The Annals of Statistics
1992, Vol. 20, No. 4, 1953–1969

FIXED SIZE CONFIDENCE REGIONS FOR PARAMETERS OF A LOGISTIC REGRESSION MODEL[1]

BY YUAN-CHIN IVAN CHANG AND ADAM T. MARTINSEK

University of Illinois

Let $(\mathbf{X}_i, Y_i)$ be independent, identically distributed observations that satisfy a logistic regression model; that is, for each i, $\log[P(Y_i = 1|\mathbf{X}_i)/P(Y_i = 0|\mathbf{X}_i)] = \mathbf{X}_i^T\beta_0$, where $Y_i \in \{0, 1\}$, $\mathbf{X}_i \in \mathbf{R}^p$ and $\beta_0 \in \mathbf{R}^p$ is the unknown parameter vector of the model. The marginal distribution of the covariate vectors $\mathbf{X}_i$ is assumed to be unknown. Sequential procedures for constructing fixed size and fixed proportional accuracy confidence regions for β_0 are proposed and shown to be asymptotically efficient as the size of the region becomes small.

1. Introduction. For any confidence set (CS), there are at least two important requirements. The first one concerns the coverage probability, that is, for a given $\alpha \in (0, 1)$, we wish to have $P_\theta(\theta \in \text{CS}) \approx 1 - \alpha$, for each $\theta \in \Theta$. The second requirement concerns the precision of the confidence set. It is undesirable to make a uselessly imprecise statement, even if it can be made with great confidence (as an extreme example, note that the entire real line is a 100% confidence interval for any real-valued parameter). For instance, suppose $X, X_1, \dots$ are i.i.d. random variables and $EX = \mu$, $\text{Var}(X) = \sigma^2 < \infty$. If σ^2 is known, then the fixed sample size confidence interval (CI) with endpoints $\overline{X}_n \pm d$ has approximate coverage probability $1 - \alpha$ provided $\sigma^2 \approx (d^2 n)/z_{\alpha/2}^2$, where $z_{\alpha/2}$ satisfies $\Phi(z_{\alpha/2}) - \Phi(-z_{\alpha/2}) = 1 - \alpha$. It is clear that if σ^2 is unknown, then there is no fixed sample size procedure (f.s.s.) that can achieve this goal—to construct a CI with approximate coverage probability $1 - \alpha$ and prescribed width $2d$ at the same time. Therefore, under such circumstances, a sequential procedure is the only way to achieve this goal (asymptotically).

The original idea of the fixed width confidence interval appears in Stein (1945, 1949). In Chow and Robbins (1965), a very useful method for constructing a CI for an unknown mean with prescribed coverage probability and precision has been presented. In their paper, they proved that the sequential procedure they presented is *asymptotically consistent* (the coverage probability converges to the prescribed probability) and *asymptotically efficient* (the ratio of the expected random sample size to the unknown best fixed sample size converges to 1 as the width of the CI goes to 0). Their ideas have been extended to higher dimensional cases and to regression models [Gleser (1965),

Received July 1991; revised December 1991.

[1]Research supported in part by NSF Grant DMS-88-02556 and by Air Force Grant AFOSR 87-0041.

AMS 1980 *subject classifications*. Primary 62L12; secondary 62F25, 62J12.

Key words and phrases. Logistic regression, fixed size confidence set, sequential estimation, stopping rule, last time, uniform integrability, asymptotic efficiency.

Albert (1966), Srivastava (1967, 1971) and Finster (1985)]. In this paper we consider fixed size confidence ellipsoids for parameters of a logistic regression model.

Assume that $(\mathbf{X}_i, Y_i)$, $i = 1, 2, \dots,$ are independent observations, where Y_i are binary variables, $\mathbf{X}_i$ are $p \times 1$ vectors and for each i, $(\mathbf{X}_i, Y_i)$ satisfies

$$\log\left\{\frac{P[Y_i = 1|\mathbf{X}_i]}{P[Y_i = 0|\mathbf{X}_i]}\right\} = \mathbf{X}_i^T\beta_0, \tag{1.1}$$

where $\beta_0 \in \mathbf{R}^p$ is unknown. Then

$$P[Y_i = 1|\mathbf{X}_i] = \frac{\exp(\mathbf{X}_i^T\beta_0)}{(1 + \exp(\mathbf{X}_i^T\beta_0))} = p_i, \quad \text{or} \quad Y_i|\mathbf{X}_i \sim \text{Ber}(1, p_i), \quad \text{for each } i.$$

Assume further that $(\mathbf{X}_i, Y_i)$, $i = 1, 2, \dots,$ are i.i.d. This model is relevant to observational studies, such as cohort studies. Then an estimate of β_0 can be obtained by maximizing the conditional likelihood function. In general, there is no explicit solution for the conditional maximum likelihood estimator (MLE) in the logistic regression problem. Therefore, the conditional MLE in this case has to be computed by an iterative method.

The logistic regression model is a very commonly used statistical tool in medical applications and other areas. For example, the response variable Y_i may be coded 1 if the ith person we observe is diseased; 0 if not. The components of $\mathbf{X}_i$ would then be covariates or risk factors.

We will assume throughout that for any vector subspace V of $\mathbf{R}^p$ with $\dim(V) < p$,

$$P(\mathbf{X}_1 \in V) < 1. \tag{1.2}$$

It can be shown that if $E\|\mathbf{X}_1\|^2 < \infty$, then the MLE $\hat{\beta}_n$ converges to β_0 a.s. and if $E\|\mathbf{X}\|^3 < \infty$, it is asymptotically normal,

$$\sqrt{n}\left(\hat{\beta}_n - \beta_0\right) \to_{\mathscr{L}} N\left(0_p, \Sigma^{-1}\right),$$

where

$$\Sigma = E\left[\frac{\exp(\mathbf{X}_1^T\beta_0)}{\left[1 + \exp(\mathbf{X}_1^T\beta_0)\right]^2}\mathbf{X}_1\mathbf{X}_1^T\right] \tag{1.3}$$

is the unknown Fisher information matrix or covariance matrix of $Y_1\mathbf{X}_1$ [Stefanski and Carroll (1985), Grambsch (1989) and Chang (1991)].

The asymptotic covariance matrix contains the unknown vector β_0, so it usually will be unknown. Therefore, there is no fixed sample size procedure that can be used for constructing a confidence set with prescribed coverage probability and precision. Only a sequential procedure can offer the possibility of achieving both goals.

It follows from the asymptotic normality result for $\hat{\beta}_n$ that

$$\left(\hat{\beta}_n - \beta_0\right)^T\hat{\Sigma}_n\left(\hat{\beta}_n - \beta_0\right) \to_{\mathscr{L}} \chi^2(p), \quad \text{as } n \to \infty,$$

where $\hat{\Sigma}_n = \sum_{i=1}^n (\exp(\mathbf{X}_i^T\hat{\beta}_n)/[1 + \exp(\mathbf{X}_i^T\hat{\beta}_n)]^2)\mathbf{X}_i\mathbf{X}_i^T$. For any $d > 0$, let

$$R_n = \left\{\mathbf{Z} \in \mathbf{R}^p: \left(\mathbf{Z} - \hat{\beta}_n\right)^T \hat{\Sigma}_n\left(\mathbf{Z} - \hat{\beta}_n\right) \le d^2\hat{\lambda}_n^{(p)}\right\}, \tag{1.4}$$

where $\hat{\lambda}_n^{(p)}$ is the smallest eigenvalue of $\hat{\Sigma}_n$. Then R_n defines an ellipsoid with maximum axis equal to $2d$ ($d > 0$), and it is in this sense that the size of the ellipsoid is fixed. Moreover, for any $\alpha \in (0, 1)$,

$$P\{\beta_0 \in \mathbf{R}_n\} \approx 1 - \alpha,$$

provided $d^2\hat{\lambda}_n^{(p)} \approx a^2(*)$, where a^2 satisfies $P[\chi^2(p) \le a^2] = 1 - \alpha$. $(1/n)\hat{\Sigma}_n \to \Sigma$ almost surely as $n \to \infty$, which implies that $(1/n)\hat{\lambda}_n^{(p)}$ converges to $\lambda^{(p)}$ almost surely as $n \to \infty$, where $\lambda^{(p)}$ is the smallest eigenvalue of the covariance matrix Σ. Hence the (unknown) sample size $n \approx a^2/d^2\lambda^{(p)}$ will achieve approximate coverage probability $1 - \alpha$.

$(*)$ suggests the stopping rule

$$T_d = \inf\left\{n \ge 1: \hat{\lambda}_n^{(p)} \ge \frac{a_n^2}{d^2}\right\} \tag{1.5}$$

and the confidence ellipsoid R_{T_d}, where $a_n \to a$ and R_n is defined by (1.4). Then R_{T_d} has maximum axis $2d$. Moreover, we have the following theorems whose proofs are given in Section 2.

THEOREM 1.1. *If* $E\|\mathbf{X}_1\|^3 < \infty$, *then*:

(i) T_d *is finite almost surely*, T_d *is increasing as* $d \to 0$ *and* $T_d \to \infty$ *a.s. as* $d \to 0$. *Moreover*, $\lim_{d \to 0} d^2T_d\lambda^{(p)}/a^2 = 1$ *a.s.*,

(ii) $\hat{\Sigma}_{T_d}^{1/2}(\hat{\beta}_{T_d} - \beta_0) \to_{\mathscr{L}} N(0_p, I_{p\times p})$ *as* $d \to 0$,

(iii) $(\hat{\beta}_{T_d} - \beta_0)^T\hat{\Sigma}_{T_d}(\hat{\beta}_{T_d} - \beta_0) \to_{\mathscr{L}} \chi^2(p)$ *as* $d \to 0$ *and* $\lim_{d\to 0} P\{\beta_0 \in R_{T_d}\} = 1 - \alpha$ *(asymptotic consistency)*.

THEOREM 1.2 (Asymptotic efficiency). *If* $E\|\mathbf{X}_1\|^4 < \infty$, *then* $\{d^2T_d: d \in (0, 1)\}$ *is uniformly integrable and* $\lim_{d \to 0} E[(d^2T_d\lambda^{(p)})/a^2] = 1$.

The third part of Theorem 1.1. states that the coverage probability of the sequential fixed size confidence ellipsoid is asymptotically, as the size of the ellipsoid approaches zero, the desired value $1 - \alpha$. Theorem 1.2 asserts that this is achieved with an expected sample size that is asymptotically equivalent to the nonrandom sample size that would have been used had $\lambda^{(p)}$ been known.

REMARKS. For the case $p = 1$ (logistic regression through the origin, with no intercept), the moment assumption for asymptotic consistency may be reduced from 3 to 2, and the moment assumption for asymptotic efficiency may be reduced from 4 to 2. See Chang (1991, Chapter 3) for details. Chang (1991) also gives results analogous to Theorems 1.1 and 1.2 for appropriate two stage procedures.

The proof of Theorem 1.1. relies on results of Chow and Robbins (1965) and Gleser (1969) and is not difficult. The proof of Theorem 1.2 is much more involved. It is natural to try to apply nonlinear renewal theory [see, e.g., Lai and Siegmund (1977, 1979) and Woodroofe (1982)] to this problem, but it appears to be very difficult to check the necessary technical conditions for the implicitly defined $\hat{\beta}_n$. The proof given in Section 2 involves use of "last times" (Chow and Lai, 1975), along with concavity and monotonicity properties of the log-likelihood function and its derivatives.

One crucial difference between our results for the logistic regression model and previous work on the general linear model, is that in the latter the asymptotic covariance matrix of the estimate depends only on the design and on an unknown nuisance parameter σ^2 [see Finster (1985)]. In the logistic regression case, by contrast, the asymptotic covariance matrix depends on the (unknown) parameter of interest β_0.

It is important to note that the ellipsoid R_{T_d} puts bounds on individual components of β_0, and hence with asymptotic confidence at least $1 - \alpha$ one can give bounds for the (multiplicative) change in the odds when the value of a particular covariate changes.

Sequential likelihood estimation problems have been considered by other authors, for example, Grambsch (1983, 1989) and Yu (1989). In Grambsch (1983), a sequential estimation procedure using the observed Fisher information to define the stopping time and to obtain the required estimation accuracy has been presented. However, the asymptotic efficiency of the stopping time has not been established. In Yu (1989), a related problem has been considered. In his work, he successfully showed that a sequential procedure for constructing a fixed width confidence interval for an unknown parameter, with its MLE as the center of the CI, is asymptotically consistent and efficient (in the i.i.d. case). In Grambsch (1989), she extends her results to the multidimensional case with some applications to logistic regression problems. She proposes a stopping rule that depends on an estimate of the smallest eigenvalue of the Fisher information matrix. In her work, she assumes the Fisher information matrix can be computed as a function of the unknown parameter, and therefore so can the eigenvalues. Then, the estimators of the Fisher information matrix and its smallest eigenvalue can be obtained by plugging in the maximum likelihood estimator of the unknown parameter. But in order to compute the Fisher information matrix, one needs to know the underlying distribution first. In the logistic regression model, this means one ought to know the distribution of the explanatory variables. In practice, this usually will be unknown, so the stopping rule in Grambsch (1989) will be very difficult to use.

The stopping rules proposed in this paper are based on the observed Fisher information matrix and its eigenvalues. Therefore, they can easily be computed by standard routines (e.g., Cholesky decomposition). In addition, by the methods used here in the proof of asymptotic efficiency, one can also prove the asymptotic efficiency of Grambsch's (1989) stopping rule [see Chang (1991)].

Section 3 gives results from a small Monte Carlo study of the sequential procedure. Section 4 contains a brief discussion of estimation with fixed proportional accuracy and estimation of a particular linear combination of the components of β_0.

There are other possible approaches to the fixed precision problem. Instead of ellipsoids, one could consider spheres, as proposed in Gleser (1965) for the linear regression case, rectangular regions or more general regions. One advantage of the approach presented here is its computational simplicity: The stopping rules corresponding to other regions are much more complicated. Moreover, the analysis in such cases does not appear to be amenable to the methods used in the proofs given here.

2. Proofs.

PROOF OF THEOREM 1.1. (i) follows directly from Chow and Robbins (1965, Lemma 1). It is clear that (ii) implies (iii). So, only (ii) needs to be proved. Let $l_n(\beta)$ be $1/n$ times the log-likelihood function,

$$l_n(\beta) = \frac{1}{n}\log[L_n(\beta)] = \frac{1}{n}\sum_{i=1}^{n}\left\{Y_i(\mathbf{X}_i^T\beta) - \log\left[1 + \exp(\mathbf{X}_i^T\beta)\right]\right\}. \tag{2.1}$$

By a Taylor series expansion and some rearranging,

$$\hat{\Sigma}_n^{1/2}(\hat{\beta}_n - \beta_0) = \hat{\Sigma}_n^{1/2}\tilde{\Sigma}_n^{-1}\tilde{\Sigma}_n(\hat{\beta}_n - \beta_0) = \hat{\Sigma}_n^{1/2}\tilde{\Sigma}_n^{-1}\left(\sum_{i=1}^{n}\mathbf{Z}_i\right), \tag{2.2}$$

where

$$\mathbf{Z}_i = Y_i\mathbf{X}_i - \frac{\exp(\mathbf{X}_i^T\beta_0)}{1 + \exp(\mathbf{X}_i^T\beta_0)}\mathbf{X}_i, \qquad i = 1, 2, \dots,$$

and

$$\tilde{\Sigma}_n = \sum_{i=1}^{n}\frac{\exp(\mathbf{X}_i^T\tilde{\beta}_n)}{\left[1 + \exp(\mathbf{X}_i^T\tilde{\beta}_n)\right]^2}\mathbf{X}_i\mathbf{X}_i^T$$

for some $\tilde{\beta}_n$ between β_0 and $\hat{\beta}_n$. Let $D_n = (n^{-1}\hat{\Sigma}_n)^{1/2}(n^{-1}\tilde{\Sigma}_n)^{-1}$. Then

$$\Sigma_{T_d}^{1/2}(\hat{\beta}_{T_d} - \beta_0) = \left(D_{T_d} - \Sigma^{-1/2}\right)T_d^{-1/2}\sum_{1}^{T_d}\mathbf{Z}_i + T_d^{-1/2}\Sigma^{1/2}\sum_{1}^{T_d}\mathbf{Z}_i. \tag{2.3}$$

Using Theorem 1.1 of Gleser (1969) with $B_n = I$ and Kolmogorov's inequality (c.f. Woodroofe, 1982, Example 1.8),

$$T_d^{-1/2}\sum_{1}^{T_d}\mathbf{Z}_i \to_{\mathscr{L}} N(0, \Sigma^{-1}) \tag{2.4}$$

and since $D_n \to \Sigma^{-1/2}$ a.s. and $T_d \to \infty$ a.s., it follows that the first term on the right side of (2.3) converges in probability to zero. From (2.4),

$$T_d^{-1/2}\Sigma^{1/2}\sum_1^{T_d}\mathbf{Z}_i \to_{\mathscr{L}} N(0, I), \tag{2.5}$$

completing the proof. □

Proving Theorem 1.2. We will sketch the proof for the case $a_n \equiv a$. There are no essential differences between this case and the more general one. The proof of Theorem 1.2 is complicated, but it involves two basic ideas. First, the MLE $\hat{\beta}_n$ will eventually lie within any given ball around β_0, and the last time $\hat{\beta}_n$ is outside the ball behaves nicely (in particular, has finite moments). Second, once $\hat{\beta}_n$ is guaranteed to stay within the ball, a monotonicity argument can be used to bound $\hat{\lambda}_n^{(p)}$ below by the smallest eigenvalue of a sum of i.i.d. matrices. The latter is much easier to handle than $\hat{\lambda}_n^{(p)}$ itself.

For any fixed $\rho > 0$, let

$$\mathbf{B}_\rho = \{\beta \in \mathbf{R}^p : \|\beta - \beta_0\| \le \rho\},$$

$$\partial\mathbf{B}_\rho = \{\beta \in \mathbf{R}^p : \|\beta - \beta_0\| = \rho\}$$

and define L_ρ to be the last time that $l_n(\tilde{\beta}) - l_n(\beta_0) \ge 0$ for some $\tilde{\beta} \in \partial\mathbf{B}_\rho$, or

$$L_\rho = \sup\left\{n \ge 1 : l_n(\tilde{\beta}) - l_n(\beta_0) \ge 0, \exists\, \tilde{\beta} \in \partial\mathbf{B}_\rho\right\}. \tag{2.6}$$

$l_n(\beta)$ is a concave function in β, $\forall\, n \in N$. [Note that $l_n(\beta)$ may not be strictly concave for small n, but here we only need the concavity of $l_n(\beta)$ for each n, not strict concavity.] By the definition of L_ρ, if $n > L_\rho$, then $l_n(\beta_0) > l_n(\tilde{\beta})$, $\forall\, \tilde{\beta} \in \partial\mathbf{B}_\rho$. Since $\hat{\beta}_n$ is an MLE, $l_n(\hat{\beta}_n) \ge l_n(\tilde{\beta})$, $\forall\, \tilde{\beta} \in \partial\mathbf{B}_\rho$. This implies, if $n > L_\rho$, $\hat{\beta}_n$ must be in $\mathbf{B}_\rho$. Otherwise, there would be a contradiction to the concavity of the log-likelihood function. Note that L_ρ is the supremum of last times of the form

$$\sup\left\{n \ge 1 : l_n(\tilde{\beta}) - l_n(\beta_0) \ge 0\right\}$$

for $\tilde{\beta} \in \partial\mathbf{B}_\rho$. Hence for $p > 1$, it is the supremum of (and not merely one of) uncountably many last times for random walks of the type considered by Chow and Lai (1975). With some stronger moment conditions on $\mathbf{X}_1$, the last time random variable L_ρ has the following nice property.

PROPOSITION 2.1. *For* $l \ge 1$, *if* $E\|\mathbf{X}_1\|^{2(l+1)} < \infty$, *then there is* $\rho > 0$ *such that* $EL_\rho^l < \infty$.

The proof of this proposition will be given at the end of this section.

The stopping time T_d depends on the smallest eigenvalue of the unknown covariance matrix. Usually, there is no explicit solution for the eigenvalues. To overcome this difficulty, we define another last time variable [see (2.10)].

Let $\beta_{0,j}$ denote the jth coordinate of β_0, for $j = 1, \dots, p$. Define

$$\beta_\rho = \beta_0 + \big(\operatorname{sgn}(\beta_{0,1})\rho, \dots, \operatorname{sgn}(\beta_{0,p})\rho\big)^T, \tag{2.7}$$

$$M_i = \frac{\exp(\|\mathbf{X}_i\| \cdot \|\beta_\rho\|)}{\big[1 + \exp(\|\mathbf{X}_i\| \cdot \|\beta_\rho\|)\big]^2} \mathbf{X}_i \mathbf{X}_i^T, \quad \text{for } i = 1, 2, \dots, \tag{2.8}$$

and

$$M = E\left[\frac{\exp(\|\mathbf{X}_1\| \cdot \|\beta_\rho\|)}{\big[1 + \exp(\|\mathbf{X}_i\| \cdot \|\beta_\rho\|)\big]^2} \mathbf{X}_1 \mathbf{X}_1^T\right]. \tag{2.9}$$

Let λ_n be the smallest eigenvalue of $\Sigma_{i=1}^n M_i$ and λ_ρ the smallest eigenvalue of M. Under the assumptions on $\mathbf{X}_1$, $\lambda_\rho > 0$. Define the last time random variable

$$L_\lambda = \sup\left\{n \ge 1 : \mathbf{Z}^T \sum_{i=1}^n \{M_i - M\} \mathbf{Z} \le \frac{-n\lambda_\rho}{2}, \exists\, \mathbf{Z} \in \mathbf{R}^p, \|\mathbf{Z}\| = 1\right\} \tag{2.10}$$

or

$$L_\lambda = \sup\left\{n \ge 1 : \mathbf{Z}^T \sum_{i=1}^n \{M - M_i\} \mathbf{Z} \ge \frac{n\lambda_\rho}{2}, \exists\, \mathbf{Z} \in \mathbf{R}^p, \|\mathbf{Z}\| = 1\right\}. \tag{2.11}$$

By using the same techniques as in the proof of Proposition 2.1, we can show that L_λ has the following nice property.

LEMMA 2.1. *For $l \ge 1$, if $E\|\mathbf{X}_1\|^{2(l+1)} < \infty$, then $EL_\lambda^l < \infty$.*

We know that the smallest eigenvalue of $\Sigma_{i=1}^n \{M_i - M\}$,

$$\Lambda_p\left(\sum_{i=1}^n \{M_i - M\}\right) = \inf_{\|\mathbf{Z}\|=1} \mathbf{Z}^T \sum_{i=1}^n \{M_i - M\} \mathbf{Z}.$$

All the matrices M_i, $i = 1, 2, \dots$, and M are symmetric and $\Sigma_{i=1}^n M_i = \Sigma_{i=1}^n \{M_i - M\} + nM$. By the results of Wilkinson (1963) and Golub and van Loan (1983), $\lambda_n \ge \Lambda_p(\Sigma_{i=1}^n \{M_i - M\}) + n\lambda_\rho$. Therefore,

$$n > L_\lambda \Rightarrow \forall\, \mathbf{Z} \in \mathbf{R}^p \text{ with } \|\mathbf{Z}\| = 1, \mathbf{Z}^T \sum_{i=1}^n \{M_i - M\} \mathbf{Z} > \frac{-n\lambda_\rho}{2}$$

$$\begin{aligned} &\Rightarrow \inf_{\|\mathbf{Z}\|=1} \mathbf{Z}^T \sum_{i=1}^n \{M_i - M\} \mathbf{Z} \ge \frac{-n\lambda_\rho}{2} \\ &\Rightarrow \Lambda_p\left(\sum_{i=1}^n \{M_i - M\}\right) \ge \frac{-n\lambda_\rho}{2} \\ &\Rightarrow \lambda_n \ge \frac{-n\lambda_\rho}{2} + n\lambda_\rho = \frac{n\lambda_\rho}{2}. \end{aligned} \tag{2.12}$$

Before we get into the proof of Theorem 1.2, we need the following monotonicity properties.

PROPERTY 1. Let β_0 be fixed. Then for any $\beta \in \mathbf{B}_\rho$ and any $\mathbf{X} \in \mathbf{R}^p$, $|\mathbf{X}^T\beta| \le |\mathbf{X}^T\beta_0| + \|\mathbf{X}\| \cdot \|\beta - \beta_0\|$.

PROPERTY 2. For $t \ge 0$, let $f(t) = e^t/(1+e^t)^2$. Then $(d/dt)f(t) = (e^t(1-e^t))/(1+e^t)^3 < 0$, $\forall\ t > 0$. That is, $f(t)$ is decreasing in t for $t > 0$.

PROPERTY 3. The function $f(t)$ (as defined in Property 2) is an even function, so $\exp(\mathbf{X}^T\beta)/(1+\exp(\mathbf{X}^T\beta))^2 = \exp(|\mathbf{X}^T\beta|)/(1+\exp(|\mathbf{X}^T\beta|))^2$.

By Theorem 1.1, $(d^2T_d\lambda^{(p)})/a^2 \to 1$ a.s. as $d \to 0$. Hence, to prove asymptotic efficiency, it is sufficient to show that $\{d^2T_d\colon d \in (0,1)\}$ is uniformly integrable.

PROOF OF THEOREM 1.2. Let $\rho > 0$ be such that $EL_\rho < \infty$. Then,

$$T_d \le \max(L_\rho, L_\lambda) + T_d I_{\{T_d > \max(L_\rho, L_\lambda)\}}. \tag{2.13}$$

If $n > L_\rho$, then $\hat{\beta}_n \in B_\rho$. Hence, $|X_i^T\hat{\beta}_n| \le \|X_i\|\,\|\hat{\beta}_n\| \le \|X_i\|\,\|\beta_\rho\|$, where β_ρ is defined as in (2.7). Let $\Sigma_{\rho,n} = \sum_{i=1}^n M_i$. Then $\hat{\Sigma}_n = \Sigma_{\rho,n} + [\hat{\Sigma}_n - \Sigma_{\rho,n}]$. From Properties 1–3 and Lemma A.3 of Chang (1991), the term inside the bracket is a nonnegative definite $p \times p$ matrix. By the results of Bellman (1960) and Srivastava (1967), $0 \le \lambda_n \le \hat{\lambda}_n^{(p)}$. It follows from this observation, together with (2.12), that on the event $\{T_d > \max(L_\rho, L_\lambda)\}$, $T_d \le 2a^2/d^2\lambda_\rho + 1$. Hence, from (2.13), we have

$$T_d \le \max(L_\rho, L_\lambda) + 2a^2/d^2\lambda_\rho + 1. \tag{2.14}$$

Since $E(L_\rho) < \infty$ and $E(L_\lambda) < \infty$, it follows from (2.14) that $\{d^2T_d\colon d \le 1\}$ is uniformly integrable and hence $\lim_{d\to 0} E(d^2T_d\lambda^{(p)})/a^2 = 1$. □

We turn now to the proof of Proposition 2.1 and Lemma 2.1. The proof of Proposition 2.1 is based on the following lemmas, and the proof of Lemma 2.1 follows by the same technique.

Let $H(\beta) = E\log(1+\exp(\mathbf{X}_1^T\beta))$. Then

$$H'(\beta) = E\left[\frac{\exp(\mathbf{X}_1^T\beta)}{1+\exp(\mathbf{X}_1^T\beta)}\mathbf{X}_1\right],$$

$$H''(\beta) = E\left[\frac{\exp(\mathbf{X}_1^T\beta)}{\left(1+\exp(\mathbf{X}_1^T\beta)\right)^2}\mathbf{X}_1\mathbf{X}_1^T\right].$$

For any $\tilde{\beta} \in \mathbf{B}_\rho$, by Taylor expansion,

$$H(\tilde{\beta}) = H(\beta_0) + H'(\beta_0)^T(\tilde{\beta} - \beta_0) + \tfrac{1}{2}(\tilde{\beta} - \beta_0)^T H''(\beta^*)(\tilde{\beta} - \beta_0), \tag{2.15}$$

where $\beta^* \in L(\tilde{\beta}, \beta_0) = \{\beta \in \mathbf{R}^p\colon \beta = t\tilde{\beta} + (1-t)\beta_0, t \in [0,1]\}$. Let $A(\beta^*, \tilde{\beta}) = (\tilde{\beta} - \beta_0)^T H''(\beta^*)(\tilde{\beta} - \beta_0)/2$. Then, under our assumptions on $\mathbf{X}_1$, $H''(\beta)$ is positive definite, $\forall\ \beta \in \mathbf{R}^p$. $A(\cdot,\cdot)$ is a continuous function on the compact set $(\mathbf{B}_\rho \times \partial\mathbf{B}_\rho)$, hence for small enough ρ, there is an $m > 0$, such that $\inf_{\tilde{\beta} \in \partial\mathbf{B}_\rho, \beta^* \in \mathbf{B}_\rho} A(\tilde{\beta}, \beta^*) = m$.

Equation (2.15) can be written as

$$\left[H'(\beta_0)^T\tilde{\beta} - H(\tilde{\beta})\right] - \left[H'(\beta_0)^T\beta_0 - H(\beta_0)\right] = -A(\beta^*, \tilde{\beta}). \tag{2.16}$$

Write

$$\begin{aligned}
l_n(\tilde{\beta}) - l_n(\beta_0) &= \frac{1}{n}\sum_{i=1}^{n}\left\{Y_i\mathbf{X}_i^T\tilde{\beta} - \log\left[1 + \exp\left(\mathbf{X}_i^T\tilde{\beta}\right)\right]\right\} \\
&\quad - \frac{1}{n}\sum_{i=1}^{n}\left\{Y_i\mathbf{X}_i^T\beta_0 - \log\left[1 + \exp\left(\mathbf{X}_i^T\beta_0\right)\right]\right\} \\
&= \left[H'(\beta_0)^T\tilde{\beta} - H(\tilde{\beta})\right] - \left[H'(\beta_0)^T\beta_0 - H(\beta_0)\right] \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\left\{Y_i\mathbf{X}_i - H'(\beta_0)\right\}^T\left(\tilde{\beta} - \beta_0\right) && \text{(2.17a)} \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\left\{\log\left[1 + \exp\left(\mathbf{X}_i^T\beta_0\right)\right] - H(\beta_0)\right\} && \text{(2.17b)} \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\left\{H(\tilde{\beta}) - \log\left[1 + \exp\left(\mathbf{X}_i^T\tilde{\beta}\right)\right]\right\}. && \text{(2.17c)}
\end{aligned}$$

Let a_n, b_n and c_n denote the terms (2.17a), (2.17b) and (2.17c), respectively. Then,

$$\begin{aligned}
l_n(\tilde{\beta}) - l_n(\beta_0) \geq 0 &\Rightarrow a_n + b_n + c_n \geq A(\beta^*, \tilde{\beta}) \\
&\Rightarrow a_n + b_n + c_n \geq m.
\end{aligned} \tag{2.18}$$

Let L_a, L_b and L_c be last time random variables defined as

$$\begin{aligned}
L_a = \sup\Bigg\{ n \geq 1 : \exists\ \tilde{\beta} \in \partial\mathbf{B}_\rho,& \\
\left\{\sum_{i=1}^{n} Y_i\mathbf{X}_i - nH'(\beta_0)\right\}^T \left(\tilde{\beta} - \beta_0\right) \geq n\frac{m}{3}\Bigg\},&
\end{aligned} \tag{2.19a}$$

$$L_b = \sup\left\{ n \geq 1 : \sum_{i=1}^{n}\left\{\log\left[1 + \exp\left(\mathbf{X}_i^T\beta_0\right)\right] - H(\beta_0)\right\} \geq n\frac{m}{3}\right\}, \tag{2.19b}$$

$$\begin{aligned}
L_c = \sup\Bigg\{ n \geq 1 : \exists\ \tilde{\beta} \in \partial\mathbf{B}_\rho,& \\
\sum_{i=1}^{n}\left\{-\log\left[1 + \exp\left(\mathbf{X}_i^T\tilde{\beta}\right)\right] + H(\tilde{\beta})\right\} \geq n\frac{m}{3}\Bigg\}.&
\end{aligned} \tag{2.19c}$$

LEMMA 2.2. *For $l \geq 1$, if $E\|\mathbf{X}_1\|^{l+1}$, then $EL_a^l < \infty$.*

LEMMA 2.3. *For $l \geq 1$, if $E\|\mathbf{X}_1\|^{l+1} < \infty$, then $EL_b^l < \infty$.*

Lemmas 2.2 and 2.3 follow from Chow and Teicher [(1978), Corollary 10.4.4] and, in the case of Lemma 2.2, $\|\tilde{\beta} - \beta_0\| = \rho$. For details, see Chang and Martinsek (1991).

To prove $EL_c^l < \infty$, we need to apply Taylor expansion several times. First,

$$H(\tilde{\beta}) = H(\beta_0) + H'(\beta_0)^T(\tilde{\beta} - \beta_0) + \tfrac{1}{2}(\tilde{\beta} - \beta_0)^T H''(\beta^*)(\tilde{\beta} - \beta_0), \tag{2.20}$$

where $\beta^* \in L(\beta^*, \tilde{\beta})$. Again, by Taylor expansion,

$$\sum_{i=1}^{n} \log\left[1 + \exp\left(\mathbf{X}_i^T\tilde{\beta}\right)\right]$$

$$= \sum_{i=1}^{n} \log\left[1 + \exp(\mathbf{X}_i^T\beta_0)\right] \tag{2.21a}$$

$$+ \sum_{i=1}^{n} \frac{\exp(\mathbf{X}_i^T\beta_0)}{1 + \exp(\mathbf{X}_i^T\beta_0)} \mathbf{X}_i^T(\tilde{\beta} - \beta_0) \tag{2.21b}$$

$$+ \frac{1}{2}(\tilde{\beta} - \beta_0)^T\left[\sum_{i=1}^{n} \frac{\exp(\mathbf{X}_i^T\beta_n^*)}{\left[1 + \exp(\mathbf{X}_i^T\beta_n^*)\right]^2}\mathbf{X}_i\mathbf{X}_i^T\right](\tilde{\beta} - \beta_0), \tag{2.21c}$$

where $\beta_n^* \in L(\tilde{\beta}, \beta_0)$ depends on $(\mathbf{X}_1, \dots, \mathbf{X}_n)$. Hence,

$$n(2.17\text{c}) = nH(\tilde{\beta}) - \sum_{i=1}^{n} \log\left[1 + \exp\left(\mathbf{X}_i^T\tilde{\beta}\right)\right]$$

$$= nH(\beta_0) - \sum_{i=1}^{n} \log\left[1 + \exp(\mathbf{X}_i^T\beta_0)\right] \tag{2.22a}$$

$$+ nH'(\beta_0)^T(\tilde{\beta} - \beta_0) - \sum_{i=1}^{n} \frac{\exp(\mathbf{X}_i^T\beta_0)}{1 + \exp(\mathbf{X}_i^T\beta_0)}\mathbf{X}_i^T(\tilde{\beta} - \beta_0) \tag{2.22b}$$

$$+ \frac{1}{2}(\tilde{\beta} - \beta_0)^T\left\{nH''(\beta^*) - \sum_{i=1}^{n} \frac{\exp(\mathbf{X}_i^T\beta_n^*)}{\left[1 + \exp(\mathbf{X}_i^T\beta_n^*)\right]^2}\mathbf{X}_i\mathbf{X}_i^T\right\}(\tilde{\beta} - \beta_0). \tag{2.22c}$$

In addition,

$$2 \times (2.22c) = (\tilde{\beta} - \beta_0)^T \Bigg\{ nH''(\beta^*) - \left[\sum_{i=1}^{n} \frac{\exp(\mathbf{X}_i^T \beta_n^*)}{[1 + \exp(\mathbf{X}_i^T \beta_n^*)]^2} \mathbf{X}_i \mathbf{X}_i^T \right] \Bigg\} (\tilde{\beta} - \beta_0)$$

$$= (\tilde{\beta} - \beta_0)^T \Bigg\{ nH''(\beta_0) - \left[\sum_{i=1}^{n} \frac{\exp(\mathbf{X}_i^T \beta_0)}{[1 + \exp(\mathbf{X}_i^T \beta_0)]^2} \mathbf{X}_i \mathbf{X}_i^T \right] \Bigg\} (\tilde{\beta} - \beta_0) \tag{2.23a}$$

$$+ n(\tilde{\beta} - \beta_0)^T [H''(\beta^*) - H''(\beta_0)](\tilde{\beta} - \beta_0) \tag{2.23b}$$

$$+ (\tilde{\beta} - \beta_0)^T \Bigg\{ \sum_{i=1}^{n} \left[\frac{\exp(\mathbf{X}_i^T \beta_0)}{[1 + \exp(\mathbf{X}_i^T \beta_0)]^2} - \frac{\exp(\mathbf{X}_i^T \beta_n^*)}{[1 + \exp(\mathbf{X}_i^T \beta_n^*)]^2} \right] \mathbf{X}_i \mathbf{X}_i^T \Bigg\} (\tilde{\beta} - \beta_0). \tag{2.23c}$$

Define

$$L_{c1} = \sup\left\{ n \geq 1: (2.22a) \geq n \frac{m}{3 \times 3} \right\}, \tag{2.24}$$

$$L_{c2} = \sup\left\{ n \geq 1: \exists\, \tilde{\beta} \in \partial \mathbf{B}_\rho,\ (2.22b) \geq n \frac{m}{3 \times 3} \right\}, \tag{2.25}$$

$$L_{c3} = \sup\left\{ n \geq 1: \exists\, \tilde{\beta} \in \partial \mathbf{B}_\rho,\ (2.23a) \geq n \frac{m}{3 \times 3 \times 2} \right\}, \tag{2.26}$$

$$L_{c4} = \sup\left\{ n \geq 1: \exists\, \tilde{\beta} \in \partial \mathbf{B}_\rho,\ (2.23b) + (2.23c) \geq n \frac{m}{3 \times 3 \times 2} \right\}. \tag{2.27}$$

By Chow and Teicher [(1978) Theorem 10.4.3], $EL_{c1}^l < \infty$ if $E\|\mathbf{X}_1\|^{l+1} < \infty$. For L_{c2}, by the same arguments as for Lemma 2.2. and by Chow and Teicher (1978), $EL_{c2}^l < \infty$, if $E\|\mathbf{X}_1\|^{l+1} < \infty$.

LEMMA 2.4. *For $l \geq 1$, if $E\|\mathbf{X}_1\|^{2(l+1)} < \infty$, then $EL_{c3}^l < \infty$.*

Lemma 2.4 can be proved using Chow and Teicher [(1978), Theorem 10.4.3] and $\|\tilde{\beta} - \beta_0\| = \rho$. For details, see Chang and Martinsek (1991).

LEMMA 2.5. *For $l \geq 1$, if $E\|\mathbf{X}_1\|^{2(l+1)} < \infty$, then $EL_{c4}^l < \infty$.*

The proof of Lemma 2.5 is much more delicate than the proofs of the other lemmas. It depends on a careful analysis of the behavior of various eigenvalues as $\rho \to 0$. For details, see Chang and Martinsek (1991).

LEMMA 2.6. *For $l \geq 1$, if $E\|\mathbf{X}_1\|^{2(l+1)} < \infty$, then $EL_c^l < \infty$.*

PROOF OF LEMMA 2.6. From previous discussion, by applying Chow and Teicher [(1978) Theorem 10.4.3] directly, $EL_{c1}^l < \infty$ and $EL_{c2}^l < \infty$. By the definitions, it is clear that

$$L_c \leq \max\{L_{ci}, i = 1, \dots, 4\},$$

hence by Lemmas 2.4 and 2.5,

$$E_{Lc}^l < \infty. \tag{2.28}$$ □

Proposition 2.1 now follows easily from Lemmas 2.2, 2.3 and 2.6, combined with the inequality

$$L_\rho \leq \max(L_a, L_b, L_c) \quad \text{a.s.}$$

The proof of Lemma 2.1 is similar to that of Lemma 2.4 and is therefore omitted [for details of the proof, see Chang (1991)].

3. Simulations. In this section, some Monte Carlo simulation results are summarized. Three different kinds of covariate variables have been studied: bivariate normal with independent components, bivariate normal with positive ($\rho = 0.5$) correlation coefficient and bivariate normal with negative ($\rho = -0.5$) correlation coefficient. In all cases, we chose $\beta_0^T = (0.1, 0.2)$ and compute the $\hat{\beta}_n$ by the Newton–Raphson method. Three choices of a_n were tried: $a_n \equiv a$, $a_n = a(1 - 1/2n)$ and $a_n = a(1 - 2/3n)$. 1000 trials were used in each case. The results are summarized in Tables 1–3.

TABLE 1
Expected sample size and coverage frequency; 95% *confidence ellipsoid*; $a_n \equiv a$*

Distributions	d	Expected sample size (s.e.)	c.p. (%)	Best f.s.s.
Normal (independent)	0.5	29.94 (0.212)	98.5	3.95
	0.3	44.40 (0.257)	97.6	10.99
	0.1	114.82 (0.423)	96.1	98.89
Normal ($\rho = 0.5$)	0.5	54.01 (0.303)	97.9	8.79
	0.3	84.89 (0.392)	96.5	24.41
	0.1	236.51 (0.642)	95.6	219.69
Normal ($\rho = -0.5$)	0.5	53.89 (0.304)	98.0	8.65
	0.3	84.31 (0.389)	96.0	24.06
	0.1	234.84 (0.665)	95.7	216.23

*$2d$, width of maximum axis of confidence ellipsoid; s.e., standard error of expected sample size based on 1000 trials; c.p., coverage probability; best f.s.s., best fixed sample size.

TABLE 2
Expected sample size and coverage frequency; 95% *confidence ellipsoid*; $a_n = a(1 - 1/2n)$*

Distributions	d	Expected sample size (s.e.)	c.p. (%)	Best f.s.s.
Normal (independent)	0.5	12.71 (0.219)	97.1	3.95
	0.3	21.48 (0.247)	96.5	10.99
	0.1	110.13 (0.381)	95.9	98.89
Normal ($\rho = 0.5$)	0.5	27.17 (0.287)	96.4	8.79
	0.3	50.26 (0.391)	95.9	24.41
	0.1	225.32 (0.671)	95.3	219.69
Normal ($\rho = -0.5$)	0.5	26.88 (0.281)	96.6	8.65
	0.3	49.91 (0.399)	95.6	24.06
	0.1	221.47 (0.669)	95.1	216.23

*$2d$, width of maximum axis of confidence ellipsoid; s.e., standard error of expected sample size based on 1000 trials; c.p., coverage probability; best f.s.s., best fixed sample size.

When the best fixed sample size is relatively small, all three choices of a_n produce expected sample sizes that are too large, although the relative discrepancy decreases as d becomes small, in accordance with the preceding theorems. The choices $a_n = a(1 - 1/2n)$ and $a_n = a(1 - 2/3n)$, especially the latter, substantially outperform $a_n \equiv a$. For larger best fixed sample sizes (i.e., smaller values of d), the agreement between asymptotic theory and simulations is much better. Again, the choices $a_n = a(1 - 1/2n)$ and $a_n = a(1 - 2/3n)$ do better than $a_n \equiv a$, and they produce results that are very close to the theoretical values.

It is known in related situations (see Muirhead and Chikuse, 1975) that the bias of the smallest sample eigenvalue is of order $1/n$, which suggests correcting a, and hence a^2, by this order, as in two of the choices above. Even such corrections may not produce good results when the best fixed sample size is

TABLE 3
Expected sample size and coverage frequency; 95% *confidence ellipsoid*; $a_n = a(1 - 2/3n)$*

Distributions	d	Expected sample size (s.e.)	c.p. (%)	Best f.s.s.
Normal (independent)	0.5	10.18 (0.234)	96.9	3.95
	0.3	16.17 (0.251)	96.4	10.99
	0.1	108.29 (0.401)	95.6	98.89
Normal ($\rho = 0.5$)	0.5	13.79 (0.291)	96.3	8.79
	0.3	30.16 (0.390)	95.6	24.41
	0.1	221.13 (0.661)	94.9	219.69
Normal ($\rho = -0.5$)	0.5	14.11 (0.301)	96.4	8.65
	0.3	29.39 (0.396)	95.5	24.06
	0.1	218.54 (0.674)	95.1	216.23

*$2d$, width of maximum axis of confidence ellipsoid; s.e., standard error of expected sample size based on 1000 trials; c.p., coverage probability; best f.s.s., best fixed sample size.

small, although it is clear from the simulations that they will help. At least in such cases the procedures are conservative, so that one can count on the coverage probability. On the other hand, simple corrections like those above produce good results for smaller values of d, that is, when greater precision is desired.

It is clear that the best choice of a_n depends on the underlying distribution and on β_0 (both assumed unknown). Note, for example, that for large d, $a_n = a(1 - 2/3n)$ provides a much bigger improvement over the other two choices when the covariates are correlated, rather than independent. Based on the simulation study, we recommend choosing a_n to increase to a at rate $1/n$, and $a_n = a(1 - 2/3n)$ seems like a reasonably good ad hoc choice. In cases where the experimenter has some idea of the distribution of the covariates, simulations similar to the ones described here, for a variety of β_0's, may be helpful in choosing the design parameters a_n.

4. Some related fixed size confidence sets.

4.1. "*Fixed proportional accuracy*" *confidence ellipsoids.* In the one dimensional case, when β_0 (assumed to be nonzero) is near the origin, one may wish to have a smaller confidence interval than that for a β_0 which is far away from the origin. One approach is to require that the confidence interval specify β_0 to within a certain fraction of its true value (fixed proportional accuracy). However, the fixed proportional accuracy problem in the higher dimensional case is a little more complicated. By modifying the idea in the one dimensional case, we may first wish to define a confidence ellipsoid as

$$\Gamma_n = \left\{\mathbf{Z} \in \mathbf{R}^p \colon \left(\mathbf{Z} - \hat{\beta}_n\right)^T \hat{\Sigma}_n\left(\mathbf{Z} - \hat{\beta}_n\right) \le d^2 \hat{\lambda}_n \|\hat{\beta}_n\|\right\}.$$

If all the coordinates of β_0 are small (in absolute value), then this could be an appropriate way to define the confidence ellipsoid. But if only some of the coordinates are small and some of them are relatively large, then the above definition of the confidence ellipsoid does not give us any improvement in accuracy of the estimates of small coordinates. To rectify this situation, first assume that all the coordinates of β_0 are nonzero. (Otherwise, we can just eliminate those coordinates from the model and reduce the dimension of the model.) Now, define

$$\Gamma_n = \left\{\mathbf{Z} \in \mathbf{R}^p \colon \left(\mathbf{Z} - \hat{\beta}_n\right)^T \hat{\Sigma}_n\left(\mathbf{Z} - \hat{\beta}_n\right) \le d^2 \hat{\lambda}_n \hat{b}_n\right\}, \quad \text{for } d \in (0,1),$$

where $\hat{b}_n = \min_{1 \le j \le p} |\hat{\beta}_{nj}|$. Γ_n defines an ellipsoid with maximum axis less than or equal to $2d\sqrt{\hat{b}_n}$.

For any given $\alpha \in (0,1)$ and $d > 0$, it is desired to have

$$P[\beta_0 \in \Gamma_n] \approx 1 - \alpha. \tag{4.1}$$

Since $\hat{\beta}_n \to \beta_0$ almost surely, $\hat{b}_n \to b = \min_{1 \le j \le p} |\beta_{0j}|$ a.s. as $n \to \infty$, and therefore

$$(4.2)\qquad (\hat{\beta}_n - \beta_0)^T \hat{\Sigma}_n(\hat{\beta}_n - \beta_0)/\hat{b}_n \to_{\mathscr{L}} \chi^2(p)/b.$$

Hence, to satisfy (4.1), one should have $n \approx a^2/\lambda^{(p)}bd^2$, where a^2 is defined as before. Here both $\lambda^{(p)}$ and b are unknown, so it is impossible to decide the sample size in advance. This suggests a stopping time

$$(4.3)\qquad \tau_d = \inf\left\{n \geq 1 \colon \hat{\lambda}_n \hat{b}_n \geq \frac{a^2}{d^2}\right\}.$$

The following theorem can be obtained by similar arguments to those in the previous section. For details, see Chang and Martinsek (1991).

THEOREM 4.1. *Suppose* $E\|\mathbf{X}_1\|^3 < \infty$ *and all the coordinates of* β_0 *are nonzero. Then*:

(i) $\lim_{d\to 0} bd^2\lambda^{(p)}\tau_d/a^2 = 1$ *a.s.*

(ii) $\lim_{d\to 0} P\{\beta_0 \in \Gamma_{\tau_d}\} = 1 - \alpha$.

Moreover, if $E\|\mathbf{X}\|^4 < \infty$,

(iii) $\lim_{d\to 0} E[bd^2\lambda^{(p)}\tau_d/a^2] = 1$.

4.2. *Confidence interval for a linear combination of* β_0. In practice, we may be interested only in a particular linear combination of the components of β_0, rather than the whole vector. That is, for some $C \in \mathbf{R}^p$, $\|C\| \neq 0$, we would like to construct a fixed width confidence interval for $C^T\beta_0$. It follows from the asymptotic normality of $\hat{\beta}_n$ that, as $n \to \infty$,

$$(4.4)\qquad \sqrt{n}\,(C^T\hat{\beta}_n - C^T\beta_0) \to_{\mathscr{L}} N(0, C^T\Sigma^{-1}C).$$

If Σ were known,then for a given $d > 0$, and $\alpha \in (0,1)$, $[C^T\hat{\beta}_n - d, C^T\hat{\beta}_n + d]$ could be used as a confidence interval for $C^T\beta_0$ with approximate coverage probability $1 - \alpha$, provided that

$$(4.5)\qquad \frac{nd^2}{z_{\alpha/2}^2} \approx C^T\Sigma^{-1}C,$$

where $z_{\alpha/2}$ is defined as before. Because Σ is unknown, a sequential procedure is needed. Equation (4.5) suggests the stopping rule

$$(4.6)\qquad T_d^{\gamma} = \inf\left\{n \geq 1 \colon C^T\hat{\Sigma}_n^{-1}C \leq \frac{d^2}{z_{\alpha/2}^2}\right\}.$$

We have the following results:

THEOREM 4.2. *If* $E\|\mathbf{X}\|^3 < \infty$, *then*:

(i) $\lim_{d\to 0} d^2T_d^{\gamma}/(z_{\alpha/2}^2C^T\Sigma^{-1}C) = 1$ *a.s.*

(ii) $\lim_{d\to 0} P\{C^T\beta_0 \in [C^T\hat{\beta}_{T_d^{\gamma}} - d, C^T\hat{\beta}_{T_d^{\gamma}} + d]\} = 1 - \alpha$.

Moreover, if $E\|\mathbf{X}\|^4 < \infty$,

(iii) $\lim_{d \to 0} E[d^2 T_d^\gamma / (z_{\alpha/2}^2 C^T \Sigma^{-1} C)] = 1.$

Theorem 4.2 is proved by dominating the stopping rule T_d^γ by a version of the stopping rule T_d. For details, see Chang and Martinsek (1991).

Acknowledgment. The authors would like to thank a referee and an Associate Editor for comments that helped improve the paper and for pointing out a flaw in our original proof of Theorem 1.1.

REFERENCES

ALBERT, A. (1966). Fixed size confidence ellipsoids for linear regression parameters. *Ann. Math. Statist.* **37** 1602–1630.

BELLMAN, R. E. (1960). *Introduction to Matrix Analysis*. McGraw-Hill, New York.

BILLINGSLEY, P. (1986). *Probability and Measure*, 2nd ed. Wiley, New York.

CHANG, Y.-C. I. (1991). Some sequential estimation problems in logistic regression models. Ph.D. dissertation, Univ. Illinois, Urbana-Champaign.

CHANG, Y.-C. I. and MARTINSEK, A. T. (1991). Fixed size confidence regions for parameters of a logistic regression model. Technical Report 61, Dept. Statistics, Univ. Illinois, Urbana–Champaign.

CHOW, Y. S. and LAI, T. L. (1975). Some one-sided theorems on the tail distribution of sample sums with applications to the last time and largest excess of boundary crossings. *Trans. Amer. Math. Soc.* **208** 51–72.

CHOW, Y. S. and ROBBINS, H. (1965). On the asymptotic theory of fixed-width confidence intervals for the mean. *Ann. Math. Statist.* **36** 457–462.

CHOW, Y. S. and TEICHER, H. (1978). *Probability Theory: Independence, Interchangeability, Martingales*. Springer, New York.

FINSTER, M. (1985). Estimation in the general linear model when the accuracy is specified before data collection. *Ann. Statist.* **13** 663–675.

GLESER, L. J. (1965). On the asymptotic theory of fixed-size sequential confidence bounds for linear regression parameters. *Ann. Math. Statist.* **36** 463–467.

GLESER, L. J. (1969). On limiting distributions for sums of a random number of independent random variables. *Ann. Math. Statist.* **40** 935–941.

GOLUB, G. H. and VAN LOAN, C. F. (1983). *Matrix Computations*. Johns Hopkins Univ. Press.

GRAMBSCH, P. (1983). Sequential sampling based on the observed Fisher information to guarantee the accuracy of the maximum likelihood estimator. *Ann. Statist.* **11** 68–77.

GRAMBSCH, P. (1989). Sequential maximum likelihood estimation with applications to logistic regression in case-control studies. *J. Statist. Plann. Inference* **22** 355–369.

LAI, T. L. and SIEGMUND, D. (1977). A nonlinear renewal theory with applications to sequential analysis I. *Ann. Statist.* **5** 946–954.

LAI, T. L. and SIEGMUND, D. (1979). A nonlinear renewal theory with applications to sequential analysis II. *Ann. Statist.* **7** 60–76.

MUIRHEAD, R. J. and CHIKUSE, Y. (1975). Asymptotic expansions for the joint and marginal distributions of the latent roots of the covariance matrix. *Ann. Statist.* **3** 1011–1017.

PRESS, W. H., FLANNERY, B. P., TEUKOLSKY, S. A. and VETTERLING, W. T. (1985). *Numerical Recipes*. Cambridge Univ. Press.

SRIVASTAVA, M. S. (1967). On fixed-width confidence bounds for regression parameters and mean vector. *J. Roy. Statist. Soc. Ser. B* **29** 132–140.

SRIVASTAVA, M. S. (1971). On fixed-width confidence bounds for regression parameters. *Ann. Math. Statist.* **42** 1403–1411.

STEFANSKI, L. A. and CARROLL, R. J. (1985). Covariance measurement error in logistic regression. *Ann. Statist.* **13** 1335–1351.

STEIN, C. (1945). A two sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.* **16** 243–252.

STEIN, C. (1949). Some problems in sequential estimation. *Econometrica* **17** 77–78.

WILKINSON, J. H. (1965). *The Algebraic Eigenvalue Problem*. Clarendon Press, Oxford.

WOODROOFE, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*. SIAM, Philadelphia.

YU, K. F. (1989). On fixed-width confidence intervals associated with maximum likelihood estimation. *J. Theoret. Probab.* **2** 193–199.

DEPARTMENT OF STATISTICS
UNIVERSITY OF ILLINOIS
101 ILLINI HALL
725 SOUTH WRIGHT STREET
CHAMPAIGN, ILLINOIS 61820