D-Optimal Design for Generalized Linear Models

Hugh A. Chipman and William J. Welch

Graduate School of Business
University of Chicago, Chicago, IL 60637, USA

Department of Statistics and Actuarial Science
University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

May, 1996

Abstract

D-optimality is one of the most commonly used design criteria for linear regression models. In industrial experiments binary or count data often arise, for example defective/nondefective or number of defects. For such data Generalized Linear Models (GLMs) are appropriate. An analogous D-optimality design criterion can be developed using the asymptotic covariance matrix. For GLMs, this matrix is a weighted version of the covariance matrix for the linear case, and the extension of existing D-optimality algorithms is thus fairly straightforward. A number of examples compare and contrast GLM D-optimal designs to linear regression D-optimal designs. In general, the GLM D-optimal design depends on the values of the parameters being estimated. We describe a minimax approach to make the design more robust to the unknown parameters. The main example considers the design of a multifactor industrial follow-up experiment with binary data.

1 Introduction

D-optimality is a standard optimal design criterion for linear regression models, partly because it corresponds to minimizing the area of confidence regions for unknown parameters, and partly because it is reasonably easy mathematically and computationally. In industrial experiments non-normal responses are often measured. For example, binary data arise if parts are classified as defective or nondefective, or count data would correspond to the number of defects. For such data Generalized Linear Models (GLMs) are appropriate (see for example McCullagh and Nelder 1989 and Dobson 1983).

For GLMs analogous optimal design criteria have been developed. The inverse of the expected information matrix is used in place of the covariance matrix in the linear case. Unfortunately, the information matrix depends on the unknown parameters, so the parameters must be known in order to optimally estimate them. Abdelbasit and Plackett (1983) considered two-point symmetric designs for a single explanatory variable and a binary response. They investigated various sequential design strategies to overcome dependence on the unknown parameters. Chaloner and Larntz (1989) used a Bayesian approach for binary logistic regression models, again with a single explanatory variable. They integrated the unknown parameters over a prior distribution. The work of Sitter and Torsney (1993) is of special interest, since they used a canonical form to express designs problems and were able to find D-optimal designs for multiple predictors. Currently such methods have difficulties with irregular design spaces. Similar dependence on unknown parameters occurs in nonlinear regression models; see Ford, Kitsos, and Titterington (1989) for a comprehensive review.

For GLMs, the form of the expected information matrix is especially simple, consisting of a weighted version of the covariance matrix for linear models. Thus, in Section 2, we illustrate how standard algorithms for D-optimal design can be quickly modified for GLMs. In doing so, all the strengths of the original methods are transferable, including the typical ability to deal with arbitrary design regions. Section 3 gives some examples that illustrate the way in which the design changes with the unknown parameters. In Section 4 we give a multifactor industrial example where molded parts were rated as acceptable or unacceptable. Parameter estimates and covariances from an initial factorial design provide a set of reasonable parameter values for design of a follow-up experiment. We use a minimax criterion over the set of values to overcome the dependence on the unknown parameters. Sitter (1992) considered a similar minimax approach for a single predictor. Conclusions are given in Section 6.

2 The D-Optimality Criterion for GLMs

The motivation for a D-optimality criterion for GLMs can be seen as a simple extension of the linear case. One way of viewing the linear case is that a D-optimal design will yield the smallest confidence regions for the parameters of interest. In the GLM setting, confidence statements are often made on an asymptotic basis, via the information matrix. The inverse expected information matrix I^{-1} is the analogue of the covariance matrix of the coefficients in the linear model. Thus to minimize the size of confidence regions for GLM parameter estimates, I^{-1} is used in place of the covariance matrix $(X'X)^{-1}$.

For members of the GLM family, the information matrix has a simple form, which is the main reason that this specific extension is considered. Before illustrating this form, we define some notation for GLMs. If Y is the response vector, and corresponding values of the predictor variables are in the matrix X, then the expected value of response Y_i is related to the predictors x_i via the given "link" function g:

$$\mathrm{E}[Y_i|x_i] = \mu_i = g^{-1}(x_i'\beta) = g^{-1}(\eta),$$

where β is a vector of parameters to be estimated. The variance of the response is given by:

$$\mathrm{Var}[Y_i|x_i] = V(\mu_i).$$

That is, the variance of Y_i is a given function its mean. A distribution for the response must also be specified. The expected information does not depend on the form of the distribution beyond the mean and variance relations stated above. Dobson (1983, Appendix 2) shows that for GLMs with the above specification, the jk^{th} element of the expected information matrix is given as follows:

$$I_{jk} = \sum_{i=1}^{N} \frac{x_{ij}x_{ik}}{V(\mu_i)} (\frac{\partial \mu_i}{\partial \eta_i})^2.$$

This can be written in matrix notation:

$$I = X'WX,$$

where W is a diagonal matrix with elements:

$$w_i = (\frac{\partial \mu_i}{\partial \eta_i})^2 \frac{1}{V(\mu_i)}.$$

Thus the asymptotic variance covariance matrix $I^{-1} = (X'WX)^{-1}$ is of a similar form to the covariance matrix $(X'X)^{-1}$ for the linear case. This similarity facilitates the adaptation of existing software. All examples in this paper are carried out using ACED (Algorithms for the Construction of Experimental Designs; Welch 1996). It is fairly easy to add new criteria to this software which can interface with all implemented design optimization algorithms. Furthermore, the capability for weighted linear regression was already an element of the code.

Another feature of this class of models is that it includes Quasi-Likelihood, introduced by Wedderburn (1974). In this situation, a distribution for the response is not actually specified, and only mean and variance relations are given. These relations yield a pseudo information matrix (called pseudo, since no likelihood is specified) of the same form as that given above. Thus the optimality criterion may be extended beyond GLMs to cover other mean-variance relations.

The main drawback of this approach is that the coefficients β must be known in order to formulate an optimal design. The presence of μ_i terms (which depend on β) in W are the cause of this dependence. We return to this issue in Section 5. The following examples consider "local" D optimality where the design is optimal near "known" parameter values.

3 Illustrative Examples

The examples given in this section illustrate how D-optimal designs for GLMs differ from those for linear regression models. Probably most important is the amount of "curvature" induced in the mean by the link function for given parameter values. Another component of the GLM model which may influence the design is the variance function. For all of the examples, the predictor variables will be restricted to the range $[-1, 1]$ and the designs for various fixed parameter values will be found.

3.1 Poisson Regression with One Predictor

If our response has a Poisson distribution, we would quite often use the log link, giving the following relations:

$$\mu_i = e^{(x_i'\beta)} \text{ and } V(\mu) = \mu.$$

The optimal design depends on the values of β, so if they are known we can find the optimal design. Suppose the linear predictor is $\beta_0 + \beta_1 x$. The first plot in Figure 1 illustrates the effect of varying the slope β_1 on the optimal design. In this plot, a single curve represents the mean as a function of x for fixed β. The slope β_1 is varied to produce different curves. The value of β_0 is unimportant and here it is held fixed at the value 1. For all six values of β_1, the optimal design has equal allocation between two support points. For curve 1, for example, the two support points are denoted by a "1". We used a large number (100) of design points and found that exactly 50 were placed at each support point. The general result seems intuitively plausible.

There are several interesting features of this plot. One of the most striking is that the design points are not always at the extreme values of the design region, as they are in linear D-optimal design. The non-linear relation between the mean and x causes this. In this case we see that as the curvature of the mean response curve increases, the left design point moves in the direction of increasing curvature.

To extend this example, consider the same model, but with a quadratic term in the linear predictor. As before, the coefficient can be varied, and the effect on the optimal design studied. This is illustrated in the second plot of Figure 1. Here, we assume that $\eta = 1 + x + \beta_2 x^2$, with β_2 taking values in the range (-2.0, 0.5). Analogous to a quadratic regression model, the optimal design has three points with equal allocations. Unlike the regression model, however, curvature from the link function results in design points that are not equally spaced nor necessarily at the extremes of the design region. The location of the center point seems to depend only mildly on the curvature, and the leftmost point moves inwards with increasing curvature.

For the sake of comparison, optimal designs for the Gamma family were calculated with the same log link and $\eta = 1 + x + \beta_2 x^2$ for the same β_2 values as the Poisson case. Since the variance

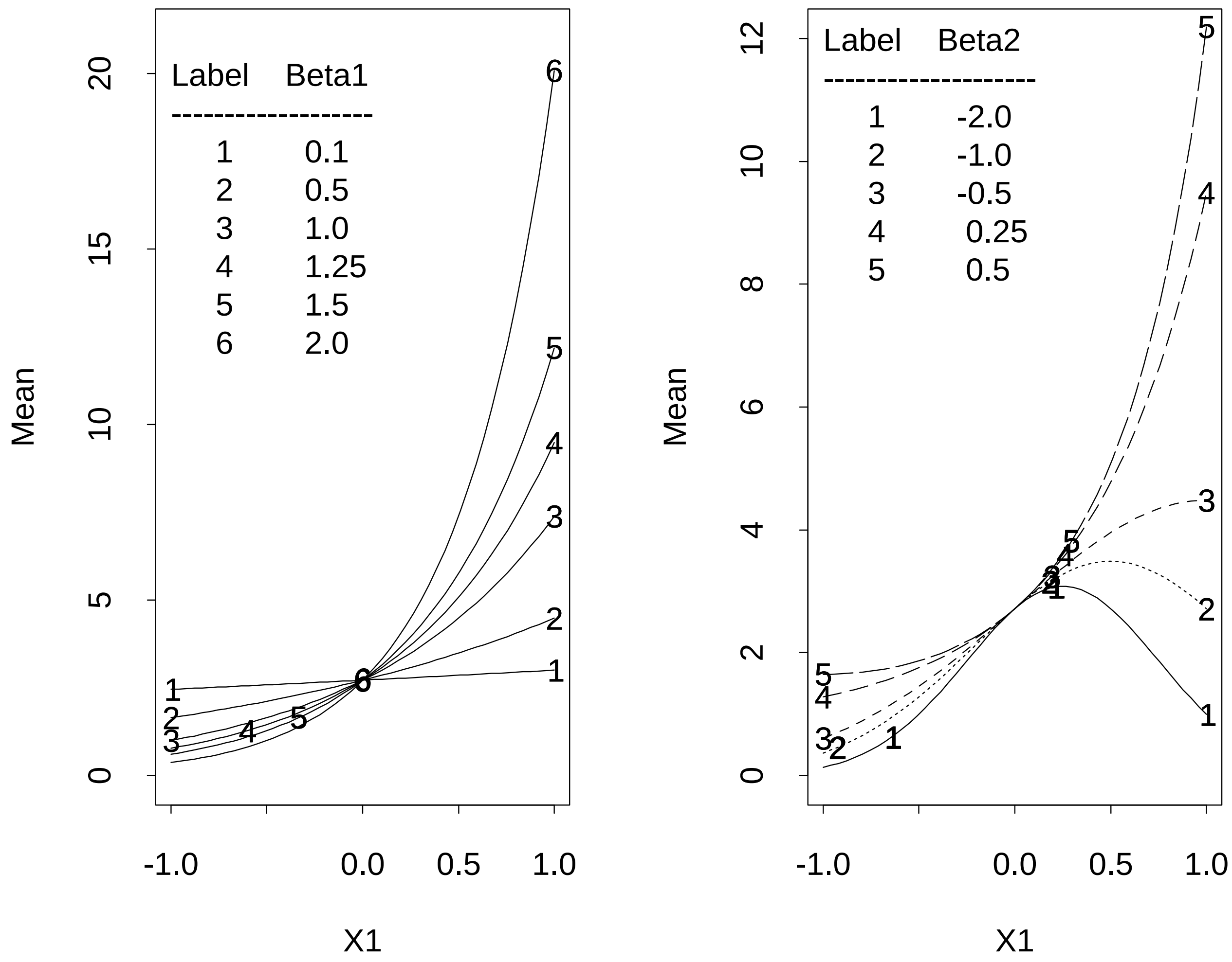

Figure 1: Optimal Designs for Poisson Regression

function is quadratic ($V(\mu) \propto \mu^2$), different design points are optimal. For the same coefficients as the Poisson quadratic model, all designs have $\frac{1}{3}$ of the points at each of -1, 0, and 1. So although the mean was related to the predictor in the same fashion, the new variance function has altered the optimal design.

3.2 Logistic Regression with Multiple Predictors

The logistic model for a binary response is given by:

$$\mu_i = \frac{e^{(x_i'\beta)}}{1 + e^{(x_i'\beta)}} \qquad\qquad V(\mu) = \mu(1-\mu),$$

where μ is the expected proportion of successful outcomes observed. In this section we consider optimal design for multiple predictor variables. Since a design with more than two dimensions is difficult to represent graphically, we limit ourselves to just two predictors. In practice, the algorithm can deal with more predictors if necessary.

We shall assume that the linear predictor has the form $\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2$, with $\beta_0 = 0$ for simplicity. Note that if nonzero values of β_0 were used, the entire curve would simply be shifted on the transformed scale. The designs produced for various values of (β_1, β_2) can be plotted on a square region representing the design space of $[-1, 1] \times [-1, 1]$. Four such plots are displayed in Figure 2 , for all combinations of $\beta_1 \in \{1, 3\}$ and $\beta_2 \in \{1, 3\}$. Note the because of the symmetry of the design region and the link function, we will get similar designs for negative values. The values of β_1 and β_2 are shown in the top and left margins, respectively. For example, the rightmost plot in the first row is the optimal design for $\beta_1 = 3$ and $\beta_2 = 1$. In all cases we optimized 12 design points; they are randomly perturbed a small amount in the plots to uncover replicates.

We notice immediately that the designs contain a larger number of distinct points than the optimal designs for one predictor variable. As many as six different support points occur for some coefficient combinations. In the linear case, there would only be the four corners of the design region. For the upper left plot, we see that this is the case. Here, the slopes for both predictors are 1, so that even with the logistic transformation, the response surface is close enough to linear that the optimal points lie at the extremes of the design space. In the other three cases, the design points

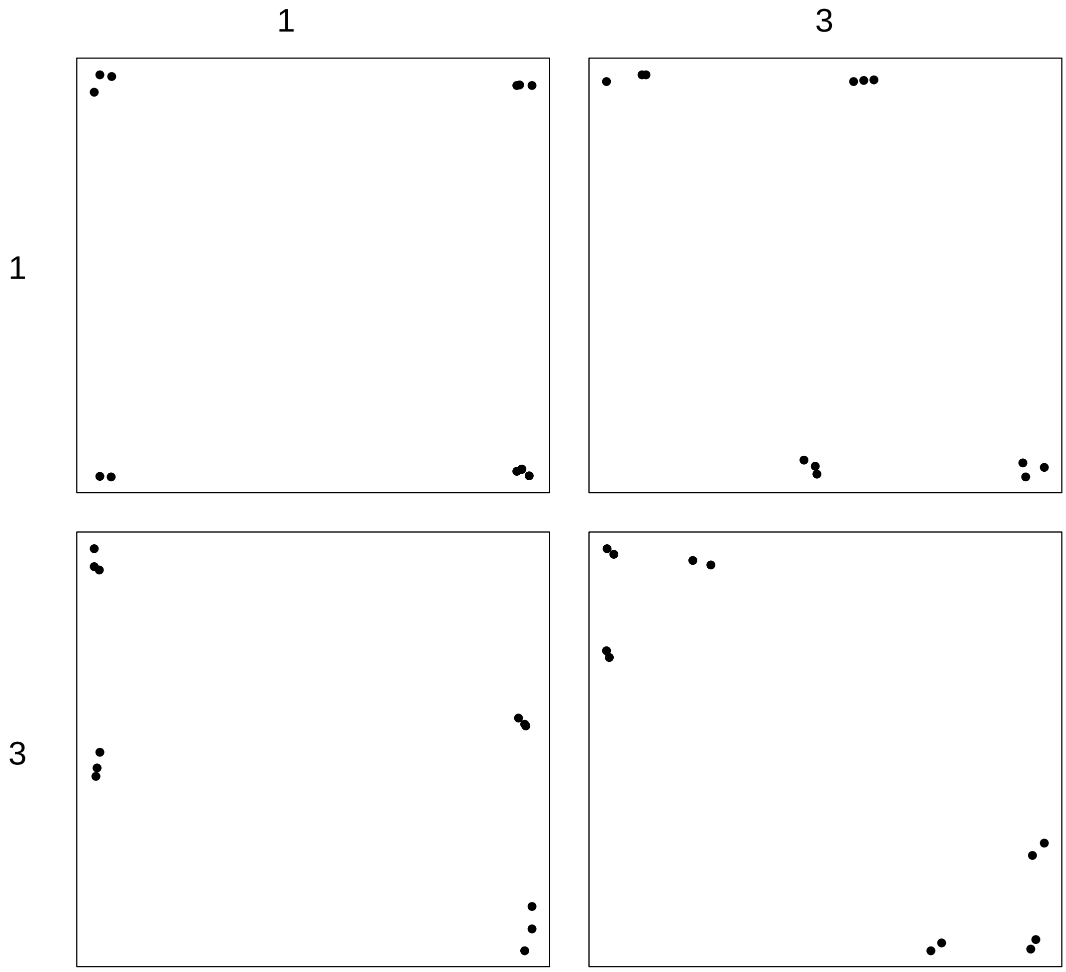

Figure 2: Optimal Designs for Logistic Regression in Two Variables

no longer lie at the extreme corners. When only one of the coefficients is small, the design points are at the extreme levels for that variable only. In the other dimension, some of the design points have "moved in", since the curvature in that direction is greater. This behavior is the same as when there is only one predictor variable. When the curvature is more extreme in both directions, more support points are used, and along either axis some lie inside the extreme values. This behavior would get more severe as the coefficients increase in absolute magnitude.

4 A Minimax Approach to Parameter Dependence

As mentioned in Section 1, the dependence of D-optimal designs for GLMs on the unknown parameters can be overcome by Bayesian and/or sequential methods. For example, if a screening experiment has been carried out, then the estimates from this experiment may be used as approximate values to design an efficient follow up study. Since the original experiment is likely to only provide vague information on the parameters, it seems reasonable to consider a collection of possible parameter values, find the optimal design for each set of parameters, and then choose one of the designs that seems best either by informal or formal properties. Alternatively, a collection of parameter values could be generated from a Bayesian prior.

Each possible parameter vector gives rise to a locally optimal design. To choose amongst these designs, we use a minimax approach suggested by Welch (1983) and Sacks et. al. (1989). It can be characterized as follows:

1. A collection B of plausible values of the parameter vector β are chosen: $B = \{\beta_1, ..., \beta_K\}$

2. For each β_k, an optimal design X_k is determined. This design minimizes the D-optimality design criterion $c(X|\beta_k)$ over all designs X for the given value of β_k.

3. For β_k, the relative efficiency of a design X relative to the optimal design X_k is defined as

$$\text{eff}(X|\beta_k) = \frac{c(X_k|\beta_k)}{c(X|\beta_k)}.$$

   This efficiency measure will be in $[0, 1]$ since the minimum value of $c(X|\beta_k)$ is attained when $X = X_k$. We then find the minimum efficiency of X over $\beta_1, \ldots, \beta_K$.

A	B	AB	C	AC	BC	D	Parts	Good Parts
-1	-1	1	-1	1	1	-1	16	0
-1	-1	1	1	-1	-1	1	16	9
-1	1	-1	-1	1	-1	1	16	7
-1	1	-1	1	-1	1	-1	16	12
1	-1	-1	-1	-1	1	1	16	3
1	-1	-1	1	1	-1	-1	16	1
1	1	1	-1	-1	-1	-1	16	2
1	1	1	1	1	1	1	16	0

Table 1: Design for preliminary molding experiment

4. The "best" design is the design $X^* \in \{X_1, ..., X_K\}$ which maximizes this minimum efficiency. Thus, we try to find a design that is efficient over all $\beta \in B$.

The intention of this method is to produce a design that is "robust" over plausible parameter values. These designs are not necessarily the best overall designs; they are merely the best among a (reasonably small) set of designs. One might also consider optimizing X over a larger set of designs than $\{X_1, ..., X_K\}$, still defining minimum efficiency over the K plausible values of $\beta \in B$. By including various compromise designs in the set of designs, we may find a design with a better overall performance.

5 An Industrial Example

To illustrate the application of optimal design for GLMs in an industrial example, data from a molding machine optimization study are considered. The experiment, which was analyzed in Moreno and Bonn (1985), has four factors and 128 runs comprised of 16 independent replications at each of eight design points from a 2^{4-1} fractional factorial design. The response is binary: good or bad part. The data are given in Table 1, and the factor descriptions in Table 2. Although the factors are at two levels, all are continuous variables.

Since this preliminary experiment has already been run, the motivation for using optimal design

Variable	Description
A	Injection Time (3/5 seconds)
B	Temperature (410/425° F)
C	Pressure (450/500 PSI)
D	Cured Time (19/22 seconds)

Table 2: Factor descriptions and levels used in robust optimization for the molding experiment

Parameter	int	A	B	C	D	AB	AC	BC
Estimate	-1.351	-0.551	0.213	0.182	0.147	-0.831	-1.023	-0.635

Table 3: Parameter estimates for the molding data.

is in the creation of a follow up experiment to further investigate the influence the main effects and interactions have on the response. We use the data from the initial experiment to identify a set of reasonable parameter values. A minimax criterion over the set of values overcomes the dependence of the design on the parameters. All the follow-up designs considered augment the original 128 design points with a further 128 points. For a given β, the optimal design will minimize the D-optimal criterion for all 256 runs, holding the (already completed) first 128 runs fixed.

We use a probit link $\Pr(Y=1) = \mu = \Phi(x'\beta)$ and binomial variance function $\mu(1-\mu)$. In the linear predictor $x'\beta$, the design allows an intercept and coefficients for four main effects and three interactions to be estimated. Because of the sparsity of the data, maximum likelihood does not produce finite estimates. To remedy this, we use a Bayesian probit model described in Chipman and Hamada (1996), and references therein. We place N(0,1) priors on the regression coefficients, and obtain posterior distribution via the Gibbs sampling technique (see Chipman and Hamada 1996 and, more generally, Gelfand and Smith 1990). As discussed in Chipman and Hamada (1996), such a choice of prior will limit the largest possible change on the probability scale that a change from -1 to 1 of one factor can produce. In this case, the largest change would be from 0.025 to 0.975. This produces finite estimates of the parameter vector β with an estimated covariance matrix as shown in Tables 3 and 4.

	int	A	B	C	D	AB	AC	BC
int	0.121	0.005	-0.002	0.001	-0.003	0.100	0.103	0.098
A	0.005	0.104	0.084	0.088	0.084	0.000	0.002	-0.002
B	-0.002	0.084	0.103	0.085	0.088	0.005	-0.004	0.002
C	0.001	0.088	0.085	0.105	0.085	-0.002	0.004	-0.001
D	-0.003	0.084	0.088	0.085	0.104	0.003	-0.001	0.005
AB	0.100	0.000	0.005	-0.002	0.003	0.118	0.098	0.100
AC	0.103	0.002	-0.004	0.004	-0.001	0.098	0.117	0.097
BC	0.098	-0.002	0.002	-0.001	0.005	0.100	0.097	0.114

Table 4: Estimated covariance matrix of the parameter estimates for the molding data.

The model fit has large terms for A, AB, AC, and BC. Although the other coefficients are not large relative to their spread, they will not be dropped. The small number of terms considered in this experiment suggests that at this initial stage variable selection may not be of interest. We treat the parameter estimates and their estimated covariance matrix as the mean and covariance of a multivariate normal posterior distribution. This posterior is in turn a prior for the follow-up experiment.

A number of different schemes could be used to obtain a set of plausible β values for the minimax approach described in Section 4. In this case, some posterior correlations are large; consequently we shall systematically choose β values that are representative of their joint distribution. One hundred β values were drawn using Latin hypercube sampling (McKay, Conover, and Beckman 1979). This ensures that the marginal empirical distributions have good agreement with the margins of the posterior. Iman and Conover (1982) showed how to also improve agreement between sample and target covariances; iterating their procedure gives sample covariances very close to those in Table 4. This leads to a more systematic coverage of the posterior than say Monte Carlo sampling.

Optimal designs were found for each of the 100 values of β. The efficiency of each design was then calculated over the 100 β values, and the one with largest minimum efficiency was selected. Relative efficiency is based on the D-optimality criterion $c(X) = |X'WX|^{-1/8}$, where W depends on β. Thus, relative efficiency is defined as

$$\text{eff}(X|\beta_k) = \frac{|X_k' W_k X_k|^{-1/8}}{|X' W_k X|^{-1/8}} = \left(\frac{|X_k' W_k X_k|}{|X' W_k X|}\right)^{-1/8}.$$

The matrices X and X_k are 256×8, with the first 128 runs fixed. The best design is displayed in Table 5. The minimax efficiencies of the ten best and ten worst designs are given in Table 6. We see a gradual decline in the minimax criterion over the 10 best designs. It is interesting to note that the best design has at most one coordinate on the interior of the design space for any design point.

For the sake of comparison, we also augment the original design with 128 runs from the complementary half replicate and evaluate under the same criterion. In Table 5, we see that this design has quite a poor worst-case efficiency. Evidently, valuable information is lost from such a standard approach.

The tendency of the optimal design to place most of the points at extreme levels can be explained by the relatively small size of the coefficients. As was illustrated in the example in Section 4.2, a moderate slope will result in designs similar to the linear case. It is only when the slopes become more extreme that the design will place more points on the interior of the design space. No interior points for factor D were included, a fact attributable to the small magnitude of its estimated effect. In contrast, factor A has 46 interior points out of 128. Note that it has a large estimated main effect and two large estimated interactions. Thus, we speculate that there is large curvature of the mean as a function of factor A. The examples in Section 3 showed that larger curvature leads to interior points.

If the design space were expanded beyond the hypercube $[-1, 1]^4$, then the relative slope would seem steeper, and optimal designs would no longer cluster most points at extreme values. That is, the fact that the points are at the extrema of the design region implies that estimation would be more efficient if the region were expanded. By doing so, we would obtain more efficient estimates of the parameters of interest. As the design region increases, our design points will not necessarily remain at the extreme values. Although this possibility could be investigated, we do not do so here, as there is no subject matter knowledge to let us know just how much we can expand the levels of the design variables.

A	B	C	D	# new replicates
-0.6	-1	-1	-1	4
0.74	-1	-1	-1	17
0.42	-1	-1	1	1
0.43	-1	-1	1	1
-0.84	-1	-1	1	3
-0.85	-1	-1	1	3
0.46	1	1	-1	7
0.47	1	1	-1	9
0.6	1	1	1	1
-1	-1	-0.86	1	3
-1	-1	-0.87	1	3
-1	-1	0.03	1	9
-1	-1	0.04	1	3
-1	-1	0.2	-1	11
-1	0.77	-1	1	1
-1	0.78	-1	1	2
1	-0.21	1	1	1
1	-0.36	-1	1	2
1	-0.43	1	-1	9
-1	1	-1	-1	15
1	1	-1	-1	5
1	1	-1	1	13
1	1	1	1	5

Table 5: Best minimax design for the molding experiment

10 Best Efficiencies				
0.639	0.627	0.625	0.619	0.617
0.609	0.609	0.605	0.603	0.599

10 Worst Efficiencies				
0.193	0.188	0.188	0.184	0.182
0.176	0.174	0.163	0.155	0.148

Efficiency for Factorial Design
0.191

Table 6: Minimax efficiencies for best, worst and factorial designs

The minimax efficiency criterion can be viewed as protecting oneself against the "worst case" scenario. Although it makes the designs more robust, it is not the only way of doing so. We might just as easily use other features of the distribution of the 100 efficiency scores for a given design. For example, we might be more interested in the average performance of the design, in which case we would look at the largest mean efficiency. Graphic techniques may also be used to compare the efficiency distributions. In Figure 3, we see boxplots of efficiencies over $\beta_1, \ldots, \beta_{100}$ for all 100 designs, as well as for the factorial design. The designs are ordered left to right by their minimum efficiencies. The middle 50% of the data lies within the boxes, and the whiskers extend to the most extreme values. We see that the best minimax design is a good performer, since it has a high mean efficiency, and the middle 50% is in the 0.83 to 0.94 efficiency range. Other designs, notably the third best, are interesting in that they appear to concentrate more efficiency values near 1 than the best design, and have only a few poor efficiencies. The factorial design (on the far right of the plot) is clearly a poor performer since the distribution of efficiency scores is comparable with the ten worst designs.

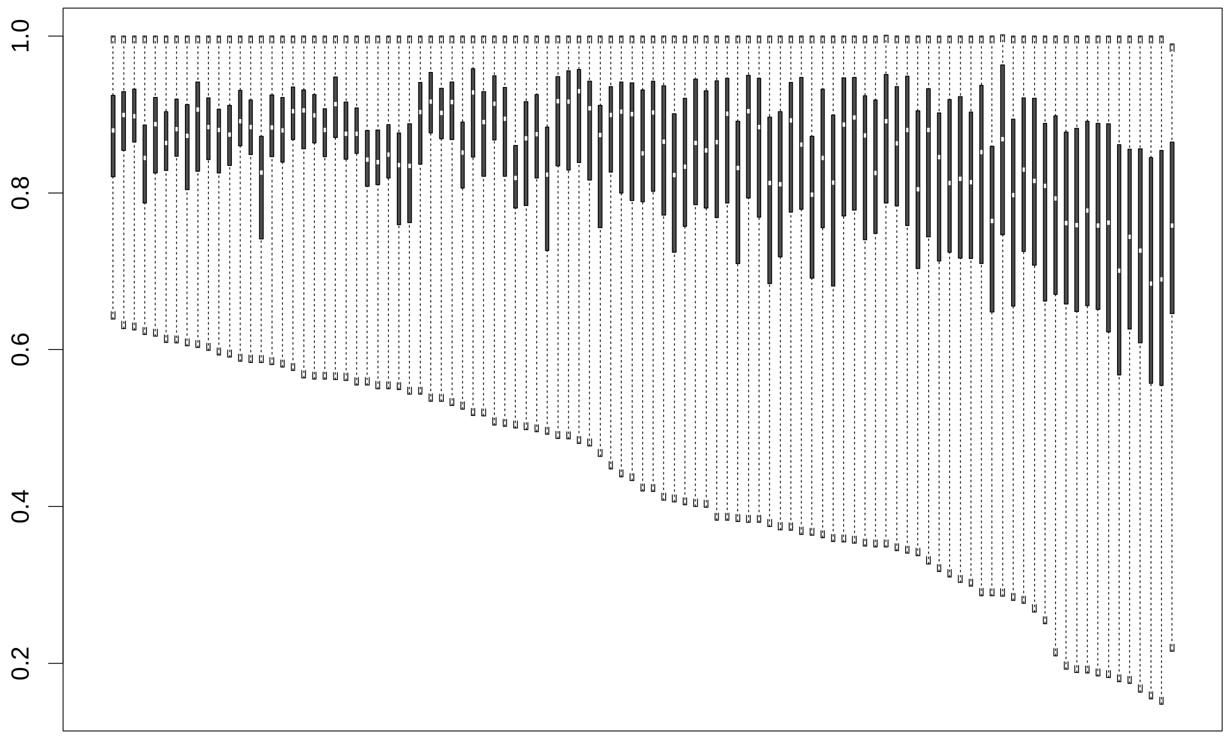

Figure 3: Efficiency score distribution (over 100 β values) of 100 designs and factorial design for the molding data

6 Conclusion

The D-optimality criterion for GLMs has behavior that is similar in many ways to D-optimality for linear models. Relatively few design points are produced, and there is a tendency to place them near the extreme values of the design region. An important difference is that the mean is no longer linearly related to the predictors, and the degree of this curvature is central in determining the departures of the optimal design from the linear case. In cases where the curvature is extreme enough within the design region, the optimal design points will "move in" from the edges of the design region.

In practice, we see that the designs produced are reasonable, but suffer from the obvious problem of needing to know the estimates before an optimal design for their estimation can be found. To overcome this, a method that uses partial knowledge about the regression coefficients is used to make the design more robust to varying values. The requirement of partial knowledge makes such procedures most useful for follow up studies, where something is known about the parameters, and we wish to find out more with a minimum expenditure.

As demonstrated by the molding experiment, the numerical methods employed here are applicable to multifactor experiments, models with interaction terms, etc. We suspect that this covers many industrial applications. As noted by Chaloner and Larntz (1989) and Sitter (1992) in their studies of single factor experiments, because these methods consider a wider collection of alternate values of parameters, the designs are typically more spread out.

References

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

Chaloner, K. and Larntz, K. (1989). Optimal Bayesian design applied to logistic regression experiments. *Journal of Statistical Planning and Inference*, 21, 191–208.

Chipman, H. and Hamada, M. (1996). Bayesian analysis of ordered categorical data from industrial experiments. *Technometrics*, 38, 1–10.

Dobson, A. J. (1983). *An Introduction to Statistical Modelling.* Chapman and Hall, London.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398 – 409.

Iman, R.L. and Conover, W.J. (1982). A Distribution-Free Approach to Inducing Rank Correlation Among Input Variables. *Communications in Statistics, Simulation and Computation*, 11, 311-334

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models, 2nd Edition.* Chapman and Hall, London.

McKay, M.D., Conover, W.J., and Beckman, R.J. (1979) A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output From a Computer Code. *Technometrics*, 21, 239–245

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

Moreno, M. and Bonn, B. (1985). Molding department optimization. In *Third Symposium on Taguchi Methods*, pages 359–366. American Suppliers Institute.

Sacks, J., Schiller, S. B., and Welch, W. J. (1989). Designs for computer experiments. *Technometrics*, 31, 41–47.

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

Sitter, R. R. and Torsney, B. (1993). D-optimal designs for generalized linear models. Technical report, Laboratory for Research in Statistics and Probability, Carleton University.

Wedderburn, R. W. M. (1974). Quasilikelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, 61, 439–447.

Welch, W. J. (1983). A mean squared error criterion for the design of experiments. *Biometrika*, 70, 205–213.

Welch, W. J. (1996). ACED (Algorithms for the Construction of Experimental Designs), unpublished software package.