

Convergence of random variables

Let (Ω, \mathcal{F}, P) be a probability space and $\{X_n\}_{n \geq 1}$ a sequence of random variables. Our objective now is double:

- I Define different ways for "notions of convergence".
- II Study "universal" limits. (Namely, study when limits of different sequences are the same).

Modes of convergence

We start defining the different modes of convergence:

Def / We say that $\{X_n\}_{n \geq 1}$ tends to X almost surely (and we write $X_n \xrightarrow{a.s.} X$) if

$$A = \{\omega \in \Omega : X_n(\omega) \xrightarrow{n \rightarrow \infty} X(\omega)\} \in \mathcal{A}, \text{ and } P(A) = 1$$

Def / We say that $\{X_n\}_{n \geq 1}$ tends to X in r th mean ($r \geq 1$) (and we write $X_n \xrightarrow{r} X$) if

$$E(|X_n - X|^r) \xrightarrow{n \rightarrow \infty} 0$$

Def / We say that $\{X_n\}_{n \geq 1}$ tends to X in probability (and we write $X_n \xrightarrow{P} X$) if

$$P(|X_n - X| > \varepsilon) \rightarrow 0 \text{ for all } \varepsilon > 0$$

Def / We say that $\{X_n\}_{n \geq 1}$ tends to X in distribution ($X_n \xrightarrow{d} X$) if

$$F_{X_n}(x) \rightarrow F_X(x)$$

at all points where $F_X(x)$ is continuous.

The first step is to justify which convergences imply which ones. Indeed, we have the following relations:

$$\begin{array}{l}
 X_n \xrightarrow{a.s.} X \implies X_n \xrightarrow{P} X \implies X_n \xrightarrow{d} X \\
 X_n \xrightarrow{S} X \stackrel{\text{Chebyshev}}{\implies} X_n \xrightarrow{r} X \implies X_n \xrightarrow{P} X
 \end{array}$$

And not other implications are true in general!

Prop / $X_n \xrightarrow{P} X$, then $X_n \xrightarrow{d} X$.

Proof / Assume that $X_n \xrightarrow{P} X$, and write $F_n(x) = P(X_n \leq x)$, $F(x) = P(X \leq x)$. Then, if $\varepsilon > 0$,

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

and similarly, for $x - \varepsilon$:

$$F(x - \varepsilon) = P(X \leq x - \varepsilon) = P(X \leq x - \varepsilon, X_n \leq x) + P(X \leq x - \varepsilon, X_n > x) \leq F_n(x) + P(|X_n - X| > \varepsilon)$$

consequently, $F(x - \varepsilon) - P(|X_n - X| > \varepsilon) \leq F_n(x) \leq F(x + \varepsilon) + P(|X_n - X| > \varepsilon)$, and so, taking $n \rightarrow \infty$ we have that $F(x - \varepsilon) \leq \liminf F_n(x) \leq \limsup F_n(x) \leq F(x + \varepsilon)$. As we are checking only points x where $F(x)$ is continuous, then $F(x) \leq \liminf F_n(x) \leq \limsup F_n(x) \leq F(x)$, as we wanted to show.

Prop/ If $X_n \xrightarrow{a.s.} X$, then $X_n \xrightarrow{p} X$.

Proof/ Pick $\epsilon > 0$, then by Markov's inequality we have $p(|X_n - X| > \epsilon) \leq \frac{1}{\epsilon} E[|X_n - X|]$, and then the result holds immediately.

Prop/ If $s > r \geq 1$, $X_n \xrightarrow{s} X$, then $X_n \xrightarrow{r} X$.

Proof/ We have already seen this before: $E[|Z|^r]^{1/r} \leq E[|Z|^s]^{1/s}$.

Prop/ If $X_n \xrightarrow{a.s.} X$, then $X_n \xrightarrow{p} X$.

Proof/ We prove it with a series of lemmas.

Lemma ①: Let $\epsilon > 0$ and $A_n(\epsilon) = \{|X_n - X| > \epsilon\}$ and $B_m(\epsilon) = \bigcup_{n=m}^{\infty} A_n(\epsilon)$. Then $X_n \xrightarrow{a.s.} X$ iff $\forall \epsilon > 0$ $p(B_m(\epsilon)) \rightarrow 0$.
write $C = \{\omega \in \Omega : X_n(\omega) \rightarrow X(\omega)\}$ and let

$$A(\epsilon) = \{\omega \in \Omega : \omega \in A_n(\epsilon) \text{ for infinitely many values } n\}$$

Then $p(C) = 1$ iff $p(A(\epsilon)) = 0$ for all $\epsilon > 0$. Observe also that $\{B_m(\epsilon)\}_{m \geq 1}$ is a decreasing sequence, and $\bigcap_{m=1}^{\infty} B_m(\epsilon) = A(\epsilon)$, so $\lim_{m \rightarrow \infty} p(B_m(\epsilon)) = p(A(\epsilon))$, and hence $p(A(\epsilon)) = 0$ iff $\lim_{m \rightarrow \infty} p(B_m(\epsilon)) = 0$.

Lemma ②: $X_n \xrightarrow{a.s.} X$ iff $\sum_n p(A_n(\epsilon)) < +\infty \forall \epsilon > 0$: we have that $p(B_m(\epsilon)) \leq \sum_{n=m}^{\infty} p(A_n(\epsilon))$, and so, if $\sum_n p(A_n(\epsilon)) < +\infty$ it means that $p(B_m(\epsilon)) \rightarrow 0$.

We conclude with the final proof: as $A_n(\epsilon) \subseteq B_n(\epsilon)$. Hence $p(|X_n - X| > \epsilon) = p(A_n(\epsilon)) \leq p(B_n(\epsilon)) \rightarrow 0$, so the condition mentioned before holds.

In the problem sheet we will see counterexamples to all the other implications. However, if we assume extra conditions, we could have that implications \Leftarrow are also true.

Theorem/ a) If $X_n \xrightarrow{d} c$, where c is a constant, then $X_n \xrightarrow{p} c$.

b) If $X_n \xrightarrow{p} X$ and $p(|X_n| \leq k) = 1$ for all n and some k , then $X_n \xrightarrow{r} X$, $r \geq 1$.

Proof/ a) We write $p(|X_n - c| > \epsilon) = p(X_n < c - \epsilon) + p(X_n > c + \epsilon) \rightarrow 0$ if $X_n \xrightarrow{d} c$.

b) If $p(|X| \leq k) = 1$ and $X_n \xrightarrow{p} X$, then for all $\epsilon > 0$ $p(|X| > k + \epsilon) = p(|X - X_n + X_n| > k + \epsilon) \leq p(|X - X_n| > \epsilon) \rightarrow 0$, so $p(|X| \leq k + \epsilon) = 1$. Let

$A_n = \{|X_n - X| > \epsilon\}$. As $|X_n - X| \leq |X_n| + |X|$ we have $|X_n - X|^r \leq \epsilon^r \mathbb{I}_{A_n^c} + (2k)^r \mathbb{I}_{A_n}$ with probability 1 (because $|X_n|, |X| \leq k$ with prob 1).

Taking then expectations, $E[|X_n - X|^r] \leq \epsilon^r + (2k)^r p(A_n) \xrightarrow{m} \epsilon^r$, and so taking ϵ arbitrarily small we get $E[|X_n - X|^r] \rightarrow 0$.

Finally we see a result concerning the relation between convergence in distribution and convergence almost surely:

Theorem/ (Skorokhod's representation theorem). Assume that $X_n \xrightarrow{d} X$, with distribution functions $\{F_n\}_{n \geq 1}$ and F . Then we can define a sequence of random variables over $\Omega' = (0,1)$, $\mathcal{F} = \mathcal{B}((0,1))$, $p = \lambda$ such that

a) $\forall n \forall \omega \in \Omega'$ and Y have distribution functions $\{F_n\}$ and F $\left\{ \begin{array}{l} X_n(\omega) = \inf\{x : \omega \leq F_n(x)\} \\ Y(\omega) = \inf\{x : \omega \leq F(x)\} \end{array} \right.$

Proof/ We explicitly build the sequence $\{Y_n\}_{n \geq 1}$ and Y . For each $w \in (0,1)$, write

$$Y_n(w) = \inf \{x \in \mathbb{R} : w \leq F_n(x)\}$$

$$Y(w) = \inf \{x \in \mathbb{R} : w \leq F(x)\}$$

In particular, $w \leq F_n(x)$ iff $Y_n(w) \leq x$ and $w \leq F(x)$ iff $Y(w) \leq x$; essentially, Y_n and Y are the inverse functions of F_n and F , respectively. Let us check the conditions:

a) Observe that $\lambda(Y \leq y) = \lambda([0, F(y)]) = F(y)$, and also $\lambda(Y_n \leq y) = \lambda([0, F_n(y)]) = F_n(y)$, hence a) is obvious.

b) This takes more arguments. Take $\varepsilon > 0$, $w \in (0,1)$ and assume that x is a point of continuity of F such that $Y(w) - \varepsilon < x < Y(w)$. Then $F(x) < w$. As $F_n(x) \rightarrow F(x)$, then $F_n(x) < w$ for $n \geq n_0$, and hence, $Y(w) - \varepsilon < x < Y_n(w)$, $n \geq n_0$. Taking now $n \rightarrow \infty$ and $\varepsilon \downarrow 0$ we get that

$$\liminf Y_n(w) \geq Y(w) \quad \forall w$$

Then, assume $w < w' < 1$, and let x be a point of continuity of F such that $Y(w') < x < Y(w') + \varepsilon$. Then, $w < w' \leq F(x)$, and so $w < F_n(x)$, for n large enough, giving that $Y_n(w) \leq x < Y(w') + \varepsilon$ for n large enough. So taking $n \rightarrow \infty$, $\varepsilon \downarrow 0$ we get

$$\limsup Y_n(w) \leq Y(w') \quad w < w'$$

So now, for all w of continuity of Y , we have that $Y(w) \leq \liminf Y_n(w) \leq \limsup Y_n(w) \leq Y(w')$, and doing $w' \rightarrow w$ we get that $Y_n(w) \rightarrow Y(w)$. As Y is increasing, the set of discontinuities D is countable, hence $\lambda(D) = 0$, as we wanted to prove.

The main application of SRT is based in proving results of convergence in distribution:

Prop/ Assume $X_n \xrightarrow{d} X$, and $g: \mathbb{R} \rightarrow \mathbb{R}$ a continuous function. Then $g(X_n) \xrightarrow{d} g(X)$.

Proof/ Let $\{Y_n\}_{n \geq 1}$ and Y given by SRT. By the continuity of g we have that

$$\{w \in (0,1) : Y_n(w) \rightarrow Y(w)\} \subseteq \{w \in (0,1) : g(Y_n(w)) \rightarrow g(Y(w))\}$$

and so $g(Y_n) \xrightarrow{d} g(Y)$, and $g(Y_n) \xrightarrow{d} g(Y)$. However, the distribution of Y_n and X_n were the same, and hence $g(Y_n)$ and $g(X_n)$ have the same distribution function, which finishes the proof.