Generalized Linear Models and Response Transformation

Anthony C. Atkinson
The London School of Economics, London WC2A 2AE, UK
a.c.atkinson@lse.ac.uk

May 28, 2004

Abstract

Response surface designs are found for models in which there is a mean variance-relationship. This may arise because generalized linear models are appropriate or because the response requires transformation to have a constant variance. The special problem of design for nonlinear response surface models is resolved through the use of structured parameters.

Keywords: Binary data; Box-Cox transformation; Gamma models; Horwitz' rule; Link function; Nonlinear least squares; Optimum design; Structured parameters

1 Introduction

Response surface models are used when the response y is a smooth function of parameters and explanatory variables. We write

$$y = h(x, \theta, \epsilon),$$

where y is a response, usually univariate, x is a vector of m explanatory variables or covariates and ϵ is an error term. In most of the large literature on designs for response surfaces from Box and Draper (1963) to Myers et al. (2004) it is assumed that the model for the ith observation can be written in the simpler form

$$y_i = \eta(x_i, \theta) + \epsilon_i, \qquad (i = 1, \ldots, n) \qquad (1)$$

where the additive errors ϵ_i are independent and identically distributed with constant variance σ^2. As a consequence, the variance does not depend on the mean of y_i. This paper is concerned with situations where this simple assumption does not hold and the variance is indeed a function of the mean.

In general, small observations have a smaller variance than large observations. In particular, with non-negative observations such as measurements of chemical concentrations, the variance increases with the mean. Two examples are mentioned in §7.1. Often the variance can be made independent of the mean by a suitable transformation, such as the family of power transformations introduced by Box and Cox (1964). The consequences of such transformations for experimental design are the subject of the latter part of the paper. In the earlier sections we consider the design of experiments for generalized linear models, when the form of the mean-variance relationship is a consequence of the assumed model.

The paper begins in §2 with a brief introduction to generalized linear models and to the theory of optimum experimental design. In §3 this theory is applied to designs for gamma models, which can be identical to standard response surface designs, and, in §4, to designs for binary responses, which are usually very different. The efficiency of standard designs for generalized linear models is investigated in §5.

Optimum designs for normal-theory response models with parameterized variance functions are derived in §6. A special case is when the variance is a function of the mean. An alternative to weighted regression to allow for these relationships is to transform the response. The effect of transformation on design for nonlinear models is indicated in §7.3. Response surface designs for estimation of the transformation parameter are described in §8. The situation is however different if the transformation is known. Then designs to estimate the parameters β of the standard polynomial response surface model (1) do not depend on the transformation that is to be applied in data analysis. However, the transformation does effect the design for the parameters of the nonlinear response surface models that are the subject of §9.

The description of response surface models at the beginning of this section is clearly over-simplified. For example, in addition to the continues variables x there may be blocking variables, or some of the variables x may be concomitant variables that are not experimentally controllable. The paper concludes in §10 with a brief discussion of some omitted topics.

2 Theory

2.1 Generalized Linear Models

In addition to the additive errors assumed in (1) it is customary to assume that the response surface model is linear in the parameters so that

$$y_i = \eta_i + \epsilon_i = f^T(x_i)\beta + \epsilon_i, \tag{2}$$

where $f(x)$ is a $p \times 1$ vector of powers and products of the m explanatory variables. Often the parameter sensitivities $f(x)$ will be the terms of a second-order polynomial, which can be thought of as coming from Taylor series expansion of the nonlinear model (1). Of course, both the assumed constant variance of the ϵ_i and the adequacy of the linear model $f(x)$ need to be checked against the data.

If the errors ϵ_i are normally distributed, (2) defines a linear multiple regression model which can be written as

$$\mathrm{E}(y) = \mu = \eta = f^T(x)\beta, \tag{3}$$

where μ, the mean of y, is equal to the linear predictor η. The generalized linear model extends (3) to any distribution belonging to the one-parameter exponential family. As well as the normal, this includes the gamma, Poisson and binomial distributions. A second extension is the introduction of a link function $g(\mu) = \eta$, relating the mean and the linear predictor. For the binomial data of §4 the link function is such that, however the values of x and β vary, the mean μ satisfies the physically meaningful constraint that $0 \leq \mu \leq 1$. The seminal work on generalized linear models is McCullagh and Nelder (1989). Myers, Montgomery, and Vining (2001) emphasize data analysis. Atkinson and Riani (2000, Chapter 6) gives a succinct introduction.

We are particularly interested in the relationship between the mean and the variance of the observations. Let the variance of y be $\mathrm{var}(y)$. Then

$$\mathrm{var}(y) = \phi V(\mu),$$

where ϕ is the dispersion parameter, equal to σ^2 for the normal distribution and one for the binomial. The variance function $V(\mu)$, which is specific to the error distribution, then determines the relationship between mean and variance.

Optimum experimental design for generalized linear models is aided by the simple form of the information matrix. Maximum likelihood estimation of the parameters β of the linear predictor reduces to iterative weighted least

squares, with the weights for individual observations given by

$$w = V^{-1}(\mu)\left(\frac{d\mu}{d\eta}\right)^2. \tag{4}$$

It follows that the weights depend both on the distribution of y and on the link function.

2.2 Optimum Experimental Design

The algorithms of optimum experimental design provide a straightforward way of constructing designs for generalized linear models, which can then be compared with regression designs. Optimum experimental design is described in the books of Pukelsheim (1993) and of Atkinson and Donev (1992), where designs for generalized linear models are discussed in §22.5.

For the nonlinear regression model (1), the parameter sensitivity for observation i is the $p \times 1$ vector

$$f(x_i, \theta) = \left\{\frac{\partial \eta(x_i, \theta)}{\partial \theta_j}\right\}. \qquad (j = 1, \ldots, p) \tag{5}$$

These sensitivities reduce to those introduced in (2) for the linear regression model.

The contribution of observation i to the information matrix for weighted least squares estimation of the parameters is proportional to $f(x_i, \theta) w_i f^T(x_i, \theta)$. Experimental designs are sought which make large suitable functions of this matrix summed over all observations. The dependence of the information on the number of experimental trials n can be removed by standardization, working instead with the fractions of trials $p_i = n_i/n$ at each of the k support points x_i of the design. More general mathematical results about designs are obtained on replacing the fractions by a measure ξ, yielding a 'continuous' design in which the n_i are no longer required to be integer. However practical designs do require that the n_i be integer.

For an experimental design represented as a measure ξ over the design region $\mathcal{X}$, the information matrix is

$$M(\xi, \theta) = \int_{\mathcal{X}} f(x, \theta) w f^T(x, \theta) \xi(dx) = E_\xi f(x) w f^T(x, \theta) = E_\xi I(x, \theta),$$

which is a function of θ. Optimum design theory is concerned with minimization of the convex function $\Psi\{M(\xi, \theta)\}$. Because this is a well-behaved optimization problem, the optimality of any design ξ can be checked by using the directional derivative $\phi(x, \xi, \theta)$. For the optimum design ξ^*, $\phi(x, \xi^*, \theta)$ is

zero at the design points and positive elsewhere. The equivalence theorems for various criteria follow from this result.

One frequently used design criterion, and that used in this chapter, is D-optimality in which the generalized variance of the parameter estimates is minimized, when

$$\Psi\{M(\xi,\theta)\} = -\log|M(\xi,\theta)|,$$

equivalent to maximizing $|M(\xi,\theta)|$. The equivalence theorem relates D-optimality to the variance of the prediction $\hat{y}(x)$ for the design ξ, standardized by taking the error variance as one. Let

$$d(x,\xi,\theta) = wf^T(x,\theta)M^{-1}(\xi,\theta)f(x,\theta). \tag{6}$$

Then the derivative $\phi(x,\xi,\theta) = p - d(x,\xi,\theta)$ and the maximum value, p, of $d(x,\xi^*,\theta)$ occurs at the points of support of the design. This condition provides a method of checking the optimality of a purported optimum design.

D-optimum designs for the linear regression model (2) with additive independent errors of constant variance do not depend on the value of β. However, the designs found in this chapter depend not only on the model and on the design region $\mathcal{X}$, but also on the value of the parameter θ. We find locally optimum designs, that is designs that are optimum for a specified value θ_0 of θ. Bayesian optimum designs where θ has a prior distribution are briefly mentioned in §10.

3 Optimum Designs for Gamma Models

The gamma model is often an alternative to response transformation. In particular, with a log link, it is often hard to distinguish the gamma from a linear regression model with logged response. Examples are given in §7.5 of Myers et al. (2001).

A useful, flexible family of links is the Box and Cox family, in which

$$g(\mu) = (\mu^\lambda - 1)/\lambda = \eta. \tag{7}$$

This is equivalent to the power family of links $g(\mu) = \mu^\lambda$ but is continuous as $\lambda \to 0$, yielding the log link. Differentiation of (7) yields

$$\frac{d\eta}{d\mu} = \mu^{\lambda-1}. \tag{8}$$

Since the variance function for the gamma distribution is

$$V(\mu) = \mu^2,$$

the combination of (4) and (8) shows that the weights for the gamma distribution with this link family are

$$w = V^{-1}(\mu)\left(\frac{d\mu}{d\eta}\right)^2 = \mu^{-2\lambda}. \tag{9}$$

When $\lambda = 0$, that is for the log link, (9) shows that the weights are equal to one. It therefore follows that optimum designs for gamma models with this link are identical to optimum designs for the regression model (2), of course when the linear predictors η are the same. Comparisons of the analyses of data from designed experiments when the data are transformed or analysed using a gamma model with log link are given by Lewis et al. (2001).

To find designs that illustrate the difference between regression and the gamma GLM we need to use a value of $\lambda \neq 0$. Burridge and Sebastiani (1994) find optimum designs when the linear predictor is first order and $\lambda > 0$. They establish conditions on the values of the parameters β under which two-level factorial designs are optimum and other conditions when the optimum designs have only factor at a time not at the lower level. Here we investigate designs for an example when the linear predictor is second-order and $\lambda \neq 0$.

Atkinson and Riani (2000, §6.9) use a gamma model to analyse data from Nelson (1981) on the degradation of insulation due to elevated temperature at a series of times. The data do not yield a particularly clean model as there seem to be some identifiable subsets which do not completely agree with the fitted response-surface model. However, for our purposes, a second-order model is required in the two continuous variables and the gamma model fits best with a power link with $\lambda = 0.5$. A theoretical difficulty with such a value of λ is that μ must be > 0, while η is, in principle, unconstrained.

We scale the variables so that the design region $\mathcal{X}$ is the unit square with vertices $(-1,-1)$, $(-1,1)$, $(1,-1)$ and $(1,1)$. The linear predictor is the quadratic

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2. \tag{10}$$

Then the standard D-optimum design for the normal theory regression model,given in Table 6 has unequally weighted support at the points of the 3^2 factorial. This design is, from what was said above, also optimum for the gamma model with log link.

For other links the design will depend on the actual values of the parameters β in (10). Any design found will therefore be locally optimum. With $\lambda = 0.5$, it follows from (9) that the weights

$$w_i = 1/\mu_i.$$

Table 1: D-optimum designs for two gamma models: parameter values for linear predictors, second-order in two variables. Power link, $\lambda = 0.5$

Design	β_0	β_1	β_2	β_{11}	β_{22}	β_{12}
G1	3.7	-0.46	-0.65	-0.19	-0.45	-0.57
G2	3.7	-0.23	-0.325	-0.095	-0.225	-0.285

Table 2: D-optimum designs for the parameter sets G1 and G2 of Table 1

Design	Design G1				Design G2			
Point i	x_{1i}	x_{2i}	p_i	μ_i	x_{1i}	x_{2i}	p_i	μ_i
1	-1.00	-1.00	0.163	12.96	-1.00	-1.00	0.149	13.32
2	-1.00	0.05	0.015	15.72	-1.00	0.00	0.057	14.71
3	-1.00	1.00	0.162	11.83	-1.00	1.00	0.150	12.74
4	0.05	-1.00	0.002	15.25	0.00	-1.00	0.053	14.44
5	0.15	0.15	0.059	12.29	0.05	0.05	0.086	13.47
6	0.25	1.00	0.145	5.43	0.10	1.00	0.108	9.59
7	1.00	-1.00	0.163	14.59	1.00	-1.00	0.149	14.14
8	1.00	0.25	0.130	7.38	1.00	0.10	0.097	10.97
9	1.00	1.00	0.161	1.90	1.00	1.00	0.151	6.45

We take β to have the values given in Table 1, G1 being rounded from an analysis of Nelson's data.

The optimum design for G1 is in Table 2. This shows that, at the points of the design, the minimum value of μ is 1.90 and the maximum 15.72. If these were normal responses that had to be non-negative, this kind of range would indicate a power transformation.

The optimum design was found by a grid search with steps of 0.05 in x_1 and x_2. For each set of m support points the optimum design weights p_i were found by numerical search using the transformation to $m-1$-dimensional polar co-ordinates described by Atkinson and Donev (1992, p.104). The variance of prediction $d(x, \xi, \theta)$ (6) was calculated at each grid point and the point with the maximum value added to the design, whilst one point was

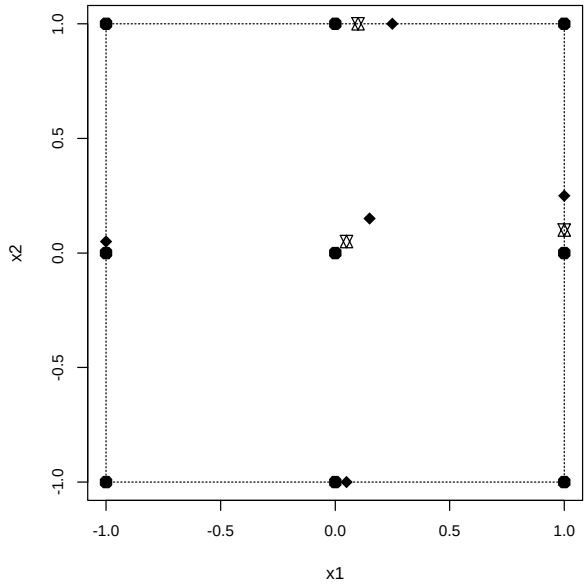

Figure 1: Support points for D-optimum designs for gamma models in Table 2: •, the points of the 3^2 factorial; rhombus, G1 and star, G2

deleted. The process continued until the point to be added was already in the design.

As Table 2 and Figure 1 show, the support points of the design are a slight, and symmetrical, distortion of those of the 3^2 factorial. Although the support points are symmetrical, the design weights are not, with a minimum weight of 0.002 on design point 4.

There are two sets of conditions under which the optimum design for the gamma model is identical to the optimum design for the second-order normal theory response surface model that has support at the points of the 3^2 factorial. One occurs when the log link is appropriate, whatever the values of the parameters β in the linear predictor. The other is when the effects of the variables are small, that is when the parameters in the linear predictor, other than β_0, tend to zero (Cox 1988). To illustrate this point we also found the D-optimum design for the set of parameter values G2 in Table 1 in which all parameters, other than β_0, have half the values they have for design G1. As Table 2 shows, the range of means at the design points is now 6.45 to 14.71, an appreciable reduction in the ratio of largest to smallest response. The support points of the design are shown in Figure 1 by open stars: all are moved in the direction of the factorial design when compared to the points of support of G1. The results in Table 2 show that the optimum design weights are now more nearly even, with a minimum value of 0.053 to be compared with the previous minimum of 0.002.

The relationship between these two-variable response surface designs for gamma models and the normal theory response surface design, is explored further in §5 where we calculate the efficiency of standard designs, including the 3^2 factorial, for our non-standard problem.

4 Models and Designs for Binomial Data

4.1 One Variable

If designs for gamma models seem much like those for normal theory models, the same is not true for designs for binary data.

The greatest interest in the literature (Chaloner and Larntz 1989) has been in logistic models for binary data, particularly with one explanatory variable. This linear logistic model is that

$$\log\{\mu/(1-\mu)\} = \eta = \alpha + \beta x, \tag{11}$$

where μ is the probability of failure, often death in, for example, insecticide studies. Other links include the probit, the complementary loglog and arcsine, which has good robustness properties. Discussion of these links and examples of their use in the analysis of data can be found in Atkinson and Riani (2000, §§6.13-6.20).

For the logistic link in (11) the variance function is

$$V(\mu) = \mu(1-\mu).$$

Differentiation of the logistic link yields

$$\frac{d\mu}{d\eta} = \mu(1-\mu).$$

The iterative weight, for this link, is thus

$$w = \mu(1-\mu).$$

As might be expected, experiments with μ near to zero or one are thus non-informative; a set of nearly all successes or failures does not provide good parameter estimates.

For a sufficiently large design region $\mathcal{X}$, the D-optimum design for $\alpha = 0, \beta = 1$ puts half the trials at $x = -1.543$ and half at $x = 1.543$. This is a special case of the result that p-point D-optimum designs for models with p parameters put weight $1/p$ at each design point. As Figure 1 shows,

Table 3: D-optimum designs for four binomial models: parameter values for linear predictors, first-order in two variables. Logistic link

Design	β_0	β_1	β_2
B1	0	1	1
B2	0	2	2
B3	2	2	2
B4	2.5	2	2

D-optimum designs may have more than this minimum number of support points; for the designs in the figure $p = 6$ and $k = 9$.

Although designs for other values of the parameters can likewise be found numerically, design problems for a single x can often be solved in a canonical form, yielding a structure for the designs independent of the particular parameter values (Ford, Torsney, and Wu 1992). The D-optimum design in this problem puts the trials where the expected response is 0.176 and 0.824, the translation into the experimental variable x depending on the values of α and β.

4.2 Response Surface Designs

The properties of designs for response surface models, that is with two or more explanatory variables, depend much more on the experimental region than those where there is only one factor.

Although it was assumed in the previous section that the experimental region $\mathcal{X}$ was effectively unbounded, the design was constrained by the weight w to lie in a region in which μ was not too close to zero or one. But with more than one explanatory variable constraints on the region are necessary. For example, for the two variable model

$$\log\{\mu/(1-\mu)\} = \eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2, \tag{12}$$

with $\beta^T = (0, \gamma, \gamma)$, all points for which $x_1 + x_2 = 0$ yield a value of 0.5 for μ, however extreme the values of x. We now explore designs for the linear predictor (12) with the logistic link for a variety of parameter values.

Four sets of parameter values are given in Table 3. In all cases we take the same design region as before, that is $\mathcal{X}$ is the square with vertices ± 1. The D-optimum designs for the sets B1 and B2 are listed in Table 4. The

Table 4: D-optimum designs for binomial models with the parameter sets B1 and B2 of Table 3

Design	Design B1				Design B2			
Point i	x_{1i}	x_{2i}	p_i	μ_i	x_{1i}	x_{2i}	p_i	μ_i
1	1	-1	0.296	0.500	-1.0	1.0	0.251	0.500
2	-1	-1	0.204	0.119	-1.0	0.1	0.142	0.142
3	1	1	0.204	0.881	-0.1	1.0	0.142	0.858
4	-1	1	0.296	0.500	0.1	-1.0	0.095	0.142
5					1.0	-0.1	0.095	0.858
6					1.0	-1.0	0.275	0.500

values of B1 $(0, 1, 1)$ are smallest and the table shows that the design has support at the points of the 2^2 factorial, although the design weights are not quite equal, as they would be for the normal theory model and also for the logistic model as β_1 and $\beta_2 \to 0$. At those factorial points for which $x_1 + x_2 = 0, \mu = 0.5$ since $\beta_1 = \beta_2$. At the other design points $\mu = 0.119$ and 0.881, slightly more extreme values than the values of 0.176 and 0.824 for the experiment with a single x.

A most interesting feature of our example is that the number of support points of the design depends upon the values of the parameters β. From Carathéodory's Theorem (Silvey 1980, Appendix 2), the maximum number of support points required by an optimum design is $p(p+1)/2$. Our second set of parameters, B2 in which $\beta^T = (0, 2, 2)$, gives an optimum design with six support points, that is the value of this bound when $p = 3$. These points are places where $\mu = 0.142, 0.5$ and 0.858.

The relationship between the support points of the design and the values of μ is highlighted in Figure 2 where the pale areas are regions in which $\mu \leq 0.15$, with the dark regions the complementary ones where $\mu \geq 0.85$. Apart from the design points where $\mu = 0.5$, all other design points are close to those boundaries of these regions where μ is around 0.15 and 0.85.

The D-optimum designs for the two remaining sets of parameters in Table 3 are given in Table 5. These designs have respectively 4 and 3 points of support. When $\beta^T = (2, 2, 2)$, the design points are where $\mu = 0.182$ and 0.818. For $\beta^T = (2.5, 2, 2)$ the values are 0.182 and 0.832. For this three-point design for a three parameter model, the design weights $p_i = 1/3$.

The relationship between the design points and the values of μ are shown,

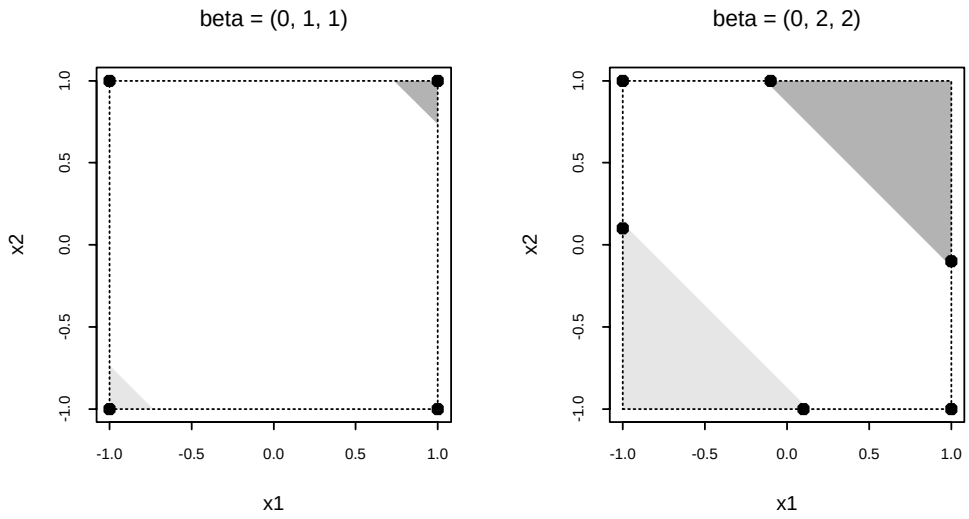

Figure 2: Support points for D-optimum designs for binomial models B1 and B2 in Table 3. In the lightly shaded area $\mu \leq 0.15$, whereas, in the darker region, $\mu \geq 0.85$

Table 5: D-optimum designs for binomial models with the parameter sets B3 and B4 of Table 3

Design	Design B3				Design B4			
Point i	x_{1i}	x_{2i}	p_i	μ_i	x_{1i}	x_{2i}	p_i	μ_i
1	-1.00	-0.75	0.168	0.182	-1.00	0.55	0.333	0.832
2	-1.00	0.75	0.332	0.818	-1.00	-1.00	0.333	0.182
3	-0.75	-1.00	0.168	0.182	0.55	-1.00	0.333	0.832
4	0.75	-1.00	0.332	0.818				

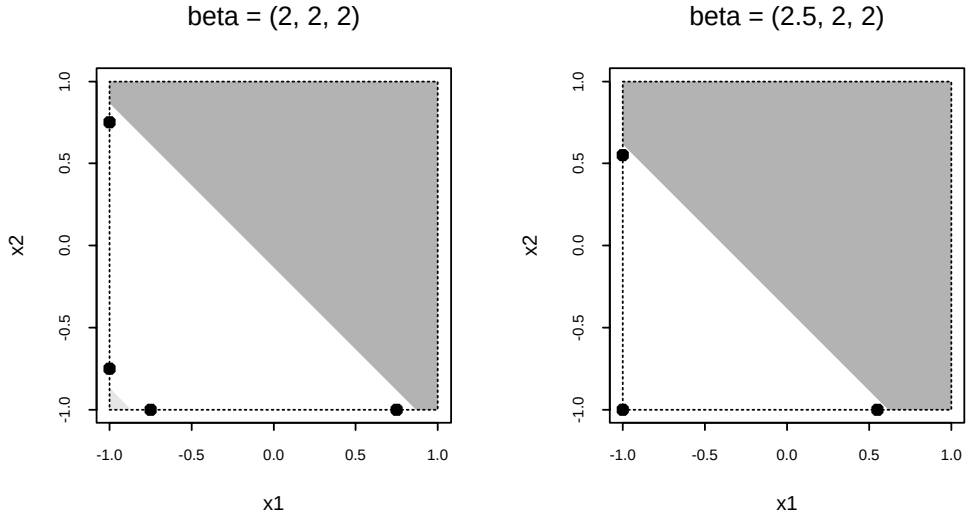

Figure 3: Support points for D-optimum designs for binomial models B3 and B4 in Table 3. In the lightly shaded area $\mu \leq 0.15$, whereas, in the darker region, $\mu \geq 0.85$

for these designs, in Figure 3. For $\beta^T = (2, 2, 2)$ the design points lie slightly away from the boundaries of the regions of high and low values of μ, as they do to a lesser extent in the right-hand panel of the figure. With $\beta^T = (2.5, 2, 2)$ the minimum value of μ, 0.182 at $(-1, -1)$, is sufficiently high that there are no experimental conditions for which $\mu = 0.15$: the panel of the figure contains only one shaded area.

The properties of some designs can be explained by the estimation procedure for generalized linear models. Because weighted least squares is used, design for the logistic model (12) is equivalent to design for the linear model

$$\eta = \beta_0\sqrt{w} + \beta_1\sqrt{w}x_1 + \beta_2\sqrt{w}x_2, = \beta_0 z_0 + \beta_1 z_1 + \beta_2 z_2. \qquad (13)$$

The design region $\mathcal{X}$ is then replaced by the induced design region $\mathcal{Z}$, the space in which the values of z can fall as x varies. Since $p = 3$, the induced design space $\mathcal{Z}$ is of dimension three. Two examples, projected onto z_1 and z_2 and so ignoring $z_0 = \sqrt{w}$, are given in Figure 4 for $\mathcal{X}$ the unit square. In the left-hand panel of the figure $\beta^T = (0, 2, 2)$ so that at the corner of $\mathcal{X}$ for which $x_1 = x_2 = 1$, $\eta = 4$ and $\mu = 0.982$. This is well beyond the range for informative experiments and the projection of the induced design space appears to be folded over. As a consequence, experiments at extreme positions in $\mathcal{Z}$ are not at extreme points in $\mathcal{X}$. The results in the other panel for $\beta^T = (2, 2, 2)$ are similar, but more extreme. For both sets of parameter values the design points lie, as they should, on the boundary of the design region.

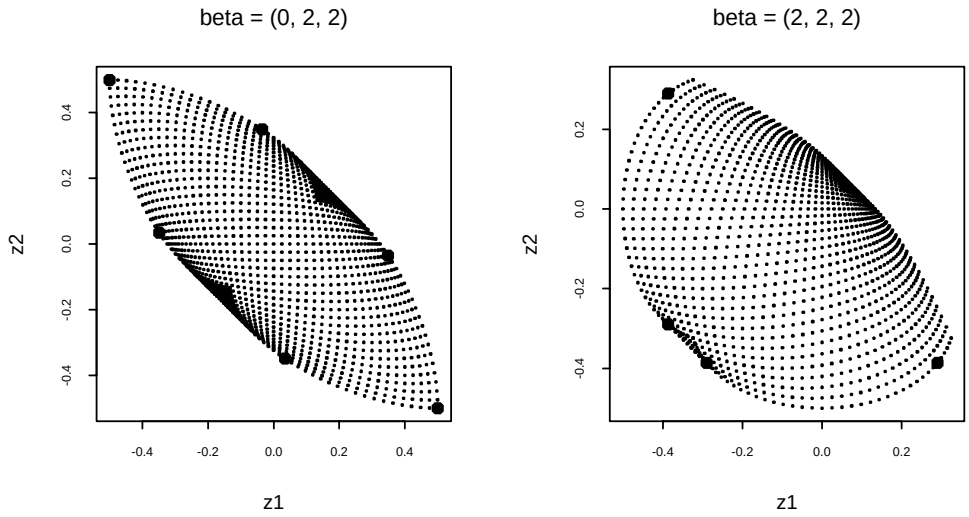

Figure 4: Support points for D-optimum designs for binomial models B2 and B3 in Table 3 plotted in the induced design region $\mathcal{Z}$.

The extension of the procedure based on (13) to second-order models such as (10) is not obvious. The difficulty is the way in which the weights enter in the transformation from $\mathcal{X}$ to $\mathcal{Z}$. If, as in (13), $z_j = \sqrt{w}x_j$, then, for example, the interaction term in the linear predictor $\sqrt{w}x_jx_k$ is not equal to z_jz_k. Thus the technique of Burridge and Sebastiani (1994) which relies on (13) cannot be applied to find the structure of optimum second-order designs.

These examples show the importance of both the design region and the value of μ in determining the optimum design. In order to reveal the structure of the designs as clearly as possible, the designs considered have all had $\beta_1 = \beta_2$, and so are symmetrical in x_1 and x_2. Both the design region and the values of μ are equally important in the asymmetric designs when the two parameter values are not equal. Asymmetric designs also arise with the log log and complementary log log links, since these links are not symmetrical.

In the next section we see how well our symmetrical designs for the logistic link can be approximated by standard designs such as the 2^2 factorial.

5 Standard Designs and GLM's

5.1 Gamma Models

The designs for second-order gamma models with the parameter sets G1 and G2 of Table 1 are both slight distortions of the 3^2 factorial. As the values of the parameters, apart from β_0, tend to zero, the design tends towards the

Table 6: D-optimum designs for two-factor normal theory response surface model with design region the unit square

x_{1i}	-1	-1	-1	0	0	0	1	1	1
x_{2i}	-1	0	1	-1	0	1	-1	0	1
p_i	p^1	p^2	p^1	p^2	p^3	p^2	p^1	p^2	p^1
where	$p^1 = 0.1458$			$p^2 = 0.0801$			$p^3 = 0.0962$		

D-optimum design for the second-order regression model. This design, given in Table 6 with minimal rounding of the weights, has unequal support at the points of the 3^2 factorial, the weights depending on the number of non-zero co-ordinates of the design point (Farrell et al. 1968; Atkinson and Donev 1992, p.130). The simplest design with this support is the 3^2 factorial in which all weights are equal to 1/9.

We compare three designs for their efficiency relative to the D-optimum design of Table 2, that is the design for the more extreme parameter set G1 of Table 1. To calculate the efficiency of the designs let the determinant of the information matrix for the optimum design be $|M(\xi^*)|$ and, for some other design be $|M(\xi)|$. Then the efficiency is

$$D_{\text{eff}} = \{|M(\xi)|/|M(\xi^*)|\}^{1/p} . \tag{14}$$

with $p = 6$. The three designs for comparison are: the D-optimum design for the less extreme parameter set G2, also given in Table 2, the D-optimum regression design of Table 6 and the equi-replicated 3^2 factorial. The resulting efficiencies are in Table 7

The main feature of these designs is how efficient they are for the gamma model, all efficiencies being greater than 90%. The designs in the table are ordered by their closeness to that for the gamma model. The design for parameter G2 is for a model with smaller effects than G1, so that the design is between that for G1 and the design for the normal theory model. The equi-weighted 3^2 factorial is furthest from the other designs, which all put greatest weight on the extreme points of the design region, that is at the points of the 2^2 factorial.

Table 7: Percentage efficiencies D_{eff} of approximations to the D-optimum design for the second-order response surface gamma model with parameters G1 of Table 1

D-optimum design G2, Table 2	98.02
D-optimum normal theory, Table 6	95.29
Unweighted 3^2 factorial	92.25

Table 8: Percentage efficiencies D_{eff} of two approximations to the D-optimum design for the first-order response surface binomial model. The design in which only one factor at a time is changed has three support points

β	2^2 Factorial	One Factor
(0, 1, 1)	98.74	92.97
(0, 2, 2)	76.56	79.19
(2, 2, 2)	71.28	92.90
(2.5, 2, 2)	67.66	88.68

5.2 Binomial Models

An indication of the example with a gamma model is that standard designs may be satisfactory for gamma models. The same conclusion does not hold for our binomial example.

Efficiencies were calculated for the four parameter sets of Table 3 for two standard designs. One was the 2^2 factorial with weight 1/4 at the four design points $(-1,-1),(-1,1),(1,-1)$ and (1,1). The other design puts weight 1/3 at the first three of these support points. This "one factor at a time" design excludes the treatment combination when both factors are simultaneously at their high level. The efficiencies are calculated as in (14), but now $p = 3$.

The eight efficiencies in Table 8 are readily interpreted with the help of Figures 2 and 3. For $\beta^T = (0,1,1)$ the optimum design is supported on the points of the 2^2 factorial. The unweighted 2^2 factorial has an efficiency of 98.74% for this parameter value, with the one factor at a time design, which has support at three of the four points of the optimum design, having a slightly lower efficiency of 92.97%. When $\beta^T = (0,2,2)$, the two standard

designs share only two points of support with the optimum design. They have efficiencies of 76.56 and 79.19%. The efficiency for the one factor at a time design is slightly higher since it puts weight 2/3 on the common points, as opposed to 0.5 for the factorial design. For the other two designs, the one-factor design provides the more efficient approximation. This is because the support of both optimum designs is such that $x_1 + x_2 \leq 0$, a condition also satisfied by the points of the one-factor design. The unique point of the 2^2 factorial, that is $(1, 1)$, is virtually non-informative, so that one quarter of the experimental effort is wasted. Indeed, the efficiencies of the factorial design for $\beta^T = (2, 2, 2)$ and (2.5,2,2) are both less than 75%. These values are to be compared with 92.9 and 88.68% for the one-factor design.

The results for the gama and binomial models have something in common. When the parameter effects are small, the D-optimum designs are virtually those for the normal model, the unequally weight 3^2 factorial of Table 6 for the second-order model or the 2^2 factorial for the first-order model. Their higher order analogues, the weighted 3^m factorial and the unweighted 2^m factorial are close to optimum when there are m factors. These designs are also optimum for the gamma model with the log link. Our results show that these factorial designs, as well as the unweighted 3^m factorial, are efficient for the gamma model with a different link. However, for the binomial response, the design depends strongly on the values of the parameters, with regions in which the value of μ is appreciably greater that 0.85 or less than 0.15 best avoided. Finding approximate designs for binomial models using regression analogues will therefore involve searching over irregular design regions where the response does not fall outside these limits. An example would be the hexagon formed by the six design points of the right-hand panel of Figure 2. Searching over the original design space to find the optimum design for the generalized linear model would both be easier and lead to a more efficient design than would trying to find such a regression approximation.

6 Structured Mean-Variance Relationships

We now return to designs for continuous responses. In §3 these responses had a gamma distribution with the variance proportional to the square of the mean. In this section we concentrate on normal, but heteroscedastic, responses, so that the distribution is symmetrical. In the remaining sections of the paper the emphasis is on response transformation, which arises from the attempt to normalize asymmetrical response distributions in which the variance is a function of the mean. In the next section we derive the mean variance relationship for the power transformation family. In this section we

outline the design consequences of a parameterized variance function. The details are given by Atkinson and Cook (1995).

Statistical models in which both means and variances are functions of explanatory variables have become increasingly important in quality control (Box 1993), although the design consequences have been less explored. The possibility of additive heteroscedastic errors, known up to a constant of proportionality, is routinely considered by, for example, Fedorov (1972). Here the model has the more general form

$$y = f^T(x)\beta + \sigma[v\{g^T(z)\alpha\}]^{1/2}\epsilon, \tag{15}$$

where x and z are design vectors of dimension m_x and m_z with $f(x)$ and $g(z)$ respectively $p \times 1$ and $q \times 1$ vectors of linearly independent continuous functions, as was $f(x)$ in (2). The error term ϵ is standardised to have expectation zero and unit variance. In order to derive information matrices it will, in addition, be taken to have a normal distribution. The unknown parameters are α, β and $\sigma > 0$. It follows from (15) that, at the point (x, y), $\mathrm{E}(y) = f^T(x)\beta$ and $\mathrm{var}(y) = \sigma^2[v\{g^T(z)\alpha\}]$.

For applications it is often useful to take v to be the exponential function and then to work with a linear model for the logarithm of the variance

$$\log\{\mathrm{var}(y)\} = \log\sigma^2 + g^T(z)\alpha. \tag{16}$$

Atkinson and Cook (1995) identify two special cases of (15) that deserve attention. One is when the design variables influencing the mean are the same as those influencing the variance, that is $x = z$, so that (15) becomes

$$y = f^T(x)\beta + \sigma[v\{g^T(x)\alpha\}]^{1/2}\epsilon.$$

A further specialization is when the variance depends on x only through the mean so that

$$y = f^T(x)\beta + \sigma[v\{\nu f^T(x)\beta\}]^{1/2}\epsilon, \tag{17}$$

where ν is an unknown real-valued parameter that allows for the strength of dependence of the variance function on the mean.

The structure of the information matrices reflects the contributions to the estimation of the parameters from information coming from the mean and from the variance. When $\alpha = \alpha_0$ and $\sigma^2 = \sigma_0^2$ are known the information per observation on β in (15) has the well-known form

$$I_m(x, z|\beta, \alpha_0, \sigma_0^2) = \frac{f(x)f(x)^T}{\sigma_0^2 v\{g^T(z)\alpha_0\}}, \tag{18}$$

leading to estimation by (non-iterative) weighted least squares. The information for known $\beta = \beta_0$ can likewise be found and does not depend on these parameters for the linear model for the mean. The $(p+q+1) \times (p+q+1)$ information matrix for all the parameters is therefore block diagonal. A consequence is that the D-optimum design criterion for α, β and σ^2 is the product of the design criteria for the two individual sets of parameters.

For (17), in which the variance is a function of the mean, there are $p+2$ parameters. When the variance has the exponential form (16) the information matrix for one observation can be written as

$$I(x|\beta, \nu, \sigma^2) = z_1 z_1^T + z_2 z_2^T, \tag{19}$$

where

$$z_1^T(x|\beta, \nu, \sigma^2) = \{\nu f^T(x), f^T(x)\beta, \sigma^2\}/\sqrt{2}$$

and

$$z_2^T(x|\beta, \nu, \sigma^2) = \{f^T(x)v^{-1/2}\{\nu f^T(x)\beta\}, 0, 0\}/\sigma^2.$$

Comparison of (19) with (18) shows the extra precision that can be obtained when information about β comes both from the structure of both the mean and of the variance.

Atkinson and Cook (1995) give examples of designs for a two-factor response surface model. For (15) the optimum design depends on all $p+q+1$ parameters in the model. Both locally optimum and Bayesian designs are found for several parameter values and variance structures. Although some of the designs are like slightly distorted versions of the 3^2 factorial, other designs can be very different, particulary when both the mean and variance are unknown and the variance changes appreciably over the experimental region. Similar derivations, a discussion of estimation and an example of a design for a single-factor nonlinear model are given by Downing et al. (2001).

7 Transformation of the Response

7.1 Empirical Evidence for Transformations

Power transformation of the response is helpful in the special case when the variance of y increases as a power of the expected value $\mathrm{E}(y)$ of y. If

$$\mathrm{var}(y) \propto \{\mathrm{E}(y)\}^{2(1-\lambda)}, \tag{20}$$

Taylor series expansion shows that the variance is approximately stabilized by using as the response

$$\begin{array}{ll} y^\lambda & \lambda \neq 0 \\ \log y & \lambda = 0. \end{array} \tag{21}$$

So, for $\lambda = 1$, the variance is independent of the mean and no transformation is necessary. When $\lambda = 0.5$, the variance is proportional to the mean and the square root transformation is indicated, whereas, when $\lambda = 0$, the standard deviation is proportional to the mean and the logarithmic transformation provides approximately constant variance.

If the power law (20) holds with $\lambda < 1$, large observations will have larger standard deviations than small ones. Taking logarithms of the square root of both sides of this relationship yields

$$\log(\text{s.d.}Y) = \gamma_0 + (1 - \lambda) \log\{E(Y)\}, \tag{22}$$

where s.d. is the standard deviation of Y. If replicate observations are available, a plot of log standard deviation against log mean will indicate whether the power law holds We check this for two examples.

Atkinson (2003) analyses the extensive data set used by Downing et al. (2001) to illustrate optimum design with a parameterized variance function. The plot of log standard deviation against log mean for the 120 sets of replicate observations has a clear linear structure. The slope of the least squares line is 0.9034 with a t value of 16.3, indicating a strong relationship between standard deviation and mean, even if there is appreciable scatter in the plot.

Similar relationships are obtained from other examples. Lindsey (2001, p.113) presents data on flosequinan, a now withdrawn drug used in the management of patients with chronic heart failure. A dose of 150mg. was given to 18 healthy volunteers and readings taken at twelve times, at the first and last of which all readings were zero. Removal of an outlying subject from the data gives 17 non-zero observations at ten times. The plot of log standard deviation against log mean again shows a linear relationship. The least squares regression line in this example has a slope of 0.6494 and a t value of 22.84.

7.2 Horwitz' Rule

Although better estimates of λ could be obtained, for example by maximum likelihood estimation, these two examples show a strong relationship between mean and standard deviation. From (22) the slope of the plots is an estimate of $1-\lambda$, so that quite strong transformations are indicated for both data sets: 0.35 for flosequinan and 0.1, not far from the log, for the inhibition data.

The linear relationship between log standard deviation and log mean is well established in analytical chemistry, where it is known as Horwitz' rule, an empirical relationship between the variability of chemical measurements and the concentration of the analyte. Lischer (1999) states that it is supported by the results of studies involving almost 10,000 individual data sets;

the average transformation is to the power 0.14. This specific value goes against the standard statistical advice of using values with simple physical interpretations, such as the square root or the one third power for volumes. However the evidence for this rule is overwhelming.

7.3 Transformations and Optimum Design for Exponential Decay

Provided the value of λ is known, transformation of the response in polynomial response surface models such as (2) has no effect on the optimum design. The response is straightforwardly transformed, the D-optimum design for the parameters β being independent of the transformation. However, if the value of λ is not known, then a design may be required that is optimum for estimating both λ and β. Such designs, in a response surface setting, are the subsect of §8. If the model is nonlinear, transforming the response often does effect the design, even for a known transformation. For instance, if a kinetic model is such that the concentrations of the components sum to one, the sum of the power-transformed components will not be one. The model has also to be transformed. Such designs are the subject of §9

A simple example of this effect of transformation of the response on experimental design comes from the nonlinear response model resulting from first-order decay

$$A \stackrel{\theta}{\rightarrow} B$$

in which the concentration of chemical A at time t is given by the nonlinear function

$$[A] = \eta_A(t, \theta) = e^{-\theta t} \qquad (\theta, t \geq 0), \tag{23}$$

if it is assumed that the initial concentration of A is 1. If the ith experiment consists of measuring the concentration of A at time t_i, the simple statistical model of the observations (1) is

$$y_i = \eta_A(t_i, \theta) + \epsilon_i,$$

where the errors ϵ_i are independently distributed with zero mean and constant variance. The variance of the least squares estimator $\hat{\theta}$ then depends on the parameter sensitivity

$$f(t_i, \theta) = \frac{d\eta_A(t_i, \theta)}{d\theta} = -t_i \exp(-\theta t_i). \tag{24}$$

The locally D-optimum design minimising the variance of $\hat{\theta}$ (Box and Lucas 1959, Atkinson and Donev 1992) consists of taking all measurements where $f(t, \psi)$ is a maximum, that is at the time $t^* = 1/\theta$.

Now suppose that the model needs to be transformed to give constant variance. If the log transformation is appropriate and $[A]$ is measured, taking logarithms of both sides of (23), combined with additive errors, yields the statistical model

$$\log y_i = \log\{\eta_A(t_i, \theta)\} + \epsilon_i = -\theta t_i + \epsilon_i.$$

The log transformation thus results in a linear statistical model with response $\log y$, for which the parameter sensitivity is just the time t. The optimum design puts all observations at the maximum possible time, when the concentration is as small as possible, an apparently absurd answer. Thus a seemingly slight assumption about the error distribution can have a huge effect on the optimum experimental design. In §9.3 we extend this model to more than one variable and study the effect of the value of λ on the design.

8 Design for a Response Transformation

To overcome the need to treat $\lambda = 0$ as a special case, as in (21), let $y(\lambda)$ denote the power transformation of the positive, univariate response variable y,

$$y(\lambda) = \frac{(y^\lambda - 1)}{\lambda}. \tag{25}$$

It is assumed that there is a value of λ for which $y(\lambda)$ follows the linear regression model (2) with normal errors.

Atkinson and Cook (1996) find D-optimum designs for simultaneous estimation of all the parameters β, σ^2 and λ as well as $D_{\rm s}$-optimum designs for the subsets β and λ. Obtaining the design criterion is complicated by the nonlinear nature of the successive derivatives of $y(\lambda)$ with λ, the expectations of which are needed to calculate the expected information matrix. Atkinson and Cook use Taylor series expansions to obtain an approximate expected information matrix per observation, $I_a(\theta)$.

For the value of λ for which the transformation holds, let

$$\mu(x|\theta) = \mathrm{E}(y^\lambda) = \lambda f^T(x)\beta + 1. \tag{26}$$

The constructed variable

$$v(\lambda) = \partial y(\lambda)/\partial\lambda$$

has expectation

$$\mathrm{E}(v) = \mathrm{E}\{v(\lambda)\} \approx \frac{\mu(x|\theta)\log\{\mu(x|\theta)\} - \mu(x|\theta) + 1}{\lambda^2}.$$

Then

$$I_a(\theta) = z_1 z_1^T + z_2 z_2^T, \tag{27}$$

where

$$\begin{array}{rcl} z_1^T & = & (f(x)/\sigma \qquad 0 \qquad\quad -\mathrm{E}(v)/\sigma), \\ z_2^T & = & (0 \qquad 1/(\sqrt{2}\sigma^2) \quad -\sqrt{2}\log(\mu)/\lambda), \end{array}$$

with a similar expression for $\lambda = 0$.

The structure of (27) is similar to that of (19) for heteroscedastic linear models. The first term in z_1 is the information for β provided by the regression function. Atkinson and Cook (1996) argue that there are two sources of information about the transformation. One comes from the constructed variable $v(\lambda)$, which is similar to the constructed variables used in the analysis of response transformations (Atkinson 1985; Cook and Weisberg 1982). The transformation information in the variance function comes from the logarithm of the regression function $\log(\mu)$, where μ is defined in (26). The sum of squares of $\log(\mu)$ over the design enters into the second factor on the right of (27) through the last term of z_2, indicating a preference for designs with relatively large changes in the variance. This condition agrees with the common knowledge that response transformations are relatively well determined when the response ranges over several orders of magnitude.

The examples of designs given by Atkinson and Cook (1996) include a second-order response surface in two variables. Many of the designs are slightly distorted versions of the 3^2 factorial. However, if the response has a sharp maximum away from these nine values, extra support points enter the design - one design in their Figure 2 has thirteen points of support.

9 Response Transformations in Nonlinear models

9.1 Transforming Both Sides of a Nonlinear Model

The simple example of §7.3 for exponential decay shows the dependence of design for nonlinear models on the transformation, even when λ is known. This section develops simple expressions for the parameter sensitivities when the response is transformed. These expressions are then applied to find optimum designs for two nonlinear response surface models. Further details are in Atkinson (2003).

When, for example, $\eta(x, \theta)$ is a mechanistic model based on chemical kinetics, the relationship between the response and the concentrations of the

other reactants needs to be preserved after transformation. This is achieved by transformation of both sides of the model, as described in Chapter 4 of Carroll and Ruppert (1988). For fixed $\lambda \neq 0$, estimation of the parameters θ after transformation does not depend on whether the response is $y(\lambda)$ (25) or straightforwardly y^λ. Simplification of the model and the introduction of observational error on this transformed scale leads to the statistical model

$$y^\lambda = \{\eta(x, \theta)\}^\lambda + \epsilon. \tag{28}$$

The notation for the parameter sensitivities has to be extended to accommodate transformation. Let (5) be written

$$f_j(1; x, \theta) = \frac{\partial \eta(x, \theta)}{\partial \psi_j}. \tag{29}$$

The parameter sensitivities in the transformed model (28) are then

$$f_j(\lambda; x, \theta) = \frac{\partial \{\eta(x, \theta)\}^\lambda}{\partial \theta_j} = \lambda \{\eta(x, \theta)\}^{\lambda - 1} \frac{\partial \eta(x, \theta)}{\partial \theta_j} \propto f_j(1; x, \theta) / \{\eta(x, \theta)\}^{1-\lambda}, \tag{30}$$

since multiplication by known λ does not affect the optimum design.

If $\lambda < 1$, the variance of the observations increases with the value of $\eta(x, \theta)$. Thus transformation of both sides for such values of λ will increase the relative value of the sensitivities for values of x where the response is small. We can expect that designs for $\lambda < 1$ will include observations at lower concentrations than those when no transformation is needed.

9.2 Structured Parameters

Examples of designs for transformations in nonlinear models are given by Atkinson (2003) for kinetic models in which the only factor is time. In this section the parametrization of these models is extended to yield nonlinear response surface models. In general let

$$\theta_j = \phi_j \exp\{g(x, \psi_j)\}, \tag{31}$$

where $g(x, \psi_j)$ is a function relating the kinetic parameter θ_j to the vector of explanatory variables x. It is simplest to take $g(.)$ as a linear function. In the examples here it is also assumed that there is only one explanatory variable, other than time, and that all parameters are similarly influenced by x. Thus (31) becomes

$$\theta_j = \phi_j \exp(\psi x). \tag{32}$$

A medical example is where x is body weight. The rate of pharmacokinetic reactions depends on the concentration of drugs, so that, when the same dose is given to all patients, the reactions proceed more slowly for larger x and ψ would be negative. In chemical processes x might be a variable such as stirring rate, catalyst activity or feedstock purity, which would again affect all rates in the same way. The approximate rule from chemistry that a ten degree centigrade rise in temperature approximately doubles the rate of chemical reactions is also modelled by (32). However, if the Arrhenius equation is to be used to model the dependence of reaction rate on temperature, (31) applies (Box and Lucas 1959).

Design to estimate the parameters θ in (28) is replaced by design to estimate the vector ϕ and the scalar ψ. If the sensitivity

$$\frac{\partial \eta}{\partial \theta_j} = f_j^{\theta}$$

then, from the chain rule operating on (32),

$$f_j^{\phi} = \frac{\partial \eta}{\partial \phi_j} = \frac{\partial \eta}{\partial \theta_j}\frac{\partial \theta_j}{\partial \phi_j} = f_j^{\theta} e^{\psi x}. \tag{33}$$

The sensitivity f^{ψ} is found by differentiation of all $p-1$ elements of θ to be

$$f^{\psi} = \sum_{j=1}^{p-1} \frac{\partial \eta}{\partial \theta_j}\frac{\partial \theta_j}{\partial \psi} = \sum_{j=1}^{p-1} f_j^{\theta}\phi_j x e^{\psi x} = x\sum_{j=1}^{p-1} f_j^{\phi}\phi_j. \tag{34}$$

Two simple examples of nonlinear models with this parametric structure are now considered. In both the response is a smooth, nonlinear function of two variables, x and the time t.

9.3 Exponential Decay

The model for exponential decay was introduced in §7.3, where designs were found when the response is $[A]$. In this section designs are found when the response is the concentration of $[B]$.

Since no material is lost during the reaction, $[A] + [B] = 1$. From (23) the concentration of B at time t is therefore

$$[B] = \eta_B(t, \theta) = 1 - e^{-\theta t} \qquad (\theta, t \geq 0).$$

In the absence of transformation the sensitivity is minus that for $[A]$ in (24) and the D-optimum design again puts all trials $t = 1/\theta$.

We start by incorporating the structured parameters of the previous section. With the single structured parameter of the form (32), the model for the concentration of B at time t becomes

$$[B] = \eta_B(t, \theta) = 1 - \exp(\phi t e^{-\psi x}) \qquad (\phi, t \geq 0).$$

From (33) the sensitivity

$$f^\phi = te^{-\theta t} e^{\psi x}.$$

Also, since θ is scalar,

$$f^\psi = x\phi te^{-\theta t} e^{\psi x}. \tag{35}$$

Usually, as in the earlier examples in this chapter, a numerical optimization would be required to find the D-optimum design over the design region of values of x and t. But here some analytical progress can be made.

First consider f^ϕ. Although θ is a function of ϕ, ψ and x, it is not a function of t. So, for fixed x, the optimum value of t is found by differentiation to be, as before,

$$t^* = 1/\theta = e^{\psi x}/\phi. \tag{36}$$

This optimum value of t can now be substituted in (35) to give

$$f^\psi = x\phi t^* e^{-\theta t^*} e^{\psi x} = xe^{-1}.$$

Thus, provided the design region allows, the optimum design consists of trials at the upper and lower values of x at a time t^* given by (36). For the numerical example of this section the design region $\mathcal{X}$ is $0 \leq t \leq 20$ and $-1 \leq x \leq 1$. With $\phi = 0.2, t^* = 5$ when $x = 0$. The value of $\psi = \log 2$ results in a doubling of the rate of reaction when $x = 1$ and a halving when $x = -1$. In the absence of transformation the optimum design consists of equal numbers of trials at $t = 2.5$, $x = 1$ and $t = 10$, $x = -1$.

Now suppose that the model needs to be transformed to give constant variance. From (30) the sensitivity for θ is

$$f_j^\theta(\lambda; x, \theta) = t \exp(-\theta t)\{1 - \exp(-\theta t)\}^{\lambda - 1},$$

which is maximized by the optimum time. The argument in the preceding paragraph shows that the optimum design puts half the trials at twice this optimum time and half at half the time. The design is shown in Figure 5. As $\lambda \to 0$, the optimum times do likewise.

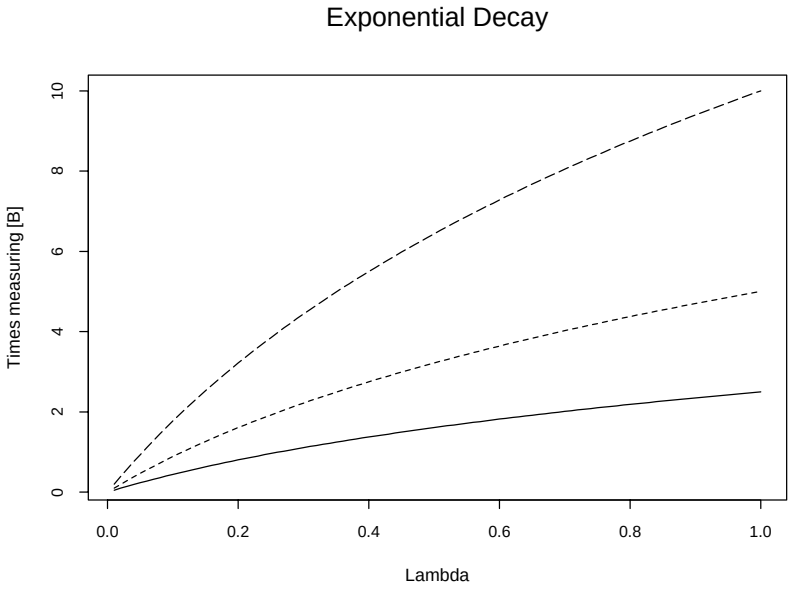

Figure 5: D-optimum designs for exponential decay as λ varies. The central line is the optimum design for the simple model with parameter $\theta = 0.2$. The optimum design for the structured parameter puts $n/2$ trials at twice this time and $x = -1$ and the other $n/2$ at half this time with $x = 1$

9.4 Two Consecutive First-Order Reactions

As a second and more complicated example, locally D-optimum designs are found in this section for the model for two consecutive first-order reactions introduced by Box and Lucas (1959).

The two reactions can be written

$$A \stackrel{\theta_1}{\rightarrow} B \stackrel{\theta_2}{\rightarrow} C.$$

When both reactions are first order, an explicit algebraic solution can be found for the concentrations of the reactants as a function of time. If the initial concentration of A is one and that of B and C are zero, $\eta_A(t, \theta)$ follows the exponential decay (23) with $\theta = \theta_1$ and

$$\eta_B(t, \theta) = \frac{\theta_1}{\theta_1 - \theta_2}(e^{-\theta_2 t} - e^{-\theta_1 t}). \tag{37}$$

As the reaction proceeds, $[B]$ rises from zero to a maximum and then decreases again to zero. With prior parameter values $\theta_1 = 0.7$ and $\theta_2 = 0.2$ and in the absence of transformation, the optimum design if just $[B]$ is measured has two equally weighted support points at times of 1.230 and 6.858.

Table 9: Two consecutive first-order reactions with a structured parameter. D-optimum designs for four values of λ

	$\lambda = 0.25$		$\lambda = 0.5$		$\lambda = 0.75$		$\lambda = 1$	
x_i	t_1	p_i	t_1	p_i	t_1	p_i	t_1	p_i
1	0.258	0.268	0.406	0.196	0.505	0.192	0.574	0.196
-1	1.018	0.084	1.621	0.177	2.019	0.192	2.299	0.196
1	10.75	0.327	5.642	0.314	4.154	0.308	3.475	0.304
-1	20.00	0.321	20.00	0.313	16.61	0.308	13.90	0.304

We now extend this model with the structured parameters (32). The effect on the concentration $[B]$ is

$$\eta_B(t, x, \phi, \psi) = \frac{\phi_1}{\phi_1 - \phi_2}\{\exp(-\phi_2 t e^{\psi x}) - \exp(-\phi_1 t e^{\psi x})\}, \qquad (38)$$

so that the effect of x can again be thought of as to multiply the time scale. With the values of x and ψ as in the previous example, the optimum points at the high level of x would be 0.615 and 3.429, whereas for the low level of x they would be 2.460 and 13.72. However, this model has three parameters, not four, so it is not clear that the optimum design will have exactly this structure.

Optimum designs for transformations are found numerically combining the sensitivities (33) and (34) for the structured parameters with those for transformations in (30). Table 9 gives optimum designs for several transformations when the maximum value of t is 20. The last two columns give the design for $\lambda = 1$. Although the design points are not far from the four suggested by elementary reasoning, the design weights are not equal. The weight on the two lower time points is 0.196, whereas the two higher times have weights of 0.304. For $x = 0$ differentiation of (37) shows that the maximum response is at $t = 2.506$. The two lower time points in the design are below this value, the two upper ones above.

The results of Table 9 show that, as λ decreases, the time points of the optimum designs move apart to regions of lower concentration. These results are illustrated for the four designs of the table in Figure 6. The dotted line in the figure shows the time of maximum yield $t_{\max}$, given from (38) by

$$t_{\max} = \frac{\log(\phi_1/\phi_2)}{\phi_1 - \phi_2} e^{-\psi x}.$$

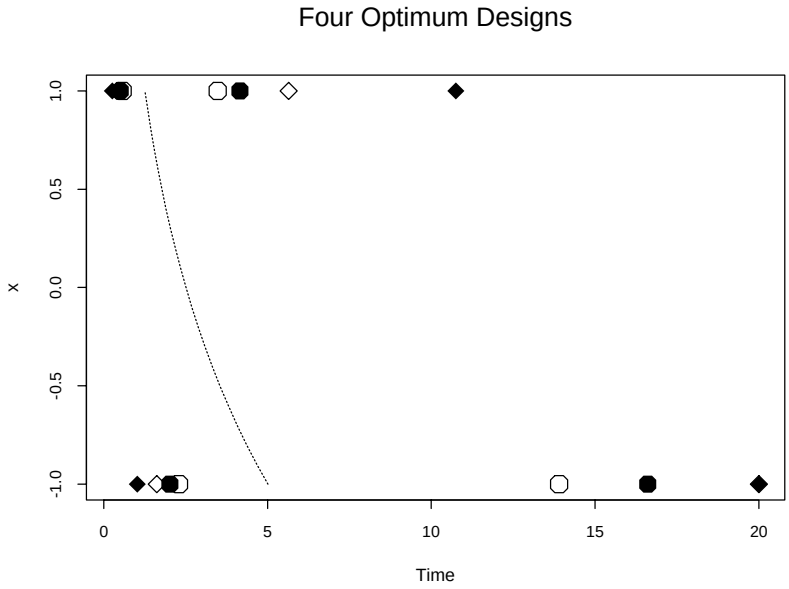

Figure 6: Consecutive first-order reactions: D-optimum designs for four values of λ. Open circle, $\lambda = 1$, filled circle, $\lambda = 0.75$, open rhombus, $\lambda = 0.5$ and filled rhombus, $\lambda = 0.25$; dotted line, time of maximum yield

For the two lower values of λ and $x = -1$, the time of reading is 20, the maximum value in the design region. Hence the symbols are overplotted.

Although all designs in Figure 6 have four points of support, Table 9 shows that the weight on the lower time point with $x = -1$ decreases as λ decreases. For lower values of λ a three-point design is optimum: for $\lambda = 0.1$, one of the support points of the design has the intermediate x value of -0.778.

9.5 Extensions

This example shows that the structure of the response surface design for nonlinear models may change appreciably with the value of the transformation parameter. More detailed results on the variation of design with criterion are available for nonlinear models with the single variable time. The variation seems to be greatest when more than one response is measured. However, although the designs may seem very different, each may be efficient for an appreciable range of values of λ. Bayesian designs, optimum for a specified prior distribution of λ values can then be found (Atkinson 2004). For some choices of prior the resulting designs are efficient for almost all values of λ in the range $0 < \lambda \leq 1$.

Transformations were introduced in this chapter to provide statistical

models in which the observations had constant variance. An alternative (Bogacka and Wright 2004) is to use weighted least squares with weights proportional to $\mathrm{E}(y)^{-2(1-\lambda)}$. The resulting parameter sensitivities, and so designs, are identical to those of §9.

10 Discussion

The emphasis in this paper is on locally D-optimum designs. Extensions would be to designs for multivariate responses, to other optimality criteria, such as the average variance of prediction, and to design when prior information is available about the parameters. However, conventional response surface methodology contains a number of other strands. For example, Box and Draper (1963) stress the importance of allowing for the biases that arise from an inadequate model and Box and Draper (1975) list 14 desiderata of a good experimental design. The application of some of these ideas to designs for generalized linear models might, for example, help in understanding and improving designs for blocking that would be produced by optimum design algorithms.

References

Atkinson, A. C. (1985). *Plots, Transformations, and Regression*. Oxford: Oxford University Press.

Atkinson, A. C. (2003). Horwitz's rule, transforming both sides and the design of experiments for mechanistic models. *Applied Statistics 52*, 261–278.

Atkinson, A. C. (2004). Some Bayesian optimum designs for response transformation in nonlinear models with nonconstant variance. In A. Di Bucchianico, H. Läuter, and H. P. Wynn (Eds.), *MODA 7 – Advances in Model-Oriented Design and Analysis*, pp. 13–22. Heidelberg: Physica-Verlag.

Atkinson, A. C. and R. D. Cook (1995). D-optimum designs for heteroscedastic linear models. *Journal of the American Statistical Association 90*, 204–212.

Atkinson, A. C. and R. D. Cook (1996). Designing for a response transformation parameter. *Journal of the Royal Statistical Society B 59*, 111–124.

Atkinson, A. C. and A. N. Donev (1992). *Optimum Experimental Designs*. Oxford: Oxford University Press.

Atkinson, A. C. and M. Riani (2000). *Robust Diagnostic Regression Analysis*. New York: Springer–Verlag.

Bogacka, B. and F. Wright (2004). A non-linear design problem in a chemical kinetic model with non-constant error variance. *Journal of Statistical Planning and Inference*. (In press).

Box, G. E. P. (1993). Quality improvement - the new industrial revolution. *International Statistical Review 61*, 1–19.

Box, G. E. P. and D. R. Cox (1964). An analysis of transformations (with discussion). *Journal of the Royal Statistical Society, Series B 26*, 211–246.

Box, G. E. P. and N. R. Draper (1963). The choice of a second order rotatable design. *Biometrika 50*, 335–352.

Box, G. E. P. and N. R. Draper (1975). Robust designs. *Biometrika 62*, 347–352.

Box, G. E. P. and H. L. Lucas (1959). Design of experiments in nonlinear situations. *Biometrika 46*, 77–90.

Burridge, J. and P. Sebastiani (1994). D–optimal designs for generalised linear models with variance proportional to the square of the mean. *Biometrika 81*, 295–304.

Carroll, R. J. and D. Ruppert (1988). *Transformation and Weighting in Regression*. London: Chapman and Hall.

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

Cook, R. D. and S. Weisberg (1982). *Residuals and Influence in Regression*. London: Chapman and Hall.

Cox, D. R. (1988). A note on design when response has an exponential family distribution. *Biometrika 75*, 161–164.

Downing, D., V. V. Fedorov, and S. Leonov (2001). Extracting information from the variance function: optimal design. In A. C. Atkinson, P. Hackl, and W. G. Müller (Eds.), *MODA 6 – Advances in Model-Oriented Design and Analysis*, pp. 45–52. Heidelberg: Physica-Verlag.

Farrell, R. H., J. Kiefer, and A. Walbran (1968). Optimum multivariate designs. In *Proc. 5th Berkeley Symposium*, Volume 1, Berkeley, CA, pp. 113–138. University of California Press.

Fedorov, V. V. (1972). *Theory of Optimal Experiments*. New York: AcademicPress.

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

Lewis, S. L., D. C. Montgomery, and R. H. Myers (2001). Confidence interval coverage for designed experiments analyzed with glms. *Journal of Quality Technology 33*, 279–292.

Lindsey, J. K. (2001). *Nonlinear Models in Medical Statistics*. Oxford: Oxford University Press.

Lischer, P. (1999). Good statistical practice in analytical chemistry. In B. Grigelionis (Ed.), *Probability Theory and Mathematical Statistics*, pp. 1–12. Dordrecht: VSP.

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

Myers, R. H., D. C. Montgomery, and G. C. Vining (2001). *Generalized Linear Models: with Applications in Engineering and the Sciences*. New York: Wiley.

Myers, R. H., D. C. Montgomery, G. C. Vining, C. M. Borror, and S. M. Kowalski (2004). Response surface methodology: a retrospective and literature survey. *Journal of Quality Technology 36*, xxx–xxx.

Nelson, W. (1981). The analysis of performance-degradation data. *IEEE Transactions on Reliability R-30*, 149–155.

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

Silvey, S. D. (1980). *Optimum Design*. London: Chapman and Hall.