Math 464/564, O. Knill, **April 1997**

Convergence of random variables

There are different definitions of convergence for random variables. Examples: The strong law of large numbers assures **almost everywhere convergence** of S_n/n to $\mathrm{E}[X]$. The weak law of large numbers assures **convergence in probability** of S_n/n to $\mathrm{E}[X]$. The central limit theorem states the **convergence in distribution** of S_n^* to a standard normal distributed random variable.

DEFINITIONS: (only convergence in probability and in distribution occur in the book).

A sequence of random variables X_n converges **almost everywhere** to a random variable X, if $\mathrm{P}[X_n \to X] = 1$.

A sequence of random variables X_n converges **in probability** to a random variable X, if $\mathrm{P}[|X_n - X| \geq \epsilon] \to 0$ for all $\epsilon > 0$.

A sequence of random variables X_n converges **in distribution** to a random variable X, if $\mathrm{P}[X_n \leq t] \to \mathrm{P}[X \leq t]$ for all $t \in \mathbb{R}$.

A sequence of random variables X_n **converges in $\mathcal{L}^1$** to a random variable X, if $\mathrm{E}[|X_n - X|] \to 0$.

A sequence of random variables X_n converges **completely** if $\sum_n \mathrm{P}[|X_n - X| \geq \epsilon] < \infty$ for all $\epsilon > 0$.

RELATIONS BETWEEN CONVERGENCE: (an arrow stands for "implies")

$$\begin{tikzcd}[column sep=small]
& \begin{array}{c}\text{In distribution}\\ \mathrm{P}[X_n \leq t] \to \mathrm{P}[X \leq t], \forall t \in \mathbb{R}.\end{array} & \\
& \begin{array}{c}\text{In probability}\\ \mathrm{P}[|X_n - X| \geq \epsilon] \to 0, \forall \epsilon > 0.\end{array} \arrow[u] & \\
\begin{array}{c}\text{Almost everywhere}\\ \mathrm{P}[X_n \to X] = 1\end{array} \arrow[ur] & & \begin{array}{c}\text{In } \mathcal{L}^1\\ \mathrm{E}[|X_n - X|] \to 0\end{array} \arrow[ul] \\
\begin{array}{c}\text{Complete}\\ \sum_{n=1}^{\infty} \mathrm{P}[|X_n - X| \geq \epsilon] < \infty, \forall \epsilon > 0\end{array} \arrow[u] & &
\end{tikzcd}$$

The convergence in distribution is the only type of convergence which does not assume that X_n, X are defined on the same probability space. The $\mathcal{L}^1$ convergence (and more generally $\mathcal{L}^p$ convergence) is used especially in more advanced topics of probability theory like stochastic processes or martingale theory. If the random variables are bounded $X_n \leq M$, the $\mathcal{L}^1$ convergence is equivalent to convergence in probability. The complete convergence can be useful to prove almost everywhere convergence. For example, the Chebychev-Markov inequality with the function $g(\epsilon) = \epsilon^4$ allows to prove that for a sequence of IID random variables with finite forth moment, S_n/n converges completely to $\mathrm{E}[X]$. This implies then the strong law of large numbers.

ALMOST EVERYWHERE $\Rightarrow$ IN PROBABILITY:

Since

$$\{X_n \to X\} = \bigcap_k \bigcup_m \bigcap_{n \geq m} \{|X_n - X| \leq 1/k\} ,$$

"almost everywhere convergence" is equivalent to

$$1 = \mathrm{P}[\bigcup_m \bigcap_{n \geq m} \{|X_n - X| \leq 1/k\}] = \lim_{n\to\infty} \mathrm{P}[\bigcap_{n \geq m} \{|X_n - X| \leq 1/k\}]$$

for all k. Therefore

$$\mathrm{P}[|X_m - X| \geq \epsilon] \leq \mathrm{P}[\bigcap_{n \geq m} \{|X_n - X| \geq \epsilon\}] \to 0$$

for all $\epsilon > 0$.

$\mathcal{L}^1$ CONVERGENCE $\Rightarrow$ IN PROBABILITY:

With Chebychev-Markov inequality for $g(\epsilon) = \epsilon^p$, one obtains

$$\mathrm{P}[|X_n - X| \geq \epsilon] \leq \mathrm{E}[|X_n - X|^p]/\epsilon^p .$$

COMPLETE $\Rightarrow$ ALMOST EVERYWHERE:

The Borel-Cantelli lemma (see seperate page) implies that for all $\epsilon > 0$

$$\mathrm{P}[|X_n - X| \geq \epsilon, \text{ infinitely often}] = 0 .$$

We get so for $\epsilon_n \to 0$

$$\mathrm{P}[\bigcup_n |X_n - X| \geq \epsilon_k, \text{ infinitely often}] \leq \sum_n \mathrm{P}[|X_n - X| \geq \epsilon_k, \text{ infinitely often}] = 0$$

from which we obtain $\mathrm{P}[X_n \to X] = 1$.

IN PROBABILITY $\Rightarrow$ IN DISTRIBUTION:

Let t be a continuity point of F_X. For $\epsilon > 0$ and $n \in \mathbb{N}$, one has

$$\begin{aligned} F_{X_n}(t) &= \mathrm{P}[X_n \leq t] \\ &= \mathrm{P}[X_n \leq t, |X_n - X| \leq \epsilon] + \mathrm{P}[X_n \leq t, |X_n - X| > \epsilon] \\ &\leq \mathrm{P}[X \leq t + \epsilon] + \mathrm{P}[|X_n - X| > \epsilon] \\ &= F_X(t+\epsilon) + \mathrm{P}[X_n - X] > \epsilon] \end{aligned}$$

which implies $\limsup_{n\to\infty} F_{X_n}(t) \leq F_X(t+\epsilon)$ and using the continuity of F at t also $\limsup_{n\to\infty} F_{X_n}(t) \leq F_X(t)$. Similarly, one gets

$$\begin{aligned} F_X(t-\epsilon) &= \mathrm{P}[X \leq t - \epsilon] \\ &= \mathrm{P}[X \leq t - \epsilon, |X_n - X| \leq \epsilon] + \mathrm{P}[X \leq t - \epsilon, |X_n - X| > \epsilon] \\ &\leq \mathrm{P}[X_n \leq t] + \mathrm{P}[|X_n - X| > \epsilon] \\ &= F_{X_n}(t) + \mathrm{P}[|X_n - X| > \epsilon] . \end{aligned}$$

For $n \to \infty$, one gets $F_X(t-\epsilon) \leq \liminf_{n\to\infty} F_{X_n}(t)$. Using the continuity of F at t one has $F_X(t) \leq \liminf_{n\to\infty} F_{X_n}(t)$.