Journal of the Korean
Data & Information Science Society
1999, Vol. 10, No 2, pp. 289 ~ 298

Large Robust Designs for Generalized Linear Model[1]

YoungIl Kim[2] · Myung-Wook Kahng[3]

Abstract

We consider a minimax approach to make a design robust to many types of uncertainty arising in reality when dealing with non-normal linear models. We try to build a design to protect against the worst case, i.e. to improve the "efficiency" of the worst situation that can happen. In this paper, we especially deal with the generalized linear model. It is a known fact that the generalized linear model is a universal approach, an extension of the normal linear regression model to cover other distributions. Therefore, the optimal design for the generalized linear model has very similar properties as the normal linear model except that it has some special characteristics. Uncertainties regarding the unknown parameters, link function, and the model structure are discussed. We show that the suggested approach is proven to be highly efficient and useful in practice. In the meantime, a computer algorithm is discussed and a conclusion follows.

Key Words and Phrases: Robust design, Generalized linear models

1. Introduction

Optimal design theory has been proven very useful in developing both the theoretical and empirical foundations for statistical experimental design since the equivalence theorem by Kiefer and Wolfowitz(1959). Among its many criteria, D-optimality is the one many researchers have investigated, but it was under normal theory regression model most of the time.

However, in reality, we confront difficulties when applying the optimal design theory developed for linear model directly to the design problem such as non-normal

[1]This research was supported by the ChungAng University research facility fund program of 1999.
[2]Professor, Department of Information System, ChungAng University, Kyunggi-Do, 456-756, Korea
[3]Associate Professor, Department of Statistics, Sookmyung Women's University, Seoul, 140-742, Korea

situations. In this paper we raise some issues that happen when D-optimality is applied to the generalized linear model(GLM). Then we describe a minimax approach to make the design robust to the dependence on the unknown parameters, the uncertainty of link function, and the choice of model structure. In Section 2, we briefly introduce some basic terms for GLMs and how the standard algorithms for D-optimal designs can be modified for GLMs. In Section 3, we take advantage of the theorem by Atwood(1969) to develop a design robust to many circumstances. We used a minimax approach over a set of possible scenarios to overcome the dependency on the unknown prameters, link function and model structure. A computer intensive algorithm is included. In Section 4, we provide several examples to implement our design criterion. It proves to be very efficient and powerful under these situations. Conclusions and recommendations follow in Section 5.

2. Design for Generalized Linear Model

The generalized linear model is a unified statistical modeling technique suggested by Nelder and Wedderburn(1972), an extension of the normal theory linear model to the distributions such as the gamma, Poisson and binomial. The model has three components according to McCullagh and Nelder(1989), which are as follows.

1) The random components: The components of $N \times 1$ random vector y which are assumed to share the same distribution from the exponential family. Its mean is μ and

$$Var(y) = a(\phi)V(\mu)$$

where $a(\phi)$ is a scale factor which does not depend upon μ.

2) The linear predictor: The linear model provides the linear predictor

$$\eta = f^T(x)\beta$$

where β is $p \times 1$ unknown vector and $f^T(x)$ is $1 \times p$ row vector of known functions of k explanatory variables.

3) The link function: This plays the role of linking the mean vector μ and the linear predictor $g(\mu) = \eta$.

Note that for the normal distribution $a(\phi) = \sigma^2$, $V(\mu) = 1$ and $g(\mu) = \mu$. The mean and the variance are not related. This is not necessarily true for other distributions. See McCullagh and Nelder(1989) for further details.

It can be shown that for GLMs with the above components specified, the Fisher information matrix at a design point x is:

$$I(x, \beta) = \omega(\eta) f(x) f^T(x)$$

where $\omega(\eta)$ is defined as $V^{-1}(\mu)(d\mu/d\eta)^2$. Thus $I(x,\beta)$ depends on η via $\omega(\cdot)$, which is a function of η. This $\omega(\eta)$ will be called efficiency function following Fedorov(1972) hereafter.

The design problem consists of selecting vectors $x_i, i = 1, \cdots, N$ from a design space χ such that the design defined by these N vectors is in some defined sense, optimal. An exact design ξ_N is a probability measure on the design space χ subject to the restriction that $\xi_N \cdot N$ can take only integers. Removing the restriction that $\xi_N \cdot N$ is a multiple of N, we can define the normalized version of information matrix $M(\xi,\eta)$ as follows

$$M(\xi,\eta) = \int I(\eta)d\xi(x), \quad \xi \in \Xi$$

where Ξ is the set of continuous design measures on the design space χ such that $\int_\chi d\xi(x) = 1$. One design criterion called D-optimality tries to maximize the determinant of this information matrix. But since the information matrix depends on the value of unknown parameter vector β via $\omega(\eta)$ in the case of generalized linear model, it is sometimes called a locally D-optimal design. The equivalence theorem for D-optimality says that

$$\max_{x\in\chi} \ \omega(\eta)f^T(x)M(\eta,x)f(x)$$

should be equal to the value of p, which is the number of parameters in the model when the D-optimality takes place. We shall denote $\omega(\eta)f^T(x)M(\eta,x)f(x)$ as $d_\omega(x)$ and call it "pseudo" variance function. The design minimizing the maximum $f^T(x)M(\eta,x)f(x)$ only without $\omega(\eta)$ in front would be called the G-optimal design in "logit" scale.

Normally this locally D-optimal design for generalized linear model depends on all three compenents specified above. Other than this annoying fact implementing the usual D-optimal design algorithm to the generalized linear model is straightforward. But the main shortcoming of this approach in practice is that the coefficients of the linear predictor must be known in advance in order to formulate an optimal design. The presence of μ in $\omega(\eta)$ is the main cause of this dependency. This dependency is unavoidable even if the model structure is fixed in advance. A natural approach to this problem would then be to attach a prior distribution to the parameters. For example, optimal Bayesian designs for the logistic regression model are pioneered by Chaloner and Larntz(1989). Nevertheless, many authors have sought analytical solutions in various generalized linear models. For example, Abdelbasti and Plackett(1983), Ford, Titterington, and Kitsos(1989), Sitter(1992), and recently Sebastiani and Settimi(1997) are the major papers to look at Ford, Torsney, and Wu(1992) stated that there is still interest in constructing locally optimal designs. Their primary reasons are as follows:

1) They provide a useful reference point in studies of the performance of sequential and other forms of design.

2) They are necessary for the construction of non-sequential designs based on efficiency and related criteria.

3) Where sequential designs can be carried out in batches, the design for batch $i+1$ might be locally optimal design based on $\hat{\theta}_i$, the estimate of θ based on the first i batches. In general, θ is the unknown parameter of the nonlinear model.

To overcome the shortcomings of locally D-optimal design, we define a design robust to the a priori misspecification of 1) values of unknown parameters, 2) link function and 3) model structure in the next section. The requirement of partial knowledge about the components of the GLMs makes such procedures most useful for followup studies, where something is known about the components.

3. Robust Design

For the moment, we consider the case of model structure uncertainty only. Given that the true model is unknown, we will consider a design to be robust to specification of model f if ξ is highly efficient for models likely to be encountered in practice. More specifically we shall assume that the model f is an unknown element of some known space of model function F. We will then attempt to characterize designs that are efficient, in sense to be described, for all possible $f \in F$. To do this we define two efficiency measures of design ξ with respect to ξ_1 for a specific model f.

Definition 1: D-efficiency :$D_f(\xi, \xi_1) = (detM^{-1}(\xi_1)detM(\xi))^{1/p}$.

Definition 2: "pseudo" G-efficiency : $G_f(\xi, \xi_1) = \max_{x \in \mathcal{X}} d_\omega(x, \xi) / \max_{x \in \mathcal{X}} d_\omega(x, \xi_1)$.

Note that the efficiency in Definition 2 is not the genuine G-efficinecy and so call it "pseudo" G-efficiency. The following theorem relating D and "pseudo" G-efficiency is from Atwood(1969).

Theorem 1. Let ξ_D be the D-optimal design for f then for any design ξ,

$$D_f(\xi, \xi_D) \geq G_f(\xi, \xi_D)$$

"pseudo" G_f-efficiency provides the lower bound for the D_f-efficiency of a design ξ with respect to the D-optimal design ξ_f^D for each $f \in F$. Therefore, it would be sufficient to consider only "pseudo" G_f-efficiency of a design ξ with respect to ξ_f for each $f \in F$. Loosely speaking, we will consider a design model-robust if its "pseudo" G_f-efficiency is high for every $f \in F$. Thus no matter what the subsequent analysis indicates regarding the choices of f, the efficiency of the design will be relatively high.

Definition 3: The design ξ^* is model-robust if and only if

$$\max_{\xi \in \Xi} \min_{f \in F} G_f(\xi, \xi_f) = \min_{f \in F} G_f(\xi^*, \xi_f).$$

Similar definitions can follow regarding the misspecifications of unknown parameters and link functions.

Definition 4: The design ξ^* is parameter-robust if and only if

$$\max_{\xi \in \Xi} \min_{\beta \in B} G_\beta(\xi, \xi_\beta) = \min_{\beta \in B} G_\beta(\xi^*, \xi_\beta).$$

Definition 5: The design ξ^* is link-robust if and only if

$$\max_{\xi \in \Xi} \min_{g \in G} G_g(\xi, \xi_g) = \min_{g \in G} G_g(\xi^*, \xi_g).$$

But for the cases of Definitions 4 and 5, the number of parameters in the model, p, does not change with β or link function, Definitions 4 and 5 can be simplified further. For example, Definition 4 is equivalent to Definition 6.

Definition 6: The design ξ^* is robust if and only if

$$\max_{\xi \in \Xi} \max_{\beta \in B} \max_{x \in \chi} d_\omega(x, \xi) = \max_{\xi \in \Xi} \max_{\beta \in B} d_\omega(x, \xi^*).$$

In most instances, analytic characterization of the misspecification of the components is impossible, and numerical methods are required. The following algorithm, which is a simple modification of the one by Fedorov(1972) can be used for computer construction of nearly robust designs. We will show only the case of model misspecification.

Algorithm

1. Specify non-singular starting design ξ_0, set $i = 0$.
2. Compute c_i such that
$$\max_{f \in F} G_f^{-1}(\xi_i, \xi_f) = c_i.$$
3. Find $x_i = \{x | x \in \chi$, and for some $f \in F$ the set of x's achieving the supremum $c_i\}$.
4. Set $\alpha_i = 1/(i+s), \quad s \geq 0$.
5. Let
$$\xi_{i+1} = (1 - \alpha_i)\xi_i + \alpha_i \xi_{x_i}$$

be the new design measure (where ξ_{x_i} places measure one at point x_i)

6. If the difference between two consecutive resulting c_i is sufficiently small then stop. Otherwise, set $i = i + 1$ and go to step 2.

Note that the sequence α_i as specified above will not in general lead to monotonically decreasing c_i.

In the next section, we apply these types of robust designs to various examples of the generalized linear model. Although the algorithm seems to converge in all cases considered in the next section, it lacks a checking condition to verify that the design is really robust. Nevertheless, this approach is heavily favored over other approaches suggested by the first author, Kim(1993) in his construction of the error-robust design.

4. Examples

Example 1. As stated before, for normal distribution, $\omega(\eta)$ is equal to 1. Suppose we have two different models in consideration, $F = [f_1, f_2]$ where $f_1^T(x) = (1, x)$ and $f_2^T(x) = (1, x, x^2)$. This classic problem is a common example referenced in many literatures. For this type of model space, a model-robust design has the following mass in the design space $\chi = [-1, 1]$:

$$\xi(-1) = \xi(1) = 0.3604, \qquad \xi(0) = 0.2792.$$

The same worst G-efficiency is obtained for both models, 0.838. In comparison, the G-efficiency of the quadratic regression D-optimal design for f_1 is only 0.80.

Example 2. For binary data with logistic link, $\eta = log\{\mu/(1-\mu)\}$ the efficiency function $\omega(\eta)$ turns out to be $\mu(1-\mu)$, where $\mu = exp(\eta)/\{1 + exp(\eta)\}$ or $\omega(\eta) = exp(\eta)/\{1 + exp(\eta)\}^2$. We consider only simple design problems in design space $\chi = [-1, 1]$ in forms of $f^T(x) = (1, x)$ and $\beta^T = (\beta_0, \beta_1)$. The D-optimum design is concentrated at two points, and we will therefore put equal weight at these two points. See Silvey(1980) for further details on analytical solution with regard to the relationship between β_0 and β_1. The following design is found to be D-optimal design for $\beta_0 = 1$ and $\beta_1 = 1$:

$$\xi(-1) = \xi(1) = 1/2.$$

The values of β are taken for simplicity. This design is the same as the usual D-optimal design for simple linear regression. Since the D-optimal designs depend on the a priori specified values of β, we notice the effect of the slope parameter β_1 on the D-optimal design, as the value of β_1 increases. The following is the D-optimum design for $\beta_0 = 1$ and $\beta_1 = 2$:

$$\xi(-1) = \xi(0.380) = 1/2.$$

Obviously, the resulting design is different from the one found above. Therefore, we naturally ask ourselves what the parameter-robust design would be if $B = [\beta_1^T, \beta_2^T, \beta_3^T]$ where $\beta_1^T = (1,1)$, $\beta_2^T = (1,2)$, and $\beta_3^T = (1,3)$. Note that $\xi(-0.849) = \xi(0.179) = 1/2$ is the D-optimum design for $\beta_0 = 1$ and $\beta_1 = 3$. The computer algorithm found the following design to be the parameter-robust design:

$$\xi(1) = 0.269, \;\; \xi(-1) = 0.313, \;\; \xi(0) = 0.316, \;\; \xi(-0.758) = 0.101.$$

Note that unlike the design in Example 1, this does not allocate the mass symmetrically around the middle points. It spreads out the design points throughout the design space taking into consideration of partial information about β. This phenomenon is quite useful for the analysis of checking the model's adequacy at a later stage. Silvey(1980) noted in his book that usual end point design $\xi(\pm 1) = 1/2$ produces extremely low efficiency and thus it is not even worth trying to calculate the max-min design. But some partial information about β gives us a chance to improve the efficiency greatly. The "pseudo" G-efficiency is as much as 0.825. Note that the accompanying D-efficiency will be higher than this "pseudo" G-efficiency. Of course, there is another approach available, the Bayesian approach, which is to put some prior distribution on the unknown parameters. The Bayesian approach has many excellent features such as a condition to check the Bayesian D-optimality of a given design, which is not provided by the robust design suggested here.

It still does give us many intuitive results and furthermore, it is far easier than the Bayesian approach. A set of reasonable values of parameter vector is not difficult to get when this approach is underway as part of sequential or follow-up design.

Example 3. To compare with Example 1, we take two different forms of model for binary data with logistic link. In other words, $F = [f_1, f_2]$, where $f_1^T = (1, x)$ and $f_2^T = (1, x, x^2)$ The values of beta are taken as $\beta_0 = 1, \beta_1 = 1$ and $\beta_2 = 1$ without loss of generality. The following two designs are D-optimum designs for each f in F respectively:

$$\xi_{f_1}(-1) = \xi_{f_1}(1) = 1/2$$

$$\xi_{f_2}(-1) = \xi_{f_2}(-0.101) = \xi_{f_2}(1) = 1/3.$$

Note that unlike the normal theory quadratic regression design in Example 1, D-optimum design for f_2 does not pick up the middle design point at 0. The middle design point for quadratic logistic regression in this case does move away in negative direction from 0. This might have been caused by the nonlinear relationship between mean and variance function of the binomial distribution. The following is found to be the model-robust design for this case:

$$\xi(1) = 0.332, \ \ \xi(-1) = 0.380, \ \ \xi(0) = 0.287.$$

The worst "pseudo" G-efficiency is as much as 0.8624. This value is higher than that achieved in Example 1, which is very surpising. Before this experiment, it was expected that the "pseudo" efficiency would get lower than one of normal theory due to the nonlinear relationship between the mean and variance of binomial distribution.

Example 4. When dealing with the binary data there is some uncertainty regarding the choice of the link function. The most popular choice would be, of course, logistic link. But there is no other reason not to choose other ones such as probit $g(\mu) = \Phi^{-1}(\mu)$, where Φ is the normal cumulative distribution function or complementary log-log link, $g(\mu) = log\{-log(1-\mu)\}$ for connecting the mean and linear predictor models. Here, we consider a situation in which we have to choose between the logistic and complementary log-log link. Like Example 2, we take only simple forms of $f^T(x) = (1, x)$ and $\beta^T = (\beta_0, \beta_1)$.

The D-optimum design is concentrated at two points for both links and we will therefore put equal weight at these two points. The following design turns out to be D-optimal design for complementary log-log link in case of $\beta_0 = 1$ and $\beta_1 = 1$:

$$\xi(-1) = \xi(0.244) = 1/2.$$

The efficiency function $\omega(\eta)$ turns out to take the form of $\{1-\mu\}/\mu \cdot log^2(1-\mu)$ or $\omega(\eta) = exp\{2\eta - exp(\eta)\}/[1-exp\{-exp(\eta)\}]$. Although it is known that two links have very common properties in most binary data analysis, the resulting designs are not the same. The following design is found to be link-robust design which has three points with unequal mass. The "pseudo" G-efficiency is as much as 0.847:

$$\xi(1) = 0.203, \ \ \xi(-1) = 0.419, \ \ \xi(0.240) = 0.378.$$

Of course, we can extend the example beyond the normal or binary distributions. If we assume Poisson regression model, the possible links would be identity link for simplicity or log link to preserve the natural characteristics of the non-negative random variable. This kind of uncertainty surrounding the choice of link functions must be carefully taken care of.

5. Conclusions

In this paper we prosposed a robust design criterion similar to the one suggested by Kim(1993) for the generalized linear regression. The GLM has very similar features as the usual linear regression model except for the "pseudo" efficiency function being attached to the model. It is not hard to implement the well-developed computer algorithms to this more broad line of models. Although it is relatively easy to implement, this does not mean that it is easy to apply the optimal design criteria of linear model directly to the GLM since the design criteria depend on all three components of the GLM. Therefore, we proposed the minimax one. Fortunately, it has been very successful for each class of uncertainty we have dealt with. Although we admit that it lacks the equivalence theorem, the computer algorithm we have suggested has been successful in finding the nearly robust design. Other approaches such as Läuter's (1974) can be implemented and other examples can be demonstrated to propose another design criteria. All these things need further research.

References

1. Abdelbasti, K. M. and Plackett, R. L. (1983). Experiemntal design for binary data, *Journal of the American Statistical Association,* 78, 90-98.

2. Atwood, C. L. (1969). Optimal and efficient designs of experiments, *Annals of Mathematical Statistics,* 40, 1570-1602.

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

4. Fedorov, V. V. (1972). *Theory of Optimal Experiments,* Academic Press, New York.

5. Ford, I. Titterington, D. M., and Kitsos, C. P. (1989). Recent advances in nonlinear experimental design, *Technometrics,* 31, 49-60.

6. Ford, I. Torsney, B., and Wu, C. F. J. (1992). The use of a canonical form in the construction of locally optimal designs for non-linear problems, *Journal of the Royal Statistical Society,* Ser. B., 54, 569-583.

7. Kiefer, J. and Wolfowitz, J. (1959). Optimal bayesian design applied to logistic regression experiments, *Canadian Journal of Mathematics,* 12, 363-366.

8. Kim, Y. I. (1993). Error-robust experimental designs: D- and Heteroscedastic G-optimalities, *The Korean Journal of Applied Statistics,* 6, 303-309.

9. Läuter, E. (1974). Experimental planning in a class of models, *Mathematische Operationsforschung und Statistik,* 5, 673-708.

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

11. Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models, *Journal of the Royal Statistical Society,* Ser. A.,135, 370-384.

12. Sebastiani. P. and Settimi, R. (1997). A note on D-optimal designs for a logistic regression model, *Journal of Statistical Planning and Inference,* 59, 359-368.

13. Silvey, S. D. (1980). *Optimal Design,* Chapman and Hall, London.

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