Convergence concepts

Connections between convergence in distribution & probability

$$X_n \xrightarrow{P} X \implies X_n \implies X \quad (X_n \xrightarrow{d} X)$$

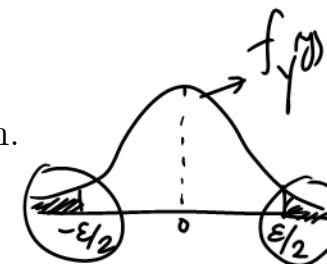- Convergence in probability *always* implies convergence in distribution.

Example

- BUT, the reverse is **not true**: take $Y \sim N(0,1)$ and $X_n = (-1)^n Y$.

$$\forall n, \ X_n = (-1)^n Y \implies X_n \sim N(0,1) \implies F_{X_n}(y) = F_Y(y) \quad \forall y \in \mathbb{R}$$

$X_n = (-1)^n$

$$F_{X_n}(y) \xrightarrow{d} F_Y(y)$$

$$\mathbb{P}(|X_n - Y| \geq \varepsilon) = \mathbb{P}(|-Y - Y| \geq \varepsilon) = \mathbb{P}(|Y| \geq \varepsilon/2) = \Phi(-\varepsilon/2) + [1 - \mathbb{P}(Y \leq \varepsilon/2)]$$

$X_n \xrightarrow{d} Y \not\Rightarrow X_n \xrightarrow{P} Y$

$$= \Phi(-\varepsilon/2) + 1 - \Phi(\varepsilon/2) \not\to 0$$

- However, convergence in probability and in distribution *to a* (constant) are equivalent. In fact, we've seen two examples of this already

  Constant

  1. $U_n \sim \text{Uniform}(0, 1/n)$, $U = 0$, then $U_n \xrightarrow{p} 0$ & $U_n \xrightarrow{d} 0$

  2. $\bar{X}_n$ from iid $N(\mu, \sigma^2)$ variables, then $\bar{X}_n \xrightarrow{p} \mu$ (WLLN) & $\bar{X}_n \xrightarrow{d} \mu$ (last example)

  Constant

Theorem

1. If $Y_n \xrightarrow{p} Y$ then $Y_n \xrightarrow{d} Y$.   (The reverse is not true; see above)

2. $Y_n \xrightarrow{p} c$ if and only if $Y_n \xrightarrow{d} c$.

Proof of 1. Pick/fix any $y \in \mathbb{R}$ at which $F_Y(\cdot)$ is continuous. Let $\epsilon > 0$. Then,

$A \subseteq B$
$\mathbb{P}(A) \leq \mathbb{P}(B)$

(I)
$$P(\underbrace{Y_n \leq y}_{A}) = P(Y_n \leq y, \underbrace{Y \leq y + \epsilon}_{B}) + P(Y_n \leq y, \underbrace{Y > y + \epsilon}_{B^C})$$
$$\leq P(Y \leq y + \epsilon) + P(|Y_n - Y| > \epsilon) \quad \text{this implies } |Y_n - Y| > \varepsilon$$

(II)
$$P(\underbrace{Y \leq y - \epsilon}) = P(Y_n \leq y, Y \leq y - \epsilon) + P(Y_n > y, Y \leq y - \epsilon)$$
$$\leq P(\underbrace{Y_n \leq y}) + P(\underbrace{|Y_n - Y| > \epsilon}) \quad \text{implies } |Y_n - Y| > \varepsilon$$

$$\mathbb{P}(Y \leq y - \varepsilon) - \mathbb{P}(|Y_n - Y| > \varepsilon) \leq \mathbb{P}(Y_n \leq y) \leq \mathbb{P}(Y \leq y + \varepsilon) + \mathbb{P}(|Y_n - Y| > \varepsilon)$$

Since $Y_n \xrightarrow{p} Y$ (II) (I) Since $Y_n \xrightarrow{p} Y$

as $n \to \infty$, $\mathbb{P}(Y \leq y - \varepsilon) \leq \mathbb{P}(Y_n \leq y) \leq \mathbb{P}(Y \leq y + \varepsilon)$ for $\varepsilon > 0$.

Since $\varepsilon > 0$ was arbitrary $\underset{219}{\implies} F_{Y_n}(y) \xrightarrow{d} F_Y(y)$

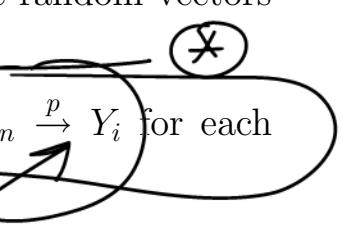# Convergence concepts

Tools of convergence in probability

1. One can extend the definition of convergence in probability to vectors:

   $\boldsymbol{Y}_n = (Y_{1,n}, \ldots, Y_{k,n})'$ **converges in probability** to $\boldsymbol{Y} = (Y_1, \ldots, Y_k)'$, denoted as $\boldsymbol{Y}_n \xrightarrow{p} \boldsymbol{Y}$ as $n \to \infty$, if

   $$\text{for any } \epsilon > 0, \quad \lim_{n \to \infty} P(\|\boldsymbol{Y}_n - \boldsymbol{Y}\| \geq \epsilon) = 0.$$

   Whenever you have vectors of random variables though, the random vectors converge in probability iff the individual components do so:

   Result: $\boldsymbol{Y}_n = (Y_{1,n}, \ldots, Y_{k,n})' \xrightarrow{p} \boldsymbol{Y} = (Y_1, \ldots, Y_k)'$ iff $Y_{i,n} \xrightarrow{p} Y_i$ for each $i = 1, \ldots, k$.

2. **A Stronger Version of Weak Law of Large Numbers (WLLN):** If $X_1, X_2, \ldots$, are iid with mean $\mu = \mathrm{E}X_1$, then

   $$\bar{X}_n \xrightarrow{p} \mu \quad \text{as } n \to \infty.$$

   (We saw one version of the WLLN already but that statement required $\mathrm{Var}(X_1) < \infty$. Actually, one does *not* need a finite variance for the WLLN to hold, only the existence of the mean μ as above (i.e., $\mathrm{E}|X_1| < \infty$) However, this is not easy to show.)

   Example: Let $X_1, X_2, \ldots$ be iid $N(0,1)$ r.v.'s. Then, by WLLN,

   $\mathbb{E}X_1 = 0, \quad \mathbb{E}X_1^3 = 0, \quad \mathbb{E}X_1^5 = 0$

   $M_X(t) = e^{t^2/2}$

   $\left.\dfrac{d^3}{dt^3}\right|_{t=0} = \left.\dfrac{d^5}{dt^5}\right|_{t=0} = 0$

   we $\begin{cases} \dfrac{1}{n}\sum_{i=1}^{n} X_i \longrightarrow 0 \\ \dfrac{1}{n}\sum_{i=1}^{n} X_i^3 \longrightarrow 0 \\ \dfrac{1}{n}\sum X_i^5 \longrightarrow 0 \end{cases}$

   strong version of WLLN

   $\Longrightarrow \begin{pmatrix} \frac{1}{n}\sum X_i \\ \frac{1}{n}\sum X_i^3 \\ \frac{1}{n}\sum X_i^5 \end{pmatrix} \xrightarrow{P} \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$

3. **Theorem (continuous mapping theorem):** Suppose $\boldsymbol{Y}_n$ are $\mathbb{R}^k$-valued random vectors such that $\boldsymbol{Y}_n \xrightarrow{p} \boldsymbol{c} \in \mathbb{R}^k$. Suppose also that $g : \mathbb{R}^k \to \mathbb{R}$ is continuous at $\boldsymbol{c}$. Then,

$$g(\boldsymbol{Y}_n) \xrightarrow{p} g(\boldsymbol{c}) \quad \text{as } n \to \infty.$$

g is Continuous at c

Proof: Pick/fix $\epsilon > 0$. Then, by the definition of continuity at $\boldsymbol{c}$, there exists some given $\delta = \delta_\epsilon > 0$ such that

$$(\ast) \quad \|\boldsymbol{y} - \boldsymbol{c}\| < \delta \implies |g(\boldsymbol{y}) - g(\boldsymbol{c})| < \epsilon \quad \text{for } \boldsymbol{y} \in \mathbb{R}^k$$

$$\|Y_n - c\| = \|Y_n - Y + Y - c\| \leq \|Y_n - Y\| + \|Y - c\| \leq \|Y_n - y\| + \delta$$

Therefore, $\|\boldsymbol{Y}_n - \boldsymbol{c}\| < \delta \implies |g(\boldsymbol{Y}_n) - g(\boldsymbol{c})| < \epsilon$ so that

$$P(\|\boldsymbol{Y}_n - \boldsymbol{c}\| < \delta) \leq P(|g(\boldsymbol{Y}_n) - g(\boldsymbol{c})| < \epsilon)$$

$$A := \{\|Y_n - c\| < \delta\} \subseteq \{|g(Y_n) - g(c)| < \epsilon\} = B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$$

$(\ast\ast)$ Example: Let $X_1, X_2, \ldots$ be iid with mean μ and variance σ^2. Then, $S^2 \xrightarrow{p} \sigma^2$ as $n \to \infty$.

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

$$\frac{n-1}{n} S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum X_i^2 - (\bar{X}_n)^2 \quad \text{(algebra)}$$

Since $\mathbb{E}X_i^2 = \operatorname{Var}X_i + [\mathbb{E}X_i]^2 = \sigma^2 + \mu^2 \xRightarrow{WLLN} \frac{1}{n} \sum_{i=1}^n X_i^2 \longrightarrow \sigma^2 + \mu^2$

$$\begin{pmatrix} \frac{1}{n} \sum X_i^2 \\ \frac{1}{n} \sum X_i \end{pmatrix} \xrightarrow{P} \begin{pmatrix} \sigma^2 + \mu^2 \\ \mu \end{pmatrix},$$

$$\frac{n-1}{n} S^2 = g\left(\frac{1}{n} \sum X_i^2, \bar{X}_n\right)$$

$$g(x,y) = x - y^2$$

221

$$\Rightarrow \quad \frac{n-1}{n}S^2 \xrightarrow[\text{Continuous mapping}]{P} g(c) = g\left(\frac{\sigma^2+\mu^2}{\mu}\right) = \sigma^2 + \mu^2 - \mu^2 = \boxed{\sigma^2}$$

Convergence concepts

Tools of convergence in distribution

$$\Rightarrow \begin{pmatrix} \frac{n-1}{n}S^2 \\ \frac{n}{n-1} \end{pmatrix} \xrightarrow{P} \begin{pmatrix} \sigma^2 \\ 1 \end{pmatrix}$$

To summarize what is to follow:

$$\Rightarrow S^2 = f\left(\frac{n-1}{n}S^2, \frac{n}{n-1}\right)$$
$$f(x,y) = xy \rightarrow \text{Continuous function.}$$

1. Establishing convergence in distribution via moment generating functions

$$\Rightarrow S^2 \xrightarrow{P} \sigma^2$$
$$\downarrow$$
$$f(*,*) \rightarrow f(c)$$

2. Central limit theorem

3. Slutsky's Theorem & Delta Method