Measure-Theoretic Econometrics for dummies

Raul Mathias Leon Petersen

September 17, 2024

This book has been done with the knowledge obtained during my time at the University of Queensland pursuing a BEcon and BEcon(Honours). I cannot thank enough Prof. Mohamad Khaled for his beautiful course in Advanced Econometric Theory. For my formation in Econometrics and Statistics I would like to thank Prof. Christiern Rose and Prof. Antonio Peyrache for the fundamental courses in Statistical Theory for Economists, Econometric Analysis and Advanced Microeconometrics.

1 Measure Theory

1.1 Why the trouble?

In essence, probability is about "measuring" the frequency of events happening. It was first formalized by Kolmogorov in the 1930s after the groundbreaking advances sorrounding the Lebesgue theory of integration. The new theory allowed to prove nice properties of the integral while only requiring mild conditions.

Specifically, this new theory defined the function $f : X \to \mathbb{R}$ in a new way. It consisted in not considering all the points $x \in X$ but only everywhere except on a countable number of points ("almost everywhere"). Later, you will find out that this set of points become irrelevant since the set in question has "measure zero". More importantly, this way of thinking will provide us with extreme flexibility when integrating since it will allows us to switch interchangeably the limit and integral signs when the limit of an integrable function isn't Riemann integrable but Lebesgue integrable. All of this thanks to some convergence theorems (Monotone convergence theorem, Fatou's lemma and Dominated convergence theorem) which we will study in detail further down the road. You have probably studied the Riemann integral where the subdivisions are done with respect to the function on the x-axis. In the Lebesgue integral, the subdivisions are made with respect to the function on the y-axis.

A direct benefit of this comes from the fact that direct images under f are not necessarily compatible with set operations. In general, one can easily show that $f(A \cap B) \neq f(A) \cap f(B)$ and $f(A \backslash B) \neq f(A) \backslash f(B)$. Nevertheless, inverse images are compatible:

$$f^{-1}(\bigcup_{n=1}^{\infty} A_n) = \bigcup_{n=1}^{\infty} f^{-1}(A_n)$$

$$f^{-1}(\bigcap_{n=1}^{\infty} A_n) = \bigcap_{n=1}^{\infty} f^{-1}(A_n)$$

$$f^{-1}(A\backslash B) = f^{-1}(A)\backslash f^{-1}(B)$$

2 Sets

The basis of probability theory is the probability space with the idea of the stabilization of relative frequencies. Imagine that one performs "independent" repetitions of a random experiment. The intuitive interpretation of the probability concept is that if the probability of some event A happening is 0.5, then one expects that the relative frequency of occurrences of A in the experiment mentioned above should be approximately 0.5. Our job is to axiomatise this.

One first observation is that the number of rules that hold for relative frequencies should also hold for probabilities. This takes us to the following question: Which is the minimal set of rules?

To answer this question one needs to introduce the probability space (i.e. the triple consisting of the sample space, a collection of events and a probability measure). In addition, for $\mathbb{P}$ to be a probability measure it will need to satisfy the three Kolmogorov axioms (to be covered later).

For now, we will need some understanding on how to operate over what we call "events". Some problems could consist on one or two things happening or not happening at all. In other words, we need some conventions on how to combine or not combine them. Thus, we will try to define a collection of sets with a certain structure. For this we will require some basic set theory.

Definition 2.1. $f : X \to Y$ is **injective** if and only if $f(x) = f(x^{'}) \Rightarrow x = x^{'}$. In turn, f is **surjective** iff $f(X) = \{f(x) \in Y : x \in X\} = Y$. Finally, f is **bijective** iff it is both surjective and injective.

Definition 2.2. Two sets A and B are equivalent if there exists a one-to-one mapping from one to the other.

A set is **finite** if it is equivalent to the empty set $\emptyset$ or to $\{1, .., n\}$ for some $n \in \mathbb{N}$.

A set is **countable** if it is either finite or equivalent to $\mathbb{N}$.

A set is **uncountable** if it is not countable.

3 Introduction to Measure Theory

Definition 3.1. $\mathcal{F}$, a set of subsets of Ω is a σ-algebra if
- $\Omega \in \mathcal{F}$.
- $A \in \mathcal{F}$ then $\Omega \backslash A \in \mathcal{F}$.
- $(A_n)_{n \geqslant 1}$ s.t. $A_n \in \mathcal{F}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Definition 3.2. A set A is measurable with respect to $\mathcal{F}$ if $A \in \mathcal{F}$.

Theorem 3.1. *1.* $\varnothing \in \mathcal{F}$.
2. $A, B \in \mathcal{F}$ *then* $A \cup B \in \mathcal{F}, A \cap B \in \mathcal{F}$.
3. $(A_n)_{n \geqslant 1}$ *s.t.* $A_n \in \mathcal{F}$, *then* $\bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$.

Definition 3.3. The σ-algebra generated by (i.e. smallest σ-algebra containing) the open sets of $\mathbb{R}$ is called the Borel σ-algebra.

Definition 3.4. A mapping $\mu : \mathcal{F} \to [0, \infty]$ is a measure if
- $\mu(\varnothing) = 0$.
- $(A_n)_{n \geqslant 1}$ pairwise disjoint sets s.t. $A_n \in \mathcal{F}$, then $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$. μ is a finite measure if $\mu(A) < \infty . \mu$ is a probability measure if $\mu(A) = 1$.

Definition 3.5. Counting measure The measure defined by

$$\mu(A) = \begin{cases} \# A & \text{if } A \text{ is finite} \\ +\infty & \text{otherwise} \end{cases}$$

Theorem 3.2 (Lebesgue measure)**.** *There exists a unique measure* λ *on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *that assigns to an interval* $[a, b)$ *its length*

$$\lambda([a, b)) = |b - a|.$$

This measure is called the Lebesgue measure.

Definition 3.6. X is a real-valued random variable if it is a measurable function from Ω to $\mathbb{R}$, that is

$$X : \Omega \to \mathbb{R}$$

such that, for every borel set A, its inverse image by X is $\mathcal{F}$-measurable

$$X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\} \in \mathcal{F}.$$

Theorem 3.3. *1. If* X_1 *and* X_2 *are two real-valued random variables then* $\min(X_1, X_2)$ *and* $\max(X_1, X_2)$ *are random variables.*
2. If (X_n) *is a sequence of real-valued random variables then* $\sup_n X_n$ *and* $\inf_n X_n$ *are random variables.*

3. If (X_n) is a sequence of real-valued random variables then $\limsup_n X_n$ and $\liminf_n X_n$ are random variables.
4. If (X_n) is a sequence of real-valued random variables and for almost all $\omega, (X_n(\omega))_{n\in\mathbb{N}}$ converges, then $\lim_{n\to\infty} X_n$ is a random variable.
5. A continuous function of a random variable is a random variable.

Construction of the Lebesgue integral Integral of a simple function A non-negative simple function is a function of the form

$$f(x) = \sum_{j=0}^{J} a_j \mathbb{1}\left(x \in A_j\right),$$

where $a_j \geqslant 0$ and A_j are disjoint Borel sets partitioning $\mathbb{R}$. Its integral with respect to a measure μ is given by

$$I_\mu(f) = \sum_{j=0}^{J} a_j \mu\left(A_j\right)$$

One magic result in probability theory concern situations in which the probability of an event can only be 0 or 1. Theorems that provide criteria for such situations are called "zero-one laws".
In order to state the celebrated Kolmogorov Zero-one law we first require the following definition.

Definition 3.7. The tail$-\sigma-$field is defined as

$$\mathcal{T} = \bigcap_{n=1}^{\infty} \mathcal{A}_n^{'}$$

where $\mathcal{A}_n^{'} = \sigma\{A_{n+1}, A_{n+2}, ...\}$for n ≥ 0 and $\mathcal{A}_n = \sigma\{A_1, A_2, ..., A_n\}$for n ≥ 1.

If we think of n as time, then $\mathcal{A}_n^{'}$ contains the information beyond time n and $\mathcal{T}$ contains the information "beyond time n for all n" i.e. the information at infinity (loosely speaking).

Theorem 3.4 (The Kolmogorov Zero-one Law)**.** *Suppose that $\mathcal{A}_n^{'}$ are independent events. If $A \in \mathcal{T}$, then*

$$\mathbb{P}(A) = 0 \ \textit{or}\ 1$$

The law states that if $\{A_n : n \geq 1\}$ are independent events, then the tail$-\sigma-$field is trivial, meaning that it only contains sets of probability measure 0 or 1.

Proof. If $A \in \mathcal{T}$, then $A \in \mathcal{A}_n' \ \forall n$ and, hence, is independent of $\mathcal{A}_n \ \forall n$.

$$\Rightarrow A \in \sigma\{A_1, A_2, A_3...\},$$

as well as,

$$\Rightarrow A \text{ is independent of} \in \sigma\{A_1, A_2, A_3...\}$$

By combining both implications one can conclude that A is independent of itself! Therefore,

$$\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A) \cdot \mathbb{P}(A) = (\mathbb{P}(A))^2$$

$\Rightarrow$ the solutions are $\mathbb{P}(A) = 0$ and $\mathbb{P}(A) = 1$.

■

In reality this is only a sketch since the passage to infinity have to be performed with greater care since the infinite union of σ-algebras need not be a σ-algebra. To correct for this one needs the Approximation lemma. But we live only once, so we will skip the proof.

Theorem 3.5 (Dominated convergence theorem)**.** *Let (f_n) be a sequence of Lebesgue integrable functions dominated by a Lebesgue integrable function g, i.e. satisfying $|f_n(x)| \leqslant |g(x)| \forall x, \forall n$ and $\int |g(x)| \mathrm{d}\mu(x) < \infty$, then $\lim_{n\to\infty} f_n(x)$ exists a.e. and is Lebesgue integrable and*

$$\int \lim_{n\to\infty} f_n(x) \mathrm{d}\mu(x) = \lim_{n\to\infty} \int f_n(x) \mathrm{d}\mu(x)$$

Theorem 3.6 (Monotone convergence for random variables)**.** *Let (X_n) be a sequence of random variables such that $X_n \geqslant 0$ a.s. and $X_n(\omega) \uparrow X(\omega)$ for almost all ω, then*

$$E[X_n] \uparrow E[X]$$

Theorem 3.7 (Fatou's lemma for random variables)**.** *Let (X_n) be a sequence of random variables satisfying $X_n \geqslant 0$ a.s. then*

$$E\left[\liminf_{n\to\infty} X_n\right] \leqslant \liminf_{n\to\infty} E[X_n]$$

Theorem 3.8 (Dominated convergence theorem for random variables)**.** *Let (X_n) be a sequence of real random variables such that $|X_n| < Y$ a.s. $\forall n$ where $E|Y| < \infty$ and let $X_n \to X$ a.s., then*

$$E|X_n - X| \to 0$$

and

$$E[X_n] \to E[X]$$

Definition 3.8. A measure ν is absolutely continuous with respect to a measure μ, denoted by $\nu \ll \mu$ iff $\mu(A) = 0 \Rightarrow \nu(A) = 0 \forall A \in \mathcal{F}$.

Definition 3.9. A measure μ is said to be σ-finite over $(\Omega, \mathcal{F})$ if

$$\Omega = \bigcup_{i=1}^{n} A_i$$

where each A_i has a finite measure $\mu(A_i) < \infty$. As an example, the Lebesgue measure over $(\mathbb{R}, \mathcal{B})$ is σ-finite as $\mathbb{R} = \bigcup_{n=-\infty}^{\infty}[n, n+1)$.

Theorem 3.9 (Radon-Nikodym theorem)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space with μ being σ-finite and let $\nu \ll \mu$ then there exists a unique measurable real positive function g such that*

$$\nu(A) = \int_A g(x)\mu(\mathrm{d}x)$$

g is called the Radon-Nikodym derivative (or density) of ν with respect to μ and is denoted by $g = \mathrm{d}\nu/\mathrm{d}\mu$.

Theorem 3.10 (Decomposition theorem)**.** *Every distribution function F on the real line could be decomposed as a convex combination of three pure type distributions 1. A discrete one F_d (jump function with at most a countable number of jumps. 2. An absolutely continuous one F_c $\Big(F_c = \int_{-\infty}^{x} f(t)\mathrm{d}t$ where $f = \mathrm{d}F_c/\mathrm{d}x)$. 3. A continuous singular one F_s ($\mathrm{d}F_s/\mathrm{d}x = 0$ a.e.)*

That is, there exists $a, b, c \geqslant 0$ such that $a + b + c = 1$ and

$$F(x) = aF_d(x) + bF_c(x) + cF_s(x)$$

Furthermore, the decomposition is unique.

4 Convergence of random variables

Definition 4.1 (Limits of monotone sequences of sets)**.** (A_n) where $A_n \in \mathcal{F}$ is a monotone sequence of sets if either $A_n \nearrow$ or $A_n \searrow$ i.e.
$A_n \nearrow$ if $A_n \subset A_{n+1} \forall n$
$A_n \searrow$ if $A_n \supset A_{n+1} \forall n$
Then, if $A_n \nearrow$ we have that

$$\lim_{n\to\infty} = \bigcup_{n=1}^{\infty} A_n$$

Similarly, if $A_n \searrow$ then

$$\lim_{n\to\infty} = \bigcap_{n=1}^{\infty} A_n$$

Theorem 4.1 (Continuity of probability measure)**.** *If* (A_n) *is a monotone sequence of sets in some* σ*-algebra* $\mathcal{F}$ *s.t.*

$$\lim_{n\to\infty} A_n = A$$

then

$$\mathbb{P}\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A)$$

Proof. Case 1: Assume that A_n is increasing. Then, $A_1 \subset A_2 \subset ... \subset A_n \subset A_{n+1} \subset ...$ and $\lim_{n\to\infty} A_n = \cup_{n=1}^{\infty} A_n$. The difficulty now is that the union of the sets are not disjoint so we cannot use the definition of a measure (countably additivity). The trick in this proof consists in constructing such a union as to utilize the definition.

Following this logic, define recursively $B_n = A_n \cap A_{n-1}^{\mathsf{c}}$ starting with $B_1 = A_1$. This means that $B_2 = A_2 \backslash A_1$, ... , $B_n = A_n \backslash A_{n-1}$. One can easily see that (B_n) is a disjoint sequence of sets since $B_m \cap B_n = \emptyset \, \forall \, m, n$ (i.). Another property of (B_n) is that $\cup_{n=1}^{\infty} B_n = \cup_{n=1}^{\infty} A_n = A$ (ii.) and that $A_n = \cup_{j=1}^{n} B_j (iii.)$.

Now for the proof,

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}(A_n) &= \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{j=1}^{n} B_j\right) && \text{(by iii.)} \\
&= \lim_{n\to\infty} \sum_{j=1}^{n} \mathbb{P}(B_j) && \text{(since Bj's are all disjoint, by i.)} \\
&= \sum_{j=1}^{\infty} \mathbb{P}(B_j) && \text{(monotonically increasing, converges to a finite number or diverges to } +\infty) \\
&= \mathbb{P}(\bigcup_{j=1}^{\infty} B_j) && \text{(countably additivity, Bj's are pairwise disjoint)} \\
&= \mathbb{P}(A) && \text{(by ii.)}
\end{aligned}$$

Case 2: Can be proven analogously. I'll present a shorter route using the result in Case 1. Assume that A_n is decreasing. Then, $A_1^{\mathsf{c}}, A_2^{\mathsf{c}}, A_3^{\mathsf{c}}$, ... are increasing. By the result in Case 1 we know that

$$
\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}(A_n^{\mathsf{c}}) = \mathbb{P}(A^{\mathsf{c}}) \\
\Rightarrow &\lim_{n\to\infty} (1 - \mathbb{P}(A_n)) = 1 - \mathbb{P}(A) \qquad\qquad (\text{since } \mathbb{P}(A^{\mathsf{c}}) = 1 - \mathbb{P}(A)) \\
\Rightarrow &\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)
\end{aligned}
$$

■

Even if sequences of sets are not monotonic we can still define a more general convergence concept.

Definition 4.2. Let (A_n) be a sequence of sets in $\mathcal{F}$. Then,

$$\liminf_n A_n := \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m$$

$$\limsup_n A_n := \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$$

Most importantly, $\lim_n A_n$ is defined to be A only if

$$\limsup_n A_n = \liminf_n A_n = A$$

Note: Interpretation of liminf and limsup for sequences of sets

If $\omega \in \liminf_n A_n$, then $\exists n$, s.t. $\forall m \geq n, \omega \in A_m$.
If $\omega \in \limsup_n A_n$, then $\forall n \ \exists m \geq n$, s.t. $\omega \in A_m$.

Furthermore, there's a correspondence with the concept of limit inferior and superior of sequences of real numbers.

Theorem 4.2. *If $\mathbb{1}_A$ is a characteristic function of a set A, that is $\mathbb{1}_A(\omega) = 1$ if $\omega \in A$ and 0 otherwise, then*

$$\limsup_{n\to\infty} \mathbb{1}_{A_n}(\omega) = \mathbb{1}_{\limsup\limits_{n\to\infty} A_n}(\omega)$$

$$\liminf_{n\to\infty} \mathbb{1}_{A_n}(\omega) = \mathbb{1}_{\liminf\limits_{n\to\infty} A_n}(\omega)$$

Proof. We will start by proving two results. In turn, they will allow us to prove the two identities of the theorem.

Claim 1: $\mathbb{1}_{\cap_{n=1}^{\infty} B_n}(\omega) = 1$ is equivalent to $\inf_{n\in\mathbb{N}} \mathbb{1}_{B_n}(\omega) = 1$.

Proof. $\mathbb{1}_{\cap_{n=1}^{\infty} B_n}(\omega) = 1$ is equivalent to $\omega \in \cap_{n=1}^{\infty} B_n$. This implies that $\forall n \in \mathbb{N}$, $\omega \in B_n$. Which in turn implies that $\forall n \in \mathbb{N}, \mathbb{1}_{B_n}(\omega) = 1$. Thus, we can conclude that $\inf_{n\in\mathbb{N}} \mathbb{1}_{B_n}(\omega) = 1$.

■

Claim 2: $\mathbb{1}_{\cup_{n=1}^{\infty} B_n}(\omega) = 1$ is equivalent to $\sup_{n\in\mathbb{N}} \mathbb{1}_{B_n}(\omega) = 1$.

Proof. $\mathbb{1}_{\cup_{n=1}^{\infty} B_n}(\omega) = 1$ is equivalent to $\omega \in \cup_{n=1}^{\infty} B_n$. This implies that $\exists n \in \mathbb{N}$, $\omega \in B_n$. Which in turn implies that $\exists n \in \mathbb{N}, \mathbb{1}_{B_n}(\omega) = 1$. Thus, we can conclude that $\sup_{n\in\mathbb{N}} \mathbb{1}_{B_n}(\omega) = 1$.

■

Continuing with the proof,

$$\begin{aligned}
\mathbb{1}_{\limsup\limits_{n\to\infty} A_n}(\omega) &= \mathbb{1}_{\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m(\omega)} && \text{(by def. of } \limsup \text{ of sets)} \\
&= \inf_{n\geq 1} \mathbb{1}_{\bigcup_{m=n}^{\infty} A_m(\omega)} && \text{(by claim 1)} \\
&= \inf_{n\geq 1} \sup_{m\geq n} \mathbb{1}_{A_m(\omega)} && \text{(by claim 2)} \\
&= \limsup_{n\to\infty} \mathbb{1}_{A_n}(\omega) && \text{(by def. of } \limsup \text{ of sequences)}
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathbb{1}_{\liminf\limits_{n\to\infty} A_n}(\omega) &= \mathbb{1}_{\bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m(\omega)} && \text{(by def. of } \liminf \text{ of sets)} \\
&= \sup_{n\geq 1} \mathbb{1}_{\bigcap_{m=n}^{\infty} A_m(\omega)} && \text{(by claim 2)} \\
&= \sup_{n\geq 1} \inf_{m\geq n} \mathbb{1}_{A_m(\omega)} && \text{(by claim 1)} \\
&= \liminf_{n\to\infty} \mathbb{1}_{A_n}(\omega) && \text{(by def. of } \liminf \text{ of sequences)}
\end{aligned}$$

■

We can also study another useful equality.

Theorem 4.3. *Let (A_n) be a sequence of sets in some σ-algebra $\mathcal{F}$. Then,*

$$\mathbb{P}\left(\liminf_{n\to\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(\bigcap_{m=n}^{\infty} A_m)$$

Proof. One should see that $(\bigcap_{m=n}^{\infty} A_m)$ increases monotonically to $\liminf_{n\to\infty} A_n$. To see this:

Define $B_n = \cap_{m=n}^{\infty} A_m$ then

$$\begin{aligned} B_n &= \bigcap_{m=n}^{\infty} A_m \\ &= \left(\bigcap_{m=n+1}^{\infty} A_m\right) \bigcap A_{n+1} \\ &\subset \bigcap_{m=n+1}^{\infty} A_m \\ &= B_{n+1} \end{aligned}$$

Thus, $B_1 \subset B_2 \subset ... \subset B_n \subset B_{n+1} \subset ...$ and thus by definition of limits of monotone sequence of sets we have that $B_n \xrightarrow{n\to\infty} \bigcup_{n=1}^{\infty} B_n$. As a result,

$$\begin{aligned} \bigcup_{n=1}^{\infty} B_n &= \bigcup_{n=1}^{\infty} \bigcap_{n=m}^{\infty} A_m \\ &= \liminf_{n\to\infty} A_n \end{aligned}$$

Finally, by the Continuity of Measure Theorem the result follows,

$$\mathbb{P}\left(\liminf_{n\to\infty} A_n\right) = \mathbb{P}\left(\lim_{n\to\infty} B_n\right) = \lim_{n\to\infty} \mathbb{P}\left(B_n\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{m=n}^{\infty} A_m\right)$$

■

Even further we have this important relationship.

Theorem 4.4. *Let (A_n) be a sequence of sets in some σ-algebra $\mathcal{F}$. Then,*

$$\mathbb{P}\left(\liminf_{n\to\infty} A_n\right) \leq \liminf_{n\to\infty} \mathbb{P}(A_n) \leq \limsup_{n\to\infty} \mathbb{P}(A_n) \leq \mathbb{P}\left(\limsup_{n\to\infty} A_n\right)$$

Proof. It is always true that

$$\bigcap_{m=n}^{\infty} A_m \subset A_n \subset \bigcup_{m=n}^{\infty} A_m$$

Hence, we have our initial inequality

$$\mathbb{P}\left(\bigcap_{m=n}^{\infty} A_m\right) \leq \mathbb{P}\left(A_n\right) \leq \mathbb{P}\left(\bigcup_{m=n}^{\infty} A_m\right)$$

Now, we don't know if $\mathbb{P}(A_n)$ converges. But we do know that the limit inferior and superior always exist. So we take limit inferior and superior in the two sides of this initial inequality, respectively. We obtain the following:

$$\begin{aligned}
\mathbb{P}\left(\liminf_{n\to\infty} A_n\right) &= \lim_{n\to\infty} \mathbb{P}(\bigcap_{m=n}^{\infty} A_m) && \text{(by the previous theorem)} \\
&= \liminf_{n\to\infty} \mathbb{P}(\bigcap_{m=n}^{\infty} A_m) && \text{(if the limit exists then it is equal to liminf)} \\
&\leq \liminf_{n\to\infty} \mathbb{P}(A_n) && \text{(by the first part of our initial inequality)} \\
&\leq \limsup_{n\to\infty} \mathbb{P}(A_n) && \text{(by } \liminf_n a_n \leq \limsup_n a_n\text{)} \\
&\leq \limsup_{n\to\infty} \mathbb{P}(\bigcup_{m=n}^{\infty} A_m) && \text{(by the second part of the initial inequality)} \\
&= \lim_{n\to\infty} \mathbb{P}(\bigcup_{m=n}^{\infty} A_m) && \text{(since } \mathbb{P}(\bigcup_{m=n}^{\infty} A_m) \text{ converges*)} \\
&= \mathbb{P}(\lim_{n\to\infty} \bigcup_{m=n}^{\infty} A_m) && \text{(monotonically decreasing, use continuity of measure theorem)} \\
&= \mathbb{P}(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m) && \text{(definition of limit of monotonically decreasing sets)} \\
&= \mathbb{P}\left(\limsup_{n\to\infty} A_n\right) && \text{(by definition of lim sup of sets)}
\end{aligned}$$

* follows since $(\bigcup_{m=n}^{\infty} A_m)$ is decreasing and the continuity of probability measure assures this convergence. ■

Now that we have build some knowledge we will try to prove two very important lemmas that we will use repeatedly from now on which are not very difficult to prove.

4.1 Borel-Cantelli

Lemma 4.5 (First Borel-Cantelli lemma)**.** *Let (A_n) be a sequence of measurable sets $A_n \in \mathcal{F}$. If*

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$$

then

$$\mathbb{P}(A_n i.o.) = 0$$

Proof.

$$\begin{aligned}
\mathbb{P}(A_n i.o.) = \mathbb{P}(\limsup_{n\to\infty} A_n) \quad &= \mathbb{P}(\bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m) \\
\leq \mathbb{P}(\bigcup_{m=n}^{\infty} A_m) & \qquad \text{(Because } \bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m \subset \bigcup_{m=n}^{\infty} A_m \text{)} \\
\leq \sum_{m=n}^{\infty} \mathbb{P}(A_m) & \qquad \text{(we don't know if they are disjoint, hence subadditivity)}
\end{aligned}$$

Now the objective is to make the left hand side go to zero. Observe now a fundamental difference between the left and right hand side. The left does not depend on n while the bound does! We can now study the tail behavior. Hence, since $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$ when we take the $n \to \infty$ the sum $\sum_{m=n}^{\infty} \mathbb{P}(A_m)$ gets smaller and smaller. Eventually it will go to zero since its finite.

Thus, by taking limits on both sides we have that

$$0 \leq \mathbb{P}(A_n i.o.) \leq \lim_{n\to\infty} \sum_{m=n}^{\infty} \mathbb{P}(A_m) = 0$$

■

The converse does not hold in general. The easiest way to fix this is by adding the additional assumption of independence.

Lemma 4.6 (Second Borel-Cantelli lemma)**.** *Let (A_n) be an independent sequence of measurable sets $A_n \in \mathcal{F}$. If*

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$$

then

$$\mathbb{P}(A_n i.o.) = 1$$

Proof. As you will see, it will be easier to work with $\liminf_{n\to\infty}$. Thus,

$$\begin{aligned}
\mathbb{P}(A_n i.o.) = \mathbb{P}(\limsup_{n\to\infty} A_n) = \mathbb{P}(\bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m) \\
&= 1 - \mathbb{P}(\bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} A_m^c) && \text{(By De Morgan's law and } \mathbb{P}(A) = 1 - \mathbb{P}(A^c)) \\
&\geq 1 - \sum_{n=1}^{\infty} \mathbb{P}(\bigcap_{m=n}^{\infty} A_m^c) && \text{(By subadditivity)} \\
&= 1 - \sum_{n=1}^{\infty}\prod_{m=n}^{\infty} \mathbb{P}(A_m^c) && \text{(By independence)} \\
&= 1 - \sum_{n=1}^{\infty}\lim_{k\to\infty}\prod_{m=n}^{k} \mathbb{P}(1 - A_m) && (\mathbb{P}(A) = 1 - \mathbb{P}(A^c)) \\
&\geq 1 - \sum_{n=1}^{\infty}\lim_{k\to\infty} e^{-\sum_{m=n}^{k}\mathbb{P}(A_m)} && \text{(since } 1 - x \leq e^{-x}) \\
&\xrightarrow{k\to\infty} 1
\end{aligned}$$

The last inequality follows because we have a sum of zeros and a one at the limit. Since an event cannot have a probability bigger than one and we have bounded it from below by one, we are done. Subtlety, we have used the fact from real analysis that a sum converges (diverges) if and only if its tail converges (diverges). ∎

This lemmas are important tools that are frequently used in connection to problems concerning almost sure convergence. We will introduce this concept after reviewing convergence in probability.

4.2 Convergence in probability

Definition 4.3 (Convergence in probability)**.** Let X_n be a sequence of random variables. X_n is said to converge in probability to a random variable X, denoted by $X_n \xrightarrow{p} X$ iff

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

I recommend watching a video in youtube to visually understand what the definition is indicating. This convergence concept is very important and jointly with Markov inequality we can proof the Weak Law of Large Numbers. Let's start by proving Markov's inequality.

Theorem 4.7 (Markov's inequality)**.** *Let g be non-negative, non-decreasing function and X be a r.v. s.t.*

$$\mathbb{E}g\big(|X|\big) < \infty$$

Then, for $\epsilon > 0$ we have that

$$\mathbb{P}\{|X| > \epsilon\} \leq \frac{\mathbb{E}g\big(|X|\big)}{g(\epsilon)}$$

Proof.

$$\begin{aligned}
\mathbb{E}g\big(|X|\big) &= \mathbb{E}[g\big(|X|\big)\big(\mathbb{1}(|X| > \epsilon) + \mathbb{1}(|X| \leq \epsilon)\big)] && \text{(Sum to 1 since only one condition can be true)} \\
&= \mathbb{E}[g\big(|X|\big)\big(\mathbb{1}(|X| > \epsilon)] + \mathbb{E}[g\big(|X|\big)\mathbb{1}(|X| \leq \epsilon)] && \text{(By linearity of expectation)} \\
&\geq \mathbb{E}[g\big(|X|\big)\big(\mathbb{1}(|X| > \epsilon)] && \text{(The second term is positive. g is non-negative)} \\
&\geq \mathbb{E}[g(\epsilon)\mathbb{1}(|X| > \epsilon)] && \text{(g is non-decreasing)} \\
&= g(\epsilon)\mathbb{E}[\mathbb{1}(|X| > \epsilon)] && \text{(Take it out since it is non-random)} \\
&= g(\epsilon)\mathbb{P}[|X| > \epsilon]
\end{aligned}$$

The proof is done. But some of you may have doubts regarding the last step. That is, $\mathbb{E}[\mathbb{1}(|X| > \epsilon)] = \mathbb{P}[|X| > \epsilon]$. Observe that:

$$\mathbb{E}[\mathbb{1}(|X| > \epsilon)] = 1 \cdot \mathbb{P}(\mathbb{1}(|X| > \epsilon) = 1) + 0 \cdot \mathbb{P}(\mathbb{1}(|X| > \epsilon) = 0) = \mathbb{P}[|X| > \epsilon]$$

■

Observation: If we let $X = Z - \mathbb{E}[Z]$ and $g(|X|) = |X|^2$ we arrive at the following inequality, known us Chebyshev's inequality:

$$\mathbb{P}\{|X - \mathbb{E}[X]| > \epsilon\} \leq \frac{\mathbb{V}|X|}{\epsilon^2}$$

Chebyshev's inequality is only valid when $\mathbb{E}[X^2] < \infty$ (which also implies $\mathbb{E}|X| < \infty$).

Now we can proof a major theorem that you have probability used a million times in introductory statistics courses.

Theorem 4.8 (Weak Law of Large Numbers)**.** *Let $X_1, ..., X_n$ be i.i.d. with $\mathbb{V}[X_i] < \infty$, then*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mathbb{E}[X_i]$$

Proof. Define u_n as the sample average and $u = \mathbb{E}[X_i]$, then:

$$\mathbb{E}[u_n] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n} \cdot n \cdot u = u$$

Similarly,

$$\mathbb{V}[u_n] = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}[X_i] = \frac{1}{n^2} \cdot n \cdot \mathbb{V}[X_i] = \frac{\mathbb{V}[X_i]}{n}$$

Now, we must show that $u_n \xrightarrow{p} u$. Fix ϵ,

$$\mathbb{P}\big(w : |u_n - u| > \epsilon\big) \leq \frac{\mathbb{V}[u_n - u]}{\epsilon^2} = \frac{\mathbb{V}[X_i]}{n \cdot \epsilon^2}$$
$$\xrightarrow{n\to\infty} 0$$

Even when ϵ is fixed, the right hand side goes to zero as n goes to ∞ as we wanted to show.

What did we use? Chebyshev's inequality to get the inequality and then we used a property of the variance with the small variance result we showed at the beggining. ■

This is the easiest form of this theorem. It requires the variance to exist. If we take certain distributions such as the Cauchy distribution or a t distribution with a very small degrees of freedoms they won't have a finite variance. Do we run into a wall that cannot be climbed? Fortunately, no. This result can be generalized to other situations, although are more complicated to proof.

As in analysis we would want that the limit be unique. Can we prove this?

Theorem 4.9 (Uniqueness of limit)**.** *Let $X_n \xrightarrow{p} X$ and $X_n \xrightarrow{p} Y$, then* $\mathbb{P}\{X = Y\} = 1$

Proof.

$$\begin{aligned}\{\omega : |X(\omega) - Y(\omega))| > \epsilon\} &= \{\omega : |X(\omega) - X_n(\omega) + X_n(\omega) - Y(\omega)| > \epsilon\} \\ &\subset \{\omega : |X(\omega) - X_n(\omega)| > \frac{\epsilon}{2}\} \bigcup |\{X_n(\omega) - Y(\omega)| > \frac{\epsilon}{2}\}\end{aligned}$$

Where does the subset appear? Fix ω. Then, if $|X(\omega) - Y(\omega)| > \epsilon$ then we must clearly have that either $|X(\omega) - X_n(\omega)| > \frac{\epsilon}{2}$ or $|X_n(\omega) - Y(\omega)| > \frac{\epsilon}{2}$. Why? Because $|X(\omega) - Y(\omega)| < |X(\omega) - X_n(\omega) + X_n(\omega) - Y(\omega)|$. There is a certain similarity to the proof in analysis since we are clearly using the triangle inequality. Then, by applying this we get

$$\begin{aligned}\mathbb{P}\{\omega : |X(\omega) - Y(\omega))| > \epsilon\} &\leq \mathbb{P}\{\omega : |X(\omega) - X_n(\omega)| > \frac{\epsilon}{2}\} \bigcup |\{X_n(\omega) - Y(\omega)| > \frac{\epsilon}{2}\} \\ &\leq \mathbb{P}\{\omega : |X(\omega) - X_n(\omega)| > \frac{\epsilon}{2}\}\} + \mathbb{P}\{|\{X_n(\omega) - Y(\omega)| > \frac{\epsilon}{2}\}\end{aligned}$$

We take limits on both extremes of this inequality. The two terms on the right hand-side converge to zero as $n \to \infty$ by convergence in probability. The left hand-side doesn't depend on n and is therefore not affected. We arrive at:

$$\mathbb{P}\{\omega : |X(\omega) - Y(\omega))| > \epsilon\} \leq 0 \text{ ... } (1)$$

This is not what we want to show but will help us later on.

For now, fix ω: If $X(\omega) \neq Y(\omega)$ then either $|X(\omega) - Y(\omega))| > 1$ or $|X(\omega) - Y(\omega))| > \frac{1}{2}$ or $|X(\omega) - Y(\omega))| > \frac{1}{3}$ or $|X(\omega) - Y(\omega))| > \frac{1}{4}$ or ... infinitely often. Following this logic we have that

$$\{\omega : X(\omega) \neq Y(\omega)\} \subset \bigcup_{j=1}^{\infty} \{\omega : |X(\omega) - Y(\omega)| > \frac{1}{j}\}$$

$$\implies 0 \leq \mathbb{P}\{\omega : X(\omega) \neq Y(\omega)\} \leq \sum_{j=1}^{\infty} \mathbb{P}\{\omega : |X(\omega) - Y(\omega)| > \frac{1}{j}\} = 0$$

where the last equality to zero follows from (1).

$$\implies \mathbb{P}\{\omega : X(\omega) = Y(\omega)\} = 1$$

■

Theorem 4.10. *If $X_n \xrightarrow{p} X$ and h is a continuous function, then $h(X_n) \xrightarrow{p} h(X)$*

Proof. Step 1: Assume that $X_n \xrightarrow{p} c$ where c is some constant (a degenerate random variable which is always equal to c) and assume that h is continuous at c. We will want to show that $h(X_n) \xrightarrow{p} h(c)$. This is a much weaker result since we are restricting to a constant and just requiring continuity of h at just one point.

Proof. We are going to approach it using the contrapositive definition of convergence of real numbers i.e. $|f(x) - f(y)| \geq \epsilon \implies |x - y| \geq \delta$. Applying this to our scenario:

$$\begin{aligned} |h(X_n(\omega)) - h(c)| \geq \epsilon &\implies |X_n(\omega) - c| \geq \delta \\ &\implies \mathbb{P}\{\omega : |h(X_n(\omega)) - h(c)| \geq \epsilon\} \leq \mathbb{P}\{\omega : |X_n(\omega) - c| \geq \delta\} \end{aligned}$$

The term in the right of the inequality goes to 0 as $n \to \infty$ (by convergence in probability). We do not know if the term on the left has a limit so in theory we should take $\limsup$ and $\liminf$ and since both are equal to zero, the limit must be equal to zero.

$$\implies h(X_n) \xrightarrow{p} h(X)$$

■

Step 2: Now we will show that if $X_n \xrightarrow{p} X$ and h is uniformly continuous then $h(X_n) \xrightarrow{p} h(X)$. This step 2 will make step 1 redundant but step 1 will help you understand this proof!!! For starters remember that uniformly continuous is a stronger concept than continuity. So it should be easier to proof. Also, the uniformity will allow us to follow the previous logic. Without it one could make an important logical mistake. It has to do with the relationship between ϵ and δ. Initially, we would have that $\delta(\omega, \epsilon)$ (where ω is the x in the standard definition of convergence). But since we will be varying ω we need to require that $\delta(\epsilon)$ so that it doesn't depend on ω.

Proof. Similarly,

$$\{\omega : |h(X_n(\omega)) - h(X(\omega))| \geq \epsilon\} \subset \{\omega : |X_n(\omega) - X(\omega)| \geq \delta\}$$
$$\implies \mathbb{P}\{\omega : |h(X_n(\omega)) - h(X(\omega))| \geq \epsilon\} \leq \mathbb{P}\{\omega : |X_n(\omega) - X(\omega)| \geq \delta\}$$

The result follows from the same logic as before. ∎

Step 3: Now we will unify this to show the initial result. We will split the support of the random variable in two sets. One that is bounded and the other unbounded. In the bounded one we will use uniform continuity (we can do this since continuity implies uniform continuity). In the other set, we will use the fact that if we make the bound in the interval large enough one can make the probability of the tail of the r.v. as close to zero as one desires.
Making the argument concrete:

$$\begin{aligned}\mathbb{P}\{\omega : |h(X_n(\omega)) - h(X(\omega))| \geq \epsilon\} &= \mathbb{P}\big(\{\omega : |h(X_n(\omega)) - h(X(\omega))| \geq \epsilon\} \cap \{\omega : |x| \leq k\}\big) \\ &\quad + \mathbb{P}\big(\{\omega : |h(X_n(\omega)) - h(X(\omega))| \geq \epsilon\} \cap \{\omega : |x| > k\}\big) \\ &\leq \mathbb{P}\big(\{\omega : |h(X_n(\omega)) - h(X(\omega))| \geq \epsilon\} \cap \{\omega : |x| \leq k\}\big) + \mathbb{P}\big(\{\omega : |x| > k\}\big) \\ &\leq \mathbb{P}\{\omega : |X_n(\omega) - X(\omega)| \geq \delta\} + \mathbb{P}\big(\{\omega : |x| > k\}\big) \\ &< \frac{\eta}{2} + \frac{\eta}{2} = \eta\end{aligned}$$

In the first equality we are using the law of total probability (can you see it?). In the inequality we are using a simple fact i.e. $\mathbb{P}(A \cap B) \leq \mathbb{P}(A)$. The first term we have a continuous function h and we are restricting it to a certain interval. Then, in that interval the function h will be uniformly continuous (right?). So by the arguments in step 2 the second inequality follows. We can make the sum of the two probabilities go to zero when $n \to \infty$. The first term goes to zero by convergence in probability. We can thus do the following argument: $\forall \eta > 0 \; \exists N(\eta) \; s.t. \; \forall n \geq N(\eta) \; \mathbb{P}\{\omega : |x_m - x| \geq \delta\} < \eta$. This will also hold for $\frac{\eta}{2}$. The second term does not depend on n and is just a positive number. Nonetheless, remember that it is us who choose k since its arbitrary. We can choose a k such that we make $\mathbb{P}\big(\{\omega : |x| > k\}\big) < \frac{\eta}{2}$.

Then, $\forall \epsilon > 0$, $\forall \eta > 0$, $\exists N(\eta), \exists k(\eta) \; s.t. \; \mathbb{P}\{\omega : |h(X_n(\omega)) - h(X(\omega))| \geq \epsilon\} \leq \mathbb{P}\{\omega : |X_n(\omega) - X(\omega)| \geq \delta\} + \mathbb{P}\big(\{\omega : |x| > k\}\big) < \frac{\eta}{2} + \frac{\eta}{2} = \eta$. With this we have show the global result. ∎

4.3 Almost sure convergence

Definition 4.4 (Almost sure convergence)**.** Let X_n be a sequence of random variables. X_n is said to converge almost surely to a random variable X, denoted by $X_n \xrightarrow{a.s.} X$ iff

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

or equivalently

$$\mathbb{P}\{\omega : \lim_{n\to\infty} X_n \neq X(\omega)\} = 0$$

The zero follows from ours previous result.

Firstly, observe that here we are utilizing pointwise convergence of functions (in a weaker form since it doesn't require it all points). That is, for each ω the functions X_n are converging to X. For the definition to hold, the set which elements are the points where the convergence holds must have probability one. In the equivalent definition, it is stating that the set made up by the points where this convergence is not happening must have probability zero. That is, pointwise convergence must hold except in a set of measure zero.

Theorem 4.11. *An equivalent definition of a.s. convergence is the following.*

$$X_n \xrightarrow{a.s.} X$$

iff $\forall \epsilon > 0$

$$\mathbb{P}\{\limsup_{n\to\infty} |X_n - X| > \epsilon\} = 0$$

Proof. This proof consists in unpacking the definition of a.s. convergence.

For this, we have to start by fixing ω. Then, $lim_{n\to\infty} X_n \neq X(\omega)$ indicates that $\exists \epsilon > 0$ s.t. $\forall N \geq 1$ $\exists n \geq N$, $|X_n - X| \geq \epsilon$. This is equivalent to the following set (in the language of set theory):

$$\bigcup_{\epsilon>0} \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} \{\omega : |X_n - X| \geq \epsilon\}$$

The logic behind this is that the exists implies a union and a for all implies an intersection in the negation of the definition of convergence. (Why?)

Observation inside a proof which is based on an answer in statexchange by FShrike: Replicate this idea for the definition of convergence without negating it. One arrives at

$$\bigcap_{\epsilon>0} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{\omega : |X_n - X| \leq \epsilon\} \text{ ... (1)}$$

The definition stated like this has a huge issue. It isn't necessarily measurable since it contains an uncountable intersection. In turn, this equivalent definition (Why?!)

$$(\bigcap_{j=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{\omega : |X_n - X| \leq \frac{1}{j}\}\} \text{ ... (2)}$$

is a measurable set! Why? Because in a general sigma algebra uncountable intersections may or may not belong to the sigma algebra, unlike countable intersections where this is assured. Hence, this definition is excluding the non-measurable sets and re-stating the convergence definition in a countable way. Now it makes sense to take the probability of that set.

So why can a generic $\epsilon > 0$ be replaced with the following sequence $(1/j)_{j\in\mathbb{N}}$? One way to prove this is by showing that the two sets are equivalent. For easier application of the definitions lets go back to our initial definitions of convergence:

Set (1) contains all ω satisfying:

For all $\epsilon > 0$ there exists some natural number N such that for all natural numbers $n \geq N$, we have $|X_n - X| \leq \epsilon$

Set (2) contains all ω satisfying:

For naturals j there exists some natural number N such that for all natural numbers $n \geq N$, we have $|X_n - X| \leq 1/j$

Clearly set (1) is contained in set (2), since by making $\epsilon = 1/j$ in the definition. Suppose ω is in set (2). Take any $\epsilon > 0$. By the "Archimedean property", there is some $j \in \mathbb{N}$ with $0 < 1/j < \epsilon$. But then because ω is in set (2), there is an N such that if $n \geq N$ we have $|X_n - X| \leq 1/j < \epsilon$. So, if $n \geq N$ we have $|X_n - X| \leq \epsilon$. Because ϵ could be any positive real, we find ω is also in set (1) and we are done.

Therefore, we have showed that (1) = (2). As a result of this, one can replace the $(1/j)_{j\in\mathbb{N}}$ with any strictly positive sequence that converges to zero.

Now continuing with the proof and to complete the whole definition we require to make the probability of the whole set equal to zero. Thus,

$$\mathbb{P}\big(\bigcup_{\epsilon>0}\bigcap_{N=1}^{\infty}\bigcup_{n=N}^{\infty}\{\omega : |X_n - X| \geq \epsilon\}\big) = 0$$

All of this can be rewritten as

$$
\begin{aligned}
0 = \mathbb{P}\big(\omega : lim_{n\to\infty} X_n \neq X(\omega)\big) &= \mathbb{P}\big(\omega : \exists \epsilon > 0 \ s.t. \ \forall N \geq 1 \ \exists n \geq N, |X_n - X| \geq \epsilon\big) \\
&= \mathbb{P}\big(\bigcup_{j=1}^{\infty}\bigcap_{N=1}^{\infty}\bigcup_{n=N}^{\infty}\{\omega : |X_n - X| \geq \frac{1}{j}\}\big) \\
&\geq^{\forall j} \mathbb{P}\big(\bigcap_{N=1}^{\infty}\bigcup_{n=N}^{\infty}\{\omega : |X_n - X| \geq \frac{1}{j}\}\big) \\
&= \mathbb{P}\big(\limsup_{n\to\infty}\{\omega : |X_n - X| \geq \frac{1}{j}\}\big) \\
&\geq 0
\end{aligned}
$$

Since this holds for all j it will also hold for all ϵ and we are done.

Now we can proof the converse. For some $j \in \mathbb{N}$ we have that

$$
\begin{aligned}
0 = \mathbb{P}\big(\limsup_{n\to\infty}\{\omega : |X_n - X| \geq \frac{1}{j}\}\big) &= \mathbb{P}\big(\bigcap_{N=1}^{\infty}\bigcup_{n=N}^{\infty}\{\omega : |X_n - X| \geq \frac{1}{j}\}\big) \\
&= \mathbb{P}\big(\omega : \forall N \geq 1 \ \exists n \geq N, |X_n - X| \geq \frac{1}{j}\big) \\
&= \mathbb{P}\big(\omega : lim_{n\to\infty} X_n \neq X(\omega)\big)
\end{aligned}
$$

■

Now to give a small proof for the basic Archimedean property proved in any basic analysis class:

Theorem 4.12 (Archimedean property). *If $x, y \in \mathbb{R}$ and $x > 0$, then there exists a positive integer n such that $nx \geq y$.*

Proof. Dividing the statement by x we arrive at:

$$\exists n \in \mathbb{N} \ s.t. \ n > \frac{y}{x} := z$$

In other words, we want to show that that $\forall z \in \mathbb{R} \ \exists n \in \mathbb{N} \ s.t. \ n > z$ i.e. z is not an upper bound for $\mathbb{N}$. In essence, this theorem is trying to tell us that the natural numbers are not bounded above. Now, we will proceed by contradiction.

Suppose $\mathbb{N}$ is bounded above. By the Least Upper Bound property (LUB), a basic axiom characterizing $\mathbb{R}$,

$$\exists s \in \mathbb{R} : S = \sup \mathbb{N}$$

In particular, $s \geq n \; \forall \; n \; \in \mathbb{N}$. By one of the Peanos axioms, we know that if $n \in \mathbb{N} \implies n+1 \in \mathbb{N}$. By our last two conclusions we have the following implication, $s \geq n+1 \; \forall \; n \in \mathbb{N}$. Which in turn implies that

$$s - 1 \geq n \; \forall \; n \; \in \mathbb{N}$$

But this just says that $s - 1$ is an upper bound for $\mathbb{N}$. This contradicts the fact that S was the LUB for $\mathbb{N}$. We can conclude that $\mathbb{N}$ is not bounded above and we are done.

$\blacksquare$

In the previous proof we let $x = \epsilon$ and $y = 1$.

Now we are ready to prove one very important theorem. It is basically stating that almost sure convergence is stronger than convergence in probability.

Theorem 4.13. *If $X_n \xrightarrow{a.s.} X$ then $X_n \xrightarrow{p} X$.*

Proof. First observe that

$$\{\omega : |X_n - X| > \epsilon\} \subset \bigcup_{m=n}^{\infty} \{\omega : |X_m - X| > \epsilon\}$$

This is obvious because the set in the left is one of the sets of the union. To show convergence in probability one needs to show that the probability of this first set goes to zero. The first step of this proof consists in bounding this set. The second part consists in making this bound go to zero (taking lim sup). Following the first step in this logic, the previous result implies that

$$\mathbb{P}\big(\{\omega : |X_n - X| > \epsilon\}\big) \leq \mathbb{P}\big(\bigcup_{m=n}^{\infty} \{\omega : |X_m - X| > \epsilon\}\big)$$

$$\Rightarrow \limsup_{n\to\infty} \mathbb{P}\big(\{\omega : |X_n - X| > \epsilon\}\big) \leq \limsup_{n\to\infty} \mathbb{P}\big(\bigcup_{m=n}^{\infty} \{\omega : |X_m - X| > \epsilon\}\big)$$

Since the sets in the union are monotonically decreasing sequence of sets they will converge to the intersection. In addition, since the probability measure is continuous one is allowed to exchange the probability of the limit as the limit of the probability. All in all, this is telling us that the the object in the right-hand side of the inequality is converging.

Thus, lim sup and lim can be used interchangeably. Try to follow the logic below:

$$\begin{aligned}
\limsup_{n\to\infty} \mathbb{P}\big(\{\omega : |X_n - X| > \epsilon\}\big) &\leq \lim_{n\to\infty} \mathbb{P}\big(\bigcup_{m=n}^{\infty} \{\omega : |X_m - X| > \epsilon\}\big) \\
&= \mathbb{P}\big(\lim_{n\to\infty} \bigcup_{m=n}^{\infty} \{\omega : |X_m - X| > \epsilon\}\big) && \text{(continuity of probability measure)} \\
&= \mathbb{P}\big(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \{\omega : |X_m - X| > \epsilon\}\big) && \text{(limit of monotone sequence)} \\
&\leq \mathbb{P}\big(\bigcup_{\epsilon>0} \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \{\omega : |X_m - X| > \epsilon\}\big) && \text{(since the set is included in the union)} \\
&= 0 && \text{(by convergence almost surely)}
\end{aligned}$$

■

Is the other direction (converse) also true ? No! I'll present two counterexamples now. Nevertheless, there exists an interesting and powerful result that is very relevant to the previous question... In essence, we can find a subsequence of the original sequence that converges almost surely! See the following theorem and how we fix the counterexamples to see what I mean by this.

Counterexaple 1: Let X_n be such that $\mathbb{P}\{X_n = 1\} = \frac{1}{n}$ and $\mathbb{P}\{X_n = 0\} = 1 - \frac{1}{n}$. In this example convergence in probability is very easy to see since $\forall \epsilon > 0, \mathbb{P}\{X_n > \epsilon\} = \frac{1}{n} \to 0$ as $n \to \infty$. However, $\sum_{n=1}^{\infty} \frac{1}{n} = \infty$. Why? It is a the harmonic series widely know to diverge. By the second Borel Cantelli lemma, $\mathbb{P}\{\limsup_{n\to\infty} X_n > \epsilon\} = 1$. We thus have shown that converse is not true. This follows by the equivalent definition of a.s. convergence. Nevertheless, take the subsequence X_{n^2}. Then, $\sum_{n=1}^{\infty} \frac{1}{n^2} = 2 < \infty$. Why? Because the reciprocals of powers of 2 produce a convergent sequence. Almost sure convergence follows by a theorem to be proved two theorems down.

Counterexaple 2: Lets try something more extravagant. Let $\Omega = [0,1], \mathcal{F} = \mathcal{B}_{[0,1]}$ and $\mathbb{P}$ Lebesgue measure. Let $Z_n^i = \mathbb{1}_{A_n^i}(\omega), i = 1, ..., n$ and $A_n^i = [\frac{i-1}{n}, \frac{i}{n}]$. Lastly, define the sequence (X_n) as:

$$\big(X_1 = Z_1^1, X_2 = Z_2^1, X_3 = Z_2^2, X_4 = Z_3^1, X_5 = Z_3^2, X_6 = Z_3^3, ...\big)$$

Clearly, $(X_n) \xrightarrow{p} 0$. Why? Observe that $A_1^1 = [0,1], A_2^1 = [0, \frac{1}{2}], A_2^2 = [\frac{1}{2}, 1]$ and so on. Continuing this logic will show that we are partitioning the interval in one, then two and increasingly more parts. One can clearly see that as $n \to \infty$ the Z_n will be equal to one in very small intervals. Thus, (X_n) is getting closer and closer to zero. However, if we fix a point in the interval, after that point infinitely often there is some Z_n that is different than zero. Then, it is clear that it is converging in probability but not almost surely. Formally, we want to show that $\lim_{n\to\infty} \mathbb{P}(X_n > \epsilon) = 0$. Clearly, $\mathbb{P}(X_n > \epsilon) < \frac{1}{n\epsilon}$. Taking limits this goes to zero. Now, try to show that the oscillation doesn't allow for almost sure convergence. Nevertheless, we can fix this by choosing the subsequence (X_{n_k}) where each X_{n_k} is equal to $Z^1_{k^2}$ i.e. $X_1 = Z^1_1, X_2 = Z^1_2, X_4 = Z^1_3$ and so on. Show that this converges almost surely.

Theorem 4.14. *If $X_n \xrightarrow{p} X$ then there exists a subsequence (X_{n_k}) such that*

$$X_{n_k} \xrightarrow{a.s.} X.$$

Proof. The definition of convergence of probability tells us that

$$\forall \epsilon > 0,\ \mathbb{P}\big(\{\omega : |X_m - X| > \epsilon\}\big) \xrightarrow{n\to\infty} 0$$

Since this holds for all ϵ' it will also hold for when $\epsilon' = 1$. Observe that this ϵ' comes from the definition of the limit that is contained in the definition of convergence in probability. Do not confuse this with the standard ϵ. Thus, let $\epsilon' = 1$. Then, $\exists n_1\ s.t.\ \forall n \geq n_1,\ \mathbb{P}\big(\{\omega : |X_m - X| > \epsilon\}\big) < 1$. Following the same logic, let $\epsilon' = \frac{1}{2}$. Then, $\exists n_2\ s.t.\ \forall n \geq n_2,\ \mathbb{P}\big(\{\omega : |X_m - X| > \epsilon\}\big) < \frac{1}{2}$. We can continue analogously.

Generalizing this argument, by letting $\epsilon' = \frac{1}{2^j}$, $\exists n_j > n_{j-1}\ s.t.\ \forall n \geq n_j, \mathbb{P}\big(\{\omega : |X_n - X| > \epsilon\}\big) < \frac{1}{2^j}$.

We can thus take the sum of the probabilities and the sum of the corresponding bounds. We arrive at

$$\sum_{k=1}^{j} \mathbb{P}\big(\{\omega : |X_{n_k} - X| > \epsilon\}\big) \leq \sum_{k=1}^{j} \frac{1}{2^k}$$

As $j \to \infty$ the bound or sum in the right converges to 1. Correspondingly,

$$\sum_{k=1}^{j} \mathbb{P}\big(\{\omega : |X_{n_k} - X| > \epsilon\}\big) = 1 < \infty \Rightarrow X_{n_k} \xrightarrow{a.s.} X$$

The last implication follows by the next theorem which is basically a corollary of First Borel-Cantelli lemma. For now, it is sufficient to prove that the series indeed converges to to 1.

Claim:

$$\sum_{n=1}^{\infty} \frac{1}{2^n} = 1$$

Proof. First we should try proving the following inequality.

$$\sum_{n=1}^{k} \frac{1}{2^n} + \frac{1}{2^k} = 1$$

For k = 1 the equality is valid. The proof follows by induction. Try it!

Now, continuing with the proof. Let s_k represent the partial sums. Then by using the little trick mentioned above,

$$|1 - s_k| = \big|1 - \sum_{n=1}^{k} \frac{1}{2^n}\big| = \big|\frac{1}{2^k}\big| = \frac{1}{2^k}$$

From basic knowledge of analysis one knows that the sequence $\{\frac{1}{2^k}\}$ converges to zero. Therefore, $\{|1 - s_k|\}$ also converges to zero. Implying that $\{|s_k|\}$ converges to 1! Which is what we wanted to proof.

■

The proof is finished. ■

Theorem 4.15. *If* $\forall \epsilon > 0$,

$$\sum_{n=1}^{\infty} \mathbb{P}(\{\omega : |X_m - X| > \epsilon\}) < \infty$$

then $X_n \xrightarrow{a.s.} X$.

Proof. One can see that we have a finite sum. This should bring to our attention the First Borel-Cantelli lemma!

Thus, by the First Borel-Cantelli lemma, we know that

$$\mathbb{P}(\limsup_{n\to\infty}\{\omega : |X_n - X| > \epsilon\}) = 0$$

Showing that the following is zero is equivalent to showing almost sure convergence.

$$\mathbb{P}\big(\bigcup_{\epsilon>0} \limsup_{n\to\infty}\{\omega : |X_n - X| > \epsilon\}\big) = 0$$

To prove this, observe that the left hand side is equivalent to

$$\mathbb{P}\big(\bigcup_{j=1}^{\infty} \limsup_{n\to\infty}\{\omega : |X_n - X| > \frac{1}{j}\}\big) = \lim_{j\to\infty} \mathbb{P}\big(\limsup_{n\to\infty}\{\omega : |X_n - X| > \frac{1}{j}\}\big) = 0$$

The equality follows by observing that the set in the left hand side is an increasing sequence of sets (Why?), the continuity of probability measure theorem and by applying the result mentioned above derived by the First Borel-Cantelli lemma! ■

Since I am a nice guy I am gonna show why the set $\limsup_{n\to\infty}\{\omega : |X_n - X| > \frac{1}{j}\}$ is increasing in j.

Lets destroy this definition by analazing it part by part. We already know that

$$\limsup_n A_n := \bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m$$

Expanding this we arrive at:

$$\bigcap_{n=1}^{\infty}(A_n \cup A_{n+1} \cup A_{n+2} \cup A_{n+3} \cup)$$

Expanding this one more time we arrive at:

$$\big(A_1 \cup A_2 \cup A_3 \cup A_4 \cup ...\big)\bigcap\big(A_2 \cup A_3 \cup A_4 \cup A_5 \cup ...\big)\bigcap\big(A_3 \cup A_4 \cup A_5 \cup A_6 \cup ...\big)\bigcap ...$$

Right?

We also know that the $A_{n,j} \subset A_{n,j+1}$. Why? As $\frac{1}{j}$ becomes smaller and smaller more omegas will satisfy the condition $|X_n(\omega) - X(\omega)| > \frac{1}{j}$.

Thus,

$$\big(A_{1,j}\cup A_{2,j}\cup A_{3,j}\cup A_{4,j}\cup ...\big)\bigcap\big(A_{2,j}\cup A_{3,j}\cup A_4\cup A_{5,j}\cup ...\big)\bigcap\big(A_{3,j}\cup A_{4,j}\cup A_{5,j}\cup A_{6.j}\cup ...\big)\bigcap ...$$

$$\subset \big(A_{1,j+1} \cup A_{2,j+1} \cup A_{3,j+1} \cup A_{4,j+1} \cup ...\big)\bigcap\big(A_{2,j+1} \cup A_{3,j+1} \cup A_{4+1} \cup A_{5,j+1} \cup ...\big)$$

$$\bigcap\big(A_{3,j+1} \cup A_{4,j+1} \cup A_{5,j+1} \cup A_{6.j+1} \cup ...\big)\bigcap ...$$

Why? Sets have the following property. If $A \subset B$ and $C \subset D$, then $A \cap C \subset B \cap D$. Applying this property inductively the desired result follows. Lets prove this property for completeness sake. The argument should go like this:

Let $x \in A \cap C$. Then, $x \in A$ and $x \in C$. But then, $x \in B$ since $A \subset B$ and $x \in A$. Similarly, $x \in C$ and $C \subset D$ implies that $x \in D$. Thus, we can conclude that $x \in B \cap D$, as desired.

Example: Let $X_1, ..., X_n$ be i.i.d. an uniform on the interval $[0, \theta]$. Define

$$\theta_n = \max\big(X_1, ..., X_n\big)$$

where θ_n is an estimator of θ. Is θ_n consistent? That is, $\theta_n \xrightarrow{p} \theta$?

As before, we have to start by fixing ϵ. Now, if $\epsilon \le \theta$

$$\begin{aligned}
\mathbb{P}\big(w : |\theta_n - \theta| > \epsilon\big) &= \mathbb{P}\big(w : \theta - \theta_n > \epsilon\big) \\
&= \mathbb{P}\big(w : \theta_n < \theta - \epsilon\big) \\
&= \mathbb{P}\big(w : \max(X_1, ..., X_n\big) < \theta - \epsilon\big) \\
&= \mathbb{P}\big(w : \big(X_1 < \theta - \epsilon\big) \wedge \big(X_2 < \theta - \epsilon\big) ... \wedge \big(X_n < \theta - \epsilon\big)\big) \\
&= \mathbb{P}\big(\bigcap\{x_i < \theta - \epsilon\}\big) \\
&= \prod_{i=1}^{n} \mathbb{P}\big(w : X_i < \theta - \epsilon\big) \\
&= \prod_{i=1}^{n} \big(\frac{\theta - \epsilon}{\theta}\big) && \text{(definition of the cdf)} \\
&= \big(\frac{\theta - \epsilon}{\theta}\big)^n \\
&\xrightarrow{n \to \infty} 0
\end{aligned}$$

Now, if $\epsilon > \theta$ then it is imposible that $|\theta_n - \theta| > \epsilon$. Thus, $\mathbb{P}\big(w : |\theta_n - \theta| > \epsilon\big) = 0$.

Now, what about a.s. convergence?

$$\sum_{n=1}^{\infty} \mathbb{P}\big(\omega : |\theta_n - \theta| > \epsilon\big) = \sum_{n=1}^{\infty} \big(\frac{\theta - \epsilon}{\theta}\big)^n = \frac{\theta - \epsilon}{\epsilon} < \infty$$

Why? It is an arithmetic series which elements are less than 1! With this we have shown that $\theta_n \xrightarrow{a.s.} \theta$.

Now we can start proving some basic properties of almost sure convergence that are similar to what we proved with convergence in probability. You will be able to observe that the proofs are easier since we are considering a stronger convergence concept (i.e. more conditions which ensure that the proof technique is easier to implement).

Theorem 4.16. *If $X_n \xrightarrow{a.s.} X$ and $X_n \xrightarrow{a.s.} Y$ then $X = Y$ almost surely.*

Proof. By the definition of almost sure convergence, we know that

$$N_X := \{\omega : \lim_{n \to \infty} X_n = X(\omega)\}^c \implies \mathbb{P}\{N_X\} = 0$$

Similarly,

$$N_Y := \{\omega : \lim_{n \to \infty} X_n = Y(\omega)\}^c \implies \mathbb{P}\{N_Y\} = 0$$

Now observe that by the triangle ineuqality:

$$|X(\omega) - Y(\omega)| \leq^{\triangle} |X(\omega) - X_n(\omega)| + |X_n(\omega) - Y(\omega)|$$

If we have convergence this means that we can make both of this terms very small. But for this to happen ω cannot be in N_X and N_Y if not at least one of the terms of the sum cannot be made small. Thus, if $\omega \notin N_X \cup N_Y$ then $|X(\omega) - Y(\omega)| = 0$.
Why?

If $\omega \notin N_X \cup N_Y$ then $\limsup|X(\omega) - Y(\omega)| \leq^{\triangle} \limsup|X(\omega) - X_n(\omega)| + \limsup|X_n(\omega) - Y(\omega)| = \lim|X(\omega) - X_n(\omega)| + \lim|X_n(\omega) - Y(\omega)| = 0$ by almost sure convergence. Similarly, the $\liminf$ can be shown to be equal to zero. Since the $\limsup = \liminf$ we have that $\limsup = \lim$. We don't even need to take $\liminf$ since the term we are bounding from above is a constant sequence (doesn't depend on n). Thus, we can conclude that $|X(\omega) - Y(\omega)| = 0$.

Thus, we have that $X = Y$ when $\omega \notin N_X \cup N_Y$. But, $\mathbb{P}\{N_X \cup N_Y\} \leq \mathbb{P}\{N_X\} \cup \mathbb{P}\{N_Y\} = 0$. This means that $X = Y$ almost surely.

■

Theorem 4.17. *If $X_n \xrightarrow{a.s.} X$ and h is continuous function then $h(X_n) \xrightarrow{a.s.} h(X)$.*

Proof. Fix ω. Then, $X_n \xrightarrow{n\to\infty} X(\omega)$. By continuity of h, we have that $h(X_n) \xrightarrow{n\to\infty} h(X)$. We just have to remember that the pointwise convergence is holding except in a set of measure zero. This implies that:

$$\{\omega : X_n(\omega) \xrightarrow{n\to\infty} X(\omega)\} \subset \{\omega : h(X_n(\omega)) \xrightarrow{n\to\infty} h(X(\omega))\}$$

$$\implies 1 = \mathbb{P}\{\omega : X_n(\omega) \xrightarrow{n\to\infty} X(\omega)\} \leq \mathbb{P}\{\omega : h(X_n(\omega)) \xrightarrow{n\to\infty} h(X(\omega))\}$$

This is just telling us that $h(X_n) \xrightarrow{a.s.} h(X)$. ■

With this knowledge we can now proof one version of the strong laws of large numbers which doesn't assume that the variance exists unlike the weak law of large numbers that we showed. It will also imply the weak law as a special case. It is requiring less things and thus will take more effort to prove it. For that we will require another inequality.

Theorem 4.18 (Kolmogorov's inequality)**.** *Let $X_1, ..., X_n$ be independent with mean 0 and variances $\sigma_1^2, ..., \sigma_n^2$. Let $S_j = \sum_{k=1}^{j} X_k$ be their partial sums. Then,*

$$\mathbb{P}\big(\max_{1\leq j\leq n}|S_j| > \epsilon\big) \leq \frac{\sum_{j=1}^{n} \sigma_j^2}{\epsilon^2}$$

Proof. As in the past, this proof will proceed by creating a sequence of disjoint events to reconstruct the term inside the probability. Let $A_j = \{\max_{1\leq k\leq j-1}|S_k| \leq \epsilon,\ |S_j| > \epsilon\}$ and let $A_1 = \{|S_1| > \epsilon\}$. One can check that $A_j \cap A_{j'} = \emptyset$ and that $\cup_{j=1}^{n} A_j = \{\max_{1\leq j\leq n}|S_j| > \epsilon\}$.

Now,

$$
\begin{aligned}
\mathbb{E}(S_n^2) &\geq \mathbb{E}(S_n^2 \sum_{j=1}^{n} \mathbb{1}(A_j)) && \text{(since only one event can happen)} \\
&= \sum_{j=1}^{n} \mathbb{E}(S_n^2 \mathbb{1}(A_j)) && \text{(linearity of expectation)} \\
&= \sum_{j=1}^{n} \mathbb{E}((S_j + S_n - S_j)^2 \mathbb{1}(A_j)) \\
&= \sum_{j=1}^{n} \mathbb{E}(\big[S_j^2 + (S_n - S_j)^2 + 2S_j(S_n - S_j)\big] \mathbb{1}(A_j)) \\
&\geq \sum_{j=1}^{n} \mathbb{E}(\big[S_j^2 + 2S_j(S_n - S_j)\big] \mathbb{1}(A_j)) && \text{(taking out a positive number)} \\
&\geq \sum_{j=1}^{n} \mathbb{E}(\big[S_j^2\big] \mathbb{1}(A_j)) && \text{($S_j - S_n$ is independent of S_j)} \\
&\geq \sum_{j=1}^{n} \mathbb{E}\big[\epsilon^2 \mathbb{1}(A_j)\big] && \text{(definition of A_j)} \\
&= \epsilon^2 \sum_{j=1}^{n} \mathbb{E}\big[\mathbb{1}(A_j)\big] \\
&= \epsilon^2 \sum_{j=1}^{n} \mathbb{P}(A_j) \\
&\geq \epsilon^2 \mathbb{P}(\bigcup_{j=1}^{n} A_j) && \text{(sub-additivity)} \\
&= \epsilon^2 \mathbb{P}\big(\max_{1\leq j\leq n} |S_j| > \epsilon\big)
\end{aligned}
$$

Then by inserting the initial definition of $S_j = \sum_{k=1}^{j} X_k$, taking the square and using independence and mean 0 of the X_i, linearity of expectation and using the definition of variance we will arrive at the desired result.

∎

This inequality is linking maxima of partial sums to the variances and creating a respective bound.

We will also need the summation by parts formula:

$$\sum_{j=1}^{n} j(b_j - b_{j-1}) = nb_n - b_0 - \sum_{j=0}^{n-1} b_j$$

We can proof this by boring algebra.

$$\sum_{j=1}^{n} j(b_j - b_{j-1}) = \sum_{j=1}^{n} j(b_j) - \sum_{j=1}^{n} j(b_{j-1}) = nb_n + \sum_{j=1}^{n-1} j(b_j) - \sum_{j=1}^{n} j(b_{j-1})$$

Now we do a change of variable: $j^{'} = j - 1$. Then the last term becomes

$$nb_n + \sum_{j=1}^{n-1} j(b_j) - \sum_{j=0}^{n-1} (j+1)(b_j) = nb_n - b_0 - \sum_{j=0}^{n-1} b_j$$

as required.

We will also need one more intermediate result.

Lemma 4.19. *Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of real numbers and $a_n \to a$. Then, $b_n \to a$ where $b_n = \frac{1}{n}\sum_{i=1}^{n} a_i$.*

Proof. Take $(a_n)_{n\in\mathbb{N}}$ converges and WLOG let $a = 0$. Then,

$$|b_n| = |\frac{1}{n}\sum_{i=1}^{n} a_i| = |\frac{1}{n}\sum_{i=1}^{N(\epsilon)} a_i + \frac{1}{n}\sum_{i=N(\epsilon)+1}^{n} a_i|$$

Then, by the traingle inequality we have that

$$\leq \frac{1}{n}|\sum_{i=1}^{N(\epsilon)} a_i| + \frac{1}{n}|\sum_{i=N(\epsilon)+1}^{n} a_i|$$

Then, by the triangle inequality once more and using the condition of epsilon in the definition of convergence we have this new upper bound

$$\leq \frac{1}{n}\sum_{n=1}^{N(\epsilon)} |a_i| + \frac{n - N(\epsilon)}{n}\epsilon$$

We can take

$$\lim_{n\to\infty}\sup|b_n| \leq 0 + \epsilon$$

since the first term is a fixed number divided by n as n goes to infinity. The second term is just $\epsilon - \frac{N(\epsilon)}{n}\epsilon$ where the first part is a constant and the second part is a constant divided by n. Since ϵ was arbitrary this shows that b_n must converge to zero. ■

Theorem 4.20. *Let $X_1, ..., X_n$ be i.i.d. with $\mathbb{E}[X_i] < \infty$, then*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s.} \mathbb{E}[X_i]$$

This is harder to proof since we are considering almost sure convergence. Even more, since we assume that the expectation exists but not the variance we are requiring a very weak condition (since $\mathbb{V}[X_i] < \infty \implies \mathbb{E}[X_i] < \infty$). In a sense, the higher the moment we select, the stronger the condition we are selecting. This theorem is saying that even if the variance does not exist this result will still hold. Incredible, no? One can actually proof another version of the theorem where one drops the assumption that the expectation exists. Nevertheless, this theorem will be meaningless since if the expectation is infinite why would one need to compute the sample average of a random variable? One could validly ask if there even exists a distribution with such a property? Yes! For example, the moments of a Cauchy distribution do not exist. Also, an example relevant to this theorem is the Pareto distribution which depending on the value of its parameter the variance will not exist but will have a finite mean. This type of distribution is used in finance to model extreme events.

Proof. We start by fixing ω and try to show that if $\sum_{n=1}^{\infty} \frac{X_n(\omega)}{n} < \infty$ then $\frac{1}{n}\sum_{i=1}^{n} X_i(\omega) \to 0$. This convergence is in the sense of a real sequence. Thus, define b_n to be the partial sum:

$$b_n = \sum_{i=1}^{n} \frac{X_i(\omega)}{i}$$

and let $b_0 = 0$ arbitrarily. This can be done since the originally sum begins at 1. Then we decompose b_n :

$$b_n = \sum_{i=1}^{n-1} \frac{X_i(\omega)}{i} + \frac{X_n(\omega)}{n} = b_{n-1} + \frac{X_n(\omega)}{n}$$
$$\implies X_n = nb_n - nb_{n-1}$$

Now by applying the summation by parts formula:

$$\frac{1}{n}\sum_{i=1}^{n} X_i(\omega) = \frac{1}{n}\sum_{i=1}^{n} i(b_i - b_{i-1}) = b_n - \frac{b_0}{n} - \frac{1}{n}\sum_{i=0}^{n-1} b_i$$

Since $b_0 = 0$ we arrive at:

$$\frac{1}{n}\sum_{i=1}^{n} X_i(\omega) = b_n - \frac{1}{n}\sum_{i=1}^{n-1} b_i$$

If the sum is finite i.e. $b_\infty < \infty$ then the right side converges to zero as n goes to infinity:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i(\omega) = b_\infty - \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n-1} b_i = b_\infty - 1 \times b_\infty = 0$$

We now have to show the equivalent result for almost sure convergence i.e. when ω is not fixed. If $\sum_{n=1}^{\infty} \frac{X_n(\omega)}{n} < \infty$ a.s., then $\frac{1}{n}\sum_{i=1}^{n} X_i(\omega) \xrightarrow{a.s.} 0$. By fixing ω and using the previous result:

$$\{\omega : \sum_{n=1}^{\infty} \frac{X_n(\omega)}{n} < \infty\} \subset \{\omega : \frac{1}{n}\sum_{i=1}^{n} X_i(\omega) \to 0\}$$

$$\implies 1 = \mathbb{P}\{\omega : \sum_{n=1}^{\infty} \frac{X_n(\omega)}{n} < \infty\} \leq \mathbb{P}\{\omega : \frac{1}{n}\sum_{i=1}^{n} X_i(\omega) \to 0\}$$

where the left hand side is equal to 1 by what the theorem tells us to assume. Then, it must be that

$$\mathbb{P}\{\omega : \frac{1}{n}\sum_{i=1}^{n} X_i(\omega) \to 0\} = 1$$

But this is nothing more than $\frac{1}{n}\sum_{i=1}^{n} X_i(\omega) \xrightarrow{a.s.} 0$. This is called a Kronecker technique that allow us to transform the property of real numbers to an almost sure result. But what happens if the sample average converges to something different to zero? Now we have to show that $\sum_{n=1}^{\infty} \frac{X_n(\omega)}{n} < \infty$. But we will try to show this for a constructed r.v.: Let $Z_n = \frac{X_n - \mathbb{E}X_n}{n}$ and show that $\sum_{n=1}^{\infty} Z_n < \infty$. For this we will use a cauchy series argument. But first lets start by studying its variance:

$$\mathbb{V}[Z_n] = \frac{1}{n^2}\mathbb{V}[X_n]$$

Then define the partial sums as:

$$S_j = \sum_{k=1}^{j} Z_k$$

Then,

$$\mathbb{P}\{\max_{n\leq j\leq m} |S_j - S_n| > \epsilon\} = \mathbb{P}\{\max_{n\leq j\leq m} |\sum_{j=n+1}^{j} Z_j| > \epsilon\}$$

Then by the Kolmogorov inequality we have that:

$$\leq \frac{1}{\epsilon^2} \sum_{j=n+1}^{m} \mathbb{V}[Z_j]$$

But observe

$$\sum_{n=1}^{\infty} \mathbb{V}[Z_n] = \sum_{n=1}^{\infty} \frac{1}{n^2} \mathbb{V}[X_n] = \mathbb{V}[X_1] \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$$

where the last equality holds by the i.i.d. assumption. If I am not mistaken one can show that $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$.

Returning to our proof we have that

$$... \leq \frac{1}{\epsilon^2} \sum_{j=n+1}^{m} \mathbb{V}[Z_j] \leq \frac{1}{\epsilon^2} \sum_{j=n+1}^{\infty} \mathbb{V}[Z_j]$$

we have bounded it with a tail of a series that converges! This means that the

$$\limsup_{n \to \infty} \mathbb{P}\{\max_{n \leq j \leq m} |S_j - S_n| > \epsilon\} = 0$$

This actually shows that the partial sums are cauchy. But we can make this more rigorous in case you don't see it.

We have shown that $\forall \epsilon$, $\mathbb{P}\{\sup_{j \geq n} |S_j - S_n| > \epsilon\} \xrightarrow{n \to \infty} 0$.

$$\mathbb{P}\{\sup_{j \geq n} |S_j - S_n| > \epsilon\} = 1 - \mathbb{P}\{\sup_{j \geq n} |S_j - S_n| \leq \epsilon\} = 1 - \mathbb{P}(\bigcap_{j=n}^{\infty} \{|S_j - S_n| \leq \epsilon\})$$

But because what is inside is a monotonic increasing sequence (i.e. converges to the union) then by the continuity of probability measure

$$\forall \epsilon, \lim_{n \to \infty} \mathbb{P}(\bigcap_{j=n}^{\infty} \{|S_j - S_n| \leq \epsilon\} = \mathbb{P}(\bigcup_{n=1}^{\infty} \bigcap_{j=n}^{\infty} \{|S_j - S_n| \leq \epsilon\})$$

Then, we can conclude that $(S_n)_{n \in \mathbb{N}}$ is almost surely a cauchy sequence. Then, $S_n \xrightarrow{a.s.} 0$. Which in turn implies that

$$\sum_{n=1}^{\infty} \frac{X_n - \mathbb{E}X_n}{n} < \infty$$

$$\implies \frac{1}{n} \sum_{n=1}^{\infty} (X_n - \mathbb{E}X_n) \xrightarrow{a.s.} 0 \implies \frac{1}{n} \sum_{n=1}^{\infty} X_n \xrightarrow{a.s.} \mathbb{E}X_n$$

This is the end of the proof when the variance is finite. How will it continue when it is not? Exercise.

We will now start studying our third convergence concept.

■

4.4 Convergence in r-mean

Definition 4.5. (Convergence in r-mean) Let $r \in \mathbb{N}$ and let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables such that $\mathbb{E}|X_n|^r < \infty$. $(X_n)_{n\in\mathbb{N}}$ is said to converge in r-mean to a random variable X, denoted by $X_n \xrightarrow{r} X$ iff

$$\lim_{n\to\infty} \mathbb{E}|X_n - X|^r = 0$$

Obs: If $r = 2$ the converge in r-mean is known as convergence in mean-squared.
We are now going to study the same properties as before:

Theorem 4.21. *$X_n \xrightarrow{r} X$ and $X_n \xrightarrow{r} Y$ then $X = Y$ a.s.*

This is an interesting proof in its own right since its almost a generalization of the triangle inequality, "the two r inequality".

Proof. We start by trying to exploiting the triangle inequality somehow:

$$|X - Y|^r = |X - X_n + X_n - Y|^r \leq (|X - X_n| + |X_n - Y|)^r$$

But we now have everything to the power r so we are gonna use the a small trick:
Imagine we have two positive numbers a and b. Then,

$$a + b \leq \max\{a + b\} + b \leq 2\max\{a + b\} \leq 2(a + b)$$

We will use this to go from the first term to the third term and from the third term to go to last during this proof.

$$\begin{aligned} ...\ (|X - X_n| + |X_n - Y|)^r &\leq (2\max\{|X - X_n| + |X_n - Y|\})^r \\ &= 2^r \max(|X - X_n|^r, |X_n - Y|^r) \\ &\leq 2^r(|X - X_n|^r + |X_n - Y|^r) \end{aligned}$$

Then we can conclude that

$$\begin{aligned} \mathbb{E}[|X - Y|^r] &\leq \mathbb{E}[2^r(|X - X_n|^r + |X_n - Y|^r)] \\ &= 2^r(\mathbb{E}[|X - X_n|^r] + \mathbb{E}[|X_n - Y|^r] \end{aligned}$$

but when we take the limit as n goes to infinity the right hand side convergence to zero by convergence in r-mean. The left hand side doesn't depend on n. We can conclude that $\mathbb{E}[|X-Y|^r]=0$. If the area under the curve (i.e. expectation) is exactly equal to zero it must be that $|X-Y|=0$ almost surely (does not imply that $|X-Y|=0$). Since we are taking an expectation which is computed using a determined measure (most probably the lebesgue measure) we can actually ignore points that have measure zero. Just imagine a function that takes the value of zero almost everywhere except in three specific points. Take the expectation with respect to the lebesgue measure. This measure assigns a set its interval length. For a point this is zero. Thus, we can conclude that $X=Y$ a.s. ■

Theorem 4.22. *$X_n \xrightarrow{r} X$ then $X_n \xrightarrow{p} X$.*

Proof. We start by using Markov's inequality applied to the function $|X|^r$ which is positive and strictly increasing:

$$\mathbb{P}\{|X_n - X| > \epsilon\} \leq \frac{\mathbb{E}|X_n - X|^r}{\epsilon^r}$$

The numerator converges to zero by r-mean. So we can take $\limsup$ on both sides and conclude that $X_n \xrightarrow{p} X$. ■

What about the relationship between r-mean and a.s. convergence? In generally, there isn't. We will require to introduce another concept called uniform integrablity and if it holds then it will imply that a.s. and r-mean convergence are equivalent. For now we will study some counterexamples:

Counterexample 1: Take

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

Then, clearly

$$\mathbb{E}[X_n] = (1-\frac{1}{n})0 + n\frac{1}{n} = 1 \; \forall n$$

But, since $\mathbb{P}\{\omega : X_n > \epsilon\} = \frac{1}{n}$ we arrive at $X_n \xrightarrow{p} 0$. Nonetheless,

$$\mathbb{E}|X_n - 0|^r = \mathbb{E}[X_n] \xrightarrow{n\to\infty} 1$$

So X_n does not converge in 1-mean to zero.

Counterexample 2: Take

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

By very similar arguments one can easily show that $X_n \xrightarrow{p} 0$ and that X_n does not converge in 2-mean to zero.

Definition 4.6. A sequence of random variables $(X_n)_{n\in\mathbb{N}}$ is uniformly integrable iff

$$\lim_{c\to\infty} \mathbb{E}|X_n|\mathbb{1}\{|X_n| > c\} = 0 \text{ uniformly in n}$$

which is equivalent to

$$\sup_n (\lim_{c\to\infty} \mathbb{E}|X_n|\mathbb{1}\{|X_n| > c\}) = 0$$

This is a hard definition to digest the first time you see it. So let me provide some intuition. If we know that for each n, $|X_n|$ is integrable (i.e. $\mathbb{E}|X_n| < \infty$ then the limit will go to zero for all n. When we add the sup we are requiring that each one of those terms go to zero at the same rate. It is a very strong concept.

Definition 4.7. A sequence of random variables $(X_n)_{n\in\mathbb{N}}$ is uniformly integrable or order r iff

$$\sup_n (\lim_{c\to\infty} \mathbb{E}|X_n|^r\mathbb{1}\{|X_n| > c\}) = 0$$

Definition 4.8. If $(X_n)_{n\in\mathbb{N}}$ is uniformly integrable or order 0 then $(X_n)_{n\in\mathbb{N}}$ is called tight.

In a sense, we are requiring that all the random variables tails go to zero at the same speed.

Theorem 4.23. *A sequence of random variables $(X_n)_{n\in\mathbb{N}}$ is uniformly integrable iff*

1. $\sup_n \mathbb{E}|X_n| < \infty$

2.$\forall\epsilon,\ \exists\delta > 0$ *s.t.* $\forall A$ *with* $\mathbb{P} < \delta,\ \mathbb{E}|X_n|\mathbb{1}\{A\} < \epsilon$ *uniformly in n.*

Proof. First lets try to prove the sufficiency part of the proof. That is uniform integrability implies 1 and 2. Once again we will divide the support into two areas, one that is bounded below c and the other above c.

$$\sup_n \mathbb{E}|X_n| = \sup_n \mathbb{E}|X_n|(\mathbb{1}\{X_n \geq c\} + \mathbb{1}\{X_n < c\})$$

$$= \sup_n \mathbb{E}|X_n|\mathbb{1}\{X_n \geq c\} + \sup_n \mathbb{E}|X_n|\mathbb{1}\{X_n < c\}$$

Here if we choose c large enough we can make the first term as small as possible uniformly in n (say smaller than ϵ). We also use the fact that $X_n < c$ in the second term. Hence, we arrive at,

$$= \sup_n \mathbb{E}|X_n|\mathbb{1}\{X_n \geq c\} + \sup_n \mathbb{E}|X_n|\mathbb{1}\{X_n < c\} \leq \epsilon + \sup_n c\mathbb{E}\mathbb{1}\{X_n < c\}$$

$$\leq \epsilon + c < \infty$$

since $\mathbb{E}\mathbb{1}\{X_n < c\}$ is just a probability and hence is between 0 and 1. This proves 1.

For proving 2 we will still split the support into two areas!

$$\mathbb{E}|X_n|\mathbb{1}\{A\} = \mathbb{E}|X_n|(\mathbb{1}\{A \bigcap \{|X_n| \geq c\}\} + \mathbb{1}\{A \bigcap \{|X_n| < c\}\})$$

$$= \mathbb{E}|X_n|\mathbb{1}\{A \bigcap \{|X_n| \geq c\}\} + \mathbb{E}|X_n|\mathbb{1}\{A \bigcap \{|X_n| < c\}\} \leq \mathbb{E}|X_n|\mathbb{1}\{|X_n| \geq c\} + c\mathbb{P}(A)$$

where the first term is obvious since we are just taking one part of the intersection and the second term follows from the same logic of what was argued before to prove that 1. holds. Now we need to make the sum of this terms small enough. The first term can be made less than $\frac{\epsilon}{2}$ if we choose a c big enough. Then, to make the second term small enough, we need to make the $c\mathbb{P}(A) < c\frac{\epsilon}{2c}$. So let $\delta = \frac{\epsilon}{2c(\epsilon)}$. And we are done.

Now we need to proof necessity: 1 and 2 imply uniform integrability. To do this we will be using Markov's inequality. We have by point 2 that $\forall \epsilon$, $\exists \delta > 0$ s.t. $\forall A$ with $\mathbb{P} < \delta$, $\sup_n \mathbb{E}|X_n|\mathbb{1}\{A\} < \epsilon$. Now take $A = \{|X_n| > c\}$. Then, by Markov's inequality

$$\mathbb{P}(A) \leq \frac{\mathbb{E}|X_n|}{c} \leq \frac{\sup \mathbb{E}|X_n|}{c}$$

Choose $\delta = \frac{\sup \mathbb{E}|X_n|}{c}$.
Then, by Markov's inequality with function absolute value:

$$\sup_n \mathbb{P}\{|X_n| > c\} \leq \frac{1}{c} \sup_n \mathbb{E}|X_n|$$

Notice that the term on the right will go to zero by point 1 if we let c go to infinity. It assures as that the expectation is finite. But then this just means that $(X_n)_{n\in\mathbb{N}}$ is uniformly integrable. ∎

Theorem 4.24. *If $X_n \xrightarrow{a.s.} X$ and $(X_n)_{n\in\mathbb{N}}$ is uniformly integrable then $\mathbb{E}[X_n] \to \mathbb{E}[X]$ and $X_n \xrightarrow{1} X$.*

This order can be generalized to uniform integrability of order r. But to make the proof simpler I decided not to since I am gonna proof the stronger option $\mathbb{E}[X_n] \to \mathbb{E}[X]$. That is, the moment converges. This also applies to higher moments, we would just need to require uniform integrability of the same order.

Proof. **Step 1:** For this statement to make sense we require that $\mathbb{E}[X_n]$ and $\mathbb{E}[X]$ exists. Therefore, step 1 will consist in showing that $\mathbb{E}[X_n]$ exists. Since X_n is uniformly integrable we know that $\sup_n \mathbb{E}|X_n| < \infty$. Hence, each of them is finite since the $\mathbb{E}|X_n| < \infty \implies \mathbb{E}[X_n] < \infty$. Why? I recommend doing a proof by picture for intuition! This proofs the required statement of this step.

Step 2: Will consist in showing that $\mathbb{E}[X] < \infty$. This is a bit harder because we are studying the behavior at the limit.

$$\begin{aligned}
\mathbb{E}|X| &= \mathbb{E}[\lim_{n\to\infty} |X_n|] && \text{(Almost sure convergence)} \\
&= \mathbb{E}[\liminf_{n\to\infty} |X_n|] && \text{(Since the limit exists)} \\
&\leq \liminf_{n\to\infty} \mathbb{E}[|X_n|] && \text{(By Fatou's lemma)} \\
&\leq \sup_n \mathbb{E}|X_n| < \infty && \text{(By Uniform Integrability)}
\end{aligned}$$

Hence, $\mathbb{E}|X| < \infty \implies \mathbb{E}[X] < \infty$.

Step 3: Here we start the main part of the proof. We will proof convergence in 1-mean. That is, we want to proof that $\mathbb{E}|X_n - X| \xrightarrow{n\to\infty} 0$. Lets define $Y_n = |X_n - X|$. Then, $Y_n \xrightarrow{a.s.} 0$. Now observe the following inequality which is obtained through the use of the triangle inequality:

$$Y_n = |Y_n| \leq |X_n| + |X| \leq 2\max\{|X_n|, |X|\} \leq 2(|X_n| + |X|)$$

Then, continuing with the proof. Observe the following result:

$$\mathbb{E}[Y_n] = \mathbb{E}[Y_n]\mathbb{1}(Y_n > c) + \mathbb{E}[Y_n]\mathbb{1}(Y_n \leq c)$$

Since only one of the indicator functions will hold at some point c we can decompose it in these two terms.

Our first job will consist in making $\mathbb{E}[Y_n]\mathbb{1}(Y_n > c) \xrightarrow{n\to\infty} 0$.

$$\mathbb{E}[Y_n]\mathbb{1}(Y_n > c) \leq 2\mathbb{E}[\max\{|X_n|, |X|\}]\mathbb{1}(2\max\{|X_n|, |X|\} > \frac{c}{2})$$

where the inequality follows by the inequality that we showed some lines ago. But we can continue using the last part of that inequality to obtain:

$$\leq 2|X_n|\mathbb{1}(\max\{|X_n|, |X|\} > \frac{c}{4}) + 2|X|\mathbb{1}(\max\{|X|, |X|\} > \frac{c}{4})$$

This inequality holds since

$$\text{If } \max\{|X_n|, |X|\} > a \text{ then either } |X_n| > \frac{a}{2} \text{ or } |X| > \frac{a}{2}$$

This shows that $\mathbb{E}[Y_n]\mathbb{1}(Y_n > c) \xrightarrow{n\to\infty} 0$ since we can choose c big enough to make both terms go to zero i.e. for any given ϵ choose a c big enough to make both terms less than ϵ.

Now we want to show that $\mathbb{E}[Y_n]\mathbb{1}(Y_n \leq c) \xrightarrow{n\to\infty} 0$

$$\mathbb{E}[Y_n]\mathbb{1}(Y_n \leq c) \leq \mathbb{E}[c]\mathbb{1}(Y_n \leq c)$$

Using the Dominated Convergence Theorem

Step 4: We want to proof that $\mathbb{E}[X_n] \to \mathbb{E}[X]$

■

5 An introduction to Time-series

6 An introduction to Combinatorics