BAYESIAN DESIGN FOR CLINICAL TRIALS WITH A CONSTRAINT ON THE TOTAL AVAILABLE DOSE

by

Mei-Mei Zen[1]
Cheng-Kung University
Republic of China
and Purdue University

and

Anirban DasGupta[2]
Purdue University

Technical Report # 97-25

Department of Statistics
Purdue University
West Lafayette, IN USA

December, 1997

[1] Research supported by a NSF R.O.C. grant
[2] Research supported by a National Security Agency grant

BAYESIAN DESIGN FOR CLINICAL TRIALS WITH A CONSTRAINT ON THE TOTAL AVAILABLE DOSE

by

Mei-Mei Zen[1]
Cheng-Kung University
Republic of China
and Purdue University

and

Anirban DasGupta[2]
Purdue University

Abstract

Bayesian optimal design of a dose response trial is considered when the response admits a linear regression relationship with the amount of dose applied. The important aspects the results are that we allow the variance of the response to depend on the mean, the number of patients to be used is itself part of the design, and in addition to a lower and an upper bound on the applicable dose, there is also a constraint on the total amount of the medication available. This last restriction is important in cases where an additional supply is not feasible due to logistic, financial, or other factors. We also avoid making any assumptions whatsoever about the form of the likelihood function by using the linear Bayes estimate of the response rate. The closed form results are illustrated on some examples by numerical computations.

[1] Research supported by a NSF R.O.C. grant
[2] Research supported by a National Security Agency grant

1 Introduction

Designing a dose response trial is one of the most important problems in practical health statistics. There is certainly a huge literature on various aspects of this problem. See [1] [4] [5] [6] [7] [8] [9] [10] [11] [12]. A very common problem in designing of clinical trials is to find optimal designs for estimating the probability of a dose response; often, this problem is of interest for binary type responses. Another common problem, with continuous responses, is to design for estimating the probability that the response exceeds a specified relevant threshold value.

In this article, we consider Bayesian optimal designs for a rather novel situation. There are some review articles for Bayesian optimal designs. See [2] [3]. We consider optimal designs when there is a regression type model between a response and the applied dose, and there are simultaneously three types of constraints: a lower bound on the applicable dose, an upper bound on the applicable dose, as well as a constraint on the total available dose that can be used. Such a situation might arise when there is a minimum dose that has to be given for any therapeutic consequences, a maximum dose beyond which toxic reactions can occur, and when the total supply of the particular medication is not easily augmentable due to logistic or budgetary reasons. In any case, this extra constraint on the total supply makes room for a new type of optimal design problems. Another novel aspect of this article is that we include the number of patients to be used in the clinical trial as a design variable itself. Consequently, we are able to address questions of this type : should a large number of patients be given a small dose of the substance, or a small number be given a high dose of the substance. Of course, in specific contexts, one may well take extraneous factors into serious consideration in designing the clinical trial and not be driven entirely by the formalism of a mathematical theorem.

The model we consider is the following : n patients are given dosage $x_1, x_2, \cdots, x_n$ to which they produce a measurable response $y_1, y_2, \cdots, y_n$. We assume $y_i = \theta f(x_i) + \varepsilon_i$, where the random errors are assumed to be independent. The goal is to estimate the regression coefficient. Commonly, one assumes in regression type models that the errors are homoscedastic. We, however, allow more flexibility. We let the variance of the errors to possibly depend on the regression coefficients. More specifically, we entertain a variance function of the form $var(\varepsilon_i) = \lambda[\theta f(x_i)]^p$; here, λ, θ and p are regarded as unknown parameters. For many applications, this should probably allow sufficient flexibility in modelling heteroscedastic variances.

At this stage, one can proceed along one of two routes. One is to assume a specific likelihood function, for example a normally distributed response, and combine it with specified priors for the parameters to compute a Bayes estimate. This avenue leads into difficulty in the kind of model we are looking at. As soon as we want to allow heteroscedastic variances of the type suggested above, the ensuing Bayesian calculation becomes impossible analytically. We take an alternative route. This alternative route has a lot to offer positively, and is open to some criticism as well.

We consider estimation of θ by the best linear Bayes estimate with respect to a specified prior for all the parameters θ, λ, and p. Let us pause momentarily to discuss the pros and cons. Here is the drawback : by settling for a linear Bayes estimate, we are settling for something less than the optimal. A linear estimate is not in general the true Bayes solution. On the other hand, here is what we gain : a) we can completely avoid making any distributional assumptions about the errors; so we gain some nonparametric flavor for the estimates, as well as the final designs; b) in certain cases the best linear estimate is already the true Bayes estimate; c) a linear estimate makes common sense for the model we have; d) by using a linear estimate, we are able to derive the optimal design and the optimal sample size in completely closed form and save on major numerical optimization; e) since the designs are derived analytically, we are able to show unifying structure in the optimal designs. Purely numerical optimization would generally preclude this.

In Section 2, we present the essential notation, and the basic calculations. In Section 3, we consider the design problem without considering a cost of sampling in the loss function. First we consider the case $f(x) = x$. The mathematics of this case is useful for more general choices of f, which we consider next in the same section. f is not completely arbitrary; we have assumed that f is convex or concave in the applicable interval for x. In Section 4, we consider the design problem by including a cost of sampling. For simplicity, we have used a linear cost. The main result in this section can be somewhat generalized to nonlinear cost functions. In Section 5, the theoretical optimal design is assessed against a reference design in order to explore if the optimal design was worth deriving. This is done by presenting two examples with the corresponding numerical illustrations. In Section 6, we give some brief concluding remarks. An interesting feature of our results is that time and again it is seen that there is a pretty common structure in the form of the Bayes optimal design and it is that contrary to intuition it is often the case that it is better to use a small number of patients at a high therapeutic dose in order to better estimate the response rate θ.

In summary,

i. we consider a popular model, namely a regression type model;
ii. we allow the variance of the observations to depend on the means;
iii. we allow a restriction on how much total medication is available;
iv. we allow the number of patients to be used to be part of the design;
v. by using a linear Bayes estimate, we can completely avoid making any assumptions about the form of the likelihood function;
vi. we consider the design problem both with and without cost of sampling;
vii. we show that there are some recurring features in the optimal designs.

2 Essential Notation

Consider the linear regression model

$$y_i = \theta f(x_i) + \varepsilon_i, \quad i = 1, 2, \cdots, n, \tag{1}$$

where $\varepsilon_i's$ are independent random variables with mean 0 and variance $var(\varepsilon_i) = \lambda[\theta f(x_i)]^p$, $0 < a \leq x_i \leq b < \infty$, $\sum_{i=1}^n x_i = T$ with a, b, T fixed, f is a positive differentiable function on $[a, b]$, and $\theta > 0$, $\lambda > 0$, $p \geq 0$ are unknown parameters. Let $f(x) = (f(x_1), f(x_2), \cdots, f(x_n))'$ and $y = (y_1, y_2, \cdots, y_n)'$ be two $n \times 1$ column vectors. Then the expectation of a linear estimator $\hat{\theta} = c'y$ is

$$E(\hat{\theta}) = E(c'y) = \theta c' f(x),$$

where $c = (c_1, c_2, \cdots, c_n)'$. Under squared-error loss, the risk function of $\hat{\theta}$ is

$$\begin{aligned} R(\hat{\theta},\ \theta) &= E(\hat{\theta} - \theta)^2 \\ &= E(c'y - \theta c' f(x) + \theta c' f(x) - \theta)^2 \\ &= \lambda \sum_{i=1}^n c_i{}^2 [\theta f(x_i)]^p + \theta^2 [c' f(x) - 1]^2. \end{aligned} \tag{2}$$

Let π be a prior on (θ, λ, p) and denote

$$\begin{aligned} \tau^2 &= E^\pi(\theta^2) > 0, \\ m(p) &= E^\pi(\theta^p | p) > 0, \\ \rho &= E^\pi(\lambda) > 0 \\ \text{and} \quad \phi(x) &= E^\pi([f(x)]^p m(p)) > 0. \end{aligned} \tag{3}$$

Then from (2), the Bayes risk of $\hat{\theta}$ equals

$$r(\pi, \hat{\theta}) = \rho \sum_{i=1}^n c_i{}^2 \phi(x_i) + \tau^2 [c' f(x) - 1]^2. \tag{4}$$

By differentiation, it can be shown that the minimum of the Bayes risk $r(\pi, \hat{\theta})$ is obtained when $\hat{\theta}_\pi = c_\pi' y$, where the i-th component of c_π is

$$c_{\pi i} = \frac{\frac{\tau^2}{\rho} \frac{f(x_i)}{\phi(x_i)}}{1 + \frac{\tau^2}{\rho} \sum_{i=1}^n \frac{f^2(x_i)}{\phi(x_i)}}. \tag{5}$$

Since

$$[c_\pi' f(x) - 1]^2 = [1 + \frac{\tau^2}{\rho} \sum_{i=1}^n \frac{f^2(x_i)}{\phi(x_i)}]^{-2},$$

(5) simplifies to

$$r(\pi, \hat{\theta}_\pi) = \frac{\tau^2}{1 + \frac{\tau^2}{\rho} \sum_{i=1}^{n} \frac{f^2(x_i)}{\phi(x_i)}}.$$

We will call this the Bayes decision risk:

$$BDR(n) = r(\pi, \hat{\theta}_\pi) = \frac{\tau^2}{1 + \frac{\tau^2}{\rho} \sum_{i=1}^{n} \frac{f^2(x_i)}{\phi(x_i)}}. \tag{6}$$

Let $C(n)$ be the cost of sampling n units. Then the total Bayes risk of $\hat{\theta}_\pi$ is defined as

$$BR(n) = BDR(n) + C(n). \tag{7}$$

A Bayesian optimal design is defined to be a probability distribution ξ over the design space $\mathcal{X} = [a, b]$ which minimizes the Bayes risk in (7). It is obvious that the Bayes risk can be minimized in two stages. First of all, find an optimal design for each fixed n. This corresponds to minimizing only $BDR(n)$. Then find the minimum of $BR(n)$ over n.

For each fixed n, minimizing $BDR(n)$ is equivalent to maximizing the quantity

$$Q = \sum_{i=1}^{n} \frac{f^2(x_i)}{\phi(x_i)}. \tag{8}$$

Note that Q/n can be treated as the mean of $h(X) = f^2(X)/\phi(X)$, where X is a random variable with probability distribution ξ over the design space $\mathcal{X}$ **satisfying the constraint**

$$E_\xi(X) = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{T}{n}.$$

Therefore, for each fixed n, we have the following problem:

$$\text{Maximize} \quad E_\xi \frac{f^2(X)}{\phi(X)} \quad \text{subject to} \quad E_\xi(X) = \frac{T}{n}, \ \ a \le X \le b. \tag{9}$$

Note that this *approximate* optimal design allows that $nP_\xi(X = x)$ is not necessarily an integer and we may take $[nP_\xi(X = x)]$ or $[nP_\xi(X = x)] + 1$ observations at each design point x. We will assume both T/a and T/b are integers.

Let us look at (9) in general. In order to find ξ which maximizes $E_\xi h(X)$ for some function h under the restriction that $E_\xi(X) = T/n$ and $a \le X \le b$, the concavity/convexity of h plays an important role. Below is a useful result.

Proposition 1.
(a) Suppose $h(x)$ is concave on $[a, b]$. Then the probability measure ξ which maximizes $E_\xi h(X)$, under the restriction that $E_\xi(X) = T/n$ and $a \le X \le b$, is obtained by

taking $\xi = \delta_{\{T/n\}}$, a point mass at T/n.
(b) Suppose $h(x)$ is convex on $[a, b]$. Then the probability measure ξ which maximizes $E_\xi h(X)$, under the restriction that $E_\xi(X) = T/n$ and $a \leq X \leq b$, is obtained by putting weights q and $1 - q$ at a and b respectively, where q satisfies

$$E_\xi(X) = qa + (1-q)b = \frac{T}{n}, \quad i.e., \quad q = \frac{b - T/n}{b - a}. \tag{10}$$

Thus the concavity/convexity properties of the function h are crucial for the design problem at hand. So in Section 2, we first investigate the sign of $h''(x)$ for the special case $f(x) = x$.

3 The optimal design without a cost function

3.1 Preliminary Calculations

In this section, we focus on the special case $f(x) = x$ first. This makes the mathematics for the case of a more general f easier. If $f(x) = x$, then

$$h(x) \stackrel{\text{def}}{=} \frac{f^2(x)}{\phi(x)} = \frac{x^2}{E^\pi[x^p m(p)]} > 0. \tag{11}$$

Let

$$g(x) = \frac{1}{h(x)} = E^\pi[x^{p-2} m(p)] = \int x^{p-2} m(p) d\mu(p), \tag{12}$$

where μ is the marginal prior for p. Differentiating both sides of the identity $h(x) \cdot g(x) = 1$ gives:

$$h'(x)g(x) + h(x)g'(x) = 0$$

and on differentiating again,

$$h''(x)g(x) + 2h'(x)g'(x) + h(x)g''(x) = 0.$$

Hence,

$$h'(x) = \frac{-h(x)g'(x)}{g(x)}$$

and

$$\begin{aligned} h''(x)g(x) &= -2h'(x)g'(x) - h(x)g''(x) \\ &= \frac{2h(x)[g'(x)]^2}{g(x)} - h(x)g''(x) \\ &= \frac{h(x)}{g(x)}\{2[g'(x)]^2 - g(x)g''(x)\}. \end{aligned} \tag{13}$$

Since h is positive, from (13) one concludes that $h''(x)$ and $2[g'(x)]^2 - g(x)g''(x)$ have the same sign. We will now derive a useful integral representation for $2[g'(x)]^2 - g(x)g''(x)$.

Using the notation $\alpha = p-2$ and $d\nu_x(\alpha) = x^\alpha m(\alpha+2)d\mu(\alpha+2)$ ($d\mu(\alpha+2)$ means the obvious thing), direct computation gives

$$\begin{aligned} & x^2\{2[g'(x)]^2 - g(x)g''(x)\} \qquad (14) \\ = \; & 2[\int \alpha d\nu_x(\alpha)]^2 - [\int 1 d\nu_x(\alpha)][\int \alpha(\alpha-1)d\nu_x(\alpha)]. \end{aligned}$$

In order to determine the sign of (14), we use double integrals to represent product intergals and split the second part into two. Then (14) becomes

$$\begin{aligned} & 2\int \alpha\beta d\nu_x(\alpha)d\nu_x(\beta) - \int \alpha(\alpha-1)d\nu_x(\alpha)d\nu_x(\beta) \qquad (15) \\ = \; & 2\int \alpha\beta d\nu_x(\alpha)d\nu_x(\beta) - 1/2[\int \alpha(\alpha-1)d\nu_x(\alpha)d\nu_x(\beta) + \int \beta(\beta-1)d\nu_x(\alpha)d\nu_x(\beta)] \\ = \; & 1/2\int [4\alpha\beta - \alpha(\alpha-1) - \beta(\beta-1)]d\nu_x(\alpha)d\nu_x(\beta) \\ = \; & 1/2\int M(\alpha,\beta)d\nu_x(\alpha)d\nu_x(\beta), \end{aligned}$$

where

$$M(\alpha,\beta) = 4\alpha\beta - \alpha(\alpha-1) - \beta(\beta-1). \qquad (16)$$

From (15), we arrive at the useful conclusion:

If $M(\alpha,\beta) \geq 0$ in its domain, then h is convex;

If $M(\alpha,\beta) \leq 0$ in its domain, then h is concave.

3.2 $f(x) = x, \quad 0 \leq p \leq 1$

Theorem 1.

Suppose $0 \leq p \leq 1$. Then $h(x)$ is convex, and for any fixed n, the optimal design ξ, which maximizes

$$E_\xi h(X) \quad \textit{subject to} \quad E_\xi(X) = T/n,$$

puts weights q and $1-q$ at a and b respectively, where $q = \frac{b-T/n}{b-a}$. Furthermore,

$$max_\xi E_\xi h(X) = qh(a) + (1-q)h(b). \qquad (17)$$

Proof. The function $M(\alpha,\beta)$ may be rewritten as

$$M(\alpha,\beta) = -(\alpha+1)(\alpha+2) - (\beta+1)(\beta+2) + 4(\alpha+1)(\beta+1). \qquad (18)$$

If p varies in $[0,1]$, then α, β are in $[-2,-1]$, and so from (18), $M(\alpha,\beta) \geq 0$. Consequently, h is convex and the optimal design follows from Proposition 1.

Theorem 1 solves the first stage of the problem, *i.e.* it provides an optimal design for fixed n. Now we proceed to the second stage: namely, how many patients should be used, *i.e.*, maximize $BDR(n)$ over n.

Recall that in this entire section, the cost function $C(n)$ is being ignored. The following result holds.

Theorem 2.
Suppose $0 \leq p \leq 1$. Then the optimal sample size is the smallest possible $n = T/b$ and the unique point in the support of the optimal design is $x \equiv b$. Furthermore, the minimum Bayes decision risk is

$$BDR = \frac{\tau^2}{1 + \frac{\tau^2}{\rho} T \frac{h(b)}{b}}.$$

Proof. From (17), we want to maximize

$$\begin{aligned} & n[qh(a) + (1-q)h(b)] \qquad (19) \\ = \; & (nb - T)\frac{h(a)}{b-a} + (T - na)\frac{h(b)}{b-a} \\ = \; & nab\frac{[h(a)/a - h(b)/b]}{b-a} + \frac{T[h(b) - h(a)]}{b-a}, \end{aligned}$$

which is a linear function in n. Applying l'Hôspital's rule twice, we have

$$\lim_{x \to 0} h(x) = 0 \qquad \text{for } \; 0 \leq p \leq 2.$$

The convexity of h for $0 \leq p \leq 1$ yields the fact that $h(u)/u$ is increasing in u. That means the linear function of n has a negative slope. Hence the maximum of (19) is obtained when n is the smallest, *i.e.*, $n = T/b$. The rest of the theorem follows on using Theorem 1.

3.3 $f(x) = x, \quad 1 \leq p \leq 2$

Theorem 3.
Suppose $1 \leq p \leq 2$. Then $h(x)$ is concave, and for any fixed n, the optimal design ξ which maximizes $E_\xi h(X)$ subject to $E_\xi(X) = T/n$ is a point mass at T/n. Furthermore,

$$max_\xi E_\xi h(X) = h(T/n). \qquad (20)$$

Proof. The function $M(\alpha, \beta)$ can be written as

$$M(\alpha, \beta) = \alpha(\alpha + 1) + \beta(\beta + 1) - 2(\alpha - \beta)^2. \tag{21}$$

From (21), It is clear that $M(\alpha, \beta) \leq 0$ if $1 \leq p \leq 2$. Consequently, h is concave and the optimal design follows from Proposition 1.

Analogous to Theorem 2, we now proceed to the second stage of the problem, *i.e.*, work out the optimal n. The following result holds.

Theorem 4.
Suppose $1 \leq p \leq 2$. Then the optimal sample size is the largest possible $n = T/a$ and the unique point in the support of the optimal design is $x \equiv a$. Furthermore, the minimum Bayes decision risk is

$$BDR = \frac{\tau^2}{1 + \frac{\tau^2}{\rho} T \frac{h(a)}{a}}.$$

Proof. Similar to the proof of Theorem 2.

In summary,
i. Suppose $0 \leq p \leq 1$. Then $h(x)$ is convex,
the optimal sample size is the smallest possible and $n = T/b$
the unique point in the support of the optimal design is $x \equiv b$.
ii. Suppose $1 \leq p \leq 2$. Then $h(x)$ is concave,
the optimal sample size is the largest possible and $n = T/a$
the unique point in the support of the optimal design is $x \equiv a$.

3.4 Generalization to more general $f(x)$

The crucial feature driving the nature of the optimal design in section 2.1 and 2.2 is the concavity/convexity of the function h. So for a generel f, the driving feature is the concavity/convexity of the function $f^2(x)/\phi(x)$; see (9). We have the following important lemma.

Lemma 1.
(a) Let f be convex. Then $f^2(x)/\phi(x)$ is convex if $0 \leq p \leq 1$.
(b) Let f be concave. Then $f^2(x)/\phi(x)$ is concave if $1 \leq p \leq 2$.

Proof. Observe that $f^2(x)/\phi(x)$ is just the composition function $(h \circ f)(x)$ where h

is as defined in (11). By chain rule,

$$\frac{d^2}{dx^2}(h \circ f)(x) = h''(f(x)) \cdot [f'(x)]^2 + h'(f(x)) \cdot f''(x). \tag{22}$$

For part (a), use the facts: h is convex and increasing and f is convex, and use (22). Part (b) follows analogously.

Application of Lemma 1 leads to the following two theorems.

Theorem 5.
(a) Suppose $0 \leq p \leq 1$ and $f(x)$ is convex on $[a, b]$. Then for each fixed n, the Bayesian optimal design that maximizes $E_\xi f^2(X)/\phi(X)$ subject to $E_\xi(X) = T/n$ puts weights q and $1-q$ at a and b respectively, where $q = \frac{b-T/n}{b-a}$. Furthermore,

$$max_\xi E_\xi \frac{f^2(X)}{\phi(X)} = \frac{\tau^2}{1 + \frac{\tau^2}{\rho} n\{q\frac{f^2(a)}{\phi(a)} + (1-q)\frac{f^2(b)}{\phi(b)}\}}. \tag{23}$$

(b) Suppose $1 \leq p \leq 2$ and $f(x)$ is concave on $[a, b]$. Then for each fixed n, the Bayes optimal design that maximizes $E_\xi f^2(X)/\phi(X)$ subject to $E_\xi(X) = T/n$ is a point mass at T/n. Furthermore,

$$max_\xi E_\xi \frac{f^2(X)}{\phi(X)} = \frac{\tau^2}{1 + \frac{\tau^2}{\rho} n \frac{f^2(T/n)}{\phi(T/n)}}. \tag{24}$$

Proof. Similar to the proof of Theorems 1 and 3. We omit the details.

The following theorem is an analog of Theorems 2 and 4.

Theorem 6.
(a) Suppose $0 \leq p \leq 1$ and $f(x)$ is convex on $[a, b]$, with $f(0) = 0$. Then the optimal sample size is the smallest possible $n = T/b$ and the optimal design is a point mass at $x = b$. Furthermore, the minimum Bayes decision risk is

$$BDR = \frac{\tau^2}{1 + \frac{\tau^2}{\rho} T \frac{f^2(b)}{b\phi(b)}}. \tag{25}$$

(b) Suppose $1 \leq p \leq 2$ and $f(x)$ is concave on $[a, b]$., with $f(0) = 0$. Then the optimal sample size is the largest possible $n = T/a$ and the optimal design is a point mass at $x = a$. Furthermore, the minimum Bayes decision risk is

$$BDR = \frac{\tau^2}{1 + \frac{\tau^2}{\rho} T \frac{f^2(a)}{a\phi(a)}}. \tag{26}$$

Proof. This is because the convexity/concavity of $(h \circ f)(x) = f^2(x)/\phi(x)$ with $(h \circ f)(0) = h(f(0)) = 0$.

4 Optimal design with a cost function

In this section, we consider the Bayesian optimal designs with a cost function $C(n)$. We assume that $C(n)$ is increasing in n. The optimal designs for different ranges of p are derived in the following subsections.

4.1 $f(x) = x, \quad 0 \leq p \leq 1$

From the proof of Theorem 2, we know that the Bayesian decision risk is increasing in n. By assumption $C(n)$ is increasing. So the total Bayes risk is increasing in n, too. So this case is trival. We have the following result immediately.

Theorem 7.
Suppose $0 \leq p \leq 1$. Then the optimal sample size is the smallest possible $n = T/b$ and the optimal design is a point mass at $x = b$. Furthermore, the minimum Bayes risk is

$$BR = \frac{\tau^2}{1 + \frac{\tau^2}{\rho} T \frac{h(b)}{b}} + C(T/b). \tag{27}$$

4.2 $f(x) = x, \quad 1 \leq p \leq 2$

This case is considerably harder because the decision risk is decreasing in n and the cost function is increasing in n, and so a pretty delicate analysis is required. It turns out that the Bayes decision risk is an increasing and convex function in $u = T/n$, if we regard u as a continuous argument as a technical device. This is useful for our subsequent analysis and this is the content of the next two lemmas. We will write $B(u)$ for

$$BDR(n) = \frac{\tau^2}{1 + \frac{\tau^2}{\rho} n h(T/n)} = \frac{\tau^2}{1 + \frac{\tau^2}{\rho} T \frac{h(u)}{u}}. \tag{28}$$

If we further define $1/\tau^2$ as k_1, T/ρ as k_2 and $h(u)/u$ as $F(u)$, then we have

$$B(u) = \frac{1}{k_1 + k_2 F(u)}. \tag{29}$$

The following lemmas are crucial for the optimal design result of Theorem 8.

Lemma 2.
$B(u)$ is an increasing and concave function in u.

Proof. Straightforward calculus.

Lemma 3.
The Bayesian decision risk $BDR(n)$ is a decreasing and convex function in n, where n is treated as a continuous argument.

Proof.

Step 1. n and u are related as $n = n(u) = T/u$ and $BDR(n) = B(u)$, with $B(u)$ as in (29). By chain rule, we have the following relationship:

$$\frac{d}{dn}BDR(n) = \frac{d}{du}B(u)(-\frac{u^2}{T})$$

and

$$\frac{d^2}{dn^2}B(n) = \frac{u^3}{T^2}[\frac{d^2}{du^2}B(u)\cdot u + 2\frac{d}{du}B(u)]. \tag{30}$$

It will therefore suffice to show that $uB''(u) + 2B'(u) > 0$.

Step 2. Denote the function $1/F(u)$ as $H(u)$; here F is as in (29). Then, on algebra,

$$\begin{aligned} & uB'' + 2B' \qquad (31)\\ = \;& k_2B^2H^{-3}\{(uH'' + 2H')H - 2k_1uB(H')^2\}\\ \geq \;& k_2B^2H^{-3}\{(uH'' + 2H')H - 2u(H')^2\} \qquad (\text{since} k_1B = 1 - k_2\frac{B}{H} \leq 1)\end{aligned}$$

Step 3. We will now show that $(uH'' + 2H')H \geq 2u(H')^2$. By applying the definition of $H(u)$, *i.e.*, $H(u) = E^\pi[u^{p-1}m(p)]$,

$$\begin{aligned} & [uH''(u) + 2H'(u)]H(u)\\ = \;& \{uE^\pi[(p-1)(p-2)u^{p-3}m(p)] + 2E^\pi[(p-1)u^{p-2}m(p)]\}E^\pi[u^{p-1}m(p)]\\ = \;& E^\pi[p(p-1)u^{p-2}m(p)]E^\pi[u^{p-1}m(p)]\\ \geq \;& \{E^\pi[\sqrt{p(p-1)}u^{p-3/2}m(p)]\}^2 \qquad (\text{Cauchy-Schwartz Inequality})\\ \geq \;& \{E^\pi[\sqrt{2}(p-1)u^{p-3/2}m(p)]\}^2 \qquad (\text{since } 1 \leq p \leq 2)\\ = \;& 2u\{E^\pi[(p-1)u^{p-2}m(p)]\}^2\\ = \;& 2u(H'(u))^2.\end{aligned}$$

This prove Lemma 3.

We are now ready to state the following theorem.

Theorem 8.
Suppose $1 \leq p \leq 2$ *and* $C(n) = c_0 + c_1 n$. *If the incremental cost* c_1 *satisfies the inequality*

$$c_1 \geq (BDR(\frac{T}{b}) - BDR(\frac{T}{b} + 1)), \tag{32}$$

then the optimal sample size is $n = T/b$ *and the optimal design is a point mass at* $x = b$.

Proof. The proof consists of showing that the total Bayes risk $BR(n) = BDR(n) + C(n)$ is nondecreasing in n for $n \geq \frac{T}{b}$. But,

$$\begin{aligned}
& BR(n+1) - BR(n) \\
= \; & BDR(n+1) - BDR(n) + c_1 \\
\geq \; & BDR(\frac{T}{b} + 1) - BDR(\frac{T}{b}) + c_1 \qquad \text{(by the convexity result in Lemma 3)} \\
\geq \; & 0 \qquad\qquad\qquad\qquad\qquad\qquad\quad \text{(by hypothesis)}
\end{aligned}$$

That the optimal design is a point mass st $x = b$ follows from this.

In summary,
i. Suppose $0 \leq p \leq 1$.
Then the optimal sample size is $n = T/b$ and
the optimal design is a point mass at $x = b$.
ii. Suppose $1 \leq p \leq 2$ and $C(n) = c_0 + c_1 n$. If $c_1 \geq (BDR(\frac{T}{b}) - BDR(\frac{T}{b} + 1))$,
then the optimal sample size is $n = T/b$ and
the optimal design is a point mass at $x = b$.

5 Numerical Illustrations

In this section we give some numerical examples to show how much better the theorectical optimal design does relative to another reference design. The reference design is described each separate example. All the comparison are for fixed values of n and in terms of Bayes risk efficiency. In each example, we have taken $a = 0.5$, $b = 2.0$, $T = 10$ (so that n is necessarily in the range $5 \leq n \leq 20$, and $\rho = 3$. θ has been given an Exponential prior with mean 1. The prior on p is described separately in each example. p and θ have been assumed independent.

Example 1. ($0 \leq p \leq 1$) We take $f(x) = x^2$ to be the regression function, note that f is convex; as needed for applying Theorem 5. Here are three priors for which numerical results are reported:

p	0	0.25	0.5	0.75	1
π_1	.2	.2	.2	.2	.2
π_2	.5	0	0	0	.5
π_3	0	0	1	0	0

In this case, the theoretical optimal design is a 2-point design (Theorem 5). So a comparison is made with a 1-point design at $x = T/n$. Below are the efficiency results. Clearly, the theorectical optimal design does much better.

Table 1: BDR for different priors, $f(x) = x^2$

n	5	7	9	11	13	15	17	19
π_1	0.0768	0.0873	0.1011	0.1201	0.1479	0.1925	0.2754	0.4841
π_2	0.0895	0.1016	0.1176	0.1395	0.1714	0.2222	0.3157	0.5454
π_3	0.0643	0.0731	0.0848	0.1009	0.1246	0.1628	0.2346	0.4200

Table 2: Efficiency of 1-point design, $f(x) = x^2$

n	5	7	9	11	13	15	17	19
π_1	1.0000	0.6533	0.4909	0.4105	0.3783	0.3865	0.4516	0.6695
π_2	1.0000	0.7008	0.5402	0.4524	0.4125	0.4150	0.4762	0.6878
π_3	1.0000	0.5983	0.4364	0.3643	0.3397	0.3535	0.4227	0.6476

Example 2. ($1 \leq p \leq 2$) We take $f(x) = log(1 + x)$ to be the regression function; note that f is concave, as needed for applying Theorem 6. Here are three priors for which numerical results are reported:

p :	1	1.25	1.5	1.75	2
π_1	.2	.2	.2	.2	.2
π_2	.5	0	0	0	.5
π_3	0	0	1	0	0

In this case, the theorectical optimal design is a 1-point design (Theorem 6). So a comparison is made with a 2-point design. Below are the efficiency results. The theoretical optimal design does a bit better.

Table 3: BDR for different priors, $f(x) = log(1+x)$

n	5	7	9	11	13	15	17	19
π_1	0.5798	0.4832	0.4217	0.3786	0.3467	0.3220	0.3021	0.2859
π_2	0.6077	0.5019	0.4352	0.3894	0.3559	0.3303	0.3102	0.2939
π_3	0.5512	0.4643	0.4080	0.3679	0.3375	0.3136	0.2941	0.2779

Table 4: Efficiency of 2-point design, $f(x) = log(1+x)$

n	5	7	9	11	13	15	17	19
π_1	1.0000	0.9534	0.9366	0.9351	0.9423	0.9550	0.9713	0.9899
π_2	1.0000	0.9489	0.9297	0.9274	0.9349	0.9488	0.9671	0.9884
π_3	1.0000	0.9588	0.9447	0.9440	0.9507	0.9619	0.9759	0.9916

6 Conluding Remarks

We considered a practical design problem and showed some common structure in the form of the optimal design and the optimal number of patients that should be used. In particular, it is interesting that often it is the best to use a small number of patients at the highest therapeutic dose.

References

[1] Armitage, P. (1992). Some topics of current interest of clinical trials. *Canadian Journal of Statistics.* **10**, No. 3, 273- 304.

[2] Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science.* **10**, No. 3, 273- 304.

[3] DasGupta, A. (1996). Review of optimal Bayes designs. S. Ghosh and C. R. Rao, eds., *Handbook of Statistics*, **13**, 1099-1147.

[4] Estelle Russek-Cohen and Richard M. Simon (1994). Selecting the best dose When a monotonic dose-response relation exists. *Statistics in Medicine.* **13**, 87-95.

[5] Faraone, S. and Simpson, J. (1992). Mathematical models of complex dose-response relationships: Implications for experimental design in psychopharmacologic research. *Statistics in Medicine.* **11**, 685-702.

[6] Kalish, L. A. (1990). For estimation of median lethal dose and quantal dose-response curves. *Biometrics.* **46**, 738-748.

[7] Minkin, S. (1993). Experimental design for clonogenic assays in chemotherapy. *Journal of the American Statistical Association.* **88**, 410-420.

[8] Müller, H. G. and Wang, J. L. (1990). For effective doses in the probit model for dose-response data. *Biometrical Journal.* **32**, 529-544.

[9] Patel, H. I. (1992). Sample size for a dose-response study. *Journal of Biopharmaceutical Statistics.* **2** , 1-8.

[10] Ruberg, S. (1995). Dose-response studies. I. Some design considerations. *Journal of Biopharmaceutical Statistics.* **5** , 1-14.

[11] Strijbosch, L., Does, R. and Albers, W. (1990). Design methods for some dose-response models. *Statistics in Medicine.* **9**, 1353-63.

[12] Wijesinha, M. C. and Piantadosi, S. (1995). Dose-response models with covariates. *Biometrics.* **51**, 977-987.