Author's personal copy

Computational Statistics and Data Analysis 54 (2010) 875–890

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Robust designs for generalized linear models with possible overdispersion and misspecified link functions

Adeniyi J. Adewale [a,1], Xiaojian Xu [b,*]

[a] Department of Public Health Sciences, University of Alberta Edmonton, Alberta, Canada T6G 2G3
[b] Department of Mathematics, Brock University, 500 Glenridge Avenue, St. Catharines, Ontario, Canada L2S 3A1

ARTICLE INFO

Article history:
Received 4 January 2009
Received in revised form 20 September 2009
Accepted 22 September 2009
Available online 6 October 2009

ABSTRACT

We discuss robust designs for generalized linear models with protection for possible departures from the usual model assumptions. Besides possible inaccuracy in an assumed linear predictor, both problems of overdispersion and misspecification in link function are addressed. For logistic and Poisson models, as examples, we incorporate the variance function prescribed by a superior model similar to a generalized linear mixed model to address overdispersion, and adopt a parameterized generalized family of link functions to deal with the problem of link misspecification. The design criterion is the average mean squared prediction error (AMSPE). The exact optimal design, which minimizes the AMSPE, is also presented using examples on the toxicity of ethylene oxide to grain beetles, and on Ames Salmonella Assay.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

In this article we discuss the construction of model-robust designs for generalized linear models (GLM). The literature is replete with works on model-robust designs for linear models but there is little work done on model-robust designs for GLMs. A possible reason for the dearth of work in this area is explained by the complexity of the design problem even when the assumed model is exactly correct. Contributions to the optimal regression design literature with the notion that the assumed GLM is exactly correct include those of Abdelbasit and Plackett (1983), Minkin (1987), Ford et al. (1992), Chaudhuri and Mykland (1993), Burridge and Sebastiani (1994), Atkinson and Haines (1996), and most recently, of Li and Majumdar (2008), but to name a few. The general approach to obtain a design is to optimize a certain real-valued function of the information matrix of the model parameters. One main difficulty is the dependency of design criteria on unknown model parameters. A traditional approach around this difficulty is to use best guesses for parameter values. This was termed as locally optimal designs by Chernoff (1953). Authors such as Chaloner and Larntz (1989) and Dette and Wong (1996), have used a Bayesian paradigm—assuming a prior distribution on the unknown parameters. There have also been minimax (or maximin) proposals for robustification of the uncertainties in model parameters (see Sitter (1992), King and Wong (2000) and Biedermann et al. (2004)).

In an article by Ford et al. (1989) on nonlinear designs, it was stated that when the assumed model is seriously in doubt, designs based on such a model may be grossly inappropriate. Sinha and Wiens (2002) also employ notions of robustness in the construction of sequential designs for the nonlinear model. Woods et al. (2006) provide a method for finding robust D-optimal designs with awareness of uncertainty in link function or linear predictor for a GLM with several

* Corresponding author. Tel.: +1 905 688 5550x3299; fax: +1 905 378 5713.
E-mail addresses: adeniyiadewale@hotmail.com (A.J. Adewale), xxu@brocku.ca (X. Xu).
1 Tel: +1 780 492 9951.

0167-9473/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2009.09.032

Author's personal copy

Table 1
Number of revertant colonies of salmonella (Y_i).

Observations	Y_i $x_i = 0$[a]	10	33	100	333	1000
1	15	16	16	27	33	20
2	21	18	26	41	38	27
3	29	21	33	60	41	42

[a] Dose of quinoline (μmg/plate).

explanatory variables. In addition, Wiens and Xu (2008) discuss the construction of robust Q-optimal static designs for a possibly misspecified nonlinear model. Adewale and Wiens (2009) recently consider robust Q-optimal designs for logistic models with an eye on possible misspecification in the fixed effects specified through the linear predictor. The current work goes beyond that of Wiens and Xu (2008) in that the designs obtained are exact so that they can be implemented by practitioner without further computation and approximation. We go beyond Adewale and Wiens (2009) in treating the design constructions for GLMs in general, including logistic and Poisson models as their special cases. The major advance is our treatment of possible overdispersion and inadequately assumed link function.

There are three ways a GLM can be potentially misspecified. First, the covariates included in the systematic component of the model, the linear predictor, may not reflect the influence of covariates correctly. This may be due to the use of a wrongly specified functional form of the covariates in the model or an omission of essential covariates. The second source of misspecification in a GLM is due to the variation in data that is not appropriately accounted for by the assumed model. For instance, in logistic regression, extra-binomial variation is a situation whereby the nominal variance prescribed by the binomial distribution does not correctly account for observed variability in the data. Overdispersion is the most common form of unexpected variation, and it occurs when the data exhibit variability that exceeds that prescribed by the assumed distribution. Underdispersion is the opposite, but it is not as common as overdispersion (see Dean, 1992). Third, the link function might not be accurate. For example, the use of the logit link, in logistic model, which is the canonical link for the binomial distribution when in fact the complementary log–log link or the probit link is more appropriate.

Our approach to robust design is to regard the true model as belonging to a broad class of models accommodating variation beyond that which is prescribed by the assumed model, and several link functions other than the assumed one. We compute the average mean squared prediction error (AMSPE) as a composition of variance error as well as bias due to model misspecification. Our robust designs are obtained by minimizing the AMSPE. The average is carried out over the parameter space defining the class of the unknown true model. This notion of averaging the mean squared error of predictions over a neighbourhood of the true model was introduced in Adewale and Wiens (2006) and dubbed "minave". We note that if the prior information is given, the methodology developed in this paper can also be adopted to incorporate the prior distribution of the parameters.

Example 1.1 (*Overdispersion Illustrated*). Margolin et al. (1981) compiled the data set from an Ames Salmonella reverse mutagenicity assay. The number of revertant colonies was observed at six dose levels of quinoline. Three replicate plates were used at each dose level. The data are presented in Table 1.

Fitting two Poisson models with mean responses below

$$\textit{Model}\ \mathrm{I} : \mu_1(x_i) = \exp\{\beta_0 + \beta_1 x_i\} \tag{1}$$

$$\textit{Model}\ \mathrm{II} : \mu_2(x_i) = \exp\{\beta_0 + \beta_1 x_i + \beta_2 \log(x_i + 10)\} \tag{2}$$

produce the following estimates: $\hat{\beta}_0 = 3.322$, $\hat{\beta}_1 = 0.0002$ (deviance $= 75.81$, df $= 16$); and $\hat{\beta}_0 = 2.173$, $\hat{\beta}_1 = -0.001$, $\hat{\beta}_2 = 0.320$ (deviance $= 43.72$, df $= 15$), respectively. The deviance and corresponding degrees of freedom for each of the two fitted models indicate lack of fit.

A plot of the data with the fitted models superimposed is presented in Fig. 1, which shows that the model with the term $\log(x_i + 10)$ in the linear predictor is an improved fit over the first model. However, the data presented variability that exceeds that prescribed by the Poisson distribution. There are many proposals in the literature for modelling overdispersed data. In order to accommodate overdispersion in Margolin's data, Lawless (1987) fitted *Model* III: the negative binomial model $Y_i \sim NB(\mu_i, k)$ with mean response (2) and variance $\mu(x_i) + k^{-1}\mu^2(x_i)$. The estimates obtained using the weighted least squares — method of moments are $\hat{\beta}_0 = 2.203$, $\hat{\beta}_1 = -0.001$, $\hat{\beta}_2 = 0.311$. A likelihood ratio test rejects the hypothesis $H : k^{-1} = 0$ and thus provides evidence of overdispersion in the data.

Example 1.2 (*Link misspecification Illustrated*). Table 2 presents Bliss's (1935) data on the numbers of dead beetles after five hours exposure to gaseous carbon disulphide at various concentrations. We fit the binomial model corresponding to a range of values of λ for a generalized link function (defined by (10) in Section 2.2 of this paper). In particular, we fit the model for a grid of λ values in $[0, 2]$. A plot of deviance versus λ values is presented in Fig. 2. Using the deviance as the model fit criterion, $\lambda = 0.1861$ corresponds to the model with the minimum deviance. Fig. 3 shows that the model with the link parameter $\lambda = 0.1861$ provides an improved fit to the data than the logistic model with $\lambda = 1$. The parameter estimates

Author's personal copy

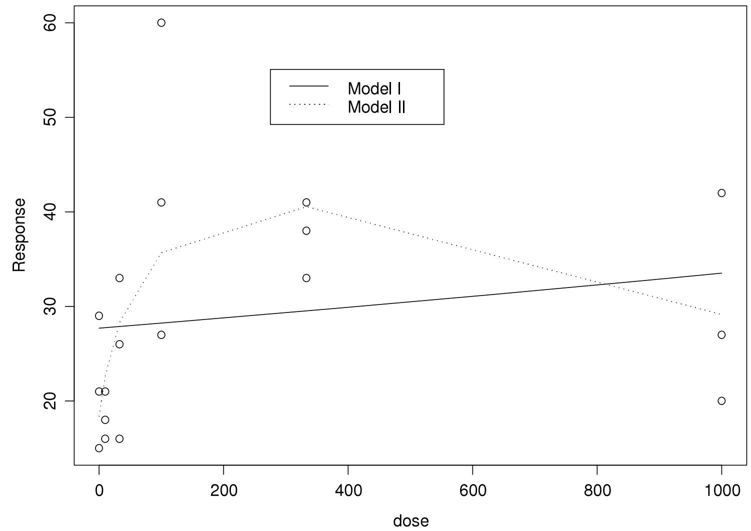

Fig. 1. Number of revertant colonies of salmonella against dose level with Model *I* and *II* superimposed.

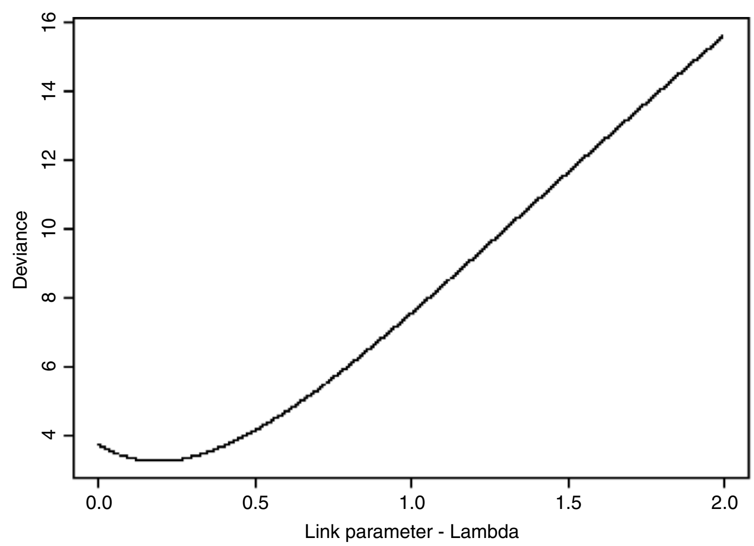

Fig. 2. Deviance versus λ for the beetle mortality data.

Table 2
Bliss's Beetle mortality data.

Dose, x_i (mgl^{-1})	49.06	52.99	56.91	60.84	64.76	68.69	72.61	76.54
Number of beetles, n_i	59	60	62	56	63	59	62	60
Number killed, $n_i y_i$	6	13	18	28	52	53	61	60

for the logistic model are $\hat{\beta}_0 = -14.808$ and $\hat{\beta}_1 = 0.249$ and the corresponding estimates for the model with $\lambda = 0.1861$ are $\hat{\beta}_0 = -10.782$ and $\hat{\beta}_1 = 0.174$.

In an investigation of designs for nonlinear models, Sinha and Wiens (2002) asserted that "Although the theoretical response functions are very similar in shape, and possibly indistinguishable if noisy data must be relied upon, the appropriate designs can be quite dissimilar". Similarly, since the link function determines the response function in a GLM, the misspecification of the link function could have consequences on the resulting designs.

2. Statistical models

We consider the case of an experimenter with a finite set $\mathcal{S} = \{\mathbf{x}_i\}_{i=1}^{N}$ of possible design points whose interest is to choose n, not necessarily distinct, points at which to observe response Y. The virtue of choosing such a finite design space is stated

Author's personal copy

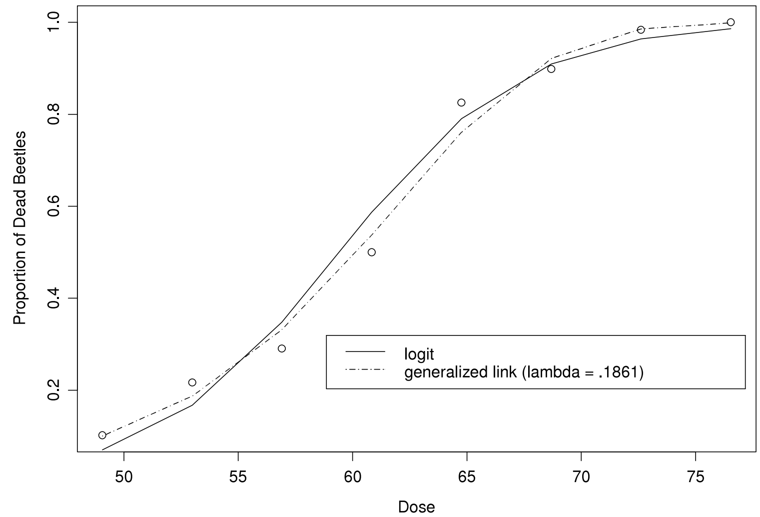

Fig. 3. Proportion of dead beetles versus dose with logit link and $\lambda = .1861$-link superimposed.

in Fang and Wiens (2000). The experimenter makes $n_i \geq 0$ observations at $\mathbf{x}_i$ such that $\sum_{i=1}^{N} n_i = n$. The design problem is to choose $n_1, \ldots, n_N$ in an optimal manner and with an eye on possible model misspecification.

In logistic regression, the experimenter intends to fit the model with mean response

$$\mu^B(\eta) = \frac{\exp(\eta)}{1+\exp(\eta)}; \tag{3}$$

and for a Poisson log linear model, one considers the mean count being

$$\mu^P(\eta) = \exp(\eta).$$

Both are nonlinear functions of the linear predictor $\eta = \mathbf{z}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta}$ with $\mathbf{z}(\mathbf{x})$ as a vector of predictors.

In view of the possible model departures, as exemplified in Examples 1.1 and 1.2, the experimenter seeks protection against possible misspecifications in the assumed model form. As highlighted in the introduction, the following are the three kinds of departures from an assumed GLM that the experimenter is worried about:

M1. The true linear predictor is $\eta = \mathbf{z}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta} + f(\mathbf{x})$ where the function $f(\mathbf{x})$ is some contamination function accounting for additional effects of covariates.
M2. The mean–variance relationship in the assumed GLM may not exactly describe the variability in data obtained.
M3. The true link function is different from the assumed link function.

The subject of designing for logistic models with possible M1 departure has been treated in Adewale and Wiens (2009). They assumed that the contamination function $f(\mathbf{x})$ belongs to a contamination neighbourhood:

$$\mathcal{F} = \left\{ \frac{1}{N}\sum_{i=1}^{N} \mathbf{z}(\mathbf{x}_i) f(\mathbf{x}_i) = 0,\ \frac{1}{N}\sum_{i=1}^{N} f^2(\mathbf{x}_i) \leq \tau^2,\ \text{with } \tau^2 = O\left(n^{-1}\right) \right\}. \tag{4}$$

It is noted that the first condition in $\mathcal{F}$ ensures identifiability of the linear predictor and the second condition ensures that the bias engendered by the misspecified linear predictor remains bounded. The average (over the design space $\mathcal{S}$) mean squared error in predicting the mean response, $\mu(\eta)$ was adopted as the loss function. This loss function depends on the unknown contamination function $f(\mathbf{x})$. In order to eliminate this dependency of the loss function on the unknown contamination function $f(\mathbf{x})$, they used the average of the loss function over the contamination neighbourhood $\mathcal{F}$ as the design criterion. In this paper, this criterion is also adopted.

2.1. Overdispersion model

Overdispersion is a very common phenomenon in data involving proportions but most of the attention it has received in the literature has been in the context of data analysis. Modelling overdispersion has been considered by, among others, Pierce and Sands (1975), Crowder (1978) and Williams (1982). However, this subject has received very scant attention in the regression design literature. In order to construct robust designs that give protection for overdispersion we adopt a mixed model which accommodates overdispersion by incorporating an additive random component to the linear predictor.

Author's personal copy

In addition, this mixed model is also required to have the fitted model as a special case. Formally, the true but unknown model belongs to a class of alternative models defined as follows:

$$E\left(Y_i|v_i\right) = \mu\left(\eta_i\right), \quad \eta_i = \mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta} + \phi v_i,$$
$$\text{with } E\left(v_i\right) = 0,\ var\left(v_i\right) = 1, \quad \text{and} \quad var\left(v_i^2\right) < \infty. \tag{5}$$

2.1.1. Overdispersion in logistic model

For binary data, in logistic regression particularly, we have

$$Y_i^B|v_i \sim \frac{1}{n_i} binomial\left(n_i, \mu^B\left(\eta_i\right)\right).$$

So the true model has $E\left(Y_i^B|v_i\right) = \mu^B\left(\eta_i\right)$, $var\left(Y_i^B|v_i\right) = \frac{E\left(Y_i^B|v_i\right)\left(1-E\left(Y_i^B|v_i\right)\right)}{n_i}$. We have $\left.\frac{\mathrm{d}\left[\mu^B(\eta_i)\right]}{\mathrm{d}\eta_i}\right|_{\phi=0} = \mu^B\left(\mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}\right) \times \left(1-\mu^B\left(\mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}\right)\right)$. Let $w\left(\mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}\right) = \mu^B\left(\mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}\right)\left(1-\mu^B\left(\mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}\right)\right)$. Then we have

$$\left.\frac{\mathrm{d}^2\left[\mu^B\left(\eta_i\right)\right]}{\mathrm{d}\eta_i^2}\right|_{\phi=0} = w\left(\mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}\right)\left(1-2\mu^B\left(\mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}\right)\right).$$

We take the following Taylor's expansion at $\phi = 0$:

$$\begin{aligned}
E\left(Y_i^B|v_i\right) &= \frac{\exp\left(\mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}\right)}{1+\exp\left(\mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}\right)} + \left.\frac{\mathrm{d}\left[\mu^B\left(\eta_i\right)\right]}{\mathrm{d}\eta_i}\right|_{\phi=0}\phi v_i \\
&\quad + \frac{1}{2}\left.\frac{\mathrm{d}^2\left[\mu^B\left(\eta_i\right)\right]}{\mathrm{d}\eta_i^2}\right|_{\phi=0}\phi^2 v_i^2 + o\left(\phi^2\right) \\
&= \mu^B\left(\mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}\right) + w\left(\mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}\right)\phi v_i \\
&\quad + \frac{1}{2} w\left(\mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}\right)\left(1-2\mu^B\left(\mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}\right)\right)\phi^2 v_i^2 + o\left(\phi^2\right).
\end{aligned}$$

Then, we have

$$\begin{aligned}
var\left(Y_i^B|v_i\right) &= \frac{1}{n_i}\left\{\begin{array}{l} w\left(\mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}\right) + \left[1-2\mu^B\left(\mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}\right)\right] w\left(\mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}\right)\phi v_i \\ + \left\{\frac{1}{2} w\left(\mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}\right)\left[1-2\mu^B\left(\mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}\right)\right]^2 - w^2\left(\mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}\right)\right\}\phi^2 v_i^2 \end{array}\right\} \\
&\quad + o\left(\phi^2\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
E\left\{var\left(Y_i^B|v_i\right)\right\} &= \frac{1}{n_i} w\left(\mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}\right)\left[1 + \left\{\begin{array}{l}\frac{1}{2}\left[1-2\mu^B\left(\mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}\right)\right]^2 \\ -w\left(\mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}\right)\end{array}\right\}\phi^2\right] \\
&\quad + o\left(\phi^2\right), \quad \text{and} \quad var\left\{E\left(Y_i^B|v_i\right)\right\} = w^2\left(\mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}\right)\phi^2 + o\left(\phi^2\right).
\end{aligned}$$

Using the identities

$$E\left(Y_i\right) = E\left\{E\left(Y_i|v_i\right)\right\}, \quad \text{and} \quad var\left(Y_i\right) = var\left\{E\left(Y_i|v_i\right)\right\} + E\left\{var\left(Y_i|v_i\right)\right\},$$

the true mean response and true variance are given by

$$E\left(Y_i^B\right) = \mu^B\left(\mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}\right) + O\left(\phi^2\right), \tag{6}$$

$$\begin{aligned}
var\left(Y_i^B\right) &= \frac{1}{n_i} w\left(\mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}\right)\left\{1 + \left(\begin{array}{l}\frac{1}{2}\left[1-2\mu^B\left(\mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}\right)\right]^2 \\ +\left(n_i-1\right) w\left(\mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}\right)\end{array}\right)\phi^2\right\} \\
&\quad + o\left(\phi^2\right).
\end{aligned} \tag{7}$$

Author's personal copy

2.1.2. Overdispersion in Poisson model

For Poisson data, Minkin (1993) consider overdispersion when constructing design. He accommodates overdispersion using a multiplicative random effect. That is to take the true model as $Y_i^P|t \sim \text{Poisson}(t\mu_i^P)$ where the random variable t has mean 1 and variance ϕ^2. Thus the unconditional distribution of Y_i^P has the mean response μ_i^P, and the variance $\mu_i^P + \phi^2\left(\mu_i^P\right)^2$. In this paper we rather accommodate overdispersion by an additive random contamination to the linear predictor such that the true model has the conditional distribution $Y_i^P|v_i \sim \text{Poisson}(\mu^P(\eta_i) = e^{\eta_i})$ in the form of (5). Thus, the true model has conditional mean and variance as follows:

$$\begin{aligned} E\left(Y_i^P|v_i\right) &= \mu_i^P = e^{\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)\boldsymbol{\beta}+\phi v_i} \\ &= e^{\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)\boldsymbol{\beta}} + e^{\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)\boldsymbol{\beta}}\phi v_i + \frac{1}{2}e^{\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)\boldsymbol{\beta}}\phi^2 v_i^2 + o\left(\phi^2\right), \\ var\left(Y_i^P|v_i\right) &= e^{\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)\boldsymbol{\beta}} + e^{\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)\boldsymbol{\beta}}\phi v_i + \frac{1}{2}e^{\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)\boldsymbol{\beta}}\phi^2 v_i^2 + o\left(\phi^2\right). \end{aligned}$$

Therefore,

$$E\left\{var\left(Y_i^P|v_i\right)\right\} = e^{\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)\boldsymbol{\beta}} + \frac{1}{2}e^{\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)\boldsymbol{\beta}}\phi^2 + o\left(\phi^2\right), \quad \text{and} \quad var\left\{E\left(Y_i^P|v_i\right)\right\} = e^{2\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)\boldsymbol{\beta}}\phi^2 + o\left(\phi^2\right).$$

Hence, the true mean response and true variance for this model are given by

$$E\left(Y_i^P\right) = e^{\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)\boldsymbol{\beta}} + O\left(\phi^2\right), \tag{8}$$

$$var\left(Y_i^P\right) = e^{\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)\boldsymbol{\beta}} + e^{\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)\boldsymbol{\beta}}\left(\frac{1}{2} + e^{\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)\boldsymbol{\beta}}\right)\phi^2 + o\left(\phi^2\right). \tag{9}$$

In Model (5), the true linear predictor has two parts — the fixed part representing the effects of covariates and a random part capturing overdispersion. We note that the overdispersion problems mainly contribute to the variance part of the response, and the resulting mean responses above approximately remain the same. However, the following subsection will show that the misspecified link function problems not only contribute to the variance part of the response but also cause the nonignorable bias in the mean response.

2.2. True link model

2.2.1. Link function misspecification in logistic model

Ponce de Leon and Atkinson (1993) investigated designing optimal experiments for the choice of link function for a binary data model. They assumed a framework where interest is in the estimation of the link function as well as the model coefficients when the link function belongs to the same parameterized generalized link family of interest in this article. Thus their approach fits into the context of classical approach to regression design in which the experimenter takes the assumed model to be exact. The difference is that the link function is rather added as an extra-parameter to be estimated from the data. Biedermann et al. (2004) introduced a robust approach in which the experimenter is considering a finite set of plausible link functions with the uncertainties in the suitability of each of them quantified with known probabilities. The probabilities reflect the preferences the experimenter attaches to each link function. The resulting robust criterion is a weighted average of the respective criterion corresponding to each link function. We take a different approach here. Given that the experimenter intends to fit the logistic model, binomial model with the logit link, we propose designs that protect against the possibility of the imprecision in the logit link specified.

The framework for the robust design in this paper is that the true but unknown link function as well as the fitted logit link belong to the generalized family of link functions

$$g\left(\mu^B, \lambda\right) = \log\left[\left\{\left(\frac{1}{1-\mu^B}\right)^{\lambda} - 1\right\} \Big/ \lambda\right], \quad (\lambda \geq 0), \tag{10}$$

parameterized by λ. The true link function corresponds to an unknown value of the link parameter that might be different from that which corresponds to the logistic model. This generalized family of links encompasses the logit and the complementary log–log links as special cases. The logit link corresponds to $\lambda = 1$ while the complementary log–log corresponds to $\lim_{\lambda\to 0}$.

In generalized linear modelling the link function connects the systematic component (the linear predictor) of the model to the mean response via

$$\eta = g\left(\mu^B, \lambda\right),$$

where η is the linear predictor representing the effects of covariates in the model on a linear scale. The reasonable choice of link functions is suggested by the distribution of the response. For example, $g\left(\mu^B, \lambda\right)$ is suggested as the logit function when the response takes the logistic distribution function, and as the complementary log–log link when it takes the extreme

Author's personal copy

value distribution. We considered the following Taylor's expansion of (10) about the parameter value, $\lambda = 1$ corresponding to the logit link that the experimenter contemplates fitting. That is,

$$\eta = \log\left(\frac{\mu^B}{1-\mu^B}\right) + h\left(\mathbf{x};\lambda\right)$$

with μ^B given by (3) and

$$h\left(\mathbf{x};\lambda\right) = \left.\frac{\partial g}{\partial \lambda}\right|_{\lambda=1} (\lambda - 1) + o\left(\lambda - 1\right). \tag{11}$$

Thus the link misspecification problem may be cast as a linear predictor misspecification problem. The true mean response is given by (3) with the linear predictor given by

$$\eta_i = \mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta} + f\left(\mathbf{x}_i, \lambda\right),$$

where the contamination function is given by

$$f\left(\mathbf{x}, \lambda\right) = -h\left(\mathbf{x}, \lambda\right). \tag{12}$$

2.2.2. Link function misspecification in Poisson model

In designing for a Poisson model with possibly misspecified link function we can employ the same framework as it for logistic regression where the true link function and the fitted canonical link belong to a generalized family of link functions. Here consider the parameterized family of link functions defined by

$$g\left(\mu^P, \lambda\right) = \begin{cases} \left(\mu^P\right)^\lambda, & \lambda \neq 0 \\ \log \mu^P, & \lambda = 0, \end{cases}$$

where λ is the link parameter. The log link corresponds to $\lambda = 0$. The strategy is to linearize this generalized link function by expanding it in a Taylor series about a fixed value $\lambda = 0$ and taking only the linear term. We then approximate the true linear predictor by

$$\eta_i = \mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta} + f\left(\mathbf{x}_i, \lambda\right),$$

with the contamination function $f\left(\mathbf{x}, \lambda\right)$ given by

$$f\left(\mathbf{x}, \lambda\right) = -\left.\frac{\partial g\left(\mathbf{x}, \lambda\right)}{\partial \lambda}\right|_{\lambda=0} \lambda = -\lambda \log \mu^P. \tag{13}$$

As discussed above, the link function misspecification can be treated as a linear predictor misspecification problem. This agrees with the work of Pregibon (1980) on goodness of link tests for generalized linear models where a misspecified link function is considered as an additive contamination in the linear predictor. On this, Pregibon (1980) commented that "The fact that the difference in link functions appears on the right side of the link defining equation should not be disturbing-indeed, this corresponds to the fact that the wrong link function is a systematic misspecification of the model". Therefore, a M3 problem can be treated as a M1 problem.

The rest of this paper is arranged as follows: in Section 3.1, we provide our general results on the asymptotic bias and asymptotic covariance matrix of the maximum likelihood estimate (MLE) of the model parameter vector, as well as the asymptotic approximation of AMSPE for GLM models with possible M1, M2, and/or M3 departures. The asymptotics for logistic and Poisson models are discussed in Sections 3.2 and 3.3. A modified simulated annealing algorithm used for obtaining the resulting robust designs in this paper is discussed in Section 3.4. We present several examples and real-life applications in Section 4 and conclude with a few remarks in Section 5. Derivations of all theorems in this article are provided in an Appendix.

3. Loss function and algorithm

We consider a GLM model:

$$\begin{aligned} &E\left(Y_i|v_i\right) = \mu\left(\eta_{T,i}\right), \\ &\eta_{T,i} = \mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta} + f\left(\mathbf{x}_i, \lambda\right) + \phi v_i, \quad \text{with} \\ &E\left(v_i\right) = 0, var\left(v_i\right) = 1, \quad \text{and} \quad var\left(v_i^2\right) < \infty. \end{aligned} \tag{14}$$

This model incorporates with all possible M1, M2, and M3 types of departures from fitted model:

$$E\left(Y_i\right) = \mu\left(\eta_i\right), \quad \eta_i = \mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}. \tag{15}$$

Author's personal copy

Since the typical focus of the design for a generalized linear model is prediction, we use the normalized average mean squared prediction error (AMSPE) $\mathcal{L}$ of the response prediction $\mu\left(\hat{\eta}_i\right)$, with $\hat{\eta}_i = \mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\hat{\boldsymbol{\beta}}$, as the design criterion,

$$\mathcal{L} \triangleq \frac{n}{N}\sum_{i=1}^{N} E\left[\left\{\mu\left(\hat{\eta}_i\right) - \mu\left(\mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta} + f\left(\mathbf{x}_i, \lambda\right) + \phi v_i\right)\right\}^2\right]. \tag{16}$$

The response Y follows a distribution from the exponential family. The mean of this response depends on the unknown parameters and the vector of explanatory variable, $\mathbf{z}\left(\mathbf{x}\right)$, through the link function $\eta = \mathbf{z}^{\mathrm{T}}\left(\mathbf{x}\right)\boldsymbol{\beta}$. The variance of this response, $Var\left(Y\right)$, depends only on this link function. Formally, the true but unknown model belongs to a class of alternative models defined as follows:

$$\begin{aligned} &Y_i|\mathbf{x}_i \sim \text{ Exponential Family, having density} \\ &\mathrm{d}(y_i) = \left\{\left[y_i\theta_i - b\left(\theta_i\right)\right]/a\left(\varphi\right) + c\left(y_i, \varphi\right)\right\}, \quad \text{with} && (17) \\ &E\left(Y_i\right) = \mu\left(\eta_{T,i}\right), \quad \eta_{T,i} = \mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta} + f\left(\mathbf{x}_i\right) + \phi v_i. && (18) \end{aligned}$$

But the fitted model is with (15).

The log-likelihood l, according to the possibly misspecified model is

$$\begin{aligned} l\left(\boldsymbol{\beta}\right) &= \sum_{i=1}^{N} l_i(\theta_i, \varphi; Y_i) = \sum_{i=1}^{N} \log \mathrm{d}\left(Y_i; \theta_i, \varphi\right) \\ &= \sum_{i=1}^{N} \left\{\left[Y_i\theta_i - b\left(\theta_i\right)\right]/a\left(\varphi\right) + c\left(Y_i, \varphi\right)\right\}. \end{aligned}$$

We employed the chain rule to obtain the first derivative of the log-likelihood l,

$$\frac{\partial l\left(\boldsymbol{\beta}\right)}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N} \frac{\partial l_i}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N} \frac{\partial l_i}{\partial \theta_i}\frac{\partial \theta_i}{\mathrm{d}\mu_i}\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \boldsymbol{\beta}},$$

where $\frac{\partial l_i}{\partial \theta_i} = \left[Y_i - b'\left(\theta_i\right)\right]/a\left(\varphi\right)$. Using $\mu_i = b'\left(\theta_i\right)$ and $var\left(Y_i\right) = a\left(\varphi\right)b''\left(\theta_i\right)$ we have

$$\begin{aligned} &\frac{\partial l_i}{\partial \theta_i} = \left[Y_i - \mu_i\right]/a\left(\varphi\right), \\ &\frac{\partial \mu_i}{\mathrm{d}\theta_i} = b''\left(\theta_i\right) = var\left(Y_i\right)/a\left(\varphi\right), \quad \text{and} \\ &\frac{\partial \eta_i}{\partial \boldsymbol{\beta}} = \mathbf{z}\left(\mathbf{x}_i\right) \quad \text{since } \eta_i = \mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}. \end{aligned}$$

Thus

$$\frac{\partial l\left(\boldsymbol{\beta}\right)}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N} \frac{\left(Y_i - \mu_i\right)}{var\left(Y_i\right)}\frac{\partial \mu_i}{\partial \eta_i}\mathbf{z}\left(\mathbf{x}_i\right),$$

and

$$\frac{\partial^2 l\left(\boldsymbol{\beta}\right)}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^{\mathrm{T}}} = \sum_{i=1}^{N}\left[\frac{\partial \mu_i/\partial \eta_i}{var\left(Y_i\right)}\mathbf{z}\left(\mathbf{x}_i\right)\frac{\partial\left\{Y_i - \mu_i\right\}}{\partial \boldsymbol{\beta}^{\mathrm{T}}} + \left(Y_i - \mu_i\right)\mathbf{z}\left(\mathbf{x}_i\right)\frac{\partial}{\partial \boldsymbol{\beta}^{\mathrm{T}}}\left\{\frac{\partial \mu_i/\partial \eta_i}{var\left(Y_i\right)}\right\}\right].$$

We have

$$\frac{\partial\left\{Y_i - \mu_i\right\}}{\partial \boldsymbol{\beta}^{\mathrm{T}}} = -\frac{\partial \mu_i}{\partial \eta_i}\mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right),$$

and for a canonical model (which the fitted model is)

$$\theta_i = \eta_i, \quad \text{and} \quad \frac{\partial}{\partial \boldsymbol{\beta}^{\mathrm{T}}}\left\{\frac{\partial \mu_i/\partial \eta_i}{var\left(Y_i\right)}\right\} = \mathbf{0}.$$

Thus,

$$\frac{\partial l\left(\boldsymbol{\beta}\right)}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N} \frac{\left(Y_i - \mu_i\right)}{a\left(\varphi\right)}\mathbf{z}\left(\mathbf{x}_i\right),$$

Author's personal copy

and -1 times the second derivative of second derivative of the possibly misspecified canonical model is

$$-\frac{\partial^2 l(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathrm{T}}} = \sum_{i=1}^{N}\frac{(\partial\mu_i/\partial\eta_i)^2}{var(Y_i)}\mathbf{z}(\mathbf{x}_i)\,\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)$$
$$= \sum_{i=1}^{N}\frac{\partial\mu_i/\partial\eta_i}{a(\varphi)}\mathbf{z}(\mathbf{x}_i)\,\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i).$$

3.1. General results

The preceding arguments cumulate into Theorem 1 which gives the expressions for the asymptotic bias and asymptotic covariance of the maximum likelihood estimates of the model parameters with awareness of any possible M1, M2, and/or M3 departure(s).

Theorem 1. *Let $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}_T$ be the $N\times 1$ vectors with elements $\mu_i = \mu\left(\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)\,\boldsymbol{\beta}_0\right)$, and $\mu_{T,i} = \mu\left(\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)\,\boldsymbol{\beta} + f(\mathbf{x}_i,\lambda) + \phi v_i\right)$ respectively, where μ_i is the fitted mean response and $\mu_{T,i}$ is the true mean response and let $\mathbf{Z}$ be the $N\times p$ matrix with rows $\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)$. Let $var_T(Y_i)$ be the true variance. Define*

$$w_i = \frac{\partial\mu_i/\partial\eta_i}{a(\varphi)}, \quad \text{and} \quad w_{T,i} = \frac{var_T(Y_i)}{a^2(\varphi)}.$$

Let $\mathbf{P}$, $\mathbf{W}$, and $\mathbf{W}_T$ be the $N\times N$ diagonal matrices with diagonal elements n_i/n, w_i, and $w_{T,i}$ respectively. Finally, define

$$\begin{aligned}\mathbf{b} &= \frac{\mathbf{Z}^{\mathrm{T}}\mathbf{P}\left(\boldsymbol{\gamma}_T - \boldsymbol{\gamma}\right)}{a(\varphi)},\\ \mathbf{H}_n &= \mathbf{Z}^{\mathrm{T}}\mathbf{PWZ},\\ \tilde{\mathbf{H}}_n &= \mathbf{Z}^{\mathrm{T}}\mathbf{PW}_T\mathbf{Z}.\end{aligned} \tag{19}$$

For any GLM model with a canonical link function, the asymptotic bias and asymptotic covariance matrix of the maximum likelihood estimate $\hat{\boldsymbol{\beta}}$ of the model parameter vector $\boldsymbol{\beta}_0$ from the misspecified model are

$$bias\left(\hat{\boldsymbol{\beta}}\right) = E\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) = \mathbf{H}_n^{-1}\mathbf{b} + o\left(n^{-1/2}\right), \tag{20}$$

$$cov\left(\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)\right) = \mathbf{H}_n^{-1}\tilde{\mathbf{H}}_n\mathbf{H}_n^{-1} + o(1), \tag{21}$$

respectively.

This is a generalization, to generalized linear models and to all three possible departures, of Theorem 1 of Adewale and Wiens (2009), which gives the asymptotic bias and asymptotic covariance matrix of the maximum likelihood estimate $\hat{\beta}$ for logistic model with possible M1 departure.

Theorem 2. *For any GLM model (17) with a canonical link function, the AMSPE has the asymptotic approximation $\mathcal{L} = \mathcal{L}(\mathbf{P};\lambda,\phi,\boldsymbol{\beta}) + o(1)$, where*

$$\mathcal{L}(\mathbf{P};\lambda,\phi,\boldsymbol{\beta}) = \frac{1}{N}\left\{tr\left[\mathbf{WZH}_n^{-1}\tilde{\mathbf{H}}_n\mathbf{H}_n^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{W}\right] + n\left\|\mathbf{W}\left(\mathbf{ZH}_n^{-1}\mathbf{b} - \mathbf{f}\right)\right\|^2\right\} \tag{22}$$

for $\mathbf{f} = (f(\mathbf{x}_1,\lambda),\ldots,f(\mathbf{x}_N,\lambda))^{\mathrm{T}}$.

An approximation $\mathcal{L}$ to AMSPE is given in Theorem 2 and the construction of a design is to minimize quantity (22). The robust designs for a GLM with a possible M1 departure can be constructed by the same approach employed in Adewale and Wiens (2009) using the result of Theorem 2 above. Therefore, in the rest of this paper we will focus on finding robust designs dealing with M2 and M3 types of departures.

3.2. Asymptotics for logistic models with possible M2 and/or M3 departure(s)

We consider a logistic model: (14) with $\mu(\cdot) = \frac{\exp(\cdot)}{1+\exp(\cdot)}$ which incorporates with any possible M2, M3 types of departures from fitted model (15). We keep the notations in (19) the same, then the elements involved in (19) have become to the

Author's personal copy

following:

$$\mu_i = \frac{\exp\left[\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)\,\boldsymbol{\beta}_0\right]}{1+\exp\left[\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)\,\boldsymbol{\beta}_0\right]}, \tag{23}$$

$$w_i = \mu_i\,(1-\mu_i) = \frac{1}{4}\mathrm{sech}^2\left(\frac{\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)\,\boldsymbol{\beta}_0}{2}\right), \tag{24}$$

$$\mu_{T,i} = \frac{\exp\left(\eta_{T,i}\right)}{1+\exp\left(\eta_{T,i}\right)} \quad \text{with } \eta_{T,i} = \mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)\,\boldsymbol{\beta}_0 + f(\mathbf{x}_i,\lambda), \quad \text{and} \tag{25}$$

$$w_{T,i} = var\left(Y_i^B\right).$$

According to (6) and (7), we have

$$E\left(Y_i^B\right) \approx \mu_{T,i}, \quad \text{and}$$

$$var\left(Y_i^B\right) \approx \frac{1}{n_i}\mu_{T,i}\left(1-\mu_{T,i}\right)\left\{1+\begin{pmatrix}\frac{1}{2}\left[1-2\mu_{T,i}\right]^2\\ +(n_i-1)\,\mu_{T,i}\left(1-\mu_{T,i}\right)\end{pmatrix}\phi^2\right\}. \tag{26}$$

We define $\tilde{\mathbf{H}}_n^B = \mathbf{Z}^{\mathrm{T}}\mathbf{P}\mathbf{W}_T^B\mathbf{Z}$, where $\mathbf{W}_T^B$ be the $N \times N$ diagonal matrices with diagonal elements $var\left(Y_i^B\right)$ in (26). For fitting a logistic version of Model (15), with consideration of possible M2, M3 departures, the asymptotic bias and asymptotic covariance matrix of the maximum likelihood estimator $\hat{\beta}$ of the logistic model parameter vector β_0 from the misspecified model can be obtained by Theorem 1 with $\tilde{\mathbf{H}}_n = \tilde{\mathbf{H}}_n^B$, where μ_i, w_i, $\mu_{T,i}$, and $w_{T,i}$ are defined as (23), (25), (25) and (26) respectively. According to Theorem 2, the AMSPE has an asymptotic approximation (22) with $\tilde{\mathbf{H}}_n = \tilde{\mathbf{H}}_n^B$.

3.3. Asymptotics for Poisson models with possible M2 and/or M3 departure(s)

We consider a Poisson model: (14) with $\mu(\cdot) = \exp(\cdot)$ which incorporates with any possible M2-3 types of departures from fitted model (15). Again keeping the notations in (19) the same, the elements involved in (19) have become to the following:

$$\mu_i = \exp\left(\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)\,\boldsymbol{\beta}_0\right), \tag{27}$$

$$w_i = \frac{\mathrm{d}\mu_i}{\mathrm{d}\eta_i} = \exp\left(\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)\,\boldsymbol{\beta}_0\right), \tag{28}$$

$$\mu_{T,i} = \exp\left(\eta_{T,i}\right), \quad \text{with } \eta_{T,i} = \mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)\,\boldsymbol{\beta}_0 + f(\mathbf{x}_i,\lambda), \quad \text{and} \tag{29}$$

$$w_{T,i} = var_T(Y_i^P).$$

According to (8) and (9), we have

$$E\left(Y_i^P\right) \approx \mu_{T,i}, \quad \text{and}$$

$$var_T(Y_i^P) \approx \mu_{T,i} + \mu_{T,i}\left(\frac{1}{2}+\mu_{T,i}\right)\phi^2. \tag{30}$$

We define $\tilde{\mathbf{H}}_n^P = \mathbf{Z}^{\mathrm{T}}\mathbf{P}\mathbf{W}_T^P\mathbf{Z}$, where $\mathbf{W}_T^P$ be the $N \times N$ diagonal matrices with diagonal elements $var\left(Y_i^P\right)$ in (30). For fitting a Poisson version of Model (15), with awareness of possible M2 and M3 departures, the asymptotic bias and asymptotic covariance matrix of the maximum likelihood estimator $\hat{\beta}$ of the logistic model parameter vector β_0 from the misspecified model can be obtained by Theorem 1 with $\tilde{\mathbf{H}}_n = \tilde{\mathbf{H}}_n^P$, where μ_i, w_i, $\mu_{T,i}$, and $w_{T,i}$ are defined as (27)–(30) respectively. According to Theorem 2, the AMSPE has the asymptotic approximation (22) with $\tilde{\mathbf{H}}_n = \tilde{\mathbf{H}}_n^P$.

3.4. Algorithm: Modified simulated annealing

In this subsection, we discuss the computation methods used in the exact design constructions based on the results above.

The loss function (22) depends on the unknown model parameter vector $\boldsymbol{\beta}_0$, the overdispersion parameter ϕ, and the link parameter λ. Our approach to dealing with this parameter dependency requires parameter spaces Θ, Φ, and Λ of plausible values of $\boldsymbol{\beta}_0$, ϕ, and λ respectively. Letting $\Omega = [\Theta \times \Phi \times \Lambda]$, the design criterion is then the average of the loss function over the given joint parameter space Ω. The averaging was carried out over a uniformly scattered set of points obtained on

Author's personal copy

Ω using number-theoretic method or quasi-Monte Carlo method (Fang and Wang, 1994). Given the parameter space Ω, we sort designs that minimize the average (over the space Ω) of the loss function:

$$\mathcal{L}_{ave}(\mathbf{P}) = aver_{\Omega}\mathcal{L}(\mathbf{P}; \lambda, \phi, \boldsymbol{\beta}), \tag{31}$$

through the matrix $\mathbf{P} = diag\,(n_1/n, \ldots, n_N/n)$. The minimization of (31) with respect to $\mathbf{P}$ is undertaken using a modification of the simulated annealing algorithm. It suffices to note here that the averaging in (31) is implemented as an integral part of the simulated annealing process — for every new design generated the value of loss $\mathcal{L}(\mathbf{P}; \lambda, \phi, \boldsymbol{\beta})$ is calculated for every point $(\lambda, \phi, \boldsymbol{\beta})$ in this scattered set in Ω. The value of $\mathcal{L}_{ave}(\mathbf{P})$ is the average of $\mathcal{L}(\mathbf{P}; \lambda, \phi, \boldsymbol{\beta})$ over the scattered points in the parameter space Ω. The resulting designs are termed minave designs.

We consider models with p regressors $(z_1(x), \ldots, z_p(x))^{\mathrm{T}}$ where $x \in [a, b]$. The design space is the set $\mathcal{S} = \left\{x_i = a + \frac{(b-a)(i-1)}{N-1}\right\}_{i=1}^{N}$ of equally spaced points in $[a, b]$. Given the desired number of observations (n) to be taken and the number of points in the design space (N), we seek designs that minimize the loss function (31). Our approach gives the experimenter the flexibility to request a symmetric design about the point $(a + b)/2$ or an asymmetric design.

Simulated annealing is a direct search optimization algorithm which has been quite successful at finding the global extremum of a function, possibly non-smooth, that has many local extrema. The simulated annealing algorithm here seeks to assign integers $n_i \geq 0$ to each of the design points x_i in such a way that $\mathcal{L}_{ave}$ is a minimum. The three steps of this algorithm are detailed in Fang and Wiens (2000). A modified simulated annealing algorithm is employed to search for optimal designs in this paper. These modifications are:

If $n < N$, then the initial design is chosen to be the design assigning one observation to each of n randomly chosen points. If $n \geq N$, we randomly distribute n observations to all N points. If our interest is in symmetric designs we randomly assign one observation to each of $\lfloor \frac{n}{2} \rfloor$ randomly chosen points in $[a, (a + b)/2)$ when $n < N$, and randomly distribute $\lfloor \frac{n}{2} \rfloor$ observations to the $\lfloor \frac{N}{2} \rfloor$ points in $[a, (a + b)/2)$ when $n \geq N$. If n is odd then N has to be odd for symmetry, in this case we assign the extra observation to the point $\left\{\frac{a+b}{2}\right\}$. A symmetric initial design is obtained by assigning the number of observations for locations in $[a, (a + b)/2)$ to their corresponding mirror image about the point $\left\{\frac{a+b}{2}\right\}$. This completes the first step of the simulated annealing algorithm.

The second step is to generate a new design. We used the perturbation scheme presented by Fang and Wiens (2000). We modify the third step as follows: Denote the maximum number of iterations by *maxit* and current iteration by *iter*. Accept the new design with probability π, defined as

$$\pi = \begin{cases} 1 & \text{if } \Delta\mathcal{L}_{ave} \leq 0, \\ uniform\,(0.5, 0.9) & \text{if } \Delta\mathcal{L}_{ave} \geq 0 \text{ and } iter < maxit/16, \\ uniform\,(0.25, 0.5) & \text{if } \Delta\mathcal{L}_{ave} \geq 0 \text{ and } maxit/16 \leq iter < maxit/8, \\ uniform\,(0, 0.1) & \text{if } \Delta\mathcal{L}_{ave} \geq 0 \text{ and } maxit/8 \leq iter < maxit^*(2/3), \\ 0 & \text{if } \Delta\mathcal{L}_{ave} \geq 0 \text{ and } maxit^*(2/3) \leq iter < maxit, \end{cases}$$

where $\Delta\mathcal{L}_{ave} = \mathcal{L}_{ave}(\text{New design chosen}) - \mathcal{L}_{ave}(\text{Initial design})$. Thus a favourable new design $(\Delta\mathcal{L}_{ave} \leq 0)$ is accepted with certainty and an unfavourable new design is accepted according to a separate Bernoulli experiment with success probability chosen such that initially such an unfavourable design is accepted with probability satisfying the inequality $0.5 < \pi < 0.9$. This follows a suggestion from Bohachevsky et al. (1986), that the tuning parameter T in step 3 of the original simulated annealing be chosen such that $0.5 < \exp(-\Delta\mathcal{L}_{ave}/T) < 0.9$. The probability of acceptance of a detrimental new design is progressively decreased to ensure that the process settles at a global minimum. In the implementations of the original simulated annealing algorithm, Fang and Wiens (2000) decrease T by a factor of 0.9 after each 100 iterations and Adewale and Wiens (2006) decrease T by a factor of 0.95 after every 20th iteration. In the implementations considered in Section 4, we found that various choice of parameter space Ω requires different scheme for decreasing T. We thus adopted this fairly generic scheme above which still imitates the characteristics of the original algorithm but obviates the demand to seek a perfect T for different parameter spaces Ω. All computations in this paper were carried out using Matlab; the relevant code is available from the authors on request.

4. Examples and applications

4.1. Example 4.1

First, we suppose an experimenter has confidence in the adequacy of the logistic model and the specification of the linear predictor for fixed effects but wants protection against overdispersion. We consider $\eta_{T,i} = \beta_0 + \beta_1 x + \phi v_i$. The design space is taken to be equally spaced points $\left\{x_i = -1 + \frac{2(i-1)}{N-1}\right\}_{i=1}^{N=41}$ in $[-1, 1]$ and the number of observations to be taken is $n = 200$. We take the range of the model parameters to be $\beta_0 \in [0.05, 1.5]$, and $\beta_1 \in [2.5, 3.5]$, i.e. $\Theta = [0.05, 1.5] \times [2.5, 3.5]$, and construct minave designs for various ranges of the overdispersion parameter ϕ. The resulting design for the overdispersion parameter space: $\Phi = [0, 0.005]$ is to place 46, 50, 66, and 38 of 200 observations at the points -0.80, -0.75, 0.10 and

Author's personal copy

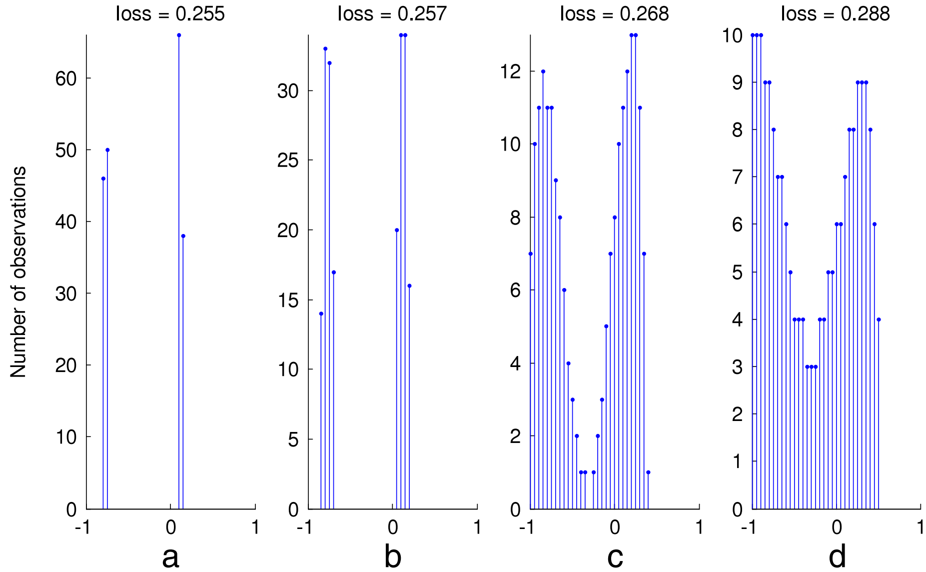

Fig. 4. Minave designs when (a) $\Phi = [0, 0.005]$, (b) $\Phi = [0, 0.05]$, (c) $\Phi = [0, 0.25]$, and (d) $\Phi = [0, 0.5]$ with $\Theta = [5, 1.5] \times [2.5, 3.5]$ and $\Lambda = \{1\}$.

0.15 respectively. This and other designs corresponding to various ranges of the overdispersion parameter are presented in Fig. 4. The range, $\Theta = [0.5, 1.5] \times [2.5, 3.5]$ of the model parameters was fixed so as to study the effect of increasing width of the overdispersion parameter space on the resulting designs. The support of the designs broadens over the design space with increasing overdispersion. Similar effects were observed when we considered other model parameter values. In general, the designs protecting against overdispersion are similar to those protecting against systematic misspecification of fixed effects, see Adewale and Wiens (2009). We observed a gradual spread of the design points such that the points that are otherwise the design points for the non-robust design which assumes no misspecification are the foci of the spread of the robust designs.

To evaluate these designs we define efficiency and robustness payoff as,

$$\text{Efficiency} = \left(\frac{aver_{\Theta} \mathcal{L}\left(\mathbf{P}_{non\text{-}robust}; \lambda = 1, \phi = 0, \boldsymbol{\beta}\right)}{aver_{\Theta} \mathcal{L}\left(\mathbf{P}_{robust}; \lambda = 1, \phi = 0, \boldsymbol{\beta}\right)} \right) \times 100\%,$$

$$\text{Robustness payoff} = \left(1 - \frac{aver_{\Omega} \mathcal{L}\left(\mathbf{P}_{robust}; \lambda, \phi, \boldsymbol{\beta}\right)}{aver_{\Omega} \mathcal{L}\left(\mathbf{P}_{non\text{-}robust}; \lambda, \phi, \boldsymbol{\beta}\right)} \right) \times 100\%$$

where $\mathbf{P}_{non\text{-}robust}$ is the design constructed on the assumption that the fitted model is correct ($\phi = 0, \lambda = 1$) and $\mathbf{P}_{robust}$ is the robust design constructed by minimizing (31). The cost of the robustness of our designs is measured by the lost efficiency: 1 − *Efficiency*. The payoff is the percentage reduction in loss due to the use of a robust design as opposed to a non-robust design which assumes the fitted model to be exactly correct. The efficiencies for the four designs are 100%, 99.8%, 98.4% and 97.0% while the payoffs are 0%, 0.21%, 11.5% and 35.2%, respectively. Thus the cost of robustness increases with the increasing width of the overdispersion parameter further away from zero which corresponds to no overdispersion. However, robustness is inexpensive here, the highest cost being 3.0% at the widest range $\Phi = [0, 0.5]$ and a corresponding payoff of 35.2% which is much higher than lost efficiency.

Now we fix the overdispersion parameter ϕ at zero and seek designs for four different ranges Λ of link parameter values with the range of model parameters kept at $\Theta = [0.5, 1.5] \times [2.5, 3.5]$ as above. In this case, $f(\mathbf{x}_i, \lambda)$ is given by (12). For all ranges of the link parameter considered the number of support points is two or three. These designs are presented in Fig. 5. The robust designs protecting against misspecification are not characterized by a spreading of the support points over the design space as is the case of robust designs protecting against overdispersion. Rather these designs have fewer support points. The robustness is provided by redistribution and shifting of the location of these few design sites. The efficiencies of these robust designs are 99.4%, 94.9%, 95.9% and 97.6% for the range of link parameters $[0.75, 1]$, $[25, 1]$, $[0, 1]$, and $[0, 2]$ with payoffs of 0.48%, 4.3%, 3.97% and 5.47%.

4.2. Example 4.2: Application to Calandra Granaria data

Busvine (1938) presented a data set for the mortality of grain beetle (*Calandra Granaria)* after exposure to ethylene oxide (C_2H_4O). A total of 290 grain beetles were exposed to 10 different levels of concentrations of C_2H_4O and the proportion killed at each concentration after a 1 h period was recorded (Table 3). Using iterated reweighted least squares as discussed in Williams (1982), we fitted the overdispersion model (5). The parameter estimates are $\hat{\phi} = 0.946$, $\hat{\beta}_0 = -3.835$, and

Author's personal copy

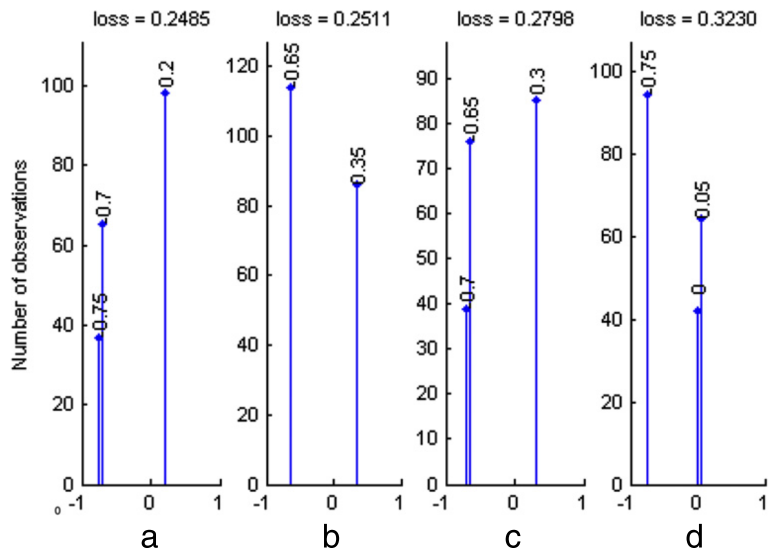

Fig. 5. Minave designs when (a) $\Lambda = [0.75, 1]$, (b) $\Lambda = [0.25, 1]$, (c) $\Lambda = [0, 1]$, and (d) $\Lambda = [0, 2]$ with $\Theta = [0.5, 1.5] \times [2.5, 3.5]$ and $\Phi = \{0\}$.

Table 3
Toxicity of ethylene oxide to the *Calandra Granaria*.

Dose x_i ($\log_{10} C_2H_4O$ mg/100 ml)	# of observations n_i	# dead n_iy_i
0.394	30	23
0.391	30	30
0.362	31	29
0.322	30	22
0.314	26	23
0.260	27	7
0.225	31	12
0.199	30	17
0.167	31	10
0.033	24	0

$\hat{\beta}_1 = 15.754$. Using the range $\Theta = [-4, -2] \times [8, 16]$ of parameter values, $\Phi = \{0\}$ and $\Lambda = \{1\}$ we construct the design minimizing (31). This is the model which assumes that there is no overdispersion nor link misspecification but still offers some robustness by incorporating a range of model parameter values. In Table 4, we label this design the Regular Optimal Design. The model-robust design that protects against overdispersion as well as link misspecification is constructed by taking $\Theta = [-4, -2] \times [8, 16]$, $\Phi = [0.75, 1]$ and $\Lambda = [0, 1]$. The regular optimal design is a 4-point design with support at dose levels.033, 0.167, 0.362, and 0.391. The model-robust design has support at all 10 dose levels in the design space. For comparison, we present two other designs in Table 4 — the original design in the data for which there is no rationale given and the design with equal number of observation at all dose levels which we labeled "naive" design, being the design that a naive experimenter would consider. The results indicated that both the original design and the "naive" design are preferable to the regular optimal design since they have good robustness payoff with small efficiency loss. The robustness payoff is 61.6%, 62.2%, and 62.6% for the original design, the "naive" design and the model-robust design, respectively. The model-robust design is superior of all four designs, it has the highest payoff at the least efficiency loss.

4.3. *Example 4.3: Application to Ames Salmonella assay*

We return to the design in the Ames Salmonella assay presented in Table 1 which has six dose levels and 3 replicates observations at each dose level. There is no rationale given for this design. In this section we consider designing for this assay when the interest is to take a total of 18 observations and the design space consists of 50 equally spaced points between 0 and 1000 µmg inclusive. As indicated in Example 1.1 that overdispersion appears in the data, we intend to fit the Poisson model with mean response (2) but seeks protection against overdispersion using approximated variance structure of $e^{\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)\boldsymbol{\beta}} + \phi^2 e^{2\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i)\boldsymbol{\beta}}$. The ranges of plausible parameter values are taken to be [2.0, 3.5], [−0.002, 0.001], and [0.30, 0.35] for β_0, β_1 and β_2, respectively. Note that these ranges of parameter values cover the estimates from the fit of *Models I*, *II* and *III* using the data in Table 1. The design protecting against overdispersion for the ranges of $\Phi = [0.05, .1]$, [0.25, 0.28], and [0.5, 1.0] of the overdispersion parameter are presented in Table 5. These designs also have few distinct support points. In this Poisson example, we have found that the number of support points is not increasing with the increasing departure of

Author's personal copy

Table 4
Comparing designs for the toxicity of ethylene oxide on *Calandra Granaria*.

Dose level	Number of observations (Total: 290)			
	Original design	"Naive" design	Model-robust design	Regular optimal design
0.033	24	29	38	5
0.167	31	29	32	132
0.199	30	29	29	–
0.225	31	29	26	–
0.260	27	29	23	–
0.314	26	29	25	–
0.322	30	29	25	–
0.362	31	29	29	65
0.391	30	29	31	88
0.394	30	29	32	–
Lost efficiency	7.3	6.9	4.5	0
Robustness payoff (%)	61.6	62.2	62.6	0

Table 5
Designs for a Poisson model with protection against overdispersion.

Φ	Support points (Number of observations)	Efficiency (%)	Payoff (%)
{0}	0 (2), 306.12 (6), 326.53 (3), 1000 (7)	100	0
[0.05, 0.1]	0 (2), 306.12 (9), 1000 (7)	98.0	3.4
[0.25, 0.28]	0 (1), 285.71 (8), 1000 (9)	88.1	26.1
[.5, 1.0]	0 (1), 285.71 (8), 1000 (9)	88	25.6

the overdispersion parameter from the value corresponding to no overdispersion. Similar to the situation for logistic model, the efficiency trade-off for robustness is inexpensive. At the widest range $\Phi = [0.5, 1]$, the corresponding payoff of 25.6% is much higher than the highest cost being 12%.

5. Concluding remarks

The design criteria proposed in this paper provide viable alternatives to classical optimal design when there is the possibility of model misspecification. The numerical examples presented show that our designs are robust against misspecification of the link function as well as overdispersion in both logistic and Poisson models. Our findings in this paper indicates the robust design that protects against overdispersion in the assumed logistic model is taking observations in clusters around the sites that would have been the design points for a classical design while that for assumed Poisson model is mostly reallocating observations on the similar sites that would have been the design points for a classical design. Generally, we found that the usual recommendation from classically optimal design would be inadequate in the presence of link misspecification and/or overdispersion. The numerical examples presented in Section 4, suggest that there is a marked difference between the characteristics of designs protecting against overdispersion and those protecting against misspecification of link functions despite the fact that the same approach has been used for their construction.

While there have been many proposals for modelling extra variation in a GLM, the approach we adopted is very attractive in that it accommodates both the random variation and the fixed effects on the same scale — somewhat akin to a generalized linear mixed modelling approach. For instance, regarding modelling possible overdispersion in binary data, another viable alternative that we consider is to specify the distribution of the true model as

$$Y_i|v_i \sim \frac{1}{n_i}\text{binomial}\left(n_i, \mu\left(\eta_i\right)\right), \quad \eta_i = \mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}$$
$$E\left(v_i\right) = \mu\left(\eta_i\right) \quad \text{and} \quad var\left(v_i\right) = \phi^2\mu\left(\eta_i\right)\left(1-\mu\left(\eta_i\right)\right).$$

The mean response from this specification is exactly the approximate true mean response (6) used in this paper but the variance is $var\left(Y_i\right) = \frac{1}{n_i} w\left(\mathbf{z}^{\mathrm{T}}\left(\mathbf{x}_i\right)\boldsymbol{\beta}\right)\left\{1+\left(n_i-1\right)\phi^2\right\}$ which is different from (7) given by the specification used in this work. Another way to obtain the same true mean response and variance prescribed by this alternative specification is to use a correlated binomial model:

$$Y_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}, \quad \text{with}$$
$$Y_{ij} \sim Bernoulli\left(\mu\left(\eta_i\right)\right) \quad \text{and}$$
$$cov\left(Y_{ij}, Y_{ik}\right) = \phi\mu\left(\eta_i\right)\left(1-\mu\left(\eta_i\right)\right), \quad j \neq k.$$

An advantage that comes with the correlated binomial model is that it can accommodate underdispersion by allowing negative correlation between the Bernoulli trials. Underdispersion, however, is a rare phenomenon in practice compared to

Author's personal copy

overdispersion. Designs protecting against this form of alternative specification when $\phi > 0$ behave the same as our specification — a spread of the design points of the design space. Interestingly, when $\phi < 0$ under this specification, the resulting optimal design is similar to the classical optimal design recommendation — choosing a few distinct points to observe the response.

We also note that the robust designs found in the paper are almost as efficient as the classical designs even when the model assume is actually correct since for all cases of the examples presented in this article the largest loss of efficiency is less than 12%. For most cases in the examples presented, the payoffs for the robust designs protecting against a relatively large uncertainty of misspecification in either overdispersion, or bad choice of a link function, or two situations combined are much large than their trade-offs in efficiency. One exception is that although the payoff of the robust design in Section 4.1 against link misspecification alone is still better than the loss of efficiency, but the gain is not as much as that for against overdispersion. Further improvements might be obtained by using a second order Taylor's expansion:

$$h(\mathbf{x};\lambda) = \left.\frac{\partial g}{\partial \lambda}\right|_{\lambda=1}(\lambda-1) + \left.\frac{\partial^2 g}{\partial \lambda^2}\right|_{\lambda=1}(\lambda-1)^2 + o\left([\lambda-1]^2\right),$$

instead of (11). However, we note that, in Section 4.2, the payoff for robust design protecting against M2 and M3 combined departures, the payoffs are even more impressive than protecting any of M2 or M3 departure alone.

While we have adopted the averaging of the loss function over uniformly scattered points from the parameter space in dealing with dependency of design criterion on unknown parameter, other possibilities remain. A proper Bayesian paradigm can be used which requires the specification of prior distribution on the parameters rather than the specification of a range of parameter values. It is our opinion that it might be easier to elicit information about plausible values of parameters from the experimenter than information leading to prior distribution assumptions. Another approach when an initial range of parameter values are available is the minimax approach. That is, an approach that seeks the design corresponding to the least loss for the worst possible parameter values in the specified range. Our experience is that the designs we presented here behave very similar to the minimax design (see Adewale and Wiens, 2006; Wiens and Xu, 2008). The approach we adopted in this paper can also be applied to construct the design for protecting possible misspecification in the linear predictor so that it is possible to unify our results for all three kinds of misspecification considered in GLMs.

Acknowledgements

The authors are grateful to Professor Douglas Wiens for his helpful suggestions, and to the Associate Editor and the reviewer for their insightful comments. The work of both authors are funded by the Natural Sciences and Engineering Research Council of Canada.

Appendix. Derivations

Proof of Theorem 1. First, we show that it is true without consideration of possible overdispersion, under conditions as in Fahrmeir (1990): the maximum likelihood estimate $\hat{\beta}$ exists and is consistent, and $\partial l\left(\hat{\boldsymbol{\beta}}\right)/\partial \beta$ is $o_p\left(n^{-1/2}\right)$. The log-likelihood l, the score function and -1 times the second derivative according to the assumed model are

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N}\sum_{j=1}^{n_i}\frac{(y_{ij}-\mu_i)}{var(y_i)}\mathbf{z}(\mathbf{x}_i)\frac{\partial \mu_i}{\partial \eta_i}, \qquad -\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^{\mathrm{T}}} = \sum_{i=1}^{N} n_i w_i \mathbf{z}(\mathbf{x}_i)\mathbf{z}^{\mathrm{T}}(\mathbf{x}_i).$$

An expansion of $\partial l(\boldsymbol{\beta})/\partial \beta_j$ around β_0 gives

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \beta_j} = \frac{\partial l\left(\boldsymbol{\beta}_0\right)}{\partial \beta_j} + \sum_k \left(\beta_k - \beta_{0,k}\right)\frac{\partial^2 l\left(\boldsymbol{\beta}_0\right)}{\partial \beta_j \partial \beta_k} + \frac{1}{2}\sum_k\sum_l \left(\beta_k-\beta_{0,k}\right)\left(\beta_l-\beta_{0,l}\right)\frac{\partial^3 l\left(\boldsymbol{\beta}_*\right)}{\partial \beta_j\partial \beta_k\partial \beta_l},$$

where β_j and $\beta_{0,j}$ are the j^{th} terms of the vectors $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_0$ respectively and $\boldsymbol{\beta}_*$ is a point on the line segment connecting $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_0$. If we replace $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}$ in this expansion, we obtain

$$\sqrt{n}\sum_k\left(\hat{\beta}_k-\beta_{0,k}\right)\left[\frac{1}{n}\frac{\partial^2 l\left(\boldsymbol{\beta}_0\right)}{\partial \beta_j\partial \beta_k} + \frac{1}{2n}\sum_l\left(\hat{\beta}_l-\beta_{0,l}\right)\frac{\partial^3 l\left(\boldsymbol{\beta}_*\right)}{\partial \beta_j\partial \beta_k\partial \beta_l}\right] = -\frac{1}{\sqrt{n}}\frac{\partial l\left(\boldsymbol{\beta}_0\right)}{\partial \beta_j}.$$

For the distributions belongs to exponential family, according to Kredler (1986), the third derivatives $\frac{\partial^3 l(\boldsymbol{\beta}_*)}{\partial \beta_j\partial \beta_k\partial \beta_l}$ are bounded, and using the consistency of $\hat{\beta}$, we have that

$$\left[\frac{1}{n}\frac{\partial^2 l\left(\boldsymbol{\beta}_0\right)}{\partial \beta_j\partial \beta_k} + \frac{1}{2n}\sum\left(\hat{\beta}_l-\beta_{0,l}\right)\frac{\partial^3 l\left(\boldsymbol{\beta}_*\right)}{\partial \beta_j\partial \beta_k\partial \beta_l}\right] \xrightarrow{p} -H_{jk},$$

Author's personal copy

where H_{jk} is the $(j,k)^{th}$ element of the matrix $\mathbf{H}_n = -\frac{1}{n}\frac{\partial^2 l(\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathrm{T}}} = \mathbf{Z}^{\mathrm{T}}\mathbf{PWZ}$. Thus the limit distribution of $\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)$ is that of the solution of the equations

$$\sum H_{jk}\sqrt{n}\left(\hat{\beta}_k - \beta_{0,k}\right) = \frac{1}{\sqrt{n}}\frac{\partial l\left(\boldsymbol{\beta}_0\right)}{\partial\beta_j},$$

i.e. is the limit distribution of $\mathbf{H}_n^{-1}\frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}}$. Using the central limit theorem for independent not identically distributed random variables we have that $\frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}}$ has a multivariate normal limit distribution with asymptotic mean

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{N} n_i E\left[y_i - \mu_i\left(\boldsymbol{\beta}_0\right)\right]\mathbf{z}\left(\mathbf{x}_i\right)/a(\varphi) = \sqrt{n}\mathbf{b},$$

and asymptotic covariance matrix $\tilde{\mathbf{H}}_n = \mathbf{Z}^{\mathrm{T}}\mathbf{PW}_T\mathbf{Z}$. From this it follows that $\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)$ is $AN\left(\sqrt{n}\mathbf{H}_n^{-1}\mathbf{b}, \mathbf{H}_n^{-1}\tilde{\mathbf{H}}_n\mathbf{H}_n^{-1}\right)$, as required. □

Proof of Theorem 2. It follows Theorem 1. The derivation from Theorem 1 to Theorem 2 is very similar to the proof of Corollary 3.1 in Adewale and Wiens (2009). So we omit it. □

References

Abdelbasit, K.M., Plackett, R.L., 1983. Experimental design for binary data. J. Amer. Statist. Assoc. 78, 90–98.
Adewale, A.J., Wiens, D.P., 2006. New criteria for robust integer-valued designs in linear models. Comput. Statist. Data Anal. 51, 723–736.
Adewale, A.J., Wiens, D.P., 2009, Robust designs for misspecified logistic models. J. Statist. Plann. Inference, 139, 3-15. Metaheuristics, Combinatorial Optimization and Design of Experiments.
Atkinson, A.C., Haines, L.M., 1996. Designs for nonlinear and generalized linear models. In: Ghosh, R., Rao, C.R. (Eds.), Handbook of Statistics. pp. 437–475.
Biedermann, S., Dette, H., Pepelyshev, A., 2004, Some robust design strategies for percentile estimation in binary response models. Universit ät Dortmund, Technical Report 2004-19.
Bliss, C.I., 1935. The calculation of the dose-mortality curve. Ann. Appl. Biology 22, 134–167.
Bohachevsky, I.O., Johnson, M.E., Stein, M.L., 1986. Generalized simulated annealing for function optimization. Technometrics 28, 209–217.
Busvine, J.R., 1938. The toxicity of ethylene oxide to calandra oryzae, C. granaria, tribolium castaneum, and cimex lectularius. Ann. Appl. Biology 25, 605–632.
Burridge, J., Sebastiani, P., 1994. D-optimal designs for generalized linear models with variance proportional to the square of the mean. Biometrika 81, 295–304.
Chaloner, K., Larntz, K, 1989. Optimal bayesian design applied to logistic regression experiments. J. Statist. Plann. Inferencce 21, 191–208.
Chaudhuri, P., Mykland, P., 1993. Nonlinear experiments: Optimal design and inference based on likelihood. J. Amer. Statist. Assoc. 88, 538–546.
Chernoff, H., 1953. Locally optimal designs for estimating parameters. Ann. Math. Statist. 24, 586–602.
Crowder, M.J., 1978. Beta-binomial anova for proportions. Appl. Stat. 27, 34–37.
Dean, C.B., 1992. Testing for overdispersion in poisson and binomial regression models. J. Amer. Statist. Assoc. 87, 451–457.
Dette, H., Wong, W.K., 1996. Optimal Bayesian designs for models with partially specified heteroscedastic structure. Ann. Stat. 24 (5), 2108–2127.
Fang, K.-T., Wang, Y., 1994. Number-Theoretic Methods in Statistics. Chapman & Hall, London.
Fang, Z., Wiens, D.P., 2000. Integer-valued, minimax robust designs for estimation and extrapolation in heteroscedastic, approximately linear models. J. Amer. Statist. Assoc. 95, 807–818.
Fahrmeir, L., 1990. Maximum likelihood estimation in misspecified generalized linear models. Statistics 21, 487–502.
Ford, I, Titterington, D.M., Kitsos, C.P., 1989. Recent advances in nonlinear experimental design. Technometrics 31, 49–60.
Ford, I., Torsney, B., Wu, C.F.J., 1992. The use of a canonical form in the construction of locally optimal designs for nonlinear problems. J. R. Stat. Soc. Ser. B 54, 569–583.
Kredler, C., 1986. Behaviour of third order terms in quadratic approximations of lr-statistics in multivariate generalized linear models. Ann. Stat. 14, 326–335.
King, J., Wong, W.K., 2000. Minimax D-optimal designs for logistic model. Biometrics 56, 1263–1267.
Lawless, J.F., 1987. Negative binomial and mixed poisson regression. Canad. J. Statist. 15, 209–225.
Li, G., Majumdar, D., 2008. D-optimal designs for logistic models with three and four parameters. J. Statist. Plann. Inferencce 138, 1950–1959.
Margolin, B.H., Kaplan, N., Zeiger, E., 1981. Statistical analysis of the ames salmonella/microsome test. Proc. Natl. Acad. Sci. USA 76, 3779–3783.
Minkin, S., 1987. Optimal design for binary data. J. Amer. Statist. Assoc. 82, 1098–1103.
Minkin, S., 1993. Experimental design for clonogenic assays in chemotherapy. J. Amer. Statist. Assoc. 88, 410–420.
Pierce, D.A., Sands, B.R., 1975, Extra-Bernoulli Variation in Binary Data. Technical Report 46, Department of Statistics, Oregon State University.
Ponce de Leon, A.M., Atkinson, A.C., 1993. Designing optimal experiments for the choice of link function for a binary data model. In: Muller, W.G., Wynn, H.P., Zhigljavsky, A.A. (Eds.), Model- Oriented Data Analysis. Physica-Verlag, Heidelberg, pp. 25–36.
Pregibon, D., 1980. Goodness of link selection in generalized linear models. Appl. Stat. 29, 15–24.
Sinha, S., Wiens, D.P., 2002. Robust sequential designs for nonlinear regression. Canad. J. Statist. 30, 601–618.
Sitter, R.R., 1992. Robust designs for binary data. Biometrics 48, 1145–1155.
Wiens, D.P., Xu, X., 2008. Robust prediction and extrapolation designs for misspecified generalized linear regression models. J. Statist. Plann. Inferencce 138, 30–46.
Williams, D.A., 1982. Extra-binomial variation in logistic linear models. Appl. Stat. 31 (2), 144–148.
Woods, D.C., Lewis, S.M., Eccleston, J.A., Russell, K.G., 2006. Designs for generalized linear models with several variables and model uncertainty. Technometrics 48 (2), 284–292.