Some Bayesian Optimum Designs for Response Transformation in Nonlinear Models with Nonconstant Variance

Anthony C. Atkinson*, Department of Statistics,
London School of Economics, London WC2A 2AE, UK

November 24, 2003

Abstract

Often the responses from mechanistic models have to be transformed to achieve error distributions that are symmetric and have constant variance. The paper develops methods for the design of experiments when the value of the power transformation λ is not precisely known. Because of the nature of the relationship between response and the mechanistic model, it is necessary to transform both sides of the model. Expressions are given for the parameter sensitivities in the transformed model and examples given of optimum designs for particular values of λ, together with the efficiency of these designs as λ varies. When the uncertainty about λ is expressed by a prior distribution, Bayesian D-optimum designs can be found. It is shown, for one example, that these designs can have high efficiency for a range of values of λ.

Keywords: Bayesian D-optimality; Box-Cox transformation; Chemical kinetics; D-optimum design; Mechanistic model; Parameter sensitivities; Transform both sides.

1 Introduction

This paper is concerned with the design of experiments for the mechanistic models which arise in pharmacokinetics and chemical kinetics. Usually the variance of the measurements of concentrations to which these models are

*a.c.atkinson@lse.ac.uk

fitted increases with the mean of the observations. Atkinson (2003) describes several examples. The variance can be stabilised, and the error distribution made symmetrical, by transforming the response. But, with the mechanistic models that are considered here, it is necessary to transform both sides of the model. The optimum design will then depend on the transformation that is appropriate as well as on the model.

The transformation is indexed by the parameter λ. Calculations and plots in Atkinson (2003) indicate that the optimum designs often change rapidly with λ. The example, for two consecutive second-order chemical reactions, in Atkinson (2004) shows that this dependence of the designs on λ can be complicated. In practice the exact value of λ will often not be known. The purposes of this paper are to calculate the efficiency of the designs for specific λ as the true value of the transformation parameter varies and to find Bayesian designs that are optimum for a set of values of λ.

The paper starts in §2 with the simple, but illuminating, example of first-order decay. The next section reviews optimum design theory for multivariate responses. In §4 the transformation model is described together with parameter sensitivities. The main numerical example of two consecutive first-order reactions is introduced in §5. The main theoretical results are in §6, where efficiencies are defined and Bayesian D-optimum designs for λ are introduced. Both are calculated for D-optimum designs for two consecutive first-order reactions. The paper concludes with a brief discussion in §9.

2 Transformations and First-order Decay

The experimental design problem considered is that of selecting a set of times t_i at which readings y_i should be taken with weights w_i in order to obtain good estimates of the parameter vector ψ in the nonlinear mechanistic model

$$y = \eta(t, \psi). \tag{1}$$

This becomes a statistical problem when error is introduced.

Let the expected value of Y be $E(Y) = \mu$ and the variance be var$Y$. If the variance of the random variable Y increases as a power of the expected value so that

$$\text{var}Y \propto \{E(Y)\}^{2(1-\lambda)}, \tag{2}$$

Taylor expansion shows that power transformation of the response provides a random variable with variance approximately independent of the mean. For transformation of just the response y in a regression model, the normalized

power transformation of Box and Cox (1964) is

$$z(\lambda) = \begin{cases} (y^\lambda - 1)/(\lambda \dot{y}^{\lambda-1}) & \lambda \neq 0 \\ \dot{y} \log y & \lambda = 0, \end{cases} \tag{3}$$

where the geometric mean of the observations is written as $\dot{y} = \exp(\Sigma \log y_i/n)$. When $\lambda = 1$, there is no transformation. The model to be fitted is (1) with response $z(\lambda)$ and additive errors of constant variance.

However, when, for example, $\eta(t, \psi)$ 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 ψ after transformation does not depend on whether the response is $z(\lambda)$ or the non-normalized y^λ. Simplification of the model and the introduction of observational error on this transformed scale leads to the statistical model

$$y^\lambda = \{\eta(t, \psi)\}^\lambda + \epsilon. \tag{4}$$

Atkinson (2003) assumed that the value of λ was known and compared designs for different λ. Designs for estimation of λ when the response of a regression model is to be transformed are described by Atkinson and Cook (1996).

A simple example of the effect of assumptions about the mean-variance relationship on experimental design comes from the model for first-order decay in which the concentration of chemical A at time t is given by the nonlinear function

$$[A] = \eta_A(t, \theta) = e^{-\theta t} \qquad (\theta, t \geq 0), \tag{5}$$

if it is assumed that the initial concentration of A is 1. If we put $\lambda = 1$ in (4) we have the usual model with additive errors of constant variance. Box and Lucas (1959) show that, for this error model, the locally D-optimum design consists of taking all measurements at a time $t^* = 1/\theta$.

However, if the log transformation is appropriate to obtain errors of constant variance and $[A]$ is measured, taking logarithms of both sides of (5), combined with additive errors, yields the statistical model

$$\log y_i = \log\{\eta_A(t_i, \theta)\} + \epsilon_i = -\theta t_i + \epsilon_i. \tag{6}$$

The optimum design now puts all observations at the maximum possible time, when the concentration is as small as possible, an implausible experimental strategy.

Analysis of two sets of pharmacokinetic data in Atkinson (2003) yields values of λ of 0.35 and 0.1. In comparison, Horwitz's rule for analytical chemistry based on the analysis of thousands of data sets (Lischer 1999) yields a value around 0.14 for λ. Clearly transformation of such data is often required.

3 Optimum Design for a Multivariate Response

The ith reading comes from running the experiment for a time t_i and taking a measurement of one or more responses. As in Box and Lucas (1959), the measurements are taken at a single time point for each run. Here only continuous designs are found, in which the design ξ is a continuous measure specifying both a set of k distinct points in a design region $\mathcal{T}$ and the proportions, w_i, of observations taken at these points. The times t_i are the points of support of the design ξ and w_i the design weights. Locally D-optimum designs maximize the logarithm of the determinant of the information matrix, $\log|M(\xi,\psi)|$, in which a best guess ψ^o is taken for the values of the parameters.

If the concentration of more than one chemical is measured, there will be a model for each expected response giving a matrix F_i of parameter sensitivities for the ith response, $i = 1, \ldots, m$,

$$F_i = \{f_{ij}(t_u, \psi)\} = \left\{\frac{\partial \eta_i(t_u,\psi)}{\partial \psi_j}\right\},$$

where $u = 1, ..., k$ denote the design points and $j = 1, ..., p$ denote the parameters. In the absence of transformation the observations follow the model

$$y_{iu} = \eta_i(t_u, \psi) + \varepsilon_{iu},$$

with

$$E(\varepsilon_{iu}) = 0, \qquad E(\varepsilon_{iu}\varepsilon_{lv}) = \begin{cases} 0 & if \quad u \neq v \\ \sigma_{il} & if \quad u = v \end{cases},$$

when the variance-covariance matrix of the responses is

$$\Sigma = \{\sigma_{il}\}_{i,l=1,\ldots,m}.$$

For normally distributed errors the information matrix is given by

$$M(\xi,\psi) = \sum_{i=1}^{m}\sum_{l=1}^{m} \sigma^{il} F_i^T W F_l, \tag{7}$$

where $\Sigma^{-1} = \{\sigma^{il}\}_{i,l=1,\ldots,m}$ and $W = \text{diag}\{w_1, \ldots, w_k\}$.

4 Parameter Sensitivities and Transforming Both Sides

The parameter sensitivities for the model in §3 with $\lambda = 1$ will be written

$$f_j^1(t,\psi) = \frac{\partial \eta(t,\psi)}{\partial \psi_j}. \tag{8}$$

The parameter sensitivities in the transformed model (4) are then

$$f_j^\lambda(t,\psi) = \frac{\partial \{\eta(t,\psi)\}^\lambda}{\partial \psi_j} = \lambda \{\eta(t,\psi)\}^{\lambda-1} \frac{\partial \eta(t,\psi)}{\partial \psi_j} = \lambda \{\eta(t,\psi)\}^{\lambda-1} f_j^1(t,\psi). \tag{9}$$

For fixed λ, multiplication by λ in (9) does not change the optimum design, so the sensitivities have the easily calculated form

$$f_j^\lambda(t,\psi) = \{\eta(t,\psi)\}^{\lambda-1} f_j^1(t,\psi) = f_j^1(t,\psi)/\{\eta(t,\psi)\}^{1-\lambda}. \tag{10}$$

If $\lambda < 1$, the variance of the observations increases with the value of $\eta(t,\psi)$. Thus transformation of both sides for such values of λ will increase the relative value of the sensitivities for times 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.

5 Two Consecutive First-Order Reactions

Section 6 of Atkinson (2003) shows how the locally optimum design for θ in the exponential decay model of §2 changes as a function of λ. In this section locally D-optimum designs are found for the model for two consecutive first order reactions introduced by Box and Lucas (1959).

The two reactions can be written

$$A \stackrel{\theta_1}{\rightarrow} B \stackrel{\theta_2}{\rightarrow} C. \tag{11}$$

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 (5) with $\theta = \theta_1$. The other concentrations are given by

$$\begin{aligned} \eta_B(t,\theta) &= \frac{\theta_1}{\theta_1 - \theta_2}(e^{-\theta_2 t} - e^{-\theta_1 t}) \\ \eta_C(t,\theta) &= 1 - \eta_A(t,\theta) - \eta_B(t,\theta). \end{aligned} \tag{12}$$

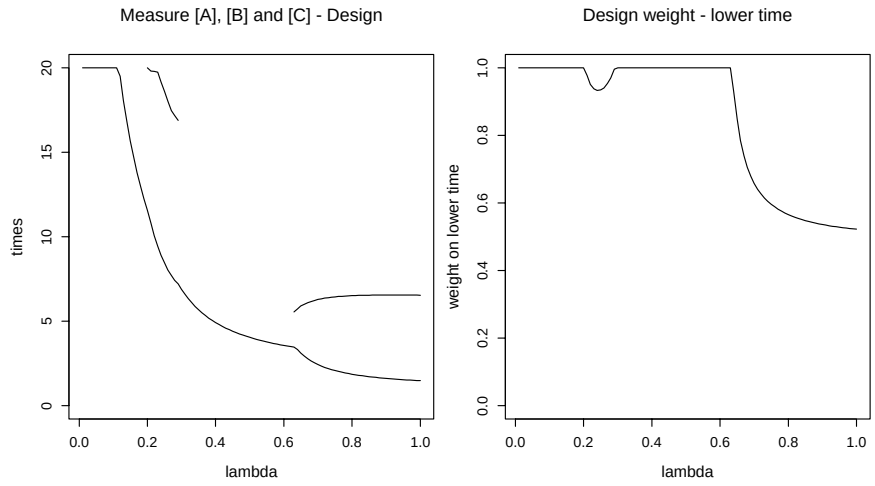

Figure 1: Two consecutive first-order reactions: locally-optimum designs when [A], [B] and [C] are measured. Left panel, times; right-panel, design weights

The parameter sensitivities $f^1_{Bj}(t,\theta)$ and $f^1_{Cj}(t,\theta) = -f^1_{Aj}(t,\theta) - f^1_{Bj}(t,\theta)$ are readily found by differentiation of the concentrations in (12). The parameter sensitivities for the transformed model are then found from (10).

The locally optimum designs when the concentrations of all three reactants are measured are plotted in Figure 1 for λ from 0.01 to one. In the calculations $\theta_1 = 0.7$, $\theta_2 = 0.2$ and the three measurements were taken as uncorrelated with equal variances. For λ close to one, the design has two approximately equally weighted design points. However, for λ below 0.64 there is only one design point, except for a region around 0.25 where there are two unequally weighted points. Below 0.12 there is a single design point at $t = 20$, the edge of the design region.

6 Efficiencies and Bayesian Optimum Designs

Since the optimum designs in Figure 1 vary so much with λ it is likely that a design for the wrong λ will be inefficient. In this section efficiencies are calculated for designs as λ varies and Bayesian optimum designs introduced that are intended to have high efficiency over a set of values of λ.

Let the optimum design for a specified λ_0 be ξ^*_0 and for some other λ be ξ^*_λ. The value of the information matrix depends not only on ξ and ψ but also on the parameter λ. When the value of the transformation parameter is λ the information matrix for the design ξ^*_0 can be written as $M(\xi^*_0, \psi, \lambda)$.

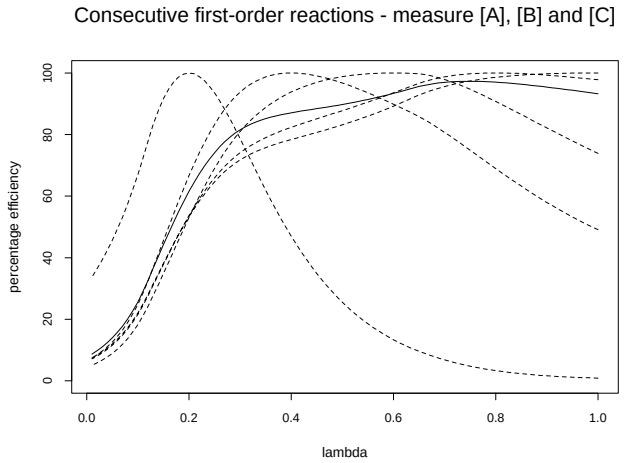

Figure 2: Two consecutive first-order reactions when [A], [B] and [C] are measured. Continuous line, efficiency of Bayesian optimum design for $\lambda_0 = (0.2, 0.4, 0.6, 0.8, 1.0)^T$; dashed lines, efficiencies of the optimum designs for individual members of λ_0

Then the efficiency of the design ξ_0^* for some λ is the ratio of determinants

$$E_f(\xi_0^*, \lambda) = \{|M(\xi_0^*, \psi, \lambda)|/|M(\xi_\lambda^*, \psi, \lambda)|\}^{1/p}, \tag{13}$$

where p is the number of parameters in the model. The dotted lines in Figures 2 and 3 are the efficiencies over a range of values of λ of the D-optimum designs for five values of λ_0: 0.2, 0.4, 0.6, 0.8 and 1. Each is, of course, 100% when $\lambda = \lambda_0$. What is particularly noticeable is that all designs except that for $\lambda_0 = 0.2$ are similarly inefficient for low values of λ. The design for $\lambda_0 = 0.2$ is itself inefficient for high values of λ.

Bayesian optimum designs are found by introducing a prior distribution with weight q_i for the ith value of λ_0. The resulting compound design criterion, when the prior is concentrated on $n(q)$ points, maximizes

$$\Phi(\xi, \psi, \lambda_0) = \sum_{i=1}^{n(q)} q_i \log |M(\xi, \psi, \lambda_i)|. \tag{14}$$

Comparison with (13) shows that ξ_B^*, the Bayesian D-optimum design, is also the design that maximizes a weighted product of the efficiencies of the designs for the elements of λ_0. The effect of the logarithm in (14) is to remove the effect of the determinants $|M(\xi_\lambda^*, \psi, \lambda)|$ since these are constants at the $n(q)$ values where optimization takes place.

Three Bayesian optimum designs are listed in Table 1, each for a five-point prior for λ with all $q_i = 0.2$. The first design has the values of λ

Table 1: Bayesian optimum designs for three five-point priors on λ

Prior λ_0	t, w	Design		
0.2, 0.4, 0.6, 0.8, 1.0	t_i	2.043	6.883	
	w_i	0.452	0.548	
0.1, 0.2, 0.3, 0.4, 0.5	t_i	5.139	17.911	
	w_i	0.591	0.409	
0.1, 0.2, 0.3, 0.9, 1.0	t_i	1.633	8.228	20.000
	w_i	0.280	0.591	0.129

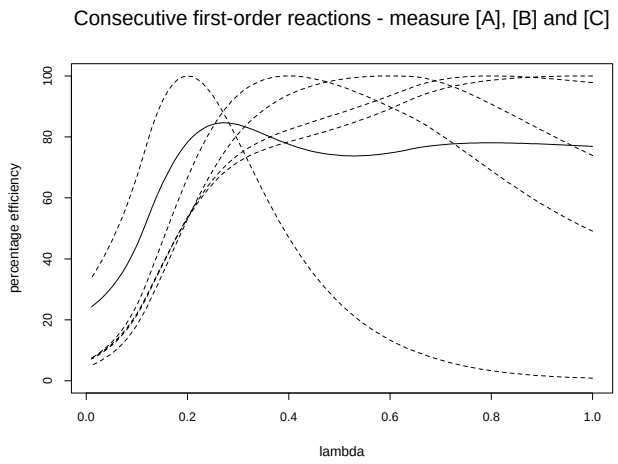

Figure 3: Two consecutive first-order reactions when [A], [B] and [C] are measured. Continuous line, efficiency of Bayesian optimum design for $\lambda_0 = (0.1, 0.2, 0.3, 0.9, 1.0)^T$; dashed lines, efficiencies of the optimum designs for individual values 0.2, 0.4, 0.6, 0.8, 1.0 for λ

uniformly distributed over 0.2, 0.4, 0.6, 0.8 and 1. It is a two-point design, the efficiency of which is shown by the continuous line in Figure 2. It has good efficiency for high values of λ, but its efficiency falls off rapidly below $\lambda = 0.3$. The second design in Table 1 is also a two-point design. The prior values of λ concentrate on a lower range of values than before, 0.1 to 0.5. Plots not given here show that the resulting design is efficient over this range, but of decreasing efficiency for higher values of λ. An efficient design for a wider range of values of λ is in the last row of Table 1. It is for a prior concentrated at the ends of the region and the resulting design has three points of support at low medium and high values of time. As the plot of the efficiency in Figure 3 shows, the design has good efficiency for values of λ as low as 0.2 and a steady efficiency between 75 and 80% for most values above

this. By suitable choice of the prior, a design has been found that is robust to the choice of λ.

7 Discussion

The plot of the design in Figure 1 shows, at least for this example, how dependent the locally D-optimum design is on the assumed value of λ. The examples in Atkinson (2003) and Atkinson (2004) for consecutive second-order reactions show even more complicated changes of design with λ, particularly if several responses are measured. The plots of efficiencies in Figures 2 and 3 reveal that designs for a single value of λ are inefficient. However, the Bayesian D-optimum designs introduced in this paper illustrate how, by a suitable choice of prior, designs can be found that have good efficiency for a range of values of λ.

Transformations were introduced in this paper in order 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 $E(Y)^{-(1-\lambda)}$. The resulting parameter sensitivities and so designs, are identical to those of §4. An important difference between such methods and the transformation studied here is in the model implied for fitting the data. In the transformation model, the original observations will have skewed distributions, which become symmetrical, with constant error variance, after the appropriate transformation. However, weighting and models with structured variances lead to symmetrical distributions of error of non-constant variance in the original scale of the observations.

Several extensions of this work are possible. One important direction is to the design of experiments for industrial processes where the dynamics of the model are described by sets of differential equations. Any correlation of errors is ignored, for example as in Bauer, Bock, Körkel, and Schlöder (2000). So a second extension is to design when the errors are correlated. Recent papers describing algorithms for optimum designs in the presence of correlated errors include Müller and Pázman (2003) and Uciński and Atkinson (2004).

References

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). Transforming both sides and optimum experimen-

tal design for a nonlinear model arising from second-order chemical kinetics. *Tatra Mountains Mathematical Publications*. (In press).

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

Bauer, I., H. G. Bock, S. Körkel, and J. P. Schlöder (2000). Numerical methods for optimum experimental design in DAE systems. *Journal of Computational and Applied Mathematics 120*, 1–25.

Bogacka, B. and F. Wright (2004). A non-linear design problem in a chemical kinetic model with non-constant error variance. (submitted).

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

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

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

Müller, W. G. and A. Pázman (2003). Measures for designs in experiments with correlated errors. *Biometrika 90*, 423–434.

Uciński, D. and A. C. Atkinson (2004). Experimental design for processes over time. *Studies in Nonlinear Dynamics and Econometrics*. (Submitted).