D-OPTIMAL DESIGNS FOR POISSON REGRESSION MODELS

K. G. RUSSELL, D. C. WOODS, S. M. LEWIS, J. A. ECCLESTON

ABSTRACT

We consider the problem of finding an optimal design under a Poisson regression model with a log link, any number of independent variables and an additive linear predictor. Local D-optimality of a class of designs is established through use of a canonical form of the problem and a general equivalence theorem. The theorem is applied in conjunction with clustering techniques to obtain a fast method of finding designs that are robust to wide ranges of model parameter values. The methods are illustrated through examples.

Southampton Statistical Sciences Research Institute
Methodology Working Paper M08/01

D-optimal designs for Poisson regression models

By K. G. RUSSELL
School of Mathematics and Applied Statistics, University of Wollongong, NSW 2522, Australia
kgr@uow.edu.au

D. C. WOODS, S. M. LEWIS
Southampton Statistical Sciences Research Institute, University of Southampton, Southampton, SO17 1BJ, UK
D.Woods@soton.ac.uk, S.M.Lewis@soton.ac.uk

and J. A. ECCLESTON
Centre for Statistics, School of Physical Sciences, University of Queensland, Brisbane QLD 4072, Australia
jae@maths.uq.edu.au

Summary

We consider the problem of finding an optimal design under a Poisson regression model with a log link, any number of independent variables and an additive linear predictor. Local D-optimality of a class of designs is established through use of a canonical form of the problem and a general equivalence theorem. The theorem is applied in conjunction with clustering techniques to obtain a fast method of finding designs that are robust to wide ranges of model parameter values. The methods are illustrated through examples.

Some key words: Clustering; Locally optimal design; Log-linear models; Robust design.

1. Introduction and notation

We consider experiments in which the ith response variable, Y_i, is described by a Poisson distribution with rate λ_i dependent on p independent variables through the log-linear model:

$$\ln(\lambda_i) = \eta_i = f(x_i)^{\mathrm{T}}\beta = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ji}\,, \quad i = 1, \ldots, n\,, \tag{1}$$

where $x_i = (x_{1i}, \ldots, x_{pi})^{\mathrm{T}}$, $f(x_i) = (1, x_i^{\mathrm{T}})^{\mathrm{T}}$, $\beta_0, \ldots, \beta_p$ are unknown constants, and $\beta_j \neq 0$ for $j > 0$ (see, for example, McCullagh & Nelder, 1989, Ch. 6). Our aim is to find a design for an experiment which enables efficient estimation of $\beta = (\beta_0, \ldots, \beta_p)^{\mathrm{T}}$ in the sense of minimizing the volume of the $100(1-\alpha)\%$ confidence ellipsoid for β; that is, a D-optimal design. A complication is that, in common with all non-linear models, the optimal design depends on the unknown values of the model parameters. Locally optimal designs can be found by assuming particular values for the parameters which can be updated in a sequence of experiments (see, for example, Atkinson et al., 2007, Ch. 17). Alternative ways of overcoming parameter dependence are through Bayesian design (Chaloner & Larntz, 1989; Firth & Hinde, 1997), maximin criteria (Sitter, 1992; Biedermann et al., 2006) and compromise or parameter-robust design (Woods et al., 2006; Dror & Steinberg, 2006).

There is little guidance available on how to design a multi-variable experiment for Poisson regression. For single variable toxicology experiments, Minkin (1993) found locally optimal designs for estimation of the slope parameter in terms of an "effective dose", and compared the performance of the optimal designs with designs having various different numbers of equally-spaced support points. For models with one or two variables, Wang et al. (2006a) investigated the dependence of locally D-optimal designs on functions of the parameter values and Wang et al. (2006b) developed sequential designs. For a single variable, Ford et al. (1992) used a transformation of the design space to a canonical form, together with geometric arguments, to find locally optimal designs for a class of nonlinear models including Poisson regression.

Our aim is to determine closed-form locally D-optimal designs for several variables and first-order Poisson regression. These include designs of Wang et al. (2006a) and Ford et al. (1992) as special cases. We demonstrate the usefulness of these designs in the construction of efficient parameter-robust designs.

2. Locally D-optimal designs

An approximate design $\xi \in \Xi$ in design space $\mathcal{X}$ with finite support is represented as

$$\xi = \left\{ \begin{array}{cccc} x_1 & x_2 & \dots & x_s \\ \delta_1 & \delta_2 & \dots & \delta_s \end{array} \right\} ,$$

where $x_i \in \mathcal{X}$, $\mathcal{X}$ is a compact subset of $\mathbb{R}^p$, $0 < \delta_i \leq 1$ and $\sum_{i=1}^{s} \delta_i = 1$. Under a first-order Poisson regression model with linear predictor (1) and log link, the information matrix for ξ is

$$\begin{aligned} M(\xi,\beta) &= \sum_{i=1}^{s} \delta_i w(x_i) f(x_i) f(x_i)^{\mathrm{T}} \\ &= X^{\mathrm{T}} W X \,, \end{aligned}$$

where $w(x_i) = \exp(\eta_i)$, $X = (f(x_1),\dots,f(x_s))^{\mathrm{T}}$ and $W = \mathrm{diag}\,\{\delta_i w(x_i)\}$, $i = 1,\dots,s$.

We want to find a D-optimal design, ξ^*, i.e. such that

$$|M(\xi^*, \beta)| = \max_{\xi \in \Xi} |M(\xi, \beta)| \,.$$

In order to suppress the dependence of this design problem on β, following Ford et al. (1992), we apply a linear transformation to $f(x_i)$ to obtain

$$f(z_i) = Bf(x_i)\,, \quad i = 1, \ldots, s\,,$$

where $z_i = (z_{1i}, \ldots, z_{pi})^{\mathrm{T}} \in \mathcal{Z}$,

$$B = \begin{pmatrix} B_{11} & 0 \\ 0 & B_{22} \end{pmatrix}, \qquad B_{11} = \begin{pmatrix} 1 & 0 \\ \beta_0 & \beta_1 \end{pmatrix},$$

$B_{22} = \mathrm{diag}\,\{\beta_2, \ldots, \beta_p\}$ and $\beta_j \neq 0$ $(j = 1, \ldots, p)$. It follows from (1) that $\eta_i = \left(B^{-1} f(z_i)^{\mathrm{T}}\right)^{\mathrm{T}} \beta = \sum_{j=1}^{p} z_{ji}$. Let $\psi \in \Psi$ be a design measure over the induced design space $\mathcal{Z}$, then

$$\psi = \left\{ \begin{array}{cccc} z_1 & z_2 & \ldots & z_s \\ \delta_1 & \delta_2 & \ldots & \delta_s \end{array} \right\}.$$

As the D-optimality design criterion is invariant to a linear transformation of the design space (see, for example, Pukelsheim, 1993, Ch. 6), it is sufficient to find a locally optimal design over Ψ, as in the following Lemma.

LEMMA. *A D-optimal design for the canonical first-order Poisson regression model with $\eta_i = \left(B^{-1} f(z_i)^{\mathrm{T}}\right)^{\mathrm{T}} \beta = \sum_{j=1}^{p} z_{ji}$, where $a_j \leq z_{ji} \leq b_j$, for a_j, b_j constants, and $b_j - a_j \geq 2$ $(j = 1, \ldots, p)$, is given by*

$$\psi^* = \left\{ \begin{array}{cccc} z_1^* & z_2^* & \cdots & z_{p+1}^* \\ 1/(p+1) & 1/(p+1) & \cdots & 1/(p+1) \end{array} \right\},$$

where $z_1^* = (b_1, \ldots, b_p)^{\mathrm{T}}$ and z_k^* has jth element $b_j - 2\Delta_{j,k}$, $k = 2, \ldots, p+1$, with $\Delta_{j,k} = 1$ if $j = k-1$, and 0 otherwise.

The proof is outlined in the Appendix.

Example 1. Wang et al. (2006a) reported D-optimal designs for (1) with $p = 1$ and $p = 2$ and support points defined in terms of $q_i = \lambda_i/\lambda_c$, where $\lambda_c = \exp(\beta_0)$ and $\lambda_i = \exp(\eta_i)$. In their context of toxicity studies, where $x_{ji} \geq 0$ and $\beta_j < 0\,(i = 1, \ldots, s;\ j = 1, 2)$, the canonical variables satisfy $z_{1i} \leq \beta_0$, $z_{2i} \leq 0$. The D-optimal support points are $\{z_1 : \beta_0, \beta_0{-}2\}$ for $p = 1$, and $\{(z_1, z_2) : (\beta_0, 0), (\beta_0{-}2, 0), (\beta_0, -2)\}$ for $p = 2$. At these support points, $q_1 = 1$, $q_2 = \exp(-2)$

for $p = 1$, and $q_1 = 1$ and $q_2 = q_3 = \exp(-2)$ for $p = 2$, matching the results from Wang et al. (2006a).

An optimal design in $\mathcal{X}$ space for finitely bounded variables follows directly from the Lemma by application of the inverse transformation to obtain $f(x_i) = B^{-1}f(z_i)$, $i = 1, \ldots, p+1$.

Theorem. *A D-optimal design for Poisson regression with $\eta_i = f(x_i)^{\mathrm{T}}\beta$, $l_j \leq x_{ji} \leq u_j$ and $|\beta_j(u_j - l_j)| \geq 2$ $(j = 1, \ldots, p)$ has $p+1$ equally weighted support points:*

$$
\begin{aligned}
x_1^* &= (-\beta_0/\beta_1 + b_1/\beta_1, b_2/\beta_2, \ldots, b_p/\beta_p)^{\mathrm{T}}, \\
x_2^* &= (-\beta_0/\beta_1 + (b_1 - 2)/\beta_1, b_2/\beta_2, \ldots, b_p/\beta_p)^{\mathrm{T}}, \\
x_i^* &= (-\beta_0/\beta_1 + b_1/\beta_1, b_2/\beta_2, \ldots, (b_{i-1} - 2)/\beta_{i-1}, b_i/\beta_i, \ldots, b_p/\beta_p)^{\mathrm{T}},
\end{aligned}
$$

where $i = 3, \ldots, p+1$ and b_j is defined in the Lemma.

Remark 1. In practice, the requirement $|\beta_j(u_j - l_j)| \geq 2$ in the Theorem is not overly restrictive. For example, the use of the standardised design space $[-1, 1]^p$ requires $|\beta_j| \geq 1$, $j = 1, \ldots, p$.

Example 2. For $p = 2$, suppose that $\beta = (1, -2, 3)^{\mathrm{T}}$, and $x_{1i} \in [0, 10]$, $x_{2i} \in [0, 12]$ $(i = 1, \ldots, s)$. Then the design with equally weighted support points $\{(x_{1i}, x_{2i}) : (0, 12), (1, 12), (0, 34/3)\}$ is D-optimal.

Remark 2. When $b_j - a_j < 2$ for some $j = 1, \ldots, p$, it can be shown that the D-optimal *saturated* design for the canonical model has equally weighted support points $z_1 = (b_1, b_2, \ldots, b_p)^{\mathrm{T}}$, $z_i = (b_1, \ldots, \max(b_{i-1} - 2, a_{i-1}), b_i, \ldots, b_p)$ $(i = 2, \ldots, p)$. Ford et al. (1992) proved that this design is D-optimal for $p = 1$ over Ψ. In general, the D-optimal design over Ψ is not saturated when $b_j - a_j < 2$ for some j. For example, if $\beta = (-0.91, 0.04, -0.69)^{\mathrm{T}}$ and $x_{ji} \in [-1, 1]$ $(i = 1, \ldots, s;\ j = 1, 2)$, then $z_1 \in [-0.95, -0.87]$, $z_2 \in [-0.69, 0.69]$ and the D-optimal design has four support points.

3. Robust design

Often experimenters have little information about parameter values prior to observing the data. Woods et al. (2006) found *compromise* designs for GLMs which are robust to wide ranges of parameter values. Dror & Steinberg (2006) approximated these methods by using a K-means clustering algorithm (see, for example, Hastie et al., 2001, Ch. 14). The design points from a large number of locally optimal designs were found by computer search for ranges of parameter values; the cluster centroids were then used as equally weighted support points of a *cluster design*. The use of the Theorem in Section 2 allows the efficient computation of a cluster design, ξ^c, by removing the need to perform computer searches, as follows.

Algorithm

1. *Define a parameter space $\mathcal{B}$ for β, and a design space $\mathcal{X}$ that satisfies the Theorem for all $\beta \in \mathcal{B}$.*

2. *Generate N_s vectors β_{sj} $(j = 1, \ldots, N_s)$ from $\mathcal{B}$ using quasi-random numbers (Gentle, 2003, Ch. 3).*

3. *For each β_{sj}, apply the Theorem to construct a locally optimal design ξ_j^*.*

4. *Apply a clustering algorithm to the total $N_s(p+1)$ design points (see below).*

5. *Use each cluster centroid as an equally weighted support point of ξ^c.*

Example 3. Suppose that $p = 2$, $x_{1i}, x_{2i} \in [-1, 1]$ $(i = 1, \ldots, p+1)$ and $\mathcal{B} = [-\alpha, +\alpha] \times [1, 1+\alpha] \times [-1-\alpha, -1]$, with $\alpha > 0$. Fig. 1 shows the ensemble of design points from the locally D-optimal designs obtained from the Theorem using $N_s = 1000$ values of β for each of $\alpha = 1, 2, 5, 10, 15, 20$. All the designs include the point $(1, 1)$. For the remaining points, the spread of values for x_{1i} and x_{2i} increases with increasing α, reflecting increasing uncertainty in the value of β.

Fig. 1 suggests that the natural clusters in the design points are not well described by spherical clusters of equal volume and hence a more flexible clustering algorithm than K-means may be advantageous. We compared K-means with the model-based clustering algorithm of Fraley & Raftery (2002), which is based on a mixture of normal distributions with possibly differing covariance matrices. We also investigated the number of support points, i.e. clusters, that should be selected. To allow the estimation of the variance components in the model-based clustering, the locally optimal design points were "jittered" slightly through the addition of a small amount of uniform random noise.

Example 3 cont. For each of K-means and model-based clustering techniques, a cluster design was formed with $s = 3, \ldots, 22$ support points for each of $\alpha = 1, 2, 5, 10, 15, 20$. The efficiency of each cluster design was calculated for each value of β as $\{|M(\xi^c, \beta_{sj})|/|M(\xi^*, \beta_{sj})|\}^{1/p}$ $(j = 1, \ldots, N_s)$. Fig. 2 shows how the median and minimum efficiencies vary with s and α. The use of model-based clustering frequently results in higher median and minimum efficiencies than K-means. This is often achieved with fewer support points, as for $\alpha = 5, 10$, where the designs with highest efficiencies are found from model-based clustering and have three support points (1.0,-1.0), (0.3,-1.0), (1.0,-0.3) and (1.0,-1.0), (0.54,-1.0), (1.0,-0.54) respectively. As α increases, the parameter space $\mathcal{B}$ increases in volume and hence efficiencies are lower relative to locally optimal designs. The more efficient designs for $\alpha = 15, 20$ have a greater number of support points than for smaller α, with K-means designs requiring more support points than the designs from model-based clustering.

This method is particularly useful for designs with large numbers of variables, as needed for screening experiments. To avoid the need to perform a post-hoc

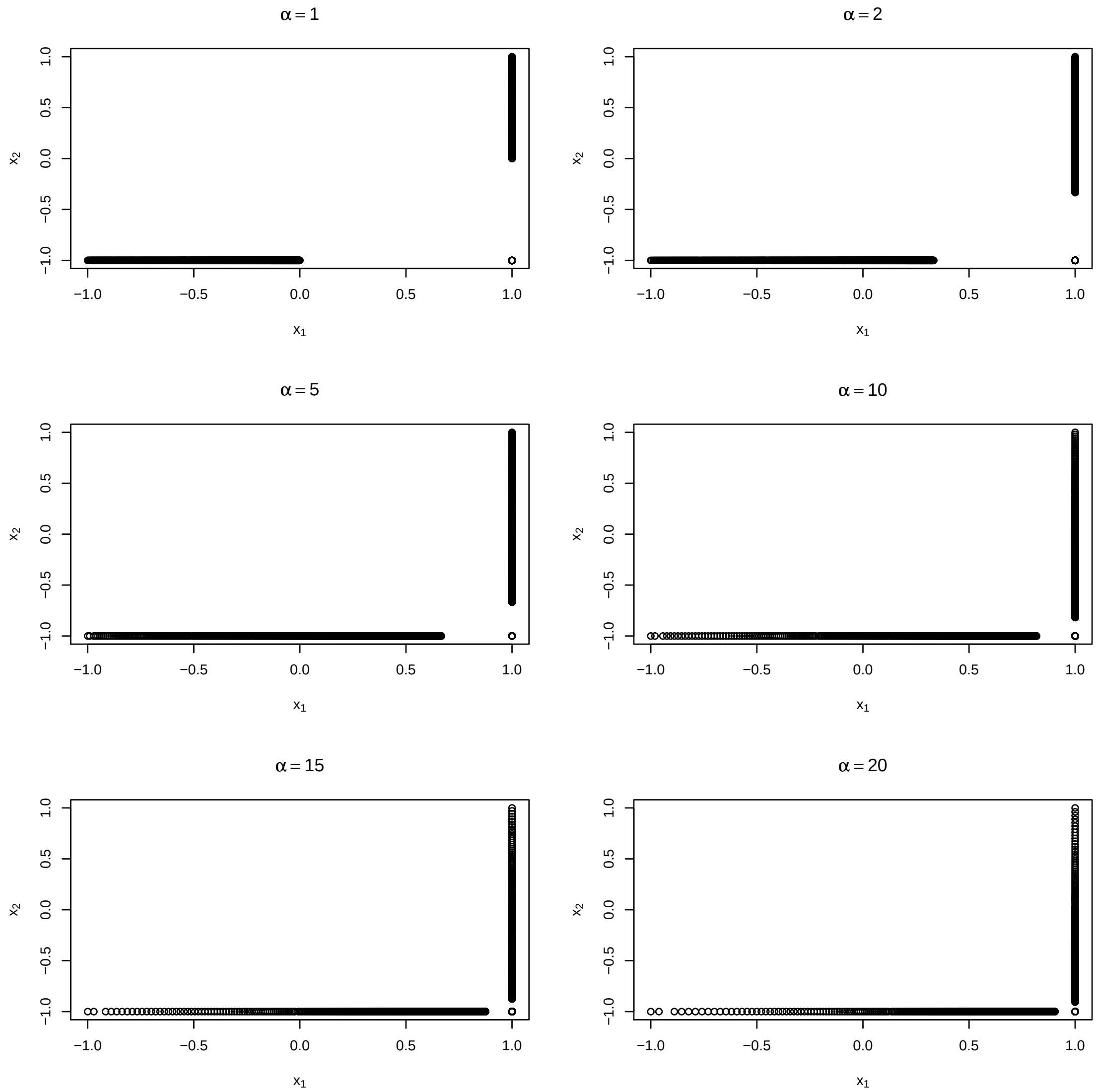

Fig. 1: Design points (∘), often overlapping, from 1000 locally optimal designs for two variables for each parameter space of Example 3.

evaluation of designs with many different numbers of support points, as in Example 3, standard metrics from the unsupervised learning literature for the selection of the number of clusters can be employed, such as the Bayesian Information Criterion (BIC; Fraley & Raftery, 2002).

Example 4. For $p = 10$ variables and a first-order Poisson regression model and log link, suppose a robust design is required across the following parameter space:

$$\beta_k \in \begin{cases} [-\alpha, +\alpha] & \text{for } k = 0\,, \\ [1, 1+\alpha] & \text{for } k = 1, 3, 5, 7, 9\,, \\ [-1-\alpha, -1] & \text{for } k = 2, 4, 6, 8, 10\,. \end{cases} \tag{2}$$

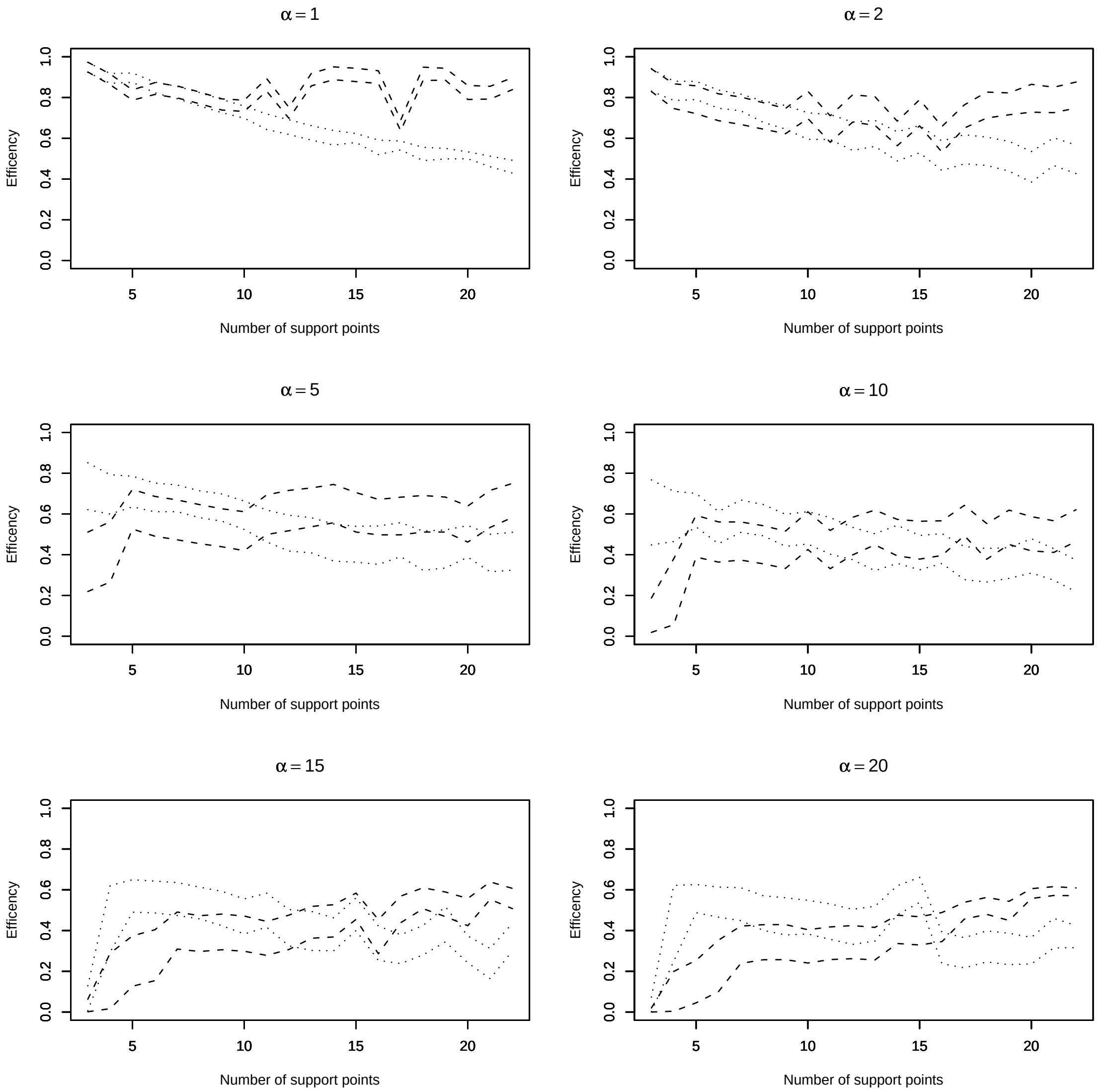

Fig. 2: Median and minimum efficiencies against number of support points ($s = 3, \ldots, 22$) for the cluster designs for Example 3: K-means (dashed); model-based (dotted).

Cluster designs were found and evaluated using model-based clustering for (2) with $\alpha = 1, 2, 3$ and $N_s = 1000$. For each robust design, the number of support points (clusters), chosen using BIC, was found to be equal to 21. Table 1 gives the median and minimum efficiencies across $\mathcal{B}$ and shows that, for each value of α, the cluster design performs well across the parameter space. As in Example 3, the median and minimum efficiencies decrease as α increases but good performance is maintained even for $\alpha = 3$. The use of the BIC statistics allows an informed choice of the number of support points without needing to evaluate more than one design.

Table 1: *Median and minimum efficiencies for the model-based cluster designs in Example 4, with parameter spaces defined through (2)*

	$\alpha = 1$	$\alpha = 2$	$\alpha = 3$
median eff.	0.936	0.877	0.748
minimum eff.	0.895	0.803	0.633

4. Conclusions

The results presented in this paper allow the analytic construction of D-optimal designs for first-order Poisson regression with a log-link, and demonstrate their use in the fast construction of designs robust to parameter values. First-order models are commonly used in practical data analysis, and are particularly appropriate for the analysis of data from experiments in the early stages of scientific investigations. Hence the design methods from this paper are particularly important for screening experiments, which may involve a large number of variables. The use of the Theorem enables much larger problems to be tackled than computer search currently allows.

Acknowledgements

This work was supported by grants from the Engineering and Physical Sciences Research Council, UK, and the Australian Research Council. We are grateful to Dr. Nema Dean (University of Glasgow) for helpful advice on model-based clustering.

Appendix

Proof of the Lemma

The result is proved using a general equivalence theorem (Atkinson et al., 2007, p. 122) and, specifically, by proving that the standardised variance of the predicted response at point z_0, $d(z_0, \psi^*) = w(z_0) f(z_0)^{\mathrm{T}} M^{-1}(\psi^*) f(z_0) \leq p + 1$ for all $z_0 \in \mathcal{Z}$, where $M(\psi^*)$ is the information matrix for the design ψ^* defined in the Lemma. Some algebra establishes that the symmetric matrix $M^{-1}(\psi^*)$ has (i, j)th entry

$$\frac{p+1}{4\exp(\sum_{i=1}^{p} b_i)} m^{ij} ,$$

where $m^{11} = (2 - \sum_{i=1}^{p} b_i)^2 + e^2 \sum_{i=1}^{p} b_i^2$; $m^{1j} = 2 - \sum_{i=1}^{p} b_i - e^2 b_{j-1}$, $m^{jj} = 1 + e^2$ $(2 \leq j \leq p+1)$; and $m^{ij} = 1$ $(2 \leq i < j \leq p+1)$. Further,

$$d(z_0, \psi^*) = \frac{p+1}{4} \exp\left(\sum_{i=1}^{p} z_{i0} - \sum_{i=1}^{p} b_i \right) g(z_0) ,$$

where

$$g(z_0) = m^{11} + 2\sum_{i=2}^{p+1} m^{1i} z_{(i-1)0} + 2\sum_{i=2}^{p}\sum_{j=i+1}^{p+1} m^{ij} z_{(i-1)0} z_{(j-1)0} + \sum_{i=2}^{p+1} m^{ii} z^2_{(i-1)0} \,.$$

It is easy to show that $d(z_i^*, \psi^*) = p+1$, $i = 1, \ldots, p+1$. The Karush-Kuhn-Tucker theorem provides the following necessary conditions for the constrained maximisation of $d(z_0, \psi^*)$ to be achieved:

$$\frac{\partial d(z_0, \psi^*)}{\partial z_{j0}} - \mu_j = 0 \,, \tag{A1}$$

subject to

$$\mu_j(b_j - z_{j0}) = 0 \,, \quad \mu_j \geq 0 \,, \quad z_{j0} \leq b_j \,, \tag{A2}$$

for $j = 1, \ldots, p$, where the μ_j are Lagrange multipliers. If $\mu_j > 0$ for every j, then (A1) and (A2) imply that $z_{i0} = b_j$ and $d(z_0, \psi^*) = p+1$. If $\mu_j = 0$ for at least one j, without loss of generality, set $\mu_1 = \ldots = \mu_r = 0$ $(1 \leq r \leq p)$. Then to satisfy (A2), $z_{(r+1)0} = b_{r+1}, \ldots, z_{p0} = b_p$. From (A1), after some algebra, $z_{jd} = b_j - c$, $i = 1, \ldots, r$, where $c = 2/r$ or $c = 4/(r + e^2)$. For each solution, $\partial d / \partial z_{i0}^* \geq 0$ $(i = r+1, \ldots, p)$ and hence, from (A1), $\mu_j \geq 0$, satisfying (A2). As $d(z_0, \psi^*) = c(p+1)/2$ if $c = 2/r$, and $d(z_0, \psi^*) = (p+1)\exp(-rc)$ if $c = 4/(r+e^2)$, the maximum value of $d(z_0, \psi^*)$ over $\mathcal{Z}$ is $p+1$ when $r = 1$. Therefore, from the general equivalence theorem, ψ^* is D-optimal. □

References

Atkinson, A. C., Donev, A. N. & Tobias, R. D. (2007). *Optimum Experimental Designs, with SAS.* Oxford: Oxford University Press.

Biedermann, S., Dette, H. & Pepelyshev, A. (2006). Some robust design strategies for percentile estimation in binary response models. *Can. J. Statist.* **34**, 603–622.

Chaloner, K. & Larntz, K. (1989). Optimal Bayesian design applied to logistic regression experiments. *J. Statist. Plan. Infer.* **21**, 191–208.

Dror, H. A. & Steinberg, D. M. (2006). Robust experimental design for multivariate generalized linear models. *Technometrics* **48**, 520–529.

Firth, D. & Hinde, J. P. (1997). Parameter neutral optimum design for non-linear models. *J. R. Statist. Soc.* B **59**, 799–811.

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. Statist. Soc.* B **54**, 569–583.

Fraley, C. & Raftery, A. E. (2002). Model-based clustering, discriminant analysis and density estimation. *J. Am. Statist. Assoc.* **97**, 611–631.

Gentle, J. E. (2003). *Random Number Generation and Monte Carlo Methods*, 2nd ed. New York: Springer.

Hastie, T., Tibshirani, R. & Friedman, J. (2001). *The Elements of Statistical Learning*. New York: Springer.

McCullagh, P. & Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed. New York: Chapman and Hall.

Minkin, S. (1993). Experimental design for clonogenic assays in chemotherapy. *J. Am. Statist. Assoc.* **88**, 410–420.

Pukelsheim, F. (1993). *Optimal Design of Experiments*. New York: Wiley.

Sitter, R. R. (1992). Robust designs for binary data. *Biometrics* **48**, 1145–1155.

Wang, Y., Myers, R. H., Smith, E. P. & Ye, K. (2006a). D-optimal designs for Poisson regression models. *J. Statist. Plan. Infer.* **136**, 2831–2845.

Wang, Y., Smith, E. P. & Ye, K. (2006b). Sequential designs for a Poisson regression model. *J. Statist. Plan. Infer.* **136**, 3187–3202.

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**, 284–292.