# Asymptotic Concepts

L. Magee January, 2010

---

## 1 Definitions of Terms Used in Asymptotic Theory

Let $a_n$ to refer to a random variable that is a function of $n$ random variables. An example is a sample mean

$$a_n = \bar{x} = n^{-1} \sum_{i=1}^{n} x_i$$

**Convergence in Probability**
The scalar $a_n$ converges in probability to a constant $\alpha$ if, for any positive values of $\epsilon$ and $\delta$, there is a sufficiently large $n^*$ such that

$$\text{Prob}(|a_n - \alpha| > \epsilon) < \delta \text{ for all } n > n^*$$

$\alpha$ is called the probability limit, or plim, of $a_n$.

**Consistency**
If $a_n$ is an estimator of $\alpha$, and plim $a_n = \alpha$, then $a_n$ is a (weakly) consistent estimator of $\alpha$.

**Convergence in Distribution**
$a_n$ converges in distribution to a random variable $y$ ($a_n \rightarrow y$) if, as $n \rightarrow \infty$, Prob$(a_n \leq b) =$ Prob$(y \leq b)$ for all b. In other words, the distribution of $a_n$ becomes the same as the distribution of $y$.

**Examples** Let $x_i \sim N[\mu, \sigma^2], \ i = 1, \ldots, n$, where the $x_i$'s are mutually independently distributed. Define the three statistics

1. $\bar{x} = n^{-1} \sum_{i=1}^{n} x_i$
2. $s^2 = (n-1)^{-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$
3. $t = \frac{\bar{x} - \mu}{(s^2/n)^{1/2}}$

Considering each statistic one-by-one,

1. As $n \to \infty$, $\text{Var}(\bar{x}) \to 0$. This implies that $\text{plim}(\bar{x}) = \mu$, and $\bar{x}$ is a consistent estimator of $\mu$. (Uses fact that $\text{Var}(\bar{x}) = \sigma^2/n$ and $E(\bar{x}) = \mu$.)

2. As $n \to \infty$, $\text{Var}(s^2) \to 0$, so $\text{plim}(s^2) = \sigma^2$ and $s^2$ is a consistent estimator of $\sigma^2$. (Uses distributional result: $(n-1)s^2/\sigma^2 \sim \chi^2_{n-1}$. Since $E(\chi^2_{n-1}) = n-1$ and $\text{Var}(\chi^2_{n-1}) = 2(n-1)$, then $Es^2 = \sigma^2$ and $\text{Var}(s^2) = 2\sigma^4/(n-1)$.)

3. $t$ converges in distribution to $z$, where $z \sim N[0,1]$. (Uses distributional result: $t \sim t_{n-1}$. Since $\text{plim}(s^2) = \sigma^2$, then as $n \to \infty$, $t \to (\bar{x} - \mu)/(\sigma^2/n)^{1/2}$. This is a standardized normal random variable.)

**Properties**

(i) if $\text{plim}(x_n) = \theta_x$, then $\text{plim}(g(x_n)) = g(\theta_x)$, for any function $g(\cdot)$ that is continuous at $\theta_x$. This is sometimes called *Slutsky's theorem.*

(ii) if $x_n$ converges in distribution to some random variable $x$, i.e. $x_n \to x$, then, for any function $g(\cdot)$, $g(x_n) \to g(x)$. That is, the distribution of $g(x_n)$ converges to the distribution of $g(x)$. (This is like property (i) but for convergence in distribution instead of convergence in probability.)

(iii) if $\text{plim}(x_n) = \theta_x$ and $\text{plim}(y_n) = \theta_y$, then $\text{plim}(x_n y_n) = \theta_x \theta_y$.

(iv) if $\text{plim}(x_n) = \theta_x$ and $y_n \to y$, then $x_n y_n \to \theta_x y$.
Often $x_n$ is a matrix and $y$ is a normally distributed vector. Similarly, $y_n'(x_n)^{-1} y_n \to y'(\theta_x)^{-1} y$, which relates to the asymptotic chi-square distribution often encountered in hypothesis testing.

## 2 Order Notation

It is useful to have notation that describes the rate that a statistic converges to zero or goes off to infinity as the sample size $n$ grows. First, consider $f(n)$, some non-random function of $n$.

**Definitions**

(i) $f(n)$ is $O(n^d)$ ("is order $n^d$") if, as $n \to \infty$, then $f(n)/n^d$ is finite. (If $d > 0$, then $f(n)$ grows to infinity at the same rate as $n^d$, and if $d < 0$, $f(n)$ shrinks to zero at the same rate as $n^d$.)

(ii) $f(n)$ is $o(n^d)$ ("is order smaller than $n^d$") if, as $n \to \infty$, then $f(n)/n^d \to 0$.

## 2.1 Examples

(i) -3 is $O(1)$, or $O(n^0)$

(ii) $5n^3$ is $O(n^3)$, and it is $o(n^4)$ and $o(n^5)$

(iii) $\frac{3}{n}$ is $O(n^{-1})$ and is $o(1)$

(iv) $\frac{3}{n} - \frac{2}{n^{3/2}}$ is $O(n^{-1})$

Example (iv) illustrates that the order of a sum depends only on the order of the highest-order term (meaning the term with the largest "$d$" in $O(n^d)$) as long as the number of terms in the sum does not depend on $n$.

If the number of terms in a sum itself depends on $n$, then the order of the sum can be affected. Let $x_i$ and $x_{ij}$ be $O(n^0)$ constants. For example, $x_i$ and $x_{ij}$ do not display a trend as $i$ or $j$ increase. Then, in general, $\sum_{i=1}^n x_i$ is $O(n)$ and $\sum_{i=1}^n \sum_{j=1}^n x_{ij}$ is $O(n^2)$. (For an example of what happens when $x_i$ is <u>not</u> an $O(n^0)$ constant, consider $\sum_{i=1}^n x_i$ when $x_i = a + bi$. Then $\sum_{i=1}^n x_i = na + bn(n+1)/2$, which is $O(n^2)$ when $b \neq 0$.)

Here are some rules for operations involving order notation:

| | | |
|---|---|---|
| $O(n^p) + O(n^q)$ | is | $O(n^{\max(p,q)})$ |
| $o(n^p) + o(n^q)$ | is | $o(n^{\max(p,q)})$ |
| $O(n^p) + o(n^q)$ | is | $O(n^p)$ if $p \geq q$ (already mentioned in example (iv)) |
| | and is | $o(n^q)$ if $p < q$ |
| $O(n^p) \times O(n^q)$ | is | $O(n^{p+q})$ |
| $O(n^p) \times o(n^q)$ | is | $o(n^{p+q})$ |
| $o(n^p) \times o(n^q)$ | is | $o(n^{p+q})$ |
| $(O(n^p))^{-1}$ | is | $O(n^{-p})$ |
| $(o(n^p))^{-1}$ | is | of unknown order without more information |

Combining these gives other results, such as $O(n^p)/O(n^q)$ is $O(n^{p-q})$.

## 2.2 Order in Probability

Order notation can be applied to random variables, using a "$p$" subscript, so that $O_p$ denotes "order in probability". Let $a_n$ be a random variable function of $n$ random variables as on page 1.

**Definitions**

(i) $a_n$ is $O_p(n^d)$ if, for every $\epsilon > 0$, there is some $K > 0$ for which, as $n \to \infty$, $\text{Prob}(|a_n/n^d| > K) < \epsilon$.

(ii) $a_n$ is $o_p(n^d)$ if as $n \to \infty$, then $\text{plim}(a_n/n^d) = 0$.

Except for special cases that usually do not apply to econometric models, (i) is equivalent to the condition that the mean and the standard deviation of $a_n/n^d$ stay bounded as $n \to \infty$. Therefore if $a_n$ is $O_p(n^d)$ then: (1) $E(a_n)$ is $O(n^d)$ and (2) $\text{Var}(a_n)$ is $O(n^{2d})$. Another way to think about it is, if $E(a_n)$ is $O(n^d)$ and $\text{Var}(a_n)$ is $O(n^f)$ then $a_n$ is $O(n^g)$, where $g = \max(d, f/2)$.

## 2.3 Examples of Order in Probability

Let $x_i,\ i = 1, \ldots, n$, be independent random variables with mean $\mu$ and variance $\sigma^2$.

(i) $x_i$ is $O_p(1)$

(ii) $\sum_{i=1}^n x_i$ is $O_p(n^{1/2})$ if $\mu = 0$, and it is $O_p(n)$ if $\mu \neq 0$

(iii) $\sum_{i=1}^n x_i^2$ is $O_p(n)$ unless $\mu = \sigma^2 = 0$

(ii) arises often in asymptotic theory. If $\mu = 0$, then $\sum_{i=1}^n x_i$ has a mean of 0 and variance $n\sigma^2$, implying $\sum_{i=1}^n x_i/n^{1/2}$ has a mean of 0 and a variance of $\sigma^2$. So if $\mu = 0$ then $\sum_{i=1}^n x_i$ is $O_p(n^{1/2})$. However, if $\mu \neq 0$, then the mean of $\sum_{i=1}^n x_i$ is $n\mu$, implying that $\sum_{i=1}^n x_i$ is $O_p(n)$, not $O_p(n^{1/2})$.

(iii) follows from the fact that $x_i^2$ has a finite non-trending mean, $m_1 = \mu^2 + \sigma^2$, and finite variance, $m_2$. Then $\sum_{i=1}^n x_i^2/n$ has a finite mean $m_1$ and variance $m_2/n$.

## 2.4 Asymptotic Expansions

Many consistent estimators are "root-$n$ consistent", meaning that the sampling error is $O_p(n^{-1/2})$, as in $\hat{\theta} - \theta = O_p(n^{-1/2})$. Asymptotic expansions simplify the analysis of the distribution of $\hat{\theta}$, by ignoring the part of $\hat{\theta} - \theta$ that is $o_p(n^{-1/2})$. This involves decomposing $\hat{\theta}$ as

$$\hat{\theta} \;=\; \theta + \xi_{-1/2} + o_p(n^{-1/2}) \qquad (1)$$

The right hand side of (1) contains three terms of declining importance as $n \to \infty$. The first term is $O(1)$, and must equal the true parameter value if $\hat{\theta}$ is consistent. The second term is $O_p(n^{-1/2})$.

It usually has mean zero, and often is simple enough to enable the derivation of $E(\xi^2_{-1/2})$, which is then used for estimating $\text{Var}(\hat{\theta})$. The third term is the remainder. It is left out in most asymptotic approximations. We hope it is not very big compared to the first two terms. Its importance often can be examined most easily by simulations.

## 3 Application to Ordinary Least Squares with Heteroskedastic Errors

Assume the true model is $y_i = x_i'\beta + u_i$, where $Ex_iu_i = 0$, $\text{Var}(u_i|x_i) = \sigma_i^2$, and the $u_i$'s are independent. Consider the OLS estimator of $\beta$, $b = (\sum x_ix_i')^{-1}\sum x_iy_i$. Unless indicated otherwise, the summations run over $i$ from 1 to $n$, where $n$ is the number of observations. Assume that the $x_i$'s are random, as in survey data where the randomness in both $x_i$ and $y_i$ derives from the random survey sampling.

Aside on notation: $x_i$ is a $k \times 1$ vector of observations on the RHS variables. $x_i'$ is the $i^{th}$ row of the usual $n \times k$ matrix $X$, and $y_i$ is the $i^{th}$ element of the usual $n \times 1$ vector $y$. So in this vector notation, a matrix product such as $X'X$ is written as $\sum x_ix_i'$.

Substituting out $y_i$ gives

$$b = \beta + (\sum x_ix_i')^{-1}\sum x_iu_i \tag{2}$$

Relating this to (1), $\beta$ is $O(1)$, and there is no remainder term. To find the order of the second RHS term, consider its two parts separately.

$\sum x_ix_i'$ is a $k \times k$ matrix. Assume that $n^{-1}\sum x_ix_i'$ converges to some finite positive definite matrix $\Sigma_{xx}$. Then $n^{-1}\sum x_ix_i' = \Sigma_{xx} + o_p(1)$.

$\sum x_iu_i$, might appear to be $O_p(n)$ since it is the sum of $n$ terms that are $O_p(1)$. But since we have assumed that $Ex_iu_i = 0$, then $E(\sum x_iu_i) = 0$ and $\text{Var}(\sum x_iu_i) = E\left((\sum_i x_iu_i)(\sum_j x_ju_j)'\right) = E(\sum x_ix_i'u_i^2) = O(n)$. (This result uses independence of the $x_iu_i$'s to get $E(x_iu_i)(x_ju_j)' = 0$ for all $i \neq j$.) Therefore we only need to multiply $\sum x_iu_i$ by $n^{-1/2}$ to give it an $O(1)$ variance, so $\sum x_iu_i$ is $O_p(n^{1/2})$. This is like example (ii) in section 2.3 when $\mu = 0$. $x_i$ in that example is replaced here by $x_iu_i$.

The second term of the RHS of (2), then, is the product of an $O(n^{-1})$ term and an $O_p(n^{1/2})$ term. It follows that this second term is $O_p(n^{-1/2})$ if $Ex_iu_i = 0$. In this case, $b$ is consistent. If it had been the case that $Ex_iu_i \neq 0$, then this term would have been $O_p(1)$, and $b$ would not have been consistent.

### 3.1 Variance of $b$

Since $Ex_iu_i = 0$, then $Eb = \beta$ from (2), the variance of $b$ is the variance of $(\sum x_i x_i')^{-1} \sum x_i u_i$. It is convenient to multiply each term by the appropriate power of $n$ so that we can work with $O(1)$ and $O_p(1)$ terms.

$$b - \beta \;=\; n^{-1/2}(n^{-1}\sum x_i x_i')^{-1}(n^{-1/2}\sum x_i u_i) \qquad (3)$$

and

$$\begin{aligned} \text{Var}(b) &= E(b-\beta)(b-\beta)' \\ &= n^{-1}E(n^{-1}\sum x_i x_i')^{-1}(n^{-1/2}\sum x_i u_i)(n^{-1/2}\sum x_i u_i)'(n^{-1}\sum x_i x_i')^{-1} \qquad (4) \end{aligned}$$

Substituting $n^{-1}\sum x_i x_i' = \Sigma_{xx} + o_p(1)$, then

$$\begin{aligned} \text{Var}(b) &= n^{-1}E\left((\Sigma_{xx} + o_p(1))^{-1}\right)\left((n^{-1/2}\sum x_i u_i)(n^{-1/2}\sum x_i u_i)'\right)\left((\Sigma_{xx} + o_p(1))^{-1}\right) \\ &= n^{-1}E\Sigma_{xx}^{-1}\left((n^{-1/2}\sum x_i u_i)(n^{-1/2}\sum x_i u_i)'\right)\Sigma_{xx}^{-1} + o(n^{-1}) \\ &= n^{-1}\Sigma_{xx}^{-1}(n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}(Ex_iu_i(x_ju_j)'))\Sigma_{xx}^{-1} + o(n^{-1}) \qquad (5) \end{aligned}$$

The assumption that the $x_iu_i$'s are not correlated allows us to set the expected values of the cross-product terms $(x_iu_i)(x_ju_j)'$ in (5) equal to zero. Then (5) can be written as

$$\begin{aligned} \text{Var}(b) &= n^{-1}\Sigma_{xx}^{-1}\left(n^{-1}(E\sum x_i x_i' u_i^2)\right)\Sigma_{xx}^{-1} + o(n^{-1}) \\ &= n^{-1}A^{-1}BA^{-1} + o(n^{-1}) \qquad (6) \end{aligned}$$

where $A = \Sigma_{xx} = Ex_ix_i'$ and $B = n^{-1}(E\sum x_i x_i' u_i^2) = Ex_ix_i'u_i^2$

A consistent estimator of Var($b$) is formed by replacing $A$ and $B$ by consistent estimators $\hat{A}$ and $\hat{B}$. Common choices are the sample means $\hat{A} = n^{-1}\sum x_i x_i'$ and $\hat{B} = n^{-1}\sum x_i x_i' e_i^2$, where $e_i =$

$y_i - x_i'b = y_i - x_i'\beta + o_p(1) = u_i + o_p(1)$ . Since $\hat{A} = A + o_p(n^{-1})$ and $\hat{B} = B + o_p(n^{-1})$, then

$$\begin{aligned}
\text{Var}(b) &= n^{-1}\hat{A}^{-1}\hat{B}\hat{A}^{-1} + o_p(n^{-1}) \\
&= n^{-1}(n^{-1}\sum x_i x_i')^{-1}(n^{-1}\sum x_i x_i' e_i^2)(n^{-1}\sum x_i x_i')^{-1} + o_p(n^{-1}) \\
&= (\sum x_i x_i')^{-1}(\sum x_i x_i' e_i^2)(\sum x_i x_i')^{-1} + o_p(n^{-1}) \\
&= \hat{V}(b) + o_p(n^{-1})
\end{aligned}$$

This $\hat{V}(b) = (\sum x_i x_i')^{-1}(\sum x_i x_i' e_i^2)(\sum x_i x_i')^{-1}$ is White's (1980) *heteroskedasticity-consistent variance-covariance matrix*. It is used in the `robust` option in Stata, and referred to as HCCME in Davidson and MacKinnon (1993, p.552) and elsewhere.

## 3.2 Variance of $b$ when errors are homoskedastic

If the errors are homoskedastic ($Eu_i^2 = \sigma^2$ for all $i$), then $Eu_i^2$ is unrelated to the elements of $x_i x_i'$, which allows for $E(x_i x_i' u_i^2)$ to be split into two separate multiplicative expectation terms, and $E(x_i x_i' u_i^2) = E(u_i^2)E(x_i x_i') = \sigma^2\Sigma_{xx}$. (The general property is that if random variables $a$ and $b$ are independently distributed, then $E(ab) = E(a)E(b)$.) The asymptotic variance (6) simplifies as

$$\begin{aligned}
\text{Var}(b) &= n^{-1}\Sigma_{xx}^{-1}\left(n^{-1}(E\sum x_i x_i' u_i^2)\right)\Sigma_{xx}^{-1} + o(n^{-1}) \\
&= n^{-1}\Sigma_{xx}^{-1}(\sigma^2\Sigma_{xx})\Sigma_{xx}^{-1} + o(n^{-1}) \\
&= n^{-1}\sigma^2\Sigma_{xx}^{-1} + o(n^{-1})
\end{aligned}$$

This can be consistently estimated by the usual OLS variance-covariance estimator, since

$$s^2 \equiv \frac{e'e}{n-k} = \sigma^2 + o_p(1) \quad \text{and} \quad n^{-1}\sum_i x_i x_i' = \Sigma_{xx} + o_p(1)$$

so that

$$(n^{-1}\sum_i x_i x_i')^{-1} = \Sigma_{xx}^{-1} + o_p(1) \quad \text{or} \quad (\sum_i x_i x_i')^{-1} = n^{-1}\Sigma_{xx}^{-1} + o_p(n^{-1})$$

and therefore

$$\begin{aligned}
s^2(\sum_i x_i x_i')^{-1} &= \left(\sigma^2 + o_p(1)\right)\left(n^{-1}\Sigma_{xx}^{-1} + o_p(n^{-1})\right) = \left(\sigma^2 + o_p(1)\right)\left(n^{-1}\Sigma_{xx}^{-1} + o_p(n^{-1})\right) \\
&= n^{-1}\sigma^2\Sigma_{xx}^{-1} + o_p(n^{-1}) = \text{Var}(b) + o_p(n^{-1})
\end{aligned}$$

# 4 Central Limit Theorem (CLT)

To know the asymptotic distribution of test statistics and compute asymptotic confidence intervals, we require the asymptotic distribution of $b$, not just a variance estimator. Fortunately, under a broad range of assumptions, including the ones here, the asymptotic distribution is normal. The main result that leads to this is the central limit theorem, originating with Laplace in 1810. It has many versions depending on the assumptions made. One version is:

*If $n$ random vectors $a_i$ are independently and identically distributed, with mean $\mu$ and variance $\Sigma$, then the distribution of $n^{1/2}(n^{-1}\sum_i a_i - \mu)$ converges to $N[0, \Sigma]$.*

In practice, we can approximate the distribution of the sample mean of $a_i$ vectors, $n^{-1}\sum_i a_i$, as $N[\mu, n^{-1}\Sigma]$. Since this result is asymptotic, it also is valid to approximate the distribution as $N[\mu, \hat{V}]$, where $n\hat{V}$ is a consistent estimator of $\Sigma$.

## 4.1 OLS with heteroskedastic errors

In the above regression example,

$$n^{1/2}(b - \beta) = (n^{-1}\sum_i x_i x_i')^{-1}(n^{-1/2}\sum_i x_i u_i)$$

The first RHS term has $\text{plim}(n^{-1}\sum_i x_i x_i') = \Sigma_{xx}$. The CLT applies to the second RHS term. To match it up with the way the CLT was presented above, express the second term as $n^{1/2}(n^{-1}\sum_i x_i u_i - 0)$. Let $\text{Var}(x_i u_i) = \Sigma_{xu}$. Then, also using $Ex_i u_i = 0$, the CLT implies that $n^{-1/2}\sum_i x_i u_i$ converges in distribution to $N[0, \Sigma_{xu}]$. Then $n^{1/2}(b - \beta)$ converges in distribution to the distribution of $\Sigma_{xx}^{-1}$ times $(n^{-1/2}\sum_i x_i u_i)$, which, using standard results for variances, is $N[0, \Sigma_{xx}^{-1}\Sigma_{xu}\Sigma_{xx}^{-1}]$. The variance, $\Sigma_{xx}^{-1}\Sigma_{xu}\Sigma_{xx}^{-1}$, is what is being estimated by $n\hat{V}(b)$, where $\hat{V}(b)$ is White's HCCME from section 3.1.

# 5 References

Davidson, R. and J.G. MacKinnon (1993), *Estimation and Inference in Econometrics*, Oxford University Press, Oxford.

White, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817-838.