Statistica Sinica Preprint No: SS-2025-0075	
Title	D-Optimal Designs for Ordinal Response Experiments
Manuscript ID	SS-2025-0075
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0075
Complete List of Authors	Huiping Dang, Jun Yu and Fasheng Sun
Corresponding Authors	Fasheng Sun
E-mails	sunfs359@nenu.edu.cn

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0075

D-OPTIMAL DESIGNS FOR ORDINAL RESPONSE EXPERIMENTS

Huiping Dang[1], Jun Yu[2], Fasheng Sun[1]

KLAS and School of Mathematics and Statistics, Northeast Normal University[1],

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

Abstract: Ordinal response data are quite common in scientific experiments, and finding the optimal design for them is a challenging task. Adjacent-category models are widely used to model ordinal response data. In this paper, we study the D-optimal designs of adjacent-categories models with general link functions concerning both quantitative and qualitative factors. Some structure characteristics, including the number of support points and a simple complete class of the locally D-optimal design, are derived. Utilizing the obtained structure characteristics, an efficient algorithm is proposed to search out corresponding D-optimal designs. The integer-valued allocations for the corresponding D-optimal design are further discussed for practical implementation. Numerical examples show the advantages of the proposed design in both statistical efficiency and computational time.

Key words and phrases: Adjacent-categories models, complete class, D-optimal designs, ordinal response.

1. Introduction

Categorical data are common in scientific experiments. Several generalized linear models have been developed to characterize the relationship between an ordinal response and explanatory factors. Readers may refer to Agresti (2019) for a comprehensive review. For categorical responses with unordered scales, such as types of music (classical, country, folk, jazz, pop, rock), researchers usually compare the probability of the outcome being equal to k ($k > 1$, conditional on experimental setting $\boldsymbol{x}$) to a baseline category where the outcome equals 1. When we move to an ordinal response, such as the Glasgow Outcome Scale (death, vegetative state, major disability, minor disability, good recovery (Jennett and Bond, 1975)) in trauma clinical trial (Chuang-Stein and Agresti, 1997), we must decide what outcomes to compare. One of the most reasonable models is to compare each response with its next or adjacent level, which is known as the adjacent-categories (AC) models (O'Connell, 2006).

It is well known that meticulously designed experiments can help augment the predictive power of a statistical model (Smucker et al., 2018; Gevertz and Kareva, 2024). However, how to collect data which can maximize its utility for model estimation or prediction is non-trivial. Planning an experiment for categorical outcomes is still in its infancy, especially for

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0075

the scenario that there are more than two categories in the response. The difficulty comes from the complicated structure of the optimization problem to obtain the maximum statistical efficiency via deciding the experimental setups (aka. the supports) and the number of runs under each setup (aka. the weights).

More precisely, when the inputs include both qualitative and quantitative factors, researchers not only need to find the optimal weights but also need to decide the possible supports among the infinite candidate setups in the design region. Besides the complex experimental inputs, the difficulty also comes from the complicated objective function which depends on the structure of the information matrix. One major obstacle is that the optimal design depends not only on the unknown parameters introduced by the nonlinearity, but also on the choice of link functions, and different parametrizations.

Due to the aforementioned reasons, research on optimal designs for AC models remains limited. To our best knowledge, existing works have predominantly focused on the AC model with the logit link. The pioneering work may date back to the optimal design for the logistic regression, such as Yang et al. (2011). Bu et al. (2020) extended the results to multiple categories under the logit link and discrete design region. Hao and Yang

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0075

(2020) considered the design problem for AC model with one continuous design factor and three-category responses. Huang et al. (2024) developed the ForLion algorithm to find locally D-optimal approximate designs for AC model numerically. It is worth mentioning that AC model is not the only model for the categorical responses. Baseline model, cumulative model and continuation ratio model are also widely adopted in practice. Some impressive progress has been achieved for the three models. Typical examples include but are not limited to Zocchi and Atkinson (1999), Perevozskaya et al. (2003), Yang et al. (2017), Bu et al. (2020) and Ai et al. (2023).

Our contributions. In this work, we study the locally optimal design for AC model with general link functions on both discrete and continuous design regions, which extends the scope of application. We carefully study the characteristics of the design structure and find a simple complete class for this complex problem. Compared with the existing results given by Yang et al. (2011) and Hao and Yang (2020), our method efficiently reduces the cost in altering the experimental settings without losing efficiency. An efficient algorithm is proposed to search out the locally D-optimal designs and the D-optimal integer-valued allocations. Numerical examples are provided to demonstrate the superior performance of the obtained optimal designs.

The rest of this paper is organized as follows. In Section 2, the AC

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0075

models with general link functions are reviewed and the explicit form of the Fisher information matrix is derived. Section 3 derives some structure characteristics, including the number of support points and a simple complete class of the locally D-optimal design. Section 4 gives algorithms for searching out the locally D-optimal designs over discrete and continuous design regions. Section 5 further illustrates our approach with several examples. Section 6 concludes this paper.

2. Statistical model and design criterion

2.1 The adjacent-categories models

Suppose $j \in \{1, \ldots, J\}$ is the categorical response of an experiment under the setting $\boldsymbol{x}$, where $\boldsymbol{x}$ is q-dimensional. After transforming j into a J-dimensional vector $\boldsymbol{Y}$ with the jth element Y_j being one and the rest being zero, the response $\boldsymbol{Y}$ can be characterized by the following multinomial distribution

$$\Pr\left(\boldsymbol{Y}|\boldsymbol{x}\right) = \pi_1\left(\boldsymbol{x}\right)^{Y_1} \cdots \pi_J\left(\boldsymbol{x}\right)^{Y_J}, \tag{2.1}$$

where $\pi_j\left(\boldsymbol{x}\right)$ denotes the probability that the response is j under experimental setting $\boldsymbol{x}$. Assume the design region $\mathcal{X} = \prod_{t=1}^{q} I_t$ with I_t being a finite set or a compact interval ranging from U_t to V_t for $1 \leq t \leq q$. To compare the jth level of the response with its next or adjacent level, the general AC proportional odds (po) model (McCullagh, 1980; Bu et al., 2020) assumes that the contrast $\pi_j(\boldsymbol{x})/(\pi_j(\boldsymbol{x})+\pi_{j+1}(\boldsymbol{x}))$ has a nonlinear relationship with $\theta_j + \boldsymbol{x}^\top\boldsymbol{\beta}$, associated with a link function $g(\cdot)$ for $j = 1, \ldots, J-1$. More precisely, for the ith design unit,

$$g\left(\frac{\pi_j(\boldsymbol{x}_i)}{\pi_j(\boldsymbol{x}_i)+\pi_{j+1}(\boldsymbol{x}_i)}\right) = \eta_{ij} = \theta_j + \boldsymbol{x}_i^\top\boldsymbol{\beta},\ j = 1, \ldots, J-1. \tag{2.2}$$

Here the levels of discrete factors are considered to be numerical and only the main effects of factors are considered.

Let $\boldsymbol{\eta}_i = (\eta_{i1}, \ldots, \eta_{i,J-1})^\top$ and denote $p = J-1+q$. Model (2.2) can be represented by $\boldsymbol{\eta}_i = H(\boldsymbol{x}_i)\boldsymbol{\gamma}$, where $H(\boldsymbol{x}_i)$ is $(J-1)\times p$ design matrix with

$$H(\boldsymbol{x}_i) = \begin{pmatrix} 1 & 0 & \cdots & 0 & \boldsymbol{x}_i^\top \\ 0 & 1 & \cdots & 0 & \boldsymbol{x}_i^\top \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \boldsymbol{x}_i^\top \end{pmatrix},$$

and $\boldsymbol{\gamma} = (\theta_1, \ldots, \theta_{J-1}, \boldsymbol{\beta}^\top)^\top$ is a vector of unknown parameters. Let $\Theta \subset \mathbb{R}^p$ be the parameter space. Clearly, when $g(\cdot)$ takes the form $g(t) = \log(t/(1-t))$, AC model (2.2) turns into the multinomial logit model under the adjacent-categories framework (Agresti, 2019). To ensure that the model is well defined, we need the following regularity assumptions mentioned in McCullagh and Nelder (1989) throughout this paper.

Assumption 1. The parameter space $\Theta \subset \mathbb{R}^p$ is compact.

Assumption 2. The link function $g(\cdot)$ is differentiable and its derivative $g'(\cdot) > 0$, $g^{-1}(\eta)$ is well-defined at each $\eta \in (-\infty, \infty)$.

Assumption 2 ensures $g(\cdot)$ and $g^{-1}(\cdot)$ are injective functions. Thus, in Model (2.2), $\pi_j(\boldsymbol{x}_i)/\left(\pi_j(\boldsymbol{x}_i) + \pi_{j+1}(\boldsymbol{x}_i)\right)$ can be represented by $g^{-1}\left(\theta_j + \boldsymbol{x}_i^\top \boldsymbol{\beta}\right)$, for $i = 1, \ldots, m$, $j = 1, \ldots, J-1$, and denote it as $g_j(\boldsymbol{x}_i)$. With the fact that $\pi_1(\boldsymbol{x}_i) + \cdots + \pi_J(\boldsymbol{x}_i) = 1$, it is clear to see that $\pi_1(\boldsymbol{x}_i), \ldots, \pi_J(\boldsymbol{x}_i)$ are completely determined by the link function $g(\cdot)$, experiment setup $\boldsymbol{x}_i$, and the parameters $\boldsymbol{\gamma}$.

2.2 Fisher information matrix and D-optimality criterion

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. Here $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ are m distinct points which lie in the design region $\mathcal{X}$, and $\omega_i > 0$ for $i = 1, \ldots, m$, with $\omega_1 + \cdots + \omega_m = 1$.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0075

One can transform the exact design with N_i runs on the setting $\boldsymbol{x}_i$ to the approximate design by letting $\omega_i = N_i/N$ with $N = \sum_{i=1}^m N_i$.

By standard statistical theory (Bu et al., 2020), the Fisher information matrix has the formula

$$M(\xi) = \sum_{i=1}^{m} \omega_i M(\boldsymbol{x}_i),$$

where $M(\boldsymbol{x}_i) = (\partial \boldsymbol{\pi}(\boldsymbol{x}_i)/\partial \boldsymbol{\gamma}^\top)^\top (\operatorname{diag} \boldsymbol{\pi}(\boldsymbol{x}_i))^{-1} (\partial \boldsymbol{\pi}(\boldsymbol{x}_i)/\partial \boldsymbol{\gamma}^\top)$, and $\boldsymbol{\pi}(\boldsymbol{x}_i) = (\pi_1(\boldsymbol{x}_i), \ldots, \pi_J(\boldsymbol{x}_i))^\top$ defined by Model (2.2). Here $\operatorname{diag}(\cdot)$ denotes a diagonal matrix with corresponding diagonal elements. To take a close look at how the design affects the Fisher information matrix, we start with introducing the following notations. For $\boldsymbol{x} \in \mathcal{X}$, define $\Delta_1(\boldsymbol{x}) = \Delta_2(\boldsymbol{x}) = 1$, $\prod_{t=l}^{k}(1 - g_t(\boldsymbol{x})) = 1$ and $\prod_{t=l}^{k} g_t(\boldsymbol{x}) = 1$ for $l > k$, and for $k > 2$

$$\begin{aligned}
\Delta_k(\boldsymbol{x}) =& \prod_{t=1}^{k-2}(1 - g_t(\boldsymbol{x})) + \Delta_{k-2}(\boldsymbol{x}) g_{k-2}(\boldsymbol{x}) g_{k-1}(\boldsymbol{x}) \\
=& \prod_{t=1}^{k-1}(1 - g_t(\boldsymbol{x})) + \prod_{t=1}^{k-3}(1 - g_t(\boldsymbol{x})) g_{k-1}(\boldsymbol{x}) + \left[\Delta_{k-2}(\boldsymbol{x}) - \prod_{t=1}^{k-3}(1 - g_t(\boldsymbol{x}))\right] \\
& \times g_{k-2}(\boldsymbol{x}) g_{k-1}(\boldsymbol{x}).
\end{aligned}$$

Simple calculation yields that

$$\Delta_J(\boldsymbol{x}) = (\pi_2(\boldsymbol{x}) \times \cdots \times \pi_{J-1}(\boldsymbol{x})) / ((\pi_1(\boldsymbol{x}) + \pi_2(\boldsymbol{x})) \times \cdots \times (\pi_{J-1}(\boldsymbol{x}) + \pi_J(\boldsymbol{x}))),$$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0075

which implies

$$\pi_j(\boldsymbol{x}) = \prod_{t=1}^{j-1}(1-g_t(\boldsymbol{x}))\prod_{t=j}^{J-1} g_t(\boldsymbol{x})\Delta_J^{-1}(\boldsymbol{x}),\ j=1,\ldots,J.$$

Let $D(\boldsymbol{x}) = \text{diag}\,(\pi_1(\boldsymbol{x}),\ldots,\pi_J(\boldsymbol{x}))$ and $U(\boldsymbol{x}) = (u_{lj}(\boldsymbol{x}))$ be a $J\times(J-1)$ matrix with (l,j)th element denoting by $u_{lj}(\boldsymbol{x})$. For $j=1,\ldots,J-1$,

$$u_{1j}(\boldsymbol{x}) = \sum_{k=1}^{\lceil\frac{J}{2}\rceil}\left(\mathbb{I}(j<J-2(k-1))\prod_{t=1,t\neq j}^{J-2k}(1-g_t(\boldsymbol{x}))\prod_{t=J-2(k-1)}^{J-1} g_t(\boldsymbol{x})\right) \times g_j'(\boldsymbol{x})\prod_{t=1,t\neq j}^{J-1} g_t(\boldsymbol{x})\Delta_J^{-2}(\boldsymbol{x}),$$

where $\mathbb{I}(\cdot)$ denotes the indicator function.

For $l=2,\ldots,J$ and $j=1,\ldots,J-1$,

$$u_{lj}(\boldsymbol{x}) = \begin{cases} -\prod_{t=1}^{j-1}(1-g_t(\boldsymbol{x}))\Delta_j(\boldsymbol{x})\left(\prod_{t=j+1}^{J-1} g_t(\boldsymbol{x})\right) g_j'(\boldsymbol{x})\prod_{t=j+1}^{J-1} g_t(\boldsymbol{x})\Delta_J^{-2}(\boldsymbol{x}), \\ l=j+1, \\ u_{l-1,j}(\boldsymbol{x})\frac{1-g_{l-1}(\boldsymbol{x})}{g_{l-1}(\boldsymbol{x})},\ l\neq j+1. \end{cases}$$

After defining $H(\boldsymbol{x})$, $D(\boldsymbol{x})$ and $U(\boldsymbol{x})$, the Fisher information matrix for Model (2.2) under a design ξ is given 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*

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0075

$$M(\xi) = \sum_{i=1}^{m} \omega_i M(\boldsymbol{x}_i) = \sum_{i=1}^{m} \omega_i H^\top(\boldsymbol{x}_i) U^\top(\boldsymbol{x}_i) D^{-1}(\boldsymbol{x}_i) U(\boldsymbol{x}_i) H(\boldsymbol{x}_i). \tag{2.4}$$

The Fisher information matrix $M(\xi)$ plays a key role in optimal design theory. A commonly used design criterion in applications is that of D-optimality, which maximizes the determinant of the Fisher information matrix $M(\xi)$. In the language of statistics, a D-optimal design minimizes the volume of the confidence ellipsoid of the maximum likelihood estimator of $\boldsymbol{\gamma}$.

3. Some structure characteristics of D-optimal design

To obtain the optimal design, we propose to identify a subclass of designs called complete class (Stufken and Yang, 2012) with a simple format so that for any given design ξ, there exists a design $\tilde{\xi}$ in that subclass with $M(\tilde{\xi}) \geq M(\xi)$, that is, the information matrix under $\tilde{\xi}$ dominates that under ξ in the Loewner ordering. To ease the presentation, we begin with the case where only one continuous quantitative factor lies in $\mathcal{X}$.

To obtain the relevant results, we give the definition of the Chebyshev system, more details can be found in Karlin and Studden (1966).

Definition 1. A set of $k+1$ continuous functions $u_0, \ldots, u_k : [U, V] \to \mathbb{R}$

doi:10.5705/ss.202025.0075

is called a Chebyshev system on $[U, V]$ if the inequality

$$\begin{vmatrix} u_0(x_0) & u_0(x_1) & \ldots & u_0(x_k) \\ u_1(x_0) & u_1(x_1) & \ldots & u_1(x_k) \\ \vdots & \vdots & \ddots & \vdots \\ u_k(x_0) & u_k(x_1) & \ldots & u_k(x_k) \end{vmatrix} > 0$$

holds for all $U \le x_0 < x_1 < \cdots < x_k \le V$.

Recall that the matrix $M(\xi)$ corresponds to the design ξ with only one continuous factor only and J response categories can be represented by

$$M(\xi) = \begin{pmatrix} \int_{\mathcal{X}} \Psi_{11}(x) d\xi(x) & \cdots & \int_{\mathcal{X}} \Psi_{1p}(x) d\xi(x) \\ \vdots & \ddots & \vdots \\ \int_{\mathcal{X}} \Psi_{1p}(x) d\xi(x) & \cdots & \int_{\mathcal{X}} \Psi_{pp}(x) d\xi(x) \end{pmatrix}$$

and $\Psi_{11}, \Psi_{12}, \ldots, \Psi_{pp}$ are functions defined on $\mathcal{X}$. Denote by $\Psi_1, \ldots, \Psi_K$ the distinct elements among the functions $\{\Psi_{ij} \mid 1 \le i \le j \le p\}$, which are not equal to the constant function. Assume $\Psi_K = \Psi_{ll}$ for some $l \in \{1, \ldots, p\}$ and $\Psi_{ij} \ne \Psi_K$ for all $(i, j) \ne (l, l)$. Note that $K \le p(p+1)/2$.

Following Theorem 3.1 of Dette and Melas (2011), one can show the following result.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0075

Lemma 1. *For model (2.2) with only one continuous factor and the functions* $\Psi_0(x) = 1$, $\Psi_1, \ldots, \Psi_{K-1}, \Psi_K$, *suppose that either* $\{\Psi_0, \Psi_1, \ldots, \Psi_{K-1}\}$ *and* $\{\Psi_0, \Psi_1, \ldots, \Psi_{K-1}, \Psi_K\}$ *are Chebyshev systems or* $\{\Psi_0, \Psi_1, \ldots, \Psi_{K-1}\}$ *and* $\{\Psi_0, \Psi_1, \ldots, \Psi_{K-1}, -\Psi_K\}$ *are Chebyshev systems, then for any design* ξ *there exists a design* $\tilde{\xi}$ *with at most* $(K+2)/2$ *support points, such that* $M(\tilde{\xi}) \geq M(\xi)$.

The conditions on the $\Psi_1, \ldots, \Psi_K$ in Lemma 1 impose more constraints on the link functions. To ease the presentation, we say an AC model with the link function satisfies the conditions in Lemma 1, if and only if the information matrix of model (2.2) with only one continuous factor and same link function meets the conditions in Lemma 1. Specifically, when Model (2.2) adopts the logit link function, such conditions are naturally satisfied. Thus we have the following result.

Corollary 1. *For model (2.2) under the logit link function, with only one continuous factor and* J *response categories, designs with at most* $2(J-1)$ *support points form a complete class.*

For the multi-factor design, we rewrite the support points $\boldsymbol{s}_i = (x_{i1}, \ldots, x_{i,q-1}, c_i)^\top$, where $c_i = \sum_{t=1}^{q} \beta_t x_{it}$ and $\beta_t \neq 0$ for all possible t. Note that the transformation does not change the complete class result. To ease the presentation, we assume the design region $\mathcal{X} = \prod_{t=1}^{q-1}[-1, 1] \times [U, V]$, where

$[U, V]$ is the relatively large interval that must contain all c_i. One can map -1 to U_k and $+1$ to V_k for the kth factor ($k = 1, \dots, q-1$) to obtain the design for the problem introduced in Section 2.1.

Let $A(c) = U^\top(c)D^{-1}(c)U(c) = U^\top(\boldsymbol{x})D^{-1}(\boldsymbol{x})U(\boldsymbol{x})$, since $U(\boldsymbol{x})$ and $D(\boldsymbol{x})$ are related only to $\boldsymbol{x}^\top\boldsymbol{\beta}$, i.e., c when the parameters are given. Denote ξ_s as the approximate design on the transformed design space. Then $M(\xi_s) = \sum_{i=1}^m \omega_i H^\top(\boldsymbol{s}_i)A(c_i)H(\boldsymbol{s}_i)$.

The following theorem introduces a structured complete class construction for multi-factor design and demonstrates that an arbitrary design ξ_s can be dominated by a design $\tilde{\xi}_s$.

Theorem 2. *For the transformed design space of AC model with at least 1 continuous factor and where the link function satisfies the conditions in Lemma 1, then for an arbitrary design* $\xi_s = \{(\boldsymbol{s}_i, \omega_i),\ i = 1, \dots, m;\ \sum_{i=1}^m \omega_i = 1\}$, *there exists a design* $\tilde{\xi}_s$ *such that the following inequality for information matrices hold:* $M(\xi_s) \le M(\tilde{\xi}_s)$, *where*

$$\tilde{\xi}_s = \left\{(\tilde{\boldsymbol{s}}_{il}, \tilde{\omega}_{il}), i = 1, \dots, (K+2)/2, l = 1, \dots, 2^{q-1}\right\},$$

and $\tilde{\boldsymbol{s}}_{il} = (b_{l1}, \dots, b_{l,q-1}, \tilde{c}_i)$. *Here* $b_{lj} = -1$ *or* 1, *and* $(b_{l1}, \dots, b_{l,q-1})$, $l = 1, \dots, 2^{q-1}$ *are all combinations of them.* $\tilde{c}_i$ *in* $\xi_c = \{(\tilde{c}_i, \tilde{\omega}_i), i = 1, \dots, (K+2)/2, \sum_{i=1}^{(K+2)/2} \tilde{\omega}_i = 1\}$ *are* $(K+2)/2$ *numbers that need to be solved.*

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0075

For an approximate design ξ given in (2.3), a key characteristic is the number of the support points, say m in (2.3). When m is too small, the resultant design can not guarantee parameter identifiability (i.e., permitting confidence in model predictions). Oppositely, for a too large m, the cost in changing the experimental settings may be unaffordable.

The following theorem provides lower and upper bounds for the number of support points in a design.

Theorem 3. *The Fisher information matrix $M(\xi)$ calculated in Equation (2.4) is positive definite only if $m \geq q+1$. Furthermore, there exists a D-optimal design with $m \leq p(p+1)/2$.*

Theorem 3 must hold as long as design region $\mathcal{X}$ is compact. When changing the experimental setting is too expensive, it is of a special value to find D-optimal design within the class where the number of support points is the smallest possible integer such that $|M(\xi)| > 0$. This is known as a minimally supported D-optimal design. To obtain such a design, we begin with characterizing the form of $|M(\xi)|$.

Following Theorem 2 in Yang et al. (2017), given a map $\tau : \{1, 2, \ldots, p\} \rightarrow \{1, 2, \ldots, m\}$, M_τ is a $p \times p$ matrix whose tth row is the same as the tth

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0075

row of $M(\boldsymbol{x}_{\tau(t)})$, $t = 1, 2, \ldots, p$. Define

$$\Phi\left(\alpha_1, \ldots, \alpha_m\right) = \left\{ \tau \middle| \sum_{t=1}^{p} I_{\{i\}}(\tau(t)) = \alpha_i, i = 1, \ldots, m \right\},$$

where $I_{\{i\}}(\tau(t)) = 1$ if $\tau(t) = i$, otherwise $I_{\{i\}}(\tau(t)) = 0$, and $\alpha_1, \ldots, \alpha_m \in \mathbb{N}$ with $\alpha_1 + \cdots + \alpha_m = p$. Armed with the aforementioned notation, $|M(\xi)|$ can be expressed as an order-p homogeneous polynomial of $\omega_1, \ldots, \omega_m$.

Theorem 4. *The determinant of the Fisher information matrix $M(\xi)$ is*

$$|M(\xi)| = \sum_{\substack{\alpha_1 \geq 0, \ldots, \alpha_m \geq 0 \\ \alpha_1 + \cdots + \alpha_m = p}} c_{\alpha_1, \ldots, \alpha_m} \omega_1^{\alpha_1} \cdots \omega_m^{\alpha_m}, \tag{3.1}$$

where

$$c_{\alpha_1, \ldots, \alpha_m} = \sum_{\tau \in \Phi(\alpha_1, \ldots, \alpha_m)} |M_\tau|. \tag{3.2}$$

Furthermore, let $n = \sum_{i=1}^{m} I\{\alpha_i > 0\}$, where $I\{\alpha_i > 0\}$ is 1 if $\alpha_i > 0$ is true, and 0 otherwise. Then the coefficients (3.2) are zero if the $(\alpha_1, \ldots, \alpha_m)$ satisfies one of the following conditions.

(1) $\max_{1 \leq i \leq m} \alpha_i \geq J$. *(2) $n \leq q$.*

For the discrete design region, finding an optimal design is a standard convex optimization problem of an order-p homogeneous polynomial of the weights $\omega_1, \ldots, \omega_m$. Thus the related subclass only contains the optimal

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0075

design itself.

Clearly, when all factors are quantitative, the design region $\mathcal{X}$ is convex, finding the D-optimal design with minimal support turns out to be a bi-concave problem. More precisely, fixed the $q+1$ support points, (3.1) is a strictly concave optimization problem with respect to $(\omega_1, \ldots, \omega_{q+1})$. When fixed $q+1$ weights, (3.1) is also a concave function with respect to $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{q+1})$. Thus, one can resort to coordinate descent (Stephen, 2015) or alternating direction method of multipliers (Ouyang et al., 2015) algorithms to solve the problem.

Another advantage of the minimally supported designs is that the optimal weights restricted to $q+1$ support points can be obtained more easily or even analytically.

Example 1. Consider the following model

$$\log\left(\frac{\pi_j(x_i)}{\pi_{j+1}(x_i)}\right) = \alpha_j + \beta x_i,\ j = 1, 2. \tag{3.3}$$

According to Theorem 3, the minimally supported design for this model has two different design points.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0075

The D-optimal design on $\{x_1, x_2\}$ of model (3.3) is

$$\xi = \begin{pmatrix} x_1 & x_2 \\ \frac{b_1 - b_2 + \sqrt{b_1^2 - b_1 b_2 + b_2^2}}{2b_1 - b_2 + \sqrt{b_1^2 - b_1 b_2 + b_2^2}} & \frac{b_1}{2b_1 - b_2 + \sqrt{b_1^2 - b_1 b_2 + b_2^2}} \end{pmatrix},$$

where $b_1 = (\pi_1(x_2)\pi_2(x_2) + 4\pi_1(x_2)\pi_3(x_2) + \pi_2(x_2)\pi_3(x_2))\pi_1(x_1)\pi_2(x_1)\pi_3(x_1) \times (x_1 - x_2)^2$, $b_2 = (\pi_1(x_1)\pi_2(x_1) + 4\pi_1(x_1)\pi_3(x_1) + \pi_2(x_1)\pi_3(x_1))\pi_1(x_2)\pi_2(x_2) \times \pi_3(x_2)(x_1 - x_2)^2$.

Therefore, finding the locally D-optimal design becomes a bi-variate optimization problem over the design region $\mathcal{X}$.

When the costs of changing experimental settings are negligible, the D-optimal design with minimal support may lead to a relatively low efficiency compared with the D-optimal design with more than $q+1$ supports. Despite the results in Theorem 3 provide the upper bound of m, there are still $p(p+1)/2$ supports and $p(p+1)/2 - 1$ weights that need to be optimized. This will also lead to a complex optimization problem, especially for the large p.

In the following, we refine the results in Theorem 2, to further reduce the cost of altering the experimental settings. Before presenting our results, we introduce the concept of orthogonal array (OA) which is the main ingredient of our results. Consider the matrix of n runs with m factors of s levels, where $2 \le s \le n$. If for every $n \times t$ submatrix of the matrix, say $\mathbf{R}$, all possible level combinations appear equally often, the matrix $\mathbf{R}$ is called an orthogonal array of strength t (Hedayat et al., 1999). We use $\mathrm{OA}(n, m, s, t)$ to denote such a matrix. A column is said to be balanced if each level appears equally often.

Theorem 5. *For an arbitrary design $\tilde{\xi}_s$ in Theorem 2, $\tilde{\xi}_s^*$ is D-optimal under $\xi_c = \{(\tilde{c}_i, \tilde{\omega}_i), i = 1, \ldots, (K+2)/2, \sum_{i=1}^{(K+2)/2} \tilde{\omega}_i = 1\}$ if*

$$\tilde{\xi}_s^* = \left\{ \left(\tilde{\boldsymbol{s}}_{il}, \tilde{\omega}_i / 2^{q'-1} \right), i = 1, \ldots, (K+2)/2, l = 1, \ldots, 2^{q'-1} \right\},$$

where $q' \le q$ and $\tilde{\boldsymbol{s}}_{il} = (b_{l1}, \ldots, b_{l,q-1}, \tilde{c}_i)$. Here $b_{lj} = -1$ or 1, and $b_{l1}, l = 1, \ldots, 2^{q'-1}$ is balanced when $q = 2$, $(b_{l1}, \ldots, b_{l,q-1})$, $l = 1, \ldots, 2^{q'-1}$ form an $\mathrm{OA}(2^{q'-1}, q-1, 2, 2)$ when $q > 2$.

According to Theorem 5, $M(\tilde{\xi}_s^*) = \sum_{i=1}^{(K+2)/2} \sum_{l=1}^{2^{q'-1}} (\tilde{\omega}_i / 2^{q'-1}) H(\tilde{\boldsymbol{s}}_{il})^\top A(\tilde{c}_i) H(\tilde{\boldsymbol{s}}_{il})$ only relevant to $\tilde{c}_i$ and $\tilde{\omega}_i$, i.e., ξ_c. Because for the design point $\tilde{\boldsymbol{s}}_{il} = (b_{l1}, \ldots, b_{l,q-1}, \tilde{c}_i)$, e.g., $\tilde{\boldsymbol{s}}_{il} = (1, -1, -1, \tilde{c}_i)$ and $J = 3$,

$$H(\tilde{\boldsymbol{s}}_{il}) = \begin{pmatrix} 1 & 0 & 1 & -1 & -1 & \tilde{c}_i \\ 0 & 1 & 1 & -1 & -1 & \tilde{c}_i \end{pmatrix}.$$

To get the D-optimal design ξ_s^*, we just need to find the design ξ_c^* that maximizes $\Phi(\xi_c) = |M(\tilde{\xi}_s^*)|$.

According to Theorems 2 and 5, finding the D-optimal design with multiple factors also turns out to be a bi-concave problem. Therefore, it can be solved by the method of the bi-concave problem mentioned after Theorem 4, except that $\{(\boldsymbol{x}_i, \omega_i), i = 1, \ldots, q+1, \sum_{i=1}^{q+1} \omega_i = 1\}$ is changed to $\{(\tilde{c}_i, \tilde{\omega}_i), i = 1, \ldots, (K+2)/2 \text{ (at most)}, \sum_{i=1}^{(K+2)/2} \tilde{\omega}_i = 1\}$.

It is worth noting that the structure covers the results in Hao and Yang (2020) and Yang et al. (2011) which studied D-optimal designs for the AC model with logit link. To be precise, Hao and Yang (2020) showed that the $(b_{l1}, \ldots, b_{l,q-1})$ takes all possible value among 2^{q-1} full factorial design. Theorem 5 improves the result by generalizing the 2^{q-1} full factorial design to an arbitrary orthogonal array of strength two. From Yang et al. (2011), it is clear to see the structure in ξ^* is naturally satisfied for the optimal design of logistic regression. The benefits of reducing the cost of altering the experimental settings are obvious. More precisely, when $q = 6$, we need to conduct design on $16(K+2)$ different settings according to Hao and Yang (2020) while our approach only need $4(K+2)$ different settings. This demonstrates that our method is more flexible for varying design run sizes.

4. Algorithm for D-optimal Design

Despite the structure characteristics presented in Section 3, it remains to show the concrete algorithm in finding the corresponding D-optimal design. In this section, we will show how the D-optimal design can be constructed with the assistance of the aforementioned structure information.

According to Theorem 4, if the support points of the design, say ξ defined in (2.3), are pre-specified, the determinant of the Fisher information matrix can be rewritten as $|M(\xi)| = f(\mathbf{w}) = f(\omega_1, \ldots, \omega_m)$, which is a polynomial function of $\mathbf{w}$. The following theorem states that the optimal allocation must exist under some mild conditions.

Theorem 6. *When m design points satisfy the condition of Theorem 3, the locally D-optimal design ξ^* that maximizes $|M(\xi)|$ must exist, i.e., $\mathbf{w}^*$ that maximizes $f(\mathbf{w})$ must exist.*

Although Theorem 6 presents the existence of the optimal design under the scenario that the design supports are pre-specified, further steps are required to identify the design. One can resort to classical optimization tools in maximizing $f(\mathbf{w})$ for qualitative factor experiments. However, when some quantitative factors are taken into account, we need additional effort to find the possible support points. To achieve this goal, one of the powerful

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0075

tools is the equivalence theorem (Silvey, 1980). Denote the Fréchet derivate of $\log|M(\xi)|$ at ξ in the direction of $\boldsymbol{x}$ by $\phi(\boldsymbol{x},\xi)$. Direct calculation yields that

$$\begin{aligned}
\phi(\boldsymbol{x},\xi) &= \lim_{\varepsilon\to 0}\left(\log|(1-\varepsilon)M(\xi)+\varepsilon M(\boldsymbol{x})| - \log|M(\xi)|\right)/\varepsilon \\
&= \lim_{\varepsilon\to 0}\left(\log|M(\xi)+\varepsilon(M(\boldsymbol{x})-M(\xi))| - \log|M(\xi)|\right)/\varepsilon \\
&= \operatorname{tr}\left(M^{-1}(\xi)\left(M(\boldsymbol{x})-M(\xi)\right)\right) \\
&= \operatorname{tr}\left(M^{-1}(\xi)\,M(\boldsymbol{x})\right) - p.
\end{aligned}$$

Based on $\phi(\boldsymbol{x},\xi)$, we can present the general equivalence theorem in searching out a D-optimal design for AC model.

Theorem 7. *The following three conclusions are equivalent:*

(1) ξ^* *is the D-optimal design, i.e.,* $|M(\xi^*)| = \max_\xi |M(\xi)|$;

(2) $\forall\ \boldsymbol{x} \in \mathcal{X}$, $\phi(\boldsymbol{x},\xi^*) \le 0$;

(3) $\phi(\boldsymbol{x},\xi^*)$ *is maximized at each design point* $\boldsymbol{x}$ *of* ξ^*, *and* $\phi(\boldsymbol{x},\xi^*) = 0$.

Armed with the equivalence theorem, we obtain a stopping rule in searching out an optimal design algorithmically. Thanks to the structure information presented in Theorems 2 and 5, the algorithm for constructing a multi-factor D-optimal design can be described by the following three steps. First, find the single-factor optimal design ξ_c^* defined after Theorem 5. A lot algorithms such as ForLion (Huang et al., 2024) can be applied here.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0075

Compared with the original multi-factor problem, this step is much simpler and can be solved more quickly. Then, the optimal design ξ^* can be obtained by inversely solving $\xi_s^* = \xi_{2^{q'-1}} \otimes \xi_c^*$ on the transformed design space, where $\xi_{2^{q'-1}}$ is a uniform measure with equal mass supported on the points of the $2^{q'-1}$ fractional factorial design. The last step is to convert the approximate design ξ^* to the exact design. In contrast to the classical rounding procedure, we suggest using an additional lift-one algorithm to obtain integer-valued allocations. Similar idea has also been adopted in Huang et al. (2025b) for D-optimal designs under Multinomial Logit models. A step-by-step algorithm is presented in Algorithm 1.

Algorithm 1 Multi-factor D-optimal design algorithm

- **1** Run Algorithm 2 to obtain the optimal design ξ_c^*.
- **2** Construct D-optimal design ξ^* step: $\xi^* = \{(\boldsymbol{x}_{il}, \omega_{il}),\ i = 1, \ldots, m_t,\ l = 1, \ldots, 2^{q'-1};\ \sum_{i=1}^{m_t} \sum_{l=1}^{2^{q'-1}} \omega_{il} = 1\}$, where $\boldsymbol{x}_{il} = (b_{l1}, \ldots, b_{l,q-1}, x_{l,q})$, $b_{lj} = -1$ or 1, $j = 1, \ldots, q-1$ and b_{l1}, $l = 1, \ldots, 2^{q'-1}$ are balanced when $q = 2$, $(b_{l1}, \ldots, b_{l,q-1})$, $l = 1, \ldots, 2^{q'-1}$ form an $\text{OA}(2^{q'-1}, q-1, 2, 2)$ when $q > 2$, $x_{l,q} = (c_i - \sum_{j=1}^{q-1} \beta_j b_{lj})/\beta_q$, $\omega_{il} = \omega_i / 2^{q'-1}$.
- **3** Round-off step: Fix the design points, let $N_i = \lfloor N\omega_i \rfloor$, the largest integer no more than $N\omega_i$, for $i = 1, \ldots, m (m = m_t \times 2^{q'-1})$, and $k = N - \sum_{i=1}^{m} N_i$.
  - **3.1** Calculate $d_i = f(N_1, \ldots, N_{i-1}, N_i + 1, N_{i+1}, \ldots, N_m)$ for $i = 1, \ldots, m$.
  - **3.2** Pick up any $i \in \text{argmax}_{i \in \{1,\ldots,m\}} d_i$.
  - **3.3** Let $N_i \leftarrow N_i + 1$ and $k \leftarrow k - 1$.
  - **3.4** Repeat steps 3.1 $\sim$ 3.3 until $k = 0$.
- **4** Output: Report ξ^* and $\mathbf{N} = (N_1, \ldots, N_m)^\top$ as the D-optimal design and the D-optimal integer-valued allocations, respectively.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0075

Algorithm 2 D-Optimal design algorithm

- **1** Initialization: Set $t = 0$, arbitrarily choose $\xi_{c,0} = \{(c_i, \omega_i),\ i = 1, 2;\ \sum_{i=1}^{2} \omega_i = 1\}$, $\varepsilon > 0$ be the predefined tolerance and N be the total number of samples. Define $\mathbf{c} = (c_1, \ldots, c_m)^\top$, $\mathbf{w} = (\omega_1, \ldots, \omega_m)^\top$ and the determinant of the Fisher information matrix corresponding to design $\xi_c = \{(c_i, \omega_i),\ i = 1, \ldots, m;\ \sum_{i=1}^{m} \omega_i = 1\}$ is $f(\mathbf{c}, \mathbf{w})$.
- **2** Optimize c step: For $\xi_{c,t}$, let m_t be the length of $\mathbf{c}_t$ and $\mathbf{w}_t = (\omega_1, \ldots, \omega_{m_t})^\top = (m_t^{-1}, \ldots, m_t^{-1})^\top$.
  - **2.1** Set up a random order of i going through $\{1, 2, \ldots, m_t\}$. For each i, do steps 2.2 $\sim$ 2.3.
  - **2.2** Denote $\mathbf{c}_t^i(y) = (c_1, \ldots, c_{i-1}, y, c_{i+1}, \ldots, c_{m_t})^\top$, and $f_i(y, \mathbf{w}_t) = f(\mathbf{c}_t^i(y), \mathbf{w}_t)$, $y \in \mathcal{X}$.
  - **2.3** Use an analytic solution or the quasi-Newton algorithm to find y^* maximizing $f_i(y, \mathbf{w}_t)$. Define $\mathbf{c}_t^{i*} = \mathbf{c}_t^i(y^*)$. If $f(\mathbf{c}_t^{i*}, \mathbf{w}_t) > f(\mathbf{c}_t, \mathbf{w}_t)$, replace $\mathbf{c}_t$ with $\mathbf{c}_t^{i*}$ and $f(\mathbf{c}_t, \mathbf{w}_t)$ with $f(\mathbf{c}_t^{i*}, \mathbf{w}_t)$.
  - **2.4** Repeat steps 2.1 $\sim$ 2.3 until convergence, that is, $f(\mathbf{c}_t^{i*}, \mathbf{w}_t) \leq f(\mathbf{c}_t, \mathbf{w}_t)$ for each i.
- **3** Lift-one step:
  - **3.1** Set up a random order of i going through $\{1, 2, \ldots, m_t\}$. For each i, do steps 3.2 $\sim$ 3.3.
  - **3.2** Denote $\mathbf{w}_t^i(z) = ((1-z)\omega_1/(1-\omega_i), \ldots, (1-z)\omega_{i-1}/(1-\omega_i), z, (1-z)\omega_{i+1}/(1-\omega_i), \ldots, (1-z)\omega_{m_t}/(1-\omega_i))^\top$, and $f_i(\mathbf{c}_t, z) = f(\mathbf{c}_t, \mathbf{w}_t^i(z))$.
  - **3.3** Use an analytic solution or the quasi-Newton algorithm to find z^* maximizing $f_i(\mathbf{c}_t, z)$ with $z \in [0, 1)$. Define $\mathbf{w}_t^{i*} = \mathbf{w}_t^i(z^*)$. If $f(\mathbf{c}_t, \mathbf{w}_t^{i*}) > f(\mathbf{c}_t, \mathbf{w}_t)$, replace $\mathbf{w}_t$ with $\mathbf{w}_t^{i*}$ and $f(\mathbf{c}_t, \mathbf{w}_t)$ with $f(\mathbf{c}_t, \mathbf{w}_t^{i*})$.
  - **3.4** Repeat steps 3.1 $\sim$ 3.3 until convergence, that is, $f(\mathbf{c}_t, \mathbf{w}_t^{i*}) \leq f(\mathbf{c}_t, \mathbf{w}_t)$ for each i. $\mathbf{c}_t$ and $\mathbf{w}_t$ form $\xi_{c,t}$. Repeat steps 2 $\sim$ 3 until convergence.
- **4** New point step: Find $c_t^* = \arg\max_{c \in \mathcal{X}} \phi(c, \xi_{c,t})$.
- **5** If $\phi(c_t^*, \xi_{c,t}) \leq \varepsilon$, go to step 6. Otherwise, set $\xi_{c,t+1} = \xi_{c,t} \cup \{(c_t^*, 0)\}$, go back to Step 2, and update $t = t + 1$.
- **6** Output: Report the optimal design $\xi_c^* = \xi_{c,t}$.

Remark 1. The number of i in Algorithm 2 must be at least two according to Theorem 3, since a design ξ_c with $i = 1$ cannot satisfy the condition

$|M(\xi)| > 0$. By Theorem 5, we also verify that the number of different $c_1, c_2, \ldots$ will not exceed $(K+2)/2$. Consequently, the algorithm will naturally be stopped in finite steps and the resulting design is still scalable. After obtaining the optimal design ξ_c^*, the results in Theorem 5 can help us produce the final D-optimal design ξ^*. Note that $\xi_{2q'-1}$ is fixed, we construct the optimal design via a univariate optimal measure ξ_c^*. This transformation effectively reduces the original multivariate optimization problem to a simpler univariate optimization problem. The general equivalence theorem guarantees the D-optimality of ξ_c^* over $\mathcal{X}$ in Step **4**.

5. Simulation studies

Example 2. Consider the developmental toxicity study described in Table 6.11 in Agresti (2019). In this study, each mouse was exposed to one of five concentration levels for ten days early in the pregnancy. The concentration levels ranged from 0 to 500, and each fetus was classified into one of three possible outcomes: dead, malformed, or normal.

The AC po logit model for this study is given by:

$$\log\left(\frac{\pi_{i,j}}{\pi_{i,j+1}}\right) = \theta_j + \beta x_i, i = 1,2,3,4,5, j = 1,2. \tag{5.1}$$

This model has three parameters, and their maximum likelihood estimates

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0075

are $(\hat{\theta}_1, \hat{\theta}_2, \hat{\beta})^\top = (-1.90055, -2.68434, 0.00593)^\top$.

Considering $(\hat{\theta}_1, \hat{\theta}_2, \hat{\beta})^\top$ as the assumed values, the locally D-optimal design of model (5.1), as well as the relative D-efficiencies of the original allocation and uniform allocation with respect to the D-optimal design, are summarized in Table 1.

Table 1: Integer and approximate allocations for toxicity study

Original x	0	62.5	125	250	500	
Original integer	297	242	312	299	285	
Original allocation $\mathbf{w}_o$	0.20697	0.16864	0.21742	0.20836	0.19861	72.52%
Uniform allocation $\mathbf{w}_u$	0.20000	0.20000	0.20000	0.20000	0.20000	72.22%
D-optimal x	236	500				
D-optimal $\mathbf{w}_d$	0.48020	0.51980				
D-optimal integer	689	746				
Bayesian D-optimal x	225	247	500			
Bayesian D-optimal $\mathbf{w}_b$	0.19051	0.29136	0.51813			
$\mathbf{w}_b$ integer-valued	274	418	743			
EW D-optimal x	225	247	500			
EW D-optimal $\mathbf{w}_e$	0.32614	0.15271	0.52115			
$\mathbf{w}_e$ integer-valued	469	219	747			

To evaluate the performance of ξ^*, we compare it with three other types of designs. The first is the uniform design ξ_U, which assigns equal weights to all design points, providing a straightforward baseline for comparison. The second is the grid-based D-optimal design ξ_B proposed by Bu et al. (2020), which focuses on predefined equidistant grid points. The third is the optimal design ξ^*_{For} obtained by the ForLion algorithm proposed by Huang et al. (2024) and implemented in Huang et al. (2025a).

To assess the efficiencies of ξ^* relative to ξ_U and ξ_B, we consider two

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0075

scenarios involving 6 and 11 grid points, and the corresponding designs are denoted by ξ_{U6}, ξ_{U11}, ξ_{B6}, and ξ_{B11}, respectively. The efficiency statistics for these comparisons are summarized in Table 2.

Table 2: Relative D-efficiencies of ξ^* against ξ_U, ξ_B and ξ^*_{For}

Design	Min	1st Quartile	Median	3rd Quartile	Max
ξ_{U6}	0.84464	0.84464	0.84464	0.84464	0.84464
ξ_{U11}	0.83796	0.83796	0.83796	0.83796	0.83796
ξ_{B6}	0.99113	0.99113	0.99113	0.99113	0.99113
ξ_{B11}	0.99813	0.99813	0.99813	0.99813	0.99813
ξ^*_{For}	0.99922	0.99999	0.99999	0.99999	1.00000

The computational time required for these designs on a Windows 10 desktop with 16GB RAM and Intel Core i7-9700 processor are reported in Table 3.

Table 3: Summary of computational times (secs)

Design	Min	1st Quartile	Median	3rd Quartile	Max
ξ^*	0.33213	0.34730	0.36155	0.39183	0.53203
ξ_{B6}	0.32766	0.65439	0.74058	0.83581	1.14581
ξ_{B11}	0.95289	4.59924	5.58286	6.83737	8.11617
ξ^*_{For}	1.01593	1.14059	1.47949	2.57269	5.42663

As shown in Table 2 and 3, none of the uniform designs achieve satisfactory efficiency. Moreover, D-optimal designs based on grid points and ξ^*_{For} require significantly more computational effort to achieve comparable efficiency.

When the parameters of the Fisher information matrix are unknown in advance, locally optimal designs cannot be directly applied. To address this issue, we adopt Bayesian optimal designs, which require prior distributions for the unknown parameters. Since these prior distributions are often unavailable, a bootstrap approach is employed. In practice, we resample 1435 observations from the initial dataset 100 times, yielding a set of parameter estimates $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_{100}$.

For a design ξ and design point x_i under parameter $\boldsymbol{\gamma}_j$, the Fisher information matrices are denoted by $M(\xi, \boldsymbol{\gamma}_j)$ and $M_i(\boldsymbol{\gamma}_j)$, respectively. Then $|M(\xi, \boldsymbol{\gamma}_j)| = |\sum_{i=1}^m \omega_i M_i(\boldsymbol{\gamma}_j)| = f(\mathbf{w}, \boldsymbol{\gamma}_j)$, where $\mathbf{w} = (\omega_1, \ldots, \omega_m)^\top$ represents the allocation weights. An estimate of the Bayesian D-optimality criterion $\psi(\mathbf{w}) = E(\log|M(\xi, \boldsymbol{\gamma})|)$ for a given design ξ with allocation $\mathbf{w}$ can be approximated as:

$$\begin{aligned}\hat{\psi}(\mathbf{w}) = \hat{E}(\log|M(\xi, \boldsymbol{\gamma})|) &= 100^{-1} \sum_{j=1}^{100} \log|M(\xi, \boldsymbol{\gamma}_j)| \\ &= 100^{-1} \sum_{j=1}^{100} \log\left|\sum_{i=1}^{m} \omega_i M_i(\boldsymbol{\gamma}_j)\right|.\end{aligned}$$

The Bayesian D-optimal design that maximizes $\hat{\psi}(\mathbf{w})$, along with its integer-valued allocation, can then be identified.

According to Yang et al. (2016), EW D-optimal designs are significantly easier to compute while maintaining high efficiency compared to Bayesian designs. The expected value of the Fisher information matrix

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0075

at design point x_i, denoted as $E(M_i)$, can be estimated using $\hat{E}(M_i) = 100^{-1}\sum_{j=1}^{100} M_i(\boldsymbol{\gamma}_j)$. An estimate of the EW criterion $\phi(\mathbf{w}) = |E(M(\xi, \boldsymbol{\gamma}))|$ for a design ξ with allocation $\mathbf{w} = (\omega_1, \ldots, \omega_m)^\top$ can then be expressed as:

$$\hat{\phi}(\mathbf{w}) = \left|\hat{E}(M(\xi, \boldsymbol{\gamma}))\right| = \left|\sum_{i=1}^{m} \omega_i \hat{E}(M_i)\right| = \left|\sum_{i=1}^{m} \omega_i \left(100^{-1}\sum_{j=1}^{100} M_i(\boldsymbol{\gamma}_j)\right)\right|.$$

Therefore, the EW D-optimal allocation that maximizes $\hat{\phi}(\mathbf{w})$ can be computed straightforwardly, making it a practical alternative to Bayesian designs.

The Bayesian D-optimal design, the EW D-optimal design, and their corresponding integer-valued allocations are summarized in Table 1. Specifically, $1435 \times \mathbf{w}_b = (273.3819,\ 418.1016, 743.5165)$ and the Bayesian D-optimal integer-valued allocation obtained using Algorithm 1 is $(274, 418, 743)$. This approach eliminates the complexity of manual rounding and avoids potentially suboptimal allocations that rounding might cause.

We next evaluate the robustness of the Bayesian D-optimal design and the EW D-optimal design under misspecified parameter values. Specifically, the fitted parameters $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_{100}$ as described earlier, are utilized for this analysis. For $j = 1, \ldots, 100$, let $\boldsymbol{\gamma}_j$ represent the assumed parameter value, we can compute the locally D-optimal approximate allocation $\mathbf{w}_j^*$ for each

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0075

$\boldsymbol{\gamma}_j$.

The efficiencies of Bayesian D-optimal design $\mathbf{w}_b$, EW D-optimal design $\mathbf{w}_e$ and uniform design $\mathbf{w}_u$ with respect to the locally D-optimal design $\mathbf{w}_j^*$ are defined as $(f(\mathbf{w}_b, \boldsymbol{\gamma}_j)/\ f(\mathbf{w}_j^*, \boldsymbol{\gamma}_j))^{1/3}$, $(f(\mathbf{w}_e, \boldsymbol{\gamma}_j)/f(\mathbf{w}_j^*, \boldsymbol{\gamma}_j))^{1/3}$, and $(f(\mathbf{w}_u, \boldsymbol{\gamma}_j)/f(\mathbf{w}_j^*, \boldsymbol{\gamma}_j))^{1/3}$ respectively. The summary statistics of these efficiencies are presented in Table 4.

Table 4: Summary of efficiency in developmental toxicity study

Design	Min	1st Quartile	Median	3rd Quartile	Max
Bayesian	0.98772	0.99796	0.99894	0.99930	0.99958
EW	0.98363	0.99807	0.99909	0.99938	0.99955
Uniform	0.65543	0.71375	0.72484	0.73673	0.78415

The results suggest that Bayesian and EW D-optimal designs are comparable in terms of efficiency and both outperform uniform design with respect to robustness. Given the significantly lower computational cost of the EW D-optimal design compared to the Bayesian D-optimal design, the EW D-optimal design is recommended for practical applications.

Example 3. Consider an experiment with $q = 6$ factors and $J = 5$ categories, where the design region is $\mathcal{X} = \prod_{t=1}^{5}[-1, 1] \times [-10^{1000}, 10^{1000}]$. The first five factors can be either quantitative or qualitative.

The AC po model under a general link function g is given by

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0075

$$g\left(\frac{\pi_{i,j}}{\pi_{i,j}+\pi_{i,j+1}}\right)=\theta_j+\beta_1x_{i1}+\beta_2x_{i2}+\beta_3x_{i3}+\beta_4x_{i4}+\beta_5x_{i5}+\beta_6x_{i6}, j=1,2,3,4, \tag{5.2}$$

and $c_i=\beta_1x_{i1}+\beta_2x_{i2}+\beta_3x_{i3}+\beta_4x_{i4}+\beta_5x_{i5}+\beta_6x_{i6}$. The parameters are specified as $(\theta_1,\theta_2,\theta_3,\theta_4,\beta_1,\beta_2,\beta_3,\beta_4,\beta_5,\beta_6)^\top=(1,2,3,4,-1,2,-3,1,-2,3)^\top$.

Using Algorithm 2, the optimal design ξ_c^* can be obtained. Through simulations, we found that the design based on 2 $OA(8,5,2,2)$ (denoted as ξ_{2OA}) also demonstrates reasonably high efficiency. Designs ξ_c^* and ξ_{2OA} for Model (5.2) under the log-log, logit, and cauchit link functions are presented in Table 5, where the first row of each block represents c_i, and the second row represents ω_i. The D-optimal design $\xi^*=\{(\boldsymbol{x}_{il},\omega_i/2^3)\}$ can be obtained by inverting $\xi_{2^3}\otimes\xi_c^*$, and is therefore not explicitly listed here (See Supplementary Material, Section S3).

Table 5: Optimal design ξ_c^* and design based on 2 $OA(8,5,2,2)$

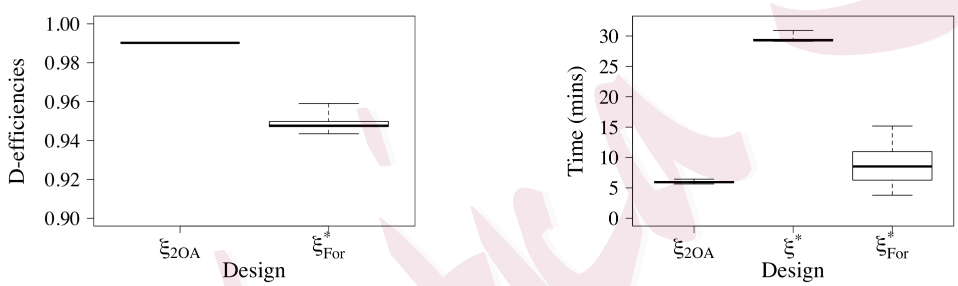
link	ξ_{2OA}		ξ_c^*			
log-log	−1.61058	−3.17952	−1.56997	−2.84867	−3.90903	
	0.56690	0.43310	0.51688	0.33735	0.14577	
logit	−1.74252	−3.25748	−1.64950	−2.50000	−3.35050	
	0.50000	0.50000	0.40925	0.18150	0.40925	
cauchit	−1.69745	−3.30255	−1.09456	−2.02094	−2.97906	−3.90544
	0.50000	0.50000	0.19694	0.30306	0.30306	0.19694

From Table 5, we can conclude that the D-optimal designs vary considerably across different link functions. For the log-log and logit link functions, the D-optimal designs require 3 $OA(8,5,2,2)$. However, this is not

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0075

the case for the cauchit link function, which requires 4 $OA(8, 5, 2, 2)$.

Furthermore, we compare Algorithm 2 with the ForLion algorithm from Huang et al. (2024) in terms of computational time and relative D-efficiency, using the log-log link function as an example. The D-optimal design obtained by ForLion algorithm is denoted as ξ^*_{For}. The relative efficiencies of ξ^* against ξ_{2OA} and ξ^*_{For}, along with their respective computational times, are summarized in Figure 1.



(a) Relative D-efficiencies (log-log) (b) Computational time (log-log)

Figure 1: Relative D-efficiencies and computational time of ξ^* against ξ_{2OA} and ξ^*_{For}

From Figure 1, it can be observed that ξ_{2OA} is highly efficient and computationally fast compared to ξ^* in this case. When compared to ξ^*_{For}, ξ_{2OA} demonstrates superior efficiency and computational advantage.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0075

6. Conclusion

In this paper, the AC models are used to model ordinal responses. We study the locally optimal designs for AC model with general link functions on both discrete and continuous design regions. The explicit form of the Fisher information matrix for the AC model is derived, along with structure characteristics, including the number of support points and a simple complete class of locally D-optimal designs. An efficient algorithm is proposed to identify locally D-optimal designs with multiple factors and D-optimal integer-valued allocations. Numerical examples are provided to demonstrate the superior performance of the obtained optimal designs.

The characteristic of the design structure largely depends on the Chebyshev system, thus we need a relatively weak constraint on the last covariate. Otherwise, the design structure presented in Theorem 5 no longer holds. Consequently, one cannot leverage the structure information to alleviate the computational burden. In this scenario, experimenters can use the general algorithms such as PSO, ForLion (Lukemire et al. (2019, 2022), Huang et al. (2024)) to search out the optimal design. It is also worth mentioning that the Chebyshev system condition in Lemma 1 needs to be handled case-by-case since it relies on the number of response categories, the design region, and the concrete formula of the link function. The results for the logit link presented in Corollary 1 benefit from the functional class composed of different exponential functions, for which the Haar condition (Cheney, 1998) naturally holds. However, such techniques cannot directly extend to other commonly used link functions without additional knowledge about J and $[U, V]$. Nevertheless, the structure remains valuable if a universal optimal design for a single continuous factor can be found. New mathematical tools beyond Chebyshev system are required to find such universal optimal designs, enabling this characteristic to accommodate more general scenarios. This represents an interesting direction for future research.

It is worth noting that, the design points of the D-optimal design of Example 2 do not include the extreme point 0 of the design region. This is a departure from the nature of D-optimal designs for univariate responses and is an aspect that warrants further investigation.

Supplementary Material

The online Supplementary Material contains all technical proofs, an extension of the AC po model and additional simulation studies.

DODORE-code. It contains codes for the simulation studies.

Acknowledgments

The authors are grateful to the editor, associate editor, and the reviewers for valuable comments and suggestions. Huiping Dang and Jun Yu contributed equally to this work. Fasheng Sun is the corresponding author. Huiping Dang and Fasheng Sun's research is supported by National Natural Science Foundation of China (No. 12371259) and the Fundamental Research Funds for the Central Universities (No. 2412023YQ003). Jun Yu's research is supported by National Natural Science Foundation of China (No. 12471244).

References

Ai, M. Y., Z. Q. Ye and J. Yu (2023). Locally D-optimal designs for hierarchical response experiments. *Statistica Sinica* **33**, 381–399.

Agresti, A. (2019). *An Introduction to Categorical Data Analysis*. 3rd Edition. Wiley, New York.

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

Cheney, E. W. (1998). *Introduction to Approximation Theory*. 2nd Edition. American Mathematical Society.

Chuang-Stein, C. and A. Agresti (1997). A review of tests for detecting a monotone dose-response relationship with ordinal response data. *Statistics in medicine* **16**, 2599–2618.

Dette, H. and V. B. Melas (2011). A note on the de la garza phenomenon for locally optimal designs. *Annals of Statistics* **39(2)**, 1266–1281.

Gevertz, J. L. and I. Kareva (2024). Minimally sufficient experimental design using identifiability analysis. *Npj Systems Biology and Applications* **10:2**.

Hao, S. and M. Yang (2020). Support point of locally optimal designs for multinomial logistic regression models. *Journal of Statistical Planning and Inference* **209**, 144–159.

Hedayat, A. S., N. J. Sloane and J. Stufken (1999). *Orthogonal Arrays: Theory and Applications*. Springer, New York.

Huang, Y. F., K. Li, A. Mandal and J. Yang (2024). ForLion: a new algorithm for D-optimal designs under general parametric statistical models with mixed factors. *Statistics and Computing* **34:157**.

Huang, Y. F., S. T. Lin, and J. Yang (2025a). *ForLion: ForLion Algorithm to Find D-Optimal Designs for Experiments*. R package version 0.1.0.

Huang, Y. F., L. P. Tong, and J. Yang (2025b). Constrained D-optimal design for paid research study. *Statistica Sinica* **35**, 1479–1498.

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

Karlin, S. and W. J. Studden (1966). Tchebycheff Systems: With Applications in Analysis and Statistics. *Pure and Applied Mathematics, Vol. XV*. Wiley, New York.

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

Lukemire, J., A. Mandal and W. K. Wong (2022). Optimal experimental designs for ordinal models with mixed factors for industrial and healthcare applications. *Journal of Quality Technology* **54(2)**, 184–196.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B* **42(2)**, 109–142.

McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models.* 2nd Edition. Chapman and Hall/CRC, Boca Raton.

O'Connell, A. A. (2006). *Logistic Regression Models for Ordinal Response Variables.* 1st Edition. Sage.

Ouyang, Y. Y., Y. M. Chen, G. H. Lan and E. Pasiliao (2015). An accelerated linearized alternating direction method of multipliers. *SIAM Journal on Imaging Sciences* **8**, 644–681.

Perevozskaya, I., W. F. Rosenberger and L. M. Haines (2003). Optimal design for the proportional odds model. *The Canadian Journal of Statistics* **31**, 225–235.

Silvey, S. D. (1980). *Optimal Design*. Chapman & Hall/CRC.

Smucker, B., M. Krzywinski and N. Altman (2018). Optimal experimental design. *Nature Methods* **15**, 559–560.

Stephen, J. W. (2015). Coordinate descent algorithms. *Mathematical Programming* **151**, 3–34.

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

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

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

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

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

KLAS and School of Mathematics and Statistics, Northeast Normal University

E-mails: danghuiping5566@sina.com, sunfs359@nenu.edu.cn

School of Mathematics and Statistics, Beijing Institute of Technology

E-mail: yujunbeta@bit.edu.cn