Revista Ciencia en Desarrollo, Vol. 7 No. 1
ISSN 0121-7488 – Enero-Junio de 2016, pp. 71-81

Optimal Population Designs for Discrimination Between Two Nested Nonlinear Mixed Effects Models

Diseños poblacionales óptimos para discriminación entre dos modelos no lineales de efectos mixtos anidados

M. E. Castañeda L.[a,*]
V. I. López-Ríos[b]

Recepción: 21-sep-2015
Aceptación: 20-ene-2016

Abstract

In this paper we consider the problem of finding optimal population designs for discrimination between two nested nonlinear mixed effects models which differ in their intra-individual covariance matrix. The criterion proposed is a generalization of the T-optimality criterion. For this criterion an equivalence theorem is provided. The application of the criterion is illustrated with an example in pharmacokinetic.

***Key words*:** Optimal Designs, Mixed Effects Model, T-Optimal Designs.

Resumen

En este artículo se considera el problema de encontrar diseños poblacionales óptimos para dicriminar entre dos modelos no lineales de efectos mixtos anidados, los cuales difieren en su matriz de covarianza intra-individual. El criterio propuesto es una generalización del criterio de T-optimalidad; para él se proporciona el respectivo teorema de equivalencia, y su aplicación se ilustra por medio de un ejemplo en farmacocinética.

***Palabras clave*:** diseños T-óptimos, diseños óptimos, modelo de efectos mixtos.

[a]Universidad de Antioquia, Medellín, Colombia.
[*]Autor de correspondencia: maria.castaneda@udea.edu.co
[b]Profesor asociado, Escuela de Estadística, Universidad Nacional de Colombia, Medellín, Colombia.

1 Introduction

In the application of optimal design theory one of the basic assumptions is to assume that the model used to describe a given phenomenon or process is the correct model. However, in the practice there may exist several candidate models. One way of to select the most adequate model among several candidates is conducting an experiment designed so that the observations obtained allow us to discriminate between the models in the best way possible. This leads to the problem of to find the optimal experimental conditions using some optimality criterion for discriminate between competing models.

In the case of fixed effects models a commonly used criterion for discriminating between two competing homoscedastic models is the T-optimality criterion, which was proposed by [1]. Under normally distributed observations, the T-optimal design provides the most powerful F-test for the lack of fit of one model when the other is assumed to be true. The criterion has been generalized for other fixed effects models see [18, 19].

The nonlinear mixed effects models are particularly useful in longitudinal studies such as population pharmacokinetics experiments, assay analysis and studies of growth, in which a limited number of samples can be obtained from each individual. These models distinguish two classes of variation: the random variation among observations within a given individual (intra-individual) and random variation among individuals (inter-individual) [3, 4]. This separation of variability allows the estimation of population characteristics from sparse samples per individual in a set of subjects without requiring individual estimation of the parameters. Depending of the application and nature of data, different covariance structures may be considered to model the two class of variation. For example, in some situations it is common practice not to assume a particular structure for the inter-individual variation, whereas for intra-individual variation can be considered different structures, among them, the usual structure which assume independent observations with constant variance, compound symmetry and autoregressive structures, constant coefficient variation structure, variance function which depends on the conditional mean response, or combinations of these structures [7, 21, 12]. Under a nonlinear mixed effects model, a population design is defined by the number of individuals to study and the individual designs to be performed in the individuals (number of samples and the sampling times) [11]. Thus, assuming that the response function and the inter-individual covariance model have been correctly specified, it may be of interest the problem of designing an experiment in a group of individuals with sparse samples per individual for discrimination between two competing intra-individual covariance models which may be nested.

Although some approaches such as those described below have been proposed for discriminate between two nonlinear mixed effects models, these may be inappropriate or less efficient in situations involving nested models as the previously considered. Waterhouse *et al.* in [22] proposed the product D-optimality criterion based on the product of the determinants of the Fisher information matrices for to find designs useful for both parameter estimation and model discrimination. For nonlinear models, such designs may be less efficient for discriminate that the T-optimal designs. Vajjah and Duffull in [20] proposed a robust T-optimal design method which does not depend on a priori selection of the true model. However, in the particular case of two nested intra-individual variation models this methodology can be not applied because the T-optimal designs are based only on the fixed effects models without residual error. Kuczewski *et al.* in [9] proposed an extension of the T-optimality criterion for heteroscedastic models, for discriminate between two multiresponse models. The criterion is derived in the case of non-nested models and can be applied directly when all individuals are observed under the same experimental conditions. Therefore, alternative methods for discrimination in the case of nested models are required. In this work, we consider the problem of to find optimal population designs for discrimination between two nested nonlinear mixed-effects models which differ in their intra-individual covariance matrix. We propose a generalization of the T-optimality criterion for this case. Our approach can be applied to population studies for groups of individuals with different sampling scheme where each sampling scheme is a multidimensional point in a finite space of admissible sampling sequences.

This paper is organized as follows. In Section 2 we present the nonlinear mixed effects model considered in this work, optimal design concepts and the model discrimination problem. In Section 3 a generalization of T-optimality criterion is defined and a necessary and sufficient condition for optimality of a design is given. In the section 4 we present an example where the criterion is applied to discriminate between two pharmacokinetic models. Finally, some conclusions and further work are given.

2 Theoretical Background

2.1 Nonlinear Mixed Effects Model

We assume that for each individual i in a population of N individuals the number of different observations available is n. Let $y_i = (y_{i1}, \ldots, y_{in})^T$ be the vector of repeated measurements for the ith individual and $x_i = (x_{i1}, \ldots, x_{in})^T$ the $n \times 1$ vector of sampling times where x_{ij} belongs to a finite set $\mathscr{X}$ consisting of t different measurement times. It is assumed that the measurements made on different subjects are independent. To model the relationship between y_i and x_i we consider a nonlinear mixed effects model which may be written as hierarchical two-stage model, see [3]:

Stage 1. Intra-Individual Model

In this stage the variability among observations *within* a given individual (intra-individual) is modeled.

Suposse that

$$y_i = f(x_i, \beta_i) + \varepsilon_i, \ i = 1, \ldots, N \tag{1}$$

where β_i is a $(p \times 1)$ vector of parameters for the ith individual; $f(x_i, \beta_i)$ is an $n \times 1$ vector function, $f(x_i, \beta_i) = (f(x_{i1}, \beta_i), \ldots, f(x_{in}, \beta_i))^T$ where f is a known nonlinear function of β_i; and ε_i is the $n \times 1$ random errors vector.

It is assumed that $\varepsilon_i | \beta_i \sim N_n(0, \mathrm{Cov}(\varepsilon_i | \beta_i))$ with

$$\mathrm{Cov}(\varepsilon_i | \beta_i) = R(x_i, \beta_i, \sigma^2, \lambda) \tag{2}$$

where $R(x_i, \beta_i, \sigma^2, \lambda)$ is an $(n \times n)$ matrix and is called the intra-individual covariance matrix, which depends on the parameters $\sigma^2 > 0$ and $\lambda \in \Omega \subset \mathbb{R}^d$.

For a given individual, the matrix R takes into account the nature of intra-individual variation and may be choosen in such a way that reflects the heterogeneity of variance, and the correlation among observations, or both. For example, in the case of data from pharmacokinetic experiments and growth studies a common model is

$$G(x_i, \beta_i, \sigma^2, \theta) = \sigma^2 \, \mathrm{diag}(f^{2\theta}(x_{i1}, \beta_i), \ldots, f^{2\theta}(x_{in}, \beta_i))$$

and $\lambda = \theta$. This matrix corresponds to uncorrelated errors with variance proportional to a power of the conditional mean. If the repeated observations are taken over time, a model for serial correlation can be considered, for example, the autoregressive (AR) model of order one for equally spaced data in time. Thus a model with this correlation structure and constant variance is $R(x_i, \beta_i, \sigma^2, \lambda) = \sigma^2 \Gamma(x_i, \beta_i, \alpha)$ where $\Gamma(x_i, \beta_i, \alpha) = \{\alpha_{\iota\upsilon}\}_{\iota\upsilon=1}^{n}$, $\alpha_{\iota\upsilon} = \alpha^{|\iota - \upsilon|}$ and $\lambda = \alpha$. Also, we can consider the case where the errors have nonconstant variance with correlation structure by the specification

$$R(x_i, \beta_i, \sigma^2, \lambda) = G^{1/2}(x_i, \beta_i, \sigma^2, \theta) \Gamma(x_i, \beta_i, \alpha) \times G^{1/2}(x_i, \beta_i, \sigma^2, \theta)$$

with $\lambda = (\theta, \alpha)^T$.

For others structures of intra-individual correlation see [12].

Stage 2. Inter-Individual Model

In model (1), the variation *among* individuals (inter-individual) is modeled through the individual specific parameters β_i. In order to account the possible dependence of this variation on individual characteristics, a model for β_i is provided in this stage.

Suposse that

$$\beta_i = \beta + b_i \tag{3}$$

where β is a $(p \times 1)$ vector of population parameters and b_i is a $(p \times 1)$ vector of random effects associated with individual i. It is assumed that the b_i are independent and normally distributed with mean 0 and variance-covariance matrix D and that the b_i and ε_i are independent. The matrix D is called the inter-individual covariance matrix. The parameter σ^2 and the distinct elements of the covariance matrix D can be arranged in a single vector ψ of covariance parameters.

First-order approximation model

An approximation to the marginal distribution of y_i can be derived taking a first-order Taylor series expansion of the model (1) about $E(b_i) = 0$. This expansion yields to the linealized model given by

$$y_i \approx f(x_i,\beta) + Z(x_i,\beta)b_i + R^{1/2}(x_i,\beta,\sigma^2,\lambda)\varepsilon_i^* \quad (4)$$

where $Z(x_i,\beta)$ is the $(n \times p)$ matrix $\left.\frac{\partial f^T(x_i,\beta_i)}{\partial \beta_i}\right|_{\beta_i=\beta}$ and $\varepsilon_i^* \sim$ i.i.d. $N_n(0,\sigma^2 I_n)$.

Thus, under (4) the approximate marginal distribution of y_i is normal with approximate mean vector and variance-covariance matrix given by

$$\begin{aligned} \mathrm{E}(y_i) &\approx f(x_i,\beta) \\ \mathrm{Cov}(y_i) &\approx Z(x_i,\beta)DZ^T(x_i,\beta) + R(x_i,\beta,\sigma^2,\lambda) \\ &\equiv \Sigma(x_i,\gamma) \end{aligned} \quad (5)$$

where $\gamma = (\beta^T, \psi^T, \lambda^T)^T \in \Gamma \subset \mathbb{R}^q$ is the full vector of unknown parameters.

2.2 Population Designs

For the given model, suppose that n_k independent observations are taken at point $x_k \in \mathscr{X}^n$, and $\sum_{k=1}^s n_k = N$, where s is the number of distinct x_k. For example, consider a study in which n_k individuals are observed under the conditions vector $x_k = (x_{k1},\ldots,x_{kn})$, with the total number of individuals N. Then the collection of x_k and n_k, represented by

$$\left\{ \begin{array}{cccc} x_1, & \ldots & , x_s \\ n_1, & \ldots & , n_s \end{array} \right\} = \{x_k, n_k\}_1^s, \ \sum_{k=1}^s n_k = N \quad (6)$$

is called population design [11, 6]. The set $\mathscr{X}^n$ is called the design region and the points x_k are called design points. We will use the term group to denote the individuals who are allocated to the same sampling sequence x_k. The collection $\zeta_N = \{x_k, \omega_k\}_1^s$ where $\omega_k = \frac{n_k}{N}$ is called normalized or exact population design with weights vector $\omega = (\omega_1,\ldots,\omega_s)$. For fixed values of the total number of individuals N and the number of sampling times n, the population optimal design problem consists in to find the design by a choice of distinct values for the sampling times vector $x \in \mathscr{X}^n$ and values for the number subjects assigned to vector x so that the resulting design maximizes some optimality criterion which will depend on the objective of experiment. The design is said to be optimal with respect to that criterion [5]. The most commonly used optimality criteria usually depend on the unknown model parameter. One approach is to construct locally optimal designs which requires to specify a prior estimate of parameter and then address the optimization problem for this specific value [5] .

Since finding an optimal exact design is a discrete optimization problem which may be difficult from both analytical and computational points of view, the corresponding approximate design should be considered one in which the weights ω_k may be any real numbers from the interval $[0,1]$. Thus, the collection

$$\zeta = \{x_k, \omega_k\}_1^s, \ 0 \le \omega_k \le 1, \ \sum_{k=1}^s \omega_k = 1 \quad (7)$$

is called approximate population design. The weight ω_k represents the proportion of total individuals that should be observed at the point x_k.

If r denotes the number of elements in the design region $\mathscr{X}^n$ then the design ζ can be specified by the vector of weights $\omega = (\omega_1,\ldots,\omega_r) \in \Xi$ where $\Xi = \{(\omega_1,\ldots,\omega_r) | \omega_k \ge 0, k = 1,\ldots,r, \sum_{k=1}^r \omega_k = 1\}$. Under this representation, if $\omega_k = 0$ this means that the corresponding design point is not used in the experiment. The set of points x_k in the design region $\mathscr{X}^n$ for which the design ω has nonzero weights ω_k is called the support set of ω and is denoted by $\mathrm{supp}(\omega)$.

After optimization, the number of individuals in each group is obtained from the optimal weights by using $n_k = N \times \omega_k$. This can yield noninteger number and therefore a rounding procedure is applied [14].

In what follows, we adopt the approximate locally optimal design approach and we use the approximate design $\omega = (\omega_1,\ldots,\omega_r)$.

2.3 The Problem of Model Discrimination

In the case of fixed effects models, these are models that do not contain the level of random effects, one most commonly used criterion for model discrimination is T-optimality proposed by Atkinson and Fedorov in [1]. For two competing homoscedastic models this criterion is based on the assumption of one model $f_t(x) = f_1(x,\beta_1)$ is the true model. The T-optimal design is a design that maximizes

$$\Delta(\xi) = \min_{\beta_2} \sum_{k=1}^s \omega_k (f_t(x_k) - f_2(x_k,\beta_2))^2 \quad (8)$$

where $\xi = \{x_k, \omega_k\}_1^s$ such that $0 \leq \omega_k \leq 1$ and $\sum_{k=1}^s \omega_k = 1$, and $f_2(x, \beta_2)$ is the rival model.

If the functions $f_j(x, \beta_j)$ depend linearly on the parameter β_j, $j = 1, 2$, then the quantity $\Delta(\xi)$ is proportional to noncentrality parameter of the χ^2 distribution of the residual sum of squares for the second model. The T-optimal design provides the most powerful F test for the lack of fit of the second model when the first is true. For nonlinear models this result is asymptotic.

For nonlinear mixed effects models, assuming that the response function f and the inter-individual model are correctly specified, we consider the discrimination problem between two nested intra-individual variation models. Specifically, let R_1 and R_2 be two alternative models to describe the intra-individual variability. It is assumed that R_2 is nested within R_1 in sense that both models involve the same structure $R(x_i, \beta_i, \sigma^2, \lambda)$ and with respect to the parameter λ the parameter space Ω_2 of R_2 is a subset of the parameter space Ω_1 of R_1 defined by the imposition of κ equality constraints. That is, $\Omega_2 = \{\lambda \in \Omega_1 \mid h_\tau(\lambda) = 0, \tau = 1, \ldots, \kappa\}$ where the functions $h_\tau(\lambda)$ are assumed to be continuously differentiable.

The objective is to find the appropriate form of the intra-individual covariance matrix R. This can be achieved by performing an experiment designed in such a way that observations obtained allow us to discriminate between R_1 and R_2 in the best way possible. For to determine such experimental design, we propose an optimality criterion which corresponds to a generalization of the T-optimality criterion. This design provides the most powerful likelihood ratio test when the largest model is assumed to be the true model. In the next section the criterion is derived.

3 Criterion for Discrimination Between Two Intra-Individual Models

The discrimination between two nested intra-individual variation models leads to the discrimination between two nested nonlinear mixed effects models $\mathbf{M}_1$ and $\mathbf{M}_2$ such that the second stage is as in (3) for both models and the first stage of each model represents a different assumption about intra-individual model, specifically:

$\mathbf{M}_1$: **Model 1**

$$\varepsilon_i \mid \beta_i \sim N_n(0, R(x_i, \beta_i, \sigma^2, \lambda)), \;\; \gamma \in \Gamma_1 \subset \mathbb{R}^q$$

where $\gamma = (\beta^T, \psi^T, \lambda^T)^T$

$\mathbf{M}_2$: **Model 2**

$$\varepsilon_i \mid \beta_i \sim N_n(0, R(x_i, \beta_i, \sigma^2, \lambda)), \;\; \gamma \in \Gamma_2 \subset \Gamma_1$$

where $\Gamma_2 = \{\gamma = (\beta^T, \psi^T, \lambda^T)^T \in \Gamma_1 \mid h_\tau(\lambda) = 0, \tau = 1, \ldots, \kappa\}$.

Assuming that the approximation (4) is exact, these models can be represented as:

$$\mathbf{M}_1 : y_i \sim N_n(f(x_i, \beta_1), \Sigma(x_i, \gamma_1))$$
$$\mathbf{M}_2 : y_i \sim N_n(f(x_i, \beta_2), \Sigma(x_i, \gamma_2))$$

In order to discriminate between these models, assuming that the largest model is completely known, we propose to find the approximate design ω^* that maximizes the following generalization of T-optimality criterion over the set Ξ:

$$\mathrm{T_W}(\omega) = \min_{\gamma_2 \in \Gamma_2} \sum_{k=1}^{r} \omega_k F(x_k, \gamma_2) \tag{9}$$

with

$$F(x_k, \gamma_2) = \log\det(\Sigma(x_k, \gamma_2)) - \log\det(\Sigma(x_k)) + \mathrm{tr}(\Sigma(x_k, \gamma_2)^{-1}\Sigma(x_k)) - n + (f(x_k) - f(x_k, \beta_2))^T \times \Sigma(x_k, \gamma_2)^{-1}(f(x_k) - f(x_k, \beta_2))$$

where $f(x) = f(x, \beta_1^0)$ and $\Sigma(x) = \Sigma(x, \gamma_1^0)$ for some known value γ_1^0. The design ω^* be called $\mathrm{T_W}$-optimal, where the letter W refers to the within-individual variance-covariance matrix. This design is locally optimal because it depends on the values of γ_1.

For this class of nonlinear mixed effects models, this criterion may be considered as an extension of proposed criterion by [9] for groups with different designs and a single response.

The justification of criterion is follows.

Let γ^0 denote the true value but unknown of parameter γ. To discriminate between the alternative models $\mathbf{M}_1$ and $\mathbf{M}_2$ we consider the likelihood ratio test for the model selection. Since $\mathbf{M}_2$ is nested within $\mathbf{M}_1$, the testing problem can be formulated as:

$$H_0 : \gamma^0 \in \Gamma_2 \;\text{ against }\; H_1 : \gamma^0 \in \Gamma_1 \tag{10}$$

where $\Gamma_2 = \{\gamma = (\beta^T, \psi^T, \lambda^T)^T \mid \gamma \in \Gamma_1, h_\tau(\lambda) = 0, \tau = 1, \ldots, \kappa\}$.

The following assumptions will be required:

(A1) Γ_1 is a compact set,

(A2) $f(x,\beta)$ is a continuous and twice continuously differentiable function in Γ_1,

(A3) $\Sigma(x,\gamma)$ is a continuous and twice continuously differentiable function in Γ_1.

Let $y_{k1}, \ldots, y_{kn_k}$ be the independent observations vectors of individuals with sampling times vector x_k, $k = 1, \ldots, s$. Assuming that the approximation (4) is exact, $y_{k1}, \ldots, y_{kn_k}$, are $N_n(f(x_k,\beta), \Sigma(x_k,\gamma))$ random vectors ($k = 1, \ldots, s$).

For the testing problem (10), the likelihood function based on the s independent samples is

$$L(\gamma) = \prod_{k=1}^{s} (2\pi)^{-n_k n/2} (\det \Sigma(x_k,\gamma))^{-n_k/2} \times \operatorname{etr}\left[-\frac{1}{2}\Sigma(x_k,\gamma)^{-1} A_k\right] \exp\left[-\frac{1}{2} n_k (\bar{y}_k - f(x_k,\beta))^T \times \Sigma(x_k,\gamma)^{-1} (\bar{y}_k - f(x_k,\beta))\right]$$

where $\bar{y}_k = n_k^{-1} \sum_{m=1}^{n_k} y_{km}$, $A_k = \sum_{m=1}^{n_k} (y_{km} - \bar{y}_k)(y_{km} - \bar{y}_k)^T$ and $\operatorname{etr}(\cdot) = \exp[\operatorname{tr}(\cdot)]$.

The log of the likelihood function is

$$\ell(\gamma) = \sum_{k=1}^{s} n_k \left[-\frac{n}{2}\log 2\pi - \frac{1}{2}\log\det(\Sigma(x_k,\gamma)) - \frac{1}{2}\operatorname{tr}(\Sigma(x_k,\gamma)^{-1} S_k) - \frac{1}{2}(\bar{y}_k - f(x_k,\beta))^T \Sigma(x_k,\gamma)^{-1} (\bar{y}_k - f(x_k,\beta))\right]$$

where $S_k = n_k^{-1} A_k$. If $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ are the maximum likelihood estimators over Γ_1 and Γ_2 respectively, the likelihood ratio test of H_0 against H_1, reject H_0 for large values of

$$-2\log\widehat{\lambda}_N = -2\{\ell(\widehat{\gamma}_2) - \ell(\widehat{\gamma}_1)\} = \sum_{k=1}^{s} n_k \{F(\bar{y}_k, S_k, \widehat{f}_{2k}, \widehat{\Sigma}_{2k}) - F(\bar{y}_k, S_k, \widehat{f}_{1k}, \widehat{\Sigma}_{1k})\}$$

with

$$F(\bar{y}, S, \widehat{f}_j, \widehat{\Sigma}_j) = \log\det(\widehat{\Sigma}_j) + \operatorname{tr}\left[\widehat{\Sigma}_j^{-1} S\right] + (\bar{y} - \widehat{f}_j)^T \widehat{\Sigma}_j^{-1} (\bar{y} - \widehat{f}_j) - \log\det(S) - n \tag{11}$$

where $\widehat{f}_j = f(x, \widehat{\beta}_j)$ and $\widehat{\Sigma}_j = \Sigma(x, \widehat{\gamma}_j)$, $j = 1, 2$.

Under H_0, the test statistic $-2\log\widehat{\lambda}_N$ has asymptotically a central χ^2 distribution with κ degrees of freedom. Therefore, an approximate test of size α of H_0 is to reject H_0 if $-2\log\widehat{\lambda}_N > c_\kappa(\alpha)$, where $c_\kappa(\alpha)$ denotes the upper $100\alpha\%$ point of the χ^2_κ distribution.

In analysis of mean and covariance structure models, the function $F(\bar{y}, S, f_j, \Sigma_j)$ is known as the maximum likelihood discrepancy function which measures the discrepancy between the sample moments and the moments based in the model which depends on the parameter γ (see [16]). The minimizing of this function leads to the maximum likelihood estimator for ith group. Extensions of discrepancy function to more than one group is straightforward [2]. Specifically, the discrepancy function for s groups is defined as

$$F(\gamma) = \sum_{k=1}^{s} \frac{n_k}{N} F(\bar{y}_k, S_k, f_{jk}, \Sigma_{jk}),\ j = 1, 2 \tag{12}$$

Thus, the test statistic (11) can be written as

$$-2\log\widehat{\lambda}_N = N(F(\widehat{\gamma}_2) - F(\widehat{\gamma}_1)) \tag{13}$$

Now, the power of this test, $P(-2\log\widehat{\lambda}_N > c_\kappa(\alpha) \mid \gamma \in \Gamma_1)$, is a function of the alternative parameter value γ. Given a specific value of N denoted by N_0 and a specific alternative parameter value $\gamma_1^0 \in \Gamma_1$ close to Γ_2, this probability can be approximated by considering the asymptotic distribution of $-2\log\widehat{\lambda}_N$ under a sequence $\{\gamma_N^0\}$ of local alternatives converging to a point γ_2^0 in Γ_2 (see [17]). It is assumed that γ_2^0 is an interior point of Γ_1. The parameter value γ_1^0 is identified then with $\gamma_{N_0}^0$. Since $F(\gamma)$ in (12) is a discrepancy function, under assumptions (A1)-(A3) and regularity conditions, this asymptotic distribution is the noncentral Chi-square distribution $\chi^2_\kappa(\delta)$ with κ degrees of freedom and noncentrality parameter δ, which can be approximated by the value:

$$\begin{aligned}\tilde{\delta} &= N \min_{\gamma_2 \in \Gamma_2} \sum_{k=1}^{s} \frac{n_k}{N} F(f_{1k}^0, \Sigma_{1k}^0, f(x_k,\beta_2), \Sigma(x_k,\gamma_2)) \\ &= N \times \min_{\gamma_2 \in \Gamma_2} \sum_{k=1}^{s} \frac{n_k}{N} F(x_k, \gamma_2) \\ &= N \times \mathrm{T_W}(\zeta_N)\end{aligned} \tag{14}$$

where $f_1^0 = f(x, \gamma_1^0)$ and $\Sigma_1^0 = \Sigma(x, \gamma_1^0)$, see [16]. Since the power of test is a monotonically increasing function of the noncentrality parameter, from (14)

the power is an increasing function of $\mathrm{T_W}(\zeta_N)$ and hence can be maximized by the choice of design ζ_N.

Finally, the exact design ζ_N can be replaced by the corresponding approximate design ζ which is represented by ω. Thus we obtain the $\mathrm{T_W}$-criterion defined in (9).

A Necessary and Sufficient Condition for $\mathrm{T_W}$-Optimality

The following definition is fundamental for characterizing of $\mathrm{T_W}$-optimal designs.

Definition 1. A design ω is called a regular design if the following set

$$\Gamma_2(\omega) = \left\{ \widetilde{\gamma}_2 : \widetilde{\gamma}_2(\omega) = \arg\min_{\gamma_2 \in \Gamma_2} \sum_{k=1}^{r} \omega_k F(x_k, \gamma_2) \right\}$$

is singleton, otherwise it is called singular design.

Hence, if ω is a regular design and $\widetilde{\gamma}_2 \in \Gamma_2(\omega)$, then $\widetilde{\gamma}_2$ is the unique solution of the equation

$$\sum_{k=1}^{r} \omega_k F(x_k, \widetilde{\gamma}_2) = \min_{\gamma_2 \in \Gamma_2} \sum_{k=1}^{r} \omega_k F(x_k, \gamma_2)$$

The following theorem is the equivalence theorem for $\mathrm{T_W}$-criterion which provides precise conditions for checking whether a particular design is $\mathrm{T_W}$-optimal.

Theorem 1. *Let ω^* be a regular design. Under the assumptions $(A1)$-$(A3)$:*

(i) A necessary and sufficient condition for the design ω^* to be T_W-optimal is $F(x,\gamma_2^*) \leq T_W(\omega^*)$, $\forall x \in \mathscr{X}^n$, where $\gamma_2^* \in \Gamma_2(\omega^*)$.

(ii) The function $F(x,\gamma_2^*)$ achieves its maximum value at the support points of the optimal designs ω^*.

Proof. The proof of this theorem is similar to the proof of Theorem 1 in [9].

(i) First, we prove that the criterion $\mathrm{T_W}$ is a concave function. To this end, suposse $\omega_1, \omega_2 \in \Xi$ and $\alpha \in [0,1]$. It is clear that Ξ is a convex set. Let $\omega = (1-\alpha)\omega_1 + \alpha\omega_2$, then

$$\begin{aligned}
\mathrm{T_W}(\omega) &= \min_{\gamma_2 \in \Gamma_2} \left[(1-\alpha) \sum_{k=1}^{r} \omega_{1k} F(x_k, \gamma_2) + \alpha \sum_{k=1}^{r} \omega_{2k} F(x_k, \gamma_2) \right] \\
&\geq (1-\alpha) \min_{\gamma_2 \in \Gamma_2} \sum_{k=1}^{r} \omega_{1k} F(x_k, \gamma_2) \\
&\quad + \alpha \min_{\gamma_2 \in \Gamma_2} \sum_{k=1}^{r} \omega_{2k} F(x_k, \gamma_2) \\
&= (1-\alpha)\mathrm{T_W}(\omega_1) + \alpha \mathrm{T_W}(\omega_2)
\end{aligned}$$

Now, the directional derivative of $\mathrm{T_W}$ at ω in the direction of $\delta_{\bar{\omega}} = \bar{\omega} - \omega$ where $\bar{\omega}$ is any design, is given by

$$\partial \mathrm{T_W}(\omega, \bar{\omega}) = \lim_{\lambda \to 0^+} \frac{\mathrm{T_W}(\omega + \lambda \delta_{\bar{\omega}}) - \mathrm{T_W}(\omega)}{\lambda}$$

Let $g(\omega, \gamma_2) = \sum_{k=1}^{r} \omega_k F(x_k, \gamma_2)$. Then

$$\mathrm{T_W}(\omega) = \min_{\gamma_2 \in \Gamma_2} g(\omega, \gamma_2)$$

Since $f(x,\beta)$ and $\Sigma(x,\gamma)$ are continuous and twice continuously differentiable in Γ_1, it follows that $g(\omega + \alpha\delta_{\bar{\omega}}, \gamma_2)$ is a continuous function at α and in Γ_2. Additionally, $\dfrac{\partial g(\omega + \alpha\delta_{\bar{\omega}}, \gamma_2)}{\partial \alpha}$ exists and is also continuous at α and in Γ_2. Hence, applying the Theorem 3.3 of [13], we get

$$\partial \mathrm{T_W}(\omega, \bar{\omega}) = \min_{\gamma_2 \in \Gamma_2(\omega)} \partial g(\omega, \gamma_2, \bar{\omega})$$

where $\partial g(\omega, \gamma_2, \bar{\omega})$ in the direction derivative of g at ω in direction of $\delta_{\bar{\omega}}$.
Note that if $\Gamma_2(\omega) = \{\widetilde{\gamma}_2\}$, then

$$\partial \mathrm{T_W}(\omega, \bar{\omega}) = \partial g(\omega, \widetilde{\gamma}_2, \bar{\omega}) \tag{15}$$

and using the definition of directional derivative

$$\begin{aligned}
&\partial g(\omega, \widetilde{\gamma}_2, \bar{\omega}) \\
&= \sum_{k=1}^{r} \bar{\omega}_k F(x_k, \widetilde{\gamma}_2) - \sum_{k=1}^{r} \omega_k F(x_k, \widetilde{\gamma}_2) \\
&= \sum_{k=1}^{r} \bar{\omega}_k \phi(x_k, \omega)
\end{aligned}$$

where $\phi(x,\omega) = F(x,\widetilde{\gamma}_2) - \mathrm{T_W}(\omega)$.
As ω^* is a regular $\mathrm{T_W}$-optimal by assumption, we have $\Gamma_2(\omega^*) = \{\gamma_2^*\}$ and from (15) it follows that

$$\partial \mathrm{T_W}(\omega^*, \omega) = \sum_{k=1}^{r} \omega_k \phi(x_k, \omega^*) \tag{16}$$

where ω is any design.

Since $T_W(\omega)$ is a concave function of ω, then the nonpositivity of the directional derivative at ω^* is a necessary and sufficient condition for the optimality of ω^*. From this fact and by (16), it follows that a necessary and sufficient condition for the optimality of ω^* is that ω^* fulfills the inequality

$$\max_{\omega\in\Xi}\left[\sum_{k=1}^{r}\omega_k\phi(x_k,\omega^*)\right]\leq 0$$

Consequently

$$\max_{x\in\mathcal{X}^n}\phi(x,\omega^*)\leq 0$$

which yields to

$$\phi(x,\omega^*)\leq 0,\ \forall x\in\mathcal{X}^n$$

(ii) We assume the contrary, this mean there is a set $\{x_1^*,\ldots,x_{s_1}^*\}\subset \text{supp}(\omega^*)$ and a scalar a such that $\sum_{k=1}^{s_1}\omega_k^*\phi(x_k^*,\omega^*)\leq a<0$ and $\phi(x^*,\omega^*)=0$ for $x^*\in\text{supp}(\omega^*)\setminus\{x_1,\ldots,x_{s_1}\}$. Then

$$\sum_{k=1}^{s}\omega_k^*\phi(x_k^*,\omega^*)\leq a<0 \qquad (17)$$

where s is the number of elements in $\text{supp}(\omega^*)$. From (15) taking $\omega=\bar{\omega}=\omega^*$, we have

$$\sum_{k=1}^{s}\omega_k^*\phi(x_k^*,\omega^*)=0 \qquad (18)$$

This contradiction proves the assertion.

□

4 An Example

In this section we present an example to illustrate the use of the criterion proposed. This is a theorical pharmacokinetics example described by [8] in a simulation study and used by [10] in the application of methods to find optimal population designs to estimate population characteristics of the pharmacokinetics of a drug in sparse-sampling experiments. We reproduce the models and parameters values from the second study.

The pharmacokinetic studies seek to understand the process of drug absortion, distribution and elimination using for example kinetic models to describe the plasma concentration as a function of time. The simplest compartmental model assumed for such a relationship is the nonlinear model given by,

$$\text{plasma concentration}=\frac{\text{Dose}}{V}\exp\left(-\frac{Cl}{V}\times\text{time}\right)$$

where V is the volume of distribution, CL is the clearance and Cl/V represents the rate of elimination; V and CL are the parameters of model which vary from individual to individual across the population under study. Suposse that the objective is to design an experiment for discriminate between two alternative models for variation within individuals. The model $\mathbf{M}_1$ assumes uncorrelated errors with variance proportional to a power of mean response and the model $\mathbf{M}_2$ also assumes uncorrelated errors but constant variance. The models are as follows.

The nonlinear mixed effects model can be written as

Stage 1. (Intra-Individual Variation)

$$\begin{aligned} y_{il} &= f(x_{il},\beta_i)+\varepsilon_{il}\\ f(x_{il},\beta_i) &= \frac{D}{V_i}\exp\left(-\frac{Cl_i}{V_i}x_{il}\right)\\ \varepsilon_i\mid\beta_i &\sim N_n(0,R(x_i,\beta_i,\sigma^2,\lambda)) \end{aligned} \qquad (19)$$

where, for the subject i, y_{il} represents the lth concentration measurement taken at time x_{il}, Cl_i is the clearance, V_i is the volume of distribution and $\beta_i=(Cl_i,V_i)^T$. The dose $D=1$ is fixed for all individuals.

Stage 2. (Inter-Individual Variation)

$$\beta_i=\beta+b_i,\ b_i\sim N_2(0,D),\ D=\begin{bmatrix}\psi_{Cl} & 0\\ 0 & \psi_V\end{bmatrix} \qquad (20)$$

where $\beta=(Cl,V)^T$ is the mean values vector.

The two alternative models for the within-individual covariance matrix $R(x_i,\beta_i,\sigma^2,\lambda)$ are:

$\mathbf{M}_1$. *Variance proportional to a power of mean response and uncorrelated errors*

$$\begin{aligned}&\varepsilon_i\mid\beta_i\sim N_n(0,R(x_i,\beta_i,\sigma^2,\lambda))\\ &R(x_i,\beta_i,\sigma^2,\lambda)=\sigma^2G(x_i,\beta_i,\theta)\end{aligned}$$

where

$$G(x_i,\beta_i,\theta)=\text{diag}(f^{2\theta}(x_{i1},\beta_i),\ldots,f^{2\theta}(x_{in},\beta_i))$$

$\mathbf{M}_2$. *Constant variance and uncorrelated errors*

$$\varepsilon_i \mid \beta_i \sim N_n(0, R(x_i, \beta_i, \sigma^2, \lambda))$$
$$R(x_i, \beta_i, \sigma^2, \lambda) = \sigma^2 I_n$$

It is assumed that the variance σ^2 is a fixed constant equal to 0.15 for both models.

Since the model $\mathbf{M}_2$ is nested within $\mathbf{M}_1$, the true model is $\mathbf{M}_1$ with the population parameter vector given by

$$\gamma_1^0 = (Cl_1, V_1, \psi_{Cl_1}, \psi_{V_1}, \theta)^T$$

where $Cl_1 = 0.5$, $V_1 = 0.2$, $\psi_{Cl_1} = 0.01$, $\psi_{V_1} = 0.0016$ and $\theta = 1$.

For the alternative model $\mathbf{M}_2$ the population parameter vector is

$$\gamma_2 = (Cl_2, V_2, \psi_{Cl_2}, \psi_{V_2}) \tag{21}$$

The set of possible sampling times considered is $\mathscr{X} = \{0.05, 0.15, 0.3, 0.6, 1\}$ hours after administration. We assume that three observations are available for each patient, $n = 3$, without replicates at an identical time point, which in a pharmacokinetic study is a sparse sampling situation. Therefore, the design region is given by the set of combinations of three sampling times from $\mathscr{X}$, that is, $\mathscr{X}^3 = \{x = (x_1, x_2, x_3); x_j \in \mathscr{X}, j = 1, 2, 3\}$ containing 10 elements. The sequences are:

$$x_1 = (0.05, 0.15, 0.3),\ x_2 = (0.05, 0.15, 0.6),$$
$$x_3 = (0.05, 0.15, 1),\ x_4 = (0.05, 0.3, 0.6),$$
$$x_5 = (0.05, 0.3, 1),\ x_6 = (0.05, 0.6, 1),$$
$$x_7 = (0.15, 0.3, 0.6),\ x_8 = (0.15, 0.3, 1),$$
$$x_9 = (0.15, 0.6, 1),\ x_{10} = (0.3, 0.6, 1).$$

To find the locally $\mathrm{T_W}$-optimal design we used the nominal values of parameters defined previously like the local parameters and the design region $\mathscr{X}^3$. The optimal design was calculated optimizing the $\mathrm{T_W}$-criterion implemented through an algorithm in R [15]. The function `nlminb` was used for the optimization in the design region $\mathscr{X}^3$.

The local $\mathrm{T_W}$-optimal design obtained is a two-point design

$$\zeta^* = \left\{ \begin{array}{cc} (0.05, 0.15, 0.3) & (0.05, 0.15, 0.6) \\ 0.82 & 0.18 \end{array} \right\} \tag{22}$$

and the parameter vector for the $\mathbf{M}_2$ model obtained in the optimization procedure is

$$\begin{aligned} \gamma_2^* &= (Cl_2^*, V_2^*, \psi_{Cl_1}^*, \psi_{V_1}^*)^T \\ &= (0.6287, 0.1998, 1.4059, 0.0027)^T \end{aligned}$$

Thus, if the objective is to design a new experiment with sparse sampling for discriminate between the $\mathbf{M}_1$ and $\mathbf{M}_2$ models then for about a 82% of the patients, blood samples should be taken at 0.05, 0.15 and 0.3 hours, and for about 18% blood samples should be taken at 0.05, 0.15 and 0.6 hours.

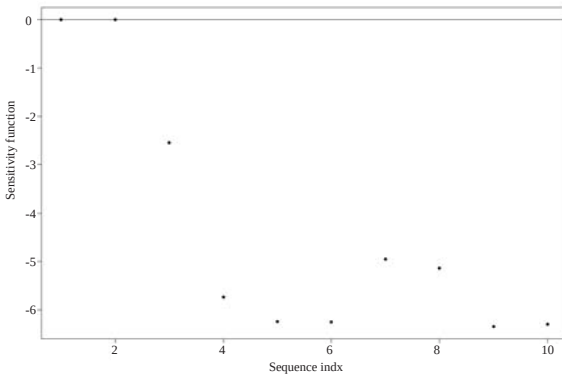

Figure 1. Sensitivity function for sampling sequences.

To check that the design obtained is $\mathrm{T_W}$-optimal we use the equivalence theorem. First, we enumerate all candidate sampling sequences, that is, the elements of $\mathscr{X}^3$ and calculate the sensitivity function $F(x_i, \gamma_2^*) - \mathrm{T_W}(\zeta^*)$ for each sampling sequence. Then plot the sensitivity function as a function of index i. The resulting plot is shown in Figure 1. From this plot it is clear that the design ζ^* consisting of the x_1 and x_2 sequences is $\mathrm{T_W}$-optimal.

5 Conclusions

A generalization of T-optimality criterion has been proposed for discriminate between two nested nonlinear mixed effects models. The first stage of each model represents a different assumption about intra-individual random variation and the second stage is the same for both models. Assuming that the response function is common for both models and it is correctly specified, we observe that the criterion development in this paper may be considered an extension of the proposed criterion by [9] for groups with different designs and a single response.

In the case of nested models an alternative criterion for discriminate between models is the $\mathrm{D_S}$-

criterion which is appropiate when interest is in estimating a subset of s parameters. Since the rival models considered in this paper are nested this criterion may be applied. The comparison between the performances of T_W- and D_S-optimal designs will be studied in future papers.

Another future work involve the study of designs with multiple objectives as the compound design. For example, the compound criteria for parameter estimation and for discrimination between models using D-optimality with T_W-optimality and D-optimality with D_S-optimality.

References

[1] A. C. Atkinson, and V. V. Fedorov, "The design of experiments for discriminating between two rival models", *Biometrika*, vol. 62, pp. 57-70, 1975.

[2] K. Bollen, "Estructural Equations with Latents Variables", John Wiley & Sons, New York, Inc, 1989.

[3] M. Davidian, and D. M. Giltinan, "Nonlinear Models for Repeated Measurement Data", Chapman & Hall, London, 1995.

[4] E. Demidenko, "Mixed Models: Theory and Applications", John Wiley & Sons Inc, New York, 2004.

[5] V. V. Fedorov, and P. Hackl, "Model-Oriented Design of Experiments", Springer, New York, 1997.

[6] R. Gagnon, and S. Leonov, "Optimal population designs for PK models with serial sampling", *Journal of Biopharmaceutical Statistics*, vol. 15, pp. 143-163, 2005.

[7] D. Hand, and M. Crowder, "Practical Longitudinal Data Analysis", Chapman & Hall, London, 1996.

[8] Y. Hashimoto, and L. Sheiner, "Designs for population pharmacodynamics: value of pharmacokinetic data and population analysis", *Journal of Pharmacokinetics and Biopharmaceutics*, vol. 19, pp. 333-353, 1991.

[9] B. Kuczewski, B. Bogacka, and D. Uciński, "Optimum designs for discrimination between two nonlinear multivariate dynamic mixed-effects models", *Biometrical Letters*, vol. 45, pp. 1-28, 2008.

[10] F. Mentré, P. Burtin, Y. Merlé, J. Bree, A. Mallet, and J. Steimer, "Sparse-sampling optimal designs in pharmacokinetics and toxicokinetics", *Drug Information Journal*, vol. 29, pp. 997-1019, 1995.

[11] F. Mentré, A. Mallet and D. Baccar, "Optimal design in random-effects regression models", *Biometrika*, vol. 84, pp. 429-442, 1997.

[12] J. C. Pinheiro and D. M. Bates, "Mixed-Effects Models in S and S-PLUS", Springer, New York, 2000.

[13] B. N. Pshenichny, "Necessary Conditions for an Extremum", Marcel Dekker, New York, 1971.

[14] F. Pukelsheim, Optimal Design of Experiments, John Wiley & Sons Inc, New York, 1993.

[15] R Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2014, ISBN 3-900051-07-0, [Online], Available: http://www.R-project.org.

[16] J. H. Steiger, A. Shapiro, and W. Browne, "On the multivariate asymptotic distribution of sequential Chi-Square statistics", *Psychometrika*, vol. 50, pp. 253-264, 1985.

[17] T. W. Stroud, "Fixed alternatives and Wald's formulation of the noncentral asymptotic behavior of the likelihood ratio statistic", *Annals of Mathematical Statistics*, vol. 43, pp. 447-454, 1972.

[18] D. Uciński, and B. Bogacka, "T-optimum designs for multiresponse dynamic heteroscedastic models", mODa 7- Advances in Model-oriented Design and Analysis, eds A. Di Bucchianico, H. Läuter and H. P. Wynn, pp. 191-199, Heidelberg: Physica, 2004.

[19] D. Uciński, and B. Bogacka, "T-optimum designs for discrimination between two multiresponse dynamic models", *Journal of the Royal Statistical Society: Series B*, vol. 67, pp.3-18, 2005.

[20] P. Vajjah, and S. Duffull, "A generalisation of T-optimality for discriminating between competing models with an application to pharmacokinetic studies", *Pharmaceutical Statistics*, vol. 11, pp. 503-510, 2012.

[21] E. F. Vonesh, and V. M. Chinchilli, "Linear and Nonlinear Models for The Analysis of Repeated Measures", Marcel Dekker, New York, 1997.

[22] T. H. Waterhouse, S. Redmann, S. B. Duffull and J. A. Eccleston, "Optimal design for model discrimination and parameter estimation for itraconazole population pharmacokinetics in cystic fibrosis patients", *Journal of Pharmacokinetics and Pharmacodynamics*, vol. 32, pp. 521-545, 2005.