Convergence Of Random Variables

Contents

1 Basic Definitions

1.1 Measure and σ-algebra

1.1.1 Properties of Measure

1.2 Integration

1.2.1 Simple Integral

1.2.2 Unsigned Integral

1.2.3 Absolutely convergent Integral

2 Convergence Theorems

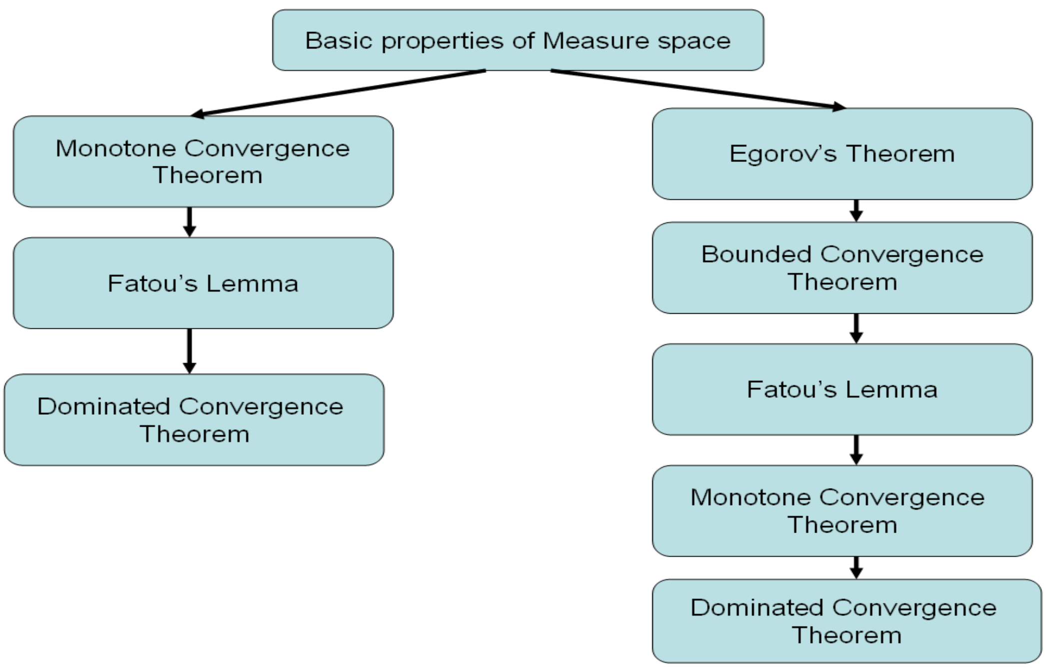


Figure 1: Convergence Theorems

2.1 Monotone Convergence Theorem

Theorem (Monotone Convergence Theorem): Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $0 \leq f_1 \leq f_2 \leq \cdots$ be a monotone nondecreasing sequence of unsigned measurable functions on X. Then we have

$$\lim_{n\to\infty} \int_X f_n \, d\mu = \int_X \lim_{n\to\infty} f_n \, d\mu \tag{1}$$

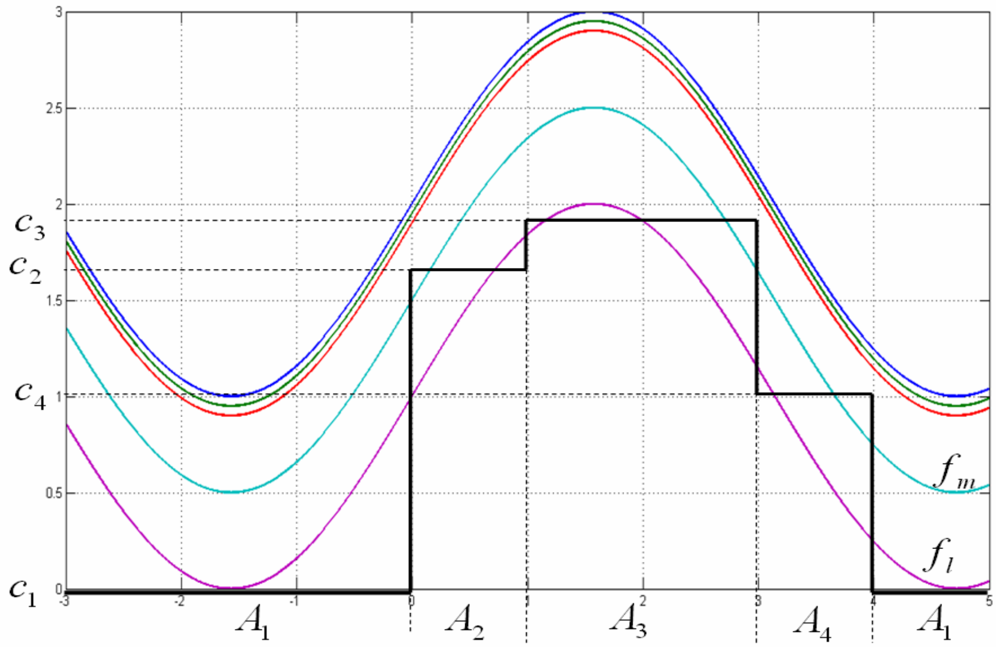


Figure 2: Monotone Convergence Theorem

Proof: To prove the above theorem we will use the standard trick of proving both the inclusions: Let $f := \lim_n f_n = \sup_n f_n$, that is pointwise supremum. We know that $f : X \to [0, +\infty]$ is measurable.

- $\lim_{n\to\infty} \int_X f_n \, d\mu \leq \int_X \lim_{n\to\infty} f_n \, d\mu$
- $\lim_{n\to\infty} \int_X f_n \, d\mu \geq \int_X \lim_{n\to\infty} f_n \, d\mu$

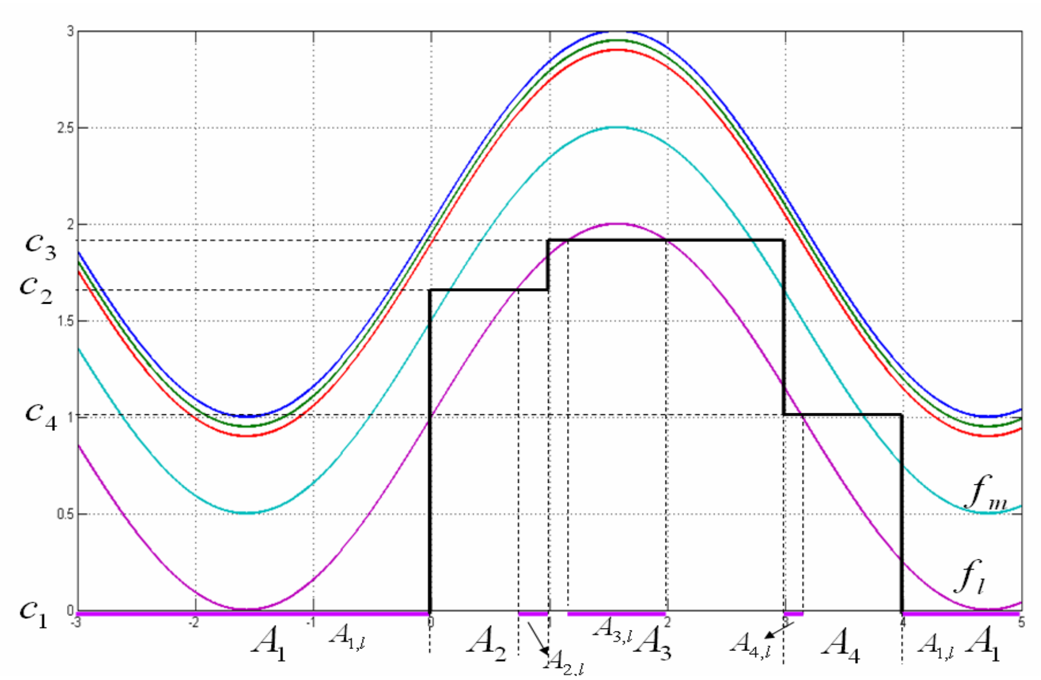


Figure 3: Monotone Convergence Theorem

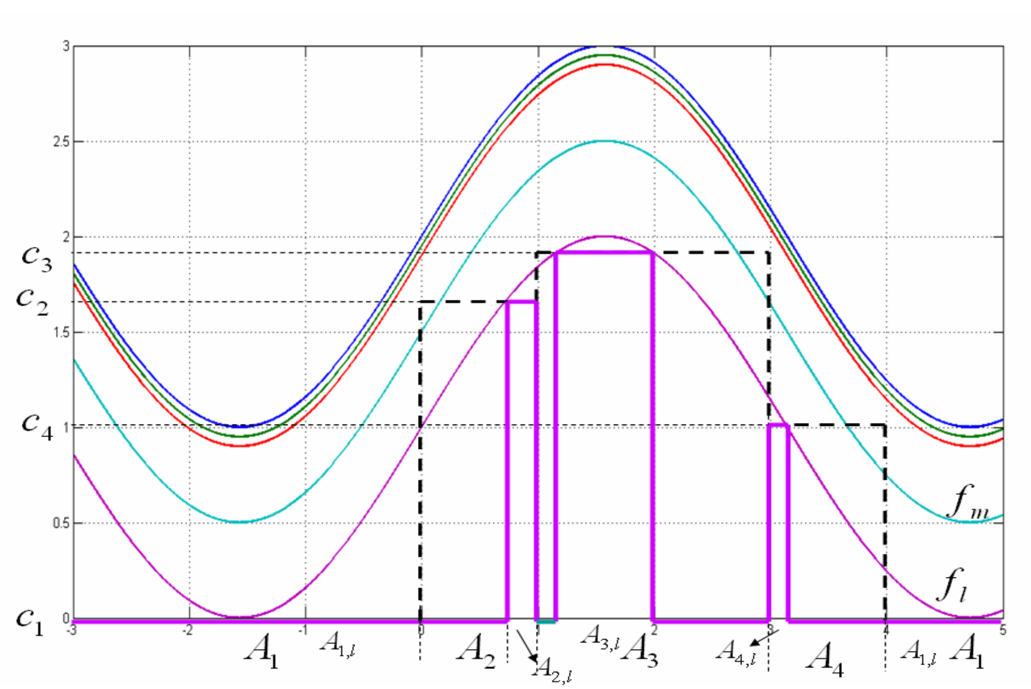


Figure 4: Monotone Convergence Theorem

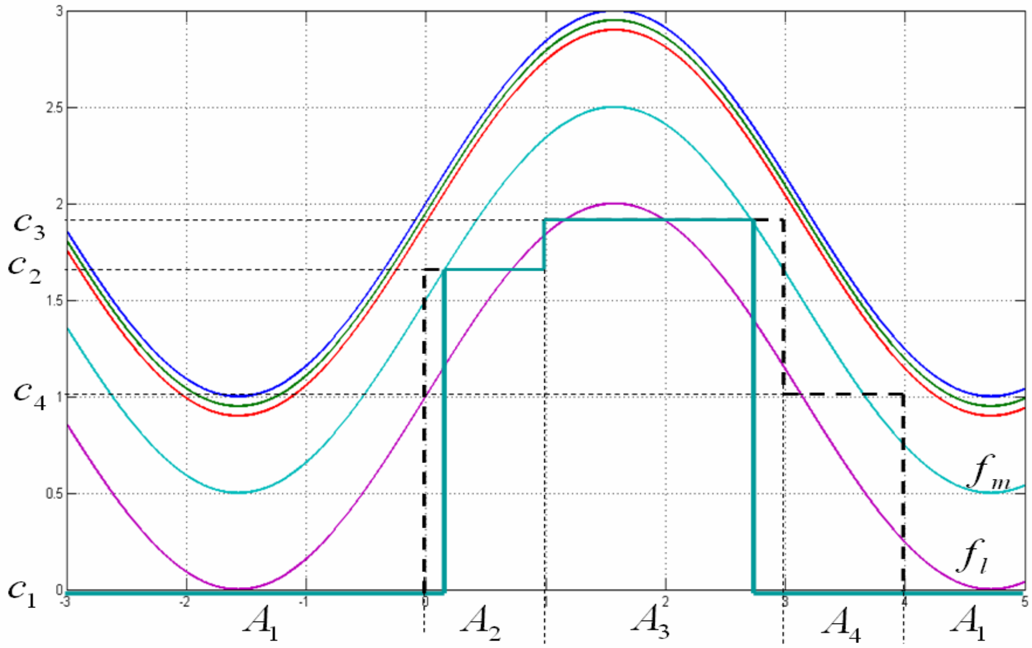


Figure 5: Monotone Convergence Theorem

The monotone convergence theorem is an essential tool in many situations, but its immediate significance for us is as follows. The definition of $\int f$ involves the supremum over a huge (usually uncountable) family of simple functions, so it may be difficult to evaluate $\int f$ directly from the definition. The monotone convergence theorem, however, assures us that to compute $\int f$ it is enough to compute $\lim \int \phi_n$ where $\{\phi_n\}$ is any sequence of simple functions that increase to f, we know that such sequences exist.

2.2 Corollaries to Monotone Convergence Theorem

2.2.1 Tonelli's theorem for exchanging sums

2.2.2 Fatou's Lemma

Monotone convergence theorem gives the condition under which the intergral of the limit functions is equal to the limit of intergral of the sequence of functions, that is when the sequence of functions are non-increasing (or non-decreasing). But when we do not have monotonicity or any kind of structure in the sequence of functions, then, what can we say about the integral of the limit function. Under such a situation, Fatou's lemma gives a general inequality, which says that, in the limit mass can only be lost and not gained.
Example: Consider X_n defined as: Each $n \geq 1$ can be written as $n = 2k + j$ where $k = \lfloor \ln_2 n \rfloor$ and $0 \leq j < 2^k$.

$$X_j(\omega) = \begin{cases} 1 & \text{when} \quad \omega \in (j2^{-k}, (j+1)2^{-k}] \\ 0 & \text{for all other values of } \omega \end{cases} \tag{2}$$

Note that in the above example, the sequence of functions do not converge even at one point. But we have $\liminf f_n = 0$. Fatou's lemma can be well appreciated with this example because the sequence of functions have no monotonicity structure. Neither do the converge to any function. But since lim inf always exists, this inequality, based on a more general quantity holds true.

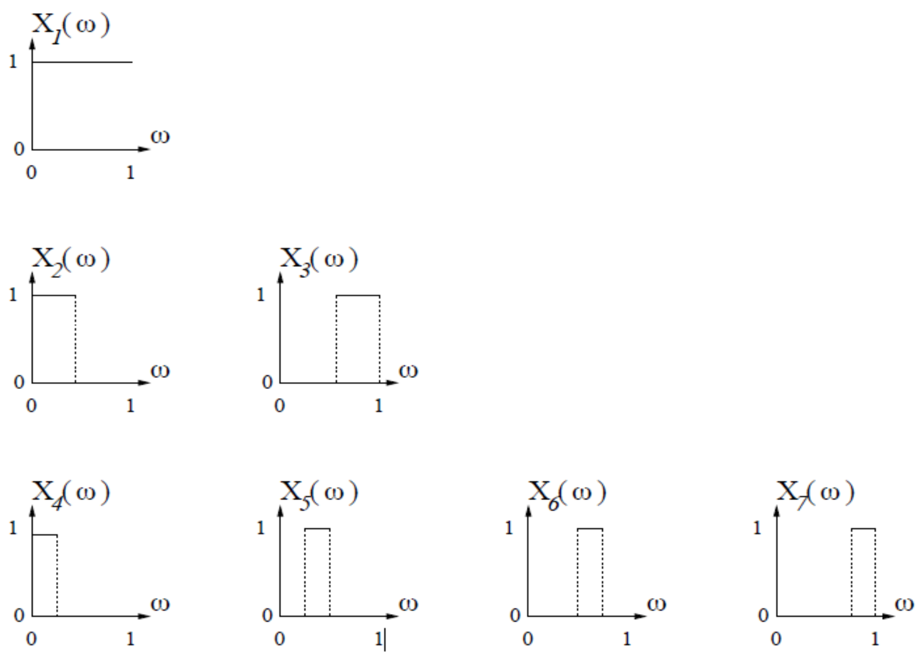


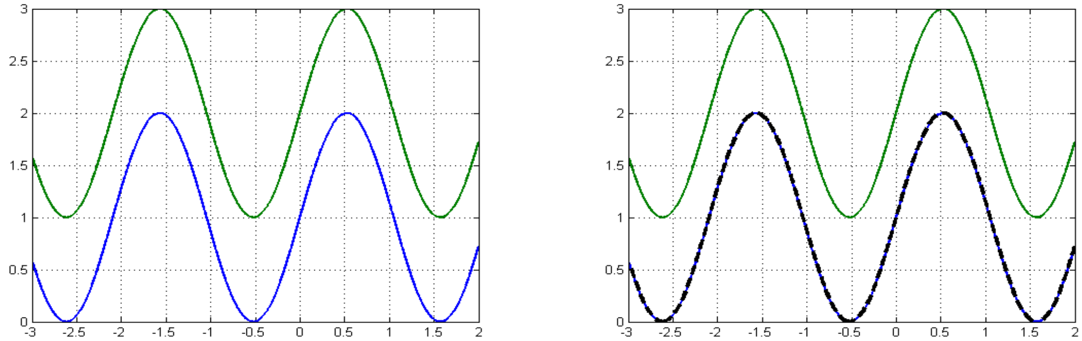
Figure 6: Example: Functions which do not convergence pointwise even at one point.

Before we go about proving the above lemma, let us consider the following 4 cases of sequence of functions and show that Fatou's lemma holds and hence is a general inequality not depending on any kind of structure of sequence of the functions.

- **Case 1:** Let

$$f_n = \begin{cases} \sin(x) + \alpha & \text{if } n \text{ is odd} \\ \sin(x) + \beta & \text{if } n \text{ is even} \end{cases} \tag{3}$$

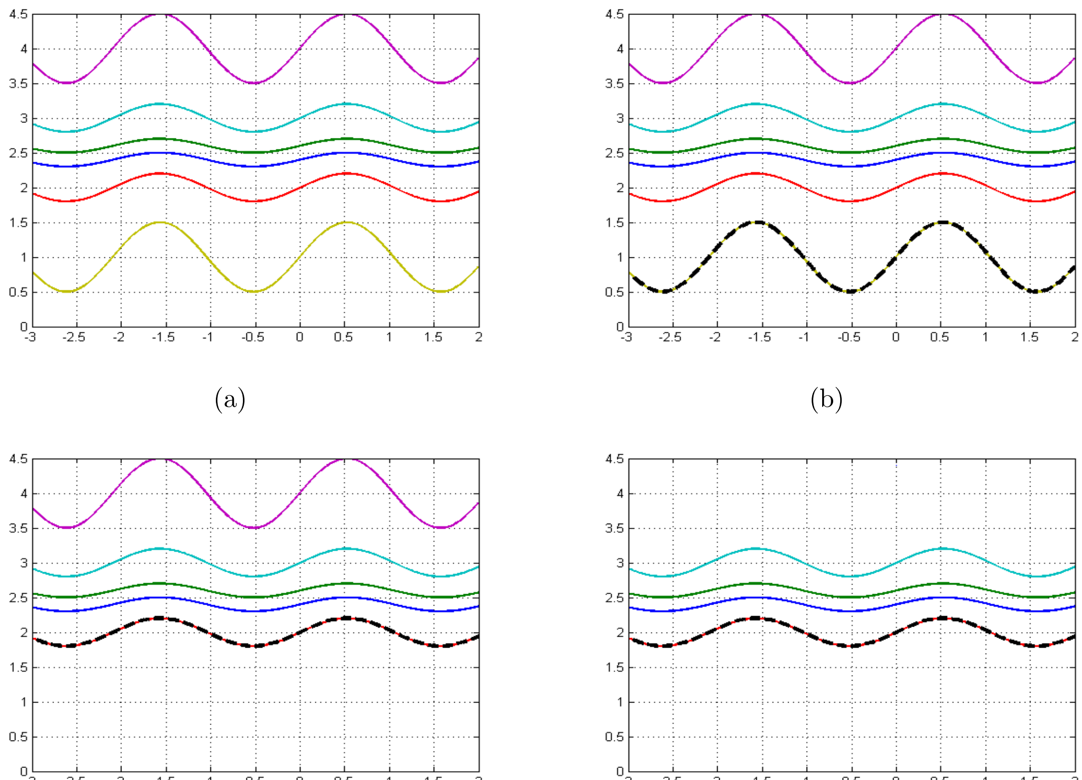
  In this case the sequence of functions alternate between these two functions and do not converge to any limit function. But the liminf exists and we take the liminf.



(a) (b)

Figure 7: Fatou's Lemma

- **Case 2:** Consider the case where the sequence of functions are not monotone but converge to a limit function.



(a) (b)

(c) (d)

Figure 8: Fatou's Lemma

- **Case 3:** Consider the case where the function have the supremum of the infimums at ∞.

$$f_n = \sin(x) + n \tag{4}$$

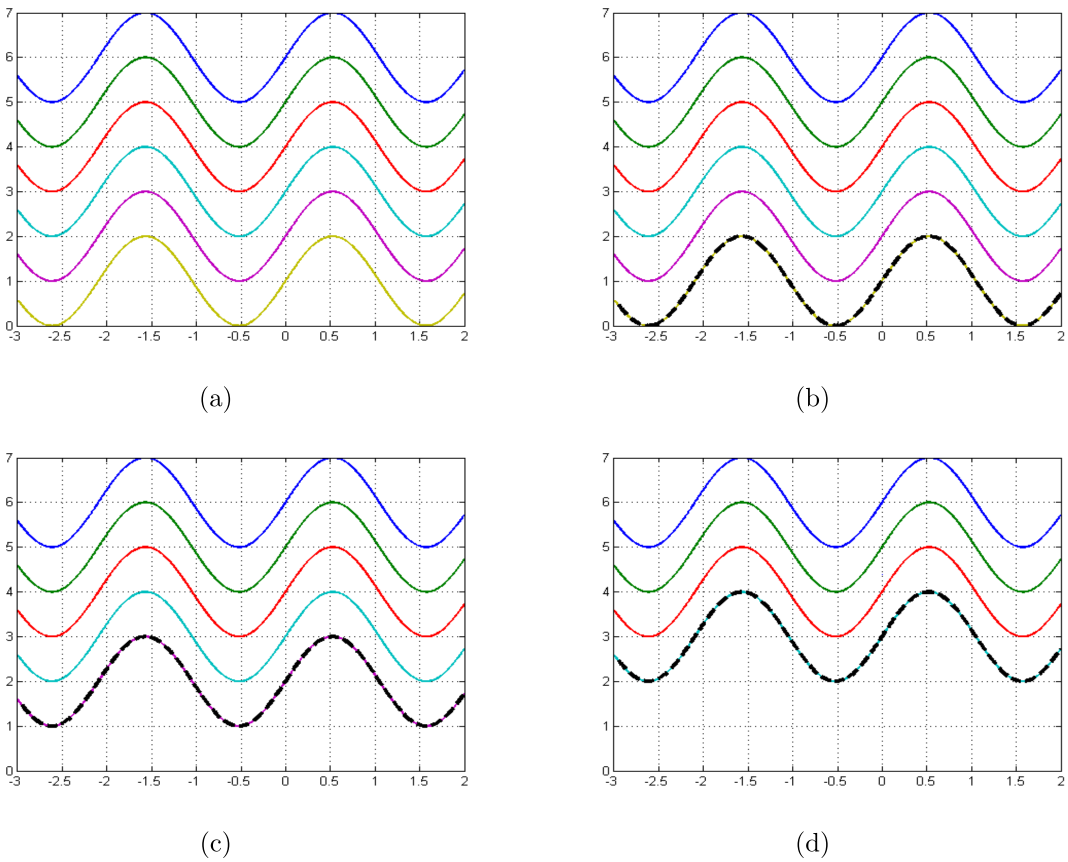
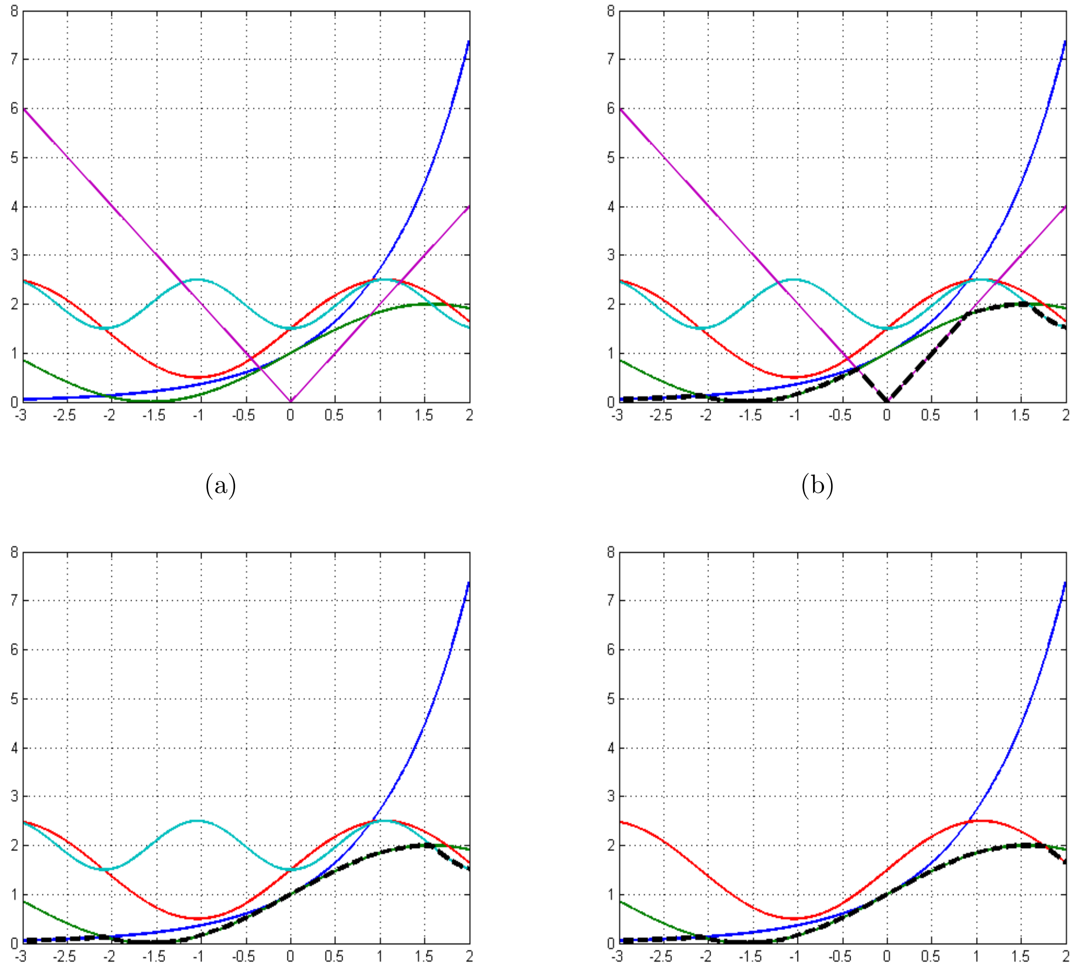


Figure 9: Fatou's Lemma

- **Case 4:** Consider the sequence of monotonically increasing functions which converge to a limit function. In this MCT itself will hold.

The lemma is stated below more formally:
Corollary (Fatou's Lemma): Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $f_1, f_2, \cdots : X \to [0, +\infty]$ be a sequence of unsigned measurable functions. Then

$$\int_X \liminf_{n\to\infty} f_n d\mu \leq \liminf_{n\to\infty} \int_X f_n d\mu \tag{5}$$



(a) (b)

(c) (d)

Figure 10: Fatou's Lemma

Proof: Let $F_N(x) = \inf_{n \geq N} f_n(x)$. As can be seen in the above figure. These sequence of pointwise infimum of the functions forms a sequence of non decreasing functions. From Monotone convergence theorem, we have[1],

$$\int_X \sup_{N>0} F_N d\mu = \sup_{N>0} \int_N F_N d\mu \tag{6}$$

By definition of $\liminf$ we have,

$$\int_X \liminf_{n\to\infty} f_n d\mu = \sup_{N>0} \int_X F_N d\mu \tag{7}$$

[1]Note that when the functions are as in case 3, the supremum will be ∞. In this case the $f = \infty$ at the same time $\lim \int f_n \to \infty$. So the equality still holds.

From the definition of F_N we have, for all $n \geq N$,

$$\int_X F_N \, d\mu \leq \int_X f_n \, d\mu \tag{8}$$
$$\int_X F_N \, d\mu \leq \inf_{n \geq N} \int_X f_n \, d\mu \tag{9}$$

Therefore, we have,

$$\int_X \liminf_{n\to\infty} f_n d\mu = \sup_{N>0} \int_X F_N d\mu \leq \sup_{N>0} \left(\inf_{n\geq N} \int_X f_n d\mu \right) \tag{10}$$
$$\int_X \liminf_{n\to\infty} f_n d\mu \leq \liminf_{n\to\infty} \int_X f_n d\mu \tag{11}$$

In the last equation we have the used the definition of $\liminf$.

Note that in the above hypothesis $f_n \geq 0$ is necessary. If we replace it with $f_n \leq 0$, then we need to replace $\liminf$ with $\limsup$ and $\leq$ with $\geq$. Consider for example one of the moving bump examples multiplied with -1.

2.3 Dominated Convergence Theorem

We now discuss another major way to shut down the loss of mass via escape to infinity, which is to dominate all the functions involved by an absolutely convergent one. This result is known as domniated convergence theorem:

Theorem (Dominated Convergence Theorem): Let $(X, \mathcal{B}, \mu)$ be a measure space and let $f_1, f_2, \cdots : X \to \mathbb{C}$ be a sequence of measurable functions that converge pointwise μ-almost everywhere to a measurable limit $f : X \to \mathbb{C}$. Suppose that there is an unsigned absolutely integrable function$G : X \to [0, +\infty]$ such that $|f_n|$ are pointwise μ-almost everywhere bounded by G for each n. Then we have

$$\lim_{n\to\infty} \int_X f_n \, d\mu = \int_X f \, d\mu \tag{12}$$

3 Characteristic Function

4 Convergence of Random Variables

4.1 Modes of Convergence

If one has a sequence $x_1, x_2, x_3, \ldots \in \mathbf{R}$ of real numbers x_n, it is unambiguous what it means for that sequence to converge to a limit $x \in \mathbf{R}$: it means that for every $\varepsilon > 0$, there exists an N such that $|x_n - x| \leq \varepsilon$ for all $n > N$. Similarly for a sequence $z_1, z_2, z_3, \ldots \in \mathbf{C}$ of complex numbers z_n converging to a limit $z \in \mathbf{C}$.

More generally, if one has a sequence $v_1, v_2, v_3, \ldots$ of d-dimensional vectors v_n in a real vector space $\mathbf{R}^d$ or complex vector space $\mathbf{C}^d$, it is also unambiguous what it means for that sequence to converge to a limit $v \in \mathbf{R}^d$ or $v \in \mathbf{C}^d$; it means that for every $\varepsilon > 0$, there exists an N such that $\|v_n - v\| \leq \varepsilon$ for all $n \geq N$. Here, the norm $\|v\|$ of a vector $v = (v^{(1)}, \ldots, v^{(d)})$ can be chosen to be the Euclidean norm $\|v\|_2 := (\sum_{j=1}^d (v^{(j)})^2)^{1/2}$, the supremum norm $\|v\|_\infty := \sup_{1 \leq j \leq d} |v^{(j)}|$, or any other number of norms, but for the purposes of convergence, these norms are all *equivalent*; a sequence of vectors converges in the Euclidean norm if and only if it converges in the supremum norm, and similarly for any other two norms on the finite-dimensional space $\mathbf{R}^d$ or $\mathbf{C}^d$.

If however one has a sequence $f_1, f_2, f_3, \ldots$ of functions $f_n : X \to \mathbf{R}$ or $f_n : X \to \mathbf{C}$ on a common domain X, and a putative limit $f : X \to \mathbf{R}$ or $f : X \to \mathbf{C}$, there can now be many different ways in which the sequence f_n may or may not converge to the limit f. (One could also consider convergence of functions $f_n : X_n \to \mathbf{C}$ on different domains X_n, but we will not discuss this issue at all here.) This is contrast with the situation with scalars x_n or z_n (which corresponds to the case when X is a single point) or vectors v_n (which corresponds to the case when X is a finite set such as $\{1, \ldots, d\}$). Once X becomes infinite, the functions f_n acquire an infinite number of degrees of freedom, and this allows them to approach f in any number of inequivalent ways.

However, pointwise and uniform convergence are only two of dozens of many other modes of convergence that are of importance in analysis. We will,however, discuss some of the modes of convergence that arise from measure theory, when the domain X is equipped with the structure of a measure space $(X, \mathcal{B}, \mu)$ and the functions are measurable with respect to this space. Furthermore, let us assume that this space has probability measure. In such a context we discuss the following modes of convergences. Consider a sequence of random variables $X_1, X_2, \cdots$ defined on the same probability space $(\Omega, \mathcal{F}, P)$. That is basically a sequence of measurable mappings from Ω to $\mathbb{R}$.

4.2 Almost Sure Convergence

Definition (Almost Sure Convergence): A sequence of random variables $(X_n : n \geq 1)$ converges almost surely to a random variable X, if all the random variables are defined on the same probability space, and

$$P\Big\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\Big\} = 1 \tag{13}$$

Note that almost sure convergence is similar to pointwise convergence of functions. But here the set on which the sequence of measurable functions do not converge should be of measure zero.

Example 1: Let $(X_n : n \geq 1)$ be the sequence of random variables on the standard unit-interval probability space defined by $X_n(\omega) = \omega^n$. This sequence converges for all $\omega \in \Omega$, with the limit

$$\lim_{n\to\infty} X_n(\omega) = \begin{cases} 0 & \text{if} \quad 0 \leq \omega < 1 \\ 1 & \text{if} \quad \omega = 1 \end{cases} \tag{14}$$

The single point set $\{1\}$ has probability zero, so it is also true (and simpler to say) that $(X_n : n \geq 1)$ converges a.s. to zero. In other words, if we let X be the zero random variable, defined by $X(\omega) = 0$ for all ω, then $X_n \rightarrow_{a.s} X$.

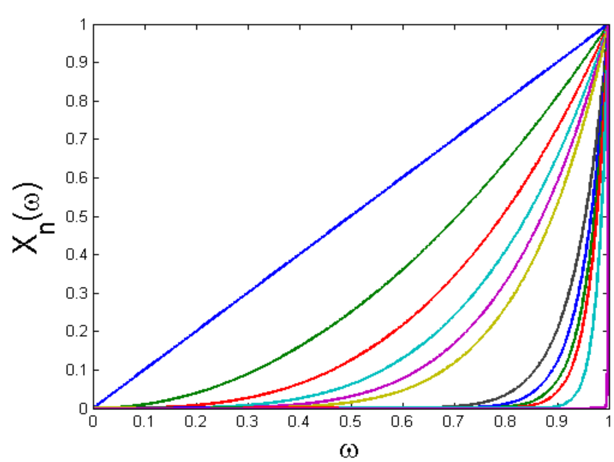


Figure 11: Example: Almost sure convergence

Example 2 (Moving, Shrinking rectangles): In this example we consider two types of moving shrinking rectangles. In one case we show that the random variables converge almost surely. In the second case they do not.

- **Case 1:** Consider X_n be defined as

$$X_n(\omega) = \begin{cases} 2n & \text{when} \quad \omega \in [1/2n, 1/n] \\ 0 & \text{for all other values of } \omega \end{cases} \tag{15}$$

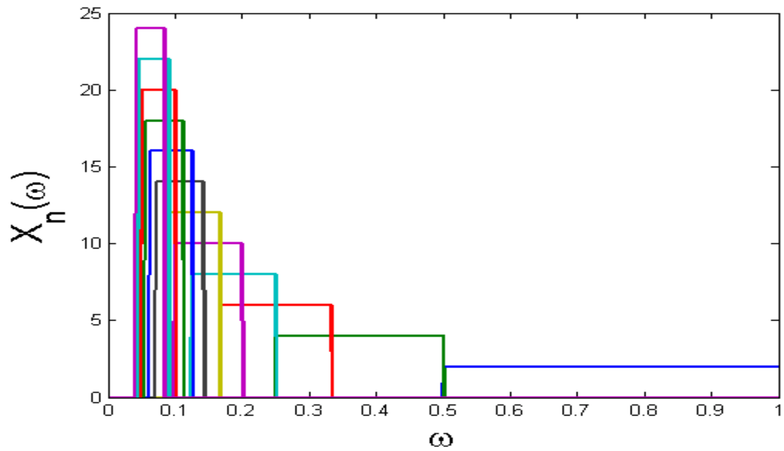


Figure 12: Example: Almost sure convergence

As in example 1, even in this case if we let X be the zero random variable, defined by $X(\omega) = 0$ for all ω, then $X_n \to_{a.s} X$. As in pointwise convergence of functions, each ω can have a different N, such that, for all $n \geq N$, $|X_n(\omega) - X(\omega)| < \epsilon$.

- **Case 2:** Consider X_n defined as: Each $n \geq 1$ can be written as $n = 2k + j$ where $k = \lfloor \ln_2 n \rfloor$ and $0 \leq j < 2^k$.

$$X_j(\omega) = \begin{cases} 1 & \text{when} \quad \omega \in (j2^{-k}, (j+1)2^{-k}] \\ 0 & \text{for all other values of } \omega \end{cases} \tag{16}$$

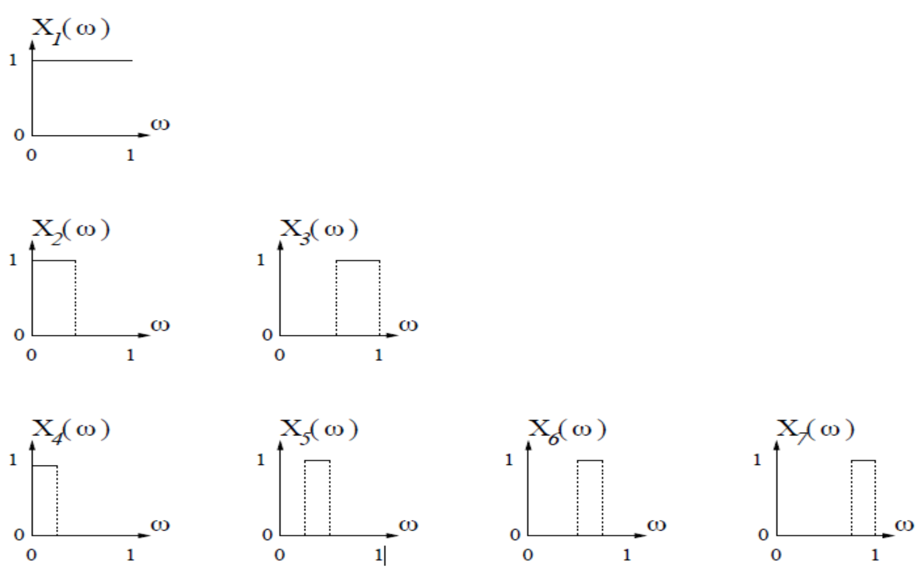


Figure 13: Example: Almost sure convergence

To investigate a.s. convergence, fix an arbitrary value for ω. Then for each $k \geq 1$, there is one value of n with $2^k \leq n < 2^{k+1}$ such that $X_n(\omega) = 1$, and $X_n(\omega) = 0$ for all other n. Therefore,$\lim_{n\to\infty} X_n$ does not exist. That is, $\{\omega : \lim_{n\to\infty} X_n \text{ exists }\} = \emptyset$, so of course, $P\{\omega : \lim_{n\to\infty} X_n \text{ exists }\} = 0$. Thus, X_n does not converge in the a.s. sense.
However, for large n, $P\{\omega : X_n(\omega) = 0\}$ is close to one. This suggests that X_n converges to the zero random variable in some weaker sense.

4.3 Convergence in Probability

The above exampe motivates us to consider the following weaker notion of convergence of a sequence of random variables.
Definition: A sequence of random variables (X_n) converges to a random variable X in probability if all the random variables are defined on the same probability space, and for any $\epsilon > 0$,

$$\lim_{n\to\infty} P\{\omega : |X_n(\omega) - X(\omega)| \geq \epsilon\} = 0 \tag{17}$$

Now, let us examine the difference between convergence in probability and almost sure convergence. Consider a sequence of events in the probability space. $\limsup A_n := \{\omega : \omega \text{ belongs to infinitely many } A_n\}$. Then we can see that

$$\mathbb{1}_{A_n} \to_P 0 \Longleftrightarrow \lim_{n\to\infty} P(A_n) = 0 \tag{18}$$

whereas

$$\mathbb{1}_{A_n} \to_{a.s} 0 \Longleftrightarrow P(\limsup_{n\to\infty} A_n) = 0 \tag{19}$$

The above equations can be seen with respect to the example 2 case 2. It can be seen that for each k, each ω belongs to some A_n. Therefore, each ω belongs to infinitely many A_n. Therefore $\limsup A_n = [0, 1]$, whose measure is not zero. But measure of each A_n is tending to zero. Therefore, it converges in probability.

4.4 Convergence in Mean squared sense

Convergence in probability requires that $|X - X_n|$ be small with high probability (to be precise, less than or equal to ϵ with probability that converges to one as $n \to \infty$), but on the small probability event that $|X - X_n|$ is not small, it can be arbitrarily large. For some applications that is unacceptable. Roughly speaking, the next definition of convergence requires that $|X - X_n|$ be small with high probability for large n, and even if it is not small, the average squared value has to be small enough.

Definition: A sequence of random variables (X_n) converges to a random variable X in the mean square sense if all the random variables are defined on the same probability space, $E[X_n^2] < \infty$ for all n, and $\lim_{n\to\infty} E[(X_n - X)^2] = 0$. Mean square convergence is denoted by $\lim_{n\to\infty} X_n = X$ m.s. or $X_n \to_{m.s} X$.

Example 3: Let us consider the example of moving shrinking rectangle below and see the different modes of convergence of the random variables.Consider the standard unit-interval probability space. Each random variable of the sequence $(X_n : n \geq 1)$ is defined as indicated in the figure below, where the value $a_n > 0$ is some constant depending on n. The graph of X_n for $n \geq 1$ has height a_n over some subinterval of length $\frac{1}{n}$. We don't explicitly identify the location of the interval, but we require that for any fixed ω, $X_n(\omega) = a_n$ for infinitely many values of n, and $X_n(\omega) = 0$ for infinitely many values of n.

- **Almost sure convergence:** $X_n \rightarrow_{a.s} 0$ if the deterministic sequence (a_n) converges to zero. However, if there is a constant $\epsilon > 0$ such that $a_n > \epsilon$ for all n (for example if $a_n = 1$ for all n), then, $\{\omega : \lim_{n\to\infty} X_n(\omega) \text{ exists}\} = \emptyset$.

- **Convergence in probability:** The sequence converges to zero in probability for any choice of the constants (a_n), because for any $\epsilon > 0$,

$$P\{\omega : |X_n(\omega) - X(\omega)| \geq \epsilon\} \leq P\{\omega : X_n(\omega) \geq 0\} = \frac{1}{n} \rightarrow 0 \tag{20}$$

- **Convergence in mean square sense:** Let us now investigate if it converges in the mean squared sense.

$$E[|X_n - 0|^2] = \int_0^1 |x_n(\omega) - 0|^2 p(\omega) d\omega \tag{21}$$

$$= \frac{a_n^2}{n} \tag{22}$$

Hence, $X_n \rightarrow_{m.s} 0$ if and only if the sequence of constants (a_n) is such that $\lim_{n\to\infty} \frac{a_n^2}{n} = 0$.

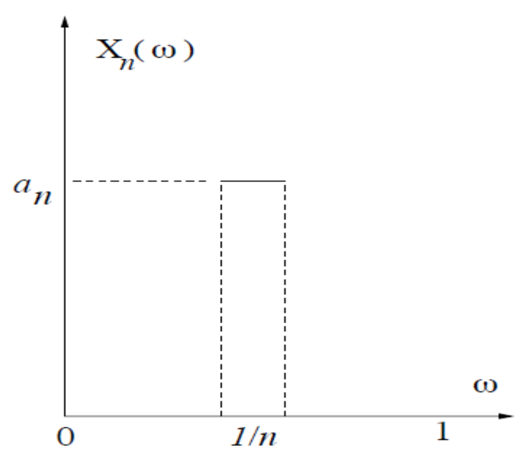


Figure 14: Example: Comparing a.s, p and m.s convergence

4.5 Convergence in Distribution

Consider the below example which motivates the definition of convergence in distribution.
Example 4 (Rearrangement of rectangles): Let $(X_n : n \geq 1)$ be a sequence of random variables defined on the standard unit-interval probability space. The first three random variables in the sequence are indicated in figure below. Suppose that the sequence is periodic, with period three, so that $X_{n+3} = X_n$ for all $n \geq 1$. Intuitively speaking, the sequence of random variables persistently jumps around. Obviously it does not converge in the a.s. sense. The sequence does not settle down to converge, even in the sense of convergence in probability, to any one random variable. Neither does it converge in the mean squared sense.

Even though the sequence fails to converge in a.s., m.s., or p. senses, it can be observed that all of the X_n's have the same probability distribution as shown in Figure.15(c). The variables are only different in that the places they take their possible values are rearranged.

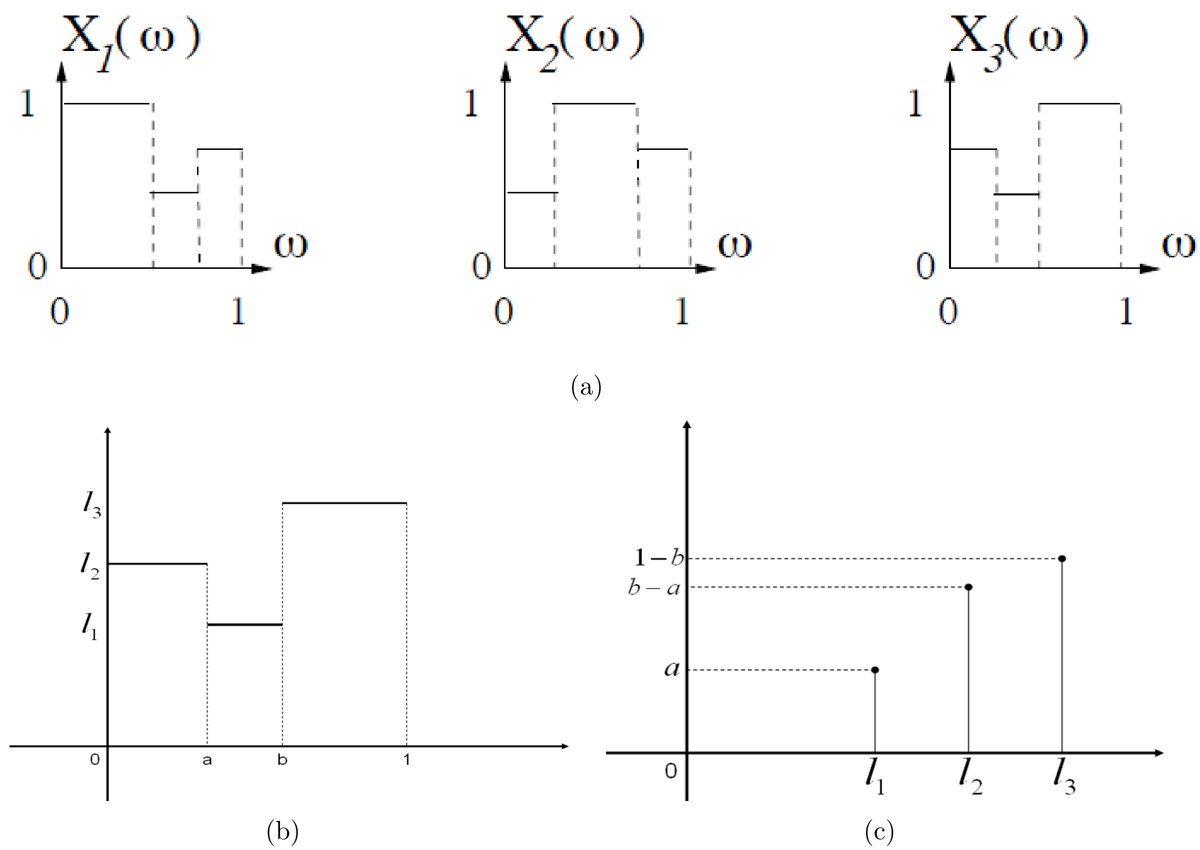


Figure 15: Example: Comparing a.s, p and m.s convergence

Example 2.1.9 suggests that it would be useful to have a notion of convergence that just depends on the distributions of the random variables. One idea for a definition of convergence in distribution is to require that the sequence of CDFs $F_{X_n}(x)$ converge as $n \to \infty$ for all n. The following example shows such a definition could give unexpected results in some cases.

Example 2.1.10 Let U be uniformly distributed on the interval $[0,1]$, and for $n \geq 1$, let $X_n = \frac{(-1)^n U}{n}$. Let X denote the random variable such that $X = 0$ for all ω. It is easy to verify that $X_n \xrightarrow{a.s.} X$ and $X_n \xrightarrow{p.} X$. Does the CDF of X_n converge to the CDF of X? The CDF of X_n is graphed in Figure 2.7. The CDF $F_{X_n}(x)$ converges to 0 for $x < 0$ and to one for $x > 0$. However,

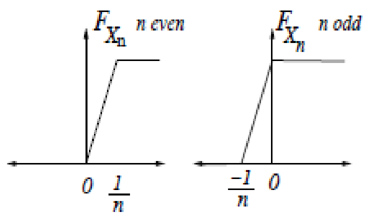


Figure 2.7: CDF of $X_n = \frac{(-1)^n}{n}$.

$F_{X_n}(0)$ alternates between 0 and 1 and hence does not converge to anything. In particular, it doesn't converge to $F_X(0)$. Thus, $F_{X_n}(x)$ converges to $F_X(x)$ for all x except $x = 0$.

Definition (Convergence in distribution): A sequence $\{X_n, n \geq 1\}$ of random variables converge in distribution to a random variable X if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \qquad \text{at all continuity points of} F_X \tag{23}$$

Let $(X, \mathcal{B}, \mu)$ be a measure space. Let $A \in \mathcal{B}$. We say that event A has occured if the outcome is $\omega \in X$ such that $\omega \in A$.
Definition: The event $\{A_n \text{infinitely often}\}$ is the set of $\omega \in X$ such that $\omega \in A_n$ for infinitely many values of n.

Another way to describe $\{A_n \text{infinitely often}\}$ is that it is the set of ω such that for any k, there is an $n \geq k$ such that $\omega \in A_n$. Therefore,

$$\{A_n \text{infinitely often}\} = \cap_{k\geq 1}(\cup_{n\geq} A_n)$$

For each k, the set $\cup_{n\geq k} A_n$ is a countable union of events, so it is an event, and $\{A_n \text{infinitely often}\}$ is an intersection of countably many such events, so that$\{A_n \text{infinitely often}\}$ is also an event.

If A_n is a sequence of subsets of Ω, we let

$$\limsup A_n = \lim_{m\to\infty} \cup_{n=m}^{\infty} A_n = \{\omega \text{ that are in infinitely many } A_n\}$$

(the limit exists since the sequence is decreasing in m) and let

$$\liminf A_n = \lim_{m\to\infty} \cap_{n=m}^{\infty} A_n = \{\omega \text{ that are in all but finitely many } A_n\}$$

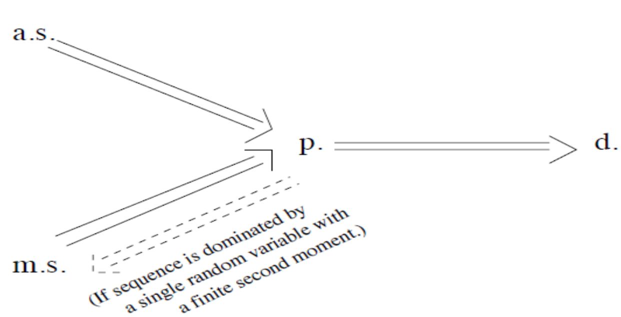
(the limit exists since the sequence is increasing in m). The names lim sup and lim inf can be explained by noting that

$$\limsup_{n\to\infty} 1_{A_n} = 1_{(\limsup A_n)} \qquad \liminf_{n\to\infty} 1_{A_n} = 1_{(\liminf A_n)}$$

It is common to write $\limsup A_n = \{\omega : \omega \in A_n \text{ i.o.}\}$, where i.o. stands for infinitely often. An example which illustrates the use of this notation is: "$X_n \to 0$ a.s. if and only if for all $\epsilon > 0$, $P(|X_n| > \epsilon \text{ i.o.}) = 0$." The reader will see many other examples below. The next result should be familiar from measure theory even though its name may not be.

4.6 Relation between different modes of convergence

The relationships among the four types of convergence discussed in this section are given in the following proposition, and are pictured in figure below. The definitions use differing amounts of information about the random variables $(X_n : n \geq 1)$ and X involved. Convergence in the a.s. sense involves joint properties of all the random variables. Convergence in the p. or m.s. sense involves only pairwise joint distributions-namely those of (X_n, X) for all n. Convergence in distribution involves only the individual distributions of the random variables to have a convergence property. Convergence in the a.s., m.s., and p. senses require the variables to all be defined on the same probability space. For convergence in distribution, the random variables need not be defined on the same probability space.



Note that the above picture is pointing out to the fact that unlike in convergence of vectors with respect to different norms, here convergence of functions in different sense are not equivalent.

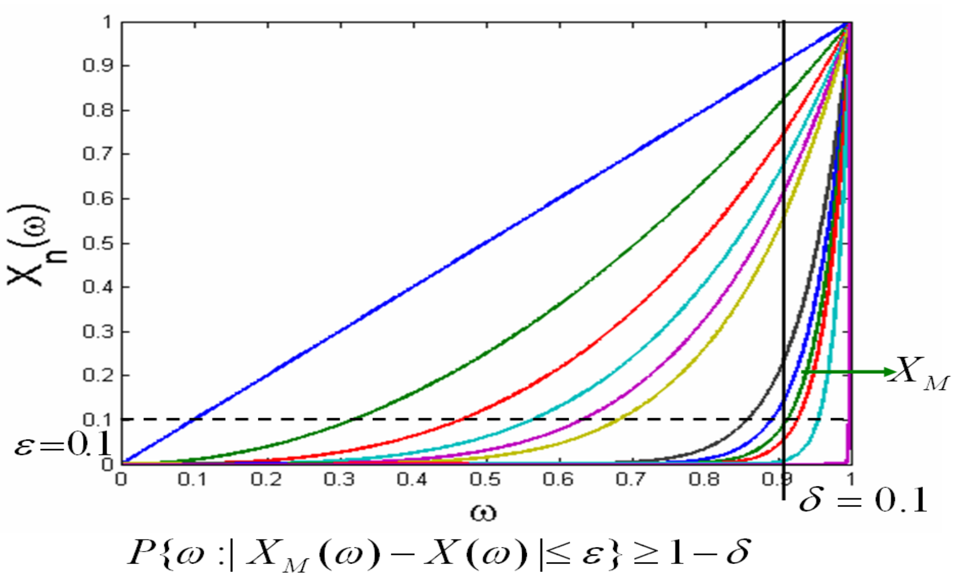
Proposition:(Relation between different modes of convergence)

1. If $X_n \rightarrow^{a.s.} X$ then $X_n \rightarrow^{p} X$.

2.

Proof:

1. Given that $X_n \rightarrow^{a.s.} X$. We need to show that $X_n \rightarrow^{p} X$.
   $X_n \rightarrow^{a.s.} X$ means that
   $$P\{\omega : \lim_{n\to\infty} X_n(\omega) = X\} = 1$$
   . That implies that,
   $$P\{\omega : \text{for every } \epsilon > 0, \text{there exists } N_\omega \text{ such that for all } n \geq N_\omega, |X_n(\omega) - X(\omega)| \leq \epsilon\} = 1$$
   Note that each ω has it own N_w associated with it.



   This means that for every $\epsilon > 0, \delta > 0$, there exists a M, such that for all $n > M$
   $$P\{\omega : |X_n(\omega) - X| \leq \epsilon\} \geq 1 - \delta$$
   . This implies that,
   $$\lim_{n\to\infty} P\{\omega : |X_n(\omega) - X| \leq \epsilon\} = 1$$
   $$\Rightarrow X_n \rightarrow^{p} X$$

2. Given that $X_n \to^{m.s.} X$. We need to show that $X_n \to^{p} X$.

3. Given that $X_n \to^{p} X$. We need to show that $X_n \to^{d} X$.

5 Product measure space

6 Weak Law of Large Numbers

7 Central Limt Theorem