LTCC/Reid: Order in probability November 8, 2012

1. Recall that if we have two sequences of real numbers $\{a_n\}$ and $\{b_n\}$, that

$$a_n = o(b_n) \iff \lim_{n\to\infty} \frac{a_n}{b_n} = 0, \text{ and}$$
$$a_n = O(b_n) \iff \lim_{n\to\infty} \frac{a_n}{b_n} = A < \infty.$$

2. Similarly, if we have a sequence of random variables, $\{X_n\}$, we say

$$X_n = o_p(a_n) \iff \frac{X_n}{b_n} \stackrel{p}{\to} 0,$$

where, you recall, $X_n/b_n \stackrel{p}{\to} 0$ means

For any ϵ, there exists a δ and n_0 such that $Pr(|X_n/b_n| > \epsilon) < \delta$ for all $n > n_0$.

3. If we say $X_n = O_p(a_n)$, then X_n/a_n is bounded in probability; more precisely there is an $M < \infty$, and for any ϵ, an n_0, such that

$$Pr(\frac{X_n}{a_n} > M) < \epsilon, \text{ for all } n > n_0.$$

4. The $o_p(\cdot)$ and $O_p(\cdot)$ notation can be used for two sequences of random variables as well, e.g. $X_n = o_p(Y_n) \iff X_n/Y_n \stackrel{p}{\to} 0$, etc.

Most often, in asymptotic theory of statistics, we have $a_n = n^{j/2}$ for $j = 1, 2, 3$. So, for example, we might say $X_n = o_p(n)$ to mean $X_n/n \stackrel{p}{\to} 0$, and $X_n = O_p(n)$ to mean X_n/n is bounded in probability. If we have an i.i.d. sample $Y_1, \ldots, Y_n$ from a distribution with finite expected value μ and finite variance σ^2, then letting $X_n = \bar{Y} = n^{-1}\Sigma Y_i$, the sample mean, we have

$$X_n \stackrel{p}{\to} \mu; \text{ i.e. } \bar{Y} - \mu = o_p(1)$$

by the weak law of large numbers, and

$$\sqrt{n}(\bar{Y} - \mu) = O_p(1)$$

by the central limit theorem, $\sqrt{n}(\bar{Y} - \mu) \stackrel{d}{\to} N(0, \sigma^2)$, and $\sigma^2 < \infty$. A random variable that is $O_p(1)$ is bounded in the limit, i.e. it has a limiting distribution. A random variable that is $O_p(\sqrt{n})$ converges to a bounded random variable after you divide it by $\sqrt{n}$.

If $U(\theta) = \sum_{i=1}^{n} U_i(\theta)$, say, where $U_i(\theta) = \partial \log f(y_i;\theta)/\partial\theta$, and we assume the y_i are i.i.d., then

$$\frac{U(\theta)}{n} \xrightarrow{p} E\{U(\theta)\} = 0,$$

where we have hidden the dependence of $U(\theta)$ on $y = (y_1, \ldots, y_n)$. And

$$\frac{U(\theta)}{\sqrt{n}} \xrightarrow{d} N\{0, i_1(\theta)\},$$

again by the central limit theorem. If $U(\theta)$ did not have mean 0, but rather had mean $n\mu$, say, then we would probably have something like

$$\frac{U(\theta) - n\mu}{\sqrt{n}} \xrightarrow{d} N(0, \text{ something finite }),$$

which we could write as

$$U(\theta) - n\mu = O_p(\sqrt{n}),$$

but $U(\theta)$ would be $O(n)$.

There is a calculus of $O(\cdot), o(\cdot), O_p(\cdot), o_p(\cdot)$, that lets one derive things like

$$o(a_n + b_n) = o(a_n) + o(b_n), \quad O(a_n b_n) = O(a_n)O(b_n),$$

and so on; this is discussed in Chapter 1 of Barndorff-Nielsen & Cox (1989).

Reference

Barndorff-Nielsen, O.E. and Cox, D.R. (1989). *Asymptotic Techniques for Use in Statistics*. Chapman & Hall, London.