THE CONVERGENCE RATE OF THE CONDITIONAL LOGIT ESTIMATOR

Koen Jochmans

May 11, 2016

We discuss the convergence rate of the conditional logit estimator and give some supporting calculations.

The Chamberlain (1992, 2010) two-period logit model has n outcome variables $y_1, \ldots, y_n$, and each outcome $y_i \equiv (y_{i1}, y_{i2})'$ is generated independently according to

$$\Pr(y_{i1} = 1) = \frac{1}{1+e^{-\alpha_i}}, \qquad \Pr(y_{i2} = 1) = \frac{1}{1+e^{-(\alpha_i+\beta)}},$$

for unknown parameters $\alpha_1, \ldots, \alpha_n$ and β.

Let $\Delta y_i \equiv y_{i2} - y_{i1}$. Because

$$\Pr(\Delta y_i = 1 | \Delta y_i \neq 0) = \frac{1}{1+e^{-\beta}}, \qquad \Pr(\Delta y_i = -1 | \Delta y_i \neq 0) = \frac{e^{-\beta}}{1+e^{-\beta}}$$

do not depend on α_i the conditional log-likelihood function

$$\ell(\beta) \equiv \sum_{i:\Delta y_i \neq 0} \left(\frac{1+\Delta y_i}{2}\right) \log\left(\frac{1}{1+e^{-\beta}}\right) + \left(\frac{1-\Delta y_i}{2}\right) \log\left(\frac{e^{-\beta}}{1+e^{-\beta}}\right)$$

separates estimation of the α_i from inference on β. The conditional-logit estimator of β (Rasch 1960) is

$$b_n = \arg\max_{\beta} \ell(\beta).$$

In practice, b_n is computed via a standard logit programme, applied to the subsample of informative units, the index set $\{i : \Delta y_i \neq 0\}$, commonly referred to as movers.

Let $n^* \equiv \sum_{i=1}^n 1\{\Delta y_i \neq 0\}$, the number of informative observations. Note that n^* is random. Furthermore, its distribution clearly depends on the α_i; we have

$$E(n^*) = \sum_{i=1}^n \Pr(\Delta y_i \neq 0) = \sum_{i=1}^n \frac{e^{-\alpha_i} + e^{-(\alpha_i+\beta)}}{(1+e^{-\alpha_i})(1+e^{-(\alpha_i+\beta)})}.$$

Consistency of b_n requires that $E(n^*) \to \infty$ as $n \to \infty$. This limits the speed at which $\Pr(\Delta y_i \neq 0)$ is allowed to shrink to zero as i grows. In particular, we require that

$$n\, p_n \to \infty$$

as $n \to \infty$, where we let $p_n = E(n^*/n)$. Thus, the expected fraction of movers is allowed to shrink with the sample size n, but at a rate no faster than n^{-1}.

The Fisher information on β is

$$I \equiv \sum_{i=1}^{n} \frac{e^{-\beta}}{(1+e^{-\beta})^2} \ \Pr(\Delta y_i \neq 0).$$

In large samples, $b_n \sim \mathcal{N}(\beta, I^{-1})$. The rate at which information accrues is $\sqrt{np_n} = \sqrt{E(n^*)}$ and may be very slow.

Note that the difference $n^* - np_n$ converges to zero in probability as $n \to \infty$. Furthermore, the inverse of the conditional information

$$I_* \equiv \sum_{i=1}^{n} \frac{e^{-\beta}}{(1+e^{-\beta})^2} \ 1\{\Delta y_i \neq 0\} = \sum_{i:\Delta y_i \neq 0} \frac{e^{-\beta}}{(1+e^{-\beta})^2}$$

is a valid large-sample variance for b_n. That is, $I_* - I$ converges to zero in probability as $n \to \infty$. The quantity I_* is delivered by any standard logit optimization programme when applied to the subsample of movers. Thus, in practice, we base inference on the approximation

$$b_n \overset{a}{\sim} \mathcal{N}(\beta, I_*^{-1}),$$

which becomes more precise at the rate $(np_n)^{-1/2}$.

The convergence rate can be interpreted as a function of the growth rate of the α_i. Moreover, because

$$\Pr(\Delta y_i \neq 0) \asymp e^{-|\alpha_i|} \text{ as } |\alpha_i| \to \infty,$$

we have

$$E(n^*) \asymp \sum_{i=1}^{n} e^{-|\alpha_i|}.$$

If α_i is finite for all i then $0 < \Pr(\Delta y_i \neq 0) < 1$. Consequently, $E(n^*)$ grows like n and the convergence rate of b_n is $n^{-1/2}$, the parametric rate. More generally, if the α_i are drawn from a distribution whose tails are sufficiently thin to ensure that p_n converges to a positive constant as $n \to \infty$ the parametric rate remains attainable. The normal distribution would be one example. On the other hand, if the α_i are allowed to become unbounded the convergence rate will decrease. For example, if $\alpha_i \asymp \log(i)$, then $E(n^*) \asymp \sum_{i=1}^{n} i^{-1}$, which is the nth harmonic number. For large n, the nth harmonic number behaves like $\log(n)$. Therefore, in this case, $\text{var}(b_n)$ shrinks like $(\log(n))^{-1}$, which is extremely slow. When $\alpha_i \asymp c \log(i)$ for a constant c, $E(n^*)$ converges to the Euler-Riemann zeta function at c, which is a finite constant for any $c > 1$. In such a case, $np_n \to c$ as $n \to \infty$ and b_n is not consistent.

REFERENCES

Chamberlain, G. (1992), Binary Response Models for Panel Data: Identification and Information,. Mimeo.

Chamberlain, G. (2010), "Binary Response Models for Panel Data: Identification and Information," *Econometrica*, 78, 159–168.

Rasch, G. (1960), "Probabilistic models for some intelligence and attainment tests,", Unpublished report, The Danish Institute of Educational Research, Copenhagen.