The Annals of Statistics
2001, Vol. 29, No. 4, 1050–1057

LOCALLY ASYMPTOTICALLY OPTIMAL DESIGNS FOR TESTING IN LOGISTIC REGRESSION

BY M.C. SPRUILL AND RENJIN TU

Georgia Institute of Technology and Columbus State University

Design measures maximizing local power of asymptotically uniformly most powerful (AUMP) tests about the value of logit$P$ outside the observation space are characterized.

1. Introduction. Suppose that for each $n < \infty$ and each collection $\{x_{n1}, \dots, x_{nn}\}$ of points in a compact set X an experiment can be performed which results in observations of the independent random variables $\{Y(x_{n,1}), \dots, Y(x_{n,n})\}$. The $Y(x)$ are Bernoulli with success probability

$$P(Y(x) = 1 \mid \theta, x) = e^{\theta' f(x)}/(1 + e^{\theta' f(x)}),$$

where $f'(x) = (f_0(x), \dots, f_{m-1}(x))$ is a vector of known continuous functions, and θ is an unknown m-vector. If, for example, $X = [a, b]$ is an interval and $c > b > a$, then to test for some fixed $k \in \{0, \dots, m-1\}$ a hypothesis such as

$$H_0 : (\theta' f(x))^{(k)}|_{x=c} \le C \quad \text{vs.} \quad H_a : (\theta' f(x))^{(k)}|_{x=c} > C, \tag{1.1}$$

about the value of the kth derivative of logit(P) at some point $x = c$ outside the interval X, where $C = (\theta_0{}' f(x))^{(k)}|_{x=c}$, what is a good choice of the x's? Such models arise in accelerated life testing and dose response studies where one would like to make decisions about the probabilities P or their rates of change outside the interval of observation.

The collection of x's determines the design probability measure ξ_n on X by $d\xi_n(x) = \#\{x_{n,i} = x\}/n$ so the question above is, "What is the optimal design for executing this test?" Actually we shall let n grow and ask what sequences of design measures are optimal for testing a hypothesis about the kth derivative of logit$P$, for k fixed, under a sequence of alternatives getting closer to the null. Optimality is in the sense of maximizing the local power of asymptotically uniformly most powerful (AUMP) tests. In the spirit of optimal approximate theory of designs we shall allow as design measures any probability measures on X.

Under widely applicable assumptions (A1–A2), the locally optimal designs to be derived here, assuming the parameter is close to θ_0, are as follows. For any given θ_0 and $k \in \{0, 1, \dots, m-1\}$ there are points $a \le x_1 < x_2 < \cdots x_m \le b$ which are the equioscillation points of the minimizer $h_k(x) - \sum_{j \ne k} d_j^* h_j(x)$ of $\|h_k - \sum_{j \ne k} d_j h_j\|^2_{[a,b],\infty}$, and the design ξ^0 which places masses at x_u, $u =$

Received April 1998; revised June 2001.
AMS 2000 *subject classifications*. Primary 62K05, 62F03; secondary 62F20, 62J12.
Key words and phrases. Pitman efficiency, Hoel-Levine designs, AUMP tests, Chebyshev system.

$1, \ldots, m$, proportional to $|F_{x_u}^{(k)}(c)|/\omega(x_u)$, is the unique locally asymptotically optimal design at θ_0 for testing the hypothesis (1.1). Here $h_j(x) = \omega(x) f_j(x)$, $\omega(x) = e^{\theta_0' f(x)/2}(1 + e^{\theta_0' f(x)})^{-1}$ is the standard deviation of Y at x under the local parameter value θ_0, and $F_{x_i}(x)$ are the Lagrange interpolation polynomials to the points x_i in terms of the f's.

To describe the tests and optimal design criterion more explicitly, we begin more generally with a sequence of rational designs ξ_n, probability measures on X placing integer weights $nd\xi_n(x_{n,j})$ at finite but not necessarily fixed collections of points $x_{n,j}$ in X, and converging to an arbitrary probability measure ξ on X. These are designs which correspond to implementable experiments. In accordance with assumption (A2) below, let $f_k^{(k)}(c) > 0$. Furthermore, let

$$\sigma_{\theta_0}^2(\xi) = \min_d \int_a^b (f_k(x) - \sum_{j \neq k} d_j f_j(x))^2 \omega^2(x) d\xi(x), \tag{1.2}$$

$d_j(\xi)$ be the d_j's in a vector d minimizing it, $\bar{Y}(x)$ be the average of the observations taken at x, and $\mu(x) = \omega(x) e^{\theta_0' f(x)/2}$ be the mean of Y at x under the local parameter value θ_0. The tests which reject H_0 of (1.1) if the test statistics

$$\tau_n = \sqrt{n} \int_a^b \left(f_k(x) - \sum_{i \neq k} d_i(\xi) f_i(x) \right) \left[\bar{Y}(x) - \mu(x) \right] d\xi_n(x) / \sigma_{\theta_0}(\xi)$$

exceed z_α are AUMP level α against the sequence of alternatives θ_n which can be written under our assumptions (A1–A2) as $\theta_n = \theta_0 + e_{k+1}/\sqrt{n} + \gamma/\sqrt{n} + o(n^{-1/2})$. Here e_{k+1} is the vector with zeros in all coordinates except a 1 in coordinate $k+1$, and γ is in the nuisance parameter space.

Justification of the claim that the sequence of designs and test statistics is optimal comes from an extension to the design setting of the results of Choi, Hall and Schick (1996) on AUMP tests. It can be shown that for a sequence of designs ξ_n which converge weakly to ξ and for parameter values θ_n converging to θ_0 above, the Pitman efficacy of the AUMP tests at θ_0 is simply $\sigma_{\theta_0}^2(\xi)$, the expression in (1.2). If the limiting design is chosen to maximize the expression (1.2) over all probability measures ξ on X, then any corresponding sequence of designs is optimal. We show under (A1-A2) that the design ξ^0 described above is the unique maximizer of $\sigma_{\theta_0}^2(\xi)$ and that $d_j^* = d_j(\xi^0)$.

Little of the optimal design literature deals with hypothesis testing and none with our problem directly. Familiar formulas reminiscent of D_1 optimal designs arise however. There are related design problems of estimation for non-linear models. See, for example, Chaloner and Verdinelli (1995), Dette and Sahm (1997) and Heiligers (1996). Using the technique here applied, for example, to the multihit dose-response model yields the same design found by Hoel and Jennrich (1979) who studied estimation.

2. Main result. We shall find the optimal design under the following conditions on the functions f_j.

(A1) $f_0, \ldots, f_{m-1} \in C^{m-1}[a, c]$ constitute a Chebyshev system on $[a, c]$ and for each $k \in \{0, \ldots, m-1\}$ any function $\sum_{j=0}^{m-1} a_j f_j^{(k)}(x)$ not identically zero has at most $m-k-1$ zeros in $[a, c]$.

Condition (A1) is satisfied by the familiar ordinary polynomial powers $f_j(x) = x^j$, $j = 0, \ldots, m-1$, on any interval. It can be shown that if $f_0, \ldots, f_{m-1}$ satisfy (A1) and $k \in \{0, \ldots, m-1\}$ then for some index j, $f_j^{(k)}(c) \neq 0$. So given $k \in \{0, \ldots, m-1\}$ we can and shall, by renumbering and with a linear transformation, if necessary, assume that

(A2) $f_k^{(k)}(c) > 0$ while $f_i^{(k)}(c) = 0$ for $i \neq k$.

For $a \leq x_1 < \cdots < x_m \leq c$, let Δ be the determinant of the matrix whose entry in the ith row and jth column is $f_{i-1}(x_j)$ and $\Delta_u(x)$ denote the determinant of the matrix just described, but altered to have the entry $f_{i-1}(x)$ in the ith row of the uth column, $i = 1, \ldots, m$. As a consequence of the f's forming a Chebyshev system on $[a, c]$, the Lagrange interpolation polynomials are $F_{x_u}(x) = \Delta_u(x)/\Delta$, for $u = 1, \ldots, m$. For any x_j the function $F_{x_j}(x)$ has at most $m-1$ zeros and since $F_{x_j}(x_i) = \delta_{ij}$, it follows that the sign of $F_{x_j}(x)$ at c is the same as it is at $x_m + 0$. One can prove that for $a \leq x_1 < \cdots < x_m \leq b < c$ and $k \in \{0, \ldots, m-1\}$, if the assumption (A1) is satisfied, then $\operatorname{sgn}(F_{x_u}^{(k)}(c)) = (-1)^{m-u}$.

Noting that $\omega(x) > 0$ on $[a, c]$ and $h_j(x) = \omega(x) f_j(x)$, by Karlin and Studden (1966a) it follows that the collection h_i is also a Chebyshev system on $[a, c]$. Let $a \leq x_1 < \cdots < x_m \leq b$ and let $H_{x_i}(x)$ be the Lagrange interpolation polynomials to the given points in terms of the h's. The proof of Lemma 2.1 of Spruill (1987) shows that the collection of functions h_i, $i \neq k$, is also a Chebyshev system on $[a, b]$. Finally, the polynomials in f are determined by their values at m points and $H_{x_i}(x)/\omega(x)$ being such a polynomial entails $H_{x_i}(x)/\omega(x) = F_{x_i}(x)/\omega(x_i)$.

THEOREM 2.1. *In logistic regression, under the assumptions* (A1–A2), *for any given* θ_0 *and* $k \in \{0, 1, \ldots, m-1\}$ *there are points* $a \leq x_1 < x_2 < \cdots x_m \leq b$ *which are the equioscillation points of the minimizer* $h_k(x) - \sum_{j \neq k} d_j^* h_j(x)$ *of* $\|h_k - \sum_{j \neq k} d_j h_j\|_{[a,b],\infty}^2$ *and the design* ξ^0 *which places masses proportional to* $|F_{x_u}^{(k)}(c)|/\omega(x_u)$ *at* x_u, $u = 1, \ldots, m$, *is the unique locally asymptotically optimal design at* θ_0 *for testing the hypothesis* (1.1).

PROOF. Since $\sigma_{\theta_0}^2(\xi) = \min_d \int_a^b (h_k(x) - \sum_{j \neq k} d_j h_j(x))^2 d\xi(x)$, using a minimax theorem [see, e.g., Karlin and Studden (1966b), page 807] it can be shown that

$$\max_{\xi} \sigma_{\theta_0}^2(\xi) = \min_{d} \|h_k(x) - \sum_{j \neq k} d_j h_j(x)\|_{[a,b],\infty}^2. \tag{2.1}$$

Since for fixed k, h_i, $i \neq k$, form a Chebyshev system and h_k is continuous, the minimizing d_j^* result in the equioscillation of $h_k - \sum_{j \neq k} d_j^* h_j$ at m points

$a \le x_1 < \cdots < x_m \le b$. Therefore, denoting the right hand side of (2.1) by B^2, we get, for some u, $h_k(x)-\sum_{j\neq k} d_j^* h_j(x) = B\sum_{j=1}^m (-1)^{j+u} H_{x_j}(x)$. Noting that $f_k(x) - \sum_{j\neq k} d_j^* f_j(x) = B\sum_{j=1}^m (-1)^{j+u} F_{x_j}(x)/\omega(x_j)$, taking k derivatives on both sides, and setting $x = c$ one gets

$$f_k^{(k)}(c) = B\sum_{j=1}^m (-1)^{j+u} F_{x_j}^{(k)}(c)/\omega(x_j) = \pm BQ \tag{2.2}$$

where $Q = \sum_{i=1}^m |F_{x_i}^{(k)}(c)|/\omega(x_i)$. We claim that the design ξ^0 which places masses proportional to $|F_{x_i}^{(k)}(c)|/\omega(x_i)$ at the equioscillation points is optimal. To check this we simply compute

$$\min_d \int_a^b \left(f_k(x) - \sum_{j\neq k} d_j f_j(x) \right)^2 \omega^2(x) d\xi^0(x) \tag{2.3}$$

and show that it equals B^2. Toward that end we employ a Lagrange multiplier λ and use the fact that any function $f_k(x) - \sum_{j\neq k} d_j f_j(x)$ can be expressed in terms of the Lagrange interpolation polynomials $F_{x_i}(x)$ so that (2.3) is the same as

$$\min_g \left\{ \int_a^b \left(\sum_{j=1}^m g_j F_{x_j}(x) \right)^2 \omega^2(x) d\xi^0(x) : \sum_{j=1}^m g_j F_{x_j}^{(k)}(c) = f_k^{(k)}(c) \right\}.$$

Thus $(\lambda/2)Q \operatorname{sgn}(F_{x_j}^{(k)}(c))/\omega(x_j) = g_j^*$ and the minimum value is

$$Q^{-1} Q^2 (\lambda/2)^2 \sum_{j=1}^m |F_{x_j}^{(k)}(c)|/\omega(x_j) = Q^2(\lambda/2)^2 = (f_k^{(k)}(c))^2/Q^2$$

which we recognize from (2.2) as B^2. This establishes the optimality of ξ^0. To see uniqueness, for any design on s points of $[a, b]$, $s < m$, we show

$$\min_d \sum_{i=1}^s \left(f_k(x_i) - \sum_{j\neq k} d_j f_j(x_i) \right)^2 \omega^2(x_i)\xi(x_i) = 0.$$

Since $s < m$ consider the m points $y_i = x_i$, $i = 1, \ldots, s$ and augment the collection with any other points y_i, $i = s+1, \ldots, m$ in $[a, b]$. Then, setting $a_i = 0$, $i = 1, \ldots, s$ and choosing the remaining a_i to achieve $(\sum_{j=1}^m a_j F_{y_j}^{(k)}(c)) = f_k^{(k)}(c)$, we have $\sum_{i=1}^s (\sum_{j=1}^m a_j F_{y_j}(x_i))^2 \omega^2(x_i)\xi(x_i) = 0$ and consequently, any design on fewer than m points is suboptimal. Now let ξ' be any optimal design. Because of (2.1), the support of ξ' must be contained in a set of points $x \in [a, b]$ at which a minimizer $h_k - \sum_{j\neq k} d_j^* h_j$ attains its extreme values whose magnitudes are the minimum in (2.1). The support must also contain at least m points. By Theorem 3.1 of Spruill (1984) (taking $\eta = \infty$ there) for example, this minimizer is unique, so we conclude that the support points of ξ' must coincide with those of ξ^0. Our previous argument in this proof shows that the masses are determined once the support points are known, so $\xi^0 = \xi'$. □

The prescription for finding the optimal design is conceptually simple. The equioscillation points are identified by solving the problem (2.1) from which the prescribed masses at the equioscillation points are computed according to Theorem 2.1.

EXAMPLE 2.2. Take $c = 1.2$, $X = [-1, 1]$, $m = 5$, $k = 0$, $f'(x) = (1, x - c, (x-c)^2, (x-c)^3, (x-c)^4)$ and $\theta_0{}' = (1, 1, 1, 1, 1)$. Thus, we are testing the null hypothesis that the probability of a favorable response at $x = 1.2$ is no greater than $e/(1+e)$ against the alternative that it is greater. The design was found numerically by replacing the interval $[-1, 1]$ by a grid of 101 points and employing an algorithm of Barrodale and Phillips (1975) to find the generalized polynomial solving (2.1) above. The support points of the optimal design are $x_1 = -0.56$, $x_2 = -0.1$, $x_3 = 0.36$, $x_4 = 0.80$ and $x_5 = 1$ and $\sigma_{\theta_0}(\xi^0) = 0.043876$. The corresponding optimal masses are 0.109, 0.078, 0.145, 0.334 and 0.334. The polynomial $p_0(x) = 1 - \sum_{j=2}^m d_j^*(x-c)^{j-1}$ satisfying $\omega(x)p_0(x) = h_1(x) - \sum_{j=2}^m d_j^* h_j(x)$ is depicted in Figure 1 below. There are no polynomials p of degree $m-1$ passing through $(c, 1)$ satisfying $-B/\omega(x) \le p(x) \le B/\omega(x)$ for all $x \in [-1, 1]$ unless $B \ge \sigma_{\theta_0}(\xi^0)$. There is one in the case of equality, and p_0 is proportional to it.

Maximization of the Pitman efficacy $\sigma_\theta^2(\xi)$ over designs ξ can also be carried out for some multiple covariate cases. Let $X = \{-1, 1\}^m$ and $\operatorname{logit} P(Y = 1|\alpha, \beta, x) = \alpha + \beta' x$, where α is an unknown scalar and β is an unknown m-vector. Suppose it is desired to test for some fixed index j, $H_j : \beta_j = 0$ against $K_j : \beta_j \ne 0$. For a vector v in R^m let $v^{(j)}$ denote the vector in

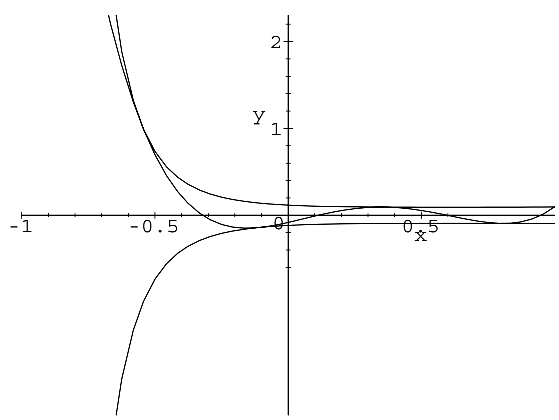

FIG. 1.

R^{m-1} formed from the coordinates of v except the jth one. Utilizing again the interplay provided by a minimax theorem one can prove the following.

LEMMA 2.3. *If S is any subset of $\{x^{(j)} : x \in X\}$ satisfying*:

(i) *$s \in S$ implies $-s \in S$ and*
(ii) *$c'^{(j)}s = 0$ for all s in S entails $c^{(j)} = 0$ then setting $W = \{x \in X : x^{(j)} \in S\}$, the design ξ^0 which places equal masses on the set W is locally optimal for testing $\beta_j = 0$ at $(\alpha_0, 0, \dots, 0)$.*

It can be shown under the assumptions of Lemma 2.3 that for testing any individual coordinate, an optimal design can be found as a uniform design on $4(m-1)$ of the 2^m points in X.

EXAMPLE 2.4. Take $m = 4$, $X = \{-1, 1\}^4$, and $\theta_0 = (\alpha_0, 0, 0, 0, 0)$. It can be checked that for testing $H_1 : \beta_1 = 0$ against $K_1 : \beta_1 \neq 0$ the conditions of Lemma 2.3 are satisfied for the set $S = \{(1,1,1)', (-1,1,1)', (1,-1,1)', (-1,-1,-1)', (1,-1,-1)', (-1,1,-1)'\}$ and the design placing masses $\frac{1}{12}$ at each point of the set $W = \{x \in X : x^{(1)} \in S\}$ is locally optimal.

3. Robustness. Robustness issues include sensitivity to the local parameter, the extrapolation point c, and sample size.

How well does a test based on a design locally optimal at some θ_0 behave at other θ? For example, when $f'(x) = (1, x-c, (x-c)^2, \dots, (x-c)^{m-1})$, $x \in [-1, 1]$ and $\theta_{01} = \log[p^0/(1-p^0)]$, $\theta_{02} = \cdots = \theta_{0m} = 0$, it follows from its expression through (2.1) that the Hoel-Levine design placing its masses at the points

$$x_j = -\cos\left(\frac{(j-1)\pi}{m-1}\right), \qquad j = 1, \dots, m, \tag{3.1}$$

is locally optimal. The performance of a design ξ at arbitrary θ is given by

$$\sigma_\theta^2(\xi) = \min_d \int_a^b \left(1 - \sum_{j=2}^m d_j (x-c)^{j-1}\right)^2 \omega^2(x) d\xi(x) = M_{11} - M_{12} M_{22}^{-1} M_{21}$$

where $M_{11} = \int_a^b \omega^2(x) d\xi(x)$, M has ijthF entry $\int_a^b (x-c)^{i+j-2} \omega^2(x) d\xi(x)$, $i, j \in \{1, \dots, m\}$, and M_{ij} are the submatrices in the usual partition of M. For $m = 3$ and θ on a grid ranging over plus or minus 0.2 in each free coordinate about θ^0 it was found that $\sigma_\theta^2(\xi^{\theta_0})/\sigma_\theta^2(\xi^\theta) \geq 0.9837$, where ξ^θ is optimal at θ. Little is lost here by using the simple design on the points (3.1).

More generally, one can prove the following based upon the proof of Theorem 2.1.

LEMMA 3.1. *Let the design ξ^0 be optimal at θ_0 and have support $x_1^0 < \cdots < x_m^0$ and θ be arbitrary. Then $\sigma_\theta^2(\xi^0) \geq \sigma_{\theta_0}^2(\xi^0) \min_i \omega_\theta^2(x_i^0)/\omega_{\theta_0}^2(x_i^0)$.*

TABLE 1
Power at $(\theta_1, 2, -1)$

θ_1	1.0	2.0	2.5	3.0
design				
Uniform, $n = 10$	.0289	.0452	.0411	.0323
Optimal, $n = 10$	.0289	.0478	.0495	.0622
Uniform, $n = 15$	.0429	.1024	.0968	.0697
Optimal, $n = 15$	.0429	.0912	.0997	.1200

LEMMA 3.2. *Let the design ξ^c be optimal at θ^0 with $k = 0$ and $c > b$ and have support $x_1 < \cdots < x_m$. Then for $c' > b$, $\sigma_{c'}(\xi^c) \geq \sigma_c(\xi^c) \min_i [F_{x_i}(c)/F_{x_i}(c')]^2$.*

The latter relates how well a test based on a design locally optimal at some θ_0 for testing the hypothesis (1.1) about $\theta' f(c)$ does at c'.

To evaluate the small sample performance a simulation was run to compare the optimal design for $c = 1.2$, $X = [-1, 1]$, $f'(x) = (1, x - c, (x - c)^2)$ and ${\theta_0}' = (1, 2, -1)$ with the uniform design on $x = -1, 0, 1$.

Table 1 shows the powers of the AUMP test, reject $H_0 : p \leq 1/(1 + e^{-1})$ if $\tau_n > 1.645$ and the corresponding AUMP tests of the same size based on the uniform design. Figures are computed based on 10^5 runs for each case. Pitman efficiency, appropriate in large samples and locally, of the optimal to the uniform is 5.43 here.

Comparisons with the UMPU tests are trivial since they reject H_0 with probability 0.03, regardless of the observations. It can be shown that the only cases in which UMPU tests are non-trivial for $n = 10$ are for $c = 2$ or 3.

Acknowledgments. The authors thank the referees and editors whose careful review of our paper resulted in many improvements.

REFERENCES

BARRODALE, I. and PHILLIPS, C. (1975). ALGORITHM 495. Solution of an overdetermined system of linear equations in the Chebyshev norm. *ACM Trans. Math. Software* **1** 264–270.

CHALONER, K. and VERDINELLI, I. (1995). Bayesian experimental design: a review. *Statist. Science* **10** 273–304.

CHOI, S., HALL, W. J. and SCHICK, A. (1996). Asymptotically uniformly most powerful tests in parametric and semiparametric models. *Ann. Statist.* **24** 841–861.

DETTE, H. and SAHM, M. (1997). Standardized optimal designs for binary response experiments. *South African Statist. J.* **31** 271–298.

HEILIGERS, B. (1996). Computing E-optimal polynomial regression designs. *J. Statist. Plann. Inference* **55** 219–233.

HOEL, P. G. and JENNRICH, R. I. (1979). Optimal designs for dose response experiments in cancer research. *Biometrika* **66** 307–316.

KARLIN, S. and STUDDEN, W. J. (1966a). *Tchebycheff Systems with Applications in Analysis and Statistics*. Interscience, New York.

KARLIN, S. and STUDDEN, W. J. (1966b). Optimal experimental designs. *Ann. Math. Statist.* **37** 783–816.

SPRUILL, M. C. (1987). Optimal extrapolation of derivatives. *Metrika* **34** 45–60.

SPRUILL, M. C. (1984). Optimal designs for minimax extrapolation. *J. Multivariate Anal.* **15** 52–62.

SCHOOL OF MATHEMATICS
GEORGIA INSTITUTE OF TECHNOLOGY
ATLANTA, GEORGIA 30332-0160
E-mail: spruill@math.gatech.edu

DEPARTMENT OF MATHEMATICS
COLUMBUS STATE UNIVERSITY
COLUMBUS, GEORGIA 31907
E-mail: Tu_Renjin@colstate.edu