New Challenges and Strategies in Robust Optimal Design for Multicategory Logit Modelling

Timothy E. O'Brien and Changwon Lim

1 Introduction

Binary logistic and multi-category logit (MCL) regression models are amongst the most popular techniques in applied research where a goal is to determine relationships between attributes and/or adjusting for covariates. As such, introductory statistics texts cover these methods, and many applications-focused students note their usefulness in basic statistical methods courses. Aside from choosing from probit-based or logit-based link functions, modelling in the logistic case is relatively straightforward. But the situation is complicated in the multi-category case since several reasonable rival models have been suggested to handle these data. In these MCL cases, the practitioner is thus faced with choosing one of these models over the others, and, more importantly, deciding which experimental design to use. As in all cases of modelling, it is desired that this design should then allow for efficient model-parameter estimation and provide for a test of goodness-of-fit of the chosen model.

Important background to quantal, logistic and multicategory modelling is given in McCullagh and Nelder (1989), Agresti (2007, 2013), and Dobson and Barnett (2008), and extensions and applications are provided in Finney (1978). Optimal design strategies are introduced and illustrated in Silvey (1980), O'Brien and Funk (2003) and Atkinson et al. (2007), and geometric and uniform designs are explored in O'Brien et al. (2009).

T.E. O'Brien (✉)
Department of Mathematics and Statistics and Institute of Environmental Sustainability, Loyola University of Chicago, Chicago, IL, USA
e-mail: tobrie1@luc.edu

C. Lim
Department of Applied Statistics, Chung Ang University, Seoul, South Korea
e-mail: changwon77@gmail.com

D.-G. Chen et al. (eds.), *New Advances in Statistics and Data Science*,
ICSA Book Series in Statistics, https://doi.org/10.1007/978-3-319-69416-0_4

In the context of typical MCL modelling situations, in what follows we provide needed background and introduce and demonstrate the usefulness of model-robust near optimal designs, highlighting extensions that allow for geometric and uniform design strategies. Thus, these results provide practitioners with useful guidelines in situations where potentially several MCL models can be chosen for a given dataset. Note that although the illustrations provided in this paper concern only three-level outcomes with a single explanatory variable, the results have been applied to numerous illustrations involving several independent variables and as many as five outcome categories.

2 Quantal Dose-Response Modelling

For the binary logistic model, where the x variable corresponds to dose or concentration, it is common that the researcher wishes to select the k dose points to run the experiment. This dose selection as well as the number of replicates at each of these points is the experimental design problem addressed here in a larger context. For n_i experimental units receiving dose x_i, the logistic model holds that the number of "successes" y_i has a binomial distribution with success probability π_i; under the assumed logit link, we obtain the generalized linear model equation, $\log\left(\frac{\pi_i}{1-\pi_i}\right) = \alpha + \beta x_i$. Also, when this model function is reparameterized so that the ED_{50} parameter $\gamma = \frac{-\alpha}{\beta}$ is a model parameter—so that the right-hand side in this expression is $\beta(x-\gamma)$—the model then becomes generalized nonlinear model. Important references for generalized linear and nonlinear models include McCullagh and Nelder (1989), Agresti (2007, 2013), and Dobson and Barnett (2008).

In contrast with binary logistic situation—where experiments result in "successes" or "failures"—often the number of outcomes is three or more. Commonly-used models for these data include the adjacent category logit (ACL), baseline category logit (BCL), continuation ratio (CR), and proportional odds (PO). For example, in the case of $K = 3$ outcomes and single predictor x, the ACL model is given by the simultaneous equations

$$\begin{cases} (i)\ \log\left(\frac{\pi_1}{\pi_2}\right) = \alpha_1 + \beta_1 x \\ (ii)\ \log\left(\frac{\pi_2}{\pi_3}\right) = \alpha_2 + \beta_2 x \end{cases} \tag{1}$$

Denoting $ex_1 = e^{\alpha_1+\beta_1 x}, ex_2 = e^{\alpha_2+\beta_2 x}, den = 1 + ex_1 + (ex_1)(ex_2)$, this expression is equivalent to $\pi_1 = (ex_1)(ex_2)/den, \pi_2 = ex_2/den, \pi_3 = 1/den$. To obtain parameter estimates, confidence regions/intervals and experimental designs, these expressions can be substituted into the log-likelihood expression. The BCL model amends the left-hand sides of the expressions in (1) with $(i)\log(\pi_1/\pi_3)$ and $(ii)\log(\pi_2/\pi_3)$. It is therefore observed that the BCL model is equivalent to the ACL model through a simple reparameterization, and it is therefore subsumed by

Table 1 Multicategory logit models for $K = 3$ outcomes

Continuation ratio A (CRA) model $\begin{cases} (i)\ \log\left(\frac{\pi_1}{\pi_2}\right) = \alpha_1 + \beta_1 x \\ (ii)\ \log\left(\frac{\pi_1+\pi_2}{\pi_3}\right) = \alpha_2 + \beta_2 x \end{cases}$	Un-proportional odds (UPO) logit model $\begin{cases} (i)\ \log\left(\frac{\pi_1}{\pi_2+\pi_3}\right) = \alpha_1 + \beta_1 x \\ (ii)\ \log\left(\frac{\pi_1+\pi_2}{\pi_3}\right) = \alpha_2 + \beta_2 x \end{cases}$
Adjacent category logit (ACL) model $\begin{cases} (i)\ \log\left(\frac{\pi_1}{\pi_2}\right) = \alpha_1 + \beta_1 x \\ (ii)\ \log\left(\frac{\pi_2}{\pi_3}\right) = \alpha_2 + \beta_2 x \end{cases}$	Continuation ratio B (CRB) model $\begin{cases} (i)\ \log\left(\frac{\pi_1}{\pi_2+\pi_3}\right) = \alpha_1 + \beta_1 x \\ (ii)\ \log\left(\frac{\pi_2}{\pi_3}\right) = \alpha_2 + \beta_2 x \end{cases}$

results for the ACL model. Additional details regarding multicategory logit models are given below as well as in Agresti (2007, 2013).

In addition to the ACL model, a listing of useful multicategory logit models is given in Table 1. For $K = 3$ outcomes, each of these models entails two equations. These expressions are easily extended to $K > 3$ outcomes where each model would then contain $(K - 1)$ equations.

As specified in Table 1, in addition to the ACL model, commonly-used models include the two variants of the Continuation Ratio model (denoted CRA and CRB here) as well as the Proportional Odds (PO) model. The PO model is derived from the UPO model imposing the equal-slope restriction, viz, $\beta_1 = \beta_2 (= \beta)$. In addition to noting similarities and differences in models, an important goal in listing these models here is to unify them under one umbrella in order to provide the researcher with near-optimal robust designs (see Sect. 5).

Example 1 Price et al. (1987) provides toxicity data involving pregnant mice in which the predictor variable is the concentration of a certain ether. The chosen concentration levels in the study were $x_i = 0, 62.5, 125, 250, 500$ mg/kg per day. With respective sample sizes of $n_i = 297, 242, 312, 299, 285$, the total sample size is $n = 1435$ mice. The response variable here encompassed the three levels relating to the status of the offspring: death, malformed, or normal. Among the model functions given in Table 1, the model with the highest log-likelihood value (and thus AIC) here is the CRB model, with maximum likelihood estimates: $\widehat{\alpha}_1 = -3.2479, \widehat{\beta}_1 = 0.0064, \widehat{\alpha}_2 = -5.7019, \widehat{\beta}_2 = 0.0174$. In terms of interpretation of these estimates, since equation (i) in the CRB model contrasts dead with alive offspring and equation (ii) contrasts malformed with normal offspring, these results are best interpreted in terms of odds ratios: as the concentration level increases by an additional 100 mg/kg/day, the odds of a dead pup (versus alive) increases by a multiplicative factor of $e^{100\widehat{\beta}_1} = 1.89$ and the odds of a malformed pup (versus normal) increases by a multiplicative factor of $e^{100\widehat{\beta}_2} = 5.68$.

We return to this illustration below to demonstrate ways to improve upon the chosen experimental design.

3 Confidence Regions and Intervals

As noted in Seber and Wild (1989), in the case of normal linear and nonlinear models involving the p-vector $\boldsymbol{\theta}$ of model parameters, $(1-\alpha)100\%$ Wald confidence regions for $\boldsymbol{\theta}$ are of the form: $\left\{\boldsymbol{\theta} \in \Theta : \left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\right)^T \widehat{\boldsymbol{V}}^T \widehat{\boldsymbol{V}} \left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\right) \leq ps^2 F_\alpha\right\}$. In this expression, $\widehat{\boldsymbol{\theta}}$ is the least-squares (i.e., maximum likelihood) estimate of $\boldsymbol{\theta}$, $\widehat{\boldsymbol{V}}$ is the $n \times p$ Jacobian matrix of first derivatives evaluated at $\widehat{\boldsymbol{\theta}}$, s^2 is the mean square error (estimator of σ^2), and F_α is a tabled F percentile with p and $n-p$ degrees of freedom with tail probability of α. The $(1-\alpha)100\%$ likelihood-based confidence region in this situation is $\left\{\boldsymbol{\theta} \in \Theta : S(\boldsymbol{\theta}) - S\left(\widehat{\boldsymbol{\theta}}\right) \leq ps^2 F_\alpha\right\}$. Here, $S(\boldsymbol{\theta}) = (\boldsymbol{y} - \boldsymbol{\eta}(x, \boldsymbol{\theta}))^T (\boldsymbol{y} - \boldsymbol{\eta}(x, \boldsymbol{\theta})) = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}$. These two regions will be nearly equivalent depending upon the degree to which the (vector) model function, $\boldsymbol{\eta}(x, \boldsymbol{\theta})$, is well-approximated by the planar expression, $\boldsymbol{\eta}\left(x, \widehat{\boldsymbol{\theta}}\right) + \widehat{\boldsymbol{V}}\left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\right)$. In normal linear models, this result is exactly met, and only approximately so for normal nonlinear, generalized linear, and generalized nonlinear models.

In non-normal situations, such as those considered here, approximate $(1-\alpha)100\%$ likelihood-based confidence regions are of the form $\left\{\boldsymbol{\theta} \in \Theta : 2\left[LL(\boldsymbol{\theta}) - LL\left(\widehat{\boldsymbol{\theta}}\right)\right] \leq \chi^2_\alpha\right\}$, where $LL(\theta)$ is the model log-likelihood and χ^2_α is a tabled χ^2 percentile with p degrees of freedom and tail probability equal to α. Wald and likelihood confidence intervals can be obtained from these regions by conditioning or profiling; further details are given in Seber and Wild (1989) and Pawitan (2013). Notably, often the researcher wishes to choose an experimental design to reduce the length of the resulting confidence interval or the volume of the resulting confidence region.

4 Optimal Design Theory

An n-point design, denoted ξ, is written

$$\xi = \begin{Bmatrix} x_1 & x_2 & \ldots & x_n \\ \omega_1 & \omega_2 & \ldots & \omega_n \end{Bmatrix} \tag{2}$$

The ω_i are non-negative design weights which sum to one, and the x_i are design points (or vectors) that belong to the design space, and which are not necessarily distinct. For the constant-variance normal setting with linear or nonlinear normal model function $\eta(x, \boldsymbol{\theta})$, the $n \times p$ Jacobian matrix is $\boldsymbol{V} = \frac{\partial \eta}{\partial \theta}$. Denoting $\boldsymbol{\Omega} = \text{diag}\{\omega_1, \omega_2, \ldots, \omega_n\}$, the $p \times p$ (Fisher) information matrix is then written

$$\boldsymbol{M}(\xi, \boldsymbol{\theta}) = \boldsymbol{V}^T \boldsymbol{\Omega} \boldsymbol{V} \tag{3}$$

In the more general case of either non-constant variance or non-normality, the corresponding information matrix is given by

$$\boldsymbol{M}(\xi, \boldsymbol{\theta}) = -E\left(\frac{\partial^2 LL}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}^T}\right) \tag{4}$$

As underscored in Atkinson et al. (2007), the information matrix for the binary logistic model has the same form as in (3) with an appropriate modification of the weight matrix $\boldsymbol{\Omega}$. Since the (asymptotic) variance of $\widehat{\boldsymbol{\theta}}_{MLE}$ is proportional to $\boldsymbol{M}^{-1}(\xi, \boldsymbol{\theta})$, in many regression settings designs are often chosen to minimize some (convex) function of $\boldsymbol{M}^{-1}(\xi, \boldsymbol{\theta})$. For example, designs which minimize its determinant are called D-optimal. As noted in Seber and Wild (1989), these designs minimize the volume of the confidence region given in the previous section. Since for nonlinear/logistic models, $\boldsymbol{M}$ depends upon $\boldsymbol{\theta}$, so-called local (or Bayesian) designs are typically obtained.

The (approximate) variance of the predicted response at the value x is

$$d(x, \xi, \boldsymbol{\theta}) = \frac{\partial \eta(x, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \boldsymbol{M}^{-1}(\xi) \frac{\partial \eta(x, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = tr\left\{\boldsymbol{M}^{-1}(\xi)\, \boldsymbol{M}(x)\right\} \tag{5}$$

Here, $\boldsymbol{M}(x) = \frac{\partial \eta(x, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \eta(x, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}$ is the information matrix evaluated at the arbitrary value x; note that in contrasting with Eq. (4) where it is highlighted that for nonlinear models the information matrix depends upon the design and parameter values, occasionally one or both of these symbols are drop in what follows merely for typographic simplicity. Designs that minimize (over ξ) the maximum (over x) of $d(x, \xi, \boldsymbol{\theta})$ in (5) are called G-optimal. As stated above, since this predicted variance depends upon $\boldsymbol{\theta}$ for logistic and nonlinear models, researchers often seek optimal designs either using a "best guess" for $\boldsymbol{\theta}$ (called a local optimal design) or by assuming a plausible prior distribution on $\boldsymbol{\theta}$ (called a Bayesian optimal design).

The General Equivalence Theorem (GET) of Kiefer and Wolfowitz (1960) establishes that D- and G-optimal designs are equivalent. This theorem also demonstrates that the variance function (5) evaluated using the D−/G-optimal design does not exceed the line $y = p$ (where p is the number of model function parameters)—but that it will exceed this line for all other designs. A corollary of the GET establishes that the maximum of the variance function is achieved for the D−/G-optimal design at the support points of this design. This result is very useful in demonstrating optimality of a given design, by substituting it into (5) and plotting the resulting variance function. Results and additional references for optimal design in binary logistic settings are given in Abdelbasit and Plackett (1983) and Minkin (1987), and in the general setting in Silvey (1980).

Example 1 Continued For the pregnant mice illustration and CRB model with the (MLE) parameter estimates given above and design points in the range [0, 2000], the local D-optimal design associates the respective weights $w = 0.4058, 0.3805, 0.2136$ with design support points (concentrations) $x = 222.59, 401.35, 767.91$. The

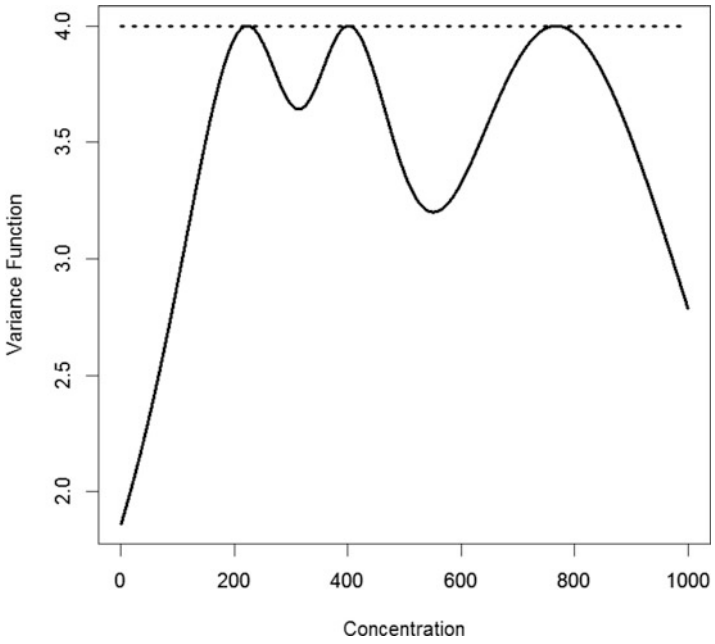

Fig. 1 Variance function for CRB model using D-optimal design—pregnant mice example

corresponding variance-function plot is shown in Fig. 1 along with the cut line, $y = 4$, since this model contains $p = 4$ parameters. D-optimality is established here by noting that the variance function does not exceed the cut-line.

Mindful that for the models considered here we are typically more interested in efficient estimation of only a subset of the model parameters, we partition the Fisher information matrix as

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{M}_{11} & \boldsymbol{M}_{12} \\ \boldsymbol{M}_{21} & \boldsymbol{M}_{22} \end{bmatrix} \tag{6}$$

In this expression, each sub-matrix $\boldsymbol{M}_{ij}$ is of dimension $p_i \times p_j$ for $i, j = 1, 2$, and $p_1 + p_2 = p$. In the current situation, the parameter vector is similarly partitioned, $\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{pmatrix}$ with $\boldsymbol{\theta}_1$ of dimension $p_1 \times 1$, $\boldsymbol{\theta}_2$ of dimension $p_2 \times 1$, and $\boldsymbol{\theta}_1$ is the parameter vector of interest and $\boldsymbol{\theta}_2$ are the nuisance parameters. Subset D-optimal designs for $\boldsymbol{\theta}_2$ in the joint model, as discussed in Atkinson et al. (2007), are obtained by maximizing

$$\left|\boldsymbol{M}_{22} - \boldsymbol{M}_{21}\boldsymbol{M}_{11}^{-1}\boldsymbol{M}_{12}\right| = \frac{|\boldsymbol{M}|}{|\boldsymbol{M}_{11}|} \tag{7}$$

Noting problems associated with subset designs, O'Brien (2005) and Atkinson et al. (2007) instead combine the subset and full-parameter criteria and suggest that designs be chosen to maximize the objective function

$$\Phi_{\phi}(\xi, \boldsymbol{\theta}) = \frac{1-\phi}{p_1}\log|\boldsymbol{M}_{11}| + \frac{\phi}{p_2}\log\left|\boldsymbol{M}_{22} - \boldsymbol{M}_{21}\boldsymbol{M}_{11}^{-1}\boldsymbol{M}_{12}\right| \tag{8}$$

For ϕ chosen in the interval $\left[0, \frac{p_2}{p}\right]$, we call designs that maximize (8) D_{ϕ}-optimal. The resulting designs range from D-optimal designs for the $\boldsymbol{\theta}_1$ parameters in the smaller model containing only the $\boldsymbol{\theta}_1$ parameters for the choice $\phi = 0$ to D-optimal designs for the full $\boldsymbol{\theta}$ parameter vector in the larger model for the choice $\phi = \frac{p_2}{p}$. The corresponding variance function associated with (8) and an extension of the General Equivalence Theorem are then used to ensure D_{ϕ}-optimality of the resulting design by plotting the variance function, with the note that this normalized variance function has cut line $y = 1$ instead of $y = p$. To illustrate using the first example given in O'Brien (2005), the subset design for the two-parameter intermediate product model comprises only a single design support point and so is a singular design, whereas the D_{ϕ}-optimal design has two support points for ϕ in $\left(0, \frac{p_2}{p}\right]$.

A measure of the distance or discrepancy between an arbitrary design ξ_C and the D-optimal design ξ_D^* is the D-efficiency discussed in O'Brien and Funk (2003) and Atkinson et al. (2007), and given by the expression

$$\left(\frac{|\boldsymbol{M}(\boldsymbol{\xi}_C)|}{|\boldsymbol{M}(\boldsymbol{\xi}_D^*)|}\right)^{1/p} \tag{9}$$

To illustrate, for an arbitrary design ξ_C with a D-efficiency of 66.7%, the researcher would need 50% more (1/0.667) experimental units to obtain the same information as the D-optimal design. Thus, in this setting, the same information would thus be achieved using the D-optimal design and only 120 experimental units as with the chosen (arbitrary) design using 180 experimental units.

The above advantage (i.e., optimality) notwithstanding, optimal designs can often only be used as a starting point in realistic situations since they often have some associated shortcomings. One important shortcoming is that often in practice, optimal designs for p-parameter model functions comprise only p support points, and so they provide little or no ability to test for lack of fit of the assumed model. Indeed, for the pregnant mice example discussed above, although the model contains $p = 4$ model parameters, the D-optimal design contains only three support points, so this design gives little or no means to check model adequacy. Further, in spite of the important theoretical optimal design results given in Zocchi and Atkinson (1999), Fan and Chaloner (2001), and Perevozskaya et al. (2003) for the CRB, CRA and PO models respectively, these works do not directly deal with the model-robustness issues raised and addressed here.

Table 2 Local D-optimal designs for pregnant mice example

Continuation ratio A (CRA) model $\xi^*_{CRA} = \begin{Bmatrix} 194.5 & 428.1 & 1682.0 \\ 0.3023 & 0.4531 & 0.2445 \end{Bmatrix}$	Un-proportional odds (UPO) logit model $\xi^*_{UPO} = \begin{Bmatrix} 0 & 353.2 & 678.2 \\ 0.3575 & 0.4066 & 0.2359 \end{Bmatrix}$
Adjacent category logit (ACL) model $\xi^*_{AC} = \begin{Bmatrix} 193.5 & 425.5 & 1554.8 \\ 0.3037 & 0.4527 & 0.2435 \end{Bmatrix}$	Continuation ratio B (CRB) model $\xi^*_{CRB} = \begin{Bmatrix} 222.6 & 401.3 & 767.9 \\ 0.4058 & 0.3805 & 0.2136 \end{Bmatrix}$

Importantly, optimal designs can also vary substantially—including the ACL, CRA, CRB and UPO models considered here. To illustrate, for the pregnant mice example and the concentration-range $[0, 2000]$ as used in Price et al. (1987), the local D-optimal designs are given in Table 2 (obtained using the respective best fitting model parameter estimates). Note that whereas one such optimal design includes a concentration level as low as 0 mg/kg (i.e., for the UPO model), the highest concentration in another design is almost 1700 mg/kg (i.e., for the CRA model). This underscores the fact that optimal designs for one model may be very inefficient for another model.

As noted above, the designs and design strategies considered to date have focused primarily on efficiently estimating parameters in the assumed model, and not focused on allowing for—or discriminating amongst—other MCL models. Since in general rival models exist, clearly designs should also highlight which model best fits the data. That is, researchers often desire near-optimal so-called "robust" designs which have extra support points that can then be used to test for model adequacy. We next give very useful means to obtain these robust near-optimal designs.

5 Near-Optimal Robust Design Strategies

The structure of the four multicategory logit models considered in Table 1 suggest the following model function, which we refer to as the generalized ordinal logit (GOL) model function:

$$\begin{cases} (i)\ \log\left(\frac{\pi_1}{\pi_2+\theta_1\pi_3}\right) = \alpha_1 + \beta_1 x \\ (ii)\ \log\left(\frac{\theta_2\pi_1+\pi_2}{\pi_3}\right) = \alpha_2 + \beta_2 x \end{cases} \tag{10}$$

In this expression, θ_1 and θ_2 are additional (or "hyper") parameters introduced to connect the above models. The ACL, CRA, CRB, and UPO models result by choosing $(\theta_1, \theta_2) = (0,0), (0,1), (1,0), (1,1)$, respectively. As a result, for the GOL model, we impose the constraints $0 \le \theta_1 \le 1, 0 \le \theta_2 \le 1$; numerically this is achieved by imposing for example for $i = 1, 2, \theta_i = \frac{e^{\psi_i}}{1+e^{\psi_i}}$ so when ψ_i varies between $-\infty$ and ∞, (θ_1, θ_2) is bounded in the unit square. Estimation of the six

model parameters (including the hyper-parameters) can easily be achieved using maximum likelihood estimation algorithms. Although none of the ACL, CRA, CRB, or UPO models are special cases of another, since each of these models is nested in the larger GOL model, differences between each of these models and the best-fitting GOL model can be evaluated using the asymptotic χ^2 test statistic (i.e., two times the change in log-likelihood) with associated 2 degrees of freedom. Further, subsets of this larger family can also be connected: an important such special case of the GOL model is the UPOCRB model, obtain for $\theta_1 = 1$. This latter model connects UPO and CRB models, and is demonstrated in the illustration below.

The key goal of our introducing the GOL model here is to facilitate our obtaining model-robust near-optimal designs. This is achieved by viewing the assumed model function chosen from one of the constituents (viz, ACL, CRA, CRB, and UPO) as an element of the GOL family and using the modified subset design procedure given in (8) to obtain D_ϕ-optimal designs. For example, if the ACL is the assume model function with given *a priori* parameter estimates for this ACL model, it is suggested to use design criterion (8) with $\boldsymbol{\theta}_2^{\boldsymbol{T}} = (\theta_1, \theta_2) = (0, 0)$ and $\boldsymbol{\theta}_1^{\boldsymbol{T}} = (\alpha_1, \beta_1, \alpha_2, \beta_2)$ fixed at the *a priori* parameter estimates. We choose the tuning parameter ϕ in (8) so that the D-efficiency given in (9) for the ACL model exceeds some lower bound such as 90%. We thereby obtain an efficient model-robust D_ϕ-optimal design. This is illustrated in the following example.

Example 1 Continued For the pregnant mice illustration, the best fitting model is the CRB model and second best fitting model is the UPO model. As highlighted in Table 2, the (local) optimal designs for these two models differ substantially, with one design containing a lowest concentration of 0 and the other containing a lower bound in excess of 200. Further, since the fit of these two models to these data is far superior to the other two models, we view the chosen CRB model as embedded in the UPOCRB model. As noted above, we envision the frequently-encountered situation in which the researcher has the CRB model in mind (with *a priori* parameter estimates), and desires a near-optimal design which satisfies the dual objectives of: (1) efficiently estimating the CRB model parameters, and (2) providing for some ability to test for lack-of-fit in the direction of the UPO model. Taking $\phi = 0.05$, the local D_ϕ-optimal design assigns the weights $w = 0.0856, 0.3635, 0.3572, 0.1937$ to the design points (concentrations) $x = 0, 230.5, 405.8, 760.9$. We underscore that the additional design support point reflects the multi-objective nature of this design. Indeed, D_ϕ-optimality of this design is established by noting that the corresponding variance function, plotted in Fig. 2, lies below the cut line $y = 1$. The associated D-efficiency for this design for the CRB model is 95.3% and for the UPO model exceeds 80%, and so it is therefore quite efficient for both models. Certainly, if the researcher was concerned with departures from the assumed CRB model in the direction of the ACL and/or CRA models in addition to the UPO model, we would easily embed the CRB model in the larger GOL model and find the associated D_ϕ-optimal design.

The structure of the design chosen in Price et al. (1987), as well as several additional examples given in O'Brien et al. (2009), underscores the popularity of

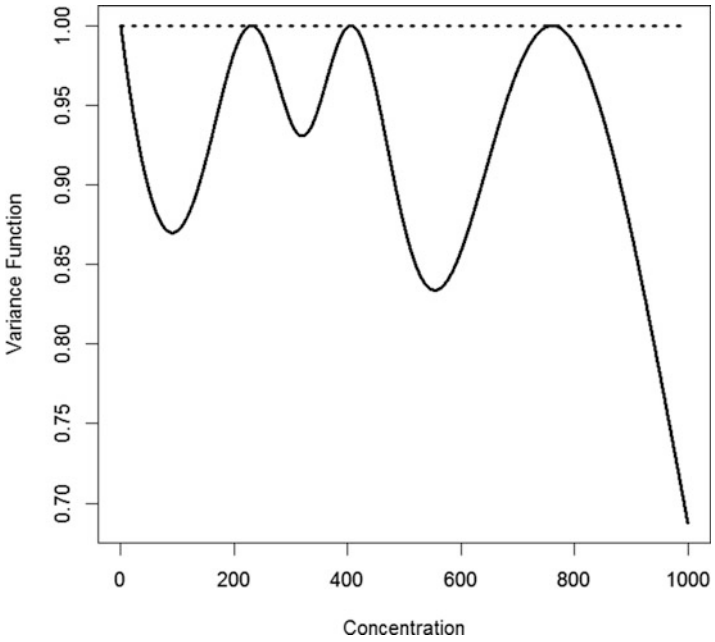

Fig. 2 Variance function for UPOCRB model using D_ϕ-optimal design—pregnant mice example

geometric and uniform designs in practical settings. Thus, we also examine here robust geometric designs of the form $x = a, ab, ab^2 \ldots ab^K$ for multicategory logit models, checking to see whether addition of the point $x = 0$ improves this geometric design. Here, K is specified by the researcher, and computer maximization algorithms are used to obtain optimal values of a and b as well as any associated information loss (as measured by the D-efficiency). We have also obtained optimal uniform designs of the form $A, A + B, A + 2B \ldots A + KB$, letting the final choice of the design structure (geometric or uniform) be the one with the higher D-efficiency or up to the researcher's discretion. So that the final design is robust to the assumed model function choice, we recommend obtaining local D_ϕ-optimal designs using the modified subset design procedure given in (8) and with ϕ chosen to yield a sufficiently-high final D-efficiency for the assumed sub-model.

Example 1 Continued For the pregnant mice illustration and now embedding the CRB model in the GOL model, we have noted somewhat higher D-efficiencies for geometric designs over uniform designs, so we highlight only geometric designs here. As such, we have sought designs which associate weights of the form ω^*, $\frac{1-\omega^*}{4}$, $\frac{1-\omega^*}{4}$, $\frac{1-\omega^*}{4}$, $\frac{1-\omega^*}{4}$ respectively with support points $x = 0, a, ab, ab^2, ab^3$. Hence, robust geometric designs have been obtained here by optimizing over ω^*, a, b. Choosing $\phi = 0.10$ yields the optimal values $\omega^* = 0.054, a = 160.2, b = 1.65$, and produces a robust optimal design with D-efficiency (for the CRB model) of 90.6%. For the total sample size used by the

authors ($n = 1435$), this design assigns $79, 339, 339, 339, 339$ mice to the respective concentrations $x = 0, 169.2, 264.2, 435.8, 718.9$. We emphasize that the original design used in Price et al. (1987) given above—with nearly uniform weights and geometric support points $x = 0, 62.5, 125, 250, 500$—has D-efficiency (for the CRB model) of only 62.8%. Therefore, with a D-efficiency in excess of 90%, the robust optimal geometric design strategy and design suggested here is strongly favored.

Some additional extensions—further demonstrating the breadth of our multiple-objective design strategy—are provided in the following illustration.

Example 2 Zocchi and Atkinson (1999) presents a dataset in which seven sets of 500 housefly pupae were exposed to one of seven doses of gamma radiation. The response variable for this study encompassed the three classes: death, opened but died before complete emergence, and complete emergence. The chosen radiation levels in the study were $x = 80, 100, 120, 140, 160, 180, 200$ Gy, and with equal replicates of $n_i = 500$ fly pupae per level, the total sample size was therefore $n = 3500$. Due to nonlinearities involved with these data, the authors suggest quadratic fits, and for the ACL, CRA, CRB and UPO models considered here, the best-fitting is the quadratic CRA model,

$$\begin{cases} (i)\ \log\left(\frac{\pi_1}{\pi_2}\right) = \alpha_1 + \beta_1 x + \gamma_1 x^2 \\ (ii)\ \log\left(\frac{\pi_1+\pi_2}{\pi_3}\right) = \alpha_2 + \beta_2 x + \gamma_2 x^2 \end{cases} \tag{11}$$

We underscore that, with design points constrained to lie in the design space $[80, 200]$, the (local) D-optimal design for this model places equal weights at only three design points: $x = 80, 125.2, 163.6$. With only three support points, this design is thus of limited use to detect lack of fit of the assumed model. This model is easily embedded in the corresponding quadratic GOL model (which then contains eight parameters), and local D_ϕ-optimal designs using the modified subset design procedure given in (8) can then be easily obtained. Here, with $\phi = 0.25$, the local D_ϕ-optimal design associates the weights $w = 0.2423, 0.0456, 0.2272, 0.2454, 0.2395$ with the five design points $x = 80, 97.8, 116.1, 147.4, 182.1$. With a D-efficiency of 93.5%, this design represents only a minor information loss but a vast improvement in terms of additional design support points and thus the ability to test for model adequacy. To justify the claim of optimality, the corresponding scaled variance function is given in Fig. 3, and D_ϕ-optimality is established by noting that this function lies below the cut-line $y = 1$. Also, among designs of the form $A, A + B, A + 2B, \ldots, A + 6B$, using $\phi = 0.10$, the local D_ϕ-optimal uniform design has the support points, $x = 80, 96.6, 113.2, 129.8, 146.4, 163.0, 179.6$. Our final recommendation would be to allocate 500 fly pupae to each of these seven radiation levels. Whereas the original (uniform) seven-point design given in Zocchi and Atkinson (1999) has a D-efficiency (viz-a-viz the CRA model) of 84.1%, this proposed design increases the D-efficiency to 92.1%, and represents a modest improvement.

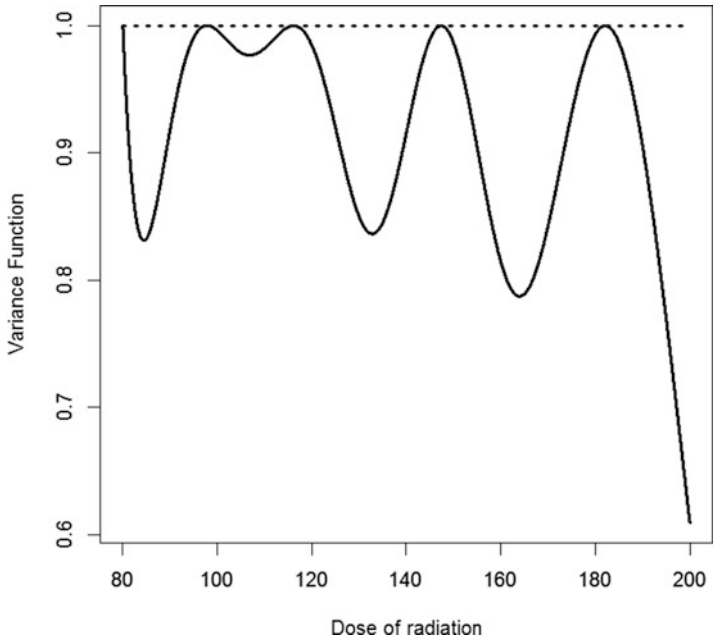

Fig. 3 Variance function for GOL model using D_ϕ-optimal design—house flies example

6 Discussion

In addition to linear and logistic regression models, researchers often find that multi-category logit models—including the adjacent category logit, baseline category logit, continuation ratio and proportional odds models considered here—are useful for modelling their data. The resulting parameter estimates then aid these researchers to make predictions or comparisons under different settings, for example using estimated odds ratios across strata. As such, practical experimental design methodologies are needed to gather the data to estimate these values and make needed predictions, and these researchers often consider using optimal designs.

But important theoretical optimal design results that are applicable only to the assumed model function are of only limited use to the practitioner. As noted, most optimal designs for models containing only p support points comprise no more than p support points, and this is certainly the case for the MCL models considered here. Underscoring this fact, Govaerts (1996) comments that this limitation prevents the use of optimal designs in most industrial settings. Therefore, the multiple-objective design strategies introduced and illustrated here for multi-category logit models—as well as in Hyun and Wong (2015) for normal nonlinear models—are paramount in applied research. Additionally, the extension of our GOL nesting strategy to incorporate geometric- and uniform-type designs gives practitioners clear suggestions as to how these designs in situations where they are desired. The

suggested designs suggested here are indeed very "near" to the optimal designs in the sense that often the resulting D-efficiency is above 90%. As such, practitioners typically find that an information loss of less than 10% is relatively small compared to the practical nature of geometric and uniform robust designs and the resulting ability to assess model goodness-of-fit.

We conclude by pointing out that beyond the MCL models considered here—viz, the PO, UPO, ACL, CRA and CRB—authors such as Agresti (2010) and others have introduced yet more models for ordinal response data, and extensions of our methods provided here to these additional cases are now under study.

Acknowledgements The first author expresses his appreciation to the J. William Fulbright Foreign Scholarship Board for ongoing grant support and to Vietnam National University (Hanoi), Kathmandu University (Nepal) and Gadjah Mada University and Islamic University of Indonesia for kind hospitality and assistance during research visits.

References

Abdelbasit, K. M., & Plackett, R. L. (1983). Experimental design for binary data. *Journal of the American Statistical Association, 78*, 90–98.

Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). Hoboken, NJ: Wiley.

Agresti, A. (2010). *Analysis of ordinal categorical data* (2nd ed.). Hoboken, NJ: Wiley.

Agresti, A. (2013). *Categorical data analysis* (3rd ed.). Hoboken, NJ: Wiley.

Atkinson, A. C. (1972). Planning experiments to detect inadequate regression models. *Biometrika, 59*, 275–293.

Atkinson, A. C., Donev, A. N., & Tobias, R. D. (2007). *Optimum experimental designs, with SAS*. New York, NY: Oxford.

Dobson, A. J., & Barnett, A. G. (2008). *An introduction to generalized linear models* (3rd ed.). Boca Raton, FL: CRC Press.

Fan, S. K., & Chaloner, K. (2001). Optimal design for a continuation-ratio model. In A. C. Atkinson, P. Hackl, & W. G. Müller (Eds.), *MODA6 – Advances in model-oriented design and analysis* (pp. 77–85). Heidelberg: Physica-Verlag.

Finney, D. J. (1978). *Statistical method in biological assay* (3rd ed.). London: Griffin.

Govaerts, B. (1996). Discussion of the papers by Atkinson, and Bates et al. *Journal of the Royal Statistical Society: Series B, 58*, 95–111.

Hyun, S. W., & Wong, W. K. (2015). Multiple-objective optimal designs for studying the dose response function and interesting dose levels. *International Journal of Biostatistics, 11*, 253–271.

Kiefer, J., & Wolfowitz, J. (1960). The equivalence of two extremum problems. *Canadian Journal of Mathematics, 12*, 363–366.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Minkin, S. (1987). Optimal designs for binary data. *Journal of the American Statistical Association, 82*, 1098–1103.

O'Brien, T. E. (2005). Designing for parameter subsets in Gaussian nonlinear regression models. *Journal of Data Science, 3*, 179–197.

O'Brien, T. E., Chooprateep, S., & Homkham, N. (2009). Efficient geometric and uniform design strategies for sigmoidal regression models. *South African Statistical Journal, 43*, 49–83.

O'Brien, T. E., & Funk, G. M. (2003). A gentle introduction to optimal design for regression models. *The American Statistician, 57*, 265–267.

Pawitan, Y. (2013). *All likelihood: Statistical modelling and inference using likelihood*. Oxford: Oxford University Press.

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

Price, C. J., Kimmel, C. A., George, J. D., & Marr, M. C. (1987). The developmental toxicity of diethylene glycol dimethyl ether in mice. *Fundamental and Applied Toxicology, 8*, 115–126.

Seber, G. A. F., & Wild, C. J. (1989). *Nonlinear regression*. New York, NY: Wiley.

Silvey, S. D. (1980). *Optimal design*. London: Chapman & Hall.

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