Lecture 12 / Week 7

Convergence in Probability

The first part of the lecture on the "Extension of the WLLN to Stationary Time Series" can be found in the handout given by Prof. Fortini.

Sofar we defined the theorems on convergence in probability for random variables, but these theorems can be extended to random vectors. Then

$$X, X_n : \text{random vectors}$$

convergence in probability can be expressed now

$$P(||X_n - X|| > \varepsilon) \to 0 \quad n \to \infty$$

where we used Euclidean norm(||x||=$\sqrt{x^T.x} = \sqrt{\Sigma x_i^2}$) instead of the absolute value. Slutsky Theorem and Law of Large Numbers carry over to random vectors.

Convergence in Distribution

We will use a completely different notation than we used before in case of convergence in probability or a.s.

Suppose X_n are random vectors and X is a random vector. We define

$$\begin{aligned} F_n &: = \text{the distribution function of } X_n \\ F &: = \text{the distribution function of } X \\ F(x) &= P(X \leq x) \end{aligned}$$

Definition X_n **converges in distribution** to X if $F_n(x) \to F(x)$ $n \to \infty$ for every x where F is continuous. Note that the distribution of X_n converges to the distribution of X, not X_n to X. ın other words, suppose X_n, X are random variables then

$$P(a < X_n \leq b) = F_n(b) - F_n(a) \to F(b) - F(a) = P(a < X \leq b)$$

if F is continuous in a and b.

$$\begin{aligned} P(a &< X_n \leq b) \to P(a < X \leq b) \\ P(X_n &\in A) \to P(X \in A) \end{aligned}$$

for most of the Borel sets A. (not every, because we have the condition that b is a continuity point, recall that the distribution function is right-continuous.)

If n is large enough

$$P(X_n \in A) \approx P(X \in A)$$

for most of the Borel sets.

Example Assume that X_n have exponential distribution(n).

$$\begin{aligned}
f_n(x) &= n.e^{-nx}.1_{[0,\infty)}(x) \\
F_n(x) &= \int_{-\infty}^{x} f_n(s)ds = (1-e^{-nx}).1_{[0,\infty)}(x) \\
X_n &\rightarrow {}^{d}0, \quad n \rightarrow \infty \\
n &\rightarrow \infty \quad \lim_{n\rightarrow\infty}(1-e^{-nx}).1_{[0,\infty)}(x) = 1_{(0,\infty)}(x)
\end{aligned}$$

Insert here Figure 1

Note that the limit distribution is a constant function $(1_{[0,\infty)}(x))$ which is not continuous at x=0, but we are not concerned.

Insert here Figure 2

$$F_n(x)_{n\rightarrow\infty} \rightarrow 1_{[0,\infty)}(x) \quad \text{where } F \text{ is continuous}$$

We will use the following notation

$$\begin{aligned}
X_n &\rightarrow {}^{d}X \\
X_n &\rightarrow {}^{d}X \sim N(0,1) \\
&\quad\ \text{or directly} \\
X_n &\rightarrow {}^{d}N(0,1)
\end{aligned}$$

Note that

$$n \rightarrow \infty \quad \lim_{n\rightarrow\infty}(1-e^{-nx}).1_{[0,\infty)}(x) = 1_{(0,\infty)}(x)$$

is not a distribution function since it is left-continous but not right-continous, i.e. $F(0^{-})$=$F(0)$, that's why we define as

$$F_n(x)_{n\rightarrow\infty} \rightarrow 1_{[0,\infty)}(x) \quad \text{where } F \text{ is continuous}$$

Insert here Figure 3

The Relation between Convergence in Distribution and Probability

$$\begin{aligned} \text{Convergence in Probability} &\Rightarrow \text{Convergence in Distribution} \\ X_n &\rightarrow {}^P X \Rightarrow X_n \rightarrow^d X \end{aligned}$$

The converse is not true in general. It is true if X is *constant*. Therefore

$$X_n \rightarrow^d c \Rightarrow X_n \rightarrow^P c$$

In other words when the limit is constant

$$\rightarrow^p = \rightarrow^d$$

Continuous Mapping Theorem If $X_n \rightarrow^d X$ and ϕ is a *continuous* function, then

$$\phi(X_n) \rightarrow^d \phi(X)$$

Example $X_n \rightarrow^d N(0,1)$

$$X_n^2 = \phi(X_n) \rightarrow^d \phi(X) \text{ with } X \sim N(0,1)$$

since $\phi(x) = x^2$ is a continuous function, the assumption of CMT holds.

$$X_n^2 \rightarrow X_1^2$$

Example Let $X_n \rightarrow^d N_k(0, \mathbf{I}_k)$ (i.e. $X \sim (0, \mathbf{I}_k)$) and M an idempotent,symmetric matrix. Consider the following transformation.

$$X_n^T.M.X_n = \phi(X_n)$$

note that ϕ is a continuous function.

$$X_n^T.M.X_n = \phi(X_n) \rightarrow^d \phi(X) = X^T.M.X \sim \mathcal{X}_r^2$$

where r=rank(M), this is an example of an asymtotic result often used in statistical inference.

Suppose we have

$$\begin{aligned} X_n &\rightarrow {}^d X \\ Y_n &\rightarrow {}^d Y \end{aligned}$$

then we have a continuous function $\phi(x, y)$. Is it true that

$$\phi(X_n, Y_n) \rightarrow^d \phi(X, Y)$$

the answer is no in general. But, the following theorem tells us under which condition it holds.

Theorem Let

$$\begin{aligned} X_n &\rightarrow {}^d X \\ Y_n &\rightarrow {}^d c \end{aligned}$$

and let $\phi(x, y)$ be a continuous function. Then $\phi(X_n, Y_n) \rightarrow^d \phi(X, c)$.

Example Let $T_n \sim Student - t(n)$. Then

$$T_n \rightarrow^d N(0,1)$$

this result is a consequence of the previous theorem. (One can check that looking at the tables at the end of any statistics book, for high n the distributions are very similar.) Consider

$$\begin{aligned} T_n &= \frac{X_n}{\sqrt{\frac{Y_n}{n}}} \\ X_n &\sim N(0,1) \\ Y_n &\sim \mathcal{X}^2(n) \end{aligned}$$

Exercise Let $X_n \rightarrow^d N(0,1)$ and show that $\frac{Y_n}{n} \rightarrow^d 1$. Hint: use the fact that $Y_n = Z_1^2 + Z_2^2 ...$, together with the law of large numbers.

So, we can express it i.t.o the previous theorem, i.e

$$\phi(x, y) = \frac{x}{\sqrt{y}}$$

note that ϕ is a continuous function. Then the theorem says

$$T_n = \phi(X_n, \frac{Y_n}{n}) \rightarrow^d \phi(X, 1) = \frac{X}{\sqrt{1}} = X \sim N(0,1)$$

note that we used the result in the previous exercise.

Central Limit Theorem

Theorem (Lévy) Let X_n be a sequence of independent and identically distributed random variables with $E(X_n) = \mu$ and $var(X_n) = \sigma^2 < \infty$. Then

$$\sqrt{n}(\bar{X}_n - \mu) \rightarrow^d N(0, \sigma^2)$$

This is one of the most important, hence mostly applied, theorems in statistics.

We woudl like to generalize this result in case of an continuous function ϕ. Does $\sqrt{n}(\phi(\bar{X}_n) - \phi(\mu))$ convergence in distribution?

Suppose ϕ is continuously differentiable function, s.t. $\phi : \mathbb{R} \to \mathbb{R}$. By properties of derivative

$$\sqrt{n}(\phi(\bar{X}_n) - \phi(\mu)) = \sqrt{n}(\phi^{'}(\mu + \lambda(\bar{X}_n - \mu))(\bar{X}_n - \mu))$$

where we used *Taylor expansion*

$$\phi(x) = \phi(x_0) + \phi^{'}(x_0 + \lambda(x - x_0)).(x - x_0)$$

Then

$$\begin{array}{rcl} \sqrt{n}(\phi(\bar{X}_n) - \phi(\mu)) & = & (\phi^{'}(\mu + \lambda(\bar{X}_n - \mu)).\sqrt{n}(\bar{X}_n - \mu)) \\ n & \to & \infty \ \ \textbf{CLT}{:}\sqrt{n}(\bar{X}_n - \mu) \to^d N(0, \sigma^2) \\ n & \to & \infty \ \ \textbf{WLLN} : \bar{X}_n \to^P \mu \\ \sqrt{n}(\phi(\bar{X}_n) - \phi(\mu)) & = & (\phi^{'}(\mu + \lambda(\bar{X}_n - \mu)).\sqrt{n}(\bar{X}_n - \mu)) \to^d \phi^{'}(\mu).Z \sim N(0, \sigma^2.\phi^{'}(\mu)^2) \end{array}$$

where $Z \sim N(0, \sigma^2)$.

This result on mean can be generalized also for other moments, using CLT+**Delta Method** as long as $\phi()$ is a continuous function.

Theorem Let X_n be a sequence of i.i.d random variables of size k with $E(X_n) = \mu$, $var(X_n) = \sum$. (i.e. the second moment is finite.) Then

$$\sqrt{n}(\bar{X}_n - \mu) \to^d N_k(0, \sum)$$

Suppose we have $\phi : \mathbb{R}^k \to \mathbb{R}^m$ continuously diffentiable. Then

$$\sqrt{n}(\phi(\bar{X}_n) - \phi(\mu)) \to^d N_m(0, \triangle\phi(\mu) \sum \triangle\phi(x)^T)$$

where $\triangle\phi(\mu)$ is computed as follows

$$\triangle\phi(x) = \begin{bmatrix} \frac{\partial\phi_1}{\partial x_1} & . & . & \frac{\partial\phi_1}{\partial x_k} \\ . & & & . \\ . & & & . \\ \frac{\partial\phi_m}{\partial x_1} & . & . & \frac{\partial\phi_m}{\partial x_k} \end{bmatrix}$$