ON THE MAXIMUM LIKELIHOOD
ESTIMATE FOR LOGISTIC ERRORS-IN-VARIABLES REGRESSION MODELS

by

R. J. Carroll*

University of North Carolina at Chapel Hill

*Research supported by the Air Force Office of Scientific Research
Grant AFOSR-F49620 82 C 0009.

ABSTRACT

Maximum likelihood estimates for errors-in-variables models are not always root-N consistent. We provide an example of this for logistic regression.

SOME KEY WORDS: Binary regression, Measurement error, Logistic regression, Maximum likelihood, Functional models.

I. INTRODUCTION

Logistic regression is a popular device for estimating the probability of an event such as the development of heart disease from a set of predictors, e.g., systolic blood pressure. The simplest form of this model is logistic regression through the origin with a single predictor:

$$\text{(1)} \qquad \Pr\{Y_i=1|c_i\} = G(\alpha_0 c_i) = \{1+\exp(-\alpha_0 c_i)\}^{-1} \quad ,$$

where α_0 and $\{c_i\}$ are scalars (i=1,...,N). In many applications, the predictors are measured with substantial error; a good example of this is systolic blood pressure, see Carroll, et al (1983). Thus, we observe

$$\text{(2)} \qquad C_i = c_i + v_i \; ,$$

where the errors $\{v_i\}$ are assumed here to be normally distributed with mean zero and variance σ^2.

The functional errors-in-variables logistic regression model is the case where (1) and (2) hold and the true values $\{c_i\}$ are unknown constants. The parameters are α_0 and $\{c_i\}$; there are (N+1) parameters with N observations, so classical maximum likelihood theory does not apply. Up to a constant, the log-likelihood is

$$-N \log_e \sigma - (2\sigma^2)^{-1} \sum_{i=1}^{N} (C_i - c_i)^2$$

$$+ \sum_{i=1}^{N} \{Y_i \log_e G(\alpha c_i) + (1-Y_i) \log_e (1-G(\alpha c_i))\} \; .$$

The linear functional errors in variables model (Kendall and Stuart (1979)) takes a form similar to (1) and (2), although of course (1) is replaced by the usual linear regression model with variance σ_ε^2. If σ^2, σ_ε^2 or $\sigma^2/\sigma_\varepsilon^2$ is known, then the linear functional maximum likelihood estimate exists and is both consistent and asymptotically normally distributed.

In this note, we show that for the functional logistic errors-in-variables model (1) and (2), even if σ^2 is known, the maximum likelihood estimate cannot be consistent and asymptotically normally distributed about α_0. The result can be extended to multiple logistic regression, and it is true even it we replicate (2) a finite number M times. If the number of replicates $M \to \infty$ as the sample size $N \to \infty$, then the functional maximum likelihood estimate can be shown to be consistent whether σ^2 is known or not.

II. THE THEOREM

In model (1) with the $\{c_i\}$ known, the ordinary maximum likelihood estimate for α_0 satisfies

$$0 = \sum_{i=1}^{N} c_i (Y_i - G(\hat{\alpha}_1 c_i)) \ .$$

In the presence of measurement error, the naive estimator would solve

(3) $$0 = \sum_{i=1}^{N} C_i (Y_i - G(\hat{\alpha}_2 C_i)) \ .$$

However, because of correlations, it turns out that

(4) $$\lim_{N\to\infty} E\, N^{-1} \sum_{i=1}^{N} C_i (Y_i - G(\alpha_0 C_i)) \neq 0 \ .$$

Condition (4) says that the defining equation (3) for $\hat{\alpha}_2$ is not even consistent at the true value α_0. Under these circumstances, it is well known from the theory of M-estimators that the usual naive estimator $\hat{\alpha}_2$ converges not to α_0 but to the value α_* satisfying

$$\lim_{N\to\infty} E\, N^{-1} \sum_{i=1}^{N} C_i (Y_i - G(\alpha_* C_i)) = 0 \ ,$$

assuming such a value α_* exists and is unique.

Assuming it exists and is unique, the functional MLE $\hat{\alpha}_0$ satisfies an equation analogous to (3):

$$(5)\qquad o = N^{-1}\sum_{i=1}^{N}\hat{c}_i(\hat{\alpha}_0)(Y_i - G(\hat{\alpha}_0\hat{c}_i(\hat{\alpha}_0))) \ ,$$

where

$$(6)\qquad \hat{c}_i(\alpha) = C_i + \alpha\sigma^2(Y_i - G(\alpha\hat{c}_i(\alpha))) \ .$$

It is easy to construct examples for which an analogue to (4) holds:

$$(7)\qquad \lim_{N\to\infty} E\, N^{-1}\sum_{i=1}^{N}\hat{c}_i(\alpha_0)(Y_i - G(\alpha_0\hat{c}_i(\alpha_0))) \neq 0.$$

One example of (7) is the extraordinarily easy problem $\sigma^2 = 1$ and $c_i = (-1)^i$. The only question is whether (7) is enough to guarantee that the functional MLE $\hat{\alpha}_0$ cannot be asymptotically normally distributed about the true value α_0. This is the case.

<u>Theorem</u> Suppose that σ^2 is known and that

(A.1) The maximum likelihood estimate $\hat{\alpha}_0$ exists;

$$(A.2)\qquad N^{-1}\sum_{i=1}^{N} c_i \to A \qquad (|A| < \infty) \ ;$$

$$(A.3)\qquad N^{-1}\sum_{i=1}^{N} c_i^2 \to B \qquad (0 < B < \infty) \ .$$

Then, if

$$(A.4)\qquad N^{\frac{1}{2}}(\hat{\alpha}_0 - \alpha_0) = O_p(1) \ ,$$

we must have that (7) fails, i.e.,

$$(8)\qquad \lim_{N\to\infty} E\, N^{-1}\sum_{i=1}^{N}\hat{c}_i(\alpha_0)(Y_i - G(\alpha_0\hat{c}_i(\alpha_0))) = 0.$$

The theorem as stated does not readily follow from the theory of M-estimators unless one assumes the existence of a unique α_* which satisfies (8), along with other regularity conditions. The proof we give avoids these complications because it exploits the form of the logistic function G.

III. PROOF OF THE THEOREM

It is most transparent to take $\sigma^2 = 1$. By formal differentiation, $\hat{\alpha}_0$ simultaneously satisfies (6) and

$$(9) \qquad N^{-1} \sum_{i=1}^{N} \hat{c}_i(\alpha)\{G(\alpha\hat{c}_i(\alpha)) - Y_i\} = 0.$$

Assumptions (A.2) and (A.3) imply that

$$(10) \qquad \max\{c_i^2/N : 1 \le i \le N\} \to 0 .$$

From (2) and (6), it follows that

$$(11) \qquad \lim_{\varepsilon\to 0} \max_{1\le i\le N} \sup_{|\alpha-\alpha_0|<\varepsilon} |\hat{c}_i(\alpha)-v_i|/(1+|\alpha_0|+|c_i|) = O_p(1) .$$

Further, since the $\{v_i\}$ are normally distributed,

$$(12) \qquad \max\{|v_i|N^{-\frac{1}{2}} : 1 \le i \le N\} \overset{p}{\to} 0 .$$

<u>Lemma</u> It follows that if (A.1)-(A.4) hold, then

$$(13) \qquad \max\{|\hat{c}_i(\alpha_0) - \hat{c}_i(\hat{\alpha}_0)|: 1 \le i \le N\} \overset{p}{\to} 0.$$

<u>Proof of the Lemma.</u> Define

$$H_i(u,\alpha) = u-c_i-v_i+\alpha\{G(\alpha u) = Y_i\} ,$$

$$H_i(c_i(\alpha), \alpha) = 0.$$

The partial derivatives of H_i are

$$D_1H_i(u,\alpha) = \frac{\partial}{\partial u} H_i(u,\alpha) = 1 = \alpha^2 G(\alpha u)\{1-G(\alpha u)\} ,$$

$$D_2H_i(u,\alpha) = \frac{\partial}{\partial \alpha} H_i(u,\alpha) = \{G(\alpha u) - Y_i\} + \alpha u G(\alpha u)\{1-G(\alpha u)\} .$$

By the chain rule,

(14) $$\frac{\partial}{\partial\alpha}\hat{c}_i(\alpha) = -[D_1H_i\{\hat{c}_i(\alpha),\alpha\}]^{-1}D_2H_i\{\hat{c}_i(\alpha),\alpha\} .$$

From (10)-(12) and (14) it follows that for every $M > 0$,

$$N^{-\frac{1}{2}} \max_{1\leq i\leq N} \sup_{|\alpha-\alpha_0|<M/N^{\frac{1}{2}}} \left|\frac{\partial}{\partial\alpha} \hat{c}_i(\alpha)\right|$$

$$= O_p\{ \max_{1\leq i\leq N} \sup_{|\alpha-\alpha_0|<M/N^{\frac{1}{2}}} |c_i(\alpha)|/N^{\frac{1}{2}}\} \xrightarrow{P} 0 .$$

This means that for every $M > 0$,

$$\max_{1<i<N} \sup_{|\alpha-\alpha_0|<M/N^{\frac{1}{2}}} |\hat{c}_i(\alpha) - \hat{c}_i(\alpha_0)| \xrightarrow{P} 0 ,$$

which by (A.4) completes the proof of the Lemma. □

We must prove that (A.1)-(A.4) imply (8). We are first going to show that

(15) $$N^{-1}\sum_{i=1}^{N} \hat{c}_i(\alpha_0)\{G(\alpha_0\hat{c}_i(\alpha_0))-Y_i\} \xrightarrow{p} 0.$$

The term in (15) can be written as $A_{1N} + A_{2N} + A_{3N}$, where

$$A_{1N} = N^{-1}\sum_{i=1}^{N} \{\hat{c}_i(\alpha_0) - \hat{c}_i(\hat{\alpha}_0)\}[G\{\alpha_0\hat{c}_i(\alpha_0)\} - Y_i] ,$$

$$A_{2N} = N^{-1}\sum_{i=1}^{N} \hat{c}_i(\hat{\alpha}_0)[G\{\alpha_0\hat{c}_i(\alpha_0)\}- G\{\hat{\alpha}_0\hat{c}_i(\hat{\alpha}_0)\}],$$

$$A_{3N} = N^{-1}\sum_{i=1}^{N} \hat{c}_i(\hat{\alpha}_0)[G\{\hat{\alpha}_0\hat{c}_i(\hat{\alpha}_0)\} - Y_i] .$$

By (9), $A_{3N} = 0$ and, since G is bounded, the Lemma and (A.4) gives $A_{1N} \xrightarrow{P} 0$ Because G and its derivative are bounded, the Lemma says that $A_{2N} \xrightarrow{P} 0$ as long as

$$N^{-1}\sum_{i=1}^{N} \{\hat{c}_i(\hat{\alpha}_0)\}^2 = O_p(1) \text{ and } N^{-1}\sum_{i=1}^{N} \{\hat{c}_i(\alpha_0)\}^2 = O_p(1),$$

which follow from (A.3), (10) and (11). Since (15) holds, to prove (8) we merely need to show that

$$N^{-1}\sum_{i=1}^{N}\left[\begin{array}{l}\hat{c}_i(\alpha_0)(Y_i-G(\alpha_0\hat{c}_i(\alpha_0)))\\ -E\{\hat{c}_i(\alpha_0)(Y_i-G(\alpha_0\hat{c}_i(\alpha_0)))\}\end{array}\right]\overset{p}{\to} 0$$

This follows from Chebychev's inequality and (A.3), completing the proof. □

The Theorem does not follow from ordinary likelihood calculations because the number of parameters increases with the sample size.

IV. A SIMULATION STUDY

To give some idea of the effect of measurement error, we conducted a small Monte-Carlo study of the logistic regression model

$$\Pr\{Y_i = 1\} = G(c_i/2 - 1), \quad i=1,\ldots,N$$

Here the values $\{c_i\}$ were randomly generated as normal random variables with mean zero and variance $3 = \sigma_c^2$, while the measurement errors were normally distributed with mean zero and variance $2 = \sigma_v^2$, with each $\{c_i\}$ being replicated twice. We chose the two sample sizes N = 200,400 and took 100 simulations for each sample size.

In Table 1 we report the Monte-Carlo efficiencies of the usual naive estimator and the functional MLE with respect to the logistic regression based on the correct values $\{c_i\}$. If the replicates of c_i are C_{i1}, C_{i2}, we used $C_i = (C_{i1}+C_{i2})/2$ and estimated the variance of $C_i - c_i$ by the sample variance of $(C_{i1}-C_{i2})/2$.

The results make it clear that neither the usual naive method nor the functional MLE are acceptable. Further work is clearly needed to identify good methods.

TABLE 1

Monte-Carlo Mean Squared Error Efficiencies Relative to Logistic Regression Based On The True Predictors

$\Pr\{Y_i=1|c_i\} = G(\alpha+\beta c_i)$, $\beta = \frac{1}{2}$, $\alpha = -1.0$

$i=1,\ldots,N$

		USUAL LOGISTIC	FUNCTIONAL MLE
α	N=200	0.74	0.25
	N=400	0.46	0.32
β	N=200	0.27	0.15
	N=400	0.09	0.24
$\alpha+\beta$	N=200	1.13	0.59
	N=400	0.60	0.53
$\alpha+2\beta$	N=200	0.38	0.28
	N=400	0.13	0.43

REFERENCES

Carroll, R.J., Spiegelman, C.H., Lan, K.K.G., Bailey, K.T. and Abbott, R.D. (1982). On errors-in-variables for binary regression models. Manuscript.

Kendall, M. and Stuart, A. The Advanced Theory of Statistics, Volume 2, pp. 399-443. Macmillan Publishing Co., New York.