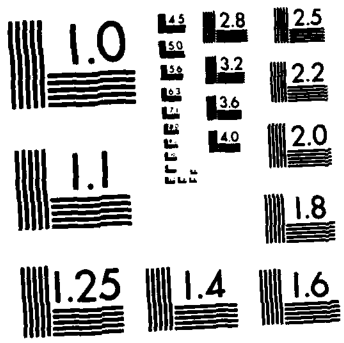
AD-A134 573 EFFICIENT MODEL-BASED SEQUENTIAL DESIGNS FOR SENSITIVITY EXPERIMENTS(U) WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER C F WU SEP 83 MRC-TSR-2563 1/1

UNCLASSIFIED DAAG29-80-C-0041 F/G 12/1 NL


1.0
4.5
5.0
5.6
6.3
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.1
1.8
1.25
1.4
1.6

AD-A134573

MRC Technical Summary Report #2563

EFFICIENT MODEL-BASED SEQUENTIAL DESIGNS FOR SENSITIVITY EXPERIMENTS

C. F. Jeff Wu

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53705

September 1983

(Received July 25, 1983)

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

EFFICIENT MODEL-BASED SEQUENTIAL DESIGNS FOR SENSITIVITY EXPERIMENTS

C. F. Jeff Wu*

Technical Summary Report #2563

September 1983

ABSTRACT

A sequential design for estimating the percentiles of a quantal response curve is proposed. Its updating rule is based on an efficient summary of all the data available via a parametric model. Its efficiency in terms of saving the number of runs and its robustness against the distributional assumption are demonstrated heuristically and in a simulation study. A linear approximation to the "logit-MLE" version of the proposed sequential design is shown to be equivalent to an asymptotically optimal stochastic approximation method, thereby providing a large sample justification. For sample size between 12 and 35, the simulation study shows that the "logit-MLE" version of the general sequential procedure substantially outperforms an adaptive (and asymptotically optimal) version of the Robbins-Monro method, which in turn outperforms the nonadaptive Robbins-Monro and Up-and-Down methods. A nonparametric sequential design, via the Spearman-Karber estimator, for estimating the median is also proposed.

AMS (MOS) Subject Classifications: 62K05, 62L05

Key Words: Logit, Optimal design, Quantal response curve, Robbins-Monro stochastic approximation, Sensitivity experiments, Spearman-Karber estimator, Up-and-Down method

Work Unit Number 4 - Statistics and Probability

*Work completed while visting the MSRI, Berkeley.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041 and ARO Grant No. DAAG29-82-K0154.

SIGNIFICANCE AND EXPLANATION

In many physical or biological experiments with binary response a quantal response curve is assumed to relate the probability of response to the corresponding level of the stimulus variable. To estimate the percentiles of the quantal response curve efficiently, a sequential design is often used in practice. We propose a new class of sequential designs with updating rules based on an efficient summary of all data available via a parametric model. This method is shown to be asymptotically as good as the optimal stochastic approximation method. More importantly, its finite sample performance in a simulation study is better than the latter method. Specifically, the percentage of runs saved by using our method ranges from 25% to 60%.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the author of this report.

EFFICIENT MODEL-BASED SEQUENTIAL DESIGNS FOR SENSITIVITY EXPERIMENTS

C. F. Jeff Wu*

1. Introduction

A sensitivity experiment is characterized by a response curve that relates the stimulus level applied to an experimental subject to the probability of response. The outcome of the experiment is assumed dichotomous, response or nonresponse. This situation arises in many fields of research. In testing the strength of materials, the stimulus level may be the level of impact energy applied to a piece of material, and the response is either "fail" or "not fail" (Wetherill, 1963). In testing explosives, the stimulus level may be the height from which a weight is dropped or the pressure directly applied to the explosive, and the response is "explode" or "not explode" (Dixon and Mood, 1948). In biological assays a test animal survives or not at a given dose level (Finney, 1978). In psycho-physical research the probability of detecting a stimulus is related to its intensity level (Rose et al., 1970). In educational testing, one may want to study the "item characteristic curve" that relates the difficulty level of the test item to the probability of "right" or "wrong" answer (Lord, 1971).

Our main interest is in estimating the percentiles of the response curve $F(x)$, which is the probability of response for a given stimulus level x. The $100p$ percentile L_p is defined as

$$F(L_p) = p. \tag{1}$$

For simplicity we assume F is monotone increasing and continuous. The median of F, $L_{0.5}$, is the most commonly used measure of a characteristic of the response curve. In some situations estimating $L_{0.5}$ is of intrinsic interest,

*Work completed while visiting the MSRI, Berkeley.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041 and ARO Grant No. DAAG29-82-K0154.

but more often it is because $L_{0.5}$ is easy to estimate. In quality assurance it may be more interesting to study the extreme percentiles, e.g., to find the impact energy level that results in the failure of material for at most 10% of the time. On the other hand $L_{0.9}$ may be more relevant in explosive research.

In this paper we will present some new sequential designs for the efficient estimation of L_p for small or moderate sized experiments. As will be explained later, our method is more appropriate for $0.2 \leq p \leq 0.8$. We consider sequential designs in such a way that all the information in the previous experimentation can be utilized in a most efficient manner for suggesting how the next experiment should be performed. When the experimental runs are very expensive, the saving of a few runs by an efficient design outweighs the extra pains taken in designing a sequential experiment. The sequential nature of the design requires quick responses so that the experiment will not be unduly prolonged. It is suitable, for example, when the experimental facility is limited so that experimental runs must be performed one after another. Many biological experiments that involve inexpensive animals and slow responses have to be ruled out. A key element of our sequential scheme is the efficient summary of all available information for suggesting the next design. This requires a certain degree of computing. As computing becomes cheaper and more personalized, the cost of automating an experimental design will be less. By taking all the factors into account, our method is more appropriate for expensive experiments with short response time, which are more often encountered in engineering research. In educational or psychological testings, if a test has to be repeated routinely on many subjects, it pays off to automate the design and to look for the most efficient ones (in terms of reducing the number of test items).

In the next section we shall review two nonparametric sequential designs and point out their inappropriateness for the scenarios described above. Our approach is to assume a parametric model for the response curve and estimate efficiently the relevant parameters in the model based on all the data available. An estimated quantal response curve (EQRC) is constructed through the current estimate of the parameters and the next design point is determined from the EQRC. Some heuristic and theoretical justifications are provided for the methodology. In particular, a linear approximation to the "logit-MLE" version of the EQRC approach is shown to be equivalent to an asymptotically optimal stochastic approximation method. The Monte Carlo results of the last section indicate that the EQRC approach is not sensitive to distributional assumptions for estimating $L_{0.5}$ and $L_{0.75}$ and substantially outperforms the Robbins-Monro stochastic approximation method and the Up-and-Down method, including an adaptive (and asymptotically optimal) Robbins-Monro method. For the particular simulation experiment, our method results in saving 25% to 60% of the total number of runs required by its nearest competitors. The empirical study also reveals that a mild degree of truncation is needed for both our method and the adaptive Robbins-Monro method to perform stably. A good guess of the initial design and the step size is more critical to the performance of the Robbins-Monro and Up-and-Down methods than the parametric assumption is to our method. For details see Section 6. A nonparametric sequential design for estimating the median $L_{0.5}$ is proposed via the Spearman-Kärber estimator. Its limitations are discussed.

2. Review and criticism of the Stochastic Approximation method and the Up-and-Down method.

The Stochastic Approximation method and the Up-and-Down method are two most commonly used nonparametric sequential designs for quantal response problems.

Stochastic Approximation Method (Robbins and Monro, 1951):

Let $y_n = 1$ or 0 as the n^{th} experiment results in a response or nonresponse. For estimting L_p, the stimulus level x_{n+1} of the $(n+1)^{th}$ run is chosen according to

$$x_{n+1} = x_n - \frac{c}{n}(y_n - p). \tag{2}$$

According to the results of Chung (1954), Hodges and Lehmann (1955) and Sacks (1958), the optimal choice of c in (1) is $(F'(L_p))^{-1}$. Procedure (2) with this choice of c is asymptotically consistent and fully efficient, i.e., $x_n \rightarrow L_p$ a.s. and its asymptotic variance is minimized. The small sample behavior of (2) depends very much on a good starting value x_1 (Wetherill, 1963). Ideally x_1 should be close to L_p. A good guess of the optimal constant c may also be hard to come by since in most practical situations the experimenters have little idea about the slope of F at L_p. Poor choice of c and x_1 will make (2) an inefficient procedure for small and even moderate samples. The Stochastic Approximation method has been used more effectively in on-line estimation wherein a large number of data have to be processed quickly.

To achieve minimal asymptotic variance, it is necessary to estimate the slope $F'(L_p)$. One such estimator is the regression slope of y_i over x_i,

(3) $$\hat{\beta}_n = \frac{\sum^n y_i(x_i-\bar{x}_n)}{\sum^n (x_i-\bar{x}_n)^2}, \quad \bar{x}_n = \frac{1}{n}\sum_1^n x_i.$$

The procedure (2) with $c = \hat{\beta}_n^{-1}$ is aptly called a stochastic Newton-Raphson method by Anbar (1978), since it can be viewed as a method of solving the equation $F(x) = p$ by the tangential approximation to F at x_n with $F(x_n)$ replaced by y_n and $F'(x_n)$ by $\hat{\beta}_n$. Under various regularity conditions, Anbar (1978) and Lai and Robbins (1981) proved that $\hat{\beta}_n \rightarrow F'(L_p)$ a.s. and that the procedure (2a) is asymptotically optimal,

(2a) $$x_{n+1} = x_n - \frac{1}{n\hat{\beta}_n}(y_n - p), \ \hat{\beta}_n \text{ in (3)}.$$

When n is small or the current guess x_n is on the tails of the response curve, $\hat{\beta}_n^{-1}$ may behave erratically. Since the tails of the response curve are flat, $\hat{\beta}_n$ with $\{x_i\}_1^n$ located on either tail tends to be closer to zero, thus making the adjustment from x_n to x_{n+1} in (2a) unreasonably large. This happens when the initial guess is poor for estimating the median or when the initial design takes too few points from the middle part of the response curve for estimating the extreme percentiles. To remedy this, we propose to truncate $\hat{\beta}_n^{-1}$, that is, to use $\max(\min(\hat{\beta}_n^{-1},c),-c)$ instead of $\hat{\beta}_n^{-1}$ for some positive constant c. The simulation study of Section 6 shows that there is considerable improvement in using this truncated version of (2a). Anbar (1978) and Lai and Robbins (1981) considered truncating $\hat{\beta}_n$ instead of $\hat{\beta}_n^{-1}$ mainly for technical reasons.

Up-and-Down method (Dixon and Mood, 1948):

(4) $$x_{n+1} = \begin{cases} x_n + \Delta \\ x_n - \Delta \end{cases} \text{if } y_n = \begin{cases} 0 \\ 1 \end{cases}$$

The method works only for $L_{0.5}$. It is very simple to implement but, for small or moderate samples, its performance depends very much on a good guess of x_1 and Δ. Unless the step size Δ is made adaptive, the large sample property of x_n can not be studied. Its empirical performance is usually not as good as the Stochastic Approximation method. See Wetherill (1963) and Section 6 of the present paper. Some modifications of the two methods can be found in Wetherill (1963, 1966).

Both methods are "Markovian" in that the choice of the next run depends sensibly on the outcome of the current one. Their simplicity was a positive factor when inexpensive computing was not accessible. Their main disadvantages are: (a) The updating rules (2) and (4) do not make use of all the data available in an efficient way, and thus making the choice of step size less flexible. (b) Their small sample behavior depends on a good choice of the relevant constants in (2) and (4), which in turn depends on the experimenter's knowledge of the unknown response curve F. For small or moderate sized experiments with expensive runs, inefficiency and lack of robustness can be quite serious. Large sample properties seem quite irrelevant in this context.

The sequential design method proposed in the following section is quite different from (2) and (4) with respect to the shortcomings (a) and (b).

3. A class of sequential designs based on the estimated quantal response curve.

Ideally we would like to have a good estimate $\hat{F}_n$ of the whole curve F, from which the next design point x_{n+1} is chosen to be its 100p percentile, i.e., $\hat{F}_n(x_{n+1}) = p$. A (smooth) nonparametric estimate $\hat{F}_n$ of F is not feasible since it requires a large number of observations for $\hat{F}_n$ to be a good estimate. A natural approach for small sample problems is to assume a parametric model

$$F(x) = F(x|\theta),\ F \text{ is continuous in } x,$$

$$\lim_{x \to -\infty} F(x|\theta) = 0,\ \lim_{x \to +\infty} F(x|\theta) = 1.$$

The general recipe of our sequential design procedure for estimating L_p is:

(i) find an efficient estimate $\hat{\theta}_n = \hat{\theta}((y_i, x_i)_1^n)$ of θ,

(5)

(ii) define the estimated quantal response curve (EQRC) $\hat{F}_n(x) = F(x|\hat{\theta}_n)$, and choose the next design x_{n+1} s.t. $\hat{F}_n(x_{n+1}) = p$.

Recall that $y_i = 1$ or 0 is the response or nonresponse at level x_i.

In general the choice of the parametric model $F(\cdot|\theta)$ should reflect the experimenter's knowledge of the problem, if there is any. Given this model, were there a reliable prior on θ, a Bayesian approach for estimating θ would be appropriate. In the absence of such information (a more typical situation in practice), we suggest to use the logit model

$$(6) \qquad F(x|\theta) = \frac{1}{1+e^{-\lambda(x-\alpha)}}, \ \lambda > 0, \ \theta = (\alpha,\lambda)$$

and the maximum likelihood estimator (MLE) $(\hat{\alpha},\hat{\lambda})$ of (α,λ). For (6), $L_p = \alpha-\frac{1}{\lambda}\ln(\frac{1}{p}-1)$ and its MLE $\hat{L}_p = \hat{\alpha} - \frac{1}{\hat{\lambda}} \ln(\frac{1}{p}-1)$. Note $\hat{L}_{0.5} = \hat{\alpha}$.

The main reason for preferring logit to its competitor, the probit model,

$$(7) \qquad G(x|\theta) = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{x} e^{-\frac{(z-\mu)^2}{2\sigma^2}}\,dz, \ \theta = (\mu,\sigma), \ \sigma > 0,$$

is computational ease. It is well known that the logit, the probit and other parametric models like the angular and the linear curves agree very closely in the range 0.2 to 0.8 (Cox, 1970, Table 2.1). We do not see any advantage in using the probit over the logit, although it is a legitimate choice. It is rarely the case that a parametric quantal response model be justifiable on biological or physical grounds. The successful use in practice of the parametric approach for quantal response problems is mainly due to this key fact that the parametric curves (after adjusted for location and scale) agree very closely in a wide range of p values. For p outside [0.1,0.9] the percentiles for different parametric models vary greatly. Therefore we can not recommend our procedure (5) for extreme p values unless there is a good reason to believe in the particular model. One may argue that the Stochastic Approximation method, being nonparametric, still works for the extreme tails. This is only so for <u>very large</u> samples. For instance, the method makes on the average nine negative moves for each positive move in the neighborhood of $L_{0.9}$. Instead of "straddling" $L_{0.9}$, the sequence makes far too many moves in one direction. This explains the much poorer empirical performance of the Stochastic Approximation method (2) even for moderately extreme tails like $L_{0.75}$ (Wetherill, 1963, §6).

The next issue is the choice of efficient estimator $\hat{\theta}_n$ in (5) (i). The minimum logit chi-square method (Berkson, 1955) is not suitable for the kind of data generated by a sequential procedure like (5), especially for small or moderate samples. This is because there are few, and typically only one or two, observations at a given x level to make the minimum logit chi-squ work. Unless we restrict the search of design levels to a small number o levels, the situation will not be much changed. The same remark applies the minimum modified chi-square method, and to a lesser extent, to the minimum chi-square method. The maximum likelihood estimate of (α,λ) in (6) is obtained by iteratively solving the equations

$$(8)\qquad \begin{aligned} \sum_{i=1}^{n} F(x_i|\alpha,\lambda) &= \sum_{1}^{n} y_i, \\ \sum_{1}^{n} x_i F(x_i|\alpha,\lambda) &= \sum_{1}^{n} y_i x_i, \end{aligned}$$

where $F(x|\alpha,\lambda) = (1 + e^{-\lambda(x-\alpha)})^{-1}$. Th MLE is a function of the sufficient statistics $(\sum y_i, \sum y_i x_i)$. It is asymptotically fully efficient given the right model and is an efficient summary of all the information available in small samples. Unless there is a reliable prior on θ so that a Bayesian approach (Freeman, 1970; Tsutakawa, 1972; Owen, 1975; Leonard, 1982) becomes effective, it might be hard to beat the MLE for small samples.

Given an efficient estimate $\hat{\theta}_n$, the sequential design (5) makes full use of all the information available and the step size $x_{n+1}-x_n$ is more flexible, i.e., it is capable of making large or small adjustment as the situation calls for. Its only "ad hockery" is in the logit assumption. As argued before, the assumption is quite robust for $0.2 \leq p \leq 0.8$. It will be further supported in the empirical study of §6.

For the implementation of procedure (5), it is important to know when the MLE exists. To avoid trivialities, assume there are at least two distinct x_i's. It is known (Silvapulle, 1981) that the MLE of the "linear" parameters $(\lambda,\lambda\alpha)$ in the logit model (6) exists uniquely iff

(9.1) $$(x^+_{min}, x^+_{max}) \cap (x^-_{min}, x^-_{max}) \text{ is non-empty}$$

or

(9.2) $$x^+_{min} < x^-_{min} = x^-_{max} < x^+_{max}$$

or

(9.3) $$x^-_{min} < x^+_{min} = x^+_{max} < x^-_{max}$$

where $x^+_{max(min)} = \max(\min)\ (x_i: y_i = 1)$, $x^-_{max(min)} = \max(\min)\ (x_i: y_i = 0)$. The same result holds for more general distributions F including the probit model (7). See Silvapulle (1981, Theorem (iii)). It is easy to see that (9), once satisfied, is always satisfied by the addition of more obversations. The change from x_n to x_{n+1} via the logit-ML⁻ method may be unduly large when the problem is "ill-posed." It happens when the data configuration $(x_i,y_i)_1^n$ is such that the first time the condition (9) is satisfied is n or n-k with very small k, that is, the existence and uniqueness of MLE has only been guaranteed in the current or last few runs. We propose a truncated version as follows. Define d_n as the solution of $x_{n+1} = x_n - \frac{d_n}{n}(y_n - p)$, where $x_{n+1} = \hat{\alpha}_n - \hat{\lambda}_n^{-1} \ln(p^{-1}-1)$ and $(\hat{\alpha}_n,\hat{\lambda}_n)$ is the solution of (8). The $(n+1)^{th}$ design level is chosen to be

(10) $$x_n - \frac{d_n^*}{n}(y_n - p),\ d_n^* = \max(-d, \min(d_n,d)),$$

where d is a given positive constant. The procedure (10) is shown in Section 6 to perform extremely well over a broad range of the truncation constant d. For very large d, like 600, which is equivalent to almost no truncation, (10) does not perform very stably.

The question of how to choose, say, the first ten runs is very difficult unless some prior knowledge is available. It may be done in an <u>ad hoc</u> manner aided with experience, or by the Stochastic Approximation procedure (2) with a reasonable guess of x_1 and a slightly <u>larger</u> c than the experimenter's guess. Wetherill (1963) showed that the procedure (2) with larger c is less susceptible to a poor choice of x_1 especially for small samples.

Since the logit (and any other parametric) assumption is vulnerable on the extreme tails, it may be desirable to use an estimation method that places less weights on the observations with more extreme x_i's. For data generated by sequential procedures like (2), (4) and (5), the x_i's in the initial runs tend to be more extreme. A simple way to achieve this is to insert weight $w_i = w(|x_i - x_n|)$ on both sides of (8) and solve iteratively the weighted version of the likelihood equation (8), where w(z) is decreasing in $z > 0$, and x_n is considered to be a good estimate of L_p. If we choose w_i to be 0 or 1, it is equivalent to performing the unweighted MLE based on a subset of data with moderate x_i's. The general question of robust estimation for quantal response data was addressed in Miller and Halpern (1980).

Let $\hat{L}_p^{(n)}$ be the MLE of L_p from n observations. Its variance $\mathrm{var}(\hat{L}_p^{(n)})$ can also be estimated via the same parametric model by a standard method. A stopping rule may be devised based on this variance estimate. This provides another advantage of the parametric approach over the nonparametric one.

4. <u>A sequential design for estimating $L_{0.5}$ based on the Spearman-Kärber estimator</u>

If the unknown response curve $F(x) = H(x-\alpha, \phi)$ is skew-symmetric about α, i.e. $H(z,\phi) + H(-z,\phi) = 2H(0,\phi)$ for any z, ϕ, α is both the median $L_{0.5}$ and the mean of F. The Spearman-Kärber estimator (Finney, 1978, p. 394) is a nonparametric estimator of the (discretized) mean of F,

$$\hat{\alpha}_{SK} = \sum_{j=1}^{J} (\hat{p}_j - \hat{p}_{j-1})\tfrac{1}{2}(x_{j-1}+x_j),$$

where $x_1 < \ldots < x_J$, n_j observations are taken at x_j with r_j responses, $\hat{p}_j = r_j/n_j$, $n = \sum_1^J n_j$. Under conditions that ensure that $\hat{\alpha}_{SK}$ is an efficient estimator of α, an alternative sequential design for estimating the median $L_{0.5} = \alpha$ is the following:

$$\text{(11)} \qquad \begin{aligned} &\text{(i)} \quad \text{compute } \hat{\alpha}_{SK}^{(n)} = \hat{\alpha}_{SK}((y_i, x_i)_1^n), \\ &\text{(ii)} \quad \text{set} \quad x_{n+1} = \hat{\alpha}_{SK}^{(n)}. \end{aligned}$$

The two distinct advantages of the procedure (11) are: 1) computational ease, 2) weak assumption on F, i.e., the functional form of H is not assumed known. But the price to pay for these is quite dear. The conditions required to ensure a proper performance of (11) are quite restrictive. First, F should be skew-symmetric so that its mean and median are equal. Since $\hat{\alpha}_{SK}$ is an unbiased estimator of the discretized mean, not the population mean, their difference becomes negligible only when the spacing $(x_i)_1^n$ is reasonably dense. A proper use of $\hat{\alpha}_{SK}$ requires that x_1 and x_J are chosen such that $F(x_1) = 0$, $F(x_J) = 1$, which may be hard to achieve in the initial stage of the type of sequential designs considered in the paper. If the experimenter has to pray for the validity of these assumptions, the procedure (11) can not be truly termed "nonparametric." Therefore it will not be included in the empirical study of §6.

5. Some theoretical results relating the sequential designs (5) to the Stochastic Approximation method.

Our efficiency claim in §3 for the "logit-MLE" version of (5) is based on the good faith in MLE for the logit model. This loose claim will be reinforced in this section by showing that, for the estimation of any percentile L_p, a linear approximation to the "logit-MLE" version of (5) is equivalent to the asymptotically fully efficient version (2a) of the Stochastic Approximation scheme. It is interesting to note that, while the former is a seemingly parametric procedure, the latter is nonparametric. This lends further support to our previous argument that the "logic-MLE" version is not sensitive to distributional assumption, at least for the middle percentiles.

The following simple approximation

$$\frac{1}{1+e^{-t}} \simeq \frac{1}{2} + \frac{1}{6}t \tag{12}$$

has a maximum error of 0.07 in the range $|t| \leqslant 3$ (Cox, 1970, p. 90). Assume that $|\lambda(x_i-\alpha)| \leqslant 3$ for all i in the likelihood equation (8). (This may be achieved by applying the MLE to the data satisfying the constraint.) By applying the approximation (12) to (8), we have

$$\begin{aligned} &\sum_1^n \frac{1}{2} + \frac{1}{6}\sum_1^n(\lambda x_i-\mu) = \sum_1^n y_i, \quad \mu = \lambda\alpha \\ &\sum_1^n \frac{x_i}{2} + \frac{1}{6}\sum_1^n(\lambda x_i^2 - \mu x_i) = \sum_1^n y_i x_i . \end{aligned} \tag{13}$$

Let $\hat{\lambda}$ and $\hat{\mu}$ be the solutions to (13). The estimator $\hat{\alpha}_n$ of the median $L_{0.5}$ = α is obtained as follows:

$$\hat{\alpha}_n = \frac{\hat{\mu}}{\hat{\lambda}} = \frac{\sum^n x_i \sum^n (y_i-\frac{1}{2})x_i - \sum^n x_i^2 \sum^n (y_i-\frac{1}{2})}{\sum^n 1 \sum^n (y_i-\frac{1}{2})x_i - \sum^n x_i \sum^n (y_i-\frac{1}{2})}, \tag{14}$$

which is the weighted average of x_i, i = 1,...,n with weight w_i proportional to $\sum^n (y_j-\frac{1}{2})x_j - x_i\sum^n(y_j-\frac{1}{2}) = \frac{1}{2}\sum_{y_j=1}(x_j-x_i) - \frac{1}{2}\sum_{y_j=0}(x_j-x_i)$.

More generally for any p, we can consider the approximation

$$(15) \qquad J(t) = \frac{1}{1+e^{-t}} \simeq p+(t-J^{-1}(p))J'(J^{-1}(p)), \; J^{-1}(p) = -\ln(\frac{1}{p} - 1),$$

which is approximately valid for x_i close to L_p. By applying (15) to (8) we obtain

$$(16) \qquad \begin{aligned} \sum^n(\lambda x_i - \lambda L_p) &= \frac{1}{J'(J^{-1}(p))}\sum^n(y_i - p) \\ \sum^n(\lambda x_i^2 - \lambda L_p x_i) &= \frac{1}{J'(J^{-1}(p))}\sum^n(y_i - p)x_i \end{aligned}$$

where the 100p percentile $L_p = \alpha - \frac{1}{\lambda}\ln(\frac{1}{p} - 1) = \alpha + \frac{1}{\lambda}J^{-1}(p)$. The estimator $\hat{L}_p^{(n)}$, (17), is obtained from $\widehat{\lambda L_p}$ and $\hat{\lambda}$ by solving (16),

$$(17) \qquad \hat{L}_p^{(n)} = \frac{\widehat{\lambda L_p}}{\hat{\lambda}} = \frac{\sum^n x_i \sum^n (y_i - p)x_i - \sum^n x_i^2 \sum^n (y_i - p)}{\sum^n 1 \sum^n (y_i - p)x_i - \sum^n x_i \sum^n (y_i - p)},$$

which is the weighted average of x_i with weight w_i proportional to $\sum_j^n(y_j - p)(x_j - x_i)$. Since $\hat{L}_p^{(n)}$ is independent of $J'(J^{-1}(p))$ of (16), $J'(J^{-1}(p))$ in the approximation (15) can be replaced by any other constant without affecting the subsequent results (17) and (18). Formula (17) extends (14).

Note that some w_i may be negative. The denominators of (14) and (17) are both equal to $n \sum_{y_i=1} x_i - \sum_{y_i=1} 1 \sum^n x_i$, which is nonzero unless $\left(\sum_{y_i=1} 1\right)^{-1} \sum_{y_i=1} x_i = n^{-1}\sum^n x_i$. From (17) and after some algebras, it is easy to show that the $(n+1)^{th}$ run, according to the procedure (5),

$$x_{n+1} = \hat{L}_p^{(n)} = \hat{L}_p^{(n-1)} - \frac{(y_n-p)\sum^n(x_i-x_n)^2}{n\sum^n y_i(x_i-\bar{x}_n)}$$

(18)

$$= x_n - \frac{c_n}{n}(y_n-p), \quad c_n = \frac{\sum^n(x_i-x_n)^2}{\sum^n y_i(x_i-\bar{x}_n)},$$

where $\bar{x}_n = n^{-1}\sum^n x_i$. Therefore the linear approximation (18) to our procedure (5) is asymptotically fully efficient if c_n in (18) converges almost surely to $[F'(L_p)]^{-1}$. To this end, note that the regression slope estimate $\hat{\beta}_n$ in (3) converges to $F'(L_p)$ a.s. By comparing (18) and (3), $c_n - \hat{\beta}_n^{-1} = n(x_n - \bar{x}_n)^2/\sum^n y_i(x_i-\bar{x}_n)$. Since both procedures converge to L_p for large n, $x_n \simeq \bar{x}_n$ and $c_n - \hat{\beta}_n^{-1} \rightarrow 0$ follows from the assumption $F'(L_p) > 0$. Therefore the asymptotic full efficiency of (18) follows from similar results of Anbar (1978) and Lai and Robbins (1981). (Their regularity conditions do not apply directly to the quantal response problem but their technique can be modified to suit our purpose.)

6. A simulation study

Under comparison are (i) the logit-MLE version of the sequential designn (5) with truncation as defined in (10) (abbreviated as MLE in the Tables), (ii) the following adaptive Robbins-Monro (ARM) design with truncation,

(19) $$x_{n+1} = x_n - \frac{c_n}{n}(y_n-p), \quad c_n = \max(-c, \min(c,\hat{\beta}_n^{-1})),$$

where $\hat{\beta}_n$ is defined in (3), (iii) the Robbins-Monro (RM) design (2), and (iv) the Up-and-Down (UD) design (4).

Since the MLE of the logit model does not often exist for very small sample size, we fix an initial design of size 10 and a parametric distribution H for the quantal response curve. For each stimulus level x_i, $i = 1,\dots,10$, in the initial design, $y_i = 0$ or 1 is generated according as $u_i >$ or $< H(x_i)$, where u_i is the i^{th} uniform random number in $[0,1]$. Let $(\hat{\alpha}_{10}, \hat{\lambda}_{10})$ be the MLE of (α,λ) in the logit model based on $\{x_i, y_i\}_1^{10}$. The common starting value for all designs under comparison is chosen to be $x_{11} = \hat{\alpha}_{10} - \hat{\lambda}_{10}^{-1} \ln(p^{-1}-1)$ according to (5)(ii). Once x_{11} is chosen, the subsequent design levels $x_{12},\dots,x_{35}$ are generated according to different design schemes, but with the same random numbers u_i. If the MLE $(\hat{\alpha}_{10}, \hat{\lambda}_{10})$ does not exist the simulation sample is discarded. On the other hand, if $(\hat{\alpha}_{10}, \hat{\lambda}_{10})$ exists and is unique, the subsequent MLE always exists as is obvious from condition (9). This is repeated for 500 times, including those discarded due to the nonexistence of MLE. (The total number of discarded samples is denoted by M in Tables I and II. The actual number of simulation samples in our study varies from 386 to 484.) For sample size n, the Monte Carlo mean square error (MSE) of a sequential design is calculated as the average of $(x_n - L_p)^2$ over the simulation samples. In Table I, $\sqrt{MSE}$ are given for the desings (i) - (iv) for estimating $L_{0.5}$. In Table II, $\sqrt{MSE}$ are given for the designs (i) - (iii) for estimating $L_{0.75}$.

For the study of robustness to distributional assumption and to the choice of starting value and other constants, we choose a variety of initial designs and response curves to reflect the degrees of the experimenter's knowledge of the response curve. (But note that for design (i) the MLE is always computed on the assumption of the logit model no matter what the true response curve is.) In Tables I(a)(c) and II(a)(c)(e), the standard logit model is used for the true response curve. In other tables, the probit models with different locations and scales are used for the true response curve. In each table L_p denotes the design level that is the 100p percentile of the corresponding response curve. Therefore, for example, the two initial designs in Tables I(a) and I(b) are identical, but correspond to different percentiles under different response curves.

Table I. Monte Carlo $\sqrt{MSE}$ (Root Mean Square Error) of Sequential Designs for Estimating the 50 Percentile of the True Quantal Response Curve

(a) Initial design:

stimulus level	$L_{0.1}$	$L_{0.3}$	$L_{0.5}$	$L_{0.7}$	$L_{0.9}$
no. of observations	1	2	4	2	1

True response curve: logit model (6) with $\alpha = 0$, $\lambda = 1$

design	n = 12	16	20	25	30	35	no. truncations
MLE-30	1.44	1.02	.72	.56	.43	.36	197
MLE-50	1.40	.78	.53	.47	.40	.36	102
MLE-100	1.40	.60	.49	.46	.40	.36	46
MLE-200	1.36	.61	.54	.48	.40	.35	15
MLE-600	1.48	.78	.56	.45	.41	.36	0
ARM-16	1.55	1.29	1.09	.92	.78	.67	321
ARM-30	1.52	1.16	.88	.69	.53	.45	97
ARM-50	1.54	1.16	.91	.69	.55	.46	32
ARM-100	1.59	1.24	.99	.80	.62	.51	12
ARM-600	1.63	2.02	1.66	1.67	1.36	1.12	4
RM-32	1.84	1.41	1.21	.94	.81	.74	
RM-16	1.58	1.31	1.14	.97	.83	.73	
RM-4	1.59	1.46	1.37	1.29	1.23	1.19	
UD-2	2.10	1.94	1.86	1.60	1.82	1.51	
UD-1	1.65	1.44	1.29	1.17	1.14	1.10	
UD-0.25	1.50	1.40	1.26	1.15	1.05	.94	

M = 114

where L_p = 100p percentile of the true response curve,
MLE-d = procedure (10) with truncation constant d
ARM-c = procudure (19) with truncation constant c
RM-c = procedure (2) with constant c
UD-Δ = procedure (4) with constant Δ
no. truncations = total number of truncations of the kind (10) or (19)
M = total number of simulation samples for which no MLE exists

(b) Initial design (same as I(a)):

stimulus level	$L_{0.3}$	$L_{0.46}$	$L_{0.56}$	$L_{0.66}$	$L_{0.80}$
no. of observations	1	2	4	2	1

True response curve: probit model (7) with $\mu = -0.5$, $\sigma = 3.1915$

design	n = 12	16	20	25	30	35	no. truncations
MLE-30	1.87	1.34	1.07	.85	.82	.77	859
MLE-50	1.84	1.10	.88	.83	.77	.73	472
MLE-100	1.95	.93	.85	.74	.83	.62	214
MLE-200	1.95	.90	.79	.71	.84	.63	62
MLE-600	1.90	1.13	.87	.75	.76	.63	4
ARM-16	2.01	1.70	1.54	1.36	1.21	1.08	1411
ARM-30	2.00	1.58	1.42	1.23	1.08	.96	625
ARM-50	2.06	1.62	1.39	1.21	1.07	.92	251
ARM-100	2.15	1.80	1.56	1.32	1.19	1.02	90
ARM-600	2.21	3.32	2.90	2.37	1.86	1.56	9
RM-32	2.16	1.77	1.51	1.35	1.17	1.05	
RM-16	2.01	1.69	1.47	1.28	1.10	.93	
RM-4	2.07	1.92	1.81	1.71	1.63	1.56	
UD-2	2.41	2.22	2.21	2.27	2.18	2.15	
UD-1	2.04	1.73	1.58	1.60	1.48	1.40	
UD-0.25	2.06	1.83	1.64	1.46	1.31	1.13	

M = 56

(c) Initial design:

stimulus level	$L_{0.3}$	$L_{0.5}$	$L_{0.7}$	$L_{0.8}$	$L_{0.9}$
no. of observations	1	2	3	3	1

True response curve: logit model (6) with $\alpha = 0$, $\lambda = 1$

designs	n: 12	16	20	25	30	35	no. truncations
MLE-30	2.00	1.26	.95	.75	.57	.45	637
MLE-50	1.92	1.06	.71	.56	.47	.41	349
MLE-100	1.89	.86	.61	.53	.46	.42	182
MLE-200	2.22	.89	.67	.56	.54	.46	66
MLE-600	2.68	1.47	.66	.58	.53	.63	10
ARM-16	2.11	1.72	1.44	1.18	.98	.84	747
ARM-30	2.09	1.60	1.21	.89	.70	.52	334
ARM-50	2.09	1.54	1.04	.82	.66	.53	165
ARM-100	2.06	1.58	1.13	.85	.71	.59	45
ARM-600	2.21	1.86	1.30	1.05	1.01	.82	2
RM-32	2.14	1.46	1.22	1.02	.84	.72	
RM-16	2.06	1.56	1.30	1.11	.96	.85	
RM-4	2.16	1.97	1.83	1.71	1.61	1.54	
UD-2	2.36	1.92	1.88	1.61	1.79	1.55	
UD-1	2.06	1.50	1.34	1.18	1.11	1.08	
UD-0.25	2.14	1.85	1.62	1.40	1.21	1.07	

M = 99

(d) Initial Design (same as I(c)):

stimulus level	$L_{0.2}$	$L_{0.38}$	$L_{0.58}$	$L_{0.71}$	$L_{0.85}$
no. of observations	1	2	3	3	1

True response curve: probit model (7) with $\mu = 0.5$, $\sigma = 1.5957$

design	n: 12	16	20	25	30	35	no. truncations
MLE-30	1.10	.60	.52	.47	.42	.37	325
MLE-50	1.02	.57	.51	.45	.42	.34	152
MLE-100	1.01	.58	.47	.43	.40	.34	70
MLE-200	1.41	.70	.49	.43	.43	.34	32
MLE-600	1.94	.92	.50	.43	.41	.34	4
ARM-16	1.26	.98	.80	.67	.57	.50	371
ARM-30	1.26	.98	.80	.65	.56	.48	106
ARM-50	1.29	1.02	.86	.69	.58	.50	50
ARM-100	1.35	1.11	.92	.74	.62	.54	14
ARM-600	1.45	1.19	.97	.95	.73	.57	1
RM-32	1.57	1.16	1.02	.90	.81	.73	
RM-16	1.28	.98	.74	.66	.59	.52	
RM-4	1.30	1.14	1.02	.91	.83	.76	
UD-2	1.95	1.86	1.90	1.59	1.83	1.45	
UD-1	1.35	1.08	1.09	1.15	1.17	1.08	
UD-0.25	1.29	1.05	.87	.71	.59	.51	

M = 76

Table II. Monte Carlo $\sqrt{MSE}$ (Root Mean Square Error) of Sequential Designs for Estimating the 75 Percentile of the True Quantal Response Curve

(a) Initial design and true response curve: same as in I(a)

design	n = 12	16	20	25	30	35	no. truncations
MLE-30	1.43	.87	.70	.61	.56	.48	561
MLE-50	1.36	.80	.64	.56	.53	.47	212
MLE-100	1.38	.77	.64	.58	.55	.50	50
MLE-200	1.43	.77	.65	.59	.56	.50	23
MLE-600	1.49	.76	.65	.59	.56	.51	3
ARM-16	1.54	1.19	1.02	.87	.78	.70	228
ARM-30	1.55	1.21	1.05	.89	.79	.72	86
ARM-50	1.60	1.26	1.09	.93	.83	.75	29
ARM-100	1.69	1.37	1.19	1.01	.90	.81	7
ARM-600	1.71	1.41	1.73	1.39	1.16	.98	1
RM-32	1.69	1.23	1.16	.93	.87	.78	
RM-16	1.51	1.13	.93	.75	.68	.61	
RM-4	1.57	1.41	1.28	1.17	1.08	1.01	

N = 114

For explanation of symbols, see the bottom of Table I(a)

(b) Initial design and true response curve: same as in I(b)

design	n = 12	16	20	25	30	35	no. truncations
MLE-30	1.97	1.47	1.21	1.16	1.04	.98	1817
MLE-50	1.95	1.51	1.18	1.12	1.08	1.01	796
MLE-100	2.09	1.38	1.22	1.15	1.11	1.03	205
MLE-200	2.33	1.43	1.27	1.14	1.11	1.06	68
MLE-600	2.33	1.41	1.55	1.16	1.12	1.06	11
ARM-16	2.02	1.80	1.56	1.40	1.26	1.12	1462
ARM-30	1.98	1.74	1.54	1.32	1.20	1.16	580
ARM-50	1.99	1.80	1.69	1.46	1.28	1.23	308
ARM-100	2.07	2.08	1.92	1.63	1.45	1.40	108
ARM-600	2.08	3.85	3.39	2.98	2.41	2.19	12
RM-32	2.01	1.52	1.45	1.29	1.19	1.07	
RM-16	1.95	1.46	1.20	1.03	.91	.82	
RM-4	2.08	1.86	1.71	1.56	1.45	1.36	

N = 56

(b1) Initial design:

stimulus level*	$L_{0.3}$	$L_{0.46}$	$L_{0.56}$	$L_{0.66}$	$L_{0.80}$
no. of observations	1	3	6	3	1

Initial sample size: 14

True response curve: same as in II(b)

design	n 16	20	25	30	35	no. truncations
MLE-30	2.23	1.63	1.37	1.13	.99	2694
MLE-50	2.23	1.42	1.20	1.06	.97	1387
MLE-100	2.27	1.31	1.16	1.07	1.01	483
MLE-200	2.76	1.40	1.18	1.11	1.03	160
MLE-600	2.91	1.48	1.17	1.25	1.07	32
ARM-16	2.26	1.97	1.67	1.48	1.31	2116
ARM-30	2.24	1.91	1.60	1.44	1.31	855
ARM-50	2.26	1.97	1.71	1.52	1.38	388
ARM-100	2.31	2.10	1.82	1.62	1.41	113
ARM-600	2.96	3.36	3.12	2.72	2.38	11
RM-32	2.15	1.67	1.36	1.20	1.04	
RM-16	2.20	1.77	1.46	1.28	1.13	
RM-4	2.31	2.15	2.01	1.90	1.81	

M = 16

* same as in II(b)

(c) Initial design and true response curve: same as in I(c)

design	n 12	16	20	25	30	35	no. truncations
MLE-30	2.58	1.96	1.57	1.28	1.08	.93	808
MLE-50	2.46	1.57	1.15	.83	.58	.44	496
MLE-100	2.22	1.02	.54	.54	.44	.43	186
MLE-200	1.89	.85	.61	.52	.45	.41	82
MLE-600	1.44	.85	.57	.54	.44	.47	19
ARM-30	2.65	2.12	1.75	1.47	1.26	1.11	330
ARM-50	2.61	1.93	1.53	1.18	.94	.82	204
ARM-100	2.62	1.74	1.36	1.09	.93	.87	82
ARM-200	2.72	1.81	1.54	1.23	1.03	.94	13
ARM-600	2.73	1.88	1.86	1.49	1.17	1.03	1
RM-32	2.77	2.13	1.79	1.42	1.24	1.07	
RM-16	2.71	2.32	2.08	1.84	1.66	1.53	
RM-4	2.78	2.65	2.56	2.47	2.40	2.35	

M = 99

(d) Initial design and true response curve: same as in I(d)

design	n: 12	16	20	25	30	35	no. truncations
MLE-30	2.59	1.72	1.16	.81	.69	.60	782
MLE-50	2.46	1.19	.82	.60	.48	.41	420
MLE-100	2.16	.91	.57	.49	.44	.41	153
MLE-200	1.62	.84	.60	.49	.45	.42	56
MLE-600	1.26	.77	.60	.50	.46	.43	7
ARM-30	2.66	1.93	1.42	1.08	.93	.81	233
ARM-50	2.60	1.76	1.29	.98	.82	.70	83
ARM-100	2.67	1.78	1.30	1.01	.88	.77	46
ARM-200	2.83	2.03	1.49	1.19	1.00	.86	5
ARM-600	2.83	2.48	1.92	1.54	1.30	1.12	1
RM-32	2.74	1.89	1.39	1.03	.94	.84	
RM-16	2.72	2.23	1.87	1.55	1.31	1.11	
RM-4	2.82	2.65	2.53	2.42	2.33	2.26	

M = 76

(e) Initial design:

stimulus level	$L_{0.1}$	$L_{0.2}$	$L_{0.3}$	$L_{0.5}$	$L_{0.7}$
no. of observations	1	3	3	2	1

True response curve: logit model (6) with $\alpha = 0$, $\lambda = 1$

design	n: 12	16	20	25	30	35	no. truncations
MLE-30	4.64	3.55	2.98	2.44	2.01	1.68	1940
MLE-50	4.53	2.98	2.11	1.36	.94	.67	1178
MLE-100	4.34	1.71	.87	.71	.64	.57	417
MLE-200	4.14	1.88	.90	.73	.70	.63	143
MLE-600	3.65	1.92	.92	.78	.80	.66	40
ARM-30	4.66	3.84	3.38	2.97	2.66	2.41	246
ARM-50	4.65	3.85	3.36	2.95	2.63	2.38	125
ARM-100	4.68	3.92	3.43	3.01	2.68	2.42	35
ARM-200	4.81	4.08	3.56	3.12	2.78	2.50	5
ARM-600	4.81	4.31	3.69	3.18	2.79	2.51	1
RM-32	4.51	3.55	2.97	2.38	1.97	1.64	
RM-16	4.66	4.01	3.65	3.33	3.08	2.89	
RM-4	4.83	4.62	4.47	4.33	4.23	4.14	

M = 96

(f) Initial design (same as II(e)):

stimulus level	$L_{0.15}$	$L_{0.29}$	$L_{0.42}$	$L_{0.62}$	$L_{0.80}$
no. of observations	1	3	3	2	1

True response curve: probit model (7) with $\mu = -0.5$, $\sigma = 1.5957$

design	n: 12	16	20	25	30	35	no. truncations
MLE-30	3.78	2.95	2.54	2.13	1.80	1.54	1334
MLE-50	3.66	2.54	1.86	1.25	.83	.57	859
MLE-100	3.43	1.56	.74	.58	.54	.46	277
MLE-200	3.06	1.82	.78	.62	.58	.51	89
MLE-600	2.44	1.88	.77	.64	.59	.53	19
ARM-30	3.82	3.18	2.81	2.48	2.22	2.01	147
ARM-50	3.82	3.19	2.80	2.46	2.21	2.00	40
ARM-100	3.87	3.27	2.85	2.50	2.24	2.02	19
ARM-200	3.99	3.51	3.00	2.60	2.32	2.09	6
ARM-600	3.99	4.67	3.35	2.80	2.42	2.15	2
RM-32	3.78	3.04	2.61	2.14	1.79	1.52	
RM-16	3.84	3.34	3.07	2.82	2.64	2.48	
RM-4	3.96	3.79	3.68	3.57	3.49	3.42	

M = 66

The results in Tables I and II are summarized in the following.

(A) General comparison of designs.

In general, the performance of the designs is in the following descending order,

$$\text{MLE} > \text{ARM} > \text{RM} > \text{UD}$$

Only in Table II(b) does RM-16 (the Robbins-Monro method (2) with $c = 16$) outperform the others. But when we increase the size of the initial design from 10 to 14 as in Table II(b1), MLE has again the best performance.

Within RM we observe the descending order of performance

$$\text{RM-32 and RM-16} > \text{RM-4} > \text{RM-1} > \text{RM-0.25}.$$

For Tables I(a)(b)(d) and II(a)(b), RM-16 > RM-32; for other tables, RM-32 > RM-16. Note that RM-4 is asymptotically fully efficient if the response curve is the standard logit (as assumed in Tables I(a)(c), II(a)(c)(e)) because $F'(0) = \frac{1}{4}$. RM-4 certainly fails to deliver this asymptotic promise of optimality for n as large as 35. Asymptotic results seem quite irrelevant in this context. Within UD, we observe the descending order of performance

$$\text{UD-0.25} > \text{UD-1} > \text{UD-2} > \text{UD-4},$$

where UD-Δ means the Up-and-Down method (4) with step size Δ. To save space, RM-1, RM-0.25 and UD-4 are not included in the tables.

Since our interest is in finding superior designs, we will confine the remaining discussion to MLE, ARM, RM-32 and RM-16. A very complete comparison of the empirical performance of RM-c and UD-Δ for different c and Δ was done in Wetherill (1963).

(B) Superiority of the logit-MLE design.

The superiority of the logit-MLE design (10) with truncation constant d, hereafter denoted as MLE-d, is broad-based. In the eleven tables, MLE-50, MLE-100, MLE-200 consistently outperform the best ARM. Except in Table II(c), MLE-30 outperforms the best ARM. The efficiency gain of MLE over ARM is more conspicuous for larger n.

What truncation constant d should be chosen? The MLE designs with $50 \leq d \leq 200$ all perform well. Within this range their difference of performance is probably negligible. MLE-30 does not perform as well, because a forceful truncation like $d = 30$ limits the potential of the MLE design in making more flexible and justifiably large moves when the design levels are not yet close enough to the target value. On the other hand, the performance of MLE-600, which involves very weak truncations, is more fluctuating. For $n \geq 20$, MLE-600 is comparable to the best MLE design. For small n, MLE-600 is comparable to the best MLE design in Tables I(a)(b) and II(a)(b), but worse than MLE-50, MLE-100, and MLE-200 in Tables I(c)(d) and II(b1). In Tables II(c)(d)(e)(f), MLE-600 is much better than the other MLE designs for $n = 12$ (an uninteresting case), and is comparable to MLE-200 for $n \geq 16$. For $n \geq 20$, the effect of truncation is negligible over $100 \leq d \leq 600$.

Since a major purpose for finding better designs is to reduce the number of runs required for satisfying an error bound, we shall measure the efficiency gain of the MLE-design over the ARM design by such numbers. In each case, we find the smallest $\sqrt{MSE}$ achieved by the best ARM design at $n = 35$. We then find m to be the <u>smallest sample size at which an MLE design achieves the same</u> $\sqrt{MSE}$. In Table III, the values of m are obtained by linear interpolation for the eleven tables in Tables I and II.

Table III. Values of m for Tables I(a)-(d), II(a)-(f)

I(a)	I(b)	I(c)	I(d)	II(a)	II(b)	II(b1)	II(c)	II(d)	II(e)	II(f)
26	16	26	20	18	25	20	17	18	15	14

The percentage of runs saved by using the best MLE design instead of the best

ARM design ranges from 25% to 60%. This surprising difference of performance can be explained by the different natures of the two approximation schemes. The adaptive Robbins-Monro design is a stochastic Newton-Raphson method which uses linear approximation to nonlinear equation for the iterative solution of nonlinear equation. It is known to be unstable unless the starting value is close to the target value. Only under this premise does the large sample results like asymptotic normality and efficiency make sense for small or moderate samples. On the other hand our logit-MLE method seems to be free from this problem since a sigmoid curve is used in the iterative solution of nonlinear equation.

C. Improvement of ARM over RM.

There is a slight but definite improvement of ARM-c (procedure (19) with truncation constant c) with c = 50, 100 over the best RM in Tables I(a)(c) II(c)(d). The best RM design (RM-16 or RM-32) is usually quite comparable to the best ARM design. In Tables II(a)(b)(b1)(e)(f) it even beats the best ARM. (But RM-4 is definitely inferior to the best ARM))

The best performance of ARM occurs with ARM-c with $16 \leq c \leq 100$ with the majority of them in the narrower range $30 \leq c \leq 50$. The ARM-600 design, which involves very weak truncation, is a real disappointment. Except for Table II(c), it is worse than the best (nonadaptive) RM design. It is consistently worse than the best ARM design, and for Tables II(b)(b1)(e)(f) much worse. Asymptotic full efficiency is a quite irrevelant concept here. Moreover the MSE of ARM-600 exhibits an erratic pattern, e.g, it sometimes increases as n increases. Generally the ARM requires more severe truncation than the MLE. This is because the ARM can make an unduly large move from x_n to x_{n+1} as explained in Section 2.

D. In Tables I and II we have counted the total number of times the truncation (defined in (10) and (19)) is invoked. For the same truncation constant, the MLE design always requires more trunctations than the ARM Design. It suggests that the MLE design makes large moves more frequently than the ARM design. Since MLE-100, MLE-200 and MLE-600 do very well in the study, such large moves are probably justified.

We have also examined the empirical behavior of the same set of designs for initial designs of size 25. The results are very similar. As the size of the initial design increases, the number of simulation samples for which no MLE exists quickly drops.

Since the MLE-d designs with $50 \leqslant d \leqslant 600$ perform extremely well in a variety of situations considered in this paper, we suggest that they may be considered seriously in practical work.

ACKNOWLEDGEMENTS

Thanks, are due to G.E.P. Box and Tom Leonard for useful discussions, and to Siu-Keung Tse for computational assistance.

References

Anbar, D. (1978), "A stochastic Newton-Raphson method," Journal of Statistical Planning and Inference, 2, 153-163.

Berkson, J. (1955), "Maximum likelihood and minimum χ^2 estimates of the logistic function," Journal of the American Statistical Association, 50, 130-162.

Cox, D. (1970), Analysis of Binary Data, London: Methuen.

Chung, K.L. (1954), "On a stochastic approximation method," Annals of Mathematical Statistics, 25, 463-483.

Dixon, W.J. and Mood, A.M. (1948), "A method for obtaining and analyzing sensitivity data," Journal of the American Statistical Associates, 43, 109-126.

Finney, D.J. (1978), Statistical Method in Biological Assay, London: Griffin.

Freeman, P.R. (1970), "Optimal Bayesian sequential estimation of the median effective dose," Biometrika, 57, 79-89.

Hodges, J.L. and Lehmann, E.L. (1955), "Two approximations to the Robbins-Monro process," Proceedings of the 3rd Berkeley Symposium, 1, 95-104.

Lai, T.L. and Robbins, H. (1981), "Consistency and asymptotic efficiency of slope estimates in stochastic approximation schemes," Z. Wahrscheinlichkeitstheorie verw. Gebiete, 56, 329-360.

Leonard, T. (1982), "An inferential approach to the bioassay design problem," Technical report, Univ. of Wisconsin, Madison.

Lord, F.M. (1971), "Tailored testing, an application of stochastic approximation," Journal of the American Statistical Association, 66, 707-711.

Miller, R.G. and Halpern, J.W. (1980), "Robust estimators for quantal bioassay," Biometrika, 67, 103-110.

Owen, R.J. (1975), "A Bayesian sequential procedure for quantal response in the context of adaptive mental testing," Journal of the American Statistical Association, 70, 351-356.

Robbins, H. and Monro, S. (1951), "A stochastic approximation method," Annals of Mathematical Statistics, 29, 400-407.

Rose, R.M., Teller, D.Y. and Rendleman, P. (1970), "Statistical properties of staircase estimates," Perception and Psychophysics, 8, 199-204.

Sacks, J. (1958), "Asymptotic distribution of stochastic approximation procedures," Annals of Mathematical Statistics, 29, 373-405.

Silvapulle, M.J. (1981), "On the existence of maximum likelihood estimators for the binomial response model," Journal of the Royal Statistical Society B, 43, 310-313.

Tsutakawa, R.K. (1972), "Design of experiment for bioassay," Journel of the American Statistical Association, 67, 584-590.

Wetherill, G.B. (1963), "Sequential estimation of quantal response curves," (with Discussions) Journal of the Royal Statistical Society B, 25, 1-48.

———— (1966), Sequential Methods in Statistics, London: Methuen.

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
2563

2. GOVT ACCESSION NO.
AD-A134 593

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
Efficient Model-Based Sequential Designs for Sensitivity Experiments

5. TYPE OF REPORT & PERIOD COVERED
Summary Report - no specific reporting period

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
C. F. Jeff Wu

8. CONTRACT OR GRANT NUMBER(s)
DAAG29-80-C-0041 &
ARO DAAG29-82-K0154

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Mathematics Research Center, University of Wisconsin
610 Walnut Street
Madison, Wisconsin 53706

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
Work Unit Number 4 -
Statistics & Probability

11. CONTROLLING OFFICE NAME AND ADDRESS
U. S. Army Research Office
P.O. Box 12211
Research Triangle Park, North Carolina 27709

12. REPORT DATE
September 1983

13. NUMBER OF PAGES
26

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report)
UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Logit, Optimal design, Quantal response curve, Robbins-Monro stochastic approximation, Sensitivity experiments, Spearman-Karber estimator, Up-and-Down method

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A sequential design for estimating the percentiles of a quantal response curve is proposed. Its updating rule is based on an efficient summary of all the data available via a parametric model. Its efficiency in terms of saving the number of runs and its robustness against the distributional assumption are demonstrated heuristically and in a simulation study. A linear approximation to the "logit-MLE" version of the proposed sequential design is shown to be equivalent to an asymptotically optimal stochastic approximation method, thereby providing a large sample justification. For sample size

ABSTRACT (Continued)

between 12 and 35, the simulation study shows that the "logit-MLE" version of the general sequential procedure substantially outperforms an adaptive (and asymptotically optimal) version of the Robbins-Monro method, which in turn outperforms the nonadaptive Robbins-Monro and Up-and-Down methods. A nonparametric sequential design, via the Spearman-Kärber estimator, for estimating the median is also proposed.

END

FILMED

1-83

DTIC