The Annals of Statistics
1983, Vol. 11, No. 3, 886–895

BAYESIAN BIOASSAY DESIGN

BY LYNN KUO

State University of New York at Stony Brook

A Bayesian treatment of the quantal bioassay design problem is given. It is assumed that the potency curve is a Dirichlet random distribution F with parameter $\alpha(t) = MF_0(t)$, and that $n_1, \ldots, n_L$ animals are treated at drug levels $t_1, \ldots, t_L$ respectively. The optimal design levels $t_1, \ldots, t_L$ that minimize the Bayes risk for weighted integrated quadratic loss functions are found in the following cases: (i) $L = 1$ and the weight function arbitrary; (ii) uniform prior guess, uniform weight and two animals treated; and (iii) uniform weight and L arbitrary, but $M \to 0$. These results disprove a conjecture of Antoniak.

1. Introduction. We can state our quantal bioassay problem as follows. The experimenter tests the potency of a certain drug by giving groups of animals injections of the drug at certain levels; namely he treats $n_1, \ldots, n_L$ animals at drug dosage levels $t_1, \ldots, t_L$ respectively, and records $k_1, \ldots, k_L$, the number of animals giving positive response at each level. It is assumed that each animal's response to a given dose of drug is either positive or negative (no response) and that each animal has a threshold which the dose given him must equal or exceed to produce a positive response. However, this threshold varies from one animal to the next, so we treat it as a random variable with unknown distribution F. In the literature this F is often called the tolerance distribution or potency curve. The drug dosage levels may be the actual dosage levels or the logarithms of these levels.

A Bayesian nonparametric approach to estimating F, in which the prior distribution of F is a process of the Dubins and Freedman type, was first proposed by Kraft and van Eeden (1964). Ramsey (1972), Antoniak (1974), Wesley (1976), and Bhattacharya (1981) have also studied inferences on F. For them F is chosen from Ferguson's Dirichlet process (1973) with parameter α, where α is a finite measure on $\mathbb{R}$. The objective of this paper is to find the drug levels $t_1, \ldots, t_L$ to be administered such that the Bayes risk $\mathscr{E} \int (F(t) - \hat{F}(t))^2 \, dW(t)$ is minimized, where W is a nonrandom measure, and $\hat{F}(t) = \mathscr{E}(F(t) | k_1, \ldots, k_L)$ is the Bayes estimator of F with respect to the Dirichlet process prior.

The Bayesian nonparametric approach has advantages over other approaches. Prior information, which is often available from previous assays using similar subjects and similar stimuli, can be incorporated into the analysis. While F is often assumed in the literature to have a specific parametric form, such as normal or logistic, such assumptions are frequently too strong. For a short review of other approaches to bioassay design problems, I would recommend Cochran (1973), Abdelbasit and Plackett (1981), and Tsutakawa (1982). A discussion of principles of good design for bioassays is also given by Finney (1978, Chapters 6 and 19).

To use the approach based on Dirichlet processes, the statistician need only specify a prior guess at F (possibly obtained from past assays) and a "prior sample size" signifying how confident he is of his initial guess. As pointed out by Ferguson (1973): "$\alpha((-\infty, t])/\alpha(\mathbb{R})$ denoted by $F_0(t)$ represents our prior guess at the shape of the unknown $F(t)$. $\alpha(\mathbb{R})$ denoted by M can be interpreted as a measure of the strength of belief in the prior guess, measured in units of sample size. If M is large compared to the sample size n, little weight is given to the observations. If M is small compared to n, little weight is given to the prior guess at F."

A few preliminary results are given in Section 2. In Section 3, we treat the design

Received January 1981; revised March 1983.

AMS 1980 *subject classifications.* Primary, 62K05, 62P10; secondary, 62C10.

Key words and phrases. Dirichlet process, mixtures of Dirichlet processes, quantal bioassay, potency curve, threshold of tolerance, Bayes risk, optimal design.

problem for the case $L = 1$, i.e. all the animals treated at the same level. In this case, we are able to derive an optimal solution for any $W(t)$, M, and continuous $F_0(t)$. For the cases $L > 1$, we make the assumptions $W(t) = t$ and $F_0(t) = t$, where $0 \le t \le 1$, for computational convenience. Actually, one of these assumptions (but not both) can be made without loss of generality by a reparametrization of t. In Section 4, we find the optimal solution for the special case: only two animals are tested, and $W(t) = t$, $F_0(t) = t$, $0 \le t \le 1$. The optimal solution to this case is given by treating one animal each at levels $\frac{1}{2} - c_0$ and $\frac{1}{2} + c_0$, where c_0 is the unique root of a rational function depending on the value of M. For example: when $M \to 0$, the two optimal levels approach $\frac{1}{3}$ and $\frac{2}{3}$; when $M = 1$, the two levels are given approximately by .365 and .635; when $M = 2$, the two levels are given approximately by .389 and .611; when $M \to \infty$, the two levels approach $\frac{1}{2}$. In Section 5, we treat the case $M \to 0$. The Bayes risk is reduced to a simple function from which the optimal design can be derived easily. This design has drug levels spaced to divide the line into intervals having equal prior probability.

A few words should be said about Ramsey's result (1972). In Example 4 of his paper, he takes the parameter $\alpha(t) = MF_0(t)$, such that $F_0(t)$ $(-\infty < t < \infty)$ is the standard normal distribution function $\Phi(t)$. He assumes that the actual potency curve is a shifted standard normal distribution $\Phi(t - \theta)$. He examines the following designs for a total of six animals:

(i) one animal at each of the six different doses;
(ii) two animals at each of three different doses;
(iii) three animals at each of the two different doses.

The spacing of the drug levels is determined by letting them divide the prior guess in equal percentages, namely, $F_0(t_i) - F_0(t_{i-1}) = C$ for all i. With each of these designs, he estimates the median of F and graphs bias, standard deviation, and the square root of the mean squared error of this estimator for each of the designs over a range of θ. These graphs reveal one conclusion; whether based on bias, standard deviation, or root mean squared error, the best design uses one observation per dose.

Antoniak shows that the posterior distribution of F for the bioassay is a mixture of Dirichlet processes when the prior on F is a Dirichlet process, and obtains the Bayes estimator of F with two doses for the integrated squared error loss. Assuming $F_0(t) = t$, $W(t) = t$, $0 \le t \le 1$, he conjectures for the design problem: (i) If L is fixed beforehand, the Bayes design is to set $t_i = i/(L + 1)$ and makes the n_i as nearly equal as possible, i.e. $|n_i - n_j| \le 1$ for all i, j. (ii) If L is not fixed, the Bayes design is to let $L = n$ (where n is the total number of animals experimented on) and take one observation each at the levels $t_i = i/(n + 1)$. Our results exhibit a counterexample to both parts of this conjecture. Even with the uniform prior guess, the equal spacing of the dosage levels is optimal only for $M \to 0$. As M increases, we can reduce the Bayes risk by taking animals at drug levels closer to the center. Antoniak's conjecture, part (ii), is correct for the case $M \to 0$. Although we have only treated very special cases, nevertheless I feel that when $0 < M < \infty$, W is an absolutely continuous measure, and the number of drug levels is not fixed beforehand, the best design is given by one observation per dose.

2. Preliminary results. Let $\mathbb{R}$ be the real line, $\mathcal{B}$ be a σ-field of Borel sets, and α be a nonnull finite measure on $(\mathbb{R}, \mathcal{B})$. Let us make the following three assumptions:

(2.1) (1) F is a random distribution function chosen from the Dirichlet process on $(\mathbb{R}, \mathcal{B})$ with parameter α;

(2) $n_1, \cdots, n_L$ animals are treated at drug levels $t_1, \cdots, t_L$ respectively, and $k_1, \cdots, k_L$ (denoted by $\mathbf{k}$) are the observed number of animals that respond positively at each level, where the t_i are ordered in nondecreasing order;

(3) The loss function $L(F, \hat{F})$ is given by $\int (F(t) - F(\hat{t}))^2 \, dW(t)$, where W (the weight function) is a nonrandom measure. The measure W and the function $W(t) = W((-\infty, t])$ will be used interchangeably.

Then we can obtain the Bayes rule of F by pointwise minimization of the function $\mathcal{E}[(F(t) - \hat{F}(t))^2 \,|\, \mathbf{k}]$ for each t.

Let us sketch the approach to estimating F for general α and L. Define y_i and β_i by the

following equations: $y_i = F(t_i) - F(t_{i-1})$, $\beta_i = MF_0(t_i) - MF_0(t_{i-1})$, $i = 1, \cdots, L+1$, where $t_0 = -\infty$, $t_{L+1} = \infty$. The likelihood function of the $\mathbf{y}$ is given by

$$\prod_{i=1}^{L} \binom{n_i}{k_i} (\textstyle\sum_{j=1}^{i} y_j)^{k_i} (1 - \sum_{j=1}^{i} y_j)^{n_i - k_i},$$

since, given F, the k_i's are independently and binomially distributed with parameters n_i and $F(t_i)$. Given the prior Dirichlet density of $\mathbf{y} = (y_1, \cdots, y_L; y_{L+1})$ with parameters $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_L; \beta_{L+1})$ and the above likelihood function, we can derive the posterior density of $\mathbf{y}$, which is a mixture of Dirichlet densities that becomes quite complicated as L increases. To minimize $\mathscr{E}[(F(t) - \hat{F}(t))^2 | \mathbf{k}]$, we want to estimate $F(t)$ by $\hat{F}(t) = \mathscr{E}(F(t) | \mathbf{k})$. Therefore, for $i = 1, \cdots, L$, we have

$$\hat{F}(t_i) = \textstyle\sum_{j=1}^{i} \mathscr{E}(y_j | \mathbf{k}).$$

For other values of t, say $t_i < t < t_{i+1}$, $i = 0, \cdots, L$, we use the following interpolation formula:

$$\hat{F}(t) = \hat{F}(t_i) + \frac{F_0(t) - F_0(t_i)}{F_0(t_{i+1}) - F_0(t_i)} [\hat{F}(t_{i+1}) - \hat{F}(t_i)]. \tag{2.2}$$

A proof of (2.2) is given in Wesley (1976, pages 53–57).

For $L = 1$, or 2, it is easy to obtain the likelihood function and posterior distribution. A few results for $L = 1$ are given here for future reference. For $L = 2$, see Antoniak's paper (1974).

Assume $F_0(t) = t$, $0 \le t \le 1$, $L = 1$, and n_1 animals are treated at $t_1 \in (0, 1)$, and k_1 of them react positively. Then, the posterior density of y_1 given k_1 is given by:

$$g(y_1 | k_1) = \frac{\Gamma(n_1 + M)}{\Gamma(k_1 + \beta_1)\Gamma(n_1 - k_1 + \beta_2)} y_1^{k_1 + \beta_1 - 1} (1 - y_1)^{n_1 - k_1 + \beta_2 - 1}$$

where $\beta_1 = Mt_1$, $\beta_2 = M - Mt_1$. Hence

$$\hat{F}(t_1) = \frac{k_1 + \beta_1}{n_1 + M} = \frac{k_1 + Mt_1}{n_1 + M}, \quad \text{and} \tag{2.3}$$

$$\hat{F}(t) = \begin{cases} \dfrac{t}{t_1} \hat{F}(t_1), & \text{if } 0 \le t \le t_1; \\ \dfrac{1 - \hat{F}(t_1)}{1 - t_1} (t - t_1) + \hat{F}(t_1), & \text{if } t_1 \le t \le 1. \end{cases} \tag{2.4}$$

For the general problem, there are L drug levels to be administered. $\hat{F}(t)$ is given as in (2.2). The design problem is to find $\mathbf{t} = (t_1, \cdots, t_L)$ such that the Bayes risk $r(\mathbf{t}) = \int \mathscr{E}(F(t) - \hat{F}(t))^2 \, dW(t)$ is minimized. Note that $\hat{F}(t)$ is a function of the observations $\mathbf{k}$ and the drug levels $\mathbf{t}$. The expectation is taken with respect to the joint distribution of $\mathbf{k}$ and $F(t)$. Since both $F(t)$ and $\hat{F}(t)$ are random, the first step is to simplify $\mathscr{E}(F(t) - \hat{F}(t))^2$ to an expectation of a more manageable function. The following theorem gives the desired simplifications.

THEOREM 2.1. *With the assumptions in* (2.1), $\hat{F}(t)$ *as in* (2.2), *and* $\int \mathscr{E}F(t)^2 \, dW(t) < \infty$, *then*

$$\begin{aligned} \mathscr{E} \int [F(t) - \hat{F}(t)]^2 \, dW(t) &= \int \operatorname{Var} F(t) \, dW(t) - \int \operatorname{Var} \hat{F}(t) \, dW(t) \\ &= \int \mathscr{E}F(t)^2 \, dW(t) - \int \mathscr{E}\hat{F}(t)^2 \, dW(t). \end{aligned}$$

PROOF. The proof is standard. We will only show the first equality. Since $\mathscr{E}(F(t)|\mathbf{k}) = \hat{F}(t)$, and $\mathscr{E}\hat{F}(t) = \mathscr{E}F(t) = F_0(t)$, we have

$$\begin{aligned}
&\mathscr{E}\int [F(t) - \hat{F}(t)]^2\, dW(t) \\
&\quad = \int (\mathscr{E}[F(t) - \mathscr{E}F(t)]^2 - 2\mathscr{E}[\mathscr{E}\{(F(t) - \mathscr{E}F(t))(\hat{F}(t) - \mathscr{E}F(t))|\mathbf{k}\}] \\
&\qquad + \mathscr{E}[\hat{F}(t) - \mathscr{E}F(t)]^2)\, dW(t) \\
&\quad = \int [\mathrm{Var}F(t) - \mathscr{E}\{\hat{F}(t) - \mathscr{E}F(t)\}^2]\, dW(t) = \int [\mathrm{Var}F(t) - \mathrm{Var}\,\hat{F}(t)]\, dW(t).
\end{aligned}$$

From Theorem 2.1, we see that minimizing the Bayes risk $\mathscr{E}\int (F(t) - \hat{F}(t))^2\, dW(t)$ over $\mathbf{t}$ is equivalent to maximizing $\int \mathrm{Var}\,\hat{F}(t)\, dW(t)$ over $\mathbf{t}$, since $\mathrm{Var}\,F(t) = F_0(t)(1 - F_0(t))/(M + 1)$ is independent of the drug levels $\mathbf{t}$. Intuitively we can make the following interpretation. If there were no observations, our best guess at $F(t)$ would be $\mathscr{E}F(t) = F_0(t)$. The Bayes risk is $\int \mathrm{Var}\,F(t)\, dW(t)$. If there are observations we can reduce the Bayes risk by $\int \mathrm{Var}\,\hat{F}(t)\, dW(t)$. The best design corresponds to testing animals at drug levels where the weighted average of $\mathrm{Var}\,\hat{F}(t)$ is big. From the second equality of Theorem 2.1, the design which maximizes $\int \mathscr{E}\hat{F}(t)^2\, dW(t)$ also solves the design problem.

3. Solution to the design problem for $L = 1$. In this section, we assume that $n(=n_1)$ animals are treated at the same drug dosage level, say t_1. Mantel (1967) has given an example in which a single dose experiment is significant when prior information is available. We assume F is taken from the Dirichlet process with parameter MF_0, where F_0 is assumed to be continuous. Then we can find a strictly increasing function ρ such that $F_0(\rho(t)) = t$, $0 \le t \le 1$. Moreover, $G(t) = F(\rho(t))$ is the random Dirichlet distribution with parameter $MF_0(\rho(t)) = Mt$, $0 \le t \le 1$. We also have

$$\begin{aligned}
L(F, \hat{F}) &= \int (F(t) - \hat{F}(t))^2\, dW(t) = \int [F(\rho(t)) - \hat{F}(\rho(t))]^2\, dW(\rho(t)) \\
&= \int (G(t) - \hat{G}(t))^2\, dW(\rho(t)).
\end{aligned}$$

Therefore, by a reparametrization of t, we can assume without loss of generality $F_0(t) = t$, $0 \le t \le 1$. We will give a general formula for solving the design problem and treat special cases such as: $W(t)$ is concentrated on the two points t_0 and $1 - t_0$, and $dW(t)/dt = 1$, t, $1/[t(1-t)]$, or $t(1-t)$.

THEOREM 3.1. *Assume* (2.1) *with* $F_0(t) = t$, $0 \le t \le 1$, $L = 1$, *and* W *is a measure such that* $\int_0^1 t(1-t)\, dW(t) < \infty$. *Then,* (i) *the solution to the design problem is given by* $t_1 \in (0, 1)$ *such that the following function of* t_1 *is maximized*:

$$h(t_1) = \frac{1 - t_1}{t_1} \int_{[0,t_1]} t^2\, dW(t) + \frac{t_1}{1 - t_1} \int_{(t_1,1]} (1 - t)^2\, dW(t);$$

(ii) *if* $W(t)$ *is continuous, then the optimal solution is necessarily a root of the following equation in* t_1:

$$\frac{1}{(1 - t_1)^2} \int_{t_1}^{1} (1 - t)^2\, dW(t) - \frac{1}{t_1^2} \int_0^{t_1} t^2\, dW(t) = 0. \tag{3.1}$$

PROOF. (i) By Theorem 2.1, we have:

$$(3.2)\qquad \mathscr{E}\int_0^1 [F(t)-\hat{F}(t)]^2\,dW(t)=\int_0^1 [t(1-t)/(M+1)]\,dW(t)-\int_0^1 \operatorname{Var}\hat{F}(t)\,dW(t).$$

(The condition $\int_0^1 t(1-t)\,dW(t)<\infty$ is imposed in the theorem to assure that the Bayes risk is finite.) It follows from (2.4) and straightforward calculation (see Kuo 1980a or b) that

$$(3.3)\qquad \begin{aligned}\int_0^1 \operatorname{Var}\hat{F}(t)\,dW(t) &= \frac{\mathscr{E}\hat{F}(t_1)^2-t_1^2}{t_1^2}\int_{[0,t_1]} t^2\,dW(t)\\ &\quad+\frac{\mathscr{E}\hat{F}(t_1)^2-t_1^2}{(1-t_1)^2}\int_{(t_1,1]}(1-t)^2\,dW(t).\end{aligned}$$

Now, from (2.3),

$$\mathscr{E}\hat{F}(t_1)^2=\mathscr{E}\left(\frac{k_1+Mt_1}{n_1+M}\right)^2=\frac{\mathscr{E}k_1^2+2Mt_1\mathscr{E}k_1+(Mt_1)^2}{(n_1+M)^2}.$$

We need

$$\begin{aligned}\mathscr{E}k_1&=\mathscr{E}(\mathscr{E}(k_1\mid F))=\mathscr{E}(n_1F(t_1))=n_1t_1 \text{ and}\\ \mathscr{E}k_1^2&=\mathscr{E}(\mathscr{E}(k_1^2\mid F))=\mathscr{E}[n_1F(t_1)(1-F(t_1))+n_1^2F(t_1)^2]\\ &=n_1Mt_1(1-t_1)/(M+1)+n_1^2t_1(Mt_1+1)/(M+1).\end{aligned}$$

It now can be verified that

$$\mathscr{E}\hat{F}(t_1)^2-t_1^2=n_1t_1(1-t_1)/[(M+1)(n_1+M)].$$

From (3.3) we have

$$(3.4)\qquad \begin{aligned}\int_0^1 \operatorname{Var}\hat{F}(t)\,dW(t)&=c\cdot\left(\int\left[\frac{1-t_1}{t_1}t^2I_{[0,t_1]}(t)+\frac{t_1}{1-t_1}(1-t)^2I_{(t_1,1]}(t)\right]dW(t)\right)\\ &=ch(t_1),\end{aligned}$$

where

$$c=n_1/[(M+1)(n_1+M)].$$

To prove the existence of an optimal design, we need to show that h has a maximum for $0<t_1<1$. It suffices to show h is continuous and nonnegative on (0, 1) and $h(t_1)\to 0$ as $t_1\to 0$ or $t_1\to 1$. It is clear that h is nonnegative. To show h is continuous, take any sequence $a_n\in(0,1)$ with $\lim_{n\to\infty}a_n=t_1$; let

$$\psi_n(t)=\frac{1-a_n}{a_n}t^2I_{[0,a_n]}(t)+\frac{a_n}{1-a_n}(1-t)^2I_{(a_n,1]}(t).$$

It can be verified that for every n, $0\le\psi_n(t)\le t(1-t)\in L^1(W)$. Therefore,

$$\lim_{n\to\infty}h(a_n)=\int\lim_{n\to\infty}\psi_n(t)\,dW(t)=h(t_1).$$

This proves that h is continuous on (0, 1). Moreover, $h(t_1)\to 0$ when $t_1\to 0$. To see this, take $a_n\in(0,1)$ with $\lim_{n\to\infty}a_n=0$. Then

$$\lim_{n\to\infty}h(a_n)=\int_0^1\lim_{n\to\infty}\psi_n(t)\,dW(t)=0.$$

Similarly, for the case $t_1 \to 1$, $h(t_1) \to 0$.

(ii) We now assume $W(t)$ (regarded as a function) is continuous, and first prove the result for $W([0, 1]) < \infty$. Then, we can rewrite $h(t_1)$ using integration by parts (see Theorem 6.30 of Rudin, 1964):

$$h(t_1) = -2\frac{1-t_1}{t_1}\int_0^{t_1} tW(t)\,dt + 2\frac{t_1}{1-t_1}\int_{t_1}^1 (1-t)W(t)\,dt.$$

Applying Theorem 6.30 of Rudin again, we obtain

$$\frac{d}{dt_1}h(t_1) = -\frac{1}{t_1^2}\int_0^{t_1} t^2\,dW(t) + \frac{1}{(1-t_1)^2}\int_{t_1}^1 (1-t)^2\,dW(t). \tag{3.5}$$

Therefore, the optimal design is given by a root of $(d/dt_1)h(t_1) = 0$.

If $W([0, 1])$ is not finite, the same result as above follows. Since that value of t_1 which maximizes the function h is shown in (i) to be bounded away from 0 or 1, we can rewrite $h(t_1)$ as:

$$\int \left(\frac{1-t_1}{t_1}t^2 I_{(\varepsilon,t_1]}(t) + \frac{t_1}{1-t_1}(1-t)^2 I_{(t_1,1-\varepsilon)}(t)\right) dW(t)$$

$$+ \frac{1-t_1}{t_1}\int_{[0,\varepsilon]} t^2\,dW(t) + \frac{t_1}{1-t_1}\int_{[1-\varepsilon,1]} (1-t)^2\,dW(t)$$

where $W((\varepsilon, 1-\varepsilon)) < \infty$ because $\int_0^1 t(1-t)\,dW(t) < \infty$, and the above integration by parts is applicable to the first integral. Then we can differentiate with respect to t_1 and obtain (3.5).

From Theorem 3.1, we can derive the optimal design for one drug level for specific W. The following corollary gives the results for $W(t) = t$. It can be verified by a straightforward computation from (3.1), (3.2), and (3.4).

COROLLARY 3.1. *Assume $W(t) = t$ in Theorem* 3.1. *Then the optimal design is given by $t_1 = \frac{1}{2}$, and the Bayes risk for it equals $(n_1 + 2M)/[12(M+1)(n_1+M)]$.*

REMARK. Let us observe this Bayes risk $\to 1/[12(M+1)]$(not 0), as $n_1 \to \infty$ for fixed M. Presumably this happens because we are only learning more about the potency curve at a single point and the loss occurs throughout. ($W(t) = t$ assigns a uniform weight.) This would suggest that it is necessary to increase the number of drug levels to make the Bayes risk arbitrarily small.

For other weight functions, a few examples are listed as follows. When $dW(t) = t\,dt$, $dW(t) = t(1-t)\,dt$, or $dW(t) = 1/[t(1-t)]\,dt$, then the corresponding optimal designs, which are derived in detail in Kuo (1980a), are given by $t_1 = (1+\sqrt{7})/6 \cong .61$, $t_1 = \frac{1}{2}$, or $t_1 = \frac{1}{2}$ respectively. When the measure $W(t)$ has mass 1 at $t_0 \in (0, \frac{1}{2})$ and mass 1 at $1 - t_0$, zero mass elsewhere, then the optimal design is given by $t_1 = t_0$ or $t_1 = 1 - t_0$ from the result of part (i) of Theorem 3.1. Note that $h(t_1)$ is not differentiable at t_0 or $1 - t_0$ in the latter example.

4. Solution to the design problem for $L \le 2$. When $L \le 2$, let t_1 and $t_2 (t_1 \le t_2)$ be the drug levels. As before, if F_0 is continuous, we can assume $F_0(t) = t$, $0 \le t \le 1$ without loss of generality. We treat only the special case $W(t) = t$, and $n_1 = n_2 = 1$. For other weight functions W, the same method could be applied to find the optimal solution. However, we cannot generalize the results to larger sample sizes because of the difficulties in computing $\mathscr{E}\hat{F}(t_1)^2$, $\mathscr{E}\hat{F}(t_2)^2$, and $\mathscr{E}\hat{F}(t_1)\hat{F}(t_2)$.

THEOREM 4.1. *Assume* (2.1) *with* $F_0(t) = t$, $W(t) = t$, $0 \le t \le 1$, $n_1 = n_2 = 1$. *Then the solution to the optimal design problem is given by* $t_1 = \frac{1}{2} - c_0$, $t_2 = \frac{1}{2} + c_0$, *where* c_0 *is the unique root of the following equation*:

$$(4.1) \quad -(2M+4)c + 1 - \frac{16c(M+1)(Mc+M+2)}{(2Mc+M+2)^2} + \frac{16c(M+1)(2Mc+M+4c)}{(M+4Mc^2+4Mc+8c)^2} = 0.$$

PROOF. From the second equality of Theorem 2.1, we want to maximize $\int_0^1 \mathscr{E}\hat{F}(t)^2\,dt$ for the optimal design. It follows from (2.2) and straightforward calculations that

$$\int_0^1 \mathscr{E}\hat{F}(t)^2\,dt = [t_2\mathscr{E}\hat{F}(t_1)^2 + (1-t_1)\mathscr{E}\hat{F}(t_2)^2 + (t_2-t_1)\mathscr{E}\hat{F}(t_1)\hat{F}(t_2) + (1-t_2^2)]/3.$$

Denote this function of t_1, t_2 by $g(t_1, t_2)$.

When $n_1 = 1$, $n_2 = 1$, the observed $\mathbf{k} = (k_1, k_2)$ can only take on the values (0, 0), (0, 1), (1, 0), (1, 1), with the following probabilities:

$$\mathscr{P}(\mathbf{k} = (0,0)) = \mathscr{E}[(1-y_1)y_3] = [\beta_2\beta_3 + \beta_3(\beta_3+1)]/[M(M+1)],$$

$$\mathscr{P}(\mathbf{k} = (0,1)) = \mathscr{E}[(1-y_1)(y_1+y_2)] = [\beta_2(M+1) + \beta_1\beta_3]/[M(M+1)],$$

$$\mathscr{P}(\mathbf{k} = (1,0)) = \mathscr{E}[y_1y_3] = \beta_1\beta_3]/[M(M+1)],$$

and

$$\mathscr{P}(\mathbf{k} = (1,1)) = \mathscr{E}[y_1(y_1+y_2)] = [\beta_1(\beta_1+1) + \beta_1\beta_2]/[M(M+1)],$$

where $\mathbf{y}$, $\boldsymbol{\beta}$ are defined in Section 2.

We can obtain the posterior distribution of $\mathbf{y}$ and $\mathbf{k}$ from the likelihood function, and estimate $F(t_1)$ and $F(t_2)$ according to the posterior distribution. For example, given $\mathbf{k} = (0, 0)$, the posterior distribution of $\mathbf{y}$ is

$$\frac{\beta_2}{\beta_2+\beta_3+1}\mathscr{D}(\beta_1, \beta_2+1, \beta_3+1) + \frac{\beta_3+1}{\beta_2+\beta_3+1}\mathscr{D}(\beta_1, \beta_2, \beta_3+2),$$

where $\mathscr{D}$ denotes the Dirichlet distribution. Therefore, we can estimate $F(t_1)$ and $F(t_2)$ respectively by $\beta_1/(M+2)$ and $(\beta_1+\beta_2)/[(M+2) + \beta_2/[(M+2)(\beta_2+\beta_3+1)]$, which are denoted by $\hat{F}(t_1; (0,0))$ and $\hat{F}(t_2; (0,0))$ for emphasis. Then we can proceed to compute $g(t_1, t_2)$ by computing the following quantities:

$$\mathscr{E}\hat{F}(t_i)^2 = \textstyle\sum_{k_2=0}^{1}\sum_{k_1=0}^{1}[\hat{F}(t_i; (k_1,k_2))]^2\mathscr{P}(\mathbf{k} = (k_1,k_2)), \quad i = 1, 2,$$

and

$$\mathscr{E}\hat{F}(t_1)\hat{F}(t_2) = \textstyle\sum_{k_2=0}^{1}\sum_{k_1=0}^{1}\hat{F}(t_1; (k_1,k_2))\hat{F}(t_2; (k_1,k_2))\mathscr{P}(\mathbf{k} = (k_1,k_2)).$$

For a detailed computation, refer to Kuo (1980a).

It will be easier to maximize the function $g(t_1, t_2)$ if we introduce the reparametrization

$$t_1 = s + \tfrac{1}{2} - c, \quad t_2 = s + \tfrac{1}{2} + c.$$

Let $h(s, c) = 3(M+2)^2(M+1)g(s+\frac{1}{2}-c, s+\frac{1}{2}+c)$. It can be verified as in Kuo (1980a) that the function $h(s, c)$ reduces to the following:

$$h(s,c) = -(2M+4)(s^2 - \tfrac{1}{4} + c^2) + 2c + (M+1)(M+2)^2 + \frac{(8Mc^3+4c^2)(2Mc+M+2)}{(M/2+Mc+1)^2 - M^2s^2} + \frac{4c^2 - 16c^2(Mc^2+Mc+2c)}{Mc^2+Mc+2c+M/4-Ms^2}.$$

Note that maximizing $g(t_1, t_2)$ with the restrictions $0 < t_1 \leq t_2 < 1$ is equivalent to maximizing $h(s, c)$ with the following restrictions:

$$0 \leq c < \tfrac{1}{2} \text{ and } -(\tfrac{1}{2} - c) < s < \tfrac{1}{2} - c. \tag{4.2}$$

It is shown in Kuo (1980a or b) that $h(s, c)$ subject to these restrictions is maximized by $(0, c_0)$ where c_0 is the unique root of (4.1). Therefore $g(t_1, t_2)$ is maximized by $(\frac{1}{2} - c_0, \frac{1}{2} + c_0)$.

Using Theorem 4.1 we can obtain the optimal design for any value of M. For example,
(1) $M \to 0$, the optimal design $\to (\frac{1}{3}, \frac{2}{3})$.
(2) $M = 1$, the optimal design $\cong (0.365, 0.635)$.
(3) $M = 2$, the optimal design $\cong (0.389, 0.611)$.
(4) $M \to \infty$, the optimal design $\to (\frac{1}{2}, \frac{1}{2})$.

REMARK. Let us note Theorem 4.1 is proved for $t_1 \leq t_2$, which includes the one dose case. The conclusion implies that except for $M \to \infty$, it is better to test one animal each at two distinct levels than both of them at the same level. If the number of drug levels L is fixed at two beforehand, then the proof of Theorem 4.1 still holds with $0 < c$ replacing $0 \leq c$ in (4.2), and the optimal design remains the same.

5. Solution to the design problem for the case $W(t) = t$, $M \to 0$. It takes a considerable amount of effort to find the optimal solution for the case $L \leq 2$, $n_1 = n_2 = 1$, and $W(t) = t$. It is much more difficult to compute the optimal solution for $L \geq 3$, even with the assumption $n_i = 1$ for all i, since $\mathscr{E}\hat{F}(t_1)^2$, $\mathscr{E}\hat{F}(t_2)^2$, $\cdots$ each is an average of 2^L terms. Nevertheless, for the case $M \to 0$, we have a simple expression for $\mathscr{E}\hat{F}(t_1)^2$, $\mathscr{E}\hat{F}(t_2)^2$, $\cdots$. Consequently, we can obtain an optimal solution in this case.

THEOREM 5.1. *Assume* (2.1) *with* $F_0(t) = t$, $W(t) = t$, $0 \leq t \leq 1$, *and* n *animals are tested. Then* (i) *if the number of drug levels is fixed beforehand at* $L \leq n$, *the optimal design converges as* $M \to 0$ *to the equally spaced design,* $t_i = i/(L + 1)$ *for* $i = 1, \cdots, L$, *and the allocation of animals to these drug levels can be arbitrary with* $\sum_{\ell=1}^{L} n_\ell = n$, $n_\ell \geq 1$; (ii) *if the number of drug levels is not fixed beforehand, the optimal design converges as* $M \to 0$ *to the equally spaced design with one animal each at the drug levels, i.e.,* $t_i = i/(n + 1)$, $n_i = 1$, *for* $i = 1, \cdots, n$.

PROOF. (i) Let $\mathbf{n} = (n_1, \cdots, n_L)$ denote an arbitrary allocation of the animals treated at doses $\mathbf{t} = (t_1, \cdots, t_L)$ and let $\mathbf{k} = (k_1, \cdots, k_L)$ denote the number of animals which react positively at the respective doses, where $0 < t_1 < t_2 < \cdots < t_L < 1$, and $\sum_{\ell=1}^{L} n_\ell = n$, $n_\ell \geq 1$, for every ℓ. From Theorem 2.1, we need to maximize $g_M(\mathbf{t}) = \int_0^1 \mathscr{E}_M\hat{F}(t)^2\, dt$ (note that $\hat{F}$ depends on $\mathbf{t}$). It can be verified that

$$g_M(\mathbf{t}) = [\textstyle\sum_{i=1}^{L} (t_{i+1} - t_i)\mathscr{E}_M\hat{F}(t_i)^2 + \sum_{i=1}^{L-1} (t_{i+1} - t_i)\mathscr{E}_M\hat{F}(t_i)\hat{F}(t_{i+1}) + 1 - t_L^2]/3, \tag{5.1}$$

where $t_0 = 0$, $t_{L+1} = 1$.

Define, for every $j = 1, \cdots, L + 1$, $\mathbf{k}_j = (0, \cdots, 0, n_j, \cdots, n_L)$, where the first $j - 1$ entries are 0; i.e. all animals react negatively at levels 1 through $j - 1$, and positively at levels j through L. It will be shown that all other events have negligible probabilities for small M. We can compute from Section 2 of the probabilities of the events $\mathbf{k}_j, j = 1, \cdots, L + 1$:

$$\mathscr{P}(\mathbf{k} = \mathbf{k}_j) = \mathscr{E}[\textstyle\prod_{i=1}^{j-1} (1 - y_1 - \cdots - y_i)^{n_i} \prod_{i=j}^{L} (y_1 + \cdots + y_i)^{n_i}] = \mathscr{E}[y_j^n + Cy_j^{n-1}y_i + \cdots]$$

where $i \neq j$, and C is a finite constant. Thus

$$\mathscr{P}(\mathbf{k} = \mathbf{k}_j) = \beta_j^{(n)}/M^{(n)} + O(M) = t_j - t_{j-1} + O(M), \tag{5.2}$$

where the superscript (n) is defined by $a^{(n)} = a(a + 1) \cdots (a + n - 1)$, and the terms

$O(M)$ are the product of M and a polynomial in $\mathbf{t}$ of degree $\leq n$. It follows from (5.2) that $\sum_{j=1}^{L+1} \mathscr{P}(\mathbf{k} = \mathbf{k}_j) \to 1$ as $M \to 0$.

The posterior distribution of $\mathbf{y}$ given $\mathbf{k}$ is a mixture of Dirichlet distributions which approaches a single Dirichlet as $M \to 0$. Given the observations, say $\mathbf{k}_j$ for example, we compute $\hat{F}(t_i; \mathbf{k}_j)$ for $i = 1, \cdots, L$ by means of the posterior distribution. It follows that

$$\begin{aligned} \hat{F}(t_i; \mathbf{k}_j) &= O(M) \qquad \text{when } 1 \leq i < j; \\ &= 1 + O(M) \quad \text{when } j \leq i \leq L. \end{aligned} \tag{5.3}$$

The quantities $\mathscr{E}_M \hat{F}(t_i)^2$ and $\mathscr{E}_M \hat{F}(t_i)\hat{F}(t_{i+1})$ can be computed from (5.2) and (5.3):

$$\mathscr{E}_M \hat{F}(t_i)^2 = \sum_{j=1}^{L+1} \hat{F}(t_i; \mathbf{k}_j)^2 \mathscr{P}(\mathbf{k}_j) + O(M) = t_i + O(M)$$

$$\begin{aligned} \mathscr{E}_M \hat{F}(t_i)\hat{F}(t_{i+1}) &= \sum_{j=1}^{L+1} \hat{F}(t_i; \mathbf{k}_j)\hat{F}(t_{i+1}; \mathbf{k}_j)\, \mathscr{P}(\mathbf{k}_j) + O(M) \\ &= t_i + O(M). \end{aligned}$$

As $M \to 0$, $g_M(\mathbf{t})$ converges to

$$g_0(\mathbf{t}) = (t_L + \sum_{i=1}^{L-1} t_i t_{i+1} - \sum_{i=1}^{L} t_i^2 + 1)/3. \tag{5.4}$$

In fact, it is not hard to show the convergence is uniform in $\mathbf{t}$. Therefore, $\lim_{M\to 0} \max_{\mathbf{t}} g_M(\mathbf{t}) = \max_{\mathbf{t}} \lim_{M\to 0} g_M(\mathbf{t})$. Moreover, $g_0(\mathbf{t})$, which is independent of $\mathbf{n} = (n_1, \ldots, n_L)$, is maximized by $t_i = i/(L+1)$ (see Kuo 1980a, b). Hence, the optimal design as $M \to 0$ is given by $t_i = i/(L+1)$, $i = 1, \ldots, L$.

(ii) When L is not fixed beforehand, the experimenter can choose L to be any number from 1 to n. Given L, we have obtained the optimal design, namely, $t_i = i/(L+1)$ for each $i = 1, \cdots, L$. Let this optimal solution be denoted by $\mathbf{t}_L^0$. We shall compare the Bayes risks for each of the optimal designs for $1 \leq L \leq n$.

$$\begin{aligned} \lim_{M\to 0} r(\mathbf{t}_L^0) &= \lim_{M\to 0} \left[\int_0^1 t(Mt+1)/(M+1)\, dt - \max_{\mathbf{t}} g_M(\mathbf{t}) \right] \\ &= \tfrac{1}{6} - L/[6(L+1)], \quad \text{from (5.4)}. \end{aligned}$$

Therefore, since $L/(L+1) \leq n/(n+1)$ for $L \leq n$, $\lim_{M\to 0} r(\mathbf{t}_n^0) \leq \lim_{M\to 0} r(\mathbf{t}_L^0)$. Hence $L = n$, one animal each at $t_i = i/(n+1)$, is the optimal design.

REMARK. Although Theorem 5.1 is proved for $F_0(t) = t$, the uniform prior, we can obtain optimal designs as $M \to 0$ for any $F_0(t)$ by using the transformation discussed in the beginning of Section 3. For example, let $F_0(t) = \Phi(t)$ be the standard normal distribution function. Then the optimal design is given by $t_i = \Phi^{-1}(i/(n+1))$, $i = 1, \ldots, n$. This confirms Ramsey's result (1972). While only $W(t) = t$ is treated here, it is rather straightforward to modify (5.1) for other weight functions and find the optimal design.

Acknowledgment. This article is revised from Chapter 2 of the author's dissertation written at UCLA under the direction of Professor Thomas Ferguson. I am deeply grateful to him for his patient guidance and encouragement throughout the project, and also for his numerous suggestions and comments which greatly improved this article. I would also like to thank Michael Cohen for very helpful discussions. Thanks also to the referee and the associate editor for their comments.

REFERENCES

ABDELBASIT, K. M. and PLACKETT, R. L. (1981). Experimental design for categorized data. *Internat. Statist. Rev.* **49** 111–126.

ANTONIAK, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.* **2** 1152–1174.

BHATTACHARYA, P. K. (1981). Posterior distribution of a Dirichlet process from quantal response data. *Ann. Statist.* **9** 803–811.

COCHRAN, W. G. (1973). Experiments for nonlinear functions. *J. Amer. Statist. Assoc.* **68** 771–781.

FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230.

FINNEY, D. J. (1978). *Statistical Method in Biological Assay*, 3rd ed. Griffin, London.

KRAFT, C. and VAN EEDEN, C. (1964). Bayesian bio-assay. *Ann. Math. Statist.* **35** 886–890.

KUO, L. (1980a). Computations and applications of mixtures of Dirichlet processes. Ph.D. thesis, Dept. of Mathematics, University of California, Los Angeles.

KUO, L. (1980b). Bayesian bioassay design. Tech. Rep. No. 101, Dept. of Statistics, University of Michigan.

MANTEL, N. (1967). Adaptation of Kärber's method for estimating the exponential parameter from quantal data, and its relationship to birth, death, and branching processes. *Biometrics* **23** 739–746.

RAMSEY, F. L. (1972). A Bayesian approach to bio-assay. *Biometrics* **28** 841–858.

RUDIN, W. (1964). *Principles of Mathematical Analysis*, 2nd ed. McGraw-Hill, New York.

TSUTAKAWA, R. K. (1982). Statistical methods in bioassay. In *Encyclopedia of Statistical Sciences*, Vol. 1 (eds., S. Kotz, N. L. Johnson, and C. B. Read) 236–243. Wiley, New York.

WESLEY, M. N. (1976). Bioassay: estimating the mean of the tolerance distribution. Tech. Rep. No. 17, Division of Biostatistics, Stanford University.

DEPARTMENT OF APPLIED MATHEMATICS
AND STATISTICS
STATE UNIVERSITY OF NEW YORK
STONY BROOK, N.Y. 11794