LECTURE 8: ASYMPTOTICS I

We are interested in the properties of estimators as 

$n \to \infty$.

Consider a sequence of random variables 

$\{ X_n, n \geq 1 \}$. 
• **Definition: (Weak convergence)** A sequence of random variables \( \{X_n, n \geq 1\} \) is said to converge weakly to a constant \( c \) if

\[
\lim_{n \to \infty} P(|X_n - c| > \varepsilon) = 0
\]

for every given \( \varepsilon > 0 \).

This is written \( \text{plim } X_n = c \) or \( X_n \overset{P}{\to} C \) and is also called convergence in probability.
Definition: (Strong convergence) A sequence of random variables is said to converge strongly to a constant \( c \) if

\[
P(\lim_{n \to \infty} X_n = c) = 1.
\]

or

\[
\lim_{N \to \infty} P(\sup_{n > N} |x_n - c| > \varepsilon) = 0
\]

Strong convergence is also called almost sure convergence or convergence with probability one and is written \( X_n \to C \) w.p.1 or \( X_n \to C \).
LAWS OF LARGE NUMBERS:

Let \( \{X_n, n \geq 1\} \) be observations and suppose we look at the sequence

\[
\bar{X}_n = \sum_{i=1}^{n} X_i / n.
\]

when does \( \bar{X}_n \to \xi \) where \( \xi \) is some parameter?

Weak Law of Large Numbers: (WLLN) Let \( E(X_i) = \mu, V(X_i) = \sigma^2, \) \( \text{cov}(X_iX_j) = 0. \)

Then \(- \mu \to 0\) in probability
Proof: Recall Chebyshev's inequality:

\[ P\left(\left| X - \mu \right| \geq k \right) \leq \frac{\sigma^2}{k^2} \]

where \( \mu = E(X) \) and \( \sigma^2 = V(X) \).

Proof of Chebyshev’s inequality

\[ \sigma^2 = \int (x - \mu)^2 dF \]

\[ = \int_{-\infty}^{\mu-\lambda \sigma} (x - \mu)^2 dF + \int_{\mu-\lambda \sigma}^{\mu+\lambda \sigma} (x - \mu)^2 dF + \int_{\mu+\lambda \sigma}^{\infty} (x - \mu)^2 dF \]

Put in the smallest value of x in the first and last integral, and drop the middle to get:

\[ \sigma^2 \geq \lambda^2 \sigma^2 P\left(\left| x - \mu \right| \geq \lambda \sigma \right) \]
Since we are interested in $X_N$, note that

$$E(X_N) = \mu \text{ and } V(X_N) = \frac{\sigma^2}{n}.$$  

Consequently,

$$\lim_{n \to \infty} P\left( \left| \overline{X}_n - \mu \right| > \varepsilon \right) = \lim_{n \to \infty} \frac{\sigma^2}{n \varepsilon^2} = 0.$$
**Notes:**

1. $E(X_i) = \mu_i$ is O.K. Consider $X_n - \bar{\mu}_n$ with $\bar{\mu}_n = n^{-1} \sum \mu_i$.

2. $V(X_i) = \sigma_i^2$ is O.K.. As long as $\lim \sum \sigma_i^2 / n^2 = 0$, our proof applies.

3. Existence of $\sigma^2$ can be dropped if we assume independent and identically distributed observations. (In this case, the proof is different.)

**Strong Law of Large Numbers: (SLLN)** $X_i$ are independent with $E(X_i) = \mu_i$, $V(X_i) = \sigma_i^2$ and $\sum \sigma_i^2 / i^2 < \infty$. Then $X_n - \bar{\mu}_n \to 0$ almost surely (a.s.).
Note: Again, we can drop assumption on the existence of $\sigma^2_i$ if we assume independent and identically distributed observations.

Some properties of plim:

1. \( \text{plim } XY = \text{plim } X \text{ plim } Y \)

2. \( \text{plim } (X+Y) = \text{plim } X + \text{plim } Y \)

3. Slutsky's theorem: If the function \( g \) is continuous at \( \text{plim } X \), then \( \text{plim } g(X) = g(\text{plim } X) \).
CENTRAL LIMIT THEOREM: 
(Asymptotic Normality)

**Definition:** The moment generating function is defined as

\[
m(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} f(x)dx.
\]

The name comes from the fact that

\[
\frac{d^r m}{dt^r} = \int_{-\infty}^{\infty} x^r e^{tx} f(x)dx = E(X^r)
\]

when evaluated at \( t = 0 \).
Note the following series expansion:

\[ m(t) = E(e^{itX}) = E(1 + Xt + \frac{1}{2!}(Xt)^2 + \ldots) \]

\[ = 1 + \alpha_1 t + \frac{1}{2} \alpha_2 t^2 + \ldots \]

where \( \alpha_r = EX^r \)
(For example: \( \alpha_1 = \mu_1, \alpha_2 = \mu^2 + \sigma^2 \)).

**Property 1:** The moment generating function of \( \sum_{i=1}^{n} X_i \) when \( X_i \) are independent is the product of the moment generating functions of \( X_i \). (*Exercise: Prove this.*)

**Property 2:** Let \( X \) and \( Y \) be random variables with continuous densities \( f(x) \) and \( g(y) \). If the moment generating function of \( X \) is equal to that of \( Y \) in an interval \(-h < t < h\), then \( f = g \).
Example: The moment generating function for $X \sim q(0,1)$

$$m(t) = E(e^{tx}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-\frac{1}{2}x^2} \, dx$$

$$= e^{t^2/2} \left( \frac{1}{\sqrt{2\pi}} \right) \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-t)^2} \, dx = ?$$

Uses of Asymptotic Distributions:

Suppose $\bar{X}_n - \mu \to 0$ in probability. (What can be said about the distribution of $\bar{X}_n - \mu$?)

In order to get distribution theory, we need to norm the random variable; we usually look at $n^{1/2} (\bar{X}_n - \mu)$.  

N. M. Kiefer, Cornell University, Economics 620
Note that the random variable sequence {\(n^{1/2}(\bar{X}_n - \mu), n \geq 1\)} does not converge in probability. (why not?)

We might be able to make probability statements like

\[
\lim_{n \to \infty} P(n^{1/2}(\bar{X}_n - \mu) < z) = F(z)
\]

for some distribution \(F\).

Then we could use \(F\) as an approximate distribution for \(n^{1/2}(\bar{X}_n - \mu)\). This, of course, implies an approximate distribution for \(\bar{X}_n\).

It is a little easier to work with

\[
Y_n = n^{1/2}(\bar{X}_n - \mu)/\sigma.
\]
Central Limit Theorem: (CLT) (Lindberg-Levy)  The distribution of \( Y_n \) (as defined above) as \( n \to \infty \) is

\[
\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2} \, dx
\]

(standard normal)

Proof: Let \( m_{X_i - \mu} (t) \) be the moment generating function of \( (X_i - \mu) \). That is,

\[
m_{X_i - \mu} (t) = 1 + \frac{\sigma^2 t^2}{2} + o(t^2)
\]

where \( o(t^2) \) is the remainder term such that \( o(t^2)/ t^2 \to 0 \) as \( t \to 0 \).
We know that

$$Y_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \frac{\sum_{i=1}^{n} (X_i - \mu)}{\sigma \sqrt{n}}.$$

The moment generating function of $Y_n$ is

$$m_{Y_n}(t) = \left[ m_{X_i-\mu} \left( \frac{t}{\sigma \sqrt{n}} \right) \right]^n$$

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

$$\Rightarrow \ln m_{Y_n}(t) = n \ln \left[ 1 + \frac{t^2}{2n} + o \left( \frac{t^2}{n} \right) \right]$$

$$\approx n \ln \left[ 1 + \frac{t^2}{2n} \right]$$
⇒ As \( n \to \infty \), \( m_{Y_n}(t) \to e^{t^2/2} \) which is the moment generating function of a standard normal random variable.

**Point of the Central Limit Theorem:** The distribution function of \( X_n \) for large \( n \) can be approximated by that of a normal with mean \( \mu \) and variance \( \sigma^2/n \).

**Qualification:** We really should use characteristic function \( C(t) = E e^{itX} \) in the proof.
Notes:
1. Identical means and variances can be dropped straightforwardly. We need some restrictions on the variance sequence though. In this case, we work with
\[ Y_n = \frac{\sum_{i=1}^{n} (X_i - \mu_i)}{\left(\sum_{i=1}^{n} \sigma_i^2\right)^{1/2}}. \]

2. Versions of the Central Limit Theorem with random vectors are also available.

3. The basic requirements is that each term in the sum should make a negligible contribution.
Examples:
1. Estimation of mean $\mu$ from a sample of normal random variables: In this case, we estimate $\mu$ by $\bar{X}$, and the asymptotic approximation for the distribution of $\bar{X}$ or $(\bar{X} - \mu)$ is exact.

2. Consider $n^{1/2}(\hat{\beta} - \beta)$ where $\hat{\beta}$ is the LS estimator.

$$n^{1/2}(\hat{\beta} - \beta) = n^{1/2} (X'X)^{-1} X' \varepsilon$$
$$= [X'X / n]^{-1} n^{1/2} [X' \varepsilon / n]$$

Where $[X'X / n]$ is the sample second moment matrix of the regressors.
Under the assumption that regressors are well-behaved (i.e. contribution of any particular observation to \([X'\varepsilon/n]\) is negligible), we can apply a Central Limit Theorem and conclude that

\[n^{1/2}(\hat{\beta} - \beta) = [X'X/n]^{-1} n^{1/2} [X'\varepsilon/n] \xrightarrow{D} q(0, \sigma^2 Q^{-1})\].
Definition: (Convergence in distribution)

A sequence of random variables \( \{Z_n, n \geq 1\} \) with distribution functions \( \{F_n(z) = P(Z_n \leq z), n \geq 1\} \) is said to converge in distribution to a random variable \( Z \) with distribution function \( F(z) \) if and only if
\[
\lim_{n \to \infty} F_n(z) = F(z)
\]

at all points of continuity of \( F(z) \).

Notation: \( Z_n \xrightarrow{D} Z \).
Some properties of convergence in probability (plim) and convergence in distribution:

1. $X_n$ and $Y_n$ are random variable sequences. If $\text{plim} (X_n - Y_n) = 0$ and $Y_n \xrightarrow{D} Y$, then $X_n \xrightarrow{D} Y$ as well.

2. If $Y_n \xrightarrow{D} Y$ and $X_n \rightarrow c$ in probability (i.e. $\text{plim} X_n = c$), then
   
   a. $X_n + Y_n \xrightarrow{D} c + Y$
   
   b. $X_n Y_n \xrightarrow{D} cY$
   
   c. $Y_n / X_n \xrightarrow{D} Y / c$, $c \neq 0$.

3. If $X_n \xrightarrow{D} X$ and $g$ is a continuous function, then $g(X_n) \xrightarrow{D} g(X)$. 