Statistica Sinica Preprint No: SS-2020-0517	
Title	Locally D-Optimal Designs for Hierarchical Response Experiments
Manuscript ID	SS-2020-0517
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202020.0517
Complete List of Authors	Mingyao Ai, Zhiqiang Ye and Jun Yu
Corresponding Author	Jun Yu
E-mail	yujunbeta@bit.edu.cn

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

LOCALLY *D*-OPTIMAL DESIGNS FOR HIERARCHICAL RESPONSE EXPERIMENTS

Mingyao Ai[1], Zhiqiang Ye[1], Jun Yu[2]

LMAM, School of Mathematical Sciences and Center for Statistical Science, Peking University[1]

School of Mathematics and Statistics, Beijing Institute of Technology[2]

Abstract: Categorical responses with a hierarchical structure are common in social sciences, public health, and marketing. The continuation ratio model is one of the most common models used to characterize such hierarchical data. Despite the wealth of research on this model, few studies have considered its design in the data collection step. Here, we study locally *D*-optimal designs for models with general link functions under the partial proportional odds assumption. The necessary and sufficient conditions for the positive definiteness of the Fisher information matrix are derived, which show that a feasible design may contain fewer supports than the number of parameters in the model. Based on some deduced characteristics of the *D*-optimal criterion, an efficient algorithm is proposed to search for optimal designs that can deal with both discrete and continuous design fields. Lastly, numerical examples illustrate the advantages of the proposed designs over some existing designs.

Key words and phrases: Approximate design; Continuation ratio model; General link functions; Multinomial response.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

1. Introduction

Categorical responses are common in scientific experiments. As such, design of experiments with categorical responses is becoming increasingly popular in many scientific disciplines. Some progress has been achieved in terms of both theory and algorithms; see, for example, Sitter and Wu (1993), Atkinson, Donev, and Tobias (2007), Yang, Zhang, and Huang (2011), Yang, Biedermann, and Tang (2013), Yang, Tong, and Mandal (2017), and Lukemire, Mandal, and Wong (2019). Among the various types of categorical responses, ordinal responses with a hierarchical structure are often used. Examples include ratings of preferences in consumer choice experiments, tumor grades in drug testing, and animal fitness in ecology; see Agresti (2006) for details.

The continuation ratio (CR) model (O'Connell, 2006) is one of the most commonly applied models, and focuses on estimating the probabilities of successive stages when the lower stages are reached first (Fullerton, 2009). Mathematically, the CR model estimates the conditional probabilities. For example, in an experiment on the emergence of house flies (Zocchi and Atkinson, 1999), seven sets of $m = 500$ pupae were exposed to one of several doses of radiation. Observations included the number of flies that died before the opening of the pupae (y_1); the number of flies out of the opened pupae but that died before completing emergence (y_2); and the

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

number of flies that emerged completely (y_3). Such a response is a typical ordinal variable. The two essential ratios, namely, the mortality of the flies inside the pupae (y_1/m) and the mortality of the flies before complete emergence to the total number of flies that started to emerge ($y_2/(m-y_1)$), can be described by the CR model.

In a CR model, the responses are assumed to be generated from a multinomial distribution. It can be regarded as a special case of a multivariate generalized linear model (Mccullagh and Nelder, 1989). Finding optimal designs on a CR model is challenging, because it involves a wide choice of link functions. Here, popular options include the logit link, probit link, log-log link, Cauchit link, and complementary log-log link. Moreover, different parameter restriction assumptions across ordinal responses need to be considered. Three kinds of assumptions are routinely used in the CR model, namely, the proportional odds assumption, non-proportional odds assumption, and partial proportional odds assumption. The proportional odds assumption (McCullagh, 1980) assumes that the effect of each predictor is invariant across ordinal responses. The non-proportional odds assumption allows different parameters for different categories (Agresti, 2006). Because a different explanatory variable structure may be required to model the different stages, the non-proportional odds assumption is more commonly used

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

in practice. The partial proportional odds assumption is proposed in Peterson and Harrell (1990), in which a subset of these explanatory variables are assumed to have proportional odds, while the rest are not. Clearly, the partial proportional odds assumption includes both the proportional odds and the non-proportional odds assumptions as special cases.

For the aforementioned reasons, relevant results in the design literature for CR models are very limited, with most being handled on a case-by-case basis. To the best of our knowledge, the most relevant works are those of Zocchi and Atkinson (1999) and Bu et al. (2020). Zocchi and Atkinson (1999) determine optimal designs for CR models with the logit link function under the non-proportional odds assumption. Bu et al. (2020) generalized optimal designs for the partial proportional odds case. However, the latter work considers only the discrete design field, which is not always applicable in practice, especially for dose-finding studies. Note that there are infinite candidate design points in the continuous design field. Therefore, the optimal designs given in Bu et al. (2020) cannot be applied directly here. When the number of grid points is not small, there is a computational bottleneck when searching for the optimal weights on the grid points over the design field. Furthermore, finding optimal designs for CR models with general link functions has not been considered in the previous works.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

To address these issues, we study D-optimal designs for CR models with general link functions under the three kinds of parameter assumptions. This study makes three main contributions to the development of optimal design problems on CR models. First, to the best of our knowledge, our method is the first to give optimal designs for CR models with general link functions on both discrete and continuous design fields. Second, similarly to Bu et al. (2020), we obtain the explicit representation of the Fisher information matrix and derive the sufficient and necessary conditions for the positive definiteness of the Fisher information matrix, which ensures the non-degeneracy of the corresponding designs. Third, the proposed algorithm can deal with multivariate responses cases. This computational aspect has rarely been considered in the literature. Comprehensive simulation results show the advantages of our methods over current designs.

The rest of this paper is organized as follows. In Section 2, we review CR models with general link functions. The Fisher information matrix is also derived. Section 3 characterizes the locally D-optimality criterion for approximate designs, and shows the sufficient and necessary conditions for the positive definiteness of the Fisher information matrix. In Section 4, an algorithm is provided to search the locally D-optimal designs under both discrete and continuous design fields. Moreover, the explicit form of the

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

optimal weights is derived for some special models. Section 5 illustrates our methods using several numerical examples. Section 6 concludes the paper. All proofs are relegated to the Supplementary Material.

2. The CR model and its Fisher information matrix

2.1 The CR model

Suppose we conduct $n_1, \ldots, n_m$ experiments under $m(m \geq 2)$ different experimental settings $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m\}$, respectively, where $\boldsymbol{x}_i \in \mathcal{X}$, for $i = 1, \ldots, m$, and $\mathcal{X}$ is the design field that we are interested in. Let the possible responses be denoted by $\{1, \ldots, J\}(J \geq 2)$, and y_{ij} be the number of responses equal to $j(j \in \{1, \ldots, J\})$ under the experimental setting $\boldsymbol{x}_i$, for $i = 1, \ldots, m$. Clearly, $y_{i1} + \cdots + y_{iJ} = n_i$. Then, $\boldsymbol{Y}_i = (y_{i1}, \ldots, y_{iJ})^T$ can be described by the following multinomial distribution:

$$p(y_{i1}, \ldots, y_{iJ}) = \frac{n_i!}{y_{i1}! \cdots y_{iJ}!} \pi_1(\boldsymbol{x}_i)^{y_{i1}} \cdots \pi_J(\boldsymbol{x}_i)^{y_{iJ}}, \tag{2.1}$$

where $\pi_j(\boldsymbol{x}_i)$ denotes the probability that the response is j under experimental setting $\boldsymbol{x}_i$. For notational simplicity, we use π_{ij} to denote $\pi_j(\boldsymbol{x}_i)$ in the rest of the paper. The relationship between the explanatory variables $\boldsymbol{x}_i$ and the probabilities $\pi_{i1}, \ldots, \pi_{iJ}$ are described through the CR model in terms of the partial proportional odds structure. Specifically, for some given

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

link function $g(\cdot)$, the conditional distribution $\boldsymbol{Y}|\boldsymbol{x}_i$ is linked to the working model through the probabilities $\pi_{i1}, \ldots, \pi_{iJ}$ by the following equation:

$$g\left(\frac{\pi_{ij}}{\pi_{ij}+\cdots+\pi_{iJ}}\right) = \boldsymbol{h}_0^T(\boldsymbol{x}_i)\boldsymbol{\beta} + \boldsymbol{h}_j^T(\boldsymbol{x}_i)\boldsymbol{\theta}_j, \tag{2.2}$$

for $i = 1, \ldots, m$ and $j = 1, \ldots, J-1$. The transformations $\boldsymbol{h}_j(\cdot)$, for $j = 0, \ldots, J-1$, are known in advance, where $\boldsymbol{h}_0(\boldsymbol{x}_i) \in \mathbb{R}^{p_0}$ stands for the predictors that are common in all categories, and $\boldsymbol{h}_j(\boldsymbol{x}_i) \in \mathbb{R}^{p_j}$, for $j = 1, \ldots, J-1$, stands for the individual predictors belonging to the jth category only. Furthermore, $\boldsymbol{\beta}$ and $\boldsymbol{\theta}_j (j = 1, \ldots, J-1)$ are unknown parameters. We further assume all $\boldsymbol{\beta}$ and $\boldsymbol{\theta}_j (j = 1, \ldots, J-1)$ lie in a compact parameter space. There are $p = p_0 + \cdots + p_{J-1}$ unknown parameters in total. Clearly, CR models under proportional odds and non-proportional odds assumptions are two special cases, with $h_j(\boldsymbol{x}_i) = 1$, for $j = 1, \ldots, J-1$, and $h_0(\boldsymbol{x}_i) = 0$, respectively. Note that Model (2.2) becomes the multinomial logit model (Agresti, 2006) when $g(\cdot)$ is the logit link function.

To ensure that the model is well defined, we need the following two regularity assumptions, mentioned in Mccullagh and Nelder (1989), throughout the remainder of this paper.

Assumption 1. For any $i = 1, \ldots, m$ and $j = 1, \ldots, J$, $0 < \pi_{ij} < 1$.

Assumption 2. The link function $g(\cdot)$ is differentiable and its derivative

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

$g'(\cdot) > 0$.

Under Assumption 2, $g(\cdot)$ is an injective function, as is $g^{-1}(\cdot)$. Thus, in Model (2.2), $\pi_{ij}/(\pi_{ij} + \cdots + \pi_{iJ})$ can be represented using $g^{-1}(\boldsymbol{h}_0^T(\boldsymbol{x}_i)\boldsymbol{\beta} + \boldsymbol{h}_j^T(\boldsymbol{x}_i)\boldsymbol{\theta}_j)$, for $i = 1, \ldots m$ and $j = 1, \ldots, J-1$. Note that $\pi_{i1}+\cdots+\pi_{iJ} = 1$, it is clear that $\pi_{i1}, \ldots, \pi_{iJ}$ are completely determined by the link function $g(\cdot)$, the predictors $\boldsymbol{h}_0(\boldsymbol{x}_i), \boldsymbol{h}_1(\boldsymbol{x}_i), \ldots, \boldsymbol{h}_{J-1}(\boldsymbol{x}_i)$, and the parameters $\boldsymbol{\beta}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{J-1}$.

2.2 Fisher information matrix

Following Kiefer (1974), we call a probability measure

$$\xi = \begin{pmatrix} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_m \\ \omega_1 & \cdots & \omega_m \end{pmatrix} \tag{2.3}$$

an approximate design, where $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \in \mathcal{X}$ and $\omega_i > 0$, for $i = 1, \ldots, m$, with $\omega_1 + \cdots + \omega_m = 1$. We further assume that the experimental region χ is compact. To obtain the Fisher information matrix, we first require some additional notation.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

For $\boldsymbol{x} \in \mathcal{X}$, $H(\boldsymbol{x})$ is a $(J-1) \times p$ matrix with

$$H(\boldsymbol{x}) = \begin{pmatrix} \boldsymbol{h}_0^T(\boldsymbol{x}) & \boldsymbol{h}_1^T(\boldsymbol{x}) & 0 & \cdots & 0 \\ \boldsymbol{h}_0^T(\boldsymbol{x}) & 0 & \boldsymbol{h}_2^T(\boldsymbol{x}) & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \boldsymbol{h}_0^T(\boldsymbol{x}) & 0 & 0 & \cdots & \boldsymbol{h}_{J-1}^T(\boldsymbol{x}) \end{pmatrix}.$$

Let $D(\boldsymbol{x}) = \text{diag}(\pi_1(\boldsymbol{x}), \ldots, \pi_J(\boldsymbol{x}))$, with $\pi_1(\boldsymbol{x}), \ldots, \pi_J(\boldsymbol{x})$ defined in Model (2.1), where $\text{diag}(\cdot)$ denotes a diagonal matrix with corresponding diagonal elements.

After defining $H(\boldsymbol{x})$ and $D(\boldsymbol{x})$, the Fisher information matrix for Model (2.2) under a design ξ is derived in the following theorem.

Theorem 1. *Suppose Assumptions 1 and 2 hold. The Fisher information matrix for Model* (2.2) *under the design* ξ *defined in* (2.3) *can be written as*

$$M(\xi) = \sum_{i=1}^{m} \omega_i H^T(\boldsymbol{x}_i) G^T(\boldsymbol{x}_i) D^{-1}(\boldsymbol{x}_i) G(\boldsymbol{x}_i) H(\boldsymbol{x}_i), \tag{2.4}$$

where $G(\boldsymbol{x}_i)$ *is a* $J \times (J-1)$ *matrix decided by the model and experimental setting* $\boldsymbol{x}_i$*, which is given in Appendix A.1.*

Example 1. Consider the following model that was used in Zocchi and Atkinson (1999):

$$\begin{aligned} \log\left(\frac{\pi_{i1}}{\pi_{i2}+\pi_{i3}}\right) &= \theta_{11} + \theta_{12} x_i + \theta_{13} x_i^2, \\ \log\left(\frac{\pi_{i2}}{\pi_{i3}}\right) &= \theta_{21} + \theta_{22} x_i, \end{aligned} \tag{2.5}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

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

Note that the above model is a special case of Model (2.2), with three ($J = 3$) categories and the logit link, under the non-proportional odds assumption. In this model, at experimental setting x_i, the matrices $H(x_i), D(x_i)$, and $G(x_i)$ in the Fisher information matrix (2.4) can be expressed as

$$H(x_i) = \begin{pmatrix} 1 & x_i & x_i^2 & 0 & 0 \\ 0 & 0 & 0 & 1 & x_i \end{pmatrix},$$

$$D(x_i) = \operatorname{diag}\left(\frac{b_{i1}}{1+b_{i1}}, \frac{b_{i2}}{(1+b_{i1})(1+b_{i2})}, \frac{1}{(1+b_{i1})(1+b_{i2})} \right),$$

$$G(x_i) = \frac{1}{(1+b_{i1})^2(1+b_{i2})^2} \begin{pmatrix} b_{i1}(1+b_{i2})^2 & 0 \\ -b_{i1}b_{i2}(1+b_{i2}) & (1+b_{i1})b_{i2} \\ -b_{i1}(1+b_{i2}) & -(1+b_{i1})b_{i2} \end{pmatrix},$$

where $b_{i1} = e^{\theta_{11}+\theta_{12}x_i+\theta_{13}x_i^2}$, and $b_{i2} = e^{\theta_{21}+\theta_{22}x_i}$, for $i = 1, \ldots, m$.

3. Determinant of the Fisher information matrix

The D-optimality criterion, which seeks to maximize the determinant of the Fisher information matrix, is one of the most popular design criteria in optimal design theory.

This criterion results in minimizing the volume of the estimator's confidence ellipsoid. To avoid trivial solutions, the Fisher information matrix is

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

required to be positive definite. The sufficient and necessary condition for the positive definiteness of the Fisher information matrix is stated in the following theorem.

Theorem 2. *The Fisher information matrix $M(\xi)$ calculated in Equation (2.4) is positive definite if and only if the matrix $(H^T(\boldsymbol{x}_1), \ldots, H^T(\boldsymbol{x}_m))$ has full row rank.*

As a direct conclusion, we obtain some equivalent conditions, which are easy to verify.

Corollary 1. *Let $H_j = (\boldsymbol{h}_j(\boldsymbol{x}_1), \ldots, \boldsymbol{h}_j(\boldsymbol{x}_m))$, for $j = 0, \ldots, J-1$. The Fisher information matrix $M(\xi)$ is positive definite if and only if the following statements hold simultaneously:*

(1) The number of support points m satisfies $m \geq \max\{p_0+r, p_1, \ldots, p_{J-1}\}$, where $r = dim\left(\cap_{j=1}^{J-1} C(H_j^T)\right)$, $C(H_j^T)$ denotes the column space of H_j^T, and $dim(\cdot)$ denotes the dimension of the corresponding space.

(2) The m support points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ satisfy the following two conditions: (a) $H_0, \ldots, H_{J-1}$ are full row rank; (b) $\cap_{j=0}^{J-1} C(H_j^T) = \{\mathbf{0}\}$.

The same conditions are also required in Theorem 3.3 for the logit link in Bu et al. (2020). From these conditions, the minimally supported designs

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

have $\max\{p_0 + r, p_1, \ldots, p_{J-1}\}$ different design points, which is different to the designs for binary responses. Note that a feasible design for a CR model with different types of links may contain fewer experimental settings than parameters. As a special case, $H_1 = \cdots = H_{J-1}$, the minimum number of support points is $p_0 + p_1$, which is strictly less than the number of parameters $p_0+(J-1)p_1$ when $J \geq 3$. Furthermore, for Model (2.2) under the proportional odds assumption, that is, $h_j(\boldsymbol{x}_i) = 1$, for $j = 1, \ldots, J-1$, the minimum number of support points is $p_0 + 1$.

Example 2. Recall Model (2.5) mentioned in Example 1. It is clear to see that $p_0 = 0, p_1 = 3$, and $p_2 = 2$. The dimension of $C(H_1^T) \cap C(H_2^T)$ is $r \leq 2$. Thus, the number of support points m satisfies $m \geq \max\{p_0+r, p_1, p_2\} = 3$. Note that when there are only three distinct design points x_1, x_2, x_3, simple algebra shows that H_0, H_1, and H_2 are full row rank, and $C(H_0^T) \cap C(H_1^T) \cap C(H_2^T) = \{\mathbf{0}\}$. Therefore, the minimum number of support points is three, according to Corollary 1.

To study the D-optimal designs and their properties, we start with a characterization of $|M(\xi)|$, where $|\cdot|$ denotes the determinant of the corresponding matrix. Theorem 3 shows that $|M(\xi)|$ is a homogeneous polynomial of $\omega_1, \ldots, \omega_m$ when the support points $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m\}$ are predefined. Before formally presenting Theorem 3, we introduce some necessary

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

notation.

Let $\delta_{\boldsymbol{x}}$ denote the single-point design and $M(\delta_{\boldsymbol{x}})$ be the corresponding Fisher information matrix. For simplicity, let τ be a map from $\{1,\ldots,p\}$ to $\{1,\ldots,m\}$ and M_τ be a $p\times p$ matrix with the kth row the same as the kth row of matrix $M(\delta_{\boldsymbol{x}_{\tau(k)}})$, for $k=1,\ldots,p$. Define

$$\Delta(\alpha_1,\ldots,\alpha_m)=\left\{\tau\left|\sum_{k=1}^{p}\mathbb{I}_{\{i\}}(\tau(k))=\alpha_i, i=1,\ldots,m\right.\right\},$$

where $\mathbb{I}_A(\cdot)$ is an indicator function defined on the set A, and $\alpha_1,\ldots,\alpha_m\in\mathbb{N}$, with $\alpha_1+\cdots+\alpha_m=p$. Then, the following theorem can be used to simplify the calculation of $|M(\xi)|$.

Theorem 3. *The determinant of the Fisher information matrix $M(\xi)$ is*

$$|M(\xi)|=\sum_{\substack{\alpha_1\geq 0,\ldots,\alpha_m\geq 0,\\ \alpha_1+\cdots+\alpha_m=p}} c_{\alpha_1,\ldots,\alpha_m}\omega_1^{\alpha_1}\cdots\omega_m^{\alpha_m},$$

where

$$c_{\alpha_1,\ldots,\alpha_m}=\sum_{\tau\in\Delta(\alpha_1,\ldots,\alpha_m)}|M_\tau|. \tag{3.1}$$

Furthermore, let $\{j_1^*,\ldots,j_k^*\}=\{j|\alpha_j>0\}$. Then, the coefficients (3.1) are zero if the $(\alpha_1,\ldots,\alpha_m)$ satisfies one of the following conditions:

(1) The maximum of $\alpha_1,\ldots,\alpha_m$ is equal to or greater than J.

(2) The number of positive elements in $\{\alpha_j\}_{j=1}^m$, say k, satisfies $k+1\leq \max\{p_0+r,p_1,\ldots,p_{J-1}\}$, where r is defined in Corollary 1.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

Theorem 3 reflects the fact that the determinants of the Fisher information matrices are homogeneous polynomials. This enables us to use the lift-one algorithm to search out the optimal weights when the support points are given in advance. This characteristic also holds for the cumulative link models. See Yang et al. (2017) for details.

Example 3. Consider Model (2.5) mentioned in Examples 1 and 2. To ensure the Fisher information matrix being positive definite, the number of support points $m \geq 3$. According to Theorem 3, $|M(\xi)|$ is an order-5 homogeneous polynomial and the coefficients of $\omega_i^5, \omega_i^4\omega_j, \omega_i^3\omega_j^2, \omega_i^3\omega_j\omega_k$ are zero. Therefore,

$$|M(\xi)| = \sum_{1\leq i<j\leq m} \sum_{\substack{1\leq k\leq m,\\ k\neq i,k\neq j}} c_{(ijk)} \cdot \omega_i^2\omega_j^2\omega_k + \sum_{i=1}^{m} \sum_{\substack{1\leq j<k<l\leq m,\\ j\neq i,k\neq i,l\neq i}} c_{(ijkl)} \cdot \omega_i^2\omega_j\omega_k\omega_l$$
$$+ \sum_{1\leq h<i<j<k<l\leq m} c_{(hijkl)} \cdot \omega_h\omega_i\omega_j\omega_k\omega_l,$$

with some coefficients $c_{(ijk)}, c_{(ijkl)}, c_{(hijkl)}$.

4. Locally D-optimal designs

As with other generalized linear models, the Fisher information matrix (2.4) depends on both the design ξ and the unknown parameters. To solve the dependence on the unknown parameters, many approaches have been proposed, such as local optimality (Chernoff, 1953), sequential procedure

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

(Ford et al., 1989), Bayesian optimality (Chaloner and Verdinelli, 1995), and maximin approach (Imhof, 2001). In this work, we focus on the locally optimal approach based on a "best guess" of the unknown parameters. Note that the locally optimal designs may be inefficient if the initial guesses of parameters are far away from the true parameters. However, it can still be a benchmark for designs chosen to satisfy experimental constraints (Ford et al., 1992; Stufken and Yang, 2012).

To search the locally D-optimal designs, we begin with the following general equivalence theorem (Atkinson, Donev, and Tobias, 2007; Atkinson, Fedorov, Herzberg, and Zhang, 2014).

Theorem 4. *For a design ξ and a point $\boldsymbol{x} \in \mathcal{X}$, define*

$$\phi(\boldsymbol{x}, \xi) = tr\left(M^{-1}(\xi)H^T(\boldsymbol{x})G^T(\boldsymbol{x})D^{-1}(\boldsymbol{x})G(\boldsymbol{x})H(\boldsymbol{x})\right) - p,$$

where $tr(\cdot)$ denotes the trace of the corresponding matrix. A design ξ^* is D-optimal if and only if the inequalitiy $\phi(\boldsymbol{x}, \xi^*) \leq 0$ holds for all $\boldsymbol{x} \in \mathcal{X}$. The equality holds if and only if $\boldsymbol{x}$ is one of the support points of ξ^*.

When the design field is discrete, we need only find the optimal weights for the finite experimental settings. The lift-one algorithm (Yang and Mandal, 2015; Yang et al., 2017; Bu et al., 2020) can be used to find the locally D-optimal designs. Details of this algorithm are given in Appendix

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

A.2. However, when the design field is continuous, we are also interested in finding possible support points. Although optimal designs found on a moderate number of grid points using the lift-one algorithm are possible, this method is often computationally infeasible when the number is large.

To solve this problem, we propose the following algorithm to find the locally D-optimal designs on the continuous design field $\mathcal{X}$.

Algorithm 1.

(i) Set $t = 0$, and initialize ξ_0 such that $|M(\xi_0)| > 0$. Let $S^{(0)}$ be the set of the support points of ξ_0, and $\varepsilon(\varepsilon > 0)$ be the predefined tolerance.

(ii) Derive the optimal weights for $S^{(t)}$ to form ξ_t using the lift-one algorithm.

(iii) Find $\boldsymbol{x}_t^* = \arg\max_{\boldsymbol{x}\in\mathcal{X}} \phi(\boldsymbol{x}, \xi_t)$, where $\phi(\boldsymbol{x}, \xi_t)$ is defined in Theorem 4.

(iv) If $\phi(\boldsymbol{x}_t^*, \xi_t) \leq \varepsilon$, then return ξ_t.

(v) Otherwise, set $S^{(t+1)} = S^{(t)} \cup \{\boldsymbol{x}_t^*\}$ and update $t = t+1$. Repeat Steps (ii)–(iv).

According to Theorem 2, the uniform-weights design with support points satisfying the conditions in Corollary 1 is a feasible initialization to start

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

the lift-one algorithm. In step (iii), the optimal solution is easy to achieve when the design field $\mathcal{X}$ is regular. For example, the Nelder-Mead method (Nelder and Mead, 1965) can be applied here. Note that our algorithm is also suitable for the discrete design case, and the maximization in step (iii) can be achieved using an exhaustive search.

The design achieved by Algorithm 1 converges to an optimal design, as stated in the following theorem.

Theorem 5. *For any initial design that satisfies* $|M(\xi_0)| > 0$*, the sequence of designs* $\{\xi_t\}$ *in Algorithm 1 converges to an optimal design that maximizes* $|M(\xi)|$ *as* $t \to \infty$*.*

We consider a CR model under the non-proportional odds assumption with $p_1 = \cdots = p_{J-1}$, and there exist p_1 distinct experimental settings $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{p_1}$ such that $H_1, \ldots, H_{J-1}$ are full row rank. The following theorem shows that the optimal weights are $1/p_1$ for all the experimental settings. This can further accelerate step (ii) in our algorithm.

Theorem 6. *If* $p_0 = 0, p_1 = \cdots = p_{J-1}$ *and there exist* p_1 *distinct experimental settings* $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{p_1}$ *such that* $H_1, \ldots, H_{J-1}$ *are full row rank, then*

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

the D-optimal design on $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{p_1}\}$ *is*

$$\xi^* = \begin{pmatrix} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_{p_1} \\ \frac{1}{p_1} & \cdots & \frac{1}{p_1} \end{pmatrix}.$$

Theorem 6 shows that for non-proportional odds models with $p_1 = \cdots = p_{J-1}$, uniform allocations are still D-optimal as the D-optimal designs studied in Corollary S.7 in the Supplementary Material of Bu et al. (2020). In general, uniform weighted designs are not always optimal. For example, consider the following CR model with two distinct experimental settings $x_1, x_2 \in \mathbb{R}$:

$$\begin{aligned} g(\pi_{i1}) &= \theta_1 + \beta x_i, \\ g\left(\frac{\pi_{i2}}{\pi_{i2} + \pi_{i3}}\right) &= \theta_2 + \beta x_i. \end{aligned} \tag{4.1}$$

The D-optimal design for Model (4.1) can be obtained by the following theorem.

Theorem 7. *For Model* (4.1)*, the D-optimal design supported on* $\{x_1, x_2\}$ *is*

$$\xi^* = \begin{pmatrix} x_1 & x_2 \\ \frac{c_1 - c_2 + \sqrt{c_1^2 - c_1 c_2 + c_2^2}}{2c_1 - c_2 + \sqrt{c_1^2 - c_1 c_2 + c_2^2}} & \frac{c_1}{2c_1 - c_2 + \sqrt{c_1^2 - c_1 c_2 + c_2^2}} \end{pmatrix},$$

where

$$c_1 = \frac{\bar{g}_{11}^2 \bar{g}_{12}^2 (\pi_{12} + \pi_{13})^2}{\pi_{11}\pi_{12}\pi_{13}} \left(\frac{\bar{g}_{21}^2}{\pi_{21}(\pi_{22} + \pi_{23})} + \frac{(\pi_{22} + \pi_{23})^2 \bar{g}_{22}^2}{\pi_{22}\pi_{23}} \right) (x_1 - x_2)^2,$$

$$c_2 = \frac{\bar{g}_{21}^2 \bar{g}_{22}^2 (\pi_{22} + \pi_{23})^2}{\pi_{21}\pi_{22}\pi_{23}} \left(\frac{\bar{g}_{11}^2}{\pi_{11}(\pi_{12} + \pi_{13})} + \frac{(\pi_{12} + \pi_{13})^2 \bar{g}_{12}^2}{\pi_{12}\pi_{13}} \right) (x_1 - x_2)^2,$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

and $\bar{g}_{ij} = (g^{-1})'(\theta_j + \beta x_i)$, *for* $i = 1, 2, j = 1, 2$.

From Theorem 7, the uniform allocation is D-optimal if and only if $c_1 = c_2$. Thus, the uniform allocations are not D-optimal, in general, for CR models under the partial proportional odds assumption.

Now, we derive the analytic solution for the following model, which generalizes the model in Example 1:

$$\begin{aligned} g(\pi_{i1}) &= \theta_{11} + \theta_{12} x_i + \theta_{13} x_i^2, \\ g\left(\frac{\pi_{i2}}{\pi_{i2} + \pi_{i3}}\right) &= \theta_{21} + \theta_{22} x_i. \end{aligned} \tag{4.2}$$

The optimal weights for the D-optimal design supported on $\{x_1, x_2, x_3\}$ are summarized in the following theorem.

Theorem 8. *For Model* (4.2)*, define*

$$c_1 = t_1 t_2 (x_1 - x_2)^2, c_2 = t_1 t_3 (x_1 - x_3)^2, c_3 = t_2 t_3 (x_2 - x_3)^2,$$

where $t_i = ((g^{-1})'(\theta_{21} + \theta_{22} x_i))^2 (\pi_{i2} + \pi_{i3})^3 (\pi_{i2} \pi_{i3})^{-1}$. *Without loss of generality, we assume* $c_1 \geq c_2 \geq c_3$. *Then, the optimal weights for the D-optimal design supported on* $\{x_1, x_2, x_3\}$ *satisfy the following statements:*

(1) If $c_1 = c_2 = c_3$, *then* $\omega_1 = \omega_2 = \omega_3 = 1/3$.

(2) If $c_1 = c_2 < c_3$, *then* $\omega_1 = \omega_2 = (-2c_1 + c_3 + \Delta_1)/\Lambda_1, \omega_3 = c_3/\Lambda_1$, *where* $\Delta_1 = \sqrt{4c_1^2 - c_1 c_3 + c_3^2}$ *and* $\Lambda_1 = -4c_1 + 3c_3 + 2\Delta_1$.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

(3) If $c_1 < c_2 = c_3$, *then* $\omega_1 = (-c_1 + 2c_3 + \Delta_2)/\Lambda_2$ *and* $\omega_2 = \omega_3 = 3c_3/\Lambda_2$, *where* $\Delta_2 = \sqrt{c_1^2 - c_1c_3 + 4c_3^2}$ *and* $\Lambda_2 = -c_1 + 8c_3 + \Delta_2$.

(4) If $c_1 < c_2 < c_3$, *then* $\omega_1 = u_1/(u_1 + u_2 + 1), \omega_2 = u_2/(u_1 + u_2 + 1)$, *and* $\omega_3 = 1/(u_1 + u_2 + 1)$, *where* u_1 *and* u_2 *satisfy*

$$\begin{aligned} c_3u_1u_2(u_1 - 2) + 2c_2u_1(u_1 - 1) &= c_1u_2(-2u_1 + 1), \\ c_3u_1u_2(u_2 - 2) + 2c_1u_2(u_2 - 1) &= c_2u_1(-2u_2 + 1). \end{aligned} \tag{4.3}$$

Theorem 8 generalizes the results for the logit link, which is studied in Bu et al. (2020). The characteristic of such a design is the same, regardless of c_1, c_2, and c_3, which are decided by the link function and the model parameters. The explicit solution of Equation (4.3) is given in the proof of Lemma S.4 of Bu et al. (2020); thus, we omit it here.

5. Simulation studies

In this section, we evaluate our method by means of several examples. CR models with five commonly used link functions, listed in Table 1, are considered throughout this section. The quality of a design ξ is evaluated by the relative D-efficiency:

$$\text{Eff}_D(\xi) = \left(\frac{|M(\xi)|}{|M(\xi^*)|}\right)^{1/p} \times 100\%,$$

where ξ^* is found using Algorithm 1 in Section 4.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

Table 1: Link functions considered in Section 5.

link name	link function	link name	link function
logit	$\log(\mu/(1-\mu))$	probit	$\Phi^{-1}(\mu)$
log-log	$-\log(-\log(\mu))$	c-log-log	$\log(-\log(1-\mu))$
Cauchit	$\tan(\pi(\mu-1/2))$		

Here, $\Phi(\cdot)$ is the cumulative distribution function of $N(0,1)$ and c-log-log is the "complementary log-log".

Two kinds of designs are compared with our method (ξ^*). The first is the uniform weighted design, in which all the given support points are assigned equal weights. The second is the design proposed by Bu et al. (2020) on the discrete design field. Note that both the uniform weighted design and the design proposed by Bu et al. (2020) are applicable only if the design field is discrete. When the design field is continuous, a common method is to consider the grid points as its discrete levels. Let ξ_U and ξ_{BMY} denote the uniform weighted design and the design proposed by Bu et al. (2020), respectively on the grid points. Note that only the optimal design for Model (2.2) with the logit link is considered in Bu et al. (2020). Hence, ξ_{BMY} is not available when the link is not the logit link, and we denote the results as "NA" in the corresponding part.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

Example 4. In an experiment on the emergence of house flies (Zocchi and Atkinson, 1999), seven sets of pupae were exposed to seven different doses of gamma radiation. After a period of time, for each set of pupae, observations included the number of flies that died before the opening of the pupae, the number of flies out of the opened pupae, but that died before completing emergence, and the number of flies that completely emerged. See Table 1 in Zocchi and Atkinson (1999) for more details.

We confirm that Model (2.5) used in Zocchi and Atkinson (1999) fits the data the best among the five link functions in terms of AIC. Therefore, we only consider the optimal design for Model (2.5).

Following Zocchi and Atkinson (1999), the design field $\mathcal{X} = [0, 200]$. We search for the locally D-optimal design ξ^* using the algorithm in Section 4, with the true parameters estimated from the original data. To show the efficiency of our design (ξ^*) against ξ_U and ξ_{BMY}, we consider five different cases, namely, 4, 6, 10, 20, and 50 grid points. The summarized efficiencies of ξ_U and ξ_{BMY} are listed in Table 2. The numbers in the first row of Table 2 indicate the numbers of grid points used in the corresponding designs.

As shown in Table 2, the uniform weighted designs are far from satisfactory for all cases. The relative D-efficiency increases as the number of grid points increases when using the grid-point search mentioned in Bu

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

Table 2: Relative D-efficiencies of our method (ξ^*) against the uniform-weights designs (ξ_U) and the designs proposed by Bu et al. (2020) (ξ_{BMY}).

		4	6	10	20	50
logit	$\text{Eff}_D(\xi_U)$	60.1%	65.8%	66.5%	66.1%	65.6%
	$\text{Eff}_D(\xi_{BMY})$	75.3%	90.3%	97.3%	99.6%	99.9%

et al. (2020) is adopted. However, the corresponding design cannot achieve satisfactory efficiency in this example when only four grid points are used. In practice, the number of grid points is difficult to determine. Therefore, our method occurs prior to the grid-point search on the continuous design field.

To evaluate the computational efficiency, we report the computing time (in seconds) for ξ^* and ξ_{BMY} with 4, 6, 10, 20, and 50 grid points. We repeat each case 1000 times, and compute the average time. All computations are carried out on a Mac Mini with an M1 chip. Table 3 shows the results, where $\xi_{BMY,k}$ indicates the design proposed by Bu et al. (2020) with k grid points. From Table 3, one can see that to achieve the optimal design (in the sense that the relative efficiency is more than 99%), ξ_{BMY} needs much more time than our method does.

Example 5. In a developmental toxicity study (Agresti, 2006), in early

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

Table 3: Average computing time (in seconds) of our method (ξ^*) and the designs proposed by Bu et al. (2020) with different grid points ($\xi_{BMY,k}$).

ξ^*	$\xi_{BMY,4}$	$\xi_{BMY,6}$	$\xi_{BMY,10}$	$\xi_{BMY,20}$	$\xi_{BMY,50}$
0.198	0.0257	0.0747	0.529	3.033	22.73

pregnancy, each mouse was exposed to one of five scores for concentration level for 10 days. Each fetus had three possible outcomes: dead, malformation, and normal. See Table 6.11 in Agresti (2006) for more details. Agresti (2006) applied Model (2.2) with the logit link under the proportional odds assumption. The model is

$$\begin{aligned} g\left(\pi_{i1}\right) &= \theta_1 + \beta x_i, \\ g\left(\frac{\pi_{i2}}{\pi_{i2}+\pi_{i3}}\right) &= \theta_2 + \beta x_i, \end{aligned} \tag{5.1}$$

where $i = 1, \ldots, m$ and $g(\cdot)$ is the logit link.

The score for the concentration level is a continuous factor ranging from 0 to 500. As in Example 4, we also compare our design with ξ_U and ξ_{BMY} under the same settings.

Actually, the Cauchit link is the best of the five link functions in terms of the AIC. Therefore, we also consider Model (5.1) with $g(\cdot)$ being the Cauchit link. The relative efficiencies of ξ_U and ξ_{BMY} are listed in Table 4. From Table 4, ξ_U and ξ_{BMY} perform similarly to Example 4, in the logit

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

link case.

Table 4: Relative D-efficiencies of our methods against the uniform-weights designs (ξ_U) and the designs proposed by Bu et al. (2020) (ξ_{BMY}).

		4	6	10	20	50
logit	$\text{Eff}_D(\xi_U)$	83.7%	84.0%	83.8%	83.3%	83.0%
	$\text{Eff}_D(\xi_{BMY})$	97.4%	99.2%	99.7%	99.9%	99.9%
Cauchit	$\text{Eff}_D(\xi_U)$	66.2%	55.5%	55.3%	53.2%	51.9%
	$\text{Eff}_D(\xi_{BMY})$	NA	NA	NA	NA	NA

Example 6. In this example, we demonstrate the case in which the parameters in the Fisher information matrix are not known in advance. Here, a locally optimal design cannot be applied directly. To overcome this problem, a Bayesian optimal design (Chaloner and Verdinelli, 1995) is adopted here. Specifically, we seek a design such that $E_{\boldsymbol{\gamma}}(\log|M(\xi,\boldsymbol{\gamma})|)$ attains its maximum, where $\boldsymbol{\gamma} = (\boldsymbol{\beta}^T, \boldsymbol{\theta}_1^T, \ldots, \boldsymbol{\theta}_{J-1}^T)^T$.

Consider the house flies experiment in Example 4. A multivariate normal prior with its mean and variance-covariance matrix estimated from the initial data set in Zocchi and Atkinson (1999) is considered in this example. This prior is also used as the prior for the EW D-optimal design in Bu et al. (2020). To avoid calculating the expectation of the Fisher

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

information matrix analytically, we use its Monte Carlo version instead. More precisely, we consider the Bayesian optimal design ξ_B that maximizes $\sum_{i=1}^{1000} \log |M(\xi, \boldsymbol{\gamma}_i)|$, where $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_{1000}$ are independent and identically generated from the prior. The relative D-efficiencies of the 1000 locally optimal designs are summarized in Figure 1, where ξ_L is the locally optimal design studied in Example 4, and $\xi_{U,k}$ indicates the uniform weighted design for k grid points. Clearly, Figure 1 illustrates that ξ_B and ξ_L are significantly more robust than the uniform-weights design.

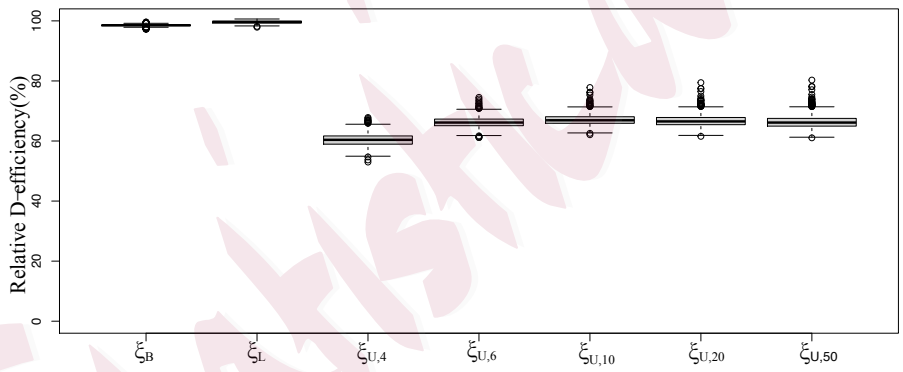

Figure 1: Summary of the relative D-efficiencies of the Bayesian optimal design (ξ_B), locally optimal design (ξ_L), and the uniform-weights designs ($\xi_{U,k}$).

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

6. Conclusion

We have address the problem of constructing locally D-optimal designs for CR models with general link functions under the partial proportional odds assumption. We derived the Fisher information matrix of the CR model to help us better understand the characteristics of D-optimal designs. We presented an algorithm to search for optimal designs on continuous design fields that is much better than the grid-point search. See Example 4 for details. Furthermore, our method can be applied to discrete design fields with a simple modification. Example 5 shows that blindly using the optimal designs under the logit link is not very efficient when other links are used to characterize the data. Several numerical results show the necessity and superiority of the proposed method over current methods.

Locally optimal designs may not be very efficient if the initial guesses are far away from the true parameters. However, they can still serve as the basis for constructing optimal designs with respect to more robust criteria, such as Bayesian optimality (Chaloner and Verdinelli, 1995). These can be regarded as possible implementations of our proposed designs.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

Supplementary Material

All technical proofs and additional simulation results are included in the online Supplementary Material.

Acknowledgments

The authors sincerely thank the editor, associate editor, and two referees for their valuable comments. Ai's work was supported by NSFC grants 12071014 and 12131001, SSFC grant 19ZDA121 and LMEQF. Yu's work was supported by NSFC grant 12001042 and the Beijing Institute of Technology research fund program for young scholars.

A. Appendix

A.1 Definition of $G(\boldsymbol{x})$

Recall $g(\cdot)$ is a general link function, $\boldsymbol{x} \in \mathcal{X}$ is a design point, and $G(\boldsymbol{x}) = (g_{jk}(\boldsymbol{x}))$ in Theorem 1 is a $J \times (J-1)$ matrix.

When $j = 1$, $g_{1k}(\boldsymbol{x})$ has the form

$$g_{1k}(\boldsymbol{x}) = \begin{cases} (g^{-1})'(\boldsymbol{h}_0^T(\boldsymbol{x})\boldsymbol{\beta} + \boldsymbol{h}_1^T(\boldsymbol{x})\boldsymbol{\theta}_1) & k = 1, \\ 0 & k = 2, \ldots, J-1. \end{cases}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

When $j = 2, \ldots, J-1$,

$$g_{jk}(\boldsymbol{x}) = \begin{cases} -g^{-1}(\boldsymbol{h}_0^T(\boldsymbol{x})\boldsymbol{\beta} + \boldsymbol{h}_j^T(\boldsymbol{x})\boldsymbol{\theta}_j)\left(\sum_{l=1}^{j-1} g_{lk}(\boldsymbol{x})\right) & k = 1, \ldots, j-1, \\ (g^{-1})'(\boldsymbol{h}_0^T(\boldsymbol{x})\boldsymbol{\beta} + \boldsymbol{h}_j^T(\boldsymbol{x})\boldsymbol{\theta}_j)\left(1 - \sum_{l=1}^{j-1} \pi_{il}\right) & k = j, \\ 0 & k = j+1, \ldots, J-1. \end{cases}$$

Finally, when $j = J$,

$$g_{Jk}(\boldsymbol{x}) = -\sum_{j=1}^{J-1} g_{jk}(\boldsymbol{x}), k = 1, \ldots, J-1.$$

Note that $G(\boldsymbol{x})$ is a lower triangular matrix, i.e., $g_{jk}(\boldsymbol{x}) = 0$ for $k > j$. By Assumption 2, $g_{jj}(\boldsymbol{x}) > 0$ for $j = 1, \ldots, J-1$, thus $G(\boldsymbol{x})$ has full column rank.

A.2 Lift-one algorithm

When the experimental settings $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m\}$ is given, lift-one algorithm gives the optimal weights $(\omega_1^*, \ldots, \omega_m^*)$ for the corresponding experiment settings. According to Theorem 3, we write the determinant of the Fisher information matrix as $|M(\xi)| = f(\omega_1, \ldots, \omega_m)$ for an order p homogeneous polynomial function f . Following Yang et al. (2016, 2017), define $S_+ = \{(\omega_1, \ldots, \omega_m)^T | \omega_i \geq 0, i = 1, \ldots, m, \sum_{i=1}^m \omega_i = 1, f(\omega_1, \ldots, \omega_m) > 0\}$, and

$$f_i(z) = f\left(\frac{1-z}{1-\omega_i}\omega_1, \ldots, \frac{1-z}{1-\omega_i}\omega_{i-1}, z, \frac{1-z}{1-\omega_i}\omega_{i+1}, \ldots, \frac{1-z}{1-\omega_i}\omega_m\right). \quad \text{(A.1)}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

with $0 \leq z \leq 1$ and $(\omega_1, \ldots, \omega_m)^T \in S_+$. Note that $f_i(z)$ in Equation (A.1) can be represented as

$$f_i(z) = \sum_{\substack{\alpha_1 \geq 0, \ldots, \alpha_m \geq 0 \\ \alpha_1 + \cdots + \alpha_m = p}} c_{\alpha_1, \ldots, \alpha_m} z^{\alpha_i} \prod_{j=1, j \neq i}^{m} \left(\frac{1-z}{1-\omega_i} \omega_j \right)^{\alpha_j}.$$

Moreover, $c_{\alpha_1, \ldots, \alpha_m} = 0$ when some $\alpha_i \geq J$, thus $f_i(z)$ can be further simplified as $f_i(z) = \sum_{k=0}^{J-1} c_{(k)} z^k (1-z)^{p-k}$, for some coefficients $c_{(0)}, \ldots, c_{(J-1)}$. The coefficients $c_{(0)}, \ldots, c_{(J-1)}$ can be calculated through solving the following J linear equations, $f_i(z_j) = \sum_{k=0}^{J-1} c_{(k)} z_j^k (1-z_j)^{p-k}$, for J distinct $z_1, \ldots, z_J \in [0, 1)$, where $f_i(z_j)$ can be directly calculated through Equation (A.1).

The lift-one algorithm is described in what follows.

Algorithm 2.

(i) Start with a uniform allocation $\boldsymbol{\omega}_0 = (1/m, \ldots, 1/m)^T$ and compute $f(\boldsymbol{\omega}_0)$.

(ii) Set up a random order of i going through $\{1, \ldots, m\}$.

(iii) For each i, determine $f_i(z)$.

(iv) Use quasi-Newton algorithm to find z^* maximizing $f_i(z)$ with $0 \leq z \leq 1$. If $f_i(z^*) \leq f_i(0)$, let $z^* = 0$. Define $\boldsymbol{\omega}_*^{(i)} = (\omega_1(1-z_*)/(1-$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

$\omega_i), \ldots, \omega_{i-1}(1 - z_*)/(1 - \omega_i), z_*, \omega_{i+1}(1 - z_*)/(1 - \omega_i), \ldots, \omega_m(1 - z_*)/(1 - \omega_i))^T$.

(v) Replace $\boldsymbol{\omega}_0$ with $\boldsymbol{\omega}_*^{(i)}$, and $f(\boldsymbol{\omega}_0)$ with $f(\boldsymbol{\omega}_*^{(i)})$.

(vi) Repeat (ii)–(v) until $f(\boldsymbol{\omega}_0) = f(\boldsymbol{\omega}_*^{(i)})$ for all i.

References

Agresti, A. (2006). *An Introduction to Categorical Data Analysis* (2nd ed.). John Wiley & Sons.

Atkinson, A., A. Donev, and R. Tobias (2007). *Optimum Experimental Designs, with SAS*, Volume 34. Oxford University Press.

Atkinson, A. C., V. V. Fedorov, A. M. Herzberg, and R. Zhang (2014). Elemental information matrices and optimal experimental design for generalized regression models. *Journal of Statistical Planning and Inference 144*, 81–91.

Bu, X., D. Majumdar, and J. Yang (2020). D-optimal designs for multinomial logistic models. *Annals of Statistics 48*(2), 983–1000.

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

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Annals of Mathematical Statistics 24*(4), 586–602.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

Ford, I., D. Titterington, and C. P. Kitsos (1989). Recent advances in nonlinear experimental design. *Technometrics 31*(1), 49–60.

Ford, I., B. Torsney, and C. Wu (1992). The use of a canonical form in the construction of locally optimal designs for non-linear problems. *Journal of the Royal Statistical Society: Series B (Methodological) 54*(2), 569–583.

Fullerton, A. S. (2009). A conceptual framework for ordered logistic regression models. *Sociological methods & research 38*(2), 306–347.

Imhof, L. A. (2001). Maximin designs for exponential growth models and heteroscedastic polynomial models. *Annals of Statistics 29*(2), 561–576.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Annals of Statistics 2*(5), 849–879.

Lukemire, J., A. Mandal, and W. K. Wong (2019). d-QPSO: A quantum-behaved particle swarm technique for finding D-optimal designs with discrete and continuous factors and a binary response. *Technometrics 61*(1), 77–87.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society: Series B (Methodological) 42*(2), 109–127.

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

Nelder, J. A. and R. Mead (1965). A simplex method for function minimization. *The Computer*

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

Journal 7(4), 308–313.

O'Connell, A. A. (2006). *Logistic regression models for ordinal response variables* (1st ed.). Sage.

Peterson, B. and F. E. Harrell (1990). Partial proportional odds models for ordinal response variables. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 39*(2), 205–217.

Sitter, R. R. and C. Wu (1993). Optimal designs for binary response experiments: Fieller, D, and A criteria. *Scandinavian Journal of Statistics 20*(4), 329–341.

Stufken, J. and M. Yang (2012). Optimal designs for generalized linear models. In K. Hinkelmann (Ed.), *Design and Analysis of Experiments, Volume 3: Special Designs and Applications*. Wiley.

Yang, J. and A. Mandal (2015). D-optimal factorial designs under generalized linear models. *Communications in Statistics-Simulation and Computation 44*(9), 2264–2277.

Yang, J., A. Mandal, and D. Majumdar (2016). Optimal designs for 2^k factorial experiments with binary response. *Statistica Sinica 26*(1), 385–411.

Yang, J., L. Tong, and A. Mandal (2017). D-optimal designs with ordered categorical data. *Statistica Sinica 27*(4), 1879–1902.

Yang, M., S. Biedermann, and E. Tang (2013). On optimal designs for nonlinear models: A general and efficient algorithm. *Journal of the American Statistical Association 108*(504),

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0517

1411–1420.

Yang, M., B. Zhang, and S. Huang (2011). Optimal designs for generalized linear models with multiple design variables. *Statistica Sinica 21*(3), 1415–1430.

Zocchi, S. S. and A. C. Atkinson (1999). Optimum experimental designs for multinomial logistic models. *Biometrics 55*(2), 437–444.

LMAM, School of Mathematical Sciences and Center for Statistical Science, Peking University, Beijing 100871, China.

E-mails: myai@math.pku.edu.cn, zqye@pku.edu.cn

School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100811, China.

E-mail: yujunbeta@bit.edu.cn