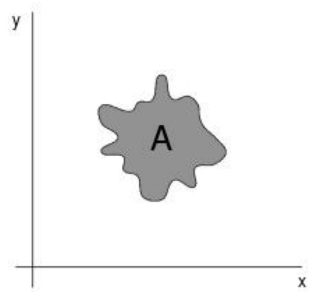
Almost Sure Convergence of a Sequence of Random Variables

(...for people who haven't had measure theory.)

1 Preliminaries

1.1 The "Measure" of a Set (Informal)

Consider the set $A \subseteq \mathbb{R}^2$ as depicted below.



How can we measure the "size" of this set?

The most intuitive answer might be to give the area of the set. This is the **Lebesgue measure** of the set A and will be denoted by $\lambda(A)$.

This is generalizable to different dimensions:

- If A is an interval in $\mathbb{R}$, $\lambda(A) =$ the length of A.
- If A is a subset of $\mathbb{R}^2$, $\lambda(A) =$ the area of A.
- If A is a subset of $\mathbb{R}^3$, $\lambda(A) =$ the volume of A.
- If A is a subset of $\mathbb{R}^n$, $\lambda(A) =$ the n-dimensional volume of A.

We know many things about Lebesgue measure. Just as a couple of examples, we know that:

- A and B disjoint $\Rightarrow \lambda(A \cup B) = \lambda(A) + \lambda(B)$

and

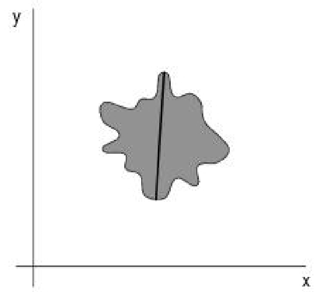
- $A \subseteq B \Rightarrow \lambda(A) \leq \lambda(B)$

Another way that we might consider "measuring" the set A above is to compute the **diameter** of A:

$$diam(A) := \sup_{a_1, a_2 \in A} d(a_1, a_2)$$

where $d(x, y)$ is the usual Euclidean distance between the points x and y.

This is the length of the line depicted as follows.



(Yes, it was a close call with the more horizontal looking line!)

This "measure" has the property that

- $A \subseteq B \Rightarrow \lambda(A) \leq \lambda(B)$

but does **not** have the property that

- A and B disjoint $\Rightarrow diam(A \cup B) = diam(A) + diam(B)$

This is sad and should not be allowed... We are going to formally define what a measure of a set should be. We will see some notation/machinery

1.2 Measurable Spaces

Let Ω be any non-empty set. It may be that $\Omega = \mathbb{R}^2$ but it may also be that Ω is not even numerical but rather a collection of objects!

Definition: Let $\mathcal{F}$ denote a collection of subsets of Ω. We say that $\mathcal{F}$ is a **σ-field** (equivalently a σ-algebra) on Ω if

(i) $\Omega \in \mathcal{F}$

(ii) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$ (Here, A^c is the complement of A in Ω.)

(iii) $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$

(Note that by (ii), we could replace (i) with $\emptyset \in \mathcal{F}$.)

(Aside: If (iii) only hold for finite unions of sets in $\mathcal{F}$, $\mathcal{F}$ is called a **field** or **algebra** without the "σ".)

We will define a function on the sets in $\mathcal{F}$ to measure the sets in $\mathcal{F}$. For this reason, the sets in $\mathcal{F}$ are called "measurable sets".

Taken together, $(\Omega, \mathcal{F})$ is called a **measurable space**.

Example:

Suppose that $\Omega = \mathbb{R}$. Let $\mathcal{A}$ be the collection of all intervals of the form (a, b) for $a < b$.

$\mathcal{A}$ is **not** a σ-field. For example,

$$(a, b)^c = (-\infty, a] \cup [b, \infty)$$

is not in A.

We could, however, "generate" a σ-field from $\mathcal{A}$ by building up a collection of subsets of $\mathbb{R}$ that

- contains all sets in $\mathcal{A}$
- contains all of the other sets we need to "throw in" in order to make a σ-field

The **Borel sets** on $\mathbb{R}$ is the σ-field on $\mathbb{R}$ that is generated by all open intervals of the form (a, b) with $a < b$. This collection is usually called $\mathcal{B}$ or $\mathcal{B}(\mathbb{R})$.

$(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a measurable space. □

1.3 Defining a Measure

Definition: Let $(\Omega, \mathcal{F})$ be a measurable space. A **measure** on this space is a function $\mu : \mathcal{F} \to [0, \infty]$ such that

(i) $\mu(\emptyset) = 0$

(ii) If $A_1, A_2, \ldots \in \mathcal{F}$ are disjoint, then

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

A measure will have many other properties as well. For example,

- (ii) holds for finite unions of disjoint sets in $\mathcal{F}$ because we can take $A_n = \emptyset$ for all n after some point.
- $A \subseteq B \Rightarrow \mu(A) \leq \mu(B)$ since

  $$A \subseteq B \qquad \Rightarrow \qquad B = A \cup (B \setminus A)$$

  and A and $B \setminus A$ are disjoint so

  $$\mu(B) = \mu(A \cup (B \setminus A)) \overset{disjoint}{=} \mu(A) + \underbrace{\mu(B \setminus A)}_{\geq 0} \geq \mu(A)$$

  ($\mu(B \setminus A) \geq 0$ since $\mu : \mathcal{F} \to [0, \infty]$.)

Definition: Let $(\Omega, \mathcal{F})$ be a measurable space. Let μ be a measure on $\mathcal{F}$. The then triple $(\Omega, \mathcal{F}, \mu)$ is called a **measure space**.

Example:

Lebesgue measure λ on $\mathbb{R}$ is a measure on $\mathbb{R}$, so $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ is a measure space.

We know that, for $a < b$, $\lambda((a, b)) = b - a$ and we can extend our definition of λ to other sets using the definition of a measure. For example, for disjoint intervals (a, b) and (c, d), the Lebesgue measure of the union would be the sum of the lengths of both intervals. While many extensions to other sets in $\mathcal{B}(\mathbb{R})$ is obvious, extensions to "weirder" sets on $\mathbb{R}$ may not be so obvious. For example, $\lambda([a, b)) = b-a$ as well but we would have to prove this by writing $[a, b) = \cap_{n=1}^{\infty}(a-1/n, b)$ and discussing how measures go through limits. There is a much longer discussion to be had here but since we are not in a measure theory course I will move on!

1.4 A Probability Measure

Let $(\Omega, \mathcal{F})$ be a measurable space. A probability measure on this space is a measure with the additional restriction that $P(\Omega) = 1$.

We will usually denote a probability measure with a capital P.

Definition: Let $(\Omega, \mathcal{F})$ be a measurable space. A **probability measure** on this space is a function $P : \mathcal{F} \to [0, \infty]$ such that

(i) $P(\emptyset) = 0$

(ii) If $A_1, A_2, \ldots \in \mathcal{F}$ are disjoint, then

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

(iii) $P(\Omega) = 1$.

Note then that $P : \mathcal{F} \to [0, 1]$.

Note also that a probability measure acts like probability as we know it. For example, take $A \in \mathcal{A}$. Then A^c is necessarily in $\mathcal{F}$ by definition of a σ-field, and

$$\Omega = A \cup A^c \qquad \Rightarrow \qquad P(\Omega) = P(A \cup A^c) \overset{disjoint}{=} P(A) + P(A^c)$$

but $P(\Omega) = 1$. So, $P(A^c) = 1 - P(A)$.

When defining a probability measure, one usually thinks of Ω as the set of all possible outcomes of an experiment that involves some randomness. (This is also known as the **sample space** of the experiment.)

For example, consider an unfair coin that has probability 3/4 of coming up "Heads". Flip the coin twice.

The sample space for this experiment is defined as

$$\Omega = \{HH, HT, TH, TT\}.$$

Subsets of Ω define "events". For example, if we let A be the event that we saw at least one Heads, then

$$A = \{HH, HT, TH\} \subseteq \Omega.$$

A σ-field on Ω is a collection of subsets of Ω (with formal properties of a σ-field). In this probability context, we can think of it as a collection of events.

Here are three different examples of σ-fields on Ω.

- $\mathcal{F}_1$ = the power set of Ω (the set of all possible subsets)
- $\mathcal{F}_2 = \{\emptyset, \{TT\}, \{HH, HT, TH\}, \Omega\}$

  (This is the σ-field "generated by the event" $\{TT\}$ since we have included $\{TT\}$ and everything else required to have a σ-field.)
- $\mathcal{F}_3 = \{\emptyset, \Omega\}$

  (This is called the "trivial σ-field".)

Let's let $\mathcal{F}$ be the power set of $\Omega = \{HH, HT, TH, TT\}$. We can define a probability measure P on the measurable space $(\Omega, \mathcal{F})$ using our "real world idea of probability". For example, we could define

$$\begin{aligned} P(\{HH\}) &= Prob(\text{ getting the outcome } HH \text{ }) \\ &= \tfrac{3}{4} \cdot \tfrac{1}{4} = \tfrac{3}{16} \end{aligned}$$

and all other probabilities like

$$P(\{HH, HT, TH\}) = Prob(\text{ getting at least one Head })$$

accordingly.

One can verify that this would produce a valid probability measure.

2 Random Variables

Let $(\Omega, \mathcal{F}, P)$ be a probability space.

A random variable X will be a function from Ω into $\mathbb{R}$. (This is not a complete definition though!)

For example, in our coin tossing example, with $\Omega = \{HH, HT, TH, TT\}$ and $\mathcal{F}$ being the power set of Ω, we could define X to be the number of Heads observed. Then

$$X(\omega) = \begin{cases} 2 & , \quad \omega = HH \\ 1 & , \quad \omega = HT \text{ or } \omega = TH \\ 0 & , \quad \omega = TT \end{cases}$$

Note that we have already assigned a probability measure to all events (sets) in $\mathcal{F}$. We can then translate these to probabilities for X.

For example, the "event" that $X = 2$ is equivalent to the event that we observed HH. The set $\{HH\}$ has already been "measured" to be 3/16.

$$Prob(X = 2) := P(\{\omega : X(\omega) = 2\}) = P(\{HH\}) = \frac{3}{16}.$$

Suppose that we replace the σ-field with $\mathcal{F}_2 = \{\emptyset, \{TT\}, \{HH, HT, TH\}, \Omega\}$ and have only defined P on the sets in $\mathcal{F}_2$. Then we would not be able to figure out some probabilities for X using the fact that we have already assigned probability measure values to all sets in $\mathcal{F}_1$ since, for example,

$$\{\omega : X(\omega) = 2\} = \{HH\}$$

but $\{HH\} \notin \mathcal{F}_2$ so $P(\{HH\})$ has not been assigned!

To combat this problem, we require that, for a given probability space $(\Omega, \mathcal{F}, P)$, a function $X : \Omega \to \mathrm{I\!R}$ is a random variable on the space if and only if "all probabilities involving X" are defined in the sense that all events involving X correspond to sets (events) in $\mathcal{F}$. (Recall that $\mathcal{F}$ consists of the sets for which P is defined.)

Formally, we require that $X : \Omega \to \mathrm{I\!R}$ be a "measurable function".

Definition: Let $(\Omega, \mathcal{F})$ be a measurable space. $X : \Omega \to \mathrm{I\!R}$ is a **measurable function** if and only if

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

for all $B \in \mathcal{B}(R)$.

We are now really to formally define a random variable!

Definition: Let $(\Omega, \mathcal{F}, P)$ be a probability space. A **random variable** is a measurable function $X : \Omega \to \mathrm{I\!R}$.

Note that the randomness for X comes from the randomness of the experiment resulting in the outcomes in Ω. In our coin tossing experiment, X, the number of heads, is a random variable but $X(HH)$ is not random– it is 2.

Example:

Let $\Omega = [0, 1]$. Let $\mathcal{F}$ be an "appropriate" σ-field on Ω. (It is not really important for our purposes.) Let P be Lebesgue measure. (Note that $P([0, 1]) = \lambda([0, 1]) = 1 - 0 = 0$ so, for this Ω, Lebesgue measure is a probability measure!)

Define a random variable $X(\omega) = \omega$ for all $\omega \in \Omega$. Then, for any $(a, b) \subseteq [0, 1]$, $X^{-1}((a, b)) = (a, b)$ and so

$$Prob(a < X < b) = P(\{\omega : a < X(\omega) < b\}) = P((a, b)) = \lambda((a, b)) = b - a.$$

This is how we define the $unif[0, 1]$ random variable in terms of measure theory!

We are ready to define the almost sure convergence of a sequence of random variables!

3 Almost Sure Convergence

Let $(\Omega, \mathcal{F}, P)$ be a probability space.

Let $X_1, X_2, \ldots$ be a sequence of random variables defined on this one common probability space.

Note that, for fixed $\omega \in \Omega$, $X_1(\omega), X_2(\omega), \ldots$ is a sequence of real numbers. We know what it means to take a limit of a sequence of real numbers.

Definition: Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $X_1, X_2, \ldots$ be a sequence of random variables on $(\Omega, \mathcal{F}, P)$. Let X be another random variable on $(\Omega, \mathcal{F}, P)$. We say that X_n **converges almost surely** (or, **with probability 1**) to X if

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

In this case, we write

$$X_n \stackrel{a.s.}{\to} X.$$

This is a really strong type of convergence for random variables in the sense that

$$X_n \stackrel{a.s.}{\to} X \quad \Rightarrow \quad X_n \stackrel{P}{\to} X \quad \Rightarrow \quad X_n \stackrel{d}{\to} X.$$

In order to prove this, we need more measure theory. Take some measure theory!

Example:

Let $\Omega = [0, 1]$. Let $\mathcal{F}$ be an "appropriate" σ-field on Ω. Let P be Lebesgue measure.

Define a sequence of random variables $\{X_n\}$ on $(\Omega, \mathcal{F})$, with $X_n(\omega) = \omega + \omega^n$ for all $\omega \in \Omega$ and for all $n = 1, 2, \ldots$.

We will show that $X_n \stackrel{a.s.}{\to} X$ where $X \sim unif[0, 1]$.

Recall that we can define X as the function $X(\omega) = \omega$.

Note that, for every $\omega \in [0, 1)$, $\omega^n \to 0$ as $n \to \infty$.

Thus, we have that $X_n(w) = \omega + \omega^n \to \omega = X(\omega)$ for $\omega \in [0, 1)$.

Note, however, that

$$X_n(1) = 2 \not\to 1 = X(1).$$

So, we do not have convergence of $X_n(w)$ to $X(\omega)$ for all $\omega \in \Omega[0, 1]$. However, we do have it for "almost all" of them. Indeed,

$$\lim_{n\to\infty} P(\{\omega : X_n(\omega) = X(\omega)\}) = P([0, 1)) = 1 - 0 = 1.$$

Example of $\mathbf{X_n \stackrel{P}{\to} X \not\Rightarrow X_n \stackrel{a.s.}{\to} X}$: Consider the same space and probability measure of the previous example.

Define a sequence of random variables $\{X_n\}$ on $(\Omega, \mathcal{F})$, with $X_n(\omega) = \omega + \omega^n$ as

$$\begin{aligned}
X_1(\omega) &= \omega + I_{[0,1]}(\omega) \\
X_2(\omega) &= \omega + I_{[0,1/2]}(\omega) \\
X_3(\omega) &= \omega + I_{[1/2,1]}(\omega) \\
X_4(\omega) &= \omega + I_{[0,1/3]}(\omega) \\
X_5(\omega) &= \omega + I_{[1/3,2/3]}(\omega) \\
X_6(\omega) &= \omega + I_{[2/3,1]}(\omega) \\
&\vdots
\end{aligned}$$

Let $X(\omega) = \omega$.

Then one can show that $X_n \xrightarrow{P} X$ since, for any $\varepsilon > 0$,

$$Prob(|X_n - X| > \varepsilon) = P(\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\})$$

is equal to the length of an interval whose length is going to zero.

On the other hand, the sequence $\{X_n(\omega)\}$ does not converge for any $\omega \in [0,1]$ at all since, for any fixed ω, $\{X_n(\omega)\}$ keeps alternating between ω and $\omega + 1$.

Thus,

$$\lim_{n\to\infty} P(\{\omega : X_n(\omega) = X(\omega)\}) = \lim_{n\to\infty} P(\{\omega : X_n(\omega) = \omega\}) = \lim_{n\to\infty} P(\emptyset) = 0 \neq 1.$$