# Lecture 2
# Limit theorems

## 1 Useful Inequalities

**Theorem 1.** *(Markov inequality) Let $X$ be any nonnegative random variable such that $E[X]$ exists. Then for any $t > 0$, we have $P\{X \geq t\} \leq E[X]/t$.*

*Proof.* Since $X$ is nonnegative,

$$E[X] = \int_0^\infty xf(x)dx = \int_0^t xf(x)dx + \int_t^\infty xf(x)dx$$
$$\geq \int_t^\infty xf(x)dx \geq t\int_t^\infty f(x)dx = tP\{X \geq t\}$$

where $f$ denotes the pdf of $X$. A similar argument works for other types of random variables (not-continuous) as well. □

**Theorem 2.** *(Chebyshev inequality) For any random variable $X$ with mean $\mu$ and finite variance and for any $t > 0$, we have $P\{|X - \mu| \geq t\} \leq Var(X)/t^2$.*

*Proof.* Note that $|X - \mu| \geq t$ if and only if $|X - \mu|^2 \geq t^2$. Thus, $P\{|X - \mu| \geq t\} = P\{|X - \mu|^2 \geq t^2\}$. Since $|X - \mu|^2$ is a nonnegative random variable, $P\{|X - \mu|^2 \geq t^2\} \leq E[|X - \mu|^2]/t^2 = Var(X)/t^2$ by Markov inequality. □

**Theorem 3.** *(Hölder's inequality) If $p > 1$ and $1/p + 1/q = 1$, and if $E|X|^p < \infty$ and $E|Y|^q < \infty$, then $E|XY| \leq (E|X|^p)^{1/p}(E|Y|^q)^{1/q}$.*

## 2 Convergence in probability and Law of Large Numbers

**Definition 4.** Let $X_1, ..., X_n, ...$ be a sequence of random variables. We say that $\{X_n\}_{n=1}^\infty$ *converges* to $X$ *in probability* if for any $\varepsilon > 0$ $P\{\,|X_n - X| > \varepsilon\} \to 0$ as $n \to \infty$. In this case we write $X_n \to^p X$.

**Theorem 5.** *If $E(X_n - X)^2 \to 0$, then $X_n \to_p X$.*

*Proof.* By Markov inequality, for any $\varepsilon > 0$

$$P\{|X_n - X| > \varepsilon\} = P\{|X_n - X|^2 > \varepsilon^2\} \leq E[|X_n - X|^2]/\varepsilon^2 \to 0$$

.

∎

**Theorem 6.** *If $\{X_i\}_{i=1}^{\infty}$ is a sequence of independent and identically distributed (i.i.d.) random variables with $E[X_i] = \mu$ and $Var(X_i) = \sigma^2 < \infty$, then $\overline{X}_n := \sum_{i=1}^n X_i/n \to^p \mu$.*

*Proof.* By linearity of expectation, $E[\overline{X}_n] = E[\sum_{i=1}^n X_i/n] = \sum_{i=1}^n E[X_i]/n = \mu$. Thus,

$$E[|\overline{X}_n - \mu|^2] = V(\overline{X}) = V(\sum_{i=1}^n X_i)/n^2 = \sum_{i=1}^n V(X_i)/n^2 = \sigma^2/n.$$

Thus, $E|\overline{X}_n - \mu|^2 \to 0$ as $n \to \infty$. ∎

Theorem 6 uses very strong i.i.d. assumption, in Econometrics we often consider cases when it is not satisfied. Limit theorems for dependent random variables are discussed in 14.384 Time Series. It is easy to get an extension for independent but non-identically distributed random variables. Assume that $\{X_i\}_{i=1}^{\infty}$ are independent random variables with $E[X_i] = \mu$ but $Var(X_i) = \sigma_i^2$. Show that LLN holds if $\frac{1}{n^2}\sum_{i=1}^n \sigma_i^2 \to 0$.

Another version of the law of large numbers is

**Theorem 7.** *If $\{X_n\}_{n=1}^{\infty}$ is a sequence of iid random variables with $EX_n = \mu$ and $E|X_n| < \infty$, then $\overline{X}_n \to_p \mu$*

# 3 Weak convergence and Central Limit Theorem

**Definition 8.** We say that $\{X_n\}_{n=1}^{\infty}$ *converges* to $X$ *in distribution* or *weakly* if $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ for all $x \in \mathbb{R}$ where $F_X(x)$ is continuous. In this case we write $X_n \Rightarrow X$.

**Theorem 9.** *If $X_n \to_p X$, then $X_n \Rightarrow X$.*

*Proof.* Note that $X_n \leq x$ and $X > x + \varepsilon$ implies $|X_n - X| > \varepsilon$. Thus,

$$\begin{aligned} F_{X_n}(x) &= P\{X_n \leq x\} \\ &= P\{X_n \leq x,\, X \leq x+\varepsilon\} + P\{X_n \leq x,\, X > x+\varepsilon\} \\ &\leq P\{X \leq x+\varepsilon\} + P\{|X_n - X| > \varepsilon\} \\ &= F_X(x+\varepsilon) + P\{|X_n - X| > \varepsilon\}. \end{aligned}$$

for any $x \in \mathbb{R}$ and $\varepsilon > 0$. Similarly,

$$F_X(x-\varepsilon) \leq F_{X_n}(x) + P\{|X_n - X| > \varepsilon\}.$$

Thus,

$$F_X(x-\varepsilon) - P\{|X_n - X| > \varepsilon\} \leq F_{X_n}(x) \leq F_X(x+\varepsilon) + P\{|X_n - X| > \varepsilon\}.$$

Next, if $x$ is a point of continuity of $F_X$, for any $\delta > 0$, there exists $\varepsilon(\delta) > 0$ such that

$$F_X(x + \varepsilon(\delta)) - \delta \le F_X(x) \le F_X(x - \varepsilon(\delta)) + \delta.$$

Therefore

$$F_X(x) - \delta - P\{|X_n - X| > \varepsilon(\delta)\} \le F_{X_n}(x) \le F_X(x) + \delta + P\{|X_n - X| > \varepsilon(\delta)\}.$$

Next, since $X_n \to_p X$,by definition:

$$\lim_n |F_{X_n}(x) - F_X(x)| \le \delta$$

So, $F_{X_n}(x) \to F_X(x)$ as $n \to \infty$ for any $x \in \mathbb{R}$ where $F_X(x)$ is continuous. □

As an exercise, prove that if $c$ is some constant and $X_n \Rightarrow c$, then $X_n \to_p c$.

**Theorem 10.** *( Central limit theorem) Let $\{X_i\}$ be a sequence of i.i.d. random variables with mean $\mu$ and variance $\sigma^2$. Then $\sum_{i=1}^n (X_i - \mu)/\sqrt{n} \Rightarrow N(0, \sigma^2)$.*

In the multivariate case, if $Var(X_i) = E[(X_i - E[X_i])(X_i - E[X_i])^T] = \Sigma$, then $\sum_{i=1}^n (X_i - \mu)/\sqrt{n} \Rightarrow N(0, \Sigma)$.

We often will need to consider non-identically distributed random variables, in such a case we should use Linderberg- Feller's CLT:

**Theorem 11.** *Let $\{X_i\}_{i=1}^\infty$ be a sequence of independent random variables with $EX_i = \mu_i$ and $V(X_i) = \sigma_i^2$. Denote $c_n^2 = V(\sum_{i=1}^n X_i) = \sum_{i=1}^n \sigma_i^2$. If for any $\varepsilon > 0$*

$$\lim_{n\to\infty} \frac{1}{c_n^2} \sum_{i=1}^n E\left((X_i - \mu_i)^2 \mathbb{I}\{|X_i - \mu_i| > \varepsilon c_n\}\right) = 0 \tag{1}$$

*then $\frac{\sum_{i=1}^n (X_i - \mu_i)}{c_n} \Rightarrow N(0, 1)$.*

Sometimes Linderberg's condition (1) is called "asymptotic negligibility" as in particular it implies that $\max_{1\le i\le n} \frac{\sigma_i^2}{c_n^2} \to 0$ and guarantees that a personal contribution of any $X_i$ to the variance of the sum is sufficiently small for large $n$. The following sufficient condition for (1) is called Lyapunov's:

$$\lim_{n\to\infty} \frac{1}{c_n^{2+\delta}} \sum_{i=1}^n E|X_i - \mu_i|^{2+\delta} = 0 \text{ for some } \delta > 0.$$

# 4 Asymptotic statements derived from basic limit theorems

**Theorem 12.** *(Slutsky theorem and Continuous mapping theorem) Let $X, X_1, ..., X_n, ...$ and $Y, Y_1, ..., Y_n, ...$ be some random variables. Let $g$ be some continuous function. Let $c$ be some constant. Then*

1. *If $X_n \to_p X$ and $Y_n \to_p Y$, then $X_n + Y_n \to_p X + Y$ and $X_n Y_n \to_p XY$.*

2. *If $X_n \Rightarrow X$ and $Y \to_p c$, then $X_n + Y_n \Rightarrow X + c$ and $X_n Y_n \Rightarrow cX$.*

*3. If $X_n \to_p X$, then $g(X_n) \to_p g(X)$.*

*4. If $X_n \Rightarrow X$, then $g(X_n) \Rightarrow g(X)$*

The first and second statements are known as the Slutsky theorem. The third and forth statements are known as the Continuous mapping theorem. These theorems are widely used in statistics.

## 4.1 Symbols $o_p$ and $O_p$

First, let us talk about some notation for **non-stochastic** sequences $x_n$ and $b_n$. We would write $x_n = o(b_n)$ when $\lim_{n\to\infty} \frac{x_n}{b_n} = 0$. Often it is described as "$x_n$ is asymptotically smaller than $b_n$". We would write $x_n = O(b_n)$ when $\sup_n \frac{x_n}{b_n} < \infty$. In general we will use some easy-to-describe sequences as $b_n$, such as $\sqrt{n}$, $\frac{1}{n}$ or 1. Notice that $x_n = o(1)$ means $x_n \to 0$ and $x_n = O(1)$ means that $x_n$ is a bounded sequence. Now let us adopt similar notations for sequences of random variables $\{X_n\}_{n=1}^{\infty}$:

**Definition 13.** We say that $X_n = o_p(b_n)$ iff $\frac{X_n}{b_n} \to_p 0$.

We say that $X_n = O_p(b_n)$ iff for any $\varepsilon > 0$ there exists constant $C < \infty$ such that $P\left\{ \frac{X_n}{b_n} > C \right\} < \varepsilon$ for all $n$.

Sometimes statement $X_n = O_p(b_n)$ is described as "sequence $\frac{X_n}{b_n}$ is stochastically bounded." Some example of the use of these symbols are given below:

- If $X_n \Rightarrow N(0,1)$ as $n \to \infty$ then $X_n = O_p(1)$;
- If $X_n \to_p X$ then $X_n = X + o_p(1)$;
- Chebyshev's inequality for i.i.d. $X_i$ with finite variance implies $\overline{X}_n = O_p(\frac{1}{\sqrt{n}})$

We will often use the following statements :

- If $X_n = O_p(n^{-\delta})$ for some $\delta > 0$ then $X_n = o_p(1)$;
- If $X_n = o_p(b_n)$ then $X_n = O_p(b_n)$;
- If $X_n = O_p(n^{\alpha})$ and $Y_n = O_p(n^{\beta})$, then $X_n Y_n = O_p(n^{\alpha+\beta})$ and $X_n + Y_n = O_p(\max\{n^{\alpha}, n^{\beta}\})$;
- If $X_n = O_p(n^{\alpha})$ and $Y_n = o_p(n^{\beta})$, then $X_n Y_n = o_p(n^{\alpha+\beta})$;
- If $X_n = O_p(n^{\alpha})$ and $Y_n = o_p(n^{\alpha})$, then $X_n + Y_n = O_p(n^{\alpha})$

# 5 Delta method

**Theorem 14.** *Assume that for a sequence of random variables $X_n$ and constants $\mu$ and $\sigma$ we have $\sqrt{n}(X_n - \mu) \Rightarrow N(0, \sigma^2)$. If $g'(\mu) \neq 0$, then $\sqrt{n}(g(X_n) - g(\mu)) \Rightarrow N(0, \sigma^2(g'(\mu))^2)$.*

*Proof.* By the mean value theorem, for any realization $X_n(\omega)$, there is some $\mu_n^\star(\omega)$ between $\mu$ and $X_n(\omega)$ such that

$$g(\overline{X}_n(\omega)) - g(\mu) = g'(\mu^*)(\overline{X}_n(\omega) - \mu). \quad (2)$$

Thus, we have defined a new sequence of random variables, $\{\mu_n^\star\}_{n=1}^\infty$. By assumptions we have $\sqrt{n}(X_n - \mu) = O_p(1)$, thus $(X_n - \mu) = o_p(1)$ and $X_n \to^p \mu$. Since $\mu_n^\star$ is between $\mu$ and $X_n$, $\mu_n^\star \to_p \mu$ as well. By the Continuous mapping theorem, $g'(\mu_n^\star) \to_p g'(\mu)$ since $g'(x)$ is continuous. Moreover, by the Slutsky theorem and by the Central limit theorem, $\sqrt{n}g'(\mu^*)(X_n(\omega) - \mu) \Rightarrow g'(\mu)N(0, \sigma^2)$. □

Note that this theorem also holds when $g'(\mu) = 0$ but in this case the asymptotic distribution will be 0 (constant), i.e. degenerate. I recommend that you remember the argument used in this theorem as it is very typical in statistics and econometrics.

The Delta method has a multidimensional extension. Let $X_1, ..., X_n, ...$ be a sequence of iid $k \times 1$ random vectors with mean $\mu$ and covariance matrix $\Sigma$. Then, by the multidimensional Central limit theorem, $\sqrt{n}(\overline{X}_n - \mu) \Rightarrow N(0, \Sigma)$. Let $g : \mathbb{R}^k \to \mathbb{R}$ be a twice continuously differentiable function. Let $\tau^2 = (\partial g(\mu)/\partial\mu)^T \Sigma (\partial g(\mu)/\partial\mu)$. Here $\partial g(\mu)/\partial\mu$ is a $k \times 1$ vector with $i$-th component equals $\partial g(\mu)/\partial\mu_i$. Then $\sqrt{n}(g(\overline{X}_n) - g(\mu)) \Rightarrow N(0, \tau^2)$.

## 5.1 Example

Let $X_1, ..., X_n, ...$ be a sequence of iid random variables with mean $\mu$ and variance $\sigma^2$. What is the limiting distribution of $(\overline{X}_n)^2$? Let $g(x) = x^2$. Then $g'(\mu) = 2\mu$. Thus, by the Delta method, $\sqrt{n}((\overline{X}_n)^2 - \mu^2) \Rightarrow N(0, 4\mu^2\sigma^2)$. Note that if $\mu = 0$, then the limit distribution is degenerate.

14.381 Statistical Method in Economics
Fall 2018