EE5110: Probability Foundations for Electrical Engineers **July-November 2015**

Lecture 28: Convergence of Random Variables and Related Theorems

Lecturer: Dr. Krishna Jagannathan *Scribe: Gopal, Sudharsan, Ajay, Swamy, Kolla*

An important concept in Probability Theory is that of convergence of random variables. Since the important results in Probability Theory are the limit theorems that concern themselves with the asymptotic behaviour of random processes, studying the convergence of random variables becomes necessary. We begin by recalling some definitions pertaining to convergence of a sequence of real numbers.

Definition 28.1 *Let $\{x_n, n \geq 1\}$ be a real-valued sequence, i.e., a map from $\mathbb{N}$ to $\mathbb{R}$. We say that the sequence $\{x_n\}$ converges to some $x \in \mathbb{R}$ if there exists an $n_0 \in \mathbb{N}$ such that for all $\epsilon > 0$,*

$$|x_n - x| < \epsilon, \; \forall \; n \geq n_0.$$

We say that the sequence $\{x_n\}$ converges to $+\infty$ if for any $M > 0$, there exists an $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $x_n > M$.
We say that the sequence $\{x_n\}$ converges to $-\infty$ if for any $M > 0$, there exists an $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $x_n < -M$.

We now define various notions of convergence for a sequence of random variables. It would be helpful to recall that random variables are after all deterministic functions satisfying the measurability property. Hence, the simplest notion of convergence of a sequence of random variables is defined in a fashion similar to that for regular functions.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\{X_n\}_{n \in \mathbb{N}}$ be a sequence of real-valued random variables defined on this probability space.

Definition 28.2 [Definition 0 (Point-wise convergence or sure convergence)]
A sequence of random variables $\{X_n\}_{n \in \mathbb{N}}$ is said to converge point-wise or surely to X if

$$X_n(\omega) \to X(\omega), \quad \forall \; \omega \in \Omega.$$

Note that for a fixed ω, $\{X_n(\omega)\}_{n \in \mathbb{N}}$ is a sequence of real numbers. Hence, the convergence for this sequence is same as the one in definition (28.1). Also, since X is the point-wise limit of random variables, it can be proved that X is a random variable, *i.e.*, it is an $\mathcal{F}$-measurable function. This notion of convergence is exactly analogous to that defined for regular functions. Since this notion is too strict for most practical purposes, and neither does it consider the measurability of the random variables nor the probability measure $\mathbb{P}(\cdot)$, we define other notions incorporating the said characteristics.

Definition 28.3 [Definition 1 (Almost sure convergence or convergence with probability 1)]
A sequence of random variables $\{X_n\}_{n \in \mathbb{N}}$ is said to converge almost surely or with probability 1 (denoted by a.s. or w.p. 1) to X if

$$\mathbb{P}\left(\{\omega | X_n(\omega) \to X(\omega)\}\right) = 1.$$

Almost sure convergence demands that the set of ω's where the random variables converge have a probability one. In other words, this definition gives the random variables "freedom" not to converge on a set of zero measure! Hence, this is a weakened notion as compared to that of sure convergence, but a more useful one.

In several situations, the notion of almost sure convergence turns out to be rather strict as well. So several other notions of convergence are defined.

Definition 28.4 [Definition 2 (convergence in probability)]
A sequence of random variables $\{X_n\}_{n\in\mathbb{N}}$ is said to converge in probability (denoted by i.p.) to X if

$$\lim_{n\to\infty} \mathbb{P}\left(|X_n - X| > \epsilon\right) = 0, \quad \forall\, \epsilon > 0.$$

As seen from the above definition, this notion concerns itself with the convergence of a sequence of probabilities!

At the first glance, it may seem that the notions of almost sure convergence and convergence in probability are the same. But the two definitions actually tell very different stories! For almost sure convergence, we collect all the ω's wherein the convergence happens, and demand that the measure of this set of ω's be 1. But, in the case of convergence in probability, there is no direct notion of ω since we are looking at a sequence of probabilities converging. To clarify this, we do away with the short-hand for probabilities (for the moment) and obtain the following expression for the definition of convergence in probability:

$$\lim_{n\to\infty} \mathbb{P}\left(\{\omega | |X_n(\omega) - X(\omega)| > \epsilon\}\right) = 0, \quad \forall\, \epsilon > 0.$$

Since the notion of convergence of random variables is a very intricate one, it is worth spending some time pondering the same.

Definition 28.5 [Definition 3 (convergence in r^{th} mean)]
A sequence of random variables $\{X_n\}_{n\in\mathbb{N}}$ is said to converge in r^{th} mean to X if

$$\lim_{n\to\infty} \mathbb{E}\left[|X_n - X|^r\right] = 0.$$

In particular, when $r = 2$, the convergence is a widely used one. It goes by the special name of *convergence in the mean-squared sense.*

The last notion of convergence, known as convergence in distribution, is the weakest notion of convergence. In essence, we look at the distributions (of random variables in the sequence in consideration) converging to some distribution (when the limit exists). This notion is extremely important in order to understand the Central Limit Theorem (to be studied in a later lecture).

Definition 28.6 [Definition 4 (convergence in distribution or weak convergence)]
A sequence of random variables $\{X_n\}_{n\in\mathbb{N}}$ is said to converge in distribution to X if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x), \quad \forall\, x \in \mathbb{R} \text{ where } F_X(\cdot) \text{ is continuous.}$$

That is, the sequence of distributions must converge at all points of continuity of $F_X(\cdot)$. Unlike the previous four notions discussed above, for the case of convergence in distribution, the random variables need not be defined on a single probability space!

Before we look at an example that serves to clarify the above definitions, we summarize the notations for the above notions.

(1) *Point-wise Convergence:* $X_n \xrightarrow{\text{p.w.}} X$.

(2) *Almost sure Convergence:* $X_n \xrightarrow{\text{a.s.}} X$ or $X_n \xrightarrow{\text{w.p. } 1} X$.

(3) *Convergence in probability:* $X_n \xrightarrow{\text{i.p.}} X$.

(4) *Convergence in r^{th} mean:* $X_n \xrightarrow{r} X$. When $r = 2$, $X_n \xrightarrow{\text{m.s.}} X$.

(5) *Convergence in Distribution:* $X_n \xrightarrow{\text{D}} X$.

Example: Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), \lambda)$ and a sequence of random variables $\{X_n, n \geq 1\}$ defined by

$$X_n(\omega) = \begin{cases} n, & \text{if } \omega \in \left[0, \frac{1}{n}\right], \\ 0, & \text{otherwise.} \end{cases}$$

Since the probability measure specified is the Lebesgue measure, the random variable can be re-written as

$$X_n = \begin{cases} n, & \text{with probability } \frac{1}{n}, \\ 0, & \text{with probability } 1 - \frac{1}{n}. \end{cases}$$

Clearly, when $\omega \neq 0$, $\lim_{n\to\infty} X_n(\omega) = 0$ but it diverges for $\omega = 0$. This suggests that the limiting random variable must be the constant random variable 0. Hence, except at $\omega = 0$, the sequence of random variables converges to the constant random variable 0. Therefore, this sequence does not converge surely, but converges almost surely.
For some $\epsilon > 0$, consider

$$\begin{aligned} \lim_{n\to\infty} \mathbb{P}\left(|X_n| > \epsilon\right) &= \lim_{n\to\infty} \mathbb{P}\left(X_n = n\right), \\ &= \lim_{n\to\infty} \left(\frac{1}{n}\right), \\ &= 0. \end{aligned}$$

Hence, the sequence converges in probability.
Consider the following two expressions:

$$\begin{aligned} \lim_{n\to\infty} \mathbb{E}\left[|X_n|^2\right] &= \lim_{n\to\infty} \left(n^2 \times \frac{1}{n} + 0\right), \\ &= \infty. \end{aligned}$$

$$\begin{aligned}\lim_{n\to\infty}\mathbb{E}\left[|X_n|\right] &= \lim_{n\to\infty}\left(n\times\frac{1}{n}+0\right),\\ &= 1.\end{aligned}$$

Since the above limits do not equal 0, the sequence converges neither in the mean-squared sense, nor in the sense of first mean.
Considering the distribution of X_n's, it is clear (through visualization) that they converge to the following distribution:

$$F_X(x)=\begin{cases}0, & \text{if } x<0,\\ 1, & \text{otherwise.}\end{cases}$$

Also, this happens at each $x\neq 0$ i.e. at all points of continuity of $F_X(\cdot)$. Hence, the sequence of random variables converge in distribution.

So far, we have mentioned that certain notions are weaker than certain others. Let us now formalize the relations that exist among various notions of convergence.

It is immediately clear from the definitions that point-wise convergence implies almost sure convergence. Figure (28.1) is a summary of the implications that hold for any sequence for random variables. No other implications hold in general. We prove these, in a series of theorems, as below.

$$\begin{tikzcd}
\text{p.w.} \arrow[r, Rightarrow] & \text{a.s.} \arrow[dr, Rightarrow] & & \\
& & \text{i.p.} \arrow[r, Rightarrow] & \text{D} \\
r^{th} \text{ mean } (r\geq 1) \arrow[urr, Rightarrow] \arrow[rrr, Rightarrow] & & & s^{th} \text{ mean } (r>s\geq 1)
\end{tikzcd}$$

Figure 28.1: Implication Diagram

Theorem 28.7 $X_n \xrightarrow{r} X \implies X_n \xrightarrow{\text{i.p.}} X, \quad \forall\, r\geq 1.$

Proof: Consider the quantity $\lim\limits_{n\to\infty}\mathbb{P}\left(|X_n-X|>\epsilon\right)$. Applying Markov's inequaliy, we get

$$\begin{aligned}\lim_{n\to\infty}\mathbb{P}\left(|X_n-X|>\epsilon\right) &\leq \lim_{n\to\infty}\frac{\mathbb{E}\left[|X_n-X|^r\right]}{\epsilon^r},\ \forall\epsilon>0,\\ &\overset{(a)}{=} 0,\end{aligned}$$

where (a) follows since $X_n \xrightarrow{r} X$. Hence proved. ■

Theorem 28.8 $X_n \xrightarrow{\text{i.p.}} X \implies X_n \xrightarrow{\text{D}} X.$

Proof: Fix an $\epsilon>0$.

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

Similarly,

$$\begin{aligned}F_X(x-\epsilon) &= \mathbb{P}(X \le x-\epsilon),\\ &= \mathbb{P}(X \le x-\epsilon, X_n \le x) + \mathbb{P}(X \le x-\epsilon, X_n > x),\\ &\le F_{X_n}(x) + \mathbb{P}(|X_n - X| > \epsilon).\end{aligned}$$

Thus,

$$F_X(x-\epsilon) - \mathbb{P}(|X_n - X| > \epsilon) \le F_{X_n}(x) \le F_X(x+\epsilon) + \mathbb{P}(|X_n - X| > \epsilon).$$

As $n \to \infty$, since $X_n \xrightarrow{\text{i.p.}} X$, $\mathbb{P}(|X_n - X| > \epsilon) \to 0$. Therefore,

$$F_X(x-\epsilon) \le \liminf_{n\to\infty} F_{X_n}(x) \le \limsup_{n\to\infty} F_{X_n}(x) \le F_X(x+\epsilon),\ \forall \epsilon > 0.$$

If F is continuous at x, then $F_X(x-\epsilon) \uparrow F_X(x)$ and $F_X(x+\epsilon) \downarrow F_X(x)$ as $\epsilon \downarrow 0$. Hence proved. ■

Theorem 28.9 $X_n \xrightarrow{r} X \implies X_n \xrightarrow{s} X$, *if* $r > s \ge 1$.

Proof: From Lyapunov's Inequality [1, Chapter 4], we see that $\left(\mathbb{E}[|X_n - X|^s]\right)^{1/s} \le \left(\mathbb{E}[|X_n - X|^r]\right)^{1/r}$, $r > s \ge 1$. Hence, the result follows. ■

Theorem 28.10 $X_n \xrightarrow{\text{i.p.}} X \centernot\implies X_n \xrightarrow{\text{r}} X$ *in general.*

Proof: Proof by counter-example:
Let X_n be an independent sequence of random variables defined as

$$X_n = \begin{cases} n^3, & \text{w.p. } \frac{1}{n^2},\\ 0, & \text{w.p. } 1\text{-}\frac{1}{n^2}.\end{cases}$$

Then, $\mathbb{P}(|X_n| > \epsilon) = \frac{1}{n^2}$ for large enough n, and hence $X_n \xrightarrow{\text{i.p.}} 0$. On the other hand, $\mathbb{E}[|X_n|] = n$, which diverges to infinity as n grows unbounded. ■

Theorem 28.11 $X_n \xrightarrow{\text{D}} X \centernot\implies X_n \xrightarrow{\text{i.p.}} X$ *in general.*

Proof: Proof by counter-example:
Let X be a Bernoulli random variable with parameter 0.5, and define a sequence such that X_i=X $\forall\ i$. Let Y=$1 - X$. Clearly, $X_i \xrightarrow{\text{D}} Y$. But, $|X_i - Y|$=1, $\forall\ i$. Hence, X_i does not converge to Y in probability. ■

Theorem 28.12 $X_n \xrightarrow{\text{i.p.}} X \centernot\implies X_n \xrightarrow{\text{a.s.}} X$ *in general.*

Proof: Proof by counter-example:
Let $\{X_n\}$ be a sequence of independent random variables defined as

$$X_n = \begin{cases} 1, & \text{w.p. } \frac{1}{n},\\ 0, & \text{w.p. } 1\text{-}\frac{1}{n}.\end{cases}$$

$\lim\limits_{n\to\infty} \mathbb{P}(|X_n| > \epsilon) = \lim\limits_{n\to\infty} \mathbb{P}(X_n = 1) = \lim\limits_{n\to\infty} \frac{1}{n} = 0$. So, $X_n \xrightarrow{\text{i.p.}} 0$.

Let A_n be the event that $\{X_n = 1\}$. Then, A_n's are independent and $\sum\limits_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$. By Borel-Cantelli Lemma 2, w.p. 1 infinitely many A_n's will occur, *i.e.*, $\{X_n = 1\}$ i.o.. So, X_n does not converge to 0 almost surely. ■

Theorem 28.13 $X_n \xrightarrow{s} X \;\not\Rightarrow\; X_n \xrightarrow{r} X$ *if* $r > s \geq 1$ *in general.*

Proof: Proof by counter-example:
Let $\{X_n\}$ be a sequence of independent random variables defined as

$$X_n = \begin{cases} n, & \text{w.p. } \frac{1}{n^{\frac{r+s}{2}}}, \\ 0, & \text{w.p. } 1\text{-}\frac{1}{n^{\frac{r+s}{2}}}. \end{cases}$$

Hence, $\mathbb{E}[|X_n^s|] = n^{\frac{s-r}{2}} \to 0$. But, $\mathbb{E}[|X_n^r|] = n^{\frac{r-s}{2}} \to \infty$. ■

Theorem 28.14 $X_n \xrightarrow{\text{m.s.}} X \;\not\Rightarrow\; X_n \xrightarrow{\text{a.s.}} X$ *in general.*

Proof: Proof by counter-example:
Let $\{X_n\}$ be a sequence of independent random variables defined as

$$X_n = \begin{cases} 1, & \text{w.p. } \frac{1}{n}, \\ 0, & \text{w.p. } 1\text{-}\frac{1}{n}. \end{cases}$$

$\mathbb{E}[X_n^2]=\frac{1}{n}$. So, $X_n \xrightarrow{\text{m.s.}} 0$. As seen previously (during the proof of Theorem (28.12)), X_n does not converge to 0 almost surely. ■

Theorem 28.15 $X_n \xrightarrow{\text{a.s.}} X \;\not\Rightarrow\; X_n \xrightarrow{\text{m.s.}} X$ *in general.*

Proof: Proof by counter-example:
Let $\{X_n\}$ be a sequence of independent of random variables defined as

$$X_n(\omega) = \begin{cases} n, & \omega \in (0, \frac{1}{n}), \\ 0, & \text{otherwise.} \end{cases}$$

We know that X_n converges to 0 almost surely. $\mathbb{E}[X_n^2]=n \longrightarrow \infty$. So, X_n does not converge to 0 in the mean-squared sense. ■

Before proving the implication $X_n \xrightarrow{\text{a.s.}} X \implies X_n \xrightarrow{\text{i.p.}} X$, we derive a sufficient condition followed by a necessary and sufficient condition for almost sure convergence.

Theorem 28.16 *If* $\forall\, \epsilon > 0$, $\sum_n \mathbb{P}(A_n(\epsilon)) < \infty$, *then* $X_n \xrightarrow{\text{a.s.}} X$, *where* $A_n(\epsilon) = \{\omega : |X_n(\omega) - X(\omega)| > \epsilon\}$.

Proof: By Borel-Cantelli Lemma 1, $A_n(\epsilon)$ occurs finitely often, for any $\epsilon > 0$ w.p. 1. Let $B_m(\epsilon) = \bigcup_{n \geq m} A_n(\epsilon)$. Therefore,

$$\mathbb{P}(B_m(\epsilon)) \leq \sum_{n=m}^{\infty} \mathbb{P}(A_n(\epsilon)).$$

So, $\mathbb{P}(B_m(\epsilon)) \to 0$ as $m \to \infty$, whenever $\sum_n \mathbb{P}(A_n(\epsilon)) < \infty$. An equivalent way of proving almost sure convergence is to first consider $\lim_{m\to\infty} \mathbb{P}\left(\bigcup_{n\geq m} \{\omega : |X_n(\omega) - X(\omega)| > \epsilon\} \right) = 0,\ \forall\, \epsilon > 0$. Hence, $\mathbb{P}(B_m(\epsilon)) \to 0$ as $m \to \infty$. This implies almost sure convergence.

■

Theorem 28.17 *$X_n \xrightarrow{\text{a.s.}} X$ iff $\mathbb{P}(B_m(\epsilon)) \to 0$ as $m \to \infty$, $\forall\ \epsilon > 0$.*

Proof:

Let $A(\epsilon) = \{\ \omega \in \Omega : \omega \in A_n(\epsilon)$ for infinitely many values of n $\} = \bigcap_m \bigcup_{n=m}^{\infty} A_n(\epsilon)$.
If $X_n \xrightarrow{\text{a.s.}} X$, then it is easy to see that $\mathbb{P}(A(\epsilon))$=0, $\forall\ \epsilon > 0$.
Then, $\mathbb{P}\left(\bigcap_{m=1}^{\infty} B_m(\epsilon)\right) = 0$.
Since $\{B_m(\epsilon)\}$ is a nested, decreasing sequence, it follows from the continuity of probability measures that $\lim_{m\to\infty} \mathbb{P}(B_m(\epsilon)) = 0$.

Conversely, let C=$\{\ \omega \in \Omega : X_n(\omega) \to X(\omega)$ as n $\to \infty\}$. Then,

$$\begin{aligned}
\mathbb{P}(C^c) &= \mathbb{P}\left(\bigcup_{\epsilon>0} A(\epsilon)\right) \\
&= \mathbb{P}\left(\bigcup_{m=1}^{\infty} A\left(\frac{1}{m}\right)\right) && \text{as } A(\epsilon) \subseteq A(\epsilon') \text{ if } \epsilon \geq \epsilon' \\
&\leq \sum_{m=1}^{\infty} \mathbb{P}\left(A\left(\frac{1}{m}\right)\right).
\end{aligned}$$

Also, $\mathbb{P}(A(\epsilon)) = \lim_{m\to\infty} \mathbb{P}(B_m(\epsilon))$=0. Consider

$$\begin{aligned}
\mathbb{P}\left(A\left(\frac{1}{k}\right)\right) &= \mathbb{P}\left(\bigcap_{m=1}^{\infty} B_m\left(\frac{1}{k}\right)\right) \\
&= \lim_{m\to\infty} \mathbb{P}\left(B_m\left(\frac{1}{k}\right)\right) \\
&= 0. && \forall\ k \geq 1, \text{ by assumption}
\end{aligned}$$

So, $\mathbb{P}(C^c) = 0$. Hence, $\mathbb{P}(C) = 1$. ■

Corollary 28.18 *$X_n \xrightarrow{\text{a.s.}} X \implies X_n \xrightarrow{\text{i.p.}} X$.*

Proof: $X_n \xrightarrow{\text{a.s.}} X \implies \lim_{m\to\infty} \mathbb{P}(B_m(\epsilon))$=0.
As $A_m(\epsilon) \subseteq B_m(\epsilon)$, it is implied that $\lim_{m\to\infty} \mathbb{P}(A_m(\epsilon))$=0.
Hence, $X_n \xrightarrow{i.p.} X$. ■

Theorem 28.19 *If $X_n \xrightarrow{i.p.} X$, then there exists a deterministic, increasing subsequence $n_1, n_2, n_3, \ldots$ such that $X_{n_i} \xrightarrow{a.s} X$ as $i \longrightarrow \infty$.*

Proof: The reader is referred to Theorem 13 in Chapter 7 of [1] for a proof. ■

Example: Let $\{X_n\}$ be a sequence of independent random variables defined as

$$X_n = \begin{cases} 1, & \text{w.p. } \frac{1}{n}, \\ 0, & \text{w.p. } 1\text{-}\frac{1}{n}. \end{cases}$$

It is easy to verify that, $X_n \xrightarrow{i.p.} 0$, but $X_n \stackrel{a.s.}{\nrightarrow} X$. However, if we consider the subsequence $\{X_1, X_4, X_9, \dots\}$, this (sub)sequence of random variables converges almost surely to 0. This can be verified as follows.
Let $n_i = i^2$, $Y_i = X_{n_i} = X_{i^2}$.
Thus, $\mathbb{P}(Y_i = 1) = \mathbb{P}(X_{i^2} = 1) = \frac{1}{i^2}$.
$\Rightarrow \sum\limits_{i \in \mathbb{N}} \mathbb{P}(Y_i) = \sum\limits_{i \in \mathbb{N}} \frac{1}{i^2} < \infty$. Hence, by BCL-1, $X_i^2 \stackrel{a.s.}{\rightarrow} 0$.
Although this is not a *proof* for the above theorem, it serves to verify the statement via a concrete example.

Theorem 28.20 [Skorokhod's Representation Theorem]
Let $\{X_n, n \geq 1\}$ and X be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ such that X_n converges to X in distribution. Then, there exists a probability space $(\Omega', \mathcal{F}', \mathbb{P}')$, and random variables $\{Y_n, n \geq 1\}$ and Y on $(\Omega', \mathcal{F}', \mathbb{P}')$ such that,

a) $\{Y_n, n \geq 1\}$ and Y have the same distributions as $\{X_n, n \geq 1\}$ and X respectively.

b) $Y_n \stackrel{a.s.}{\rightarrow} Y$ as $n \rightarrow \infty$.

Theorem 28.21 [Continuous Mapping Theorem]
If $X_n \stackrel{D}{\rightarrow} X$, and $g : \mathbb{R} \longrightarrow \mathbb{R}$ is continuous, then $g(X_n) \stackrel{D}{\rightarrow} g(X)$.

Proof: By Skorokhod's Representation Theorem, there exists a probability space $(\Omega', \mathcal{F}', \mathbb{P}')$, and $\{Y_n, n \geq 1\}$, Y on $(\Omega', \mathcal{F}', \mathbb{P}')$ such that, $Y_n \stackrel{a.s.}{\rightarrow} Y$. Further, from continuity of g,
$\{\omega \in \Omega' \mid g(Y_n(\omega)) \rightarrow g(Y(\omega))\} \supseteq \{\omega \in \Omega' \mid Y_n(\omega) \rightarrow Y(\omega)\}$,
$\Rightarrow \mathbb{P}(\{\omega \in \Omega' \mid g(Y_n(\omega)) \rightarrow g(Y(\omega))\}) \geq \mathbb{P}(\{\omega \in \Omega' \mid Y_n(\omega) \rightarrow Y(\omega)\})$,
$\Rightarrow \mathbb{P}(\{\omega \in \Omega' \mid g(Y_n(\omega)) \rightarrow g(Y(\omega))\}) \geq 1$,
$\Rightarrow g(Y_n) \stackrel{a.s.}{\rightarrow} g(Y)$,
$\Rightarrow g(Y_n) \stackrel{D}{\rightarrow} g(Y)$.
This completes the proof since, $g(Y_n)$ has the same distribution as $g(X_n)$, and $g(Y)$ has the same distribution as $g(X)$. ■

Theorem 28.22 *$X_n \stackrel{D}{\rightarrow} X$ iff for every bounded continuous function $g : \mathbb{R} \longrightarrow \mathbb{R}$, we have $\mathbb{E}[g(X_n)] \rightarrow \mathbb{E}[g(X)]$.*

Proof: Here, we present only a partial proof. For a full treatment, the reader is referred to Theorem 9 in chapter 7 of [1].
Assume $X_n \stackrel{D}{\rightarrow} X$. From Skorokhod's Representation Theorem, we know that there exist random variables $\{Y_n, n \geq 1\}$ and Y, such that $Y_n \stackrel{a.s.}{\rightarrow} Y$. From Continuous Mapping Theorem, it follows that $g(Y_n) \stackrel{a.s.}{\rightarrow} g(Y)$, since g is given to be continuous. Now, since g is bounded, by DCT, we have $\mathbb{E}[g(Y_n)] \rightarrow \mathbb{E}[g(Y)]$. Since, $g(Y_n)$ has the same distribution as $g(X_n)$, and $g(Y)$ has the same distribution as $g(X)$, we have $\mathbb{E}[g(X_n)] \rightarrow \mathbb{E}[g(X)]$. ■

Theorem 28.23 *If $X_n \xrightarrow{D} X$, then $C_{X_n}(t) \longrightarrow C_X(t)$, $\forall t$.*

Proof: If $X_n \xrightarrow{D} X$, from Skorokhod's Representation Theorem, there exist random variables $\{Y_n\}$ and Y such that $Y_n \xrightarrow{a.s.} Y$.
So,

$$\cos(Y_n t) \longrightarrow \cos(Y t), \ \cos(X_n t) \longrightarrow \cos(X t), \ \forall t.$$

As $\cos(\cdot)$ and $\sin(\cdot)$ are bounded functions,

$$\mathbb{E}[\cos(Y_n t)] + i\mathbb{E}[\sin(Y_n t)] \longrightarrow \mathbb{E}[\cos(Yt)] + i\mathbb{E}[\sin(Yt)], \ \ \forall t.$$

$$\Rightarrow C_{Y_n}(t) \longrightarrow C_Y(t), \ \ \forall t.$$

We get,

$$C_{X_n}(t) \longrightarrow C_X(t), \ \ \forall t,$$

since distributions of $\{X_n\}$ and X are same as those of $\{Y_n\}$ and Y respectively, from Skorokhod's Representation Theorem. ■

Theorem 28.24 *Let $\{X_n\}$ be a sequence of RVs with characteristic functions, $C_{X_n}(t)$ for each n, and let X be a RV with characteristic function $C_X(t)$. If $C_{X_n}(t) \longrightarrow C_X(t)$, then $X_n \xrightarrow{D} X$.*

Theorem 28.25 *Let $\{X_n\}$ be a sequence of RVs with characteristic functions $C_{X_n}(t)$ for each n, and suppose $\lim_{n\to\infty} C_{X_n}(t)$ exists $\forall t$, and is denoted by $\phi(t)$. Then, one of the following statements is true:*

(a) $\phi(\cdot)$ is discontinuous at $t = 0$, and in this case, X_n does not converge in distribution.

(b) $\phi(\cdot)$ is continuous at $t = 0$, and in this case, ϕ is a valid characteristic function of some RV X. Then $X_n \xrightarrow{D} X$.

Remark 28.26 *In order to prove that the $\phi(t)$ above is indeed a valid characteristic function, we need to verify the three defining properties of characteristic functions. However, in the light of Theorem (28.25), it is sufficient to verify the continuity of $\phi(t)$ at $t = 0$. After all $\phi(t)$ is not an arbitrary function; it is the limit of the characteristic functions of X_ns, and therefore inherits some nice properties. Due to these inherited properties, it turns out it is enough to verify continuity at $t = 0$, instead of verifying all the conditions of Bochner's theorem!*

Note: Theorems (28.24) and (28.25) together are known as Continuity Theorem. For proof, refer to [1].

28.1 Exercises

1. (a) Prove that convergence in probability implies convergence in distribution, and give a counter-example to show that the converse need not hold.
   (b) Show that convergence in distribution to a constant random variable implies convergence in probability to that constant.

2. Consider the sequence of random variables with densities

$$f_{X_n}(x) = 1 \text{ - } \cos(2\pi n x),\ x \in (0, 1).$$

   Do X_n' s converge in distribution? Does the sequence of densities converge?

3. *[Grimmett]* A sequence $\{X_n, n \geq 1\}$ of random variables is said to be *completely convergent* to X if

$$\sum_n \mathbb{P}(|X_n - X| > \epsilon) < \infty, \forall \epsilon > 0$$

Show that, for sequences of independent random variables, complete convergence is equivalent to almost sure convergence. Find a sequence of (dependent) random variables that converge almost surely but not completely.

4. Construct an example of a sequence of characteristic functions $\phi_n(t)$ such that the limit $\phi(t) = \lim_{n\to\infty} \phi_n(t)$ exists for all t, but $\phi(t)$ is not a valid characteristic function.

References

[1] G. G. D. Stirzaker and D. Grimmett. Probability and random processes. Oxford Science Publications, ISBN 0, 19(853665):8, 2001.