Optimal Designs under Logistic Mixed Models

by

Yao Shi

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved November 2022 by the
Graduate Supervisory Committee:

John Stufken, Co-Chair
Ming-hung Kao, Co-Chair
Shiwei Lan
Rong Pan
Mark Reiser

ARIZONA STATE UNIVERSITY

December 2022

ABSTRACT

Longitudinal data involving multiple subjects is quite popular in medical and social science areas. I consider generalized linear mixed models (GLMMs) applied to such longitudinal data, and the optimal design searching problem under such models. In this case, based on optimal design theory, the optimality criteria depend on the estimated parameters, which leads to local optimality. Moreover, the information matrix under a GLMM doesn't have a closed-form expression. My dissertation includes three topics related to this design problem.

The first part is searching for locally optimal designs under GLMMs with longitudinal data. I apply penalized quasi-likelihood (PQL) method to approximate the information matrix and compare several approximations to show the superiority of PQL over other approximations. Under different local parameters and design restrictions, locally D- and A- optimal designs are constructed based on the approximation. An interesting finding is that locally optimal designs sometimes apply different designs to different subjects. Finally, the robustness of these locally optimal designs is discussed.

In the second part, an unknown observational covariate is added to the previous model. With an unknown observational variable in the experiment, expected optimality criteria are considered. Under different assumptions of the unknown variable and parameter settings, locally optimal designs are constructed and discussed.

In the last part, Bayesian optimal designs are considered under logistic mixed models. Considering different priors of the local parameters, Bayesian optimal designs are generated. Bayesian design under such a model is usually expensive in time. The running time in this dissertation is optimized to an acceptable amount with accurate results. I also discuss the robustness of these Bayesian optimal designs, which is the motivation of applying such an approach.

ACKNOWLEDGMENTS

I have had an amazing five-year PhD experience in ASU, and this dissertation is part of the proof. During these years, Dr. Stufken gave me the most support, and working with him is really encouraging and inspiring. I also want to thank my other committee members, who gave me many interesting ideas in my research.

I want to thank my parents, Xuean Shi and Yan Zhuang for their love and support through these years. Also, I want to say thank you and I love you to my fiancé, Beini Zhou.

Wish all the people live happily and healthily.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

Chapter 1

LITERATURE REVIEW

1.1 Optimal Design Approach

In an experiment, denote the response variable as Y and the design matrix as X. For a parametric model with parameter vector θ, an estimator of $\theta = (\theta_1, \ldots, \theta_d)'$, say $\hat{\theta}$, is usually derived by the maximum likelihood (ML) method. We want the estimator to be as precise as possible by wisely choosing X, which is achieved by a judicious selection of the design.

First, we introduce the definition of an experimental design. Let the total number of observations in this experiment be N, and each observation corresponds to a function of combination of factor levels, say $f(x_i^0) = (f_1(x_i^0), \ldots, f_d(x_i^0))'$. Note that $f(x_i^0)$, which is actually the ith row of X, is a vector of dimension d, with d being the number of parameters in the experiment. Usually x_i^0 takes values in R^t, where t is the number of covariates. With $\{x_1, \ldots, x_k\}$ as the set of unique elements in the set $\{x_1^0, \ldots, x_N^0\}$, and n_i, $1 \leq i \leq k$, as the number of times that x_i appears in $\{x_1^0, \ldots, x_N^0\}$. In this case, $\xi = \begin{Bmatrix} x_1, & \ldots, & x_k \\ n_1, & \ldots, & n_k \end{Bmatrix}$ is an exact design, and each x_i, $1 \leq i \leq k$, is a design point. We also have $\sum_{i=1}^{k} n_i = N$. Discrete values could cause difficulties in optimization, so we also use the concept of approximate design, which is in the form of $\xi' = \begin{Bmatrix} x_1, & \ldots, & x_k \\ w_1, & \ldots, & w_k \end{Bmatrix}$ with $\sum_{i=1}^{k} w_i = N$, allowing w_i's to be any positive real numbers less than or equal to N. To be more general, any probability measure defined on the design region U could be considered as a design, where w_i is

actually scaled in that case. In this report, approximate designs are not scaled unless specified. Approximate designs can be helpful but sometimes we still prefer exact designs, especially when we don't have much design resources. Meanwhile, under some models, like generalized linear mixed models, the form of approximate designs do not help us evaluate the information matrix, which will be discussed later.

There are many optimality criteria to evaluate the estimator obtained from the experiment, which is also evaluating the design itself. In these optimality criteria, an aim function is maximized on all possible designs, and such a function often depends on the information matrix of the model likelihood in the experiment. Considering the variance-covariance matrix of a ML estimator, denoted by $Cov(\hat{\theta}|\xi)$, by asymptotic normality, $Cov(\hat{\theta}|\xi)/N \to (I(\theta|\xi))^{-1}$ as $N \to \infty$, where $I(\theta|\xi)$ is the Fisher information of the model likelihood for one observation. Thus, we can maximize a function of $I(\theta|\xi)$, which can be theoretically derived under many statistical models, for optimization of the designs.

There are some commonly used function ϕ to maximized in different optimality criteria. In D-optimality criterion, $\phi(I(\theta|\xi)) = \prod_a \lambda_{\xi,a} = |I(\theta|\xi)|$ is maximized, where $\lambda_{\xi,a}$ is the ath eigenvalue of $I(\theta|\xi)$. For the advantage of optimization, a concave version $\log|I(\theta|\xi)|$ is usually used instead. We can also maximize $\phi(I(\theta|\xi)) = (\sum_a \lambda_{\xi,i}^{-1})^{-1} = (trace(I^{-1}(\theta|\xi)))^{-1}$ to get A-optimality criterion. As a generalization, we can consider the function $\Phi_p(I(\theta|\xi)) = (\frac{1}{d}\sum_{a=1}^{d} \lambda_{\xi,a}^{p})^{1/p}$, $p \in (-\infty, 1]$, where d is the dimension of θ, which is maximized in Φ_p-optimality. We can see that when $p = 0$, it corresponds to D-optimality, and when $p = -1$, it corresponds to A-optimality. More generally, we can also consider a function Φ that is nonnegative, nonconstant, concave, and smooth on the set of all positive definite matrices of dimension d. The optimality

criterion maximizing Φ covers many other optimality criteria, including Φ_p-optimality for $p \in (-\infty, 1]$, cf. Pukelsheim 2006, and Hu, Yang, and Stufken 2015.

Under some models, like linear models, the information matrix does not depend on the true value of θ, i.e., $I(\theta|\xi)$ is not a function of θ, so, we can derive a best design that maximizes some optimality criterion not depending on the true value of model parameter θ. However, under some models, like generalized linear models, the information matrix $I(\theta|\xi)$ is a function of θ, which means a design can never be best for all θ in the parameter space. In this case, we will consider an optimal design given a value of θ, which is called a locally optimal design. For a locally optimal design, the guess of the parameter is very important in deciding a really optimal design for the practice. To make sure that a bad guess will not ruin a locally optimal design, we must study the robustness of a locally optimal design ξ^* derived under assumed true parameter θ_{true}. Different guesses of the parameter can be used to get different locally optimal designs, and their values of the optimality criteria could be compared to that of ξ^*, to see how bad a locally optimal design could be if we made a wrong guess. For a design ξ, which may be an optimal design derived under mis-specified parameter θ_{mis}, the efficiency compared to the A-optimal design ξ_A^* and D-optimal design ξ_D^* which are derived under θ_{true} separately is defined as

$$\text{eff}(\xi) = \begin{cases} \frac{tr\{I^{-1}(\theta_{true}|\xi_A^*)\}}{tr\{I^{-1}(\theta_{true}|\xi)\}}, & \text{A} - \text{optimality} \\ \left[\frac{\det\{I^{-1}(\theta_{true}|\xi_D^*)\}}{\det\{I^{-1}(\theta_{true}|\xi)\}}\right]^{\frac{1}{d}}, & \text{D} - \text{optimality} \end{cases}$$

Different from locally optimal designs, Bayesian optimal designs are not restricted to one guessed value, but use a prior distribution for θ. Usually, normal distributions are considered. Then, different Bayesian optimality criteria are considered by maximizing different utility functions. Since we can choose different utility functions, the designs

can be optimized with respect to some loss functions we are interested in to fit in different real cases.

In practice, the set of all feasible designs usually contains too many elements, which causes difficulty in finding a design that maximizes $I(\theta|\xi)$. Complete class approach is shown to be very powerful. A complete class Ξ is defined as a set of designs such that for any design ξ, there exists a design $\xi^* \in \Xi$ satisfying $I(\theta|\xi^*) \geq I(\theta|\xi)$ under Loewner ordering. There are some short but very useful conclusions about complete class. Yang (2010), Yang and Stufken (2012) showed that for many commonly used models with d-dimensional parameter vector and one continuous covariate $x \in [U_1, U_2]$, including three-parameter Emax model, loglinear model, four-parameter LINEXP model, biexponential model and polynomial regression model,

Condition A. A complete class Ξ can be formed by one of the following:

1. Designs with at most d points, including point U_1; or
2. Designs with at most d points, including point U_2; or
3. Designs with at most d points; or
4. Designs with at most d points, including point U_1 and U_2.

Like U_1 in case above, if a design point exists in all designs in a complete class, we call it a fixed design point.

In some design problems, we are focusing on estimating a transformation of θ, for example, a smooth function of θ, say $g(\theta) : \mathbb{R}^d \to \mathbb{R}^v$, $v \leq d$. In this case, by the asymptotic normality and the property of normal distribution, we know the information for $g(\theta)$ is $\mathfrak{I}(g(\theta)|\xi) = (K' I^-(\theta|\xi) K)^{-1}$, where $K = (\partial g(\theta)/\partial\theta)'$ being full rank. Thus same discussions as above can be applied to this design problem.

1.2 Generalized Linear Mixed Models (GLMMs)

1.2.1 Brief Introduction to GLMMs

In this section, GLMMs and the corresponding design problem are discussed. A GLMM is a mixed-effects model as an extension to a GLM. So, first, it is necessary to discuss mixed-effects models. In contrast to fixed-effects models, where parameters are considered constant, mixed-effects models may have random coefficients. This can be a good description of different individuals involved in clinical trials, which shows both similarity and uniqueness of the individuals. For example, in discrete choice experiment, we have choices made by multiple individuals, and it is necessary to consider the individuals as different, while they may share some same features as belonging to our aim group. In longitudinal studies, several measurements are taken sequentially on a subject, so correlation between observations can be described by shared parameters among them, while the parameters themselves are random and change between different subjects. Also, in block designs with random block effects, an additive random effect is considered for each block. These models are also known as hierarchical models, namely the parameter is again modeled by some distribution within the whole model. One difficulty in analyzing mixed-effects models could be the lack of a closed-form likelihood function, involving an integration with respect to the random parameters. Especially in searching for optimal designs, it is hard to evaluate a design without a closed-form information matrix.

There are quite a few results for efficient experimental designs under mixed-effects models. For a linear mixed model on serial data

$$y_{mj} = \mathbf{x}_{mj}\boldsymbol{\beta} + \mathbf{z}_{mj}\mathbf{b}_m + \varepsilon_{mj}, \tag{1.1}$$

responses are grouped by the indicator m, $\mathbf{b}_m \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, and j indicates the observations in the same group. Note that the ML estimator of $\boldsymbol{\beta}$ and thus the covariance of the estimator both depend on the true value of $\boldsymbol{\Sigma}$. For such a linear local optimality problem, the closed-form expression of $cov(\hat{\boldsymbol{\beta}})$ can be obtained. For more complicated mixed models, the PhD dissertation of Pinheiro 1994 provided results for approximations to $cov(\hat{\boldsymbol{\beta}})$. As an application of the approximation under nonlinear mixed model, Hu and Stufken 2017 constructed a complete class for longitudinal experimental designs. Briefly, the model is approximated by a first-order Taylor expansion of the nonlinear part, so that it can be analysed in a linear way. The information matrix is derived on this linearized model. By the result of Schmelter 2007, this form can be linked with the information matrix under nonlinear fixed models, which means complete class results for nonlinear fixed models can be applied here, too. The optimal designs found in this type of complete classes are also shown to be robust with respect to mis-specification of variance-covariance matrix of the random effects.

Now, we formulate a GLMM for further discussion. The general form for a generalized linear model (GLM) with link function $\eta(\cdot)$ is

$$\mu = E(y) = \eta^{-1}(\mathbf{f}(x)^T\boldsymbol{\beta}). \tag{1.2}$$

With an exact design $\xi = \{(x_l), l = 1, \ldots, k\}$, the information matrix for $\boldsymbol{\beta}$ is $M(\xi) = \mathbf{F}^T\mathbf{V}\mathbf{F}$, where $\mathbf{F} = (\mathbf{f}(x_1), \ldots, \mathbf{f}(x_k))^T$, and $\mathbf{V} = \text{diag}(\text{var}(y_1), \ldots, \text{var}(y_k))$. For a longitudinal study with N subjects, denote $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^T$ as the response vector for subject i, $i = 1, \ldots, N$, using design $\xi_i = \{(x_{il}), l = 1, \ldots, n_i\}$, where n_i is the number of measurements on subject i. The total number of observations over all subjects is then $K = \sum_{i=1}^{N} n_i$. For this longitudinal data, extending the GLM (1.2) to a GLMM, the conditional mean for y_{ij}, $j = 1, \ldots, n_i$, is

$$\mu_{ij}^{\mathbf{b}_i} = E(y_{ij}|\mathbf{b}_i) = \eta^{-1}(\mathbf{f}(x_{ij})^T\mathbf{b}_i), \tag{1.3}$$

where $\mathbf{b}_i = (b_{i0}, b_{i1}, \ldots, b_{i,q-1})^T = \boldsymbol{\beta} + \boldsymbol{\alpha}_i$ consists of the subject-independent fixed-effects vector $\boldsymbol{\beta}$ and the random vector $\boldsymbol{\alpha}_i \sim N_q(0, \boldsymbol{\Sigma})$, i.e., $\mathbf{b}_i \sim N_q(\boldsymbol{\beta}, \boldsymbol{\Sigma})$. The covariance matrix $\boldsymbol{\Sigma}$ can be singular, allowing some effects to be fixed effects. The conditional variance of y_{ij} given $\mathbf{b}_i$ is $v_{ij}^{\mathbf{b}_i} = \phi a v(\mu_{ij}^{\mathbf{b}_i})$, where $v(\cdot)$ is a known function depending on $\eta(\cdot)$, a is a known constant, and ϕ is a dispersion parameter.

For GLMMs under different model assumptions added to (1.3), many optimal design conclusions were derived. For panel mixed logit models, which is a special case of GLMMs, Zhang 2018 discussed the corresponding optimal designs in her PhD dissertation. Tekle, Tan, and Berger 2008 compared extended generalized estimation equation and penalized quasi-likelihood (PQL) in searching for maximin D-optimal designs under logistic mixed-effects models, Abebe et al. 2014 studied Bayesian D-optimality under a same model by PQL, Waite and Woods 2015 proposed locally D-optimal and Bayesian D-optimal designs for GLMMs with random intercept in a block design, through marginal quasi-likelihood (MQL) and an outcome-enumeration method. Ueckert and Mentré 2017 applied Monte Carlo and adaptive Gaussian quadrature method to approximate the Fisher information matrix under discrete mixed effect models, which is shown to be superior than MQL in accuracy. Sequential D-optimal designs for GLMMs were investigated by Sinha and Xu 2011. Niaparast 2009 studied optimal designs for poisson mixed-effects models.

1.2.2 Approximate Parameters Estimation for Hierarchical and Marginal Models

Now, we introduce the method of estimating parameters in a GLMM, by which the approximated information matrix is generated for further optimal design searching. For a mixed study using a GLMM, denote $\mathbf{y} = (y_1, \ldots, y_n)^T$ as the response vector,

as the conditional mean for y_i satisfies

$$\mu_i^{\boldsymbol{\alpha}} = E(y_i|\boldsymbol{\alpha}) = \eta^{-1}(x_i'\boldsymbol{\beta} + z_i'\boldsymbol{\alpha}), \tag{1.4}$$

where $\boldsymbol{\beta}$ is the vector of fixed-effects and $\boldsymbol{\alpha}$ is the vector of random effects following multivariate $N(0, \mathbf{D})$, covariance matrix $\mathbf{D} = \mathbf{D}(\theta)$ depending on some unknown variance components. (1.3) also has a matrix form $E(\mathbf{y}|\boldsymbol{\alpha}) = g^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha})$, with $\mathbf{X}$ and $\mathbf{Z}$ formed by rows x_i' and z_i', respectively. The conditional variance of y_i given $\boldsymbol{\alpha}$ is $v_i^{\boldsymbol{\alpha}} = \phi a_i v(\mu_i^{\boldsymbol{\alpha}})$, where $v(\cdot)$ is a known function depending on $\eta(\cdot)$, a_i is a known constant, and ϕ is a dispersion parameter.

For estimating $(\boldsymbol{\beta}, \theta)$, the integrated quasi-likelihood function is defined by

$$e^{ql(\boldsymbol{\beta},\theta)} \propto |\mathbf{D}|^{-1/2} \int \exp[-\frac{1}{2\phi}\sum_i d_i(y_i; \mu_i^{\boldsymbol{\alpha}}) - \frac{1}{2}\boldsymbol{\alpha}'\mathbf{D}^{-1}\boldsymbol{\alpha}]d\boldsymbol{\alpha} \tag{1.5}$$

where

$$d_i(y, \mu) = -2\int_y^{\mu} \frac{y-u}{a_i v(u)} du.$$

Writing (1.5) as $c|\mathbf{D}|^{-1/2}\int \exp(-q(\boldsymbol{\alpha}))d\boldsymbol{\alpha}$, Breslow and Clayton 1993 applied Laplace's method to the above integral approximation, and ignored constant c, yielding

$$ql(\boldsymbol{\beta}, \theta) \approx -\frac{1}{2}\log|\mathbf{D}| - \frac{1}{2}\log|q''(\tilde{\boldsymbol{\alpha}})| - q(\tilde{\boldsymbol{\alpha}}), \tag{1.6}$$

where $\tilde{\boldsymbol{\alpha}}$ minimizes $q(\boldsymbol{\alpha})$ by $q'(\tilde{\boldsymbol{\alpha}}) = 0$. $q''(\boldsymbol{\alpha}) \approx \mathbf{Z}'\mathbf{W}\mathbf{Z} + \mathbf{D}^{-1}$, and $\mathbf{W}$ is the GLM iterated weights, which is $diag\{\{\phi a_1 v(\mu_1^{\boldsymbol{\alpha}})[\eta'(\mu_1^{\boldsymbol{\alpha}})]^2\}^{-1}, \ldots, \{\phi a_n v(\mu_n^{\boldsymbol{\alpha}})[\eta'(\mu_n^{\boldsymbol{\alpha}})]^2\}^{-1}\}$. This term can be assumed varying very slowly, so the final aim function can be simplified as

$$-\frac{1}{2\phi}\sum_i d_i(y_i; \mu_i^{\boldsymbol{\alpha}}) - \frac{1}{2}\boldsymbol{\alpha}'\mathbf{D}^{-1}\boldsymbol{\alpha}, \tag{1.7}$$

as penalized quasi-likelihood. By taking derivatives w.r.t. $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, the score equations for maximizing PQL is derived. These equations have iterative solutions

based on the iterated weighted least squares (IWLS) by Green 1987. One can first solve $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})\boldsymbol{\beta} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$, where $\mathbf{V} = \mathbf{W}^{-1} + \mathbf{Z}\mathbf{D}\mathbf{Z}'$ and $\mathbf{Y}$ has components $Y_i = x_i'\boldsymbol{\beta} + z_i'\boldsymbol{\alpha} + (y_i - \mu_i^{\boldsymbol{\alpha}})\eta'(\mu_i^{\boldsymbol{\alpha}})$, then find $\hat{\boldsymbol{\alpha}} = \mathbf{D}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})$.

Coming back to our design situation, under locally optimality, we have some knowledge of the parameters, say some prior estimation or a fair guess of $(\boldsymbol{\beta}, \mathbf{D})$. Then the conditional estimator of $\boldsymbol{\beta}$ given $\boldsymbol{\alpha}$ is subject to $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})\boldsymbol{\beta} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$, which yields an approximated covariance matrix of $\hat{\boldsymbol{\beta}}$, as suggested by Breslow and Clayton 1993,

$$cov(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}. \tag{1.8}$$

It is worth noting that, the previous analysis can be applied directly to a longitudinal study with multiple subjects in model (1.3). Since the responses from different subjects are assumed to be independent, we just need to replace $\mathbf{X}$ and $\mathbf{Z}$ with stacked ones, and $\mathbf{V}$ with diagonally stacked $\mathbf{V}_i$. $\mathbf{V}_i$ is corresponding to subject i, dependent on the design assigned to it. By matrix calculation, it is easy to show that for model (1.3),

$$cov(\hat{\boldsymbol{\beta}}) = (\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i)^{-1}. \tag{1.9}$$

Other than the hierarchical model (1.3), one can also build marginal models on the desire of estimating covariate effects on population averages, which is also discussed by Breslow and Clayton 1993. The marginal mean is defined as

$$E(y_i) = \mu_i = g^{-1}(x_i'\boldsymbol{\beta}). \tag{1.10}$$

For longitudinal designs, Zeger, Liang, and Albert 1998 showed that the mean in (1.10) needs to be altered in the regression variables or coefficients. After that, by McCullagh and Nelder 1989 (sec. 9.3), quasi-likelihood equations $\frac{\partial \boldsymbol{\mu}'}{\partial \boldsymbol{\beta}} var^{-1}(\mathbf{y})(\mathbf{y} - \boldsymbol{\mu})$ can be used. IWLS regression on these score equations gives solutions in a similar form as for hierarchical models. Note that the expression of $\mathbf{W}$ changes along with μ_i

in (1.10), i.e., $diag\{\{\phi a_1 v(\mu_1)[\eta'(\mu_1)]^2\}^{-1}, \ldots, \{\phi a_n v(\mu_n)[\eta'(\mu_n)]^2\}^{-1}\}$. Applying the same discussion, $cov(\hat{\boldsymbol{\beta}})$ is derived as $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$, which is in the same form as (1.8), except the change in $\mathbf{W}$. Similarly, for multiple subjects, we have the same form of $cov(\hat{\beta})$ as (1.9).

1.3 Support Points: a Good Representative Points Set

To evaluate the value of the optimality criteria based on approximated information matrices, for example, using expression (1.9), with a given distribution of $\mathbf{b}_i$, an average can be calculated from a sample of $\mathbf{b}_i$. Which is to say, the estimation of $E(g(X))$ for a function g of the random variable X can be

$$\frac{1}{n}\sum_i g(x_i), \tag{1.11}$$

with a size n random sample $\{x_i\}_{i=1}^n$ from the distribution of X. In design problems, the prior of $\mathbf{b}_i$ is usually assumed to be a multivariate normal distribution, thus such a method is considered a good solution here, and the accuracy of this estimation is highly dependent on sample size n.

To decrease n in terms of shorter running time, other than a random sample, we can turn to a better representative points set. As one of better representative sets, Mak and Joseph 2018 introduced support points, which performs especially well in estimating expectations. We now mention a little bit about how support points are constructed for a better understanding. For a random variable $K \sim F$, a size n set of support points $\{x_i\}_{i=1}^n$ minimizes the energy distance

$$E(F, F_n) = \frac{2}{n}\sum_{i=1}^{n} E\|x_i - K\|_2 - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|x_i - x_j\|_2,$$

where F_n is the empirical distribution function of $\{x_i\}_{i=1}^n$, and $\|\cdot\|_2$ is the L_2 norm. This points set has many good properties, including distributional convergence and consistency. To be specific, for a continuous and bounded function $g(K)$ and $K_n \sim F_n$,

$$\lim_{n\to\infty} E[g(K_n)] = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} g(x_i) = E[g(K)].$$

Compared to using a random sample, support points provides faster convergence in estimation of the expectation as the size $n \to \infty$. Without specific methods to be used in further analysis, support points are also good sub-samples of a big data set. To define support points in a data set, the only change to be made is replacing F by the empirical distribution of the data set.

1.4 Design Searching Algorithm: Particle Swarm Optimization (PSO)

In optimal design problems, after deciding a criterion and settling a good way to evaluate it on computer, we also need an efficient way to search for the best design in the design space. The design space is defined as the space containing all possible designs, or all designs that we are interested in. Here we are interested in a popular meta-heuristic algorithm, particle swarm optimization (PSO). The advantages of meta-heuristic algorithms includes gradient-free (compared to gradient descent algorithm), have good efficiency, and are easy to apply. On the other hand, such algorithms usually don't have any promises on convergence of the solution found. For our design problems, we can use the equivalence theorem as a stopping criterion, but even if we don't have one, PSO can also give us quite efficient designs in a decent amount of time. A small difference between the efficient designs found by PSO and the optimal one may be not important at all, in other words, we may consider an efficient design found by a proper run of PSO as optimal.

Kennedy and Eberhart 1995 first raised this algorithm in a bare-bone version and since then it attracted a lot of attention by researchers. Bare-bone PSO (BPSO) optimizes the aim function through the behavior of a group of particles, called the swarm. Through iterations, each particle moves based on knowledge of both its own and the whole swarm, and finally gets to an optimum. Denote the ith particle as x_i, and a corresponding velocity as v_i, its movement can be described as follows.

$$v_{i+1} = wv_i + c_1 rand_1(p_i - x_i) + c_2 rand_2(g_i - x_i), \tag{1.12}$$

$$x_{i+1} = x_i + v_i, \tag{1.13}$$

where w is the inertia weight, c_i's are two step parameters, p_i is the best position the ith particle ever achieved (pbest), and g_i is the best position the whole swarm ever achieved (gbest). Pseudo code of BPSO is shown in Algorithm 1. In this algorithm, the number of particles, the maximum iteration number n_{max}, the inertia weight w, and two step parameters, c_1 and c_2, are user defined parameters. Parameter choices are usually dependent on the optimization problem. For a default setting, we use several dozens of particles with several thousands of iterations, with w being .9, and c_1 and c_2 being 2. The algorithm stops when the maximum number of iterations is reached or the some stopping criterion is satisfied. In my simulation, I always use a fixed maximum number of iterations, since the design problems in my dissertation don't have equivalence theorems to verify the optimality of the designs. Some detailed discussion is in Chapter 2.

There are also variety of variations of BPSO. In my dissertation, unless specified, BPSO will be used in optimal design searching.

Algorithm 1 BPSO Algorithm

Initialization of swarm;
Maximum iteration number n_{max};
$i = 1$;
while Stopping criteria not satisfied or $i \leq n_{max}$ **do**
    Update ***pbest*** and ***gbest***;
    Update velocity using 1.12;
    Update particle positions using 1.13;
end while

1.5 Bayesian Design Approach

The quality of a locally optimal design highly depends on the true value of the parameter, which is unknown and could be far from our guess. To obtain more robust optimal designs, one possibility is to assume that the parameters are following some prior distributions, which leads to a Bayesian approach. Here we briefly introduce Bayesian A- and D-optimality criteria based on results of Chaloner and Verdinelli. 1995.

As in other Bayesian approaches, we assume that parameters are random variables that follow a prior distribution. First we consider a linear regression model $y|\boldsymbol{\theta}, \sigma^2 \sim N(X\boldsymbol{\theta}, \sigma^2 I)$, and the prior information is $\boldsymbol{\theta}|\sigma^2 \sim N(\boldsymbol{\theta}_0, \sigma^2 R^{-1})$ with hyper parameters $\boldsymbol{\theta}_0$ and matrix $\mathbf{R}$. σ^2 is assumed to be fixed and known. A Bayesian optimal design for this model would maximize an expected utility function based on some decision rule over a class of designs that we are interested in. One of the options could be maximizing the expected gain in Shannon information as an expected utility, which is the expected Kullback-Leibler distance between the posterior and the prior distribution

$$\int \log \frac{p(\boldsymbol{\theta}|y, \xi)}{p(\boldsymbol{\theta})} p(y, \boldsymbol{\theta}|\xi) d\boldsymbol{\theta} dy.$$

This reduces to maximizing

$$U_1(\xi) = -\frac{d}{2}\log(2\pi) - \frac{d}{2} + \frac{1}{2}\log\det\{\sigma^{-2}(nI(\boldsymbol{\theta}|\xi) + \mathbf{R})\},$$

recall that d is the dimension of $\boldsymbol{\theta}$. This lead to the Bayes D-optimality criterion of maximizing $\phi_1(\xi) = \det\{nI(\boldsymbol{\theta}|\xi) + \mathbf{R}\}$, where $nI(\boldsymbol{\theta}|\xi) = X'X$, which is just the information matrix.

When the purpose of the experiment is to find an estimation of the parameter $\boldsymbol{\theta}$ or a function $g(\boldsymbol{\theta})$, a proper loss function might be applied, for example, quadratic loss. In this case, the expected utility turns out to be

$$U_2(\xi) = -\int (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\mathbf{C}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})p(y, \boldsymbol{\theta}|\xi)d\boldsymbol{\theta}dy,$$

where $\mathbf{C}$ is a symmetric nonnegative definite matrix related to the quadratic loss function. Then, after Bayes procedure calculation, we obtain that $U_2(\xi) = -\sigma^2 tr\{\mathbf{C}(nI(\boldsymbol{\theta}|\xi) + \mathbf{R})^{-1}\}$. Thus, we have the Bayesian A-optimality criterion that maximizes $\phi_2(\xi) = -tr\{\mathbf{C}(nI(\boldsymbol{\theta}|\xi) + \mathbf{R})^{-1}\}$ over design ξ.

Note that there are also other useful utility functions and even more than one method to equivalently derive the same Bayesian optimality.

Next, we discuss the Bayesian design approach under mixed models. In a mixed model, the random effect is assumed to be following some distribution with some parameters denoted by vector $\boldsymbol{\theta}$. A locally optimal design for such a model is dependent on $\boldsymbol{\theta}$. In Bayesian approach, $\boldsymbol{\theta}$ is assumed to be following a given prior distribution $p(\boldsymbol{\theta})$. For the first step, a normal approximation to the posterior distribution of $\boldsymbol{\theta}$ under any mixed models could be

$$\boldsymbol{\theta}|y, \xi \sim N(\hat{\boldsymbol{\theta}}, [n\hat{I}(\boldsymbol{\theta}|\xi)]^{-1}),$$

where $\hat{\theta}$ denotes the maximum likelihood estimator of $\boldsymbol{\theta}$, $\hat{I}(\boldsymbol{\theta}|\xi)$ is the expected Fisher information matrix of parameter $\boldsymbol{\theta}$ for that mixed model under design ξ. For mixed

models, a closed-form expression of the expected Fisher information matrix is usually infeasible, thus an approximation could be used in the evaluation here. When the expected gain in Shannon information is considered, an approximation to the expected utility is

$$U_1(\xi) = -\frac{d}{2}\log(2\pi) - \frac{d}{2} + \frac{1}{2}\int \log\det\{(n\hat{I}(\boldsymbol{\theta}|\xi))\}p(\boldsymbol{\theta})d\boldsymbol{\theta},$$

so, Bayesian D-optimality criterion maximizes

$$\phi_1(\xi) = \int \log\det\{n\hat{I}(\boldsymbol{\theta}|\xi)\}p(\boldsymbol{\theta})d\boldsymbol{\theta}.$$

Then we look at the case where we consider loss functions. When it comes to the estimation of a function $g(\boldsymbol{\theta})$ and a squared error loss is appropriate, following $U_2(\xi)$, the approximate expected utility turns out be

$$\phi_2(\xi) = -\int c(\boldsymbol{\theta})'\{n\hat{I}(\boldsymbol{\theta}|\xi)\}^{-1}c(\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta},$$

where $c(\boldsymbol{\theta}) = (c_1(\boldsymbol{\theta}), \ldots, c_d(\boldsymbol{\theta}))'$, with $c_i(\boldsymbol{\theta}) = \partial g(\boldsymbol{\theta})/\partial\theta_i$.

If $\mathbf{C}(\boldsymbol{\theta})$ is a weighted sum of different matrices $c(\boldsymbol{\theta})c(\boldsymbol{\theta})'$ corresponding to different $g(\boldsymbol{\theta})$, $\phi_2(\xi)$ changes to

$$\phi_2(\xi) = -\int tr\{C(\boldsymbol{\theta})[n\hat{I}(\boldsymbol{\theta}|\xi)]^{-1}\}p(\boldsymbol{\theta})d\boldsymbol{\theta},$$

which is referred to as Bayesian A-optimality.

Sometimes we are only interested in estimation of a subset of $\boldsymbol{\theta}$, say $\boldsymbol{\theta}_s$, then the current expected information matrix becomes a sub matrix of $\hat{I}(\boldsymbol{\theta}|\xi)$. For example, $\phi_1(\xi)$ would become $\int \log\det\{n\hat{I}(\boldsymbol{\theta}_s|\xi)\}p(\boldsymbol{\theta})d\boldsymbol{\theta}$.

1.6 Designs with Experimental and Observational Variables

In this section, we discuss another important aspect in experimental designs, which is observational covariates, also called uncontrolled covariates. We will first go over existing results on this topic and briefly discuss this situation under GLMMs.

In clinical trials, responses usually do not only depend on the treatment and control, but also on many covariates, like age, gender and so on. These covariates could be able to be balanced at the subject selection stage, or only accessible after the subject selection but known before the treatments, or even unknown before the experiment is carried out. Some designs could even fail because of unknown or ignored but important covariates (Leyland-Jones 2003). To deal with each case properly, different methods and optimality criteria have been introduced and will be discussed here.

1.6.1 Designs with Known Observational Covariates

Based on the very fundamental idea of randomization theory in clinical trials, some covariate-adaptive and covariate-adaptive response-adaptive (CARA) procedures were introduced and are becoming more popular, Rosenberger and Sverdlov 2008 already contains a thorough introduction to such methods. Here we introduce the main framework of existing studies briefly. A covariate-adaptive randomization procedure, which belongs to the class of marginal procedures, can be described as follows for a two-treatment sequential clinical trial. Note that this sequential procedure will lead to an exact experimental design. Let $N_{iml}(n)$ be the number of patients on treatment l in level m of categorical covariate Z_i, $i = 1, \ldots, K, m = 1, \ldots, k_i, l = 1, 2$, after n

patients have been sequentially randomized. Denote the response of patient j by Y_j with treatment T_j, and $\boldsymbol{Y}_j = (Y_1, \ldots, Y_j)'$, $\boldsymbol{T}_j = (T_1, \ldots, T_j)'$. For the next patient, patient $n+1$, the covariates are observed as $(z_1, \ldots, z_K)$. Define $D(n) = \sum w_i D_i(n)$, where $D_i(n) = N_{iz_i1}(n) - N_{iz_i2}(n)$, and w_i be user-defined weights, then the treatment assignment of patient $n+1$ follows the general rule formulated by Pocock and Simon 1975 using the biased coin design of Efron 1971:

$$\phi_{n+1} = \begin{cases} 1/2, & \text{if } D(n) = 0; \\ p, & \text{if } D(n) < 0; \\ 1-p, & \text{if } D(n) > 0, \end{cases}$$

where ϕ_{n+1} is the probability of assigning treatment 1 to patient $n+1$. The choice $p = 3/4$ is discussed by Pocock and Simon 1975, and for $p = 1$ this procedure is deterministic (Taves 1974). Other variations of this procedure are suggested by Wei 1978 and Efron 1980, based on similar ideas. There is some concern about such procedures among statisticians, including that the procedure is not theoretically proved optimal and that it is ethically unattractive due to assigning an inferior treatment too often. Some statisticians even argue that we should not use such methods (Grouin, Day, and Lewis 2004).

Considering some optimality criteria like the discussion in the Section 1.1 about general optimal design approach, Harville 1974 suggested another method to balance the covariates through the covariance matrix of the estimated parameters, but only those associated with controlled variables. Considering exact designs, say a saturated design consists of t design points, then for $0 \leq s < t$, we generate the $(s+1)$th design point by balancing the number of design points in each covariate level group, like in the last paragraph. After t points have been determined, the incoming patients are assigned with treatments maximizing criteria based on the resulting information

matrix, like the determinant of the information matrix for D-optimality. Based on this idea, Atkinson 1982 suggested a randomized procedure and illustrated it for D_A optimality under linear models with L treatments as follows. Say the mean of the responses is modeled as $E(Y_i) = \boldsymbol{x}_i'\boldsymbol{\beta}, i = 1, \ldots, n$, where $\boldsymbol{x}_i$ includes both treatments and q covariates, and the variance of a given contrast $A'\hat{\boldsymbol{\beta}}$ is of interest, an optimality criterion would maximize $|A'(X'X/n)^{-1}A|^{-1}$. X is the design matrix formed by $\boldsymbol{x}_i$'s, determined by the design ξ. Then, the randomization procedure is based on the Fréchet derivative, $d_A(\boldsymbol{x}, \xi)$, of the determinant above. Given n patients assigned in design ξ_n, the $(n+1)$th patient will be assigned to treatment l with probability $\frac{\psi(d_A(l,\xi_n))}{\sum_{l=1}^{L}\psi(d_A(l,\xi_n))}$, where ψ is some monotone increasing function. This randomization procedure can be modified to be deterministic by always assigning treatment l when $d_A(l, \xi_n)$ is the largest for $l \in \{1, \ldots, L\}$. As stated by Rosenberger and Sverdlov 2008, the following CARA procedures can be superior to those covariate-adaptive procedures in some ways. Generally, a CARA procedure can be formulated by

$$\phi_j = P(T_{j+1} = 1 | \boldsymbol{T}_j, \boldsymbol{Y}_j, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_j, \boldsymbol{Z}_{j+1}),$$

where $\boldsymbol{Z}_j$ is the vector of Z_i values for patient j. Here the assignment of treatments also depends on the observed outcomes of previous patients. Rosenberger and Sverdlov 2008 indicated that with heterogeneous variances, CARA procedures can be more efficient and more ethically attractive (more assignments on the superior treatment) than covariate-adaptive procedures.

Besides these sequentially adaptive methods, there are also studies on non-sequential methods. In such studies, the number of patients or tested subjects is fixed and all covariate values are known. In such cases, we have all the uncontrolled covariate values known and can decide our treatments based on those, as a result, there are some good theoretical results proved. One of the earliest results is by Cook

and Thibodeau 1980, who consider optimality criteria based on information matrix under the restriction of those uncontrolled covariates and an equivalence theorem for D-optimal designs in such cases is presented. Harville 1974's sequential method mentioned above can be modified to fit such cases, by maximizing the criterion subject to both treatment assignments and order of patients coming into the experiment. Such a modified optimization starts from a saturated design, which could be generated by design-based adaptive randomization or just randomization. Following patients are added one-by-one or batch-by-batch into the optimization problem to maximize the criteria based on the information. Note that the "sequential" property only shows in optimization, and since we can alter the order of such a sequence, the resulting design is not a sequential design. In mathematics, partition the design matrix X into (X_1, X_2), the information of interested parameters can be expressed as $X_1'(I - X_2(X_2'X_2)^{-1}X_2')X_1$, so we maximize the determinant of this matrix subject to the choice of the next patient and his/her treatment. For observations with correlated errors, López-Fidalgo, Martín-Martín, and Stehlík 2008 discussed the solution for D-optimality, and constructed optimal designs for several models in real cases. The information matrix is derived based on different error structures, and then the determinant is maximized subject to the controlled variables, with the uncontrolled variables fixed and known.

The discussion above is based on the assumption that the values of the covariates are all known at the design stage, one-by-one or in the very beginning, and such an optimality is usually called marginally restricted (MR) optimality.

1.6.2 Designs with Unknown Observational Covariates

In practice, some important covariates are unknown until the experiment is carried out, like some real-time measurements of the patients in clinical trials. To model these covariates properly, the analysis is usually considered under conditioned values on these covariates and then try to properly average them. Such an optimization problem is usually called conditionally restricted (CR) optimality.

Tsukanov 1981 and Gupta and Richards 1985 first considered such optimization problems with the linear regression model

$$E(\boldsymbol{Y}_n|\boldsymbol{T}_n, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n) = X_k p_{(l_1\times 1)} + X_u q_{(l_2\times 1)},\ Cov(\boldsymbol{Y}_n|\boldsymbol{T}_n, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n) = \sigma^2 I.$$

X_k stands for the design matrix of the controlled variables, with ith row being $(f_1(T_i), \ldots, f_{l_1}(T_i))$. On the other hand, the ith row of X_u stands for the outcome of the ith uncontrolled variable, which is just the transpose of $\boldsymbol{Z}_i$. p and q are the corresponding parameter vectors.

First, each $\boldsymbol{Z}_i$ is assumed to be following the same normal distribution with mean μ_u and covariance Σ_u, which is independent of X_k. In both papers, the expectation of the aim functions in the optimality criteria with respect to the unknown covariates is optimized, which is called the optimality in the mean. For example, D-optimality in the mean (mD-optimality) minimizes $E(\det H)$ where H is the covariance matrix of the estimators conditioned on X_u. In this independent case, mA-optimality and mD-optimality coincides with A- and D-optimality, respectively, for estimation of p and $(p', q')'$.

Then, a linear dependence $X_u = X_k L + E$ is considered for a known linear-transformation matrix L and random normal error E. In this case, for estimating p,

by mA-optimality, the aim function satisfies

$$E\,tr\,H_p = \frac{n-l_1-1}{n-l_1-l_2-1} tr(X_k'X_k)^{-1} + \frac{1}{n-l_1-l_2-1} trL\Sigma_u^{-1}L',$$

where H_p is the covariance matrix of the estimator of p, conditioned on X_u. For mD-optimality and estimation of $(p', q')'$, mD-optimality still coincides with D-optimality, while for estimating p with $l_2 = 1$, by Tsukanov 1981, the aim function would satisfiy

$$E\,det\,H_p = det(X_k'X_k)^{-1}(\frac{L'X_k'X_kL\Sigma^{-1}}{n-l_1-2} + \frac{n-2}{n-l_1-2}). \tag{1.14}$$

Furthermore, Gupta and Richards 1985 extended equation 1.14 to all positive l_2 cases such that $n > l_1 + l_2 + 1$.

In the previous discussion, the estimation of q is not mentioned. For both the independent and dependent case with either mA- or mD-optimality criterion, the estimation of q is only dependent on the number of trials n and Σ_u.

Other than taking expectations on optimality criteria, López-Fidalgo and Garcet-Rodríguez 2004 considers the average design given known priors of the covariates for fixed effects models including nonlinear models, and derives a general equivalence theorem for ϕ-optimality. The model considered here is

$$E(Y_j|T_j, \boldsymbol{Z}_j) = f'(T_j, \boldsymbol{Z}_j)\theta,\ Cov(Y_j|T_j, \boldsymbol{Z}_j) = \sigma^2,$$

where f is a vector-valued function with mutually independent continuous components. For a simple example, a two-variable model, let x_1 be the known experimental variable, corresponding to T, and x_2 be a dependent unknown covariate, corresponding to $\boldsymbol{Z}$, but $\boldsymbol{Z}$ is actually a scalar now. The relationship of the two is defined by a known prior $p(x_2|x_1)$. Since we can only control x_1, an approximate design to be carried out is

$$\xi_1 = \begin{Bmatrix} x_{11}, & \ldots, & x_{1k} \\ w(x_{11}), & \ldots, & w(x_{1k}) \end{Bmatrix},$$

and the actual design in expectation would be

$$\xi_{12} = \{(x_{1i}, x_2) \text{ with probability } w(x_{1i})p(x_2|x_{1i}), i = 1, \ldots, k\}.$$

Based on ξ_{12}, the information matrix of θ is

$$\sum_{i=1}^{k} w(x_{1i}) \int p(x_2|x_{1i}) I(\theta|x_{1i}, x_2) dx_2,$$

where $I(\theta|x_{1i}, x_2)$ is the information matrix of θ given data (x_{1i}, x_2).

López-Fidalgo and Garcet-Rodríguez 2004 also discussed and derived a general equivalence theorem for the combination of MR and CR optimality, called MCR optimality. Another covariate X_3 is now added to the model and is known in the design stage. The discussion is following the last paragraph.

Chapter 2

TOPIC I: LOCALLY OPTIMAL DESIGNS FOR LOGISTIC MIXED MODELS WITH ONE COVARIATE

2.1 Locally Optimal Design Problem Illustration

In this topic, we consider model (1.3) in Section 1.2.2 for one covariate and a linear regressor. In a longitudinal study, for response from subject i under jth measurement, y_{ij}, $j = 1, \ldots, n_i$, satisfies

$$\mu_{ij}^{\mathbf{b}_i} = E(y_{ij}|\mathbf{b}_i) = \eta^{-1}(\mathbf{f}(x_{ij})^T\mathbf{b}_i) = \eta^{-1}(b_{i0} + b_{i1}x_{ij}), \tag{2.1}$$

where $\mathbf{b}_i = (b_{i0}, b_{i1})^T = \boldsymbol{\beta} + \boldsymbol{\alpha}_i$ consists of the subject-independent fixed-effects vector $\boldsymbol{\beta}$ and the random vector $\boldsymbol{\alpha}_i \sim N_2(0, \boldsymbol{\Sigma})$, i.e., $\mathbf{b}_i \sim N_2(\boldsymbol{\beta}, \boldsymbol{\Sigma})$. The covariance matrix $\boldsymbol{\Sigma}$ can be singular, allowing some effects to be fixed effects. The conditional variance of y_{ij} given $\mathbf{b}_i$ is $v_{ij}^{\mathbf{b}_i} = \phi a v(\mu_{ij}^{\mathbf{b}_i})$, where $v(\cdot)$ is a known function depending on $\eta(\cdot)$, a is a known constant, and ϕ is a dispersion parameter.

For different choices of the link function, η,

$$\eta[E(y_{ij}|\mathbf{b}_i)] = \begin{cases} \log(\frac{P(y_{ij}=1|\mathbf{b}_i)}{1-P(y_{ij}=1|\mathbf{b}_i)}), & \text{for logistic model;} \\ \Phi^{-1}(P(y_{ij}=1|\mathbf{b}_i)), & \text{for probit model;} \\ \log[E(y_{ij}|\mathbf{b}_i)], & \text{for Poisson model,} \end{cases} \tag{2.2}$$

with corresponding conditional variances

$$v_{ij}^{\mathbf{b}_i} = \text{var}(y_{ij}|\mathbf{b}_i) = \begin{cases} \frac{\exp(\mathbf{f}(x_{ij})^T\mathbf{b}_i)}{\{1+\exp(\mathbf{f}(x_{ij})^T\mathbf{b}_i)\}^2}, & \text{for logistic model;} \\ \frac{\{\Phi'(\mathbf{f}(x_{ij})^T\mathbf{b}_i)\}^2}{\Phi(\mathbf{f}(x_{ij})^T\mathbf{b}_i)\{1-\Phi(\mathbf{f}(x_{ij})^T\mathbf{b}_i)\}}, & \text{for probit model;} \\ \exp(\mathbf{f}(x_{ij})^T\mathbf{b}_i), & \text{for Poisson model,} \end{cases} \tag{2.3}$$

where $\Phi(\cdot)$ is the cumulative density function of the standard normal distribution. For my dissertation, the logistic link is mainly focused, but the methodology can also be applied to other link functions.

Model (2.1) can also be used for experiments that fit a GLM (with fixed effects) but that are run in a block design with random block effects (Waite and Woods 2015). Thus, the number of random components in $\mathbf{b}_i$ is equal to the number of blocks, and the matrix $\boldsymbol{\Sigma}$ would typically be a diagonal matrix. What is different in this setting is that the values of $x_{i1}, \ldots, x_{in_i}$ need not all be distinct. In contrast, in a longitudinal study, where these values refer to time, they cannot be the same. For now, we refrain from requiring that all values must be different, but we will return to this later. For simplicity, we will continue to refer to subjects rather than blocks.

With the goal of estimation of $\boldsymbol{\beta}$, we seek a population design $\xi = (\xi_1, \ldots, \xi_N)$ that is locally optimal for this goal. The likelihood function for $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ is given by $L_{Total}(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\mathbf{y}) = \prod_{i=1}^{N} L_i(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\mathbf{y}_i) = \prod_{i=1}^{N} \int \prod_{j=1}^{n_i} h(y_{ij}|\mathbf{b}_i = \boldsymbol{\beta} + \boldsymbol{\alpha}_i) g(\boldsymbol{\alpha}_i|\boldsymbol{\Sigma}) d\boldsymbol{\alpha}_i$, where $h(y_{ij}|\mathbf{b}_i)$ is the conditional density function of y_{ij} and $g(\boldsymbol{\alpha}_i|\boldsymbol{\Sigma})$ is the density function of random effects $\boldsymbol{\alpha}_i$ Molenberghs and Verbeke 2005. For simplicity, we will just write $L_{Total}(\boldsymbol{\beta}, \boldsymbol{\Sigma})$ and $L_i(\boldsymbol{\beta}, \boldsymbol{\Sigma})$. An approximation to the information matrix for $\boldsymbol{\beta}$ can be obtained by approximating $\text{cov}(\hat{\boldsymbol{\beta}})$ based on the estimation methods we are using. In general, for GLMMs there are no analytical expressions for $\text{cov}(\hat{\boldsymbol{\beta}})$. One exception is when y_{ij} follows a Poisson distribution with a log link function and only a random intercept (Niaparast 2009). Methods for approximating $\text{cov}(\hat{\boldsymbol{\beta}})$ include generalized estimating equations (GEE, cf. Liang and Zeger 1986 or Dean et al. 2015, Chapter 13), penalized quasi-likelihood (PQL) (Breslow and Clayton 1993), and marginal quasi-likelihood (MQL, cf. Breslow and Clayton 1993 or Dean

et al. 2015, Chapter 13) Ueckert and Mentré 2017 also applied adaptive Gaussian quadrature to approximate the information matrix directly.

Breslow and Clayton 1993 defined the integrated quasi-likelihood function for estimating $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, and applied the Laplace method for integral approximation to obtain PQL. Using their method, for model (2.1), an approximation of the variance-covariance matrix $\text{cov}(\hat{\boldsymbol{\beta}})$ is given by (Breslow and Clayton 1993, Tekle, Tan, and Berger 2008)

$$\text{cov}(\hat{\boldsymbol{\beta}}) = \text{cov}(\hat{\boldsymbol{\beta}}|\mathbf{b}) \approx \left(\sum_{i=1}^{N} \mathbf{F}_i^T \mathbf{U}_i^{-1} \mathbf{F}_i \right)^{-1}, \tag{2.4}$$

where $\mathbf{b} = (\mathbf{b}_1^T, \ldots, \mathbf{b}_N^T)^T$, $\mathbf{U}_i$, a weight matrix for the ith subject, equals $\mathbf{V}_i^{-1} + \mathbf{F}_i \boldsymbol{\Sigma} \mathbf{F}_i^T$, $\mathbf{F}_i$ is the design matrix of design ξ_i, i.e., $\mathbf{F}_i = \mathbf{F}_i(\xi_i) = (\mathbf{f}(x_{i1}), \ldots, \mathbf{f}(x_{in_i}))^T$, and $\mathbf{V}_i = \mathbf{V}_i(\xi_i) = \text{diag}(v_{i1}^{\mathbf{b}_i}, \ldots, v_{in_i}^{\mathbf{b}_i})$. For the design choice, "optimal choices" for the number of distinct covariate values x_{ij} for each subject, the values of these covariates, and the number of measurements at each of these values need to be determined. Since $\text{cov}(\hat{\boldsymbol{\beta}})$ in (2.4) depends on the unknown $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ through the $\mathbf{b}_i$'s, we will substitute values based on prior knowledge. This will eventually result in locally optimal designs. The inverse of (2.4) is denoted by $\mathfrak{M}_{Total}(\xi|\mathbf{b})$, and taking its expectation,

$$E_{\mathbf{b}}(\mathfrak{M}_{Total}(\xi|\mathbf{b}) = \sum_{i=1}^{N} E_{\mathbf{b}}(\mathbf{F}_i^T (\mathbf{V}_i^{-1} + \mathbf{F}_i \boldsymbol{\Sigma} \mathbf{F}_i^T)^{-1} \mathbf{F}_i), \tag{2.5}$$

gives the expected information matrix. Before going on, note that this expected information degenerates to the information of the corresponding GLM when the variance covariance matrix $\boldsymbol{\Sigma}$ goes to 0.

For convenience, we make a change in the notation for the population design $\xi = (\xi_1, \ldots, \xi_N)$. Since some subjects may receive the same design, we will write $\xi = \{(\xi_p, m_p), p = 1, \ldots, N_s\}$, where $\xi_p = (x_{pl}, l = 1, \ldots, n_p)$, $N_s \leq N$ is the number of distinct ξ_p's, m_p is the number of subjects receiving ξ_p, and $\sum_p m_p = N$. By the

independence between subjects, we can consider the approximate design approach. An approximate population design can be represented as

$$\xi = \{(\xi_p, w_p), p = 1, \ldots, N_s\}, w_p > 0, \sum_p w_p = 1. \tag{2.6}$$

The sub-designs ξ_p's are also known as individual designs. One could use the same individual design for each subject, i.e., $N_s = 1$. But as noted by Schmelter 2007, for exact individual designs, optimal designs may use more than a single individual design. The number of measurements on a subject, n_p, is in practice often the same for all subjects, and we will assume from now on that $n_p = n$. Based on (2.5), in the approximate design setting we define

$$\mathfrak{M}^{PQL}(\xi) = \mathfrak{M}^{PQL}(\xi|\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \sum_{p=1}^{N_s} w_p E_{\mathbf{b}}(\mathbf{F}_p^T(\mathbf{V}_p^{-1} + \mathbf{F}_p \boldsymbol{\Sigma} \mathbf{F}_p^T)^{-1}\mathbf{F}_p) \tag{2.7}$$

as an approximation to the information matrix for $\boldsymbol{\beta}$ under population design ξ. We will approximate the sum in (2.7) by approximating each expectation in that sum. Since all $\mathbf{b}_p$'s are assumed to be $N(\boldsymbol{\beta}, \boldsymbol{\Sigma})$, we approximate such an expectation by sampling S (e.g., $S = 1000$) $\mathbf{b}_p$'s from $N(\boldsymbol{\beta}, \boldsymbol{\Sigma})$ and taking the average.

The expression for $\mathfrak{M}^{PQL}(\xi)$, with its inverses, must be simplified for computational purposes. Miller 1981 proved the following theorem about the inverse of the sum of matrices.

Lemma 1 *(Miller 1981) Let A and $A + B$ be invertible matrices, with the rank of B equal to $r > 0$. Let $B = B_1 + \cdots + B_r$, where each B_i has rank 1 and, for $k = 1, \ldots, r$, each $C_{k+1} = A + \sum_{i=1}^{k} B_i$ is invertible. Setting $C_1 = A$, then*

$$C_{k+1}^{-1} = C_k^{-1} - g_k C_k^{-1} B_k C_k^{-1},$$

where $g_k = \frac{1}{1+tr(C_k^{-1}B_k)}$. In particular,

$$(A + B)^{-1} = C_r^{-1} - g_r C_r^{-1} B_r C_r^{-1}.$$

Based on Lemma 2.1, we can partition $\boldsymbol{\Sigma}$ into a sum of $\boldsymbol{\Sigma}_i$'s, where $\boldsymbol{\Sigma}_i$ is the matrix obtained from $\boldsymbol{\Sigma}$ by replacing all its elements by 0 except for those in the ith row. Then, the inverse of $\mathbf{V}_p^{-1} + \mathbf{F}_p\boldsymbol{\Sigma}\mathbf{F}_p^T$ in (2.7) can be computed by multiplication and summation of matrices, which is much faster when searching for optimal designs.

Focusing more on the population averages of the covariate effects, marginal model can also be applied to estimate the fixed effects. We assume $n_p \equiv n$ in the following discussion. Breslow and Clayton 1993 introduced MQL approximation to the information matrix for subject i as

$$\mathfrak{M}^{MQL}(\xi) = \sum_{p=1}^{N_s} w_p \mathbf{F}_p^T (\mathbf{V}_p^{-1} + \mathbf{F}_p\boldsymbol{\Sigma}\mathbf{F}_p^T)^{-1}\mathbf{F}_p,$$

where $\mathbf{V}_p = \text{diag}(v_{p1}^{\boldsymbol{\beta}}, \ldots, v_{pn}^{\boldsymbol{\beta}})$, and $v_{pj}^{\boldsymbol{\beta}} = \text{var}(y_{pj}|\boldsymbol{\beta})$. Zeger, Liang, and Albert 1998 showed that a better approximation to the marginal mean for logistic link has attenuated coefficients, which means that an attenuated value $\boldsymbol{\beta}_{pj}^a = \boldsymbol{\beta} \cdot c_{pj} = \boldsymbol{\beta} \cdot |1 + c^2 \boldsymbol{f}^T(x_{pj})\boldsymbol{\Sigma}\boldsymbol{f}(x_{pj})|^{-1/2}$ should be used in the expression, where $c = 16\sqrt{3}/(15\pi)$. Based on this result, adjusted MQL approximation is given by

$$\mathfrak{M}^{aMQL}(\xi) = \sum_{p=1}^{N_s} w_p \mathbf{F}_p^T (\tilde{\mathbf{V}}_p^{-1} + \mathbf{F}_p\boldsymbol{\Sigma}\mathbf{F}_p^T)^{-1}\mathbf{F}_p,$$

where $\tilde{\mathbf{V}}_p = \text{diag}(v_{p1}^{\boldsymbol{\beta}_{p1}^a}, \ldots, v_{pn}^{\boldsymbol{\beta}_{pn}^a})$. Note that for random-intercept model, which is more common for block designs, c_i could be irrelevant of the values of the covariates, leading to a same transform for all design points, i.e., we can set $\boldsymbol{\beta}_p^a \equiv \boldsymbol{\beta}_{pj}^a$.

Other than approximations, the information matrix can also be obtained numerically from the definition. The expectation in Fisher information can be obtained by enumerating all possible outcomes based on a given design, with the probabilities being integrated out numerically. The contribution to the likelihood function by a single subject can also be obtained by enumerating all possible outcomes. For the

logistic link in model (2.1), the likelihood function for a subject, say i, who is assigned to individual design ξ_p is

$$L_i(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \iint \prod_{l=1}^{n} \frac{\exp[\mathbf{f}(x_{pl})^T \mathbf{b}_i]^{y_{il}}}{1 + \exp[\mathbf{f}(x_{pl})^T \mathbf{b}_i]} \Phi(\mathbf{b}_i; \boldsymbol{\beta}, \boldsymbol{\Sigma}) d\mathbf{b}_i.$$

This corresponds to a contribution to the information matrix given by

$$\begin{aligned} \mathfrak{I}_i(\xi_p|\boldsymbol{\beta}, \boldsymbol{\Sigma}) &= E_{\mathbf{y}_i}\Big(\frac{\partial \log L_i(\boldsymbol{\beta}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\beta}} \frac{\partial \log L_i(\boldsymbol{\beta}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\beta}'}\Big) \\ &= \sum_{\mathbf{y}_i} \frac{1}{L_i(\boldsymbol{\beta}, \boldsymbol{\Sigma})} \frac{\partial L_i(\boldsymbol{\beta}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\beta}} \frac{\partial L_i(\boldsymbol{\beta}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\beta}'}. \end{aligned} \tag{2.8}$$

Using numerical integration, this expression can serve as a benchmark for other approximations of the information matrix.

2.2 Comparison of Approximations to Information Matrices under GLMM

Before searching for optimal designs, we compare different approximations numerically to ensure that we use a good proxy for the exact information matrix. We do this for a fixed set of designs and multiple parameter settings for a logistic model. The model has $\eta(x_{ij}) = b_{i0} + b_{i1}x_{ij}$ with parameter of interest $\beta = (\beta_0, \beta_1)^T$. We focus on the D- and A-optimality criteria, say ϕ_D and ϕ_A, respectively, where $\phi_D = \log\det(\mathfrak{I}(\xi))$, and $\phi_A = 1/\mathrm{tr}(\mathfrak{I}^{-1}(\xi))$. If we use an approximation for $\mathfrak{I}$, then the approximation, such as $\mathfrak{M}^{PQL}$, replaces $\mathfrak{I}$ in these expressions. To use a sufficiently broad spectrum of "guesses" for $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, we consult the complete class result for fixed effects models in Yang and Stufken 2009.

Theorem 2 *(Yang and Stufken 2009) For a fixed effects GLM formulated by* $\mu = E(y) = \eta^{-1}(\boldsymbol{f}(x)^T\boldsymbol{\beta})$, *set* $c_j = \eta(\mu_j)$. *Suppose that the design region in the c-space takes one of the following forms:*

1. $[-D, D]$ *for some* $D > 0$*; or*
2. $[D_1, D_2]$ *with (a)* $D_1 \geq 0$ *or (b)* $D_2 \leq 0$*; or*
3. $[D_1, D_2]$ *with* $D_1 < 0 < D_2$ *and* $-D_1 < D_2$*; or*
4. $[D_1, D_2]$ *with* $D_1 < 0 < D_2$ *and* $-D_1 > D_2$*.*

For these four cases, a complete class can be formed by designs that are

1. *supported on two symmetric points;*
2. *supported on two points, including* D_1 *if* $D_1 \geq 0$ *and including* D_2 *otherwise;*
3. *supported on two symmetric points, or on* D_1 *and a point in* $(-D_1, D_2]$*; and*
4. *supported on two symmetric points, or on* D_2 *and a point in* $[D_1, -D_2)$*,*

respectively.

For our comparisons we use a design region of $[1, 6]$ in x-space. We select values for $\boldsymbol{\beta}$, as shown in Table 1, that represent each case in Theorem 2. For $\boldsymbol{\Sigma}$, we use a fixed correlation $\rho = .5$ and 3 choices for the diagonals (I, II, and III in Table 1). The population designs that we use all consist of a single individual design, that is, $N_s = 1$. But, based on the additive expressions for the approximations, a good approximation for $N_s = 1$ implies a good approximation for moderately larger values of N_s. For the single individual design, we use every design that is supported on exactly two points from $\{1, 2, 3, 4, 5, 6\}$ with a total number of measurements equal to $n = 10$. There are $9 \times \binom{6}{2} = 135$ such designs. The sample size S for approximating the expectation in (2.7) by a random sample is set to 1000, corresponding to the red line in Figure 1 and 2. The sample size for support points is set to 100, corresponding to the dark blue line in the two figures.

For $n = 10$, there are $2^{10} = 1024$ possible outcomes of $\mathbf{y}$. Using Mathematica, it takes about 800 seconds of CPU time to evaluate $\mathfrak{I}(\xi_i|\boldsymbol{\beta}, \boldsymbol{\Sigma})$ defined in (2.8) once on

Choices for (β_0, β_1)	$(1, -1)$	$(3, -1)$	$(7/2, -1)$	$(4, -1)$	$(6, -1)$
Case in Theorem 2	2 (b)	4	1	3	2 (a)
Diagonals of $\boldsymbol{\Sigma}$	I: (1.7145, 1.05)		II: (6, .3)		III: (.3, 6)

Table 1. Parameter Choices for Comparisons of the Approximations

a 3.9 GHz Intel CPU and 24 GB RAM. We can also get rid of the slow numerical integration by using a sample from the density to approximately evaluate the result, which is much faster, but still takes 9 seconds with a size 500 sample. This is too slow for finding optimal designs, where this computation would have to be performed many times. In comparison, the PQL and MQL or adjusted MQL approximations take .1075 and 6.3×10^{-5} CPU-seconds, respectively.

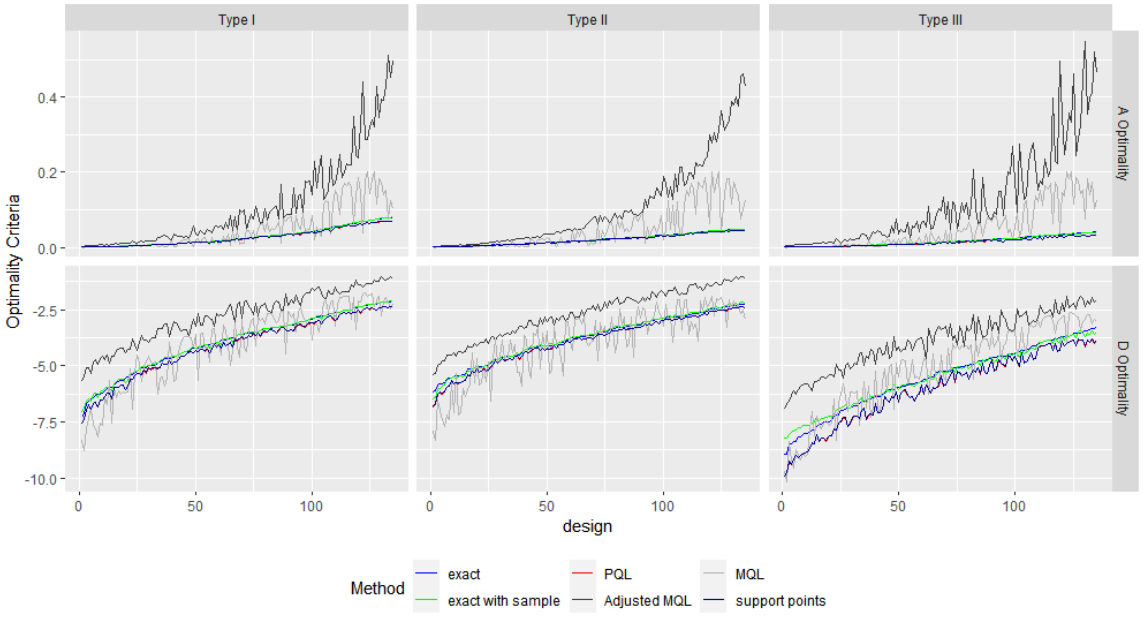

Figure 1. Comparing PQL, MQL and Adjusted MQL Approximations to the Exact Information Matrix for $\boldsymbol{\beta} = (1, -1)'$

Figures 1 and 2 show results of the comparisons for two of the five cases for $\boldsymbol{\beta}$ in Table 1. For both figures, the labels on the horizontal axis represent the 135 designs,

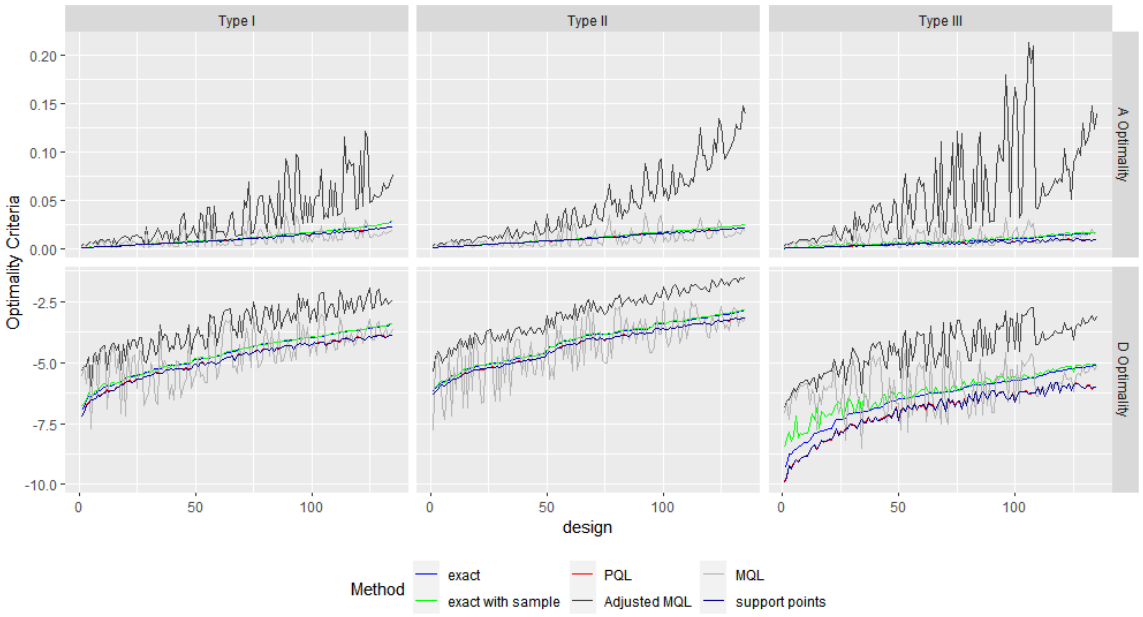

Figure 2. Comparing PQL, MQL and Adjusted MQL Approximations to the Exact Information Matrix for $\boldsymbol{\beta} = (6, -1)'$

ordered by the criterion value (A-optimality or D-optimality) for the exact information matrix (blue line). Among the three approximations, PQL (red line) clearly has the smallest absolute error and orders the designs approximately correctly based on both the D- and A-optimality criterion. For MQL (light grey line) and adjusted MQL (dark grey line), while they show the correct trends, the approximations are relatively poor and, more importantly, the ordering of the designs that they provide is also rather poor with the possible exception of adjusted MQL for D-optimality. Thus, by far the best approximation is provided by $\mathfrak{M}^{PQL}(\xi)$. Only for cases when this is computationally too expensive might one want to consider $\mathfrak{M}^{aMQL}(\xi)$.

2.3 Simulation: Locally Optimal Designs under GLMMs

Based on Section 2.2, for given $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, we use the PQL approximation instead of the information matrix and Particle Swarm Optimization (PSO) to find locally A- or D-optimal designs for the mixed-effects logistic model. PSO is widely used in optimal design research (Zhou, Wang, and Yue 2021). To illustrate the results, we use the same settings for $\boldsymbol{\beta}$ as in Table 1. We take $\boldsymbol{\Sigma} = r \cdot \boldsymbol{\Sigma}_0$ as a diagonal matrix, where $r \in \{5, 7, 10, 15, 25\}$ and $\boldsymbol{\Sigma}_0 = \text{diag}(\sigma_1^2, \sigma_2^2)$ with $(\sigma_1^2, \sigma_2^2) = (.1143, .07), (.4, .02),$ or, $(.02, .4)$. We refer to these three choices for (σ_1^2, σ_2^2) as type I, II and III, respectively. Note that, for fixed r, the generalized variance, $\det(r \cdot \boldsymbol{\Sigma}_0)$, is approximately the same for all three types. The three cases in Section 2.2 correspond to the choice $r = 15$. Finally, we use the same design region of $[1, 6]$ as in Section 2.2.

Due to correlated observations, neither the information matrix nor its approximations are additive for observations on the same subject. Additivity does hold when combining information from different subjects. Therefore, using an approximate design approach, we use a population design $\xi = \{(\xi_p, w_p), p = 1, \ldots, N_s\}$, where $w_p > 0$, $\sum_p w_p = 1$, and each individual design ξ_p is an n-point design for a fixed value of n. We do not insist that the n points in the design region $[1, 6]$ are distinct. In a situation where it is not possible to have repeated values (for example when the design variable denotes the time at which an observation is to be made in a longitudinal study), we can distribute the design points around the repeated value. Alternatively, if there is a specified minimum distance d_0 between any two design points, we can build this constraint into the PSO algorithm. We considered optimal designs for $n = 2$

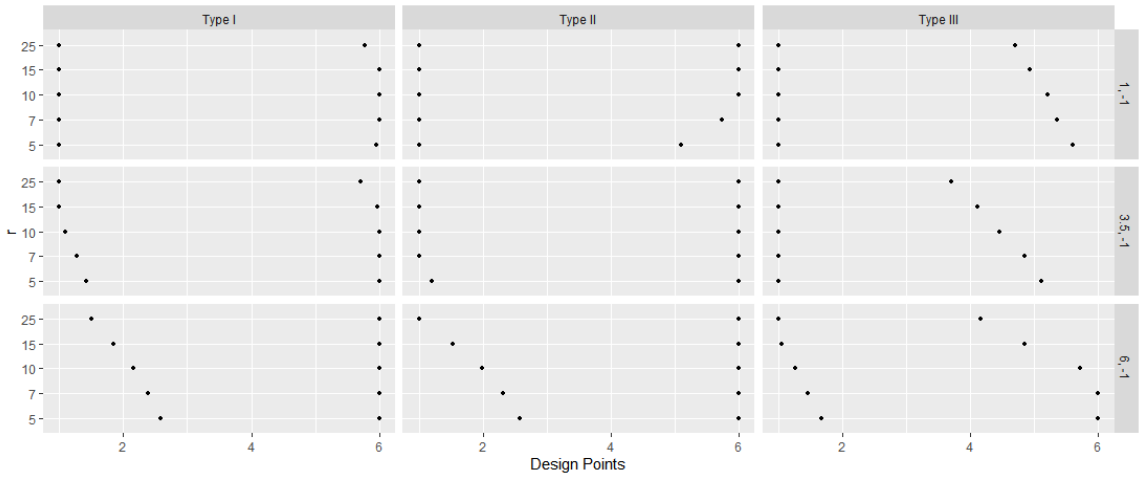

Figure 3. D-optimal Designs with $n = 2$ Observations per Subject. The Covariance Type Is Shown in the Top Bar and the Choice for $\boldsymbol{\beta}$ in the Right-hand Bar. The Value of r for the Covariance Matrix $\boldsymbol{\Sigma}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.

through 6, but will only show those for $n = 2$ and 5 here. Other designs can be found in Appendix A.1.

Since we use exact individual designs, the optimization problem is a discrete problem, and we can no longer use a general equivalence theorem to verify optimality of a design. Even though, we are not able to guarantee that our designs are indeed optimal; by running the PSO algorithm multiple times with different starting designs and a large number of iterations, we are confident that the population designs are very efficient.

In Figures 3 through 6 we show D- and A-optimal population designs for $n = 2$ and 5. In all cases, the design either has $N_s = 1$ or 2. When $N_s = 1$, which occurs most often, we simply present the n points of the corresponding individual design ξ_1. For $n = 5$, repeated points are presented slightly apart so that the number of points

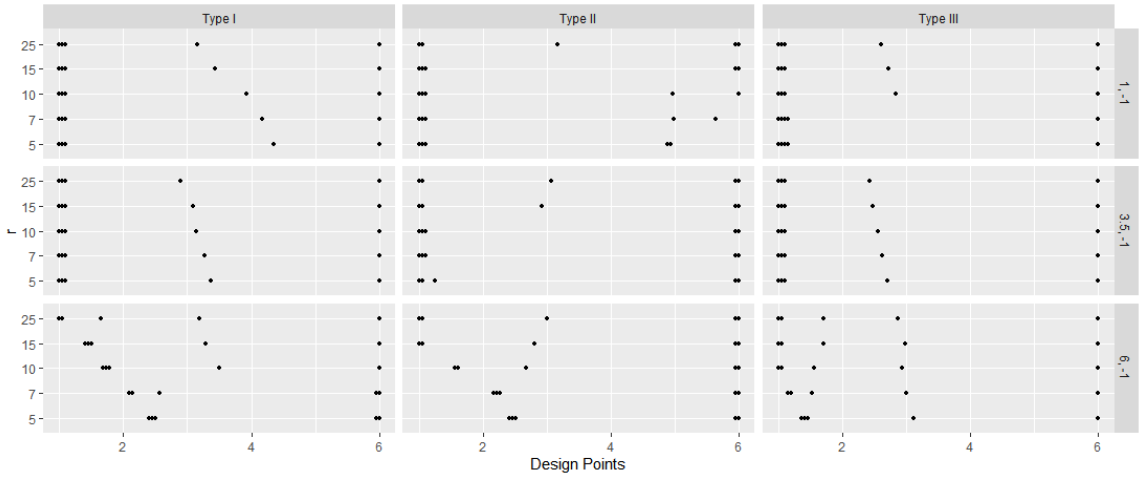

Figure 4. D-optimal Designs with $n = 5$ Observations per Subject. The Covariance Type Is Shown in the Top Bar and the Choice for $\boldsymbol{\beta}$ in the Right-hand Bar. The Value of r for the Covariance Matrix $\boldsymbol{\Sigma}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.

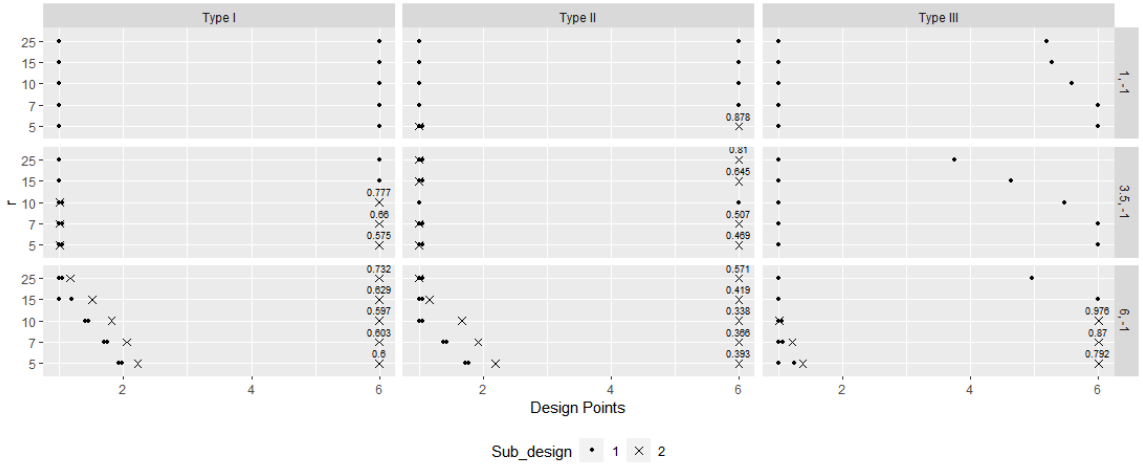

Figure 5. A-optimal Designs with $n = 2$ Observations per Subject. The Covariance Type Is Shown in the Top Bar and the Choice for $\boldsymbol{\beta}$ in the Right-hand Bar. The Value of r for the Covariance Matrix $\boldsymbol{\Sigma}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.

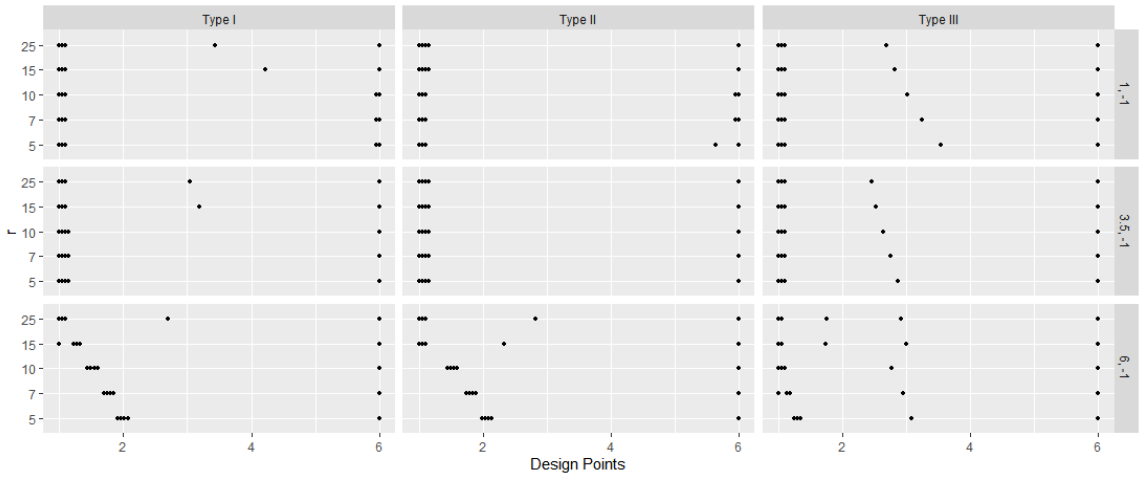

Figure 6. A-optimal Designs with $n = 5$ Observations per Subject. The Covariance Type Is Shown in the Top Bar and the Choice for $\boldsymbol{\beta}$ in the Right-hand Bar. The Value of r for the Covariance Matrix $\boldsymbol{\Sigma}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.

	(1, -1)	(3.5, -1)	(6, -1)
D	(1,3.399;.5,.5)	(1.957,5.043;.5,.5)	(3.601,6;.5,.5)
A	(1,4.009;.553,.447)	(1.211,5.789;.789,.211)	(2.920,6;.828,.172)

Table 2. Optimal Designs for (Fixed Effects) GLM under Corresponding Parameters

at a location is discernible. For $n = 2$ and A-optimality, some cases yield $N_s = 2$. Different plotting symbols are then used for the two individual designs ξ_1 and ξ_2, and weights are shown for ξ_2 next to one of the design points.

There are several noteworthy features in the figures. First, for $n = 5$, the number of distinct design points increases with the value of r. For small r the optimal designs are 2-point designs. This is not surprising because this is known for the limiting case $r \to 0$, which corresponds to a (fixed-effects) GLM (Yang and Stufken 2009). Optimal designs for GLMs are shown in Table 2 for reference.

For both A- and D-optimality, the number of distinct design points tends to

increase with r, albeit it that the increase can be limited depending on the design region. This increase can, for example, be seen in results for $\boldsymbol{\beta} = (6, -1)'$ in Figures 4 and 6. For other $\boldsymbol{\beta}$'s in these figures, a larger design region would exhibit a similar pattern. Second, for $\boldsymbol{\beta} = (1, -1)'$ or $(6, -1)'$, corresponding to cases 2(b) and 2(a) in Theorem 2, for a GLM one of the endpoints of the design region must be included in an optimal design; this need not be the case for the mixed-effects model (see, for example, Figure 3 with $\boldsymbol{\beta} = (6, -1)'$ for type III and r=10). Third, while the results in Schmelter 2007 do not apply here, optimal population designs still consist often (but not always) of the same individual design for all subjects. For A-optimality, Figure 7 shows efficiencies of the best one-sequence designs for $n = 2$ as measured by $\frac{\phi_A(\mathfrak{M}^{PQL}(\xi))}{\phi_A(\mathfrak{M}^{PQL}(\xi^*))}$, where ϕ_A denotes the A-optimality criterion and ξ^* is an A-optimal design. For large enough r, one-sequence designs are optimal. They are however also quite efficient for smaller values of r, although the efficiencies vary with $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$.

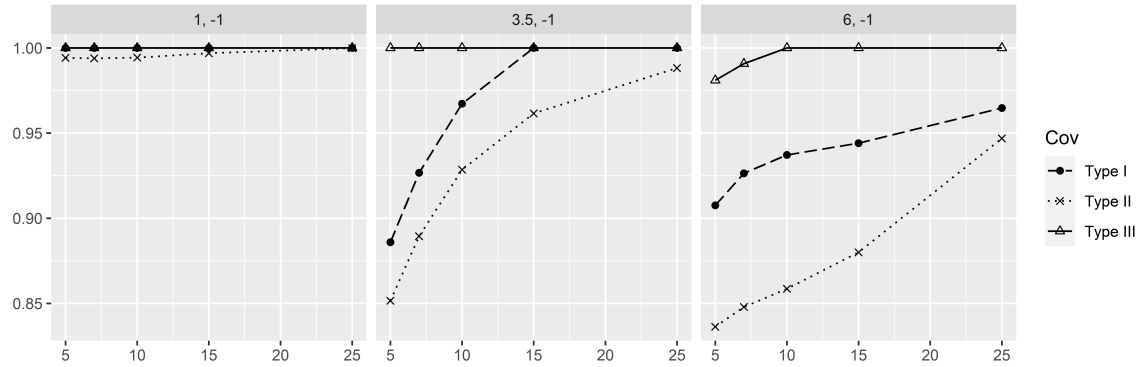

Figure 7. Efficiencies of the Best One-sequence Designs for $n = 2$ for the Conditions in Figure 5. The Axes Show the Values of r (Horizontal) and the Efficiency (Vertical). The Top Bar in Each Panel Shows $\boldsymbol{\beta}$, and Different Line Types Correspond to the Three Covariance Types.

Finally, for our PSO algorithm, the number of measurements n per subject must be specified. For the optimal designs reported here, we use 20 particles and 500 iterations

to get a near optimal design ξ_a, and then use a new random start, keeping ξ_a as the initial global optimum, with another 1500 iterations. The final design may contain replicated points. While not done here, if these replicated points are undesirable, we can re-optimize using PSO again, keeping one replicate of the distinct design points and forcing a minimum distance d_0 between any two points. For our examples, using $d_0 = .25$ or .5, this results in replacing the replicated points by the nearest points with distance d_0 as in Hu and Stufken 2017. We return to this in the next section.

2.4 An Application and Robustness Analysis for Locally Optimal Designs under GLMM

For illustration of the methodology, we focus on the French EPIDOS study (Epidétiology de l'ostéoporose), a prospective multi-centre study of the risk factors for hip fractures in women who were 75 years or older in 1992-1993. The participating women completed health-related questionnaires annually for six years. Carrière and Bouyer 2002 analyzed the data from Montpellier, one of the 5 participating centers, using a generalized linear mixed model with a logistic link function as in model (2.1). After testing the significance of random effects, the authors finally determined the covariance structure with random intercept and random slope. Let y_{ij} be the jth indicator of disability, "needing help to go outdoors or home-confined", for woman i ($i = 1, \ldots, 1548$; $j = 1, \ldots, 6$). The proposed logistic mixed-effects model is

$$\text{logit}[E(y_{ij}|\mathbf{b}_i)] = \log(\frac{P(y_{ij} = 1|\mathbf{b}_i)}{1 - P(y_{ij} = 1|\mathbf{b}_i)}) = b_{i0} + b_{i1}x_{ij}, \tag{2.9}$$

where, in this study, the design points are $x_{ij} = j,\ j = 1, \ldots, 6$; $\mathbf{b}_i = (b_{i0}, b_{i1})^T \sim N_2(\boldsymbol{\beta}, \boldsymbol{\Sigma})$, $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$, and $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix}$. The estimated parameters are displayed in Table 3.

Parameter	β_0	β_1	σ_{11}	σ_{22}
Estimate	-3.61	0.17	7.25	0.18

Table 3. Parameter Estimates for French EPIDOS Study

But is an equally spaced design optimal? Using a PSO algorithm and the estimated parameters in Table 3 with design space $[1, 6]$ (consistent with the original study), locally optimal designs for this model are presented in Table 4. Both A- and D-optimal designs are one-sequence designs.

Optimality	x_1	x_2	x_3	x_4	x_5	x_6
A	1	1	1	1	1	6
D	1	1	1	6	6	6

Table 4. A- and D-optimal Designs ξ^* for Self-reported Disability Study

From Table 4, the A-optimal design requires each woman to complete 5 questionnaires at the first time point, which is clearly infeasible. We could simply replace replicated points by nearest neighbors, with all design points at least separated by a specified distance d_0 (Hu and Stufken 2017). Alternatively, as already suggested, we can use PSO again keeping one copy of the replicated point and forcing a distance of d_0 between any two design points. In this example, these two methods provide answers. Using half a year, a quarter of a year, and one month for d_0, the obtained

designs are ξ_1, ξ_2 and ξ_3, respectively, where

$$\xi_1 = \begin{cases} \{1.0, 1.5, 2.0, 2.5, 3.0, 6.0\}, & \text{A-optimality} \\ \{1.0, 1.5, 2.0, 5.0, 5.5, 6.0\}, & \text{D-optimality,} \end{cases}$$

$$\xi_2 = \begin{cases} \{1.00, 1.25, 1.50, 1.75, 2.00, 6.00\}, & \text{A-optimality} \\ \{1.00, 1.25, 1.50, 5.50, 5.75, 6.00\}, & \text{D-optimality,} \end{cases}$$

and

$$\xi_3 = \begin{cases} \{1.00, 1.08, 1.17, 1.25, 1.33, 6.00\}, & \text{A-optimality} \\ \{1.00, 1.08, 1.17, 5.83, 5.92, 6.00\}, & \text{D-optimality.} \end{cases}$$

The relative efficiencies of these designs and design $\xi_0 = \{1, 2, 3, 4, 5, 6\}$ in the original study relative to the optimal design ξ^* are shown in Table 5.

Optimality	ξ_0	ξ_1	ξ_2	ξ_3
A	82.57	92.60	97.03	99.25
D	94.22	98.12	98.98	99.62

Table 5. Efficiencies in Percentage

Table 5 indicates that designs will remain highly efficient if d_0 is not large. There is more loss of efficiency under A-optimality because the proposed replacement for the A-optimal design in Table 4 moves us further from that design since it places more emphasis on one of the endpoints.

Locally optimal designs depend on "guessed" parameters, and poor guesses may lead to poor designs. An important part to consider is robustness to misspecification of the variance-covariance matrix $\mathbf{\Sigma}$. With the notation from Section 2.3, let ξ_r denote a locally optimal design under A- or D-optimality for $\mathbf{\Sigma}_r = r \cdot \mathbf{\Sigma}_0$ and a given vector $\boldsymbol{\beta}$. The relative efficiencies e_{lm} of design ξ_{r_l} to design ξ_{r_m} when $\boldsymbol{\beta} = (1, -1)'$ or $(6, -1)'$ and $\mathbf{\Sigma}_{r_m}$, with $r_l, r_m \in \{0, .01, .05, .1, .5, 1, 3, 5, 7, 9, 11, 13, 15, 17,$
$19, 21\}$, were calculated. For clarity of the figures, results are shown for only six

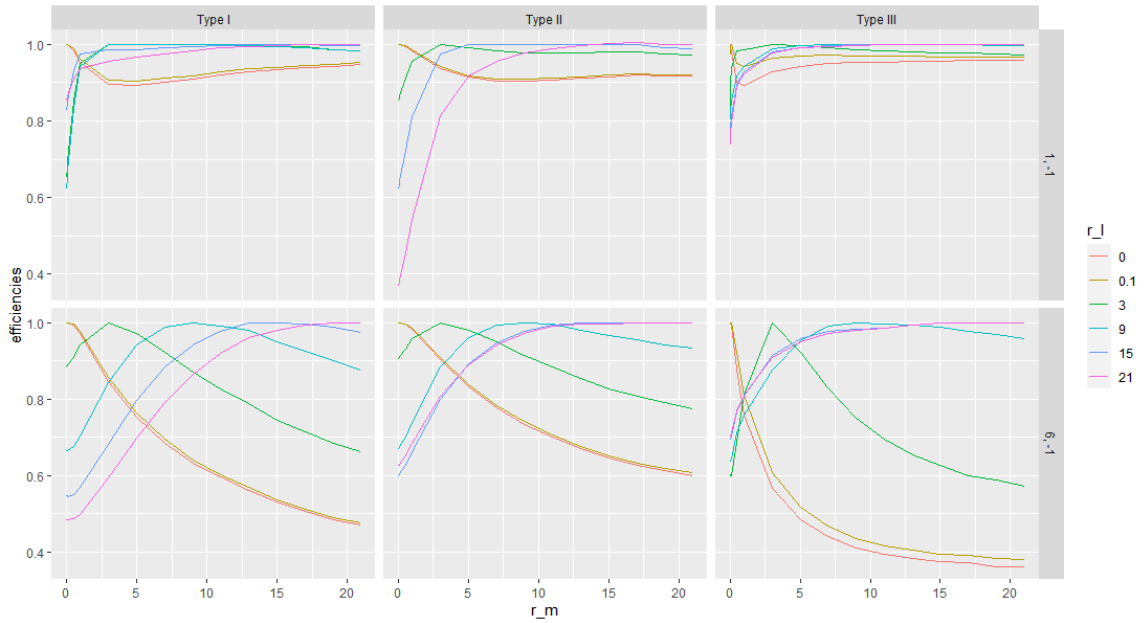

Figure 8. Robustness Study: Efficiencies of A-optimal Designs, $n = 5$. The Covariance Type Is Shown in the Top Bar and the Choice for $\boldsymbol{\beta}$ in the Righthand Bar. Value of r_m Is along the Horizontal Axis, Which Corresponds to the True Covariance Matrix, and Values of r_l Are Represented by the Six Different Lines, Each Corresponds to the Optimal Design under Such r_l. The Efficiencies of These Designs Are Shown along the Vertical Axis.

designs ($r_l = 0, .1, 3, 9, 15, 21$) in Figures 8 and 9 using $n = 5$. Notice that $e_{ll} = 1$ for all l. From the figures we can get some useful findings. When the true r, which is r_m, is very small, like 0.01, an overestimated r_l will always cause significant loss in efficiency. In this case we are actually misspecifying a GLM with a GLMM, the huge error is expected. Interestingly, in some of the cases with $\boldsymbol{\beta} = (1, -1)'$, if we use GLM as the model for design, from the line for r_l =0.01, we can see it is actually performing well, with the efficiency even going up when r_m increases. If the true model is very different from a GLM, overestimating or underestimating will not cause much loss in efficiency. See the case where r_l is 9, if we ignore the beginning part of x-axis, this line almost never go under 80 percent. Then, we can look at some specific cases. For $\boldsymbol{\beta} = (6, -1)'$, we should be very careful about too small or too large guesses to avoid

potential loss when the underlying truth is just opposite. On the other hand, for $\boldsymbol{\beta} = (1, -1)'$, we can trust a small guess without worrying about potential loss. In both figures, some of the lines go slightly higher than 1, like the Type II, $\boldsymbol{\beta} = (1, -1)'$ case in Figure 8, this is because the designs given by PSO are actually near optimal designs, and there is no guarantee of the optimality without generalized equivalence theorem.

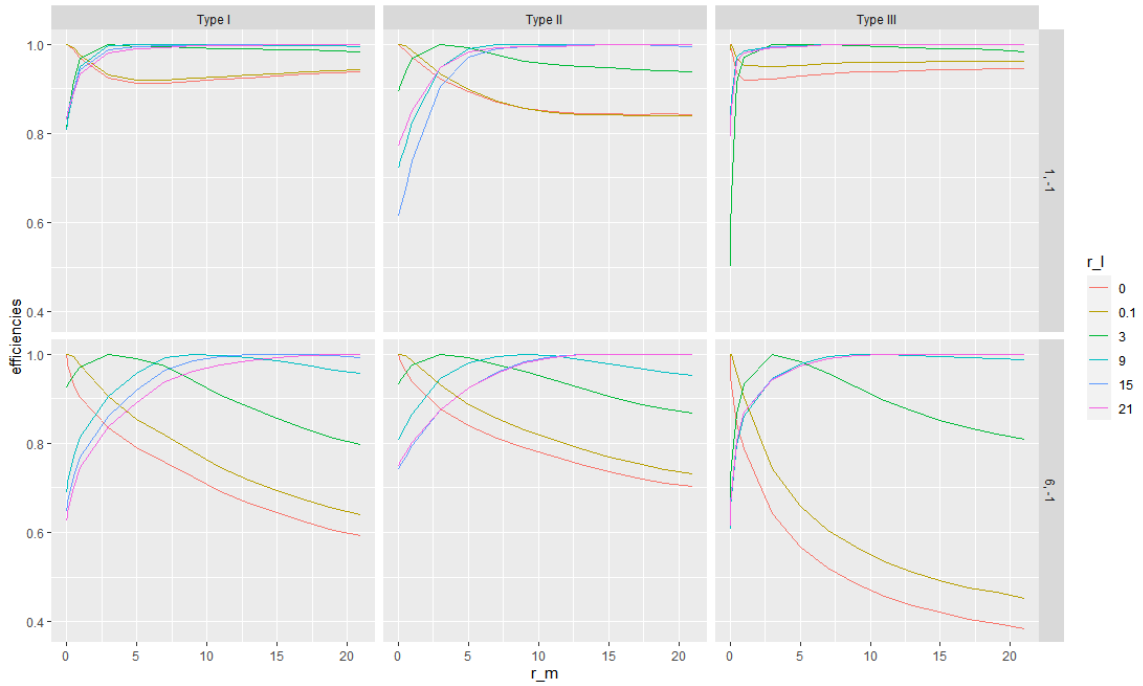

Figure 9. Robustness Study: Efficiencies of D-optimal Designs, $n = 5$. The Covariance Type Is Shown in the Top Bar and the Choice for $\boldsymbol{\beta}$ in the Right-hand Bar. Value of r_m Is along the Horizontal Axis, Which Corresponds to the True Covariance Matrix, and Values of r_l Are Represented by the Six Different Lines, Each Corresponds to the Optimal Design under Such r_l. The Efficiencies of These Designs Are Shown along the Vertical Axis.

Another type of mis-specification we study here is based on a Wishart distribution. Assume that a matrix $\mathbf{G}$ is generated from from a Wishart distribution $G_2(\boldsymbol{\Sigma}, df)$ with $\boldsymbol{\Sigma}$ being the true covariance, set to be $\text{diag}(1.7145, 1.05)$, and selected degrees

of freedom df, df =2, 4, 7 and 15. Then, use $\mathbf{G}/r$ as the guess of the covariance in searching for an optimal design. For each df, 100 $\mathbf{G}$ are generated. Note that the coefficient variation of the diagonals of $\mathbf{G}/r$ is $\sqrt{2/r}$, so the sample has a smaller variation when r is larger. The fixed effects guesses are set to be $\boldsymbol{\beta} = (1, -1)'$, then the efficiencies of these designs comparing to the true optimal design are shown in Figure 10. Based on this box plot, we can see that the designs are quite robust. When the coefficient variation is largest, corresponding to $r = 2$, the minimal efficiency is still around .875, with median above .975, for both optimality criteria.

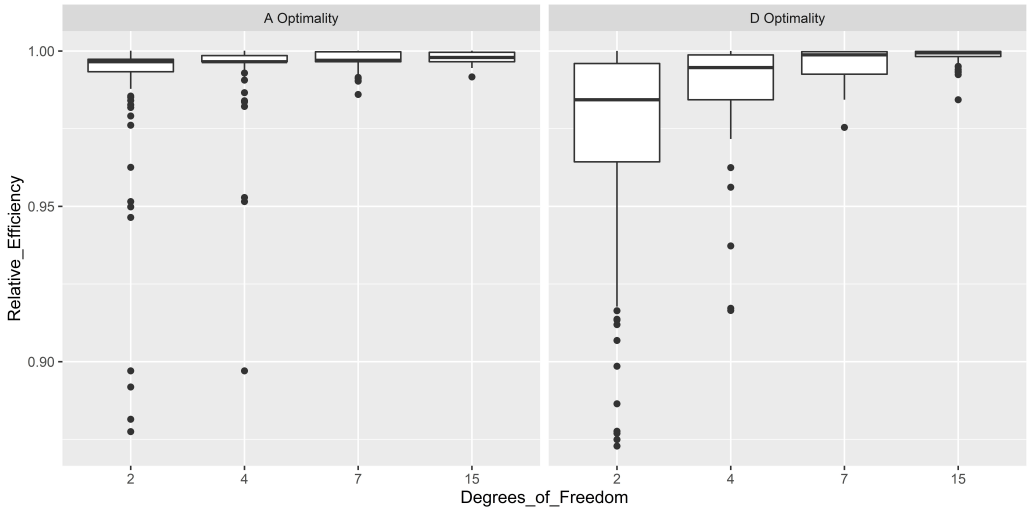

Figure 10. Efficiencies of D- and A-optimal Designs with Mis-specified $\boldsymbol{\Sigma}$ from Wishart Distributions. The Optimality Criterion Is Shown on the Top Bar. The Degrees of Freedom in the Wishart Distribution Are Shown along the Horizontal Axis, and Efficiency Is along the Vertical Axis.

2.5 Summary and Discussion

We found optimal designs under logistic mixed model by approximating the information matrices using the PQL method. After evaluating the approximation

by comparing it to the information matrix by basic definition and MQL, based on logistic link, optimal designs are found so that we can see some characteristics and also the relationship with optimal designs under corresponding GLMs. More-than-one-sequence designs are suggested for some parameter settings under GLMMs, which is different from the GLM case, but two sequences are quite enough to get optimal for exact designs.

Chapter 3

TOPIC II: LOCALLY OPTIMAL DESIGNS FOR LOGISTIC MIXED MODELS WITH ONE EXPERIMENTAL VARIABLE AND ONE DEPENDENT OBSERVATIONAL COVARIATE

3.1 Model Introduction

In many design circumstances, variables are not all subject to control, known as observational variables, and it is important to consider them. Section 1.6 introduced such cases where uncontrolled variables need to be carefully dealt with. Following the model description in Section 2.1, the model discussed here can be described as following.

In a longitudinal study, assume that the responses depend on two explanatory variables, x_1 and x_2. x_1 is subject to control, while x_2 is not and unknown until the experiment is carried out. Assume that x_2 is dependent on x_1, through density $p(x_2|x_1)$. The response from subject i under jth measurement, y_{ij}, $j = 1, \ldots, n$, satisfies

$$\mu_{ij}^{\mathbf{b}_i} = E(y_{ij}|\mathbf{b}_i) = \eta^{-1}(b_{i0} + b_{i1}x_{1,ij} + b_{i2}x_{2,ij}). \tag{3.1}$$

Write $\mathbf{b}_i = (b_{i0}, b_{i1}, b_{i,2})^T = \boldsymbol{\beta} + \boldsymbol{\alpha}_i$ consists of the subject-independent fixed-effects vector $\boldsymbol{\beta}$ and the random vector $\boldsymbol{\alpha}_i \sim N_3(0, \boldsymbol{\Sigma})$, i.e., $\mathbf{b}_i \sim N_3(\boldsymbol{\beta}, \boldsymbol{\Sigma})$. The covariance matrix $\boldsymbol{\Sigma}$ can be singular, allowing some effects to be fixed effects. In this model, we are interested in estimating all coefficients in the regressor, i.e., β_0, β_1, and β_2.

Discussion in Section 1.6 is only based on fixed-effects models. Under the current mixed model 3.1, like the case in Chapter 2, we should stick to the exact design

for each subject, and the approximate design approach only works between subjects. By adjusting the average design approach suggested by López-Fidalgo and Garcet-Rodríguez 2004, we consider the information matrix given the uncontrolled variable x_2.

For N subjects, given the value of $\boldsymbol{x}_2^i = \{x_{2,i1}, \ldots, x_{2,in}\}, i = 1, \ldots, N$, and the value of the mixed effects $\mathbf{b} = (\mathbf{b}_1, \ldots, \mathbf{b}_N)$, analogous to (2.4), the conditional covariance matrix of $\hat{\boldsymbol{\beta}}$ is

$$Cov(\hat{\boldsymbol{\beta}}|\boldsymbol{X}_2, \mathbf{b}) \approx \left(\sum_{i=1}^{N} \mathbf{F}_i^T \mathbf{U}_i^{-1} \mathbf{F}_i \right)^{-1}. \tag{3.2}$$

$\mathbf{U}_i$ equals $\mathbf{V}_i^{-1} + \mathbf{F}_i \boldsymbol{\Sigma} \mathbf{F}_i^T$, and $\mathbf{F}_i$ is the design matrix of design ξ_i, i.e., $\mathbf{F}_i = \mathbf{F}_i(\xi_i) = (\mathbf{f}(x_{1,i1}, x_{2,i1}), \ldots, \mathbf{f}(x_{1,in}, x_{2,in_i}))^T = ((1, x_{1,i1}, x_{2,i1})^T, \ldots, (1, x_{1,in_i}, x_{2,in})^T)^T$, and $\mathbf{V}_i = \mathbf{V}_i(\xi_i) = \text{diag}(v_{i1}^{\mathbf{b}_i}, \ldots, v_{in}^{\mathbf{b}_i})$. Note that $v_{ij}^{\mathbf{b}_i}$ now depends on both $x_{1,ij}$ and $x_{2,ij}$.

Using the same notation in (2.6), considering designs should only depend on the controlled variable x_1, the approximate population design is now formulated as

$$\xi = \{(\xi_p(x_1), w_p), p = 1, \ldots, N_s\}, w_p > 0, \sum_p w_p = 1. \tag{3.3}$$

Based on this notation, we first take the expectation of the inverse of (3.2) with respect to the random effects $\mathbf{b}$ to get the information matrix given uncontrolled variable $\boldsymbol{X}_2$, and then take the expectation with respect to the density $p(\boldsymbol{X}_2|\boldsymbol{X}_1)$, to obtain the PQL approximation to the information matrix under model (3.1)

$$\mathfrak{M}^{PQL}(\xi) = \mathfrak{M}^{PQL}(\xi|\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \sum_{p=1}^{N_s} w_p E_{\mathbf{b}, \boldsymbol{X}_2}(\mathbf{F}_p^T (\mathbf{V}_p^{-1} + \mathbf{F}_p \boldsymbol{\Sigma} \mathbf{F}_p^T)^{-1} \mathbf{F}_p). \tag{3.4}$$

This expression is similar to (2.7), but the matrices involved now depend on both x_1 and x_2, and the expectation should be taken with respect to both $\mathbf{b}$ and $\boldsymbol{X}_2$. Since $\mathbf{b}$ and $\boldsymbol{X}_2$ are assumed to be independent, the order of the expectations doesn't matter in computation. For density $p(\boldsymbol{X}_2|\boldsymbol{X}_1)$, this is just the multiplication of $p(x_{2,ij}|x_{1,ij})$ by independence.

3.2 Simulation: Locally Optimal Designs Searching for GLMMs with One Experimental Variable and One Dependent Observational Covariate

In this section, locally optimal designs for model (3.1) are obtained using (3.4) as an approximation to the information matrix. Since this is a locally design problem, and expression (3.4) is dependent on the distribution of the observational covariate, we set a comprehensive set of parameter choices in Table 6 in the following optimal design searching. For $\mathbf{\Sigma}$, we still take $\mathbf{\Sigma} = r \cdot \mathbf{\Sigma}_0$ as in 2.3, but with redefined $\mathbf{\Sigma}$ as in Table 6. Moreover, the dependency relationship of x_2 on x_1 is defined as $x_2 \sim N(qu(x_1), \sigma_u)$. We choose $qu(\cdot)$ to be a quadratic function, $(x - mid)^2$, where mid is the mid-point of the design region.

Choices for (β_0, β_1)	$(1, -1)$	$(6, -1)$		
Choices for β_2	1	-1		
Choices for σ_u	.6	3		
Diagonals of $\mathbf{\Sigma}_0$	I: (.1143, .07, .09)		II: (.4, .02,.09)	III: (.02, .4,.09)

Table 6. Parameter Choices for Locally Optimal Design Searching with Unknown Observational Covariate

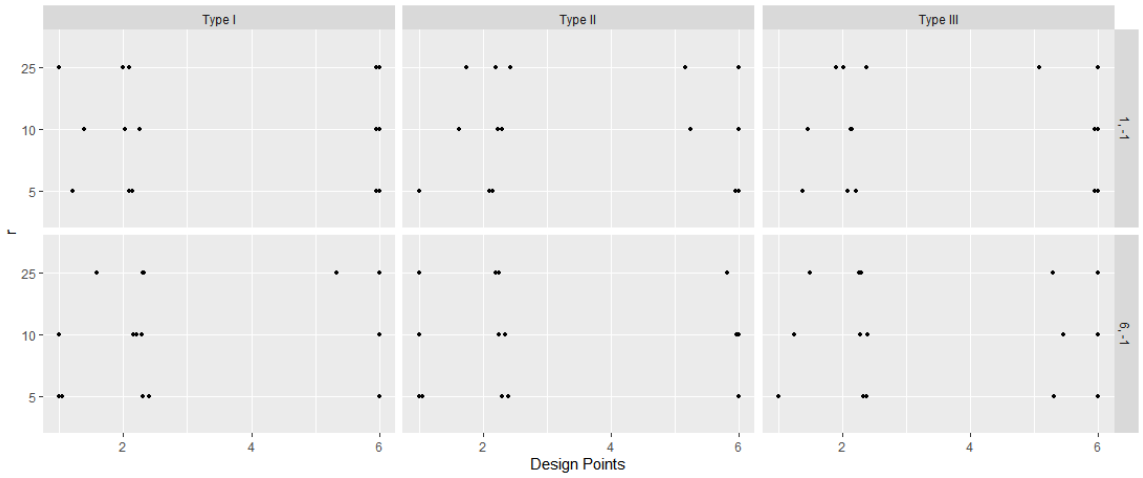

Figure 11. A-optimal Designs with $n = 5$ Observations per Subject, with $\beta_2 = 1$, $\sigma_u = .6$. The Covariance Type of $\mathbf{\Sigma}$ Is Shown in the Top Bar and the Choice for (β_0, β_1) in the Right-hand Bar. The Value of r for the Covariance Matrix $\mathbf{\Sigma}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.

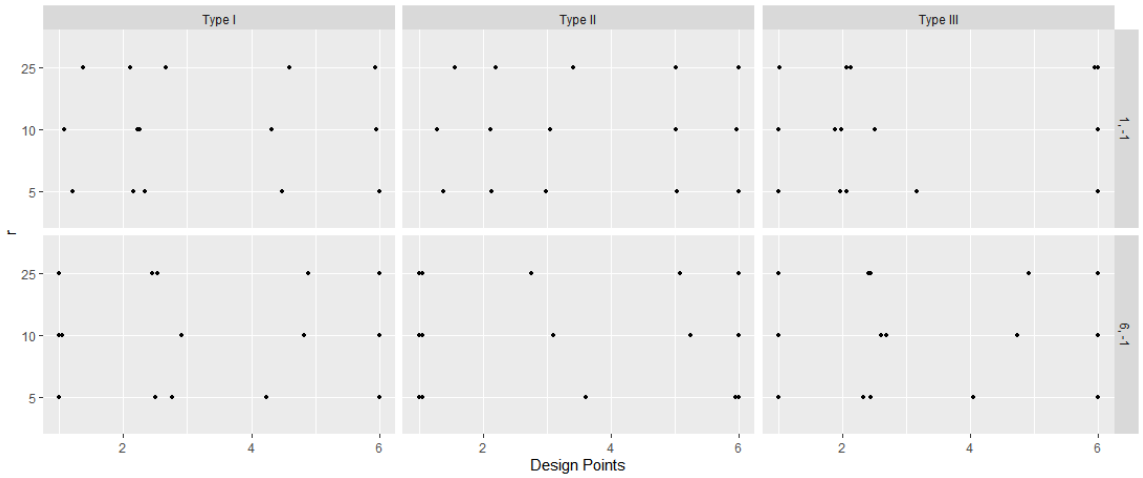

Figure 12. D-optimal Designs with $n = 5$ Observations per Subject, with $\beta_2 = 1$, $\sigma_u = .6$. The Covariance Type of $\mathbf{\Sigma}$ Is Shown in the Top Bar and the Choice for (β_0, β_1) in the Right-hand Bar. The Value of r for the Covariance Matrix $\mathbf{\Sigma}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.

We show some of the optimal designs found in Figure 11 and 12, designs for other parameter settings are shown in A.2. This is enough for us to discuss some findings

about optimal designs. Compared to designs in Section 2.3, we observe more distinct design points. When the variance of the observational variable is larger (taking value 3), we tend to have less distinct design points in a locally optimal design, and this trend is more obvious for A-optimality. The designs don't change much only for different σ_u or β_2 value. Overall, the designs are quite robust with respect to different assumption of the observational variable. This suggests that in practice, the error in the prior of the observational variable doesn't hurt the experimental design too much.

Chapter 4

TOPIC III: BAYESIAN OPTIMAL DESIGNS FOR LOGISTIC MIXED MODELS WITH ONE COVARIATE

4.1 Bayesian Design Problem Illustration

In the locally optimal design approach, the optimal designs must be generated with given values of the parameter to be estimated in the experiments. When we don't have a very convincing guess of the design parameters, Bayesian designs are often attractive, especially when some prior is considered reasonable for the design problem. Following the model description in Section 2.1, we briefly describe the model settings.

In a longitudinal study, the response from subject i under jth measurement, y_{ij}, $j = 1, \ldots, n_i$, satisfies

$$\mu_{ij}^{\mathbf{b}_i} = E(y_{ij}|\mathbf{b}_i) = \eta^{-1}(b_{i0} + b_{i1}x_{1,ij}). \tag{4.1}$$

Write $\mathbf{b}_i = (b_{i0}, b_{i1})^T = \boldsymbol{\beta} + \boldsymbol{\alpha}_i$ consists of the subject-independent fixed-effects vector $\boldsymbol{\beta}$ and the random vector $\boldsymbol{\alpha}_i \sim N_2(0, \boldsymbol{\Sigma})$, i.e., $\mathbf{b}_i \sim N_2(\boldsymbol{\beta}, \boldsymbol{\Sigma})$. In the Bayesian design approach, the unknown parameters $(\boldsymbol{\beta}, \boldsymbol{\Sigma})$ are assumed to be following some prior distributions, and the design optimality criteria will be an expectation with respect to these distributions. In this model, we are interested in estimating all coefficients in the regressor, i.e., β_0, β_1.

This model doesn't differ from model (1.3) in estimation, we can obtain the PQL approximation to the conditional information matrix under model (3.1) as

$$\mathfrak{M}^{PQL}(\xi) = \mathfrak{M}^{PQL}(\xi|\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \sum_{p=1}^{N_s} w_p E_{\mathbf{b}}(\mathbf{F}_p^T(\mathbf{V}_p^{-1} + \mathbf{F}_p\boldsymbol{\Sigma}\mathbf{F}_p^T)^{-1}\mathbf{F}_p), \tag{4.2}$$

which is in the same form as (2.7).

By Chaloner and Verdinelli. 1995, Bayesian optimality criteria can be expressed as

$$\phi_B(\xi) = \begin{cases} E_{\boldsymbol{\beta},\boldsymbol{\Sigma}}(\log\det(I(\xi|\boldsymbol{\beta},\boldsymbol{\Sigma}))), & \text{for Bayesian D-optimality;} \\ E_{\boldsymbol{\beta},\boldsymbol{\Sigma}}(-tr(I^{-1}(\xi|\boldsymbol{\beta},\boldsymbol{\Sigma}))), & \text{for Bayesian A-optimality;} \end{cases} \tag{4.3}$$

where $I(\xi|\boldsymbol{\beta},\boldsymbol{\Sigma})$ is the expected information matrix for design ξ and parameter $\boldsymbol{\beta},\boldsymbol{\Sigma}$. Since there are two expectations in the criterion, which makes the simulation more complex, we describe the process of calculating Bayesian criteria step by step in Algorithm 2.

Algorithm 2 Calculating Bayesian Criterion using expectation of the Conditional Information

Get n paired samples from the prior of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, $\{(\boldsymbol{\beta}_i,\boldsymbol{\Sigma}_i)\}_{i=1}^n$;
For $(\boldsymbol{\beta}_i,\boldsymbol{\Sigma}_i)$, $i=1,\ldots,n$
Get m samples from $N(\boldsymbol{\beta}_i,\boldsymbol{\Sigma}_i)$, $\{(\mathbf{b}_j)\}_{j=1}^m$ to calculate $M^{PQL}(\boldsymbol{\beta}_i,\boldsymbol{\Sigma}_i)$;
$\frac{1}{n}\phi^1_{B-D} \approx \sum_{i=1}^n \log det(M^{PQL}(\boldsymbol{\beta}_i,\boldsymbol{\Sigma}_i))$;
$\frac{1}{n}\phi^1_{B-A} \approx -\sum_{i=1}^n tr(M^{-1^{PQL}}(\boldsymbol{\beta}_i,\boldsymbol{\Sigma}_i))$;

Using this approach, n needs to be sufficiently large to get a stable estimation to such an expectation. By some simulation, $n = 10000$ is still not large enough, making this approach too slow in finding an optimal design.

To have a shorter running time, we try to derive the expected information matrix from the inverse of the expectation of the covariance matrix as following.

For m_0 subjects, given the value of $\boldsymbol{x}_2^i = \{x_{2,i1},\ldots,x_{2,in_i}\}, i = 1,\ldots,m_0$, and the value of the mixed effects $\mathbf{b} = (\mathbf{b}_1,\ldots,\mathbf{b}_{m_0})$, analogous to (2.4), the conditional covariance matrix of $\hat{\beta}$ is

$$Cov(\hat{\boldsymbol{\beta}}|\mathbf{b}) \approx \left(\sum_{i=1}^{m_0} \mathbf{F}_i^T \mathbf{U}_i^{-1} \mathbf{F}_i\right)^{-1}. \tag{4.4}$$

$\mathbf{U}_i$ equals $\mathbf{V}_i^{-1} + \mathbf{F}_i\boldsymbol{\Sigma}\mathbf{F}_i^T$, and $\mathbf{F}_i$ is the design matrix of design ξ_i, i.e., $\mathbf{F}_i = \mathbf{F}_i(\xi_i) = (\mathbf{f}(x_{1,i1}), \ldots, \mathbf{f}(x_{1,in_i}))^T = ((1, x_{1,i1})^T, \ldots, (1, x_{1,in_i})^T)^T$, and $\mathbf{V}_i = \mathbf{V}_i(\xi_i) = \text{diag}(v_{i1}^{\mathbf{b}_i}, \ldots, v_{in_i}^{\mathbf{b}_i})$.

It is intuitive to compute the expectation of (4.4) with respect to $\mathbf{b}$ as $C_{m_0}^{PQL}(\boldsymbol{\beta}, \boldsymbol{\Sigma})$. Given the estimator $\hat{\boldsymbol{\beta}}$ being unbiased, this expectation is an unbiased estimator of the covariance matrix of the estimator. Then, taking the inverse, we have the value for the expected information matrix. This expression is dependent on m_0, the total number of subjects in the experiment, which suggests different designs for different numbers of subjects. The simulation details for this approach is showed in Algorithm 3.

Algorithm 3 Calculating Bayesian Criterion using Expectation of the Conditional Covariance Matrix

Get n paired samples from the prior of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, $\{(\boldsymbol{\beta}_i, \boldsymbol{\Sigma}_i)\}_{i=1}^n$;
For $(\boldsymbol{\beta}_i, \boldsymbol{\Sigma}_i)$, $i = 1, \ldots, n$
Get $m \cdot m_0$ samples from $N(\boldsymbol{\beta}_i, \boldsymbol{\Sigma}_i)$, $\{(\mathbf{b}_j)\}_{j=1}^{m \cdot m_0}$ to calculate $C_{m_0}^{PQL}(\boldsymbol{\beta}_i, \boldsymbol{\Sigma}_i)$;
$\phi_{B-D}^2 \approx \frac{1}{n}\sum_{i=1}^n \log det^{-1}(C_{m_0}^{PQL}(\boldsymbol{\beta}_i, \boldsymbol{\Sigma}_i))$;
$\phi_{B-A}^2 \approx -\frac{1}{n}\sum_{i=1}^n tr(C_{m_0}^{PQL}(\boldsymbol{\beta}_i, \boldsymbol{\Sigma}_i))$;

These two approaches have a direct connection. It is not hard to show that $\mathbf{F}_i^T\mathbf{U}_i^{-1}\mathbf{F}_i$ has finite expectation with respect to $\mathbf{b}_i$. When m_0 goes to infinity, $\frac{1}{m_0}\sum_{i=1}^{m_0}\mathbf{F}_i^T\mathbf{U}_i^{-1}\mathbf{F}_i$ has a finite limit, which is just $E_{\mathbf{b}_i}(\mathbf{F}_i^T\mathbf{U}_i^{-1}\mathbf{F}_i)$. When we have enough subjects, these two approaches should match. For longitudinal studies, especially in clinical trials, available subjects are usually limited, so a relatively smaller m_0 is focused on. By simulation, the second approach indeed provides much smaller error for $m_0 = 10$, and the later optimal design searching is based on this $m_0 = 10$ setting.

4.2 Simulation: Bayesian Optimal Designs Searching

In this section, Bayesian optimal designs for model (4.1) are obtained using the second approach to calculate the optimality criteria. For Bayesian optimal designs, evaluations are made on priors of the model parameters. In the following simulation, we consider sufficiently comprehensive cases of the priors of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ to illustrate the behavior of these optimal designs. $\boldsymbol{\beta}$ is assumed to be following a normal distribution $N(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_{\boldsymbol{\beta}})$, where $\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = r \cdot \boldsymbol{\Sigma}_{\boldsymbol{\beta},0}$, with three different choices of $\boldsymbol{\Sigma}_{\boldsymbol{\beta},0}$. $\boldsymbol{\Sigma}$ is assumed to a diagonal matrix $diag\{\sigma_0^2, \sigma_1^2\}$, and the two diagonals are assumed to be following inverse-Gamma distributions. The detailed priors and other settings are shown in Table 7.

Choices for $\boldsymbol{\beta}_0$	$(1,-1)'$	$(6,-1)'$		
Choices for $\boldsymbol{\Sigma}_{\boldsymbol{\beta},0}$	I: (.1143, .07)		II: (.4, .02)	III: (.02, .4)
Inverse Gamma for σ_0^2	a: (1,4/3)	b: (3,1)		
Inverse Gamma for σ_1^2	a: (1,4/3)	b: (3,1)		

Table 7. Parameter Choices for Bayesian Optimal Design Searching

We show some of the optimal designs found in Figure 13 and 14, designs for other parameter settings are shown in Appendix A.3. The choices of the parameters of the inverse-Gamma distribution of $\sigma_{\boldsymbol{\beta},0}^2$ and $\sigma_{\boldsymbol{\beta},1}^2$ are represented as (a, a), (a, b), etc. For the inverse-Gamma priors of σ_0^2 and σ_1^2, case a corresponds to an inverse-Gamma distribution with a quite concentrate distribution. For case b, the distribution is spreading out with a larger variance.

We can conclude some characteristic of the designs from the plots. In many cases for number of observations $n = 2$, optimal designs contain 2 sub-designs. For $\boldsymbol{\beta} = (6,-1)'$ and Type III for $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$, we can see some beautiful symmetry of the two sub-designs. Each of the sub-designs is focusing on one of the end points of the design

region. For $n = 5$ case, the number of distinct design points is 3 or 4, which is more than the typical value 3 for locally optimal design. This is a reasonable result, that with more variation of the model assumption, more distinct values of the design points are needed.

The choices of prior of $\mathbf{\Sigma}$ is then discussed. For cases involving b, comparing to case (a, a), we tend to need more distinct designs points to form a Bayesian optimal design. But overall, the designs are quite robust with respect to such a choice. This suggests that, in practice, the prior of $\boldsymbol{\beta}$ has more influence on the optimal design than that of $\mathbf{\Sigma}$.

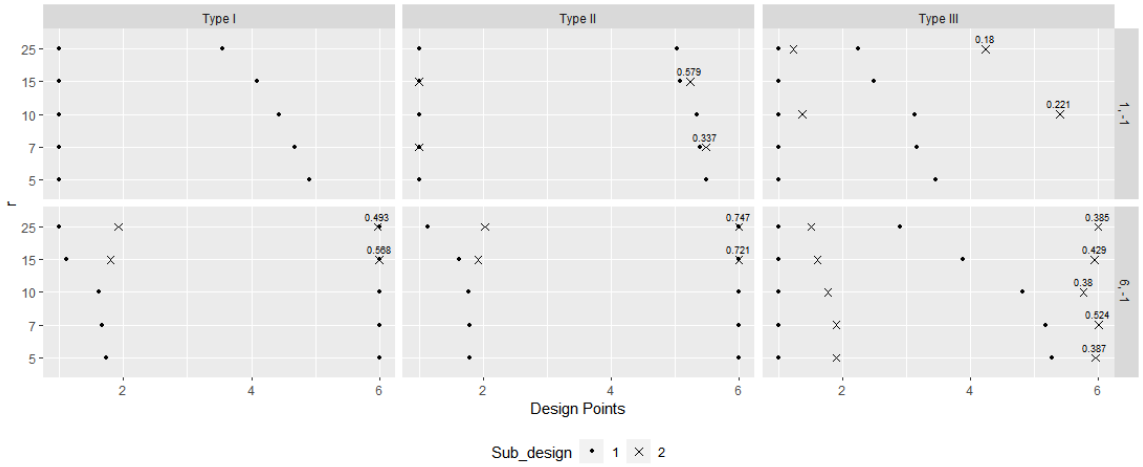

Figure 13. Bayesian D-optimal Designs with $n = 2$ Observations per Subject. The Prior Settings of $\mathbf{\Sigma}$ Are Case (a, a). The Covariance Type of $\mathbf{\Sigma}_{\boldsymbol{\beta}}$ Is Shown in the Top Bar and the Choice for $\boldsymbol{\beta}_0$ in the Right-hand Bar. The Value of r for the Covariance Matrix $\mathbf{\Sigma}_{\boldsymbol{\beta}}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.

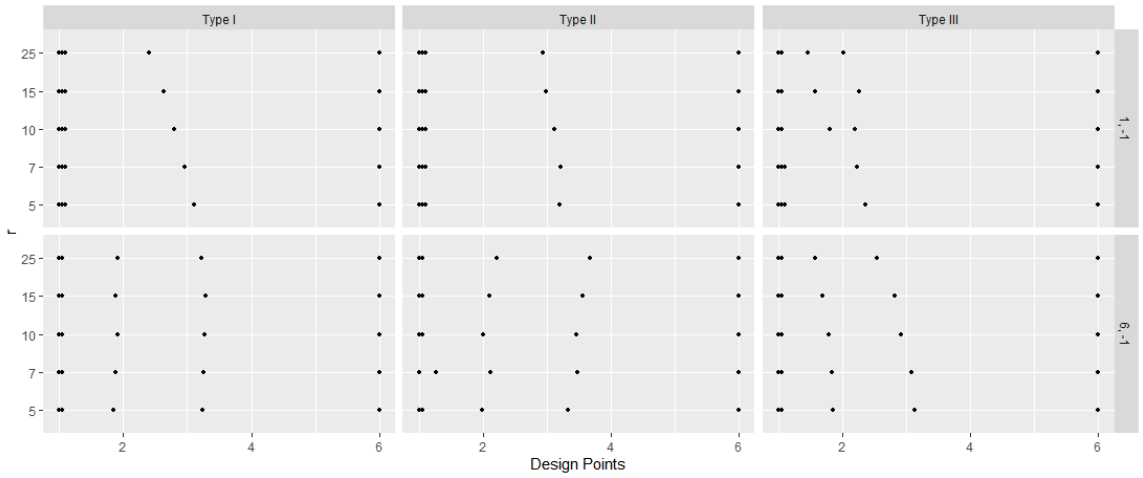

Figure 14. Bayesian D-optimal Designs with $n = 5$ Observations per Subject. The Prior Settings of $\boldsymbol{\Sigma}$ Are Case (a, a). The Covariance Type of $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ Is Shown in the Top Bar and the Choice for $\boldsymbol{\beta}_0$ in the Right-hand Bar. The Value of r for the Covariance Matrix $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.

4.3 Robustness Study on Bayesian Optimal Designs

Compared to locally optimal designs, one of the main advantages of Bayesian optimal designs is better performance in robustness. In Bayesian optimality criteria, the aim function is evaluated over a prior, instead of the point guess in locally optimality criteria. In this subsection, we discuss the robustness of the Bayesian optimal designs under model (4.1).

The robustness can be evaluated by relative efficiencies comparing Bayesian designs and others. Under a given prior, a joint sample of size $n_b = 50$ can be generated, and we can search for locally optimal designs corresponding to each value in the sample. In the beginning, under these local parameters, efficiencies of Bayesian designs can be computed by comparing to the corresponding locally optimal designs. For a Bayesian optimal design, we are interested in the lowest efficiency E_{bL}, which stands for the

worst performance of a Bayesian design. On the other hand, we can also study the robustness of these locally optimal designs as a comparison. The relative efficiencies of these locally optimal designs comparing to each other can be computed. Then, we can compute the lowest efficiency achieved by each of the 50 locally optimal designs under these 50 parameter values. For these 50 minimums, we look at their maximum E_{lM} and minimum E_{lL}.

The results of efficiencies are summarized in Table 8 and 9. For each prior case, we show the lowest efficiency E_{bL} of the Bayesian optimal design. Also, we show both the maximum E_{lM} and the minimum E_{lL} of the 50 lowest efficiencies achieved by the 50 locally optimal designs. In these tables, by comparing E_{bL} and E_{lM}, we can see that Bayesian designs offer much higher or not much smaller lowest efficiencies compared to all randomly generated locally optimal designs, which means some rare but disastrous cases can be avoided. To be specific, in some cases, some extremely low efficiencies are avoided by applying Bayesian optimal design, see the red numbers in the tables. That means in those cases, each of the locally optimal designs may have achieved such a low value for at least one time, for all 50 parameter values, but a Bayesian design can offer a much better lowest efficiency. By comparing E_{bL} and E_{lL}, we can also see some of the locally optimal designs are performing badly under mis-specified parameter values. Overall, Bayesian optimal designs are robust with respect to different parameter values, especially compared with locally optimal designs.

$\boldsymbol{\Sigma}$ Case		(a,a)			(a,b)			(b,a)			(b,b)		
	r	5	10	25	5	10	25	5	10	25	5	10	25
	E_{bL}	.83	.84	.89	.74	.80	.78	.79	.56	.77	.50	.70	.63
I	E_{lM}	.85	.80	.87	.78	.84	.78	.78	.70	.77	.53	.70	.51
	E_{lL}	.44	.47	.46	.39	.24	.14	.36	.10	.06	0	0	0
	E_{bL}	.62	.80	.75	.61	.73	.68	.65	.66	.69	.63	.37	.58
II	E_{lM}	.70	.79	.75	.71	.74	.69	.72	.62	.72	.65	.54	.54
	E_{lL}	.31	.44	.30	.30	.27	.28	.23	.22	.16	.04	.05	.01
	E_{bL}	.53	.54	.74	.74	.39	.56	.64	.63	.43	.45	.58	.42
III	E_{lM}	.72	.74	.75	.66	.71	.42	.70	.41	.54	.18	.53	.19
	E_{lL}	0	.02	.16	.02	0	0	0	0	0	0	0	0

Table 8. Some Lowest Efficiencies Achieved by Bayesian D-Optimal Designs and Locally D-Optimal Designs, for $\boldsymbol{\beta}_0 = (6, -1)'$. See more Details in the Related Discussion. 0 Means the Value Is Less Than .01.

$\boldsymbol{\Sigma}$ Case		(a,a)			(a,b)			(b,a)			(b,b)		
	r	5	10	25	5	10	25	5	10	25	5	10	25
	E_{bL}	.77	.60	.79	.68	.72	.63	.63	.57	.63	.34	.37	.45
I	E_{lM}	.79	.63	.79	.61	.76	.70	.67	.57	.67	.25	.59	.19
	E_{lL}	.25	.21	.33	.16	.07	.03	.17	0	0	0	0	0
	E_{bL}	.57	.58	.69	.37	.70	.58	.63	.28	.40	.35	.51	.29
II	E_{lM}	.55	.66	.66	.50	.64	.50	.58	.42	.56	.40	.32	.29
	E_{lL}	.14	.21	.15	.10	.11	.10	.07	.05	.05	.01	.02	0
	E_{bL}	.63	.52	.71	.72	.49	.62	.61	.59	.53	.48	.56	.43
III	E_{lM}	.72	.74	.75	.66	.71	.42	.70	.41	.54	.20	.53	.12
	E_{lL}	0	.04	.12	.03	0	.14	.07	0	0	0	0	0

Table 9. Some Lowest Efficiencies Achieved by Bayesian A-Optimal Designs and Locally A-Optimal Designs, for $\boldsymbol{\beta}_0 = (6, -1)'$. See more Details in the Related Discussion. 0 Means the Value Is Less Than .01.

Chapter 5

CONCLUSION AND DISCUSSION

In this dissertation, we investigate optimal designs for logit mixed models.

In Chapter 2, we search for locally optimal designs under one-covariate logit mixed models. To approximate the information matrix, we apply penalized quasi-likelihood (PQL) method and show the superiority of it. Under different local parameters and design restrictions, locally D- and A-optimal designs are constructed based on the approximation. We have several conclusions. By the expression of the PQL approximation, the locally optimal designs are dependent on the number of measurements for each subject. For these locally optimal designs, we can still see some similarity to the locally optimal designs for GLMs, in both number and positions of distinct design points. Moreover, for many cases, we should apply different designs to different subjects. We also study the robustness of these locally optimal designs with respect to mis-specified variance-covariance matrix of the random effects. Based on the line plot of the robustness, we should carefully make the guess of the parameters to decrease potential loss in efficiency.

In Chapter 3, an unknown observational covariate is added to the previous model. With such an unknown observational variable in the experiment, expected optimality criteria based on the PQL approximation are considered. Under different assumptions of the unknown variable and parameter settings, locally optimal designs are constructed and discussed. In this case, the number of distinct design points increases. Moreover, the distribution of these points is more even in the design region. Comparing these

locally optimal designs under different parameter settings, the designs are quite robust to many of the settings, especially to the settings related of the observational variable.

In Chapter 4, we study Bayesian optimal designs based on the model in Chapter 2. Bayesian design under such a model is usually expensive in time. The running time of simulations in this dissertation is optimized to an acceptable amount with accurate result. Typically, we need more distinct design points in this case. For Bayesian design, the choice of the prior is also important. Based on the results, the Bayesian designs are quite robust with respect to the prior of $\boldsymbol{\Sigma}$, while the prior of $\boldsymbol{\beta}$ has much more influence. We must be careful about the prior of $\boldsymbol{\beta}$, other than that of $\boldsymbol{\Sigma}$. We also discuss the robustness of Bayesian optimal designs with respect to different parameters from the priors and got much better results compared with locally optimal designs.

There are still lots of interesting topics related to optimal designs under logit mixed models. Using other approximations to the information matrix, or considering simpler models, it is possible to pursue complete class results for locally optimal designs. It is also possible to consider more assumptions on the unknown observational variable to fit more realistic design problems.

REFERENCES

Abebe, H., F. Tan G., Van Breukelen, and M. Berger. 2014. "Robustness of Bayesian D-optimal design for the logistic mixed model against misspecification of autocorrelation." *Computational Statistics* 29 (6): 1667–1690.

Atkinson, A. C. 1982. "Optimum biased coin designs for sequential clinical trials with prognostic factors." *Biometrika* 69:61–67.

Breslow, N. E., and D. G. Clayton. 1993. "Approximate inference in generalized linear mixed models." *Journal of the American Statistical Association* 88 (421): 9–25.

Carrière, I., and J. Bouyer. 2002. "Choosing marginal or random-effects models for longitudinal binary responses: application to self-reported disability among older persons." *BMC Medical Research Methodology* 2 (1): 1.

Chaloner, K., and I. Verdinelli. 1995. "Bayesian Experimental Design: A Review." *Statistical Science* 10 (3): 273–304.

Cook, R. D., and L. A. Thibodeau. 1980. "Marginally Restricted DOptimal Designs." *Journal of the American Statistical Association* 75 (370): 366–371.

Dean, A., M. Morris, J. Stufken, and D. Bingham. 2015. *Handbook of design and analysis of experiments.* CRC Press.

Efron, B. 1971. "Forcing a sequential experiment to be balanced." *Biometrika* 58:403–417.

———. 1980. "Randomizing and balancing a complicated sequential experiment." In *Biostatistics Casebook,* edited by R. G. Miller, B. Efron, B. Brown, and L. E. Moses, 19–30. New York: Wiley.

Green, P. J. 1987. "Penalized likelihood for general semi-parametric regression models." *International Statistical Review* 55:245–259.

Grouin, J.-M., S. Day, and J. Lewis. 2004. "Adjustment for baseline covariates: An introductory note." *Statistics in Medicine* 23:697–699.

Gupta, R., and D. Richards. 1985. "Testing Optimality of Experimental Designs for a Regression Model With Random Variables." *Journal of Statistical Planning and Inference* 11:75–80.

Harville, D. A. 1974. "Nearly optimal allocation of experimental units using observed covariate values." *Technometrics* 16:589–599.

Hu, L., and J. Stufken. 2017. "Optimal designs for mixed effects models." *submitted to Statistica Sinica.*

Hu, L., M. Yang, and J. Stufken. 2015. "Saturated locally optimal designs under differentiable optimality criteria." *Ann. Statist* 43:30–56.

Kennedy, J., and R. Eberhart. 1995. "Particle swarm optimization." *Proceedings of ICNN'95 - International Conference on Neural Networks.*

Leyland-Jones, B. 2003. "Breast cancer trial with erythropoietin terminated unexpectedly." *Lancet Oncology* 4:459–460.

Liang, K.-Y., and S. L. Zeger. 1986. "Longitudinal data analysis using generalized linear models." *Biometrika* 73 (1): 13–22.

López-Fidalgo, J., and S. A. Garcet-Rodríguez. 2004. "Optimal Experimental Designs When Some Independent Variables Are Not Subject to Control." *Journal of the American Statistical Association* 99 (468): 1190–1199.

López-Fidalgo, J., R. Martín-Martín, and M. Stehlík. 2008. "Marginally restricted D-optimal designs for correlated observations." *Journal of Applied Statistics* 35 (6): 617–632.

Mak, S., and V. R. Joseph. 2018. "Support points." *The Annals of Statistics* 46:2562–2592.

McCullagh, P., and J. A. Nelder. 1989. *Generalized linear models (2nd ed.)* London: Chapman / Hall.

Miller, K. S. 1981. "On the Inverse of the Sum of Matrices." *Mathematics Magazine* 54 (2): 67–72.

Molenberghs, G., and G. Verbeke. 2005. *Models for discrete longitudinal data.* Springer.

Niaparast, M. 2009. "On optimal design for a poisson regression model with random intercept." *Statistics & Probability Letters* 79 (6): 741–747.

Pinheiro, J. C. 1994. "Topics in Mixed Effects Models, University of Wisconsin-Madison."

Pocock, S. J., and R. Simon. 1975. "Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial." *Biometrics* 31:103–115.

Pukelsheim, F. 2006. *Optimal design of experiments.* Society for Industrial / Applied Mathematics.

Rosenberger, W. F., and O. Sverdlov. 2008. "Handling Covariates in the Design of Clinical Trials." *Statistical Science* 23 (3): 404–419.

Schmelter, T. 2007. "The optimality of single-group designs for certain mixed models." *Metrika* 65 (2): 183–193.

Sinha, S. K., and X. Xu. 2011. "Sequential D-optimal designs for generalized linear mixed models." *Journal of Statistical Planning and Inference* 141 (4): 1394–1402.

Taves, D. R. 1974. "Minimization: A new method of assigning patients to treatment and control groups." *Clinical Pharmacology and Therapeutics* 15:443–453.

Tekle, F. B., F. E. Tan, and M. P. Berger. 2008. "Maximin D-optimal designs for binary longitudinal responses." *Computational Statistics & Data Analysis* 52 (12): 5253–5262.

Tsukanov, A. V. 1981. "On Test of Optimality of Experimental Designs for a Regression Model With Random Variables." *Theory of Probability and Its Applications* 26:173–177.

Ueckert, S., and F. Mentré. 2017. "A new method for evaluation of the Fisher information matrix for discrete mixed effect models using Monte Carlo sampling and adaptive Gaussian quadrature." *Comput Stat Data Anal* 111:203–219.

Waite, T. W., and D. C. Woods. 2015. "Designs for generalized linear models with random block effects via information matrix approximations." *Biometrika* 102 (3): 677–693.

Wei, L. J. 1978. "An application of an urn model to the design of sequential controlled clinical trials." *Journal of the American Statistical Association* 72:382–386.

Yang, M., and J. Stufken. 2009. "Support points of locally optimal designs for nonlinear models with two parameters." *The Annals of Statistics* 37 (1): 518–541.

Zeger, S. L., K.-Y. Liang, and P. S. Albert. 1998. "Models for longitudinal data: a generalized estimating equation approach." *Biometrics* 44 (4): 1049–1060.

Zhang, Wei. 2018. "Optimal Designs for The Panel Mixed Logit Model." *PhD dissertation, The University of Georgia. Athens, Georgia, US.*

Zhou, X., Y. Wang, and R. Yue. 2021. "Optimal designs for discrete-time survival models with random effects." *Lifetime Data Analysis.*

APPENDIX A

OPTIMAL DESIGNS FOR ALL PARAMETER SETTINGS

Here are all design results not shown in the previous chapters.

A.1 Locally A- and D-optimal designs for all parameter settings

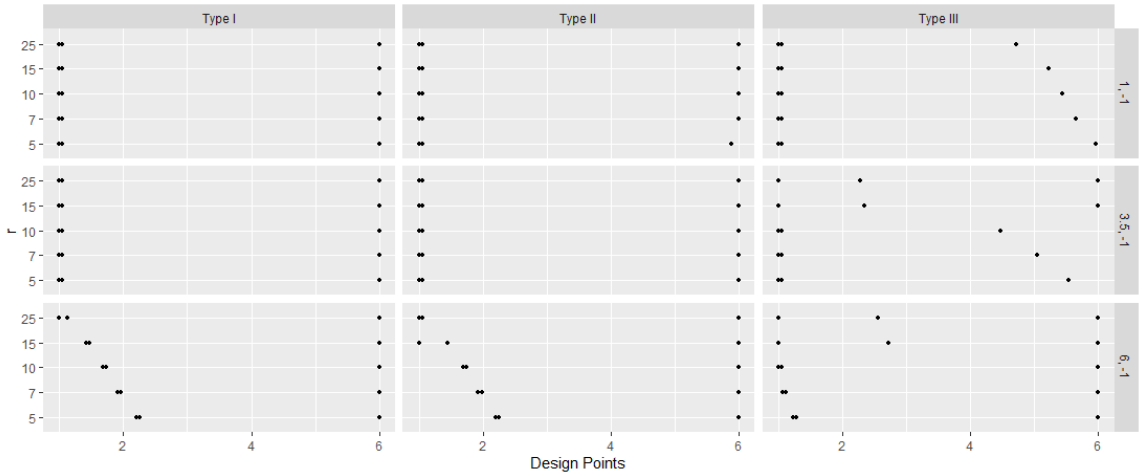

Figure 15. A-optimal Designs with $n = 3$ Observations per Subject. The Covariance Type Is Shown in the Top Bar and the Choice for $\boldsymbol{\beta}$ in the Right-hand Bar. The Value of r for the Covariance Matrix $\boldsymbol{\Sigma}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.

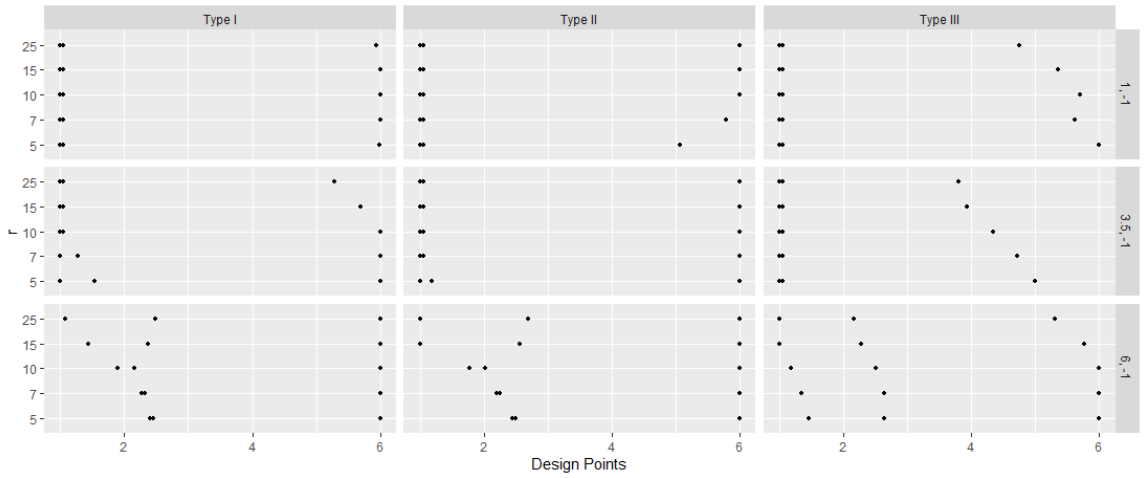

Figure 16. D-optimal Designs with $n = 3$ Observations per Subject. The Covariance Type Is Shown in the Top Bar and the Choice for $\boldsymbol{\beta}$ in the Right-hand Bar. The Value of r for the Covariance Matrix $\boldsymbol{\Sigma}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.

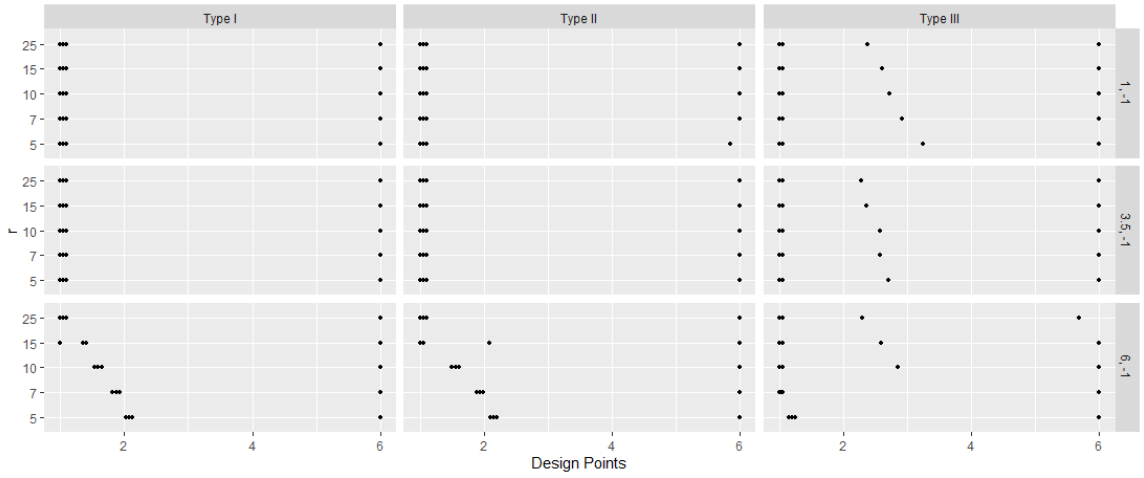

Figure 17. A-optimal Designs with $n = 4$ Observations per Subject. The Covariance Type Is Shown in the Top Bar and the Choice for $\boldsymbol{\beta}$ in the Right-hand Bar. The Value of r for the Covariance Matrix $\boldsymbol{\Sigma}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.

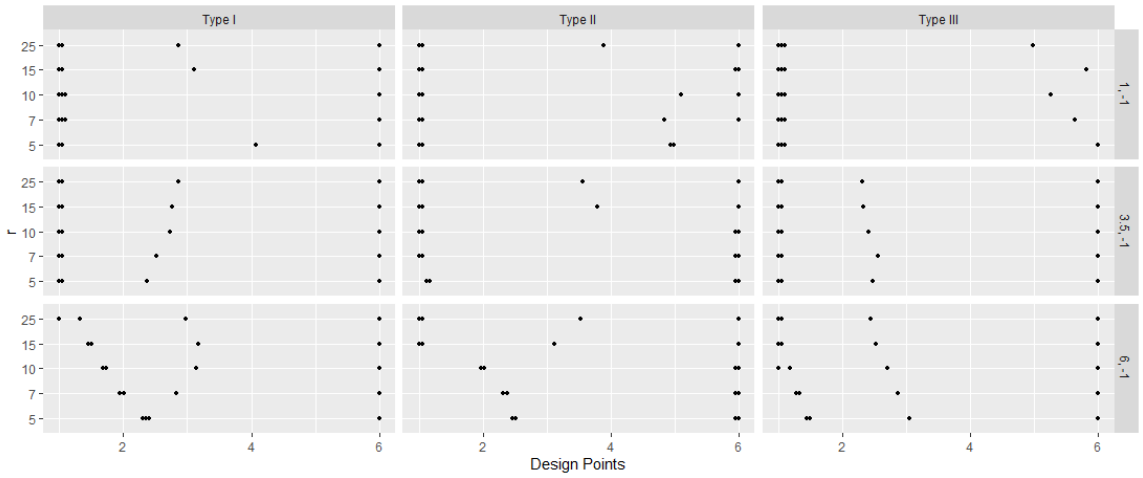

Figure 18. D-optimal Designs with $n = 4$ Observations per Subject. The Covariance Type Is Shown in the Top Bar and the Choice for $\boldsymbol{\beta}$ in the Right-hand Bar. The Value of r for the Covariance Matrix $\boldsymbol{\Sigma}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis..

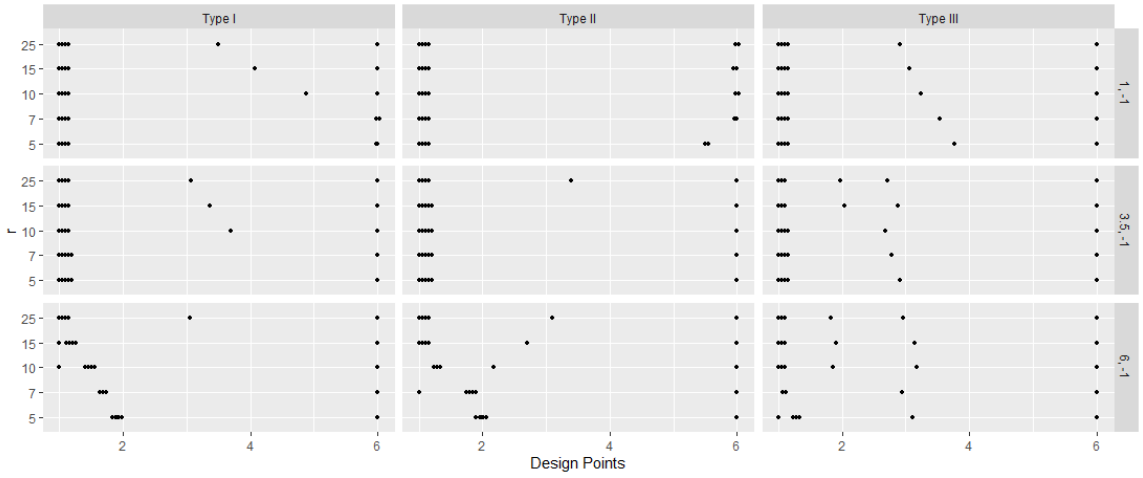

Figure 19. A-optimal Designs with $n = 6$ Observations per Subject. The Covariance Type Is Shown in the Top Bar and the Choice for $\boldsymbol{\beta}$ in the Right-hand Bar. The Value of r for the Covariance Matrix $\boldsymbol{\Sigma}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.

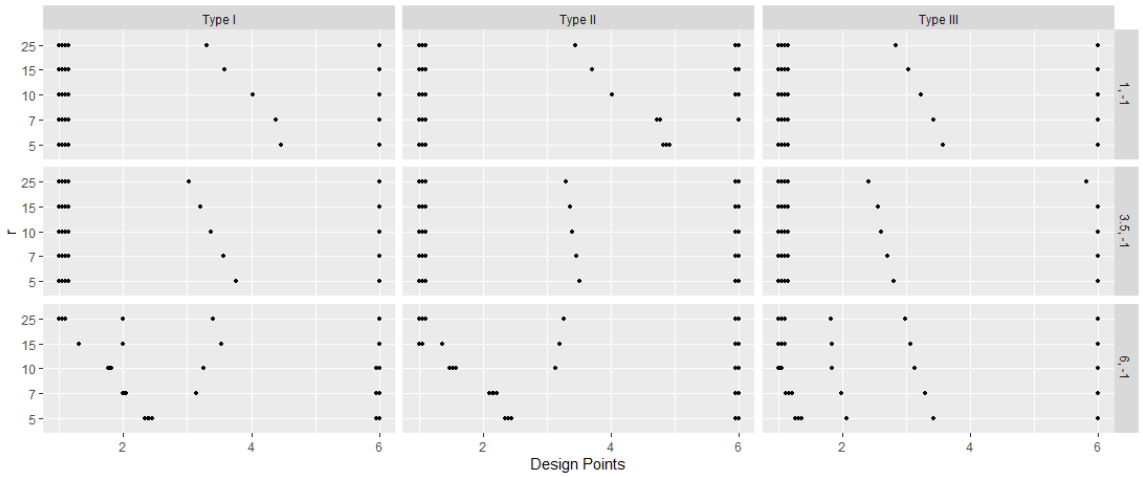

Figure 20. D-optimal Designs with $n = 6$ Observations per Subject. The Covariance Type Is Shown in the Top Bar and the Choice for $\boldsymbol{\beta}$ in the Right-hand Bar. The Value of r for the Covariance Matrix $\boldsymbol{\Sigma}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.

A.2 Locally A- and D-optimal designs with an unknown observational covariate for all parameter settings

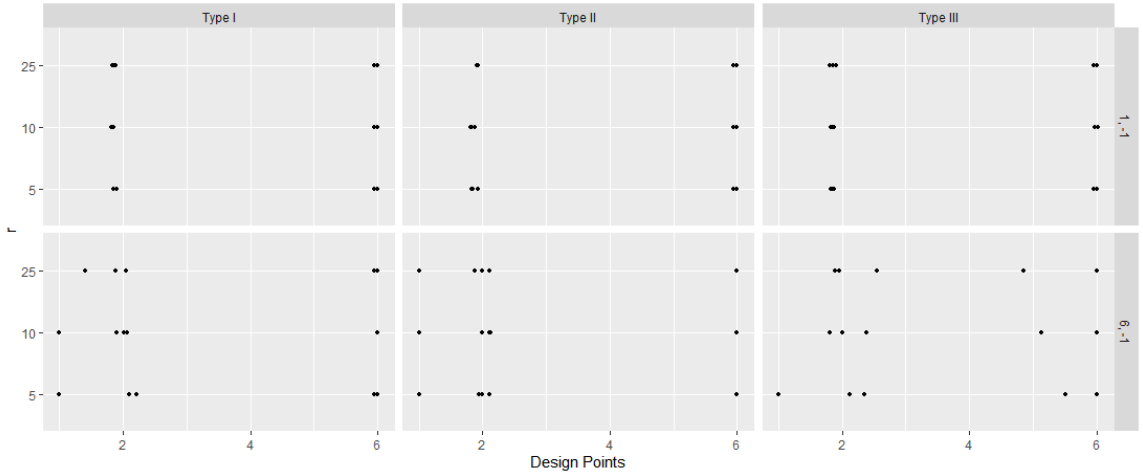

Figure 21. A-optimal Designs with $n = 5$ Observations per Subject, with $\beta_2 = 1$, $\sigma_u = 3$. The Covariance Type of $\boldsymbol{\Sigma}$ Is Shown in the Top Bar and the Choice for (β_0, β_1) in the Right-hand Bar. The Value of r for the Covariance Matrix $\boldsymbol{\Sigma}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.

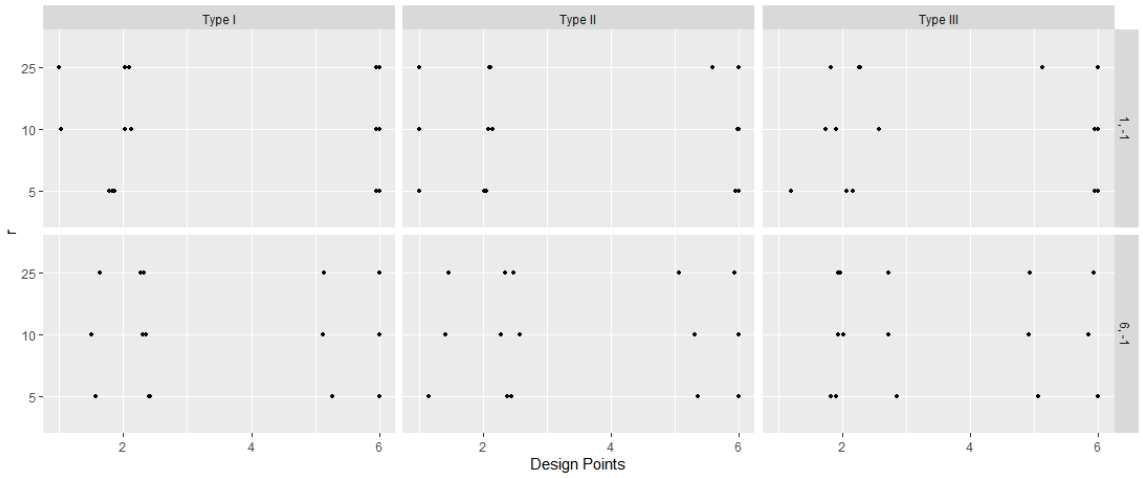

Figure 22. A-optimal Designs with $n = 5$ Observations per Subject, with $\beta_2 = -1$, $\sigma_u = .6$. The Covariance Type of $\mathbf{\Sigma}$ Is Shown in the Top Bar and the Choice for (β_0, β_1) in the Right-hand Bar. The Value of r for the Covariance Matrix $\mathbf{\Sigma}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.

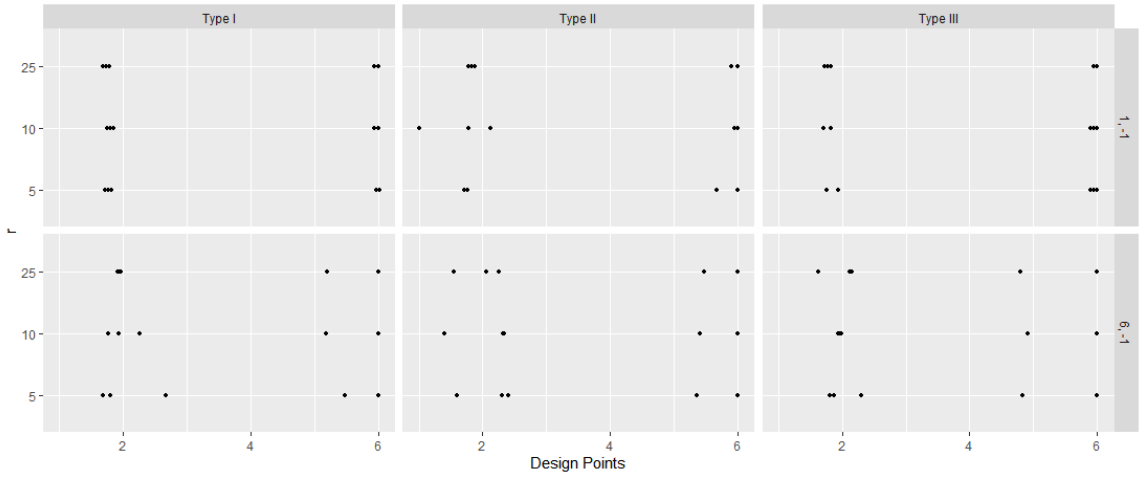

Figure 23. A-optimal Designs with $n = 5$ Observations per Subject, with $\beta_2 = -1$, $\sigma_u = 3$. The Covariance Type of $\mathbf{\Sigma}$ Is Shown in the Top Bar and the Choice for (β_0, β_1) in the Right-hand Bar. The Value of r for the Covariance Matrix $\mathbf{\Sigma}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.

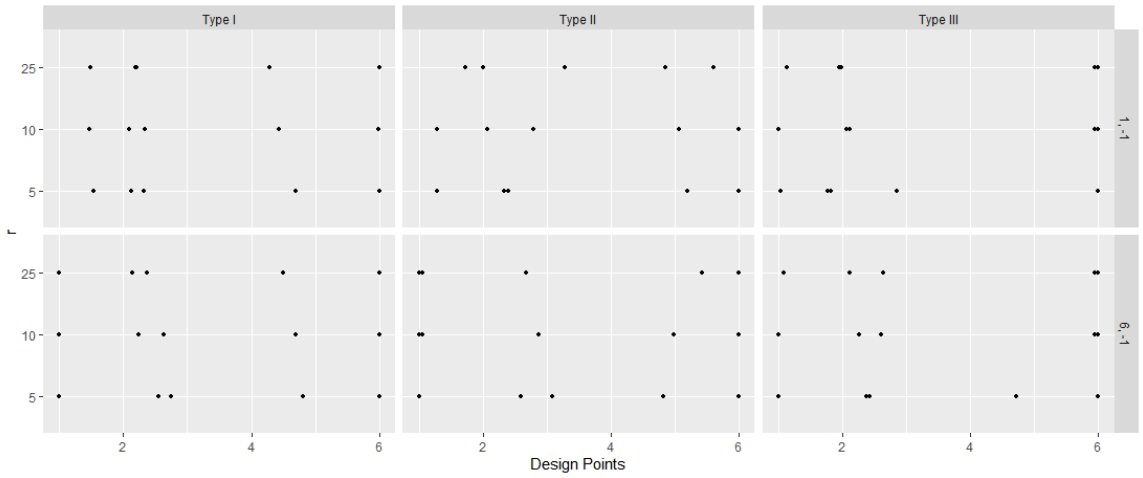

Figure 24. D-optimal Designs with $n = 5$ Observations per Subject, with $\beta_2 = 1$, $\sigma_u = 3$. The Covariance Type of $\mathbf{\Sigma}$ Is Shown in the Top Bar and the Choice for (β_0, β_1) in the Right-hand Bar. The Value of r for the Covariance Matrix $\mathbf{\Sigma}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.

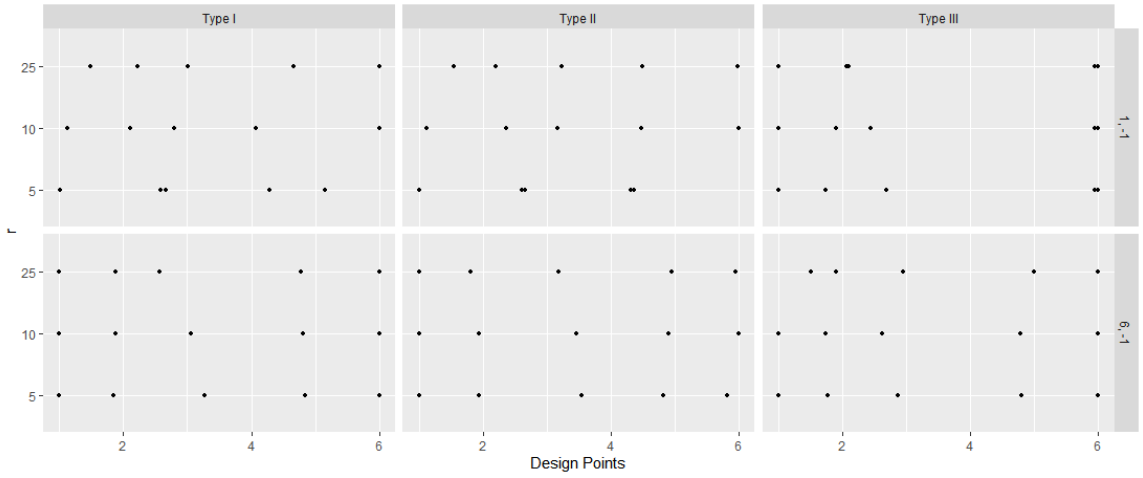

Figure 25. D-optimal Designs with $n = 5$ Observations per Subject, with $\beta_2 = -1$, $\sigma_u = .6$. The Covariance Type of $\mathbf{\Sigma}$ Is Shown in the Top Bar and the Choice for (β_0, β_1) in the Right-hand Bar. The Value of r for the Covariance Matrix $\mathbf{\Sigma}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.

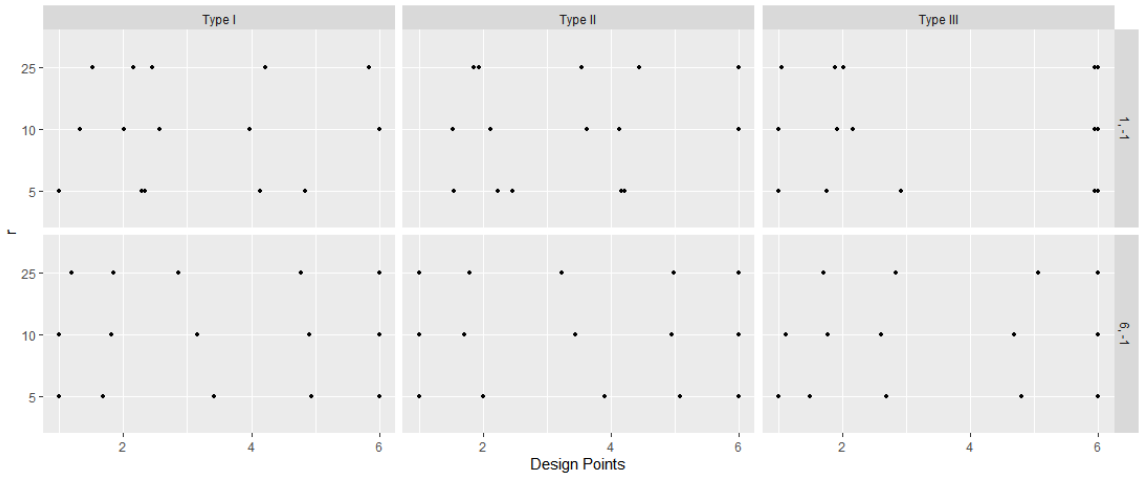

Figure 26. D-optimal Designs with $n = 5$ Observations per Subject, with $\beta_2 = -1$, $\sigma_u = 3$. The Covariance Type of of $\mathbf{\Sigma}$ Is Shown in the Top Bar and the Choice for (β_0, β_1) in the Right-hand Bar. The Value of r for the Covariance Matrix $\mathbf{\Sigma}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.

A.3 Bayesian A- and D-optimal designs for all parameter settings and priors

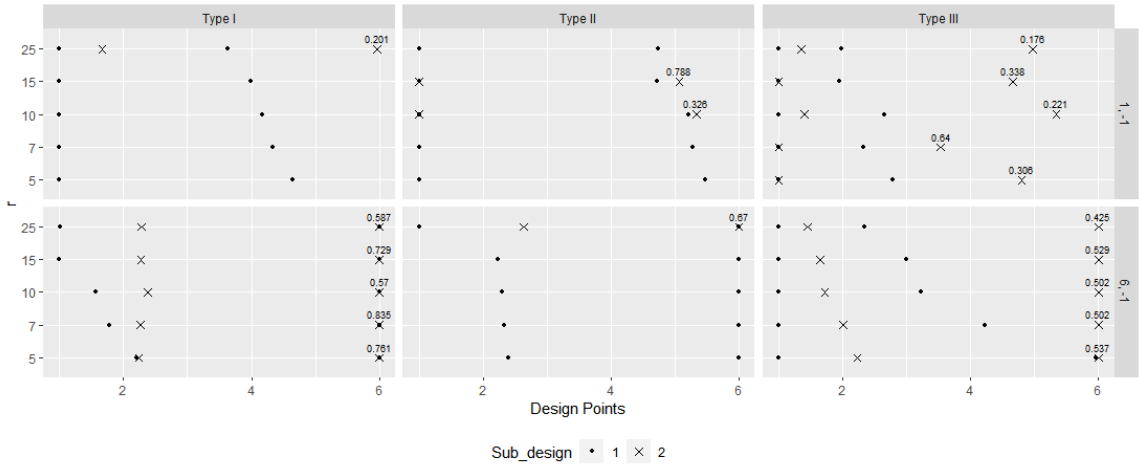

Figure 27. Bayesian D-optimal Designs with $n = 2$ Observations per Subject. The Prior Settings of $\mathbf{\Sigma}$ Are Case (a, b). The Covariance Type of $\mathbf{\Sigma}_{\boldsymbol{\beta}}$ Is Shown in the Top Bar and the Choice for $\boldsymbol{\beta}_0$ in the Right-hand Bar. The Value of r for the Covariance Matrix $\mathbf{\Sigma}_{\boldsymbol{\beta}}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.

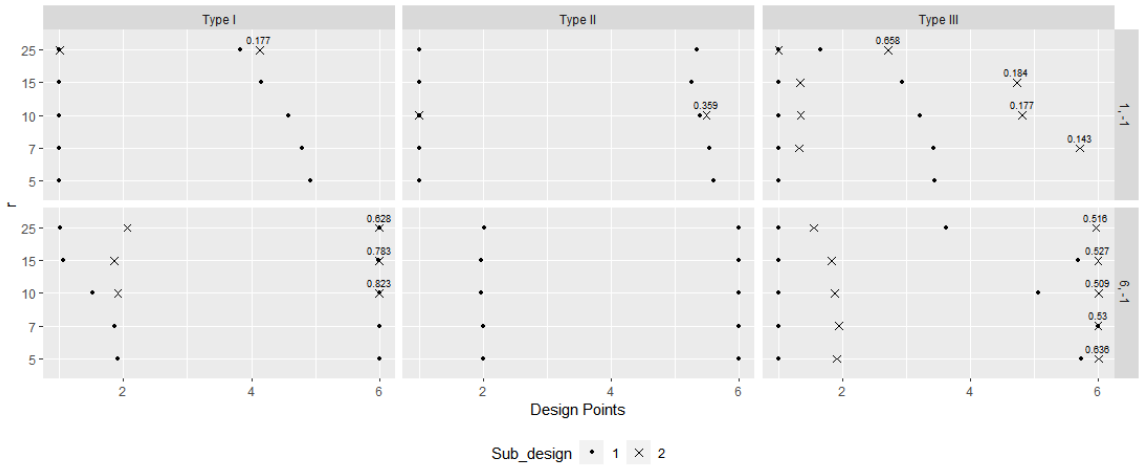

Figure 28. Bayesian D-optimal Designs with $n = 2$ Observations per Subject. The Prior Settings of $\boldsymbol{\Sigma}$ Are Case (b, a). The Covariance Type of $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ Is Shown in the Top Bar and the Choice for $\boldsymbol{\beta}_0$ in the Right-hand Bar. The Value of r for the Covariance Matrix $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.

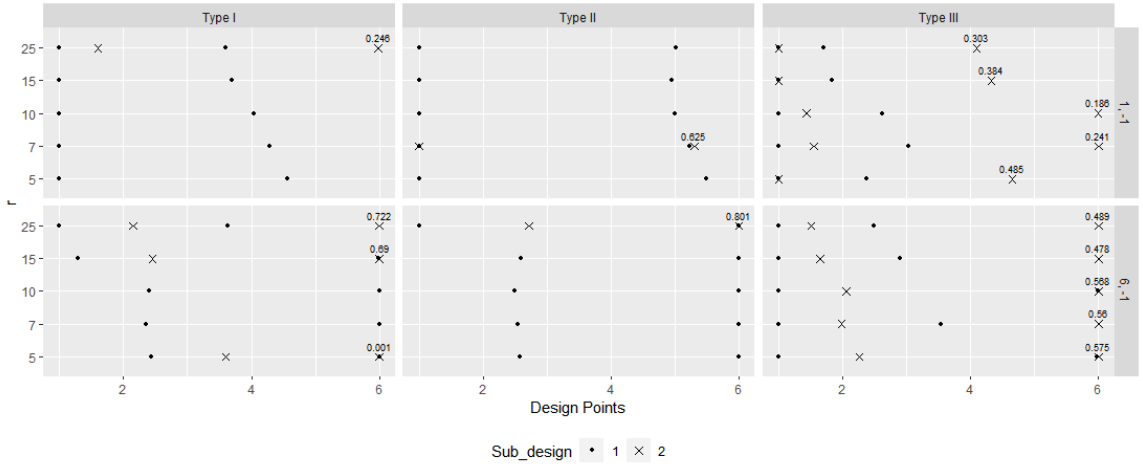

Figure 29. Bayesian D-optimal Designs with $n = 2$ Observations per Subject. The Prior Settings of $\boldsymbol{\Sigma}$ Are Case (b, b). The Covariance Type of $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ Is Shown in the Top Bar and the Choice for $\boldsymbol{\beta}_0$ in the Right-hand Bar. The Value of r for the Covariance Matrix $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.

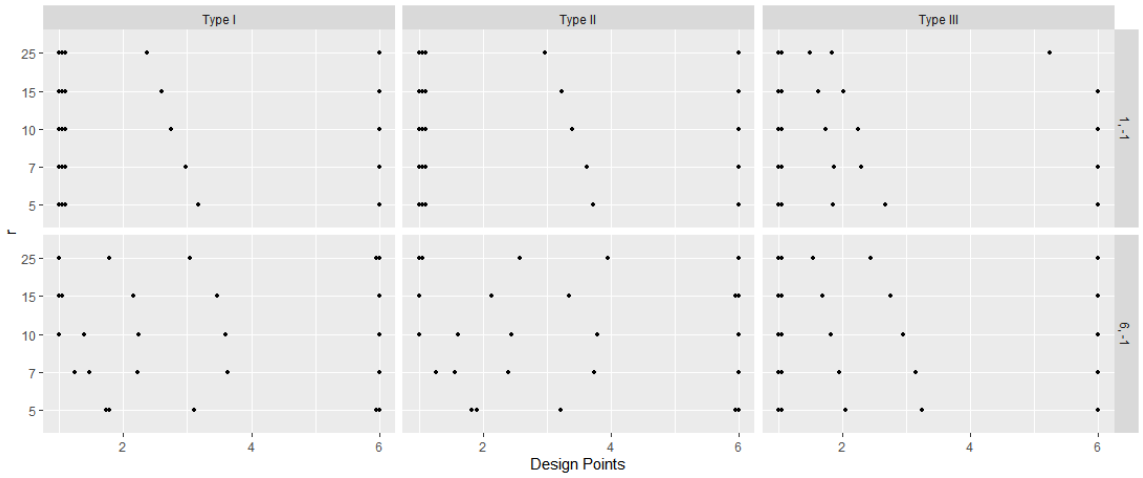

Figure 30. Bayesian D-optimal Designs with $n = 5$ Observations per Subject. The Prior Settings of $\boldsymbol{\Sigma}$ Are Case (a, b). The Covariance Type of $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ Is Shown in the Top Bar and the Choice for $\boldsymbol{\beta}_0$ in the Right-hand Bar. The Value of r for the Covariance Matrix $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.

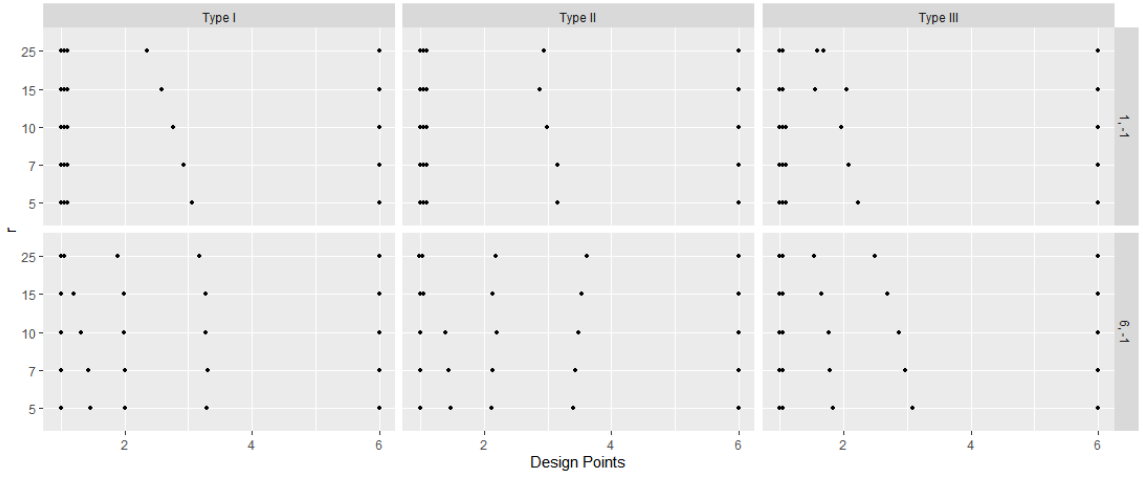

Figure 31. Bayesian D-optimal Designs with $n = 5$ Observations per Subject. The Prior Settings of $\boldsymbol{\Sigma}$ Are Case (b, a). The Covariance Type of $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ Is Shown in the Top Bar and the Choice for $\boldsymbol{\beta}_0$ in the Right-hand Bar. The Value of r for the Covariance Matrix $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.

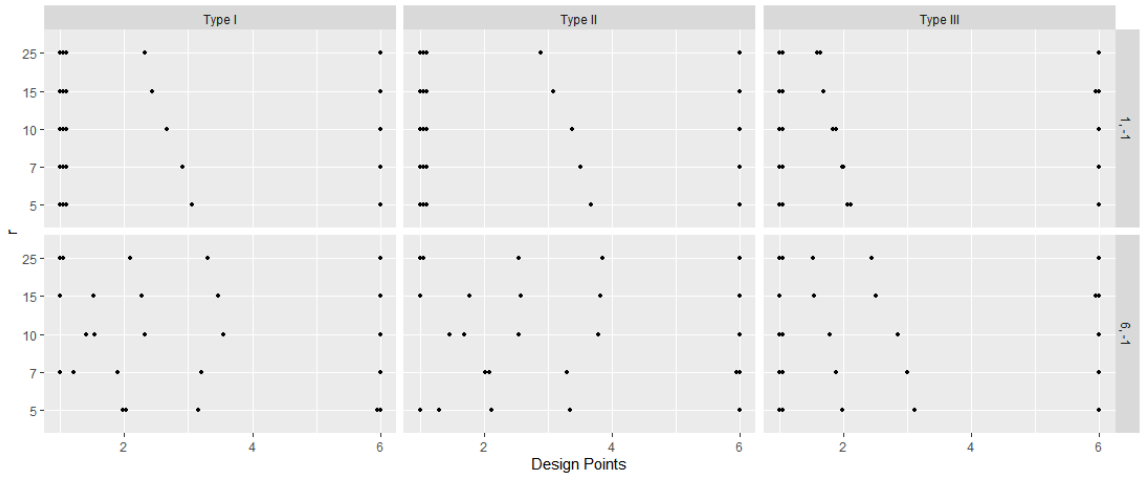

Figure 32. Bayesian D-optimal Designs with $n = 5$ Observations per Subject. The Prior Settings of $\boldsymbol{\Sigma}$ Are Case (b, b). The Covariance Type of $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ Is Shown in the Top Bar and the Choice for $\boldsymbol{\beta}_0$ in the Right-hand Bar. The Value of r for the Covariance Matrix $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.

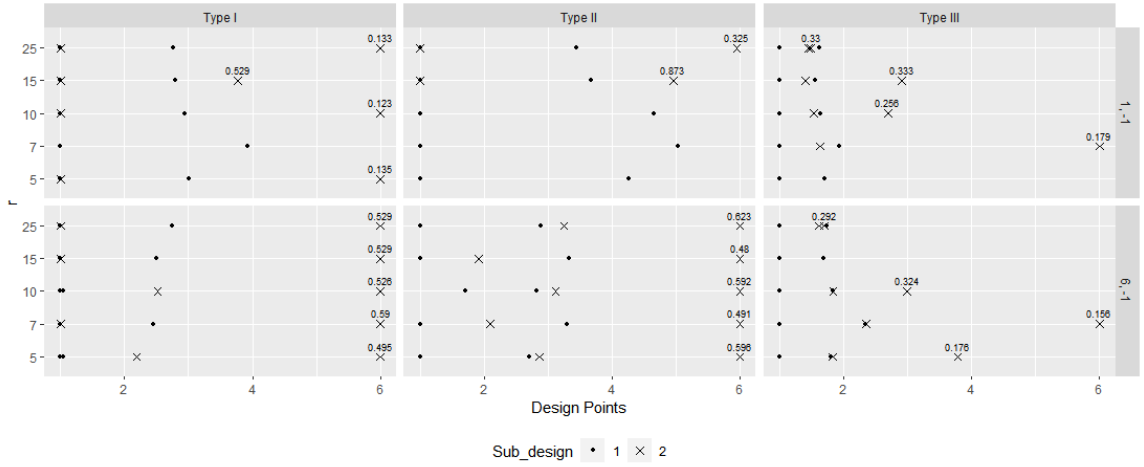

Figure 33. Bayesian A-optimal Designs with $n = 2$ Observations per Subject. The Prior Settings of $\boldsymbol{\Sigma}$ Are Case (a, b). The Covariance Type of $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ Is Shown in the Top Bar and the Choice for $\boldsymbol{\beta}_0$ in the Right-hand Bar. The Value of r for the Covariance Matrix $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.

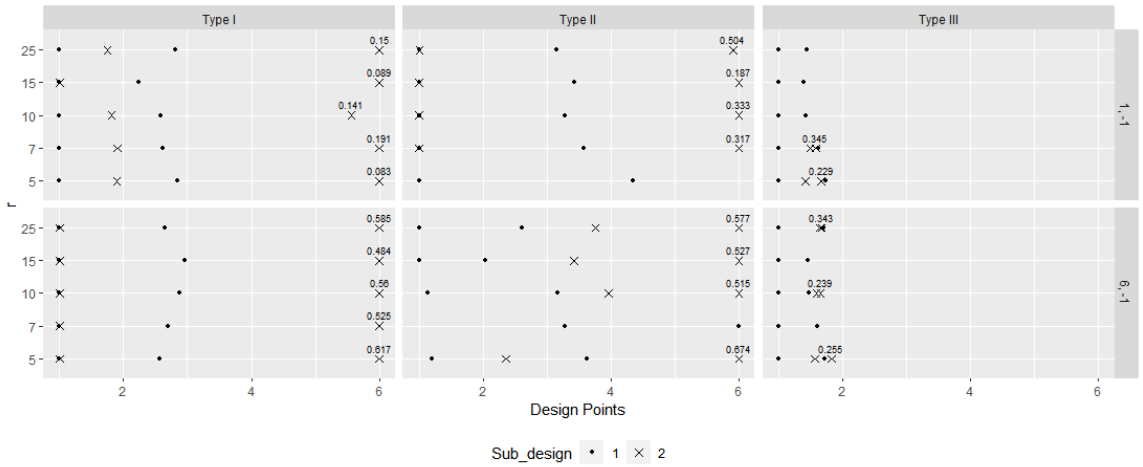

Figure 34. Bayesian A-optimal Designs with $n = 2$ Observations per Subject. The Prior Settings of $\boldsymbol{\Sigma}$ Are Case (b, a). The Covariance Type of $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ Is Shown in the Top Bar and the Choice for $\boldsymbol{\beta}_0$ in the Right-hand Bar. The Value of r for the Covariance Matrix $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.

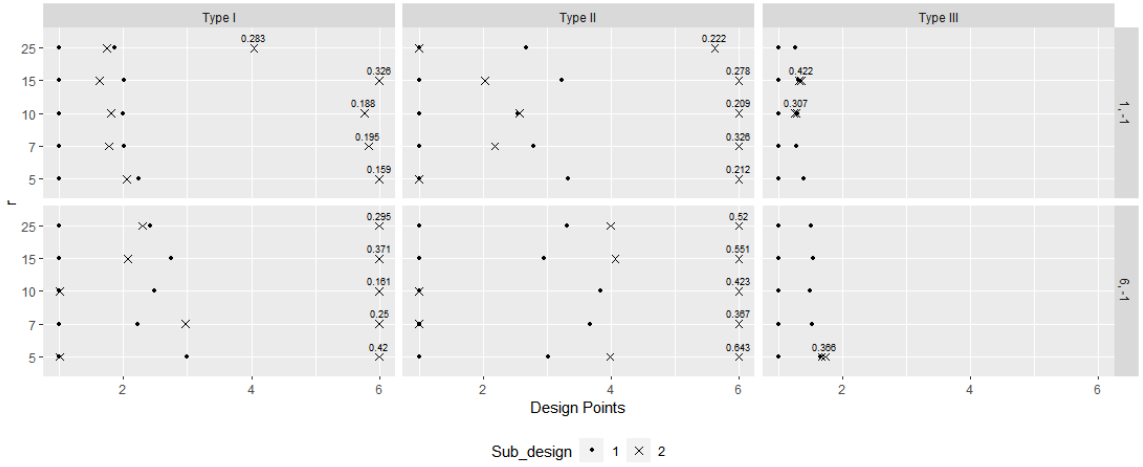

Figure 35. Bayesian A-optimal Designs with $n = 2$ Observations per Subject. The Prior Settings of $\boldsymbol{\Sigma}$ Are Case (b, b). The Covariance Type of $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ Is Shown in the Top Bar and the Choice for $\boldsymbol{\beta}_0$ in the Right-hand Bar. The Value of r for the Covariance Matrix $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.

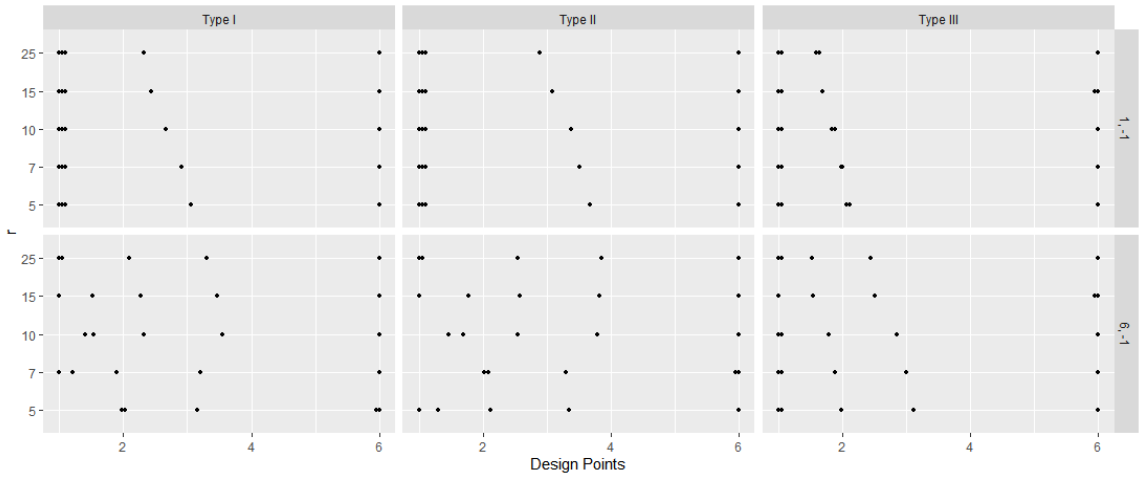

Figure 36. Bayesian A-optimal Designs with $n = 5$ Observations per Subject. The Prior Settings of $\boldsymbol{\Sigma}$ Are Case (a, b). The Covariance Type of $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ Is Shown in the Top Bar and the Choice for $\boldsymbol{\beta}_0$ in the Right-hand Bar. The Value of r for the Covariance Matrix $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.

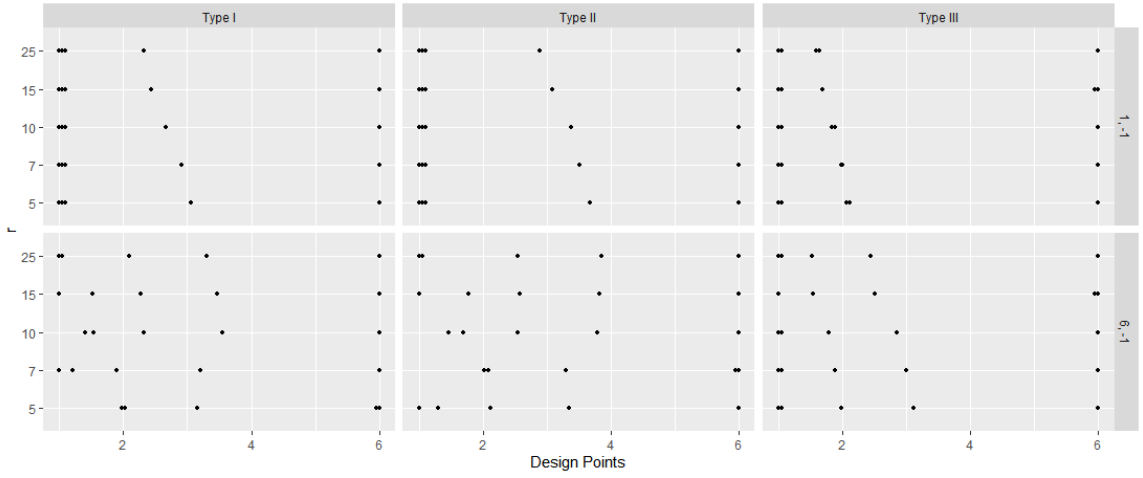

Figure 37. Bayesian A-optimal Designs with $n = 5$ Observations per Subject. The Prior Settings of $\boldsymbol{\Sigma}$ Are Case (b, a). The Covariance Type of $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ Is Shown in the Top Bar and the Choice for $\boldsymbol{\beta}_0$ in the Right-hand Bar. The Value of r for the Covariance Matrix $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.

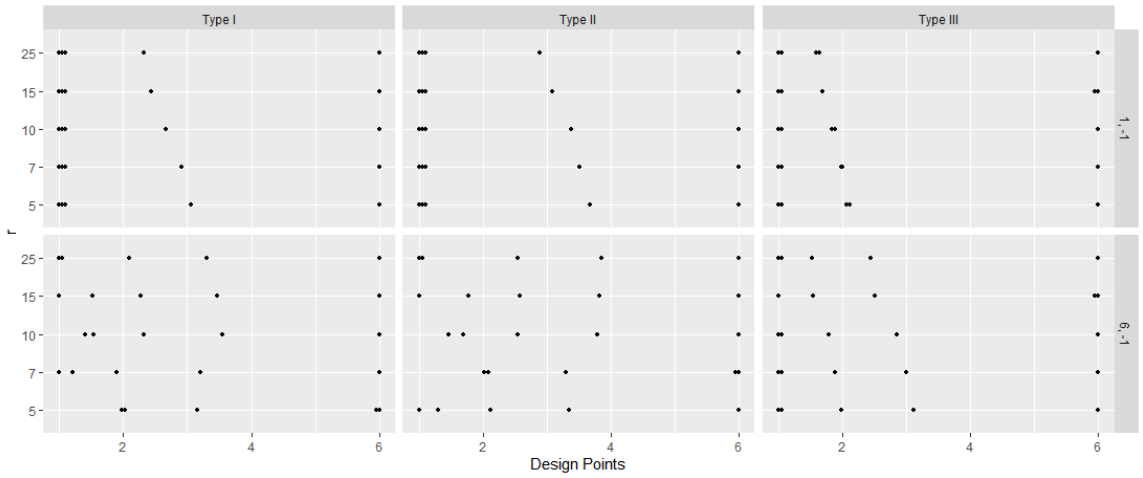

Figure 38. Bayesian A-optimal Designs with $n = 5$ Observations per Subject. The Prior Settings of $\boldsymbol{\Sigma}$ Are Case (b, b). The Covariance Type of $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ Is Shown in the Top Bar and the Choice for $\boldsymbol{\beta}_0$ in the Right-hand Bar. The Value of r for the Covariance Matrix $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ Is along the Vertical Axis, and the Design Region $[1, 6]$ Is Shown along the Horizontal Axis.