The Annals of Statistics
2020, Vol. 48, No. 2, 983–1000
https://doi.org/10.1214/19-AOS1834

D-OPTIMAL DESIGNS FOR MULTINOMIAL LOGISTIC MODELS

BY XIANWEI BU[1], DIBYEN MAJUMDAR[2,*] AND JIE YANG[2,**]

[1]*AbbVie Inc., sherry.bu@abbvie.com*
[2]*Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, [*] dibyen@uic.edu; [**] jyang06@math.uic.edu*

We consider optimal designs for general multinomial logistic models, which cover baseline-category, cumulative, adjacent-categories and continuation-ratio logit models, with proportional odds, nonproportional odds or partial proportional odds assumption. We derive the corresponding Fisher information matrices in three different forms to facilitate their calculations, determine the conditions for their positive definiteness, and search for optimal designs. We conclude that, unlike the designs for binary responses, a feasible design for a multinomial logistic model may contain less experimental settings than parameters, which is of practical significance. We also conclude that even for a minimally supported design, a uniform allocation, which is typically used in practice, is not optimal in general for a multinomial logistic model. We develop efficient algorithms for searching D-optimal designs. Using examples based on real experiments, we show that the efficiency of an experiment can be significantly improved if our designs are adopted.

1. Introduction. Experiments with categorical responses arise naturally in a rich variety of scientific disciplines (Christensen (2019)). While there is a sizable and growing literature for optimal designs with binary response, the literature for experiments with more than two categories is limited. On the other hand, there are many applications where three or more categories arise naturally. For instance, the response of an odor removal study (Yang, Tong and Mandal (2017)) has three levels, `serious odor`, `medium odor`, `almost no odor` for manufactured bio-plastics; and the outcome of trauma clinical trial (Chuang-Stein and Agresti (1997)) has five categories known as the Glasgow Outcome Scale (Jennett and Bond (1975)), `death`, `vegetative state`, `major disability`, `minor disability` and `good recovery`. In some experiments, responses are hierarchical, such as the outcome of emergence of house flies (Zocchi and Atkinson (1999)) with categories, `died before opening of the pupae`, `died before complete emergence` and `completely emerged`. Other examples include a wine bitterness study (Randall (1989)), a polysilicon deposition study (Wu (2008)) and a toxicity study (Agresti (2013)). Our aim is to obtain foundational results for the identification of optimal and efficient designs for experiments with three or more response categories and explore the properties of these designs.

When the response is binary, generalized linear models have been used widely (McCullagh and Nelder (1989), Dobson and Barnett (2018)) for analyzing the experimental data. For optimal designs of experiments with generalized linear models for *univariate* responses, there is a growing body of literature (Atkinson, Donev and Tobias (2007), Khuri et al. (2006), Stufken and Yang (2012)). In this case, the minimum number of distinct experimental settings required by a nondegenerate Fisher information matrix is equal to the number of parameters (Fedorov (1972), Yang and Mandal (2015)). Moreover, for the widely studied *D-optimal*

Received August 2018; revised February 2019.
MSC2010 subject classifications. Primary 62K05; secondary 62J12.
Key words and phrases. Approximate design, exact design, Fisher information matrix, multinomial response, minimally supported design, lift-one algorithm.

approximate designs, the experimental units are uniformly allocated when a minimally supported design, that is, a design with the minimum number of distinct experimental settings, is adopted (Yang and Mandal (2015), Yang, Mandal and Majumdar (2016)).

For responses with three or more categories, these results do not directly apply. The models used in the literature are special cases of the *multivariate* generalized linear model (McCullagh (1980)). According to the relationship among categories, categorical responses can be of three types: nominal, ordinal and hierarchical (Zocchi and Atkinson (1999)). In practice, a multinomial distribution is typically used to model the responses for a specified experimental setting. In the statistical literature, four kinds of logit models have been commonly used to link the categorical probabilities to experimental settings or values of covariates, the *baseline-category logit model* for nominal responses (Agresti (2013), Zocchi and Atkinson (1999)), the *cumulative logit model* for ordinal responses (McCullagh (1980), Christensen (2019)), the *adjacent-categories logit model* for ordinal responses (Liu and Agresti (2005), Agresti (2013)) and the *continuation-ratio logit model* for hierarchical responses (Agresti (2013), Zocchi and Atkinson (1999)). Also, three different assumptions on parameter structures have been proposed for the four logit models, *proportional odds* (*po*) assuming the same parameters for different categories (McCullagh (1980)), *nonproportional odds* (*npo*) allowing parameters to change across categories (Agresti (2013)) and *partial proportional odds* (*ppo*) containing both *po* and *npo* components (Peterson and Harrell (1990)) as special cases. The four logit models and three odds assumptions generate 12 different models for multinomial responses including, for example, a cumulative logit model with proportional odds. All 12 models can be fitted using `SAS` (Stokes, Davis and Koch (2012)) or `R` (Yee (2015)), and AIC or BIC criterion may be used for model selection.

Despite this rich complexity of models, the relevant results in the design literature for multinomial responses are limited to special classes. Zocchi and Atkinson (1999) constructed a general framework of optimal designs for multinomial logistic models with nonproportional odds. Perevozskaya, Rosenberger and Haines (2003) discussed a special class of cumulative logit models with proportional odds. Yang, Tong and Mandal (2017) obtained results for the cumulative link model, which is an extension of the cumulative logit models with proportional odds. Although the cumulative logit model with proportional odds is the most popular model in practice for ordinal responses, Agresti (2010) found strong evidence against the assumption of proportional odds for the trauma clinical trial data. For the emergence of house flies data (Zocchi and Atkinson (1999)), a continuation-ratio logit model with nonproportional odds fits better. It is also known that for certain applications, partial proportional odds may have a better performance than *po* or *npo* models (Lall et al. (2002)).

The goal of this work is a comprehensive approach to the study of optimal designs for multinomial responses. We work in a general framework, which covers all of the 12 models. The traditional starting point for obtaining an optimal design, given a model, is the Fisher information matrix for an arbitrary design. This information matrix is then "maximized" using a criterion function; for instance, the criterion we use, D-optimality, maximizes the determinant of the Fisher information matrix. Our first step is to derive explicit representations of the Fisher information matrix, thereby providing the foundation for the quest of all optimal designs. Our next step is to derive conditions for the positive definiteness of the information matrix, which is necessary for the nondegeneracy of the design and essential for formulating the criteria functions. This also allows us to determine the minimum number of experimental conditions needed for positive definiteness, which is necessary to understand the cost-efficiency of the study. Then we proceed to study D-optimal designs. The Fisher information matrix for the multinomial logistic models, as in all models except linear models, depends on the (unknown) model parameters. There are two approaches in the optimal design literature to deal with this. The first approach, local optimality, which we pursue, uses assumed values

of the parameters. Difficulties arise when no prior information, such as prior experimental result, is available; we will address this issue in Section 5.3. The second approach that requires specification of a prior distribution of the parameters is Bayesian optimality (Chaloner and Verdinelli (1995)). Using the D-criterion, these approaches lead to locally D-optimal and Bayesian D-optimal designs. Instead of Bayesian optimality, however, we study a variant, EW optimality (Atkinson, Donev and Tobias (2007), Yang, Mandal and Majumdar (2016), Yang, Tong and Mandal (2017)), which maximizes the determinant of the expected Fisher information matrix under a prior on the parameters; this is a good surrogate of Bayesian D-optimal designs, with substantially reduced computational cost.

Our work shows that the optimal designs for multinomial responses with three or more categories are remarkably different from the ones for binary responses in two major aspects: (i) the required minimum number of experimental settings is less than the number of parameters; (ii) even among minimally supported designs, unlike D-optimal designs for binary responses, uniform allocation is not D-optimal, except for *regular npo* models (defined in Section 5.6). Indeed, we find that uniform designs can be quite inefficient. Theoretically, this work reveals significant new features of optimal designs for general multinomial models. It also provides a way to accurately compute efficiency of designs of experiments based on multinomial models, and shows that widely used designs, like uniform designs, may not be efficient in these models. More generally, this work indicates that as we go from univariate to multivariate responses, some widely-applicable tenets of optimal design theory, like uniform allocation on a minimal set of experimental conditions, may no longer hold.

Similar conclusions for minimally supported designs were observed by Yang, Tong and Mandal (2017) for cumulative link models and proportional odds. Our results confirm and extend these to the general setup.

The rest of this paper is organized as follows: In Section 2, we formulate the general multinomial logistic model and its Fisher information matrix; in Section 3, we derive a necessary and sufficient condition for the Fisher information matrix to be positive definite, which determines the minimal number of required experimental settings; in Section 4, we formulate the determinant of the Fisher information matrix as a homogeneous polynomial of allocations of experimental units and simplify its structure for D-optimality; in Section 5, we develop numerical algorithms for searching D-optimal designs, as well as some analytical results for minimally supported designs; we conclude with discussions in Section 6. Although we focus on D-optimality, our basic results on Fisher information matrix are useful for other criteria as well, such as A-optimality, E-optimality, etc. We also study two examples extensively, the trauma clinical trial (Chuang-Stein and Agresti (1997)) and the emergence of house flies (Zocchi and Atkinson (1999)). These are used to illustrate the results and explore the properties of the designs, such as efficiency and robustness.

One important point to note is that throughout this paper the major results are expressed in terms of *ppo* models only, as this includes both *po* and *npo* models as special cases. The special cases are of considerable interest, however, especially in many applications, and all of the specific formulae for *po* and *npo* models are provided in the Supplementary Material (Bu, Majumdar and Yang (2020)), so they are readily available to the reader. The Supplementary Material also contain results that are of further interest.

2. Multinomial logistic model and its Fisher information matrix. This section is dedicated to the description of the model and the Fisher information matrix, the starting points of the study of optimal designs. We consider an experiment with $d \geq 1$ factors and $m \geq 2$ distinct experimental settings $\mathbf{x}_i = (x_{i1}, \ldots, x_{id})^T$, $i = 1, \ldots, m$ that could be used in the experiment. For the ith experimental setting, $n_i \geq 0$ categorical responses are collected i.i.d. from a discrete distribution with $J \geq 2$ categories, with $n_i = 0$ indicating no experimental unit assigned to this experimental setting. When $n_i > 0$, the responses

associated with the ith experimental setting are summarized into a multinomial response $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iJ})^T \sim \text{Multinomial}(n_i; \pi_{i1}, \ldots, \pi_{iJ})$, where π_{ij} is the probability that the response falls into the jth category at the ith experimental setting. Note that π_{ij}'s are functions of the experimental settings and the model parameters (to be introduced next); however, for simplicity we do not show the dependence in the notation of π_{ij}. Throughout this paper, we assume $\pi_{ij} > 0$ for all $i = 1, \ldots, m$ and $j = 1, \ldots, J$, which is necessary for a multinomial logistic model (in Section 5.1 we examine the implications of this on the *design space*).

The general structure of models that are considered is a linear regression of log odds on two components: one specific to the category and the other common to all categories. As mentioned in the Introduction, we will work with the most general model—the partial proportional odds (*ppo*) model. For this, we write the four logit models (baseline-category, cumulative, adjacent-categories and continuation-ratio) in terms of *ppo* structure as follows:

$$\log\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) = \mathbf{h}_j^T(\mathbf{x}_i)\boldsymbol{\beta}_j + \mathbf{h}_c^T(\mathbf{x}_i)\boldsymbol{\zeta}, \quad \text{baseline-category},$$

$$\log\left(\frac{\pi_{i1} + \cdots + \pi_{ij}}{\pi_{i,j+1} + \cdots + \pi_{iJ}}\right) = \mathbf{h}_j^T(\mathbf{x}_i)\boldsymbol{\beta}_j + \mathbf{h}_c^T(\mathbf{x}_i)\boldsymbol{\zeta}, \quad \text{cumulative},$$

$$\log\left(\frac{\pi_{ij}}{\pi_{i,j+1}}\right) = \mathbf{h}_j^T(\mathbf{x}_i)\boldsymbol{\beta}_j + \mathbf{h}_c^T(\mathbf{x}_i)\boldsymbol{\zeta}, \quad \text{adjacent-categories},$$

$$\log\left(\frac{\pi_{ij}}{\pi_{i,j+1} + \cdots + \pi_{iJ}}\right) = \mathbf{h}_j^T(\mathbf{x}_i)\boldsymbol{\beta}_j + \mathbf{h}_c^T(\mathbf{x}_i)\boldsymbol{\zeta}, \quad \text{continuation-ratio},$$

where $i = 1, \ldots, m$, $j = 1, \ldots, J-1$, $\mathbf{h}_j^T(\cdot) = (h_{j1}(\cdot), \ldots, h_{jp_j}(\cdot))$ are known functions to determine the p_j predictors associated with the p_j unknown parameters $\boldsymbol{\beta}_j = (\beta_{j1}, \ldots, \beta_{jp_j})^T$ for the jth response category, and $\mathbf{h}_c^T(\cdot) = (h_1(\cdot), \ldots, h_{p_c}(\cdot))$ are known functions to determine the p_c predictors associated with the p_c unknown parameters $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_{p_c})^T$ that are common for all categories. As special cases of *ppo*, $\mathbf{h}_j^T(\mathbf{x}_i) \equiv 1$ leads to *po* models, and $\mathbf{h}_c^T(\mathbf{x}_i) \equiv 0$ leads to *npo* models. The corresponding expressions for *po* and *npo* models are provided in the Supplementary Material (Sections S.7 and S.8).

Following Glonek and McCullagh (1995) and Zocchi and Atkinson (1999), we rewrite these four logit models into a unified form

$$\mathbf{C}^T \log(\mathbf{L}\boldsymbol{\pi}_i) = \boldsymbol{\eta}_i = \mathbf{X}_i\boldsymbol{\theta}, \quad i = 1, \ldots, m, \tag{1}$$

where $\boldsymbol{\pi}_i = (\pi_{i1}, \ldots, \pi_{iJ})^T$ satisfying $\sum_{j=1}^J \pi_{ij} = 1$, $\boldsymbol{\eta}_i = (\eta_{i1}, \ldots, \eta_{iJ})^T$,

$$\mathbf{C}^T = \begin{pmatrix} \mathbf{I}_{J-1} & -\mathbf{I}_{J-1} & \mathbf{0}_{J-1} \\ \mathbf{0}_{J-1}^T & \mathbf{0}_{J-1}^T & 1 \end{pmatrix}$$

is a $J \times (2J-1)$ constant matrix, where $\mathbf{I}_k$ is the identity matrix of order k and $\mathbf{0}_k$ is a vector of k zeros, and $\mathbf{L}$ is a $(2J-1) \times J$ constant matrix taking different forms across the four logit models (see Appendix A.1 for details). The model matrix is

$$\mathbf{X}_i = \begin{pmatrix} \mathbf{h}_1^T(\mathbf{x}_i) & & & \mathbf{h}_c^T(\mathbf{x}_i) \\ & \ddots & & \vdots \\ & & \mathbf{h}_{J-1}^T(\mathbf{x}_i) & \mathbf{h}_c^T(\mathbf{x}_i) \\ \mathbf{0}_{p_1}^T & \cdots & \mathbf{0}_{p_{J-1}}^T & \mathbf{0}_{p_c}^T \end{pmatrix}_{J \times p} \tag{2}$$

and the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_{J-1}^T, \boldsymbol{\zeta}^T)^T$ consists of $p = p_1 + \cdots + p_{J-1} + p_c$ unknown parameters in total. Note that $\pi_{i1} + \cdots + \pi_{iJ} = 1$ implies that $\eta_{iJ} = 0$, and thus the last row of $\mathbf{X}_i$ is all 0's. We keep η_{iJ} and the last row of $\mathbf{X}_i$ in (1) for convenience following

Glonek and McCullagh (1995). Model (1) covers all four logit models and all three odds structures (*po*, *npo* and *ppo*).

EXAMPLE 2.1. Yee (2015) considered a *ppo* model with $d = 4$ factors, $J = 3$ response categories, and parameter vector $\boldsymbol{\theta} = (\beta_{11}, \beta_{12}, \beta_{13}, \beta_{21}, \beta_{22}, \beta_{23}, \zeta_1, \zeta_2)^T$. At the ith experimental setting $\mathbf{x}_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4})^T$,

$$\mathbf{X}_i = \begin{pmatrix} 1 & x_{i1} & x_{i2} & 0 & 0 & 0 & x_{i3} & x_{i4} \\ 0 & 0 & 0 & 1 & x_{i1} & x_{i2} & x_{i3} & x_{i4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

We will revisit this example in Section 4.

Using matrix differentiation formulae (see, e.g., Seber (2008), Chapter 17), we obtain the Fisher information matrix for model (1) as follows.

THEOREM 2.1. *Consider the multinomial logistic model* (1) *with independent observations. The Fisher information matrix*

$$\mathbf{F} = \sum_{i=1}^{m} n_i \mathbf{F}_i, \tag{3}$$

where

$$\mathbf{F}_i = \left(\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T}\right)^T \operatorname{diag}(\boldsymbol{\pi}_i)^{-1} \frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T} \tag{4}$$

with $\partial \boldsymbol{\pi}_i / \partial \boldsymbol{\theta}^T = (\mathbf{C}^T \mathbf{D}_i^{-1} \mathbf{L})^{-1} \mathbf{X}_i$ *and* $\mathbf{D}_i = \operatorname{diag}(\mathbf{L}\boldsymbol{\pi}_i)$.

Theorem 2.1 is a special case of Glonek and McCullagh (1995) who built a more general framework for multiple categorical responses. We provide independent proofs in the Supplementary Material (Section S.15), as well as a result, Lemma S.5, for use later on. Our results apply to more general models than Zocchi and Atkinson (1999).

REMARK 2.1. The Fisher information matrix $\mathbf{F}$ plays a key role in optimal design theory. For example, a D-optimal design maximizes the determinant of $\mathbf{F}$, an A-optimal design minimizes the trace of $\mathbf{F}^{-1}$ and an E-optimal design maximizes the minimum eigenvalue of $\mathbf{F}$. Given experimental settings $\mathbf{x}_1, \ldots, \mathbf{x}_m$ and the parameter vector $\boldsymbol{\theta}$, one can calculate $\mathbf{F}_1, \ldots, \mathbf{F}_m$ using (4). Then Theorem 2.1 provides a convenient way for calculating $\mathbf{F}$ as a function of the allocation $(n_1, \ldots, n_m)$. The nonsingularity of the key matrix $\mathbf{C}^T \mathbf{D}_i^{-1} \mathbf{L}$ was guaranteed by Glonek and McCullagh ((1995), Theorem 1). To facilitate calculations, we derive explicit forms of $(\mathbf{C}^T \mathbf{D}_i^{-1} \mathbf{L})^{-1}$ for all of the four logit models in the Supplementary Material (Section S.3).

3. Positive definiteness of the Fisher information matrix. Positive definiteness of the information matrix is essentially a necessary condition for the existence of unbiased estimators of parameters with finite variance (Stoica and Marzetta (2001)). In this section, our focus is the determination of conditions for the positive definiteness. We also determine the minimal number of experimental settings required for positive definiteness, which is a basic question, since it deals with the cost of experimentation. In order to do that, we first derive a more amenable representation of $\mathbf{F}$.

To study the structure of $\mathbf{F}$, we first denote $(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})^{-1} = (\mathbf{c}_{i1}, \ldots, \mathbf{c}_{iJ})$, with $\mathbf{c}_{ij}$ denoting the $J \times 1$ column j. For the ith experimental setting, we define $u_{st}(\boldsymbol{\pi}_i) = \mathbf{c}_{is}^T \operatorname{diag}(\boldsymbol{\pi}_i)^{-1}\mathbf{c}_{it}$, for $s, t = 1, \ldots, J-1$ and then a $J \times J$ matrix

$$\mathbf{U}_i = \begin{pmatrix} u_{11}(\boldsymbol{\pi}_i) & \cdots & u_{1,J-1}(\boldsymbol{\pi}_i) & 0 \\ \vdots & \ddots & \vdots & \vdots \\ u_{J-1,1}(\boldsymbol{\pi}_i) & \cdots & u_{J-1,J-1}(\boldsymbol{\pi}_i) & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}. \tag{5}$$

Then we can rewrite the Fisher information at the ith experimental setting $\mathbf{F}_i$ into a simpler form as a corollary of Theorem 2.1.

COROLLARY 3.1. *Under the setup of Theorem 2.1,* $\mathbf{F}_i = \mathbf{X}_i^T \mathbf{U}_i \mathbf{X}_i$.

In order to reformulate the Fisher information matrix $\mathbf{F}$ into a form that facilitates the discussion of the positive definiteness of $\mathbf{F}$, we further define an $m(J-1) \times m(J-1)$ matrix $\mathbf{U} = (\mathbf{U}_{st})_{s,t=1,\ldots,J-1}$ with $\mathbf{U}_{st} = \operatorname{diag}\{n_1 u_{st}(\boldsymbol{\pi}_1), \ldots, n_m u_{st}(\boldsymbol{\pi}_m)\}$, and a $p \times m(J-1)$ matrix

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_1 & & \\ & \ddots & \\ & & \mathbf{H}_{J-1} \\ \mathbf{H}_c & \cdots & \mathbf{H}_c \end{pmatrix}, \tag{6}$$

where $\mathbf{H}_j = (\mathbf{h}_j(\mathbf{x}_1), \ldots, \mathbf{h}_j(\mathbf{x}_m))$ and $\mathbf{H}_c = (\mathbf{h}_c(\mathbf{x}_1), \ldots, \mathbf{h}_c(\mathbf{x}_m))$.

With the aid of Corollary 3.1, we obtain the theorem below, whose proof is relegated to the Supplementary Material (Section S.15).

THEOREM 3.1. *Consider the multinomial logistic model* (1) *with independent observations. The Fisher information matrix* $\mathbf{F} = \mathbf{H}\mathbf{U}\mathbf{H}^T$.

It can be verified that the matrix $\mathbf{U}$ is positive definite if $n_i > 0$ for all $i = 1, \ldots, m$ (see Section S.4). For general cases, if $n_i = 0$ for some i, one can denote $k = \#\{i : n_i > 0\}$ and $\mathbf{U}_{st}^* = \operatorname{diag}\{n_i u_{st}(\boldsymbol{\pi}_i) : n_i > 0\}$, then the reduced $k(J-1) \times k(J-1)$ matrix $\mathbf{U}^* = (\mathbf{U}_{st}^*)_{s,t=1,\ldots,J-1}$ is still positive definite. In this case, one can remove all columns of $\mathbf{H}$ associated with $n_i = 0$ and denote the leftover as $\mathbf{H}^*$, which is a $p \times k(J-1)$ matrix.

THEOREM 3.2. *The Fisher information matrix* $\mathbf{F}$ *is positive definite if and only if* $\mathbf{H}^*$ *is of full row rank. Furthermore, if* $n_i > 0$ *for all* $i = 1, \ldots, m$, *then* $\mathbf{F}$ *is positive definite if and only if* $\mathbf{H}$ *is of full row rank.*

Detailed discussion on the row rank of the matrix $\mathbf{H}$ is relegated to the Supplementary Material (Section S.5). As a direct conclusion, we obtain the main result for the positive definiteness of the Fisher information matrix $\mathbf{F}$.

THEOREM 3.3. *Consider the multinomial logistic model* (1) *with* m *distinct experimental settings* $\mathbf{x}_i$ *and* $n_i > 0$ *experimental units,* $i = 1, \ldots, m$. *The Fisher information matrix* $\mathbf{F}$ *is positive definite if and only if*:

(1) $m \geq \max\{p_1, \ldots, p_{J-1}, p_c + p_H\}$, *where* $p_H = \dim(\bigcap_{j=1}^{J-1} \mathcal{M}(\mathbf{H}_j^T))$ *and* $\mathcal{M}(\mathbf{H}_j^T)$ *stands for the column space of* $\mathbf{H}_j^T$; *and*

(2) $\mathbf{x}_i$*'s keep* $\mathbf{H}_j$ *of full row rank* p_j, $j = 1, \ldots, J-1$; *and keep* $\mathbf{H}_c$ *of full row rank* p_c; *as well as* $\mathcal{M}(\mathbf{H}_c^T) \cap (\bigcap_{j=1}^{J-1} \mathcal{M}(\mathbf{H}_j^T)) = \{0\}$.

As a special case, if $\mathbf{H}_1 = \cdots = \mathbf{H}_{J-1}$, *then* $\mathbf{F}$ *is positive definite if and only if* $m \geq p_c + p_1$ *and the extended matrix* $(\mathbf{H}_1^T, \mathbf{H}_c^T)$ *is of full rank* $p_c + p_1$.

Since the determination of $\mathcal{M}(\mathbf{H}_c^T) \cap (\bigcap_{j=1}^{J-1} \mathcal{M}(\mathbf{H}_j^T))$ in Theorem 3.3 is not straightforward, we provide a formula for its calculation in Appendix A.2.

Theorem 3.3 implies that the number m of distinct experimental settings could be as low as $\max\{p_1, \ldots, p_{J-1}, p_c + p_H\}$, denoted as $k_{\min}$, which is strictly less than the number of parameters $p = p_1 + \cdots + p_{J-1} + p_c$ if $J \geq 3$. If the model is constructed with the same set of predictors for different categories, for example, a main-effects model, then $\mathbf{H}_1 = \cdots = \mathbf{H}_{J-1}$ and $k_{\min}$ is just $p_c + p_1$. This confirms the finding by Yang, Tong and Mandal ((2017), Theorem 3) on cumulative link models and extends it to the general case covering all the 12 multinomial logistic models. We examine examples in Section 5, and revisit the issue of minimal support in Section 5.6. For the applications that we have examined, $k_{\min}$ experimental settings ensure the positive definiteness of $\mathbf{F}$. We provide more examples in the Supplementary Material (Section S.14).

4. Determinant of the Fisher information matrix. Until now, we have derived basic results that are needed for the identification of optimal designs irrespective of the specific criterion. We also derived a result on the minimal number of experimental settings. Now we turn our focus to D-optimal designs. The D-criterion for optimal designs maximizes the determinant of the Fisher information matrix $\mathbf{F}$. In order to study D-optimal designs and their properties, we start with a characterization of the structure of $|\mathbf{F}|$. Recall that n is the total number of observations with n_i of them assigned to the ith experimental setting $\mathbf{x}_i$. We further define a $mJ \times mJ$ diagonal matrix $\mathbf{W} = \operatorname{diag}\{w_1 \operatorname{diag}(\boldsymbol{\pi}_1)^{-1}, \ldots, w_m \operatorname{diag}(\boldsymbol{\pi}_m)^{-1}\}$ with proportions $w_i = n_i/n$, and a $mJ \times p$ matrix

$$\mathbf{G} = \begin{pmatrix} \mathbf{c}_{11}\mathbf{h}_1^T(\mathbf{x}_1) & \cdots & \mathbf{c}_{1,J-1}\mathbf{h}_{J-1}^T(\mathbf{x}_1) & \sum_{j=1}^{J-1} \mathbf{c}_{1j} \cdot \mathbf{h}_c^T(\mathbf{x}_1) \\ \mathbf{c}_{21}\mathbf{h}_1^T(\mathbf{x}_2) & \cdots & \mathbf{c}_{2,J-1}\mathbf{h}_{J-1}^T(\mathbf{x}_2) & \sum_{j=1}^{J-1} \mathbf{c}_{2j} \cdot \mathbf{h}_c^T(\mathbf{x}_2) \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{c}_{m1}\mathbf{h}_1^T(\mathbf{x}_m) & \cdots & \mathbf{c}_{m,J-1}\mathbf{h}_{J-1}^T(\mathbf{x}_m) & \sum_{j=1}^{J-1} \mathbf{c}_{mj} \cdot \mathbf{h}_c^T(\mathbf{x}_m) \end{pmatrix}. \tag{7}$$

Then we can reformulate $\mathbf{F}$ into the form of $\mathbf{G}^T\mathbf{W}\mathbf{G}$ so that $|\mathbf{F}|$ becomes a homogeneous polynomial of allocations.

THEOREM 4.1. *Consider the multinomial logistic model* (1) *with independent observations. The Fisher information matrix* $\mathbf{F} = n\mathbf{G}^T\mathbf{W}\mathbf{G}$.

In order to find D-optimal designs, we need to maximize $|\mathbf{G}^T\mathbf{W}\mathbf{G}|$. Since $\mathbf{W}$ is diagonal, we obtain the following theorem as a direct consequence of Theorem 1.1.2 of Fedorov (1972) or Lemma 3.1 of Yang and Mandal (2015).

THEOREM 4.2. *The determinant of the Fisher information matrix is*

$$|\mathbf{F}| = n^p |\mathbf{G}^T\mathbf{W}\mathbf{G}| = n^p \sum_{\alpha_1 \geq 0, \ldots, \alpha_m \geq 0 : \sum_{i=1}^m \alpha_i = p} c_{\alpha_1, \ldots, \alpha_m} \cdot w_1^{\alpha_1} \cdots w_m^{\alpha_m} \tag{8}$$

with

$$
\begin{aligned}
&c_{\alpha_1,\ldots,\alpha_m} \\
&= \sum_{(i_1,\ldots,i_p)\in\Lambda(\alpha_1,\ldots,\alpha_m)} |\mathbf{G}[i_1,\ldots,i_p]|^2 \prod_{k:\alpha_k>0} \prod_{l:(k-1)J<i_l\le kJ} \pi_{k,i_l-(k-1)J}^{-1} \qquad (9)\\
&\ge 0,
\end{aligned}
$$

where $\alpha_1,\ldots,\alpha_m$ *are nonnegative integers,* $\Lambda(\alpha_1,\ldots,\alpha_m) = \{(i_1,\ldots,i_p) \mid 1 \le i_1 < \cdots < i_p \le mJ; \#\{l : (k-1)J < i_l \le kJ\} = \alpha_k, k = 1,\ldots,m\}$, *and* $\mathbf{G}[i_1,\ldots,i_p]$ *is the submatrix consisting of the* i_1*th*, $\ldots$, i_p*th rows of* $\mathbf{G}$.

It follows from Theorem 4.2, that the determinant of the Fisher information matrix is an order-p homogeneous polynomial of the proportions $w_1,\ldots,w_m$, with coefficients $c_{\alpha_1,\ldots,\alpha_m}$, which must be nonnegative. As a matter of fact, the structure of the determinant can be further simplified quite significantly. This is given in Lemma S.3, Theorem S.6 and Corollaries S.3 and S.4, in the Supplementary Material (Section S.6).

EXAMPLE 2.1 (continued). In this example, the number of factors is $d = 4$, and the experimental settings are $\mathbf{x}_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4})^T$, $i = 1,\ldots,m$. Since $p_1 = p_2 = 3$, $p_c = 2$, and the number of parameters $p = p_1 + p_2 + p_c = 8$, the minimal number of experimental settings is $m = p_1 + p_c = 5$ by Theorem 3.3. We consider the simplest case $m = 5$. That is,

$$
\mathbf{H}_1^T = \mathbf{H}_2^T = \begin{pmatrix} 1 & x_{11} & x_{12} \\ \vdots & \vdots & \vdots \\ 1 & x_{51} & x_{52} \end{pmatrix}, \qquad \mathbf{H}_c^T = \begin{pmatrix} x_{13} & x_{14} \\ \vdots & \vdots \\ x_{53} & x_{54} \end{pmatrix}.
$$

By (8) of Theorem 4.2, $|\mathbf{F}|$ is an order-8 homogeneous polynomial of allocations to the 5 experimental settings, which may contain up to $(8+5-1)!/(8!(5-1)!) = 465$ terms. However, Lemma S.3 implies $c_{\alpha_1,\ldots,\alpha_5} \ne 0$ only if $\alpha_i \in \{0,1,2\}$. On the other hand, Corollary S.4 says $c_{\alpha_1,\ldots,\alpha_5} \ne 0$ only if $\#\{i \mid \alpha_i > 0\} \ge p_1 + p_c = 5$, that is, $\alpha_i > 0$ for each i. Therefore, $|\mathbf{F}|$ contains only $5!/(3!2!) = 10$ nonzero terms with $\alpha_i \in \{1,2\}$. That is,

$$
|\mathbf{F}| = n^8 \cdot \prod_{i=1}^{5} w_i \cdot \sum_{1\le i_1<i_2<i_3\le 5} e_{i_1,i_2,i_3} w_{i_1} w_{i_2} w_{i_3}
$$

for some coefficients e_{i_1,i_2,i_3}. Actually, in terms of the notation $c_{\alpha_1,\ldots,\alpha_m}$ in (9) of Theorem 4.2, $e_{i_1,i_2,i_3} = c_{\alpha_1,\ldots,\alpha_5}$ with $\alpha_i = 2$ for $i \in \{i_1,i_2,i_3\}$ and 1 otherwise. For example, $e_{1,2,3} = c_{2,2,2,1,1}$.

REMARK 4.1. The reformulation of $\mathbf{F}$ in Theorem 4.1 enables us to conclude that $|\mathbf{F}|$ is an order-p homogeneous polynomial of allocations or proportions in Theorem 4.2. It establishes the foundation for an efficient numerical algorithm for searching D-optimal designs (Section 5). On the other hand, the simplification of $|\mathbf{F}|$ based on Lemma S.3 and Corollary S.4 allows us to obtain D-optimal designs analytically for minimally supported designs, which is critical for investigating their theoretical properties (Section 5.6).

With the aid of Theorem 4.1, Theorem 4.2 is derived in a more straightforward way. It is broader and with more detailed structures than Theorem 2 in Yang, Tong and Mandal (2017), which focused on *po* models only. Lemma S.3 and Corollary S.4 also extend the corresponding results in Yang, Tong and Mandal ((2017), Lemmas S.4 and S.5 in their Supplementary Material).

5. D-optimal designs. This section is dedicated to D-optimal designs. Our main focus is on locally D-optimal designs, in which the information matrix is computed at an assumed value of the model parameter $\boldsymbol{\theta}$. Later, in Section 5.4, we study EW D-optimal designs under a prior distribution of the parameters. We consider both *approximate designs*, which specify proportions of observations at experimental settings without regard to the number of observations n, as well as *exact designs* where allocation of the n observations to the locations is determined. To start with, we fix a given set of experimental settings $\mathbf{x}_i$, $i = 1, \ldots, m$, and consider D-optimal designs $(n_1, \ldots, n_m)$ or $(w_1, \ldots, w_m)$ that maximizes the determinant of the Fisher information matrix $|\mathbf{F}|$. Then we consider the optimization problem over experimental settings as well, using a grid-point search algorithm. While we are thus treating the factor space as *discrete* in analytical and computational results, in Section 5.5 we will address the problem of determining D-optimal designs in a *continuous* factor space.

5.1. *Design space for multinomial logistic models.* The experimental settings $\mathbf{x}_i$'s appear in the multinomial logistic model through the categorical probabilities π_{ij}'s, which are functions of the parameter vector $\boldsymbol{\theta}$ and the $\mathbf{x}_i$'s. Our first result examines the restriction imposed on the $\mathbf{x}_i$'s by the requirement $0 < \pi_{ij} < 1$, $j = 1, \ldots, J$. This gives us the collection of all feasible experimental settings, known as the *design space*,

$$\mathcal{X} = \{\mathbf{x} = (x_1, \ldots, x_d)^T \mid 0 < \pi_j < 1, j = 1, \ldots, J\}.$$

THEOREM 5.1. *Let* $a_j = \mathbf{h}_j^T(\mathbf{x})\boldsymbol{\beta}_j + \mathbf{h}_c^T(\mathbf{x})\boldsymbol{\zeta}$, $j = 1, \ldots, J-1$. *The design space is* $\mathcal{X} = \{\mathbf{x} = (x_1, \ldots, x_d)^T \mid a_j \in (-\infty, \infty), j = 1, \ldots, J-1\}$ *for baseline-category, adjacent-categories and continuation-ratio logit models*; *and* $\mathcal{X} = \{\mathbf{x} = (x_1, \ldots, x_d)^T \mid -\infty < a_1 < a_2 < \cdots < a_{J-1} < \infty\}$ *for cumulative logit models.*

Theorem 5.1 essentially places no restriction on the design space for models other than cumulative logit models, regardless of the odds structure. Its proof and an illustrative example are provided in the Supplementary Material (Sections S.15 and S.14, resp.).

5.2. *D-optimal approximate designs.* Given distinct experimental settings $\mathbf{x}_i \in \mathcal{X}$, $i = 1, \ldots, m$, we look for a D-optimal approximate design $\mathbf{w} = (w_1, \ldots, w_m)^T$ that maximizes the quantity $|\mathbf{G}^T\mathbf{W}\mathbf{G}|$ defined in Theorem 4.1. The collection of all feasible approximate designs $S = \{(w_1, \ldots, w_m)^T \in \mathbb{R}^m \mid w_i \geq 0, i = 1, \ldots, m; \sum_{i=1}^m w_1 = 1\}$ is a bounded closed convex set. The objective function

$$f(\mathbf{w}) = |\mathbf{G}^T\mathbf{W}\mathbf{G}|$$

is an order-p homogeneous polynomial by Theorem 4.2. Therefore, a D-optimal approximate design that maximizes $f(\mathbf{w})$ must exist. For typical applications, we need designs coming from $S_+ = \{\mathbf{w} \in S \mid f(\mathbf{w}) > 0\}$ to avoid degenerate cases. Due to Theorem 2.1 and the log-concavity of the determinant on positive semidefinite matrices, we know $f(\mathbf{w})$ is log-concave (Silvey (1980), Yang, Tong and Mandal (2017)) and S_+ is convex. A useful result as a corollary of Theorem 3.2, with a proof provided in the Supplementary Material (Section S.15), is the following.

COROLLARY 5.1. *S_+ is nonempty if and only if $f(\mathbf{w}_u) > 0$, where $\mathbf{w}_u = (1/m, \ldots, 1/m)^T$ is the uniform allocation. In this case, $f(\mathbf{w}) > 0$ for any $\mathbf{w} = (w_1, \ldots, w_m)^T$ such that $0 < w_i < 1$, $i = 1, \ldots, m$.*

In order to avoid trivial cases, we assume $f(\mathbf{w}_u) > 0$ from now on. Following Yang, Tong and Mandal ((2017), Section 3), we define

$$f_i(z) = f\left(\frac{w_1(1-z)}{1-w_i}, \ldots, \frac{w_{i-1}(1-z)}{1-w_i}, z, \frac{w_{i+1}(1-z)}{1-w_i}, \ldots, \frac{w_m(1-z)}{1-w_i}\right)$$

with $0 \le z \le 1$ and $\mathbf{w} = (w_1, \ldots, w_m)^T \in S_+$. As a special case of the general equivalence theorem (Atkinson, Donev and Tobias (2007), Fedorov and Leonov (2014), Kiefer (1974), Pukelsheim (1993), Stufken and Yang (2012), Yang, Mandal and Majumdar (2016), Yang, Tong and Mandal (2017)), $\mathbf{w}$ is D-optimal if and only if $f_i(z)$ attains its maximum at $z = w_i$ for each $i = 1, \ldots, m$. Actually, $f_i(z) = n^{-p}|\mathbf{F}|$ for the approximate design $\mathbf{w} + (\mathbf{e}_i - \mathbf{w}) \cdot (z - w_i)/(1 - w_i)$, where $\mathbf{e}_i$ is an $m \times 1$ vector with the ith coordinate 1 and all others 0. Parallel to Theorem 6 in Yang, Tong and Mandal (2017), we obtain an explicit formula for $f_i(z)$ as an order-p polynomial of z (see Theorem S.9 in the Supplementary Material).

In order to find D-optimal designs numerically, we use a lift-one algorithm (Section S.10 in the Supplementary Material), which is essentially the same as the one in Yang, Tong and Mandal (2017) for cumulative link models. The lift-one algorithm is of general-equivalence-theorem type (Yang, Mandal and Majumdar (2016)). Its convergence to a global maximum is guaranteed (Yang and Mandal (2015)).

We use the following real experiment as an illustration. In the application (Zocchi and Atkinson (1999)), a uniform design and a continuation-ratio logit model with *npo* component was used for the analysis. As noted earlier, the theoretical results for the *npo* model, as special cases of the results for the *ppo* model, are given in the Supplementary Material (Section S.8).

EXAMPLE 5.1 (Emergence of house flies). Zocchi and Atkinson (1999) discussed an experiment on emergence of house flies originally reported by Itepan (1995). In this experiment, $n = 3500$ pupae were grouped evenly into seven sets and exposed to seven doses of radiation (in units Gy), $x_i = 80, 100, 120, 140, 160, 180, 200$, respectively. After a period of time, for each set of $n_i = 500$ pupae, the summarized responses are the number y_{i1} of flies that died before the opening of the pupae (`unopened pupae`), the number y_{i2} of flies out of `opened pupae` but died before complete emergence and the number $y_{i3} = n_i - y_{i1} - y_{i2}$ of flies out of `opened pupae` and completely emerged. Following Zocchi and Atkinson (1999), we assume that the responses of the $n_i = 500$ pupae in the ith set are independent and follow the same distribution. Then the summary responses (y_{i1}, y_{i2}, y_{i3}) follow a multinomial distribution and have a clearly nested or hierarchical structure (see Table 1 of Zocchi and Atkinson (1999) for the experimental data). We confirm that the continuation-ratio logit model with *npo* component fits the data the best in terms of AIC and BIC (see Table 5 in the Supplementary Material); this model was adopted by Zocchi and Atkinson (1999) as follows:

$$\log\left(\frac{\pi_{i1}}{\pi_{i2} + \pi_{i3}}\right) = \beta_{11} + \beta_{12}x_i + \beta_{13}x_i^2, \qquad \log\left(\frac{\pi_{i2}}{\pi_{i3}}\right) = \beta_{21} + \beta_{22}x_i.$$

The model has 5 parameters with fitted values $\hat{\boldsymbol{\beta}} = (\hat{\beta}_{11}, \hat{\beta}_{12}, \hat{\beta}_{13}, \hat{\beta}_{21}, \hat{\beta}_{22})^T = (-1.935, -0.02642, 0.0003174, -9.159, 0.06386)^T$.

Considering a followup experiment with the fitted parameter values as the assumed values, we obtain the locally D-optimal approximate design (Table 1) using the lift-one algorithm. The efficiency of the original uniform allocation is $(|\mathbf{F}_{\text{original}}|/|\mathbf{F}_{D\text{-opt}}|)^{1/5} = 83.1\%$. The D-optimal approximate design only requires 4 rounds (doses) of experiments. It is interesting to note that in this example the D-optimal design does not allocate observations to high doses of radiation; we will revisit this issue in Section 6. The D-optimal exact design, Bayesian and EW D-optimal designs listed in Table 1 will be described in Sections 5.3 and 5.4, respectively. We will return to this example later.

TABLE 1
Exact and approximate allocations for house flies experiment

Dose of radiation (Gy)	80	100	120	140	160	180	200
Original allocation	500	500	500	500	500	500	500
D-optimal exact	1091	0	1021	374	1014	0	0
Original proportion	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429
D-optimal approximate	0.3116	0	0.2917	0.1071	0.2896	0	0
Bayesian D-optimal	0.3159	0.0000	0.2692	0.1160	0.2990	0.0000	0.0000
EW D-optimal	0.3120	0	0.2911	0.1087	0.2882	0	0

5.3. *D-optimal exact designs*. In practice, a design should specify an integer-valued allocation of the n experimental units to the m distinct experimental settings $\mathbf{x}_i$, $i = 1, \ldots, m$. An integer-valued allocation $\mathbf{n} = (n_1, \ldots, n_m)^T$ that maximizes $|\sum_{i=1}^m n_i \mathbf{F}_i|$ as defined in (3) with $\sum_{i=1}^m n_i = n$ is known as a *D-optimal exact design*.

Although different rounding algorithms have been proposed in the literature to obtain an exact allocation from an optimal approximate allocation, an algorithm searching for optimal exact designs directly is still needed (see Yang, Tong and Mandal (2017), Section 4 and reference therein). For simplicity, we denote the objective function as $f(\mathbf{n}) = f(n_1, \ldots, n_m) = |\sum_{i=1}^m n_i \mathbf{F}_i|$ if there is no ambiguity. Following Yang, Mandal and Majumdar (2016) and Yang, Tong and Mandal (2017), we define $f_{ij}(z) = f(n_1, \ldots, n_{i-1}, z, n_{i+1}, \ldots, n_{j-1}, n_i + n_j - z, n_{j+1}, \ldots, n_m)$ with $z = 0, 1, \ldots, n_i + n_j$ given $1 \le i < j \le m$ and $\mathbf{n} = (n_1, \ldots, n_m)^T$. Similar to Theorem 9 in Yang, Tong and Mandal (2017), we can obtain an explicit formula for $f_{ij}(z)$ (see Theorem S.10 in the Supplementary Material) from Theorem 4.2, Lemma S.3 and Corollary S.4.

Given $n > 0$, we assume that there exists an exact allocation $\mathbf{n} = (n_1, \ldots, n_m)^T$ such that $f(\mathbf{n}) > 0$. An exchange algorithm (Section S.10 in the Supplementary Material), which is essentially the same as the one in Yang, Tong and Mandal (2017), is used for obtaining an exact design. Note that the integer-valued allocation found by the exchange algorithm is not guaranteed to be D-optimal, especially when n is relatively small compared with m (Yang, Mandal and Majumdar (2016)).

The following real experiment shows that a minimally supported design can be D-optimal for a cumulative logit model with *npo* assumption.

EXAMPLE 5.2 (Trauma clinical trial). Chuang-Stein and Agresti (1997) studied a dataset of trauma patients that has five ordered response categories, `death`, `vegetative state`, `major disability`, `minor disability` and `good recovery`, describing their clinical outcomes. These five categories are often called the Glasgow Outcome Scale (GOS) in the literature of critical care (Jennett and Bond (1975)). Agresti (2010) found strong evidence against the assumption of proportional odds for the trauma data. We fit the data with different logit models, as well as different odds assumptions, and confirm that the cumulative logit model with *npo* assumption fits the trauma data the best in terms of AIC and BIC (see Table 4 in the Supplementary Material). Thus the model is

$$\log\left(\frac{\pi_{i1} + \cdots + \pi_{ij}}{\pi_{i,j+1} + \cdots + \pi_{i5}}\right) = \beta_{j1} + \beta_{j2} x_i, \quad i = 1, 2, 3, 4;\ j = 1, 2, 3, 4. \tag{10}$$

The fitted parameter values are $\hat{\boldsymbol{\beta}} = (\hat{\beta}_{11}, \hat{\beta}_{12}, \hat{\beta}_{21}, \hat{\beta}_{22}, \hat{\beta}_{31}, \hat{\beta}_{32}, \hat{\beta}_{41}, \hat{\beta}_{42})^T = (-0.865, -0.113, -0.094, -0.269, 0.706, -0.182, 1.909, -0.119)^T$. The clinical trial with 802 patients assigned 210, 190, 207, 195 patients to the four treatment groups, `Placebo`, `Low`

`dose`, `Medium dose` and `High dose`, respectively, which is roughly uniform on dosage. If we treat the fitted parameter values as the assumed values for local optimality, using the exchange algorithm, we find the D-optimal exact design (401, 0, 0, 401), which is minimally supported and uniform on its supporting points. If a followup experiment is considered and the fitted parameter values are the true values, the efficiency of the original allocation compared with our D-optimal design is only 74.7%.

EXAMPLE 5.1 (Emergence of house flies (*continued*)). In this experiment, the allocation of experimental units used by the scientists is 500 for each of the seven doses of radiation. Using the exchange algorithm, we obtain the D-optimal exact allocation (Table 1), which is similar to the D-optimal approximate design. Actually, the proportions of the D-optimal exact design $(1091, 1021, 374, 1014)/3500 = (0.3117, 0.2917, 0.1069, 0.2897)$ on the four support points are roughly the same as the ones in the D-optimal approximate design. If the fitted parameter values are the true values, the efficiency of the allocation used in the experiment is 83.1% compared with the D-optimal exact allocation.

REMARK 5.1. Our D-optimal exact designs are "locally" D-optimal since they require assumed parameter values. When there is no pilot study available for a good parameter estimate, D-optimal designs can still be used in a multistage design (see Khuri et al. (2006) for a review). In the trauma clinical trial, for example, we may consider a two-stage design with 802 patients in total. At Stage 1, we may assign 240 patients (about 30%) randomly and uniformly to the four treatment groups. After collecting the outcomes of the patients from Stage 1, we obtain the estimated coefficients and then find the corresponding D-optimal exact design for the rest 562 patients, known as Stage 2. The parameter estimates from the two-stage design are based on the outcomes of all the 802 patients. In terms of root mean squared error (RMSE) from the assumed parameter values, the two-stage design achieves 0.149 for the intercepts $\beta_{11}, \ldots, \beta_{41}$ and 0.053 for the slopes $\beta_{12}, \ldots, \beta_{42}$, on average over 100 simulations. The corresponding average RMSEs from the original design are 0.190 and 0.070, respectively. The reductions in RMSEs by using our D-optimal exact designs in the two-stage design are both significant.

5.4. *EW D-optimal designs*. The D-optimal approximate and exact designs discussed thus far are locally D-optimal designs using assumed parameter values. Bayesian D-optimality (Chaloner and Verdinelli (1995)), which maximizes $E(\log|\mathbf{F}|)$ with a given prior distribution on the unknown parameters, provides an alternative approach. A drawback of the Bayesian approach is its computational intensity since the objective function deals with multiple integrals. An alternative solution is the EW D-optimality (Atkinson, Donev and Tobias (2007), Yang, Mandal and Majumdar (2016), Yang, Tong and Mandal (2017)), which maximizes $\log|E(\mathbf{F})|$ or $|E(\mathbf{F})|$ instead. Among all of the criteria proposed by Atkinson, Donev and Tobias ((2007), Table 18.1) as surrogates for Bayesian D-optimality, including $-\log E(|\mathbf{F}|^{-1})$, $-\log|E(\mathbf{F}^{-1})|$, and $\log E|\mathbf{F}|$, EW D-optimal design requires the minimum computation. Yang, Mandal and Majumdar (2016) showed that an EW D-optimal design could be highly efficient in terms of the Bayesian criterion in comparison to the Bayesian D-optimal design, while computational time is essentially the same as a locally D-optimal design. Yang, Tong and Mandal (2017) also used EW-criterion for cumulative link models with proportional odds and confirmed its high-efficiency.

By Theorem 2.1 and Corollary 3.1, the Fisher information matrix in our case is $\mathbf{F} = \sum_{i=1}^{m} n_i \mathbf{X}_i^T \mathbf{U}_i \mathbf{X}_i$, where $\mathbf{U}_i$ consists of $u_{st}(\boldsymbol{\pi}_i)$, the only components involving model parameters. In order to calculate $E(\mathbf{F})$ with respect to a prior distribution on parameters, we first calculate $E(u_{st}(\boldsymbol{\pi}_i))$ and then use the results and algorithms developed for locally D-optimal designs to get EW D-optimal designs.

We provide formulae in both the proof of Theorem 5.1 and Section S.11 (in matrix form) in the Supplementary Material for calculating π_{ij}'s given $\mathbf{X}_i$'s and the parameter values. We also provide formulae in Appendix A.3 for calculating $u_{st}(\boldsymbol{\pi}_i) = \mathbf{c}_{is}^T \operatorname{diag}(\boldsymbol{\pi}_i)^{-1}\mathbf{c}_{it}$'s given π_{ij}'s.

By Theorem 5.1, for baseline-category, adjacent-categories and continuation-ratio logit models, there is essentially no restriction on values of parameters for a given design point $\mathbf{x} = (x_1, \ldots, x_d)^T$. The corresponding $E(u_{st}(\boldsymbol{\pi}_i))$ can be obtained by integrating over the same range of parameter values for different design points. However, for cumulative logit models, where the parameters should satisfy $a_1 < a_2 < \cdots < a_{J-1}$ with $a_j = \mathbf{h}_j^T(\mathbf{x})\boldsymbol{\beta}_j + \mathbf{h}_c^T(\mathbf{x})\boldsymbol{\zeta}$, the domain of integration is typically not rectangular.

EXAMPLE 5.2 (Trauma clinical trial (*continued*)). In this example, the cumulative logit model (10) with nonproportional odds was adopted. Given the predetermined set $\mathcal{X} = \{1, 2, 3, 4\}$ consisting of $m = 4$ design points, the feasible parameter space is $\boldsymbol{\Theta} = \{\boldsymbol{\theta} = (\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}, \beta_{31}, \beta_{32}, \beta_{41}, \beta_{42})^T \mid \beta_{11} + \beta_{12}x < \beta_{21} + \beta_{22}x < \beta_{31} + \beta_{32}x < \beta_{41} + \beta_{42}x, \text{ for } x \in \mathcal{X}\}$, which is not rectangular. Zocchi and Atkinson (1999) used a multivariate normal prior with its variance-covariance matrix estimated from an initial dataset. For the computations in this example, we bootstrap the 802 observations from the initial dataset for 1000 times and denote the corresponding fitted parameters as $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{1000}$. Then an estimate of the Bayesian criterion $\phi(\mathbf{w}) = E(\log|\mathbf{F}(\mathbf{w}, \boldsymbol{\theta})|)$ for design $\mathbf{w} = (w_1, \ldots, w_m)^T$ is

$$\hat{\phi}(\mathbf{w}) = \frac{1}{1000}\sum_{j=1}^{1000} \log|\mathbf{F}(\mathbf{w}, \boldsymbol{\theta}_j)| = \frac{1}{1000}\sum_{j=1}^{1000} \log\left|\sum_{i=1}^{m} w_i \mathbf{F}_i(\boldsymbol{\theta}_j)\right|.$$

The Bayesian D-optimal design listed in Table 2 maximizes $\hat{\phi}(\mathbf{w})$, while the EW D-optimal design maximizes $|\sum_{i=1}^m w_i \hat{E}(\mathbf{F}_i)|$ with estimated entry-by-entry expectation $\hat{E}(\mathbf{F}_i) = \sum_{j=1}^{1000} \mathbf{F}_i(\boldsymbol{\theta}_j)/1000$. The two designs are essentially the same in this case.

EXAMPLE 5.1 (Emergence of house flies (*continued*)). Similar to the trauma example, we bootstrap the original observations 1000 times and obtain the corresponding Bayesian and EW D-optimal designs (see Table 1). In this case, EW D-optimal design is close to the Bayesian design. In order to check the robustness of these designs toward misspecified parameter values, we consider the same set of bootstrapped parameter vectors $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{1000}$. For $j = 1, \ldots, 1000$, we apply our lift-one algorithm to find the corresponding D-optimal allocation $\mathbf{p}_j$ with $\boldsymbol{\theta}_j$ as the assumed parameter value. The efficiencies of a target design $\mathbf{p}$ with respect to $\mathbf{p}_j$ is defined as $(|\mathbf{F}(\mathbf{p}, \boldsymbol{\theta}_j)|/|\mathbf{F}(\mathbf{p}_j, \boldsymbol{\theta}_j)|)^{1/p}$ with $p = 5$ parameters in this case. The summary statistics of efficiencies in Table 3 show that the Bayesian and EW D-optimal designs are highly robust in terms of parameter misspecification; the EW design is slightly better. Both of them are much better than the uniform design used in the study.

TABLE 2
D-optimal designs for trauma clinical trial

Design point x	1	2	3	4
Treatment group	Placebo	Low	Medium	High
Original allocation	210	190	207	195
D-optimal exact	401	0	0	401
Original proportion	0.2618	0.2369	0.2581	0.2431
D-optimal approximate	0.5	0	0	0.5
Bayesian D-optimal	0.4997	1.0×10^{-8}	2.5×10^{-8}	0.5003
EW D-optimal	0.5	0	0	0.5

TABLE 3
Efficiencies of designs for house flies experiment

Design	Min	1st Quartile	Median	3rd Quartile	Max
Bayesian D-optimal	0.9912	0.9985	0.9989	0.9992	0.9998
EW D-optimal	0.9934	0.9991	0.9996	0.9998	1.0000
Uniform design	0.7762	0.8181	0.8304	0.8445	0.8861

5.5. *Grid point search for continuous factors*. Both the lift-one and exchange algorithms are appropriate for searching optimal designs with a predetermined finite set of experimental settings. When a factor is continuous, one common practice is to partition the continuous region of the factor levels into finite subintervals and consider only the grid points as its discrete levels. It works reasonably well with a moderate number of continuous factors (see Yang, Biedermann and Tang (2013) for a discussion on the efficiency of optimal designs based on grid-point experimental settings).

Once a set of grid points is chosen for each continuous factor, the design problem becomes an allocation problem on a finite set of design points. Hence, all of the previous algorithms and results can be applied. We use the example of house flies for the purpose of illustration.

EXAMPLE 5.1 (Emergence of house flies (*continued*)). The sole factor in this experiment, dose of radiation, is by nature a continuous factor. In the emergence of house flies experiment, seven levels ranging equidistantly from 80 to 200 were used, that is, with grid distance 20. Suppose a followup experiment is considered and dose levels with grid distance 5 are feasible (which technically depends on the sensitivity of the radiation device). Then there are 25 dose levels available as 80, 85, 90, ..., 195, 200. The D-optimal approximate design given the 25 grid-5 design concentrates on five design points 80, 120, 125, 155, 160 with optimal allocation 0.3163, 0.1429, 0.2003, 0.1683, 0.1723, respectively. If we further consider the set of 121 grid-1 dose levels, the D-optimal design is supported on five design points 80, 122, 123, 157, 158 with allocations 0.3163, 0.0786, 0.2636, 0.2206, 0.1209. The optimal design seems to converge to a three-point design as the grid points become finer and finer. Actually, if we reallocate the grid-1 D-optimal design, denoted as $\mathbf{p}_1$, into a minimally supported design $\mathbf{p}_{1m}$ at dose levels 80, 123, 157 with weights 0.3163, 0.3422, 0.3415, then the efficiency of $\mathbf{p}_{1m}$ compared with $\mathbf{p}_1$ is as high as 99.99%.

5.6. *Minimally supported designs*. An important question in design theory is, what is the least number of design points we need to keep $\mathbf{F}$ positive definite? In other words, what is the number of distinct experimental settings for a *minimally supported design*? Theorem 3.3 provides a lower bound, which has been denoted as $k_{\min}$. For typical models proposed in the literature, for example, main-effects models (see Example S.3), $k_{\min}$ is the minimal number of experimental settings since the conditions in Theorem 3.3 are satisfied. Answers to general cases rely on the specific forms of the predictor functions $\mathbf{h}_j$ and $\mathbf{h}_c$ if applicable.

Another question is whether a uniform allocation is D-optimal among minimally supported designs. The answer is known to be "Yes" for $J = 2$ since all of the four logit models are equivalent to the usual logistic model for binary responses.

Nevertheless, for *po* models, Yang, Tong and Mandal (2017) showed analytically that uniform allocations are not D-optimal among minimally supported designs for cumulative link. We use Example S.7 in the Supplementary Material to show that the conclusion is the same

for *po* models with baseline-category, adjacent-categories and continuation-ratio logit links as well. Actually, the objective functions take the same form for the four different logit models. Since *po* models are special cases of *ppo* models, we conclude that uniform allocations are not D-optimal in general for *ppo* models as well.

As for *npo* models with $J \geq 3$, in general, uniform allocations are not D-optimal either (see Section S.13 in the Supplementary Material). Nevertheless, for "regular" *npo* models, that is, when $p_1 = \cdots = p_{J-1}$, uniform allocations are D-optimal among minimally supported designs (see Corollary S.7 in the Supplementary Material), which theoretically confirms the D-optimal design for trauma clinical trial in Example 5.2.

6. Discussion. A criticism of optimal designs is that their rigid reliance on the assumed model does not allow exploration of different models; this is mainly due to the frugality of the number of experimental settings. Nevertheless, the results can be used to identify efficient designs that allow flexibility. For instance, in the trauma clinical trial example (Example 5.2), the D-optimal exact design assigns 401 observations to the placebo (1) and high dose (4), but none to the low (2) or medium (3) levels. If instead we allocate 361 observations to each of 1 and 4, and 40 observations to each of 2 and 3, we gain some flexibility with a loss of efficiency (in comparison to the D-optimal design) of only 5%. This is further discussed in the Supplementary Material (Table 6).

In Example 5.1, the support of the D-optimal design did not include some of the extreme points of the experimental region. This is a departure from the nature of D-optimal designs for univariate responses, and is something that will be studied in the future.

For EW and Bayesian D-optimal designs, the choice of the prior on the parameter space $\boldsymbol{\Theta}$ is critical. Suppose the parameters are $\theta_1, \ldots, \theta_p$ with individual ranges $\theta_i \in I_i$. As a common practice for generalized linear models, a uniform prior or independent normal prior may be assumed for a rectangular domain $I_1 \times \cdots \times I_p$ (Yang, Mandal and Majumdar (2016)). One issue with multinomial logit models is that the feasible domain of $\boldsymbol{\Theta}$ may not be rectangular, at least for cumulative logit models. For the two examples in Section 5.4, we use bootstrapping to obtain an empirical prior. Zocchi and Atkinson (1999) adopted a multivariate normal prior with mean vector and covariance structure estimated from previous experimental data. For the two examples that we study, our results are essentially the same when we use the multivariate normal prior. However, when we use uniform or independent normal prior, the EW D-optimal design is not as good as the Bayesian D-optimal design in terms of robustness. We will further investigate the choice of prior and criteria other than EW optimality.

When a pilot study or experimental data is not available for locally or EW optimal designs, sequential design or multistage design may be used to obtain an initial guess or a reasonable prior for unknown parameters. In Remark 5.1, we provide a two-stage design for the trauma clinical trial as an example. Further investigation needs to be done toward best multistage designs for multinomial logistic models.

When confronted with model uncertainty, *npo* versus *ppo* for instance, our work may provide the experimenter the option to choose a design, for instance, that is highly efficient for both *ppo* and *npo* models. Further investigations along this line would be practically useful.

In the literature, model parameters other than $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_{J-1}^T, \boldsymbol{\zeta})^T$ for linear predictors $\eta_{ij} = \mathbf{h}_j^T(\mathbf{x}_i)\boldsymbol{\beta}_j + \mathbf{h}_c^T(\mathbf{x}_i)\boldsymbol{\zeta}$ were also used, for example, in Perevozskaya, Rosenberger and Haines (2003). It can be verified that D-optimal designs will not be affected by the choice of parameters (see Section S.12 in the Supplementary Material).

APPENDIX

A.1. Constant matrix L in equation (1).

$$\mathbf{L}_{\text{baseline}} = \begin{pmatrix} 1 & & & & 0 \\ & 1 & & & 0 \\ & & \ddots & & \vdots \\ & & & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}, \qquad \mathbf{L}_{\text{cumulative}} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ 1 & 1 & \cdots & 1 & 0 \\ 0 & 1 & \cdots & 1 & 1 \\ 0 & 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix},$$

$$\mathbf{L}_{\text{continuation}} = \begin{pmatrix} 1 & & & & 0 \\ & 1 & & & 0 \\ & & \ddots & & \vdots \\ & & & 1 & 0 \\ 0 & 1 & \cdots & \cdots & 1 \\ 0 & 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}, \qquad \mathbf{L}_{\text{adjacent}} = \begin{pmatrix} 1 & & & & 0 \\ & 1 & & & 0 \\ & & \ddots & & \vdots \\ & & & 1 & 0 \\ 0 & 1 & & & \\ 0 & & 1 & & \\ \vdots & & & \ddots & \\ 0 & & & & 1 \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}.$$

A.2. Formula toward $\mathcal{M}(\mathbf{H}_c^T) \cap (\bigcap_{j=1}^{J-1} \mathcal{M}(\mathbf{H}_j^T))$ in Theorem 3.3.

THEOREM A.1. *Suppose* $\mathbf{H}_i$ *is of* $p_i \times m$ *with rank* r_i, $i = 1, \ldots, n$. *Denote* $\operatorname{rank}((\mathbf{H}_{i_1}^T, \ldots, \mathbf{H}_{i_k}^T)) = r_{i_1,\ldots,i_k}$ *for any* $2 \le k \le n$ *and* $1 \le i_1 < \cdots < i_k \le n$. *Then*

$$\begin{aligned} \dim\left(\bigcap_{i=1}^{n} \mathcal{M}(\mathbf{H}_i^T)\right) &= \sum_{i=1}^{n} r_i - \sum_{i_1<i_2} r_{i_1,i_2} \\ &\quad + \sum_{i_1<i_2<i_3} r_{i_1,i_2,i_3} - \cdots + (-1)^{n-1} r_{1,2,\ldots,n}. \end{aligned} \tag{11}$$

The proof of Theorem A.1 is relegated to the Supplementary Material. Note that $\mathcal{M}(\mathbf{H}_c^T) \cap (\bigcap_{j=1}^{J-1} \mathcal{M}(\mathbf{H}_j^T)) = \{0\}$ if and only if its dimension is 0.

A.3. Formulae for calculating $u_{st}(\boldsymbol{\pi}_i)$'s given π_{ij}'s.

THEOREM A.2. *Consider the multinomial logit model* (1). *For* $i = 1, \ldots, m$:

(i) $u_{st}(\boldsymbol{\pi}_i) = u_{ts}(\boldsymbol{\pi}_i)$, $s, t = 1, \ldots, J$;
(ii) $u_{sJ}(\boldsymbol{\pi}_i) = 0$ *for* $s = 1, \ldots, J-1$ *and* $u_{JJ}(\boldsymbol{\pi}_i) = 1$;
(iii) *For* $s = 1, \ldots, J-1$,

$$u_{ss}(\boldsymbol{\pi}_i) = \begin{cases} \pi_{is}(1-\pi_{is}) & \text{for baseline-category}, \\ \gamma_{is}^2(1-\gamma_{is})^2(\pi_{is}^{-1} + \pi_{i,s+1}^{-1}) & \text{for cumulative}, \\ \gamma_{is}(1-\gamma_{is}), & \text{for adjacent-categories}, \\ \pi_{is}(1-\gamma_{is})(1-\gamma_{i,s-1})^{-1} & \text{for continuation-ratio}; \end{cases}$$

(iv) *For* $1 \le s < t \le J-1$,

$$u_{st}(\boldsymbol{\pi}_i) = \begin{cases} -\pi_{is}\pi_{it} & \text{for baseline-category}, \\ -\gamma_{is}\gamma_{it}(1-\gamma_{is})(1-\gamma_{it})\pi_{it}^{-1} & \text{for cumulative}, t-s=1, \\ 0 & \text{for cumulative}, t-s>1, \\ \gamma_{is}(1-\gamma_{it}) & \text{for adjacent-categories}, \\ 0 & \text{for continuation-ratio}; \end{cases}$$

where $\gamma_{ij} = \pi_{i1} + \cdots + \pi_{ij}$, $j = 1, \ldots, J-1$; $\gamma_{i0} \equiv 0$ *and* $\gamma_{iJ} \equiv 1$.

Acknowledgments. The first and third authors were supported by the LAS Award for Faculty of Science at UIC.

SUPPLEMENTARY MATERIAL

Supplementary material (DOI: 10.1214/19-AOS1834SUPP; .pdf). It contains: (1) list of notation; (2) formulae of matrix differentiation; (3) explicit forms of $(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})^{-1}$ for all the four logit models; (4) positive definiteness of $\mathbf{U}$; (5) row rank of $\mathbf{H}$ matrix; (6) results on the coefficient $c_{\alpha_1,\ldots,\alpha_m}$ for simplifying $|\mathbf{F}|$; (7) expressions for *po* models; (8) expressions for *npo* models; (9) model selections for trauma clinical trial and emergence of house flies; (10) lift-one and exchange algorithms; (11) formulae for calculating π_{ij}'s from $\mathbf{X}_i$'s; (12) reparametrization and D-optimality; (13) more discussion on D-optimality of uniform designs; (14) more examples; (15) proofs of Theorems 2.1, 3.1, S.3, S.4, S.5, A.1, 4.1, S.6, 5.1, S.10, S.11, A.2, Corollaries S.2, 5.1, S.7 and Lemmas S.5, S.9, S.3, S.10, S.11, S.12, S.13, S.4.

REFERENCES

AGRESTI, A. (2010). *Analysis of Ordinal Categorical Data*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2742515 https://doi.org/10.1002/9780470594001

AGRESTI, A. (2013). *Categorical Data Analysis*, 3rd ed. Wiley, New York.

ATKINSON, A. C., DONEV, A. N. and TOBIAS, R. D. (2007). *Optimum Experimental Designs, with SAS*. *Oxford Statistical Science Series* **34**. Oxford Univ. Press, Oxford. MR2323647

BU, X., MAJUMDAR, D. and YANG, J. (2020). Supplement to "D-optimal designs for multinomial logistic models." https://doi.org/10.1214/19-AOS1834SUPP.

CHALONER, K. and VERDINELLI, I. (1995). Bayesian experimental design: A review. *Statist. Sci.* **10** 273–304. MR1390519

CHRISTENSEN, R. (2019). Cumulative link models for ordinal regression with the R package ordinal. Available at https://cran.r-project.org/web/packages/ordinal/vignettes/clm_article.pdf.

CHUANG-STEIN, C. and AGRESTI, A. (1997). Tutorial in biostatistics—a review of tests for detecting a monotone dose-response relationship with ordinal response data. *Stat. Med.* **16** 2599–2618.

DOBSON, A. J. and BARNETT, A. G. (2018). *An Introduction to Generalized Linear Models*, 4th ed. *Texts in Statistical Science Series*. CRC Press, Boca Raton, FL. MR3890007

FEDOROV, V. V. (1972). *Theory of Optimal Experiments*. Academic Press, New York. MR0403103

FEDOROV, V. V. and LEONOV, S. L. (2014). *Optimal Design for Nonlinear Response Models*. *Chapman & Hall/CRC Biostatistics Series*. CRC Press, Boca Raton, FL. MR3113978

GLONEK, G. and MCCULLAGH, P. (1995). Multivariate logistic models. *J. Roy. Statist. Soc. Ser. B* **57** 533–546.

ITEPAN, N. (1995). Aumento do periodo de aceitabilidade de pupas de Musca domestica L., 1758 (Diptera: Muscidae), irradiadas com raios gama, como hospedeiras de parasitoides (Hymenoptera: Pteromalidae). Master's thesis, Centro de Energia Nuclear na Agricultura/USP, Piracicaba, SP, Brazil.

JENNETT, B. and BOND, M. (1975). Assessment of outcome after severe brain damage. *Lancet* **305** 480–484.

KHURI, A. I., MUKHERJEE, B., SINHA, B. K. and GHOSH, M. (2006). Design issues for generalized linear models: A review. *Statist. Sci.* **21** 376–399. MR2339137 https://doi.org/10.1214/088342306000000105

KIEFER, J. (1974). General equivalence theory for optimum designs (approximate theory). *Ann. Statist.* **2** 849–879. MR0356386

LALL, R., CAMPBELL, M., WALTERS, S. and MORGAN, K. (2002). A review of ordinal regression models applied on health-related quality of life assessments. *Stat. Methods Med. Res.* **11** 49–67.

LIU, I. and AGRESTI, A. (2005). The analysis of ordered categorical data: An overview and a survey of recent developments. *TEST* **14** 1–73. MR2203424 https://doi.org/10.1007/BF02595397

MCCULLAGH, P. (1980). Regression models for ordinal data. *J. Roy. Statist. Soc. Ser. B* **42** 109–142. MR0583347

MCCULLAGH, P. and NELDER, J. A. (1989). *Generalized Linear Models. Monographs on Statistics and Applied Probability*. CRC Press, London. MR3223057 https://doi.org/10.1007/978-1-4899-3242-6

PEREVOZSKAYA, I., ROSENBERGER, W. F. and HAINES, L. M. (2003). Optimal design for the proportional odds model. *Canad. J. Statist.* **31** 225–235. MR2016228 https://doi.org/10.2307/3316068

PETERSON, B. and HARRELL, F. (1990). Partial proportional odds models for ordinal response variables. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **39** 205–217.

PUKELSHEIM, F. (1993). *Optimal Design of Experiments. Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, New York. MR1211416

RANDALL, J. (1989). The analysis of sensory data by generalised linear model. *Biom. J.* **31** 781–793.

SEBER, G. A. F. (2008). *A Matrix Handbook for Statisticians. Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR2365265

SILVEY, S. D. (1980). *Optimal Design*. CRC Press, London. MR0606742

STOICA, P. and MARZETTA, T. L. (2001). Parameter estimation problems with singular information matrices. *IEEE Trans. Signal Process.* **49** 87–90. MR1890585 https://doi.org/10.1109/78.890346

STOKES, M., DAVIS, C. and KOCH, G. (2012). *Categorical Data Analysis Using SAS*, 3rd ed. SAS Institute, Cary, NC.

STUFKEN, J. and YANG, M. (2012). Optimal designs for generalized linear models. In *Design and Analysis of Experiments. Volume* 3: *Special Designs and Applications* (K. Hinkelmann, ed.). *Wiley Ser. Probab. Stat.* 137–164. Wiley, Hoboken, NJ. MR2932061 https://doi.org/10.1002/9781118147634.ch4

WU, F.-C. (2008). Simultaneous optimization of robust design with quantitative and ordinal data. *Int. J. Ind. Eng. Theory, Appl. Pract.* **5** 231–238.

YANG, M., BIEDERMANN, S. and TANG, E. (2013). On optimal designs for nonlinear models: A general and efficient algorithm. *J. Amer. Statist. Assoc.* **108** 1411–1420. MR3174717 https://doi.org/10.1080/01621459.2013.806268

YANG, J. and MANDAL, A. (2015). D-optimal factorial designs under generalized linear models. *Comm. Statist. Simulation Comput.* **44** 2264–2277. MR3359169 https://doi.org/10.1080/03610918.2013.815773

YANG, J., MANDAL, A. and MAJUMDAR, D. (2016). Optimal designs for 2^k factorial experiments with binary response. *Statist. Sinica* **26** 385–411. MR3468357

YANG, J., TONG, L. and MANDAL, A. (2017). D-optimal designs with ordered categorical data. *Statist. Sinica* **27** 1879–1902. MR3726770

YEE, T. W. (2015). *Vector Generalized Linear and Additive Models*: *With an Implementation in R. Springer Series in Statistics*. Springer, New York. MR3408425 https://doi.org/10.1007/978-1-4939-2818-7

ZOCCHI, S. S. and ATKINSON, A. C. (1999). Optimum experimental designs for multinomial logistic models. *Biometrics* **55** 437–444. MR1705110 https://doi.org/10.1111/j.0006-341X.1999.00437.x