D-optimal Designs for Multinomial Logistic Models

Xianwei Bu[1], Dibyen Majumdar[2] and Jie Yang[2]

[1]*AbbVie Inc. and* [2]*University of Illinois at Chicago*

Supplementary Materials

S.1. List of notations.

$\mathbf{0}_k$	A vector of k zeros
a_j	$\mathbf{h}_j^T(\mathbf{x})\boldsymbol{\beta}_j + \mathbf{h}_c^T(\mathbf{x})\boldsymbol{\zeta}$, $j = 1, \ldots, J-1$, given $\mathbf{x} = (x_1, \ldots, x_d)^T$
b_j	Coefficients in representing $f_i(z)$, $j = 0, \ldots, J-1$
$\mathbf{B}_J$	$J \times J$ constant matrix used for deriving the coefficients of $f_i(z)$, $(s^{t-1})_{st}$
$\mathbf{C}$	$J \times (2J-1)$ constant matrix, same for all the four logit models
$\mathbf{c}$	Vector used for deriving coefficients of $f_i(z)$, $(c_1, \ldots, c_{J-1})^T$
$\mathbf{c}_{ij}$	$J \times 1$ vectors such that $(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})^{-1} = (\mathbf{c}_{i1}, \ldots, \mathbf{c}_{iJ})$
c_j	$(j+1)^p j^{J-1-p} f_i(1/(j+1)) - j^{J-1} f_i(0)$, $j = 1, \ldots, J-1$
$c_{\alpha_1,\ldots,\alpha_m}$	Coefficient of $w_1^{\alpha_1} \cdots w_m^{\alpha_m}$ in the determinant of $\mathbf{G}^T\mathbf{W}\mathbf{G}$
d	Total number of design factors
d_s	$d_s = (f_{ij}(s) - f_{ij}(0))/s$, $s = 1, \ldots, q$, for coefficients in f_{ij}
$\mathbf{D}_i$	$\mathrm{diag}(\mathbf{L}\boldsymbol{\pi}_i)$
$\mathbf{e}_i$	$m \times 1$ vector with the ith coordinate 1 and all others 0
$\mathbf{F}$	Fisher information matrix of the design, $\mathbf{F} = \sum_{i=1}^m n_i \mathbf{F}_i$
f	$f(\mathbf{w}) = f(w_1, \ldots, w_m) = \lvert\mathbf{G}^T\mathbf{W}\mathbf{G}\rvert$ which is proportional to $\lvert\mathbf{F}\rvert$; or $f(\mathbf{n}) = f(n_1, \ldots, n_m) = \lvert\sum_{i=1}^m n_i \mathbf{F}_i\rvert = \lvert\mathbf{F}\rvert$
$\mathbf{F}_i$	Fisher information matrix at the ith design point
f_i	$f_i(z) = f(w_1(1-z)/(1-w_i), \ldots, w_{i-1}(1-z)/(1-w_i), z, w_{i+1}(1-z)/(1-w_i), \ldots, w_m(1-z)/(1-w_i))$ with $0 \le z < 1$
f_{ij}	$f_{ij}(z) = f(n_1, \ldots, n_{i-1}, z, n_{i+1}, \ldots, n_{j-1}, n_i + n_j - z, n_{j+1}, \ldots, n_m)$ with $z = 0, 1, \ldots, n_i + n_j$
$\mathbf{G}$	Matrix component for Fisher information matrix such that $\mathbf{F} = n\mathbf{G}^T\mathbf{W}\mathbf{G}$, $mJ \times p$
g_s	$g_0 = f_{ij}(0)$ and $(g_1, \ldots, g_q)^T = \mathbf{B}_q^{-1}(d_1, \ldots, d_q)^T$
$\mathbf{H}$	Matrix component for Fisher information matrix such that $\mathbf{F} = \mathbf{H}\mathbf{U}\mathbf{H}^T$, consisting of $\mathbf{H}_1, \ldots, \mathbf{H}_{J-1}$ and possibly $\mathbf{H}_c$, $p \times m(J-1)$
$\mathbf{H}_c$	Matrix for the common component of $J-1$ categories, $(\mathbf{h}_c(\mathbf{x}_1), \ldots, \mathbf{h}_c(\mathbf{x}_m))$, $p_c \times m$
$\mathbf{h}_c(\mathbf{x}_i)$	Vector of p_c predictors associated with the p_c parameters $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_{p_c})^T$ that are common for all of the response categories as known functions of the ith experimental setting, $(h_1(\mathbf{x}_i), \ldots, h_{p_c}(\mathbf{x}_i))^T$
$\mathbf{H}_j$	Matrix for the jth category only, $(\mathbf{h}_j(\mathbf{x}_1), \ldots, \mathbf{h}_j(\mathbf{x}_m))$, $p_j \times m$
$\mathbf{h}_j(\mathbf{x}_i)$	Vector of p_j predictors associated with the p_j parameters $\boldsymbol{\beta}_j = (\beta_{j1}, \ldots, \beta_{jp_j})^T$ for the jth response category as known functions of the ith experimental setting, $(h_{j1}(\mathbf{x}_i), \ldots, h_{jp_j}(\mathbf{x}_i))^T$

$\mathbf{I}_k$	The identity matrix of order k
J	Total number of response categories
$k_{\min}$	Smallest possible $\#\{i \mid \alpha_i > 0\}$ such that $c_{\alpha_1,\ldots,\alpha_m} > 0$
$\mathbf{L}$	Constant $(2J-1) \times J$ matrix, different for the four logit models
m	Total number of distinct experimental settings or design points
$\mathcal{M}(\mathbf{H})$	Column space of matrix $\mathbf{H}$, that is, the linear subspace spanned by the columns of $\mathbf{H}$
n	Total number of experimental units, $n = n_1 + \cdots + n_m$
$\mathbf{n}$	Allocation of experimental units, $(n_1, \ldots, n_m)^T$, $n_i \geq 0$, $\sum_i n_i = n$
n_i	Number of replicates at the ith experimental setting
p	Total number of parameters
p_c	Number of common parameters for $J-1$ categories
p_H	$dim\left(\cap_{j=1}^{J-1} \mathcal{M}(\mathbf{H}_j^T)\right)$
p_j	Number of parameters for the jth category only
q	$\min\{2J-2, p-k_{\min}+2, p\}$, upper bound of order of $f_{ij}(z)$
S	Collection of all feasible approximate allocations, $\{(w_1, \ldots, w_m)^T \in \mathbb{R}^m \mid w_i \geq 0, i = 1, \ldots, m; \sum_{i=1}^m w_1 = 1\}$
S_+	Collection of approximate allocations, $\{\mathbf{w} \in S \mid f(\mathbf{w}) > 0\}$
$\mathbf{U}$	Block matrix $(\mathbf{U}_{st})_{s,t=1,\ldots,J-1}$, $m(J-1) \times m(J-1)$
$\mathbf{U}_{st}$	$\text{diag}\{n_1 u_{st}(\boldsymbol{\pi}_1), \ldots, n_m u_{st}(\boldsymbol{\pi}_m)\}$, $m \times m$
$u_{st}(\boldsymbol{\pi}_i)$	$\mathbf{c}_{is}^T \text{diag}(\boldsymbol{\pi}_i)^{-1} \mathbf{c}_{it}$ for $s, t = 1, \ldots, J-1$
$\mathbf{w}$	Real-valued allocation of experimental units, $(w_1, \ldots, w_m)^T$, $w_i \geq 0$, $\sum_i w_i = 1$
$\mathbf{W}$	$\text{diag}\{w_1 \text{diag}(\boldsymbol{\pi}_1)^{-1}, \ldots, w_m \text{diag}(\boldsymbol{\pi}_m)^{-1}\}$, $mJ \times mJ$
w_i	Proportion of experimental units assigned to the ith experimental setting, n_i/n
$\mathbf{w}_u$	Uniform allocation, $(1/m, \ldots, 1/m)^T$
$\mathcal{X}$	Design space, the collection of all design points yielding strictly positive categorical probabilities of response; or a predetermined set of design points considered
$\mathbf{x}_i$	The ith distinct experimental setting or design point, $(x_{i1}, \ldots, x_{id})^T$
$\mathbf{X}_i$	Model matrix at the ith design point, $J \times p$, the last row is all 0's
$\boldsymbol{\beta}_j$	Vector of parameters for the jth response category only, $(\beta_{j1}, \ldots, \beta_{jp_j})^T$
γ_{ij}	The cumulative probability from the 1st to jth categories at the ith experimental setting, $\gamma_{ij} = \pi_{i1} + \cdots + \pi_{ij}$
$\boldsymbol{\zeta}$	Vector of common parameters for all of the response categories, $(\zeta_1, \ldots, \zeta_{p_c})^T$
$\boldsymbol{\eta}_i$	Vector of linear predictors at the ith experimental setting, $\boldsymbol{\eta}_i = (\eta_{i1}, \ldots, \eta_{iJ})^T = \mathbf{X}_i \boldsymbol{\theta}$ with $\eta_{iJ} \equiv 0$
$\boldsymbol{\theta}$	Vector of all parameters, $p \times 1$
$\boldsymbol{\Theta}$	Parameter space, the collection of all feasible parameter vectors
$\boldsymbol{\pi}_i$	Vector of response category probabilities at the ith experimental setting. $\boldsymbol{\pi}_i = (\pi_{i1}, \ldots, \pi_{iJ})^T$, $\pi_{i1} + \cdots + \pi_{iJ} = 1$
$\Lambda(\alpha_1, \ldots, \alpha_m)$	$\{(i_1, \ldots, i_p) \mid 1 \leq i_1 < \cdots < i_p \leq mJ; \#\{l : (k-1)J < i_l \leqslant kJ\} = \alpha_k, k = 1, \ldots, m\}$
π_{ij}	Probability that the response falls into the jth category at the ith experimental setting
ϕ	Bayesian D-optimal criterion, $\phi(\mathbf{p}) = E(\log \lvert\mathbf{F}\rvert)$

S.2. Formulae of matrix differentiation. According to Seber (2008, Chapter 17)),

$$\begin{aligned}
\frac{\partial \mathbf{y}}{\partial \mathbf{x}^T} &= \left(\frac{\partial y_i}{\partial x_j}\right)_{ij} \\
\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}^T} &= \mathbf{A} \\
\frac{\partial \mathbf{z}}{\partial \mathbf{x}^T} &= \frac{\partial \mathbf{z}}{\partial \mathbf{y}^T} \cdot \frac{\partial \mathbf{y}}{\partial \mathbf{x}^T} \\
\frac{\partial \log \mathbf{y}}{\partial \mathbf{x}^T} &= [\mathrm{diag}(\mathbf{y})]^{-1} \frac{\partial \mathbf{y}}{\partial \mathbf{x}^T}
\end{aligned}$$

where $\mathbf{x} = (x_i)_i$, $\mathbf{y} = (y_i)_i$, $\mathbf{z} = (z_i)_i$, and thus $\log \mathbf{y} = (\log y_i)_i$ are vectors, and $\mathbf{A}$ is a constant matrix.

S.3. Explicit forms of $(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})^{-1}$ for all the four logit models. There are the four different kinds of multinomial logistic models in the literature: *baseline-category logit model* for nominal responses, *cumulative logit model* for ordinal responses, *adjacent-categories logit model* for ordinal responses, and *continuation-ratio logit model* for hierarchical responses. According to Theorem 2.1, $(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})^{-1}$ is a key matrix that we must calculate.

Recall that $\pi_{i1} + \cdots + \pi_{iJ} = 1$, $i = 1, \ldots, m$. Then

$$(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})_{baseline} = \begin{pmatrix} \frac{1}{\pi_{i1}} & 0 & \cdots & 0 & -\frac{1}{\pi_{iJ}} \\ 0 & \frac{1}{\pi_{i2}} & \ddots & \vdots & -\frac{1}{\pi_{iJ}} \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & \frac{1}{\pi_{i,J-1}} & -\frac{1}{\pi_{iJ}} \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}_{J\times J}$$

$$(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})_{cumulative} = \begin{pmatrix} \frac{1}{\gamma_{i1}} & -\frac{1}{1-\gamma_{i1}} & -\frac{1}{1-\gamma_{i1}} & \cdots & -\frac{1}{1-\gamma_{i1}} \\ \frac{1}{\gamma_{i2}} & \frac{1}{\gamma_{i2}} & -\frac{1}{1-\gamma_{i2}} & \cdots & -\frac{1}{1-\gamma_{i2}} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \frac{1}{\gamma_{i,J-1}} & \frac{1}{\gamma_{i,J-1}} & \cdots & \frac{1}{\gamma_{i,J-1}} & -\frac{1}{1-\gamma_{i,J-1}} \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}_{J\times J}$$

$$(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})_{continuation} = \begin{pmatrix} \frac{1}{\pi_{i1}} & -\frac{1}{1-\gamma_{i1}} & -\frac{1}{1-\gamma_{i1}} & \cdots & -\frac{1}{1-\gamma_{i1}} \\ 0 & \frac{1}{\pi_{i2}} & -\frac{1}{1-\gamma_{i2}} & \cdots & -\frac{1}{1-\gamma_{i2}} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \frac{1}{\pi_{i,J-1}} & -\frac{1}{1-\gamma_{i,J-1}} \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}_{J\times J}$$

$$(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})_{adjacent} = \begin{pmatrix} \frac{1}{\pi_{i1}} & -\frac{1}{\pi_{i2}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\pi_{i2}} & -\frac{1}{\pi_{i3}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{\pi_{i,J-1}} & -\frac{1}{\pi_{iJ}} \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}_{J\times J}$$

where $\gamma_{ij} = \pi_{i1} + \cdots + \pi_{ij}$ is the cumulative categorical probability, $j = 1, \ldots, J-1$. The corresponding inverse matrices are

$$
\begin{aligned}
&(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})^{-1}_{baseline} \\
&= \begin{pmatrix}
-\pi_{i1}^2 + \pi_{i1} & -\pi_{i1}\pi_{i2} & \cdots & -\pi_{i1}\pi_{i,J-1} & \pi_{i1} \\
-\pi_{i1}\pi_{i2} & -\pi_{i2}^2 + \pi_{i2} & \cdots & -\pi_{i2}\pi_{i,J-1} & \pi_{i2} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
-\pi_{i1}\pi_{i,J-1} & -\pi_{i2}\pi_{i,J-1} & \cdots & -\pi_{i,J-1}^2 + \pi_{i,J-1} & \pi_{i,J-1} \\
-\pi_{i1}\pi_{iJ} & -\pi_{i2}\pi_{iJ} & \cdots & -\pi_{i,J-1}\pi_{iJ} & \pi_{iJ}
\end{pmatrix}_{J\times J} \\
&\triangleq \begin{pmatrix} \mathbf{c}_{i1} & \mathbf{c}_{i2} & \cdots & \mathbf{c}_{iJ} \end{pmatrix}_{baseline}
\end{aligned}
$$

where $(\mathbf{c}_{ij})_{baseline} = \pi_{ij}(\mathbf{e}_j - \boldsymbol{\pi}_i)$, $j = 1, \ldots, J-1$, $(\mathbf{c}_{iJ})_{baseline} = \boldsymbol{\pi}_i$, and $\mathbf{e}_j$ here is the $J \times 1$ vector with the jth coordinate 1 and all others 0. Recall that $\boldsymbol{\pi}_i = (\pi_{i1}, \ldots, \pi_{iJ})^T$.

$$
\begin{aligned}
&(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})^{-1}_{cumulative} \\
&= \begin{pmatrix}
\gamma_{i1}(1-\gamma_{i1}) & 0 & \cdots & 0 & \pi_{i1} \\
-\gamma_{i1}(1-\gamma_{i1}) & \gamma_{i2}(1-\gamma_{i2}) & \ddots & \vdots & \pi_{i2} \\
0 & -\gamma_{i2}(1-\gamma_{i2}) & \ddots & 0 & \vdots \\
\vdots & \ddots & \ddots & \gamma_{i,J-1}(1-\gamma_{i,J-1}) & \pi_{i,J-1} \\
0 & \cdots & 0 & -\gamma_{i,J-1}(1-\gamma_{i,J-1}) & \pi_{iJ}
\end{pmatrix}_{J\times J} \\
&\triangleq \begin{pmatrix} \mathbf{c}_{i1} & \mathbf{c}_{i2} & \cdots & \mathbf{c}_{iJ} \end{pmatrix}_{cumulative}
\end{aligned}
$$

where $(\mathbf{c}_{ij})_{cumulative} = \gamma_{ij}(1-\gamma_{ij})(\mathbf{e}_j - \mathbf{e}_{j+1})$ with $\mathbf{e}_j$ defined as above; and $(\mathbf{c}_{iJ})_{cumulative} = \boldsymbol{\pi}_i$.

$$
\begin{aligned}
&(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})^{-1}_{continuation} \\
&= \begin{pmatrix}
\pi_{i1}(1-\gamma_{i1}) & 0 & \cdots & 0 & \pi_{i1} \\
-\pi_{i1}\pi_{i2} & \frac{\pi_{i2}(1-\gamma_{i2})}{1-\gamma_{i1}} & \ddots & \vdots & \pi_{i2} \\
\vdots & \vdots & \ddots & 0 & \vdots \\
-\pi_{i1}\pi_{i,J-1} & -\frac{\pi_{i2}\pi_{i,J-1}}{1-\gamma_{i1}} & \cdots & \frac{\pi_{i,J-1}(1-\gamma_{i,J-1})}{1-\gamma_{i,J-2}} & \pi_{i,J-1} \\
-\pi_{i1}\pi_{iJ} & -\frac{\pi_{i2}\pi_{iJ}}{1-\gamma_{i1}} & \cdots & -\frac{\pi_{i,J-1}\pi_{iJ}}{1-\gamma_{i,J-2}} & \pi_{iJ}
\end{pmatrix}_{J\times J} \\
&= \begin{pmatrix} \mathbf{c}_{i1} & \mathbf{c}_{i2} & \cdots & \mathbf{c}_{iJ} \end{pmatrix}_{continuation}
\end{aligned}
$$

where $(\mathbf{c}_{i1})_{continuation} = \pi_{i1}(1-\gamma_{i1}, -\pi_{i2}, \ldots, -\pi_{iJ})^T$,
$(\mathbf{c}_{ij})_{continuation} = \frac{\pi_{ij}}{1-\gamma_{i,j-1}}(0, \ldots, 0, 1-\gamma_{ij}, -\pi_{i,j+1}, \ldots, -\pi_{iJ})^T$ with "$1-\gamma_{ij}$" being the

jth coordinate, $j = 2, \ldots, J-1$, and $(\mathbf{c}_{iJ})_{continuation} = \boldsymbol{\pi}_i$.

$$
\begin{aligned}
&(\mathbf{C}^T \mathbf{D}_i^{-1} \mathbf{L})^{-1}_{adjacent} \\
&= \begin{pmatrix}
(1-\gamma_{i1})\pi_{i1} & (1-\gamma_{i2})\pi_{i1} & \cdots & (1-\gamma_{i,J-1})\pi_{i1} & \pi_{i1} \\
-\gamma_{i1}\pi_{i2} & (1-\gamma_{i2})\pi_{i2} & \cdots & (1-\gamma_{i,J-1})\pi_{i2} & \pi_{i2} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
-\gamma_{i1}\pi_{i,J-1} & -\gamma_{i2}\pi_{i,J-1} & \cdots & (1-\gamma_{i,J-1})\pi_{i,J-1} & \pi_{i,J-1} \\
-\gamma_{i1}\pi_{iJ} & -\gamma_{i2}\pi_{iJ} & \cdots & -\gamma_{i,J-1}\pi_{iJ} & \pi_{iJ}
\end{pmatrix}_{J\times J} \\
&= \begin{pmatrix} \mathbf{c}_{i1} & \mathbf{c}_{i2} & \cdots & \mathbf{c}_{iJ} \end{pmatrix}_{adjacent}
\end{aligned}
$$

where $(\mathbf{c}_{ij})_{adjacent} = ((1-\gamma_{ij})\pi_{i1}, \ldots, (1-\gamma_{ij})\pi_{ij}, -\gamma_{ij}\pi_{i,j+1}, \ldots, -\gamma_{ij}\pi_{iJ})^T$, $j = 1, \ldots, J-1$, and $(\mathbf{c}_{iJ})_{adjacent} = \boldsymbol{\pi}_i$.

For certain applications, we need to know $|\mathbf{C}^T \mathbf{D}_i^{-1} \mathbf{L}|$ (see, for example, Lemma S.9). Since adding a multiple of one row (column) to another row (column) does not change the determinant (see, for example, 4.28(f) in Seber (2008, page 58)), we may (1) do row operations on $(\mathbf{C}^T \mathbf{D}_i^{-1} \mathbf{L})_{baseline}$ and change it into an upper triangular matrix with diagonal entries $\pi_{i1}^{-1}, \ldots, \pi_{iJ}^{-1}$; (2) do row operations on $(\mathbf{C}^T \mathbf{D}_i^{-1} \mathbf{L})^{-1}_{cumulative}$ and change it into an upper triangular matrix with diagonal entries $\gamma_{i1}(1-\gamma_{i1}), \ldots, \gamma_{i,J-1}(1-\gamma_{i,J-1}), 1$; (3) do column operations on $(\mathbf{C}^T \mathbf{D}_i^{-1} \mathbf{L})_{adjacent}$ and change it into a lower triangular matrix with diagonal entries $\pi_{i1}^{-1}, \ldots, \pi_{iJ}^{-1}$; and (4) do column operations on $(\mathbf{C}^T \mathbf{D}_i^{-1} \mathbf{L})_{continuation}$ and change it into a lower triangular matrix with diagonal entries $\pi_{i1}^{-1}, \ldots, \pi_{iJ}^{-1}$. Therefore,

$$
\text{(S.1)} \qquad |\mathbf{C}^T \mathbf{D}_i^{-1} \mathbf{L}| = \begin{cases}
\prod_{j=1}^{J} \pi_{ij}^{-1} & \text{for baseline-category,} \\
 & \text{adjacent-categories,} \\
 & \text{and continuation-ratio logit models} \\
\prod_{j=1}^{J-1} \gamma_{ij}^{-1}(1-\gamma_{ij})^{-1} & \text{for cumulative logit models}
\end{cases}
$$

As a direct conclusion, $|\mathbf{C}^T \mathbf{D}_i^{-1} \mathbf{L}| > 0$ as long as $\pi_{ij} > 0$ for all $j = 1, \ldots, J$.

S.4. Positive definiteness of U.

In order to determine the positive definiteness of $\mathbf{F}$, we first investigate the $m(J-1) \times m(J-1)$ matrix $\mathbf{U}$ defined for Theorem 3.1, which is symmetric since $u_{st}(\boldsymbol{\pi}_i) = u_{ts}(\boldsymbol{\pi}_i)$ and thus $\mathbf{U}_{st} = \mathbf{U}_{ts}$.

THEOREM S.3. *If $n_i > 0$ for all $i = 1, \ldots, m$, then $\mathbf{U}$ is positive definite.*

THEOREM S.4. $|\mathbf{U}| = (\prod_{i=1}^m n_i)^{J-1} \cdot \prod_{i=1}^m (\prod_{j=1}^J \pi_{ij})^{-1} |\mathbf{C}^T \mathbf{D}_i^{-1} \mathbf{L}|^{-2}$.

The proofs of Theorems S.3 and S.4 are relegated to Section S.15. Note that Theorem S.3 is not a corollary of Theorem S.4 since nonsingularity itself does not mean positive definiteness. Theorem S.4 implies that $\mathbf{U}$ is singular if $n_i = 0$ for some $i = 1, \ldots, m$. Note that $\mathbf{F}$ can still be positive definite even if $\mathbf{U}$ is singular, as long as $\mathbf{H}$ is of full row rank. In general, given an allocation $(n_1, \ldots, n_m)$ of the n experimental units with $n_i \geq 0$ and $\sum_{i=1}^m n_i = n$, if we denote $k = \#\{i : n_i > 0\}$ and $\mathbf{U}^*_{st} = \text{diag}\{n_i u_{st}(\boldsymbol{\pi}_i) : n_i > 0\}$, then $\mathbf{U}^* = (\mathbf{U}^*_{st})_{s,t=1,\ldots,J-1}$ is a $k(J-1) \times k(J-1)$ matrix. After removing all columns of $\mathbf{H}$ associated with $n_i = 0$, we denote the leftover as $\mathbf{H}^*$, which is a $p \times k(J-1)$ matrix. It can be verified that

LEMMA S.1. $\mathbf{H}\mathbf{U}\mathbf{H}^T = \mathbf{H}^*\mathbf{U}^*(\mathbf{H}^*)^T$.

LEMMA S.2. $|\mathbf{U}^*| = (\prod_{i:n_i>0} n_i)^{J-1} \cdot \prod_{i:n_i>0}(\prod_{j=1}^J \pi_{ij})^{-1}|\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L}|^{-2}$.

Since $\mathbf{U}^*$ is simply $\mathbf{U}$ if all $n_i > 0$, we have the following corollary of Theorem S.3:

COROLLARY S.1. *$\mathbf{U}^*$ is positive definite.*

S.5. Row rank of H matrix. According to Theorem 3.2, the positive definiteness of the Fisher information matrix $\mathbf{F}$ depends on the row rank of $\mathbf{H}$ or $\mathbf{H}^*$. To simplify the notations, we assume $n_i > 0, i = 1, \ldots, m$ throughout this section. In this case, $\mathbf{H} = \mathbf{H}^*$ and $\mathbf{U} = \mathbf{U}^*$. We also assume that

$$m \geq p_j, \quad j = 1, \ldots, J-1 \quad \text{and} \quad m \geq p_c \text{ if applicable} \tag{S.2}$$

since $\mathbf{H}$ is of full row rank only if $rank(\mathbf{H}_j) = p_j$, $j = 1, \ldots, J-1$ and $rank(\mathbf{H}_c) = p_c$ if applicable.

Since $\mathbf{H}$ takes different forms for *ppo*, *npo*, and *po* models, we investigate its row rank case by case.

THEOREM S.5. *Consider the $p \times m(J-1)$ matrix $\mathbf{H}$ in Theorem 3.1.*

(1) For npo models, $rank(\mathbf{H}) = rank(\mathbf{H}_1) + \cdots + rank(\mathbf{H}_{J-1})$.
(2) For po models, $rank(\mathbf{H}) = rank((\mathbf{1}, \mathbf{H}_c^T)) + J - 2$, where $\mathbf{1}$ is a vector of all 1's.
(3) For ppo models, $rank(\mathbf{H}) = rank(\mathbf{H}_1)+\cdots+rank(\mathbf{H}_{J-1})+rank(\mathbf{H}_c)-dim[\mathcal{M}(\mathbf{H}_c^T)\cap(\cap_{j=1}^{J-1}\mathcal{M}(\mathbf{H}_j^T))]$, where $\mathcal{M}(\mathbf{H}_c^T)$ stands for the column space of $\mathbf{H}_c^T$ or the row space of $\mathbf{H}_c$.

The proof of Theorem S.5 is relegated to Section S.15. In order to apply it to *ppo* models, we need an efficient way to calculate $dim[\mathcal{M}(\mathbf{H}_c^T)\cap(\cap_{j=1}^{J-1}\mathcal{M}(\mathbf{H}_j^T))]$. We provide a formula for calculating $dim(\bigcap_j \mathcal{M}(\mathbf{H}_j^T))$ for general matrices, Theorem A.1 in the Appendix, and relegated its proof to Section S.15.

Recall that $p_H = dim(\cap_{j=1}^{J-1}\mathcal{M}(\mathbf{H}_j^T))$. As a direct conclusion of Theorem S.5, we have

COROLLARY S.2. *For ppo models, $|F| > 0$ only if $m \geqslant p_c + p_H$.*

S.6. Results on the coefficient $c_{\alpha_1,\ldots,\alpha_m}$ for simplifying $|\mathbf{F}|$.

LEMMA S.3. *If $\max_{1\leq i\leq m} \alpha_i \geq J$, then $|\mathbf{G}[i_1, \ldots, i_p]| = 0$ for any $(i_1, \ldots, i_p) \in \Lambda(\alpha_1, \ldots, \alpha_m)$. Therefore, $c_{\alpha_1,\ldots,\alpha_m} = 0$ in this case.*

THEOREM S.6. *The coefficient $c_{\alpha_1,\ldots,\alpha_m}$ as defined in (9) is nonzero only if the restricted Fisher information matrix $\mathbf{F}_{res} = \sum_{i:\alpha_i>0}\mathbf{F}_i$ is positive definite, where $\mathbf{F}_i$ is defined as in (4).*

The proofs for Lemma S.3 and Theorem S.6 are relegated to Section S.15. Combining Theorems 3.2 and S.6, Theorems 3.3 and S.6, respectively, we obtain the following corollaries:

Corollary S.3. *The coefficient* $c_{\alpha_1,\ldots,\alpha_m}$ *is nonzero only if* $\mathbf{H}_{\alpha_1,\ldots,\alpha_m}$ *is of full row rank* p*, where* $\mathbf{H}_{\alpha_1,\ldots,\alpha_m}$ *is the submatrix of* $\mathbf{H}$ *after removing all columns associated with* $\mathbf{x}_i$ *for which* $\alpha_i = 0$*.*

Corollary S.4. *The coefficient* $c_{\alpha_1,\ldots,\alpha_m} = 0$ *if* $\#\{i \mid \alpha_i > 0\} \leq k_{\min} - 1$*, where* $k_{\min} = \max\{p_1, \ldots, p_{J-1}, p_c + p_H\}$*. If* $\mathbf{H}_1 = \cdots = \mathbf{H}_{J-1}$*,* $k_{\min} = p_c + p_1$*.*

We provide an example (Example S.6) in Section S.14 to illustrate that $c_{\alpha_1,\ldots,\alpha_m}$ could be nonzero for *ppo* models with $\#\{i \mid \alpha_i > 0\} = p_c + p_H$.

S.7. Expressions for proportional odds (*po*) models. As special cases of *ppo*, *po* models are degenerate cases of *ppo* models with $\mathbf{h}_j^T(\mathbf{x}_i)$ replaced by 1, $j = 1, \ldots, J-1$, and thus $p_1 = \cdots = p_{J-1} = 1$.

In Section 2, the four logit models in the literature with proportional odds are:

$$\begin{aligned}
\log\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) &= \beta_j + \mathbf{h}_c^T(\mathbf{x}_i)\boldsymbol{\zeta}\ , \text{ baseline-category} \\
\log\left(\frac{\pi_{i1} + \cdots + \pi_{ij}}{\pi_{i,j+1} + \cdots + \pi_{iJ}}\right) &= \beta_j + \mathbf{h}_c^T(\mathbf{x}_i)\boldsymbol{\zeta}\ , \text{ cumulative} \\
\log\left(\frac{\pi_{ij}}{\pi_{i,j+1}}\right) &= \beta_j + \mathbf{h}_c^T(\mathbf{x}_i)\boldsymbol{\zeta}\ , \text{ adjacent-categories} \\
\log\left(\frac{\pi_{ij}}{\pi_{i,j+1} + \cdots + \pi_{iJ}}\right) &= \beta_j + \mathbf{h}_c^T(\mathbf{x}_i)\boldsymbol{\zeta}\ , \text{ continuation-ratio}
\end{aligned}$$

where $i = 1, \ldots, m$, $j = 1, \ldots, J-1$, β_j is an unknown parameter for the jth response category, $\mathbf{h}_c^T(\cdot) = (h_1(\cdot), \ldots, h_{p_c}(\cdot))$ 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.

In equation (1), the corresponding model matrix is

$$\text{(S.3)} \qquad \mathbf{X}_i = \begin{pmatrix} 1 & 0 & \cdots & 0 & \mathbf{h}_c^T(\mathbf{x}_i) \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \mathbf{h}_c^T(\mathbf{x}_i) \\ 0 & \cdots & 0 & 1 & \mathbf{h}_c^T(\mathbf{x}_i) \\ 0 & 0 & \cdots & 0 & \mathbf{0}^T \end{pmatrix}_{J \times p}$$

and the parameter vector $\boldsymbol{\theta} = (\beta_1, \beta_2, \cdots, \beta_{J-1}, \boldsymbol{\zeta})^T$ consists of $p = J - 1 + p_c$ unknown parameters in total. The previous $\boldsymbol{\beta}_j$ reduces to β_j serving as the cut-off point in this case.

In Section 3, the $p \times m(J-1)$ matrix

$$\text{(S.4)} \qquad \mathbf{H} = \begin{pmatrix} \mathbf{1}^T & & \\ & \ddots & \\ & & \mathbf{1}^T \\ \mathbf{H}_c & \cdots & \mathbf{H}_c \end{pmatrix}$$

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

As a special case of Theorem 3.3,

THEOREM S.7. *Consider the multinomial logistic model (1) with m distinct experimental settings $\mathbf{x}_i$ with $n_i > 0$ experimental units, $i = 1, \ldots, m$. For proportional odds models, the Fisher information matrix $\mathbf{F}$ is positive definite if and only if $m \geq p_c + 1$ and the extended matrix $(\mathbf{1}, \mathbf{H}_c^T)$ is of full rank $p_c + 1$.*

In Section 4, for proportional odds models, the $mJ \times p$ matrix

$$\mathbf{G} = \begin{pmatrix} \mathbf{c}_{11} & \cdots & \mathbf{c}_{1,J-1} & \sum_{j=1}^{J-1} \mathbf{c}_{1j} \cdot \mathbf{h}_c^T(\mathbf{x}_1) \\ \mathbf{c}_{21} & \cdots & \mathbf{c}_{2,J-1} & \sum_{j=1}^{J-1} \mathbf{c}_{2j} \cdot \mathbf{h}_c^T(\mathbf{x}_2) \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{c}_{m1} & \cdots & \mathbf{c}_{m,J-1} & \sum_{j=1}^{J-1} \mathbf{c}_{mj} \cdot \mathbf{h}_c^T(\mathbf{x}_m) \end{pmatrix} \tag{S.5}$$

As a special case of Corollary S.4,

COROLLARY S.5. *The coefficient $c_{\alpha_1,\ldots,\alpha_m} = 0$ if $\#\{i \mid \alpha_i > 0\} \leq k_{\min} - 1$, where $k_{\min} = p_c + 1$ for po models.*

As special cases of *ppo* models, *po* models imply $p_1 = \cdots = p_{J-1} = p_H = 1$, and $\mathbf{H}_1 = \cdots = \mathbf{H}_{J-1}$ implies $p_1 = \cdots = p_{J-1} = p_H$. That is, $k_{\min}$'s are consistent across different odds models.

S.8. Expressions for non-proportional odds (*npo*) models.

As special cases of *ppo*, $\mathbf{h}_c^T(\mathbf{x}_i) \equiv 0$ leads to *npo* models. Therefore, $p_c = 0$.

In Section 2, the four logit models in the literature with non-proportional odds are:

$$\begin{aligned} \log\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) &= \mathbf{h}_j^T(\mathbf{x}_i)\boldsymbol{\beta}_j \text{ , baseline-category} \\ \log\left(\frac{\pi_{i1} + \cdots + \pi_{ij}}{\pi_{i,j+1} + \cdots + \pi_{iJ}}\right) &= \mathbf{h}_j^T(\mathbf{x}_i)\boldsymbol{\beta}_j \text{ , cumulative} \\ \log\left(\frac{\pi_{ij}}{\pi_{i,j+1}}\right) &= \mathbf{h}_j^T(\mathbf{x}_i)\boldsymbol{\beta}_j \text{ , adjacent-categories} \\ \log\left(\frac{\pi_{ij}}{\pi_{i,j+1} + \cdots + \pi_{iJ}}\right) &= \mathbf{h}_j^T(\mathbf{x}_i)\boldsymbol{\beta}_j \text{ , continuation-ratio} \end{aligned}$$

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.

In equation (1), the corresponding model matrix is

$$\mathbf{X}_i = \begin{pmatrix} \mathbf{h}_1^T(\mathbf{x}_i) & \mathbf{0}^T & \cdots & \mathbf{0}^T \\ \mathbf{0}^T & \mathbf{h}_2^T(\mathbf{x}_i) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0}^T \\ \mathbf{0}^T & \cdots & \mathbf{0}^T & \mathbf{h}_{J-1}^T(\mathbf{x}_i) \\ \mathbf{0}^T & \cdots & \cdots & \mathbf{0}^T \end{pmatrix}_{J \times p} \tag{S.6}$$

and the parameter vector reduces to $\boldsymbol{\theta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_{J-1})^T$, which consists of $p = p_1 + \cdots + p_{J-1}$ unknown parameters in total. Note that we always use p to represent the total number of parameters.

In Section 3, the $p \times m(J-1)$ matrix

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_1 & & \\ & \ddots & \\ & & \mathbf{H}_{J-1} \end{pmatrix} \tag{S.7}$$

where $\mathbf{H}_j = (\mathbf{h}_j(\mathbf{x}_1), \cdots, \mathbf{h}_j(\mathbf{x}_m))$, $j = 1, \ldots, J-1$.

As a special case of Theorem 3.3, we have

Theorem S.8. *Consider the multinomial logistic model (1) with m distinct experimental settings $\mathbf{x}_i$ with $n_i > 0$ experimental units, $i = 1, \ldots, m$. For non-proportional odds (npo) models, the Fisher information matrix $\mathbf{F}$ is positive definite if and only if $m \geq \max\{p_1, \ldots, p_{J-1}\}$ and $\mathbf{x}_i$'s keep $\mathbf{H}_j$ of full row rank p_j, $j = 1, \ldots, J-1$.*

In Section 4, for non-proportional odds models, the $mJ \times p$ matrix

$$\mathbf{G} = \begin{pmatrix} \mathbf{c}_{11}\mathbf{h}_1^T(\mathbf{x}_1) & \cdots & \mathbf{c}_{1,J-1}\mathbf{h}_{J-1}^T(\mathbf{x}_1) \\ \mathbf{c}_{21}\mathbf{h}_1^T(\mathbf{x}_2) & \cdots & \mathbf{c}_{2,J-1}\mathbf{h}_{J-1}^T(\mathbf{x}_2) \\ \cdots & \cdots & \cdots \\ \mathbf{c}_{m1}\mathbf{h}_1^T(\mathbf{x}_m) & \cdots & \mathbf{c}_{m,J-1}\mathbf{h}_{J-1}^T(\mathbf{x}_m) \end{pmatrix} \tag{S.8}$$

As a special case of Corollary S.4, we have

Corollary S.6. *The coefficient $c_{\alpha_1,\ldots,\alpha_m} = 0$ if $\#\{i \mid \alpha_i > 0\} \leq k_{\min} - 1$, where $k_{\min} = \max\{p_1, \ldots, p_{J-1}\}$ for npo models.*

As special cases of *ppo* models, *npo* models imply $p_c = 0$ and $p_H \leq \min\{p_1, \ldots, p_{J-1}\}$. That is, $k_{\min}$'s are consistent across different odds models.

S.9. Model selection.

See Tables 4 and 5.

Table 4
Model Comparison for Trauma Clinical Trial Data

	Cumulative *po*	Cumulative *npo*	Continuation *po*	Continuation *npo*	Adjacent *po*	Adjacent *npo*
AIC	107.75	**99.41**	108.98	101.36	107.67	101.54
BIC	104.68	**94.51**	105.91	96.45	104.60	96.63

S.10. Lift-one and exchange algorithms.

Following Yang et al. (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 $0 \leq z \leq 1$ and $\mathbf{w} = (w_1, \ldots, w_m)^T \in S_+$. Parallel to Theorem 6 in Yang et al. (2017), we obtain the following result by Theorem 4.2:

TABLE 5

Model Comparison for Emergence of House Flies Data

	Cumulative *po*	Cumulative *npo*	Continuation *po*	Continuation *npo*	Adjacent *po*	Adjacent *npo*
AIC	195.87	121.17	116.40	**114.42**	209.64	194.47
BIC	195.71	120.96	116.24	**114.20**	209.47	194.25

THEOREM S.9. *Given an approximate allocation* $\mathbf{w} = (w_1, \ldots, w_m)^T \in S_+$ *and an* $i \in \{1, \ldots, m\}$, *for* $0 < z < 1$,

$$f_i(z) = (1-z)^{p-J+1} \sum_{j=0}^{J-1} b_j z^j (1-z)^{J-1-j} \tag{S.9}$$

$$f_i'(z) = (1-z)^{p-J} \sum_{j=1}^{J-1} b_j (j - pz) z^{j-1} (1-z)^{J-1-j} - p b_0 (1-z)^{p-1} \tag{S.10}$$

where $b_0 = f_i(0)$, $(b_{J-1}, \ldots, b_1)^T = \mathbf{B}_{J-1}^{-1}\mathbf{c}$, $\mathbf{B}_{J-1} = (s^{t-1})_{s,t=1,\ldots,J-1}$ *is a* $(J-1)\times(J-1)$ *constant matrix, and* $\mathbf{c} = (c_1, \ldots, c_{J-1})^T$ *with* $c_j = (j+1)^p j^{J-1-p} f_i(1/(j+1)) - j^{J-1} f_i(0)$, $j = 1, \ldots, J-1$.

Theorem S.9 shows that $f_i(z)$ is an order-p polynomial of z. Since $f_i(1) = 0$, the solution to maximization of $f_i(z), 0 \le z \le 1$ can occur only at $z = 0$ or $0 < z < 1$ such that $f_i'(z) = 0$, that is,

$$\sum_{j=1}^{J-1} j b_j z^{j-1} (1-z)^{J-j-1} = p \sum_{j=0}^{J-1} b_j z^j (1-z)^{J-j-1}, \quad 0 < z < 1. \tag{S.11}$$

This is an order-$(J-1)$ polynomial equation in z. For $J \le 5$, (S.11) is a polynomial equation of order-4 or less, which can be solved analytically. For $J \ge 6$, a quasi-Newton algorithm can be applied for searching numerical solutions.

Lift-one algorithm for D-optimal allocation $\mathbf{w} = (w_1, \ldots, w_m)^T$*:*

1° Start with an arbitrary allocation $\mathbf{w}_0 = (w_1, \ldots, w_m)^T$ satisfying $0 < w_i < 1$, $i = 1, \ldots, m$ and compute $f(\mathbf{w}_0)$.
2° Set up a random order of i going through $\{1, 2, \ldots, m\}$.
3° For each i, determine $f_i(z)$ according to Theorem S.9. In this step, J determinants $f_i(0), f_i(1/2), f_i(1/3), \ldots, f_i(1/J)$ are calculated.
4° Use quasi-Newton algorithm to find z_* maximizing $f_i(z)$ with $0 \le z \le 1$. If $f_i(z_*) \le f_i(0)$, let $z_* = 0$. Define $\mathbf{w}_*^{(i)} = (w_1(1-z_*)/(1-w_i), \ldots, w_{i-1}(1-z_*)/(1-w_i), z_*, w_{i+1}(1-z_*)/(1-w_i), \ldots, w_m(1-z_*)/(1-w_i))^T$. Note that $f(\mathbf{w}_*^{(i)}) = f_i(z_*)$.
5° Replace $\mathbf{w}_0$ with $\mathbf{w}_*^{(i)}$, and $f(\mathbf{w}_0)$ with $f(\mathbf{w}_*^{(i)})$.
6° Repeat 2° $\sim$ 5° until convergence, that is, $f(\mathbf{w}_0) = f(\mathbf{w}_*^{(i)})$ for each i.

Following Yang et al. (2016, 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 $\mathbf{n} = (n_1, \ldots, n_m)^T$. As a conclusion of Theorem 4.2, Lemma S.3 and Corollary S.4, we obtain the following result:

THEOREM S.10. *Suppose* $\mathbf{n} = (n_1, \ldots, n_m)^T$ *satisfies* $f(\mathbf{n}) > 0$ *and* $n_i + n_j \ge q$ *for given* $1 \le i < j \le m$*, where* $q = \min\{2J-2, p - k_{\min} + 2, p\}$*. Then*

$$f_{ij}(z) = \sum_{s=0}^{q} g_s z^s, \quad z = 0, 1, \ldots, n_i + n_j \tag{S.12}$$

where $g_0 = f_{ij}(0)$*, and* $g_1, \ldots, g_q$ *can be obtained using* $(g_1, \ldots, g_q)^T = \mathbf{B}_q^{-1}(d_1, \ldots, d_q)^T$ *with* $\mathbf{B}_q = (s^{t-1})_{s,t=1,\ldots,q}$ *as a* $q \times q$ *constant matrix and* $d_s = (f_{ij}(s) - f_{ij}(0))/s$*.*

Exchange algorithm for D-optimal allocation $(n_1, \ldots, n_m)^T$ *given* $n > 0$*:*

1° Start with an initial allocation $\mathbf{n} = (n_1, \ldots, n_m)^T$ such that $f(\mathbf{n}) > 0$.
2° Set up a random order of (i, j) going through all pairs $\{(1, 2), (1, 3), \ldots, (1, m), (2, 3), \ldots, (m-1, m)\}$.
3° For each (i, j), let $c = n_i + n_j$. If $c = 0$, let $\mathbf{n}^*_{ij} = \mathbf{n}$. Otherwise, there are two cases. *Case one:* $0 < c \le q$, we calculate $f_{ij}(z)$ for $z = 0, 1, \ldots, c$ directly and find z^* which maximizes $f_{ij}(z)$. *Case two:* $c > q$, we first calculate $f_{ij}(z)$ for $z = 0, 1, \ldots, q$; secondly determine $g_0, g_1, \ldots, g_q$ in (S.12) according to Theorem S.10; thirdly calculate $f_{ij}(z)$ for $z = q+1, \ldots, c$ based on (S.12); fourthly find z^* maximizing $f_{ij}(z)$ for $z = 0, \ldots, c$. For both cases, we define

$$\mathbf{n}^*_{ij} = (n_1, \ldots, n_{i-1}, z^*, n_{i+1}, \ldots, n_{j-1}, c - z^*, n_{j+1}, \ldots, n_m)^T$$

Note that $f(\mathbf{n}^*_{ij}) = f_{ij}(z^*) \ge f(\mathbf{n}) > 0$. If $f(\mathbf{n}^*_{ij}) > f(\mathbf{n})$, replace $\mathbf{n}$ with $\mathbf{n}^*_{ij}$, and $f(\mathbf{n})$ with $f(\mathbf{n}^*_{ij})$.
4° Repeat 2° ∼ 3° until convergence, that is, $f(\mathbf{n}^*_{ij}) = f(\mathbf{n})$ in step 3° for all (i, j).

S.11. Formulae for calculating $\boldsymbol{\pi_{ij}}$'s from $\mathbf{X}_i$'s. Following the notations in model (1), $\boldsymbol{\eta}_i = \mathbf{X}_i\boldsymbol{\theta} = \mathbf{C}^T \log(\mathbf{L}\boldsymbol{\pi}_i)$. The formulae towards calculating π_{ij}'s are listed as follows:
(1) Baseline-category logit model

$$\log(\boldsymbol{\pi}_i) = \begin{pmatrix} 1 & & & & -1 \\ & 1 & & & -1 \\ & & \ddots & & \vdots \\ & & & 1 & -1 \\ & & & & -1 \end{pmatrix}_{J\times J} \cdot \log\left(\begin{pmatrix} 1 & & & & 0 \\ & 1 & & & 0 \\ & & \ddots & & \vdots \\ & & & 1 & 0 \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}_{J\times J} \cdot \exp(\boldsymbol{\eta}_i)\right)$$

(2) Adjacent-categories logit model

$$\log(\boldsymbol{\pi}_i) = \begin{pmatrix} 1 & & & & -1 \\ & 1 & & & -1 \\ & & \ddots & & \vdots \\ & & & 1 & -1 \\ & & & & -1 \end{pmatrix}_{J\times J} \cdot \log\left(\begin{pmatrix} 1 & & & & 0 \\ & 1 & & & 0 \\ & & \ddots & & \vdots \\ & & & 1 & 0 \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}_{J\times J} \cdot \exp\left(\begin{pmatrix} 1 & 1 & \cdots & 1 & 0 \\ & 1 & \cdots & 1 & 0 \\ & & \ddots & \vdots & \vdots \\ & & & 1 & 0 \\ & & & & 1 \end{pmatrix}_{J\times J} \cdot \boldsymbol{\eta}_i\right)\right)$$

(3) Continuation-ratio logit model

$$\log(\boldsymbol{\pi}_i) = \boldsymbol{\eta}_i - \begin{pmatrix} 1 & & & & \\ 1 & 1 & & & \\ \vdots & & \ddots & & \\ 1 & 1 & \cdots & 1 & \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}_{J\times J} \cdot \log\left(\begin{pmatrix} 1 & 0 & \cdots & 0 & 1 \\ 0 & 1 & \cdots & 0 & 1 \\ & & \ddots & & \\ 0 & 0 & \cdots & 1 & 1 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}_{J\times J} \cdot \exp(\boldsymbol{\eta}_i)\right)$$

(4) Cumulative logit model

$$\log\left(\begin{pmatrix} 1 & & & & \\ 1 & 1 & & & \\ \vdots & & \ddots & & \\ 1 & 1 & \cdots & 1 & \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}_{J\times J} \cdot \boldsymbol{\pi}_i\right) = \begin{pmatrix} 1 & & & -1 & & \\ & \ddots & & & \ddots & \\ & & 1 & & & -1 \\ 0 & \cdots & 0 & 0 & \cdots & -1 \end{pmatrix}_{J\times 2(J-1)} \cdot \log\left(\begin{pmatrix} 1 & & & 0 \\ & \ddots & & \vdots \\ & & 1 & 0 \\ 1 & & & 1 \\ & \ddots & & \vdots \\ & & 1 & 1 \end{pmatrix}_{2(J-1)\times J} \cdot \exp(\boldsymbol{\eta}_i)\right)$$

Note that $\mathbf{X}_i\boldsymbol{\theta}$ in the above models could be *po*, *npo*, or *ppo*.

S.12. Reparametrization and D-optimality. In general, let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T$ be one set of parameters and $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_p)^T$ be another set of parameters, such that, $\theta_l = h_l(\boldsymbol{\vartheta})$, $l = 1, \cdots, p$; the map $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\vartheta}) = (h_1(\boldsymbol{\vartheta}), \ldots, h_p(\boldsymbol{\vartheta}))^T$ is one-to-one; h_l's are differentiable; and the $p \times p$ Jacobian matrix $\mathbf{J} = (h_i(\boldsymbol{\vartheta})/\partial\vartheta_j)_{ij}$ is nonsingular.

Consider a design $\xi = \{(\mathbf{x}_i, w_i), i = 1, \ldots, m\}$ with the distinct experimental settings $\mathbf{x}_i$'s and the corresponding proportions $w_i \in [0, 1]$. According to Schervish (1995, page 115), the Fisher information matrix $\mathbf{F}_\xi(\boldsymbol{\vartheta})$ at $\boldsymbol{\vartheta}$ and the Fisher information matrix $\mathbf{F}_\xi(\boldsymbol{\theta})$ at $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\vartheta})$ satisfy $\mathbf{F}_\xi(\boldsymbol{\vartheta}) = \mathbf{J}^T \mathbf{F}_\xi(\boldsymbol{\theta}(\boldsymbol{\vartheta}))\mathbf{J}$. Then $|\mathbf{F}_\xi(\boldsymbol{\vartheta})| = |\mathbf{J}|^2 \cdot |\mathbf{F}_\xi(\boldsymbol{\theta}(\boldsymbol{\vartheta}))|$, where $\mathbf{J}$ contains no design points but parameters. A locally D-optimal design maximizing $|\mathbf{F}_\xi(\boldsymbol{\vartheta})|$ also maximizes $|\mathbf{F}_\xi(\boldsymbol{\theta}(\boldsymbol{\vartheta}))|$. That is, it is mathematically equivalent to find D-optimal designs for parameters $\boldsymbol{\vartheta}$ or $\boldsymbol{\theta}$.

In terms of Bayesian D-optimal criterion, if a prior distribution of $\boldsymbol{\vartheta}$ is available, it induces a prior distribution of $\boldsymbol{\theta}$ since $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\vartheta})$ is one-to-one. Then $E_{\boldsymbol{\vartheta}} \log |\mathbf{F}_\xi(\boldsymbol{\vartheta})| = E_{\boldsymbol{\vartheta}} \log \left|\mathbf{J}^T \mathbf{F}_\xi(\boldsymbol{\theta}(\boldsymbol{\vartheta}))\mathbf{J}\right| = E_{\boldsymbol{\vartheta}} \log |\mathbf{J}|^2 + E_{\boldsymbol{\vartheta}} \log |\mathbf{F}_\xi(\boldsymbol{\theta}(\boldsymbol{\vartheta}))| = E_{\boldsymbol{\vartheta}} \log |\mathbf{J}|^2 + E_{\boldsymbol{\theta}} \log |\mathbf{F}_\xi(\boldsymbol{\theta})|$. Therefore, a Bayesian D-optimal design that maximizes $E_{\boldsymbol{\theta}} \log |\mathbf{F}_\xi(\boldsymbol{\theta})|$ also maximizes $E_{\boldsymbol{\vartheta}} \log |\mathbf{F}_\xi(\boldsymbol{\vartheta})|$.

EXAMPLE S.1. Perevozskaya et al. (2003) considered the *po* model:

$$\log \frac{\gamma_j(x)}{1 - \gamma_j(x)} = \frac{x - \alpha'_j}{\beta'} \qquad j = 2, \ldots, J \tag{S.13}$$

where $\gamma_j(x) = P(Y \geq j|x)$. Let us reparametrize this model as

$$\log \frac{\gamma_j(x)}{1 - \gamma_j(x)} = \alpha_j + \beta x \qquad j = 2, \ldots, J \tag{S.14}$$

Let $\boldsymbol{\theta} = (\alpha_2, \alpha_3, \beta)^T$ be the parameters in (S.13), and $\boldsymbol{\vartheta} = (\alpha'_2, \alpha'_3, \beta')^T$ be the parameters in (S.14). Then $\beta = 1/\beta', \alpha_2 = -\alpha'_2/\beta', \alpha_3 = -\alpha'_3/\beta'$, and the Jacobian matrix

$$\mathbf{J} = \begin{pmatrix} -\frac{1}{\beta'} & 0 & \frac{\alpha'_2}{\beta'^2} \\ 0 & -\frac{1}{\beta'} & \frac{\alpha'_3}{\beta'^2} \\ 0 & 0 & -\frac{1}{\beta'^2} \end{pmatrix}$$

Based on Theorem 2.1, the Fisher information $I_i(\boldsymbol{\theta})$ at x_i is

$$\begin{pmatrix} \frac{\pi_{i1}\pi_{i2,3}^2\pi_{i1,2}}{\pi_{i2}} & -\frac{\pi_{i1}\pi_{i1,2}\pi_{i2,3}\pi_{i3}}{\pi_{i2}} & \pi_{i1}\pi_{i1,2}\pi_{i2,3}x_i \\ -\frac{\pi_{i1}\pi_{i1,2}\pi_{i2,3}\pi_{i3}}{\pi_{i2}} & \frac{\pi_{i1,2}^2\pi_{i2,3}\pi_{i3}}{\pi_{i2}} & \pi_{i3}\pi_{i1,2}\pi_{i2,3}x_i \\ \pi_{i1}\pi_{i1,2}\pi_{i2,3}x_i & \pi_{i3}\pi_{i1,2}\pi_{i2,3}x_i & (\pi_{i1}\pi_{i2,3}^2 + \pi_{i2}(\pi_{i1} - \pi_{i3})^2 + \pi_{i1,2}^2\pi_{i3})x_i^2 \end{pmatrix}$$

where $\pi_{ij,k} = \pi_{ij} + \pi_{ik}$. It can be verified that $I_i(\boldsymbol{\vartheta}) = \mathbf{J}^T I_i(\boldsymbol{\theta})\mathbf{J}$ equals to the corresponding one given by Perevozskaya et al. (2003). For any given design $\xi = \{(\mathbf{x}_i, w_i), i = 1, \ldots, m\}$ with proportions $w_i \in [0, 1]$, the Fisher information matrix $I_\xi(\boldsymbol{\vartheta}) = \sum_{i=1}^m w_i I_i(\boldsymbol{\vartheta}) = \mathbf{J}^T I_\xi(\boldsymbol{\theta})\mathbf{J}$. Then $|I_\xi(\boldsymbol{\vartheta})| = |\mathbf{J}|^2 \cdot |I_\xi(\boldsymbol{\theta})|$ and the D-optimal design maximizing $|I_\xi(\boldsymbol{\theta})|$ also maximizes $|I_\xi(\boldsymbol{\vartheta})|$. That is, the D-optimal designs for Models (S.13) and (S.14) are the same. □

S.13. More discussion on D-optimality of uniform designs.

THEOREM S.11. *Consider Multinomial logit model* (1) *with only two response categories* ($J = 2$)*. In this case, the minimum number of support points is* $m = p$*. The objective function* $f(\mathbf{w}) \propto w_1 \cdots w_m$ *and the D-optimal allocation among minimally supported designs is* $\mathbf{w} = (1/m, \ldots, 1/m)^T$*.*

It can be verified that with $J = 2$ all of the four logit models are equivalent to the usual logistic model for binary response. In this case, *po*, *npo*, or *ppo* are essentially the same. Theorem S.11 confirms the corresponding results for binary responses in the literature (see, for example, Yang and Mandal (2015)). We provide an independent proof in Section S.15.

Besides the cases with $J = 2$, for certain *npo* models with $J \geq 3$, uniform allocations could still be D-optimal among minimally supported designs if $p_1 = \cdots = p_{J-1}$.

COROLLARY S.7. *Consider multinomial logit models* (1) *with npo assumption. Suppose* $p_1 = \cdots = p_{J-1}$ *and there exist* p_1 *distinct experimental settings such that* $\mathrm{rank}(\mathbf{H}_1) = \cdots = \mathrm{rank}(\mathbf{H}_{J-1}) = p_1$*. Then the minimal number of experimental settings is* $m = p_1$ *and the uniform allocation is D-optimal among minimally supported designs.*

According to Corollary S.7, for "regular" *npo* models (that is, $p_1 = \cdots = p_{J-1}$), uniform allocations are still D-optimal among minimally supported designs even with $J \geq 3$. However, the following lemma and example further represent that, if the condition $p_1 = \cdots = p_{J-1}$ is violated, uniform allocations are not D-optimal in general even for *npo* models.

LEMMA S.4. *Given* $0 < c_1 \leq c_2 \leq c_3$*, we consider the maximization problem* $f(w_1, w_2, w_3) = w_1 w_2 w_3 (c_1 w_2 w_3 + c_2 w_1 w_3 + c_3 w_1 w_2)$ *with respect to* $0 \leq w_i \leq 1$ *and* $w_1 + w_2 + w_3 = 1$*. Then the solution is* $w_1 = w_2 = w_3 = 1/3$ *if and only if* $c_1 = c_2 = c_3$*.*

The proof of Lemma S.4 is relegated to Section S.15, where analytical solutions are provided for (w_1, w_2, w_3) for general values of c_1, c_2 and c_3.

EXAMPLE S.2. Consider the *npo* model adopted by Zocchi and Atkinson (1999) with $\mathbf{h}_1(x_i) = (1, x_i, x_i^2)^T$, $\mathbf{h}_2(x_i) = (1, x_i)^T$, $J = 3$, $p_1 = 3, p_2 = 2$, and $p = 5$. According to Corollary S.4, the minimum number of support points is $m = \max\{p_1, p_2\} = 3$, which is feasible. The objective function $f(\mathbf{w})$ is an order-5 polynomial with terms $c_{\alpha_1,\alpha_2,\alpha_3} w_1^{\alpha_1} w_2^{\alpha_2} w_3^{\alpha_3}$. Lemma S.3 implies that $\alpha_i \in \{0, 1, 2\}, i = 1, 2, 3$ in order to keep $c_{\alpha_1,\alpha_2,\alpha_3} \neq 0$. Combined with Corollary S.4, we further know $\alpha_i \in \{1, 2\}, i = 1, 2, 3$. According to Theorem 4.2, the objective function is

$$f(w_1, w_2, w_3) = w_1 w_2 w_3 (c_{122} w_2 w_3 + c_{212} w_1 w_3 + c_{221} w_1 w_2) \tag{S.15}$$

for all the four logit models. Rewriting $(c_{122}, c_{212}, c_{221}) = C \cdot (c_1, c_2, c_3)$, it can be verified that for the continuation-ratio logit model adopted by Zocchi and Atkinson (1999) for the house flies experiment (Example 5.1), $C = (x_1 - x_2)^2 (x_1 - x_3)^2 (x_2 - x_3)^2 \prod_{i=1}^3 \prod_{j=1}^3 \pi_{ij}$, $c_1 = (x_2 - x_3)^2 (\pi_{12}^{-1} + \pi_{13}^{-1})$, $c_2 = (x_1 - x_3)^2 (\pi_{22}^{-1} + \pi_{23}^{-1})$, $c_3 = (x_1 - x_2)^2 (\pi_{32}^{-1} + \pi_{33}^{-1})$; for a cumulative logit model (see, for example, Example 5.2), $C = (x_1 - x_2)^2 (x_1 - x_3)^2 (x_2 - x_3)^2 \prod_{i=1}^3 \pi_{i1} \pi_{i2}^{-1} \pi_{i3} (\pi_{i1} + \pi_{i2})^2 (\pi_{i2} + \pi_{i3})^2$, $c_1 = (x_2 - x_3)^2 \pi_{13}^{-1} (\pi_{11} + \pi_{12})^{-1}$, $c_2 = (x_1 - x_3)^2 \pi_{23}^{-1} (\pi_{21} + \pi_{22})^{-1}$, and $c_3 = (x_1 - x_2)^2 \pi_{33}^{-1} (\pi_{31} + \pi_{32})^{-1}$. According to Lemma S.4, $w_1 = w_2 = w_3 = 1/3$ is D-optimal if and only if $c_1 = c_2 = c_3$, which is in general not true for both continuation-ratio and cumulative logit models with non-proportional odds. □

S.14. More examples.

EXAMPLE S.3. *(For Section 3)* Consider an experiment with a main-effects multinomial logistic model with d factors and m distinct experimental settings $\mathbf{x}_1, \ldots, \mathbf{x}_m$, where $\mathbf{x}_i = (x_{i1}, \ldots, x_{id})^T$, $i = 1, \ldots, m$.

For a main-effects model, the linear predictors may take the form of

$$\eta_{ij} = \beta_{j1} + \beta_{j2} x_{i1} + \cdots + \beta_{j,k+1} x_{ik} + \zeta_1 x_{i,k+1} + \cdots + \zeta_{d-k} x_{id} \tag{S.16}$$

where $i = 1, \ldots, m$, $j = 1, \ldots, J-1$. In other words, the intercept and the coefficients of the first k factors depend on j, while the coefficients of the last $d-k$ factors do not.

We claim that the minimum number of experimental settings is simply $d+1$ for the main-effects multinomial logistic model (S.16) with $0 \leq k \leq d$, regardless of J.

Actually, first we consider $1 \leq k \leq d-1$. It is a *ppo* model. In this case, $p_1 = \cdots = p_{J-1} = k+1$, $p_c = d-k$,

$$\mathbf{H}_1 = \cdots = \mathbf{H}_{J-1} = \begin{pmatrix} 1 & \cdots & 1 \\ x_{11} & \cdots & x_{m1} \\ \vdots & \vdots & \vdots \\ x_{1k} & \cdots & x_{mk} \end{pmatrix}, \mathbf{H}_c = \begin{pmatrix} x_{1,k+1} & \cdots & x_{m,k+1} \\ \vdots & \vdots & \vdots \\ x_{1d} & \cdots & x_{md} \end{pmatrix}$$

According to the special case of Theorem 3.3, the Fisher information matrix $\mathbf{F}$ is positive definite if and only if $m \geq p_c + p_1 = d+1$ and the matrix

$$(\mathbf{H}_1^T, \mathbf{H}_c^T) == \begin{pmatrix} 1 & x_{11} & \cdots & x_{1d} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{m1} & \cdots & x_{md} \end{pmatrix}$$

is of full rank $d+1$.

Now we let $k = 0$. The model (S.16) leads to a *po* model. By applying Theorem S.7, we obtain the same conditions as for the *ppo* model. Similarly, if we let $k = d$ and apply Theorem S.8, we get the same conditions for *npo* models. □

EXAMPLE S.4. *(For Section 3)* Consider an experiment with four factors ($d = 4$), three response categories ($J = 3$), and four distinct experimental settings ($m = 4$). Then the experimental settings are $\mathbf{x}_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4})^T$, $i = 1, 2, 3, 4$. Consider a multinomial logistic model with *ppo* such that

$$\mathbf{H}_1^T = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{13} \\ 1 & x_{21} & x_{22} & x_{23} \\ 1 & x_{31} & x_{32} & x_{33} \\ 1 & x_{41} & x_{42} & x_{43} \end{pmatrix}, \mathbf{H}_2^T = \begin{pmatrix} 1 & x_{11} \\ 1 & x_{21} \\ 1 & x_{31} \\ 1 & x_{41} \end{pmatrix}, \mathbf{H}_c^T = \begin{pmatrix} x_{14} \\ x_{24} \\ x_{34} \\ x_{44} \end{pmatrix}$$

That is, $p_1 = 4, p_2 = 2, p_c = 1, p_H = 2$, $\max\{p_1, p_2, p_c + p_H\} = p_1 = 4$, and there are $p = p_1 + p_2 + p_c = 7$ parameters. In this case,

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_1 & \\ & \mathbf{H}_2 \\ \mathbf{H}_c & \mathbf{H}_c \end{pmatrix}$$

is 7×8 with rank 7. That is, the minimum number in Theorem 3.3, $m = \max\{p_1, \ldots, p_{J-1}, p_c + p_H\} = 4$, is attained in this case. □

EXAMPLE S.5. *(For Section 3)* Consider an experiment with three factors ($d = 3$), three response categories ($J = 3$), and three distinct experimental settings ($m = 3$). Denote the experimental settings as $\mathbf{x}_i = (x_{i1}, x_{i2}, x_{i3})^T$, $i = 1, 2, 3$. Consider a multinomial logistic model with *ppo* such that

$$\mathbf{H}_1^T = \begin{pmatrix} 1 & x_{11} \\ 1 & x_{21} \\ 1 & x_{31} \end{pmatrix}, \mathbf{H}_2^T = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \mathbf{H}_c^T = \begin{pmatrix} x_{12} & x_{13} \\ x_{22} & x_{23} \\ x_{32} & x_{33} \end{pmatrix}$$

That is, $p_1 = 2, p_2 = 1, p_c = 2, p_H = 1$, $\max\{p_1, p_2, p_c + p_H\} = p_c + p_H = 3$, and there are $p = p_1 + p_2 + p_c = 5$ parameters. In this case,

$$\mathbf{H} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ x_{11} & x_{21} & x_{31} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ x_{12} & x_{22} & x_{32} & x_{12} & x_{22} & x_{32} \\ x_{13} & x_{23} & x_{33} & x_{13} & x_{23} & x_{33} \end{pmatrix}$$

is 5×6. It can be verified that $\text{rank}(\mathbf{H}) = 5$ using Theorem S.5. That is, the minimal number of experimental settings in this case is $m = \max\{p_1, \ldots, p_{J-1}, p_c + p_H\} = 3$. □

EXAMPLE S.6. *(For Section 4)* Consider an example with responses in $J = 4$ categories, $d = 5$ factors, and $m = 5$ distinct experimental settings $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,5})^T$, $i = 1, \ldots, 5$. Suppose a multinomial logistic model with

$$\mathbf{H}_1^T = \begin{pmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & x_{51} & x_{52} \end{pmatrix}, \mathbf{H}_2^T = \begin{pmatrix} 1 & x_{11} \\ 1 & x_{21} \\ \vdots & \vdots \\ 1 & x_{51} \end{pmatrix}, \mathbf{H}_3^T = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \mathbf{H}_c^T = \begin{pmatrix} x_{13} & x_{14} & x_{15} \\ x_{23} & x_{24} & x_{25} \\ \vdots & \vdots & \vdots \\ x_{53} & x_{54} & x_{55} \end{pmatrix}$$

is used. That is, $p_1 = 3, p_2 = 2, p_3 = 1$, $p_H = 1$, $p_c = 3$, and $p = 9$. In this case, $\mathbf{G}$ defined in Theorem 4.1 is 20×9 and $p_c + p_H = 4$ is the minimum number of $\#\{i \mid \alpha_i > 0\}$ to keep $|G[i_1, \ldots, i_p]| \neq 0$ if $(i_1, \ldots, i_p) \in \Lambda(\alpha_1, \ldots, \alpha_m)$. Actually, $(i_1, \ldots, i_9) = (1, 2, 3, 6, 7, 8, 10, 11, 12) \in \Lambda(3, 3, 3, 0, 0)$ leads to $rank(\mathbf{G}[i_1, \ldots, i_9]) = 8$, while $(1, 2, 5, 6, 9, 10, 13, 14, 15) \in \Lambda(2, 2, 2, 3, 0)$ leads to $rank(\mathbf{G}[i_1, \ldots, i_9]) = 9$. Therefore, $|G[i_1, \ldots, i_9]| \neq 0$ in general if $(i_1, \ldots, i_9) \in \Lambda(2, 2, 2, 3, 0)$ for such a *ppo* model. □

EXAMPLE 5.2. (*continued, for Section 5.1*) Recall that there are eight parameters with fitted values $\hat{\boldsymbol{\beta}} = (\hat{\beta}_{11}, \hat{\beta}_{21}, \hat{\beta}_{31}, \hat{\beta}_{41}, \hat{\beta}_{12}, \hat{\beta}_{22}, \hat{\beta}_{32}, \hat{\beta}_{42})^T = (-0.865, -0.094, 0.706, 1.909, -0.113, -0.269, -0.182, -0.119)^T$. If we treat the fitted parameter values as the assumed values, the design space is $\mathcal{X} = \{x \geq 0 \mid \beta_{11} + \beta_{12}x < \beta_{21} + \beta_{22}x < \beta_{31} + \beta_{32}x < \beta_{41} + \beta_{42}x\} = \{x \geq 0 \mid -9.195 < x < 4.942\} = [0, 4.942)$. It is not a surprise that the four levels $\{1, 2, 3, 4\}$ in the original dataset are included in the design space. □

EXAMPLE S.7. *(For Section 5.6)* Consider a multinomial logistic model with proportional odds for responses with $J = 3$ categories, $d = 1$ factors, and $m = 2$ distinct experimental settings x_1, x_2. Same as in Example S.1, the parameters are $\beta_1, \beta_2, \zeta_1$ and the linear predictors

$$\eta_{i1} = \beta_1 + \zeta_1 x_i,\ \ \eta_{i2} = \beta_2 + \zeta_1 x_i,\ \ i = 1, 2.$$

According to Theorem 4.2, the objective function of allocation (w_1, w_2) is an order-3 homogeneous polynomial of w_1, w_2 consisting of monomials $c_{\alpha_1,\alpha_2} w_1^{\alpha_1} w_2^{\alpha_2}$ with coefficients $c_{\alpha_1,\alpha_2} \geq 0$. Based on Lemma S.3 and Corollary S.4, $c_{\alpha_1,\alpha_2} \neq 0$ only if $\max\{\alpha_1, \alpha_2\} \leq 2$ and $\#\{i \mid \alpha_i > 0\} = 2$, which implies (α_1, α_2) is either $(2, 1)$ or $(1, 2)$. That is, the objective function is

$$f(w_1, w_2) = w_1 w_2 (c_{21} w_1 + c_{12} w_2),$$

which takes the same form as in Corollary 5.2 in Yang et al. (2017). If we rewrite $c_{21} = C \cdot c_2$ and $c_{12} = C \cdot c_1$, that is, $f(w_1, w_2) = C \cdot w_1 w_2 (c_2 w_1 + c_1 w_2)$, then for a baseline-category logit model, $C = \pi_{13}\pi_{23}(x_1 - x_2)^2$, $c_2 = \pi_{11}\pi_{12}(1 - \pi_{23})$, $c_1 = \pi_{21}\pi_{22}(1 - \pi_{13})$; for a cumulative logit model, $C = \pi_{12}^{-1}(1 - \pi_{13})(1 - \pi_{11})\pi_{22}^{-1}(1 - \pi_{23})(1 - \pi_{21})(x_1 - x_2)^2$, $c_2 = \pi_{11}(1 - \pi_{11})\pi_{13}(1 - \pi_{13})\pi_{22}(1 - \pi_{22})$, $c_1 = \pi_{12}(1 - \pi_{12})\pi_{21}(1 - \pi_{21})\pi_{23}(1 - \pi_{23})$; for an adjacent-categories logit model, $C = (x_1 - x_2)^2$, $c_2 = \pi_{11}\pi_{12}\pi_{13}(\pi_{21}\pi_{22} + \pi_{22}\pi_{23} + 4\pi_{21}\pi_{23})$, $c_1 = \pi_{21}\pi_{22}\pi_{23}(\pi_{11}\pi_{12} + \pi_{12}\pi_{13} + 4\pi_{11}\pi_{13})$; for a continuation-ratio logit model, $C = (1 - \pi_{11})^{-1}(1 - \pi_{21})^{-1}(x_1 - x_2)^2$, $c_2 = \pi_{11}\pi_{12}\pi_{13}(1 - \pi_{11})[\pi_{22}\pi_{23} + \pi_{21}(1 - \pi_{21})^2]$, $c_1 = \pi_{21}\pi_{22}\pi_{23}(1 - \pi_{21})[\pi_{12}\pi_{13} + \pi_{11}(1 - \pi_{11})^2]$. According to Corollary 5.2 in Yang et al. (2017), the uniform allocation $w_1^* = w_2^* = 1/2$ is D-optimal if and only if $c_1 = c_2$, which is not true in general for all the four logit models. □

Example 5.2. (*continued, for Section 6*) In practice, we may use designs not as extreme as the D-optimal design. Here are some alternative allocations of subjects, along with efficiencies:

Table 6
Alternative Designs for Trauma Clinical Trial

Design point x	1	2	3	4	Efficiency(%)
D-optimal design	401	0	0	401	100.0
Original design	210	190	207	195	74.7
1% reallocated	397	4	4	397	99.4
2.5% reallocated	391	10	10	391	98.8
5% reallocated	381	20	20	381	97.6
10% reallocated	361	40	40	361	95.3

We may recommend 2.5% or 5% reallocated design, which is not so extreme but still highly efficient.

S.15. Proofs.

Proof of Theorem 2.1:

Suppose for distinct $\mathbf{x}_i, i = 1, \cdots, m$, we have independent multinomial responses

$$\mathbf{Y}_i = (Y_{i1}, \cdots, Y_{iJ})^T \sim \text{Multinomial}(n_i; \pi_{i1}, \cdots, \pi_{iJ})$$

where $n_i = \sum_{j=1}^{J} Y_{ij}$. Then the log-likelihood for the multinomial model is

$$\begin{aligned} l(\boldsymbol{\theta}) &= \log L(\boldsymbol{\theta}) \\ &= \log \prod_{i=1}^{m} \frac{n_i!}{Y_{i1}! \cdots Y_{iJ}!} \pi_{i1}^{Y_{i1}} \cdots \pi_{iJ}^{Y_{iJ}} \\ &= \text{constant} + \sum_{i=1}^{m} \mathbf{Y}_i^T \log \boldsymbol{\pi}_i \end{aligned}$$

where $\log \boldsymbol{\pi}_i = (\log \pi_{i1}, \cdots, \log \pi_{iJ})^T$. Then the score vector

$$\begin{aligned} \frac{\partial l}{\partial \boldsymbol{\theta}^T} &= \sum_{i=1}^{m} \mathbf{Y}_i^T \text{diag}(\boldsymbol{\pi}_i)^{-1} \frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T} \\ \frac{\partial l}{\partial \boldsymbol{\theta}} &= (\frac{\partial l}{\partial \boldsymbol{\theta}^T})^T = \sum_{i=1}^{m} (\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T})^T \text{diag}(\boldsymbol{\pi}_i)^{-1} \mathbf{Y}_i \end{aligned}$$

Using the formulae of matrix differentiation, we get

$$\begin{aligned} \frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T} &= \frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\eta}_i^T} \cdot \frac{\partial \boldsymbol{\eta}_i}{\partial \boldsymbol{\theta}^T} \\ &= \left(\frac{\partial \boldsymbol{\eta}_i}{\partial \boldsymbol{\pi}_i^T} \right)^{-1} \cdot \mathbf{X}_i \\ &= \left(\frac{\partial [\mathbf{C}^T \log(\mathbf{L}\boldsymbol{\pi}_i)]}{\partial [\log(\mathbf{L}\boldsymbol{\pi}_i)]^T} \cdot \frac{\partial [\log(\mathbf{L}\boldsymbol{\pi}_i)]}{\partial [\mathbf{L}\boldsymbol{\pi}_i]^T} \cdot \frac{\partial [\mathbf{L}\boldsymbol{\pi}_i]}{\partial \boldsymbol{\pi}_i^T} \right)^{-1} \cdot \mathbf{X}_i \\ &= \left(\mathbf{C}^T [\text{diag}(\mathbf{L}\boldsymbol{\pi}_i)]^{-1} \mathbf{L} \right)^{-1} \mathbf{X}_i \end{aligned}$$

LEMMA S.5.

$$\boldsymbol{\pi}_i{}^T \text{diag}(\boldsymbol{\pi}_i)^{-1} (\mathbf{C}^T \mathbf{D}_i^{-1} \mathbf{L})^{-1} \mathbf{X}_i = \mathbf{0}^T$$

Proof of Lemma S.5: Recall that $\mathbf{1}^T \boldsymbol{\pi}_i = \pi_{i1} + \cdots + \pi_{iJ} = 1$ for each i; the last row of $\mathbf{X}_i$ is all 0; and

$$\mathbf{C}^T = \begin{pmatrix} * & * & \cdots & 0 \\ * & * & \cdots & 0 \\ & \ddots & & \\ 0 & 0 & \cdots & 1 \end{pmatrix}, \qquad \mathbf{L} = \begin{pmatrix} * & * & \cdots & * \\ * & * & \cdots & * \\ & \ddots & & \\ 1 & 1 & \cdots & 1 \end{pmatrix}$$

Then

$$\mathbf{D}_i^{-1} = \text{diag}(\mathbf{L}\boldsymbol{\pi}_i)^{-1} = \begin{pmatrix} * & 0 & \cdots & 0 \\ 0 & * & \cdots & 0 \\ & \ddots & & \\ 0 & 0 & \cdots & \frac{1}{\mathbf{1}^T \boldsymbol{\pi}_i} \end{pmatrix} = \begin{pmatrix} * & 0 & \cdots & 0 \\ 0 & * & \cdots & 0 \\ & \ddots & & \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

and

$$\mathbf{D}_i^{-1} \mathbf{L} = \begin{pmatrix} * & \cdots & * \\ * & \cdots & * \\ & \ddots & \\ & \mathbf{1}^T & \end{pmatrix} \text{ and } \mathbf{C}^T \mathbf{D}_i^{-1} \mathbf{L} = \begin{pmatrix} * & \cdots & * \\ * & \cdots & * \\ & \ddots & \\ & \mathbf{1}^T & \end{pmatrix}$$

Rewrite $(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})^{-1} = (\boldsymbol{c}_{i1}, \cdots, \boldsymbol{c}_{iJ})$. Then $\mathbf{1}^T\boldsymbol{c}_{i1} = \cdots = \mathbf{1}^T\boldsymbol{c}_{i,J-1} = 0$ and $\mathbf{1}^T\boldsymbol{c}_{iJ} = 1$ (just check the last row of $\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L}$). Since $\boldsymbol{\pi}_i^T\text{diag}(\boldsymbol{\pi}_i)^{-1} = (1, \cdots, 1)$, then

$$\boldsymbol{\pi}_i{}^T\text{diag}(\boldsymbol{\pi}_i)^{-1}(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})^{-1} = (1, \cdots, 1)(\boldsymbol{c}_{i1}, \cdots, \boldsymbol{c}_{iJ}) = (0, \cdots, 0, 1)$$

Since the last row of $\mathbf{X}_i$ is all 0, then $\boldsymbol{\pi}_i{}^T\text{diag}(\boldsymbol{\pi}_i)^{-1}(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})^{-1}\mathbf{X}_i = \mathbf{0}^T$. □

As a direct conclusion of Lemma S.5,

$$E(\frac{\partial l}{\partial \boldsymbol{\theta}^T}) = \sum_{i=1}^m n_i\boldsymbol{\pi}_i{}^T\text{diag}(\boldsymbol{\pi}_i)^{-1}(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})^{-1}\mathbf{X}_i = \mathbf{0}^T$$

Then the Fisher information matrix (see, for example, Schervish (1995, Section 2.3.1))

$$\begin{aligned}
\mathbf{F} &= \text{Cov}\left(\frac{\partial l}{\partial \boldsymbol{\theta}}, \frac{\partial l}{\partial \boldsymbol{\theta}}\right) = E\left(\frac{\partial l}{\partial \boldsymbol{\theta}} \cdot \frac{\partial l}{\partial \boldsymbol{\theta}^T}\right) \\
&= E\left(\sum_{i=1}^m (\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T})^T\text{diag}(\boldsymbol{\pi}_i)^{-1}\mathbf{Y}_i \cdot \sum_{j=1}^m \mathbf{Y}_j^T\text{diag}(\boldsymbol{\pi}_j)^{-1}\frac{\partial \boldsymbol{\pi}_j}{\partial \boldsymbol{\theta}^T}\right) \\
&= E\left(\sum_{i=1}^m\sum_{j=1}^m (\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T})^T\text{diag}(\boldsymbol{\pi}_i)^{-1}\mathbf{Y}_i\mathbf{Y}_j^T\text{diag}(\boldsymbol{\pi}_j)^{-1}\frac{\partial \boldsymbol{\pi}_j}{\partial \boldsymbol{\theta}^T}\right)
\end{aligned}$$

Since $\mathbf{Y}_i$'s follow independent multinomial distributions, then

$$\begin{aligned}
E(\mathbf{Y}_i\mathbf{Y}_i^T) &= \begin{pmatrix} n_i(n_i-1)\pi_{i1}^2 + n_i\pi_{i1} & \cdots & n_i(n_i-1)\pi_{is}\pi_{it} \\ \vdots & \ddots & \vdots \\ n_i(n_i-1)\pi_{is}\pi_{it} & \cdots & n_i(n_i-1)\pi_{iJ}^2 + n_i\pi_{iJ} \end{pmatrix} \\
&= n_i(n_i-1)\boldsymbol{\pi}_i\boldsymbol{\pi}_i^T + n_i\text{diag}(\boldsymbol{\pi}_i)
\end{aligned}$$

On the other hand, for $i \neq j$,

$$E(\mathbf{Y}_i\mathbf{Y}_j^T) = E(\mathbf{Y}_i) \cdot E(\mathbf{Y}_j^T) = n_in_j\boldsymbol{\pi}_i\boldsymbol{\pi}_j^T$$

Then the Fisher information matrix

$$\begin{aligned}
\mathbf{F} &= \sum_{i=1}^m (\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T})^T\text{diag}(\boldsymbol{\pi}_i)^{-1}n_i(n_i-1)\boldsymbol{\pi}_i\boldsymbol{\pi}_i^T\text{diag}(\boldsymbol{\pi}_i)^{-1}\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T} \\
&+ \sum_{i=1}^m (\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T})^T\text{diag}(\boldsymbol{\pi}_i)^{-1}n_i\text{diag}(\boldsymbol{\pi}_i)\text{diag}(\boldsymbol{\pi}_i)^{-1}\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T} \\
&+ \sum_{i \neq j} (\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T})^T\text{diag}(\boldsymbol{\pi}_i)^{-1}n_in_j\boldsymbol{\pi}_i\boldsymbol{\pi}_j^T\text{diag}(\boldsymbol{\pi}_j)^{-1}\frac{\partial \boldsymbol{\pi}_j}{\partial \boldsymbol{\theta}^T} \\
&\triangleq (a) + (b) + (c)
\end{aligned}$$

where

$$(b) = \sum_{i=1}^{m} (\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T})^T \mathrm{diag}(\boldsymbol{\pi}_i)^{-1} \frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T} n_i$$

$$(a) + (c) = \left[\sum_{i=1}^{m} (\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T})^T \mathrm{diag}(\boldsymbol{\pi}_i)^{-1} \boldsymbol{\pi}_i n_i\right] \left[\sum_{i=1}^{m} (\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T})^T \mathrm{diag}(\boldsymbol{\pi}_i)^{-1} \boldsymbol{\pi}_i n_i\right]^T - \sum_{i=1}^{m} (\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T})^T \mathrm{diag}(\boldsymbol{\pi}_i)^{-1} n_i \boldsymbol{\pi}_i {\boldsymbol{\pi}_i}^T \mathrm{diag}(\boldsymbol{\pi}_i)^{-1} \frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T}$$

Actually, let

$$\mathbf{E}_i = {\boldsymbol{\pi}_i}^T \mathrm{diag}(\boldsymbol{\pi}_i)^{-1} \frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T} = {\boldsymbol{\pi}_i}^T \mathrm{diag}(\boldsymbol{\pi}_i)^{-1} (\mathbf{C}^T \mathbf{D}_i^{-1} \mathbf{L})^{-1} \mathbf{X}_i$$

which is $\mathbf{0}^T$ for each i according to Lemma S.5. Then

$$(a) + (c) = \left[\sum_{i=1}^{m} n_i \mathbf{E}_i^T\right] \left[\sum_{i=1}^{m} n_i \mathbf{E}_i^T\right]^T - \sum_{i=1}^{m} n_i \mathbf{E}_i^T \mathbf{E}_i = \mathbf{0}_{J \times J}$$

The arguments above have proved Theorem 2.1. □

Proof of Theorem 3.1: Because the last row of $\mathbf{X}_i$ consists of all zeros, the entries in the last row and last column of $\mathbf{U}_i$ actually won't make any difference. In order to simplify the notations in this proof, we rewrite

$$\begin{array}{lcll}
\mathbf{h}_{ji} & \triangleq & \mathbf{h}_j(\mathbf{x}_i) & j = 1, \ldots, J-1; \quad i = 1, \ldots, m \\
\mathbf{h}_{ci} & \triangleq & \mathbf{h}_c(\mathbf{x}_i) & i = 1, \ldots, m \\
u_{sti} & \triangleq & u_{st}(\boldsymbol{\pi}_i) & s, t = 1, \ldots, J-1; \quad i = 1, \ldots, m \\
u_{s\cdot i} & \triangleq & \sum_{t=1}^{J-1} u_{sti} & s = 1, \ldots, J-1; \quad i = 1, \ldots, m \\
u_{\cdot ti} & \triangleq & \sum_{s=1}^{J-1} u_{sti} & t = 1, \ldots, J-1; \quad i = 1, \ldots, m \\
u_{\cdot\cdot i} & \triangleq & \sum_{s=1}^{J-1} \sum_{t=1}^{J-1} u_{sti} & i = 1, \ldots, m
\end{array}$$

Based on Corollary 3.1, when $\mathbf{X}_i$ takes partial proportional odds form (2), the Fisher information $\mathbf{F}_i = \mathbf{X}_i^T \mathbf{U}_i \mathbf{X}_i =$

$$\begin{pmatrix}
u_{11i} \mathbf{h}_{1i} \mathbf{h}_{1i}^T & \cdots & u_{1,J-1,i} \mathbf{h}_{1i} \mathbf{h}_{J-1,i}^T & u_{1\cdot i} \mathbf{h}_{1i} \mathbf{h}_{ci}^T \\
\vdots & \ddots & \vdots & \vdots \\
u_{J-1,1,i} \mathbf{h}_{J-1,i} \mathbf{h}_{1i}^T & \cdots & u_{J-1,J-1,i} \mathbf{h}_{J-1,i} \mathbf{h}_{J-1,i}^T & u_{J-1\cdot i} \mathbf{h}_{J-1,i} \mathbf{h}_{ci}^T \\
u_{\cdot 1i} \mathbf{h}_{ci} \mathbf{h}_{1i}^T & \cdots & u_{\cdot J-1,i} \mathbf{h}_{ci} \mathbf{h}_{J-1,i}^T & u_{\cdot\cdot i} \mathbf{h}_{ci} \mathbf{h}_{ci}^T
\end{pmatrix}$$

Then the Fisher information matrix $\mathbf{F} = \sum_{i=1}^m n_i \mathbf{F}_i =$

$$\begin{pmatrix} \sum_{i=1}^m n_i u_{11i} \mathbf{h}_{1i} \mathbf{h}_{1i}^T & \cdots & \sum_{i=1}^m n_i u_{1,J-1,i} \mathbf{h}_{1i} \mathbf{h}_{J-1,i}^T & \sum_{i=1}^m n_i u_{1\cdot i} \mathbf{h}_{1i} \mathbf{h}_{ci}^T \\ \vdots & \ddots & \vdots & \vdots \\ \sum_{i=1}^m n_i u_{J-1,1,i} \mathbf{h}_{J-1,i} \mathbf{h}_{1i}^T & \cdots & \sum_{i=1}^m n_i u_{J-1,J-1,i} \mathbf{h}_{J-1,i} \mathbf{h}_{J-1,i}^T & \sum_{i=1}^m n_i u_{J-1\cdot i} \mathbf{h}_{J-1,i} \mathbf{h}_{ci}^T \\ \sum_{i=1}^m n_i u_{\cdot 1i} \mathbf{h}_{ci} \mathbf{h}_{1i}^T & \cdots & \sum_{i=1}^m n_i u_{\cdot J-1,i} \mathbf{h}_{ci} \mathbf{h}_{J-1,i}^T & \sum_{i=1}^m n_i u_{\cdot\cdot i} \mathbf{h}_{ci} \mathbf{h}_{ci}^T \end{pmatrix}$$

or simply

$$\begin{pmatrix} \mathbf{H}_1 & & \\ & \ddots & \\ & & \mathbf{H}_{J-1} \\ \mathbf{H}_c & \cdots & \mathbf{H}_c \end{pmatrix} \begin{pmatrix} \mathbf{U}_{11} & \cdots & \mathbf{U}_{1,J-1} \\ \vdots & \ddots & \vdots \\ \mathbf{U}_{J-1,1} & \cdots & \mathbf{U}_{J-1,J-1} \end{pmatrix} \begin{pmatrix} \mathbf{H}_1^T & & & \mathbf{H}_c^T \\ & \ddots & & \vdots \\ & & \mathbf{H}_{J-1}^T & \mathbf{H}_c^T \end{pmatrix}$$

□

Proof of Theorem S.3: Recall that $(\mathbf{C}^T \mathbf{D}_i^{-1} \mathbf{L})^{-1} = (\mathbf{c}_{i1} \cdots \mathbf{c}_{iJ})$ and $u_{st}(\boldsymbol{\pi}_i) = \mathbf{c}_{is}^T \text{diag}(\boldsymbol{\pi}_i)^{-1} \mathbf{c}_{it}$, for $s, t = 1, \ldots, J-1$ and $i = 1, \ldots, m$. Denote

$$\tilde{\mathbf{C}} = \begin{pmatrix} \mathbf{c}_{11}^T & & \\ & \ddots & \\ & & \mathbf{c}_{m1}^T \\ \mathbf{c}_{12}^T & & \\ & \ddots & \\ & & \mathbf{c}_{m2}^T \\ \vdots & \ddots & \vdots \\ \mathbf{c}_{1,J-1}^T & & \\ & \ddots & \\ & & \mathbf{c}_{m,J-1}^T \end{pmatrix}_{m(J-1) \times mJ}$$

$$\text{and } \tilde{\mathbf{W}} = \begin{pmatrix} n_1 \text{diag}(\boldsymbol{\pi}_1)^{-1} & & \\ & \ddots & \\ & & n_m \text{diag}(\boldsymbol{\pi}_m)^{-1} \end{pmatrix}_{mJ \times mJ}$$

We claim that $\mathbf{U} = \tilde{\mathbf{C}} \tilde{\mathbf{W}} \tilde{\mathbf{C}}^T$. Actually

$$\tilde{\mathbf{C}} \tilde{\mathbf{W}} = \begin{pmatrix} n_1 \mathbf{c}_{11}^T \text{diag}(\boldsymbol{\pi}_1)^{-1} & & \\ & \ddots & \\ & & n_m \mathbf{c}_{m1}^T \text{diag}(\boldsymbol{\pi}_\text{m})^{-1} \\ \vdots & \ddots & \vdots \\ n_1 \mathbf{c}_{1,J-1}^T \text{diag}(\boldsymbol{\pi}_1)^{-1} & & \\ & \ddots & \\ & & n_m \mathbf{c}_{m,J-1}^T \text{diag}(\boldsymbol{\pi}_\text{m})^{-1} \end{pmatrix}$$

and

$$\tilde{\mathbf{C}}\tilde{\mathbf{W}}\tilde{\mathbf{C}}^T = \tilde{\mathbf{C}}\tilde{\mathbf{W}} \begin{pmatrix} \mathbf{c}_{11} & & & \cdots & \mathbf{c}_{1,J-1} & & \\ & \ddots & & \ddots & & \ddots & \\ & & \mathbf{c}_{m1} & \cdots & & & \mathbf{c}_{m,J-1} \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{U}_{11} & \cdots & \mathbf{U}_{1,J-1} \\ \vdots & \ddots & \vdots \\ \mathbf{U}_{J-1,1} & \cdots & \mathbf{U}_{J-1,J-1} \end{pmatrix} = \mathbf{U}$$

Note that $\tilde{\mathbf{W}}$ is diagonal with positive diagonal entries. Thus $\tilde{\mathbf{W}}$ is positive definite. By adjusting the rows, we can verify that $rank(\tilde{\mathbf{C}})$ is the same as $rank(\tilde{\mathbf{C}}')$, where

$$\tilde{\mathbf{C}}' = \begin{pmatrix} \mathbf{c}_{11}^T & & & \\ \vdots & & & \\ \mathbf{c}_{1,J-1}^T & & & \\ & \mathbf{c}_{21}^T & & \\ & \vdots & & \\ & \mathbf{c}_{2,J-1}^T & & \\ & & \ddots & \\ & & & \mathbf{c}_{m1}^T \\ & & & \vdots \\ & & & \mathbf{c}_{m,J-1}^T \end{pmatrix}$$

That is, $\tilde{\mathbf{C}}$ has full row rank and thus $\mathbf{U}$ is positive definite. □

Proof of Theorem S.4:

LEMMA S.6. $|\mathbf{U}| = (\prod_{i=1}^m n_i)^{J-1} |\mathbf{V}|$, *where*

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \cdots & \mathbf{V}_{1,J-1} \\ \vdots & \ddots & \vdots \\ \mathbf{V}_{J-1,1} & \cdots & \mathbf{V}_{J-1,J-1} \end{pmatrix}$$

$$= \begin{pmatrix} u_{111} & & & \cdots & u_{1,J-1,1} & & \\ & \ddots & & & & \ddots & \\ & & u_{11m} & & & \cdots & u_{1,J-1,m} \\ \vdots & & & \ddots & & & \vdots \\ u_{J-1,1,1} & & & \cdots & u_{J-1,J-1,1} & & \\ & \ddots & & & & \ddots & \\ & & u_{J-1,1,m} & & & \cdots & u_{J-1,J-1,m} \end{pmatrix}$$

Kovacs et al. (1999) generalized Schur's Formula (Gantmacher (1960)) as follows:

Lemma S.7. *(Kovacs et al., 1999, Theorem 1)*
Assume that $\mathbf{M}$ *is a* $k \times k$ *block matrix with each block element* $\mathbf{A}_{ij}$ *as an* $n \times n$ *matrix.*

$$\mathbf{M} = \begin{pmatrix} \mathbf{A}_{11} & \cdots & \mathbf{A}_{1k} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{k1} & \cdots & \mathbf{A}_{kk} \end{pmatrix}$$

If all of $\mathbf{A}_{ij}$*'s commute pairwise, that is,* $\mathbf{A}_{ij}\mathbf{A}_{lm} = \mathbf{A}_{lm}\mathbf{A}_{ij}$ *for all possible pairs of indices* i, j *and* l, m*. Then*

$$|\mathbf{M}| = \left| \sum_{\pi \in S_k} (sgn\pi)\mathbf{A}_{1\pi(1)}\mathbf{A}_{2\pi(2)} \cdots \mathbf{A}_{k\pi(k)} \right| \tag{S.17}$$

Here the sum is computed over all permutations π *of* $\{1, 2, ..., k\}$*.*

In our case, all of $\mathbf{V}_{ij}$'s are diagonal matrices, so they commute pairwise. Moreover, the sum of product matrices in Equation (S.17) is a diagonal matrix, in which each element is the sum of products of the corresponding elements in those matrices. If we apply the above lemma, we get

$$\begin{aligned} |\mathbf{V}| &= \left| \sum_{\pi \in S_{J-1}} (sgn\pi)\mathbf{V}_{1\pi(1)}\mathbf{V}_{2\pi(2)} \cdots \mathbf{V}_{J-1,\pi(J-1)} \right| \\ &= \prod_{i=1}^{m} \left| \sum_{\pi \in S_{J-1}} (sgn\pi) u_{1\pi(1)i} u_{2\pi(2)i} \cdots u_{J-1,\pi(J-1),i} \right| \end{aligned}$$

Then the following result is obtained:

Lemma S.8. $|\mathbf{V}| = \prod_{i=1}^{m} |\mathbf{V}_i|$, *where*

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

Note that $\mathbf{V}_i$ defined above is very similar to $\mathbf{U}_i$ define in equation (5).

Lemma S.9. $|\mathbf{V}_i| = \left(\prod_{j=1}^{J} \pi_{ij}\right)^{-1} \cdot |\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L}|^{-2}$.

Proof of Lemma S.9: It can be verified that $\mathbf{c}_{iJ} = \boldsymbol{\pi}_i$.
Since $\mathbf{c}_{ij}^T \text{diag}(\boldsymbol{\pi}_i)^{-1}\mathbf{c}_{iJ} = \mathbf{c}_{ij}^T\mathbf{1} = 0$ for $j = 1, \ldots, J-1$ and 1 for $j = J$, then

$$\left[\left(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L}\right)^{-1}\right]^T \text{diag}(\boldsymbol{\pi}_i)^{-1} \left[\left(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L}\right)^{-1}\right] = \begin{bmatrix} \mathbf{V}_i & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix}$$

□

Combining Lemmas S.6, S.8, and S.9, we obtain Theorem S.4.

Remark S.1. Actually, we provide an explicit formula for $|\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L}|$ in (S.1), which can further clarify Lemma S.9 as (1) $|\mathbf{V}_i| = \prod_{j=1}^{J}\pi_{ij}$ for baseline-category, adjacent-categories, and continuation-ratio logit models; (2) $|\mathbf{V}_i| = \pi_{iJ}^{-1}\prod_{j=1}^{J-1}\pi_{ij}^{-1}\gamma_{ij}^2(1-\gamma_{ij})^2$ for cumulative logit models.

Proof of Theorem S.5:

The simplest case is the *npo* model whose conclusion is straightforward.

The *ppo* model is the most general case. In this case, we consider a sequence of linear subspaces

$$\{0\} \subset \mathcal{M}(\mathbf{H}_c^T) \cap (\cap_{j=1}^{J-1}\mathcal{M}(\mathbf{H}_j^T)) \subset M(\mathbf{H}_c^T)$$

with corresponding dimensions $0 \le r_c - r_0 \le r_c \triangleq rank(\mathbf{H}_c)$, where $r_0 = rank(\mathbf{H}_c) - dim[\mathcal{M}(\mathbf{H}_c^T)\cap(\cap_{j=1}^{J-1}\mathcal{M}(\mathbf{H}_j^T))]$. Then there exist $\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_{r_c-r_0}, \boldsymbol{\alpha}_{r_c-r_0+1}, \cdots, \boldsymbol{\alpha}_{r_c} \in \mathbb{R}^m$ s.t. $\{\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_{r_c-r_0}\}$ forms a basis of $\mathcal{M}(\mathbf{H}_c^T)\cap(\cap_{j=1}^{J-1}\mathcal{M}(\mathbf{H}_j^T))$ and $\{\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_{r_c}\}$ forms a basis of $\mathcal{M}(\mathbf{H}_c^T)$. By simple operations $\mathbf{H}_c$ can be transformed into $\mathbf{H}_c^* = (\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_{r_c}, \mathbf{0}, \cdots, \mathbf{0})^T$ and $\mathbf{H}_j$ can be transformed into

$$\mathbf{H}_j^* = (\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_{r_c-r_0}, \boldsymbol{\alpha}_{r_c-r_0+1}^{(j)}, \cdots, \boldsymbol{\alpha}_{r_j}^{(j)}, \mathbf{0}, \cdots, \mathbf{0})^T$$

where $r_j = rank(\mathbf{H}_j)$, $j = 1, 2, \cdots, J-1$. Then $rank(\mathbf{H}_{ppo}) = rank(\mathbf{H}_{ppo}^*)$ with

$$\mathbf{H}_{ppo}^* = \begin{pmatrix} \mathbf{H}_1^* & & \\ & \ddots & \\ & & \mathbf{H}_{J-1}^* \\ \mathbf{H}_c^* & \cdots & \mathbf{H}_c^* \end{pmatrix}_{p\times m(J-1)}$$

Since the first $r_c - r_0$ rows of $(\mathbf{H}_c^*, \cdots, \mathbf{H}_c^*)$ can be eliminated by applying row operations of $\mathbf{H}_j^*$ onto it separately, then $rank(\mathbf{H}_{ppo}^*) = rank(\mathbf{H}_{ppo}^{**})$ where

$$\mathbf{H}_{ppo}^{**} = \begin{pmatrix} \mathbf{H}_1^* & & \\ & \ddots & \\ & & \mathbf{H}_{J-1}^* \\ \mathbf{H}_c^{**} & \cdots & \mathbf{H}_c^{**} \end{pmatrix}_{p\times m(J-1)}$$

and $\mathbf{H}_c^{**} = (\mathbf{0}, \cdots, \mathbf{0}, \boldsymbol{\alpha}_{r_c-r_0+1}, \cdots, \boldsymbol{\alpha}_{r_c}, \mathbf{0}, \cdots, \mathbf{0})^T$. Therefore, $rank(\mathbf{H}_{ppo}) = rank(\mathbf{H}_{ppo}^{**}) \leqslant r_1 + \cdots + r_{J-1} + r_0$.

We claim that the nonzero rows of $\mathbf{H}_{ppo}^{**}$ are linearly independent which will lead to the final conclusion. Actually, let's denote those nonzero rows of $\mathbf{H}_{ppo}^{**}$ as $\boldsymbol{\Lambda}_i^{(j)}, i = 1, 2, \cdots, r_j, j = 1, 2, \cdots, J-1$ and $\boldsymbol{\Lambda}_{r_c-r_0+1}, \cdots, \boldsymbol{\Lambda}_{r_c}$, where $\boldsymbol{\Lambda}_i^{(j)}$ is the ith row of $(\mathbf{0}, \cdots, \mathbf{0}, \mathbf{H}_j^*, \mathbf{0}, \cdots, \mathbf{0})$, and $\boldsymbol{\Lambda}_i$ is the ith row of $(\mathbf{H}_c^{**}, \cdots, \mathbf{H}_c^{**})$. Suppose there exist $a_i^{(j)} \in \mathbb{R}, i = 1, 2, \cdots, r_j, j = 1, 2, \cdots, J-1$ and $a_i \in \mathbb{R}, i = r_c - r_0 + 1, \cdots, r_c$ s.t.

$$\mathbf{0} = \sum_{j=1}^{J-1}\sum_{i=1}^{r_j} a_i^{(j)}\boldsymbol{\Lambda}_i^{(j)} + \sum_{i=r_c-r_0+1}^{r_c} a_i\boldsymbol{\Lambda}_i$$

then for $j = 1, \ldots, J-1$,

$$\mathbf{0} = \sum_{i=1}^{r_c - r_0} a_i^{(j)} \boldsymbol{\alpha}_i + \sum_{i=r_c-r_0+1}^{r_j} a_i^{(j)} \boldsymbol{\alpha}_i^{(j)} + \sum_{i=r_c-r_0+1}^{r_c} a_i \boldsymbol{\alpha}_i$$

which implies for $j = 1, \ldots, J-1$,

$$\sum_{i=r_c-r_0+1}^{r_c} a_i \boldsymbol{\alpha}_i = -\sum_{i=1}^{r_c-r_0} a_i^{(j)} \boldsymbol{\alpha}_i - \sum_{i=r_c-r_0+1}^{r_j} a_i^{(j)} \boldsymbol{\alpha}_i^{(j)} \in \mathcal{M}(\mathbf{H}_c^T) \cap \mathcal{M}(\mathbf{H}_j^T)$$

Thus, $\sum_{i=r_c-r_0+1}^{r_c} a_i \boldsymbol{\alpha}_i \in \mathcal{M}(\mathbf{H}_c^T) \bigcap \left(\cap_{j=1}^{J-1} \mathcal{M}(\mathbf{H}_j^T)\right)$. Then we must have $\sum_{i=r_c-r_0+1}^{r_c} a_i \boldsymbol{\alpha}_i = \mathbf{0}$ since $\{\boldsymbol{\alpha}_{r_c-r_0+1}, \ldots, \boldsymbol{\alpha}_{r_c}\}$ and $\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{r_c-r_0}\}$ are linearly independent. Therefore, $a_i = 0$ for $i = r_c - r_0 + 1, \ldots, r_c$ and thus

$$\mathbf{0} = \sum_{i=1}^{r_c-r_0} a_i^{(j)} \boldsymbol{\alpha}_i + \sum_{i=r_c-r_0+1}^{r_j} a_i^{(j)} \boldsymbol{\alpha}_i^{(j)}$$

It implies $a_i^{(j)} = 0$, $i = 1, \ldots, r_c - r_0, r_c - r_0 + 1, \ldots, r_j$ since $\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{r_c-r_0}, \boldsymbol{\alpha}_{r_c-r_0+1}^{(j)}, \ldots, \boldsymbol{\alpha}_{r_j}^{(j)}\}$ are linear independent.

Therefore, the conclusion on *ppo* models is justified.

Since *po* models are special cases of *ppo* models, the corresponding result is a direct conclusion. □

Proof of Theorem A.1:

Recall that $dim(\mathcal{M}(\mathbf{H}_i^T)) = rank(\mathbf{H}_i^T) = r_i$ and $dim(\mathcal{M}(\mathbf{H}_{i_1}^T) + \cdots + \mathcal{M}(\mathbf{H}_{i_k}^T)) = dim(\mathcal{M}((\mathbf{H}_{i_1}^T, \cdots, \mathbf{H}_{i_k}^T))) = rank((\mathbf{H}_{i_1}^T, \cdots, \mathbf{H}_{i_k}^T)) = r_{i_1,\ldots,i_k}$, for $i_1 < \cdots < i_k$ and $k = 2, \ldots, n$, where "+" stands for the sum of two linear subspaces.

First of all, $dim(\mathcal{M}(\mathbf{H}_1^T) \cap \mathcal{M}(\mathbf{H}_2^T)) = dim(\mathcal{M}(\mathbf{H}_1^T)) + dim(\mathcal{M}(\mathbf{H}_2^T)) - dim(\mathcal{M}(\mathbf{H}_1^T) + \mathcal{M}(\mathbf{H}_2^T)) = r_1 + r_2 - r_{12}$. That is, (11) is true for $n = 2$.

Suppose (11) is true for $n = k$. Then for $n = k+1$,

$$\begin{aligned} & dim(\cap_{i=1}^{k+1} \mathcal{M}(\mathbf{H}_i^T)) = dim(\cap_{i=1}^{k} \mathcal{M}(\mathbf{H}_i^T) \cap \mathcal{M}(\mathbf{H}_{k+1}^T)) \\ = \;& dim(\cap_{i=1}^{k} \mathcal{M}(\mathbf{H}_i^T)) + dim(\mathcal{M}(\mathbf{H}_{k+1}^T)) - dim(\cap_{i=1}^{k} \mathcal{M}(\mathbf{H}_i^T) + \mathcal{M}(\mathbf{H}_{k+1}^T)) \\ = \;& \sum_{i=1}^{k} r_i - \sum_{1 \leqslant i_1 < i_2 \leqslant k} r_{i_1 i_2} + \cdots + (-1)^{k-1} r_{12\cdots k} + r_{k+1} - \triangle \end{aligned}$$

where

$$\begin{aligned} & \triangle = dim(\cap_{i=1}^{k} \mathcal{M}(\mathbf{H}_i^T) + \mathcal{M}(\mathbf{H}_{k+1}^T)) = dim(\cap_{i=1}^{k} \mathcal{M}((\mathbf{H}_i^T, \mathbf{H}_{k+1}^T))) \\ = \;& \sum_{i=1}^{k} rank((\mathbf{H}_i^T, \mathbf{H}_{k+1}^T)) - \sum_{1 \leqslant i_1 < i_2 \leqslant k} rank((\mathbf{H}_{i_1}^T, \mathbf{H}_{k+1}^T, \mathbf{H}_{i_2}^T, \mathbf{H}_{k+1}^T)) \\ & + \cdots + (-1)^{k-1} rank((\mathbf{H}_1^T, \mathbf{H}_{k+1}^T, \cdots, \mathbf{H}_k^T, \mathbf{H}_{k+1}^T)) \\ = \;& \sum_{i=1}^{k} r_{i,k+1} - \sum_{1 \leqslant i_1 < i_2 \leqslant k} r_{i_1,i_2,k+1} + \cdots + (-1)^{k-1} r_{1,2,\ldots,k+1} \end{aligned}$$

Therefore,

$$\begin{aligned}
& dim(\cap_{i=1}^{k+1}\mathcal{M}(\mathbf{H}_i^T)) \\
= & \sum_{i=1}^{k} r_i - \sum_{1\leqslant i_1<i_2\leqslant k} r_{i_1 i_2} + \cdots + (-1)^{k-1} r_{12\cdots k} + r_{k+1} \\
& - \sum_{i=1}^{k} r_{i,k+1} + \sum_{1\leqslant i_1<i_2\leqslant k} r_{i_1,i_2,k+1} + \cdots + (-1)^k r_{1,2,\ldots,k+1} \\
= & \sum_{i=1}^{k+1} r_i - \sum_{1\leqslant i_1<i_2\leqslant k+1} r_{i_1 i_2} + \cdots + (-1)^{(k+1)-1} r_{1,2,\ldots,k+1}
\end{aligned}$$

That is, (11) is true for $n = k+1$ as well. By mathematical induction, (11) is true for general n. □

Proof of Corollary S.2:

Suppose $p_H > 0$. Then there exist $m \times 1$ vectors $\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_{p_H}$, which form a basis of $\cap_{j=1}^{J-1}\mathcal{M}(\mathbf{H}_j^T)$. Write $\mathbf{H}_c = (\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_{p_c})^T$. According to Theorem S.5, if $|\mathbf{F}| > 0$, then $r_0 = rank(\mathbf{H}_c) = p_c$, or equivalently, $\mathcal{M}(\mathbf{H}_c^T) \cap \left(\cap_{j=1}^{J-1}\mathcal{M}(\mathbf{H}_j^T)\right) = \{\mathbf{0}\}$. Then $\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_{p_H}, \boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_{p_c}$ are linearly independent. Thus $m \geq p_c + p_H$. □

Proof of Theorem 4.1:

Actually, according to Theorem 3.1, $\mathbf{F} = \mathbf{HUH}^T$. From the proof of Theorem S.3, $\mathbf{U} = \tilde{\mathbf{C}}\tilde{\mathbf{W}}\tilde{\mathbf{C}}^T$, where $\tilde{\mathbf{W}}$ is a diagonal matrix. Therefore, $\mathbf{F} = \mathbf{H}\tilde{\mathbf{C}}\tilde{\mathbf{W}}\tilde{\mathbf{C}}^T\mathbf{H}^T$. Let $\mathbf{W} = \tilde{\mathbf{W}}/n$ and $\mathbf{G} = \tilde{\mathbf{C}}^T\mathbf{H}^T$. Then $\mathbf{F} = n\mathbf{G}^T\mathbf{W}\mathbf{G}$, which leads to the final result. □

Proof of Lemma S.3: Actually, $\max_{1\leq i\leq m}\alpha_i \leq J$. Suppose $\max_{1\leq i\leq m}\alpha_i \geq J$, which means $\max_{1\leq i\leq m}\alpha_i = J$. Without any loss of generality, we assume $\alpha_1 = J$. Then $i_j = j$ for $j = 1, \ldots, J$.

According to the proof of Lemma S.5, we have $\mathbf{1}^T\mathbf{c}_{ij} = 0$ for $i = 1, \ldots, m$ and $j = 1, \ldots, J-1$. Then $\mathbf{1}^T(\mathbf{c}_{11} + \cdots + \mathbf{c}_{1,J-1}) = 0$ and thus $\mathbf{1}^T\mathbf{G}[i_1, \ldots, i_J] = 0$. That is, $\text{rank}(\mathbf{G}[i_1, \ldots, i_J]) \leq J-1$. Therefore, $\text{rank}(\mathbf{G}[i_1, \ldots, i_p]) \leq p-1$ and $|\mathbf{G}[i_1, \ldots, i_p]| = 0$. □

Proof of Theorem S.6: Suppose $c_{\alpha_1,\ldots,\alpha_m} \neq 0$ for some $(\alpha_1, \ldots, \alpha_m)$. Therefore, there exist $(i_1, \ldots, i_p) \in (\alpha_1, \ldots, \alpha_m)$ such that $\mathbf{G}[i_1, \ldots, i_p]$ is of full rank p. Without any loss of generality, we assume $\alpha_1 \geq \cdots \geq \alpha_k > 0 = \alpha_{k+1} = \cdots = \alpha_m$, that is, $\{i \mid \alpha_i > 0\} = \{1, \ldots k\}$. Consider the submatrix $\tilde{\mathbf{G}} := \mathbf{G}[1, \ldots, kJ]$ which is $kJ \times p$ and contains $\mathbf{G}[i_1, \ldots, i_p]$ as a submatrix. Then $\tilde{\mathbf{G}}$ is of rank p or $\tilde{\mathbf{G}}^T$ is of full row rank p. Write $\tilde{\mathbf{W}} = k^{-1}\text{diag}\{\text{diag}(\boldsymbol{\pi}_1)^{-1}, \ldots, \text{diag}(\boldsymbol{\pi}_k)^{-1}\}$. Then the restricted matrix $\mathbf{F} := n\ \tilde{\mathbf{G}}^T\tilde{\mathbf{W}}\tilde{\mathbf{G}}$ is positive definite. On the other hand, $\mathbf{F}$ is the Fisher information matrix $n\mathbf{G}^T\mathbf{W}\mathbf{G}$ as defined in Theorem 4.1 with $w_1 = \cdots = w_k = 1/k$ and $w_{k+1} = \cdots = w_m = 0$. According to Theorem 4.1 and Theorem 2.1, $\mathbf{F} = nk^{-1}\sum_{i=1}^k \mathbf{F}_i$. Therefore, $\mathbf{F}_{res} := \sum_{i=1}^k \mathbf{F}_i$ is positive definite. □

Proof of Theorem 5.1:

Case 1: Baseline-category logit model

The baseline-category logit model for nominal response (Agresti, 2013; Zocchi and Atkinson, 1999) can be extended in general as follows

$$\log\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) = \mathbf{h}_j^T(\mathbf{x}_i)\boldsymbol{\beta}_j + \mathbf{h}_c^T(\mathbf{x}_i)\boldsymbol{\zeta}\ , \quad j = 1,\ldots,J-1 \tag{S.18}$$

Lemma S.10. *Fixing $\mathbf{x}_i$, $\boldsymbol{\beta}_j, j = 1,\cdots,J-1$ and $\boldsymbol{\zeta}$ in Model (S.18), let $a_j = \mathbf{h}_j^T(\mathbf{x}_i)\boldsymbol{\beta}_j + \mathbf{h}_c^T(\mathbf{x}_i)\boldsymbol{\zeta}$, $j = 1,\ldots,J-1$. Then $0 < \pi_{ij} < 1, j = 1,\ldots,J$ exist uniquely if and only if $-\infty < a_j < \infty$, $j = 1,\ldots,J-1$. In this case,*

$$\pi_{ij} = \begin{cases} \frac{e^{a_j}}{e^{a_1}+\cdots+e^{a_{J-1}}+1} & 1 \le j \le J-1 \\ \frac{1}{e^{a_1}+\cdots+e^{a_{J-1}}+1} & j = J \end{cases} \tag{S.19}$$

Proof of Lemma S.10: Write $y_j = \log \pi_{ij}$, $j = 1,\ldots,J$. Then $0 < \pi_{ij} < 1, j = 1,\ldots,J$ if and only if $y_j \in (-\infty, 0)$, $j = 1,\ldots,J$. In this case, Model (S.18) implies $a_j = y_j - y_J \in (-\infty,\infty)$, $j = 1,\ldots,J-1$.
On the other hand, for any given $a_1,\ldots,a_{J-1} \in (-\infty,\infty)$, $y_j = a_j + y_J$, $j = 1,\ldots,J-1$. Note that

$$\begin{aligned} 1 &= \pi_{i1} + \pi_{i2} + \cdots + \pi_{i,J-1} + \pi_{iJ} \\ &= e^{y_1} + e^{y_2} + \cdots + e^{y_{J-1}} + e^{y_J} \\ &= e^{a_1+y_J} + e^{a_2+y_J} + \cdots + e^{a_{J-1}+y_J} + e^{y_J} \\ &= e^{y_J}\left(e^{a_1} + e^{a_2} + \cdots + e^{a_{J-1}} + 1\right) \end{aligned}$$

Since $\pi_{ij} = e^{y_j}$, we get solutions of π_{ij} given in (S.19), and thus $\pi_{ij} \in (0,1)$ exists and is unique, $j = 1,\ldots,J$. #

Case 2: Cumulative logit model

The cumulative logit model for ordinal responses (McCullagh, 1980; Christensen, 2015) can be described in general as follows:

$$\log\left(\frac{\pi_{i1}+\cdots+\pi_{ij}}{\pi_{i,j+1}+\cdots+\pi_{iJ}}\right) = \mathbf{h}_j^T(\mathbf{x}_i)\boldsymbol{\beta}_j + \mathbf{h}_c^T(\mathbf{x}_i)\boldsymbol{\zeta}\ , \quad j = 1,\ldots,J-1 \tag{S.20}$$

Lemma S.11. *Fixing $\mathbf{x}_i$, $\boldsymbol{\beta}_j, j = 1,\cdots,J-1$ and $\boldsymbol{\zeta}$ in Model (S.20), let $a_j = \mathbf{h}_j^T(\mathbf{x}_i)\boldsymbol{\beta}_j + \mathbf{h}_c^T(\mathbf{x}_i)\boldsymbol{\zeta}$, $j = 1,\ldots,J-1$. Then $0 < \pi_{ij} < 1, j = 1,\ldots,J$ exist and are unique if and only if $-\infty < a_1 < a_2 < \cdots < a_{J-1} < \infty$. In this case,*

$$\pi_{ij} = \begin{cases} \frac{\exp(a_1)}{1+\exp(a_1)} & j = 1 \\ \frac{\exp(a_j)}{1+\exp(a_j)} - \frac{\exp(a_{j-1})}{1+\exp(a_{j-1})} & 1 < j < J \\ \frac{1}{1+\exp(a_{J-1})} & j = J \end{cases} \tag{S.21}$$

Proof of Lemma S.11: Taking $j = 1$ in Model (S.20), then $\log(\pi_{i1}/(1-\pi_{i1})) = a_1$ and $\pi_{i1} = \exp(a_1)/[1+\exp(a_1)]$. Then $0 < \pi_{i1} < 1$ if and only if $-\infty < a_1 < \infty$. For $j = 2,\cdots,J-1$,

$$\pi_{ij} = \frac{\exp(a_j)}{1+\exp(a_j)} - \frac{\exp(a_{j-1})}{1+\exp(a_{j-1})}$$

which implies that $\pi_{ij} > 0$ if and only if $a_j > a_{j-1}$. Therefore, $\pi_{iJ} = 1 - (\pi_{i1} + \cdots + \pi_{i,J-1}) = 1 - \exp(a_{J-1})/[1 + \exp(a_{J-1})] = 1/[1 + \exp(a_{J-1})]$, which indicates $0 < \pi_{iJ} < 1$ if and only if $-\infty < a_{J-1} < \infty$. Given $\pi_{i1} + \cdots + \pi_{iJ} = 1$, we have

$$-\infty < a_1 < a_2 < \cdots < a_{J-1} < \infty \Leftrightarrow \pi_{ij} \in (0,1), \quad j = 1, \ldots, J$$

#

COROLLARY S.8. *For the cumulative logit model with proportional odds*

$$\log\left(\frac{\pi_{i1} + \cdots + \pi_{ij}}{\pi_{i,j+1} + \cdots + \pi_{iJ}}\right) = \beta_j + \mathbf{h}_c^T(\mathbf{x}_i)\boldsymbol{\zeta}\ , \quad j = 1, \ldots, J-1 \tag{S.22}$$

The design space has no restriction since $-\infty < \beta_1 < \beta_2 < \cdots < \beta_{J-1} < \infty$ is part of the model assumptions, which implies $\pi_{ij} \in (0,1)$, $j = 1, \ldots, J$.

Case 3: Adjacent-categories logit model

The adjacent-categories logit model for ordinal responses (Liu and Agresti, 2005; Agresti, 2013) can be extended as follows:

$$\log\left(\frac{\pi_{ij}}{\pi_{i,j+1}}\right) = \mathbf{h}_j^T(\mathbf{x}_i)\boldsymbol{\beta}_j + \mathbf{h}_c^T(\mathbf{x}_i)\boldsymbol{\zeta}\ , \quad j = 1, \ldots, J-1 \tag{S.23}$$

LEMMA S.12. *Fixing $\mathbf{x}_i$, $\boldsymbol{\beta}_j, j = 1, \cdots, J-1$ and $\boldsymbol{\zeta}$ in Model (S.23), let $a_j = \mathbf{h}_j^T(\mathbf{x}_i)\boldsymbol{\beta}_j + \mathbf{h}_c^T(\mathbf{x}_i)\boldsymbol{\zeta}$, $j = 1, \ldots, J-1$. Then $0 < \pi_{ij} < 1, j = 1, \ldots, J$ exist uniquely if and only if $-\infty < a_j < \infty$, $j = 1, \ldots, J-1$. In this case,*

$$\pi_{ij} = \begin{cases} \frac{\exp(a_{J-1}+\cdots+a_j)}{\exp(a_{J-1}+\cdots+a_1)+\exp(a_{J-1}+\cdots+a_2)+\cdots+\exp(a_{J-1})+1} & j = 1, \ldots, J-1 \\ \frac{1}{\exp(a_{J-1}+\cdots+a_1)+\exp(a_{J-1}+\cdots+a_2)+\cdots+\exp(a_{J-1})+1} & j = J \end{cases} \tag{S.24}$$

Proof of Lemma S.12: Let $y_j = \log \pi_{ij}$. Then $0 < \pi_{ij} < 1, j = 1, \ldots, J$ if and only if $y_j \in (-\infty, 0)$. In this case, Model (S.23) implies $a_j = y_j - y_{j+1} \in (-\infty, \infty)$, $j = 1, \ldots, J-1$. On the other hand, for any given $a_1, \ldots, a_{J-1} \in (-\infty, \infty)$, $y_j = (a_{J-1} + \cdots + a_j) + y_J$, $j = 1, \ldots, J-1$. Note that

$$\begin{aligned} 1 &= \pi_{i1} + \pi_{i2} + \cdots + \pi_{i,J-1} + \pi_{iJ} \\ &= e^{y_1} + e^{y_2} + \cdots + e^{y_{J-1}} + e^{y_J} \\ &= e^{y_J}\left(e^{a_{J-1}+\cdots+a_1} + e^{a_{J-1}+\cdots+a_2} + \cdots + e^{a_{J-1}} + 1\right) \end{aligned}$$

Since $\pi_{ij} = e^{y_j}$, we get solutions of π_{ij} given in (S.24), and thus $\pi_{ij} \in (0,1)$ exists and is unique, $j = 1, \ldots, J$. #

Case 4: Continuation-ratio logit model

The continuation-ratio logit model for hierarchical responses (Agresti, 2013; Zocchi and Atkinson, 1999) can be rewritten in general as follows:

$$\log\left(\frac{\pi_{ij}}{\pi_{i,j+1} + \cdots + \pi_{iJ}}\right) = \mathbf{h}_j^T(\mathbf{x}_i)\boldsymbol{\beta}_j + \mathbf{h}_c^T(\mathbf{x}_i)\boldsymbol{\zeta}\ , \quad j = 1, \ldots, J-1 \tag{S.25}$$

LEMMA S.13. *Fixing* $\mathbf{x}_i$, $\boldsymbol{\beta}_j, j = 1, \cdots, J-1$ *and* $\boldsymbol{\zeta}$ *in Model (S.25), let* $a_j = \mathbf{h}_j^T(\mathbf{x}_i)\boldsymbol{\beta}_j + \mathbf{h}_c^T(\mathbf{x}_i)\boldsymbol{\zeta}$, $j = 1, \ldots, J-1$. *Then* $0 < \pi_{ij} < 1, j = 1, \ldots, J$ *exist uniquely if and only if* $-\infty < a_j < \infty$, $j = 1, \ldots, J-1$. *In this case,*

$$\pi_{ij} = \begin{cases} e^{a_j} \prod_{s=1}^{j} (e^{a_s} + 1)^{-1} & j = 1, \ldots, J-1 \\ \prod_{s=1}^{J-1} (e^{a_s} + 1)^{-1} & j = J \end{cases} \tag{S.26}$$

Proof of Lemma S.13: Let $y_j = \log \pi_{ij}$. Then $0 < \pi_{ij} < 1, j = 1, \ldots, J$ if and only if $y_j \in (-\infty, 0)$. In this case, Model (S.25) implies $a_j = y_j - \log(e^{y_{j+1}} + \cdots e^{y_J}) \in (-\infty, \infty)$, $j = 1, \ldots, J-1$.
On the other hand, for any given $a_1, \ldots, a_{J-1} \in (-\infty, \infty)$, it can be verified by induction that

$$\begin{aligned} e^{y_{J-1}} &= e^{y_J} e^{a_{J-1}} \\ e^{y_{J-2}} &= e^{y_J} e^{a_{J-2}} (e^{a_{J-1}} + 1) \\ e^{y_j} &= e^{y_J} e^{a_j} (e^{a_{j+1}} + 1) \cdots (e^{a_{J-1}} + 1), \; j = J-3, J-4, \cdots, 1 \end{aligned}$$

Therefore, it can be verified that

$$\begin{aligned} 1 &= \pi_{i1} + \pi_{i2} + \cdots + \pi_{i,J-1} + \pi_{iJ} \\ &= e^{y_1} + e^{y_2} + \cdots + e^{y_{J-1}} + e^{y_J} \\ &= e^{y_J} (e^{a_1} + 1)(e^{a_2} + 1) \cdots (e^{a_{J-1}} + 1) \end{aligned}$$

Since $\pi_{ij} = e^{y_j}$, we get solutions of π_{ij} given in (S.26), and thus $\pi_{ij} \in (0, 1)$ exists and is unique, $j = 1, \ldots, J$. #

Theorem 5.1 is obtained as a summary of Lemmas S.10, S.11, S.12, and S.13. □

Proof of Corollary 5.1: We only need to verity the "only if" part. According to Theorem 3.2, if $f(\mathbf{w}) > 0$ for some $\mathbf{w} = (w_1, \ldots, w_m)^T = (n_1, \ldots, n_m)^T/n$, then the corresponding $\mathbf{H}^*$ is of full row rank. Note that $\mathbf{H}^*$ can be obtained from $\mathbf{H}$ after removing the columns of $\mathbf{H}$ corresponding to $n_i = 0$. Thus $\mathbf{H}$ is of full row rank too, which corresponds to the uniform allocation. That is, $f(\mathbf{w}_u) > 0$.
In this case, any $\mathbf{w} = (w_1, \ldots, w_m)^T$ such that $0 < w_i < 1, i = 1, \ldots, m$ leads to $f(\mathbf{w}) > 0$ since it corresponds to the same $\mathbf{H}$ matrix. □

Proof of Theorem S.10: According to Theorem 4.2,

$$f_{ij}(z) = \sum_{\alpha_i \geq 0, \alpha_j \geq 0, \alpha_i + \alpha_j \leq p} \text{coefficient} \cdot z^{\alpha_i} (n_i + n_j - z)^{\alpha_j}$$

is a polynomial with nonnegative coefficients, whose order depends on the largest possible $\alpha_i + \alpha_j$. Lemma S.3 implies that $\max\{\alpha_i, \alpha_j\} \leq J - 1$ for positive coefficients and Corollary S.4 further implies that $\alpha_i + \alpha_j \leq p - (k_{\min} - 2) = p - k_{\min} + 2$ for positive coefficients. Therefore, $f_{ij}(z)$ is at most an order-q polynomial of z. □

Proof of Theorem S.11: In this case, the model is essentially a generalized linear model for binomial response with logit link. Theorem 4.2 says that the objective function $f(\mathbf{w}) = |\mathbf{G}^T \mathbf{W} \mathbf{G}|$ is an order-p polynomial consisting of terms $c_{\alpha_1, \ldots, \alpha_m} w_1^{\alpha_1} \cdots w_m^{\alpha_m}$. According to Lemma S.3, $c_{\alpha_1,\ldots,\alpha_m} \neq 0$ only if $\alpha_i \in \{0,1\}, i = 1,\ldots,m$. Therefore, in order to keep $f(\mathbf{w}) > 0$, we must have $m \geq p$. In other words, a minimally supported design may contain exactly $m = p$ distinct design points or experimental settings. In this case, the objective function $f(\mathbf{w}) \propto w_1 \cdots w_m$ and the D-optimal allocation is $\mathbf{w} = (1/m,\ldots,1/m)^T$. □

Proof of Corollary S.7: According to Theorem 3.1, $\mathbf{F} = \mathbf{HUH}^T$. In this case, there exist $m = p_1$ experimental settings such that $\text{rank}(\mathbf{H}) = p_1(J-1) = p$. On the other hand, the minimum number of experimental settings is at least $\max\{p_1,\ldots,p_{J-1}\} = p_1$ based on Corollary S.4. Therefore, the minimal number is $m = p_1$. In this case, $\mathbf{H}$ is a square matrix and

$$|\mathbf{F}| = |\mathbf{H}|^2 \cdot |\mathbf{U}| \propto \left(\prod_{i=1}^{m} w_i\right)^{J-1}$$

according to Theorem S.4. Thus, the uniform allocation $\mathbf{w}_u = (1/m,\ldots,\ 1/m)^T$ is D-optimal in this case. Note that $m = p_1 < p_1(J-1) = p$. □

Proof of Lemma S.4: We actually claim more detailed conclusions as follows:

(i) If $c_1 = c_2 = c_3$, then the solution is $w_1 = w_2 = w_3 = 1/3$.
(ii) If $c_1 = c_2 < c_3$, then $w_1 = w_2 > w_3 > 0$. Actually, $w_1 = w_2 = (-2c_1 + c_3 + \Delta_1)/D_1$ and $w_3 = c_3/D_1$, where $\Delta_1 = \sqrt{4c_1^2 - c_1c_3 + c_3^2}$ and $D_1 = -4c_1 + 3c_3 + 2\Delta_1$.
(iii) If $c_1 < c_2 = c_3$, then $w_1 > w_2 = w_3 > 0$. Actually, $w_1 = (-c_1 + 2c_3 + \Delta_2)/D_2$ and $w_2 = w_3 = 3c_3/D_2$, where $\Delta_2 = \sqrt{c_1^2 - c_1c_3 + 4c_3^2}$ and $D_2 = -c_1 + 8c_3 + \Delta_2$.
(iv) If $c_1 < c_2 < c_3$, then $w_1 > w_2 > w_3 > 0$. The procedure of obtaining analytic solutions of w_1, w_2, w_3 is as follows: (1) obtain y_1 from (S.33); (2) obtain y_2 from (S.31); (3) $w_1 = y_1/(y_1 + y_2 + 1)$, $w_2 = y_2/(y_1 + y_2 + 1)$, $w_3 = 1/(y_1 + y_2 + 1)$.

First of all, we only need to consider the cases of $0 < w_i < 1$, $i = 1, 2, 3$ (otherwise, $f(w_1, w_2, w_3) = 0$). It can also be verified that $0 < c_1 \leq c_2 \leq c_3$ implies that $w_1 \geq w_2 \geq w_3 > 0$ (otherwise, for example, if $w_1 < w_2$, one may replace w_1, w_2 both with $(w_1+w_2)/2$ and strictly increase f). The same argument implies that if $c_i = c_j$, then $w_i = w_j$ in the solution.

According to Theorem 5.10 in Yang et al. (2017), $(w_1, w_2, w_3)^T$ maximizes $f(w_1, w_2, w_3)$ if and only if

$$\frac{\partial f}{\partial w_1} = \frac{\partial f}{\partial w_2} = \frac{\partial f}{\partial w_3}$$

which is equivalent to $\partial f/\partial w_1 = \partial f/\partial w_3$ and $\partial f/\partial w_2 = \partial f/\partial w_3$ and thus equivalent to

$$c_3 w_1 w_2 (w_1 - 2w_3) + 2c_2 w_1 w_3 (w_1 - w_3) = c_1 w_2 w_3 (-2w_1 + w_3) \tag{S.27}$$

$$c_3 w_1 w_2 (w_2 - 2w_3) + 2c_1 w_2 w_3 (w_2 - w_3) = c_2 w_1 w_3 (-2w_2 + w_3) \tag{S.28}$$

Following Yang et al. (2016b, Section 5.2), we denote $y_1 = w_1/w_3 > 0$ and $y_2 = w_2/w_3 > 0$. Actually, $w_1 \geq w_2 \geq w_3 > 0$ implies $y_1 \geq y_2 \geq 1$. Since $w_1 + w_2 + w_3 = 1$, it implies $w_3 = 1/(y_1 + y_2 + 1)$, $w_1 = y_1/(y_1 + y_2 + 1)$, and $w_2 = y_2/(y_1 + y_2 + 1)$. Then (S.27) and (S.28) are equivalent to

$$c_3 y_1 y_2 (y_1 - 2) + 2c_2 y_1 (y_1 - 1) = c_1 y_2 (-2y_1 + 1) \tag{S.29}$$

$$c_3 y_1 y_2 (y_2 - 2) + 2c_1 y_2 (y_2 - 1) = c_2 y_1 (-2y_2 + 1) \tag{S.30}$$

From (S.29) we get $y_2[c_3y_1^2 - 2(c_3-c_1)y_1 - c_1] = 2c_2y_1(1-y_1)$. If $y_1 = 1$, then we must have $y_2 = 1$ and $c_3 - 2(c_3 - c_1) - c_1 = 0$, which implies $w_1 = w_2 = w_3 = 1/3$ and $c_1 = c_2 = c_3$. Actually, we can also verify that $c_1 = c_3$ implies $y_1 = 1$.

Now we assume $y_1 > 1$, which implies $c_1 < c_3$. Then

$$y_2 = \frac{2c_2(1-y_1)y_1}{c_3y_1^2 - 2(c_3-c_1)y_1 - c_1} \tag{S.31}$$

After plugging (S.31) into (S.30), we get

$$a_0 + a_1y_1 + a_2y_1^2 + a_3y_1^3 + y_1^4 = 0 \tag{S.32}$$

where $a_0 = c_1^2/c_3^2 > 0$, $a_1 = 4c_1(-2c_1+c_2+2c_3)/(3c_3^2) > 0$, $a_2 = 2(2c_1^2 - 2c_1c_2 - 7c_1c_3 - 2c_2c_3 + 2c_3^2)/(3c_3^2)$, and $a_3 = 4(2c_1 + c_2 - 2c_3)/(3c_3)$.

Denote $h(y_1) = a_0 + a_1y_1 + a_2y_1^2 + a_3y_1^3 + y_1^4$. Note that $h(\infty) = \infty$, $h(-c_1/c_3) = -c_1^2(c_1^2 + 8c_1c_2 - 2c_1c_3 + 8c_2c_3 + c_3^2)/(3c_3^4) < 0$, $h(0) = c_1^2/c_3^2 > 0$, $h(1) = -(c_1-c_3)^2/(3c_3^2) < 0$, and $h(\infty) = \infty$. Then $h(y_1) = 0$ yields four real roots in $(\infty, -c_1/c_3)$, $(-c_1/c_3, 0)$, $(0,1)$, and $(1,\infty)$, respectively. That is, there is one and only one $y_1 \in (1,\infty)$.

According to Tong et al. (2014, equation (12)),

$$y_1 = -\frac{a_3}{4} + \frac{\sqrt{A_1}}{2} + \frac{\sqrt{C_1}}{2} \ , \tag{S.33}$$

where

$$\begin{aligned}
A_1 &= -\frac{2a_2}{3} + \frac{a_3^2}{4} + \frac{G_1}{3\times 2^{1/3}} \ , \\
C_1 &= -\frac{4a_2}{3} + \frac{a_3^2}{2} - \frac{G_1}{3\times 2^{1/3}} + \frac{-8a_1 + 4a_2a_3 - a_3^3}{4\sqrt{A_1}} \ , \\
G_1 &= \left(F_1 - \sqrt{F_1^2 - 4E_1^3}\right)^{1/3} + \left(F_1 + \sqrt{F_1^2 - 4E_1^3}\right)^{1/3} \ , \\
E_1 &= 12a_0 + a_2^2 - 3a_1a_3 \ , \\
F_1 &= 27a_1^2 - 72a_0a_2 + 2a_2^3 - 9a_1a_2a_3 + 27a_0a_3^2 \ .
\end{aligned}$$

The calculation of G_1, A_1, C_1, and y_1 are operations among complex numbers, while y_1 at the end would be a real number.

The procedure of obtaining analytic solutions of w_1, w_2, w_3 would be, (1) obtain y_1 from (S.33); (2) obtain y_2 from (S.31); (3) $w_1 = y_1/(y_1+y_2+1)$, $w_2 = y_2/(y_1+y_2+1)$, $w_3 = 1/(y_1+y_2+1)$.

Now we discuss some special cases.

(i) If $c_1 = c_2 < c_3$, then $w_1 = w_2$ and thus $y_1 = y_2$. Both (S.29) and (S.30) yield $y_1 = c_3^{-1}(-2c_1 + c_3 + \sqrt{4c_1^2 - c_1c_3 + c_3^2})$, which implies

$$w_1 = w_2 = \frac{-2c_1 + c_3 + \Delta_1}{-4c_1 + 3c_3 + 2\Delta_1}, \quad w_3 = \frac{c_3}{-4c_1 + 3c_3 + 2\Delta_1}$$

where $\Delta_1 = \sqrt{4c_1^2 - c_1c_3 + c_3^2}$. Note that $w_1 > w_3$ since $\Delta_1 > 2c_1$.

(ii) If $c_1 < c_2 = c_3$, then $w_2 = w_3$ and thus $y_2 = 1$. From (S.29) we get $y_1 = 3c_3^{-1}(-c_1 + 2c_3 + \sqrt{c_1^2 - c_1 c_3 + 4c_3^2})$, which implies

$$w_1 = \frac{-c_1 + 2c_3 + \Delta_2}{-c_1 + 8c_3 + \Delta_2}, \quad w_2 = w_3 = \frac{3c_3}{-c_1 + 8c_3 + \Delta_2}$$

where $\Delta_2 = \sqrt{c_1^2 - c_1 c_3 + 4c_3^2}$. Note that $w_1 > w_2$ since $\Delta_2 > c_1 + c_3$.

(iii) If $c_1 < c_2 < c_3$, then y_1, y_2 and thus w_1, w_2, w_3 can be obtained analytically. We have proven $y_1 \geq y_2 \geq 1$. Using (S.29) and (S.30), it can be verified that $y_1 \neq y_2$ unless $c_1 = c_2$; and $y_2 \neq 1$ unless $c_2 = c_3$. That is, $y_1 > y_2 > 1$ and $w_1 > w_2 > w_3$.

□

Proof of Theorem A.2: (i) is straightforward. (ii) follows from the facts in the proof of Lemma S.5, $\mathbf{c}_{iJ} \equiv \boldsymbol{\pi}_i$; $\mathbf{1}^T \mathbf{c}_{ij} = 0, j = 1, \ldots, J-1$; and $\mathbf{1}^T \mathbf{c}_{iJ} = 1$. (iii) and (iv) can be verified using the formulae of $\mathbf{c}_{ij}$ in Section S.3. □

AbbVie, Inc.
Data and Statistical Sciences
Dept R436, Bldg AP9A-2
1 North Waukegan Road
North Chicago, IL 60064, USA
E-mail: sherry.bu@abbvie.com

Department of Mathematics, Statistics,
and Computer Science
University of Illinois at Chicago
322 Science and Engineering Offices (M/C 249)
851 S. Morgan Street, Chicago, IL 60607, USA
E-mail: dibyen@uic.edu
jyang06@math.uic.edu