§ 1.3. Convergence modes.

Unlike convergence of a sequence of numbers, the convergence of a sequence of r.v.s at least has four commonly used modes: almost sure convergence, in probability convergence, L_p convergence and in distribution convergence. The first is sometimes called convergence almost everywhere or almost certain and the last convergence in law.

(i). Definitions

In what follows, we give definitions. Suppose $X_1, X_2, ...$ are a sequence of r.v.s.

$X_n \to X$ almost surely, (*a.s.*) if $P(\{\omega : X_n(\omega) \to X(\omega)\}) = P(X_n \to X) = 1$. Namely, *a.s.* convergence is a point-wise convergence "everywhere" except for a null set.

$X_n \to X$ in probability, if $P(|X_n - X| > \epsilon) \to 0$ for any $\epsilon > 0$.

$X_n \to X$ in L_p, if $E(|X_n - X|^p) \to 0$.

$X_n \to X$ in distribution. There are four equivalent definitions:

1). For every continuity point t of F, $F_n(t) \to F(t)$, where F_n and F are c.d.f of X_n and X.

2). For every closed set B, $\limsup_n P(X_n \in B) \leq P(X \in B)$.

3). For every open set B, $\liminf_n P(X_n \in B) \geq P(X \in B)$.

4). For every continuous bounded function $g(\cdot)$, $E(g(X_n)) \to E(g(X))$.

Remark. The L_p convergence preclude the limit X taking values of infinity with positive chances. Sometimes in some textbooks, a sequence of numbers going to infinity is called convergence to infinity rather than divergence to infinity. If this is the case, the limit X can be ∞ or $-\infty$, for a.s. convergence and, by slightly modifying the definition, for in probability convergence. For example, $X_n \to \infty$ in probability is naturally defined as, for any $M > 0$, $P(X_n > M) \to 1$. Convergence in distribution only has to do with distributions.

(ii). Convergence theorems.

The following three theorems/lemma, tantamount to their analogues in real analysis, play important role in the technical development of probability theory.

(1). *Monotone convergence theorem.* If $X_n \geq 0$, and $X_n \uparrow X$, then $E(X_n) \uparrow E(X)$.

Proof. $E(X_n) \leq E(X)$. For any $a < E(X)$, there exists a N and m such that $\sum_{i=0}^N \frac{i}{2^m} P\Big(\frac{i}{2^m} < X(w) \leq \frac{i+1}{2^m}\Big) > a$. But $P\Big(\frac{i}{2^m} < X_n(w) \leq \frac{i+1}{2^m}\Big) \to P\Big(\frac{i}{2^m} < X(w) \leq \frac{i+1}{2^m}\Big)$ (why?). Therefore, $\lim E(X_n) \geq a$. Hence, $E(X_n) \to E(X)$. □

(2). *Fatou's lemma.* If $X_n \geq 0$, a.s., then

$$E(\liminf X_n) \leq \liminf E(X_n)$$

Proof. Let $X_n^* = \inf(X_k : k \geq n)$, then $X_n^* \uparrow \liminf X_n$, so the Monotone convergence theorem, $E(X_n^*) \uparrow E(\liminf X_n)$. On the other hand, $X_n^* \leq X_n$ so, $E(X_n^*) \leq E(X_n)$. As a result, $E(\liminf X_n) \leq \liminf E(X_n)$. □

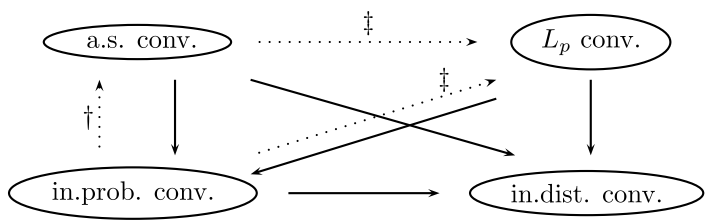
(3). *Dominated convergence theorem.* If $|X_n| \leq Y$, $E(Y) < \infty$, and $X_n \to X$ a.s., then $E(X_n) \to E(X)$.

Proof. Observe that $Y - X_n \geq 0 \leq Y + X_n$. By Fatou's lemma, $E(Y - \lim X_n) \leq \liminf E(Y - X_n)$, leading to $E(X) \geq \limsup E(X_n)$. Likewise $E(Y + \lim X_n) \leq \liminf E(Y + X_n)$, leading to $E(X) \leq \liminf E(X_n)$. Consequently, $E(X_n) \to E(X)$. □

The essence of the above convergence theorems is to use a bound, upper or lower, to ensure the desired convergence in expectation. These bounds, lower bounds as 0 in the monotone convergence theorem and the Fatou lemma, and both lower and upper bounds in the dominated convergence theorem, can actually be relaxed; see DIY exercises. The most general extension is through the concept of uniform integral r.v.s, which shall be introduced later if necessary.

(iii). Relations between convergence modes.

The relations are partly illustrated in the following diagram:



†: exist a subsequence that converges a.s. ‡: if $|X_n| \le Y$ where $Y \in L_p$.

(iv) Some examples.

We use following examples to clarify the above diagram.

a). in prob. conv. but not a.e. conv.

Let $\xi \sim Unif[0,1]$. Set $X_{2^j+k} = 1$ if $\xi \in [k/2^j, (k+1)/2^j]$ and 0 otherwise, for all $0 \le k \le 2^j - 1$ and $j = 0, 1, 2, ...$ Then, $X_n \to 0$ in probability as $n \to \infty$, but $X_n \nrightarrow 0$, a.e.. In fact, $P(X_n \to 0) = 0$.

Let ξ_n be i.i.d $\sim Unif[0,1]$. Let $X_n = 1$ if $\xi_n \le 1/n$ and 0 otherwise. Then $X_n \to 0$ in probability, but $X_n \nrightarrow 0$, a.e. by Borel-Contelli lemma.

b). in distribution conv. but not in probability conv..

This is in fact quite trivial. Any sequence of (non-constant) i.i.d. random variables converge in distribution, but not in probability. Observe that convergence in distribution only concerns the distribution. The variables even do not have to be in the same probability space.

c). a.s. but not L^p conv.

Let $\xi \sim Unif[0,1]$. Let $X_n = e^n$ if $\xi \le 1/n$ and 0 otherwise. Then $X_n \to 0$ a.s. but $E(|X_n|^p|) = e^{np}/n \to \infty$.

(v). Technical proofs.

①. a.s. convergence $\Longrightarrow$ in probability convergence.

Proof. Let $A_n = \{|X_n - X| > \epsilon\}$. a.s. convergence implies $P(A_n, i.o.) = 0$. But $\{A_n, i.o.\} = \cap_{n=1}^\infty \cup_{k=n}^\infty A_k$. So $0 = P(A_n, i.o.) = \lim_n P(\cup_{k=n}^\infty A_k) \ge \limsup_n P(A_n)$.

②. L^p convergence $\Longrightarrow$ in prob convergence.

Proof. $0 \leftarrow E(|X_n - X|^p) \ge E(|X_n - X|^p 1_{\{|X_n - X| > \epsilon\}}) \ge \epsilon^p P(|X_n - X| > \epsilon)$.

③. in prob convergence $\Longrightarrow$ in distribution convergence.

Proof. For any t, and any $\epsilon > 0$, $\limsup P(X_n \le t) \le \limsup P(\{X_n \le t\} \cap \{X \le X_n + \epsilon\}) \le P(X \le t + \epsilon)$. Let $\epsilon \downarrow 0$, we have $\limsup P(X_n \le t) \le P(X \le t)$. Likewise $\limsup P(-X_n \le -t) \le P(-X \le -t)$. (Why?) Then $\liminf P(X_n < t) \ge P(X < t)$. Suppose now, t is a continuity point of X. Then $P(X < t) = P(X \le t)$. As a result, $\lim_n P(X_n \le t) = P(X \le t)$.

④. in prob convergence $\Longrightarrow$ existence of a subsequence that converges a.s.

Proof. Let $\epsilon_k \downarrow 0$. Since $P(|X_n - X| > \epsilon_k) \to 0$ as $n \to \infty$, there exists an n_k such that $P(|X_{n_k} - X| > \epsilon_k) < 2^{-k}$. Therefore $\sum_{k=1}^{\infty} P(|X_{n_k} - X| > \epsilon_k) < \infty$, which implies by the Borel-Contelli lemma, which is introduced in the next section, that $P(|X_{n_k} - X| > \epsilon_k, i.o.) = 0$. This means that, with probability 1, $|X_{n_k} - X| \le \epsilon_k$ for all large k. This is tantamount to $X_{n_k} \to X$ a.s..

⑤. L^p convergence $\Longrightarrow$ L^q convergence for $p > q > 0$.

Proof. Let $Y_n = |X_n - X|$. For any $\epsilon > 0$, $E(Y_n^q) \le \epsilon + E(Y_n^q 1_{\{Y_n \ge \epsilon\}}) \le \epsilon + E(Y_n^q 1_{\{Y_n \ge 1\}} + P(\epsilon \le Y_n \le 1) \le \epsilon + E(Y_n^p 1_{\{Y_n \ge 1\}} + P(\epsilon \le Y_n) \to \epsilon$ as $n \to \infty$. Since $\epsilon > 0$ is arbitrary, it follows that $X_n \to X$ in L^q.

⑥. Suppose $|X_n| \le c > 0$ a.s., then, in probability convergence $\Longleftrightarrow$ L^p convergence for all (any) $p > 0$.

Proof. $\Longleftarrow$ follows from ②. And $\Longrightarrow$ follows from the dominated convergence theorem.

⑦. The four equivalent definitions of in distribution convergence.

Proof. 2) $\Longleftrightarrow$ 3). The complement of any closed set is open. Likewise, the complement of any closed set is open.

1) $\Longrightarrow$ 3). Continuity points of F are dense (why?). Consider interval $(-\infty, t)$, there exists continuity points $t_k \uparrow t$. Then,

$$\liminf_n P(X_n \in (-\infty, t)) \ge \liminf_n P(X_n \in (-\infty, t_k]) = P(X \in (-\infty, t_k]) \to P(X \in (-\infty, t)).$$

The result can be extended for general open sets. We omit the proof.

3) $\Longrightarrow$ 1). Suppose t is a continuity point. Then $\limsup_n F_n(t) \le F(t)$ by 2) and the equivalency of 2) and 3). $\liminf_n F_n(t) \ge \liminf_n P(X_n < t) \ge P(X < t) = F(t)$ as t is a continuity point. So 1) follows.

4) $\Longrightarrow$ 1). Let t be a continuity point of F. For any small $\epsilon > 0$, choose a non-increasing continuous function f of x which is 1 for $x < t$, and is 0 for $x > t + \epsilon$. Then, $P(X_n \le t) \le E(f(X_n)) \to E(f(X)) \le P(X \le t + \epsilon)$. Therefore the $\limsup P(X_n \le t) \le P(X \le t)$. Likewise (how?), one can show $\liminf P(X_n \le t) \ge P(X \le t)$. The desired convergence follows.

1) $\Longrightarrow$ 4). Continuity points of the cdf of X are dense (why?). Suppose $|f(t)| < c$. Choose continuity points $-\infty = t_0 < t_1, ... < t_K < t_{K+1} = \infty$ such that $F(t_1) < \epsilon > 1 - F(t_K)$, and $|f(t) - f(s)| < \epsilon$ for any $t, s \in [t_j, t_{j+1}]$ for $j = 1, ..., K-1$. Then,

$$\begin{aligned}
& |E(f(X_n)) - E(f(X))| = |\int f(t) dF_n(t) - \int f(t) dF(t)| \\
\le \quad & \sum_{j=0}^{K} |\int_{t_j}^{t_{j+1}} f(t)[dF_n(t) - dF(t)]| \\
\le \quad & 2c\epsilon + \sum_{j=1}^{K-1} |\int_{t_j}^{t_{j+1}} f(t)[dF_n(t) - dF(t)]| \\
\le \quad & 2c\epsilon + + \sum_{j=1}^{K-1} |\int_{t_j}^{t_{j+1}} f(t_j)[dF_n(t) - dF(t)]| \\
& + \sum_{j=1}^{K-1} |\int_{t_j}^{t_{j+1}} [f(t) - f(t_j)][dF_n(t) - dF(t)]|
\end{aligned}$$

$$\begin{aligned}
&\leq 2c\epsilon + \sum_{j=1}^{K-1} c\left|\int_{t_j}^{t_{j+1}} [dF_n(t) - dF(t)]\right| + \sum_{j=1}^{K-1} \epsilon \int_{t_j}^{t_{j+1}} [dF_n(t) + dF(t)] \\
&\to 2c\epsilon + 2\epsilon \int_{t_1}^{t_K} dF(t) \qquad \text{as } n \to \infty. \\
&\leq (2c+1)\epsilon,
\end{aligned}$$

which can be arbitrarily small.

DIY Exercises.

Exercise 1.9 ⋆⋆ Suppose $X_n \geq \eta$, with $E(\eta^-) < \infty$. Show $E(\liminf X_n) \leq \liminf E(X_n)$.

Exercise 1.10 ⋆⋆ Show the dominated convergence theorem still holds if $X_n \to X$ in probability or in distribution.

Exercise 1.11 ⋆⋆⋆ Let $S_n = \sum_{i=1}^n X_i$. Raise a counter-example to show $S_n/n \not\to 0$ in probability but $X_n \to 0$ in probability.

Exercise 1.12 ⋆⋆⋆ Let $S_n = \sum_{i=1}^n X_i$. Show that $S_n/n \to 0$ *a.s.* if $X_n \to 0$ *a.s.*, and $S_n/n \to 0$ in L_p if $X_n \to 0$ in L_p for $p \geq 1$.