

# Optimal Designs and Limiting Optimal Designs for a Trinomial Response

BY SHENGHUA KELLY FAN

*Department of Statistics and Applied Probability, National University of Singapore,  
Singapore 117543, Singapore*

AND KATHRYN CHALONER

*Department of Biostatistics, University of Iowa, Iowa City, IA 52242, U.S.A.*

## SUMMARY

Designs for the continuation-ratio model for a trinomial response will be described. For the situation where the three response categories are: “no response”, “efficacy” and “adverse reaction”, both D-optimal designs and c-optimal designs for estimating the dose with the maximum probability of efficacy are found.

Optimal designs are not available in closed form but designs with closed form expression are found which are approximately optimal. Motivated by these designs, a new concept is defined, “limiting optimality”, where a sequence of designs is said to be optimal in an asymptotic sense for a sequence of prior distributions. A member of the sequence is approximately optimal for the corresponding prior distribution. Algebraic forms of limiting optimal designs are derived for a special case of the model where the slopes are equal. They are shown to be very efficient, provide insight, and also provide starting designs for numerical algorithms.

*AMS Subject Classifications:* 62K05.

*Key words:* Continuation-ratio model, Optimal design, Trinomial responses.

## 1 Introduction

There are many important design problems where the dose response is a multinomial distribution: Zocchi and Atkinson (1999) give a toxicology example with a trinomial

response of pupae to dose (survival after emergence/death after emergence/death before emergence); Thall and Russell (1998) present a strategy for phase I/II trials where the response of subjects to dose is trinomial (no reaction/efficacy/adverse outcome); Heise and Myers (1996) discuss a clinical trial with a bivariate binary response to dose, efficacy (yes/no) and toxicity (yes/no). This last example can be thought of as a multinomial response with 4 cells; the two cells corresponding to the occurrence of toxicity can be collapsed together to give a trinomial response. Other examples are given in Glonek and McCullagh (1995), Glonek (1996), and Zhu, Krewski and Ross (1994). This literature provides algorithms for conducting certain types of clinical trials with multinomial responses and also provides optimal designs for a few sets of specified parameter values. In this paper, a particular model for multinomial responses is used and general properties of its optimal design for trinomial responses are given.

An experimental design will be regarded as a probability measure on the dose domain  $\mathcal{X}$ . For a design  $\eta$  putting weight  $m_i$  at the dose  $x_i$  for  $i = 1, 2, \dots, k$ , the  $m_i$ 's are non-negative and sum to one and will be written as  $\eta = (m_1, m_2, \dots, m_k)$  at  $(x_1, x_2, \dots, x_k)$ . For a sample size  $n$ , the values  $nm_i$  can be rounded to integers in a systematic way (Pukelsheim, 1993, Chapter 12).

Designs which optimize concave criteria based on the Fisher information matrix are considered. The inverse of the Fisher information matrix gives the asymptotic variance-covariance matrix for the maximum likelihood estimate (MLE) of the unknown parameters. The information matrix,  $M(\theta, \eta)$ , typically depends on the parameters  $\theta$  and the design  $\eta$ .

Local D-optimality chooses the design which maximizes  $\log \det M(\theta, \eta)$  for a specified value  $\theta$ . If the specified value of  $\theta$  is close to the true value then maximizing this criterion should make the asymptotic variance-covariance matrix of the MLE of  $\theta$  small. The same criterion averaged over a "prior" distribution,  $\pi$ , on the parameters gives "Bayesian" D-optimality: maximizing  $\int \log \det M(\theta, \eta) d\pi(\theta)$ . This criterion is typically a more robust criterion than local D-optimality in that it is often more efficient for  $\theta$  values close to the

best guess and includes local D-optimality as a special case (Chaloner and Larntz, 1989; Chaloner and Verdinelli, 1995). Applications and examples of locally optimal designs are given in Wu (1988), Kitsos, Titterington and Torsney (1988) and Ford, Torsney and Wu (1992).

A comparison between the continuation-ratio model and the proportional odds model is given in Section 2 for one data set. Optimal designs for the continuation-ratio model and some approximately optimal designs are presented in Section 3.

The optimal design is not always appropriate to use in practice, but it can provide insight for practical use. For example, in a phase I/II dose-finding clinical trial, in which patients are allocated to doses in sequential cohorts, the conventional approach is to assign an entire cohort to a single dose. Allocating a cohort to more than one dose might, however, be more efficient since the optimal designs have more than one design point. See Kpamegan and Flournoy (2001) for recent work in sequentially allocating a cohort to two successive doses.

The general algebraic form of the optimal design can be very difficult to find. A new design concept, “limiting optimality”, motivated by the approximately optimal designs in Section 3, is therefore introduced in Section 4. A sequence of designs is said to be limiting optimal for a sequence of prior distributions if the limiting efficiency is one.

The algebraic form of a sequence of limiting optimal designs is typically easier to find than that of the corresponding optimal designs. Algebraic forms of limiting c-optimal, limiting D-optimal, and Bayesian limiting D-optimal designs for the continuation-ratio model are given as examples and are shown to be very efficient. Note that the efficiency of a design  $\eta$  is defined to be the sample size required for an experiment using the optimal design to reach the same value of the criterion as an experiment using the design  $\eta$  with sample size one. This “limiting optimality” concept can also be applied to other models. They may or may not be very efficient for other models, but they can at least provide starting designs for finding optimal designs numerically. Possible future work is discussed in Section 5.

## 2 The Trinomial Response Model

In a phase I/II clinical trial, the response of a patient may be classified as: “no reaction” when neither toxicity nor efficacy occurs; “efficacy” for efficacy without toxicity; and “adverse reaction” for toxicity. The response when  $n_i$  subjects are given a dose  $x_i$  (often in log units) is therefore trinomial,  $(y_{1i}, y_{2i}, y_{3i})$ ,  $y_{1i} + y_{2i} + y_{3i} = n_i$  and the corresponding cell probabilities are  $(p_1(\theta, x_i), p_2(\theta, x_i), p_3(\theta, x_i))$  where  $\theta$  denotes the parameters of the model. For any  $x$  and  $\theta$ ,  $p_1(\theta, x) + p_2(\theta, x) + p_3(\theta, x) = 1$ . The values of  $x$  can be chosen from some set  $\mathcal{X}$ . Designs for the proportional odds model (Agresti, 1990, Chapter 9) are discussed in Thall and Russell, (1998). The proportional odds model is:

$$\begin{aligned}\log[p_3(\theta, x)/(1 - p_3(\theta, x))] &= a + bx \\ \log[(p_2(\theta, x) + p_3(\theta, x))/p_1(\theta, x)] &= a + u + bx\end{aligned}$$

where  $u \geq 0, b > 0$ . This model assumes that the effect of dose is the same across the cumulative logits. The continuation-ratio model does not have this assumption (Agresti, 1990, Chapter 9). Assuming  $u \geq 0$  and  $b_1, b_2 > 0$ , the continuation-ratio model is:

$$\log[p_3(\theta, x)/(1 - p_3(\theta, x))] = a_1 + b_1x \tag{1}$$

$$\log[p_2(\theta, x)/p_1(\theta, x)] = a_1 + u + b_2x. \tag{2}$$

A special case is when  $b_1 = b_2$  and this will be referred to as the “constant slopes continuation-ratio model”, and is not the same as the proportional odds model.

The assumption of proportional odds is unlikely to be valid if the number of categories of response is greater than 3. Even for a trinomial response the assumption might sometimes fail. Agresti (1990, p. 320) gives a data set where the continuation-ratio model fits much better than the proportional odds model: the data, together with the fit of the two models are shown in Figure 1. In this paper, we will focus on the continuation-ratio model.

### 3 Optimal Designs

Figure 2 gives plots of  $(p_1(\theta, x), p_2(\theta, x), p_3(\theta, x))$  against  $x$  for 3 values of  $\theta$  for the constant slopes continuation-ratio model, where  $b_1 = b_2$ . The plots indicate that a larger value of  $u$  gives a wider range of dose levels where the probability of adverse reaction is low and the probability of efficacy is high. In the following discussion,  $\theta$  will be rewritten as  $(u, a_1, b_1)$  for convenience but the information matrix with  $\theta = (a_1, b_1, a_2 = u + a_1)$  in Appendix A is used to find the optimal designs. All optimal designs in this paper have been verified by the Equivalence Theorem of Whittle (1973) or Silvey (1980). The compact design space,  $\mathcal{X}$ , is assumed to be large enough to include the optimal designs in the interior.

If  $\eta_0^* = (m_1^*, m_2^*, \dots, m_k^*)$  at  $(x_1^*, x_2^*, \dots, x_k^*)$  is locally D-optimal for  $\theta_0 = (u, 0, 1)$  then it can be shown that  $\eta^* = (m_1^*, m_2^*, \dots, m_k^*)$  at  $(\frac{x_1^* - a_1}{b_1}, \frac{x_2^* - a_1}{b_1}, \dots, \frac{x_k^* - a_1}{b_1})$  is also locally D-optimal for  $\theta = (u, a_1, b_1)$ , providing that  $m_i^*$ 's are all in  $\mathcal{X}$ . The proof uses the result that  $\det(M(\theta, \eta)) = b_1^{-2} \det(M(\theta_0, \eta_0))$ . (See Fan, 1999, for details).

For Bayesian D-optimality let the prior distribution  $\pi$  put weight  $w_i$  at  $\theta_i = (u_i, a_1, b_1)$ ,  $i = 1, 2, \dots, h$  and let the prior distribution  $\pi_0$  put the same weight  $w_i$  at  $\theta_i = (u_i, 0, 1)$ ,  $i = 1, 2, \dots, h$ . It can be shown by straightforward algebra that if design  $\eta_0^* = (m_1^*, m_2^*, \dots, m_k^*)$  at  $(x_1^*, x_2^*, \dots, x_k^*)$  is the Bayesian D-optimal design for prior distribution  $\pi_0$  then design  $\eta^* = (m_1^*, m_2^*, \dots, m_k^*)$  at  $(\frac{x_1^* - a_1}{b_1}, \frac{x_2^* - a_1}{b_1}, \dots, \frac{x_k^* - a_1}{b_1})$  is the Bayesian D-optimal design for prior distribution  $\pi$ , providing all design points are in  $\mathcal{X}$ .

In this paper, therefore,  $a_1 = 0, b_1 = 1$  will be assumed for deriving locally and Bayesian D-optimal designs.

Locally D-optimal designs are found numerically to have two, three, or four design points; the number of design points increases with  $u$ . Table 1 (numbers are after rounding) illustrates that for  $u = 0$ , the locally D-optimal design has 2 points; for  $u = 5$ , it has 3 points; and for  $u = 10$ , it has 4 points.

Because no closed form expression can be found for the locally optimal designs it is

$u = 0$		$u = 5$		$u = 10$	
dose	weight	dose	weight	dose	weight
-1.54	0.525	-5.63	0.306	-11.2	0.251
0.73	0.475	-2.71	0.392	-8.75	0.251
		0.61	0.301	-1.24	0.249
				1.19	0.250

Table 1: Locally D-optimal designs for  $u = 0, 5$ , and  $10$ .

helpful to have some approximately optimal designs that can be expressed in closed form. One such class of approximately optimal designs can be heuristically motivated: the probability plot for  $u = 10$  in Figure 2 looks like two separate, single logistic regressions in different regions of dose. One logistic regression is on the left: for no reaction with probability  $p_1(x, \theta)$ , and adverse reaction has a negligible probability of  $p_3(x, \theta)$ . The other is on the right: for adverse reaction with probability  $p_3(x, \theta)$  and no reaction has a negligible probability of  $p_1(x, \theta)$ . The value of  $x$  at which the probability of  $p_1(x, \theta)$  and  $p_3(x, \theta)$  is  $\frac{1}{2}$  (the LD50's) is  $-u$  for  $p_1(x, \theta)$  and  $0$  for  $p_3(x, \theta)$ . From results for a single logistic regression therefore (White, 1975, p. 43), for a large  $u$ , a design putting equal weights at  $\pm a, -u \pm a$  where  $a > 0$  might be a good initial guess of the locally D-optimal design. Such a strategy is approximately locally D-optimal as will be shown in Section 4.

Bayesian D-optimal designs for simple prior distributions are also found. Consider a class of simple prior distributions  $\pi_u$  with two equally weighted support points:  $\theta = (0, 0, 1)$  and  $\theta = (u, 0, 1)$  when  $u > 0$ . Bayesian D-optimal designs typically have 3 design points: for  $u = 5$  the design puts weight 0.170, 0.442, 0.388 at  $x = -5.24, -1.84, 0.68$ , respectively, and for  $u = 10$  the design puts weight 0.168, 0.448, 0.384 at  $x = -10, -1.49, 1.12$ , respectively. An approximately Bayesian D-optimal design for this simple prior distribution was suggested by numerical exploration and will be shown in Section 4.

The criterion of c-optimality minimizes the average asymptotic variance of the MLE of a function of interest. When  $\pi$  is the prior distribution of  $\theta$  and  $c(\theta)$  is the gradient

vector of the function of interest, c-optimality minimizes  $\phi(\pi, \eta) = \int c(\theta)^T M(\theta, \eta)^{-1} c(\theta) d\pi(\theta)$ . Suppose that the goal of the experiment is to find the target dose,  $x_{max}$ , where  $p_2(\theta, x)$ , the probability of efficacy, is maximized;  $x_{max} = -(a_1 + a_2)/2b_1$ . Then  $c(\theta)$  is the gradient vector of  $x_{max}$ ,  $c(\theta) = \nabla x_{max}$ . The asymptotic variance of the MLE of  $x_{max}$ , for a design  $\eta$ , with a sample size  $n$ , is then  $n^{-1}c(\theta)^T M(\theta, \eta)^{-1}c(\theta)$ . See Silvey (1980) for an extension of c-optimality to singular information matrices.

Similarly to local D-optimality, define two designs:  $\eta_0 = (m_1, m_2, \dots, m_k)$  at  $(x_1, x_2, \dots, x_k)$  and  $\eta_1 = (m_1, m_2, \dots, m_k)$  at  $(\frac{x_1 - a_1}{b_1}, \frac{x_2 - a_1}{b_1}, \dots, \frac{x_k - a_1}{b_1})$ , and two sets of parameters:  $\theta_0 = (u, 0, 1)$  and  $\theta_1 = (u, a_1, b_1)$ . It can be shown that  $c(\theta_1)^T M(\theta_1, \eta_1)^{-1}c(\theta_1) = c(\theta_0)^T M(\theta_0, \eta_0)^{-1}c(\theta_0)$ . Hence if the design  $\eta_0$  is locally c-optimal for  $\theta_0$  the design  $\eta_1$  is also locally c-optimal for  $\theta_1$ , providing all design points are in  $\mathcal{X}$ . Only c-optimal designs for  $a_1 = 0, b_1 = 1$  will therefore be explored here.

Locally c-optimal designs for estimating  $x_{max}$  found numerically are typically two-point designs. It will be shown in Section 4 that the design putting weight 1/2 at each of  $x = 0$  and  $x = -u$  is approximately c-optimal for estimating  $x_{max}$ . The mid-point of the two design points is the dose  $x_{max}$ . In practice, the value of  $u$  is unknown but can be estimated from a design with two or more design points. When  $u$  is 0, the c-optimal design becomes a singular one-point design at  $x_{max}$ . Singular designs are of limited direct practical use as the parameters are not all estimable. They are however useful in sequential strategies and as benchmarks. Note that even though the model has 3 parameters, a 2-point design leads to an information matrix of full rank as the response is multivariate.

For a more general model where  $b_1$  is not necessarily equal to  $b_2$ , locally D-optimal and Bayesian D-optimal designs have been described in Fan and Chaloner (2001). Similarly to the model with  $b_1 = b_2$  locally D-optimal designs have no closed form expression and also have 2, 3, or 4 design points. The criterion for c-optimality for estimating  $x_{max}$  is however not straightforward to implement in this more general model as there is no closed form expression for  $x_{max}$ . The probability  $p_2(\theta, x)$  is maximized at dose  $x$  if, and

only if

$$g(x, \theta) = b_2(1 + e^{-a_1 - b_1 x}) - b_1(1 + e^{a_2 + b_2 x}) = 0. \quad (3)$$

The solution exists and is unique because  $b_2(1 + e^{-a_1 - b_1 x})$  is strictly decreasing and  $b_1(1 + e^{a_2 + b_2 x})$  is strictly increasing. The solution  $x_{max}$ , however, cannot be expressed explicitly as in the constant slopes case. Thus the vector  $c(\theta) = \nabla x_{max}$  cannot be calculated directly.

Atkinson and Haines (1996) provide a solution which can be applied to any function defined implicitly as a solution to an equation. Suppose the function of interest,  $x(\theta)$ , is a solution of  $g(x, \theta) = 0$ . If the function  $g(x, \theta)$  has continuous first derivatives, and  $x(\theta)$  is continuous, then by the implicit function theorem,

$$\left. \frac{\partial g}{\partial \theta} \right|_{(x(\theta), \theta)} = \left. \frac{\partial g}{\partial x} \right|_{(x(\theta), \theta)} \left. \frac{\partial x}{\partial \theta} \right|_{\theta},$$

and so,

$$\nabla x(\theta) = \left. \frac{\partial x}{\partial \theta} \right|_{\theta} = \left[ \left. \frac{\partial g}{\partial x} \right|_{(x(\theta), \theta)} \right]^{-1} \left. \frac{\partial g}{\partial \theta} \right|_{(x(\theta), \theta)}.$$

The calculation of this expression of  $\nabla x(\theta)$  does not require an explicit  $x(\theta)$ . The implicit function theorem is also used to study E-optimal designs for polynomial regression on implicit segments (Melas, 1998).

For finding the c-optimal designs for estimating  $x_{max}$  let  $\theta = (a_1, b_1, a_2, b_2)$ ,  $x(\theta) = x_{max}$ , and  $g(x, \theta)$  be as in (3). Then the vector  $c(\theta)$  is:

$$\nabla x_{max} = \begin{pmatrix} e^{-a_1 - b_1 x_{max}} / [b_1 (e^{-a_1 - b_1 x_{max}} + e^{a_2 + b_2 x_{max}})] \\ x_{max} e^{-a_1 - b_1 x_{max}} / [b_1 (e^{-a_1 - b_1 x_{max}} + e^{a_2 + b_2 x_{max}})] \\ e^{a_2 + b_2 x_{max}} / [b_2 (e^{-a_1 - b_1 x_{max}} + e^{a_2 + b_2 x_{max}})] \\ x_{max} e^{a_2 + b_2 x_{max}} / [b_2 (e^{-a_1 - b_1 x_{max}} + e^{a_2 + b_2 x_{max}})] \end{pmatrix}.$$

To find the (locally) c-optimal design, the value of  $x_{max}$  can first be found numerically using the prior guess of the parameters. For example, if  $(-3.3, 0.5, 3.4, 1)$  is the prior

Parameter ( $a_1, b_1, a_2, b_2$ )	c-optimal design		D-optimal design		
	dose	weight	dose	weight	efficiency
(-3.3, 0.5, 3.4, 1)	-5.67	0.0012	-4.63	0.2922	55.95%
	-0.64	0.8003	-1.32	0.4164	
	4.84	0.1985	4.19	0.0557	
			8.64	0.2357	
(-2.76, 0.8, 2.8, 1)	-1.07	0.6365	-3.92	0.2753	56.65%
	2.24	0.3635	-0.38	0.4744	
			4.69	0.2503	
(-1.6, 0.2, 2, 1)	0.12	0.7792	-3.56	0.3676	59.90%
	13.62	0.2208	-0.49	0.3864	
			14.92	0.2460	
(-1, 0.5, 2, 1)	-1.26	0.6318	-3.54	0.3662	67.20%
	4.11	0.3682	-0.59	0.4030	
			4.80	0.2308	
(-1.04, 0.8, 1.2, 1)	-1.30	0.5494	-2.67	0.3704	77.22%
	2.37	0.4506	-0.00	0.3980	
			2.88	0.2316	
(0.4, 0.2, 2, 1)	-14.00	0.1004	-13.00	0.0696	62.15%
	-1.14	0.6278	-4.11	0.3996	
	9.99	0.2718	-0.77	0.3717	
			9.08	0.1591	

Table 2: Optimal designs for different values of parameters.

guess of  $\theta$ , then the corresponding prior guess of  $x_{max}$  is 0.4104, found numerically.

Table 2 gives the c-optimal designs and, for comparison, corresponding D-optimal designs. The c-optimal designs have two or three design points while their corresponding D-optimal designs have an additional design point. The last column in Table 2 gives the efficiencies of D-optimal designs under the c-optimality criterion which are mostly about or below 60%. This indicates these D-optimal designs are not efficient for estimating  $x_{max}$ .

## 4 Limiting Optimal Designs

In finding optimal designs, numerical problems often arise in optimization, especially if we have no idea of a starting design. Often the Equivalence Theorem can help, but not

always. For example, the locally c-optimal designs for the model with constant slopes for a very small value of  $u$  are very hard to find using algorithms based on the Equivalence Theorem alone. These c-optimal designs can be found easily using the approximately optimal designs in Section 3 as starting designs. In this section, approximately optimal designs are given a formal definition and called “limiting optimal designs”. They can provide insight on the optimal design and serve as good starting designs for optimization algorithms.

**Definition.** For a concave criterion  $\phi$  on a set of design measures  $\mathcal{H}$ , a sequence of designs,  $\{\eta_i, i \in \mathbb{R}\}$ , is called a sequence of limiting  $\phi$ -optimal designs for a sequence of prior distributions,  $\{\pi_i, i \in \mathbb{R}\}$ , if

$$\sup_{\eta \in \mathcal{H}} F_\phi(\eta_i, \eta, \pi_i) \rightarrow 0 \text{ as } i \rightarrow \infty. \quad (4)$$

The design  $\eta_i$  is called a limiting  $\phi$ -optimal design for  $\pi_i$ .

**Theorem 4.1.** *Consider a concave criterion  $\phi$ , a sequence of prior distributions,  $\pi_i$ 's, and the corresponding limiting optimal designs,  $\eta_i$ 's. Let  $\eta_i^*$  be the  $\phi$ -optimal design for prior  $\pi_i$ . Then  $\phi(\pi_i, \eta_i^*) - \phi(\pi_i, \eta_i)$ , the difference in the value of the criterion at design  $\eta_i^*$  and at design  $\eta_i$ , goes down to zero as  $i \rightarrow \infty$ .*

*Proof.* By concavity of  $\phi$ , it is easy to show that  $F_\phi(\eta_i, \eta_i^*, \pi_i) \geq \phi(\pi_i, \eta_i^*) - \phi(\pi_i, \eta_i)$ . Therefore,  $\phi(\pi_i, \eta_i^*) - \phi(\pi_i, \eta_i) \leq F_\phi(\eta_i, \eta_i^*, \pi_i) \leq \sup_{\eta} F_\phi(\eta_i, \eta, \pi_i)$ . Since  $\sup_{\eta} F_\phi(\eta_i, \eta, \pi_i) \rightarrow 0$  as  $i \rightarrow \infty$ ,  $\phi(\pi_i, \eta_i^*) - \phi(\pi_i, \eta_i)$  goes down to zero as  $i \rightarrow \infty$ .  $\square$

If a criterion  $\phi$  is concave and differentiable at  $\eta_i$  for each  $i$ , then (4) can be replaced by:

$$\sup_{x \in \mathcal{X}} F_\phi(\eta_i, \eta_x, \pi_i) \rightarrow 0 \text{ as } i \rightarrow \infty, \quad (5)$$

where  $\eta_x$  is point mass at  $x$ .

Note that there is no restriction on  $\{\pi_i, i \in \mathbb{R}\}$ , the sequence of prior distributions, in the definition and no concept of convergence. This sequence  $\{\pi_i, i \in \mathbb{R}\}$  may be

suggested after calculating optimal designs or examining the probability curves as in the examples of Section 3. It is often easier to find a general algebraic expression for limiting optimal designs than for optimal designs.

Intuition can often be used to generate candidate optimal designs and the limiting optimality concept formalizes a concept of approximate optimality that can lead to a general strategy. For example, in the continuation-ratio model with constant slopes, the probability curves (Figure 2) look like two single logistic regressions when  $u$  is large. In addition, the optimal design for a single logistic regression is already known. Thus  $u$  is selected to be the index  $i$ .

**Example 1.** Let the prior distribution  $\pi_i$  be point mass at  $\theta_i = (u = i, 0, 1)$ . The design  $\eta^*$  putting equal weight at  $x = 1.223, -1.223, -i + 1.223, -i - 1.223$  is a locally limiting D-optimal design for  $\pi_i$ .

*Proof.* The best choice of  $a$  among designs putting 0.25 at each of  $\pm a, -u \pm a$  is the value of  $x$  maximizing  $x^2 e^{3x} / (4(1 + e^x)^6)$ , found numerically as  $a = 1.223$ . Let  $I(\theta_i, x)$  and  $M(\theta_i, \eta^*)$  be as in Appendix A. It can be shown that the directional derivative,  $\sup_{x \in \mathcal{X}} \text{Tr}[I(\theta_i, x)M^{-1}(\theta_i, \eta^*)] - 3$ , goes to 0 as  $i \rightarrow \infty$ . For details, see Fan (1999, Chapter 5). □

The limits of probabilities of the responses at the four design points as  $i \rightarrow \infty$ , are given in Table 3. When the value of  $u (= i)$  is large, the probabilities of adverse reaction at  $x = -u \pm 1.223$  are almost 0 and the probabilities of no reaction at  $x = \pm 1.223$  are also almost 0. The values 0.227 and 0.773 are not quite the same as the probabilities of response at the two design points of a locally D-optimal design in a simple logistic regression (White, 1975, p. 43) which are 0.176 and 0.824.

Note that in Example 1, and elsewhere,  $\mathcal{X}$  is assumed to be large enough to contain the design points. For example, as the index  $u$  gets large the design space  $\mathcal{X}$  could be the closed interval  $[-2u, 2u]$ .

For a simple prior distribution putting uncertainty only on  $u$  with probability 1/2

$x$	$p_1(\theta, x)$	$p_2(\theta, x)$	$p_3(\theta, x)$
1.223	0	0.227	0.773
-1.223	0	0.773	0.227
$-i + 1.223$	0.227	0.773	0
$-i - 1.223$	0.773	0.227	0

Table 3: The limits of probabilities at design points of the limiting design, as  $i \rightarrow \infty$ .

on 0 and some other value (for fixed  $a_1$  and  $b_1$ ), limiting Bayesian D-optimal designs are found in Example 2.

**Example 2.** Let the prior distribution  $\pi_i$  put equal weight at  $\theta = (0, 0, 1)$  and  $\theta = (i, 0, 1)$ . The design  $\eta$  putting weight 0.167, 0.448, 0.385 at  $x = -i, -1.47, 1.14$ , respectively, is a limiting Bayesian D-optimal design for  $\pi_i$ .

See Fan (1999, Chapter 5) for a proof similar to that of Example 1. For unequally weighted prior distributions, Fan (1999) gives some examples and conjectures for the limiting Bayesian optimal design.

For estimating the most efficacious dose  $x_{max}$  defined in Section 3 a limiting locally c-optimal design is found.

**Example 3.** Let the prior distribution  $\pi_i$  be point mass at  $\theta_i = (u = i, 0, 1)$ . The design  $\eta$  putting weight 1/2 at each of  $x = 0$  and  $x = -i$  is a limiting locally c-optimal design for estimating  $x_{max}$  for  $\pi_i$ .

The proof is in Fan (1999, Chapter 5). Because these limiting designs are not optimal, it is important to examine their efficiencies. For large  $i$  values Theorem 4.1 shows that these limiting designs have extremely high efficiency compared to the optimal designs and the efficiencies of a few values larger than 10 were observed to be almost 100%, we therefore examined the efficiencies of the limiting designs for  $i \leq 10$ . The limiting D-optimal designs in Examples 1 and 2 are surprisingly efficient, as shown in Figure 3: all are higher than 97%.

The limiting locally c-optimal designs in Example 3 are not as efficient as the limiting

D-optimal designs, but they are quite efficient for moderate  $u$  values. As shown in Figure 3, for  $u$  between 2.5 and 4 these limiting designs are over 75% efficient, and for  $u$  greater than 4 over 95% efficient. For  $0 < u < 2.5$  numerical problems occurred in finding c-optimal designs and thus the efficiencies of the limiting c-optimal designs were not calculated.

Fan and Chaloner (2001) find local and Bayesian D-optimal designs for the general non-constant slopes continuation-ratio model. An example of a sequence of designs with limiting efficiency one is given but these are not, however, very efficient, even for large  $u$ . In some cases, the efficiency can be as little as 60% even for  $u = 10$ .

## 5 Discussion

This paper has presented methods and designs for a trinomial response model. In addition, a new tool for studying designs is provided: the concept of a sequence of limiting optimal designs. A member of a sequence of limiting optimal designs can serve as an initial design for finding the optimal design and a sequence of limiting optimal designs can provide general insight into the design problem. Limiting optimal designs can also provide ideas for designing experiments in practice. For example in a phase I/II trial to find  $x_{max}$ , using the limiting c-optimal designs to allocate a cohort of patients to a pair of doses may be more efficient than the conventional approach of allocating a cohort to a single dose. The limiting optimal design also provides a way to select the pair of doses, one at the right and the other at the left of the estimate of the dose  $x_{max}$  (the dose with the highest probability of efficacy). There is more work in this area to be done: see Kpamegan and Flournoy (2001).

In a phase I/II trial, the sample size is usually quite small and thus fewer parameters can be used. Design for the constant slopes model has been studied in this paper extensively and the model has only three parameters. Fan (1999), therefore, also explored the robustness of the constant slopes assumption in the following three ways:

1. examining the efficiency of the optimal design for the constant slopes model when the slopes are actually not equal.
2. calculating the values to which the MLE's of the constant slopes model converge, and by how much the fitted probability curves differ from the true probability curves when the slopes are assumed equal but are not (using results from Huber, 1967).
3. studying the performance of the optimal design for the constant slopes model for a small sample of  $n = 20$  when the slopes are not equal when the goal is to estimate  $x_{max}$ . Simulations of data from both D-optimal and c-optimal designs were used and the distributions of  $\hat{x}_{max} - x_{max}$  were compared.

The results (Fan, 1999) have generally confirmed that the constant slopes model performs reasonably well if the slopes are not equal but close. These c-optimal designs also perform reasonably well for a small sample size of 20. The results also indicate that, with  $n = 20$ , for estimating  $x_{max}$ , extreme outlying estimates sometimes occur. In general, the c-optimal designs, designed to estimate  $x_{max}$ , are indeed better than the corresponding D-optimal designs in estimating  $x_{max}$  if the slopes  $b_1$  and  $b_2$  are close. Moreover, a useful guideline for how close is "close" is found: if  $b_1$  is within 10% of  $b_2$ , this is close enough. Hence the results in the paper for the constant slopes model can be used safely when the slopes are within 10% difference.

The same approach can be applied to other models and problems. Exact optimal designs for complicated models are usually difficult to find, even numerically. The concept of a sequence of limiting optimal design can be found in algebraic form and reduce the computational complexity as well as provide insight into the practical design problem.

## Acknowledgments

This research was supported in part by a grant from the National Security Agency. We are grateful to the editor, associate editor and referees for very helpful suggestions.

## Appendix A: Information Matrix

For a design  $\eta$  putting mass  $m_i$  at  $x_i$ ,  $i = 1, 2, \dots, k$ , and  $\sum m_i = 1$  the Fisher information matrix is  $M(\theta, \eta) = \sum_i m_i I(\theta, x_i)$  with  $I(\theta, x)$  given below.

1. For the constant slopes model ( $b_1 = b_2$ ):

$$I(\theta, x) = \left\{ p_3(\theta, x)(p_1(\theta, x) + p_2(\theta, x)) \begin{bmatrix} 1 & x & 0 \\ x & x^2 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \frac{p_1(\theta, x)p_2(\theta, x)}{p_1(\theta, x) + p_2(\theta, x)} \begin{bmatrix} 0 & 0 & 0 \\ 0 & x^2 & x \\ 0 & x & 1 \end{bmatrix} \right\}$$

where  $\theta$  is  $(a_1, b_1, a_2)$ .

2. For the different-slope model ( $b_1 \neq b_2$ ):

$$I(\theta, x) = \frac{e^{a_2+b_2x}}{(1 + e^{a_2+b_2x})^2(1 + e^{a_1+b_1x})} \begin{bmatrix} 1 & x & 0 & 0 \\ x & x^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \frac{e^{a_1+b_1x}}{(1 + e^{a_1+b_1x})^2} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & x \\ 0 & 0 & x & x^2 \end{bmatrix}$$

where  $\theta = (a_2, b_2, a_1, b_1)$ .

Note that if  $k \geq 2$  then  $M(\theta, \eta)$  is non-singular.

## REFERENCES

- Agresti, A. (1990). *Categorical Data Analysis*. John Wiley & Sons Inc., New York.
- Atkinson, A.C. and Haines, L.M. (1996). Designs for nonlinear and generalized linear models. *Handbook of Statistics*, 13, 437-475.
- Chaloner, K. and Larntz, K. (1989). Optimal Bayesian 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.
- Fan, S.K. (1999). *Multivariate Optimal Designs*. PhD Thesis, School of Statistics, University of Minnesota.
- Fan, S.K. and Chaloner, K. (2001). Optimal Design for a Continuation-ratio Model. In *mODa6-Advances in Model-Oriented Design and Analysis*, eds. Atkinson, A.C., Hackl, P., and Muller, W.G. Physica-Verlag. 77-86.
- Ford, I., Torsney B. and Wu C.F.J. (1992). The use of a canonical form in the construction of locally optimal designs for non-linear problems. *J. Roy. Stat. Soc. Ser. B*, 54, 569-583.
- Glonek, G.F.V. (1996). A class of regression models for multivariate categorical responses. *Biometrika*, 83, 15-28.
- Glonek, G.F.V. and McCullagh, P. (1995). Multivariate logistic models. *J. Roy. Stat. Soc. Ser. B*, 57, 533-546.
- Heise, M.A. and Myers, R.H. (1996). Optimal designs for bivariate logistic regression. *Biometrics*, 52, 613-624.
- Huber, P.J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium*. Berkeley: University of California Press, 221-233.
- Kitsos, C.P., Titterington, D.M. and Torsney, B. (1988). An optimal design problem in rhythmometry. *Biometrics*, 44, 657-671.
- Kpamegan, E.E. and Flournoy, N. (2001). An optimizing up-and-down design. *Optimum Design 2000*, eds. Atkinson, A.C., Bogacka, B., and Zhigljavsky, A. Kluwer Academic Publishers. 437-475.
- Melas, V.B. (1998). Analytical theory of E-optimal designs for polynomial regression on a segment. *mODa5-Advances in Model-Oriented Data Analysis and Experimental Design*, eds. Atkinson, A.C., Pronzato, L., and Wynn, H.P. Physica-Verlag. 51-58.
- Pukelsheim, F. (1993). *Optimal Design of Experiments*. John Wiley & Sons Inc, New York.

- Silvey, S.D. (1980). *Optimal Design*. Chapman and Hall.
- Thall, P.F. and Russell, K.E. (1998). A strategy for dose-finding and safety monitoring based on efficacy and adverse outcomes in phase I/II clinical trials. *Biometrics*, 54, 251-264.
- White, L.V. (1975). *The Optimal Design of Experiments for Estimation in Nonlinear Models*. PhD Thesis, University of London.
- Whittle, P. (1973). Some general points in the theory of optimal experimental design. *J. Roy. Stat. Soc. Ser. B*, 35, No. 1, 123-130.
- Wu, C.F.J. (1988). Optimal design for percentile estimation of a quantal response curve. In *Optimal Design and Analysis of Experiments*, eds. Y. Dodge, V.V. Fedorov, H.P. Wynn. North-Holland, Amsterdam, 213-223.
- Zhu, Y., Krewski, D., and Ross, W.H. (1994). Dose-response models for correlated multinomial data from developmental toxicity studies. *Applied Statistics*, 43, 583-598.
- Zocchi, S.S. and Atkinson, A.C. (1999). Optimum experimental designs for multinomial logistic models. *Biometrics*, 55, 437-444.

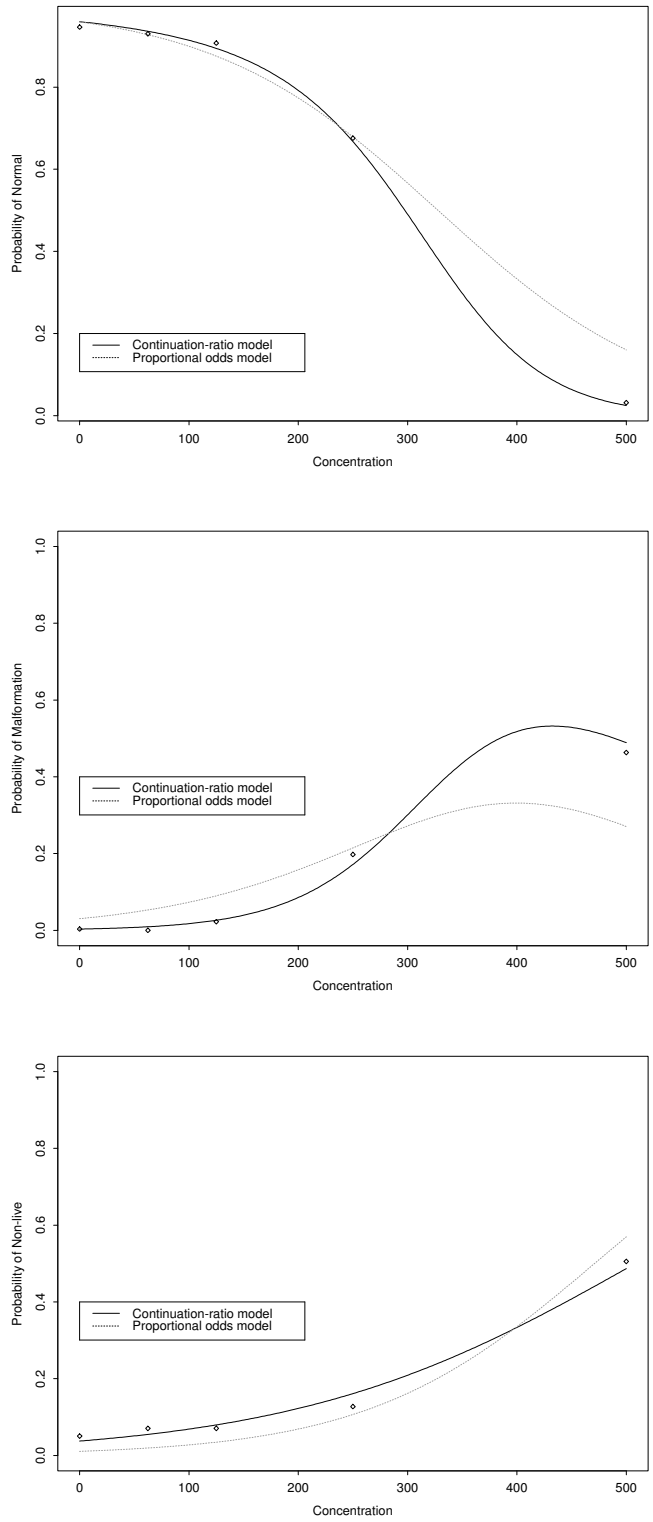


Figure 1: Fitted probability curves for the three responses: solid lines for the continuation-ratio model and dotted lines for the proportional odds model.

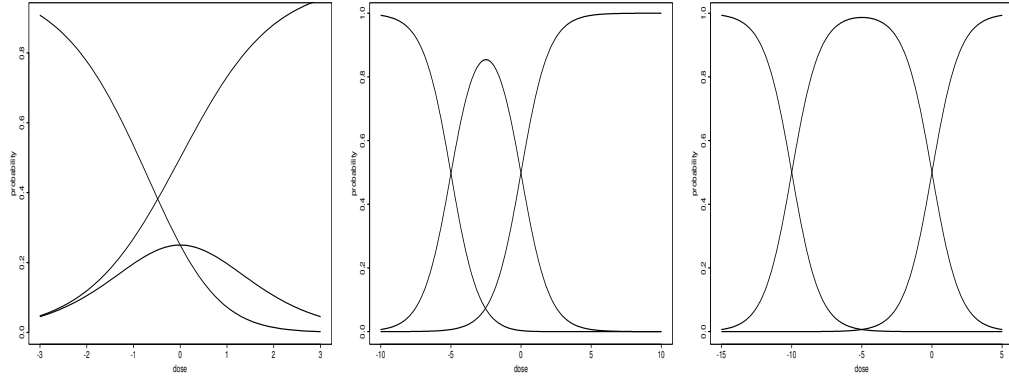


Figure 2: From the left to the right: probability plot, probability vs. dose, for  $a_1 = 0$ ,  $b_1 = 1$ , and  $u = 0, 5$ , and  $10$ , respectively.

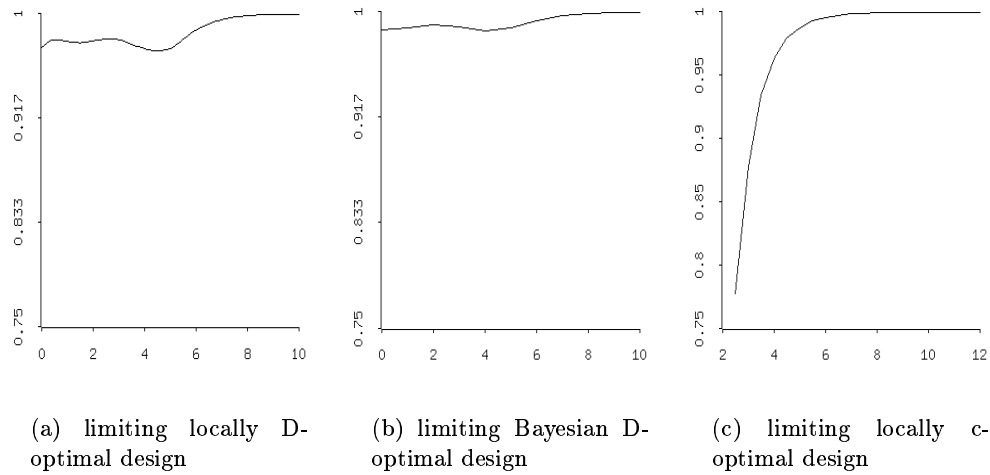


Figure 3: Efficiency plots of limiting optimal designs: efficiency vs.  $u$ ,  $b_1 = b_2$ .