Chapter 5

EQUIVALENCE THEORY FOR BAYESIAN OPTIMAL DESIGNS

§5.1 Introduction

The majority of the Bayesian designs detailed in the previous chapter were obtained via numerical optimization and simulation. Since the designs were not found by exact mathematical methods, it is reasonable for one to question their optimality. When faced with this problem in the non-Bayesian case, equivalence theory is a common tool used to verify the optimality of a design. Equivalence theory functions in the same capacity for Bayesian designs as well. Chaloner and Larntz (1989) have applied Bayesian equivalence theory to D-optimal designs in the single regressor logistic setting (1989). Heise and Myers (1995) extended it to the bivariate logistic case. In addition, this technique has been used to demonstrate optimality for designs based on a Bayesian linear criterion function for compartmental models (Atkinson, et. al, 1993). This chapter focuses on verifying the optimality of the two level Bayesian D-and F-optimal designs discussed in Chapters 3 and 4.

Traditional equivalence theory combines techniques from measure theory and calculus to show that a particular design satisfies a given optimality criterion. Kiefer and Wolfowitz first developed equivalence theory in 1959 for the case of D-optimality. Their results were generalized to non-linear models by Federov(1972). In turn, Silvey (1980) extended equivalence theory to cover more optimality criteria for both linear and non-linear models. These include D-optimality and optimality criteria known as "linear criteria functions". F-optimality falls into the latter category. Equivalence theory will be applied to selected D- and F-optimal Bayesian two level designs based on both the uniform and normal priors to demonstrate their optimality. Keep in mind that the only truly optimal designs are two level designs. Without restrictions, the three level design criterion chose two level designs in all cases. While all two level designs were verified as optimal, only select cases are shown for the sake of brevity.

§5.2 Equivalence Theory

In order to understand equivalence theory for Bayesian designs an overview of more traditional equivalence theory is necessary. This exposition on equivalence theory is based on Silvey's results and will consequently use his notation. Consider the model

$$y = f(\mathbf{x}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon} \tag{5.2.1}$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed unknown model parameters and $f(\mathbf{x}, \boldsymbol{\beta})$ is a linear or non-linear function. Suppose there is interest in designing an experiment to estimate $\boldsymbol{\beta}$ using a particular design criterion ϕ. The ϕ–optimal design is defined as the design which maximizes the criterion ϕ. (Note that if the design criterion is one which suggests the minimization of some function ψ then ϕ is taken to be the negative of ψ.) Thus, the formulation of the optimal design requires one to select a design with N observations which maximizes ϕ.

An N-observation design is defined as containing m distinct points $\mathbf{x}_1, \ldots, \mathbf{x}_m$ where each point is replicated $r_1, \ldots, r_m$ times respectively where $\sum r_i = N$. This design can then be characterized by a discrete probability distribution on $\mathcal{X}$, the set of all vectors, **x**, in the design space. This distribution denoted by η_N weights $\mathbf{x_i}$ with probability $p_i = \frac{r_i}{N}$ for i=1, . . . , m.

Now consider a random vector **x** with discrete distribution η_N and define the Fisher information matrix of a *single observation* taken at the point **x** as $\mathbf{J}(\mathbf{x},\beta)$. Define

$$\mathbf{M}(\eta_N,\beta) = \mathrm{E}(\mathbf{J}(\mathbf{x},\beta)) = \frac{1}{N}\mathbf{I}(\mathbf{X},\beta) \tag{5.2.2}$$

which implies that $N\mathbf{M}(\eta_N,\beta) = \mathbf{I}(\mathbf{X},\beta)$. The N-observation design problem then becomes that of finding a probability distribution η_N^* which maximizes $\phi(\mathbf{M}(\eta_N,\beta))$. However, this is not an easy task. Due to its discrete nature, the set $\mathcal{M}_N = \{\mathbf{M}(\eta_N,\beta)\}$ for which ϕ is defined does not lend itself to standard numerical analysis optimization techniques (Silvey, pg. 13). This problem can be remedied by defining H as the set of all η on $\mathcal{X}$. This modification allows the use of calculus techniques to find the f-optimal design η^* which maximizes $\phi(\mathbf{M}(\eta,\beta))$ over all H. This is known as the approximate theory problem meaning that the "aim [of optimal design] is to approximate η^* by a design measure corresponding to an N-observational design" (Silvey, pg. 16).

Let us now shift to the mathematical details which lead to equivalence theory. The first topic to be discussed is that of the Frechet derivative. The Frechet directional derivative of ϕ at $\mathbf{M}_1$ in the direction of $\mathbf{M}_2$ is

$$F_\phi(\mathbf{M}_1,\mathbf{M}_2) = \lim_{\varepsilon\to 0^+}\frac{1}{\varepsilon}\left[\phi\{(1-\varepsilon)\mathbf{M}_1 + \varepsilon\mathbf{M}_2\} - \phi(\mathbf{M}_1)\right]. \tag{5.2.3}$$

Silvey proves the following theorem using the Frechet derivative and defining $\mathcal{M} = \{M(\eta,\beta) : \eta \in H\}$:

Theorem 1: If β is fixed and if ϕ is increasing and concave on $\mathcal{M}$ and differentiable at $\mathbf{M}(\eta^*,\beta)$, then η^* is f-optimal if and only if $F_\phi\left[\mathbf{M}(\eta^*,\beta),\mathbf{J}(\mathbf{x},\beta)\right]\le 0$ for all $\mathbf{x}\in\mathcal{X}$ (Silvey, pg. 54).

In order to apply this to Bayesian optimal designs, the substitution must be made for the appropriate criterion. One may question the application of the theorem to the Bayesian case because it restricts b to be fixed and all Bayesian designs are based on the premise that β may vary across some range. However, since optimization occurs after integration over the prior, β is effectively fixed.

Another important result of equivalence theory pertains to the value of the Frechet derivative at the design points. If $\mathbf{x}_1,\ldots,\mathbf{x}_m$ are design points that make up the support of η^* then in order for the design to be optimal,

$$F_\phi\left[\mathbf{M}(\eta^*,\beta),\mathbf{J}(\mathbf{x}_i,\beta)\right]=0 \qquad \forall\ i,\ i=1,\ldots,m. \tag{5.2.4}$$

In other words, the Frechet derivative must be equal to zero at the design points. In summary, in order to verify the optimality of a particular design, η^*, one must evaluate the Frechet derivative at that design for all points in the design space. In turn, the Frechet derivative must be less than or equal to zero at all points in the design space and equal to zero at the design points.

§5.3 Equivalence Theory for Bayesian D-Optimal Designs in the Poisson Case

The first step in verifying the optimality of a design is the definition of the criterion, ϕ. Typically, the Bayesian D-optimality criterion is

$$\phi=\int R(\delta,\beta)\pi(\beta)d\beta=\int\det\mathbf{M}(\eta,\beta)\pi(\beta)d\beta\,. \tag{5.3.1}$$

In order to ensure the concavity required by Theorem 1, an equivalent form of the criterion is used,

$$\phi=\int R(\delta,\beta)\pi(\beta)d\beta=\log\int\det\mathbf{M}(\eta,\beta)\pi(\beta)d\beta\,. \tag{5.3.2}$$

With f defined, the Frechet derivative is given by

$$F_{\phi}\left[\mathbf{M}(\eta,\beta),\mathbf{J}(\mathbf{x},\beta)\right]=\lim_{\varepsilon\to 0^{+}}\frac{1}{\varepsilon}\left[\log\frac{\int\det\{(1-\varepsilon)\mathbf{M}(\eta,\beta)+\varepsilon\mathbf{J}(\mathbf{x},\beta)\}\pi(\beta)d\beta}{\int\det\mathbf{M}(\eta,\beta)\pi(\beta)d\beta}\right]. \tag{5.3.3}$$

where $\mathbf{M}(\eta,\beta)=\begin{bmatrix}\sum\lambda_i & \sum\lambda_i\mathbf{x}_i\\ \sum\lambda_i\mathbf{x}_i & \sum\lambda_i\mathbf{x}_i^2\end{bmatrix}$ and $\mathbf{J}(\mathbf{x},\beta)=\begin{bmatrix}\lambda_j & \lambda_j\mathbf{x}_j\\ \lambda_j\mathbf{x}_j & \lambda_j\mathbf{x}_j^2\end{bmatrix}$. The numerator of this argument of the log in (5.3.3) can be expressed as

$$\int\det\ \mathbf{M}(\eta,\beta)\pi(\beta)d\beta+\varepsilon\int Q_1\pi(\beta)d\beta+\varepsilon^2\int Q_2\pi(\beta)d\beta \tag{5.3.4}$$

where

$$Q_1=n_1\lambda_1\lambda_j\left(x_j-x_1\right)^2+n_2\lambda_2\lambda_j\left(x_j-x_2\right)^2-2n_1n_2\lambda_1\lambda_2\left(x_1-x_2\right)^2 \tag{5.3.5}$$

and

$$Q_2=n_1n_2\lambda_1\lambda_2\left(x_1-x_2\right)^2-n_1\lambda_1\lambda_j\left(x_j-x_1\right)^2-n_2\lambda_2\lambda_j\left(x_j-x_2\right)^2. \tag{5.3.6}$$

Equation (5.3.3) then becomes

$$F_{\phi}\left[\mathbf{M}(\eta,\beta),\mathbf{J}(\mathbf{x},\beta)\right]=\lim_{\varepsilon\to 0^{+}}\frac{1}{\varepsilon}\log\left[1+\frac{\varepsilon\int Q_1\pi(\beta)d\beta+\varepsilon^2\int Q_2\pi(\beta)d\beta}{\int\det\mathbf{M}(\eta,\beta)\pi(\beta)d\beta}\right]. \tag{5.3.7}$$

Taking the limit of the above expression yields the indeterminate form of $\frac{0}{0}$. Applying L'Hopital's rule results in the final form of the Frechet derivative shown in (5.3.8).

$$F_{\phi}\left[\mathbf{M}\left(\eta^{*},\beta\right),\mathbf{J}(\mathbf{x},\beta)\right]=\frac{\int Q_1\pi(\beta)d\beta}{\int \det \mathbf{M}\left(\eta^{*},\beta\right)\pi(\beta)d\beta} \tag{5.3.8}$$

$$=\frac{\int n_1\lambda_1\lambda_j\left(x_j-x_1\right)^2\pi(\beta)d\beta+\int n_2\lambda_2\lambda_j\left(x_j-x_2\right)^2\pi(\beta)d\beta}{\int n_1n_2\lambda_1\lambda_2\left(x_1-x_2\right)^2\pi(\beta)d\beta}-2$$

Substituting $\lambda_i = e^{\beta_0+\beta_1 x_i}$ in (5.3.8) gives the following:

$$F\left(\mathbf{M}\left(\eta^{*},\beta\right),\mathbf{J}(\mathbf{x},\beta)\right)=\frac{\int n_1 e^{\beta_1\left(x_1+x_j\right)}\left(x_j-x_1\right)^2\pi\left(\beta_1\right)d\beta+\int n_2 e^{\beta_1\left(x_2+x_j\right)}\left(x_j-x_2\right)^2\pi\left(\beta_1\right)d\beta_1}{\int n_1n_2 e^{\beta_1\left(x_1+x_2\right)}\left(x_1-x_2\right)^2\pi\left(\beta_1\right)d\beta_1}-2 \tag{5.3.9}$$

After integration over the uniform prior on β_1, the optimal design must satisfy the relationship in (5.3.10) derived from the Frechet derivative in (5.3.3). Let the left hand side of this equation be known as *f*.

$$\frac{n_1\left(x_1+x_2\right)\left(x_2+x_j\right)\left(x_j-x_1\right)^2\left(e^{d\left(x_1+x_j\right)}-e^{c\left(x_1+x_j\right)}\right)+n_2\left(x_1+x_2\right)\left(x_1+x_j\right)\left(x_j-x_2\right)^2\left(e^{d\left(x_2+x_j\right)}-e^{c\left(x_2+x_j\right)}\right)}{n_1n_2\left(x_1-x_2\right)^2\left(x_1+x_j\right)\left(x_2+x_j\right)\left(e^{d\left(x_1+x_2\right)}-e^{c\left(x_1+x_2\right)}\right)}\leq 2 \tag{5.3.10}$$

Two examples of the applications of equivalence theory in verifying optimality of designs will now be presented. From Section 5.2, recall that $F_{\phi}\left[\mathbf{M}\left(\eta^{*},\beta\right),\mathbf{J}(\mathbf{x},\beta)\right]\leq 0$ for all $\mathbf{x}\in\mathcal{X}$ and $F_{\phi}\left[\mathbf{M}\left(\eta^{*},\beta\right),\mathbf{J}(\mathbf{x}_i,\beta)\right]=0 \quad \forall\ i,\ i=1,\ldots,m$. For the D-optimal single regressor case, this means that $f \leq 2$ for all $\mathbf{x}\in\mathcal{X}$ and $f = 2$ at the design points. With this in mind, recall the D-optimal design where $\beta \sim U(-2.4,-0.6)$ from Table 3.5.1. One half of the observations are placed at the $EC_{6.07}$ and the remaining half are placed at the control. The graph of f for this design in Figure 5.3.1 verifies its optimality since $f = 2$ at the $EC_{6.07}$ and the control, and $f \leq 2$ throughout the design space.

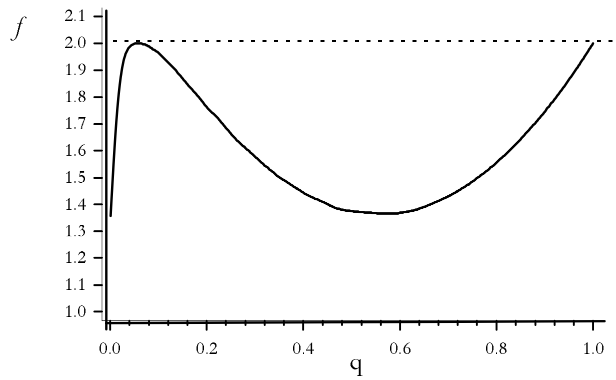

Figure 5.3.1 Graph Of f for a D-Optimal Design Based on a Uniform Prior with a Ratio of 4.

An example of equivalence theory in practice is also shown for a Bayesian D-optimal design in the normal case. However, in this situation the integrals for the equivalence theory function in (5.3.9) cannot be found in closed form. Hence, Monte Carlo integration is used to simulate their values. Figure 5.3.2 is the graph of f which results from the design where $\sigma = 0.3214$. This design places 50% of the experimental units at the $EC_{11.64}$ and 50% at the control.

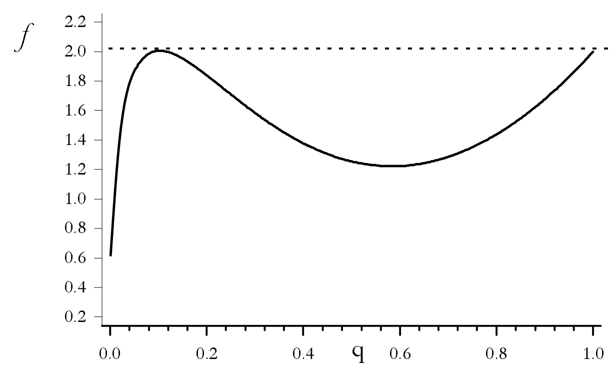

Figure 5.3.2 Graph of f for a D-Optimal Design Based on a Normal Prior where σ_b=0.3214.

Again, $f \leq 2$ for all ECs as required in Theorem 1 and $f = 2$ at the design points as characterized by (5.2.4). Thus, the design is optimal.

§5.4 Equivalence Theory for Bayesian F-Optimal Designs

As with the D-case, traditional F-optimality will be explored prior to its application to the Bayesian case. Many optimality functions, including that of F-optimality, can be described as linear criteria functions (Silvey, pg. 48). Consider the linear criterion function

$$\phi\left(\mathrm{M}(\eta,\beta)\right) = -\mathrm{tr}\left(\mathbf{A}'\mathbf{M}(\eta,\beta)^{-1}\mathbf{A}\right). \qquad (5.4.1)$$

where $\mathbf{A}$ is a $p \times s$ matrix of rank s. If $\mathbf{A} = \mathbf{a}'$, a vector, the value of f is the variance of the linear form $\mathbf{a}'\mathbf{b}$. In the F-optimality case, the goal is to select a design which minimizes var(b_1). Consider the following expression for the var(b_1)

$$\begin{aligned} \mathrm{var}(b_1) &= [0 \;\; 1]\mathbf{I}(\mathbf{X},\beta)^{-1}\begin{bmatrix}0\\1\end{bmatrix} \\ &= [0 \;\; 1]N^{-1}\mathbf{M}(\eta,\beta)^{-1}\begin{bmatrix}0\\1\end{bmatrix} \\ &= \mathbf{a}'\mathbf{M}(\eta,\beta)^{-1}\mathbf{a} \end{aligned} \qquad (5.4.2)$$

where $\mathbf{a}' = \left[\mathbf{0} \;\; N^{\frac{-1}{2}}\right]$. Of course, (5.4.2) is the criterion that would be minimized for a traditional F-optimal design whereas the current interest lies in the Bayesian design. The Bayesian F-optimality criterion can then be expressed as

$$\phi = \int -\,\mathrm{tr}\left(\mathbf{a}'\mathbf{M}(\eta,\beta)^{-1}\mathbf{a}\right)\pi(\beta)d\boldsymbol{\beta}\,. \qquad (5.4.3)$$

Now, the Frechet derivative must be found. This time the Frechet derivative will be pursued in a slightly different manner to facilitate the mathematical manipulations involved in the derivation. To do this, a new directional derivative must be introduced. The general form of the Gateaux derivative of ϕ at $\mathbf{M}_1$ in the direction of $\mathbf{M}_2$ is

$$G_{\phi}(\mathbf{M}_1, \mathbf{M}_2) = \lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon}\left\{\phi(\mathbf{M}_1 + \varepsilon \mathbf{M}_2) - \phi(\mathbf{M}_1)\right\} \tag{5.4.4}$$

(Silvey, pg. 17). Substituting the Bayesian F-optimality criterion, the Gateaux derivative is

$$G_{\phi}(\mathbf{M}(\eta,\beta), \mathbf{J}(\mathbf{x},\beta)) = \lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon}\left\{\int -\operatorname{tr}\left(\mathbf{a}'\left(\mathbf{M}(\eta,\beta) + \varepsilon \mathbf{J}(\mathbf{x},\beta)\right)^{-1}\mathbf{a}\right)\pi(\beta)d\beta - \int -\operatorname{tr}\left(\mathbf{a}'\mathbf{M}(\eta,\beta)^{-1}\mathbf{a}\right)\pi(\beta)d\beta\right\}. \tag{5.4.5}$$

Taking the limit yields an indeterminate form so L'Hopital's rule is employed once again to obtain the limit. Thus,

$$G_{\phi}(\mathbf{M}(\eta,\beta), \mathbf{J}(\mathbf{x},\beta)) = \int \operatorname{tr}\left(\mathbf{a}'\left(\mathbf{M}(\eta,\beta)^{-1}\mathbf{J}(\mathbf{x},\beta)\mathbf{M}(\eta,\beta)^{-1}\right)\mathbf{a}\right)\pi(\beta)d\beta\,. \tag{5.4.6}$$

The translation to the Frechet derivative is easily made through the following relationship

$$F_{\phi}(\mathbf{M}_1, \mathbf{M}_2) = G_{\phi}(\mathbf{M}_1, \mathbf{M}_2 - \mathbf{M}_1). \tag{5.4.7}$$

(Silvey, pg. 21). So the Frechet derivative is given by

$$F_{\phi}(\mathbf{M}(\eta,\beta), \mathbf{J}(\mathbf{x},\beta)) = \int \operatorname{tr}\left(\mathbf{a}'\left(\mathbf{M}(\eta,\beta)^{-1}\mathbf{J}(\mathbf{x},\beta)\mathbf{M}(\eta,\beta)^{-1}\right)\mathbf{a}\right)\pi(\beta)d\beta - \int \operatorname{tr}\left(\mathbf{a}'\mathbf{M}(\eta,\beta)^{-1}\mathbf{a}\right)\pi(\beta)d\beta\,.$$

According to Theorem 1 the optimal design will then satisfy the following relationship

$$\int \operatorname{tr}\left(\mathbf{a}'\left(\mathbf{M}(\eta,\beta)^{-1}\mathbf{J}(\mathbf{x},\beta)\mathbf{M}(\eta,\beta)^{-1}\right)\mathbf{a}\right)\pi(\beta)d\beta - \int \operatorname{tr}\left(\mathbf{a}'\mathbf{M}(\eta,\beta)^{-1}\mathbf{a}\right)\pi(\beta)d\beta \le 0\,. \tag{5.4.9}$$

Now, for the Poisson exponential model, the first term of the inequality in (5.4.9) is

$$\int \mathrm{tr}\Big(\mathbf{a}'\big(\mathbf{M}(\eta,\beta)^{-1}\mathbf{J}(\mathbf{x},\beta)\big)\mathbf{M}(\eta,\beta)^{-1}\mathbf{a}\Big)\pi(\beta)d\beta = \int \frac{\lambda_j \beta_1^2\left(n_1\lambda_1 \ln\frac{\lambda_j}{\lambda_1} + n_1\lambda_1 \ln\frac{\lambda_j}{\lambda_2}\right)^2}{n_1^2 n_2^2 \lambda_1^2 \lambda_2^2\left(\ln\frac{\lambda_j}{\lambda_2}\right)^4}\pi(\beta)d\beta \tag{5.4.10}$$

.

Substituting, $\lambda_i = e^{\beta_0+\beta_1 x_i}$, the following expression is obtained

$$\int \mathrm{tr}\Big(\mathbf{a}'\big(\mathbf{M}(\eta,\beta)^{-1}\mathbf{J}(\mathbf{x},\beta)\big)\mathbf{M}(\eta,\beta)^{-1}\mathbf{a}\Big)\pi(\beta)d\beta = \int \frac{e^{\beta_1 x_j}\left(n_1 e^{\beta_1 x_1}\left(x_j - x_1\right)^2 + n_2 e^{\beta_1 x_2}\left(x_j - x_2\right)^2\right)^2}{\beta_1^2 n_1^2 n_2^2 e^{\beta_0} e^{2\beta_1 x_1} e^{2\beta_1 x_2}\left(x_1 - x_2\right)^4}\pi(\beta)d\beta$$

(5.4.11)

The second term on the left hand side of (5.4.9) is

$$\int \mathrm{var}(b_1)\pi(\beta)d\beta = \int \frac{\beta_1^2\left(n_1\lambda_1 + n_2\lambda_2\right)}{n_1 n_2 \lambda_1 \lambda_2\left(\ln\frac{\lambda_1}{\lambda_2}\right)^2}\pi(\beta)d\beta\,. \tag{5.4.12}$$

Substituting $\lambda_i = e^{\beta_0+\beta_1 x_i}$ yields

$$\int \mathrm{var}(b_1)\pi(\beta)d\beta = \int \frac{\left(n_1 e^{\beta_1 x_1} + n_2 e^{\beta_1 x_2}\right)}{n_1 n_2 e^{\beta_0} e^{\beta_1 x_1} e^{\beta_1 x_2}\left(x_1 - x_2\right)^2}\pi(\beta)d\beta\,. \tag{5.4.13}$$

Expression (5.4.13) is simply the F-criterion evaluated at the design points. So, the final form expression that the Bayesian F-optimal design must satisfy is

$$\int \frac{e^{\beta_1 x_j}\left(n_1 e^{\beta_1 x_1}\left(x_j - x_1\right)^2 + n_2 e^{\beta_1 x_2}\left(x_2 - x_1\right)^2\right)^2}{\beta_1^2 n_1^2 n_2^2 e^{2\beta_1 x_1} e^{2\beta_1 x_2}\left(x_1 - x_2\right)^4}\pi(\beta)d\beta - \int \frac{\left(n_1 e^{\beta_1 x_1} + n_2 e^{\beta_1 x_2}\right)}{n_1^2 n_2^2 e^{\beta_1 x_1} e^{\beta_1 x_2}\left(x_1 - x_2\right)^4}\pi(\beta)d\beta \le 0. \tag{5.4.14}$$

Let the left hand side of the equation be called g. Since the integral of the first term of g cannot be found in closed form, verification of the optimal design was pursued via Monte Carlo integration for both the uniform and the normal cases. Optimality of a uniform and a normal Bayesian F-optimal design are verified in Figures 5.4.1 and 5.4.2. In both of the figures, the function satisfies the criteria set forth in Section 5.2. The function g is less than or equal to zero in the design region and the function takes on the value of zero at the design points. Figure 5.4.1 verifies the optimality for the Bayesian design based on a uniform prior with a ratio of 4. This design places 77% of the experimental units at the $EC_{12.00}$ and 23% at the control.

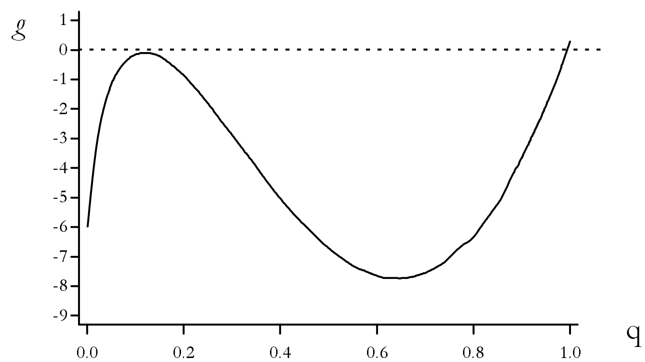

Figure 5.4.1 Graph of g for an F-Optimal Design Based on a Uniform Prior with a Ratio of 4.

The next graph, Figure 5.4, demonstrates the design based on a normal prior with s=0.3214 is optimal. In this design, 79% of the observations are allocated to the $EC_{8.94}$ and 21% are allocated at the control.

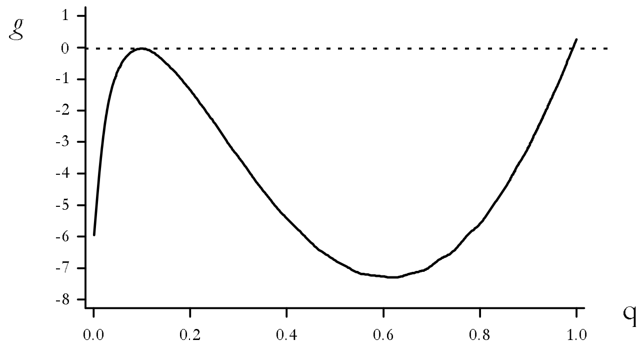

Figure 5.4.2 Graph of g for an F-Optimal Design Based on a Normal Prior where s_b=0.3214.

In conclusion, one should note that it is standard practice to verify numerical results with equivalence theory since numerical rounding errors and convergence problems can often be serious. However, there is also an aesthetic nature of equivalence theory which makes it pleasing and particularly satisfying when it is developed for a specific design/model.