Optimal Designs of Discrete Response Experiments in Contingent Valuation Studies

Hans Nyquist

The Review of Economics and Statistics, Vol. 74, No. 3. (Aug., 1992), pp. 559-563.

Stable URL:
http://links.jstor.org/sici?sici=0034-6535%28199208%2974%3A3%3C559%3AODODRE%3E2.0.CO%3B2-S

The Review of Economics and Statistics is currently published by The MIT Press.

ratios at the 0.05 level. Furthermore, the $\chi^2_{\min}n$ of two of the ratios is a random walk.

In summary, P & S's findings adjustments of long-term ratios stem from a statistical bias; the evidence points to the conclusion that the ratios follow a random walk. Further, the adjustment of the short-term ratios reported by P & S is slightly biased downward.

REFERENCES

Kendall, Sir Maurice, Alan Stuart, and J. Keith Ord, *The Advanced Theory of Statistics Volume 3* (London: Charles Griffen and Company, 1983).

Peles, Yoram C., and Meir I. Schneller, "The Duration of the Adjustment Process of Financial Ratios," this REVIEW 71 (Nov. 1989), 527–532.

OPTIMAL DESIGNS OF DISCRETE RESPONSE EXPERIMENTS IN CONTINGENT VALUATION STUDIES

Hans Nyquist*

Abstract—Optimal designs for estimating model parameters and other characteristics such as mean and median willingness to pay are discussed when a logistic or a probit regression model is used for analyzing a contingent valuation study with discrete questions. A numerical example, related to a study of the value of preserving some virgin forests in Sweden, illustrates the efficiencies of different designs and how a sequential procedure can be applied.

I. Introduction

The contingent valuation method is an approach for estimation of environmental benefits. The approach dates back to Davis (1964), who measured the benefits accruing from visiting a particular recreational area. The early development of the contingent valuation method is reviewed in Mitchell and Carson (1989). In recent years the approach has received a rapidly growing interest. A review of some recent applications is found in Kriström (1990).

In the contingent valuation method a hypothetical market for trading environmental goods is created. Valuation questions are then put to individuals in order to reveal their preferences. There are several ways of framing the valuation questions (see, e.g., Kriström (1990) for a review), one of which is the discrete valuation question, introduced by Bishop and Heberlein (1979).

When discrete valuation questions are used m different costs, $x_1, \ldots, x_m$, are selected and distributed over n individuals in groups of $n_1, \ldots, n_m$ individuals, respectively, $n = \Sigma n_j$. Individual i_j, $i = 1, \ldots, n_j$, $j = 1, \ldots, m$, is asked whether he accepts or rejects the cost x_j for receiving the benefit under study. We define a binary response variable Y_{ij} to take the value 1 if he accepts the cost and 0 otherwise. According to the random maximization utility model each individual is characterized by a willingness to pay, *WTP*, randomly distributed over the population of individuals. Individual i_j will therefore accept the cost x_j presented to him if x_j is less than or equal to his particular willingness to pay, WTP_{ij}. For a randomly selected individual the probability for acceptance is therefore

$$P(Y_{ij} = 1) = \pi_j = P(x_j \le WTP_{ij}) = 1 - H(x_j)$$

where H is the cumulative distribution function of *WTP*. This suggests a generalized linear model (McCullagh and Nelder, 1989) with binary response

$$Y_{ij} \sim \text{bin}(1, \pi_j) \quad i = 1, \ldots, n_j, j = 1, \ldots, m \tag{1a}$$

$$\pi_j = g(\eta_j) \tag{1b}$$

$$\eta_j = \alpha + \beta x_j \tag{1c}$$

where α and β are unknown parameters and the link function g is related to the distributional assumption of *WTP*. Specifically, if H is the logistic distribution, then $\pi_j = \{1 + \exp(\eta_j)\}^{-1}$ and (1) defines a logistic regression model. Another choice of H that has been used in applications is the normal distribution, which leads to $\pi_j = 1 - \phi(\eta_j)$, ϕ being the cumulative distri-

Received for publication September 18, 1990. Revision accepted for publication August 21, 1991.

* University of Umeå.

This research was carried out while the author was a visiting fellow at the Imperial College, University of London, and was financially supported by the Swedish Research Council for the Humanities and the Social Sciences and by The British Council. The author is grateful to Dr. Bengt Kriström for introducing the author to the problem and to Professor John A. Nelder for fruitful discussions on the problem and valuable comments on the manuscript.

bution function of the standard normal distribution. Under these assumptions (1) defines a probit regression model.

The maximum likelihood estimator of α and β is asymptotically normal with covariance $\{I(\alpha, \beta)\}^{-1}$, where

$$I(\alpha, \beta) = X^T W X$$

is the Fisher information matrix, X is the $n \times 2$ matrix with 1 in the first column and the x-values in the second column, and W is the $n \times n$ diagonal matrix of weights $w = (d\pi/d\eta)^2/\{\pi(1-\pi)\}$ (Fahmeir and Kaufman, 1988). Note here that the weights depend on the linear predictor, and hence on the unknown model parameters α and β.

The purpose of an optimal experimental design is to select values $x_1, \ldots, x_m$ at which to take observations $Y_{ij}(x_j)$ so that estimates of quantities of interest are "as good as possible." This means that when a single parameter is of interest the values of x are selected so as to minimize the variance or the asymptotic variance of the estimator used. In multiparameter cases, however, there is no unique criterion for what is good; several functions of the information matrix have been proposed (see Fedorov (1972) and Silvey (1980) for useful reviews). One criterion that is often used is to minimize the volume of confidence ellipsoids, leading to D-optimality. Asymptotically, the volume of the ellipsoids is inversely proportional to $[\det\{I(\alpha, \beta)\}]^{1/2}$. Thus, a D-optimal design maximizes the determinant of the information matrix.

The aim of this paper is to discuss and contrast D-optimal designs of the costs when estimation of α and β versus estimation of the mean *WTP*, μ, is of primary concern. A problem with the designs is that they depend on unknown parameters. Information about the parameters prior to the experiment is therefore of vital importance. One way of mitigating this restriction is to conduct the experiment sequentially. In a sequential design the n individuals are arranged into batches. Observations are made at one batch before the design for the next batch is set. Therefore, a sequential design allows adjustments of the design as more information about the parameters is gained. Whether a sequential procedure is feasible in practice or not depends to a large extent on the data collection method. A sequential procedure is hardly possible if data are collected by mailed questionnaires, while centrally administrated telephone interviews would provide perfect conditions for a sequential design.

II. Optimal Designs

As shown by Ford et al. (1992), D-optimal designs for symmetrical distributions (like the logistic and the normal) maximize $w^2(\eta)\eta^2$ over $(0, \infty)$. The particular cases of logistic regression and probit regression were originally derived by White (1975) and Abdelbasit and Plackett (1983). See also Minkin (1987). The D-optimal designs are of the form $\eta = \pm c$, for a constant c, with equal number of observations at each design point. In the logistic case we have $c \approx 1.543$ and in the probit case $c \approx 1.138$.

Three important remarks should be made here. First, the design is concentrated at only two points, $m = 2$. If n is even, half of the observations should be taken at $x = (c - \alpha)/\beta$ and half at $x = -(c + \alpha)/\beta$. If n is odd, the optimal design cannot be achieved. Secondly, the pair of x-values that define an optimal design depends on the model parameters. Thus, an optimal design in one application can be far from optimal in another. Finally, the parameters that determine the design points are unknown, implying that the optimal designs cannot be derived in practice. However, the results can be useful if some prior information about the model parameters is available, or if it is possible to use a sequential design. A suitable algorithm for computation of sequential designs in the logistic case is indicated in Minkin (1987).

In many applications the mean *WTP*, μ, is of primary interest. For all symmetric distributions the mean *WTP* and median *WTP*, ρ, coincides and is related to α and β as $\mu = \alpha/\beta$. The use of this relation for estimating μ has been criticized because it allows for negative *WTP* values. However, the criticism should in that case not be on the formula, but on the model itself. Once the distribution of *WTP* over the population is accepted, the formula relating mean *WTP* to the model parameters follows.

A simple way of estimating μ is to take the quotient of the estimates of α and β. For the two-point design the information matrix I is diagonal with elements $(nw(c), nw(c)(c^2 + \alpha^2)/\beta^2)$. The asymptotic variance of $\hat{\alpha}/\hat{\beta}$ is therefore

$$V(\hat{\alpha}/\hat{\beta}) = \{1 + \alpha^2/(c^2 + \alpha^2)\}/(nw(c)\beta^2). \quad (2)$$

However, as will be seen, the optimal design for estimating α and β may be far from optimal when estimating μ.

The asymptotic variance of $\hat{\mu}$ is found as

$$V(\hat{\mu}) = (\nabla\mu)^T\{I(\alpha, \beta)\}^{-1}(\nabla\mu)$$

where $\nabla\mu = (\partial\mu/\partial\alpha, \partial\mu/\partial\beta)^T = (1/\beta, -\alpha/\beta^2)^T$ is the gradient vector of $\mu = \alpha/\beta$. The local D-optimal design is now obtained by selecting design points such that $V(\hat{\mu})$ is minimized.

As shown by Wu (1988) the optimal design for estimating μ (and ρ) in the logistic regression and probit regression models is the one-point design concentrating at μ. The asymptotic variance for this design is in the logistic case

$$\{2\pi'(\mu)\}^{-2} = 4/(n\beta^2). \tag{3}$$

A comparison of (2) and (3) shows that the optimal two-point design has in the logistic case an efficiency ranging from 29% to 58%, depending on α. Thus, the gain in efficiency when selecting a design can be considerable.

As for the two-parameter case also this design depends on unknown quantities. In addition, a one-point design does not allow estimation of other parameters (such as α and β or other percentiles). This implies that if a one-point design is allocated at a point different from μ, the mean *WTP* cannot be estimated. A sequential approach is therefore the only feasible approach. A sequential approach requires some initial information, which in practice is obtained from observations made at a few different design points in an initial stage of the experiment. Having such an initial stage would also alleviate the concern about the impossibility of estimating other parameters.

III. Sequential Procedures for Estimating Mean and Median Willingness to Pay

Our models imply $E[Y|x] = \pi(x)$ and we wish to estimate $\mu(=\rho)$, implicitly defined by $\pi(\rho) = 0.5$. The optimal design is the one-point design allocating x at the unknown ρ, the quantity we wish to estimate. Given observations at $x_1, \ldots, x_r$ with responses $y_1, \ldots, y_r$, respectively, the purpose of a sequential procedure is to suggest an estimate on ρ which can be used as the next design point, x_{r+1}. Robbins and Monro (1951) proposed the recursion

$$x_{r+1} = x_r - (d/r)(y_r - 0.5)$$

where d is a constant. Sacks (1958) has shown that the asymptotic distribution of $r^{1/2}(x_r - \rho)$ is normal with mean zero and variance $-(d/2)^2/\{2d\pi'(\rho) + 1\}$, provided $d\pi'(\rho) < 1/2$, where $\pi'(\rho)$ is the derivative of $\pi(x)$ with respect to x evaluated at $x = \rho$. In order to minimize the asymptotic variance the optimal choice of d is therefore $d = -\{\pi'(\rho)\}^{-1}$. In the logistic case $\pi'(\rho) = -\beta/4$ and in the probit case $\pi'(\rho) = -\beta/\sqrt{2\pi}$, so the optimal value on d is $-4/\beta$ and $-\sqrt{2\pi}/\beta$ in the two cases, respectively.

Several approaches for selecting d in cases where $\pi'(\rho)$ is unknown have been proposed. One approach is to use a fixed value given in advance, thus defining a non-adaptive Robbins-Monro procedure. Unfortunately, a poor choice of d and starting value x_1 will make the procedure remarkably inefficient for small and moderate samples. It is therefore natural to define a procedure in which the value on d is updated as more information about $\pi'(\rho)$ is gained. A first order Taylor approximation of $\pi(x)$ around $\pi(\rho)$ is

$$\begin{aligned}\pi(x) &\approx \pi(\rho) + (x-\rho)\pi'(\rho)\\ &= a + bx\end{aligned}$$

where $a = \pi(\rho) - \rho\pi'(\rho)$ and $b = \pi'(\rho)$. The least squares estimator of b is

$$b_r = \sum y_i(x_i - \bar{x}_r) \Big/ \Big\{\sum (x_i - \bar{x}_r)^2\Big\};$$

$$\bar{x}_r = r^{-1}\sum x_i$$

which gives the adaptive Robbins-Monro procedure

$$x_{r+1} = x_r - 1/(rb_r)(y_r - 0.5).$$

Under some regularity conditions Anbar (1978) and Lai and Robbins (1981) show that the sequence $\{x_r\}$ defined by the adaptive Robbins-Monro procedure has the same asymptotic distribution as the optimal non-adaptive Robbins-Monro procedure with $d = -\{\pi'(\rho)\}^{-1}$.

Wu (1985) proposed a maximum likelihood approach in which the next design point x_{r+1} is chosen to be the current maximum likelihood estimate $\hat{\rho}_r = \hat{\alpha}_r/\hat{\beta}_r$ of ρ. He also gave a truncated version with a rule for avoiding unduly large changes from x_r to x_{r+1}. Under rather restrictive conditions consistency has been established by Wu (1985). Assuming consistency, the maximum likelihood recursion approach applied to the logistic regression model is asymptotically equivalent to the adaptive Robbins-Monro procedure, and hence, is asymptotically optimal.

The procedures described in this section treat the case where the response variable Y_j obeys a Bernoulli distribution. For practical reasons it may be better in a contingent valuation study to take several observations, say n_r, at each design point before the results are evaluated and a new point is selected. The response variable is then binomial. The generalizations of the sequential procedures to this case are, however, immediate.

We finally remark that the sequential procedures presented use only the behavior of $\pi(x)$ in a vicinity of $x = \rho$. They are therefore robust against misspecifications of the distribution of *WTP*. In particular, x_r is a consistent estimator of ρ under very weak conditions on $\pi(x)$. In addition, as long as the distribution is symmetric ρ and μ are identical, implying that μ is also consistently estimated. If, on the other hand, the distribution is skewed ρ and μ are no longer equal and a modified version of the procedure is required. If, for example, the distribution of *WTP* is such that μ equals the $(100\nu)^{\text{th}}$ quantile, i.e., $\pi(\mu) = \nu$, then the non-

TABLE 1.—RESULTS OBTAINED FROM THE STATIC TEN POINT DESIGN

x		100	400	700	1000	1500	2000	2500	3000	5000	7000
y	First 10	6	4	6	4	4	6	5	5	2	2
	Total	50	48	53	45	53	40	39	34	15	10

TABLE 2.—RESULTS FROM THE SEQUENTIAL NONADAPTIVE ROBBINS-MONRO PROCEDURE

r	1	2	3	4	5	6	7	8	
x_r	1158	1206	1214	1235	1243	1253	1261	1266	
y_r	28	26	22	27	28	28	27	22	
r	9	10	11	12	13	14	15	16	final
x_r	1260	1267	1273	1270	1269	1268	1268	1265	1268
y_r	29	29	23	24	24	25	23	28	

adaptive Robbins-Monro recursion is

$$x_{r+1} = x_r - (d/N_r)\{(y_r/n_r) - \nu\} \tag{4}$$

where $N_r = \Sigma n_r$.

IV. A Numerical Illustration

Kriström (1990) presents a contingent valuation study for estimating the value of preserving some virgin forests in Sweden. Discrete valuation questions were asked to a sample of 900 households, equally distributed over the design points 100, 400, 700, 1000, 1500, 2000, 2500, 3000, 5000 and 7000 SEK. Answers from 562 households were obtained and analyzed in a logistic regression model yielding the estimates $\hat{\alpha} = 0.52$ and $\hat{\beta} = 3.7 \cdot 10^{-4}$, and hence the estimate of mean and median willingness to pay is 1424 with a standard deviation of 411.7.

We now illustrate the effects of using different designs in this experiment. For making results comparable, we use simulated responses based on a logistic distribution for *WTP* with $\alpha = 0.5$, $\beta = 4 \cdot 10^{-4}$ and $\mu = \rho = 1250$ as true values on the parameters. The first design in the comparison is static, i.e., all bid amounts are determined at one time, before observations are made. The design assigns 90 observations at each of the ten design points used in Kriström (1990). Simulation results when using this design are summarized in table 1. Estimates of the parameters are $\hat{\alpha} = 0.4858$, $\hat{\beta} = 3.6\ 10^{-4}$, and $\hat{\mu} = \hat{\rho} = 1355$. The estimated standard deviation of $\hat{\mu}$ is 424.9. These results agree fairly well with the results based on the real data.

In the second design we first use a pilot study consisting of the ten first observations at each point in the previous design as reported in table 1. The estimates from the pilot study, $\hat{\alpha} = 0.2613$, $\hat{\beta} = 2.3\ 10^{-4}$, and $\hat{\mu} = \hat{\rho} = 1158$, respectively, are then used as initial values in a nonadaptive Robbins-Monro sequential design with $d = -40000$ (four times the optimal value). Sixteen steps, each containing $n_r = 50$ observations, are made, so that the total number of observations is 900 as for the static design. In the first step of the sequential procedure 50 observations are taken at $x_1 = 1158$. The simulated response is $y_1 = 28$. This is larger than expected so an adjustment upward (recall that d is negative) of the estimate of ρ is necessary. The recursion (4) suggests the adjusted estimate to be $x_2 = 1158 - (-40000/50)(28/50 - 0.5) = 1206$, which is used as the next design point. The process continues similarly as reported in table 2. After 16 steps of the Robbins-Monro procedure the estimate of ρ is found to be 1268 with an estimated standard deviation of 267.3.

In this illustration the estimate of ρ is much closer to the true value when the sequential procedure is used. Also the precision is greatly in favour of the sequential procedure. The relative efficiency of the static ten point design compared to the nonadaptive sequential procedure being only 40%. However, a comparison with the optimal sequential design reveals that there is room for further improvement. Relative asymptotic efficiencies of the static design and the nonadaptive procedure compared to the optimal adaptive sequential design are 15% and 39%, respectively.

REFERENCES

Abdelbasit, K. M., and R. L. Plackett, "Experimental Design for Binary Data," *Journal of the American Statistical Association* 78 (1983), 90–98.

Anbar, Dan, "A Stochastic Newton-Raphson Method," *Journal of Statistical Planning and Inference* 2 (1978), 153–163.

Bishop, Richard C., and Thomas A. Heberlein, "Measuring Values of Extra-Market Goods, Are Indirect Methods

Biased?," *American Journal of Agricultural Economics* 61 (1979), 926–930.

Davis, R., "The Value of Big Game Hunting in a Private Forest," in *Transactions of the Twenty-ninth North-American Wildlife Conference* (Washington, D.C.: Wildlife Management Institute, 1964).

Fahrmeir, L., and H. Kaufman, "Consistency and Asymptotic Normality of the Maximum Likelihood Estimator in Generalized Linear Models," *Annals of Statistics* 13 (1988), 342–368.

Fedorov, V. V., *Theory of Optimal Experiments* (New York: Academic Press, 1972).

Ford, Ian, B. Torsney, and C. F. Jeff Wu, "The Use of a Canonical Form in the Construction of Locally Optimal Designs for Nonlinear Problems," *Journal of the Royal Statistical Society*, ser. B 54 (1992), 569–583.

Kriström, Bengt, "Valuing Environmental Benefits Using the Contingent Valuation Method: An Econometric Analysis," Doctoral thesis, University of Umeå (1990).

Lai, T. L., and H. Robbins, "Iterated Least Squares in Multiperiod Control," *Advances in Applied Mathematics* 3 (1981), 50–73.

McCullagh, Peter, and John A. Nelder, *Generalized Linear Models*, 2nd ed. (London: Chapman and Hall, 1989).

Minkin, Salomon, "Optimal Designs for Binary Data," *Journal of the American Statistical Association* 82 (1987), 1098–1103.

Mitchell R. C., and Richard T. Carson, *Using Surveys to Value Public Goods: The Contingent Valuation Method* (Washington, D.C.: Resources for the Future, 1989).

Robbins, H., and S. Monro, "A Stochastic Approximation Method," *Annals of Mathematical Statistics* 22 (1951), 400–407.

Sacks, Jerome, "Asymptotic Distribution of Stochastic Approximation Procedures," *Annals of Mathematical Statistics* 29 (1958), 373–405.

Silvey, S. D., *Optimal Design* (London: Chapman and Hall, 1980).

White, Lynda V., "The Optimal Design of Experiments for Estimation in Nonlinear Models," Doctoral thesis, Imperial College, University of London (1975).

Wu, C. F. Jeff, "Efficient Sequential Designs with Binary Data," *Journal of the American Statistical Association* 80 (1985), 974–984.

______, "Optimal Design for Percentile Estimation of a Quantal Response Curve," in *Optimal Design and Data Analysis of Experiments*, Y. Dodge, V. V. Fedorov, and H. P. Wynn (eds.) (Amsterdam: North-Holland, 1988).

FINITE SAMPLE BEHAVIOR OF TESTS FOR GROUPED HETEROSKEDASTICITY

James K. Binkley*

Abstract—When the regression error variance is constant within subsets of data but differs across the subsets, grouped heteroskedasticity exists. Most tests for grouped heteroskedasticity are based on an asymptotic χ^2 variable, and how well they perform in small samples is unknown. In this study the performance of several of these, along with Breusch-Pagan tests, is examined using Monte Carlo methods. Generally, it is found that the tests which tend to be the easiest to apply and use the fewest assumptions are at least as good as somewhat more elaborate procedures.

I. Introduction

Consider the model

$$Y_i = X_i\beta + e_i \quad i = 1, 2, \ldots, m \tag{1}$$

where the dimensions of Y_i, X_i, β, and e_i are $(n_i \times 1)$, $(n_i \times k)$, $(k \times 1)$, and $(n_i \times 1)$, respectively. There are m groups, each with n_i observations, all with the same β vector. We assume that the errors are normally distributed and uncorrelated either within or across groups, but the possibility exists that $\text{var}(e_i) \neq \text{var}(e_j)$, $i \neq j$. If the variances are unequal, this is termed the grouped heteroskedasticity model. In a recent paper, Binkley (1989) considered estimation of the variances in this model. In the usual case, estimation is preceded by testing, and that is the concern of this note. Relative to testing for other kinds of heteroskedasticity, this is straightforward, because the nature of the alternative hypothesis is unambiguous. However, only when $m = 2$ is an exact test available, the Goldfeld-Quandt test. For other cases, the most common approach is to employ a test based on the likelihood ratio principle. There are alternative ways to do this, an important factor being whether the prior information on β is employed. In addition, the Breusch-Pagan test can be adapted to the grouped heteroskedasticity case, as Breusch and Pagan pointed out in their original paper (1979).

The question of interest here is the small sample performance of these large sample tests. We report the results of a Monte Carlo study examining this performance. By and large, we do not find major differences among the tests examined, although we identify pitfalls associated with some procedures. There does not appear to be a significant advantage to use of a full likelihood ratio procedure, which requires iterative estimation, over some simpler alternatives. More generally, the study results provide little if any evidence of gains (and illustrates the possibility of loss) due to

Received for publication November 28, 1990. Revision accepted for publication May 31, 1991.

* Purdue University.

The author wishes to thank Deb Brown and the referees for helpful suggestions on an earlier version of this paper.

References

Experimental Design for Binary Data
K. M. Abdelbasit; R. L. Plackett
Journal of the American Statistical Association, Vol. 78, No. 381. (Mar., 1983), pp. 90-98.
Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28198303%2978%3A381%3C90%3AEDFBD%3E2.0.CO%3B2-7

Measuring Values of Extramarket Goods: Are Indirect Measures Biased?
Richard C. Bishop; Thomas A. Heberlein
American Journal of Agricultural Economics, Vol. 61, No. 5, Proceedings Issue. (Dec., 1979), pp. 926-930.
Stable URL:
http://links.jstor.org/sici?sici=0002-9092%28197912%2961%3A5%3C926%3AMVOEGA%3E2.0.CO%3B2-X

Consistency and Asymptotic Normality of the Maximum Likelihood Estimator in Generalized Linear Models
Ludwig Fahrmeir; Heinz Kaufmann
The Annals of Statistics, Vol. 13, No. 1. (Mar., 1985), pp. 342-368.
Stable URL:
http://links.jstor.org/sici?sici=0090-5364%28198503%2913%3A1%3C342%3ACAANOT%3E2.0.CO%3B2-Y

LINKED CITATIONS

Optimal Designs for Binary Data
Salomon Minkin
Journal of the American Statistical Association, Vol. 82, No. 400. (Dec., 1987), pp. 1098-1103.
Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28198712%2982%3A400%3C1098%3AODFBD%3E2.0.CO%3B2-6

A Stochastic Approximation Method
Herbert Robbins; Sutton Monro
The Annals of Mathematical Statistics, Vol. 22, No. 3. (Sep., 1951), pp. 400-407.
Stable URL:
http://links.jstor.org/sici?sici=0003-4851%28195109%2922%3A3%3C400%3AASAM%3E2.0.CO%3B2-W

Asymptotic Distribution of Stochastic Approximation Procedures
Jerome Sacks
The Annals of Mathematical Statistics, Vol. 29, No. 2. (Jun., 1958), pp. 373-405.
Stable URL:
http://links.jstor.org/sici?sici=0003-4851%28195806%2929%3A2%3C373%3AADOSAP%3E2.0.CO%3B2-U

Efficient Sequential Designs With Binary Data
C. F. Jeff Wu
Journal of the American Statistical Association, Vol. 80, No. 392. (Dec., 1985), pp. 974-984.
Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28198512%2980%3A392%3C974%3AESDWBD%3E2.0.CO%3B2-M