Statistica Sinica **6**(1996), 899-916

SEQUENTIAL FIXED SIZE CONFIDENCE REGIONS FOR REGRESSION PARAMETERS IN GENERALIZED LINEAR MODELS

Yuan-chin Ivan Chang

Academia Sinica

Abstract: Sequential procedures for constructing fixed size confidence regions for regression parameters in generalized linear models using maximum likelihood estimators are proposed in this paper. We consider the cases of natural link function (l.f.) and nonnatural l.f., separately. Stopping times are proposed when the scale parameter is known and unknown. In either case, the asymptotic consistency and efficiency of the sequential procedures are established under regularity conditions similar to those in Fahrmeir and Kaufmann (1985). Moreover, when the scale parameter is known, we establish the asymptotic normality of the appropriately standardized stopping time.

Key words and phrases: Generalized linear models, fixed size confidence set, sequential estimation, stopping rule, last time, uniform integrability, asymptotic efficiency.

1. Introduction

Generalized linear models (Nelder and Wedderburn (1972)) are regression models for a number of cases where the classical assumptions are not met. Consider data $\{(x_i, y_i), i = 1, \ldots, n\}$, where a case consists of a response y accompanied a $p \times 1$ vector x of observed explanatory variables. A generalized linear model (McCullagh and Nelder (1989)) for these data has two parts: (a) the random component, specifying a probability density function for y of the form

$$P(y|\theta, \phi) = \exp\Big\{\frac{[y\theta - b(\theta)]}{a(\phi)} + c(y, \phi)\Big\}, \tag{1.1}$$

where θ and ϕ are scale parameters, and $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$ are known functions, and (b) the systematic component, linking y to the explanatory variable, $\theta = u(x^T\beta)$, where β is a $p \times 1$ vector of regression coefficients. The function $u(\cdot)$ will be called a θ-link to distinguish it from the conventional link function relating x to the mean of y; the θ-link function is equivalent, and more convenient for differentiating with respect to β. Assume Θ to be the natural parameter space; then Θ is convex, and in the interior Θ^0 of Θ, all derivatives of $b(\theta)$ and all moments of y exist (assume $\Theta^0 \neq \emptyset$). In this case, (1.1) is an exponential-family with natural parameter θ. The log-likelihood of regression coefficients for a sample of size n is

$$l_n(\beta) = \frac{1}{a(\phi)} \sum_{i=1}^{n} [y_i u(x_i^T \beta) - b(u(x_i^T \beta))], \tag{1.2}$$

omitting terms that do not involve β. The maximum likelihood estimate $\hat{\beta}_n$ is usually computed by iteratively reweighted least square. In Fahrmeir and Kaufmann (1985, 1986), they proved that $\hat{\beta}_n$, the MLE, is strongly consistent and asymptotically normal with asymptotic covariance matrix $a \cdot \Sigma^{-1}$, where

$$\Sigma(\beta) = E\{b''(u(x^T\beta))[u'(x^T\beta)]^2 xx^T\}, \tag{1.3}$$

is usually a complicated function of β, with $u'(\cdot)$ and $b''(\cdot)$ denoting the first and the second derivatives of $u(\cdot)$ and $b(\cdot)$, respectively.

Based on the asymptotic normality of $\hat{\beta}_n$, we can construct a confidence set for the regression coefficient β. For simplicity, we assume that the scale parameter ϕ is known or can be estimated by $\hat{\phi}_n$. Let $\hat{a}_n \equiv a(\hat{\phi}_n)$, and

$$\hat{\Sigma}_n = \frac{1}{n} \sum_{i=1}^{n} b''(u(x_i^T \hat{\beta}_n))[u'(x_i^T \hat{\beta}_n)]^2 x_i x_i^T. \tag{1.4}$$

A large sample $100(1-\alpha)\%$ confidence ellipsoid for the $p \times 1$ vector of regression coefficients is given by

$$A_n = A(\hat{\beta}_n, \alpha) = \{\beta : n(\beta - \hat{\beta}_n)^T \, \hat{a}_n^{-1} \, \hat{\Sigma}_n(\beta - \hat{\beta}_n) \leq c^2\}, \tag{1.5}$$

where c is a constant satisfying $P\{\chi_p^2 \geq c^2\} = \alpha$. The length of the maximum axis of A_n is $2(c^2 \hat{a}_n/(n\hat{\lambda}_n))^{1/2}$, where $\hat{\lambda}_n$ is the smallest eigenvalue of $\hat{\Sigma}_n$. Therefore, whether the scale parameter ϕ is known or unknown, the length of the maximum axis of A_n will still depend on $\hat{\lambda}_n$ and, hence, be random.

For any confidence set (CS), there are at least two important requirements: coverage probability and precision. That is, for a given $\alpha \in (0, 1)$, we wish to have $P_\theta(\theta \in CS) \approx 1-\alpha$, for each $\theta \in \Theta$. On the other hand, it is undesirable to make an imprecise statement, even if it can be made with great confidence (as an extreme example, note that the entire parameter space Θ is a 100% confidence set). Now, suppose we require further that the length of the maximum axis of A_n is no greater than $2d$, for some given $d > 0$; i.e. $(c^2\hat{a}_n/(n\hat{\lambda}_n)) \leq d^2$. Assume that the smallest eigenvalue of Σ, λ_p, and the scale parameter ϕ are known for the moment. Then, the best fixed sample size is

$$n_{opt} \approx \Big[\frac{c^2 a(\phi)}{d^2 \lambda_p}\Big]. \tag{1.6}$$

Since Σ depends on the unknown regression coefficients, λ_p is also unknown. Hence, there is no fixed sample size that can be used to construct a confidence ellipsoid for β with prescribed coverage probability ($\alpha \in (0,1)$) and precision ($d > 0$), simultaneously. We will show that a sequential procedure can be used to achieve our goals (asymptotically).

Suppose that ϕ is known for the moment. It follows from the strong consistency of the MLE, that $\hat{\Sigma}_n$ can also be proved to be a strongly consistent estimate of Σ. This implies that $\hat{\lambda}_n$ will converge to λ_p almost surely as n goes to infinity. Hence, by replacing λ_p in (1.6) by its strongly consistent estimate $\hat{\lambda}_n$, we have the following stopping rule:

$$T_{d,1} = \inf\Big\{n \geq 1 : n\hat{\lambda}_n \geq \frac{c^2 a(\phi)}{d^2}\Big\}. \tag{1.7}$$

When ϕ is known, we rewrite

$$A_n = \{\beta : n(\beta - \hat{\beta}_n)^T\, a^{-1}(\phi)\, \hat{\Sigma}_n(\beta - \hat{\beta}_n) \leq c^2\}. \tag{1.8}$$

When the sampling is stopped, $A_{T_{d,1}}$ will be used as a confidence ellipsoid for β. Sufficient conditions to establish the asymptotic properties of the stopping rule $T_{d,1}$ will be based on the Taylor expansions Theorem.

Let $b'(\cdot)$, $b''(\cdot)$ and $u'(\cdot)$, $u''(\cdot)$ denote the first and the second derivatives of $b(\cdot)$ and $u(\cdot)$, respectively. For each i, let's $\theta_i = u(x_i^T\beta)$ and

$$\begin{aligned} S_i(\beta) &= [y_i - b'(\theta_i)]u'(x_i^T\beta)x_i, \\ R_i(\beta) &= [y_i - b'(\theta_i)]u''(x_i^T\beta)x_i x_i^T, \\ F_i(\beta) &= b''(\theta_i)[u'(x_i^T\beta)]^2 x_i x_i^T. \end{aligned} \tag{1.9}$$

Then it follows from (1.2) that the first and the second derivatives of the log-likelihood function are

$$\frac{\partial}{\partial\beta} l_n(\beta) = a^{-1}(\phi)\sum_{i=1}^n S_i(\beta) \tag{1.10}$$

and

$$\frac{\partial^2}{\partial\beta^2} l_n(\beta) = a^{-1}(\phi)\sum_{i=1}^n [R_i(\beta) - F_i(\beta)]. \tag{1.11}$$

It is clear that if $u(\cdot) = \textit{identity}$, then $R_i(\beta) = 0$ for each i and $\partial^2 l_n(\beta)/\partial\beta^2 = -a^{-1}(\phi)\sum_{i=1}^n F_i(\beta)$. This property of the natural link function is of great advantage in proving the asymptotic properties of the MLE as well as in proving the asymptotic properties of the sequential procedures proposed here. The conditions give here are similar to those of Fahrmeir and Kaufmann (1985), for establishing the strong consistency and the asymptotic normality of ML estimates in generalized linear models. As in Fahrmeir and Kaufmann (1985), we also treat natural- and nonnatural-link function cases separately.

Now turn to the unknown scale parameter case. By (1.6), we need a strongly consistent estimator of $a(\phi)$ in order to have a useful stopping time that will be finite almost surely under this circumstance. The most commonly used estimator of $a(\phi)$ is $\hat{a}_n = a(\hat{\phi}_n) = n^{-1}\sum_{i=1}^{n}[y_i - b'(x_i^T\hat{\beta}_n)]^2/b''(x_i^T\hat{\beta}_n)$, which is usually called a "generalized Pearson χ^2" statistic (see McCullagh and Nelder (1989) or Hinkley, Reid and Snell (1991)). Now, let

$$A'_n = \{\beta : n(\beta - \hat{\beta}_n)^T\, \hat{a}_n^{-1}\, \hat{\Sigma}_n(\beta - \hat{\beta}_n) \le c^2\}$$

and

$$T_{d,2} = \inf\Big\{n \ge n_d : n\hat{\lambda}_n \cdot \hat{a}_n^{-1} \ge \frac{c^2}{d^2}\Big\}, \tag{1.12}$$

where n_d is some positive integer and depends on d. Then the asymptotic consistency still holds.

Chow and Robbins (1965) gave a very useful method for constructing a confidence interval for an unknown mean with prescribed coverage probability and precision. They also proved that the sequential procedure which they presented in their paper 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 confidence interval approaches 0). Their ideas have been extended to many different models (Gleser (1965), Albert (1966), Srivastava (1967), Mukhopadhyay (1974) and Finster (1985)). Recently, Chang and Martinsek (1992) proposed a sequential procedure for constructing fixed size confidence regions for logistic regression models.

In this paper, we extend the idea of Chang and Martinsek (1992) and consider a unifying approach for constructing fixed size confidence regions for the regression coefficients in some generalized linear models under both natural link function ($u(\cdot) \equiv$ *identity*) and non-natural link function ($u(\cdot) \neq$ *identity*) setups; and for either case, both known and unknown scale parameter cases are considered. The stopping rules proposed here are shown to be "asymptotically consistent and efficient." Furthermore, we also show the asymptotic normality of our stopping rule for the model with known scale parameter case.

This paper is organized in the following way. We state the results for generalized linear models with and without natural link functions in Section 2 and Section 3, respectively. In Section 4, we discuss some commonly used generalized linear models as examples. Proofs of theorems and lemmas will be given in Section 5.

2. Natural Link Function

Suppose (x_i, y_i), $i = 1, 2, \ldots$, are independent observations from a generalized linear model; that is, for each i, the conditional density of y_i given x_i has the form (1.1) with $\theta_i = u(x_i^T\beta_0) = x_i^T\beta_0$, where β_0 denote the unknown true regression coefficients of interest. (Note that when y_i is a discrete random variable, (1.1) denotes a conditional mass function. For convenience, we still call this a density function throughout this paper.) Assume further that

(R_s) (i) $EF_1(\beta)$ exists and is positive definite,

(ii) $E\max_{\beta\in N}\|F_1(\beta)\|$ exists for a compact neighborhood N of β_0,

where $F_1(\beta) = b''(x_1^T\beta)x_1x_1^T$;

(M_s) (i) for some $\delta \geq 0$, $E\sup_{\beta\in N}\|F_1(\beta)\|^{2+\delta}$ exists for a compact neighborhood N of β_0,

(ii) there exists a real-valued function $\underline{b}''(\cdot)$, which is symmetric about 0 and non-increasing in R^+ such that $0<\underline{b}''(\theta)\leq b''(\theta)$, for all $\theta \in \Theta$.

Remark. If $\Theta = \mathcal{R}^1$ or Θ is compact then we can construct a real-valued function $\underline{b}''(t)$ as follows: Extend the support of $b''(\theta)$ to $\mathcal{R}^1$ by letting

$$b''(s) = \begin{cases} b''(s), & \text{if } s \in \Theta, \\ \inf\limits_{\theta\in\Theta} b''(\theta), & \text{if } s \in \mathcal{R}^1 - \Theta. \end{cases} \tag{2.1}$$

Define

$$\underline{b}''(t) = \inf\{b''(s) : |s| \leq |t|,\ t \in \mathcal{R}'\}; \tag{2.2}$$

then, by definition, $\underline{b}''(t)$ is symmetric about 0 and non-increasing in $\mathcal{R}^+$.

2.1. Known scale parameter

Theorem 2.1. *Suppose (R_s) holds. Then, for a given $\alpha \in (0,1)$, the stopping rule $T_{d,1}$ is finite almost surely for any $d > 0$, and*
(i) $\lim_{d\to 0}[\frac{T_{d,1}}{n_{opt}}] = 1$, *almost surely,*
(ii) $\lim_{d\to 0} P\{\beta_0 \in A_{T_{d,1}}\} = 1 - \alpha$. (*asymptotic consistency*)
If, in addition, (M_s) is satisfied, then
(iii) $\lim_{d\to 0} E[\frac{T_{d,1}}{n_{opt}}] = 1$. (*asymptotic efficiency*)

The terms "asymptotic consistency" and "asymptotic efficiency" are due to Chow and Robbins (1965).

We now turn to the asymptotic normality of the stopping rule $T_{d,1}$. Let $\Sigma_0 = \Sigma(\beta_0)$ and assume the scale parameter ϕ is known. Then the asymptotic normality of $T_{d,1}$ will follow from the asymptotic result of $\hat{\lambda}_n$ (see Waternaux (1976) and Davis (1977)). Without loss of generality, let us assume $a(\phi) = 1$. Suppose that $\lambda_1 > \cdots > \lambda_p$ are eigenvalues of Σ_0. By definition of Σ_0, there exists an orthonormal matrix Γ such that $\Gamma^T\Sigma_0\Gamma = diag(\lambda_1, \ldots, \lambda_p)$. Let $\gamma_i = \Gamma^T[u(x_i^T\beta_0)^{\frac{1}{2}}x_i]$, $i = 1, 2, \ldots$ Assume that γ_1 has finite fourth cummulant and let κ denote the fourth cummulant of the pth component of γ_1. Then, we have

Theorem 2.2. *Assume γ_1 has finite fourth cummulant and all the eigenvalues of Σ_0 are distinct. Then, as $d \to 0$,*

$$n_{opt}^{-\frac{1}{2}}(T_{d,1} - n_{opt}) \xrightarrow{d} N\Big(0, \frac{\kappa}{\lambda_p^2} + 2\Big),$$

where κ is defined as above. Moreover, if the $x_i's$ are normally distributed, then

$$n_{opt}^{-\frac{1}{2}}(T_{d,1} - n_{opt}) \xrightarrow{d} N(0, 2),$$

as $d \to 0$.

2.2. Unknown scale parameter

For the case with unknown scale parameter, suppose that there exists a real-valued function $h(\theta)$, which is symmetric about 0 and non-decreasing in $\mathcal{R}^+$, such that

(P_s) (i) $b''(\theta) \le h(\theta)$, $\forall\theta \in \Theta$, and

(ii) $E \sup_{\beta \in N} h^2(x_1^T\beta)\|x_1\|^2$ exists for a compact neighborhood N of β_0;

then we have asymptotic efficiency of the stopping rule $T_{d,2}$.

Theorem 2.3. *Under the assumptions of Theorem 2.1, the stopping rule $T_{d,2}$ is finite almost surely for any $d > 0$. Moreover,*

(i) $\lim_{d\to 0}[\frac{T_{d,2}}{n_{opt}}] = 1$, *almost surely,*

(ii) $\lim_{d\to 0} P\{\beta_0 \in A'_{T_{d,2}}\} = 1 - \alpha$.

If, in addition, (M_s) and (P_s) are satisfied, then

(iii) $\lim_{d\to\infty} E[\frac{T_{d,2}}{n_{opt}}] = 1$.

Remark. In the proof of asymptotic efficiency of the stopping rule here, we use the "last time" technique. Unlike the last time considered in Chow and Lai (1975), the last time we use here is the supremum of uncountably many last times. The same idea has been used in Chang and Martinsek (1992) for logistic regression models. Here we extend their idea to a more general setup.

3. Non-Natural Link Function

In this section, we discuss generalized linear models for the non-natural link function case. Although we only give theorems for the case when the scale parameter is known, these results can be extended to the case with unknown scale parameter by similar arguments as in the proof of Theorem 2.3.

As in the previous section, the strong consistency and asymptotic normality of the MLE of β_0 will be sufficient for proving the "asymptotic consistency" of the proposed sequential procedure here. But the proof of asymptotic property of the sequential procedure under the non-natural link function setup will need some extra conditions.

The strong consistency and asymptotic normality of the MLE of a generalized linear model with non-natural link function can also be found in Fahrmeir and Kaufmann (1985, 1986). They proved that, if

(R_s^*) (i) $EF_1(\beta)$ exists and is positive definite,
(ii) $E\max_{\beta\in N}\|F_1(\beta)\|$ and $E\max_{\beta\in N}\|R_1(\beta)\|$ exist for a compact neighborhood N of β_0,

then $\hat\beta_n \to \beta_0$, almost surely, as $n\to\infty$ and $\hat\Sigma_n^{1/2}(\hat\beta_n-\beta_0) \xrightarrow{d} N(0,a(\phi)I)$, as $n\to\infty$, where $\hat\Sigma_n$ is defined in (1.4). From (1.6), we know that the best fixed sample size n_{opt} will depend on both λ_p and ϕ (and together with some given constants $\alpha\in(0,1)$ and $d>0$). The formulation of the stopping rule in the non-natural link function case will have a similar form to that in the natural link function case (but estimators $\hat\lambda_n$ and $\hat a_n$ are different). The proof of the Theorem under the non-natural link function case can be done by similar arguments as long as there exists an appropriate last time random variable which can be shown to be integrable under the current setup.

Remark. By the strong consistency of $\hat\beta_n$, we have that $n^{-1}\sum_{i=1}^n R_i(\hat\beta_n)\to 0$ almost surely as $n\to\infty$. Hence, $\hat\lambda_n = \Lambda_{\min}(n^{-1}\sum_{i=1}^n F_i(\hat\beta_n))$ converges to λ_p and $\hat a_n$ converges to $a(\phi)$.

We assume that ϕ is known throughout this section. For a detailed discussion about the cases with unknown scale parameter ϕ, see Chang (1992). Now, let $T_{d,1}$ be a stopping rule with a form (1.7) and A_n be a confidence ellipsoid with a form (1.8). Then for any $\alpha\in(0,1)$ and $d>0$, we have

Theorem 3.1. *Suppose (R_s^*) holds. Then $T_{d,1}$ is finite almost surely for any $d>0$ and $T_{d,1}$ will go to infinity as d approaches 0. Moreover,*
(i) $\lim_{d\to 0}(T_{d,1}/n_{opt})=1$, *almost surely, and*
(ii) $\lim_{d\to 0}P\{\beta_0\in A_{T_{d,1}}\}=1-\alpha$.

Assume further that $u(t)$ is three times continuously differentiable for all $t \in R$ and let $u^{(i)}$, $i = 1, 2, 3$, denote the first, the second and the third derivatives of u, respectively. For any $t \in \mathcal{R}$ and $i = 0, \ldots, 3$, let

$$\bar{u}^{(i)}(t) = \sup\{u^{(i)}(s) : |s| \leq |t|, s \in \mathcal{R}\},$$

where $u^{(0)} = u$. Then, $\bar{u}^{(i)}(t)$ is non-decreasing in $t \in R^+$ and symmetric about 0. Suppose that

(M_s^*) (i) $E\left(|b'(x_1^T\beta_0)u''(x_1^T\beta_0)|\,\|x_1\|^2\right)^{2+\delta} < \infty$,

(ii) $E[h(\eta_1)[\bar{u}^{(1)}]^2\|x_1\|^2]^{2+\delta} < \infty$, and

(iii) $E\left(|b'(x_1^T\beta)|\,|\bar{u}^{(3)}(\eta_1)|\,\|x_1\|^3\right)^{2+\delta} < \infty$,

(iv) $E\left[h(\bar{u}(\eta_1))|\bar{u}^{(1)}(\eta_1)\bar{u}^{(2)}(\eta_1)|\,\|x_1\|^3\right]^{2+\delta} < \infty$,

(v) there exists a real-valued function $\underline{b}''(\cdot)$, which is sysmetric about 0 and non-increasing in R^+ such that $0 < \underline{b}''(\theta) \leq b''(\theta)$, for all $\theta \in \Theta$,

where $\eta_1 = |x_1^T\beta_0| + \|x_1\| \cdot \rho$. Then, we have asymptotic efficiency of $T_{d,1}$ under the non-natural link function case; i.e.

Theorem 3.2. *Let (R_s^*) hold and (M_s^*) be satisfied for some $\delta \geq 0$. Then*

$$\lim_{d \to 0} E[T_{d,1}/n_{opt}] = 1.$$

4. Examples

In this section, we give three commonly used generalized linear models (with natural link function) and verify the conditions of the theorems stated in the above sections. We will only concentrate on the conditions required for the asymptotic efficiency of the stopping time for these models. (For strong consistency and asymptotic normality of MLE, see Fahrmeir and Kaufmann (1985, 1986).)

Example 1. Poisson Regression Model.

Suppose $\{y_n\}_{n \in N}$ are random variables having mass functions

$$P(y_n = y) = \exp(y\theta_n - e^{\theta_n})/y!, \quad y = 0, 1, 2, \ldots, \quad n = 1, 2, \ldots,$$

where $\theta_n = x_n^T\beta_0$. Then $b(\theta) = e^\theta$ and $F_n(\beta) = \exp(x_n^T\beta)x_n x_n^T$, $n \in N$. Assumption (M_s) becomes $E\max_{\beta \in N}\|\exp(x_n^T\beta)x_n x_n^T\|^{2+\delta} < \infty$, for some $\delta \geq 0$. Let $h(\theta) = e^{|\theta|}$, for all $\theta \in R$, and suppose that $E\sup_{\beta \in N} h(x_1^\beta)\|x_1\|^2 < \infty$. Then Assumption (P_s) is satisfied.

Example 2. Gamma Model.

Consider a family of gamma densities

$$F(y|\theta,\gamma)=\Gamma(\gamma)^{-1}(-\theta)^{\gamma}y^{\gamma-1}\exp(\theta y),\quad y\geq 0,$$

for a fixed shape parameter $\gamma>0$. The natural parameter space is $\Theta=(-\infty,0)$ and $b(\theta)=-\gamma l_n(-\theta)$, $b'(\theta)=-\gamma\theta^{-1}$, $b''(\theta)=\gamma\theta^{-2}$. Let $\{y_n\}_{n\in N}$ be a sequence of gamma distributed, independent random variables with natural parameter $\theta_n = x_n^T\beta_0$ for each $n\in N$. $F_n(\beta)=\gamma(x_n^T\beta)^{-2}x_nx_n^T$. Then, by setting $\underline{b}''(\theta)=b''(\theta)$, Assumption (M_s) will be satisfied provided that $E\max_{\beta\in N}(|x_n^T\beta|^{-2}\|x_n\|^2)^{2+\delta}<\infty$, for some $\delta\geq 0$. Hence, Theorems 2.1 and 2.2 can be applied. (This requirement restricts the possible sequence of regressors. For a detailed discussion about this, see Fahrmeir and Kaufmann (1985, 1986).)

Example 3. Binomial Model.

Let $\{y_n\}_{n\in N}$ be a random sample. For each n, y_n has a binomial(m,π_n) mass function. Under the natural link setup, $\theta_n=\log\frac{\pi_n}{1-\pi_n}=x_n^T\beta_0$ and $b(\theta_n)=\log(1+e^{\theta_n})$, $b'(\theta_n)=e^{\theta_n}/(1+e^{\theta_n})$, $b''(\theta_n)=e^{\theta_n}/(1+e^{\theta_n})^2$. $F_n(\beta)=x_nx_n^Te^{x_n^T\beta}/(1+e^{x_n^T\beta})^2$. Note that $|b''(\theta_n)|<1\ \forall\theta_n\in\Theta=\mathcal{R}^1$. Hence, for Assumption (M_s) to hold, it suffices to have $E\|x_1\|^{4+\delta}<\infty$, for some $\delta\geq 0$. This is consistent with results obtained by Chang and Martinsek (1992).

5. Proofs

Proofs of the asymptotic consistency for both natural- and nonnatural-link function cases are similar. Basically, both of them will follow from the strong consistency of the MLE, $\hat{\beta}_n$, of β_0 and by applying results of Gleser (1965), Chow and Robbins (1965) and similar arguments which were used in Chang and Martinsek (1992). In this paper, we concentrate only on the asymptotic efficiency and the asymptotic normality of the proposed sequential procedures. From theorems about asymptotic consistency of both cases, we have $\lim_{d\to 0}T_{d,i}/n_{opt}=1$, almost surely, for $i=1,2$. Therefore, to prove the asymptotic efficiency of our stopping rule $T_{d,i}$, $i=1$ or 2, it is sufficient to show that $\{d^2T_{d,i}:d\in(0,1)\}$, $i=1$ or 2, is uniformly integrable.

Our stopping rules depend on the estimate of the smallest eigenvalue of the unknown covariance matrix, which is also a function of the unknown regression coefficients of interest. Because of this, the sufficient conditions for the nonlinear renewal theorem (see Woodroofe (1982), Theorem 4.4) will be very difficult to check. Thus, to find an appropriate integrable last time for each case becomes an important part of this kind of method. Note that the integrability of such a last time depends on the supremum of uncountable last times and it cannot be shown to be integrable by applying Chow and Lai's (1975) results directly.

5.1. Known scale parameter

As mentioned above, we only prove Theorem 2.1 (iii) and Theorem 2.2 here.

Proof of Theorem 2.1 (iii)

For natural-link function case with known scale parameter ϕ, and for some fixed $\rho > 0$, define a last time random variable

$$L_\rho = \sup\{n \geq 1 : l_n(\beta) - l_n(\beta_0) \geq 0,\ \exists \beta \in \partial\mathcal{B}_\rho\},$$

where $\mathcal{B}_\rho = \{\beta : \|\beta - \beta_0\| \leq \rho\}$ and $\partial\mathcal{B}_\rho$ denote the boundary of $\mathcal{B}_\rho$. Then, under Assumption (R_s) and the concavity of the log-likelihood function, we have the following relationship between the MLE, $\hat{\beta}_n$, and the last time, L_ρ; i.e.

$$\{n > L_\rho\} \subset \{\hat{\beta}_n \in \mathcal{B}_\rho\},\ \forall n \geq 1. \tag{5.1}$$

Write

$$\begin{aligned} T_{d,1} &= T_{d,1} I_{\{T_{d,1} > L_\rho\}} + T_{d,1} I_{\{T_{d,1} \leq L_\rho\}}, \\ &\leq T_{d,1} I_{\{T_{d,1} > L_\rho\}} + L_\rho, \end{aligned} \tag{5.2}$$

where I is the indicator function. It follows from (5.1) that $\hat{\beta}_{T_{d,1}} \in \mathcal{B}_\rho$, when $T_{d,1} > L_\rho$. Let $\underline{b}''(\theta)$ be a real-valued function that satisfies (M_s) (ii), then

$$\underline{b}''(\theta) \leq b''(\theta),\ \forall \theta \in \Theta. \tag{5.3}$$

For any $\rho > 0$, if $\beta \in \mathcal{B}_\rho$ then $\eta_i \equiv |x_i^T \beta_0| + \|x_i\| \cdot \rho \geq |x_i^T \beta|$, for each $i \geq 1$. Hence, by (5.3) and by applying an eigenvalue inequality (see Wilkinson (1988)), it is easy to see that

$$\lambda_n^* \equiv \Lambda_{\min}(\Sigma_n^*) \leq \hat{\lambda}_n, \tag{5.4}$$

where $\Sigma_n^* = n^{-1} \sum_{i=1}^n \underline{b}''(\eta_i) x_i x_i^T$ and λ_n^* denotes the smallest eigenvalue of Σ_n^*. Note that Σ_n^* is now a sample mean of i.i.d. random matrices. Let

$$\underline{F}_i = \underline{b}''(\eta_i) x_i x_i^T \tag{5.5}$$

and

$$L_F = \sup\Big\{n \geq 1 : Z^T \sum_{i=1}^n (\underline{F}_i - \underline{F}) Z \leq -\frac{n\lambda_F}{2},\ \exists Z \in \mathcal{R}^p,\ \|Z\| = 1\Big\}, \tag{5.6}$$

where $\underline{F} = E\{\underline{F}_1\}$ and $\lambda_F = \Lambda_{\min}(\underline{F})$. By Wilkinson (1988), for $n \geq 1$,

$$\lambda_n^* \geq \Lambda_{\min}\Big(n^{-1} \sum_{i=1}^n (\underline{F}_i - \underline{F})\Big) + \lambda_F.$$

Then, by the definition of L_F, it follows that

$$\{n > L_F\} \subset \Big\{\lambda_n^* > \frac{\lambda_F}{2}\Big\}.$$

Therefore, by definition of last times L_ρ and L_F, if $n > \max(L_\rho, L_F)$ then $\hat\beta_n \in \mathcal{B}$ and $\hat\lambda_n \geq \lambda_F/2$. This implies that for any $d \in (0,1)$,

$$d^2 T_{d,1} \leq \frac{2c^2 a}{\lambda_F} + 1 + \max(L_\rho, L_F) \tag{5.7}$$

(See also Chang and Martinsek (1992)). Now, by Lemmas 5.1 and 5.2 below, it can be shown that both L_ρ and L_F are integrable, as well as $\{d^2 T_{d,1} : d \in (0,1)\}$. This implies that $\lim_{d\to 0} E[T_{d,1}/n_{opt}] = 1$.

Lemma 5.1. *Assume that for some $\rho \geq 0$, both (R_s) and (M_s) are satisfied. Then $EL_F^{1+\delta/2} < \infty$, for some $\delta \geq 0$.*

Lemma 5.2. *Under the same assumptions of Lemma* 5.1*, $EL_F^{1+\delta/2} < \infty$, for some $\delta \geq 0$.*

(Proofs of Lemma 5.1 and 5.2 will be given at the end of Section 5.)

Proof of Theorem 2.2.

By Waternaux (1976) and Davis (1977), we know that $\sqrt{n}(\hat\lambda_n - \lambda_p) \xrightarrow{d} N(0, \kappa + 2\lambda_p^2)$, as $n \to \infty$, where κ is defined as in Section 2. Moreover, by equation (3) of Lawley (1956), example 1.8 of Woodroofe (1982) and the strong consistency of $\hat\beta_n$, it can be shown that $\{\sqrt{n}(\hat\lambda_n - \lambda_p) : n \in N\}$ is uniformly continuous in probability. Hence, it follows from Anscombe's Theorem (see Woodroofe (1982)) that

$$\sqrt{T_{d,1}}(\hat\lambda_{T_{d,1}} - \lambda_p) \xrightarrow{d} N(0, \kappa + 2\lambda_p^2), \tag{5.8}$$

as $d \to 0$. Then, the asymptotic normality of $T_{d,1}$ follows from the inequalities below.

By definition of $T_{d,1}$, $T_{d,1}\hat\lambda_{T_{d,1}} \geq \frac{c^2}{d^2}$. This implies that

$$\sqrt{T_{d,1}}(\hat\lambda_{T_{d,1}} - \lambda_p) \geq T_{d,1}^{-\frac{1}{2}}\lambda_p(n_{opt} - T_{d,1}). \tag{5.9}$$

Note that $n_{opt} = c^2/(d^2\lambda_p)$. On the other hand,

$$(T_{d,1} - 1)\hat\lambda_{T_{d,1}-1} < \frac{c^2}{d^2}. \tag{5.10}$$

Therefore,

$$\sqrt{T_{d,1}}(\hat\lambda_{T_{d,1}-1} - \lambda_p) - \frac{\hat\lambda_{T_{d,1}-1}}{\sqrt{T_{d,1}}} \leq T_{d,1}^{-\frac{1}{2}}\lambda_p(n_{opt} - T_{d,1}). \tag{5.11}$$

Note that the second term of the LHS converges to 0 almost surely as $d \to 0$. Putting (5.9) and (5.11) together, and by (5.8), we have, as $d \to 0$,

$$T_{d,1}^{-\frac{1}{2}} \lambda_p (T_{d,1} - n_{opt}) \xrightarrow{d} N(0, \kappa + 2\lambda_p^2); \tag{5.12}$$

or equivalently, it can be rewritten as

$$n_{opt}^{-\frac{1}{2}} (T_{d,1} - n_{opt}) \xrightarrow{d} N\Big(0, \frac{\kappa}{\lambda_p^2} + 2\Big). \tag{5.13}$$

Moreover, if the $x_i's$ are normally distributed then $\kappa = 0$, and (5.13) becomes

$$n_{opt}^{-\frac{1}{2}} (T_{d,1} - n_{opt}) \xrightarrow{d} N(0, 2), \tag{5.14}$$

as $d \to 0$.

5.2. Unknown scale parameter

Now, let us turn to the proof of Theorem 2.3; i.e. for models with unknown scale parameter ϕ. By replacing the unknown ϕ by its estimator, we can have the stopping rule $T_{d,2}$, which depends on the estimate of the smallest eigenvalue of covariance matrix as well as $\hat{a}_n$, where $T_{d,2}$ and $\hat{a}_n$ are defined in Section 2. It follows from the SLLN, that $\hat{a}_n \to a$ almost surely as $n \to \infty$. Hence, (i) and (ii) of Theorem 2.3 can be proved easily by usual techniques. So, only the proof of Theorem 2.3 (iii) will be given below.

Proof of Theorem 2.3 (iii)

Both last times random variables L_ρ and L_F do not depend on the unknown scale parameter ϕ, so we still have following inequality:

$$\begin{aligned} T_{d,2} &= T_{d,2} I_{\{T_{d,2} > \max(L_\rho, L_F)\}} + T_{d,2} I_{\{T_{d,2} \leq \max(L_\rho, L_F)\}} \\ &\leq T_{d,2} I_{\{T_{d,2} > \max(L_\rho, L_F)\}} + \max(L_\rho, L_F). \end{aligned} \tag{5.15}$$

It is easy to see that if $n > \max(L_\rho, L_F)$, then $\hat{\beta}_n \in \mathcal{B}_\rho$ and $\hat{\lambda}_n \geq \frac{\lambda_F}{2}$. Define a new stopping time

$$T_{d,2}^* = \inf\Big\{n \geq n_d : n \frac{\lambda_F}{2} \hat{a}_n^{-1} \geq \frac{c^2}{d^2}\Big\} = \inf\Big\{n \geq n_d : n \hat{a}_n^{-1} \geq \frac{2c^2}{d^2 \cdot \lambda_F}\Big\}. \tag{5.16}$$

Hence

$$\{T_{d,2} > \max(L_\rho, L_F)\} \subset \{T_{d,2} \leq T_{d,2}^*\}.$$

For $a > 0$, $b > 0$ and $p > 0$, $(a+b)^p \leq (a^p + b^p) \max(1, 2^{p-1})$. Hence, under Assumption (P_s), and by the Taylor expansion Theorem, if $n > L_\rho$, then for $i = 1, \ldots, n$,

$$\begin{aligned} [y_i - b'(x_i^T \hat{\beta}_n)]^2 &\leq [y_i - b'(x_i^T \beta_0)]^2 + h^2(x_i^T \beta_n^*) \|x\|^2 \cdot \rho^2 \\ &\leq [y_i - b'(x_i^T \beta_0)]^2 + h^2(\eta_i) \|x_i\|^2 \cdot \rho, \end{aligned} \tag{5.17}$$

where η_i is the same as before. To simplify the notation, let $v_i = \underline{b}_i''(\eta_i)$ and w_i =RHS of (5.17), then

$$\hat{a}_n = a(\hat{\phi}_n) \leq n^{-1} \sum_{i=1}^{n} \frac{w_i}{v_i}, \tag{5.18}$$

for $n > L_\rho$. Note that $\{w_i/v_i : i \in N\}$ is a sequence of i.i.d. random variables with $E[w_i/v_i] < \infty$, under Assumption (P_s). Now, let

$$T_{d,2}^{**} = \inf\Big\{n \geq n_d : \Big[n^{-1} \sum_{i=1}^{n} \frac{w_i}{v_i}\Big]^{-1} \leq \frac{n\lambda_F \cdot d^2}{2c^2}\Big\}.$$

It follows from the definition that $\{n > \max(L_\rho, L_F)\} \subset \{T_d^* \leq T_d^{**}\}$. By applying Chow and Yu (1981) Lemma 2 (by setting parameters in their lemma as follows: $\alpha = 1,\ p = 1,\ \lambda = d^2,\ a_n = n$, and $b_n = 0$), it can be proved that $\{d^2 T_{d,2}^{**} : d \in (0,1)\}$ is uniformly integrable, provided that $n_d = O(d^{-1})$. By the definitions of $T_{d,2}^*$ and $T_{d,2}^{**}$, we have

$$T_{d,2} I_{\{T_{d,2} > \max(L_F, L_\lambda)\}} \leq T_{d,2}^{**} I_{\{T_{d,2} > \max(L_F, L_\lambda)\}}. \tag{5.19}$$

From (5.15) and (5.19) and for $d \in (0,1)$, $d^2 T_{d,2} \leq d^2 T_{d,2}^{**} + \max(L_F, L_\lambda)$. This implies the uniform integrability of $\{d^2 T_{d,2} : d \in (0,1)\}$ and completes the proof of the asymptotic efficiency part of Theorem 2.3.

5.3. Non-natural link function

We now prove Theorem 3.2. In Section 3, we have already mentioned that to show the integrability of the last time is a crucial step. In order to prove it for the non-natural link function case, we need the following lemma.

Lemma 5.3. *Suppose that (P_s), (R_s^*), and (M_s^*) are true for some $\delta \geq 0$; then $EL_\rho^{1+\delta/2} < \infty$.*

(Proof of Lemma 5.3 will be given at the end of Section 5.)

Proof of Theorem 3.2.

Assume that ϕ is known. Let $\mathcal{B}$ and $\partial\mathcal{B}$ be defined as before. By assumption $-\partial^2 \ell_n(\beta)/\partial\beta^2$ will converge to a positive definite matrix for all $\beta \in N \subset \mathcal{R}^p$, where N is some compact neighborhood of β_0. Hence, $\{n > L_\rho\} \subset \{\hat{\beta}_n \in \mathcal{B}_\rho\}$. Recall that the stopping rule

$$T_{d,1} = \inf\Big\{n \geq 1 : n\hat{\lambda}_n \geq \frac{c^2 a(\phi)}{d^2}\Big\}$$

and

$$\hat{\lambda}_n = \Lambda_{\min}\Big(n^{-1} \sum_{i=1}^{n} F_i(\hat{\beta}_n)\Big).$$

Then, by definition of $\underline{F}_i$ and the eigenvalue inequality,

$$\Lambda_{\min}(\hat{\Sigma}_n) \geq \Lambda_{\min}\Big(\frac{1}{n}\sum_{i=1}^{n}\underline{F}_i\Big), \tag{5.20}$$

provided that $\hat{\beta}_n \in \mathcal{B}_\rho$. Note that $\{\underline{F}_i : i \in N\}$ is again a sequence of i.i.d. random matrices. Then, the rest of the proof will follow by applying Lemma 5.3 and by similar arguments used in the proof of Theorem 2.1 (iii).

Remark. Since the last time is unchanged when ϕ is unknown, similar results for the non-natural link function case with unknown scale parameter can be easily obtained by slightly modifying the proof of Theorem 2.3.

The following are the proofs of Lemma 5.1 and 5.3. Proof of Lemma 5.2 is similar to that of Lemma 5.1, so it will omitted.

Proof of Lemma 5.1.

By definition, $a(\phi) > 0$. Hence, we can redefine L_ρ as below:

$$L_\rho = \sup\{n \geq 1 : e_n(\beta) - e_n(\beta_0) \geq 0,\ \exists\beta \in \partial\mathcal{B}_\rho\}, \tag{5.21}$$

where $e_n(\beta) = \sum_{i=1}^{n}[y_i u(x_i^T\beta) - b(u(x_i^T\beta))]$. That is L_ρ does not depend on scale parameter ϕ. By the Mean-value Theorem,

$$e_n(\beta) - e_n(\beta_0) = \sum_{i=1}^{n} S_i(\beta_0)^T(\beta-\beta_0) - \frac{1}{2}(\beta-\beta_0)^T\sum_{i=1}^{n}F_i(\beta_n^*)(\beta-\beta_0), \tag{5.22}$$

where $\beta_n^* \in \mathcal{B}_\rho$ is between β and β_0. By (5.5), Condition (R_s) and the concavity of the log-likelihood function, it follows that

$$F_i(\beta_n^*) \geq \underline{F}_i(\eta_i),\ \forall\, i = 1,\ldots,n, \tag{5.23}$$

where η_i is defined as before. Notation "$A \geq B$" for any two square matrices A, B means that $A - B$ is positive semi-definite. Thus,

$$\ell_n(\beta) - \ell_n(\beta_0) \leq \sum_{i=1}^{n} S_i(\beta_0)^T(\beta-\beta_0) - \frac{1}{2}(\beta-\beta_0)^T\sum_{i=1}^{n}\underline{F}_i(\eta_i)(\beta-\beta_0). \tag{5.24}$$

Note that for each fixed β, the RHS of (5.24) is a sum of i.i.d. random variables. Let

$$L_\rho^* = \sup\{n \geq 1 :\ \text{RHS of (5.24)} \geq 0,\ \exists\beta \in \partial\mathcal{B}_\rho\};$$

then $L_\rho \leq L_\rho^*$ almost surely (by definition of L_ρ and L_ρ^*). Under Assumption (M_s), it can be shown that $EL_\rho^{*\,1+\delta/2} < \infty$, for some $\delta > 0$, by similar arguments used in Chang and Martinsek (1992). This implies that $EL_\rho^{1+\delta/2} < \infty$, for some $\delta > 0$.

Proof of Lemma 5.3.

Let $e_n(\beta) = \sum_{i=1}^n [u(x_i^T\beta)y_i - b(u(x_i^T\beta))]$. By assumption, $a(\phi)$ is positive and

$$\ell_n(\beta) - \ell_n(\beta_0) = a(\phi)^{-1}[e_n(\beta) - e_n(\beta_0)];$$

therefore, the last time

$$\begin{aligned} L_\rho &= \sup\{n \geq 1 : \ell_n(\beta) - \ell_n(\beta_0) \geq 0\} \\ &= \sup\{n \geq 1 : e_n(\beta) - e_n(\beta_0) \geq 0\}. \end{aligned} \tag{5.25}$$

Recall that $u = (g \circ \mu)^{-1}$, where $g(\cdot)$ is the conventional link function which is defined in McCullagh and Nelder (1989). Usually, $g(\cdot)$ is assumed to be a monotone function. In addition, by the properties of exponential family, $\mu(\cdot)$ is also a monotone function, so is $u(\cdot)$. By Taylor's Expansion Theorem, we have

$$\begin{aligned} e_n(\beta) - e_n(\beta_0) = \sum_{i=1}^n S_i(\beta_0)^T(\beta - \beta_0) + \frac{1}{2}(\beta - \beta_0)^T \sum_{i=1}^n R_i(\beta_n^*)(\beta - \beta_0) \\ -\frac{1}{2}(\beta - \beta_0)^T \sum_{i=1}^n F_i(\beta_n^*)(\beta - \beta_0), \end{aligned}$$

where S_i, R_i and F_i are defined as in (1.9), and β_n^* is in the line segment of β and β_0. For any $t \in \mathcal{R}$, let

$$\underline{u}(t) = \inf\{u(s) : |s| \leq |t|, s \in \mathcal{R}\} \tag{5.26}$$

and

$$\underline{u}^{(1)}(t) = \inf\{u^{(1)}(s) : |s| \leq |t|, s \in \mathcal{R}\}. \tag{5.27}$$

Let $\eta_i = |x_i\beta_0| + \|x_i\| \cdot \rho$ and $\underline{F}_i = \underline{b}''(\underline{u}(\eta_i)) \left[\underline{u}^{(1)}(\eta_i)\right]^2 x_i x_i^T$, $i \geq 1$. Hence, for any $\beta \in \mathcal{B}_\rho$ and for all i, $|x_i\beta| \leq \eta_i$. By Assumption (M_s), for any $\beta \in B_\rho$, we have, for $i = 1, \ldots, n$,

$$F_i(\beta_n^*) \geq \underline{F}_i(\eta_i).$$

Then, it follows from (P_s), (M_s^*) and by the Taylor expansion theorem,

$$\begin{aligned} &\left[y_i - b'\left(u(x_i^T\beta_n^*)\right)\right]u^{(2)}(x_i^T\beta_n^*) \\ &= \left[y_i - b'\left(u(x_i^T\beta_0)\right)\right]u^{(2)}(x_i^T\beta_0) + b'\left(u(x_i^T\beta_0)\right)|u^{(2)}(x_i^T\beta_0) - u^{(2)}(x_i\beta_n^*)| \\ &\quad + \left[b'\left(u(x_i^T\beta_0)\right) - b'\left((x_i^T\beta_n^*)\right)\right]u^{(2)}(x_i^T\beta_n^*) \\ &\leq \left[y_i - b'\left(u(x_i^T\beta_0)\right)\right]u^{(2)}(x_i^T\beta_0) + |b'\left(u(x_i^T\beta_0)\right)\bar{u}^{(3)}(\eta_i)|\, \|x_i\| \cdot \rho \\ &\quad + h(\bar{u}(\eta_i))\bar{u}^{(1)}(\eta_i)|\bar{u}^{(2)}(\eta_i)|\, \|x_i\| \cdot \rho. \end{aligned}$$

Let

$$R_{i,1} = \left[y_i - b'\left(u(x_i^T\beta_0)\right)\right] u^{(2)}(x_i^T\beta_0),$$
$$R_{i,2} = |b'\left(u(x_i^T\beta_0)\right) \bar{u}^{(3)}(\eta_i)| \, \|x_i\| \cdot \rho$$

and

$$R_{i,3} = h\left(\bar{u}(\eta_i)\right) \bar{u}^{(1)}(\eta_i) |\bar{u}''(\eta_i)| \, \|x_i\| \cdot \rho. \tag{5.28}$$

Then, for all $i = 1, \ldots, n$,

$$R_i(\beta) \le R_{i,1} + R_{i,2} + R_{i,3}.$$

Hence,

$$e_n(\beta) - e_n(\beta_0) \le \sum_{i=1}^{n} S_i(\beta_0)(\beta - \beta_0) - \frac{1}{2}(\beta - \beta_0)^T \Big[\sum_{i=1}^{n} D_i\Big](\beta - \beta_0), \tag{5.29}$$

where $D_i = \underline{F}_i(\eta_i) - (R_{i,1} + R_{i,2} + R_{i,3})$ for all $i = 1, \ldots, n$. Note that $\{D_i : i = 1, \ldots, n\}$ is a sequence of i.i.d. random matrices. Moreover, for all $i \ge 1$, $ER_{i,1} = 0_{p\times p}$ and $\lim_{\rho\to 0} ER_{i,2} = \lim_{\rho\to 0} ER_{i,3} = 0_{p\times p}$. By assumption, we know that $E\underline{F}_1(\eta_1)$ is positive definite for any $\rho > 0$. Hence we can choose a small enough ρ such that $D = ED_1$ is positive definite. This implies that, for such a small $\rho > 0$, we have $\lambda_D = \Lambda_{\min}(D) > 0$. Let ρ be fixed throughout this section. Then, by an eigenvalue inequality, it is easy to see that

$$(\beta - \beta_0)^T D(\beta - \beta_0) \ge \rho^2 \cdot \lambda_D.$$

Define a last time

$$L_D = \sup\Big\{n \ge 1 : \sum_{i=1}^{n} S_i(\beta - \beta_0) - \frac{1}{2}(\beta - \beta_0)^T \sum_{i=1}^{n}[D_i - D](\beta - \beta_0) \ge \frac{n\rho^2\lambda_D}{2}, \; \exists\, \beta \in \partial\mathcal{B}_\rho\Big\}.$$

It can be shown that $EL_D^{1+\delta/2} < \infty$ for some $\delta \ge 0$, by similar arguments used in Chang and Martinsek (1992) and by applying a theorem in Chow and Lai (1975). By definition, $L_\rho \le L_D$ almost surely. This implies that there exists a $\rho > 0$ such that $EL_\rho^{1+\delta/2} < \infty$ for some $\delta \ge 0$.

Acknowledgement

The author would like to thank the referee for his efforts and suggestions which helped to improve the presentation of this paper.

References

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

Chang, Y. C. (1992). Fixed size confidence regions for the parameters in some generalized linear models. Technical report C-92-22, Institute of Statistical Science, Academia Sinica, Taipei, R.O.C.

Chang, Y. C. and Martinsek, A. T. (1992). Fixed size confidence regions for parameters of a logistic regression model. *Ann. Statist.* **20**, 1953-1969.

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

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

Chow, Y. S. and Teicher, H. (1978). *Probability Theory: Independence, Interchangeability, Martingales.* Springer-Verlag, New York.

Chow, Y. C. and Yu, K. F. (1981). The performance of a sequential procedure for the estimation of the mean. *Ann. Statist.* **9**, 184-189.

Davis, A. W. (1977). Asymptotic theory for principal component analysis: non-normal case. *Austral. J. Statist.* **19**, 206-212.

Fahrmeir, L. and Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Ann. Statist.* **13**, 342-368.

Fahrmeir, L. and Kaufmann, H. (1986). Correction. Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Ann. Statist.* **14**, 1643.

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 the fixed-size sequential confidence bounds for linear regression parameters. *Ann. Math. Statist.* **36**, 463-467.

Hinkley, D. V., Reid, N. and Snell, E. J. (1991). *Statistical theory and modeling : in honour of Sir David Cox, FRS.* Chapman and Hall, London, New York.

Lawley, D. N. (1956). Tests of significance for the latent roots of covariance and correlation matrices. *Biometrika* **43**, 128-136.

McCullagh, P. and Nelder, FRS J. A. (1989). *Generalized Linear Models*, 2nd Edition, Chapman and Hall, New York.

Mukhopadhyay, N. (1974). Sequential estimation of regression parameters in Gauss-Markoff setup. *J. Indian. Statist. Assoc.* **12**, 39-443.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *J. Roy. Statist. Soc. Ser.A* **135**, 370-384.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd Edition, John Wiley and Sons, Inc., New York.

Rockafellar, R. T. (1970). *Convex Analysis.* Princeton University Press, Princeton.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, John Wiley, New York.

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

Stein, C. (1945). A two-sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.* **16**, 243-258.

Waternaux, C. M. (1976). Asymptotic distribution of the sample roots for a nonnormal population. *Biometrika* **63**, 639-645.

Wilkinson, J. H. (1988). *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford.
Woodroofe, R. J. (1982). *Nonlinear Renewal Theory in Sequential Analysis*, Philadephia: Society for Industrial and Applied Mathematics.

Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan.

(Received September 1994; accepted October 1995)