Appendix A
Asymptotic theory

Approximate or asymptotic results are an important foundation of statistical inference. Some of the main ideas are discussed below. The ideas center around the fundamental theorem of statistics, laws of large numbers (LLN), and central limit theorems (CLT). The discussion includes definitions of convergence in probability, almost sure convergence, convergence in distribution and rates of stochastic convergence.

The *fundamental theorem of statistics* states that if we sample randomly with replacement from a population, the empirical distribution function is consistent for the population distribution function (Davidson and MacKinnon [1993], p. 120-122). The fundamental theorem sets the stage for the remaining asymptotic theory.

A.1 Convergence in probability (laws of large numbers)

Definition A.1 *Convergence in probability.*
x_n *converges in probability to constant* c *if* $\lim_{n\to\infty} \Pr\left(|x_n - c| > \varepsilon\right) = 0$ *for all* $\varepsilon > 0$*. This is written* $p\lim\left(x_n\right) = c$*.*

A frequently employed special case is convergence in quadratic mean.

Theorem A.1 *Convergence in quadratic mean (or mean square).*
If x_n *has mean* μ_n *and variance* σ_n^2 *such that ordinary limits of* μ_n *and*

σ_n^2 are c and 0, respectively, then x_n converges in mean square to c and $p\lim(x_n) = c$.

A proof follows from Chebychev's Inequality.

Theorem A.2 *Chebychev's Inequality.*
If x_n is a random variable and c_n and ε are constants then

$$Pr\left(|x_n - c_n| > \varepsilon\right) \leq E\left[(x_n - c_n)^2\right] / \varepsilon^2$$

A proof follows from Markov's Inequality.

Theorem A.3 *Markov's Inequality.*
If y_n is a nonnegative random variable and δ is a positive constant then

$$Pr\left(y_n \geq \delta\right) \leq E\left[y_n\right] / \delta$$

Proof.

$$E\left[y_n\right] = Pr\left(y_n < \delta\right) E\left[y_n \mid y_n < \delta\right] + Pr\left(y_n \geq \delta\right) E\left[y_n \mid y_n \geq \delta\right]$$

Since $y_n \geq 0$ both terms are nonnegative.
Therefore, $E\left[y_n\right] \geq Pr\left(y_n \geq \delta\right) E\left[y_n \mid y_n \geq \delta\right]$.
Since $E\left[y_n \mid y_n \geq \delta\right]$ must be greater than δ, $E\left[y_n\right] \geq Pr\left(y_n \geq \delta\right) \delta$. ■

Proof. To prove Theorem A.2, let $y_n = (x_n - c)^2$ and $\delta = \varepsilon^2$ then

$$(x_n - c)^2 > \delta$$

implies $|x - c| > \varepsilon$. ■

Proof. Now consider a special case of Chebychev's Inequality. Let $c = \mu_n$, $Pr\left(|x_n - \mu_n| > \varepsilon\right) \leq \sigma^2 / \varepsilon^2$. Now, if $\lim_{n\to\infty} E\left[x_n\right] = c$ and $\lim_{n\to\infty} Var\left[x_n\right] = 0$, then $\lim_{n\to\infty} Pr\left(|x_n - \mu_n| > \varepsilon\right) \leq \lim_{n\to\infty} \sigma^2 / \varepsilon^2 = 0$. The proof of Theorem A.1 is completed by Definition A.1 $p\lim(x_n) = \mu_n$. ■

We have shown convergence in mean square implies convergence in probability.

A.1.1 Almost sure convergence

Definition A.2 *Almost sure convergence.*
$\overline{z}_n \xrightarrow{as} z$ *if* $\Pr\left(\lim_{n\to\infty} |\overline{z}_n - z| < \varepsilon\right) = 1$ *for all* $\varepsilon > 0$.
That is, there is large enough n such that the probability of the joint event $\Pr\left(|\overline{z}_{n+1} - z| > \varepsilon, |\overline{z}_{n+2} - z| > \varepsilon, \ldots\right)$ *diminishes to zero.*

Theorem A.4 *Markov's strong law of large numbers.*
If $\{zj\}$ is sequence of independent random variables with $E[z_j] = \mu_j < \infty$ and if for some $\delta > 0$, $\frac{E\left[\left|z_j - \mu_j\right|^{1+\delta}\right]}{j^{1+\delta}} < \infty$ then $\overline{z}_n - \overline{\mu}_n$ converges almost surely to 0, where $\overline{z}_n = n^{-1} \sum_{j=1}^{n} z_j$ and $\overline{\mu}_n = n^{-1} \sum_{j=1}^{n} \mu_j$.
This is denoted $\overline{z}_n - \overline{\mu}_n \xrightarrow{as} 0$.

Kolmogorov's law is somewhat weaker as it employs $\delta = 1$.

Theorem A.5 *Kolmogorov's strong law of large numbers.*
If $\{z\}$ is sequence of independent random variables with $E[z_j] = \mu_j < \infty$, $Var[z_j] = \sigma_j^2 < \infty$ and $\sum_{j=1}^{n} \frac{\sigma_j^2}{j^2} < \infty$ then $\overline{z}_n - \overline{\mu}_n \xrightarrow{as} 0$.

Both of the above theorems allow variances to increase but slowly enough that sums of variances converge. Almost sure convergence states that the behavior of the mean of sample observations is the same as the behavior of the average of the population means (not that the sample means converge to anything specific).

The following is a less general result but adequate for most econometric applications. Further, Chebychev's law of large numbers differs from Kinchine's in that Chebychev's does not assume *iid* (independent, identical distributions).

Theorem A.6 *Chebychev's weak law of large numbers.*
If $\{z\}$ is sequence of uncorrelated random variables with $E[z_j] = \mu_j < \infty$, $Var[z_j] = \sigma_j^2 < \infty$, and $\lim_{n \to \infty} n^{-2} \sum_{j=1}^{\infty} \sigma_j^2 < \infty$, then $\overline{z}_n - \overline{\mu}_n \xrightarrow{p} 0$.

Almost sure convergence implies convergence in probability (but not necessarily the converse).

A.1.2 Applications of convergence

Definition A.3 *Consistent estimator.*
An estimator $\hat{\theta}$ of parameter θ is a consistent estimator iff $p\lim\left(\hat{\theta}\right) = \theta$.

Theorem A.7 *Consistency of sample mean.*
The mean of a random sample from any population with finite mean μ and finite variance σ^2 is a consistent estimator of μ.

Proof. $E[\bar{x}] = \mu$ and $Var[\bar{x}] = \frac{\sigma^2}{n}$, therefore by Theorem A.1 (convergence in quadratic mean) $p\lim(\bar{x}) = \mu$. ■
An alternative theorem with weaker conditions is Kinchine's weak law of large numbers.

Theorem A.8 *Kinchine's theorem (weak law of large numbers).*
Let $\{x_j\}$*,* $j = 1, 2, ..., n$*, be a random sample (iid) and assume* $E[x_j] = \mu$ *(a finite constant) then* $\bar{x} \xrightarrow{p} \mu$*.*

The Slutsky Theorem is an extremely useful result.

Theorem A.9 *Slutsky Theorem.*
For continuous function $g(x)$ *that is not a function of* n*,* $p\lim(g(x_n)) = g(p\lim(x_n))$*.*

A proof follows from the implication rule.

Theorem A.10 *The implication rule.*
Consider events E *and* F_j*,* $j = 1, ..., k$*, such that* $E \supset \cap_{j=1,k} F_j$*.*
Then $\Pr(\bar{E}) \leq \sum_{j=1}^{k} \Pr(\bar{F}_j)$*.*

Notation: $\bar{E}$ is the complement to E, $A \supset B$ means event B implies event A (inclusion), and $A \cap B \equiv AB$ means the intersection of events A and B.

Proof. A proof of the implication rule is from Lukacs [1975].
1. $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(AB)$.
2. $\Pr(\bar{A}) = 1 - \Pr(A)$.

from 1

3. $\Pr(A \cup B) \leq \Pr(A) + \Pr(B)$
4. $\Pr(\cup_{j=1,\infty} A_j) \leq \sum_{j=1}^{k} \Pr(A_j)$

1 and 2 imply $\Pr(AB) = \Pr(A) - \Pr(B) + 1 - \Pr(A \cup B)$. Since $1 - \Pr(A \cup B) \geq 0$, we obtain

5. $\Pr(AB) \geq \Pr(A) - \Pr(\bar{B}) = 1 - \Pr(\bar{A}) - \Pr(\bar{B})$ (*Boole's Inequality).*
$\Pr(\cap_j A_j) \geq 1 - \Pr(\bar{A}_1) - \Pr\left(\overline{\cap_{j=2,\infty} A_j}\right) = 1 - \Pr(A_1) - \Pr(\cup_{j=2,\infty} \bar{A}_j)$.

This inequality and 4 imply

6. $\Pr(\cap_{j=1,k} A_j) \geq 1 - \sum_{j=1}^{k} \Pr(\bar{A}_j)$ (*Boole's Generalized Inequality*).

5 can be rewritten as

7. $\Pr(\bar{A}) + \Pr(\bar{B}) \geq 1 - \Pr(AB) = \Pr(\overline{AB}) = \Pr(\bar{A} \cup \bar{B})$.

Now let C be an event implied by AB, that is $C \supset AB$, then $\bar{C} \subset \bar{A} \cup \bar{B}$ and

8. $\Pr(\bar{C}) \leq \Pr(\bar{A} \cup \bar{B})$.

Combining 7 and 8 obtains

The Implication Rule.

Let A, B, and C be three events such that $C \supset AB$, then
$\Pr\left(\bar{C}\right) \leq \Pr\left(\bar{A}\right) + \Pr\left(\bar{B}\right)$. ■

Proof. Slutsky Theorem (White [1984])
Let $g_j \in g$. For every $\varepsilon > 0$, continuity of g implies there exists $\delta(\varepsilon) > 0$ such that if $|x_{nj}(w) - x_j| < \delta(\varepsilon)$, $j = 1, ..., k$, then $|g_j(x_n(w)) - g_j(x)| < \varepsilon$. Define events

$$Fj \equiv [w : |x_{nj}(w) - x_j| < \delta(\varepsilon)]$$

and

$$E \equiv [w : |g_j(x_{nj}(w)) - g_j(x)| < \varepsilon]$$

Then $E \supset \cap_{j=1,k} F_j$, by the implication rule, leads to $\Pr\left(\bar{E}\right) \leq \sum_{j=1}^{k} \Pr\left(\bar{F}_j\right)$. Since $x_n \xrightarrow{p} x$ for arbitrary $\eta > 0$ and all n sufficiently large, $\Pr(F_j) \leq \eta$. Thus, $\Pr\left(\bar{E}\right) \leq k\eta$ or $\Pr(E) \geq 1 - k\eta$. Since $\Pr[E] \leq 1$ and η is arbitrary, $\Pr(E) \longrightarrow 1$ as $n \longrightarrow \infty$. Hence, $g_j(x_n(w)) \xrightarrow{p} g_j(x)$. Since this holds for all $j = 1, ..., k$, $g(x_n(w)) \xrightarrow{p} g(x)$. ■

Comparison of Slutsky Theorem with Jensen's Inequality highlights the difference between the expectation of a random variable and probability limit.

Theorem A.11 *Jensen's Inequality.*
If $g(x_n)$ is a concave function of x_n then $g(E[x_n]) \geq E[g(x)]$.

The comparison between the Slutsky theorem and Jensen's inequality helps explain how an estimator may be consistent but not be unbiased.[1]

Theorem A.12 *Rules for probability limits.*
If x_n and y_n are random variables with $p\lim(x_n) = c$ and $p\lim(y_n) = d$ then
a. $p\lim(x_n + y_n) = c + d$ (sum rule)
b. $p\lim(x_n y_n) = cd$ (product rule)
c. $p\lim\left(\frac{x_n}{y_n}\right) = \frac{c}{d}$ if $d \neq 0$ (ratio rule)

If W_n is a matrix of random variables and if $p\lim(W_n) = \Omega$ then

d. $p\lim\left(W_n^{-1}\right) = \Omega^{-1}$ (matrix inverse rule)

If X_n and Y_n are random matrices with $p\lim(X_n) = A$ and $p\lim(Y_n) = B$ then

e. $p\lim(X_n Y_n) = AB$ (matrix product rule).

[1] Of course, Jensen's inequality is exploited in the construction of concave utility functions to represent risk aversion.

A.2 Convergence in distribution (central limit theorems)

Definition A.4 *Convergence in distribution.*
x_n converges in distribution to random variable x with CDF $F(x)$ if $\lim_{n\to\infty} |F(x_n) - F(x)| = 0$ at all continuity points of $F(x)$.

Definition A.5 *Limiting distribution.*
If x_n converges in distribution to random variable x with CDF $F(x)$ then $F(x)$ is the limiting distribution of x_n; this is written $x_n \xrightarrow{d} x$.

Example A.1 $t_{n-1} \xrightarrow{d} N(0,1)$.

Definition A.6 *Limiting mean and variance.*
The limiting mean and variance of a random variable are the mean and variance of the limiting distribution assuming the limiting distribution and its moments exist.

Theorem A.13 *Rules for limiting distributions.*
(a) If $x_n \xrightarrow{d} x$ and $p\lim(y_n) = c$, then $x_n y_n \xrightarrow{d} xc$.
Also, $x_n + y_n \xrightarrow{d} x + c$, and
$\frac{x_n}{y_n} \xrightarrow{d} \frac{x}{c}$, $c \neq 0$.
(b) If $x_n \xrightarrow{d} x$ and $g(x)$ is a continuous function then $g(x_n) \xrightarrow{d} g(x)$ (this is the analog to the Slutsky theorem).
(c) If y_n has limiting distribution and $p\lim(x_n - y_n) = 0$, then x_n has the same limiting distribution as y_n.

Example A.2 $F(1,n) \xrightarrow{d} \chi^2(1)$.

Theorem A.14 *Lindberg-Levy Central Limit Theorem (univariate).*
If $x_1, ..., x_n$ are a random sample from probability distribution with finite mean μ and finite variance σ^2 and $\bar{x} = n^{-1}\sum_{t=1}^{n} x_t$, then $\sqrt{n}(\bar{x} - \mu) \xrightarrow{d} N(0, \sigma^2)$.

Proof. (Rao [1973], p. 127)

Let $f(t)$ be the characteristic function of $X_t - \mu$.[2] Since the first two moments exist,

$$f(t) = 1 - \frac{1}{2}\sigma^2 t^2 + o\left(t^2\right)$$

The characteristic function of $Y_n = \frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n}(X_i - \mu)$ is

$$f_n(t) = \left[f\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n = \left[1 - \frac{1}{2}\sigma^2 t^2 + o\left(t^2\right)\right]^n$$

And

$$\log\left[1 - \frac{1}{2}\sigma^2 t^2 + o\left(t^2\right)\right]^n = n\log\left[1 - \frac{1}{2}\sigma^2 t^2 + o\left(t^2\right)\right]^n \to -\frac{t^2}{2}$$

That is, as $n \to \infty$

$$f_n(t) \to e^{-\frac{t^2}{2}}$$

Since the limiting distribution is continuous, the convergence of the distribution function of Y_n is uniform, and we have the more general result

$$\lim_{n\to\infty}\left[F_{Y_n}(x_n) - \Phi(x_n)\right] \to 0$$

where x_n may depend on n in any manner. This result implies that the distribution function of $\overline{X}_n$ can be approximated by that of a normal random variable with mean μ and variance $\frac{\sigma^2}{n}$ for sufficiently large n. ■

Theorem A.15 *Lindberg-Feller Central Limit Theorem (unequal variances). Suppose* $\{x_1, ..., x_n\}$ *is a set of random variables with finite means* μ_j *and finite variance* σ_j^2. *Let* $\bar{\mu} = n^{-1}\sum_{t=1}^{n}\mu_t$ *and* $\bar{\sigma}_n^2 = n^{-1}\left(\sigma_1^2 + \sigma_2^2 + ...\right)$. *If no single term dominates the average variance* ($\lim_{n\to\infty}\frac{\max(\sigma_j)}{n\bar{\sigma}_n} = 0$), *if the average variance converges to a finite constant* ($\lim_{n\to\infty}\bar{\sigma}_n^2 = \bar{\sigma}^2$), *and* $\bar{x} = n^{-1}\sum_{t=1}^{n}x_t$, *then* $\sqrt{n}(\bar{x} - \bar{\mu}) \xrightarrow{d} N\left(0, \bar{\sigma}^2\right)$.

Multivariate versions apply to both; the multivariate version of the Lindberg-Levy *CLT* follows.

[2] The characteristic function $f(t)$ is the complex analog to the moment generating function

$$\begin{aligned} f(t) &= \int e^{itx}dF(x) \\ &= \int \cos(tx)\,dF(x) + i\int \sin(tx)\,dF(x) \end{aligned}$$

where $i = \sqrt{-1}$ (Rao [1973], p. 99).

Theorem A.16 *Lindberg-Levy Central Limit Theorem (multivariate).* *If $X_1, ..., X_n$ are a random sample from multivariate probability distribution with finite mean vector μ and finite covariance matrix Q, and $\bar{x} = n^{-1} \sum_{t=1}^{n} x_t$, then $\sqrt{n}\left(\bar{X} - \mu\right) \xrightarrow{d} N(0, Q)$.*

Delta method.

The "Delta method" is used to justify usage of linear Taylor series approximation to analyze distributions and moments of a function of random variables. It combines Theorem A.9 Slutsky's probability limit, Theorem A.13 limiting distribution, and the Central Limit Theorems A.14-A.16.

Theorem A.17 *Limiting normal distribution of a function.* *If $\sqrt{n}(z_n - \mu) \xrightarrow{d} N\left(0, \sigma^2\right)$ and if $g(z_n)$ is a continuous function not involving n, then $\sqrt{n}\left(g(z_n) - g(\mu)\right) \xrightarrow{d} N\left(0, \{g'(\mu)\}^2 \sigma^2\right)$.*

A key insight for the Delta method is the mean and variance of the limiting distribution are the mean and variance of a *linear* approximation evaluated at μ, $g(z_n) \approx g(\mu) + g\prime(\mu)(z_n - \mu)$.

Theorem A.18 *Limiting normal distribution of a set of functions (multivariate).* *If z_n is a $K \times 1$ sequence of vector-valued random variables such that $\sqrt{n}(z_n - \mu_n) \xrightarrow{d} N(0, \Sigma)$ and if $c(z_n)$ is a set of J continuous functions of z_n not involving n, then $\sqrt{n}\left(c(z_n) - c(\mu_n)\right) \xrightarrow{d} N\left(0, C\Sigma C^T\right)$ where C is a $J \times K$ matrix with jth row a vector of partial derivatives of jth function with respect to z_n, $\frac{\partial c(z_n)}{\partial z_n^T}$.*

Definition A.7 *Asymptotic distribution.* *An asymptotic distribution is a distribution used to approximate the true finite sample distribution of a random variable.*

Example A.3 *If $\sqrt{n}\left[\frac{x_n - \mu}{\sigma}\right] \xrightarrow{d} N(0, 1)$, then approximately or asymptotically $\bar{x}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$. This is written $\bar{x}_n \xrightarrow{d} N\left(\mu, \frac{\sigma^2}{n}\right)$.*

Definition A.8 *Asymptotic normality and asymptotic efficiency.* *An estimator $\widehat{\theta}$ is asymptotically normal if $\sqrt{n}\left(\widehat{\theta} - \theta\right) \xrightarrow{d} N(0, V)$. An estimator is asymptotically efficient if the covariance matrix of any other consistent, asymptotically normally distributed estimator exceeds $n^{-1}V$ by a nonnegative definite matrix.*

Example A.4 *Asymptotic inefficiency of median in normal sampling.* *In sampling from a normal distribution with mean μ and variance σ^2, both the sample mean $\bar{x}$ and median M are consistent estimators of μ. Their asymptotic properties are $\bar{x}_n \xrightarrow{a} N\left(\mu, \frac{\sigma^2}{n}\right)$ and $M \xrightarrow{a} N\left(\mu, \frac{\pi}{2}\frac{\sigma^2}{n}\right)$. Hence,* *the sample mean is a more efficient estimator for the mean than the median by a factor of* $\pi/2 \approx 1.57$.

This result for the median follows from the next theorem (see Mood, Graybill, and Boes [1974], p. 257).

Theorem A.19 *Asymptotic distribution of order statistics.*
Let $x_1, ..., x_n$ *be iid random variables with density* f *and cumulative distribution function* F. F *is strictly monotone. Let* ξ_p *be a unique solution in* x *of* $F(x) = p$ *for some* $0 < p < 1$ *(*ξ_p *is the pth quantile). Let* p_n *be such that* np_n *is an integer and* $n|p_n - p|$ *is bounded. Let* $y_{np_n}^{(n)}$ *denote (np)th order statistic for a random sample of size* n. *Then* $y_{np_n}^{(n)}$ *is asymptotically distributed as a normal distribution with mean* ξ_p *and variance* $\frac{p(1-p)}{n[f(\xi_p)]^2}$.

Example A.5 *Sample median.*
Let $p = \frac{1}{2}$ *(implies* ξ_p *= sample median). The sample median* $M \xrightarrow{a} N\left(\xi_p, \frac{1}{4n[f(1/2)]^2}\right)$. *Since* $\xi_{\frac{1}{2}} = \mu$, $f\left(\xi_{\frac{1}{2}}\right)^2 = \left(2\pi\sigma^2\right)^{-1}$, *and the variance is* $\frac{\frac{1}{2}\left(\frac{1}{2}\right)}{nf\left(\xi_{\frac{1}{2}}\right)^2} = \frac{\pi}{2}\frac{\sigma^2}{n}$ *— the result asserted above.*

Theorem A.20 *Asymptotic distribution of nonlinear function.*
If $\widehat{\theta}$ *is a vector of estimates such that* $\widehat{\theta} \xrightarrow{a} N\left(\theta, n^{-1}V\right)$ *and if* $c(\theta)$ *is a set of* J *continuous functions not involving* n, *then*

$$c\left(\widehat{\theta}\right) \xrightarrow{a} N\left(c(\theta), n^{-1}C(\theta)VC(\theta)^T\right)$$

where $C(\theta) = \frac{\partial c(\theta)}{\partial \theta^T}$.

Example A.6 *Asymptotic distribution of a function of two estimates.*
Suppose b *and* t *are estimates of* β *and* θ *such that*

$$\begin{bmatrix} b \\ t \end{bmatrix} \xrightarrow{a} N\left(\begin{bmatrix} \beta \\ \theta \end{bmatrix}, \begin{bmatrix} \sigma_{\beta\beta} & \sigma_{\beta\theta} \\ \sigma_{\theta\beta} & \sigma_{\theta\theta} \end{bmatrix}\right)$$

We wish to find the asymptotic distribution for $c = \frac{b}{1-t}$. *Let* $\gamma = \frac{\beta}{1-\theta}$ *– the true parameter of interest. By the Slutsky Theorem and consistency of the sample mean,* c *is consistent for* γ. *Let* $\gamma_\beta = \frac{\partial\gamma}{\partial\beta} = \frac{1}{1-\theta}$ *and* $\gamma_\theta = \frac{\partial\gamma}{\partial\theta} = \frac{\beta}{(1-\theta)^2}$. *The asymptotic variance is*

$$Asy.Var[c] = \begin{bmatrix} \gamma_\beta & \gamma_\theta \end{bmatrix} \Sigma \begin{bmatrix} \gamma_\beta \\ \gamma_\theta \end{bmatrix} = \gamma_\beta\sigma_{\beta\beta} + \gamma_\theta\sigma_{\theta\theta} + 2\gamma_\beta\gamma_\theta\sigma_{\beta\theta}$$

Notice this is simply the variance of a linear approximation $\widehat{\gamma} \approx \gamma + \gamma\beta(b - \beta) + \gamma\theta(t - \theta)$.

Theorem A.21 *Asymptotic normality of MLE Theorem*
MLE, $\hat{\theta}$, for strongly asymptotically identified model represented by log-likelihood function $\ell(\theta)$, when it exists and is consistent for θ, is asymptotically normal if
(i) contributions to log-likelihood $\ell_t(y,\theta)$ are at least twice continuously differentiable in θ for almost all y and all θ,
(ii) component sequences $\left\{D^2_{\theta\theta}\ell_t(y,\theta)\right\}_{t=1,\infty}$ satisfy WULLN (weak uniform law) on θ,
(iii) component sequences $\{D_\theta \ell_t(y,\theta)\}_{t=1,\infty}$ satisfy CLT.

A.3 Rates of convergence

Definition A.9 *Order $1/n$ (big-O notation).*
If f and g are two real-valued functions of positive integer variable n, then the notation $f(n) = O(g(n))$ (optionally as $n \to \infty$) means there exists a constant $k > 0$ (independent of n) and a positive integer N such that $\left|\frac{f(n)}{g(n)}\right| < k$ for all $n < N$. ($f(n)$ is of same order as $g(n)$ asymptotically).

Definition A.10 *Order less than $1/n$ (little-o notation).*
If f and g are two real-valued functions of positive integer variable n, then the notation $f(n) = o(g(n))$ means the $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$ ($f(n)$ is of smaller order than $g(n)$ asymptotically).

Definition A.11 *Asymptotic equality.*
If f and g are two real-valued functions of positive integer variable n such that $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$, then $f(n)$ and $g(n)$ are asymptotically equal. This is written $f(n) \overset{a}{=} g(n)$.

Definition A.12 *Stochastic order relations.*
If $\{a_n\}$ is a sequence of random variables and $g(n)$ is a real-valued function of positive integer argument n, then
(1) $a_n = o_p(g(n))$ means $\lim_{n\to\infty} \frac{a_n}{g(n)} = 0$,
(2) similarly, $a_n = O_p(g(n))$ means there is a constant k such that (for all $\varepsilon > 0$) there is a positive integer N such that $\Pr\left(\left|\frac{a_n}{g(n)}\right| > k\right) < \varepsilon$ for all $n > N$, and
(3) If $\{b_n\}$ is a sequence of random variables, then the notation $a_n \overset{a}{=} b_n$ means $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$.

Comparable definitions apply to almost sure convergence and convergence in distribution (though these are infrequently used).

Theorem A.22 *Order rules:*

$$O(n^p) \pm O(n^q) = O\left(n^{max(p,q)}\right)$$

$$o(n^p) \pm o(n^q) = o\left(n^{max(p,q)}\right)$$

$$\begin{aligned} O(n^p) \pm o(n^q) &= O\left(n^{max(p,q)}\right) && \text{if } p \geq q \\ &= o\left(n^{max(p,q)}\right) && \text{if } p < q \end{aligned}$$

$$O(n^p)\, O(n^q) = O\left(n^{p+q}\right)$$

$$o(n^p)\, o(n^q) = o\left(n^{p+q}\right)$$

$$O(n^p)\, o(n^q) = o\left(n^{p+q}\right)$$

Example A.7 *Square-root n convergence.*
(1) If each $x = O(1)$ *has mean* μ *and the central limit theorem applies* $\sum_{t=1}^{n} x_t = O(n)$ *and* $\sum_{t=1}^{n} (x_t - \mu) = O(\sqrt{n})$.
(2) Let $\Pr(y_t = 1) = 1/2$, $\Pr(y_t = 0) = 1/2$, $z_t = y_t - 1/2$, *and* $b_n = \sqrt{n} \sum_{t=1}^{n} z_t$. $Var[b_n] = n^{-1} Var[z_t] = n^{-1}(1/4)$. $\sqrt{n} b_n = n^{-\frac{1}{2}} \sum_{t=1}^{n} z_t$.

$$E\left[\sqrt{n} b_n\right] = 0$$

and

$$Var\left[\sqrt{n} b_n\right] = 1/4$$

Thus, $\sqrt{n} b_n = O(1)$ *which implies* $b_n = O\left(n^{-\frac{1}{2}}\right)$.

These examples represent common econometric results. That is, the average of n centered quantities is $O\left(n^{-\frac{1}{2}}\right)$, and is referred to as square-root n convergence.

A.4 Additional reading

Numerous books and papers including Davidson and MacKinnon [1993, 2003], Greene [1997], and White [1984] provide in depth review of asymptotic theory. Hall and Heyde [1980] reviews limit theory (including laws of large numbers and central limit theorems) for martingales.