The New England Journal of Statistics in Data Science Volume 0, 1–15 (2023)
DOI: https://doi.org/10.51387/23-NEJSDS30

Bayesian D-Optimal Design of Experiments with Quantitative and Qualitative Responses

Lulu KANG*, Xinwei DENG, and Ran JIN

Abstract

Systems with both quantitative and qualitative responses are widely encountered in many applications. Design of experiment methods are needed when experiments are conducted to study such systems. Classic experimental design methods are unsuitable here because they often focus on one type of response. In this paper, we develop a Bayesian D-optimal design method for experiments with one continuous and one binary response. Both noninformative and conjugate informative prior distributions on the unknown parameters are considered. The proposed design criterion has meaningful interpretations regarding the D-optimality for the models for both types of responses. An efficient point-exchange search algorithm is developed to construct the local D-optimal designs for given parameter values. Global D-optimal designs are obtained by accumulating the frequencies of the design points in local D-optimal designs, where the parameters are sampled from the prior distributions. The performances of the proposed methods are evaluated through two examples.

KEYWORDS AND PHRASES: Bayesian D-optimal design, conjugate prior, generalized linear model, multivariate responses, noninformative prior, point-exchange.

1. INTRODUCTION

In many applications, both quantitative and qualitative responses are often collected for evaluating the quality of the system. Often, the two types of responses are mutually dependent. We call such a system with both types of quality responses quantitative-qualitative system. Such systems are widely encountered in practice [27, 26, 25]. In [25], the authors studied an experiment of the lapping stage of the wafer manufacturing process. The qualitative response is the conformity of the site total indicator reading (STIR) of the wafer, which has two possible outcomes: whether or not the STIR of a wafer is within the tolerance. The quantitative response is the total thickness variation (TTV) of the wafer. [26] focused on the birth records and examined the mutual dependency of birth weight and preterm birth. The birth weight of an infant is a quantitative outcome and the preterm birth is a binary indicator of whether an infant is born before 36 gestational weeks. The two types of outcomes are correlated as an infant is usually underweight if the infant is born preterm. In [27], two case studies of quantitative-qualitative systems from material sciences and gene expressions are illustrated. In the gene expression study, the qualitative response has three possible outcomes: healthy individuals, patients with Crohn's disease, and patients with Ulcerative colitis.

This work is motivated by a study of an etching process in a wafer manufacturing process. In the production of silicon wafers, the silicon ingot is sliced into wafers in fine geometry parameters. Inevitably, this step leaves scratches on the wafers' surface. An etching process is used to improve the surface finish, during which the wafers are submerged in the container of etchant for chemical reaction. The quality of the wafers after etching is measured by two response variables: the total thickness variation of the wafer (TTV) and the binary judgment that whether the wafer has cloudy stains in its appearance. The two responses measure the quality from different but connected aspects. There is a hidden dependency between the continuous TTV and binary judgment of stains. To improve the etching quality, expensive experiments are to be carried out to reveal this hidden dependency and to model the quality-process relationship. Therefore, an ideal experimental design for continuous and binary responses should be able to extract such useful information with economic run size.

The classic design of experiments methods mainly focus on experiments with a single continuous response. There have been various methods developed for a single discrete response too, including [40, 44, 38, 43, 42]. For multiple responses, [11] proposed the seminal work for continuous responses, [8] developed a design method for bivariate binary responses modeled by Copula functions. In the case of mixed types of responses, the literature is very scarce. A naive design method is to combine the two designs that are separately constructed for each type of response. However, such a naive strategy could be reasonable for one type of response but problematic for the other by ignoring the dependency between the types of responses, as shown in Example 1.

arXiv: 2304.08701
*Corresponding author.

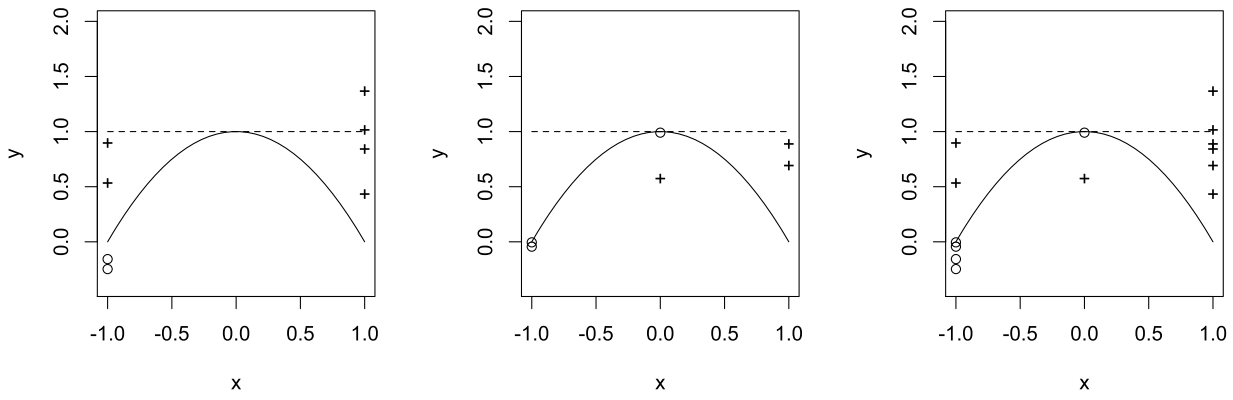

Figure 1: (a) Observations from design for Z; (b) Observations from design for Y; (c) Observations from the combined design. Dashed line "- - -" denotes $\mathbb{E}(Y|Z=1)=1$; solid line "—" denotes $\mathbb{E}(Y|Z=0)=1-x^2$; point "+" denotes (x_i, y_i) with $z_i = 1$; point "o" denotes (x_i, y_i) with $z_i = 0$.

Example 1. Denote by Y and Z a continuous response and a binary response, respectively. Assume that the true model of the binary response Z is $\mathbb{E}(Z|x) = \pi(x) = \exp(1+x)/(1+\exp(1+x))$. The true model of Y is related to Z in the form $Y|Z = z \sim N(1-(1-z)x^2, 0.3^2)$. Thus, $\mathbb{E}(Y|Z=1)=1$ and $\mathbb{E}(Y|Z=0)=1-x^2$. Using the naive design method, a 14-point design is constructed, which consists of an 8-point local D-optimal design for the model of Z with $\log(\pi(x)/(1-\pi(x)) = \eta_0 + \eta_1 x$, and a 6-point D-optimal design for linear regression with a quadratic model of x. Given the design, we generate the responses from the true models of Y and Z. Figure 1 (a)-(c) show the data (x_i, y_i, z_i) from the 8-point, 6-point and their combined 14-point design, respectively.

In this example, there is a strong dependency between the two responses since the true underlying models of $\mathbb{E}(Y|Z)$ are different when $Z=1$ and $Z=0$. In both designs for a single response shown in Figure 1 (a) and (b), the design points are balanced and reasonably distributed for the targeted response. However, since there are no Y observations for $Z=0$ at $x=1.0$ shown in Figure 1 (c), the quadratic model for $Y|Z=0$ is not estimable. Clearly, the combined design is not suitable here. Note that this problem is not caused by outliers, since all the points for $Z=1$ (with "+") are varied around $Y=1$ and the points for $Z=0$ (with "o") are around $Y=1-x^2$. In fact $P(Z=0|x=1)=0.12$, which is relatively small. Thus it is less likely to observe Y with $Z=0$ at $x=1.0$. A simple solution is to add more replications at $x=1.0$, but it is not clear how many replications should be sufficient. It becomes more difficult to spot a direct solution when the experiments get more complicated.

Such experiments call for new experimental design methods to account for both continuous and binary responses. Note that under the experimental design framework, the linear model is often considered for modeling the continuous response, and the generalized linear model (GLM) is often considered for modeling the qualitative response. A joint model must be developed to incorporate both types of responses. Compared to the classic design methods for linear models or GLMs, design for the joint model is more challenging due to the following aspects. First, the design criterion for the joint model is more complicated, as the joint model is more complicated than the separate ones. Second, experimental design for the GLM itself is more difficult than that for the linear model, which is naturally inherited by the design for the joint model. Third, efficient design construction algorithms are needed to handle the complexity of the design criterion based on the joint model. [23] proposed an A-optimal design for the experiments with both quantitative and qualitative responses. The A-optimality was derived under a Bayesian framework proposed in [25]. Although [23] addressed the three challenges to a degree, the A-optimality is not a commonly used criterion. More importantly, only informative prior is considered, which circumvented some difficulties brought by noninformative prior of the parameters.

In this paper, we choose the most commonly used D-optimal design criterion and propose a novel Bayesian design method for the continuous and binary responses. The proposed method considers both cases of noninformative priors and informative priors. With the noninformative priors, the Bayesian framework is equivalent to the frequentist approach. In this case, we also establish some regularity conditions on the experimental run sizes. With the informative priors, we develop the D-optimal design using conjugate priors. The derived design criterion has meaningful interpretations in terms of the D-optimality criteria for the models of both continuous and binary responses. Moreover, we develop an efficient point-exchange algorithm to construct the proposed designs. The construction algorithm can be applied to more general settings other than factorial designs.

The rest of the paper is organized as follows. Section 2 reviews the general Bayesian quantitative-qualitative (QQ) model and the optimal design criterion. The Bayesian D-optimal design criterion is derived using noninformative

prior distributions in Section 3. In Section 4, the design criterion is derived with conjugate informative priors. Efficient algorithms for constructing optimal designs are elaborated in Section 5. One artificial example and the etching experimental design are shown in Section 6. Section 7 concludes this paper with some discussions.

2. GENERAL BAYESIAN QQ MODEL AND DESIGN

We first review the general Bayesian QQ model introduced in [25] and focus on the scenario that Y is a continuous response and Z is a binary response. The input variable $\boldsymbol{x} = (x_1, \ldots, x_p)' \in \mathbb{R}^p$ contains p dimensions. Denote the data as $(\boldsymbol{x}_i, y_i, z_i), i = 1, \ldots, n$, where $y_i \in \mathbb{R}$ and $z_i \in \{0, 1\}$. The vectors $\boldsymbol{y} = (y_i, \ldots, y_n)'$ and $\boldsymbol{z} = (z_1, \ldots, z_n)'$ are the vectors of response observations. To jointly model the continuous response Y and the binary response Z given $\boldsymbol{x}$, consider the joint probability of $Y|Z$ and Z. The conditional model on $Y|Z$ is assumed to be a linear regression model, while the model of Z is a logistic regression model. Specifically, we consider joint modeling of Y and Z as follows,

$$Z = \begin{cases} 1, & \text{with probability } \pi(\boldsymbol{x}) \\ 0, & \text{with probability } 1 - \pi(\boldsymbol{x}) \end{cases}$$
$$\text{with } \pi(\boldsymbol{x}, \boldsymbol{\eta}) = \frac{\exp(\boldsymbol{f}(\boldsymbol{x})'\boldsymbol{\eta})}{1 + \exp(\boldsymbol{f}(\boldsymbol{x})'\boldsymbol{\eta})}, \tag{2.1}$$

where $\boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}), \ldots, f_q(\boldsymbol{x}))$ contains q modeling effects including the intercept, the main, interaction and quadratic effects, etc., and $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_q)'$ is a vector of parameter coefficients. Conditioning on $Z = z$, the quantitative variable Y has the distribution

$$Y|Z = z \sim N(\mu_0 + z\boldsymbol{f}(\boldsymbol{x})'\boldsymbol{\beta}^{(1)} + (1-z)\boldsymbol{f}(\boldsymbol{x})'\boldsymbol{\beta}^{(2)}, \sigma^2), \tag{2.2}$$

where $\boldsymbol{\beta}^{(i)} = (\beta_1^{(i)}, \ldots, \beta_q^{(i)})'$, $i = 1, 2$ are the corresponding coefficients of the model effects. The parameter μ_0 is the mean and σ^2 is the noise variance. The above conditional model (2.2) indicates that $Y|Z = 1 \sim N(\mu_0 + \boldsymbol{f}(x)'\boldsymbol{\beta}^{(1)}, \sigma^2)$ and $Y|Z = 0 \sim N(\mu_0 + \boldsymbol{f}(\boldsymbol{x})'\boldsymbol{\beta}^{(2)}, \sigma^2)$. We assume the same variance σ^2 for the two conditional distributions of $Y|Z = 1$ and 0. The design method developed in the paper can be easily adapted to the case with different variances.

The association between the two responses Y and Z is represented using the conditional model $Y|Z$. When the two linear models for $Y|Z = 0$ and $Y|Z = 1$ are different, i.e., $\boldsymbol{\beta}^{(1)} \neq \boldsymbol{\beta}^{(2)}$, then it is important to take account of the influence of the qualitative response Z when modeling the quantitative response Y. Let $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$ be the $n \times p$ design matrix with $\boldsymbol{x}_i$ as the ith design point. Based on the CB model, we can express the sampling distributions as

$$\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}, \mu_0, \sigma^2, \boldsymbol{X} \sim N(\mu_0\mathbf{1} + \boldsymbol{V}_1\boldsymbol{F}\boldsymbol{\beta}^{(1)} + \boldsymbol{V}_2\boldsymbol{F}\boldsymbol{\beta}^{(2)}, \sigma^2\boldsymbol{I}_n),$$
$$\boldsymbol{z}|\boldsymbol{\eta}, \boldsymbol{X} \sim \text{Bernoulli}(\pi(\boldsymbol{x}_i, \boldsymbol{\eta})) \text{ for } i = 1, \ldots, n, \text{ and}$$
$$p(\boldsymbol{z}|\boldsymbol{\eta}, \boldsymbol{X}) \propto \exp\left\{\sum_{i=1}^{n}\left(z_i\boldsymbol{f}(\boldsymbol{x}_i)'\boldsymbol{\eta} - \log(1 + e^{\boldsymbol{f}(\boldsymbol{x}_i)'\boldsymbol{\eta}})\right)\right\},$$

where $p(\cdot)$ denotes a general density function. Here $\boldsymbol{V}_1 = \text{diag}\{z_1, \ldots, z_n\}$ is a diagonal matrix, $\boldsymbol{I}_n$ is the $n \times n$ identity matrix and $\boldsymbol{V}_2 = \boldsymbol{I}_n - \boldsymbol{V}_1$, $\boldsymbol{F}$ is the model matrix with the ith row as $\boldsymbol{f}(\boldsymbol{x}_i)'$, and $\mathbf{1}$ is a vector of ones.

Denote $p(\boldsymbol{\beta}^{(1)})$, $p(\boldsymbol{\beta}^{(2)})$, and $p(\boldsymbol{\eta})$ as the prior distributions of the parameters $\boldsymbol{\beta}^{(1)}$, $\boldsymbol{\beta}^{(2)}$, and $\boldsymbol{\eta}$. Note that we focus on the estimation accuracy of the three groups of parameters. The mean μ_0 and variance σ^2 are considered nuisance parameters and thus excluded from the optimal design criterion. In this work, we assume that the priors of $\boldsymbol{\beta}^{(1)}$, $\boldsymbol{\beta}^{(2)}$, and $\boldsymbol{\eta}$ are independent. Under this assumption, the conditional posterior distribution of $\boldsymbol{\eta}$, $\boldsymbol{\beta}^{(1)}$, and $\boldsymbol{\beta}^{(2)}$ are also independent as explained in Sections 3 and 4. Under the Bayesian framework, the conditional posterior distribution of the parameters $(\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}, \boldsymbol{\eta})$ can be derived as

$$p(\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}, \boldsymbol{\eta}|\boldsymbol{y}, \boldsymbol{z}, \mu_0, \sigma^2, \boldsymbol{X}) \tag{2.3}$$
$$\propto p(\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}, \mu_0, \sigma^2, \boldsymbol{X})p(\boldsymbol{\beta}^{(1)})p(\boldsymbol{\beta}^{(2)})p(\boldsymbol{z}|\boldsymbol{\eta}, \boldsymbol{X})p(\boldsymbol{\eta}).$$

Using (2.3) we develop the general Bayesian optimal design criterion. Let ψ be a criterion function on the conditional posterior distribution of the parameters. For example, it can be the Shannon information (or equivalently, the Kullback-Leibler distance), A/I-optimality [16], or other design criteria. However, $\psi(\cdot)$ cannot be directly used as the final optimal design criterion because its value depends on the random parameters $(\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}, \boldsymbol{\eta})$ and the experimental outputs $(\boldsymbol{y}, \boldsymbol{z})$ that are not yet observed. The randomness of $(\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}, \boldsymbol{\eta})$ can be removed by calculating the mean of ψ with respect to these parameters. The uncertainty of $(\boldsymbol{y}, \boldsymbol{z})$ can be removed by calculating the mean $\mathbb{E}(\mathbb{E}(\psi|\boldsymbol{y}, \boldsymbol{z}))$. Therefore, the general Bayesian optimal design criterion on the design matrix $\boldsymbol{X}$ is

$$\Psi(\boldsymbol{X}|\mu_0, \sigma^2) \tag{2.4}$$
$$= \int p(\boldsymbol{y}, \boldsymbol{z}|\mu_0, \sigma^2, \boldsymbol{X}) \times \left(\int \psi(p(\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}, \boldsymbol{\eta}|\boldsymbol{y}, \boldsymbol{z}, \mu_0, \sigma^2, \boldsymbol{X})) \times p(\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}, \boldsymbol{\eta}|\boldsymbol{y}, \boldsymbol{z}, \mu_0, \sigma^2, \boldsymbol{X}) \mathrm{d}\boldsymbol{\beta}^{(1)} \mathrm{d}\boldsymbol{\beta}^{(2)} \mathrm{d}\boldsymbol{\eta}\right) \mathrm{d}\boldsymbol{y} \mathrm{d}\boldsymbol{z}.$$

It is well-known that the Bayesian D-optimal design is equivalent to the Shannon information criterion [3], omitting the constant terms to $\boldsymbol{X}$. The criterion function $\psi(\cdot)$ of Shannon information is $\log(\cdot)$. Next, we develop the Bayesian D-optimal design criteria (2.4) under different prior distributions.

3. OPTIMAL DESIGN UNDER NONINFORMATIVE PRIORS

When lacking domain knowledge or proper historical data, experimenters often favor the frequentist approach as no priors need to be specified. The frequentist approach can be seen as the Bayesian approach using noninformative priors. In this section, we derive the optimal design criterion and the regularity conditions for noninformative priors.

3.1 Design Criterion

Assume the non-informative priors $p(\boldsymbol{\beta}^{(i)}) \propto 1$ for $i = 1, 2$ and $p(\boldsymbol{\eta}) \propto 1$. The conditional posterior distribution in (2.3) is the same as the joint distribution of the data. It can be further factorized into

$$p(\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}, \boldsymbol{\eta} | \boldsymbol{y}, \boldsymbol{z}, \mu_0, \sigma^2, \boldsymbol{X}) \propto p(\boldsymbol{\eta}|\boldsymbol{z}, \boldsymbol{X}) \prod_{i=1}^{2} p(\boldsymbol{\beta}^{(i)}|\boldsymbol{y}, \boldsymbol{z}, \mu_0, \sigma^2, \boldsymbol{X}),$$

with the posterior distributions

$$\boldsymbol{\beta}^{(i)}|\boldsymbol{y}, \boldsymbol{z}, \mu_0, \sigma^2, \boldsymbol{X} \sim N\left((\boldsymbol{F}'\boldsymbol{V}_i\boldsymbol{F})^{-1}\boldsymbol{F}'\boldsymbol{V}_i(\boldsymbol{y} - \mu_0\boldsymbol{1}), \sigma^2(\boldsymbol{F}'\boldsymbol{V}_i\boldsymbol{F})^{-1}\right) \quad (3.1)$$

for $i = 1, 2$, and

$$p(\boldsymbol{\eta}|\boldsymbol{z}, \boldsymbol{X}) \propto \exp\left\{\sum_{i=1}^{n}\left(z_i \boldsymbol{f}(\boldsymbol{x}_i)'\boldsymbol{\eta} - \log(1 + e^{\boldsymbol{f}(\boldsymbol{x}_i)'\boldsymbol{\eta}})\right)\right\}. \quad (3.2)$$

Conditioning on $\boldsymbol{z}$, the posterior distributions of $(\boldsymbol{\beta}^1, \boldsymbol{\beta}^2)$ and $\boldsymbol{\eta}$ are independent. Note that the noninformative prior $p(\boldsymbol{\eta}) \propto 1$ is proper because it leads to proper posterior $p(\boldsymbol{\eta}|\boldsymbol{z}, \boldsymbol{X})$. Under the noninformative priors, the Bayesian estimation is identical to the frequentist estimation.

Using the posterior distributions (3.1)–(3.2) and the criterion function $\psi(\cdot) = \log(\cdot)$ in the general Bayesian optimal design criterion (2.4), we obtain the Bayesian D-optimal design criterion (3.3).

$$\Psi(\boldsymbol{X}|\mu_0, \sigma^2) = \mathbb{E}_{\boldsymbol{z},\boldsymbol{\eta}}\left\{\log(p(\boldsymbol{\eta}|\boldsymbol{z}, \boldsymbol{X}))\right\} + \frac{1}{2}\sum_{i=1}^{2}\mathbb{E}_{\boldsymbol{\eta}}\mathbb{E}_{\boldsymbol{z}|\boldsymbol{\eta}}\left\{\log\det\{(\boldsymbol{F}'\boldsymbol{V}_i\boldsymbol{F})\}\right\} + \text{constant}. \quad (3.3)$$

The derivation is in Supplement S1. The first additive term in (3.3) is exactly the Bayesian D-optimal design criterion for GLMs. Unfortunately, its exact integration is not tractable. The common approach in experimental design for GLMs is to use a normal approximation for the posterior distribution $p(\boldsymbol{\eta}|\boldsymbol{z}, \boldsymbol{X})$ [3, 28]. Such an approximation leads to

$$\mathbb{E}_{\boldsymbol{z},\boldsymbol{\eta}}\left\{\log(p(\boldsymbol{\eta}|\boldsymbol{z}, \boldsymbol{X}))\right\} \approx \mathbb{E}_{\boldsymbol{\eta}}\{\log\det \boldsymbol{I}(\boldsymbol{\eta}|\boldsymbol{X})\} + \text{constant}, \quad (3.4)$$

where $\boldsymbol{I}(\boldsymbol{\eta}|\boldsymbol{X})$ is the Fisher information matrix. We can easily show that

$$\boldsymbol{I}(\boldsymbol{\eta}|\boldsymbol{X}) = -\mathbb{E}_{\boldsymbol{z}}\left(\frac{\partial^2 l(\boldsymbol{z}, \boldsymbol{\eta}|\boldsymbol{X})}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^T}\right) = \sum_{i=1}^{n}\boldsymbol{f}(\boldsymbol{x}_i)\boldsymbol{f}(\boldsymbol{x}_i)'\pi(\boldsymbol{x}_i, \boldsymbol{\eta})(1 - \pi(\boldsymbol{x}_i, \boldsymbol{\eta})) = \boldsymbol{F}'\boldsymbol{W}_0\boldsymbol{F},$$

where $\boldsymbol{W}_0$ is a diagonal weight matrix

$$\boldsymbol{W}_0 = \text{diag}\{\pi(\boldsymbol{x}_1, \boldsymbol{\eta})(1-\pi(\boldsymbol{x}_1, \boldsymbol{\eta})), \ldots, \pi(\boldsymbol{x}_n, \boldsymbol{\eta})(1-\pi(\boldsymbol{x}_n, \boldsymbol{\eta}))\}.$$

Omitting the irrelevant constant, we approximate the exact criterion $\Psi(\boldsymbol{X}|\mu_0, \sigma^2)$ in (3.3) as follows.

$$\Psi(\boldsymbol{X}|\mu_0, \sigma^2) \approx \mathbb{E}_{\boldsymbol{\eta}}\{\log\det(\boldsymbol{F}'\boldsymbol{W}_0\boldsymbol{F})\} + \frac{1}{2}\sum_{i=1}^{2}\mathbb{E}_{\boldsymbol{\eta}}\mathbb{E}_{\boldsymbol{z}|\boldsymbol{\eta}}\left\{\log\det(\boldsymbol{F}'\boldsymbol{V}_i\boldsymbol{F})\right\}. \quad (3.5)$$

To construct the optimal design, we consider maximizing the approximated $\Psi(\boldsymbol{X}|\mu_0, \sigma^2)$ in (3.5). But this is not trivial, because it involves the expectation on Z_i's in the second additive term. To overcome this challenge, [18] constructed optimal designs by simulating samples from the joint distribution of responses and the unknown parameters. But this method can be computationally expensive for even slightly larger dimensions of experimental factors. Instead of simulating Z_i's, we derive the following Theorem 1 that gives a tractable upper bound $Q(\boldsymbol{X})$. Thus we propose using the upper bound $Q(\boldsymbol{X})$ as the optimal criterion.

Theorem 1. *Assume that the matrices $\boldsymbol{F}'\boldsymbol{W}_0\boldsymbol{F}$, $\boldsymbol{F}'\boldsymbol{V}_1\boldsymbol{F}$, and $\boldsymbol{F}'\boldsymbol{V}_2\boldsymbol{F}$ are all nonsingular. Omitting the irrelevant constant, an upper bound of the approximated $\Psi(\boldsymbol{X}|\mu_0, \sigma^2)$ is*

$$Q(\boldsymbol{X}) = \mathbb{E}_{\boldsymbol{\eta}}\left\{\log\det(\boldsymbol{F}'\boldsymbol{W}_0\boldsymbol{F}) + \frac{1}{2}\sum_{i=1}^{2}\log\det(\boldsymbol{F}'\boldsymbol{W}_i\boldsymbol{F})\right\}, \quad (3.6)$$

where $\boldsymbol{W}_1 = diag\{\pi(\boldsymbol{x}_1, \boldsymbol{\eta}), \ldots, \pi(\boldsymbol{x}_n, \boldsymbol{\eta})\}$ and $\boldsymbol{W}_2 = \boldsymbol{I}_n - \boldsymbol{W}_1$.

The proof of Theorem 1 is in Supplement S1. Note that Theorem 1 requires that $\boldsymbol{F}'\boldsymbol{W}_i\boldsymbol{F}$ for $i = 0, 1, 2$ are all nonsingular. Obviously $\boldsymbol{W}_0 = \boldsymbol{W}_1\boldsymbol{W}_2$. It is easy to see that $\boldsymbol{F}'\boldsymbol{W}_0\boldsymbol{F}$ is nonsingular if and only if both $\boldsymbol{F}'\boldsymbol{W}_1\boldsymbol{F}$ and $\boldsymbol{F}'\boldsymbol{W}_2\boldsymbol{F}$ are nonsingular.

The matrices $\boldsymbol{F}'\boldsymbol{V}_1\boldsymbol{F}$ and $\boldsymbol{F}'\boldsymbol{V}_2\boldsymbol{F}$ involve the responses Z_i's that are not yet observed at the experimental design stage. We can only choose the experimental run size and the design points to avoid the singularity problem with a larger probability for given values of $\boldsymbol{\eta}$. Once the run size is chosen, the design points can be optimally arranged by maximizing $Q(\boldsymbol{X})$. The weight matrix $\boldsymbol{W}_1$ (or $\boldsymbol{W}_2$) gives more weight to the feasible design points that are more likely to lead to $Z = 1$ (or $Z = 0$) observations so that the parameters $\boldsymbol{\beta}^{(1)}$

(or $\boldsymbol{\beta}^{(2)}$) of the linear model $Y|Z=1$ (or $Y|Z=0$) are more likely to be estimable. Next, we introduce some regularity conditions on the run size and number of replications to alleviate the singularity problem.

3.2 Regularity Conditions

Let m be the number of distinct design points in the design matrix $\boldsymbol{X}$, n_i be the number of repeated point $\boldsymbol{x}_i$ in $\boldsymbol{X}$ for $i=1,\ldots,m$. Thus $n=\sum_{i=1}^{m} n_i$ and $n_i \geq 1$ for $i=1,\ldots,m$. First, it is necessary that $m \geq q$ for the linear regression model to be estimable under the noninformative priors. The if-and-only-if condition for $\boldsymbol{F}'\boldsymbol{W}_0\boldsymbol{F}$ to be nonsingular is that $\text{rank}(\boldsymbol{F}'\boldsymbol{W}_0\boldsymbol{F}) \geq q$. If $m \geq q$ and $\pi(\boldsymbol{x}_i,\boldsymbol{\eta}) \in (0,1)$ for $i=1,\ldots,m$, then $\boldsymbol{F}'\boldsymbol{W}_i\boldsymbol{F}$ for $i=0,1,2$ are all nonsingular and thus positive definite. To make sure $\pi(\boldsymbol{x}_i,\boldsymbol{\eta}) \in (0,1)$ for $i=1,\ldots,m$, it is sufficient to assume that $\boldsymbol{\eta}$ is finitely bounded. This condition is typically used for the frequentist D-optimal design for GLMs. For instance, [44] suggested using the centroids of the finite bounded space of $\boldsymbol{\eta}$ to develop the local D-optimal design for GLMs. [13] clustered different local D-optimal designs with $\boldsymbol{\eta}$ randomly sampled from its bounded space.

The if-and-only-if condition for $\boldsymbol{F}'\boldsymbol{V}_1\boldsymbol{F}$ and $\boldsymbol{F}'\boldsymbol{V}_2\boldsymbol{F}$ being nonsingular is

$$\sum_{i=1}^{m} I\left(\sum_{j=1}^{n_i} Z_{ij} > 0\right) \geq q \quad \text{and} \quad \sum_{i=1}^{m} I\left(\sum_{j=1}^{n_i} Z_{ij} < n_i\right) \geq q, \tag{3.7}$$

where Z_{ij} is the jth random binary response at the unique design point $\boldsymbol{x}_i$ and $I(\cdot)$ is the indicator function. In the following, we discuss how to choose sample sizes n_i and n under two scenarios (i) $m=q$ and (ii) $m>q$.

Proposition 1. *Assume $m=q$. Both $\boldsymbol{F}'\boldsymbol{V}_1\boldsymbol{F}$ and $\boldsymbol{F}'\boldsymbol{V}_2\boldsymbol{F}$ are nonsingular if and only if $I(0<\sum_{j=1}^{n_i} Z_{ij}<n_i)=1$ for $i=1,2,\ldots,m$. For any given $\kappa \in (0,1)$, a sufficient condition on n_i for $\Pr(0<\sum_{j=1}^{n_i} Z_{ij}<n_i) \geq \kappa$ is*

$$n_i \geq 1 + \left\lceil \frac{\log(1-\kappa)}{\log\left(\max\left\{\pi(\boldsymbol{x}_i,\boldsymbol{\eta}), 1-\pi(\boldsymbol{x}_i,\boldsymbol{\eta})\right\}\right)} \right\rceil \tag{3.8}$$

for $i=1,2,\ldots,m$, and a necessary condition is

$$n_i \geq \left\lceil \frac{2\log\left(\frac{1-\kappa}{2}\right)}{\log \pi(\boldsymbol{x}_i,\boldsymbol{\eta}) + \log(1-\pi(\boldsymbol{x}_i,\boldsymbol{\eta}))} \right\rceil \tag{3.9}$$

for $i=1,2,\ldots,m$.

Proposition 2. *Assume $m>q$. To make both $\boldsymbol{F}'\boldsymbol{V}_1\boldsymbol{F}$ and $\boldsymbol{F}'\boldsymbol{V}_2\boldsymbol{F}$ nonsingular with large probability, or equivalently,*

$$\mathbb{E}\left\{\sum_{i=1}^{m} I\left(\sum_{j=1}^{n_i} Z_{ij} > 0\right)\right\} \geq q$$

and

$$\mathbb{E}\left\{\sum_{i=1}^{m} I\left(\sum_{j=1}^{n_i} Z_{ij} < n_i\right)\right\} \geq q,$$

(i) a sufficient condition is

$$n_0 \geq \max\left\lceil \left\{1, \frac{\log(1-q/m)}{\log(1-\pi_{\min})}, \frac{\log(1-q/m)}{\log \pi_{\max}}\right\}\right\rceil, \tag{3.10}$$

which is the same as

$$n \geq \left\lceil m \cdot \max\left\{1, \frac{\log(1-q/m)}{\log(1-\pi_{\min})}, \frac{\log(1-q/m)}{\log \pi_{\max}}\right\}\right\rceil, \tag{3.11}$$

where $n_0=\min\{n_1,\ldots,n_m\}$, $\pi_{\min}=\min_{i=1}^{m}\pi(\boldsymbol{x}_i,\boldsymbol{\eta})$, $\pi_{\max}=\max_{i=1}^{m}\pi(\boldsymbol{x}_i,\boldsymbol{\eta})$ and $\boldsymbol{x}_i$'s are the unique design points;

(ii) a necessary condition is

$$n_0 \geq \left\lceil \max\left\{1, \frac{\log(1-q/m)}{\log(1-\pi_{\max})}, \frac{\log(1-q/m)}{\log \pi_{\min}}\right\}\right\rceil \tag{3.12}$$

which is the same as

$$n \geq \left\lceil m \cdot \max\left\{1, \frac{\log(1-q/m)}{\log(1-\pi_{\max})}, \frac{\log(1-q/m)}{\log \pi_{\min}}\right\}\right\rceil. \tag{3.13}$$

Proposition 1 gives a sufficient condition on the lower bound of n_i when saturated design ($m=q$) is used. Under the sufficient condition, the nonsingularity of $\boldsymbol{F}\boldsymbol{V}_1\boldsymbol{F}$ and $\boldsymbol{F}\boldsymbol{V}_2\boldsymbol{F}$ holds with a probability larger than κ^m. For Example 1 in Section 1, suppose that if the possible values of x can only be $-1,0,1$, then $m=q=3$. Let $\boldsymbol{\eta}=(1,1)'$. If $\kappa=0.5$, then the numbers of replications for $x=-1,0,1$ need to satisfy $n_1 \geq 2$, $n_2 \geq 4$, and $n_3 \geq 7$, respectively. If $\kappa=0.9$, then $n_1 \geq 5$, $n_2 \geq 9$, and $n_3 \geq 20$. Proposition 1 is useful in Step 1 to construct the initial design in Algorithm 2 in Section 5.

Proposition 2 provides one sufficient condition and one necessary condition when $m>q$ on the smallest number of replications and the overall run size. But these conditions only examine the nonsingularity of the two matrices with large probability, which is weaker than Proposition 1. For given $\boldsymbol{\eta}$ value, Algorithm 2 in Section 5.1 can return the local D-optimal design. Proposition 2 can be useful to check the local D-optimal design, as $\pi_{\min}$ and $\pi_{\max}$ depend on $\boldsymbol{\eta}$. Take the artificial example in Section 6.1 for instance. The local D-optimal design for $\rho=0$ (Table S1 in Supplement S2) has $m=50$ unique design points and there are $q=22$ effects. According to Proposition 2, the sufficient condition requires $n_0 \geq 7$ and the necessary condition requires $n_0 \geq 1$. The local D-optimal design in Table S1 only satisfies the necessary condition. To meet the sufficient condition n has to be much larger. For the global optimal design considering all possible $\boldsymbol{\eta}$ values, Proposition 2 can provide some

guidelines for the design construction when $\boldsymbol{\eta}$ is varied in a relatively small range.

4. OPTIMAL DESIGN UNDER CONJUGATE PRIORS

When prior information for parameters $\boldsymbol{\beta}^{(1)}$, $\boldsymbol{\beta}^{(2)}$, and $\boldsymbol{\eta}$ is available, it would be desirable to consider the optimal design under the informative priors. In this section, we detail the proposed Bayesian D-optimal design using the conjugate priors.

4.1 Design Criterion

For the parameters $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$, the conjugate priors are normal distribution since $Y|Z$ follows normal distribution. Thus we consider their priors as

$$\boldsymbol{\beta}^{(1)} \sim N(\mathbf{0}, \tau^2 \boldsymbol{R}_1), \quad \boldsymbol{\beta}^{(2)} \sim N(\mathbf{0}, \tau^2 \boldsymbol{R}_2).$$

where τ^2 is the prior variance and $\boldsymbol{R}_i$ is the prior correlation matrix of $\boldsymbol{\beta}^{(i)}$ for $i = 1, 2$. Here we use the same prior variance τ^2 only for simplicity. The matrix $\boldsymbol{R}_i$ can be specified flexibly such as using $(\boldsymbol{F}'\boldsymbol{F})^{-1}$, or those in [20] for factorial designs.

For the parameters $\boldsymbol{\eta}$, we choose the conjugate prior derived in [4]. It takes the form

$$\begin{aligned}\boldsymbol{\eta} &\sim D(s, \boldsymbol{b}) \\ &\propto \exp\left\{\sum_{i=1}^{n} s\left(b_i \boldsymbol{f}(\boldsymbol{x}_i)'\boldsymbol{\eta} - \log(1 + e^{\boldsymbol{f}(\boldsymbol{x}_i)'\boldsymbol{\eta}})\right)\right\}, \quad (4.1)\end{aligned}$$

where $D(s, \boldsymbol{b})$ is the distribution with parameters $(s, \boldsymbol{b})$. Here s is a scalar factor and $\boldsymbol{b} \in (0, 1)^n$ is the marginal mean of $\boldsymbol{z}$ as shown in [10]. The value of $\boldsymbol{b}$ can be interpreted as a prior prediction (or guess) for $\mathbb{E}(\boldsymbol{Z})$. Based on the priors for $(\boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}, \boldsymbol{\eta})$ we can derive the posteriors as follows.

Proposition 3. *For priors $\boldsymbol{\beta}^{(i)} \sim N(\mathbf{0}, \tau^2 \boldsymbol{R}_i)$, $i = 1, 2$, and $\boldsymbol{\eta} \sim D(s, \boldsymbol{b})$, the posterior distributions of $\boldsymbol{\beta}^{(1)}$, $\boldsymbol{\beta}^{(2)}$ and $\boldsymbol{\eta}$ are independent of each other with the following forms,*

$$\boldsymbol{\beta}^{(i)}|\boldsymbol{y}, \boldsymbol{z}, \mu_0, \sigma^2, \boldsymbol{X} \sim N\left(\boldsymbol{H}_i^{-1}\boldsymbol{F}'\boldsymbol{V}_i(\boldsymbol{y} - \mu_0\mathbf{1}), \sigma^2\boldsymbol{H}_i^{-1}\right),$$

for $i = 1, 2$ and

$$\boldsymbol{\eta}|\boldsymbol{z}, \boldsymbol{X} \sim D\left(1 + s, \frac{\boldsymbol{z} + s\boldsymbol{b}}{1 + s}\right),$$

where $\boldsymbol{H}_i = \boldsymbol{F}'\boldsymbol{V}_i\boldsymbol{F} + \rho\boldsymbol{R}_i^{-1}$ with $\rho = \frac{\sigma^2}{\tau^2}$.

The proof of Proposition 3 can be derived following the standard Bayesian framework, thus is omitted. To derive the Bayesian D-optimal design criterion, we take the posterior distributions in Proposition 3 to (2.4) and set $\psi(\cdot) = \log(\cdot)$. The derivation is very similar to that in (3.3), and thus we obtain the exact design criterion as

$$\begin{aligned}\Psi(\boldsymbol{X}|\mu_0, \sigma^2) &= \mathbb{E}_{\boldsymbol{z},\boldsymbol{\eta}}\left\{\log(p(\boldsymbol{\eta}|\boldsymbol{z}, \boldsymbol{X}))\right\} \quad (4.2) \\ &+ \frac{1}{2}\sum_{i=1}^{2} \mathbb{E}_{\boldsymbol{\eta}}\mathbb{E}_{\boldsymbol{z}|\boldsymbol{\eta}}\left\{\log\det(\boldsymbol{F}'\boldsymbol{V}_i\boldsymbol{F} + \rho\boldsymbol{R}_i^{-1})\right\} + \text{constant}.\end{aligned}$$

As the integration of $\mathbb{E}_{\boldsymbol{z},\boldsymbol{\eta}}\left\{\log(p(\boldsymbol{\eta}|\boldsymbol{z}, \boldsymbol{X}))\right\}$ is not tractable, we adopt the same normal approximation of the posterior distribution $p(\boldsymbol{\eta}|\boldsymbol{z}, \boldsymbol{X})$ as in (3.4). A straightforward calculation leads to getting Fisher information matrix $\boldsymbol{I}(\boldsymbol{\eta}|\boldsymbol{X}) = (1 + s)\boldsymbol{F}'\boldsymbol{W}_0\boldsymbol{F}$. Thus we have

$$\mathbb{E}_{\boldsymbol{z},\boldsymbol{\eta}}\left\{\log(p(\boldsymbol{\eta}|\boldsymbol{z}, \boldsymbol{X}))\right\} \approx \mathbb{E}_{\boldsymbol{\eta}}\{\log\det(\boldsymbol{F}'\boldsymbol{W}_0\boldsymbol{F})\} + \text{constant}. \quad (4.3)$$

Disregarding the constant, we can approximate the exact $\Psi(\boldsymbol{X}|\mu_0, \sigma^2)$ by

$$\begin{aligned}\Psi(\boldsymbol{X}|\mu_0, \sigma^2) &\approx \mathbb{E}_{\boldsymbol{\eta}}\{\log\det(\boldsymbol{F}'\boldsymbol{W}_0\boldsymbol{F})\} \\ &+ \frac{1}{2}\sum_{i=1}^{2} \mathbb{E}_{\boldsymbol{\eta}}\mathbb{E}_{\boldsymbol{z}|\boldsymbol{\eta}}\left\{\log\det(\boldsymbol{F}'\boldsymbol{V}_i\boldsymbol{F} + \rho\boldsymbol{R}_i^{-1})\right\}.\end{aligned}$$

The following Theorem 2 gives an upper bound of the approximated criterion $\Psi(\boldsymbol{X}|\mu_0, \sigma^2)$ to avoid the integration with respect to $\boldsymbol{z}$, which plays the same role as Theorem 1.

Theorem 2. *Assume that the prior distributions of $\boldsymbol{\beta}^{(i)}$ are $\boldsymbol{\beta}^{(i)} \sim N(\mathbf{0}, \tau^2\boldsymbol{R}_i)$ for $i = 1, 2$ and $\boldsymbol{\eta}$ has either the conjugate prior $\boldsymbol{\eta} \sim D(s, \boldsymbol{b})$ or the noninformative prior $p(\boldsymbol{\eta}) \propto 1$. If $\boldsymbol{F}'\boldsymbol{W}_0\boldsymbol{F}$ is nonsingular, an upper bound of the approximated $\Psi(\boldsymbol{X}|\mu_0, \sigma^2)$ is*

$$\begin{aligned}&Q(\boldsymbol{X}) = \\ &\mathbb{E}_{\boldsymbol{\eta}}\left(\log\det(\boldsymbol{F}'\boldsymbol{W}_0\boldsymbol{F}) + \frac{1}{2}\sum_{i=1}^{2}\log\det(\boldsymbol{F}'\boldsymbol{W}_i\boldsymbol{F} + \rho\boldsymbol{R}_i^{-1})\right). \quad (4.4)\end{aligned}$$

For the same argument as in Section 3, we use the upper bound in (4.4) as the optimal design criterion. Note that since $\rho\boldsymbol{R}_i^{-1}$ is added to $\boldsymbol{F}'\boldsymbol{V}_i\boldsymbol{F}$ and $\boldsymbol{F}'\boldsymbol{W}_i\boldsymbol{F}$, $\boldsymbol{F}'\boldsymbol{V}_i\boldsymbol{F} + \rho\boldsymbol{R}_i^{-1}$ and $\boldsymbol{F}'\boldsymbol{W}_i\boldsymbol{F} + \rho\boldsymbol{R}_i^{-1}$ are nonsingular. The derivation of $\Psi(\boldsymbol{X}|\mu_0, \sigma^2)$ and $Q(\boldsymbol{X})$ only needs $\boldsymbol{F}'\boldsymbol{W}_0\boldsymbol{F}$ to be nonsingular, which requires $m \geq q$ and $\boldsymbol{\eta}$ to be finitely bounded as in Section 3.

4.2 Interpretation

Note that the criterion in (4.4) has a similar formulation with $Q(\boldsymbol{X})$ in (3.6). The only difference is that (3.6) does not involve $\rho\boldsymbol{R}_i^{-1}$. For consistency, we use the formula (4.4) as the design criterion $Q(\boldsymbol{X})$ for both cases. When noninformative priors for $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$ are used, we set $\rho = 0$. From another point of view, as $\tau^2 \to \infty$, $\rho \to 0$, the variances in the priors $p(\boldsymbol{\beta}_1)$ and $p(\boldsymbol{\beta}_2)$ diffuse and result in a noninformative priors.

The criterion $Q(\boldsymbol{X})$, consisting of three additive terms, can be interpreted intuitively. The first additive term $\mathbb{E}_{\boldsymbol{\eta}}\{\log\det(\boldsymbol{F}'\boldsymbol{W}_0\boldsymbol{F})\}$ is known as the Bayesian D-optimal criterion for logistic regression and $\mathbb{E}_{\boldsymbol{\eta}}\{\log\det(\boldsymbol{F}'\boldsymbol{W}_i\boldsymbol{F} +$

$\rho \boldsymbol{R}_i^{-1})\}$ is the Bayesian D-optimal criterion for the linear regression model of Y. To explain the weights, we rewrite $Q(\boldsymbol{X})$ as follows.

$$Q(\boldsymbol{X}) = 1 \cdot \mathbb{E}_{\boldsymbol{\eta}} \left(\log\det(\boldsymbol{F}'\boldsymbol{W}_0\boldsymbol{F})\right) + 1 \cdot \left(\frac{1}{2}\sum_{i=1}^{2} \mathbb{E}_{\boldsymbol{\eta}}(\log\det(\boldsymbol{F}'\boldsymbol{W}_i\boldsymbol{F} + \rho\boldsymbol{R}_i^{-1}))\right).$$

Since there are equal numbers of binary and continuous response observations, the design criterion should put the same weight (equal to 1) on both design criteria for Z and Y. For the two criteria for the linear regression models, the same weight $1/2$ is used. This is also reasonable because we assume $\pi_i \in (0,1)$. Then none of the diagonal entries of $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are zero, so the two terms should split the total weight 1 assigned for the entire linear regression part. Therefore, even though $Q(\boldsymbol{X})$ are derived analytically, all the additive terms and their weights make sense intuitively.

4.3 Prior Parameters

Note that the conjugate prior $p(\boldsymbol{\eta})$ requires prior parameters $(s, \boldsymbol{b})$ to be specified. Moreover, the prior distribution (4.1) contains $f(\boldsymbol{x}_i)$, which depends on the design points. When sampling $\boldsymbol{\eta}$ from the prior (4.1), it does not matter whether $f(\boldsymbol{x}_i)$'s are actually from the design points. If relevant historical data is available, we can simply sample $\boldsymbol{\eta}$ from the likelihood of the data. Alternatively, one can adopt the method in [4] to estimate the parameters $(s, \boldsymbol{b})$. Without the relevant data, we would use the noninformative prior for $\boldsymbol{\eta}$, i.e., $p(\boldsymbol{\eta}) \propto 1$ in the bounded region for $\boldsymbol{\eta}$.

The design criterion $Q(\boldsymbol{X})$ contains some unknown parameters, including the noise-to-signal ratio $\rho = \sigma^2/\tau^2$ and the correlation matrices $\boldsymbol{R}_i$'s for $i = 1, 2$. The value of ρ has to be specified either from the historical data or from the domain knowledge. Typically we would assume $\rho < 1$ such that the measurement error has a smaller variance than the signal variance.

The setting of $\boldsymbol{R}_i$ can also be specified flexibly. If historical data are available, $\boldsymbol{R}_i$ can be set as the estimated correlation matrix of $\boldsymbol{\beta}^{(i)}$. Otherwise, we can use the correlation matrix in [20] and [24], which is targeted for factorial design. Specifically, let $\boldsymbol{\beta}$ be the unknown coefficients of the linear regression model and the prior distribution is $\boldsymbol{\beta} \sim N(\boldsymbol{0}, \tau^2\boldsymbol{R})$. For 2-level factor coded in -1 and 1, [20] suggests that $\boldsymbol{R}$ is a diagonal matrix and the priors for individual β_j is

$$\begin{aligned} &\beta_0 \sim N(0, \tau^2), \\ &\beta_j \sim N(0, \tau^2 r), \qquad i = 1, \ldots, p, \\ &\beta_j \sim N(0, \tau^2 r^2), \qquad i = p+1, \ldots, p + \binom{p}{2}, \\ &\vdots \\ &\beta_{2^p-1} \sim N(0, \tau^2 r^p), \end{aligned} \tag{4.5}$$

where β_j $i = 1, \ldots, p$ are main effects, β_j $j = p+1, \ldots, p+\binom{p}{2}$ are 2-factor-interactions and up to the p-factor-interaction β_{2^p-1}. The variance of β_j decreases exponentially with the order of their corresponding effects by $r \in (0,1)$, thus it incorporates the *effects hierarchy principle* [46]. [20] showed that if $\boldsymbol{f}(\boldsymbol{x})$ contains all the 2^p effects of all p orders, $\tau^2\boldsymbol{R}$ can be represented alternatively by Kronecker product as $\tau^2\boldsymbol{R} = \varsigma^2 \bigotimes_{j=1}^{p} \boldsymbol{F}_j(x_j)^{-1}\boldsymbol{\Psi}_j(x_j)(\boldsymbol{F}_j(x_j))^{-1}$. The model matrix for the 2-level factor and the correlation matrix are

$$\boldsymbol{F}_j(x_j) = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \text{ and } \boldsymbol{\Psi}_j(x_j) = \begin{pmatrix} 1 & \zeta \\ \zeta & 1 \end{pmatrix}. \tag{4.6}$$

To keep the two different presentations equivalent, let $\zeta = \frac{1-r}{1+r}$ and $\tau^2 = (\frac{1+\zeta}{2})^p\varsigma^2$. For the mixed-level of 2- and 3-level experiments, [24] have extended the 2-level case to the format

$$\tau^2\boldsymbol{R} = \varsigma^2 \bigotimes_{j=1}^{p_2+p_{3,c}+p_{3,q}} \boldsymbol{F}_j(x_j)^{-1}\boldsymbol{\Psi}_j(x_j)(\boldsymbol{F}_j(x_j)^{-1})', \tag{4.7}$$

where p_2 is the number of 2-level factors, $p_{3,c}$ is the number of 3-level qualitative (categorical) factors, and $p_{3,q}$ is the number of 3-level quantitative factors. For all the 3-level factors, the model matrix is

$$\boldsymbol{F}_j(x_j) = \begin{pmatrix} 1 & -\sqrt{\frac{3}{2}} & \sqrt{\frac{1}{2}} \\ 1 & 0 & -\sqrt{2} \\ 1 & \sqrt{\frac{3}{2}} & \sqrt{\frac{1}{2}} \end{pmatrix}. \tag{4.8}$$

An isotropic correlation function is recommended for the 3-level qualitative factors and a Gaussian correlation function for quantitative factors. Thus, the correlation matrices for the 3-level qualitative and quantitative factors are

$$\boldsymbol{\Psi}_j(x_j) = \begin{pmatrix} 1 & \zeta & \zeta \\ \zeta & 1 & \zeta \\ \zeta & \zeta & 1 \end{pmatrix} \text{ and } \boldsymbol{\Psi}_j(x_j) = \begin{pmatrix} 1 & \zeta & \zeta^4 \\ \zeta & 1 & \zeta \\ \zeta^4 & \zeta & 1 \end{pmatrix}, \tag{4.9}$$

respectively. To keep the covariance (4.7) consistent with the 2-level case we still set $\zeta = \frac{1-r}{1+r}$. To keep the variance of the intercept equal to τ^2 [24], we set

$$\tau^2 = \varsigma^2 \left(\frac{1+\zeta}{2}\right)^{p_2} \left(\frac{1+2\zeta}{3}\right)^{p_{3,c}} \left(\frac{3+4\zeta+2\zeta^4}{9}\right)^{p_{3,q}},$$

and thus

$$\boldsymbol{R} = \left(\left(\frac{1+\zeta}{2}\right)^{p_2} \left(\frac{1+2\zeta}{3}\right)^{p_{3,c}} \left(\frac{3+4\zeta+2\zeta^4}{9}\right)^{p_{3,q}}\right)^{-1} \times \bigotimes_{j=1}^{p_2+p_{3,c}+p_{3,q}} \boldsymbol{F}_j(x_j)^{-1}\boldsymbol{\Psi}_j(x_j)(\boldsymbol{F}_j(x_j)^{-1})'.$$

It is straightforward to prove that $\boldsymbol{R}$ is a diagonal matrix if only 2-level and 3-level qualitative factors are involved, but

not so if any 3-level quantitative factors are involved, and the first diagonal entry of $\boldsymbol{R}$ is always 1.

To specify different prior distributions for $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$, we only need to use different values r_1 (or ζ_1) and r_2 (or ζ_2) to construct the prior correlation matrix. If the prior knowledge assumes that the two responses Z and Y are independent, one can set $r_1 = r_2 = r$ so that the two correlation matrices $\boldsymbol{R}_i$'s are the same, denoted as $\boldsymbol{R}$. [24] has used $r = 1/3$ (equivalently $\zeta = 1/2$) according to a meta-analysis of 113 data sets from published experiments [30]. Thus we also use $r = 1/3$ in all the examples. The readers can specify different values for r_1 and r_2 if needed.

In computation, we construct $\boldsymbol{R}$ using the Kronecker product in (4.7). But such $\boldsymbol{R}$ is for $\boldsymbol{f}(\boldsymbol{x})$ containing effects of all possible orders. Usually, we would assume the model just contains lower-order effects. So we just pick the rows and columns that correspond to the lower-order effects assumed in the model as the correlation matrix.

5. DESIGN SEARCH ALGORITHM

In this work, we focus on the construction of optimal design based on factorial design, which is suited for the prior distribution introduced in Section 4.3. For optimizing the design criterion $Q(\boldsymbol{X})$ we consider two cases. First, for fixed $\boldsymbol{\eta}$ value, we develop a point-exchange algorithm to construct a *local optimal design* that maximizes the criterion $Q(\boldsymbol{X}|\boldsymbol{\eta})$. Second, we construct a *global optimal design* based on the prior distribution of $\boldsymbol{\eta}$. Specifically, we construct the local optimal designs for different $\boldsymbol{\eta}$'s sampled from its prior distribution. Then the global optimal continuous design is obtained by accumulating the frequencies of design points selected into those local optimal designs.

5.1 Local Optimal Design for Fixed $\boldsymbol{\eta}$

For a fixed $\boldsymbol{\eta}$, we adapt the point-wise exchange algorithm to maximize the criterion

$$Q(\boldsymbol{X}|\boldsymbol{\eta}) = \log\det(\boldsymbol{F}'\boldsymbol{W}_0\boldsymbol{F}) + \frac{1}{2}\sum_{i=1}^{n}\log\det(\boldsymbol{F}'\boldsymbol{W}_i\boldsymbol{F} + \rho\boldsymbol{R}_i^{-1}).$$

The point-wise exchange algorithm is commonly used to construct D-optimal designs. It was first introduced by [16] and then widely used in many works [5, 35].

The point-wise exchange algorithm finds the optimal design from a candidate set. Here the candidate set is chosen to be the full factorial design without replicates. For now, we develop the method for 2- and 3-level factors, but it can be generalized to factors of more levels. Use previous notation that p_2, $p_{3,c}$, $p_{3,q}$ as the number of 2-level, 3-level categorical, and 3-level quantitative factors. The total number of full factorial design points is $N = 2^{p_2}3^{p_{3,c}+p_{3,q}}$, which can be large if the experiment involves many factors. To make the algorithm efficient, we filter out the candidate points that are unlikely to be the optimal design points. Following the suggestion from [12], we exclude the candidate design points whose corresponding probabilities $\pi(\boldsymbol{x}, \boldsymbol{\eta})$ is outside of $[0.15, 0.85]$. This range is used because the approximate variance of $\log\left(\frac{\pi_i}{1-\pi_i}\right)$ is nearly constant for $\pi_i \in (0.2, 0.8)$ but increases rapidly if π_i is outside that range [45]. Denote the reduced candidate set as $\boldsymbol{X}_c$ with size N'.

Next we construct the initial design of size n, such that $\boldsymbol{F}'\boldsymbol{W}_0\boldsymbol{F}$ is nonsingular, and so should be $\boldsymbol{F}'\boldsymbol{W}_i\boldsymbol{F}$ if $\rho = 0$ for $i = 1, 2$. If $N' \geq q$, we construct the initial design by reduction. Starting the initial design as $\boldsymbol{X}_c$, we remove the design points one by one until there are q points left. The remaining $n - q$ design points are then sampled from these q initial design points with probabilities proportional to the lower bounds in the sufficient condition in Proposition 1. For removing one design point, we select the one having the smallest deletion function $d(\boldsymbol{x})$ defined in (5.1). Shortcut formulas are developed in Supplement S1 for updating the inverse of the matrices $\boldsymbol{F}'\boldsymbol{W}_0\boldsymbol{F}$ and $\boldsymbol{F}'\boldsymbol{W}_i\boldsymbol{F}+\rho\boldsymbol{R}$ for $i = 1, 2$ after one design point is removed. If $N' \leq q$, we have to restore the candidate set back to the full factorial design and construct the initial design in the same reduction fashion.

To simplify the notation for $d(\boldsymbol{x})$, we define $v_i(\boldsymbol{x}_1, \boldsymbol{x}_2) = \boldsymbol{f}(\boldsymbol{x}_1)'\boldsymbol{M}_i\boldsymbol{f}(\boldsymbol{x}_2)$ and $v_i(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})'\boldsymbol{M}_i\boldsymbol{f}(\boldsymbol{x})$ for $i = 0, 1, 2$, where $\boldsymbol{M}_0 = (\boldsymbol{F}'\boldsymbol{W}_0\boldsymbol{F})^{-1}$ and $\boldsymbol{M}_i = \left(\boldsymbol{F}'\boldsymbol{W}_i\boldsymbol{F} + \rho\boldsymbol{R}_i^{-1}\right)^{-1}$ for $i = 1, 2$. Denote $\boldsymbol{X}$ as the current design and $\boldsymbol{X}_{-i}$ the design of $\boldsymbol{X}$ with the ith row removed. Then the deletion function can be derived as

$$\begin{aligned} d(\boldsymbol{x}_i) &= Q(\boldsymbol{X}|\boldsymbol{\eta}) - Q(\boldsymbol{X}_{-i}|\boldsymbol{\eta}) \\ &= -\log\left[1 - \pi(\boldsymbol{x}_i, \boldsymbol{\eta})(1 - \pi(\boldsymbol{x}_i, \boldsymbol{\eta}))v_0(\boldsymbol{x}_i)\right] \\ &\quad - \frac{1}{2}\log\left[1 - \pi(\boldsymbol{x}_i, \boldsymbol{\eta})v_1(\boldsymbol{x}_i)\right] - \frac{1}{2}\log\left[1 - (1 - \pi(\boldsymbol{x}_i, \boldsymbol{\eta}))v_2(\boldsymbol{x}_i)\right]. \end{aligned} \tag{5.1}$$

The smaller $d(\boldsymbol{x}_i)$ is, the less contribution the corresponding point makes for the overall objective $Q(\boldsymbol{X}|\boldsymbol{\eta})$.

One key of the point-wise exchange algorithm is to compute $\Delta(\boldsymbol{x}, \boldsymbol{x}_i) = Q(\boldsymbol{X}^*|\boldsymbol{\eta}) - Q(\boldsymbol{X}|\boldsymbol{\eta})$, the change in the criterion after the candidate design point $\boldsymbol{x}$ replaces $\boldsymbol{x}_i$ in the current design $\boldsymbol{X}$. Here $\boldsymbol{X}^*$ is the new design matrix after the exchange. To compute $\Delta(\boldsymbol{x}, \boldsymbol{x}_i)$ efficiently, we can obtain the following formula.

$$\begin{aligned} \Delta(\boldsymbol{x}, \boldsymbol{x}_i) &= Q(\boldsymbol{X}^*|\boldsymbol{\eta}) - Q(\boldsymbol{X}|\boldsymbol{\eta}) \\ &= \log\Delta_0(\boldsymbol{x}, \boldsymbol{x}_i) + \frac{1}{2}\sum_{i=1}^{2}\log\Delta_i(\boldsymbol{x}, \boldsymbol{x}_i), \end{aligned} \tag{5.2}$$

where $\Delta_i(\boldsymbol{x}, \boldsymbol{x}_i)$ for $i = 0, 1, 2$ are derived in Supplement S1. The matrices $\boldsymbol{M}_i$ for $i = 0, 1, 2$ need to be updated after the exchange of design points. Denote the updated matrices as $\boldsymbol{M}_i^*$ for the updated design $\boldsymbol{X}^*$. We derive the shortcut formulas to easily compute $\boldsymbol{M}_i^*$ for $i = 0, 1, 2$ as shown in Supplement S1.

Given the initial design, we can iteratively exchange the current design points with candidate design points to im-

prove the objective $Q(\boldsymbol{X}|\boldsymbol{\eta})$. The details are listed in the following Algorithm 1.

Algorithm 1 Exchange-Point Algorithm for Local D-Optimal Design.

Step 0 Generate the candidate design set from full factorial design. Filter out the points with probabilities $\pi(\boldsymbol{x}, \boldsymbol{\eta})$ outside of $[0.15, 0.85]$.

Step 1 Generate the initial design. Based on the initial design $\boldsymbol{X}$, update the matrices $\boldsymbol{F}$, $\boldsymbol{W}_i$, and $\boldsymbol{M}_i$ for $i = 0, 1, 2$. Compute the current objective value $Q(\boldsymbol{X}|\boldsymbol{\eta})$.

Step 2 Compute the deletion function $d(\boldsymbol{x}_i)$ for each $\boldsymbol{x}_i$ in $\boldsymbol{X}$. Randomly sample one design point with probability inversely proportional to $d(\boldsymbol{x}_i)$'s. Denote it as $\boldsymbol{x}_{i_0}$.

Step 3 Find $\boldsymbol{x}^*$ as the candidate point having the largest $\Delta(\boldsymbol{x}, \boldsymbol{x}_{i_0})$. If $\Delta(\boldsymbol{x}^*, \boldsymbol{x}_{i_0}) > 0$, exchange $\boldsymbol{x}^*$ with $\boldsymbol{x}_{i_0}$ in $\boldsymbol{X}$ and update the objective function value to $Q(\boldsymbol{X}|\boldsymbol{\eta}) + \Delta(\boldsymbol{x}^*, \boldsymbol{x}_{i_0})$.

Step 4 Repeat *Step 2* and *3* until the objective function has been stabilized or the maximum number of iterations is reached.

The Algorithm 1 can return different optimal designs due to different initial designs and the random sampling in Step 2. Thus, we run Algorithm 1 a few times and return the design with the best optimal value. We have several remarks regarding the algorithm. (1) The initial design generated via reduction does not have singularity issues. (2) The updated design from point-exchange does not have the singularity problem either, based on the way $\boldsymbol{x}^*$ is selected and $\boldsymbol{M}_i^*$ for $i = 0, 1, 2$ are computed. (3) To avoid being trapped in a local maximum, in Step 2 we randomly sample the design point for an exchange instead of deterministically picking the "worst" point. (4) Different from some other point-exchange algorithms, the candidate set here remains the same through Steps 1-4 since no points are deleted if they are selected in the design. It enables the resultant optimal design having replicated design points.

5.2 Global Optimal Design

Based on Algorithm 1 for local D-optimal design, we can use the following Algorithm 2 to construct global optimal design.

In Step 1 of generating $\boldsymbol{\eta}$ uniformly, we can use uniform design [14], maximin Latin hypercube design [34], or other space filling design methods [21, 31, 37] to select samples $\boldsymbol{\eta}_j$ for $j = 1, \ldots, B$. From Algorithm 2, it is likely that the discrete design obtained in Step 3 has some design points with $n_i = 1$. When experimenters prefer to have replications at every design point, they can choose a saturated design by sampling $m = q$ unique design points in Step 3. Then sample some $\boldsymbol{\eta}$ values as in Step 0. Compute the lower bounds for n_i for every $\boldsymbol{\eta}$ sample according to Proposition 1 and use the averaged lower bounds to set n_i. If $\sum_{i=1}^{m} n_i$ exceeds n, the experimenters have to either increase the experiment budget or reduce the κ value.

Algorithm 2 Algorithm for Global D-Optimal Design.

Step 0 If $p(\boldsymbol{\eta})$ is informative, simulate $\boldsymbol{\eta}_j \sim p(\boldsymbol{\eta})$ for $j = 1, \ldots, B$. Otherwise, $\boldsymbol{\eta}$ is uniformly distributed in a rectangular high-dimensional space.

Step 1 For each $\boldsymbol{\eta}_j$, call *Algorithm 1* to construct the local optimal design $\boldsymbol{X}_j$.

Step 2 For each point in the candidate set, count its frequency of being selected in the local optimal designs. The continuous optimal design is formed by the normalized frequency as a discrete distribution.

Step 3 To obtain a discrete optimal design, sample n design points from the continuous optimal design.

6. EXAMPLES

In this section, we use two examples to demonstrate the proposed Bayesian D-optimal design and the construction method. For both examples, we set $r = 1/3$ (equivalently $\zeta = 1/2$) as explained in Section 4.3. Since there are few existing works on experimental design for continuous and binary responses, we compare the proposed method with three alternative designs: the optimal designs for the quantitative-only response, the optimal design for the binary-only response, and the naively combined design method as mentioned in Example 1.

6.1 Artificial Example

In this artificial experiment, there are three 2-level factors $x_1 \sim x_3$, one 3-level categorical factor x_4, and one 3-level quantitative factor x_5. The underlying model assumed is the complete quadratic model and $\boldsymbol{f}(\boldsymbol{x})$ contains $q = 22$ model effects including the intercept and the following model effects.
First order effects:

$$x_1, x_2, x_3, x_{4,1}, x_{4,2}, x_{5,l},$$

Second order effect:

$$x_1x_2, x_1x_3, x_1x_{4,1}, x_1x_{4,2}, x_1x_{5,l}, x_2x_3, x_2x_{4,1}, x_2x_{4,2},$$
$$x_2x_{5,l}, x_3x_{4,1}, x_3x_{4,2}, x_3x_{5,l}, x_{4,1}x_{5,l}, x_{4,2}x_{5,l}, x_{5,\text{quad}}.$$

Here for the 3-level factors x_4 and x_5, the effects $x_{4,1}$ (1st comparison) and $x_{5,l}$ (linear effect) have values $\left\{-\sqrt{\frac{3}{2}}, 0, \sqrt{\frac{3}{2}}\right\}$, and $x_{4,2}$ (2nd comparison) and $x_{5,\text{quad}}$ (quadratic effect) have values $\left\{-\sqrt{\frac{1}{2}}, \sqrt{2}, \sqrt{\frac{1}{2}}\right\}$. For the 2-level factors, the effects x_i $i = 1, 2, 3$ have the same values as the design settings $\{-1, 1\}$. We consider independent uniform distribution for each η_i. Specifically, $\eta_i \sim$ Uniform$[-1, 1]$ for the intercept and the first order effects

Table 1. An example of $\boldsymbol{\eta}$ value for the local D-optimal design.

Effect	η	Effect	η	Effect	η	Effect	η
Intercept	-0.0153	x_1	-0.6067	x_2	0.7212	x_1x_2	0.0080
x_3	-0.1682	x_1x_3	0.0010	x_2x_3	0.1349	$x_{4,1}$	0.0283
$x_1x_{4,1}$	0.0594	$x_2x_{4,1}$	-0.1719	$x_3x_{4,1}$	0.1492	$x_{4,2}$	-0.1468
$x_1x_{4,2}$	0.0553	$x_2x_{4,2}$	-0.0634	$x_3x_{4,2}$	-0.2629	$x_{5,l}$	-0.0660
$x_1x_{5,l}$	-0.1054	$x_2x_{5,l}$	-0.0857	$x_3x_{5,l}$	-0.0807	$x_{4,1}x_{5,l}$	-0.1198
$x_{4,2}x_{5,l}$	-0.0292	$x_{5,\text{quad}}$	-0.1336				

and Uniform$[-0.5, 0.5]$ for the second order effects. The ranges of η_i's satisfy the effect hierarchy principle.

We set the experimental run size to be $n = 66$. Table 1 illustrates the values of a randomly chosen $\boldsymbol{\eta}$. Using Algorithm 2 with this $\boldsymbol{\eta}$, we construct the proposed local D-optimal designs for QQ model D_{QQ} for $\rho = 0$ and $\rho = 0.3$, respectively. For comparison, we also generate three alternative designs via R package *AlgDesign* developed by [41]. Specifically, they are (i) the 66-run classic D-optimal design for linear regression model, denoted as D_L, (ii) the 66-run local D-optimal design for logistic regression model given the $\boldsymbol{\eta}$, denoted as D_G, (iii) and the naively combined design of 44-run local D-optimal design for logistic regression model and 22-run D-optimal design for the linear regression model, denoted as D_C. The details of these designs can be found in Table S1 in Supplement S2.

To evaluate the performance of the proposed design in comparison with alternative designs, we consider the efficiency between two designs [44] as

$$\text{eff}(D_1, D_2|\boldsymbol{\eta}) = \exp\left\{\frac{1}{q}\left(Q(D_1|\boldsymbol{\eta}) - Q(D_2|\boldsymbol{\eta})\right)\right\}. \quad (6.1)$$

Table 2 reports the efficiency of D_{QQ} compared with D_L, D_G, and D_C, respectively. The proposed QQ optimal design D_{QQ} gains the best efficiency over the three alternative designs. It appears that the combined design D_C has the second-best design efficiency.

Table 2. Design efficiency between the proposed local design D_{QQ} and three alternative designs.

ρ	$\text{eff}(D_{QQ}, D_L\|\boldsymbol{\eta})$	$\text{eff}(D_{QQ}, D_G\|\boldsymbol{\eta})$	$\text{eff}(D_{QQ}, D_C\|\boldsymbol{\eta})$
0	1.08	1.11	1.05
0.3	1.10	1.14	1.07

Next, we focus on the comparison of D_{QQ} with D_C under different $\boldsymbol{\eta}$ values. We generate a maximin Latin hypercube design of $B = 500$ runs (R package *lhs* by [2]) for $\boldsymbol{\eta}$ with the lower and upper bounds specified earlier. For each of the $\boldsymbol{\eta}$ values, we construct a local QQ optimal design D_{QQ} and the combined design D_C. Figure 2 shows the histogram of the $\text{eff}(D_{QQ}, D_C|\boldsymbol{\eta})$ for different $\boldsymbol{\eta}$ value. All $\text{eff}(D_{QQ}, D_C|\boldsymbol{\eta})$ values are larger than 1, indicating that the local QQ optimal design outperforms the combined design.

Based on Algorithm 2, we accumulate frequencies of locally optimal designs and obtain the global D-optimal designs shown in Figure 3. Denote d_{QQ} and d_C are the proposed global optimal design for the QQ model and the global optimal combined design, respectively. The bar plots show the normalized frequencies for all the candidate points with the largest 22 frequencies colored blue. From Figure 3, for d_{QQ}, the points in the middle have much smaller frequencies than the other points. It is known that these points in the middle correspond to the points with $x_5 = 0$ in Table S1. Note that these points are only necessary for estimating the coefficient for $x_{5,\text{quad}}$, whose variances are the smallest in the prior for $\boldsymbol{\beta}^{(i)}$'s based on the effects hierarchy principle. In contrast, such a pattern is not observed for points with $x_4 = 0$. The reason is that x_4 is a categorical variable and $x_4 = -1, 0, 1$ are equally necessary to estimate the effects $x_{4,1}$ and $x_{4,2}$. For d_C, the points with the largest 22 frequencies correspond to the 22-run D-optimal design for the linear regression model, which is independent of $\boldsymbol{\eta}$ and remains the same every time. The points with $x_5 = 1$ and -1 only have slightly higher frequencies than the ones with $x_5 = 0$, due to the way we specify the prior distribution of $\boldsymbol{\eta}$.

To compare the performances of the global designs, the design efficiencies in (6.1) is used with $Q(d|\boldsymbol{\eta})$ adapted as

$$\begin{aligned} Q(d|\boldsymbol{\eta}) = & \\ & \log\det\left(n\sum_{i=1}^{N} d(\boldsymbol{x}_i)\pi(\boldsymbol{x}_i,\boldsymbol{\eta})(1-\pi(\boldsymbol{x}_i,\boldsymbol{\eta}))\boldsymbol{f}(\boldsymbol{x}_i)f(\boldsymbol{x}_i)'\right) \\ & +\frac{1}{2}\log\det\left(n\sum_{i=1}^{N} d(\boldsymbol{x}_i)\pi(\boldsymbol{x}_i,\boldsymbol{\eta})\boldsymbol{f}(\boldsymbol{x}_i)\boldsymbol{f}(\boldsymbol{x}_i)' + \rho\boldsymbol{R}\right) \\ & +\frac{1}{2}\log\det\left(n\sum_{i=1}^{N} d(\boldsymbol{x}_i)(1-\pi(\boldsymbol{x}_i,\boldsymbol{\eta}))\boldsymbol{f}(\boldsymbol{x}_i)\boldsymbol{f}(\boldsymbol{x}_i)' + \rho\boldsymbol{R}\right) \end{aligned}$$

for a global optimal design d given $\boldsymbol{\eta}$ value. Here $d(\boldsymbol{x}_i)$ is the probability frequency for candidate design point $\boldsymbol{x}_i$ specified by the design d and $\sum_{i=1}^{N} d(\boldsymbol{x}_i) = 1$. For the global optimal designs d_{QQ} and d_C obtained previously, Figure 4 shows the histograms of the $\text{eff}(d_{QQ}, d_C|\boldsymbol{\eta})$ values, where the $\boldsymbol{\eta}$ values are generated from another 100-run maximin Latin hypercube design. It is clear that d_{QQ} is universally better than d_C, thus the proposed design is more robust to different values of $\boldsymbol{\eta}$.

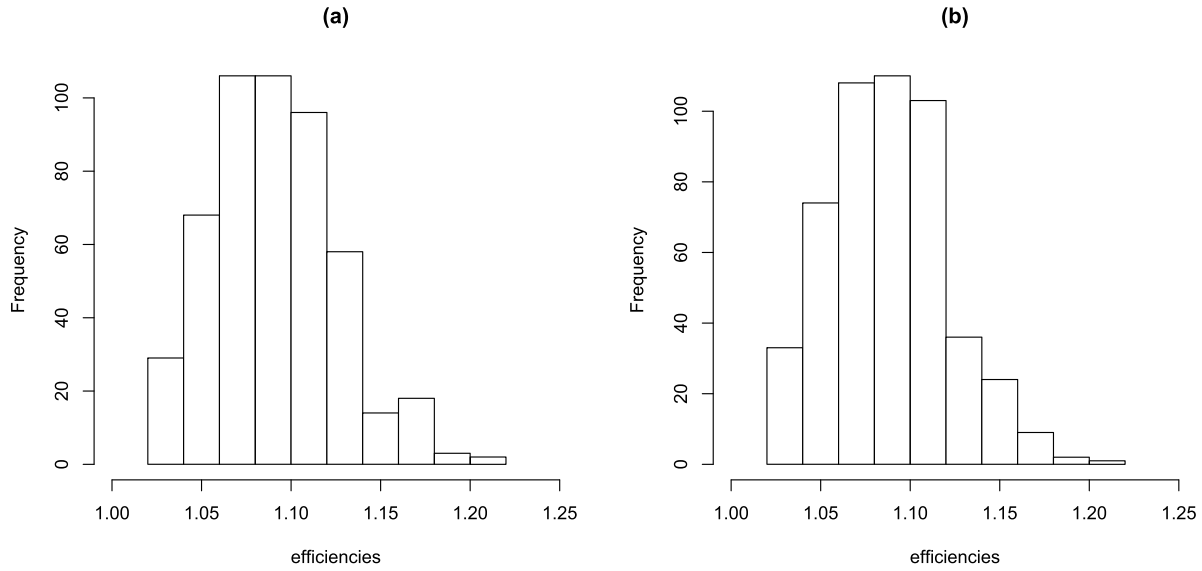

Figure 2: Artificial example: the efficiency between each local D_{QQ} and D_C under different $\boldsymbol{\eta}$ values: (a) $\rho = 0$ (b) $\rho = 0.3$.

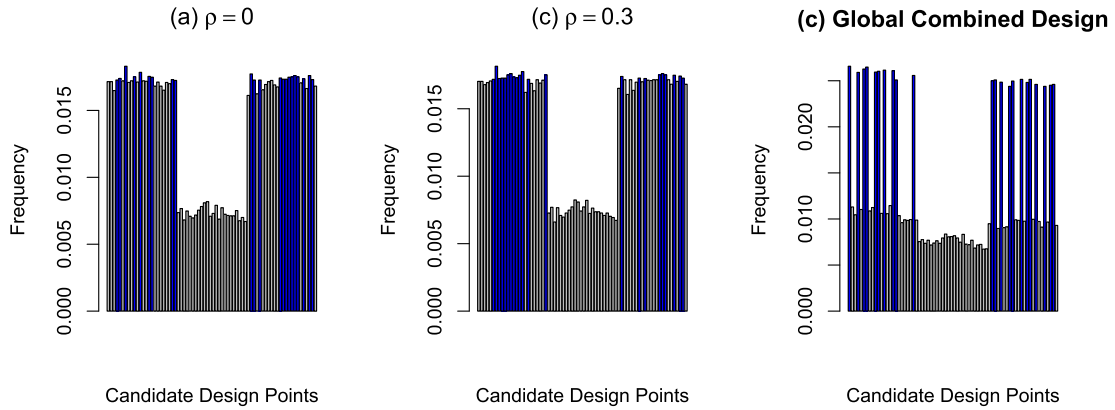

Figure 3: Artificial example: global QQ optimal designs for (a) $\rho = 0$ and (b) $\rho = 0.3$ and (c) global combined design.

6.2 Etching Experiment

In the etching process described in Section 1, the etchant is circulating at a certain flow rate. The wafers are rotated and swung horizontally and vertically. Meanwhile, the air is blown in the etchant with certain pressure. There are five factors involved in the etching process, the wafer rotation speed (x_1), the pressure for blowing the bubbles (x_2), the horizontal and vertical frequencies for swinging wafers (x_3, x_4), and the flow rate of circulating the etchant (x_5). The engineers intend to experiment to study the relationship between these factors and the two QQ responses.

Because of the newly developed process, the historical data on similar processes are not directly applicable to this experiment. Based on some exploratory analysis, we set $\rho = 0.5$. Both domain knowledge and data have shown that the wafer appearance is the worst when both the rotating speed (x_1) and bubble pressure (x_2) are low. Accordingly, we set the prior of $\boldsymbol{\eta}$ as follows. For intercept $\eta_0 \sim \text{Uniform}[0, 6]$. The linear effects of rotating speed and bubble pressure follow Uniform$[1, 5]$. The other linear effects follow Uniform$[-1, 1]$ and the second order interactions and quadratic effects Uniform$[-0.3, 0.3]$. The experimental run size is set to be $n = 21 \times 6 = 126$, 6 times the number of effects $q = 21$.

We generate a maximin Latin hypercube design of $B = 500$ runs for $\boldsymbol{\eta}$ values. For each $\boldsymbol{\eta}$ value, we obtain the local optimal designs D_{QQ} and D_C. Here the local combined design D_C has 2/3 of the runs generated from the local D-optimal design for logistic regression and 1/3 of the runs from the D-optimal design for linear regression. The efficiency between each pair of local designs D_{QQ} and D_C is reported in Figure 6(a). We can see that almost every local design D_{QQ} is better than D_C. Moreover, we obtain the global optimal designs d_{QQ} and d_C by accumulating the frequencies of the local designs. To compare d_{QQ} and d_C, we generate another 100-run maximin Latin hypercube design

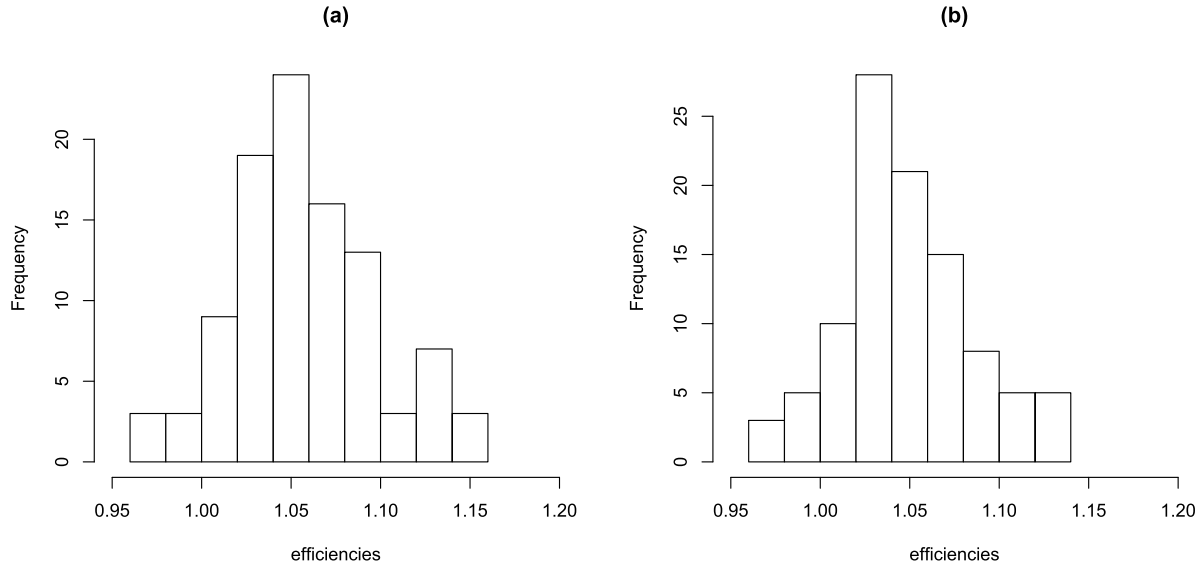

Figure 4: Artificial example: the efficiency between each global QQ optimal design and combined design under different $\boldsymbol{\eta}$ values: (a) $\rho = 0$ (b) $\rho = 0.3$.

for $\boldsymbol{\eta}$ values and compute the efficiencies between d_{QQ} and d_C under different $\boldsymbol{\eta}$ values, which are shown in Figure 6 (b). Clearly, d_{QQ} is universally better and more robust to $\boldsymbol{\eta}$ than d_C.

Fractional factorial design [46] is another commonly used design in practice. We compare the proposed design with a 3^{5-2} minimum aberration (MA) fractional factorial design by defining the contrast subgroup as $I = ABD^2 = AB^2CE^2 = AC^2DE = BCDE^2$. Each design point is replicated 5 times and the overall run is $3^{5-2} \times 5 = 135$. Figure 6(c) shows the histogram of the efficiencies between d_{QQ} and the MA design, and the proposed global optimal design is still superior.

7. DISCUSSION

In this paper, we propose the Bayesian D-optimal design criterion for QQ responses. The adoption of the Bayesian approach allows us to consider both the non-informative priors as the frequentist approach and informative priors when domain knowledge or historical data are available. A new point-exchange algorithm is developed for efficiently constructing the proposed designs. This algorithm can also be used to construct non-factorial designs when the candidate set is not a full factorial design. Moreover, the proposed method can be directly generalized for the sequential design with the QQ response. In the following, we discuss some other scenarios for the proposed method that are not considered in detail previously.

Non-conjugate Prior for $\boldsymbol{\eta}$

Other than the conjugate prior $p(\boldsymbol{\eta})$, we can also use a non-conjugate prior distribution $\boldsymbol{\eta} \sim N(\mathbf{0}, \tau_0^2 \boldsymbol{R}_0)$. In this situation, one can consider the normal approximation in the posterior distribution $p(\boldsymbol{\eta}|\boldsymbol{z})$. Then the design criterion for the binary response becomes [3] $\mathbb{E}_{\boldsymbol{z},\boldsymbol{\eta}}\{\log(p(\boldsymbol{\eta}|\boldsymbol{z})\} \approx \mathbb{E}_{\boldsymbol{\eta}}\left\{\log\det(\boldsymbol{F}'\boldsymbol{W}_0\boldsymbol{F} + \rho_0\boldsymbol{R}_0^{-1})\right\}$. The overall design criterion $Q(\boldsymbol{X})$ can be updated as

$$Q(\boldsymbol{X}) = \sum_{i=0}^{2} \mathbb{E}_{\boldsymbol{\eta}}\left\{\log\det\left(\boldsymbol{F}'\boldsymbol{W}_i\boldsymbol{F} + \tau_i\boldsymbol{R}_i^{-1}\right)\right\},$$

where $\rho_i = \sigma^2/\tau_i$, τ_i and $\boldsymbol{R}_i$ are the prior variance and correlation of $\boldsymbol{\eta}$, $\boldsymbol{\beta}^{(1)}$, and $\boldsymbol{\beta}^{(2)}$ respectively. The proposed design construction algorithm can still be applied with minor modifications.

Multiple QQ Responses

In this paper, we focus on optimal designs for one quantitative response and one qualitative response. The proposed method can also be generalized to accommodate the QQ models with multiple quantitative responses $Y_1, \ldots, Y_l$ and binary responses $Z_1, \ldots, Z_k$. For example, a multi-level qualitative response can be transformed into a set of dummy binary responses. One idea is to generalize the QQ models in (2.1) and (2.2) for both $l \geq 1$ and $k \geq 1$. For $l \geq 1$ and $k = 1$, we generalize the QQ model by introducing the correlation matrix between $Y_1, \ldots, Y_l$ as in the standard multi-response regression [1]. Then the corresponding optimal design can be established by studying its likelihood function. For $l \geq 1$ and $k > 1$ with multiple binary responses, considering all 2^k conditional models for $(Y_1, \ldots, Y_l|Z_1 = 1, \ldots, Z_k = 1), \ldots,$ $(Y_1, \ldots, Y_l|Z_1 = 0, \ldots, Z_k = 0)$ only works for a small k. Moreover, the construction algorithm can be more complicated as it needs to involve the multi-logit model [32] for modeling the multiple binary responses. When k is relatively large, we are going to pursue an alternative QQ model and

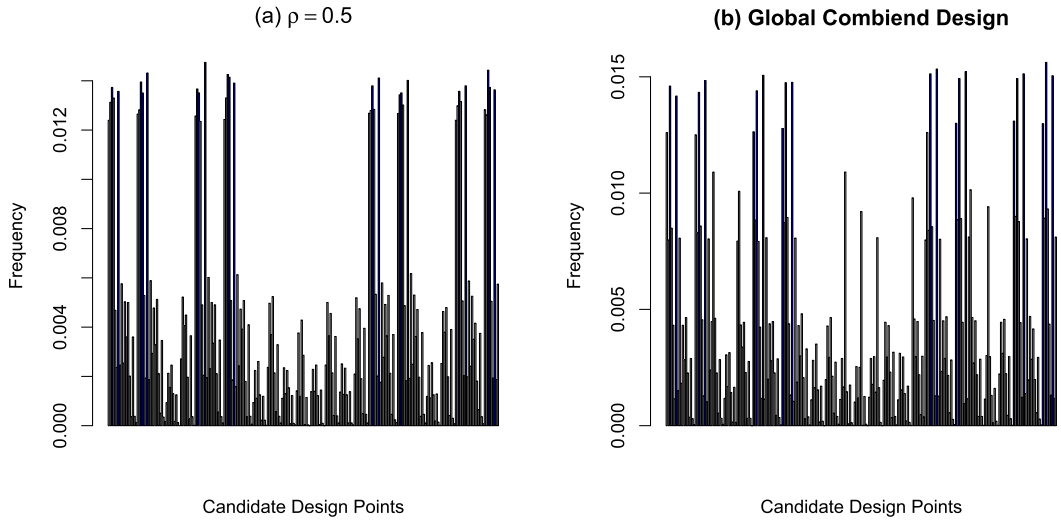

Figure 5: Etching experiment: (a) the global Bayesian QQ D-optimal design for $\rho = 0.5$ and (b) the global combined design.

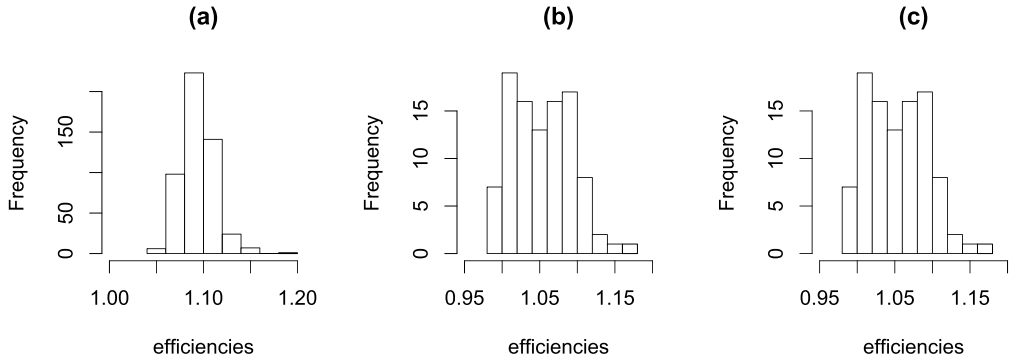

Figure 6: Etching experiment: histograms of the efficiencies (a) efficiencies between local designs D_{QQ} and D_C for different $\boldsymbol{\eta}$'s; (b) efficiencies between global designs d_{QQ} and d_C; (c) efficiencies between global design d_{QQ} and the 3^{5-2} MA fractional factorial design.

develop its corresponding optimal design method as a future research topic.

Continuous Design

The point-exchange algorithm is to construct the exact discrete optimal designs, which are different from the theoretical continuous optimal designs. As described in Sections 5 and 6, the way of generating the frequency as the local optimal design is heuristic. The rigorous definition of the local continuous D-optimal design criterion is the probability measure ξ on the design space Ω that maximizes

$$Q(\boldsymbol{X}|\boldsymbol{\eta}) = \log\det\left(\int \pi(\boldsymbol{x})(1-\pi(\boldsymbol{x}))f(\boldsymbol{x})f(\boldsymbol{x})'\mathrm{d}\xi(\boldsymbol{x})\right)$$
$$+\log\det\left(\int (\pi(\boldsymbol{x})f(\boldsymbol{x})f(\boldsymbol{x})' + \rho\boldsymbol{R}_1^{-1})\,\mathrm{d}\xi(\boldsymbol{x})\right)$$
$$+\log\det\left(\int ((1-\pi(\boldsymbol{x}))f(\boldsymbol{x})f(\boldsymbol{x})' + \rho\boldsymbol{R}_2^{-1})\,\mathrm{d}\xi(\boldsymbol{x})\right).$$

[49] developed a method to obtain the optimal ξ for the nonlinear models. It will be interesting to extend their framework and develop the method to obtain the optimal ξ for QQ models.

Different QQ Models

The proposed design is not restricted to the logit model for the binary response. For example, if the probit model is used, the Bayesian D-optimal design criterion can be directly obtained by replacing the logit transformation with the probit transformation in both $p(\boldsymbol{z}|\boldsymbol{\eta})$ and $p(\boldsymbol{\eta})$. The design criterion can be derived similarly with minor modifications. The criterion formula remains the same with the following different diagonal matrices,

$$\boldsymbol{W}_0 = \mathrm{diag}\left\{\frac{\phi^2(\boldsymbol{f}(\boldsymbol{x}_1)'\boldsymbol{\eta})}{\Phi(\boldsymbol{f}(\boldsymbol{x}_1)'\boldsymbol{\eta})\,(1-\Phi(\boldsymbol{f}(\boldsymbol{x}_1)'\boldsymbol{\eta}))}, \ldots,\right.$$
$$\left.\frac{\phi^2(\boldsymbol{f}(\boldsymbol{x}_n)'\boldsymbol{\eta})}{\Phi(\boldsymbol{f}(\boldsymbol{x}_n)'\boldsymbol{\eta})\,(1-\Phi(\boldsymbol{f}(\boldsymbol{x}_n)'\boldsymbol{\eta}))}\right\},$$

$$\boldsymbol{W}_1 = \text{diag}\left\{\Phi\left(\boldsymbol{f}(\boldsymbol{x}_1)'\boldsymbol{\eta}\right), \ldots, \Phi\left(\boldsymbol{f}(\boldsymbol{x}_n)'\boldsymbol{\eta}\right)\right\},$$
$$\boldsymbol{W}_2 = \boldsymbol{I} - \boldsymbol{W}_1,$$

where Φ and ϕ are CDF and PDF of the standard normal distribution.

It is worth pointing out that the design criterion in the work is based on the QQ model constructed by the joint model of $Y|Z$ in (2.1) and Z in (2.2). [26] created a new QQ model based on $Z|U$ where U is a latent continuous variable that is assumed to be correlated with the observed continuous response variable Y. Besides the conditional model structures, other model structures such as mixed graphical models [47] can also be used as long as the D-optimality can be derived.

SUPPLEMENTARY MATERIAL

The supplementary material contains the proofs and derivations for equations (3.3), Theorem 1, Proposition 1 and 2, Theorem 2, the shortcut formulas, update formulas, and the $\Delta(\boldsymbol{x}, \boldsymbol{x}_i)$ function in Section 5.1. The supplement material also includes the table of five different designs for the artificial example in Section 6.1. The codes and data for all the algorithms and examples are available from https://github.com/lulukang/BayesianQQDoE.git.

FUNDING

The authors were partly supported by U.S. National Science Foundation for this research project. Dr. Lulu Kang was supported by grants CMMI-1435902, DMS-1916467, and DMS-2153029, Dr. Xinwei Deng by CMMI-1233571 and CMMI-1435996, and Dr. Ran Jin by CMMI-1435996.

Accepted 17 April 2023

REFERENCES

[1] Breiman, L. and Friedman, J. H. (1997). Predicting multivariate responses in multiple linear regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **59**(1) 3–54. https://doi.org/10.1111/1467-9868.00054. MR1436554

[2] Carnell, R. (2012). lhs: Latin Hypercube Samples. R package version 0.10. http://CRAN.R-project.org/package=lhs.

[3] Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science* **10**(3) 273–304. MR1390519

[4] Chen, M. H. and Ibrahim, J. G. (2003). Conjugate priors for generalized linear models. *Statistica Sinica* **13**(2) 461–476. MR1977737

[5] Cook, R. D. and Nachtrheim, C. J. (1980). A comparison of algorithms for constructing exact D-optimal designs. *Technometrics* **22**(3) 315–324.

[6] Cox, D. R. and Wermuth, N. (1992). Response models for mixed binary and quantitative variables. *Biometrika* **79**(3) 441–461. https://doi.org/10.1093/biomet/79.3.441. MR1187603

[7] Deng, X. and Jin, R. (2015). QQ Models: Joint Modeling for Quantitative and Qualitative Quality Responses in Manufacturing Systems. *Technometrics* **57**(3) 320–331. https://doi.org/10.1080/00401706.2015.1029079. MR3384947

[8] Denman, N., McGree, J., Eccleston, J. and Duffull, S. B. (2011). Design of experiments for bivariate binary responses modelled by Copula functions. *Computational Statistics & Data Analysis* **55**(4) 1509–1520. https://doi.org/10.1016/j.csda.2010.07.025. MR2748657

[9] Dette, H., Pepelyshev, A., Zhigljavsky, A. (2013). Optimal design for linear models with correlated observations. *The Annals of Statistics* **41**(1) 143–176. https://doi.org/10.1214/12-AOS1079. MR3059413

[10] Diaconis, P. and Ylvisaker, D. (1979). Conjugate priors for exponential families. *The Annals of Statistics* **7**(2) 269–281. MR0520238

[11] Draper, N. R. and Hunter, W. G. (1966). Design of experiments for parameter estimation in multiresponse situations. *Biometrika* **53**(3-4) 525–533. https://doi.org/10.1093/biomet/53.3-4.525. MR0214240

[12] Dror, H. A. and Steinberg, D. M. (2005). Approximate Local D-optimal Experimental Design for Binary Response. Technical Report No. RP SOR-0501, Tel Aviv University.

[13] Dror, H. A. and Steinberg, D. M. (2006). Robust experimental design for multivariate generalized linear models. *Technometrics* **48**(4) 520–529. https://doi.org/10.1198/004017006000000318. MR2328621

[14] Fang, K. -T., Lin, D. K., Winker, P. and Zhang, Y. (2000). Uniform design: theory and application. *Technometrics* **42**(3) 237–248. https://doi.org/10.2307/1271079. MR1801031

[15] Fang, X. and Hedayat, A. (2008). Locally D-optimal designs based on a class of composed models resulted from blending Emax and one-compartment models. *The Annals of Statistics* **36** 428–444. https://doi.org/10.1214/009053607000000776. MR2387978

[16] Fedorov, V. V. (1972). *Theory of Optimal Experiments* **12**. Academic Press, New York. MR0403103

[17] Girard, P. and Parent, E. (2001). Bayesian analysis of autocorrelated ordered categorical data for industrial quality monitoring. *Technometrics* **43**(2) 180–191. https://doi.org/10.1198/004017001750386297. MR1954138

[18] Hainy, M., Müller, W. G. and Wynn, H. P. (2013). Approximate Bayesian computation design (ABCD), an introduction. In *mODa 10–Advances in Model-Oriented Design and Analysis* 135–143. Springer, Switzerland. https://doi.org/10.3390/e16084353. MR3255991

[19] Herzberg, A. M. and Andrews, D. F. (1976). Some considerations in the optimal design of experiments in non-optimal situations. *Journal of the Royal Statistical Society. Series B (Methodological)* **38**(3) 284–289. MR0654007

[20] Joseph, V. R. (2006). A Bayesian approach to the design and analysis of fractionated experiments. *Technometrics* **48**(2) 219–229. https://doi.org/10.1198/004017005000000652. MR2277676

[21] Joseph, V. R. and Hung, Y. (2008). Orthogonal-maximin Latin hypercube designs. *Statistica Sinica* **18**(1) 171–186. MR2416907

[22] Kang, L. (2014). Improved Coordiante-Exchange Algorithm for Constrained Optimal Design. Revision submitted to *Technometrics*.

[23] Kang, L. and Huang, X. (2019). Bayesian A-Optimal Design of Experiment with Quantitative and Qualitative Responses. *Journal of Statistical Theory and Practice* **13**(4) 64. https://doi.org/10.1007/s42519-019-0063-6. MR4021858

[24] Kang, L. and Joseph, V. R. (2009). Bayesian optimal single arrays for robust parameter design. *Technometrics* **51**(3) 250–261. https://doi.org/10.1198/tech.2009.08057. MR2751071

[25] Kang, L., Kang, X., Deng, X. and Jin, R. (2018). A Bayesian hierarchical model for quantitative and qualitative responses. *Journal of Quality Technology* **50**(3) 290–308. https://doi.org/10.1080/00224065.2018.1489042.

[26] Kang, X., Ranganathan, S., Kang, L., Gohlke, J. and Deng, X. (2021). Bayesian auxiliary variable model for birth records data with qualitative and quantitative responses. *Journal of Statistical*

Computation and Simulation **91**(16) 3283–3303. https://doi.org/10.1080/00949655.2021.1926459. MR4332587

[27] Kang, X., Kang, L., Chen, W. and Deng, X. (2022). A generative approach to modeling data with quantitative and qualitative responses. *Journal of Multivariate Analysis* **190** 104952. https://doi.org/10.1016/j.jmva.2022.104952. MR4379329

[28] Khuri, A. I., Mukherjee, B., Sinha, B. K. and Ghosh, M. (2006). Design issues for generalized linear models: A review. *Statistical Science* **21**(3) 376–399. https://doi.org/10.1214/088342306000000105. MR2339137

[29] Li, W. W. and Wu, C. F. J. (1997). Columnwise-Pairwise Algorithms with Applications to the Construction of Supersaturated Designs. *Technometrics* **39**(2) 171–179. https://doi.org/10.2307/1270905. MR1452345

[30] Li, X., Sudarsanam, N. and Frey, D. D. (2006). Regularities in data from factorial experiments. *Complexity* **11**(5) 32–45.

[31] Lin, C. D., Bingham, D., Sitter, R. R. and Tang, B. (2010). A new and flexible method for constructing designs for computer experiments. *The Annals of Statistics* **38** 1460–1477. https://doi.org/10.1214/09-AOS757. MR2662349

[32] McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society. Series B (Methodological)* **42** 109–142. MR0583347

[33] Meyer, R. K. and Nachtsheim, C. J. (1995). The Coordinate-Exchange Algorithm for Constructing Exact Optimal Experimental Designs. *Technometrics* **37**(1) 60–69. https://doi.org/10.2307/1269153. MR1322047

[34] Morris, M. D. and Mitchell, T. J. (1995). Exploratory designs for computational experiments. *Journal of Statistical Planning and Inference* **43**(3) 381–402.

[35] Nguyen, N. -K. and Miller, A. J. (1992). A review of some exchange algorithms for constructing discrete D-optimal designs. *Computational Statistics and Data Analysis* **14**(4) 489–498. https://doi.org/10.1016/0167-9473(92)90064-M. MR1192218

[36] Ning, X. and Tsung, F. (2012). A density-based statistical process control scheme for high-dimensional and mixed-type observations. *IIE transactions* **44**(4) 301–311.

[37] Qian, P. Z. (2012). Sliced Latin hypercube designs. *Journal of the American Statistical Association* **107**(497) 393–399. https://doi.org/10.1080/01621459.2011.644132. MR2949368

[38] Russell, K. G., Woods, D. C., Lewis, S. and Eccleston, J. (2009). D-optimal designs for Poisson regression models. *Statistica Sinica* **19** 721–730. MR2514184

[39] Silvey, S. (1980). *Optimal designs: an introduction to the theory for parameter estimation*. London: Chapman and Hall. MR0606742

[40] Sitter, R. R. and Torsney, B. (1995). Optimal designs for binary response experiments with two design variables. *Statistica Sinica* **5** 405–419. MR1347597

[41] Wheeler, B. (2014). AlgDesign: Algorithmic Experimental Design. R package version 1.1-7.2. http://CRAN.R-project.org/package=AlgDesign.

[42] Woods, D. C. and Van de Ven, P. (2012). Blocked designs for experiments with correlated non-normal response. *Technometrics* **53**(2) 173–182. https://doi.org/10.1198/TECH.2011.09197. MR2808343

[43] Woods, D. (2010). Robust designs for binary data: applications of simulated annealing. *Journal of Statistical Computation and Simulation* **80**(1) 29–41. https://doi.org/10.1080/00949650802445367. MR2757033

[44] Woods, D., Lewis, S., Eccleston, J. and Russell, K. (2006). Designs for generalized linear models with several variables and model uncertainty. *Technometrics* **48**(2) 284–292. https://doi.org/10.1198/004017005000000571. MR2277681

[45] Wu, C. F. J. (1985). Efficient Sequential Designs With Binary Data. *Journal of the American Statistical Association* **80**(392) 974–984. MR0819603

[46] Wu, C. J. and Hamada, M. S. (2011). *Experiments: planning, analysis, and optimization* **552**. John Wiley & Sons, New Jersey. MR2583259

[47] Yang, E., Baker, Y., Ravikumar, P., Allen, G. and Liu, Z. (2014). Mixed Graphical Models via Exponential Families. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics* 1042–1050.

[48] Yang, M. and Stufken, J. (2009). Support points of locally optimal designs for nonlinear models with two parameters. *The Annals of Statistics* **37**(1) 518–541. https://doi.org/10.1214/07-AOS560. MR2488361

[49] Yang, M., Biedermann, S. and Tang, E. (2013). On optimal designs for nonlinear models: a general and efficient algorithm. *Journal of the American Statistical Association* **108**(504) 1411–1420. https://doi.org/10.1080/01621459.2013.806268. MR3174717

Lulu Kang. Department of Applied Mathematics, 10 W 32nd Street, Chicago, IL, Illinois Institute of Technology, United States.
E-mail address: lkang2@iit.edu

Xinwei Deng. Department of Statistics, 250 Drillfield Drive, Blacksburg, VA, Virginia Polytechnic Institute and State University, United States.
E-mail address: xdeng@vt.edu

Ran Jin. Department of Statistics, 1145 Perry Street, Blacksburg, VA, Virginia Polytechnic Institute and State University, United States.
E-mail address: jran5@vt.edu