# Optimal Bayesian Design for a Logistic Regression Model:

## Geometric and Algebraic Approaches

By

Marilyn Agin and Kathryn Chaloner

School of Statistics, University of Minnesota
Technical Report No. 622
August, 1997

OPTIMAL BAYESIAN DESIGN
FOR A LOGISTIC REGRESSION MODEL:
GEOMETRIC AND ALGEBRAIC APPROACHES

Marilyn Agin and Kathryn Chaloner
*School of Statistics, University of Minnesota,*
*St. Paul, MN 55108, U.S.A.*

**Abstract**

A simple logistic regression model with a known slope parameter provides a simple model for understanding optimal Bayesian designs. For this model the geometric approach of Haines (1995) and the algebraic approach of Chaloner (1993) for finding optimal designs for a prior distribution with just two support points are reviewed, compared and discussed. A result is given, using the algebraic approach, for a three point prior distribution and difficulties of the geometric approach are illustrated.

*Key Words:* Bayesian design; logistic regression; nonlinear models; optimal design.

# 1 Introduction

Haines (1995) and Chaloner (1993) give closed form results for Bayesian designs for nonlinear problems. Both papers use, among other examples, the logistic regression model with a known slope parameter. Both papers derive some of the same results for prior distributions with two points of support, but with very different methods. Haines uses a novel geometric approach and Chaloner uses a more traditional algebraic approach using an equivalence theorem. Prior distributions with a small number of support points are not of much practical use, but when closed form solutions can be found they give an understanding of more general problems in which designs must be found numerically.

The two different approaches are compared and contrasted here in Sections 2 and 3. A new result for a three point prior distribution is given in Section 4.

## 2 Bayesian Design for Nonlinear Problems

An extensive review of Bayesian approaches to design is given in Chaloner and Verdinelli (1995). A Bayesian approach to design in nonlinear problems is to think of design as a decision problem and to maximize a criterion representing an approximation to the expected utility. One of the earliest implementations of this approach was by Tsutakawa (1972). Criteria similar to $D$- and $A$-optimality in linear problems can be derived which are related to the local optimality approach of Chernoff (1953), but which involve averaging over a prior distribution on the unknown parameters. Silvey (1980) provides a review of local optimality.

Let the design region $\mathcal{X}$ be a compact subset of $\Re^{\mathbf{k}}$, and assume the explanatory variable $x$ is an element of $\mathcal{X}$. Define $\mathcal{H}$ to be the set of probability measures over $\mathcal{X}$. Let $\eta$ denote both a probability measure and its density function.

For a fixed sample size $n$, the design problem is to choose values of the explanatory variable $x$ from $\mathcal{X}$, the proportion of observations $w_i$ to be taken at each value $x_i$, $i = 1, \ldots, k$ and the number of values $k$. This corresponds to choosing a measure $\eta$ from $\mathcal{H}$ that maximizes a criterion function $\phi(\eta)$. It is typically easier to optimize directly over the set of measures $\mathcal{H}$ and assume that the $w_i$ can take any values such that $\sum_{i=1}^{k} w_i = 1$. Thus the $nw_i$ are not constrained to be positive integers.

Define $D(\eta_1, \eta_2)$ to be the directional derivative of $\phi(\eta)$ at $\eta_1$ in the direction of $\eta_2$. Specifically:

$$D(\eta_1, \eta_2) = \lim_{\epsilon \to 0^+} \frac{1}{\epsilon} \left[ \phi((1-\epsilon)\eta_1 + \epsilon\eta_2) - \phi(\eta_1) \right]. \tag{1}$$

Also define $\eta_x$ to be the probability measure which is point mass at $x$ in $\mathcal{X}$. Then the directional derivative $D(\eta, \eta_x)$, as a function on $\mathcal{H} \times \mathcal{X}$, is denoted as $d(\eta, x)$.

If the criterion function $\phi(\eta)$ is concave on $\mathcal{H}$, the equivalence theorem of Whittle (1973) can be extended, as in Chaloner and Larntz (1989), to verify that a particular design is optimal. Whittle gave the theorem in the context of linear design problems but it also holds in nonlinear problems with the additional assumptions that the design region $\mathcal{X}$ is a compact subset of $\Re^{\mathbf{k}}$, the derivatives exist and are continuous in $x$, there is at least one measure in $\mathcal{H}$ for which $\phi$ is finite, and that $\phi$ is such that if $\eta_i \to \eta$ in weak convergence then $\phi(\eta_i) \to \phi(\eta)$.

**Equivalence Theorem**

*(a.) If $\phi$ is concave, then a $\phi$-optimal design $\eta_*$ can be equivalently characterized by any of the three conditions*

*(i) $\eta_*$ maximizes $\phi(\cdot)$,*

*(ii) $\eta_*$ minimizes $\sup_{x\in\mathcal{X}} d(\eta, x)$,*

*(iii) $\sup_{x\in\mathcal{X}} d(\eta_*, x) = 0$.*

*(b.) The point $(\eta_*, \eta_*)$ is a saddlepoint of $D$ in that $D(\eta_*, \eta_1) \leq 0 = D(\eta_*, \eta_*) \leq D(\eta_2, \eta_*)$ for all $\eta_1, \eta_2 \in \mathcal{H}$.*

*(c.) If $\phi$ is differentiable, then the support of $\eta_*$ is contained in the set of $x$ for which $d(\eta_*, x) = 0$ almost everywhere in $\eta_*$ measure.*

## 2.1 The Logistic Regression Problem

Consider independent Bernoulli responses observed at a value $x$ of the explanatory variable. Let the probability of success $p(x)$ be

$$p(x) = \frac{1}{[1 + e^{\{-(x-\theta)\}}]}.$$

The design problem is to choose values $x_i$ of the explanatory variable, and corresponding weights $w_i$, $i = 1, \ldots, k$ subject to the constraint $\sum_{i=1}^{k} w_i = 1$. The sample size is fixed and equal to $n$.

Consider a one point design at $x$. Denote the expected Fisher information for the design putting one observation at $x$ as $I(\theta, x)$. Then

$$I(\theta, x) = \frac{e^{-(x-\theta)}}{(1 + e^{-(x-\theta)})^2}.$$

Consider the design, $\eta$, with weight $w_i$ at $x_i, i = 1, ..., k$, and define

$$\begin{aligned} I(\theta, \eta) &= \int I(\theta, x) d\eta(x) \\ &= \sum_{i=1}^{k} w_i \frac{e^{-(x_i-\theta)}}{(1 + e^{-(x_i-\theta)})^2}. \end{aligned}$$

Then the Fisher information for a design measure $\eta$ in $\mathcal{H}$ is $nI(\theta, \eta)$.

As in Chaloner (1987, 1993) and Chaloner and Larntz (1989), under an asymptotic normal approximation to the posterior distribution of $\theta$, the expected Shannon information gives $\phi_1(\cdot)$ as a criterion function for Bayesian D-optimality where

$$\begin{aligned} \phi_1(\eta) &= E[\ln I(\theta, \eta)] \\ &= \int_{\Theta} \ln I(\theta, \eta) d\pi(\theta), \end{aligned}$$

with the expectation taken with respect to the prior distribution, $\pi$, on $\theta$. Similarly, the criterion function $\phi_2$ for Bayesian c-optimality approximates minus the squared error loss, so

$$\begin{aligned} \phi_2(\eta) &= -E[I(\theta, \eta)]^{-1} \\ &= -\int_{\Theta} \frac{1}{I(\theta, \eta)} d\pi(\theta), \end{aligned}$$

and the $\phi_2$-optimal design measure $\eta$ in $\mathcal{H}$ maximizes $-E[I(\theta, \eta)]^{-1}$.

Both $\phi_1(\eta)$ and $\phi_2(\eta)$ are concave functions on $\mathcal{H}$.

Consider, for illustration, a prior distribution with only two support points, at $\theta = \pm g$, with mass $\frac{1}{2}$ at each point. The following result was proved algebraically in Chaloner (1993) for $\phi_1$-optimality and proved geometrically and generalized in Haines (1995) for both $\phi_1$-optimality and $\phi_2$-optimality.

**Theorem 1** (Haines, 1995)

*For a prior distribution which puts probability $\frac{1}{2}$ on each of $\theta = \pm g$, the $\phi_1$-optimal and $\phi_2$-optimal designs are:*

*(1) If $|g| \leq \ln(2 + \sqrt{3})$, both optimal design measures put mass 1 at $x = 0$.*

*(2) If $|g| > ln(2+\sqrt{3})$ define*

$$B(g) = ln\frac{e^{4g} - 6e^{2g} + 1 + (e^{2g}-1)\sqrt{e^{4g} - 14e^{2g} + 1}}{2(e^{3g} + e^{g})}.$$

*Assuming that $\pm B(g) \in \mathcal{X}$, both optimal design measures put mass $\frac{1}{2}$ at each of $x = \pm\, B(g)$.*

## 2.2 The Algebraic Approach

For illustration and completeness, an algebraic proof is given here for $\phi_2$-optimality using the equivalence theorem.

The derivative for $\phi_2$-optimality has a simple expression:

$$d_2(\eta, x) = E\left[\frac{I(\theta,x)}{I(\theta,\eta)^2} - \frac{1}{I(\theta,\eta)}\right].$$

First consider $|g| \le ln(2+\sqrt{3})$ and define $\eta_0$ to be the one point design putting mass 1 at $x = 0$. Define

$$Q1 = 2e^{2x+2g} - e^{x+4g} + 2e^{x+3g} + 2e^{x+2g} + 2e^{x+g} - e^{x} + 2e^{2g}$$

then

$$d_2(\eta_0, x) = -\frac{e^{-g}(e^{g}+1)^2(e^{x}-1)^2(Q1)}{2(e^{x}+e^{g})^2(e^{x+g}+1)^2}.$$

$Q1$ is a quadratic in $e^{x}$ and it is straightforward, but somewhat tedious, to show that if $|g| \le \ln(2+\sqrt{3})$, the quadratic is positive and the derivative has a single root at $x = 0$ and so $\eta_0$ is the $\phi_2$-optimal design.

Now consider $|g| > \ln(2+\sqrt{3})$ and let $\eta_2$ be the symmetric design which puts mass $\frac{1}{2}$ at each of $\pm B(g)$, where $\pm B(g)$ is defined in Theorem 1. Define

$$Q2 = e^{2x+3g} + e^{2x+g} - e^{x+4g} + 6e^{x+2g} - e^{x} + e^{3g} + e^{g}.$$

Then the directional derivative $d_2(\eta_2, x)$ at the design $\eta_2$ is

$$d_2(\eta_2, x) = -\frac{8(e^{g}-1)^2(e^{g}+1)^2(Q2)^2}{(e^{2g}+1)^4(e^{x}+e^{g})^2(e^{x+g}+1)^2},$$

which is clearly nonpositive for all $x$. It can be shown that the derivative equals zero only at $x = \pm B(g)$, and so the design $\eta_2$ is $\phi_2$-optimal. The proof is therefore complete for $\phi_2$-optimality.

The algebraic approach requires a candidate optimal design for which the derivative can be calculated and optimality verified. The proof is straightforward although algebraically tedious. The algebra can be made considerably less tedious by the use of a symbolic algebra system such as Mathematica or MACSYMA. The geometric approach of Haines, in contrast, gives a more direct approach to finding an optimal design.

## 2.3 The Geometric Approach

In the geometric approach slighly different notation is required. Let the design space $\mathcal{X} = \Re$. Then for fixed $g$ the set of all possible design points, $x \in \Re$, can be represented by a curve $C(x)$ in $\Re^2$ with horizontal and vertical co-ordinates $(I(-g,x), I(g,x))$. Define $S$ to be the closed convex hull of $C(x)$. $S$ is the set of all convex combinations of points on $C$ and is the convex set over which the criterion function is maximized. $S$ represents the set of all possible designs

$$S = conv(C) = \{(I(-g,\eta), I(g,\eta)) \mid \eta \in \mathcal{H}\}.$$

$C$ is assumed to be compact, so $S$ is compact.

As $x$ varies from $-\infty$ to $+\infty$ the point $(I(-g,x), I(g,x))$ moves, from the origin, along the curve $C(x)$, and then back to the origin. $C(x)$ lies within the first quadrant of $\Re^2$ and defines a closed and bounded parametric curve representing the set of all possible design points.

Since $C(x)$ has a parametric representation, the formula for the signed curvature $\kappa_x$ of $C(x)$ has a simple form. Let $c_1 = I(-g,x)$ and $c_2 = I(g,x)$, and for $i = 1,2$ define $\dot{c}_i = \frac{\partial}{\partial x} c_i$ and $\ddot{c}_i = \frac{\partial^2}{\partial x^2} c_i$. Then also define

$$\kappa_x = \frac{\dot{c}_1\ddot{c}_2 - \dot{c}_2\ddot{c}_1}{(\dot{c}_1^2 + \dot{c}_2^2)^{\frac{3}{2}}}.$$

First consider $|g| \le ln(2+\sqrt{3})$. If $\kappa_x$ does not change sign as $x$ varies from $-\infty$ to $\infty$, then the boundary of $S$, $\partial S$, and the curve $C$ will coincide and $C$ will enclose a convex set. Since $\kappa_x$ is a continuous quadratic function of $e^x$, it may be shown that $\kappa_x$ does not change sign if and only if the equation $\kappa_x = 0$ has no real roots or one real root. This condition holds if and only if $|g| \le ln(2+\sqrt{3})$. The solid line in Figure 1 shows the curve $C(x)$ for $g=1$ and Figure 2 shows $C(x)$ for $g=2$.

For the prior distribution with mass $\frac{1}{2}$ at $\theta = \pm g$ and the design with all mass at $x$ the $\phi_2$-criterion function for a one point design $\eta_x$ is

$$\phi_2(\eta_x) = -[\frac{1}{2}\,I(-g,x)^{-1} + \frac{1}{2}\,I(g,x)^{-1}].$$

Denote the horizontal and vertical coordinates as $I_1$ and $I_2$ respectively and let

$$z = -[\frac{1}{2I_1} + \frac{1}{2I_2}]. \qquad (2)$$

For $I_1 > 0$ and $I_2 > 0$ and a fixed value of $z \in \Re$, the set of points $\{(I_1, I_2)\}$ satisfying equation 2 describes a strictly convex contour curve. Let $T_z$ be such a curve for a fixed value of $z$. From standard arguments in convex programming, in maximizing over $S$, when $z$ is a maximum, say $z^*$, there is one $T_{z^*}$ which intersects $\partial S$ at one point, $P$. This point represents the $\phi_2$-optimal design. When $|g| \le ln(2+\sqrt{3})$ it is easy to see, by symmetry and from Figure 1, that the $\phi_2$-optimal design is to put mass 1 at $x=0$.

It is also easy to see that $\phi_1$-optimality gives the same optimal design, as does any criterion $\phi(\cdot)$ with strictly convex contours which is also symmetric in the sense that $\phi(\eta_x) = \phi(\eta_{-x})$. Figure 1 shows the contour $T_{z^*}$ and the curve $C(x)$ for $\phi_2$-optimality and $g=1$.

If $|g| > ln(2+\sqrt{3})$, $C(x)$ and $\partial S$ no longer coincide, and $C(x)$ does not enclose a convex set. The solid line in Figure 2 is the curve $C(x)$ for $g=2$. $\partial S$ can be constructed by first constructing the unique line segment $L$ which joins two points

$C(-x)$ and $C(x)$, and which is also tangent to the curve at those points. $C(x)$ is symmetric with respect to the line $I_1 = I_2$ which has slope equal to 1, so $L$ must have a slope of -1. See Figure 2 for $g = 2$.

Setting the slope of the line tangent to $C(x)$ at $x$ equal to -1 and solving for $x$ results in the two roots $x = \pm B(g)$ as defined earlier, so $L$ is the line segment from the point $C(-B(g))$ to the point $C(B(g))$. Thus $\partial S$ consists of the arc of $C$ from the origin to the point $C(-B(g))$, the line segment $L$, and the arc of $C$ from $C(B(g))$ to the origin.

Using similar arguments as in the case when $|g| \leq ln(2+\sqrt{3})$, it is clear that for any criterion with convex contours and the symmetry property that $\phi(\eta_x) = \phi(\eta_{-x})$ the point $P$ where the optimal contour $T_{z^*}$ intersects $\partial S$ corresponds to a design with mass 1/2 at each of $x = \pm B(g)$. Figure 2 shows the contour $T_{z^*}$ for $g = 2$.

## 2.4 Discussion

In this special case of logistic regression with a symmetric two point prior distribution on $\theta$, the Bayesian $\phi_1$- and $\phi_2$-optimal designs are the same. This is not immediately clear using the algebraic approach. It is clear after noting that the algebraic forms of the directional derivatives at the optimal design are closely related. When $|g| \leq ln(2+\sqrt{3})$, the quadratic term $Q1$ appears in the numerators of the directional derivatives for both $\phi_1$- and $\phi_2$-optimality. Specifically, denote $d_2$ to be the directional derivative for $\phi_2$-optimality and $d_1$ to be the directional derivative for $\phi_1$-optimality. Then

$$d_2(\eta_0, x) = e^{-g}(e^g + 1)^2 \, d_1(\eta_0, x),$$

and the directional derivatives must have the same roots. Similarly, when $|g| > ln(2+\sqrt{3})$ the quadratic term $Q2$ appears in the numerators of both directional derivatives, and

$$d_2(\eta_2, x) = \frac{8(e^g - 1)^2(e^g + 1)^2}{(e^{2g} + 1)^2} \, d_1(\eta_2, x).$$

The symmetry of the information and of the prior distribution for $\theta$ account for much of the algebraic simplification.

In general, however, it is not the case that $\phi_1$- and $\phi_2$-optimal designs are the same. Haines (1995) gives conditions on the prior distribution for $\theta$ under which the $\phi_1$- and $\phi_2$-optimal designs are identical for this special case of logistic regression, and using the geometric approach these similarities are clearly apparent.

A connection between the geometric approach and the algebraic approach is now explained here for $\phi_2$-optimality in this example.

Since the line $L$ is tangent to $T_{z^*}$ at $P$ and is perpendicular to the gradient of $T_{z^*}$ at $P$, the points $(I_1, I_2)$ on $L$ satisfy:

$$\frac{-1}{2}\frac{1}{I(-g,\eta^*)^2}(I_1 - I(-g,\eta^*)) + \frac{-1}{2}\frac{1}{I(g,\eta^*)^2}(I_2 - I(g,\eta^*)) = 0.$$

As the set $S$ is contained in the lower half plane determined by $L$ substituting $I_1 = I(-g,x)$ and $I_2 = I(g,x)$ into the left hand side will result in an expression which must be nonpositive. But the resulting expression on the left hand side is $d_2(\eta^*, x)$ and so we have $d_2(\eta^*, x) \leq 0$ as required by the equivalence theorem.

Similar connections can be shown for $\phi_1$-optimality.

# 3 Asymmetric Prior Distributions

Now consider the case when the two-point prior distribution for $\theta$ is asymmetric in the sense that $\theta$ takes values which are symmetric about zero but the probabilities $\beta_1$ and $\beta_2$ are not equal.

The following theorem is Haines' theorem.

**Theorem 2** (Haines, 1995) *Let the prior distribution $\pi(\theta)$ be such that $\theta = -g$ with probability $\beta_1$ and $\theta = +g$ with probability $\beta_2 = 1 - \beta_1$. The $\phi_1$-optimal design is:*

*(1) If $|g| \leq ln(2+\sqrt{3})$, define*

$$R(g) = ln\left[\frac{\sqrt{(\beta_2-\beta_1)^2(e^{2g}-1)^2+4e^{2g}}}{2e^g}+\frac{(e^{2g}-1)(\beta_2-\beta_1)}{2e^g}\right].$$

*The $\phi_1$-optimal design measure puts mass 1 at the point $R(g)$.*

*(2) If $|g| > ln(2+\sqrt{3})$, define $\rho(g)$ to be the slope of the line from the origin to the point $C(B(g))$ and $w = \frac{\beta_2\rho(g)-\beta_1}{\rho(g)-1}$. If $\frac{\beta_2}{\beta_1}$ satisfies*

$$\frac{1}{\rho(g)} < \frac{\beta_2}{\beta_1} < \rho(g). \tag{3}$$

*the $\phi_1$-optimal design puts mass at $+B(g)$ and $-B(g)$ (the same $\pm B(g)$ as in Theorem 1) with associated weights $w$ and (1-w), respectively. If the inequality is not satisfied then the optimal design puts mass 1 at the point $R(g)$ above.*

**Haines' Geometric Proof of Theorem 2:**

**(1)** Let $|g| \leq ln(2+\sqrt{3})$.

When $|g| \leq ln(2+\sqrt{3})$, $C(x) = \partial S$, so $T^1_{z^*}$ touches $\partial S$ at exactly one point. So the optimal design must be a one point design and it is straightforward to show algebraically that $\phi_1(\eta_x)$ is maximized at $x = R(g)$.

**(2)** Let $|g| > ln(2+\sqrt{3})$.

When $|g| > ln(2+\sqrt{3})$, $C(x)$ and $\partial S$ do not coincide. Note that $C(x)$ and $\partial S$ depend only on the support points, $\pm g$, of the prior distribution and not on the probabilities, $\beta_1$ and $\beta_2$. The geometric argument is therefore very similar to that for the symmetric prior distribution, the only difference being the shape of the contours $T_z$. The points on $C(x)$, $C(B(g))$ and $C(B(-g))$, are the same two points as in the symmetric case.

The optimal contour curve $T_{z^*}$ must touch $\partial S$ at one point. So either the point corresponds to $x = R(g)$ or the optimal design has two support points $\pm B(g)$. All that remains is to find conditions on $\beta_1$ and $\beta_2$ which determine whether the optimal design is a one- or two-point design.

By symmetry considerations of $\partial S$, we know that the slope of the line $L$ is -1. So the slope of the line tangent to $T_{z^*}$ at that point where it touches $\partial S$ on $L$ is also -1. The locus of all points on the family of curves $\{T_z \mid z \in \Re\}$ must be found at which the slope of the line tangent to $T_z$ is -1.

For a fixed value of $z$, and $\phi_1$-optimality, a convex contour curve is defined by the equation

$$z = \beta_1 \, lnI_1 + \beta_2 \, lnI_2. \tag{4}$$

The slope of the line tangent to this curve is

$$-\frac{\beta_1}{\beta_2} e^{\frac{z}{\beta_2}} I_1^{-\frac{\beta_1}{\beta_2}-1}.$$

Setting this slope equal to -1 and substituting (4) for $z$ results in the following equation:

$$I_2 = \frac{\beta_2}{\beta_1} \, I_1.$$

Therefore, the locus of all points at which the slope of the line tangent to $T_z$ is -1 is a line through the origin with slope $\frac{\beta_2}{\beta_1}$. Denote this line as $L''$.

Let $\rho(g)$ be the slope of a line from the origin to the point C(B(g)). Then the line $L''$ will intersect the line segment $L'$ if and only if (3) holds. The point of intersection is a weighted combination of the points $C(+B(g))$ and $C(-B(g))$. The weight $w$ is easily found to be $\frac{\beta_2 \rho(g) - \beta_1}{\rho(g) - 1}$. Therefore the $\phi_1$-optimal design is to put weight $w$ at the optimal design point $B(g)$ and weight $1 - w$ at the optimal design point $-B(g)$.

If condition (3) does not hold, then $T_{z^*}$ and $\partial S$ intersect at the point $R(g)$ on $C$, so the $\phi_1$-optimal design is to put mass 1 at $R(g)$. This completes the geometric proof.

An algebraic proof for an asymmetric prior distribution has been found to be intractable so far. The geometric approach has provided optimal designs when the algebraic approach has not.

For $\phi_2$-optimality or any other criterion with convex contours the geometric argument shows that the optimal design is either a one point design or a two point design with support at $\pm B(g)$, see Haines (1995).

## 3.1 Discussion

The shape of the curve $C(x)$ depends only on the support points of the prior distribution and not the probabilities. The contour curves $T_z$, however, depend on the probabilities for $\theta$. The curves will be moved closer to the axis which represents the Fisher information with the larger probability for $\theta$. Figure 3 illustrates the geometric proof for $\phi_1$-optimality, $\theta = \pm 1$, and $\beta_2 = 0.75$, and Figure 4 illustrates the proof when $\theta = \pm 2$ and $\beta_2 = 0.75$.

Consider Figure 4 for $\theta = \pm 2$ and consider $\beta_2$ getting larger. As $\beta_2$ gets larger the curve $T_{z^*}$ will tilt toward the $I_2$ axis, eventually touching $\partial S$ at one point on $C$ outside of the line segment $L$. In that case, a one-point design will be optimal and it can be shown algebraically that this is the point $C(R(2))$ as defined in the statement of Theorem 2.

Suppose $\beta_2$ is large enough so that a one-point design is optimal. As $\beta_2$ gets still larger, the slope of the line tangent to $T_{z^*}$ must approach 0 since $T_{z^*}$ must be strictly convex. It can also be shown that when $\frac{\beta_2}{\beta_1} = \rho(g)$, $R(g) = B(g)$ so the change from a one-point $\phi_1$-optimal design to a two-point $\phi_1$-optimal design occurs in a continuous manner.

This example is interesting because when a two point design is optimal, the design is a weighted combination of the same two optimal design points $\pm B(g)$ found with a symmetric prior, a fact which is not apparent when using algebraic methods but is obvious with the geometric approach. The strength of the geometric approach is that it contributes to an intuitive understanding of the problem.

# 4 Three Point Prior Distribution

In this section, the optimal design problem for the special case of logistic regression is examined for a three-point, symmetric, prior distribution on $\theta$. A algebraic results are presented and the geometric argument discussed.

## 4.1 The Algebraic Approach

**Theorem 3** *Let the prior distribution be such that* $\theta = \{\pm g, 0\}$ *each with probability* $\frac{1}{3}$. *If* $|g| \leq ln(3+2\sqrt{2})$ *and* 0 *is in* $\mathcal{X}$, *the* $\phi_1$*-optimal design measure puts mass 1 at 0.*

**Proof of Theorem 3:**

By symmetry the best one-point design puts mass 1 at $x = 0$. Define

$$\begin{aligned} Q3 &= (3e^{2g})(e^{4x}) + (-e^{4g} + 4e^{3g} + 6e^{2g} + 4e^{g} - 1)(e^{3x}) \\ &\quad + (-e^{4g} + 4e^{3g} + 12e^{2g} + 4e^{g} - 1)(e^{2x}) \\ &\quad + (-e^{4g} + 4e^{3g} + 6e^{2g} + 4e^{g} - 1)(e^{x}) + (3e^{2g}). \end{aligned}$$

Then the derivative at $\eta_0$ in the direction $x$ is

$$d(\eta_0, x) = \frac{-(e^x - 1)^2(Q3)}{3(e^x+1)^2(e^x+e^g)^2(e^{x+g}+1)^2}.$$

$Q3$ is a quartic in $e^x$. If $Q3$ is strictly positive, the derivative will have one root at x = 0.

Let the standard form of a quartic in $y$ be

$$ay^4 + 4by^3 + 6cy^2 + 4dy + f.$$

Further let $y = e^x$ and write $Q3$ in standard form with:

$$\begin{aligned} a &= (3e^{2g}) \\ 4b &= 4[\frac{(-e^{4g} + 4e^{3g} + 6e^{2g} + 4e^g - 1)}{4}] \\ 6c &= 6[\frac{(-e^{4g} + 4e^{3g} + 12e^{2g} + 4e^g - 1)}{6}] \\ 4d &= 4[\frac{(-e^{4g} + 4e^{3g} + 6e^{2g} + 4e^g - 1)}{4}] \\ f &= +(3e^{2g}). \end{aligned}$$

Then from standard theory of equations (for example Barnard and Child, 1964, p. 186), let $H = ac - b^2$, $I = af - 4bd + 3c^2$, $J = acf + 2bcd - ad^2 - c - fb^2$, and $\Delta = I^3 - 27J^2$. Conditions on $g$ must be satisfied so that the quartic is strictly positive; that is, all four roots are imaginary. This will happen if and only if $\Delta > 0$ and at least one of $H$ and $2HI - 3aJ$ is positive.

Now

$$H = -\frac{(e^g - 1)^4 (e^g + 1)^2 (e^{2g} - 6e^g + 1)}{16},$$

so $H$ has roots at $g = 0$ and at $g = \pm ln(3 + 2\sqrt{2})$. When $|g| < ln(3 + 2\sqrt{2})$, then $(e^{2g} - 6e^g + 1)$ will be negative and $H$ will be positive. Similarly,

$$\Delta = -\frac{3(e^g - 1)^{12}(e^g + 1)^2(e^{2g} - 6e^g + 1)(e^{4g} - 4e^{3g} - 6e^{2g} - 4e^g + 1)^2}{256},$$

so $\Delta$ will also be positive when $|g| < ln(3+2\sqrt{2})$. That is, when $|g| < ln(3+2\sqrt{2})$, both $\Delta$ and $H$ will be strictly positive, so the quartic will have four imaginary roots and thus be strictly positive.

Therefore, when $|g| \le ln(3 + 2\sqrt{2})$ the derivative $d(\eta_0, x)$ is non-positive and has a single root at $x = 0$, thus verifying that the best one-point design is the $\phi_1$-optimal design.

This completes the proof of Theorem 3.

Three point prior distributions with $|g| > ln(3 + 2\sqrt{2})$ are examined in more detail in Agin (1997). She shows that for $|g| > ln(3 + 2\sqrt{2})$ and $|g|$ simultaneously less than about 2.29 a two point design is $\phi_1$-optimal. She derives a lengthly closed form algebraic expression for the optimal design points and shows that the transition from one to two points happens continuously. She also shows numerically that for $|g|$ greater than about 2.29 a three point design is $\phi_1$-optimal, with the weights not necessarily being equal. The transition from two to three points is continuous.

For $\phi_2$-optimality Agin (1997) also gives the following result and proves it using the algebraic approach.

**Theorem 4** *Let the prior distribution be such that $\theta = \{\pm g, 0\}$ each with probability $\frac{1}{3}$. If $|g| \leq ln(\frac{5+\sqrt{21}}{2})$ and 0 is in $\mathcal{X}$, the $\phi_2$-optimal design measure puts mass 1 at 0.*

The proof again involves finding the roots of a quartic polynomial in $e^x$ and is a straightforward application of the equivalence theorem but algebraically rather cumbersome.

## 4.2 The Geometric Approach

For this prior distribution with three support points, at $\theta = 0$ and $\theta = \pm g$, a geometric approach seems initially promising. Figure 5 shows the curve $C(x)$ for $g = 1$ and it is immediately clear that a one point design must be optimal. Similarly, Figure 6 shows $C(x)$ for $g = 3.5$ and it is clear that a one point design is not necessarily optimal. Deriving results geometrically, however, has proved intractable. For $|g| > ln\,(3 + 2\sqrt{2})$, it is not easy to define the convex hull of $C$ or specify its boundary. Both the algebraic and geometric approach are difficult to use in three dimensions.

# 5 Conclusion

In this paper the geometric approach and the algebraic approach to finding closed form optimal designs for a one parameter logistic regression model have been reviewed and discussed. Although closed form solutions are important in understanding the problem in general, they are surprisingly intractable to find. The two different approaches complement each other and help give an intuitive understanding of the problem. Agin (1997) numerically examines prior distributions with up to $m = 30$ support points, evenly spaced around zero with equal probability at each point. She shows that as the number of support points, $m$, gets larger the range for which a one point design at zero is optimal also gets larger. This is consistent with the results for $m = 2$ and $m = 3$ presented here. Algebraic results, however, have proved intractable for other than these two cases.

## Acknowledgement

This work was partly supported by a grant from the National Security Agency.

## REFERENCES

Agin, M.A. (1997). *Optimal Bayesian Designs for Nonlinear Models.* University of Minnesota, School of Statistics, PhD thesis.

Barnard, S. and J.M. Child. (1964) *Higher Algebra.* Macmillan & Co., Ltd., London.

Chaloner, K. (1984). Optimal Bayesian experimental design for linear models. *Ann. Statist.* 7, 283-300.

Chaloner, K. (1987). An approach to design for generalized linear models. In *Model Oriented Data Analysis*, V.V. Fedorov and H. Läuter eds.. Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, 3-12.

Chaloner, K. (1993). A note on optimal Bayesian design for nonlinear problems. *Journal of Statistical Planning and Inference* **37** 229-235.

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

Chaloner, K. and Verdinelli, I. (1995). "Bayesian experimental design: a review", *Statistical Science*, 10, 273–304.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Ann. Math. Statist.* 24, 586-602.

Haines, L.M. (1995). A geometric approach to optimal design for one-parameter non-linear models. *J. Roy. Statist. Soc. Ser. B* **57** 575-598.

Marlow, W.H. (1978). *Mathematics for Operations Research.* John Wiley and Sons, Inc., New York.

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

Tsutakawa, R.K. (1972). Design of an experiment for bioassay. *J. Amer. Statist. Assoc.* 67, 584-590.

Whittle, P. (1973). Some general points in the theory and construction of D-optimum experimental designs. *J. Roy. Statist. Soc. Ser. B* 35, 123-130.

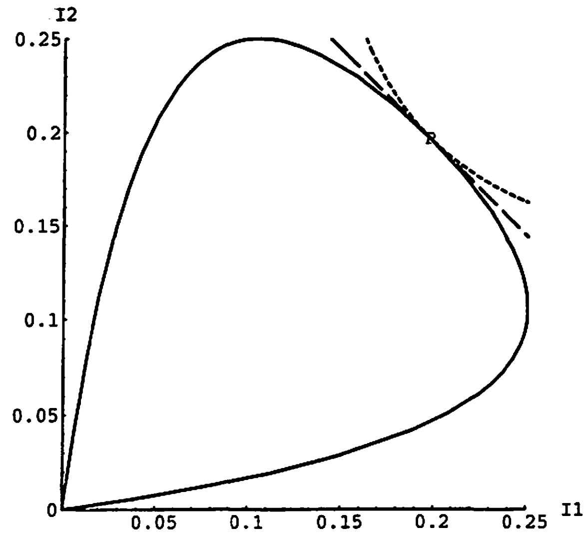


Figure 1: Symmetric Prior Distribution, $\theta = \pm 1$ (a special case of $|g| \leq ln(2+\sqrt{3})$) and $\phi_2$-optimality: $C$ is the solid line, $L$ the dashed line and $T_{z^*}$ is the dotted line. For $z^* = -5.08616$, $C$ and $T_{z^*}$ intersect at $P$.

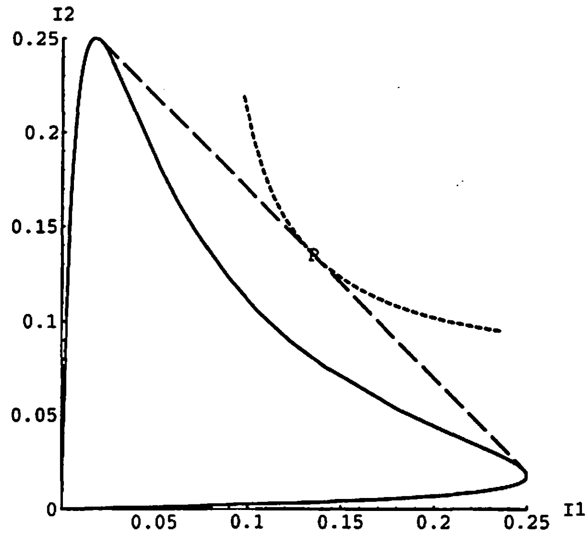


Figure 2: Symmetric Prior Distribution, $\theta = \pm 2$, a special case of $|g| > ln(2+\sqrt{3})$ and $\phi_2$-optimality: $C$ is the solid line, $L$ the dashed line and $T_{z^*}$ is the dotted line. For $z^* = -7.43479$ the maximum value of $\phi_2(\cdot)$, $T_{z^*}$ and $\partial S$ intersect at $P$.

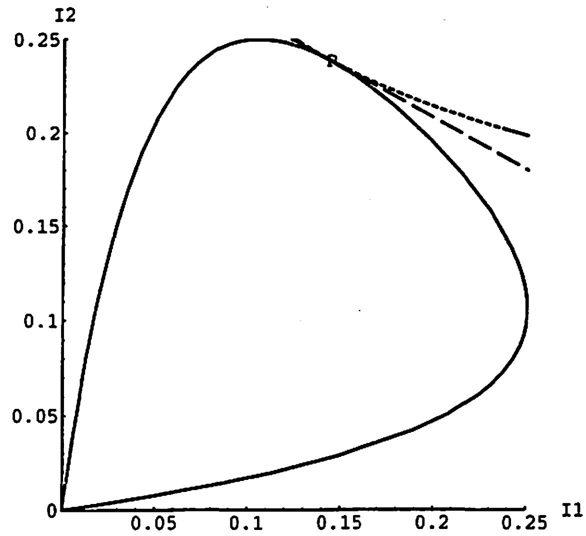


Figure 3: Asymmetric Prior Distribution, $\beta_1 = P(\theta = -1) = .25$: $C$ is the solid line, $L$ the dashed line and $T_{z^*}$ is the dotted line. The contour $T_{z^*}$ corresponds to $z^* = -4.85365$.

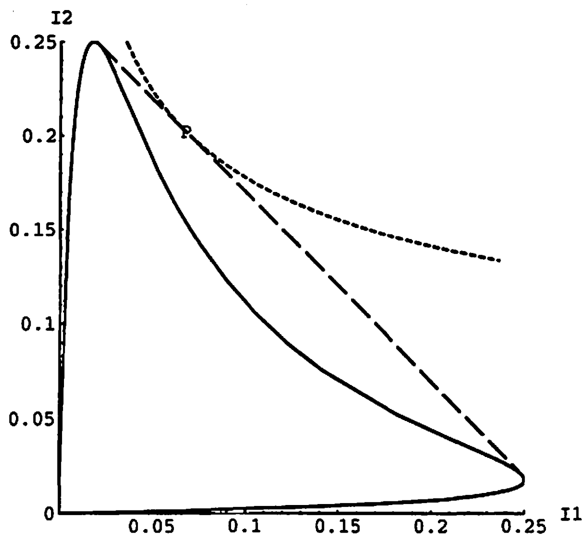


Figure 4: Asymmetric Prior Distribution, $\beta_1 = P(\theta = -2) = .25$: $C$ is the solid line, $L$ the dashed line and $T_{z^*}$ is the dotted line. The contour corresponds to $z^* = -6.93676$.

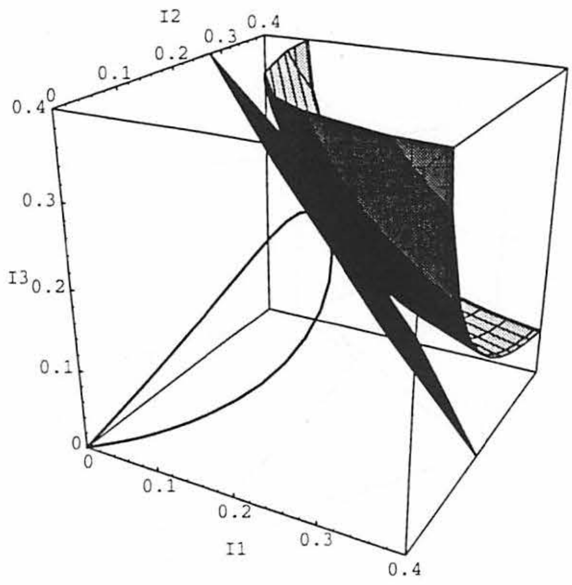


Figure 5: The curve $C$, supporting hyperplane $H$, and convex surface $T_{z^*}$ for a three point symmetric prior distribution with $g = 3$ under $\phi_1$-optimality. The optimal design is a one point design at $x = 0$.

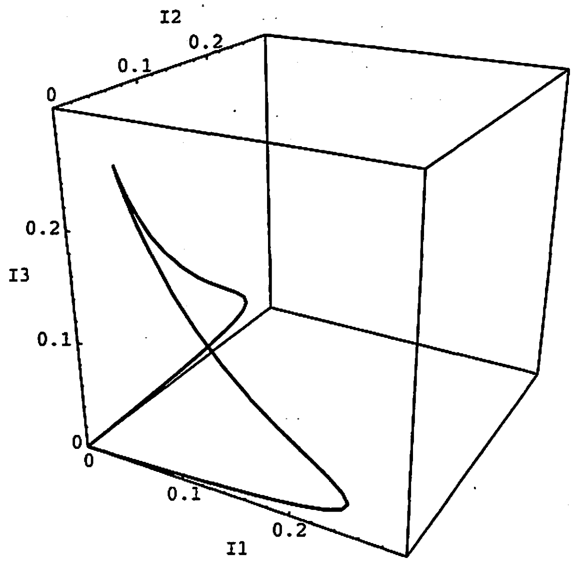


Figure 6: The curve $C(x)$ for a three point symmetric prior distribution with $g = 3.5$.