Journal of Econometrics 17 (1981) 83–97. North-Holland Publishing Company

ASYMPTOTIC PROPERTIES OF THE MAXIMUM LIKELIHOOD ESTIMATOR IN DICHOTOMOUS LOGIT MODELS

Christian GOURIEROUX

University of Paris IX, 75116 Paris, France

Alain MONFORT

E.N.S.A.E., 92240 Malakoff, France

Received August 1980, final version received June 1981

The existence and strong consistency of the maximum likelihood estimator are analyzed in the context of dichotomous logit models. Sufficient conditions are given for the asymptotic normality of this estimator.

1. Introduction

The strong consistency of the maximum likelihood estimator in qualitative response models has been previously studied in the context of repeated samples [see for instance Amemiya (1976) and Morimune (1979)]; more precisely, it has been assumed that the exogenous variables can take a finite number of values and that the number of observations goes to infinity for each set of possible values for these variables. These assumptions which are adapted to the case of controlled experiments, are not satisfactory in other fields such as most of the econometric applications. In this paper we shall examine the asymptotic properties of the maximum likelihood estimator in dichotomous logit model under less restrictive assumptions. These assumptions are basically those used by Drygas (1976) and Anderson–Taylor (1979) to prove the consistency of least squares estimators. However, for the models we shall consider, a new problem occurs, namely the existence of the maximum likelihood estimator. In section 2 we give a necessary and sufficient condition for this existence, in the case of one exogenous variables. Section 3 is devoted to the problem of the existence and of the strong consistency of the maximum likelihood estimator in the general case. In section 4 the asymptotic normality is shown under assumption which are different from those used by McFadden (1975) and Nordberg (1980).

2. Existence of the maximum likelihood estimator in the case of one explanatory variable

Let us consider T independent variables y_t, $t=1,\dots,T$, such that

$$P(y_t=1)=\frac{1}{1+e^{-x_t b}}=F(x_t b),$$

$$P(y_t=0)=\frac{e^{-x_t b}}{1+e^{-x_t b}}=1-F(x_t b),$$

where b is an unknown real parameter and x_t, $t=1,\dots,T$, are the observations of an explanatory variable.

The likelihood function of this model is

$$L_T(b,y)=\prod_{t=1}^{T}[F(x_t\, b)]^{y_t}[1-F(x_t\, b)]^{1-y_t}.$$

Note that if $x_t=0$ for some t, the corresponding terms of the previous product do not contain b; therefore these terms have no influence on the maximization of L_T and we can assume, without loss of generality, that all the x_t's are different from zero.

The log-likelihood function is given by

$$\begin{aligned}\log L_T(b,y)&=\sum_{t=1}^{T}\{y_t\log F(x_t\, b)+(1-y_t)\log[1-F(x_t\, b)]\}\\&=\sum_{t=1}^{T}\left\{y_t\log\frac{F(x_t\, b)}{1-F(x_t\, b)}+\log[1-F(x_t\, b)]\right\}\\&=\sum_{t=1}^{T}y_t\, x_t\, b+\sum_{t=1}^{T}\log[1-F(x_t\, b)].\end{aligned}$$

The derivative f of function F [defined by $F(x)=1/(1+e^{-x})$] is equal to

$$f(x)=\frac{e^{-x}}{(1+e^{-x})^2}=F(x)[1-F(x)],$$

and therefore, the likelihood equation can be written as

$$\sum_{t=1}^{T}y_t\, x_t=\sum_{t=1}^{T}F(x_t\, b)\, x_t. \tag{1}$$

Let us denote by G_T the function

$$G_T(b)=\sum_{t=1}^{T} F(x_t\, b)\, x_t.$$

This function is continuous, strictly increasing since its derivative,

$$\frac{\mathrm{d}G_T}{\mathrm{d}b}(b)=\sum_{t=1}^{T} f(x_t\, b)\, x_t^2,$$

is strictly positive. Function G_T is also such that

$$\lim_{b\to\infty} G_T(b)=\sum_{t:x_t>0} x_t=\bar{G} \qquad \text{(say)},$$

$$\lim_{b\to-\infty} G_T(b)=\sum_{t:x_t<0} x_t=\underline{G}. \qquad \text{(say)}$$

The properties of G_T imply that any number in $]\underline{G},\bar{G}[$ is the image of a unique b.

Proposition 1. A solution of the likelihood equation exists almost surely as T goes to $+\infty$, if and only if

$$\prod_{t:\,x_t>0} F(x_t\, b_0) \prod_{t:x_t>0} [1-F(x_t\, b_0)]$$

$$+\prod_{t:x_t>0} F(x_t\, b_0) \prod_{t:x_t<0} [1-F(x_t\, b_0)] \xrightarrow[T\to\infty]{} 0,$$

where b_0 is the true value of b.

Proof. We are going to show that

$$\forall\varepsilon>0,\quad \exists\, T_0,\quad \exists\,\Omega_\varepsilon\subset\{0,1\}^N,$$

such that

$$P(\Omega_\varepsilon)>1-\varepsilon,$$

and

$$\forall T>T_0,\quad \forall y=(y_1,y_2,\ldots,y_T,\ldots)\in\Omega_\varepsilon,$$

the equation

$$\sum_{t=1}^{T} y_t x_t = \sum_{t=1}^{T} F(x_t b) x_t$$

has a solution $\hat{b}_T(y)$.

Let us first show the following inequalities:

$$\underline{G} \leqq \sum_{t=1}^{T} y_t x_t \leqq \bar{G}.$$

Since the y_t's are positive we have

$$\sum_{t=1}^{T} y_t x_t \leqq \sum_{t: x_t > 0} y_t x_t,$$

and, since $y_t \leqq 1$,

$$\sum_{t=1}^{T} y_t x_t \leqq \bar{G}.$$

The proof of the other inequality is similar.

Let us now denote by Ω_T the set

$$\left\{ y \mid \underline{G} < \sum_{t=1}^{T} y_t x_t < \bar{G} \right\}.$$

There exists a solution of the likelihood equation, if and only if y belongs to Ω_T. The set Ω_T can be written as

$$^{c}\Omega_T = \left\{ y \,\middle|\, \sum_{t=1}^{T} y_t x_t = \underline{G} \right\} \cup \left\{ y \,\middle|\, \sum_{t=1}^{T} y_t x_t = \bar{G} \right\},$$

or

$$^{c}\Omega_T = \{ y \mid y_t = 1 \text{ if } x_t < 0,\ y_t = 0 \text{ if } x_t > 0;\ t = 1, \dots, T \}$$
$$\cup \{ y \mid y_t = 1 \text{ if } x_t > 0,\ y_t = 0 \text{ if } x_t < 0;\ t = 1, \dots, T \}.$$

The sequence $^{c}\Omega_T$ is decreasing as T goes to infinity. Therefore if a solution exists for a given y and for $T = T_0$, there exists a solution for this y and for any $T \geqq T_0$. Moreover the probability of $P(^{c}\Omega_T)$ is equal to

$$P({}^c\Omega_T) = \prod_{t:x_t<0} F(x_t b_0) \prod_{t:x_t>0} [1-F(x_t b_0)]$$

$$+ \prod_{t:x_t>0} F(x_t b_0) \prod_{t:x_t<0} [1-F(x_t b_0)],$$

and a solution exists almost surely as T goes to infinity, if and only if this probability converges to zero.

Proposition 2. If the sequence $|x_t|$, $t=1,2,\ldots$, does not converge to $+\infty$, a solution of the likelihood equation exists a.s. for any b_0 as T goes to $+\infty$. If the sequence $|x_t|$, $t=1,2,\ldots$, converges to $+\infty$, a solution of the likelihood equation exists a.s. as T goes to $+\infty$, if and only if $\sum_{t=1}^{\infty} \mathrm{e}^{-|x_t b_0|} = +\infty$.

Proof. Note that if $b_0=0$, $P({}^c\Omega_T)=1/2^{T-1}$ converges to zero as T goes to $+\infty$. Therefore we can only consider the case $b_0 \neq 0$ and, without loss of generality, we can assume $b_0>0$.

Let us first consider the case where the sequence $|x_t|$, $t=1,2,\ldots$, does not converge to $+\infty$. There exists $M>0$ such that an infinite number of x_t's satisfy $|x_t|<M$. Therefore we have

$$\lim_{T\to\infty} P({}^c\Omega_T) \leqq \lim_{T\to\infty} \left\{ 1/2^T + \prod_{t:0<x_t<M} F(x_t\, b_0) \right.$$

$$\left. \times \prod_{t:-M<x_t<0} [1-F(x_t\, b_0)] \right\}$$

$$\leqq \lim_{T\to\infty} \left\{ \prod_{t:0<x_t<M} F(M\, b_0) \prod_{t:-M<x_t<0} [1-F(-M\, b_0)] \right\}$$

$$= \lim_{T\to\infty} \left\{ \prod_{t:-M<x_t<M} F(M\, b_0) \right\} = 0.$$

Let us now consider the case where the sequence $|x_t|$, $t=1,2,\ldots$, converges to $+\infty$. Since the first term of the sum appearing in $P({}^c\Omega_T)$ is still smaller than $1/2^T$, we only have to determine conditions under which the second term converges to zero. This term can be written as

$$\prod_{t:x_t>0} F(x_t b_0) \prod_{t:x_t<0} [1-F(x_t b_0)] = \prod_{t:x_t>0} F(x_t b_0) \prod_{t:x_t<0} F(-x_t b_0)$$

$$= \prod_{t=1}^{T} F(|x_t b_0|).$$

$P({}^c\Omega_T)$ converges to zero, if and only if the series whose general term is $\log F(|x_t b_0|) = -\log(1+\mathrm{e}^{-|x_t b_0|})$ tends to $-\infty$. Since $|x_t| \to \infty$, the latter condition is equivalent to

$$\sum_{t=1}^{\infty} \mathrm{e}^{-|x_t b_0|} = +\infty.$$

Remark 1. Using the previous proposition, we see that if $x_t = t^a$, with $a > 0$, the set $\lim_T \uparrow \Omega_T$ of 'asymptotic existence' of a solution has a probability strictly smaller than one. Similarly we can see that if $x_t = \log t$ the solution exists a.s. in the case $|b_0| \leqq 1$.

Remark 2. If a solution of the likelihood equation exists, it provides the unique maximum of the likelihood function, since

$$\frac{\mathrm{d}^2 \log L_T}{\mathrm{d}b^2} = -\sum_{t=1}^{T} x_t^2 f(x_t b)$$

is strictly negative and $\log L_T$ is strictly concave.

3. Strong consistency of the maximum likelihood estimator in the general case

3.1. The model

y_t, $t = 1, \dots, T$, are independent variables such that

$$P\{y_t = 1\} = F(x_t b) = F(x_{t1} b_1 + \dots + x_{tK} b_K),$$

$$P\{y_t = 0\} = 1 - F(x_t b),$$

where $F(x) = 1/(1 + \mathrm{e}^{-x})$ and x_{tk} is the tth observation of the kth exogenous variable.

The likelihood function of this model is

$$L_T(b, y) = \sum_{t=1}^{T} y_t x_t b + \sum_{t=1}^{T} \log[1 - F(x_t b)],$$

and we have

$$\frac{\mathrm{d} \log L_T(b, y)}{\mathrm{d}b} = \sum_{t=1}^{T} y_t x_t' - \sum_{t=1}^{T} F(x_t b) x_t',$$

$$\frac{\mathrm{d}^2 \log L_T(b, y)}{\mathrm{d}b \, \mathrm{d}b'} = -\sum_{t=1}^{T} x_t' x_t f(x_t b).$$

We shall assume that the Hessian matrix of $\log L_T$ is invertible.

3.2. Existence and strong consistency of the maximum likelihood estimator

We shall make the following assumptions:

(A_1) The exogenous variables are uniformly bounded, i.e., $\exists M_0 : |x_{tk}| \leqq M_0$, $\forall t$, $\forall k$.

(A_2) Let λ_{1T} and λ_{KT} be respectively the smallest and the largest eigenvalue of $\sum_{t=1}^{T} x_t' x_t\ f(x_t\, b_0)$, there exists M_1 such that $\lambda_{KT}/\lambda_{1T} < M_1$, $\forall T$.

It is worth noting that the latter assumption is implied by the usual condition: there exists a function $h(T)$ such that

$$\frac{1}{h(T)} \sum_{t=1}^{T} x_t' x_t\, f(x_t\, b_0)$$

converges to a positive definite matrix. Note that A_2 is also satisfied if there is only one exogenous variable since, in this case, $\lambda_{1T} = \lambda_{KT}$.

Proposition 3. If A_1 *and* A_2 *are satisfied, the maximum likelihood estimator* $\hat{b}_T$ *of* b *exists almost surely as* T *goes to* $+\infty$, *and* $\hat{b}_T$ *converges almost surely to the true value* b_0 *if and only if* $\lim_{T\to\infty} \lambda_{1T} = +\infty$.

Proof. The proof will be split up into four steps:

First step. Let ϕ_T be the function defined by

$$\phi_T(b, y) = b + \left[\sum_{t=1}^{T} x_t' x_t\, f(x_t\, b) \right]^{-1} \frac{\mathrm{d} \log L_T}{\mathrm{d} b} (b, y).$$

Lemma 1. There exists an open ball $B(b_0, r)$ *with center at* b_0 *and radius* r, *such that* ϕ_T *satisfies the Lipschitz condition* $\forall b$, $\tilde{b} \in B(b_0, r)$,

$$|\phi_T(b, y) - \phi_T(\tilde{b}, y)| \leqq c\,|b - \tilde{b}|,$$

where c *is a real number independent of* T *and* y *and such that* $0 < c < 1$.

Proof. The property of Lemma 1 is true if

$$\forall b \in B(b_0, r), \qquad \left\| \frac{\mathrm{d}\phi_T}{\mathrm{d}b} (b, y) \right\| \leqq c, \quad 0 < c < 1,$$

and we are going to prove this condition.

We have

$$\left\| \frac{\mathrm{d}\phi_T}{\mathrm{d}b}(b, y) \right\|$$

$$= \left\| \left[\sum_{t=1}^{T} x_t' x_t f(x_t b_0) \right]^{-1} \sum_{t=1}^{T} x_t' x_t [f(x_t b) - f(x_t b_0)] \right\|$$

$$\leqq \left\| \left[\sum_{t=1}^{T} x_t' x_t f(x_t b_0) \right]^{-1} \right\|$$

$$\times \left\| \sum_{t=1}^{T} x_t' x_t f(x_t b_0) \frac{f(x_t b) - f(x_t b_0)}{f(x_t b_0)} \right\|.$$

The first term of the product is equal to $1/\lambda_{1T}$ and, since the x_{kt} are uniformly bounded, there exists $m > 0$ such that $f(x_t b_0) > m$, $\forall t$. Moreover, f is uniformly continuous on any compact set; therefore

$$\forall \varepsilon > 0, \quad \exists r : \forall b \in B(b_0, r) \Rightarrow |x_t| \, |b - b_0| < M_0 r, \quad \forall t,$$

$$\Rightarrow |f(x_t b) - f(x_t b_0)| < \varepsilon,$$

and

$$\left\| \sum_{t=1}^{T} x_t' x_t f(x_t b_0) \frac{f(x_t b) - f(x_t b_0)}{f(x_t b_0)} \right\|$$

$$\leqq \sum_{t=1}^{T} \| x_t' x_t f(x_t b_0) \| \left| \frac{f(x_t b) - f(x_t b_0)}{f(x_t b_0)} \right|$$

$$\leqq \frac{\varepsilon}{m} \sum_{t=1}^{T} \| x_t' x_t f(x_t b_0) \|$$

$$\leqq \frac{\varepsilon}{m} \sum_{t=1}^{T} \operatorname{tr} [x_t' x_t f(x_t b_0)] \quad \text{since } x_t' x_t f(x_t b_0) \text{ is a positive matrix}$$

$$= \frac{\varepsilon}{m} \operatorname{tr} \sum_{t=1}^{T} [x_t' x_t f(x_t b_0)]$$

$$\leqq \frac{\varepsilon}{m} K \lambda_{KT}.$$

Finally, for any b in $B(b_0, r)$ we have

$$\left\|\frac{\mathrm{d}\phi_t}{\mathrm{d}b}(b, y)\right\| \leqq \frac{\varepsilon}{m}\frac{\lambda_{KT}}{\lambda_{1T}} \leqq K\, M_1 \frac{\varepsilon}{m}.$$

Choosing $\varepsilon = cm/K\, M_1$, with $0<c<1$, we obtain the required condition.

Second step

Lemma 2. If $\tilde{\psi}_T$ is a function from $B(b_0, r)$ into R^K, such that $\tilde{\phi}_T(b) = b - \tilde{\psi}_T(b)$ satisfies the Lipschitz condition with $0<c<1$, then any element of $B[\tilde{\psi}_T(b_0), (1-c)r]$ is the image by $\tilde{\psi}_T$ of an element of $B(b_0, r)$.

Proof. See Cartan (1967, p. 58).

Third step

Lemma 3. The estimator $\hat{b}_T$ exists almost surely as T goes to $+\infty$, and converges almost surely to the true value b_0, if and only if

$$\left[\sum_{t=1}^{T} x_t' x_t f(x_t b_0)\right]^{-1} \sum_{t=1}^{T} [y_t - F(x_t b_0)]\, x_t' \xrightarrow[T\to\infty]{\text{a.s.}} 0.$$

Proof. From Lemmas 1 and 2,

$$\psi_T(b, y) = -\left[\sum_{t=1}^{T} x_t' x_t f(x_t b_0)\right]^{-1} \frac{\mathrm{d}\log L_T(b, y)}{\mathrm{d}b}$$

is such that any element of $B[\psi_T(b_0, y), (1-c)r]$ is the image by $\psi_T(., y)$ of an element of $B(b_0, r)$. Also note that

$$\psi_T(b_0, y) = -\left[\sum_{t=1}^{T} x_t' x_t f(x_t b_0)\right]^{-1} \sum_{t=1}^{T} [y_t - F(x_t b_0)]\, x_t'.$$

Let us first prove the sufficient part of Lemma 3. If $\psi_T(b_0, y)$ converges a.s. to zero, for almost any y there exists $T_0(y, r)$ such that for any $T \geqq T_0(y, r)$, zero belongs to $B[\psi_T(b_0, y), (1-c)r]$. Therefore, zero is the image by $\psi_T(., y)$ of an element $\hat{b}_T(y)$ belonging to $B(b_0, r)$. Since $\psi_T[\hat{b}_T(y), y] = 0$, we also have

$$\frac{\mathrm{d}\log L_T}{\mathrm{d}b}[\hat{b}_T(y), y] = 0,$$

$\hat{b}_T(y)$ is a solution of the likelihood equation and, since $(\mathrm{d}^2 \log L_T)/(\mathrm{d}b\,\mathrm{d}b')$ is negative definite, $\hat{b}_T(y)$ is the unique maximum of the likelihood function. In summary, we have shown that, for almost any y,

$$\forall r>0, \quad \exists T_0(y,r), \quad \forall T \geqq T_0(y,r), \quad \hat{b}_T(y) \in B(b_0, r).$$

This means that $\hat{b}_T(y)$ exists almost surely as Y goes to $+\infty$, and that $\hat{b}_T(y)$ converges a.s. to b_0.

The necessary part of Lemma 3 will be shown by supposing that $\hat{b}_T$ exists a.s. as $T \to \infty$, and that $\psi_T(b_0, y)$ does not converge a.s. to zero, and then by exhibiting a contradiction. In the previous context, there exists Ω_0, with $P(\Omega_0)>0$, such that

$$\forall y \in \Omega_0, \quad \exists \varepsilon>0, \quad \exists T_1, \ldots, T_n, \ldots : \|\psi_{T_n}(b_0, y)\| > \varepsilon, \quad \forall n.$$

Choosing $r=\varepsilon/2(1+c)$, we have

$$\begin{aligned} \forall b \in B(b_0, r), \quad & \|\psi_{T_n}(b,y) - \psi_{T_n}(b_0,y)\| \\ & \leqq \|b-b_0\| + \|\phi_{T_n}(b,y) - \phi_{T_n}(b_0,y)\| \\ & \leqq r + c\,r = \varepsilon/2. \end{aligned}$$

This implies that

$$\forall b \in B(b_0, r), \quad \|\psi_{T_n}(b,y)\| > \varepsilon/2,$$

and, since the maximum likelihood estimator $\hat{b}_{T_n}(y)$ satisfies $\psi_{T_n}[\hat{b}_{T_n}(y), y]=0$ for n sufficiently large, we see that $\hat{b}_{T_n}(y)$ cannot belong to $B(b_0, y)$. Therefore the subsequence $\hat{b}_{T_n}(y)$ does not converge to b_0, which implies that the sequence $\hat{b}_T(y)$ does not converge to b_0 for any $y \in \Omega$.

Note that, from the results of section 2, $\hat{b}_T(y)$ exists a.s. under less restrictive assumption if there is only one exogenous variable.

Fourth step

Lemma 4. The maximum likelihood estimator $\hat{b}_T(y)$ exists a.s. as $T \to \infty$ and converges a.s. to b_0, if and only if $\lambda_{1T} \xrightarrow[T\to\infty]{} \infty$.

Proof. Let us introduce the following notations:

$$z_t = x_t \sqrt{f(x_t\, b_0)} \quad \text{and} \quad u_t = \frac{y_t - F(x_t\, b_0)}{\sqrt{f(x_t\, b_0)}}.$$

The u_t variables are independent, and

$$Eu_t = 0, \quad V(u_t) = 1, \quad \forall t.$$

From a result on the strong consistency of the ordinary least squares estimator [see Anderson–Taylor (1979, theorem 1)], we deduce that if $\lambda_{KT}/\lambda_{1T} < M_1$, $\forall t$, and if $\lambda_{1t} \xrightarrow[T\to\infty]{} +\infty$, then

$$\left[\sum_{t=1}^{T} z_t' z_t\right]^{-1} \sum_{t=1}^{T} z_t' u_t \xrightarrow[T\to\infty]{\text{a.s.}} 0,$$

$$\left[\sum_{t=1}^{T} x_t' x_t f(x_t b_0)\right]^{-1} \sum_{t=1}^{T} [y_t - F(x_t b_0)] x_t' \xrightarrow[T\to\infty]{\text{a.s.}} 0,$$

and Lemma 3 implies the sufficient part of Lemma 4.

Conversely, if λ_{1T} does not tend to $+\infty$, we conclude, using a result by Drygas (1976), that the ordinary least squares estimator,

$$\left[\sum_{t=1}^{T} z_t' z_t\right]^{-1} \sum_{t=1}^{T} z_t' u_t,$$

does not converge weakly to zero and, from Lemma 3, $\hat{b}_T$ does not converge strongly to b_0.

This completes the proof of Proposition 3.

Note that an advantage of the logistic model is the possibility of using the existing convergence results concerning the ordinary least squares estimator. In the case of a non-logistic distribution function F, the approach should be modified.

Corollary 1. If there exists a function $h(T)$ such that

$$\frac{1}{h(T)} \sum_{t=1}^{T} x_t' x_t f(x_t b_0)$$

converges to a positive definite matrix, and

$$h(T) \xrightarrow[T\to\infty]{} \infty,$$

then the maximum likelihood estimator $\hat{b}_T$ exists a.s. as T goes to ∞, and $\hat{b}_T$ converges a.s. to b_0.

Corollary 2. If there is only one exogenous variable x_t, and if the x_t''s are bounded, $\hat{b}_T$ converges a.s. to b_0, if and only if

$$\sum_{t=1}^{\infty} x_t^2 f(x_t b_0) = +\infty,$$

or, equivalently, if and only if

$$\sum_{t=1}^{\infty} x_t^2 = +\infty.$$

4. Asymptotic normality of the maximum likelihood estimator in the general case

Proposition 4. Under assumption A_1 and A_2 and if $\hat{b}_T$ converges a.s. to b_0, then

$$\left[\sum_{t=1}^{T} x_t' x_t f(x_t \hat{b}_T)\right]^{\frac{1}{2}} (\hat{b}_T - b_0) \xrightarrow[T\to\infty]{D} N(0, I),$$

(where $\xrightarrow{D}$ means convergence in distribution, and $Q^{\frac{1}{2}}$ is a notation for the unique symmetric positive matrix associated with a symmetric positive matrix Q and such that $[Q^{\frac{1}{2}}]^2 = Q$).

Proof. The likelihood equation is

$$\sum_{t=1}^{T} y_t x_t' = \sum_{t=1}^{T} F(x_t \hat{b}_T) x_t',$$

or

$$\sum_{t=1}^{T} [y_t - F(x_t b_0)] x_t' = \sum_{t=1}^{T} [F(x_t \hat{b}_T) - F(x_t b_0)] x_t'$$

$$= \left[\sum_{t=1}^{T} x_t' x_t f(x_t b_{tT}^*)\right] (\hat{b}_T - b_0),$$

where b_{tT}^* belongs to the segment whose extremities are $\hat{b}_T$ and b_0.

Denoting, respectively, $\sum_{t=1}^{T} x_t' x_t f(x_t b_0)$, $\sum_{t=1}^{T} x_t' x_t f(x_t \hat{b}_T)$, and $\sum_{t=1}^{T} x_t' x_t f(x_t b_{tT}^*)$ by A_T°, $\hat{A}_T$, and A_T^*, we have

$$[\hat{A}_T]^{\frac{1}{2}}(\hat{b}_T - b_0) = \{[\hat{A}_T]^{\frac{1}{2}}[A_T^*]^{-\frac{1}{2}}\}\{[A_T^*]^{-\frac{1}{2}}[A_T^\circ]^{\frac{1}{2}}\}$$

$$\times [A_T^\circ]^{-\frac{1}{2}} \sum_{t=1}^{T} [y_t - F(x_t\, b_0)]x_t',$$

It can be seen that the first two terms between brackets converge a.s. to the unit matrix. Let us consider for instance the second term $[A_T^*]^{-\frac{1}{2}}[A_T^\circ]^{\frac{1}{2}}$, we have

$$\|[A_T^*]^{-\frac{1}{2}}[A_T^\circ]^{\frac{1}{2}} - I\| \leqq \|[A_T^*]^{-\frac{1}{2}}\| \, \|[A_T^\circ]^{\frac{1}{2}} - [A_T^*]^{\frac{1}{2}}\|$$

$$= \lambda_{KT}^{\frac{1}{2}} \|[A_T^*]^{-\frac{1}{2}}\| \, \|[A_T^\circ/\lambda_{KT}]^{\frac{1}{2}} - [A_T^*/\lambda_{KT}]^{\frac{1}{2}}\|.$$

The term $\lambda_{KT}^{\frac{1}{2}}\|[A_T^*]^{-\frac{1}{2}}\|$ is a.s. smaller than $C_1(\lambda_{KT}/\lambda_{1T})^{\frac{1}{2}}$, where C_1 does not depend on T; therefore from A_2 this term is bounded. Using the same kind of arguments as in the proof of Lemma 1, it can be shown that $\|(A_T^\circ - A_T^*)/\lambda_{KT}\|$ converges a.s. to zero as T_1 goes to ∞. Then, from the continuity of the mapping $Q \to Q^{\frac{1}{2}}$, we deduce that

$\|[A_T^\circ/\lambda_{KT}]^{\frac{1}{2}} - [A_T^*/\lambda_{KT}]^{\frac{1}{2}}\|$ converges a.s. to zero, and the result follows.

What remains to be shown is

$$[A_T^\circ]^{-\frac{1}{2}} \sum_{t=1}^{T} [y_t - F(x_t\, b_0)]\, x_t' \xrightarrow[T\to\infty]{D} \mathrm{N}(0, I),$$

or, using the same notations as in Lemma 4,

$$\left[\sum_{t=1}^{T} z_t' z_t\right]^{-\frac{1}{2}} \sum_{t=1}^{T} z_t' u_t \xrightarrow[T\to\infty]{D} \mathrm{N}(0, I).$$

This is simply the condition for the asymptotic normality of the ordinary least squares estimator. To show that this property is true in our context we can check, for instance, that the assumptions used by Eicker (1966) are fulfilled. These assumptions are

(I) $$\max_{t=1,\dots,T} z_t \left[\sum_{t=1}^{T} z_t' z_t\right]^{-1} z_t' \xrightarrow[T\to\infty]{} 0,$$

(II) $$\sup_t E[u_t^2 \, \mathbf{1}_{\{|u_t| > c\}}] \xrightarrow[c\to\infty]{} 0,$$

(III) $\inf_t Eu_t^2 > 0.$

where, in assumption II, $\mathbf{1}_{\{|u_t|>c\}} = 1$ if $|u_t| > c$ and 0 otherwise.

Assumption I is satisfied since

$$\max_{t=1,\dots,T} z_t \left[\sum_{t=1}^{T} z_t' z_t\right]^{-1} z_t' \leqq \max_{t=1,\dots,T} \left\| \left[\sum_{t=1}^{T} z_t' z_t\right]^{-1} \right\| \| z_t \|^2$$

$$= (1/\lambda_{1T}) \max_{t=1,\dots,T} \|z_t\|^2$$

$$= (1/\lambda_{1T}) \max_{t=1,\dots,T} \|x_t\|^2 f(x_t b_0)$$

$$\leqq (KM^2/\lambda_{1T}).$$

Since $\hat{b}_T$ is strongly consistent, λ_{1T} tends to $+\infty$ and assumption I is verified.

Assumption II is satisfied since

$$|u_t| = \frac{|y_t - F(x_t b_0)|}{\sqrt{f(x_t b_0)}} \leqq \frac{1}{\sqrt{m}},$$

where $m > 0$ is a lower bound for the $f(x_t b_0)$, $t = 1, 2, \dots$, assumption III is true since $Eu_t^2 = 1$ for any t.

5. Conclusion

There are two kinds of existing results concerning the convergence of maximum likelihood estimator in logit models. On the one hand, McFadden (1975) established the weak consistency for polytomous logit models; on the other hand, Amemiya (1976) showed the strong consistency in the context of repeated samples. In this paper we considered the dichotomous logit model and we proved strong consistency and asymptotic normality under simple assumptions which are implied by McFadden's assumptions (Axiom 7). Finally it should be stressed that all these asymptotic results give little indication on the properties of the estimators in finite sample and it would be interesting to clarify this point by means of Monte Carlo studies.

References

Amemiya, T., 1976, The maximum likelihood, the minimum chi-square and the nonlinear weighted least squares estimator in the general qualitative response model, Journal of the American Statistical Association 71, 347–351.

Anderson, T.W. and J.B. Taylor, 1979, Strong consistency of least squares estimates in dynamic models, The Annals of Statistics 7, 484–489.

Cartan, H., 1967, Calcul differentiel (Hermann, Paris).

Drygas, H., 1976, Weak and strong consistency of the least squares estimators in regression models 2, Wahrscheinlichkeitstheorie und Verwandte Gebiete 34, 119–127.

Eicker, F., 1966, A multivariate central limit theorem for random linear vector forms, Annals of Mathematical Statistics, 1825–1828.

McFadden, D., 1975, Conditional logit analysis of qualitative choice behaviour, in: P. Zarembka, ed., Frontiers of econometrics, 105–142.

Morimune, K., 1959, Comparisons of normal and logistic models in the bivariate dichotomous analysis, Econometrica 47, 957–976.

Nordberg, L., 1980, Asymptotic normality of maximum likelihood estimators based on independent, unequally distributed observations in exponential family models, Scandinavian Journal of Statistics 7, 27–32.