The Annals of Statistics
1999, Vol. 27, No. 4, 1155–1163

STRONG CONSISTENCY OF MAXIMUM QUASI-LIKELIHOOD ESTIMATORS IN GENERALIZED LINEAR MODELS WITH FIXED AND ADAPTIVE DESIGNS

BY KANI CHEN,[1] INCHI HU[1] AND ZHILIANG YING[2]

Hong Kong University of Science and Technology and Rutgers University

Strong consistency for maximum quasi-likelihood estimators of regression parameters in generalized linear regression models is studied. Results parallel to the elegant work of Lai, Robbins and Wei and Lai and Wei on least squares estimation under both fixed and adaptive designs are obtained. Let $y_1, \dots, y_n$ and $x_1, \dots, x_n$ be the observed responses and their corresponding design points ($p \times 1$ vectors), respectively. For fixed designs, it is shown that if the minimum eigenvalue of $\sum x_i x_i'$ goes to infinity, then the maximum quasi-likelihood estimator for the regression parameter vector is strongly consistent. For adaptive designs, it is shown that a sufficient condition for strong consistency to hold is that the ratio of the minimum eigenvalue of $\sum x_i x_i'$ to the logarithm of the maximum eigenvalues goes to infinity. Use of the results for the adaptive design case in quantal response experiments is also discussed.

1. Introduction. Since the fundamental work of Nelder and Wedderburn (1972), there has been continued interest in the development of theory and methodology related to generalized linear models. Wedderburn (1974) noted that many likelihood-based procedures are still valid provided that the mean and variance functions are correctly specified. In that connection, the concept of quasi-likelihood function was introduced. We refer to McCullagh and Nelder (1989) for a comprehensive account of generalized linear models and (quasi-)likelihood-based inference procedures.

An important theoretical issue regarding the maximum quasi-likelihood estimator of a regression parameter is the following: under what condition(s) will the estimator converge to the true parameter? In the case of linear regression models, certain minimum conditions have been found that ensure strong consistency for least squares estimators, under both fixed and adaptive designs. In particular, for a fixed design, Lai, Robbins and Wei (1979) showed that a necessary and sufficient condition for strong consistency of least squares estimators is that the minimum eigenvalue of the "information matrix" goes to ∞ as sample size $n \to \infty$. Note that weak consistency under

Received April 1996; revised August 1998.
[1]Supported in part by Hong Kong RGC Grants HKUST DAG95/96.SC28 and 292/94H.
[2]Supported in part by the National Science Foundation DMS-96-26750 and DMS-99-71791 and the National Security Agency MDA 904-98-1-0027.
AMS 1991 *subject classifications*. Primary 62F12; secondary 62J05.
Key words and phrases. Generalized linear models, maximum quasi-likelihood estimator, exponential family, strong consistency, martingale difference, fixed design, adaptive design.

the same condition is a somewhat trivial matter since it can be argued from the Chebychev inequality and a covariance matrix calculation. For an adaptive design, Lai and Wei (1982) showed via a counterexample that such a condition is not sufficient and they further proved that if the ratio of the minimum eigenvalue to the logarithm of the maximum eigenvalue goes to ∞, then the least squares estimator is strongly consistent. Note also that with an adaptive design, even weak consistency is quite challenging.

For generalized linear models with *fixed* designs, conditions to ensure convergence of maximum quasi-likelihood estimators of regression parameters have been studied by many authors: see Haberman (1977), Anderson (1980), Nordberg (1980) and Fahrmeir and Kaufmann (1985), among others. In the aforementioned papers and related references, conditions assumed for strong consistency are typically more than the minimal assumption of the divergence of the information matrix in the neighborhood of the true parameter. For example, strong consistency obtained in Fahrmeir and Kaufmann (1985) requires the boundedness of the ratio of the largest eigenvalue to the power of $1/2 + \delta$ over the smallest eigenvalue of the information matrices for some $\delta > 0$.

Adaptive design in linear models arises frequently in econometric and engineering control, where strong consistency of system parameter estimators is of great concern: see Anderson and Taylor (1979), Åström and Wittenmark (1973), Box and Jenkins (1970) and Moore (1978). Besides extensive use of adaptive designs in linear systems, where ill-conditioned design matrices arise naturally, there have also been important applications of such designs in nonlinear systems. Dixon and Mood (1948), in their analysis of quantal responses in sensitivity experiments, proposed what they called the up-and-down method for adjusting the level of stimulus at which the probability of a response is approximately at a targeted value. A fundamental breakthrough for adaptively approximating an optimal design point was due to Robbins and Monro (1951), who coined the now well-known stochastic approximation method. Cochran and Davis (1965) and Finney (1978) discussed how such adaptive designs might be used for obtaining efficient sequential sampling schemes in biological assays. Lord (1971a, b) showed that the Robbins–Monro approach may be used to design tailored tests in educational measurement. Wetherill (1963) and Wu (1985) contain comprehensive discussions on the up-to-date developments in the area and propose their own approaches.

In this paper we are concerned with the issue of strong consistency for maximum quasi-likelihood estimators of regression parameters in generalized linear models. The key difference of a generalized linear model from the usual linear model is obviously the nonlinearity in the link function. We handle this by establishing a local inverse function theorem. It turns out that the inverse function theorem also handles the global behavior of the estimator when we utilize convexity of the quasi-likelihood function and certain sharp probability bounds available in the literature. In doing so, we show that the minimum conditions of Lai, Robbins and Wei (1979) for fixed designs and Lai and Wei (1982) for stochastic designs in conjunction with an additional assumption, essentially to offset the nonlinearity, also ensure strong consistency for maximum quasi-likelihood estimators.

Generalized linear models with fixed designs are dealt with in Section 2, where strong consistency of maximum quasi-likelihood estimators is established under certain minimum conditions. A parallel result in the case of adaptive designs is presented in Section 3, where the relevance of such a result to quantal response experiments is also discussed.

2. Strong consistency for maximum quasi-likelihood estimators in the fixed design case. To fix notation, let (y_i, x_i'), $i = 1, \dots, n$ be n pairs of responses and design vectors. By a fixed design, we mean that either the x_i are nonrandom p-dimensional vectors or the discussion is valid when conditioning on the x_i, which may then be regarded as nonrandom. Thus throughout this section, the x_i are assumed to be nonrandom p-vectors.

Let μ be a continuously differentiable function such that $\dot{\mu}(t) = d\mu(t)/dt > 0$ for all t. A generalized linear model with μ as the link function specifies the regression relation between response y_i and design vector x_i through

$$E(y_i) = \mu(\beta_0' x_i), \tag{2.1}$$

where we use β_0 to denote the true value of regression parameter vector β. The maximum quasi-likelihood estimator, denoted by $\hat{\beta}_n$, will be the solution to

$$\sum_{i=1}^{n} x_i \left[y_i - \mu\left(\hat{\beta}_n' x_i \right) \right] = 0. \tag{2.2}$$

Note that if such $\hat{\beta}_n$ exists, it must be unique. This is easily seen from the positivity of $\dot{\mu}$.

Let $\varepsilon_i = y_i - \mu(\beta_0' x_i)$. The following two conditions will be used to establish strong consistency for $\hat{\beta}_n$:

(C1) The minimum eigenvalue of $\sum_{i=1}^{n} x_i x_i'$ goes to ∞ as $n \to \infty$.
(C2) $\sum_{i=1}^{\infty} c_i \varepsilon_i$ converges a.s. for any sequence of constants $\{c_i\}$ satisfying $\sum_{i=1}^{\infty} c_i^2 < \infty$.

Condition (C1) is certainly a minimum requirement for $\hat{\beta}_n$ to be consistent. Condition (C2) is also extremely mild. In fact, in view of the Khintchine–Kolmogorov convergence theorem [Chow and Teicher (1988), page 113] for the sum of independent zero-mean random variables, (C2) is implied by either of the following two conditions:

(C2′) The ε_i's are independent and $\sup_i E\varepsilon_i^2 < \infty$.
(C2″) The x_i's are bounded and y_i's are independent with $\mathrm{Var}(y_i) = \sigma^2 \dot{\mu}(\beta_0' x_i)$ for some $\sigma^2 > 0$.

It is clear that (C2″) implies (C2′). Since condition (C2) does not require the observations be independent, it can be used for highly stratified observations such as those discussed in Liang and Zeger (1986). Condition (C2′) handles the situation in which only the link, not the variance function, is specified. On the other hand, (C2″) implies that the variance function is also correctly specified.

Lai, Robbins and Wei (1979) proved that the least squares estimator for linear regression models is strongly consistent under (C1) and (C2). The following theorem extends their result to generalized linear models.

THEOREM 1. *For a generalized linear model specified by* (2.1) *with a fixed design*, *suppose that*

$$\sup_i \|x_i\| < \infty \tag{2.3}$$

and that (C1) *and* (C2) *are satisfied. Then* $\hat{\beta}_n \to \beta_0$ *a.s. In fact*,

$$\|\hat{\beta}_n - \beta_0\| = o\left(\left\{\left[\log \lambda_{\min}(n)\right]^{1+\delta}/\lambda_{\min}(n)\right\}^{1/2}\right) \quad a.s. \tag{2.4}$$

for any $\delta > 0$, *where* $\lambda_{\min}(n)$ *denotes the minimum eigenvalue of* $\sum_{i=1}^n x_i x_i'$.

REMARK 1. That condition (2.3) is required is due to the nonlinearity of μ. To see why, we give a counterexample to show that $\hat{\beta}_n$ may not be consistent when (2.3) is dropped. Consider a logistic regression model specified by $P(y_i = 1) = (1 + \exp(-\beta_0 x_i))^{-1}$ and $x_i = i$. Set $\beta_0 = 1$. Clearly both (C1) and (C2) are satisfied. Choose k_0 to be large enough so that

$$\sum_{i=k_0+1}^{\infty} \frac{i}{1 + e^{i/2}} < \frac{1}{2} \sum_{i=1}^{k_0} i. \tag{2.5}$$

Define the event $A = \{y_i = 0,\ i \le k_0 \text{ and } y_i = 1,\ i \ge k_0 + 1\}$. Then from the Borel–Cantelli lemma it follows that $P(A) > 0$. Furthermore, on A, $\hat{\beta}_n$ satisfies

$$\sum_{i=k_0+1}^{n} \frac{i}{1 + \exp(\hat{\beta}_n i)} = \sum_{i=1}^{k_0} \frac{i \exp(\hat{\beta}_n i)}{1 + \exp(\hat{\beta}_n i)}. \tag{2.6}$$

From (2.5) and (2.6), it follows that $\hat{\beta}_n \le 1/2$ on A. Hence $\hat{\beta}_n$ cannot be consistent.

REMARK 2. If, however, $\dot{\mu}$ is bounded away from 0, as in the case of linear regression, then it is easily seen from our proof of Theorem 1 that $\hat{\beta}_n$ is still strongly consistent without assuming (2.3).

PROOF OF THEOREM 1. The proof is built upon the key result of Lai, Robbins and Wei (1979) and an inverse function theorem, stated as Lemma A in the Appendix. Define

$$\varepsilon_i = y_i - \mu(\beta_0' x_i), \tag{2.7}$$

$$G_n(\beta) = \sum_{i=1}^{n} x_i [\mu(\beta' x_i) - \mu(\beta_0' x_i)]. \tag{2.8}$$

Because of (C1) and (C2), we can apply Theorem 1 of Lai, Robbins and Wei (1979) to get, for any $\delta > 0$,

$$\begin{aligned} a_n &\equiv \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1} \sum_{i=1}^{n} x_i \varepsilon_i \\ &= o\left(\max_{1 \le j \le p} \left\{ v_{jj}^{(n)} |\log v_{jj}^{(n)}|^{1+\delta} \right\}^{1/2} \right) \\ &= o\left(\left\{ [\log \lambda_{\min}(n)]^{1+\delta} / \lambda_{\min}(n) \right\}^{1/2} \right) \quad a.s., \end{aligned} \tag{2.9}$$

where $v_{jj}^{(n)}$ is the jth diagonal element of the matrix $(\sum_{i=1}^{n} x_i x_i')^{-1}$ and the last equality follows from the fact that $v_{jj}^{(n)} \le \lambda_{\min}^{-1}(n)$ for all j. Note that $(\sum_{i=1}^{n} x_i x_i')^{-1}$ exists for all large n since $\lambda_{\min}(n) \to \infty$.

By the mean-value theorem, for any p-vectors β_1 and β_2,

$$G_n(\beta_1) - G_n(\beta_2) = \sum_{i=1}^{n} \dot{\mu}\left(\tilde{\beta}_i' x_i \right) x_i x_i' (\beta_1 - \beta_2), \tag{2.10}$$

where for each i, $\tilde{\beta}_i$ lies on the line segment between β_1 and β_2. Define

$$\tilde{G}_n(\beta) = \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1} G_n(\beta). \tag{2.11}$$

Since $\dot{\mu}(t) > 0$ for any t, (2.10) implies that G_n is an injection from R^p to R^p. Thus, $\tilde{G}_n$ is also an injection from R^p to R^p and $\tilde{G}_n(\beta_0) = 0$.

For any $\eta > 0$, let $m_\eta = \inf_{i, \|\beta - \beta_0\| \le \eta} \dot{\mu}(\beta' x_i)$. The boundedness of the x_i and the continuity of $\dot{\mu}$ entail $m_\eta > 0$. It follows from (2.10) and (2.11) that, for any β such that $\|\beta - \beta_0\| \le \eta$,

$$\begin{aligned} \left\| \tilde{G}_n(\beta) \right\|^2 &= \left\| \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1} (G_n(\beta) - G_n(\beta_0)) \right\|^2 \\ &= (\beta - \beta_0)' \sum_{i=1}^{n} \dot{\mu}\left(\tilde{\beta}_i' x_i \right) x_i x_i' \left(\sum_{i=1}^{n} x_i x_i' \right)^{-2} \\ &\qquad \times \sum_{i=1}^{n} \dot{\mu}\left(\tilde{\beta}_i' x_i \right) x_i x_i' (\beta - \beta_0) \end{aligned} \tag{2.12}$$

$$\geq (\beta-\beta_0)'\sum_{i=1}^{n}\dot{\mu}(\tilde{\beta}_i'x_i)x_ix_i'\left(\sum_{i=1}^{n}\frac{\dot{\mu}(\tilde{\beta}_i'x_i)}{m_\eta}x_ix_i'\right)^{-2}$$
$$\times\sum_{i=1}^{n}\dot{\mu}(\tilde{\beta}_i'x_i)x_ix_i'(\beta-\beta_0)$$
$$= m_\eta^2\|\beta-\beta_0\|^2.$$

In particular, for all β such that $\|\beta-\beta_0\|=\eta$,

$$\left\|\tilde{G}_n(\beta)\right\|^2 \geq m_\eta^2\eta^2 > 0. \tag{2.13}$$

Let $\gamma = m_\eta\eta$. By applying Lemma A in the Appendix, we get

$$\tilde{G}_n^{-1}(\{b: \|b\|\leq\gamma\}) \subseteq \{\beta: \|\beta-\beta_0\|\leq\eta\}. \tag{2.14}$$

We now prove the existence and consistency of $\hat{\beta}_n$. By definition, $\|a_n\|\to 0$ a.s. Therefore, (2.14) implies that, for all large n, $\tilde{G}_n^{-1}(a_n)$ is well defined and lies in $\{\beta: \|\beta-\beta_0)\|\leq\eta\}$. However,

$$\sum_{i=1}^{n}\left[\mu\left(x_i'\tilde{G}_n^{-1}(a_n)\right)-y_i\right]x_i = \sum_{i=1}^{n}x_ix_i'\left[\tilde{G}_n\left(\tilde{G}_n^{-1}(a_n)\right)-a_n\right] = 0. \tag{2.15}$$

So, $\hat{\beta}_n = \tilde{G}_n^{-1}(a_n)$ exists and $\|\hat{\beta}_n-\beta_0\|\leq\eta$ for all large n. Since η can be chosen to be arbitrarily small, $\hat{\beta}_n\to\beta_0$ a.s.

To derive the rate, that is, (2.4), we can set $\eta=\eta_n=\varepsilon\max_{1\leq j\leq p}\{v_{jj}^{(n)}|\log v_{jj}^{(n)}|^{1+\delta}\}^{1/2}$, where ε is any fixed positive number. By (2.9) we still have $\tilde{G}_n(\hat{\beta}_n)=a_n=o(\eta_n)$ a.s. Thus, also in view of (2.14), we have $\|\hat{\beta}_n-\beta_0\|\leq\eta_n$ for all large n. The desired rate of convergence (2.4) follows since ε can be arbitrarily small.

Finally, we note that the boundedness of x_i is only used to ensure $m_\eta>0$. If $\dot{\mu}$ is bounded away from zero, then the theorem holds without this condition. In addition, in view of the preceding paragraph, the condition may be relaxed to $\max_{i\leq n}\|x_i\|^2v_{jj}^{(n)}|\log v_{jj}^{(n)}|^{1+\delta}=O(1)$ for every $1\leq j\leq p$. □

3. Strong consistency for maximum quasi-likelihood estimators in the adaptive design case. In this section, we consider the adaptive design case, in which for each i, x_i may depend on y_j, $j\leq i-1$. Formally, let $\{\mathscr{F}_i, i\geq 1\}$ be a sequence of increasing σ-fields (σ-filtration) such that $y_i\in\mathscr{F}_i$ and $x_i\in\mathscr{F}_{i-1}$. The regression relation (2.1) now becomes

$$E(y_i|\mathscr{F}_{i-1}) = \mu(\beta_0'x_i), \tag{3.1}$$

where again β_0 is the true value for β and μ the link function which is assumed to be continuously differentiable with a positive derivative function. As in the fixed design case, we similarly define errors

$$\varepsilon_i = y_i-\mu(\beta_0'x_i) = y_i - E(y_i|\mathscr{F}_{i-1}).$$

Clearly $\{\varepsilon_i, i\geq 1\}$ forms a martingale difference sequence with respect to $\{\mathscr{F}_i, i\geq 1\}$. As in Section 2, define $\lambda_{\min}(n)$ and $\lambda_{\max}(n)$ to be the minimum and maximum eigenvalues of $\sum_{i=1}^n x_ix_i'$.

In the case of linear regression, that is, $\mu(z) = z$, such adaptive designs arise naturally in feedback control of linear systems in which all or part of the x_i are set sequentially. It turns out that most optimal control rules result in asymptotically ill-conditioned (singular) design matrices $\sum_{i=1}^n x_i x_i'$ in the sense that $\lambda_{\min}(n)/\lambda_{\max}(n) \to 0$. It is for these cases that the result of Lai and Wei (1982) becomes very useful.

Many adaptive designs with generalized linear regression models face the same singularity problem in terms of ill-conditioned design matrices. This may be illustrated with a logistic regression for the so-called dose-response problem where dose z_i and response y_i for the ith experiment satisfy

$$P(y_i = 1|z_i) = 1 - P(y_i = 0|z_i) = \exp(\alpha + \gamma z_i)/[1 + \exp(\alpha + \gamma z_i)].$$

An important type of design in biological assay [Finney (1978)] is one in which z_i's are sequentially selected so that they converge to the median lethal dose, or LD50. The LD50 here is equal to $-\alpha/\gamma$ because $P(y_i = 1|z_i = -\alpha/\gamma) = 50\%$. Since for such designs z_i converges to $-\alpha/\gamma$ as $i \to \infty$, $\sum_{i=1}^n x_i x_i'$, where $x_i' = (1, z_i)$, must be asymptotically singular. It is this singularity and the dependency among the y_i due to the sequential selection rule that make the usual approach based on the law of large numbers difficult to apply. Detailed discussions of such designs may be found in Wetherill (1963) and Wu (1985, 1986).

Below, we shall derive the strong consistency result for the generalized linear models that parallels that of Lai and Wei (1982). To do so, we first introduce the following two conditions:

(C3) $\lim_{n \to \infty} \lambda_{\min}(n)/\log \lambda_{\max}(n) = \infty$ a.s., where $\lambda_{\min}(n)$ and $\lambda_{\max}(n)$ are, respectively, the minimum and the maximum eigenvalues of $\sum_{i=1}^n x_i x_i'$.

(C4) $\sup_{i \ge 1} E(|\varepsilon_i|^{\alpha} \mid \mathscr{F}_{i-1}) < \infty$ a.s. for some $\alpha > 2$.

THEOREM 2. *For the generalized linear model as specified by* (3.1) *with an adaptive design, suppose*

$$\sup_{i \ge 1} \|x_i\| < \infty \quad a.s. \tag{3.2}$$

Then $\hat{\beta}_n$ *is strongly consistent under* (C3) *and* (C4). *More precisely*,

$$\|\hat{\beta}_n - \beta_0\| = O\left(\{\log \lambda_{\max}(n)/\lambda_{\min}(n)\}^{1/2}\right) \quad a.s. \tag{3.3}$$

REMARK 3. Suppose that in a quantal response experiment with a logistic link, that is, $\mu(z) = [1 + \exp(-\alpha - \gamma z)]^{-1}$, the design sequence is generated by the stochastic approximation algorithm of form

$$z_n = z_{n-1} + \frac{c}{n}(y_n - 1/2)$$

to approximate the optimal design $z^* = -\alpha/\gamma$. If $c < 2/\gamma$, then it follows from Lai and Robbins (1979), Theorem 2, that $z_n \to z^*$ a.s., which implies $\lambda_{\max}(n)/\lambda_{\min}(n) \to \infty$ a.s., and that $\lambda_{\min}(n)$ is of order $n^{1-c\gamma/2}$, which implies (C3). So the preceding theorem implies that the maximum likelihood estimator is strongly consistent.

REMARK 4. Condition (C3) was shown to be sharp by Lai and Wei (1982) for linear regression with an adaptive design.

REMARK 5. As in the fixed design case, the boundedness of x_i is not needed if $\dot{\mu}$ is bounded away from 0. Furthermore, (3.2) can be replaced by a slightly weaker condition, $\max_{i \le n}\|x_i\|^2 = o(\lambda_{\min}(n)/\log \lambda_{\max}(n))$.

PROOF OF THEOREM 2. The proof is similar in spirit to the proof of Theorem 1. Specifically, define $a_n = (\sum x_i x_i')^{-1}\sum x_i \varepsilon_i$, the same as in (2.9). Let $G_n(\beta)$ be the vector-valued function defined by (2.8) and $\tilde{G}_n(\beta) = (\sum x_i x_i')^{-1} G_n(\beta)$. As we argued in the proof of Theorem 1, G_n and $\tilde{G}_n$ are injections from R^p to R^p and $\tilde{G}_n(\beta_0) = 0$. Thus, letting $m_\eta = \inf_{i, \|\beta - \beta_0\| \le \eta} \dot{\mu}(\beta' x_i)$, we have, by the argument of (2.12),

$$\begin{aligned} \left\|\tilde{G}_n(\beta)\right\|^2 &\ge m_\eta^2 \|\beta - \beta_0\|^2 \quad \text{for } \|\beta - \beta_0\| \le \eta \\ &= m_\eta^2 \eta^2 > 0 \quad \text{for } \|\beta - \beta_0\| = \eta. \end{aligned} \tag{3.4}$$

So, letting $\gamma = m_n \eta$, we know that $\tilde{G}_n^{-1}$ is well defined on $\{b\colon \|b\| \le \gamma\}$ and $\tilde{G}_n^{-1}(\{b\colon \|b\| \le \gamma\}) \subseteq \{\beta\colon \|\beta - \beta_0\| \le \eta\}$. Now condition (C3) and Lai and Wei (1982), Theorem 1, imply that $a_n \to 0$ a.s. So, by the argument of (2.15), $\hat{\beta}_n = \tilde{G}_n^{-1}(a_n)$ exists and falls into $\{\beta\colon \|\beta - \beta_0\| \le \eta\}$ for all large n. The arbitrariness of η entails $\hat{\beta}_n \to \beta_0$ a.s. Furthermore, (3.3) follows by choosing η to be of order $\log \lambda_{\max}(n)/\lambda_{\min}(n)$. □

APPENDIX

LEMMA A. *Let H be a smooth injection from R^p to R^p with $H(x_0) = y_0$. Define $B_\delta(x_0) = \{x \in R^p, \|x - x_0\| \le \delta\}$ and $S_\delta(x_0) = \partial B_\delta(x_0) = \{x \in R^p, \|x - x_0\| = \delta\}$. Then $\inf_{x \in S_\delta(x_0)}\|H(x) - y_0\| \ge r$ implies*:

(i) $B_r(y_0) = \{y \in R^p, \|y - y_0\| \le r\|\} \subseteq H(B_\delta(x_0))$;
(ii) $H^{-1}(B_r(y_0)) \subseteq B_\delta(x_0)$.

PROOF. Part (i), which clearly implies (ii), is a straightforward consequence of H being a homeomorphism from $B_\delta(x_0)$ to $H(B_\delta(x_0))$ and a standard result in topology [Dugundji (1966), page 359, Corollary 3.2]. □

Acknowledgment. The authors thank the reviewers for their helpful comments that led to numerous improvements.

REFERENCES

ANDERSEN, E. B. (1980). *Discrete Statistical Models with Social Science Applications*. North-Holland, Amsterdam.

ANDERSON, T. W. and TAYLOR, J. B. (1979). Strong consistency of least squares estimates in dynamic models. *Ann. Statist.* **7** 484–489.

ÅSTROM, K. L. and WITTENMARK, B. (1973). On self-tuning regulators. *Automatica* **9** 185–199.

BOX, G. E. P. and JENKINS, G. (1970). *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco.

CHOW, Y. S. and TEICHER, H. (1988). *Probability Theory*, 2nd ed. Springer, New York.

COCHRAN, W. G. and DAVIS, M. (1965). The Robbins–Monro method for estimating the median lethal dose. *J. Roy. Statist. Soc. Ser. B* **27** 28–44.

DIXON, W. J. and MOOD, A. M. (1948). A method for obtaining and analyzing sensitivity data. *J. Amer. Statist. Assoc.* **43** 109–127.

DUGUNDJI, J. (1966). *Topology*. Allyn and Bacon, Boston.

FAHRMEIR, L. and KAUFMANN (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Ann. Statist.* **13** 342–368.

FINNEY, D. J. (1978). *Statistical Methods in Biological Assay*. Griffin, London.

HABERMAN, S. J. (1977). Maximum likelihood estimates in exponential response models. *Ann. Statist.* **5** 815–841.

LAI, T. L. and ROBBINS, H. (1979). Adaptive design and stochastic approximation. *Ann. Statist.* **7** 1196–1221.

LAI, T. L., ROBBINS, H. and WEI, C. Z. (1979). Strong consistency of least squares estimates in multiple regression II. *J. Multivariate Anal.* **9** 343–361.

LAI, T. L. and WEI, C. Z. (1982). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Ann. Statist.* **10** 154–166.

LIANG, K.-Y. and ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73** 13–22.

LORD, M. F. (1971a). Tailored testing, an application of stochastic approximation. *J. Amer. Statist. Assoc.* **66** 707–711.

LORD, M. F. (1971b). Robbins–Monro procedures for tailored testing. *Educational and Psychological Measurement* **31** 3–31.

MCCULLAGH, P. and NELDER, J. A. (1989). *Generalized Linear Models*, 2nd ed. Chapman and Hall, London.

MOORE, J. B. (1978). On strong consistency of least squares identification algorithms. *Automatica* **14** 505–509.

NELDER, J. A. and WEDDERBURN, R. W. M. (1972). Generalized linear models. *J. Roy. Statist. Soc. Ser. A* **135** 370–384.

NORDBERG, L. (1980). Asymptotic normality of maximum likelihood estimators based on independent, unequally distributed observations in exponential family models. *Scand. J. Statist.* **7** 27–32.

ROBBINS, H. and MONRO, S. (1951). A stochastic approximation method. *Ann. Math. Statist.* **22** 400–407.

WEDDERBURN, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. *Biometrika* **61** 439–447.

WETHERILL, G. B. (1963). Sequential estimation of quantal response curves (with discussion). *J. Roy. Statist. Soc. Ser. B* **25** 1–48.

WU, C. F. J. (1985). Efficient sequential designs with binary data. *J. Amer. Statist. Assoc.* **80** 974–984.

WU, C. F. J. (1986). Maximum likelihood recursion and stochastic approximation in sequential designs. In *Adaptive Statistical Procedures and Related Topics* (J. Van Ryzin, ed.) IMS, Hayward, CA.

K. CHEN
DEPARTMENT OF MATHEMATICS
HONG KONG UNIVERSITY OF
SCIENCE AND TECHNOLOGY
CLEAR WATER BAY
KOWLOON
HONG KONG

I. HU
DEPARTMENT OF ISMT
HONG KONG UNIVERSITY OF
SCIENCE AND TECHNOLOGY
CLEAR WATER BAY
KOWLOON
HONG KONG

Z. YING
DEPARTMENT OF STATISTICS
HILL CENTER, BUSCH CAMPUS
RUTGERS UNIVERSITY
PISCATAWAY, NEW JERSEY 08855
E-MAIL: zying@stat.rutgers.edu