $\boldsymbol{\mu}_\star = (\mu(\boldsymbol{\theta}; x_1), \dots, \mu(\boldsymbol{\theta}; x_d))'$ be the mean vector of the first d observations from the d distinct doses. Clearly $\boldsymbol{\mu}_\star = C\boldsymbol{\mu}$, where C is the first d rows of the n -dimensional identity matrix I_n .

Let $\hat{\boldsymbol{\theta}}_n$ be the maximum quasi-likelihood estimator (MqLE) of $\boldsymbol{\theta}$, let $D'_n(\boldsymbol{\theta}) = (\dot{\mu}_1, \dots, \dot{\mu}_n)$ be the full rank $r \times n$ matrix and let $V_n^{-1}(\boldsymbol{\mu})$ be the inverse of the $n \times n$ diagonal matrix $V_n(\boldsymbol{\mu}) = \text{diag}[\nu(\mu_1), \dots, \nu(\mu_n)]$. By definition, $\hat{\boldsymbol{\theta}}_n$ satisfies

$$\sum_{i=1}^n \dot{q}(\mu_i; y_i) = \sum_{i=1}^n \frac{y_i - \mu_i}{\phi \nu(\mu_i)} \frac{\partial \mu_i}{\partial \boldsymbol{\theta}} = \frac{1}{\phi} D'_n(\boldsymbol{\theta}) V_n^{-1}(\boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu}) = 0,$$

assuming that $\sum_{i=1}^n \ddot{q}(\mu_i; y_i)$ is positive definite. The MqLE of $\boldsymbol{\theta}$ is then obtained iteratively via the quasi Newton-Raphson method, i.e., at the $(k+1)^{\text{th}}$ iteration with $\hat{\mu}_i^{(k)} = \mu(\hat{\boldsymbol{\theta}}_n^{(k)}; x_i)$,

$$\hat{\boldsymbol{\theta}}_n^{(k+1)} = \hat{\boldsymbol{\theta}}_n^{(k)} + \{D'_n(\hat{\boldsymbol{\theta}}_n^{(k)}) V_n^{-1}(\hat{\boldsymbol{\mu}}^{(k)}) D_n(\hat{\boldsymbol{\theta}}_n^{(k)})\}^{-1} D'_n(\hat{\boldsymbol{\theta}}_n^{(k)}) V_n^{-1}(\hat{\boldsymbol{\mu}}^{(k)})(\mathbf{y} - \hat{\boldsymbol{\mu}}^{(k)}).$$

Observe that the MqLE is a Z-estimator with properties analogous to a MLE. Under similar regularity assumptions required for the MLE of $\boldsymbol{\theta}$, cf., Lehmann (1999), §7.1 & §7.3, (C1) – (C7), the MqLE $\hat{\boldsymbol{\theta}}_n$ as an M-estimator of $\boldsymbol{\theta}$ is also asymptotically linear and unbiased satisfying

$$\hat{\boldsymbol{\theta}}_n = \boldsymbol{\theta} + (D'_n V_n^{-1} D_n)^{-1} D'_n V_n^{-1}(\mathbf{y} - \boldsymbol{\mu}) + o_p(n^{-1/2}). \quad (1)$$

Here and elsewhere, we have suppressed the arguments in $D'_n(\boldsymbol{\theta})$ and $V_n^{-1}(\boldsymbol{\mu})$ and write them as D'_n and V_n^{-1} , for simplicity, when there is no confusion. The symbol $o_p(n^{-1/2})$ denotes a random vector whose L^∞ norm is $o_p(n^{-1/2})$ in probability. We continue to use this notation for vectors and matrices for their L^∞ norm with the specified order when there is no confusion.

Eq(1) implies that $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} N(0, \phi \Delta^{-1})$ if $\Delta = \lim_{n \rightarrow \infty} n^{-1} (D'_n V_n^{-1} D_n)$ exists, viz. Wedderburn (1974) or McCullagh (1983). Consequently, the MqLE of $\boldsymbol{\mu}_\star$ is $\hat{\boldsymbol{\mu}}_\star = \boldsymbol{\mu}_\star(\hat{\boldsymbol{\theta}}_n)$. If we expand $\mu(\boldsymbol{\theta})$ using a Taylor expansion about a neighborhood of $\boldsymbol{\theta}_0$, assuming $\sup_{\boldsymbol{\theta}} \|\dot{\mu}(\boldsymbol{\theta})\| < \infty$ and $\mu(\boldsymbol{\theta})$ is twice differentiable, we have

$$\mu(\boldsymbol{\theta}) = \mu(\boldsymbol{\theta}_0) + \dot{\mu}'(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + O(\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|^2) \quad (2)$$

and $\frac{\partial \mu_{\star}}{\partial \theta} = D_{\star}$, where $D'_{\star} = (\dot{\mu}_1, \dots, \dot{\mu}_d)$ is a $r \times d$ matrix. By the delta method, it follows that

$$\sqrt{n}(\hat{\mu}_{\star} - \mu_{\star}) \xrightarrow{d} N(0, \phi D_{\star} \Delta^{-1} D'_{\star}).$$

Here we state an asymptotic optimality of the MqLE and in A.1 of the appendix, we provide its justifications with technical details beyond that sketched out in McCullagh (1983).

Theorem 1—Let $\tilde{\theta}_n$ be an asymptotically linear unbiased estimator of θ_0 satisfying

$$\tilde{\theta}_n = \theta_0 + (L'_n L_n)^{-1} L'_n V_n^{-1/2} [y - \mu(\theta_0)] + o_p(n^{-1/2}), \quad (3)$$

where $L'_n = (l_1, \dots, l_n)$ is an arbitrary $r \times n$ full rank matrix. If $n^{-1} \sum_{i=1}^n l_i l'_i$ tends to a positive definite matrix Λ ($r \times r$ and $\max_{1 \leq i \leq n} \|l_i\| = o(n^{1/2})$), then $\hat{\mu}_{\star}$ is asymptotically optimal in the sense that $n\{Var(\tilde{\mu}_{\star}) - Var(\hat{\mu}_{\star})\}$ is asymptotically a non-negative definite matrix.

The implication of Theorem 1 is that under mild conditions, the MqLE $\hat{\mu}_{\star}$ is the asymptotically best linear unbiased estimator of $\mu_{\star}$. This result may not be surprising because McCullagh (1983) had noted that if y have a distribution from the exponential family with pdf $f(y; \eta) = \exp\{(y\eta - b(\eta))/a(\phi) - c(y; \phi)\}$ satisfying $\eta = \eta(\mu)$, $E(y) = \mu = \mu(\theta)$, x) and $Var(y) = v = a(\phi) \nu(\mu)$, the MLE of θ , $\check{\theta}_n$, has the approximation

$$\check{\theta}_n = \theta + (D'_n V_n^{-1} D_n)^{-1} D'_n V_n^{-1} (y - \mu) + o_p(n^{-1/2}). \quad (4)$$

It follows that the MLE $\check{\theta}_n$ has the same linear approximation as the MqLE $\hat{\theta}_n$ when the observations come from an exponential family with canonical parameter $\eta(\mu)$ and variance function $\nu(\mu)$, see Eq(1). By the delta method, this implies that the MqLE $\hat{\mu}_{\star}$ of $\mu_{\star}$ is asymptotically as efficient as the MLE $\check{\mu}_{\star} = \mu(\check{\theta})$ of $\mu_{\star}$. Consequently, asymptotically linear unbiased estimators of a function of θ are not more accurate than MqLE for estimating the same function of θ .

Corollary 1—MqLE $\hat{\theta}$ is asymptotically superior to LSE $\check{\theta}^{LSE}$ for any designs under the regularity conditions.

Proof: Under the regularity conditions, LSE $\check{\theta}^{LSE}$ of θ satisfies

$$\check{\theta}_n^{LSE} = \theta + (D'_n D_n)^{-1} D'_n (y - \mu) + o_p(n^{-1/2}),$$

so $\check{\theta}^{LSE}$ is a member of the class of asymptotic best linear unbiased estimators with $(L'_n L_n)^{-1} L'_n = (D'_n D_n)^{-1} D'_n V_n^{-1/2}$ in Eq(4). By Theorem 1, $n\{Var(\check{\theta}^{LSE}) - Var(\hat{\theta})\}$ is non-negative definite.

This implies that $n\{Var(\check{\mu}_\star^{LSE}) - Var(\hat{\mu}_\star)\}$ is non-negative definite by virtue of the delta method, and so the MqLE $\hat{\mu}_\star$ of $\mu_\star$ is superior to the LSE $\check{\mu}_\star^{LSE}$ of $\mu_\star$.

Therefore the optimal designs based on the MqLE and the MLE have the same efficiency when the distribution of y is a member of exponential family distributions, and the optimal design based on the MqLE asymptotically performs just as well or better than optimal designs based on any other asymptotically linear unbiased estimators such as the least square estimator (LSE).

3. Locally optimal designs based on the MqLE

Optimal designs for nonlinear models are often constructed using strong parametric model assumptions and it is well known that they can be sensitive to the assumptions. Optimal designs based on the MqLE require less stringent assumptions and they can be found once the variance function $Var(y)$ is formulated as a function of $E(y)$. In this section, we first discuss optimality design criteria and approximate designs. We then construct locally optimal designs based on the MqLE and compare their performances with locally optimal design based on the MLE and a uniform design; the latter is popular because they are intuitive appealing and easy to construct by taking equal number of observations at equally spaced doses. As an illustrative application, we construct and compare optimal designs based on the MLE and MqLE for the 4-parameter logistic model, which is widely used in biometry.

The setup assumes that we have resources to take n observations from a model with

$$E(y_{ij}) = \mu_i = \mu(\theta, x_i), \quad Var(y_{ij}) = v_i = \phi v(\mu_i); \quad j=1, 2, \dots, n_i, \quad i=1, 2, \dots, d.$$

Here θ is a r -dimensional vector of parameters and we have n_i replicates at each doses (or log doses) x_i with $\sum_{i=1}^d n_i = n$. The observation from the j^{th} replicate at the i^{th} dose is y_{ij} and its distribution is unknown except for the relationship between its mean and variance. All the observations are assumed to be independent and observations at the same dose are identically distributed. Given the design criterion and model assumptions, the design questions are how many doses are required, where these doses are and how many replications are required at each dose. In general, such questions are very difficult to answer, where the optimal designs can depend sensitively on the value of n , the regression model and the optimality criterion. To facilitate practitioners to use such designs, we would require an endless table of optimal designs! For nonlinear models, a longer list is required because for each set of nominal values of the model parameters, the locally optimal designs also depend on the set.

3.1. Approximate designs and design criteria

An increasingly common and effective approach to solve design questions is to use approximate designs, which are essentially probability measures defined on the given dose interval. For our problem, we denote such a design with d doses by $\xi = \{(x_i, w_i)\}_1^d$, where x_i is the i^{th} dose selected from a given dose interval and $w_i = n_i/n$ is the proportion of the total observations allocated to x_i , $i = 1, \dots, d$. Given an optimality criterion, the design problem requires us to determine the optimal value of d and optimal choices for (x_i, w_i) , $i = 1, \dots, d$. The implemented design takes roughly nw_i observations at x_i , subject to the constraint that each nw_i is rounded to a positive integer and they sum to n .

The advantages of working with approximate designs are well documented in the literature, see for example, the voluminous work of Kiefer (1985). The main ones are that the approach provides a unified framework for finding and verifying optimal designs under any concave (or convex) optimality criterion, and when a design is not optimal, the approach provides an easy evaluation of the proximity of the design to the optimum without knowing the optimum (Pazman, 1986). In addition, algorithms for finding a broad class of optimal designs are available. A typical application of this approach is to first calculate the Fisher information matrix for a given design and formulate the design criterion as a concave (or convex) function of this matrix.

For our setup, we recall that the quasi-information matrix is obtained from the negative of the expectation of the second derivatives of the log quasi-likelihood function with respect to the model parameters. Specifically, the normalized quasi-information matrix for θ is

$$\Delta(\theta; \xi) = n^{-1} \phi^{-1} (D_n' V_n^{-1} D_n), \text{ which can be written as } \Delta(\theta; \xi) = \frac{1}{\phi} \sum_{i=1}^d w_i g(x_i) g'(x_i) \text{ with } \phi$$

as the dispersion parameter and $g(x_i) = v(\mu_i)^{-1/2} \left(\frac{\partial \mu_i}{\partial \theta_1}, \frac{\partial \mu_i}{\partial \theta_2}, \dots, \frac{\partial \mu_i}{\partial \theta_r} \right)'$ evaluated at the nominal values for θ . This quasi-information matrix is derived based only on the variance-mean relationship of the response distribution but has the same form as the Fisher information matrix. Both matrices become equal when responses come from a normal distribution with constant variance. As we show shortly, if the design criterion is a concave (convex) function of the quasi-information matrix, an equivalence theorem can be similarly developed for checking whether a design based on the MqLE is optimal among all designs on the given design interval. Equivalence theorems for verifying whether a design based on MLEs is optimal or not are available in design monographs such as Fedorov (1972), Pazman (1986), Pukelsheim (2006) or Fedorov and Leonov (2013).

A common objective is to find an efficient design to estimate the vector of parameters θ or some functions of model parameters. D-optimality seeks to minimize the volume of the confidence ellipsoid for θ by choice of a good design and so a D-optimal design provides the most accurate estimates for the model parameters. When there are several functions of the model parameters to estimate, C-optimality seeks to minimize the asymptotic average variance of the estimated functions. Consequently, the C-optimal design provides the most accurate inference for estimating the user-selected set of functions of the model parameters. If there is only a single function of the model parameters to estimate, C-optimality becomes

c-optimality and a c-optimal design minimizes the asymptotic variance of the estimated function of interest.

We also consider designs for estimating $s(s-r)$ linear combinations of the parameters $A'\theta$, where A is a $r \times s$ matrix of rank s . Here we are only interested in approximate designs ξ for which $A'\theta$ is estimable. The quasi-information matrix of a design ξ for estimating $A'\theta$ is $M = (A'[\Delta(\theta, \xi)]^-A)^{-1}$, where $[\Delta(\theta, \xi)]^-$ is a generalized inverse of $\Delta(\theta, \xi)$.

A useful class of concave design criteria, indexed by a scalar t , for models with r parameters and formulated in terms of information matrix M is:

$$\phi_t(M) = \begin{cases} \lambda_{\min}(M) & t = -\infty; \\ (\det[M])^{1/r} & t = 0; \\ (\frac{1}{r} \text{Tr}[M^t])^{1/t} & t < 0, \end{cases}$$

where $\lambda_{\min}(M)$ denotes the minimum eigenvalue of M . Because the criteria are concave, standard arguments using the directional derivatives provide us with a tool to verify if a design is optimal among all approximate designs on the given design space. For example, it can be shown that for any $t > -\infty$, if A is user-selected matrix, the design ξ_t is optimal among all designs on the given dose interval for estimating $A'\theta$ if and only if there is $G = [\Delta(\theta; \xi_t)]^-$ such that

$$g'(x)GA(A'GA)^{-(t+1)}A'G'g(x) \leq \text{Tr}[(A'GA)^{-t}], \quad (5)$$

with equality at the dose levels of the design ξ_t (Pukelsheim, 2006). The locally optimal design becomes a locally D-optimal design when $t = 0$ and becomes a locally C-optimal design for estimating functions of the model parameters when $t = -1$. Locally c-optimal designs for estimating a function of the model parameters is a special case of the C-optimal design when $s = 1$. The function on the left hand side of (5) is sometimes called the sensitivity function and by examining whether its graph satisfies the conditions of the equivalence theorem, one can confirm the optimality of the design. All our optimal designs reported in this paper have been verified by an appropriate equivalence theorem.

D-optimal design maximizes $\phi_0(M)$. If we set $A = I_r$ and $t = 0$ in (5), the equivalence theorem for D-optimality is obtained and the design ξ_D is D-optimal design if and only if for any dose x in the dose interval,

$$d(x, \xi_D) = g'(x)[\Delta(\theta; \xi_D)]^{-1}g(x) \leq r,$$

with equality at the dose levels of the design ξ_D . Alternatively, we may want to have an efficient design for estimating several functions of the model parameters at the same time. In this case, we use the C-optimality criterion to find a design to minimize $\text{Tr}[M^{-1}]$, which is equivalent to maximize $\phi_{-1}(M)$. If the known functions of interest are $C_1(\theta), C_2(\theta), \dots$,

$C_s(\boldsymbol{\theta})$, we set $A = (\dot{C}_1(\boldsymbol{\theta}), \dot{C}_2(\boldsymbol{\theta}), \dots, \dot{C}_s(\boldsymbol{\theta}), \dot{C}_i(\boldsymbol{\theta}) = \frac{\partial C_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ and $t = -1$. From (5), we have the design ξ_C is C-optimal design among all designs on the dose interval if and only if for any dose x in the dose interval,

$$d(x, \xi_C) = \|g'(x) [\Delta(\boldsymbol{\theta}; \xi_C)]^{-1} A\|^2 \leq \text{Tr}[A' [\Delta(\boldsymbol{\theta}; \xi_C)]^{-1} A],$$

with equality at the dose levels of the design ξ_C .

Let the vector of weights $\boldsymbol{w} = (w_1, \dots, w_{d-1})'$ because $w_d = 1 - \sum_{i=1}^{d-1} w_i$. The quasi-information matrix has the same form as the Fisher information matrix and we show in A.2 and A.3 of the appendix that the vector of weights $\boldsymbol{w}$ of the D and C-optimal designs can be obtained as follows:

Theorem 2—The vector of weights $\boldsymbol{w}$ of the D-optimal design are the non-negative roots of the equation $\frac{\partial \det[\Delta(\boldsymbol{\theta}; \xi)]}{\partial \boldsymbol{w}} = 0$. If the number of distinct dose levels in the locally D-optimal design is the same as the dimension of $\boldsymbol{\theta}$ in $\mu(\boldsymbol{\theta}, x)$ (i.e., $d = r$), the D-optimal is equally weighted.

Theorem 3—The vector of weights $\boldsymbol{w}$ of the C-optimal design are the non-negative roots of the equation $\frac{\partial \text{Tr}(A' [\Delta(\boldsymbol{\theta}; \xi)]^{-1} A)}{\partial \boldsymbol{w}} = 0$.

Knowing a good upper bound on the number of support points of the optimal design can frequently simplify the search for the optimal design considerably. Ideas in Hyun et al. (2013) and Yang (2010) may be applicable to identify an upper bound on the number of points for the optimal design based on the MqLE. In our study, we provide an algorithm for finding optimal designs based on the MqLE without requiring knowledge of a useful upper bound.

3.2. Design efficiency and algorithms

Following convention, the efficiencies of a design ξ for estimating the r -dimensional vector $\boldsymbol{\theta}$ and a set of interesting functions $C(\boldsymbol{\theta})$ based on the MqLE are, respectively, measured by

$$e_D(\xi) = \left(\frac{\det[\Delta(\boldsymbol{\theta}; \xi)]}{\det[\Delta(\boldsymbol{\theta}; \xi_D)]} \right)^{\frac{1}{r}}$$

and

$$e_C(\xi) = \frac{\text{Tr}(A' [\Delta(\boldsymbol{\theta}; \xi_C)]^{-1} A)}{\text{Tr}(A' [\Delta(\boldsymbol{\theta}; \xi)]^{-1} A)}.$$

If this ratio is one-half, the interpretation is that the design ξ needs to be replicated twice for it to do as well as the optimal design. In practice, we want designs with high efficiencies.

Standard design algorithms such as those proposed by Fedorov (1972) and their many modifications thereof, can be used to search for a D-optimal or a C-optimal design based on the MLE. An improved version is the Yang-Biedermann-Tang algorithm (Yang et al., 2013), where they showed their algorithm can more effectively generate several types of optimal designs for nonlinear models with a few independent variables. Hyun et al. (2015) extended their algorithm to find multiple-objective optimal designs and this is the algorithm we used to find optimal designs in this work, after specializing our algorithm to find single-objective optimal designs. A distinguishing feature of our algorithm is that it first selects the starting design by optimizing dose levels via the Fedorov's algorithm and obtains the optimal weights for the selected optimal doses using the Newton Raphson's method based on Theorem 2 and 3. Hyun's algorithm to search multiple-objective optimal designs is available in a R-package called **VNM** (Hyun et al., 2015) and the package can be freely downloaded from the R-archive. Interested readers may also write to the second author for the codes.

We now give examples using a popular model where heteroscedasticity is simply modeled via $v(\mu) = e^{h\mu}$, and h is some known constant; other functional forms for modeling the variance in terms of the mean, such as $v(\mu) = h\mu^2$, can also be directly incorporated in our algorithm. We note that when $h = 0$, the quasi-information matrix is identical to the Fisher information matrix when errors are independent and normally distributed with a constant variance. More generally, since the Fisher information matrix for θ based on the MLE under

normal distribution is $F(\theta; \xi) = \sum_{i=1}^d w_i f(x_i) f'(x_i)$ with

$$f(x_i) = \left[v_i^{-1} + \frac{1}{2} \left\{ \frac{\dot{v}_i}{v_i} \right\}^2 \right]^{1/2} \left(\frac{\partial \mu_i}{\partial \theta_1}, \frac{\partial \mu_i}{\partial \theta_2}, \dots, \frac{\partial \mu_i}{\partial \theta_r} \right)', \text{ where } \dot{v}_i = \frac{\partial v_i}{\partial \mu_i}.$$

Let ξ_D^0 and ξ_C^0 , respectively, be the locally D- and C-optimal designs based on the MLE when we have normally distributed responses with homoscedastic errors. Let ξ_D^h and ξ_C^h , respectively, be the locally D- and C-optimal designs based on the MqLE for a heteroscedastic model. When errors are normally distributed with constant variance, we note that the latter two optimal designs based on the MqLE are identical to the optimal designs based on the LSE since MLE and LSE are the same.

3.3. Four-parameter logistic(4PL) model

The 4PL model is variously called the Hill model or the Emax model and is widely used across several disciplines. They include education testing as in item response theory (Hambleton and Swaminathan, 1985; Harris, 1989), enzyme kinetic studies (Dette et al., 2005), pharmaceutical studies (Bretz et al., 2010; Reeve and Turner, 2013) and in toxicology (Dinse, 2011; Gadakar and Call, 2015), to name a few. It is also used for curve-fitting analysis in bioassays or immunoassays such as ELISAs or dose-response curves. When some parameters of the model are set equal to certain values, the model reduces to the usual 2 parameter logistic model or the 3-parameter logistic model, which are also frequently used in the biomedical sciences. Because of the flexibility of the model, we select it as an

exemplary nonlinear model in biometrics to study the performance of optimal designs based on the MqLE.

The 4-parameter logistic model assumes symmetry around the inflection point and the mean of its continuous response is modeled by

$$\mu(\theta, x) = \frac{\theta_1}{1 + e^{\theta_2 x + \theta_3}} + \theta_4, \quad \theta_1 > 0, \theta_2 \neq 0. \quad (6)$$

There are different parametrization of the models depending on the field and application at hand. Sometimes, the variable x may influence the choice of the parametrization. For example, the 4PL model (6) seems to be more widely used when the dose is measured in logarithmic units; otherwise, the Hill model seems to be the preferred choice. We choose the formulation in (6) because (i) its form provides more stable parameter estimates than other forms (Reeve and Turner, 2013) and (ii) it facilitates the numerical search of the optimal designs over a shorter dose interval.

An interesting dose to estimate is the ED_{50} , which is the dose expected to result in having a response halfway between the upper limit and the lower limit of the mean response. In (6), x is the log dose, θ_1 is the difference between the upper limit and the lower limit of the expected response, θ_2 is the negative value of the Hill's slope that controls the steepness of the curve, θ_3 is $-\theta_2 \log(ED_{50})$, and θ_4 is the lower limit of the response. Another interesting dose to estimate is $MED(\delta)$, the dose that is expected to result in a change of δ units in the mean response from the minimum dose and δ is user-specified. For example, when working with an inhibitory drug, it may be desirable to estimate the dose that would shrink the tumor size by some user-specified δ units. It is straightforward to show that for the 4PL model, $ED_{50} = -\theta_3/\theta_2$ and $MED(\delta) = \{\log(\frac{-\delta}{\theta_1 + \delta}) - \theta_3\}/\theta_2$ if $\theta_2 > 0$ or $MED(\delta) = \{\log(\frac{\theta_1 - \delta}{\theta_1}) - \theta_3\}/\theta_2$ if $\theta_2 < 0$, where $0 < |\delta| < \theta_1$ and δ must have an opposite sign from θ_2 (Hyun and Wong, 2015).

In what is to follow, we find locally D-optimal designs for estimating θ and locally C-optimal designs for estimating both ED_{50} and $MED(\delta)$ simultaneously for the 4PL model. For illustrative purpose, we set $\delta = -1$ for MED for the rest of the paper. A direct calculation shows that the quasi-information matrix for the 4PL model is

$$\Delta(\theta; \xi) = \sum_{i=1}^d w_i g(x_i) g'(x_i) / \phi, \text{ where}$$

$$g(x) = [e^{h\mu(x, \theta)}]^{-1/2} \left(\frac{1}{1 + e^{\theta_2 x + \theta_3}}, \frac{-\theta_1 x e^{\theta_2 x + \theta_3}}{(1 + e^{\theta_2 x + \theta_3})^2}, \frac{-\theta_1 e^{\theta_2 x + \theta_3}}{(1 + e^{\theta_2 x + \theta_3})^2}, 1 \right).$$

The 4PL model has a continuous response and we assume heteroscedasticity is modeled by $\nu(\mu) = e^{h\mu}$ after noting that this assumption is not appropriate for the classical logistic model with binary responses. The locally optimal design does not depend on the of ϕ but on the value of h for the given nominal values of θ . The parameter ϕ enters the information matrix

as a multiplicative constant and so our optimal designs do not depend on its nominal value. The nominal values for the remaining parameters were taken from Khinkis et al. (2003) who found locally D-optimal designs for studying characteristics of seven anti-cancer drugs using the above model. The dose interval used in their study was $[\log(.001), \log(1000)]$.

Equivalence theorems for the D- and C-optimality criteria can be obtained by setting $A' = I_4$

and $A' = (ED_{50}, M, D)'$, respectively in (5). Here $ED_{50} = \left(0, \frac{\theta_3}{\theta_2^2}, -\frac{1}{\theta_2}, 0 \right)$ and

$$M, E, D = \begin{cases} \left(\frac{-1}{(\theta_1 + \delta)\theta_2}, \frac{\theta_3 - \log(\frac{-\delta}{\theta_1 + \delta})}{\theta_2^2}, -\frac{1}{\theta_2}, 0 \right), & \text{if } \theta_2 > 0 \\ \left(\frac{1}{(\theta_1 - \delta)\theta_2}, \frac{\theta_3 - \log(\frac{\theta_1 - \delta}{\delta})}{\theta_2^2}, -\frac{1}{\theta_2}, 0 \right), & \text{if } \theta_2 < 0. \end{cases}$$

Table 1 shows the locally D- and C-optimal designs based on the MqLE for various values of h in the variance function when the nominal value for the model parameters is $\theta = \theta_1 = (1.563, 1.790, 8.442, 0.137)'$. All locally D-optimal designs, as expected, are equally weighted with 4 doses but locally C-optimal designs may require 3 or 4 doses, depending on the values of heteroscedasticity parameter h . When $h = 0$, the locally optimal designs based on the MqLE ξ_D^h and ξ_C^h coincide with the corresponding optimal designs for the homoscedastic model. All locally optimal designs from our algorithm can be verified by examining the plot of the sensitivity function. For example, Figure 1 is the sensitivity plot $d(x, \xi_C^h)$ of the locally C-optimal design for estimating both the ED_{50} and $MED(\delta = -1)$ in the 4PL model when $\theta = \theta_1$ and $h = 1$. The graph satisfies the conditions of the equivalence theorem and so confirms the optimality of the design based on the MqLE for estimating the ED_{50} or MED in the 4PL model.

Table 1 also displays the D- and C-efficiencies of optimal designs based on MqLEs and the uniform design ξ_{U10} with 10 points evenly spread out across the dose range $[\log(.001), \log(1000)]$. A uniform design is a popular choice among practitioners because it is simple to implement and intuitively appealing. An issue in the choice of a uniform design is the number of points it has. Larger number of points have been shown to result in lower efficiencies (Wong and Lachenbruch, 1996) and so an informed decision on the number of points it has is required. We choose 10 doses on a log scale here because this design allows us to conduct a lack of fit test to assess model adequacy. Another reason is that there are real studies that use 10 doses to fit toxicology data using the 4PL model. The take home message from Table 1 is that the homoscedastic optimal designs ξ_D^0 and ξ_C^0 have high efficiencies when there is only small mis-specification in the heteroscedastic assumption, via the values of h , and they become inefficient when the amount of heteroscedasticity is increased. In addition, the ED_{50} is always located between the middle two points when there are 4 optimal design points or close to the middle points when there are 3 optimal design points. It seems that the middle two design points or the middle points move toward the right of the ED_{50} dose as the value of h increases. The 10-point uniform design ξ_{U10} performs much worse under the same situation. For the range of values of h considered for the 10-point uniform design, the C-efficiencies for estimating both ED_{50} and $MED(\delta = -1)$

average about 19%, with some as low as 12%. For D-efficiencies, they average 59% with some as low as 31.5%. This reaffirms that the choice of an appropriate uniform design is important; in this case, a uniform design with fewer points such as 7 or 8 points is likely to perform better than one with 10 points (Wong and Lachenbruch, 1996).

We recall that MLE provides the best performance when the error distribution is correctly specified. The MqLE is asymptotically as efficient as the MLE and optimal designs based on MLE or MqLE are the same if responses are from a member of the exponential family distribution. We now investigate the robustness properties of optimal designs based on the MqLE when responses do not come from a member of the exponential class of family distributions. We do this by using an asymptotic efficiency measure and consider two distributions with heteroscedasticity. Let $\mu_i = \mu(\boldsymbol{\theta}, x_i)$ and $v_i = \phi e^{h\mu_i}$. The two distributions are

- D1= a normal distribution with $E(y_{ij}) = \mu_i$ and $Var(y_{ij}) = v_i$;
- D2= a gamma distribution with the shape parameter $\alpha = \mu_i^2 / v_i$ and scale parameter $\beta = \mu_i / v_i$.

The asymptotic efficiency compares the covariance matrix of the MqLE $\hat{\boldsymbol{\theta}}$ with the one based on MLE $\check{\boldsymbol{\theta}}$, and so it is measured by

$$ARE(\hat{\boldsymbol{\theta}}, \check{\boldsymbol{\theta}}) = (\det[\Delta(\boldsymbol{\theta}; \xi)] / \det[M^*(\boldsymbol{\theta}; \xi)])^{1/4},$$

where $M^*(\boldsymbol{\theta}, \xi)$ is the Fisher information matrix based on the MLE $\check{\boldsymbol{\theta}}$ under the given distribution. For D1, $M^*(\boldsymbol{\theta}, \xi) = R(\boldsymbol{\theta}, \xi)$ and for D2, $M^*(\boldsymbol{\theta}, \xi)$ is given in the appendix A.4. The given Fisher information matrices for D1 and D2 are generic versions for any functional forms of v_i . We first compare the asymptotic efficiencies of the MqLE and MLE using the same design, i.e. the uniform design ξ_{U10} . Then the locally D-optimal design ξ_D^h is used to construct the MqLE and its asymptotic efficiency is evaluated again to see how much gain was realized using the optimal design compared to efficiency gains from the MLE using ξ_{U10} .

Table 2 shows the asymptotic efficiencies of the MqLE for various ϕ and h values under D1 and D2 with $\boldsymbol{\theta} = \boldsymbol{\theta}_1$. The asymptotic efficiency of the MqLE depends on the heteroscedasticity of the responses since it is changed by both ϕ and h . When the same design ξ_{U10} is used for both estimators, the MLE performs better than the MqLE as we expected, but for D1, the MqLE performs close to the MLE when $\phi = 0.2$ and $h = 1$, and as discussed earlier, the MqLE becomes identical to the MLE when $h = 0$. The asymptotic efficiency of the MqLE is increased almost double fold when its optimal design ξ_D^h is used. It is also observed that the MqLE obtained from ξ_D^h is more efficient than the MLE found from ξ_{U10} for selected values of h and ϕ under D1. For example, when $\phi = 0.1$, $h = -2$, and ξ_D^h is used to obtain the MqLE, $ARE(\hat{\boldsymbol{\theta}}, \check{\boldsymbol{\theta}}) = 1.718$ and this tells that the MLE needs about 72% more subjects to do as well as the MqLE. For D2, we did not observe the same

outperformance of MqLE obtained from ξ_D^h , but the optimal designs produce the MqLE that performs about as well as the MLE when $\phi < 0.2$ and $h < 1$. The overall results suggest that when the distribution of the response is not a member of exponential family distributions, the MqLE still performs reasonably well compared with the MLE for some cases. In addition, the performance of the MqLE can be dramatically increased by using the optimal designs based on the MqLE.

Table 3 compares robustness properties of locally D-optimal designs based on the MqLE and MLE when there is mis-specifications in the nominal values for θ . They displays D-efficiencies of ξ_D^h with $h = -2, 0, 2, 5$ in the variance function assuming that the true values of the nominal values for the 4PL model are $\theta = \theta_1$. We recall that ξ_D^h with $h = 0$ is the optimal design based on the MLE. Both types of optimal designs seem to be sensitive to nominal values and seem more so when there is large heteroscedasticity in the model, as in the case when $h = 5$. As expected, they perform poorly when there is a large deviation from the true values, especially for θ_2 and θ_3 . The optimal designs seem to be more sensitive to θ_2 than θ_3 . Since θ_2 is the slope controlling parameter with a very narrow range of possible values, locally optimal designs become much less efficient for a small departure from the true value of θ_2 compared to the case for θ_3 . For this model, it is known that ξ_D^0 does not depend on θ_1 and θ_4 (Li and Majumdar, 2008). However, we observe that when $h \neq 0$, ξ_D^h depends on θ_1 and θ_4 and ξ_D^h has quite high D-efficiencies for a broad of values for these parameters. The performance of the design ξ_D^h becomes less efficient as the value of h increases. In addition, Table 4 shows the robustness properties of the same designs when there is a mis-specification in the ED_{50} under the assumption that $\theta = \theta_1$. The same patterns are observed here like Table 3. The optimal designs work very poorly when there is a large departure from the true value of the ED_{50} .

4. Conclusions

In this work, we showed how to construct optimal designs based on the MqLE, which are derived under less stringent model assumptions than those for the MLE. Using only the functional relationship between the mean and variance of the response variable, we showed that the MqLE is asymptotically best linear unbiased estimator and when the response has a distribution belonging to the exponential family, MqLE is as efficient as MLE. Consequently, the proposed locally optimal designs based on the MqLE have asymptotic optimality properties comparable to those based on the MLE. Often, optimal designs are obtained based on the LSE which is the same as the MLE under normal distribution with homoscedasticity. We proved that the optimal design based on the MqLE is asymptotically more efficient than the optimal designs based on the LSE.

Although the MqLE has good asymptotic optimality, it is rarely used to construct optimal designs. We provided some theoretical results and show these optimal designs based the MqLE are quite generally robust to distributional assumptions in a dose-response study settings. In particular, we found two types of locally optimal designs based on the MqLE for the 4PL model: D-optimal design for estimating the model parameters and C-optimal design

for estimating both the ED_{50} and MED at the same time. They were compared with the corresponding optimal designs based on the MLE for a homoscedastic model and with a uniform design with 10 doses. Our results show that optimal designs based on the MqLE can produce estimators that perform well under various model mis-specifications. We focused on the 4PL model but the methodology can be easily applied to find other types of optimal designs for different nonlinear models, as long as the optimality criterion is a concave or convex function of the information matrix. We also showed such designs can be constructed using current algorithms.

In conclusion, the robustness properties of a design to model assumptions should always be assessed before implementation. Ideally the implemented design has acceptable efficiencies under a variety of violations in the model assumptions and a change in the design criterion. While broad conclusions are usually difficult to establish, especially when there are multiple objectives in the study, it is important to carry out such an investigation for specific violations in the model assumptions to better understand sensitivities of the design to model perturbations before implementation. Our proposed designs here provide an alternative to optimal designs based on MLE when model assumptions are questionable.

Acknowledgments

The research of Wong reported in this paper was partially supported by the National Institute Of General Medical Sciences of the National Institutes of Health under Award Number R01GM107639.

References

- Bickel PJ. Using residuals robustly. i: Tests for heteroscedasticity, non-linearity. *Annals of Statistics*. 1978; 6:266–291.
- Box GEP, Hill WJ. Correcting inhomogeneity of variance with power transformation weighting. *Technometrics*. 1974; 16:385–389.
- Bretz F, Dette H, Pinheiro J. Practical considerations for optimal designs in clinical dose finding studies. *Statistics in Medicine*. 2010; 29:731–742. [PubMed: 20213708]
- Burrige J, Sebastiani P. D-optimal designs for generalized linear models with variance proportional to the square of the mean. *Biometrika*. 1994; 81:295–304.
- Chen RB, Wong WK, Li KY. Optimal minimax designs over a prespecified interval in a heteroscedastic polynomial model. *Statistics and Probability Letters*. 2008; 78:1914–1921. [PubMed: 19759807]
- Chernoff H. Locally optimal designs for estimating parameters. *The Annals of Mathematical Statistics*. 1953; 24:586–602.
- Dette H, Melas VB, Wong WK. Optimal designs for goodness of fit of the michaelis-menten enzyme kinetic function. *Journal of American Statistical Association*. 2005; 100:1370–1381.
- Dette H, Wong WK. Bayesian optimal designs for models with partially specified heteroscedastic structure. *The Annals of Statistics*. 1996; 24:2108–2127.
- Dette H, Wong WK. Optimal designs when the variance is a function of the mean. *Biometrics*. 1999; 55:925–929. [PubMed: 11315030]
- Dinse GE. An em algorithm for fitting a 4-parameter logistic model to binary dose-response data. *Journal of Agricultural, Biological, and Environmental Statistics*. 2011; 16:221–232.
- Fedorov V. Theory of optimal experiments. Studden, WJ., Klimko, EM., translators and editors. New York: Academic, New York; 1972.
- Fedorov V, Leonov S. Optimal Design for Nonlinear Response Models. Chapman and Hall/CRC Biostatistics Series. 2013

- Gadakar SR, Call GB. Computational tools for fitting the hill equation to dose-response curves. *Journal of Pharmacological and Toxicological Methods*. 2015; 71:68–76. [PubMed: 25157754]
- Hambleton, RK., Swaminathan, H. Item response theory: Principles and applications. Boston: Kluwer-Nijhoff; 1985.
- Harris D. Comparison of 1–2 and 3-parameter irt models. *Educational Measurement: Issues and Practice*. 1989; 8:35–41.
- Hyun SW, Wong WK. Multiple objective optimal designs to study the interesting features in a dose-response relationship. *International Journal of Biostatistics*. 2015; 11:253–271. [PubMed: 26565557]
- Hyun, SW., Wong, WK., Yang, Y. Vnm: An r package for finding multiple-objective optimal designs for the 4-parameter logistic model. r package version 2.0. *Journal of Statistical Software*. 2015. url <http://cran.r-project.org/package=vnm> Under review
- Hyun SW, Yang M, Flournoy N. Optimal designs for response functions with a downturn. *Journal of Statistical Planning and Inference*. 2011; 141:559–575.
- Hyun SW, Yang M, Flournoy N. A procedure for finding an improved upper bound on the number of optimal design points. *Computational Statistics and Data Analysis*. 2013; 58:276–282.
- Jobson JD, Fuller WA. Least squares estimation when the covariance matrix and parameter vector are functionally related. *Journal of the American Statistical Society*. 1980; 75:176–181.
- Khinkis LA, Levasseur L, Faessel H, Greco WR. Optimal design for estimating parameters of the 4-parameter hill model. *Nonlinearity in Biology, Toxicology, and Medicine*. 2003; 1:363–377.
- Kiefer, J. Jack Carl Kiefer collected papers III, Design of Experiments. Springer-Verlag; New York: 1985.
- Läuter E. Optimal multipurpose designs for regression models. *Mathematische Operationsforschung und Statistik*. 1974; 7:51–68.
- Lee CMS. Constrained optimal designs for regression models. *Communication and Statistics-Theory and Methods*. 1987; 16:765–783.
- Lee CMS. Constrained optimal designs. *Journal of Statistical Planning and Inference*. 1988; 18:377–389.
- Lehmann, EL. Elements of large-sample theory. Springer-Verlag; New York: 1999.
- Li G, Majumdar D. D-optimal designs for logistic models with 3 and four parameters. *Journal of Statistical Planning and Inference*. 2008; 138:1950–1959.
- McCullagh P. Quasi-likelihood functions. *The Annals of Statistics*. 1983; 11:59–67.
- McCullagh, P., Nelder, JA. Generalized linear models. 2. Chapman and Hall; London: 1989.
- Niaparast M. On optimal design for a poisson regression model with random intercept. *Statistics and Probability Letters*. 2009; 79:741–747.
- Niaparast M, Schwabe R. Optimal design for quasi-likelihood estimation in poisson regression with random coefficients. *Journal of Statistical Planning and Inference*. 2013; 143:296–306.
- Pazman, A. Mathematics and Its Applications. D. Reidel Publishing Company, Kluwer Academic Publishers; Hingham MA, USA: 1986. Foundations of optimum experimental design.
- Pukelsheim, F. Optimal design of experiment. Society for Industrial and Applied Mathematics (SIAM); Philadelphia, PA: 2006.
- Reeve R, Turner JR. Pharmacodynamic models: parameterizing the hill equation, michaelis-menten, the logistic curve, and relationships among these models. *Journal of Biopharmaceutical Statistics*. 2013; 23:648–661. [PubMed: 23611201]
- Song D, Wong WK. On the construction of G_{rm} -optimal designs. *Statistica Sinica*. 1998a; 9:263–272.
- Song D, Wong WK. Optimal two-point designs for the michaelis-menten model with heteroscedastic errors. *Communications in Statistics- Theory and Methods*. 1998b; 27:1503–1516.
- Stigler S. Optimal experimental design for polynomial regression. *Journal of the American Statistical Association*. 1971; 66:311–318.
- Studden WJ. Some robust-type d-optimal designs in polynomial regression. *Journal of the American Statistical Association*. 1982; 77:916–921.
- Wedderburn RWM. Quasi-likelihood functions, generalized linear models, and the gauss-newton method. *Biometrika*. 1974; 61:439–447.

- Wong WK, Lachenbruch PA. Designing studies for dose response. *Statistics in Medicine*. 1996; 15:343–360. [PubMed: 8668866]
- Yang M. On the de la garza phenomenon. *Annals of Statistics*. 2010; 38:2499–2524.
- Yang M, Biedermann S, Tang E. On optimal designs for nonlinear models: a general and efficient algorithm. *Journal of the American Statistical Association*. 2013; 108:1411–1420.

5. Appendix

A.1. Proof of Theorem 1

Proof

Clearly $\hat{\theta}_n$ in Eq(1) belongs to the class of asymptotically linear unbiased estimators. Each component of the n -dimensional random vector $V_n^{-1/2}[\mathbf{y}_n - \boldsymbol{\mu}(\boldsymbol{\theta}_0)]$ has a univariate distribution with mean 0 and variance ϕ , so by the Lindeberg-Feller CLT,

$$\sqrt{n}(\tilde{\theta}_n - \theta_0) = \sqrt{n}(L'_n L_n)^{-1} L'_n V_n^{-1/2} [\mathbf{y}_n - \boldsymbol{\mu}(\boldsymbol{\theta}_0)] + o_p(1) \xrightarrow{d} N(0, \phi \Lambda^{-1}).$$

By the delta method, $\sqrt{n}[\boldsymbol{\mu}_*(\tilde{\theta}_n) - \boldsymbol{\mu}_*(\theta_0)] \xrightarrow{d} N(0, D_* \phi \Lambda^{-1} D_*')$ and along with Eq(2), we have $\boldsymbol{\mu}(\tilde{\theta}_n) = \boldsymbol{\mu}(\theta_0) + D(\theta_0)(\tilde{\theta}_n - \theta_0) + \boldsymbol{\varepsilon}$ with $\|\boldsymbol{\varepsilon}\|_\infty = O_p(\|\tilde{\theta}_n - \theta_0\|^2)$, which implies that

$$(L'_n L_n)^{-1} L'_n V_n^{-1/2} [\boldsymbol{\mu}(\tilde{\theta}_n) - \boldsymbol{\mu}(\theta_0)] = (L'_n L_n)^{-1} L'_n V_n^{-1/2} D(\tilde{\theta}_n - \theta_0) + (L'_n L_n)^{-1} L'_n V_n^{-1/2} \boldsymbol{\varepsilon}.$$

Since $n^{-1} \sum_{i=1}^n l_i l_i' \rightarrow \Lambda$, $\max_{1 \leq i \leq n} \|l_i\| = o(n^{1/2})$ and $\inf_{\mu} v(\mu) > 0$, we have

$$\|(L'_n L_n)^{-1} L'_n V_n^{-1/2} \boldsymbol{\varepsilon}\| \lesssim \left\{ \frac{1}{\sqrt{n}} \max_{1 \leq i \leq n} \|l_i\| \right\} \{ \|\sqrt{n}(\tilde{\theta}_n - \theta_0)\| \} \{ \|\tilde{\theta}_n - \theta_0\| \} = o_p(\|\tilde{\theta}_n - \theta_0\|).$$

We observe that Eq(2) and Eq(3) imply $(L'_n L_n)^{-1} L'_n V_n^{-1/2} [\boldsymbol{\mu}(\tilde{\theta}_n) - \boldsymbol{\mu}(\theta_0)] = \tilde{\theta}_n - \theta_0 + o(1)$, so $(L'_n L_n)^{-1} L'_n V_n^{-1/2} D_n = I_p + o(1)$, where I_p is the $p \times p$ identity matrix. To complete the proof,

let $M_n = (V_n^{-1/2} D_n) [(V_n^{-1/2} D_n)' (V_n^{-1/2} D_n)]^{-1} (V_n^{-1/2} D_n)'$, which is an idempotent and symmetric matrix. It follows that $\{I_n - M_n\}$ is also symmetric and idempotent, and hence non-negative definite. Further,

$$\lim_{n \rightarrow \infty} n(\text{Var}(\tilde{\boldsymbol{\mu}}_*) - \text{Var}(\hat{\boldsymbol{\mu}}_*)) = \phi \lim_{n \rightarrow \infty} n \{ D_* (L' L)^{-1} L' \} \{ I_n - M_n \} \{ D_* (L' L)^{-1} L' \}',$$

and the desired conclusion follows.

The justifications for the next two results rely on a result from Hyun et al. (2011) who showed that if $M(\boldsymbol{\theta}, \xi)$ is the Fisher information matrix, then for any matrix B , the matrix with $(i, j)^{th}$ element equal to

$$\frac{\partial^2 \det[B[M(\boldsymbol{\theta}; \xi)]^{-1} B']}{\partial w_i \partial w_j}$$

is a non-negative definite matrix.

A.2. Proof of Theorem 2

Proof

Maximizing the D-optimality criterion $\det[\Delta(\boldsymbol{\theta}, \xi)]$ is equivalent to minimizing $\det[\Delta(\boldsymbol{\theta}, \xi)]^{-1}$. The optimal design points are non-negative roots of the equation

$$\frac{\partial \det[\Delta(\boldsymbol{\theta}; \xi)]^{-1}}{\partial w_i} = 0$$

provided that they minimize the inverse of the determinant of the quasi-information matrix, which is true if the matrix with $(i, j)^{th}$ element equal to

$$\frac{\partial^2 \det[\Delta(\boldsymbol{\theta}; \xi)]^{-1}}{\partial w_i \partial w_j}$$

is non-negative definite.

The key step is to recognize that the quasi-information matrix $\Delta(\boldsymbol{\theta}, \xi)$ has the same form as the Fisher information matrix, and it is also a full rank non-negative definite matrix. By setting $B = I_n$, the above result from Hyun et al. (2011) implies the determinant of the quasi-information matrix is maximized.

A.3. Proof of Theorem 3

Proof

The C-optimality criterion is $\text{Tr}(A'[\Delta(\boldsymbol{\theta}, \xi)]^{-1}A)$ and we want to minimize it by choice of a design ξ . The proof is similar to Theorem 2 after we first note that the weights of the C-

optimal design are non-negative roots of the equation $\frac{\partial \text{Tr}(A'[\Delta(\boldsymbol{\theta}; \xi)]^{-1}A)}{\partial w_i} = 0$. Then we use the invariance property of the Tr function under cyclic permutation and the same result from Hyun et al. (2011) to show the matrix with $(i, j)^{th}$ element equal

$$\frac{\partial^2 Tr(A'[\Delta(\boldsymbol{\theta};\xi)]^{-1}A)}{\partial w_i \partial w_j}$$

is non-negative definite.

A.4. The Fisher information matrix for the MLE $\check{\boldsymbol{\theta}}$ under Gamma distribution

Let the shape parameter and scale parameter of the Gamma distribution be $\alpha = \mu_i^2/v_i$ and $\beta = \mu_i/v_i$ respectively. If y has such a distribution, then $E(y) = \alpha/\beta$ and $Var(y) = \alpha/\beta^2$. For the gamma distribution, the log likelihood function is the following:

$$\sum_{i=1}^n l(\mu_i; y_i) = \sum_{i=1}^n \left\{ \frac{\mu_i^2}{v_i} \log\left(\frac{\mu_i}{v_i}\right) - \log\left(\Gamma\left(\frac{\mu_i^2}{v_i}\right)\right) + \left(\frac{\mu_i^2}{v_i} - 1\right) \log(y_i) - \frac{\mu_i^2 y_i}{v_i} \right\}.$$

Assuming we have a random sample of size n at d distinct log doses, a direct calculation shows the Fisher information matrix for $\check{\boldsymbol{\theta}}$ is

$$M^*(\boldsymbol{\theta}; \xi) = \sum_{i=1}^d w_i f(x_i) f'(x_i),$$

where $f(x_i) = [P_1 + P_2 - P_3]^{1/2} \left(\frac{\partial \mu_i}{\partial \theta_1}, \frac{\partial \mu_i}{\partial \theta_2}, \dots, \frac{\partial \mu_i}{\partial \theta_r} \right)'$ with

$$P_1 = \frac{2v_i^2 - \mu_i v_i \ddot{v}_i - 4\mu_i v_i \dot{v}_i + 2(\mu_i \dot{v}_i)^2}{v_i^3} (\log(v_i) + \psi_0 - 2\log(\mu_i));$$

$$P_2 = \frac{2\mu_i v_i - \mu_i^2 \dot{v}_i}{v_i^2} \left(\frac{\dot{v}_i}{v_i} + \psi_1 \frac{2\mu_i v_i - \mu_i^2 \dot{v}_i}{v_i^2} - \frac{1}{\mu_i} \right);$$

and

$$P_3 = \frac{1}{v_i} \left(1 - \frac{\mu_i \dot{v}_i}{v_i} \right).$$

Here $\dot{v}_i = \frac{\partial v_i}{\partial \mu_i}$; $\ddot{v}_i = \frac{\partial^2 v_i}{\partial \mu_i^2}$; $\psi_0 = \frac{\partial}{\partial \mu_i} \log\left(\Gamma\left(\frac{\mu_i^2}{v_i}\right)\right)$ and $\psi_1 = \frac{\partial}{\partial \mu_i} \psi_0$.

Highlights

- Propose optimal designs based on the MqLE that require less stringent model assumptions.
- Discuss asymptotic optimality of the MqLE.
- Develop the theory and an algorithm to find various optimal designs based on the MqLE.
- The proposed designs here provide an alternative to optimal designs based on the MLE when model assumptions are questionable.

Figure 1.

Plot of $d(x, \xi_C^1)$ for the 4PL model when $\theta = (1.563, 1.790, 8.442, 0.137)'$ and $v(\mu) = e^{h\mu}$ with $h = 1$ confirms optimality of the design ξ_C^1 .

Locally D-optimal designs (ξ_D^h) for estimating θ and locally C-optimal designs (ξ_C^h) for estimating both ED_{50} and $MED(\delta = -1)$ based on the MqLE when $\theta = \theta_1$ and $ED_{50} = -4.716$ for the 4PL model with selected values of h in $\nu(\mu) = e^{\mu}$. The optimal designs are obtained using the dose interval $[\log(0.01), \log(1000)] = [-6.91, 6.91]$. The last two columns display the D-efficiencies (top rows) and C-efficiencies (bottom rows) of the locally homoscedastic D- and C-optimal designs (ξ_D^0, ξ_C^0) and the uniform design at 10 points for different values of h .

Table 1

h	ξ_D^h	$e_D(\xi_D^0)$	$e_D(\xi_{U10})$
$h = -2$	$\begin{pmatrix} -6.91 & -5.55 & -4.50 & 6.91 \\ 0.250 & 0.250 & 0.250 & 0.250 \end{pmatrix}$	0.900	0.547
$h = -1$	$\begin{pmatrix} -6.91 & -5.39 & -4.28 & 6.91 \\ 0.250 & 0.250 & 0.250 & 0.250 \end{pmatrix}$	0.974	0.596
$h = 0$	$\begin{pmatrix} -6.91 & -5.21 & -4.08 & 6.91 \\ 0.250 & 0.250 & 0.250 & 0.250 \end{pmatrix}$	1.000	0.617
$h = 1$	$\begin{pmatrix} -6.91 & -5.01 & -3.88 & 6.91 \\ 0.250 & 0.250 & 0.250 & 0.250 \end{pmatrix}$	0.973	0.607
$h = 2$	$\begin{pmatrix} -6.91 & -4.80 & -3.71 & 6.91 \\ 0.250 & 0.250 & 0.250 & 0.250 \end{pmatrix}$	0.899	0.896
$h = 5$	$\begin{pmatrix} -5.43 & -4.22 & -3.23 & 6.91 \\ 0.250 & 0.250 & 0.250 & 0.250 \end{pmatrix}$	0.532	0.530
h	ξ_C^h	$e_C(\xi_C^0)$	$e_C(\xi_{U10})$
$h = -2$	$\begin{pmatrix} -6.91 & -5.62 & -4.71 & 6.91 \\ 0.179 & 0.058 & 0.442 & 0.321 \end{pmatrix}$	0.763	0.226

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

ξ_D^h	ξ_D^h	ξ_D^h	ξ_D^h	ξ_D^h
$h = -1$	0.962	0.241	$\begin{pmatrix} -6.91 & -4.90 & -4.49 & -2.02 \\ 0.275 & 0.241 & 0.255 & 0.229 \end{pmatrix}$	0.126
$h = 0$	1.000	0.202	$\begin{pmatrix} -6.91 & -4.67 & -2.70 \\ 0.378 & 0.445 & 0.177 \end{pmatrix}$	0.158
$h = 1$	0.937	0.126	$\begin{pmatrix} -6.91 & -4.63 & -2.91 \\ 0.496 & 0.391 & 0.113 \end{pmatrix}$	0.222
$h = 2$	0.792	0.126	$\begin{pmatrix} -6.91 & -4.65 & -2.85 \\ 0.611 & 0.322 & 0.067 \end{pmatrix}$	0.222
$h = 5$	0.457	0.222	$\begin{pmatrix} -6.91 & -4.29 & -3.20 & 2.11 \\ 0.821 & 0.119 & 0.045 & 0.016 \end{pmatrix}$	0.222

Table 2

Asymptotic efficiencies of the MqLE when ξ_{U10} or ξ_D^h is used under D1 and D2 with $q = -2, -1, 0, 1, 2$, $\phi = 0.1, 0.2, 0.5, 1.0$, and $\theta = \theta_1 = (1.563, 1.790, 8.442, 0.137)'$.

Distr.	ϕ	ξ_{U10}					ξ_D^h				
		$h = -2$	$h = -1$	$h = 0$	$h = 1$	$h = 2$	$h = -2$	$h = -1$	$h = 0$	$h = 1$	$h = 2$
D1	0.1	0.939	0.975	1	0.875	0.388	1.718	1.637	1.621	1.441	0.676
	0.2	0.889	0.953	1	0.781	0.259	1.625	1.599	1.621	1.286	0.451
	0.5	0.772	0.891	1	0.596	0.134	1.410	1.495	1.621	0.981	0.233
	1.0	0.634	0.806	1	0.432	0.075	1.177	1.352	1.621	0.711	0.131
	0.1	0.529	0.493	0.446	0.373	0.249	0.967	0.867	0.734	0.614	0.434
D2	0.2	0.409	0.371	0.321	0.244	0.138	0.747	0.622	0.520	0.402	0.240
	0.5	0.272	0.237	0.189	0.122	0.057	0.497	0.397	0.306	0.201	0.098
	1.0	0.192	0.162	0.118	0.066	0.028	0.351	0.271	0.191	0.109	0.049

Table 3

D-efficiencies of locally D-optimal designs (ξ_D^h) based on the MqLE for $h = -2, 0, 2, 5$ when their nominal values for the 4-PL Model are mis-specified to be $\theta_2, \dots, \theta_9$ and the vector of true value is $\theta_1 = (1.563, 1.790, 8.442, 0.137)'$.

$\theta = (\theta_1, \theta_2, \theta_3, \theta_4)$	$e_D(\xi_D^0)$	$e_D(\xi_D^{-2})$	$e_D(\xi_D^2)$	$e_D(\xi_D^5)$
$\theta_1 = (1.563, 1.790, 8.442, 0.137)$	1.000	1.000	1.000	1.000
$\theta_2 = (1.563, 2, 10, 0.137)$	0.933	0.945	0.909	0.867
$\theta_3 = (1.563, 3, 8, 0.137)$	0.017	0.010	0.022	0.020
$\theta_4 = (1.563, 1.5, 6, 0.137)$	0.769	0.786	0.718	0.595
$\theta_5 = (1.563, 1.5, 4, 0.137)$	0.192	0.195	0.154	0.072
$\theta_6 = (1.563, 1, 8, 0.137)$	0.593	0.686	0.483	0.108
$\theta_7 = (1.563, 0.5, 1, 0.137)$	0.306	0.381	0.212	0.069
$\theta_8 = (2, 1.790, 8.442, 1)$	1.000	0.993	0.993	0.918
$\theta_9 = (3, 1.790, 8.442, 0.137)$	1.000	0.930	0.960	0.637

Table 4

D-efficiencies of locally D-optimal designs (ξ_D^h) based on the MqLE for $h = -2, 0, 2, 5$ when there is a mis-specification in ED_{50} and the true value is $ED_{50} = -4.716$ under $\theta_1 = (1.563, 1.790, 8.442, 0.137)'$.

ED_{50}	$e_D(\xi_D^0)$	$e_D(\xi_D^{-2})$	$e_D(\xi_D^2)$	$e_D(\xi_D^5)$
-5.5	0.732	0.731	0.695	0.600
-5	0.955	0.955	0.949	0.932
-4	0.744	0.743	0.716	0.661
-3	0.221	0.209	0.196	0.085
1	<0.001	<0.001	<0.001	<0.001