D-optimal Designs for Multinomial Logistic Models

Jie Yang

University of Illinois at Chicago

Joint with Xianwei Bu and Dibyen Majumdar

October 12, 2017

1. Multinomial Logistic Models
   - Cumulative logit model: Odor removal study
   - Continuation-ratio logit model: Emergence of house flies
   - Four logit models with three types of odds
   - Relevant literature in optimal design theory
2. Fisher Information Matrix and D-optimal Designs
   - Fisher information matrix for multinomial logistic models
   - Determinant of Fisher information matrix
   - Locally D-optimal designs
3. Minimally Supported Designs
   - Positive definiteness of Fisher information matrix
   - Minimally supported designs for multinomial logistic models

Odor Removal Study (Yang, Tong, and Mandal, 2017)

A 2^2 factorial experiment with two factors:
X_1: *types of algae* $(-, +)$; X_2: *synthetic resins* $(-, +)$.

Three categories of the response Y:
serious odor $(Y = 1)$, medium odor $(Y = 2)$ and no odor $(Y = 3)$.

Group	X_1	X_2	Responses y_{i1}	y_{i2}	y_{i3}	# of replicates	Model
$i=1$	$+$	$+$	2	6	2	$n_1 = \sum y_{1j} = 10$	$\text{logit}(\gamma_{1j}) = \theta_j - \beta_1 - \beta_2$
$i=2$	$+$	$-$	7	2	1	$n_2 = \sum y_{2j} = 10$	$\text{logit}(\gamma_{2j}) = \theta_j - \beta_1 + \beta_2$
$i=3$	$-$	$+$	0	0	10	$n_3 = \sum y_{3j} = 10$	$\text{logit}(\gamma_{3j}) = \theta_j + \beta_1 - \beta_2$
$i=4$	$-$	$-$	0	2	8	$n_4 = \sum y_{4j} = 10$	$\text{logit}(\gamma_{4j}) = \theta_j + \beta_1 + \beta_2$

where $\gamma_{ij} = P(Y \le j \mid \mathbf{x}_i)$ is a cumulative probability. The model

$$\text{logit}(\gamma_{ij}) = \theta_j - \boldsymbol{\beta}^T \mathbf{x}_i, \ \ j = 1, 2$$

is known as a **proportional odds model** (McCullagh, 1980) or **cumulative logit model** for ordinal response.

Emergence of House Flies (Zocchi and Atkinson, 1999)

Hierarchical responses: Numbers of unopened pupae (y_1), flies died before emergence (y_2), and flies completed emergence (y_3)

Dose of radiation (Gy) x	y_1	Response categories y_2	y_3	Total number of pupae
80	62	5	433	500
100	94	24	382	500
120	179	60	261	500
140	335	80	85	500
160	432	46	22	500
180	487	11	2	500
200	498	2	0	500

A **continuation-ratio logit model** with **non-proportional odds**:

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

Multinomial Logistic Models in the Literature

Four kinds of logit models used for multinomial responses:

- **baseline-category logit model** for nominal responses (Agresti, 2013; Zocchi and Atkinson, 1999)
- **cumulative logit model** for ordinal responses (McCullagh, 1980; Christensen, 2015)
- **adjacent-categories logit model** for ordinal responses (Liu and Agresti, 2005; Agresti, 2013)
- **continuation-ratio logit model** for hierarchical responses (Agresti, 2013; Zocchi and Atkinson, 1999)

In practice, three types of additional assumptions were made

- **proportional odds (po)**: the parameters for different levels of logits are the same, widely assumed for ordinal responses
- **non-proportional odds (npo)**: the parameters for different levels of logits are differet, usually used for nominal responses
- **partial proportional odds (ppo)**: incorporating both po and npo components, proposed by Peterson and Harrell (1990)

Four Logit Models with Three Types of Odds

In terms of ppo, the four logit models can be expressed as

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

where $\pi_{ij} = P(Y = j \mid \mathbf{x}_i)$,
$\mathbf{x}_i = (x_{i1}, \ldots, x_{id})^T$ is the ith experimental setting,
$\mathbf{h}_j^T(\mathbf{x}_i) \equiv 1$ leads to proportional odds (po) models,
$\boldsymbol{\zeta} = \mathbf{0}$ leads to non-proportional odds (npo) models.

Unified Form for All Four Logit Models with Different Odds

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

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

where $\boldsymbol{\pi}_i = (\pi_{i1}, \ldots, \pi_{iJ})^T$, $\boldsymbol{\eta}_i = (\eta_{i1}, \ldots, \eta_{iJ})^T$,

$$\mathbf{C}^T = \begin{pmatrix} 1 & & & & -1 & & & & 0 \\ & 1 & & & & -1 & & & 0 \\ & & \ddots & & & & \ddots & & \vdots \\ & & & 1 & & & & -1 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}_{J\times(2J-1)}$$

$\mathbf{L}$ takes different forms for the four models, and the $J \times p$ matrix X_i and $p \times 1$ parameter vector $\boldsymbol{\theta}$ depend on po, npo, or ppo.

Relevant literature in optimal design theory

- Two categories ($J = 2$): a generalized linear model for binary data (McCullagh and Nelder, 1989).
  A growing body of design literature: Khuri, Mukherjee, Sinha, and Ghosh (2006); Atkinson, Donev, and Tobias (2007); Stufken and Yang (2012), and references therein.

- Three or more categories ($J \geq 3$): a special case of the multivariate generalized linear model (Glonek and McCullagh, 1995).
  Limited design results: Zocchi and Atkinson (1999); Perevozskaya, Rosenberger, and Haines (2003); Yang, Tong, and Mandal (2017).

Two types of optimal design problems

- Optimal design with quantitative or continuous factors: Identify design points $\mathbf{x}_1, \ldots, \mathbf{x}_m$ (m is not fixed) from a continuous region, and the corresponding weights $p_1, \ldots, p_m$. See, for example, Atkinson, Donev, and Tobias (2007); Stufken and Yang (2012).
- Optimal design with pre-determined design points $\mathbf{x}_1, \ldots, \mathbf{x}_m$ (m is fixed): Find the optimal weights $p_1, \ldots, p_m$. See Yang, Mandal, and Majumdar (2012, 2016); Yang and Mandal (2015); Tong, Volkmer, and Yang (2014).

One connection between the two types is through grid points of continuous region.
Tong, Volkmer, and Yang (2014) also bridged the gap in a way that the results involving discrete factors can be applied to the cases with continuous factors as well.

Fisher Information Matrix, First Form

Theorem 1 (Glonek and McCullagh, 1995)

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

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

where

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

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

Theorem 1 provides an explicit way of calculating the Fisher information matrix. It is actually valid for a more general framework for multiple categorical responses.

Fisher Information Matrix, Second Form

In order to simplify the determinant of **F**, we need

Theorem 2 (Bu, Majumdar, and Yang, 2017)

The Fisher information matrix of the multinomial logistic model (1) is

$$\mathbf{F} = n\mathbf{G}^T\mathbf{W}\mathbf{G} \tag{2}$$

where $\mathbf{W} = \mathrm{diag}\{w_1 \mathrm{diag}(\boldsymbol{\pi}_1)^{-1}, \ldots, w_m \mathrm{diag}(\boldsymbol{\pi}_m)^{-1}\}$ *is an* $mJ \times mJ$ *matrix with* $w_i = n_i/n$, $\mathbf{G}$ *is an* $mJ \times p$ *matrix which takes the forms of*

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

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

Determinant of Fisher Information Matrix

Theorem 3 (Bu, Majumdar, and Yang, 2017)

Up to the constant n^p, the determinant of Fisher information matrix is

$$|\mathbf{G}^T\mathbf{W}\mathbf{G}| = \sum_{\alpha_1\geq 0,\ldots,\alpha_m\geq 0 \,:\, \sum_{i=1}^m \alpha_i = p} c_{\alpha_1,\ldots,\alpha_m} \cdot w_1^{\alpha_1} \cdots w_m^{\alpha_m} \qquad (3)$$

with $c_{\alpha_1,\ldots,\alpha_m} =$

$$\sum_{(i_1,\ldots,i_p)\in\Lambda(\alpha_1,\ldots,\alpha_m)} |\mathbf{G}[i_1,\ldots,i_p]|^2 \prod_{k:\alpha_k>0} \prod_{l:(k-1)J<i_l\leqslant kJ} \pi^{-1}_{k,i_l-(k-1)J} \geq 0 \qquad (4)$$

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

Simplification of $|\mathbf{F}|$

Theorem 4 (Bu, Majumdar, and Yang, 2017)

The coefficient $c_{\alpha_1,\ldots,\alpha_m} = 0$ *if*

(1) $\max_{1\le i\le m} \alpha_i \ge J$; *or*

(2) $\#\{i \mid \alpha_i > 0\} \le k_{\min} - 1$, *where*

$$k_{\min} = \begin{cases} \max\{p_1,\ldots,p_{J-1}\} & \text{for npo models;} \\ p_c + 1 & \text{for po models;} \\ \max\{p_1,\ldots,p_{J-1}, p_c + p_H\} & \text{for ppo models;} \\ p_c + p_1 & \text{for ppo with same } \mathbf{H}_j \end{cases}.$$

Here $k_{\min}$ is actually the minimal number of experimental settings to keep $|\mathbf{F}| > 0$. Recall that the number of parameters is $p = p_1 + \cdots + p_{J-1} + p_c$.
Note that npo models imply $p_c = 0$ and $p_H \le \min\{p_1,\ldots,p_{J-1}\}$, po models imply $p_1 = \cdots = p_{J-1} = p_H = 1$.

Locally D-optimal designs maximizing $f(n_1, \ldots, n_m) = |F|$

The *D-optimal exact design* problem is to solve

$$\begin{aligned} & \max f(n_1, n_2, \ldots, n_m) \\ \text{subject to} \quad & n_i \in \{0, 1, \ldots, n\}, \; i = 1, \ldots, m \\ & n_1 + n_2 + \cdots + n_m = n \end{aligned}$$

Denote $p_i = n_i/n$, $i = 1, \ldots, m$.

$$f(n_1, \ldots, n_m) = \left| \sum_{i=1}^{m} n_i A_i \right| = \left| n \sum_{i=1}^{m} p_i A_i \right| = n^{d+J-1} f(p_1, \ldots, p_m)$$

The *D-optimal approximate design* problem is to solve

$$\begin{aligned} & \max f(p_1, p_2, \ldots, p_m) \\ \text{subject to} \quad & 0 \le p_i \le 1, \; i = 1, \ldots, m \\ & p_1 + p_2 + \cdots + p_m = 1 \end{aligned}$$

Theorems for D-optimality

Karush-Kuhn-Tucker type (Karush, 1939; Kuhn and Tucker, 1951):

Theorem 5

$\mathbf{p} = (p_1^*, \ldots, p_m^*)^T$ *is D-optimal if and only if there exists a* $\lambda \in \mathbb{R}$ *such that for* $i = 1, \ldots, m$, *either* $\partial f(\mathbf{p})/\partial p_i = \lambda$ *if* $p_i^* > 0$ *or* $\partial f(\mathbf{p})/\partial p_i \leq \lambda$ *if* $p_i^* = 0$.

General-equivalence-theorem type (Kiefer, 1974; Pukelsheim, 1993; Atkinson et al., 2007; Stufken and Yang, 2012; Fedorov and Leonov, 2014; Yang, Mandal and Majumdar, 2016):

Theorem 6

$\mathbf{p} = (p_1^*, \ldots, p_m^*)^T$ *is D-optimal if and only if for each* $i = 1, \ldots, m$, $f_i(z), 0 \leq z \leq 1$ *attains it maximum at* $z = p_i^*$, *where* $f_i(z) = f\left(\frac{1-z}{1-p_i} p_1, \ldots, \frac{1-z}{1-p_i} p_{i-1}, z, \frac{1-z}{1-p_i} p_{i+1}, \ldots, \frac{1-z}{1-p_i} p_m\right)$

Emergence of house flies revisited (Bu, Majumdar, and Yang, 2017)

Consider a followup experiment with 3500 pupae again. Using our numerical algorithms, we obtain various D-optimal designs.

Dose of radiation (Gy)	80	100	120	140	160	180	200
Original allocation	500	500	500	500	500	500	500
D-optimal exact	1091	0	1021	374	1014	0	0
Original proportion	.1429	.1429	.1429	.1429	.1429	.1429	.1429
D-optimal approximate	.3116	0	.2917	.1071	.2896	0	0
Bayesian D-optimal	.3159	.0000	.2692	.1160	.2990	.0000	.0000
EW D-optimal	.3120	0	.2911	.1087	.2882	0	0

Compared with the D-optimal approximate design, the efficiency of the original uniform allocation is
$(|\mathbf{F}_{original}|/|\mathbf{F}_{D-opt}|)^{1/5} = (585317/1480378)^{1/5} = 83.1\%$.

Minimally supported designs

A *minimally supported design* is a design with the minimal number of support/design points while keeping $|F| > 0$.

- $J = 2$: It is actually binomial with $d + 1$ parameters, $\theta_1, \beta_1, \ldots, \beta_d$.
  It is known that the minimal number is $d + 1$; and
  the uniform allocation is D-optimal on a minimally supported design.
- $J \geq 3$: There are $d + J - 1$ parameters, $\theta_1, \ldots, \theta_{J-1}, \beta_1, \ldots, \beta_d$.
  According to Yang, Tong, and Mandal (2017), the minimal number is still $d + 1$; and
  the uniform allocation is NOT D-optimal in general.

Odor removal study revisited (Yang, Tong, and Mandal, 2017)

Suppose we want to conduct a followup experiment with n runs. Using some numerical algorithms we proposed, we obtain the D-optimal exact designs, as well as the D-optimal approximate design $\mathbf{p}_o$.

n	n_1	n_2	n_3	n_4	$n^{-4}\|F\|$	# iterations	Time(sec.)
3	1	1	0	1	0.0002872	2	< 0.01
10	4	3	0	3	0.0003093	3	0.01
40	18	11	0	11	0.0003137	2	0.03
100	44	29	0	27	0.0003141	2	0.08
1000	445	287	0	268	0.0003141	4	0.33
10000	4456	2869	0	2675	0.0003141	6	3.21
$\mathbf{p}_o$	0.4455	0.2867	0	0.2678	0.0003141	3	0.05

Compared with the D-optimal exact design $\mathbf{n}_o = (18, 11, 0, 11)^T$ at $n = 40$, the relative efficiency of the uniform exact design $\mathbf{n}_u = (10, 10, 10, 10)^T$ is $(f(\mathbf{n}_u)/f(\mathbf{n}_o))^{1/4} = 79.7\%$.

Fisher information matrix for multinomial logistic models, Third Form

Theorem 7 (Bu, Majumdar, and Yang, 2017)

The Fisher information matrix $\mathbf{F} = \mathbf{H}\mathbf{U}\mathbf{H}^T$, *where* $\mathbf{H}$ *is*

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

for ppo, npo, and po models respectively,
$\mathbf{H}_j = (\mathbf{h}_j(\mathbf{x}_1), \cdots, \mathbf{h}_j(\mathbf{x}_m))$, $j = 1, \ldots, J-1$,
$\mathbf{H}_c = (\mathbf{h}_c(\mathbf{x}_1), \cdots, \mathbf{h}_c(\mathbf{x}_m))$, *and*

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

with block matrices $\mathbf{U}_{st} = \mathrm{diag}\{n_1 u_{st}(\boldsymbol{\pi}_1), \ldots, n_m u_{st}(\boldsymbol{\pi}_m)\}$.

Positive Definiteness of Fisher Information Matrix

Towards the positive definiteness of $\mathbf{F} = \mathbf{H}\mathbf{U}\mathbf{H}^T$, we have

Theorem 8 (Bu, Majumdar, and Yang, 2017)

Assume $\pi_{ij} > 0$, $n_i > 0$ for $i = 1, \ldots, m$ and $j = 1, \ldots, J$. Then
(i) $\mathbf{U}$ is positive definite;
(ii) $\mathbf{F}$ is positive definite if and only if $\mathbf{H}$ is of full row rank.

Remark: In general, we may denote $k := \#\{i : n_i > 0\} \leq m$, $\mathbf{U}^*_{st} = \mathrm{diag}\{n_i u_{st}(\boldsymbol{\pi}_i) : n_i > 0\}$, $\mathbf{U}^* = (\mathbf{U}^*_{st})_{s,t=1,\ldots,J-1}$, and remove all columns of $\mathbf{H}$ associated with $n_i = 0$ and denote the leftover as $\mathbf{H}^*$. Then

$$\mathbf{H}\mathbf{U}\mathbf{H}^T = (\mathbf{H}^*)(\mathbf{U}^*)(\mathbf{H}^*)^T$$

Some Key Findings for Multinomial Logistic Models

- The minimal number of experimental settings is

$$k_{\min} = \begin{cases} \max\{p_1, \ldots, p_{J-1}\} & \text{for npo models;} \\ p_c + 1 & \text{for po models;} \\ \max\{p_1, \ldots, p_{J-1}, p_c + p_H\} & \text{for ppo models;} \\ p_c + p_1 & \text{for ppo with same } \mathbf{H}_j \text{ .} \end{cases}$$

  which is less than the number of parameters $p_1 + \cdots + p_{J-1} + p_c$.
- With $J \geq 3$, the uniform allocation for a minimally supported design is NOT D-optimal in general.
- For "regular" npo models (that is, $p_1 = \cdots = p_{J-1}$), a uniform allocation is still D-optimal if restricted on a minimally supported design even with $J \geq 3$.

References

- Bu, X., Majumdar, D. and Yang, J. (2017). D-optimal designs for multinomial logistic models, submitted for publication, available at `https://arxiv.org/pdf/1707.03063.pdf`.
- Glonek, G.F.V. and McCullagh, P. (1995). Multivariate logistic models, *Journal of the Royal Statistical Society, Series B*, **57**, 533-546.
- McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, **42**, 109-142.
- Tong, L., Volkmer, H.W., and Yang, J. (2014). Analytic solutions for D-optimal factorial designs under generalized linear models. *Electronic Journal of Statistics*, **8**, 1322−1344.
- Yang, J., Mandal, A. and Majumdar, D. (2016). Optimal designs for 2^k factorial experiments with binary response, *Statistica Sinica*, **26**, 385-411.
- Yang, J., Tong, L. and Mandal, A. (2017). D-optimal designs with ordered categorical data, *Statistica Sinica*, **27**, 1879-1902.
- Zocchi S.S. and Atkinson A.C. (1999). Optimum experimental designs for multinomial logistic models, *Biometrics*, **55**, 437-444.