Intermediate Probability

Madhusoodan Gunasingam

University of Toronto

September 23, 2024

Contents

1 Background

1.1 Set Theory

A subset $A \subset \mathbb{R}$ is said to be bounded from above (resp. below) if there exists $M \in \mathbb{R}$ such that $x \leq M$ (resp. $x \geq M$) for every $x \in A$. Such a quantity is called an *upper bound* (resp. *lower bound*). Clearly if $M' \geq M$, then M' is also an upper bound and thus upper bounds are not unique for sets that are bounded above. If A is bounded from both above and below we say it is bounded, equivalently, there exists $M \in \mathbb{R}$ such that $|x| \leq M$ for every $x \in A$,

Theorem 1. (Completeness of $\mathbb{R}$) If $A \subset \mathbb{R}$ is bounded above (resp. below) then there exists a least upper bound (resp. greatest lower bound). That is, there exists a unique upper (resp. lower) bound that is smaller (resp. larger) than or equal to any other upper (resp. lower) bound.

The least upper bound of a set A, bounded from above, is denoted $\sup A$ and is denoted as the *supremum* of A. Similarly, the greatest lower bounded of a set A, bounded from below. is denoted $\inf A$ and is denoted the *infimum* is A. We may extend the definition of supremum and infimum of unbounded sets by setting $\sup A := \infty$ and $\inf A := -\infty$ when A is respectively unbounded above or unbounded below. With this, observe that for any set A, the quantities $\sup A$ and $\inf A$ always exist and are unique.

Sets can be finite or infinite, and in the case of the latter it can be countably infinite or uncountable. We say that a set is *countably infinite* if it's elements can be ordered in a sequence. Heuristically, looking at one element, one can immediately deduce the next element and so on.

in this section we explore the different types of sample spaces we will encounter. Namely we will study the various sizes of different sample spaces, this notion is referred to as *cardinality.* A set is *finite* if it has finitely many elements (as expected) and a set is infinite if it is not finite (also as expected).

Proposition 1. Suppose Ω is a finite set with n elements. Then there are exactly 2^n different possible subsets of Ω.

Proof. Suppose $A \subset \Omega$. Observe that for all $\omega \in \Omega$, exactly one of $\omega \in A$ or $\omega \notin \Omega$ is true. That is, we may provide a complete description of A by specifying exactly which elements of Ω are in A and which are not in A. We will proceed by induction. If $n = 0$, then we have that $\Omega = \varnothing$. The only subset of the empty set is the empty set itself and thus there is only 1 (or 2^0) subset of Ω, verifying the base case. Now we have to show that if sets of size n have 2^n different subsets, then it follows that sets of size $n+1$ have 2^{n+1} different subsets. Suppose Ω has $n+1$ elements. Fix $\omega \in \Omega$ and let us remove this element from the set, resulting in a new set $A := \Omega \backslash \{\omega\}$. Observe that this new set A has n elements and thus by our inductive hypothesis, there are exactly 2^n subsets of A. In other words, there are 2^n possible subsets of Ω that exclude the particular element ω. By adding the element ω to each of these we will produce 2^n new subsets of Ω. Since every subset of Ω either contains the particular ω or does not contain ω, we are certain we have identified all possible subsets. Thus, the total number of subsets of Ω is given by $2^n + 2^n = 2^{n+1}$. By induction, the claim holds for all $n \in \mathbb{N}$. □

Motivated by the above, we define the *power set* of Ω to be the collection of all subsets of Ω and is denoted by 2^Ω. Likewise when Ω is an infinite set, then it also has infinitely many subsets. Unfortunately, not all infinities are created equal and some infinities are greater than other infinities (in fact, there are more distinct infinities then there are elements in any infinite set). A set with infinitely many elements is called *countably infinite* if it's elements can be arranged in a sequence.

Example 1. The most natural example of the a countable set is the collection of natural numbers. It is automatically ordered for us with $\mathbb{N} := \{1, 2, 3, \dots\}$.

Example 2. Another countable set is the collection of integers, $\mathbb{Z}$. Observe that the standard order of $\mathbb{Z}$ has no beginning as this set is unbounded below. However we can instead impose an order similar to the natural numbers by interlacing the naturals with the negative naturals and then add in the element 0. This order appear as

$$0, -1, 1, -2, 2, -3, 3, \dots$$

Example 3. A more complicated example of a countable set is the collection of rational numbers. That is the set of all real numbers that can be expressed as a ratio of two integers (such as 1 or $-2/9$). Recall that the rational numbers are dense in the real line and so every interval (regardless of how small it is) must contain infinitely many rational numbers. With this, given a rational number it is not clear what the "next" rational number is that succeeds it. Their placement on the real number line does not reveal a way to order them in a sequence. For simplicity, let us order the positive rationals first and then appeal to the interlacing method used for the integers.

With this, we may produce a table where the rows denote the possible numerators and the columns denote the possible denominators. Then we may linearly order this table by running through the diagonals that form going from the top-right to the bottom-left. The resulting order is as follows:

$$\frac{1}{1}, \frac{1}{2}, \frac{2}{1}, \frac{1}{3}, \frac{2}{2}, \frac{3}{1}, \frac{1}{4}, \frac{2}{3}, \frac{3}{2}, \frac{3}{1}, \frac{4}{1}, \frac{1}{5}, \frac{2}{4}, \frac{3}{3}, \frac{4}{2}, \frac{5}{1}, \frac{1}{6}, \ldots$$

There will of course be duplicates which we may remove that do no affect the ability to order our collection. Hence we have obtained an ordered sequence that captures every rational number. Therefore we conclude the rationals are indeed countable.

Proposition 2. Let A_1, A_2, A_3 all be countable sets. Then $\bigcup_{n=1}^{\infty} A_n$ is also countable. That is, the countable union of countable sets is countable.

Example 4. Our final example is the collection of real numbers, of which there are infinitely many and more than there are countable. Hence the real numbers cannot be ordered in a sequence. Let us consider just the interval $(0,1)$ for simplicity (which of course is a subset of $\mathbb{R}$). A property of the real numbers is that every real number $x \in (0,1)$ has a decimal representation (which may be infinite) given by $0.x_1x_2x_3\ldots$ where each of the x_j are digits from 0 to 9. Suppose all the real numbers in $(0,1)$ can be ordered in a sequence $(x_n)_{n=1}^{\infty}$ with respective decimal representation

$$\begin{aligned}
x_1 &= 0.x_{1,1}x_{1,2}x_{1,3}x_{1,4}\ldots \\
x_2 &= 0.x_{2,1}x_{2,2}x_{2,3}x_{2,4}\ldots \\
x_3 &= 0.x_{3,1}x_{3,2}x_{3,3}x_{3,4}\ldots \\
&\vdots
\end{aligned}$$

Now consider a new number $y \in (0,1)$ with decimal expansion $0.y_1, y_2y_3y_4\ldots$ constructed in such a way that $y_i \neq x_{i,i}$ (the digits identified in blue) for all $i \in \mathbb{N}$. Observe that due to the difference in digits found in y, we have that $y \neq x_n$ for all $n \in \mathbb{N}$. Thus, y is not a member of the sequence $(x_n)_{n=1}^{\infty}$. This contradiction shows that our assumption that the real numbers in $(0,1)$ can be ordered in a sequence cannot hold.

1.2 Limits

1.2.1 Sequences

Let $(a_n)_{n\geq 1}$ be a sequence of real numbers, we say that the sequence is increasing if $a_n \geq a_{n+1}$ for each $n \in \mathbb{N}$ and called decreasing if $a_n \leq a_{n+1}$ for each $n \in \mathbb{N}$. A sequence is bounded if the set $\Omega := \{a_n : n \geq 1\}$ is a bounded set. The quantity $a \in \mathbb{R}$ is a limit point of the sequence $(a_n)_{n\geq 1}$ if for all $\varepsilon > 0$ the interval $(a-\varepsilon, a+\varepsilon)$ contains infinitely many elements of the sequence. That is, a is a limit point if for all $\varepsilon > 0$ and $N \in \mathbb{N}$, there exists $n \geq N$ such that $a_n \in (a-\varepsilon, a+\varepsilon)$. Observe that the definition of a limit point extends naturally to include ∞ and $-\infty$ for unbounded sequences. Therefore, for any sequence $(a_n)_{n\geq 1}$, there always exists a limit point in $[-\infty, \infty]$. If a sequence has a unique limit point, then we call this quantity the limit of the sequence and denote it as $\lim\limits_{n\to\infty} a_n$. More formally, we say that a sequence $(a_n)_{n\geq 1}$ has a limit (or converges to) some $a \in \mathbb{R}$ if for all $\varepsilon > 0$, the interval $(a-\varepsilon, a+\varepsilon)$ contains all but finitely many elements of the sequence $(a_n)_{n\geq 1}$. Of course, this means that if a sequence is convergent then it is also bounded. We can characterize all the limit points of a sequence $(a_n)_{n\geq 1}$ with a respective sub-sequence $(a_{n_k})_{k\geq 1}$ that is convergent (to the respective limit point).

If the sequence $(a_n)_{n\geq 1}$ is increasing (resp. decreasing) then it is convergent to a limit equal to $\sup_{n\geq 1} a_n$ (resp. $\inf_{n\geq 1} a_n$) which includes the case where the sequence is unbounded.

Theorem 2. (Cauchy Criterion) Let $(a_n)_{n\geq 1}$ be a sequence of real numbers. For each $n \in \mathbb{N}$, define the quantity $b_n := \sup_{p,q} |a_{n+p} - a_{n+q}|$. Then the sequence $(a_n)_{n\geq 1}$ is convergent (to a finite number) if and only if $\lim\limits_{n\to\infty} b_n = 0$.

1.2.2 Limit Superior and Limit Inferior

Let $(a_n)_{n\geq 1}$ denote an arbitrary sequence of real numbers and define the sequence $A_m := \sup_{a_n : n\geq m}$ for each $m \in \mathbb{N}$. It is immediate that the sequence $(A_m)_{m\geq 1}$ is decreasing, and thus has a limit (which may be possibly infinite). We call the limit of this sequence as the *limit superior* of the sequence $(a_n)_{n\geq 1}$ and is denoted by:

$$\limsup_{n\to\ \infty} a_n := \lim_{n\to\infty} \sup_{m\geq n} a_m = \lim_{m\to\infty} A_m.$$

An anologous definition can be made for the quantity $\liminf_{n\to infty} a_n$, which is called the *limit inferior* of the sequence $(a_n)_{n\geq 1}$. Observe that both of these quantities exist always, and if they are equal to each other they are also equal to the limit of the sequence. That is, a sequence $(a_n)_{n\geq 1}$ is convergent if and only if $\limsup_{n\to\infty} a_n = \liminf_{n\to\infty} a_n$.

Proposition 3. (Properties of limsup and liminf) Let $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$ be arbitrary sequences of real numbers. Then it follows that

1. $\limsup_{n\to\infty} c\cdot a_n = c\limsup_{n\to\infty} a_n$ and for all $c \in \mathbb{R}$.
2. $\liminf_{n\to\infty} c\cdot a_n = c\liminf_{n\to\infty} a_n$ and for all $c \in \mathbb{R}$.
3. $\limsup_{n\to\infty}(a_n + b_n) \leq \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n$.
4. $\liminf_{n\to\infty}(a_n + b_n) \geq \liminf_{n\to\infty} a_n + \liminf_{n\to\infty} b_n$.

The above constructions are given all with respect to sequences, which can be viewed as a function $a(n) : \mathbb{N} \to \mathbb{R}$. This can readily extended to functions defined on the real line $f : \mathbb{R} \to \mathbb{R}$, and all the above definitions have a natural analogue to general functions instead of just sequences.

2 Probability Spaces

2.1 Set Theory

The symbol $\varnothing$ denotes the empty set, the unique set without any elements. If A and B are subsets of Ω, we denote this by $A \subset \Omega$ and $B \subset \Omega$ or just $A, B \subset \Omega$. Note that $A \subset B$ if and only if for all $\omega \in A$ we also have $\omega \in B$. Hence $A = B$ if and only if $A \subset B$ and $B \subset A$. Some common set operations are

1. union: $A \cup B := \{\omega \in \Omega : \omega \in A \text{ or } \omega \in B\}$
2. intersect: $A \cap B := \{\omega \in \Omega : \omega \in A \text{ and } \omega \in B\}$
3. set-difference $A\backslash B := \{\omega \in \Omega : \omega \in A \text{ and } \omega \notin B\}$
4. symmetric-difference $A\Delta B := \{\omega \in \Omega : \omega \in A \cup B \text{ and } \omega \notin A \cap B\}$.

We can take the union of more than 2 events. If $A_1, \ldots, A_n$ are events, then we denote the union by $\bigcup_{i=1}^n A_i$. Similarly if we have an arbitary collection of events $\{A_i\}_{i\in I}$, where I is an arbitrary index set, then this union is denoted $\bigcup_{i\in I} A_i$. Observe that the order of the sets we take union of does not matter. We may derive a similar definition for set intersection.

Proposition 4. (DeMorgan's Law) For an arbitrary collection of events $\{A_n\}_{n=1}^\infty$, prove the Demorgan Laws:

$$\bigcup_{i=1}^\infty A_i^c = \left(\bigcap_{i=1}^\infty A_i\right)^c, \qquad \bigcap_{i=1}^\infty {A_i}^c = \left(\bigcup_{i=1}^\infty A_i\right)^c$$

Proof. For the first law, fix $x \in \bigcup_{i=1}^{\infty} A_i^c$. Then $x \in A_j^c$ for some j, or simply $x \notin A_j$. However observe that $\bigcap_{i=1}^{\infty} A_i \subseteq A_j$ for any $j \in \mathbb{N}$ and so we have that $x \notin \bigcap_{i=1}^{\infty} A_i$. Equivalently $x \in \left(\bigcap_{i=1}^{\infty} A_i\right)^c$. For the reverse inclusion let $x \in \left(\bigcap_{i=1}^{\infty} A_i\right)^c$. Thus $x \notin \bigcap_{i=1}^{\infty} A_i$, which means that x is not in every A_i. That is, there exists some j such that $x \in A_j^c \subseteq \bigcup_{i=1}^{\infty} A_i^c$.

For the second law, fix $x \in \left(\bigcup_{i=1}^{\infty} A_i\right)^c$. Then $x \notin \bigcup_{i=1}^{\infty} A_i$, and hence for all i, $x \notin A_i$. That is, for all i, $x \in A_i^c$ showing that $x \in \bigcap_{i=1}^{\infty} A_i{}^c$. For the reverse inclusion, suppose that $x \in \bigcap_{i=1}^{\infty} A_i{}^c$. By definition, $x \in A_i^c$ for all i. Hence for any such i, $x \notin A_i$ thus $x \notin \bigcup_{i=1}^{\infty} A_i$. Equivalently, $x \in \left(\bigcup_{i=1}^{\infty} A_i\right)^c$.

□

Observe that functions are just mappings between sets. Let X and Y be two sets and let $f : X \to Y$ be a function from X to Y. For any subset $A \subset X$ we define the image of A under f to be $f(A) := \{f(a) : a \in A\}$. Similarly for $B \subset Y$, we define the preimage of B under f to be $f^{-1}(B) := \{x \in X : f(x) \in B\}$.

Proposition 5. Let $\{A_i\}_{i \in I}$ and $\{B_j\}_{j \in J}$ be arbitrary collections of sets, where I and J are arbitrary index sets. Then,

1. $f\left(\bigcup_{i \in I} A_i\right) = \bigcup_{i \in I} f(A_i)$;
2. $f\left(\bigcap_{i \in I} A_i\right) \subset \bigcap_{i \in I} f(A_i)$;
3. $f^{-1}\left(\bigcup_{j \in J} B_j\right) = \bigcup_{j \in J} f^{-1}(B_j)$;
4. $f^{-1}\left(\bigcap_{j \in J} B_j\right) = \bigcap_{j \in J} f^{-1}(B_j)$.

Proof. For (1), fix $b \in f\left(\bigcup_{i \in I} A_i\right)$. By definition, there exists $i \in I$ such that $f(a) = b$ for some $a \in A_i$. That is, $b \in f(A_i)$ for some $i \in I$ showing that $f\left(\bigcup_{i \in I} A_i\right) \subset \bigcup_{i \in I} f(A_i)$. For the reverse inclusion, fix $b \in \bigcup_{i \in I} f(A_i)$. Then there exists $i \in I$ such that $b \in f(A_i)$. Choose $a \in A_i$ such that $f(a) = b$, since $A_i \subset \bigcup_{i \in I} A_i$ we have that $b \in f\left(\bigcup_{i \in I} A_i\right)$ which shows $\bigcup_{i \in I} f(A_i) \subset f\left(\bigcup_{i \in I} A_i\right)$.

Fix $b \in f\left(\bigcap_{i \in I} A_i\right)$, then there exists $a \in \bigcap_{i \in I} A_i$ such that $f(a) = b$. Since $a \in A_i$ for all $i \in I$, we have that $b \in f(A_i)$ for all $i \in I$ and so $b \in \bigcap_{i \in I} f(A_i)$. This shows (2) (Note that the reverse inclusion does not hold in general, for instance if the intersection is empty we have a counterexample).

Now let $a \in f^{-1}\left(\bigcup_{j \in J} B_j\right)$. Then there exists $b \in \left(\bigcup_{j \in J} B_j\right)$ such that $a = f^{-1}(b)$. That is, for some $j \in J$ there exists $b \in B_j$ such that $a = f^{-1}(b)$. Therefore, $a \in f^{-1}(B_j)$ and thus $a \in \bigcup_{i \in I} f^{-1}(B_j)$. For the reverse inclusion, let $a \in \bigcup_{i \in I} f^{-1}(B_j)$. Choose $j \in J$ such that $a \in f^{-1}(B_j)$, and so $a = f^{-1}(b)$ for some $b \in B_j \subset \bigcup_{j \in J} B_j$. Thus $a \in f^{-1}\left(\bigcup_{j \in J} B_j\right)$.

Fix $a \in f^{-1}\left(\bigcap_{j \in J} B_j\right)$. Equivalently, $f(a) \in \left(\bigcap_{j \in J} B_j\right)$ and by definition of the intersection we have $f(a) \in B_j$ for all j. Equivalently, $a \in f^{-1}(B_j)$ for all j and so $a \in \bigcap_{j \in J} f^{-1}(B_j)$. Since we have bi-directional implications throughout, we have equality, $f^{-1}\left(\bigcap_{j \in J} B_j\right) = \bigcap_{j \in J} f^{-1}(B_j)$.

□

2.2 σ-Algebras

The core object of any random occurrence is what is known as the *sample space*, which is a non-empty set Ω, that is the collection of all *outcomes* $\omega \in \Omega$ of some experiment under uncertainty. For example if we roll a dice our sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$ and if we wish to measure the height of an object, our sample space can be is $\Omega = [0, \infty)$. In probability, we are interested in the likelihood a collection of desired outcomes will occur, such a collection of outcomes are called *events* which are the observable subsets $A \subset \Omega$. For instance, we may be interested in seeing if a particular object is 2 metres tall. However in reality, we cannot measure length with infinite precision and what we are really interested in is if an object perhaps has height contained within the interval $[199.5, 200.5]$. It is for this reason, we are concerned with events instead of the elementary outcomes of an experiment. The collection of all observable events of a sample space, is called a *σ-algebra* and is denoted by $\mathcal{F}$.

Definition 1. Let $\Omega \neq \varnothing$, a collection $\mathcal{F}$ of subsets $A \subset \Omega$ is called a *σ-algebra* on Ω if

1. $\varnothing \in \mathcal{F}$,
2. $A \in \mathcal{F}$ implies that $A^c \in \mathcal{F}$, and
3. If $A_n \in \mathcal{F}$ for each $n \in \mathbb{N}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

The final condition in the above is what is called closed under countable unions. Together with the second condition it can readily be shown that $\mathcal{F}$ is actually closed under all countable set operations.

Example 5. Let $\Omega \neq \varnothing$.

1. $\mathcal{F} := 2^{\Omega}$ is the largest possible σ-algebra that can be defined on a sample space Ω.
2. Let $\mathcal{F} := \{\varnothing, \Omega\}$. This is the smallest σ-algebra on Ω. It is contained in every other σ-algebra on Ω.
3. Fix a subset $A \subset \Omega$, then $\mathcal{F} := \{\varnothing, A, A^c, \Omega\}$ is a σ-algebra on Ω.

Proposition 6. Let $\{\mathcal{F}_i\}_{i \in I}$ be a family of σ-algebras on a nonempty set Ω, where I is an arbitrary index set. Then $\mathcal{F} := \bigcap_{i \in I} \mathcal{F}_i$ is also a σ-algebra.

Proposition 7. Let $\mathcal{F}$ denote an arbitrary collection of subsets of Ω. Then there exists a unique minimal σ-algebra, denoted $\sigma(\mathcal{F})$, on Ω such that $\mathcal{F} \subset \sigma(\mathcal{F})$.

Proof. Let $\mathscr{F}$ be the set of all σ-algebras on Ω that contain the sets in $\mathcal{F}$, of course $\mathscr{F} \neq \varnothing$ since $2^{\Omega} \in \mathscr{F}$. Define

$$\sigma(\mathcal{F}) := \bigcap_{\mathcal{G} \in \mathscr{F}} \mathcal{G}.$$

By the previous result we know that $\sigma(\mathcal{F})$ is a σ-algebra. By definition of set intersection $\sigma(\mathcal{F})$ is minimal. For uniqueness, suppose there exists another minimal σ-algebra, $\mathcal{F}'$ containing all the sets in $\mathcal{F}$. However this implies that $\mathcal{F}' \in \mathscr{F}$ and thus $\sigma(\mathcal{F}) \subset \mathcal{F}'$. However $\sigma(\mathcal{F}) \in \mathscr{F}$ and thus by minimality of $\mathcal{F}'$, we have $\mathcal{F}' \subset \sigma(\mathcal{F})$ completing the proof.

□

We refer to $\sigma(\mathcal{F})$ as the σ-algebra generated by the collection of sets in $\mathcal{F}$, and the family of sets $\mathcal{F}$ is called the *generator* of $\sigma(\mathcal{F})$. Unfortunately this construction (although elegant) does not provide a systematic way to explicitly formulate all the elements of $\sigma(\mathcal{F})$. We now look at an example of a σ-algebra defined on the real line (or subsets of the real line), which is one of the most important examples of a σ-algebra.

Definition 2. (Borel σ-algebra) Let $\mathcal{U}$ denote the collection of all open subsets of $\mathbb{R}$. The collection of sets $\mathcal{B}(\mathbb{R}) := \sigma(\mathcal{U})$ is said to be called the *Borel σ-algebra* defined on $\mathbb{R}$. An analogous definition can be defined for any Euclidean space, $\mathbb{R}^n$.

The open subsets of $\mathbb{R}$, satisfy the convenient property that they can be represented as a countable union of disjoint open intervals. That is for all open sets $U \subset \mathbb{R}$, there exists a sequence of constants $-\infty \leq a_n \leq b_n \leq \infty$ for $n \in \mathbb{N}$, such that $U = \bigcup_{n=1}^{\infty}(a_n, b_n)$. With this representation, along with the definition of a σ-algebra. It can be shown that the Borel σ-algebra can be generated in various equivalent ways.

Proposition 8. Consider the following subsets of $\mathbb{R}$:

1. let $\mathcal{F}_1$ denote the collection of closed subsets of $\mathbb{R}$,

2. let $\mathcal{F}_2$ denote the collection of all intervals $(-\infty, b]$ for $b \in \mathbb{R}$,

3. let $\mathcal{F}_3$ denote the collection of all intervals $(-\infty, b)$ for $b \in \mathbb{R}$,

4. let $\mathcal{F}_4$ denote the collection of all intervals $(a, b]$ for $-\infty < a < b < \infty$,

5. let $\mathcal{F}_5$ denote the collection of all intervals (a, b) for $-\infty < a < b < \infty$.

Then it follows that $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{F}_1) = \sigma(\mathcal{F}_2) = \sigma(\mathcal{F}_3) = \sigma(\mathcal{F}_4) = \sigma(\mathcal{F}_5)$.

Proof. Since the elements of $\mathcal{F}_3$ are all open subsets of $\mathbb{R}$, it is immediate that $\sigma(\mathcal{F}_3) \subset \mathcal{B}(\mathbb{R})$. Now fix an open interval (a, b) for some $a, b \in \mathbb{R}$, and observe that

$$(a, b) = \bigcap_{n=1}^{\infty} [a + 1/n, b) = \bigcap_{n=1}^{\infty} ((-\infty, b) \backslash (-\infty, a + 1/n)) \in \sigma(\mathcal{F}_3).$$

Hence $\mathcal{F}_5 \subset \sigma(\mathcal{F}_3)$ and thus $\sigma(\mathcal{F}_5) \subset \sigma(\mathcal{F}_3)$. Now let $U \subset \mathbb{R}$ be open. Choose the disjoint interval representation given by $U = \bigcup_{n=1}^{\infty} (a_n, b_n)$ (Observe that we may let $a_n = -\infty$ and $b_n = \infty$ for some n, as these are already contained in $\sigma(\mathcal{F}_5)$). This shows that $\mathcal{B}(\mathbb{R}) \subset \sigma(\mathcal{F}_5)$. Combining all of the above we have that

$$\mathcal{B}(\mathbb{R}) \subset \sigma(\mathcal{F}_5) \subset \sigma(\mathcal{F}_3) \subset \mathcal{B}(\mathbb{R})$$

showing that

$$\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{F}_5) = \sigma(\mathcal{F}_3).$$

If $U \subset \mathbb{R}$ is open, then $U^c \in \mathcal{F}_1$ and thus $U = (U^c)^c \in \sigma(\mathcal{F}_1)$. Hence, $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{F}_1)$. The reverse inclusion is proved identically as every open set is the complement of a closed and vice-versa. Using $(a, b] = \bigcap_{n=1}^{\infty} (a, b+1/n)$, we see that $\mathcal{F}_4 \subset \sigma(\mathcal{F}_5)$, and and we conclude that $\sigma(\mathcal{F}_4) \subset \sigma(\mathcal{F}_5) = \mathcal{B}(\mathbb{R})$. Using $(a, b) = \bigcup_{n=1}^{\infty} (a, b - 1/n]$, we have $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{F}_5) \subset \sigma(\mathcal{F}_4)$. Thus $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{F}_4)$. Since $(a, b] = (-\infty, b] \backslash (-\infty, a]$, then $\mathcal{F}_4 \subset \sigma(\mathcal{F}_2)$. Furthermore, $(-\infty, b] = \bigcup_{n=1}^{\infty} (b - n, b]$ showing that $\mathcal{F}_2 \subset \sigma(\mathcal{F}_4)$. By the same reasoning above, we see that $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{F}_2)$. Hence we have set equality all throughout. $\square$

Example 6. The collecton of rational numbers, $\mathbb{Q}$, is a Borel set.

Proof. Fix $x \in \mathbb{R}$, and observe that $[x - 1/n, x + 1/n] \in \mathcal{B}(\mathbb{R})$ for each $n \in \mathbb{N}$. As $\mathcal{B}(\mathbb{R})$ is closed under countable intersection, we have that $\{x\} = \bigcap_{n=1}^{\infty} [x - 1/n, x + 1/n] \in \mathcal{B}(\mathbb{R})$. Hence single-tons are Borel sets and thus $\{q\} \in \mathcal{B}(\mathbb{R})$ for all $q \in \mathbb{Q}$. Since $\mathbb{Q}$ is countable, and $\mathcal{B}(\mathbb{R})$ is closed under countable unions, we have that $\mathbb{Q} = \bigcup_{q \in \mathbb{Q}} \{q\} \in \mathcal{B}(\mathbb{R})$ as required. $\square$

2.3 Probability Measures

Now that we have defined what a sample space is and its corresponding collection of events, the σ-algebra, it remains to assign a likelihood to the occurrence of the events in study. This is what is called the probability measure.

Definition 3. (Probability Measure) Let Ω denote a non-empty set, and suppose $\mathcal{F}$ is a σ-algebra on Ω. A function $\mathbb{P} : \mathcal{F} \to [0, 1]$ is called a *probability measure* on Ω if:

1. $\mathbb{P}(\Omega) = 1$.

2. For any collection of disjoint sets $\{A_n\}_{n=1}^{\infty}$ in $\mathcal{F}$, we have

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

We can relax the first condition to $\mathbb{P}(\Omega) < \infty$, in which case $\mathbb{P}$ is called a (finite) *measure*. General measures can also be infinite and even take non-positive values, but for the purposes of our work we will not consider these. We define a general non-negative measure below for completeness:

Definition 4. (Measure) Let Ω denote a non-empty set, and suppose $\mathcal{F}$ is a σ-algebra on Ω. A function $\mu : \mathcal{F} \to [0, \infty]$ is called a *measure* on Ω if:

1. $\mu(\varnothing) = 0$.
2. For any collection of disjoint sets $\{A_n\}_{n=1}^{\infty}$ in $\mathcal{F}$, we have
$$\mu\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mu(A_n).$$

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *probability space*. We list some basic properties of a probability measure below:

Proposition 9. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Then the following properties hold:

1. $\mathbb{P}(\varnothing) = 0$
2. If $A_1, \ldots, A_n \in \mathcal{F}$ are disjoint sets, then $\mathbb{P}\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{i=1}^{n} \mathbb{P}(A_i)$;
3. If $A, B \in \mathcal{F}$ then $\mathbb{P}(A \backslash B) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$;
4. if $A, B \in \mathcal{F}$, with $B \subset A$ then $\mathbb{P}(B) \leq \mathbb{P}(A)$;
5. (Boole's Inequality): If $\{A_n\}_{n=1}$ is an arbitrary collection of events then
$$\mathbb{P}\Big(\bigcup_{n=1}^{\infty} A_n\Big) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n);$$
6. (Left Continuity): If $\{A_n\}_{n=1}^{\infty}$ is an increasing sequence of sets in $\mathcal{F}$, then
$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\Big(\bigcup_{i=1}^{\infty} A_i\Big);$$
7. (Right Continuity): If $\{A_n\}_{n=1}^{\infty}$ is a decreasing sequence of sets in $\mathcal{F}$, then
$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\Big(\bigcap_{i=1}^{\infty} A_i\Big).$$

In the definition of the probability measure, it is important to note that the empty set is not unique in being assigned a probability of 0. That is, if $\mathbb{P}(A) = 0$, it is not necessarily the case that $A = \varnothing$. Such sets A are called *null sets*.

Consider a sequence of subsets $\{A_n\}_{n=1}^{\infty}$ of some sample space Ω. We define the set $\liminf_{n\to\infty} A_n$ to be the collection of all elements $\omega \in \Omega$ that are in A_n for *all but finitely many* $n \in \mathbb{N}$. Analogously we define the set $\limsup_{n\to\infty} A_n$ be denote the set of all $\omega \in \Omega$ that are in A_n for *infinitely many* of the $n \in \mathbb{N}$. In set-theoretic formulation, they are defined as follows:
$$\liminf_{n\to\infty} A_n := \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m \quad \text{and} \quad \limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m.$$

Proposition 10. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\{A_n\}_{n=1}^{\infty}$ a sequence of events in $\mathcal{F}$. Then it follows that
$$\mathbb{P}(\liminf_{n\to\infty} A_n) \leq \liminf_{n\to\infty} \mathbb{P}(A_n) \leq \limsup_{n\to\infty} \mathbb{P}(A_n) \leq \mathbb{P}(\limsup_{n\to\infty} A_n).$$

Proof. Observe that the collection of sets $\{\bigcup_{m=n}^{\infty} A_m\}_{n=1}^{\infty}$ is decreasing as $n \to \infty$ to $\limsup_n A_n$. Hence by continuity of probability,

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) := \mathbb{P}\left(\bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m\right) = \limsup_{n\to\infty} \mathbb{P}\left(\bigcup_{m=n}^{\infty} A_m\right) \geq \limsup_{n\to\infty} \mathbb{P}(A_n).$$

Similarly, the collection of sets $\{\bigcap_{m=n}^{\infty} A_m\}$ is increasing and thus

$$\liminf_{n\to\infty} \mathbb{P}(A_n) \geq \liminf_{n\to\infty} \mathbb{P}\left(\bigcap_{m=n}^{\infty} A_m\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} A_m\right) = \mathbb{P}\left(\liminf_{n\to\infty} A_n\right).$$

$\square$

A sequence of sets $\{A_n\}_{n=1}^{\infty}$ is called *convergent* if we have $\liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n$, and the limit of the sets is denoted $\lim\limits_{n\to\infty} A_n$ and is equal to both $\liminf_{n\to\infty} A_n$ and $\limsup_{n\to\infty} A_n$. When a sequence of sets is convergent, we have the property that

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

which means that the probability measure is *continuous.*

Example 7. Let $A_1, \ldots, A_n$ be n events in Ω. Consider the sequence

$$A_1, A_2, \ldots, A_n, A_1, A_2, \ldots, A_n, A_1, A_2, \ldots, A_n, \ldots$$

Determine $\liminf_{n\to\infty} A_n$ and $\limsup_{n\to\infty} A_n$, and if the sequence is convergent.

Proof. By inspection, we claim $\limsup_{m\to\infty} A_m = \bigcup_{k=1}^{n} A_k$. Fix $\omega \in \limsup_{n\to\infty} A_n$. Then $\omega \in A_m$ for infinitely many $m \in \mathbb{N}$, and thus for $N \in \mathbb{N}$ there exists $m > N$ such that $\omega \in A_m$. However by construction of our sequence we know that $A_m = A_k$ for some $1 \leq k \leq n$. Therefore $\omega \in \bigcup_{k=1}^{n} A_k$. Conversely let $\omega \in \bigcup_{k=1}^{n} A_k$, then $\omega \in A_k$ for some $1 \leq k \leq n$. However note that $A_k = A_{k+ln}$ for each $l \in \mathbb{N}$, and since there are infinitely many such l we conclude that $\omega \in A_m$ for infinitely many m. That is $\omega \in \limsup_{m\to\infty} A_m$. On the other hand, we conjecture that $\liminf_{m\to\infty} A_m = \bigcap_{k=1}^{n} A_k$. To this end, fix $\omega \in \liminf_{m\to\infty} A_m$. Choose $N \in \mathbb{N}$ so that $\omega \in A_m$ whenever $m > N$. However by exhaustion, it is immediate that $\omega \in A_k$ for all $1 \leq k \leq n$ and thus $\omega \in \bigcap_{k=1}^{n} A_k$. Conversely if $\omega \in \bigcap_{k=1}^{n} A_k$ then $\omega \in A_k$ for all $1 \leq k \leq n$. However by construction, this means ω is in every element in the sequence of sets and thus $\omega \in \liminf_{m\to\infty} A_m$ is immediate. As these two sets are different, we conclude that the sequence is not convergent.

$\square$

Example 8. Let $A_1, \ldots, A_n$ be n events in Ω. Consider the sequence

$$A_1, A_2, \ldots, A_n, \varnothing, \varnothing, \varnothing, \ldots$$

Determine $\liminf_{n\to\infty} A_n$ and $\limsup_{n\to\infty} A_n$, and if the sequence is convergent.

Proof. Observe that this sequence is eventually constant, and so $A_n \to \emptyset$. Therefore $\liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n = \lim_{n\to\infty} A_n = \emptyset$. $\square$

Now that we have the basic properties of a probability measure, we now look at probability measures that are defined on the Borel σ-algebra introduced above. Before we proceed, we present the following result (without proof) that uniquely identifies a set function with a corresponding right-continuous, non-decreasing function.

Theorem 3. Let $F : \mathbb{R} \to \mathbb{R}$ be an arbitrary right-continuous, non-decreasing function. Then there exists a unique measure $\mathbb{P} : \mathcal{B}(\mathbb{R}) \to [0, \infty]$ such that

$$\mathbb{P}((a, b]) = F(b) - F(a)$$

for all $-\infty < a \leq b < \infty$.

Example 9. (Lebesgue Measure) Let $F : \mathbb{R} \to \mathbb{R}$ be the identity function $F(x) = x$. Then F satisfies the conditions of Theorem 3, and the unique measure λ on $\mathcal{B}(\mathbb{R})$ that satisfies

$$\lambda((a, b]) = b - a$$

is called the *Lebesgue measure* on $\mathbb{R}$.

The next result showcases some properties of this set function $\mathbb{P}$. Furthermore it shows that if we include an additional limit condition on F, then the set function $\mathbb{P}$ is actually a probability measure. As the function F is non-decreasing, we know that it has a left-limit at every point (that need not be attained by the function). For each $b \in \mathbb{R}$, we define the *left limit* by

$$F(b-) = \lim_{x \to b-} F(x).$$

Theorem 4. Let F be any non-decreasing, right-continuous function $F : \mathbb{R} \to \mathbb{R}$. Define $\mathbb{P} : \mathcal{B}(\mathbb{R}) \to \mathbb{R}$ by

$$\mathbb{P}((a, b]) = F(b) - F(a).$$

for $-\infty \leq a \leq b \leq \infty$. Then it follows that

1. $\mathbb{P}((-\infty, b]) = F(b) - \lim_{x \to -\infty} F(x)$;
2. $\mathbb{P}((-\infty, b)) = F(b-) - \lim_{x \to -\infty} F(x)$;
3. $\mathbb{P}((a, b)) = F(b-) - F(a)$;
4. $\mathbb{P}([a, b)) = F(b-) - F(a-)$;
5. $\mathbb{P}([a, b]) = F(b) - F(a-)$.
6. $\mathbb{P}((a, \infty)) = \lim_{x \to \infty} F(x) - F(a)$;
7. $\mathbb{P}([a, \infty)) = \lim_{x \to \infty} F(x) - F(a-)$;
8. $\mathbb{P}(\{a\}) = F(a) - F(a-)$;
9. If, in addition, the function F is continuous, then
   (a) $\mathbb{P}((-\infty, b]) = \mathbb{P}((-\infty, b)) = F(b) - \lim_{x \to -\infty} F(x)$;
   (b) $\mathbb{P}((a, b)) = \mathbb{P}([a, b)) = \mathbb{P}((a, b]) = \mathbb{P}([a, b]) = F(b) - F(a)$;
   (c) $\mathbb{P}(\{a\}) = 0$.
10. If $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$, then $\mathbb{P}$ is a probability measure.

Proof. For 1, Note that

$$\mathbb{P}((-\infty, b]) = \mathbb{P}\Big(\bigcup_{n=1}^{\infty} (-n, b]\Big) = \lim_{n \to \infty} \mathbb{P}((-n, b])) = F(b) - \lim_{n \to \infty} F(-n) = F(b) - \lim_{x \to -\infty} F(x)$$

For 2,

$$\mathbb{P}((-\infty, b)) = \lim_{n \to \infty} \mathbb{P}((-\infty, b - \frac{1}{n}]) = \lim_{n \to \infty} F(b - 1/n) - \lim_{x \to -\infty} F(x) = F(b-) - \lim_{x \to -\infty} F(x),$$

by 1. For 3, we see that:

$$F(b-) - F(a) = \lim_{n \to \infty} (\mathbb{P}((-\infty, b - \frac{1}{n}]) - \mathbb{P}((-\infty, a])) = \lim_{n \to \infty} \mathbb{P}((a, b - \frac{1}{n}]) = \mathbb{P}((a, b)).$$

For 4 we see that,

$$\begin{aligned}
F(b-) - F(a-) &= \lim_{n\to\infty}\big(\mathbb{P}((-\infty, b-\frac{1}{n}]) - \mathbb{P}((-\infty, a-\frac{1}{n}])\big)\\
&= \lim_{n\to\infty}\mathbb{P}\big((-\infty, b-\frac{1}{n}]\backslash(-\infty, a+\frac{1}{n}]\big)\\
&= \lim_{n\to\infty}\mathbb{P}\big((-\infty, b-\frac{1}{n}]\cap(a-\frac{1}{n},\infty)\big) = \lim_{n\to\infty}\mathbb{P}((a-\frac{1}{n}, b-\frac{1}{n}]) = \mathbb{P}([a,b)).
\end{aligned}$$

For 5, observe that $[a,b] = (-\infty, b]\backslash(-\infty, a)$. Taking measures of both sides we see that $\mathbb{P}([a,b]) = F(b) - F(a-)$. For 6, it suffices to show that $\lim_{x\to\infty} F(x) = 1$. It follows that

$$\lim_{x\to\infty} F(x) = \lim_{x\to\infty}\mathbb{P}((-\infty, x]) = \mathbb{P}(\bigcup_{n=1}^{\infty}(-\infty, n]) = \mathbb{P}(\mathbb{R}) = 1$$

and so,

$$\lim_{x\to\infty} F(x) - F(a) = 1 - F(a) = 1 - \mathbb{P}((-\infty, a]) = \mathbb{P}((a,\infty)).$$

Analogously for 7,

$$\lim_{x\to\infty} F(x) - F(a-) = 1 - F(a) = 1 - \mathbb{P}((-\infty, a)) = \mathbb{P}([a,\infty)).$$

Observe that $(-\infty, a)\cup\{a\} = (-\infty, a]$ is a disjoint union. Taking measures of both sides and re-arranging yields $\mathbb{P}(\{a\}) = F(a) - F(a-)$. Now if F is continuous, then namely it is left continuous and so $F(b-) = \lim_{x\to b^-} F(x) = F(b)$. Thus we have that

$$F(b) - \lim_{x\to-\infty} F(x) = F(b) = F(b-) = \mathbb{P}((-\infty, b)) = \mathbb{P}((-\infty, b]).$$

It is immediate that $\mathbb{P}((a,b]) = \mathbb{P}((a,b))$ and $\mathbb{P}((a,b]) = \mathbb{P}(([a,b])$, hence $\mathbb{P}((a,b]) = ([a,b))$ as well. Finally we have that $\mathbb{P}((a,b]) = F(b) - F(a)$ by definition and thus we have equality throughout. Now $\{a\}\cup(-\infty, a) = (-\infty, a]$ is a disjoint union. Taking measures and re-arranging we have that $\mathbb{P}(\{a\}) = F(a) - F(a-) = F(a) - F(a) = 0$ proving 9. Finally for 10 suppose that $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$. Then,

$$\mathbb{P}(\varnothing) = \mathbb{P}(\bigcap_{n=1}^{\infty}(-\infty, -n])) = \lim_{x\to-\infty}\mathbb{P}((-\infty, x]) = \lim_{x\to-\infty} F(x) = 0,$$

and

$$\mathbb{P}(\mathbb{R}) = \mathbb{P}((-\infty,\infty)) = \mathbb{P}\left(\bigcup_{n=1}^{\infty}(-\infty, n]\right) = \lim_{n\to\infty}\mathbb{P}((-\infty, n]) = \lim_{x\to\infty} F(x) = 1.$$

Now suppose that $\{A_n\}_{n=1}^{\infty}$ is a disjoint collection in $\mathcal{F}$. Then by the same argument used in the proof of Proposition 8 we have that $\mathbb{P}\big(\bigcup_{n=1}^{\infty} A_n\big) = \sum_{n=1}^{\infty}\mathbb{P}(A_n)$ showing that $\mathbb{P}$ is countably additive. Therefore $\mathbb{P}$ is a probability measure. □

2.4 Independent Events

Definition 5. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We say that the events $\{A_i\}_{i\in I}$ are *independent* if for any finite subcollection $i_1, \dots, i_n \in I$ where $n\in\mathbb{N}$, we have

$$\mathbb{P}(A_{i_1}\cap A_{i_2}\cap\cdots\cap A_{i_n}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2})\cdots\mathbb{P}(A_{i_n}).$$

It is immediate that if $A_1, \dots, A_n$ are independent then their complements $A_1^c\cdots, A_n^c$ are also independent. From this, it can be readily shown that any combination of events and their complements will be independent from one another. If events are not independent on one another, then they are said to be *dependent*. That is, the occurrence of one event will influence the likelihood of the occurrence of another. With this we compute the second event's *conditional probability* given the first event.

Definition 6. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and fix $B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$. Then we define the *conditional probability* if A given B by

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

for all $A \in \mathcal{F}$.

The following is an immediate result from the definition of conditional probability, that is well known in applied probability

Theorem 5. (Baye's Theorem) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $A_1, \ldots, A_n \in \mathcal{F}$ be a partition of Ω. Then for all $B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$, and $j = 1, 2, \ldots, n$ we have that

$$\mathbb{P}(A_j|B) = \frac{\mathbb{P}(B|A_j)\mathbb{P}(A_j)}{\sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i)}.$$

The quantities $\mathbb{P}(A_j)$ are called *prior probabilities* and the $\mathbb{P}(A_j|B)$ are called *posterior probabilities* of A_j.

The following result will play a crucial role in proving results pertaining to the convergence of random variables.

Proposition 11. (Borel-Cantelli Lemma) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\{A_n\}_{n=1}^{\infty}$ be a collection of events in $\mathcal{F}$. Then,

1. If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$ then $\mathbb{P}(\limsup_{n\to\infty} A_n) = 0$.
2. If $\{A_n\}_{n=1}^{\infty}$ are independent and $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, then $\mathbb{P}(\limsup_{n\to\infty} A_n) = 1$.

Proof. Set $B_n := \bigcup_{m=n}^{\infty} A_n$ and note that $B_{n+1} \subset B_n$ for each n. Since

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m = \bigcap_{n=1}^{\infty} B_n,$$

we have that by continuity of $\mathbb{P}$, and Boole's inequality

$$\mathbb{P}(\limsup_{n\to\infty} A_n) = \mathbb{P}\Big(\bigcap_{n=1}^{\infty} B_n\Big) = \lim_{n\to\infty} \mathbb{P}(B_n) = \lim_{n\to\infty} \mathbb{P}\Big(\bigcup_{m=n}^{\infty} A_m\Big) \le \lim_{n\to\infty} \sum_{m=n} \mathbb{P}(A_m) = 0.$$

For the second part, by taking complements, we show that $\mathbb{P}(\liminf_{n\to\infty} A_n^c) = 0$. Observe that

$$(\limsup_{n\to\infty} A_n)^c = \Big(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m\Big)^c = \bigcup_{n=1}^{\infty} \Big(\bigcup_{m=n}^{\infty} A_m\Big)^c = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m^c = \liminf_{n\to\infty} A_n^c.$$

Set $B_n := \bigcap_{m=n} A_n^c$, which satisfies $B_n \subset B_{n+1}$ for each n. Then once again by continuity of $\mathbb{P}$, we have that $\mathbb{P}(\liminf_{n\to\infty} A_n^c) = \lim\limits_{n\to\infty} \mathbb{P}(B_n)$. Since the A_n are all independent, so are all the A_n^c. Thus for each $n \in \mathbb{N}$,

$$\begin{aligned}
\mathbb{P}(B_n) &= \mathbb{P}\Big(\bigcap_{m=n}^{\infty} A_m^c\Big) \\
&= \lim_{m\to\infty} (\mathbb{P}(A_n^c)(\mathbb{P}(A_{n+1}^c) \cdots \mathbb{P}(A_m^c)) \\
&= \lim_{m\to\infty} ((1 - \mathbb{P}(A_n))(1 - \mathbb{P}(A_{n+1}) \cdots (1 - \mathbb{P}(A_m))) \\
&\le \lim_{m\to\infty} e^{-\sum_{i=n}^{m} \mathbb{P}(A_i)} = 0.
\end{aligned}$$

$\square$

2.5 Problems

Problem 1. Let Ω_1, Ω_2 be two non-empty sets and quipped with respective σ-algebras $\mathcal{F}$ and $\mathcal{G}$ and let $\mathcal{G}_0$ be a generator of $\mathcal{G}$. Let $X : \Omega_1 \to \Omega_2$ be a function between them and suppose $\mathcal{A}$ is an arbitrary family of subsets of Ω_2. Show that:

(1) The collection of sets $X^{-1}(\mathcal{G}) := \{X^{-1}(A) : A \in \mathcal{G}\}$ defines a σ-algebra on Ω_1.

(2) $X^{-1}(\sigma(\mathcal{A})) = \sigma(X^{-1}(\mathcal{A}))$.

(3) If $X^{-1}(B) \in \mathcal{F}$ for all $B \in \mathcal{G}_0$, then $X^{-1}(B) \in \mathcal{F}$ for all $B \in \mathcal{G}$

Problem 2. (Countable Generators) A σ-field is said to be *countably generated* if there is a countable collection $\mathcal{C} \subset \mathcal{F}$ such that $\sigma(\mathcal{C}) = \mathcal{F}$. Prove that $\mathcal{B}(\mathbb{R})$ is countably generated.

Problem 3. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability measure space and let $\{A_n\}_{n=1}^{\infty} \subset \mathcal{F}$ be arbitrary. Prove:

(1) $\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{m\to\infty} \mathbb{P}\left(\bigcup_{n=1}^{m} A_n\right)$.

(2) $\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{m\to\infty} \mathbb{P}\left(\bigcap_{n=1}^{m} A_n\right)$.

Problem 4. Give an example where the union of two σ-algebras is not a σ-algebra.

Problem 5. Let $\mathcal{F}$ be a σ-algebra on Ω and let $A \in \mathcal{F}$. Show that the collection of sets

$$\mathcal{G} := \{B \cap A : B \in \mathcal{F}\}$$

is a σ-algebra on A.

Problem 6. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and fix $B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$. Show that $(\Omega, \mathcal{F}, \mathbb{P}_B)$ is a probability space where $\mathbb{P}_B(A) := \mathbb{P}(A|B)$ for all $A \in \mathcal{F}$.

Problem 7. Consider the following game: If we roll a die it shows each of the numbers $\{1, 2, 3, 4, 5, 6\}$ with probability $1/6$. Let $(k_n)_{n=1}^{\infty}$ be a given sequence in $\mathbb{N}$.

(a) First go: We roll the die k_1 times. We win if all k_1 times the die shows 6.
(b) Second go: We roll the die k_2 times. Again, we win if all k_2 times the die shows 6.

And so on... Show that

(a) The probability of winning infinitely often is 1 if and only if $\sum_{n=1}^{\infty} (1/6)^{k_n} = \infty$.
(b) The probability of losing infinitely often is always 1.

Problem 8. Let $\mathbb{P}$ be a finite measure on $\mathcal{B}(\mathbb{R})$. Define the function $F(x) := \mathbb{P}((-\infty, x])$. Show that F is increasing, right-continuous, and $\mathbb{P}((a, b]) = F(b) - F(a)$ for all $-\infty < a \leq b < \infty$.

Problem 9. Define the function $F(x)$ equal to 0 if $x < 0$; to $\frac{1}{2}$ if $0 \leq x < 2$ and to 1 if $2 \leq x$. What is the unique probability measure on $\mathcal{B}(\mathbb{R})$ determined by F.

Problem 10. Suppose the sequence $q_1, q_2, \ldots$ contains all the rational numbers, each one exactly once. Let $\lambda_n \geq 0$ be such that $\sum_{n=1}^{\infty} \lambda_n = 1$. Define the function

$$F(x) = \sum_{n=1}^{\infty} \lambda_n \mathbf{1}_{[q_n, \infty)}(x).$$

Show that $F(x)$ is right-continuous and increasing. What is the unique probability measure on $\mathcal{B}(\mathbb{R})$ determined by $F(x)$?

Problem 11. Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of events.Let $C_n := \cap_{m=n}^{\infty} A_m$, $B_n := \cup_{m=n}^{\infty} A_m$, for each $n \in \mathbb{N}$, and $C := \liminf_{n\to\infty} A_n$ and $B := \limsup_{n\to\infty} A_n$. Show that

1. If $A_n \to A$, then $\mathbb{P}(A_n) \to \mathbb{P}(A)$.
2. If B_n is independent of C_n for every n, then B is independent of C.
3. If B_n is independent of C_n for every n and if $A_n \to A$ then A is independent of itself.

Problem 12. (Counting Measure) Let Ω be a non-empty set, equipped with a σ-algebra $\mathcal{F}$. Define $|A|$ to denote the number of elements in A (which may be infinite). Prove that $|\cdot|$ defines a measure on Ω and give an example of a sequence of sets $\{A_n\}_{n=1}^{\infty}$ such that $A_n \to \varnothing$, while $|A_n| \to \infty$.

Problem 13. Suppose you have a collection of billiard balls, $\{B_n\}_{n=0}^{\infty}$. At each step of the game you add two balls to a bag and you withdraw one. More precisely, at one minute before noon, balls numbered 0 and 1 are placed in the bag and ball number 0 is removed. At 1/2 of a minute before noon, balls numbered 2 and 3 are added, and ball number 1 is removed. At 1/3 of a minute before noon, balls 4 and 5 are added and ball number 2 is taken out. This process is continued, and the question is asked: How many balls are in the bag at noon? Prove that your answer is correct.

Problem 14. Let Ω be countable and fix $0 < p < 1$. Show that there does not exist a sequence of independent events $\{A_n\}_{n=1}^{\infty}$, such that $\mathbb{P}(A_n) = p$ for all $n \in \mathbb{N}$.

Problem 15. Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of events. Show that if $\lim\limits_{n\to\infty} \mathbb{P}(A_n) = 0$ and the events satisfy $\sum_{n=1}^{\infty} \mathbb{P}(A_n^c \cap A_{n+1}) < \infty$ then $\mathbb{P}(\limsup_{n\to\infty} A_n) = 0$.

3 Random Variables

3.1 Introduction

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a subset $A \subset \Omega$ is called *measurable* if $A \in \mathcal{F}$. We can also define functions on probability spaces, which are called *random variables.*

Definition 7. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let S denote an arbitrary set, equipped with a σ-algebra $\mathcal{A}$. A function $X : \Omega \to S$ is called a *measurable* if $X^{-1}(B) \in \mathcal{F}$ for all $B \in \mathcal{A}$. If $S = \mathbb{R}$ and $\mathcal{A} = \mathcal{B}(\mathbb{R})$, we say X is a *random variable.*

For the purpose of this course, we will freely call all such measurable functions as random variables with respect to their state space for simplicity. That is, functions of the form $X : \Omega \to S$ will be referred to as S-valued random variables. Without any further distinction, random variables are assumed to be real-valued random variables that are Borel measurable. Observe that we can have multiple σ-algebras defined on the same sample spaces Ω and S, and thus we often express the random variable mapping $X : (\Omega, \mathcal{F}) \to (S, \mathcal{A})$ to explicitly state which σ-algebras X is measurable with respect to. Observe that in the above definition, we never needed to use the probability measure $\mathbb{P}$, and so a random variable only requires a sample space, equipped with a σ-algebra to be defined on.

Example 10. (Indicator Random Variable) Fix $A \in \mathcal{F}$, then the function

$$X(\omega) := \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A \end{cases}$$

is a random variable, and is denoted by $\mathbf{1}_A$.

Example 11. (Constant Random Variable) Let Ω be a non-empty set, equipped with the trivial σ-algebra $\{\varnothing, \Omega\}$. What are the possible random variables defined on this space? Lets say $X : \Omega \to \mathbb{R}$ takes on two different values x and y. Then of course $X^{-1}(\{x\})$ is a proper, non-empty subset of Ω and hence not in $\mathcal{F}$. Hence if $X(\omega) = x$ for some ω, it must be the case that $X(\omega) = x$ for all $\omega \in \Omega$. That is, X is the constant function that takes on the value x.

Showing that every set is measurable can be quite difficult, Problem 1 shows that it suffices to show that any generating set has a measurable pre-image in order for the function to be measurable. Combining the result of Problem 1 and Proposition 8, it suffices to check $X^{-1}((-\infty, x]) \in \mathcal{F}$ for all $x \in \mathbb{R}$ to see if X is random variable. Given a random variable X and some set $B \in \mathcal{B}(\mathbb{R})$, often for brevity, the set $\{\omega \in \Omega : X(\omega) \in A\}$ is going to be denoted as $\{X \in A\}$. Furthermore if A is an interval, we may re-write $\{\omega \in \Omega : a < X(\omega) < b\}$ as $\{a < X < b\}$ for brevity.

Proposition 12. (Composition of Random Variables) If $X : (\Omega, \mathcal{F}) \to (S, \mathcal{A})$ and $Y : (S, \mathcal{A}) \to (T, \mathcal{B})$ are random variables, then $Y \circ X : (\Omega, \mathcal{F}) \to (T, \mathcal{B})$ is a random variable.

Proposition 13. (Random Vectors) If $X_1, X_2, \ldots, X_n$ are random variables on $(\Omega, \mathcal{F})$, then the *random vector* $(X_1, X_2, \ldots, X_n) : \Omega \to \mathbb{R}^n$ is an $\mathbb{R}^n$-valued random variable.

Proof. $\mathcal{B}(\mathbb{R}^n)$ is generated by the rectangle sets $A_1 \times A_2 \times \cdots \times A_n$ where $A_i \in \mathcal{B}(\mathbb{R})$ for each $i = 1, 2, \ldots, n$. Thus, it suffices to check that the preimage of rectangles are measurable under $(X_1, X_2, \ldots, X_n)$. It follows that

$$\begin{aligned}
\left\{\omega \in \Omega : (X_1, X_2, \ldots, X_n)(\omega) \in \prod_{i=1}^{n} A_i\right\} &= \{\omega \in \Omega : (X_1(\omega), X_2(\omega), \ldots, X_n(\omega)) \in \prod_{i=1}^{n} A_i\} \\
&= \bigcap_{i=1}^{n} \{\omega \in \Omega : X_i(\omega) \in A_i\} \\
&= \bigcap_{i=1}^{n} X_i^{-1}(A_i) \in \mathcal{F}
\end{aligned}$$

since X_i are random variables and $\mathcal{F}$ is a σ-algebra. □

With the previous two results, we have the following immediate consequence

Proposition 14. If $X_1, X_2, \ldots, X_n$ are random variables on $(\Omega, \mathcal{F})$ and $f : (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is measurable, then $f(X_1, X_2, \ldots, X_n)$ is a random variable.

Example 12. (Random Sums) If $X_1, X_2, \ldots, X_n$ are random variables on $(\Omega, \mathcal{F})$ then so is $X_1 + X_2 + \cdots + X_n$ by choosing $f(x_1, x_2, \ldots, x_n) := x_1 + x_2 + \cdots + x_n$ in the previous result.

It is a common result from real analysis, that a continuous function $f : A \to B$ is characterized by the property that $f^{-1}(B)$ is open for every open set A. From this, it is immediate that every continuous function is measurable. Together with the above proposition, we have that if X is a random variable so is aX for all $a \in \mathbb{R}$, e^X, X^2, $\cos(X)$, and so on.

3.2 Sequences of Random Variables

We say that a sequence of random variables $(X_n)_{n=1}^{\infty}$ is increasing (resp. decreasing) if for each $n \in \mathbb{N}$, we have $X_n(\omega) \le X_{n+1}(\omega)$ (resp. $X_n(\omega) \ge X_{n+1}(\omega)$) for all $\omega \in \Omega$. Let $X : \Omega \to \mathbb{R}$ be a random variable, By $X_n \uparrow X$ we note that $(X_n)_{n\ge 1}$ is increasing and $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ for all $\omega \in \Omega$. Similarly, by $X_n \downarrow X$ we note that $(X_n)_{n=1}^{\infty}$ is decreasing and $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ for all $\omega \in \Omega$. For an arbitrary sequence of random variables $(X_n)_{n=1}^{\infty}$, we define the following functions:

$$\begin{aligned}
&\inf_{n\in\mathbb{N}} X_n : \Omega \to \mathbb{R} \quad \text{by} && \inf_{n\in\mathbb{N}} X_n(\omega) := \inf_{n\in\mathbb{N}} \{X_n(\omega) : n \in \mathbb{N}\} \\
&\sup_{n\in\mathbb{N}} X_n : \Omega \to \mathbb{R} \quad \text{by} && \sup_{n\in\mathbb{N}} X_n(\omega) := \sup_{n\in\mathbb{N}} \{X_n(\omega) : n \in \mathbb{N}\} \\
&\liminf_{n\to\infty} X_n : \Omega \to \mathbb{R} \quad \text{by} && \liminf_{n\to\infty} X_n(\omega) := \lim_{n\to\infty} \inf_{m\ge n} X_m(\omega) \\
&\limsup_{n\to\infty} X_n : \Omega \to \mathbb{R} \quad \text{by} && \limsup_{n\to\infty} X_n(\omega) := \lim_{n\to\infty} \sup_{m\ge n} X_m(\omega)
\end{aligned}$$

It is clear that all four of these functions defined on Ω are random variables. Fix $\omega \in \Omega$, and note that $\inf_{n\in\mathbb{N}} X_n \leq x$ if and only if there exists some n, such that $X_n(\omega) \leq x$. Hence,

$$\begin{aligned}\left(\inf_{n\in\mathbb{N}} X_n\right)^{-1}((-\infty, x]) &= \{\omega \in \Omega : \inf_{n\in\mathbb{N}} X_n(\omega) \leq x\} \\ &= \bigcup_{n=1}^{\infty} \{\omega \in \Omega : X_n(\omega) \leq x\} \\ &= \bigcup_{n=1}^{\infty} X_n^{-1}((-\infty, x]) \in \mathcal{F}.\end{aligned}$$

Since intervals of the form $(-\infty, x]$ generate the Borel σ-algebra, we know that $\inf_{n\in\mathbb{N}} X_n$ is a random variable as required. A similar argument can be used to show that $\sup_{n\in\mathbb{N}} X_n$ is a random variable. Now define the sequence of random variables $Y_n := \inf_{m\geq n} X_m$ for each n, and note that $(Y_n)_{n=1}^{\infty}$ forms an increasing sequence of random variables. Thus it follows that

$$\liminf_{n\to\infty} X_n = \lim_{n\to\infty} \inf_{m\geq n} X_m = \sup_{m\in\mathbb{N}} \inf_{m\geq n} X_m.$$

which is once again a random variable, by the first part. The argument for $\limsup_{n\to\infty} X_n$ is similar. Observe that these four defined random variables can take on infinite values, even if the original sequence of random variables only take on finite values. However this does not affect the measurability of the functions and we leave this as an exercise to verify why this is so.

Definition 8. A function $X : \Omega \to \mathbb{R}$ is called a *simple random variable* if it can be represented as

$$X = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i}$$

for some constants $a_1, \ldots, a_n \in \mathbb{R}$ and sets $A_1, \ldots, A_n \in \mathcal{F}$. That is X is a simple random variable if it is a linear combination of indicator random variables.

It can shown that simple random variables are indeed random variables, and their representation is not unique. Let X be a simple random variable and consider the representation $X(\omega) = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i}(\omega)$ where $A_1, A_2, ..., A_n \in \mathcal{F}$ and $a_1, a_2, ..., a_n \in \mathbb{R}$. Choose the sets $C_1, C_2, \ldots C_m \in \mathcal{F}$ were $C_i \cap C_j = \emptyset$ for each $i \neq j$, and $\bigcup_{j=1}^{m} C_j = \bigcup_{i=1}^{n} A_i$. Furthermore, for each A_i there exists $I_i \subseteq \{1, 2, \ldots, n\}$ such that $A_i = \bigcup_{j\in I_i} C_j$. Since we know $\mathbf{1}_{C_j}$ is a random variable, observe that

$$X = \sum_{i=1}^{n} \alpha_i \mathbf{1}_{A_i} = \sum_{j=1}^{m} \mathbf{1}_{C_j} \sum_{i=1, j\in I_i}^{n} \alpha_i = \sum_{j=1}^{m} \beta_j \mathbf{1}_{C_j}.$$

Now consider the function $Y : \mathbb{R}^N \to \mathbb{R}$ defined by $Y(x_1, x_2, \ldots, x_m) := \sum_{j=1}^{m} \beta_j x_j$, this is a linear function and thus Borel measurable. Furthermore, the evaluation map

$$(\mathbf{1}_{C_1}, \mathbf{1}_{C_2}, \ldots, \mathbf{1}_{C_m}) : (\Omega, \mathcal{F}) \to (\mathbb{R}^m, \mathcal{B}(\mathbb{R})^m)$$

defined by

$$(\mathbf{1}_{C_1}, \mathbf{1}_{C_2}, \ldots, \mathbf{1}_{C_m})(\omega) = (\mathbf{1}_{C_1}(\omega), \mathbf{1}_{C_2}(\omega), \ldots, \mathbf{1}_{C_m}(\omega))$$

is a random vector, as characteristic functions of measurable sets are random variables. Hence the composition $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ defined by

$$X(\omega) := Y \circ (\mathbf{1}_{C_1}, \mathbf{1}_{C_2}, \ldots, \mathbf{1}_{C_m})(\omega)$$

is a random variable, completing the proof.

Simple random variables are important in the theory or probability, namely in the developement of expected value as we will see in the next section. We present a result that shows that every random variable can be approximated by simple random variables.

Theorem 6. The function $X : \Omega \to \mathbb{R}$ is a random variable if and only if there exists a sequence $(X_n)_{n=1}^{\infty}$, of simple random variables such that

$$X(\omega) = \lim_{n\to\infty} X_n(\omega)$$

for all $\omega \in \Omega$.

Recall that point-wise limits of random variables are random variables, and simple random variables are also random variables. Therefore the limit of simple random variables are random variables, making one of the implications in the above result immediate. For the reverse-implication we prove the result constructively by first considering non-negative random variables and then the general case.

Theorem 7. Let the function $X : \Omega \to \mathbb{R}$ be a non-negative random variable. Then there exists a sequence $(X_n)_{n=1}^{\infty}$, of simple random variables $X_n : \Omega \to \mathbb{R}$ such that for all $\omega \in \Omega$

$$0 \leq X_n(\omega) \leq X_{n+1}(\omega) \leq X(\omega),$$

for each $n \in \mathbb{N}$ and

$$X(\omega) = \lim_{n\to\infty} X_n(\omega),$$

for all $\omega \in \Omega$.

Proof. Consider the collection of sets $\{A_{i,n}\}_{i,n\in\mathbb{N}}$, defined by

$$A_{i,n} := \{\omega \in \Omega : (i-1)/2^n \leq X(\omega) < i/2^n\} = \{(i-1)/2^n \leq X < i/2^n\}$$

for each $n \in \mathbb{N}$ and $1 \leq i \leq n2^n$. Now consider the sets $\{B_n\}_{n=1}^{\infty}$, defined by

$$B_n := \{\omega \in \Omega : X(\omega) \geq n\} = \{X \geq n\}$$

for each $n \in \mathbb{N}$. Define the sequence of functions, $(X_n)_{n=1}^{\infty}$ defined by

$$X_n = \sum_{i=1}^{n2^n} \frac{i-1}{2^n} \mathbf{1}_{A_{i,n}} + n\mathbf{1}_{B_n},$$

for each $n \in \mathbb{N}$. It is clear that X_n is a simple random variable for each n. Fix $\omega \in \Omega$ and let $n \in \mathbb{N}$ be arbitrary. Observe that the collection $\{A_{i,n}\}_{i=1}^{n2^n}$ is disjoint by construction, thus without loss of generality suppose $\omega \in A_{j,n}$ for some $1 \leq j \leq n2^n$. Then $X_n(\omega) = \frac{j-1}{2^n}$, since $I_{A_{i,n}}(\omega) = I_{B_n}(\omega) = 0$ for all $i \neq j$. Observe that $\omega \in A_{j,n}$ only if $\omega \in A_{2j,n+1}$ and so

$$X_{n+1}(\omega) = \frac{2j-1}{2^{n+1}} \geq \frac{j-1}{2^n} = X_n(\omega).$$

Now consider the case where $\omega \in B_n$. Then, $X_n(\omega) = n$. if $\omega \in B_{n+1}$, then $X_{n+1}(\omega) = n+1 > X_n(\omega)$. If $\omega \notin B_{n+1}$, then $\omega \in A_{(n+1)2^{(n+1)},n+1}$ and once again $X_{n+1}(\omega) = n \geq X_n(\omega)$. Since this exhausts all possible $\omega \in \Omega$, we conclude that $(X_n(\omega))_{n=1}^{\infty}$ is increasing for all $\omega \in \Omega$. Fix $n \in \mathbb{N}$ and suppose $X(\omega) \leq n$ for some $\omega \in \Omega$. If $X(\omega) = n$, then

$$X_n(\omega) = X(\omega) < X_n(\omega) + \frac{1}{2^n}$$

is immediate. If $X(\omega) < n$, then choose i such that $\omega \in A_{i,n}$. Then

$$X_n(\omega) = \frac{i-1}{2^n} \leq X(\omega) < \frac{i}{2^n} = \frac{i-1}{2^n} + \frac{1}{2^n} = X_n(\omega) + \frac{1}{2^n}.$$

Thus for all $\omega \in \Omega$, $X_n(\omega) \leq X(\omega) < X_n(\omega) + \frac{1}{2^n}$. Fix $\omega \in \Omega$, and let $\varepsilon > 0$ be arbitrary. Since X is a random variable, we know that $X(\omega) \in \mathbb{R}$ and so choose $N_1 \in \mathbb{N}$ such that $X_n(\omega) \leq N_1$ (hence $X_n(\omega) \leq X(\omega) < X_n(\omega) + \frac{1}{2^n}$ for $n > N_1$). Now choose $N_2 \in \mathbb{N}$ so that $\frac{1}{2^n} < \varepsilon$ for all $n > N_2$. Choose $N := \max(N_1, N_2)$, then for $n > N$ we know that by the previous argument

$$|X_n(\omega) - X(\omega)| < \frac{1}{2^n} < \varepsilon.$$

Since ε was arbitrary, we have that $X_n(\omega) \to X(\omega)$ for all $\omega \in \Omega$. $\square$

The above construction of the X_n is not unique, however care must be taken to ensure that the selection of the simple random variables yields the desired result as the following example shows.

Example 13. Let $X : \Omega \to \mathbb{R}$ be a non-negative random variable. Define the sequence of simple random variables by

$$X_n := \sum_{i=1}^{n^2} \frac{i-1}{n} \mathbf{1}_{\{(i-1)/n \leq X < i/n\}} + n\mathbf{1}_{X \geq n},$$

for each $n \in \mathbb{N}$. Determine if $X(\omega) = \lim_{n\to\infty} X_n(\omega)$ for all $\omega \in \Omega$.

We now relax the condition that X must be a non-negative random variable, by considering the following result.

Theorem 8. Let the function $X : \Omega \to \mathbb{R}$ be a random variable. Then there exists a sequence $(X_n)_{n=1}^{\infty}$, of simple random variables $X_n : \Omega \to \mathbb{R}$ such that for all $\omega \in \Omega$

$$0 \leq |X_n(\omega)| \leq |X_{n+1}(\omega)| \leq |X(\omega)|,$$

for each $n \in \mathbb{N}$ and

$$X(\omega) = \lim_{n\to\infty} X_n(\omega),$$

for all $\omega \in \Omega$.

Proof. For any arbitrary random variable X, we can decompose it into it's positive and negative components. That is, $X = X^+ - X^-$ where $X^+ := \frac{|X|+X}{2}$ and $X^- := \frac{|X|-X}{2}$ are non-negative random variables. By the Theorem 7, we can choose increasing sequences of simple random variables, $(X_n^+)_{n=1}^{\infty}$ and $(X_n^-)_{n=1}^{\infty}$ that converge point-wise to X^+ and X^- respectively. By linearity of limits, we have that

$$\lim_{n\to\infty} (X_n^+(\omega) - X_n^-(\omega)) = X^+(\omega) - X^-(\omega) = X(\omega)$$

for all $\omega \in \Omega$. Furthermore, we have for all $n \in \mathbb{N}$ and $\omega \in \Omega$:

$$|X_n^+(\omega) - X_n^-(\omega)| \leq X_n^+(\omega) + X_n^-(\omega) \leq X_{n+1}^+(\omega) + X_{n+1}^-(\omega) \leq X^+(\omega) + X^-(\omega) = |X(\omega)|.$$

Therefore, choosing the sequence of step functions $(X_n)_{n=1}^{\infty}$ defined by $X_n := X_n^+ + X_n^-$ achieves the desired result.

$\square$

With this theorem 6 follows directly. Observe that theorem 6 also has a mathematical interpretation, in that the space of random variables, is the smallest collection containing simple random variables and is closed under pointwise limits. More precisely, Let Ω be a non-empty sample space equipped with σ-algebra $\mathcal{F}$. Let $\mathscr{R}$ be an arbitrary collection of functions from Ω to $\mathbb{R}$ containing all the simple random variables. In addition suppose $\mathscr{R}$ has the property, that if $(X_n)_{n=1}^{\infty}$ is a sequence of functions converging to X, then $X \in \mathscr{R}$. Now let X denote an arbitrary random variable. By theorem 6, we may choose a sequence of simple random variables $(X_n)_{n=1}^{\infty}$ that converges pointwise to X. Of course for each $n \in \mathbb{N}$, we have $X_n \in \mathscr{R}$ and since the sequence converges to X, we have that $X \in \mathscr{R}$ by construction of $\mathscr{R}$. This shows that $\mathscr{R}$ contains all the random variables as X was arbitrary.

3.3 Law of Random Variables

We know that σ-algebras contain all events that are of interest on a sample space, and that random variables are functions of these events. However we can also indentify the events of interest purely based on the random variable that is defined on the sample space Ω.

Theorem 9. Let $X : \Omega \to S$ denote an arbitrary function between non-empty sets Ω and S. Suppose $\mathcal{A}$ is a σ-algebra on S. It follows that

1. The collection of subsets $\sigma(X) := \{X^{-1}(A) : A \in \mathcal{A}\}$ is a σ-algebra on Ω. This σ-algebra is called the *σ-algebra generated by X*.

2. If $\mathcal{F}$ is a σ-algebra on Ω such that $X : (\Omega, \mathcal{F}) \to (S, \mathcal{A})$ is an S-valued random variable, then $\sigma(X) \subset \mathcal{F}$.

3. Let $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a random variable and set $\mathcal{A}_0 := \{(-\infty, x] : x \in \mathbb{R}\}$. Then $\sigma(X) = \sigma(X^{-1}(\mathcal{A}_0))$.

Proof. Observe that $\varnothing = X^{-1}(\varnothing) \in \mathcal{A}$, so $\varnothing \in \sigma(X)$. If $B \in \sigma(X)$, then $B = X^{-1}(A)$ for some $A \in \mathcal{A}$. Then $B^c = (X^{-1}(A))^c = X^{-1}(A^c) \in \sigma(X)$ showing that $\sigma(X)$ is closed under complements. Finally suppose $\{B_n\}_{n=1}^{\infty}$ is a collection in $\sigma(X)$. Choose $\{A_n\}_{n=1}^{\infty}$ in $\mathcal{A}$ such that $X^{-1}(A_n) = B_n$ for each $n \in \mathbb{N}$. Then

$$\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} X^{-1}(A_n) = X^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right) \in \sigma(X),$$

since $\mathcal{A}$ is closed under denumerable unions. Thus $\sigma(X)$ is a σ-algebra on Ω as required. Fix $B \in \sigma(X)$ and since $X : (\Omega, \sigma(X)) \to (S, \mathcal{A})$, we have that $A := X(B) \in \mathcal{A}$. However by measurability of $X : (\Omega, \mathcal{F}) \to (S, \mathcal{A})$ it follows that $B = X^{-1}(A) = X^{-1}(X(B)) \in \mathcal{F}$ and thus $\sigma(X) \subset \mathcal{F}$. Observe that $\sigma(\mathcal{A}_0) = \mathcal{B}(\mathbb{R})$ and $\sigma(X) = X^{-1}(\mathcal{B}(\mathbb{R}))$ by construction of the Borel σ-algebra and the definition of a σ-algebra generated by a function. We require that $\sigma(X^{-1}(\mathcal{A}_0)) = \sigma(X)$, to that end we have $X^{-1}(\sigma(\mathcal{A}_0)) = \sigma(X^{-1}(\mathcal{A}_0))$ by the first problem of the first problem set. Thus, $\sigma(X^{-1}(\mathcal{A}_0)) = X^{-1}(\sigma(\mathcal{A}_0)) = X^{-1}(\mathcal{B}(\mathbb{R})) = \sigma(X)$ completing the proof.

□

Definition 9. (Law of a Random Variable) Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and let S denote a non-empty set equipped with a σ-algebra $\mathcal{A}$. Suppose $X : (\Omega, \mathcal{F}) \to (S, \mathcal{A})$ is an S-valued random variable. For all $B \in \mathcal{A}$ define

$$\mathbb{P}_X(B) := \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\}) = \mathbb{P}(X \in B) = \mathbb{P}(X^{-1}(B)).$$

It is clear that $\mathbb{P}_X$ defines a probability measure on $\mathcal{A}$, and is called the *law of X*. In particular if $S = \mathbb{R}$ and $\mathcal{A} = \mathcal{B}(\mathbb{R})$, then any random variable X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ defines a probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$. The law of a random variable is uniquely identified by it's distribution function.

Definition 10. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and suppose $X : \Omega \to \mathbb{R}$ is a random variable. X defines a function $F_X : \mathbb{R} \to [0, 1]$ by

$$F_X(x) := \mathbb{P}(X \le x)$$

called the *cumulative distribution function* of X.

It can easily be verified that the cumulative distribution function is non-decreasing, right-continuous and satisfies $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$. If an arbitrary function satisfies these four properties, we call such a function a distribution function. The distribution function of a random variable satisfies the conditions of the function described in theorem 3, and thus defines a unique probability measure on $\mathcal{B}(\mathbb{R})$. This measure is precisely the law of X.

Proposition 15. If $F : \mathbb{R} \to \mathbb{R}$ is a distribution function, then there exists a unique random variable X such that it's cumulative distribution function is $F_X = F$.

Proof. Consider the probability space $([0, 1], \mathcal{B}([0, 1]), \lambda)$ where λ is the Lebesgue measure on $[0, 1]$. Define the random variable

$$X(\omega) := \sup\{t \in \mathbb{R} : F(t) < \omega\}.$$

It remains to show that X is a random variable, and that $F_X = F$. That is, it suffices to show that

$$\{\omega \in (0, 1) : X(\omega) \le x\} = \{\omega \in (0, 1) : \omega \le F(x)\}.$$

Fix $x \in \mathbb{R}$ and choose $0 < \omega < 1$ such that $X(\omega) \leq x$. Since $X(\omega) = \sup\{t \in \mathbb{R} : F(t) < \omega\}$, this means that for all $\delta > 0$, we have that $x + \delta \notin \{t : F(t) < \omega\}$. Therefore $F(x + \delta) \geq \omega$. Since F is right-continuous we have as $\delta \to 0$, that $F(x) \geq \omega$. Conversely choose $0 < \omega < 1$ such that $\omega \leq F(x)$. Then $x \notin \{y : F(y) < \omega\}$. As F is non-decreasing we have that x is an upper bound of the set $\{y : F(y) < \omega\}$. Hence $X(\omega) = \sup\{y \in \mathbb{R} : F(y) < \omega\} \leq x$. Therefore

$$\{\omega \in (0,1) : X(\omega) \leq x\} = \{\omega \in (0,1) : \omega \leq F(x)\}$$

and thus $F_X(x) = \mathbb{P}(X \leq x) = F(x)$ for all $x \in \mathbb{R}$ as required.

□

This random variable defined above is caled the *quantile function* and is often denoted by

$$F^{-1}(\omega) := \sup\{t \in \mathbb{R} : F(t) < \omega\}$$

for all $\omega \in (0,1)$. Note that F^{-1} is simply convention and does not mean it is the inverse of F (F may not even have an inverse). When F is continuous, it turns out it's quantile function is indeed the inverse. We can see that F^{-1} is a random variable on the probability space $([0,1], \mathcal{B}([0,1]), \lambda)$ where λ is the Lebesgue measure.

3.4 Independent Random Variables

Recall that events $A_1, \ldots, A_n$ are independent if

$$\mathbb{P}\Big(\bigcap_{i=1}^{n} A_i\Big) = \prod_{i=1}^{n} \mathbb{P}(A_i).$$

We can generalize this to collections of sets, and in particular σ-algebras as well. That is, we say that the σ-algebras $\mathcal{F}_1, \mathcal{F}_2, \dot{;} \mathcal{F}_n$ are independent if

$$\mathbb{P}\Big(\bigcap_{i=1}^{n} A_i\Big) = \prod_{i=1}^{n} \mathbb{P}(A_i),$$

for all $A_i \in \mathcal{F}_i$ and $i = 1, 2, \ldots, n$. A collection of σ-algebras $\{\mathcal{F}_i\{_{i \in I}$ where I is an index set, is independent if any finite subcollection of the $\{\mathcal{F}_i\{_{i \in I}$ are independent. Since random variables identify a unique σ-algebra in which they are measurable with respect to, we can now consider a notion of independence of random variables.

Definition 11. We say that the random variables $X_1, X_2, \ldots, X_n$ are independent if the σ-algebras $\sigma(X_1), \sigma(X_2), \ldots, \sigma(X_n)$ are independent. An arbitrary collection of random variables $(X_i)_{i \in I}$, are independent if and only if each finite sub-collection of $(X_i)_{i \in I}$ are independent.

Of course if we consider the indicator random variables $(\mathbf{1}_{A_i})_{i \in I}$ where $\{A_i\}_{i \in I}$ is a collection of events, the above definition reduces to the definition of independent events. Furthermore it can be shown that the random variables $X_1, \ldots, X_n$ are independent if and only if for all $B_i \in \mathcal{B}(\mathbb{R})$, $i = 1, 2, \ldots, n$ we have

$$\mathbb{P}(X_1 \in B_1, X_2 \in B_2, \ldots, X_n \in B_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in B_i).$$

We present the following result without proof, which allows us an easy way to verify whether σ-algebras are independent by checking if their generators are independent.

Proposition 16. Suppose the collections of sets $\mathcal{F}_1, \ldots, \mathcal{F}_n$ are independent. If $\Omega \in \mathcal{F}_i$ and $\mathcal{F}_i$ is closed under intersection for each $i = 1, 2, \ldots, n$ then $\sigma(\mathcal{F}_1) \ldots, \sigma(\mathcal{F}_n)$ are independent.

Proposition 17. (Independent Random Vectors)

1. Let $X_1, ..., X_n$ be random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and consider the random vector $(X_1, ..., X_n) : \Omega \to (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. Then it follows that

$$\sigma((X_1, ..., X_n)) = \sigma(\cup_{i=1}^{n} \sigma(X_i)).$$

2. Suppose $X_{i,j}$ for each $1 \le i \le n$, $1 \le j \le m$ are independent random variables. Define the random vectors $X_i = (X_{i,1}, \dots, X_{i,m})$ for each $i = 1, 2, \dots, n$. Then it follows $X_1, \dots, X_n$ are independent random vectors.

3. Suppose $X_{i,j}$ for each $1 \le i \le n$, $1 \le j \le m$ are independent random variables. Suppose the functions $f_i : \mathbb{R}^m \to \mathbb{R}$ are measurable, and let $Y_i := f_i(X_{i,1}, ..., X_{i,m})$ for each $i = 1, 2..., n$. Then the random variables $Y_1, ..., Y_n$ are independent.

Proof. Consider the n-dimensional rectangles of the form $\prod_{i=1}^n A_i$ where each $A_i \in \mathcal{B}(\mathbb{R})$ for $i = 1, 2, \dots, n$. Recall that rectangles of this form generate $\mathcal{B}(\mathbb{R}^n)$. Observe that these sets are the inverse images of the random vector $X = (X_1, X_2, \dots X_n)$. That is, sets of the form $\bigcap_{i=1}^n X_i^{-1}(A_i)$ generate $\sigma((X_1, X_2, \dots X_n))$. However, $X_i^{-1}(A_i) \in \sigma(X_i)$ for each $i = 1, 2, \dots, n$, so $\bigcap_{i=1}^n X_i^{-1}(A_i) \in \sigma(\bigcup_{i=1}^n \sigma(X_i))$, showing

$$\sigma((X_1, X_2, \dots X_n)) \subset \sigma\Big(\bigcup_{i=1}^n \sigma(X_i)\Big).$$

For the reverse inclusion, observe that the inverse image of the rectangle, $A_1 \times \Omega^{n-1}$ under the random vector $(X_1, X_2, \dots X_n)$ is $X_1^{-1}(A_1)$ which satisfies $X_1^{-1}(A_1) \in \sigma((X_1, X_2, ..., X_n))$. However the collection $\{X_1^{-1}(A_1) : A_1 \in \mathcal{B}(\mathbb{R})\}$ generate the σ-algebra $\sigma(X_1)$ and so

$$\sigma(X_1) \subset \sigma((X_1, X_2, \dots X_n)).$$

By identical arguments, it can be shown that $\sigma(X_i) \subset \sigma((X_1, X_2 \dots X_n))$ for all $i = 1, 2, ..., n$. This shows that

$$\sigma\Big(\bigcup_{i=1}^n \sigma(X_i)\Big) \subset \sigma((X_1, X_2, \dots X_n)),$$

completing the proof of (1).

For (2) suppose $\sigma(X_{i,j})$ for $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, m$ are independent. Define $\mathcal{F}_i$ to be the collection of sets of the form $\cap_{j=1}^m A_{i,j}$ where $A_{i,j} \in \sigma(X_{i,j})$. Observe $\mathcal{F}_i \subset \sigma(\bigcup_{i=1}^m \sigma(X_{i,j}))$ and that $\mathcal{F}_i$ is closed under intersections. By inspection it is clear that each $\mathcal{F}_i$ contains Ω, and $\bigcup_{j=1}^m \sigma(X_{i,j})$. Hence $\sigma(\mathcal{F}_i) = \sigma(X_i)$ for all $i = 1, 2, ..., n$. Since the $\mathcal{F}_1, \dots, \mathcal{F}_n's$ are independent, so are the random vectors $X_1, \dots, X_n$ by Proposition 16.

Finally we let $X_i := (X_{i,1}, \dots, X_{i,m})$ and choose $\mathcal{G}_i := \sigma(\bigcup_{i=1}^m \sigma(X_{i,j}))$ for $i = 1, 2 \dots, n$. By (1), $\mathcal{G}_1, \mathcal{G}_2, \dots \mathcal{G}_n$ are independent σ-algebras. From (2), we deduce that $\sigma(Y_i) \subset \mathcal{G}_i$ for each i and so $\sigma(Y_1), ..., \sigma(Y_n)$ are independent σ-algebras. Equivalently we have that $Y_1, Y_2, \dots Y_n$ are independent random variables. ☐

Definition 12. If $X_1, \dots, X_n$ are random variables defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ then the function $F : \mathbb{R}^n \to \mathbb{R}$ defined by

$$F(x_1, \dots, x_n) := \mathbb{P}(X_1 \le x_1, X_2 \le x_2, \dots, X_n \le x_n)$$

is called the *joint distribution function* of $X_1, X_2, \dots, X_n$, and is often denoted by $F_{X_1,\dots,X_n}$

Proposition 18. If $X_1, \dots, X_n$ are random variables defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then they are independent if and only if

$$F(x_1, \dots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n)$$

for all $x_1, \dots, x_n \in \mathbb{R}$.

We have looked at measures that are defined on the Borel σ-algebra. However it should be noted that we can do the same for the Borel σ-algebra on any Euclidean space. Let us start with the $\mathbb{R}^2$ case. Let $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ and $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ denote two probability spaces. We construct their *product space* given by

$$(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mathbb{P}_1 \times \mathbb{P}_2)$$

as follows. The sample space $\Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$ is the usual Cartesian product between sets. Now consider the collection of *rectangle* sets of the form $A \times B \subset \Omega_1 \times \Omega_2$ where $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$. We define the product σ-algebra by

$$\mathcal{F}_1 \otimes \mathcal{F}_2 := \sigma(\{A \times B : A \in \mathcal{F}_1, B \in \mathcal{F}_2\}).$$

Finally we have the unique probability measure $\mathbb{P}_1 \times \mathbb{P}_2$ defined on this product σ-algebra $\mathcal{F}_1 \otimes \mathcal{F}_2$ (whose existence and uniqueness is beyond the scope of this course) that satisfies the property $(\mathbb{P}_1 \times \mathbb{P}_2)(A \times B) = \mathbb{P}_1(A)\mathbb{P}_2(B)$ for all $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$. Observe that we can re-iterate this procedure indefinitely to construct probability measures of any finite dimension. With this we have the Borel σ-algebra on $\mathbb{R}^n$ defined by $\mathcal{B}(\mathbb{R}^n) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) \otimes \cdots \otimes \mathcal{B}(\mathbb{R})$ along with the corresponding product probability measure $\mathbb{P}_1 \times \cdots \times \mathbb{P}_n$.

.

Proposition 19. Let $\mathcal{F}_1$ be a σ-algebra on Ω_1 and let $\mathcal{F}_2$ be a σ-algebra on Ω_2. Let $\mathcal{G}_1$ be a collection of subsets of Ω_1 such that $\sigma(\mathcal{G}_1) = \mathcal{F}_1$. Let $\mathcal{G}_2$ be a collection of subsets of Ω_2 such that $\sigma(\mathcal{G}_2) = \mathcal{F}_2$. Suppose in addition, that $\Omega_1 \in \mathcal{G}_1$ and $\Omega_2 \in \mathcal{G}_2$. Prove that:

$$\mathcal{G}_1 \otimes \mathcal{G}_2 = \mathcal{F}_1 \otimes \mathcal{F}_2.$$

Proof. By definition, we have that:

$$\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\{A \times B : A \in \mathcal{F}_1, B \in \mathcal{F}_2\})$$

Observe that $\mathcal{G}_1 \subset \mathcal{F}_1$ and $\mathcal{G}_2 \subset \mathcal{F}_2$ where both $\mathcal{F}_1$ and $\mathcal{F}_2$ are both σ-algebras. Hence we immediately get

$$\mathcal{G}_1 \otimes \mathcal{G}_2 = \sigma(\{A \times B : A \in \mathcal{G}_1, B \in \mathcal{G}_2\}) \subset \mathcal{F}_1 \otimes \mathcal{F}_2.$$

Conversely, define $\mathcal{F}_{\Omega_1} := \{A \subset \Omega_1 : A \times \in \mathcal{G}_1 \otimes \mathcal{G}_2\}$. It is immediate that $\Omega_1 \in \mathcal{F}_{\Omega_1}$. Now if $A \in \mathcal{F}_{\Omega_1}$ then $A \times \Omega_2 \in \mathcal{G}_1 \otimes \mathcal{G}_2$ and since $\mathcal{G}_1 \otimes \mathcal{G}_2$ is a σ-algebra we have that

$$(\Omega_1 \backslash A) \times \Omega_2 = (\Omega_1 \times \Omega_2) \backslash (A \times \Omega_2) \in \mathcal{G}_1 \otimes \mathcal{G}_2$$

and thus $(\Omega_1 \backslash A) \in \mathcal{F}_{\Omega_1}$. Now suppose $\{A_n\}_{n=1}^{\infty}$ is a collection in $\mathcal{F}_{\Omega_1}$, then $A_n \times \Omega_2 \in \mathcal{G}_1 \otimes \mathcal{G}_2$ for each $n \in \mathbb{N}$. As $\mathcal{G}_1 \otimes \mathcal{G}_2$ is a σ-algebra we have

$$\left(\bigcup_{n=1}^{\infty} A_n \right) \times \Omega_2 = \bigcup_{n=1}^{\infty} (A_n \times \Omega_2) \in \mathcal{G}_1 \otimes \mathcal{G}_2$$

and thus $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}_{\Omega_1}$ showing that $\mathcal{F}_{\Omega_1}$ is a σ-algebra. Fix $A \in \mathcal{G}_1$, then $A \times \Omega_2 \in \mathcal{G}_1 \otimes \mathcal{G}_2$ by construction and thus $A \in \mathcal{F}_{\Omega_1}$. Hence $\mathcal{G}_1 \subset \mathcal{F}_{\Omega_1}$, and so $\mathcal{F} = \sigma(\mathcal{G}_1) \subset \mathcal{F}_{\Omega_1}$ by minimality and the fact that $\Omega_1 \subset \mathcal{G}_1$. By symmetry we can conclude that $\mathcal{F}_2 = \sigma(\mathcal{G}_2) \subset \mathcal{F}_{\Omega_2}$ where $\mathcal{F}_{\Omega_2} := \{B \subset \Omega_2 : \Omega_1 \times B \in \mathcal{G}_1 \otimes \mathcal{G}_2\}$ is a σ-algebra. Now fix $A \times B \in \mathcal{F}_1 \times \mathcal{F}_2$. Then by the above arguments we have that $A \in \mathcal{F}_{\Omega_1}$ and $B \in \mathcal{F}_{\Omega_2}$ and thus $(A \times \Omega_2), (\Omega_1 \times B) \in \mathcal{G}_1 \otimes \mathcal{G}_2$. Hence,

$$A \times B = (A \times \Omega_1) \cap (\Omega_1 \times B) \in \mathcal{G}_1 \otimes \mathcal{G}_2$$

showing that $\mathcal{F}_1 \otimes \mathcal{F}_2 \subset \mathcal{G}_1 \otimes \mathcal{G}_2$. We conclude that $\mathcal{F}_1 \otimes \mathcal{F}_1 \subset \mathcal{G}_1 \otimes \mathcal{G}_2$ and thus

$$\mathcal{F}_1 \otimes \mathcal{F}_1 = \mathcal{G}_1 \otimes \mathcal{G}_2$$

as required. $\square$

We can also provide an alternate proof using the fact that projection mappins are Borel-measurable.

Proof. Consider the projection function $\pi_1 : \Omega_1 \times \Omega_2 \to \Omega_1$ defined by $\pi_1(\omega_1, \omega_2) = \omega_1$. Then for each $A \subseteq \Omega_1$, we have that

$$\pi_1^{-1}(A) = \{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 : \pi_1(\omega_1, \omega_2) = \omega_1 \in A\} = A \times \Omega_2.$$

By the final result in the first problem set, we know that the pre-image and the σ-algebra generator operators are commutative. Hence,

$$\sigma(\pi_1^{-1}(\mathcal{G}_1)) = \pi_1^{-1}(\sigma(\mathcal{G}_1)) = \pi_1^{-1}(\mathcal{F}_1) = \{A \times \Omega_2 : A \in \mathcal{F}_1\}.$$

Hence $\sigma(\{A \times \Omega_2 : A \in \mathcal{G}_1\}) = \{A \times \Omega_2 : A \in \mathcal{F}_1\} =: \mathcal{F}_{\Omega_1}$. Similarly define the projection map $\pi_2 : \Omega_1 \times \Omega_2 \to \Omega_2$ by $\pi_2(\omega_1, \omega_2) = \omega_2$. Then for each $B \subseteq \Omega_2$, we have that

$$\pi_2^{-1}(B) = \{(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 : \pi_1(\omega_1, \omega_2) = \omega_2 \in B\} = \Omega_1 \times B.$$

Similarly we conclude that $\sigma(\{\Omega_1 \times B : B \in \mathcal{G}_2\}) = \{\Omega_1 \times B : B \in \mathcal{F}_2\} =: \mathcal{F}_{\Omega_2}$. Observe that

$$\{A \times \Omega_2 : A \in \mathcal{G}_1\} = \{A \times B : A \in \mathcal{G}_1, B = \Omega_2\} \subseteq \{A \times B : A \in \mathcal{G}_1, B \in \mathcal{G}_2\}$$
$$\{\Omega_1 \times B : B \in \mathcal{G}_2\} = \{A \times B : A = \Omega_1, B \in \mathcal{G}_2\} \subseteq \{A \times B : A \in \mathcal{G}_1, B \in \mathcal{G}_2\}$$

as $\Omega_1 \in \mathcal{G}_1$ and $\Omega_2 \in \mathcal{G}_2$. Furthermore

$$\mathcal{H} := \sigma\left(\{A \times \Omega_2 : A \in \mathcal{G}_1\} \cup \{\Omega_1 \times B : G \in \mathcal{G}_2\}\right) \subseteq \sigma(\{A \times B : A \in \mathcal{G}_1, B \in \mathcal{G}_2\}) =: \mathcal{G}_1 \otimes \mathcal{G}_2.$$

By properties of σ-algebras we have $\sigma(\{A \times \Omega_2 : A \in \mathcal{G}_1\}) \subseteq \sigma(\mathcal{H}) \subseteq \mathcal{H}$, and
$\sigma(\{\Omega_1 \times B : B \in \mathcal{G}_2\}) \subseteq \sigma(\mathcal{H}) \subseteq \mathcal{H}$. Therefore we conclude:

$$\sigma(\{A \times \Omega_2 : A \in \mathcal{F}_1\} \cup \{\Omega_1 \times B : B \in \mathcal{G}_2\}) \subseteq \mathcal{H}.$$

Fix $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$ and observe that

$$\begin{aligned} A \times B &= \{(a, b) : a \in A, b \in B\} \\ &= \{(a, b) : a \in A\} \cap \{(a, b) : b \in B\} \\ &= (A \times \Omega_2) \cap (\Omega_1 \times B) \\ &= [(A \times \Omega_2) \cup (\Omega_1 \times B)]^c \in \mathcal{H} \subseteq \mathcal{G}_1 \otimes \mathcal{G}_2. \end{aligned}$$

This shows that $\{A \times B : A \in \mathcal{F}_1, B \in \mathcal{F}_2\} \subseteq \sigma(\{A \times B : A \in \mathcal{G}_1, B \in \mathcal{G}_2\})$. Hence we conclude that

$$\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\{A \times B : A \in \mathcal{F}_1, B \in \mathcal{F}_2\}) = \sigma(\sigma(\{A \times B : A \in \mathcal{G}_1, B \in \mathcal{G}_2\})) \subseteq \mathcal{G}_1 \otimes \mathcal{G}_2.$$

Conversely, observe that $\mathcal{G}_1 \subseteq \mathcal{F}_1$ and $\mathcal{G}_2 \subseteq \mathcal{F}_2$ where both $\mathcal{F}_1$ and $\mathcal{F}_2$ are both σ-algebras. Hence we immediately get

$$\mathcal{G}_1 \otimes \mathcal{G}_2 = \sigma(\{A \times B : A \in \mathcal{G}_1, B \in \mathcal{G}_2\}) \subseteq \mathcal{F}_1 \otimes \mathcal{F}_2.$$

□

The above result readily generalizes for the product of finitely many σ-algebras. Hence, we can have alternative representations for the generating sets of the Borel σ-algebra. Note that every open set $U \subset \mathbb{R}^n$ is a countable union of n-dimensional open rectangle sets $\prod_{i=1}^n (a_i, b_i)$, hence the $\mathcal{B}(\mathbb{R}^n)$ is generated by the open rectangles in $\mathbb{R}^n$. In this course, we will take for granted that if two measures coincide on a generator of a σ-algebra, then the measures coincide on the entire σ-algebra as well. This result is known as the Carathéodory Extension Theorem and it's proof is beyond the scope of this course.

Given random variables $X_1, \ldots, X_n$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we obtain n probability spaces $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_{X_1}), (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_{X_2}), \ldots, (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_{X_n})$. Their product probability space is given by $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathbb{P}_{X_1} \times \mathbb{P}_{X_2} \times \cdots \times \mathbb{P}_{X_n})$. On the other hand, the random vector $(X_1, \ldots, X_n)$ induces a unique law, and defines a probablity space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathbb{P}_{(X_1,\ldots,X_n)})$, where

$$\mathbb{P}_{(X_1,\ldots,X_n)}(B) = \mathbb{P}((X_1, \ldots, X_n) \in B).$$

for all $B \in \mathcal{B}(\mathbb{R}^n)$ In general these two measures are not equal, and it is the latter measure that corresponds to the joint distribution function $F_{X_1,\ldots,X_n}$. However it turns out the measures coincide when the random variables are independent.

Theorem 10. Let $X_1, \dots, X_n$ be random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then $X_1, \dots, X_n$ are independent if and only if

$$(\mathbb{P}_{X_1} \times \cdots \times \mathbb{P}_{X_n})(B) = \mathbb{P}_{(X_1,\dots,X_n)}(B)$$

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

Proof. Suppose $X_1, \dots, X_n$ are independent and fix $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$. Then it follows that $\{X_i \in B_i\} \in \sigma(X_i)$ for each $i = 1, 2, \dots, n$, and these sets are independent events. Thus,

$$\begin{aligned}
\mathbb{P}_{(X_1,\dots,X_n)}(B_1 \times \cdots \times B_n) &= \mathbb{P}((X_1, \dots, X_n) \in B_1 \times \cdots \times B_n) \\
&= \mathbb{P}\Big(\bigcap_{i=1}^{n} \{X_i \in B_i\}\Big) \\
&= \mathbb{P}(X_1 \in B_1) \cdots \mathbb{P}(X_n \in B_n) \\
&= \mathbb{P}_{X_1}(B_1) \cdots \mathbb{P}_{X_n}(B_n) \\
&= (\mathbb{P}_{X_1} \times \cdots \times \mathbb{P}_{X_n})(B_1 \times \cdots \times B_n)
\end{aligned}$$

Since, rectangles of the form $B_1 \times \cdots \times B_n$ generate the Borel σ-algebra on $\mathbb{R}^n$, we have that the above holds for all $B \in \mathcal{B}(\mathbb{R}^n)$. Conversely, suppose that the product measure coincides with the joint law of the random vector. That is, $(\mathbb{P}_{X_1} \times \cdots \times \mathbb{P}_{X_n})(B) = \mathbb{P}_{(X_1,\dots,X_n)}(B)$ for all $B \in \mathcal{B}(\mathbb{R}^n)$. Fix $A_i \in \sigma(X_i)$ for each $i = 1, 2, \dots, n$. Choose $B_i \in \mathcal{B}(\mathbb{R})$ such that $A_i = \{X_i \in B_i\}$, for each $i = 1, 2, \dots, n$. It follows that,

$$\begin{aligned}
\mathbb{P}\Big(\bigcap_{i=1}^{n} A_i\Big) &= \mathbb{P}\Big(\bigcap_{i=1}^{n} \{X_i \in B_i\}\Big) \\
&= \mathbb{P}((X_1, \dots, X_n) \in B_1 \times \cdots \times B_n) \\
&= \mathbb{P}_{(X_1,\dots,X_n)}(B_1 \times \cdots \times B_n) \\
&= (\mathbb{P}_{X_1} \times \cdots \times \mathbb{P}_{X_n})(B_1 \times \cdots \times B_n) \\
&= \mathbb{P}_{X_1}(B_1) \times \cdots \times \mathbb{P}_{X_n}(B_n) \\
&= \mathbb{P}(A_1) \cdots \mathbb{P}(A_n).
\end{aligned}$$

Since this holds for all $A_i \in \sigma(X_i)$ for each $i = 1, 2, \dots, n$, we have that $\sigma(X_1), \dots, \sigma(X_n)$ are independent as desired.

□

Example 14. Let $X_1, \dots, X_n$ be random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If the c.d.f. $F(x_1, \dots, x_n)$ has **density** $f(x_1, \dots, x_n)$ that is

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f(y_1, \dots, y_n)\, dy_1 ... dy_n$$

and if $f(x_1, \dots, x_n) = g_1(x_1) \cdots g_n(x_n)$, where $g_i \geq 0$ is a measurable function $i = 1, \dots, n$, then the $X_1, \dots, X_n$ are independent.

Proof. Let $p_i := \int_{-\infty}^{\infty} g_i(y_i) dy_i$ for each $i = 1, 2, \dots, n$. By Fubini's Theorem, we have

$$\begin{aligned}
1 &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(y_1, ..., y_n)\, dy_1 ... dy_n \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g_1(y_1) \cdots g_n(y_n)\, dy_1 ... dy_n \\
&= \prod_{i=1}^{n} \int_{-\infty}^{\infty} g_i(y_i) dy_i = \prod_{i=1}^{n} p_i.
\end{aligned}$$

Furthermore for fixed $i = 1, 2, \dots, n$,

$$F_{X_i}(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{x_i} \cdots \int_{-\infty}^{\infty} g_1(y_1) \cdots g_n(y_n) dy_1 \dots dy_n.$$
$$= \int_{-\infty}^{x_i} g_i(y_i) dy_i \prod_{j \neq i} p_j.$$

Putting it all together,

$$\begin{aligned} F(x_1, x_2, \dots, x_n) &= \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f(y_1, \dots, y_n) dy_1 \dots y_n. \\ &= \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} g_1(y_1) \cdots g)_2(y_n) dy_1 \dots y_n. \\ &= \prod_{i=1}^{n} \int_{-\infty}^{x_i} g_i(y_i) dy_i \\ &= \prod_{i=1}^{n} \frac{F_{X_i}(x_i)}{\prod_{j \neq i} p_j} \\ &= \prod_{i=1}^{n} F_{X_i}(x_i) \left(\prod_{j \neq i} p_j \right)^{1-n} \\ &= \prod_{i=1}^{n} F_{X_i}(x_i), \end{aligned}$$

Showing that $X_1, X_2, \dots, X_n$ are independent random variables. □

3.5 Problems

Problem 16. Let $(X_n)_{n=1}^{\infty}$ be a sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Prove that the following sets are measurable.

1. $\{\omega \in \Omega : \sup_{n \in \mathbb{N}} X_n(\omega) = \infty\}$
2. $\{\omega \in \Omega : \inf_{n \in \mathbb{N}} X_n(\omega) = -\infty\}$
3. $\{\omega \in \Omega : \limsup_{n \to \infty} X_n(\omega) = \infty\}$
4. $\{\omega \in \Omega : \liminf_{n \to \infty} X_n(\omega) = -\infty\}$

Problem 17. Let $(X_n)_{n=1}^{\infty}$ be a sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let X be a random variable defined on the same probability space and prove that the following sets are measurable:

1. $\{\omega \in \Omega : \lim_{n \to \infty} X_n(\omega) = X(\omega)\}$
2. $\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon\}$ for all $\varepsilon > 0$.

Problem 18. Let $(X_n)_{n=1}^{\infty}$ be a sequence of independent random variables, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

1. If $\inf_{n \in \mathbb{N}} P(X_n > M) > 0$ for some $M > 0$, prove that $\mathbb{P}(\sum_{n=1}^{\infty} X_n = \infty) = 1$.
2. Prove that $\sum_{n=1}^{\infty} \mathbb{P}(X_n > M) < \infty$ for some $M > 0$ if and only if $\mathbb{P}(\sup_{n \in \mathbb{N}} X_n < \infty) = 1$.

Problem 19. Let $(X_n)_{n=1}^{\infty}$ be a sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Derive a necessary and sufficient condition for $\mathbb{P}(\limsup_{n \to \infty} X_n < \infty) = 1$. Prove that your answer is correct.

Problem 20. Let $(X_n)_{n=1}^{\infty}$ be a sequence of independent and identically distributed exponential random variables with parameter $\lambda = 1$. Show that:

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{X_n}{\ln n} = 1\right) = 1$$

Problem 21. Let $(Z_n)_{n=1}^{\infty}$ be independent and identically distributed standard normal random variables. Prove that

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{|Z_n|}{\sqrt{2\ln(n)}} = 1\right) = 1.$$

Hint: Prove, and use the bound

$$\left(\frac{1}{x} - \frac{1}{x^3}\right)e^{-\frac{x^2}{2}} \leq \int_x^{\infty} e^{-\frac{y^2}{2}}\,dy \leq \frac{1}{x}e^{-\frac{x^2}{2}}.$$

Problem 22. Let $(X_n)_{n=1}^{\infty}$ be a sequence of independent and identically distributed exponential random variables with parameter $\lambda = 1$. Show that:

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{\max_{1\leq j\leq n} X_j}{\ln n} = 1\right) = 1$$

Hint: Problem 15 may be useful here.

Problem 23. (Doob-Dynkin Lemma) Let $X, Y : \Omega \to \mathbb{R}$ be two random variables. Prove that there exists a measurable function $f : \mathbb{R} \to \mathbb{R}$ such that $Y = f(X)$ if and only if $\sigma(Y) \subseteq \sigma(X)$.

Problem 24. Show that there are at most countably many $x \in \mathbb{R}$ where a distribution function F is not continuous.

Problem 25. Let F be a distribution function, and let $r \in \mathbb{N}$ be arbitrary. Show that the following are distribution functions:

1. $F(x)^r$
2. $1 - (1 - F(x))^r$
3. $F(x) + (1 - F(x))\log(1 - F(x))$
4. $(F(x) - 1)e + e^{1-F(x)}$

Problem 26. Let F be a distribution function, and F^{-1} denote it's corresponding quantile function. For each $\omega \in [0, 1]$, prove that

$$F^{-1}(\omega) = \inf\{t \in \mathbb{R} : F(t) \geq \omega\}.$$

Problem 27. Let F be a distribution function, and F^{-1} denote it's corresponding quantile function. Prove that

1. F^{-1} is increasing, left-continuous function on $(0, 1)$
2. If $x > F^{-1}(\omega)$ then $F(x) \geq \omega$
3. If F^{-1} is continuous at ω, then $x > F^{-1}(\omega)$ implies that $F(x) > \omega$.
4. If $F(x) > \omega$ then $x \geq F^{-1}(\omega)$
5. If F is continuous at x, then $F(x) > \omega$ implies that $x > F^{-1}(\omega)$
6. If $x < F^{-1}(\omega)$ then $F(x) < \omega$
7. If $F(x) < \omega$ then $x < F^{-1}(\omega)$

8. $F^{-1} \circ F(x) \le x$ for all $x \in \mathbb{R}$ such that $0 < F(x) < 1$ (what goes wrong when $F(x)$ is 0 or 1?)
9. $F \circ F^{-1}(\omega) \ge \omega$ for all $\omega \in (0,1)$
10. $F \circ F^{-1} \circ F(x) = F(x)$ for all $x \in \mathbb{R}$ such that $0 < F(x) < 1$
11. $F^{-1} \circ F \circ F^{-1}(\omega) = F^{-1}(\omega)$ for all $\omega \in (0,1)$. Consider the case, when $F \circ F^{-1}(\omega) = 1$ separately.
12. If F has a jump at x then F^{-1} is constant on $(F(x-), F(x)]$
13. If F^{-1} has a jump at ω then F is constant on $[F^{-1}(\omega), F^{-1}(\omega+))$
14. The right limit of F^{-1} at ω is $F^{-1}(\omega+) = \inf\{t : F(t) > \omega\}$
15. $Y(\omega) := F^{-1}(\omega)$ is a random variable on $\big([0,1], \mathcal{B}[0,1], \lambda\big)$ where λ is the Lebesgue measure, and show that $F_Y = F$.

Problem 28. Let U be a random variable uniformly distributed on $[0,1]$ and let F be a distribution function. Recall that F^{-1} can be viewed as a random variable on $([0,1], \mathcal{B}[0,1], \lambda)$ where λ the Lebesgue measure. Let $X = F^{-1}(U)$ and recall that this is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. Show that $F_X = F$ for all $x \in \mathbb{R}$.

Problem 29. Let X be a random variable with continuous c.d.f F_X. Prove that $F_X(X)$ is uniformly distributed on $[0,1]$. Conversely, let X be a random variable and let F be a continuous distribution function. If $F(X)$ is uniformly distributed on $[0,1]$ then $F_X = F$.

Problem 30. Let $X_1, X_2, \dots, X_n$ be random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Prove that $X_1, X_2, \dots, X_n$ are independent if and only if for all $B_i \in \mathcal{B}(\mathbb{R})$, $i = 1, 2, \dots, n$ we have

$$\mathbb{P}(X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in B_i).$$

4 Expectation

4.1 Properties of Expected Value

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $X = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i}$ denote a simple random variable for some $a_1, \dots, a_n \in \mathbb{R}$ and $A_1, \dots, A_n \in \mathcal{F}$. We define the expectation (or expected value) of X, by

$$\mathbb{E}X = \sum_{i=1}^{n} a_i \mathbb{P}(A_i).$$

Observe that for this to be a well defined quantity, the expected value must be the same for all possible representations of a given simple random variable (which we know is not unique). If $\sum_{i=1}^{n} a_i \mathbf{1}_{A_i}$ and $\sum_{j=1}^{m} b_j \mathbf{1}_{B_j}$ are two possible representations, it can be shown that $\sum_{i=1}^{n} a_i \mathbb{P}(A_i) = \sum_{j=1}^{m} b_j \mathbb{P}(B_j)$ and thus the expectation operator is a well-defined operator.

A special case of this are indicator random variables. If $A \in \mathcal{F}$, then we have that $\mathbb{E}\mathbf{1}_A = \mathbb{P}(A)$. That is, the expected value of an indicator of an event is simply the probability that specific event will occur.

Proposition 20. If X and Y are simple random variables, then the following statements hold.

1. If $X \ge 0$, then $\mathbb{E}X \ge 0$;
2. $\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y$;
3. If $X \le Y$, then $\mathbb{E}X \le \mathbb{E}Y$;

4. If $X = Y$, then $\mathbb{E}X = \mathbb{E}Y$;
5. $|\mathbb{E}X| \leq \mathbb{E}|X|$.

Proof. Since X is a simple random variable, it takes on finitely many values $x_1, \dots, x_n$. Consider the representation $X = \sum_{i=1}^n x_i \mathbf{1}_{\{X=x_i\}}$ and note that $\mathbb{E}X = \sum_{i=1}^n x_i \mathbb{P}(X = x_i)$. If $X \geq 0$ a.s., then $\mathbb{P}(X = x_i) = 0$ whenever $x_i < 0$. Hence $\mathbb{E}X \geq 0$. For 2, let $X = \sum_{i=1}^n \mathbf{1}_{\{X=x_i\}}$ and $Y = \sum_{j=1}^m \mathbf{1}_{\{Y=y_j\}}$. Set $C_{i,j} = \{X = x_i, Y = y_j\}$ for each $i = 1, 2, \dots, n, j = 1, 2, \dots, m$. Then it follows that

$$\begin{aligned}
\mathbb{E}(X+Y) &= \sum_{i=1}^n \sum_{j=1}^m (x_i + y_j)\mathbb{P}(C_{i,j}) \\
&= \sum_{i=1}^n x_i \sum_{j=1}^m \mathbb{P}(X = x_i, Y = y_j) + \sum_{j=1}^m y_j \sum_{j=1}^m \mathbb{P}(X = x_i, Y = y_j) \\
&= \sum_{i=1}^n x_i \sum_{j=1}^m \mathbb{P}(X = x_i) + \sum_{j=1}^m y_j \sum_{j=1}^m \mathbb{P}(Y = y_j) = \mathbb{E}X + \mathbb{E}Y.
\end{aligned}$$

If $X \leq Y$ a.s., then $Y - X \geq 0$ a.s, and so $\mathbb{E}(X - Y) \geq 0$ by 1. By 2. it is immediate that $\mathbb{E}X \leq \mathbb{E}Y$. If $X = Y$ a.s, then we have both $X \leq Y$ and $Y \leq X$ a.s., simultaneously. Hence $EX \leq \mathbb{E}Y$ and $\mathbb{E}Y \leq \mathbb{E}X$ showing that $\mathbb{E}X = \mathbb{E}Y$. Finally, we have that $\max(X, -X) \leq |X|$ and thus $\max(\mathbb{E}X, -\mathbb{E}X) = \max(\mathbb{E}X, \mathbb{E}(-X)) \leq \mathbb{E}|X|$ showing that $|\mathbb{E}X| \leq \mathbb{E}|X|$.

□

We now define the expectation of a non-negative random variable X, as follows: Let $\mathcal{R}_0$ denote the collection of all simple random variables defined on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and define

$$\mathbb{E}X = \sup_{Y \leq X, Y \in \mathscr{R}} \mathbb{E}Y.$$

Finally, recall that we can decompose any general random variable into their positive and negative components $X = X^+ - X^-$ where $X^+ = \max(X, 0)$ and $X^- = -\min(X, 0)$ are non-negative random variables. In this case we define the expectation of a general random variable by

$$\mathbb{E}X = \mathbb{E}X^+ - \mathbb{E}X^-$$

whenever $\mathbb{E}X^+ < \infty$ or $\mathbb{E}X^- < \infty$ and in this case, we say the expected value *exists*.

Example 15. Consider the probability space $([0,1], \mathcal{B}([0,1]), \lambda)$ and define the random variable

$$X := \begin{cases} 0 & \text{if } 2^{-1} \leq \omega < 1, \\ 2^{n+1}/n & \textbf{if } 2^{-(n+1)} \leq \omega < 3(2^{-(n+2)}), \\ -2^{n+1}/n & \textbf{if } 3(2^{-(n+2)}) \leq \omega < 2^{-n}, \end{cases}$$

where $n \in \mathbb{N}$. Observe that $\mathbb{P}(X = 0) = 1/2$ and $\mathbb{P}(X = 2^{n+1}/n) = \mathbb{P}(X = -2^{n+1}/n) = 2^{-(n+2)}$ for the distribution of X, and thus $\mathbb{E}X^+ = \sum_{n=1}^\infty \frac{2^{n+1}}{n} 2^{-(n+2)} = \frac{1}{2} \sum_{n=1}^\infty \frac{1}{n} = \infty$. Similarly, it can be shown that $\mathbb{E}X^- = \infty$ and thus $\mathbb{E}X$ does not exist.

Proposition 21. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $(X_n)_{n=1}^\infty$ be an increasing sequence of non-negative simple random variables such that $X_n \to X$ for some non-negative simple random variable X. Then $\mathbb{E}X_n \to \mathbb{E}X$.

Proof. We first suppose X is an indicator function. That is $X = \mathbf{1}_A$ for some $A \in \mathcal{F}$. Fix $0 < \varepsilon < 1$, and set

$$B_n := \{\omega \in A : X_n(\omega) > 1 - \varepsilon\}$$

for each $n \in \mathbb{N}$, and note that $(1 - \varepsilon)\mathbf{1}_{B_n}(\omega) \leq X_n(\omega) \leq \mathbf{1}_A(\omega)$ for all $\omega \in \Omega$. Therefore,

$$(1 - \varepsilon)\mathbb{P}(B_n) = \mathbb{E}((1 - \varepsilon)\mathbf{1}_{B_n}) \leq \mathbb{E}X_n \leq \mathbb{E}\mathbf{1}_A = \mathbb{P}(A).$$

Since $\{B_n\}_{n=1}^{\infty}$ is an increasing sequence such that $\bigcup_{n=1}^{\infty} B_n = A$, by continuity of measure we ahve that $\mathbb{P}(B_n) \to \mathbb{P}(A)$. Hence as $n \to \infty$ we get

$$(1-\varepsilon)\mathbb{P}(A) \le \lim_{n\to\infty} \mathbb{E}X_n \le \mathbb{P}(A).$$

Since ε was arbitrary, we have $\mathbb{P}(A) \le \lim_{n\to\infty} \mathbb{E}X_n \le \mathbb{P}(A)$ showing that $\lim_{n\to\infty} \mathbb{E}X_n = \mathbb{P}(A) = \mathbb{E}X$. By linearity of limits and expectations of simple random variables, the result immediately generalizes to arbitrary simple random variables.

$\square$

Proposition 22. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X : \Omega \to \mathbb{R}$ be a non-negative random variable. There exists an increasing sequence of non-negative simple random variables $(X_n)_{n=1}^{\infty}$ with $X_n \to X$ such that $\mathbb{E}X_n \to \mathbb{E}X$.

Proof. Observe that $\mathbb{E}X$ is a supremum of a subset of real numbers. Thus, there exists a sequence of non-negative simple random variables $(Y_n)_{n=1}^{\infty}$ such that $(\mathbb{E}Y_n)_{n=1}^{\infty}$ is an increasing sequence of real numbers and $\mathbb{E}Y_n \to \mathbb{E}X$. However we note that the random variables themselves need not be increasing nor do they need to converge pointwise almost surely to X. By theorem 7, choose an increasing sequence of non-negative simple random variables $(Z_n)_{n=1}^{\infty}$ such that $Z_n \to X$. Set $X_n := \max(Z_n, \max_{1\le j\le n} Y_j)$ for each $n \in \mathbb{N}$ and note that this is also an increasing sequence of non-negative simple random variables satisfying $X_n \to X$. Furthermore, since $Y_n \le X_n \le X$, we have that $\mathbb{E}Y_n \le \mathbb{E}X_n \le \mathbb{E}X$. Finally as $\mathbb{E}Y_n \to \mathbb{E}X$ we have that $\mathbb{E}X_n \to \mathbb{E}X$. Hence $(X_n)_{n=1}^{\infty}$ is the desired sequence that has all the required properties.

$\square$

Proposition 23. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X : \Omega \to \mathbb{R}$ be a non-negative random variable. For all increasing sequence of non-negative simple random variables $(X_n)_{n=1}^{\infty}$ and $X_n \to X$, then $\mathbb{E}X_n \to \mathbb{E}X$.

Proof. Let $(X_n)_{n=1}^{\infty}$ be such an increasing sequence of non-negative simple random variables such that $X_n \to X$. Choose an increasing sequence of simple random variables $(Y_n)_{n=1}^{\infty}$ that converges to X, and that $\mathbb{E}Y_n \to \mathbb{E}X$. Now set $Z_{m,n} = \min(X_m, Y_n)$ for all $m, n \in \mathbb{N}$. Note that for each fixed m (resp. n) the sequence $(Z_{m,n})_{n=1}^{\infty}$ (resp. $(Z_{m,n})_{m=1}^{\infty}$ is increasing). Observe that for any fixed m, we have that $Z_{m,n} \to X_m$ and for any fixed n, we have that $Z_{m,n} \to Y_n$. Finally putting it all together, we have that

$$\mathbb{E}X = \lim_{n\to\infty} Y_n = \lim_{n\to\infty}\lim_{m\to\infty} \mathbb{E}Z_{m,n} = \lim_{m\to\infty}\lim_{n\to\infty} \mathbb{E}Z_{m,n} = \lim_{m\to\infty} \mathbb{E}X_m.$$

$\square$

Example 16. (Additivity of Expected Value) Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of non-negative random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$. Show that

$$\sum_{n=1}^{\infty} \mathbb{E}X_n = \mathbb{E}\sum_{n=1}^{\infty} X_n.$$

Proof. Assume first that $\{X_n\}_{n=1}^{\infty}$ is a sequence of step functions. Consider the representation $X_n = \sum_{i=1}^{N_n} a_{n,i}\mathbf{1}_{A_{i,n}}$, the $a_{i,n} > 0$ and the $A_{i,n} \in \mathcal{F}$ are pairwise disjoint for all $i = 1, 2, \dots, N_n$ and $n \in \mathbb{N}$.

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathbb{E}X_n &= \sum_{n=1}^{\infty}\sum_{i=1}^{N_n} a_{i,n}\mathbb{P}(A_{i,n}) \\
&= \mathbb{E}\left[\sum_{n=1}^{\infty}\sum_{i=1}^{N_n} a_{i,n}\mathbf{1}_{A_{i,n}}\right] \\
&= \mathbb{E}\sum_{n=1}^{\infty} X_n.
\end{aligned}$$

Now we relax our hypothesis, and assume $(X_n)_{n=1}^{\infty}$ are arbitrary non-negative random variables. For each $n \in \mathbb{N}$, choose an increasing sequence of simple random variables $(X_{m,n})_{m=1}^{\infty}$ such that $X_{m,n} \to X_n$. Therefore,

$$0 \leq \lim_{m\to\infty} \lim_{N\to\infty} \sum_{n=1}^{N} X_{m,n} = \lim_{m\to\infty} \sum_{n=1}^{\infty} X_{m,n} = \sum_{n=1}^{\infty} X_n.$$

Therefore,

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathbb{E}X_n &= \lim_{N\to\infty} \sum_{n=1}^{N} \mathbb{E}X_n \\
&= \lim_{N\to\infty} \sum_{n=1}^{N} \lim_{m\to\infty} \mathbb{E}X_{m,n} \\
&= \sup_{N\in\mathbb{N}} \sup_{m\in\mathbb{N}} \sum_{n=1}^{N} \mathbb{E}X_{m,n} \\
&= \sup_{m\in\mathbb{N}} \sup_{N\in\mathbb{N}} \sum_{n=1}^{N} \mathbb{E}X_{m,n} \\
&= \lim_{m\to\infty} \sum_{n=1}^{\infty} \mathbb{E}X_{m,n} \\
&= \mathbb{E} \sum_{n=1}^{\infty} X_n
\end{aligned}$$

as required.

□

With this construction, the properties in Proposition 20 can be generalized to non-negative random variables and can be proved by first considering simple random variables and then moving on to non-negative random variables through a limiting argument. We now generalize this further, by considering increasing sequences of non-negative random variables and it's limit. Additionally here we take the notion of increasing and convergence to be in the almost surely case instead of point-wise.

Proposition 24. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $(X_n)_{n=1}^{\infty}$ be an increasing sequence of non-negative random variables such that $X_n \to X$. Then there exists a sequence of simple random variables $(Y_n)_{n=1}^{\infty}$ such that $0 \leq Y_n \leq X_n$ for each $n \in \mathbb{N}$ and $Y_n \to X$.

Proof. Since each X_n is a non-negative random variable, by theorem 7 there exists an increasing sequence of non-negative simple random variables $(X_{m,n})_{m=1}^{\infty}$ such that $X_{m,n} \to X_n$. Now define the simple random variables $Y_k = \max_{1\leq m,n,\leq k} X_{m,n}$ for each $k \in \mathbb{N}$, and observe that $Y_{k-1} \leq Y_k \leq \max_{1\leq n\leq k} X_n \leq X_k$. Now define $Y := \lim_{k\to\infty} Y_k$ and note that for each $1 \leq n \leq k$ that $X_{n,k} \leq Y_k \leq X_k$. Hence as $k \to \infty$ we have $X_n \leq Y \leq X$ for each $n \in \mathbb{N}$. Hence $X = \lim_{n\to\infty} X_n \leq Y \leq X$, showing that $Y = X$. Thus for the increasing sequence of simple random variables $(Y_n)_{n=1}^{\infty}$, we have $Y_n \to X$ as desired.

□

4.2 Limit Theorems for Expected Value

In probability theory, we are often concerned about events that happen with probability 1 and do not concern ourselves with undesirable outcomes if they occur in an event with probability 0. For example, we had considered point-wise convergence of random variables where a sequence $(X_n)_{n=1}^{\infty}$ converges to a random variable X if $X_n(\omega) \to X(\omega)$ for every outcome $\omega \in \Omega$. However if instead $X_n(\omega) \to X(\omega)$ for specific $\omega \in A$ for some event $A \in \mathcal{F}$ with $\mathbb{P}(A) = 1$, we say that the sequence converges $\mathbb{P}$-*almost surely* or just *almost surely* (a.s.) if the probability measure is specified. This can be generalized to any property of interest in which occurs with probability 1.

Theorem 11. (Monotone Convergence Theorem) Let $(X_n)_{n=1}^{\infty}$ be an increasing sequence of non-negative random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, such that $X_n \overset{\text{a.s.}}{\to} X$, then $\mathbb{E}X_n \to \mathbb{E}X$.

Proof. We first suppose that $X_n(\omega) \leq X_{n+1}(\omega) \leq X(\omega)$ and $X_n(\omega) \to X(\omega)$ for all $\omega \in \Omega$. Choose an increasing sequence of non-negative step functions $(Y_n)_{n=1}^{\infty}$, such that $Y_n \leq X_n$ for each $n \in \mathbb{N}$, and converges to X. Then the previous propositions imply $\mathbb{E}Y_n \to \mathbb{E}X$. Furthermore by monotonicity of expectation, we have that $\mathbb{E}Y_n \leq \mathbb{E}X_n$ since $Y_n \leq X_n$. Thus it follows that,

$$\mathbb{E}X = \lim_{n\to\infty} \mathbb{E}Y_n = \liminf_{n\to\infty} \mathbb{E}Y_n \leq \liminf_{n\to\infty} \mathbb{E}X_n.$$

On the other hand, since $X_n \leq X$ we have that $\mathbb{E}X_n \leq \mathbb{E}X$ for all $n \in \mathbb{N}$. Therefore passing through the limit superior we have that $\limsup_{n\to\infty} \mathbb{E}X_n \leq \mathbb{E}X$. Combining all of the above we have that

$$\mathbb{E}X \leq \liminf_{n\to\infty} \mathbb{E}X_n \leq \limsup_{n\to\infty} \mathbb{E}X_n \leq \mathbb{E}X,$$

and so $\mathbb{E}X_n \to \mathbb{E}X$. We now suppose that that $X_n(\omega) \leq X_{n+1}(\omega) \leq X(\omega)$ and $X_n(\omega) \to X(\omega)$ for all $\omega \in A$ for some $A \in \mathcal{F}$ with $\mathbb{P}(A) = 1$. Hence $(X_n\mathbf{1}_A)_{n=1}^{\infty}$ is an increasing sequence of non-negative random variables such that $X_n\mathbf{1}_A \to X\mathbf{1}_A$. With the argument above we have $\mathbb{E}(X_n\mathbf{1}_A) \to \mathbb{E}(X\mathbf{1}_A)$. Since, $X_n\mathbf{1}_A = X_n$ a.s., and $X\mathbf{1}_A = X$ a.s., it follows that $\mathbb{E}(X_n\mathbf{1}_A) = \mathbb{E}X_n$ and $\mathbb{E}(X\mathbf{1}_A) = \mathbb{E}X$ completing the proof.

□

Example 17. Let $(X_n)_{n=1}^{\infty}$ be an decreasing sequence of non-positive random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, such that $X_n \overset{\text{a.s.}}{\to} X$, then $\mathbb{E}X_n \to \mathbb{E}X$.

Proof. Choose $Y := -X$ and $Y_n := -X_n$ for all $n \in \mathbb{N}$. Then it follows that $0 \leq Y_n \leq Y$ and $Y_n \to Y$. Therefore, by the Monotone Convergence Theorem, we have that $\lim\limits_{n\to\infty} \mathbb{E}Y_n = \mathbb{E}Y$. However by linearity of expectation, it follows that $\lim\limits_{n\to\infty} \mathbb{E}X_n = \mathbb{E}X$. □

An consequence of the monotone convergence theorem, is that the expectation of a product of independent random variables is the product of the individial expectations.

Theorem 12. Let $X_1, X_2 \ldots, X_n$ be independent random variables defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then it follows that:

$$\mathbb{E}\Big(\prod_{i=1}^{n} X_i\Big) = \prod_{i=1}^{n} \mathbb{E}X_i.$$

Proof. We provide a proof for two random variables X and Y, and the general result follows inductively. We first take X and Y to be simple random variables. That is,

$$X = \sum_{i=1}^{m} x_i \mathbf{1}_{\{X=x_i\}} \quad \text{and} \quad Y = \sum_{j=1}^{m} y_j \mathbf{1}_{\{Y=y_j\}}.$$

It follows that

$$\begin{aligned}
\mathbb{E}(XY) &= \mathbb{E}\big(\sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j \mathbf{1}_{\{X_i=x, Y=y_j\}}\big) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j \mathbb{E}(\mathbf{1}_{\{X_i=x, Y=y_j\}}) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j \mathbb{P}(X_i = x, Y = y_j) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j \mathbb{P}(X_i = x)\mathbb{P}(Y = y_j) \\
&= \sum_{i=1}^{n} x_i \mathbb{P}(X = x_i) \sum_{j=1}^{m} y_j \mathbb{P}(Y = y_j) = \mathbb{E}X\mathbb{E}Y.
\end{aligned}$$

Now suppose X and Y are non-negative random variables. By theorem 7, we choose increasing sequences of non-negative simple random variables $(X_n)_{n=1}^\infty$ and $(Y_n)_{n=1}^\infty$ that converge to X and Y respectively. It follows that $(X_nY_n)_{n=1}^\infty$ is an increasing sequence of non-negative random variables, that converge to the product XY, and so by monotone convergence we have $\mathbb{E}(X_nY_n) \to \mathbb{E}(XY)$. Since X_n and Y_n are simple random variables, we also have by monotone convergence, $\mathbb{E}(X_nY_n) = \mathbb{E}X_n\mathbb{E}Y_n \to \mathbb{E}X\mathbb{E}Y$. Hence $\mathbb{E}(XY) = \mathbb{E}X\mathbb{E}Y$ for non-negative random variables. The general case follows from the decomposition

$$\mathbb{E}(XY) = \mathbb{E}(X^+Y^+) + \mathbb{E}(X^+Y^-) - \mathbb{E}(X^-Y^+) + \mathbb{E}(X^-Y^-)$$

and applying the non-negative case to each of the four terms in the decomposition. $\square$

The independence requirement is necessary in the above result. To see why, consider the probability space $([0,1], \mathcal{B}([0,1]), \lambda)$ and set $X(\omega) = Y(\omega) = \omega^{-1/2}$ for all $\omega \in [0,1]$. It can be seen that $\mathbb{E}X = \mathbb{E}Y = 2$ while $\mathbb{E}(XY) = \infty$.

Theorem 13. (Fatou's Lemma) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $(X_n)_{n=1}^\infty$ be a sequence of non-negative random variables. Then

$$\liminf_{n\to\infty} \mathbb{E}X_n \geq \mathbb{E}\big(\liminf_{n\to\infty} X_n\big).$$

Proof. Define $Y_n := \inf_{m\geq n} X_m$ for each $n \in \mathbb{N}$, and set $Y := \lim_{n\to\infty} Y_n$. By definition of the limit inferior we have $Y = \liminf_{n\to\infty} X_n$. Since $0 \leq Y_n \leq Y_{n+1}$ and $Y_n \to Y$ we have by monotone convergence

$$\lim_{n\to\infty} \mathbb{E}Y_n = \mathbb{E}Y = \mathbb{E}\big(\liminf_{n\to\infty} X_n\big).$$

On the other hand, since $X_n \geq Y_n$, we have $\mathbb{E}X_n \geq \mathbb{E}Y_n$ and passing through the limit inferior gives

$$\liminf_{n\to\infty} \mathbb{E}X_n \geq \liminf_{n\to\infty} \mathbb{E}Y_n = \mathbb{E}Y = \mathbb{E}\big(\liminf_{n\to\infty} X_n\big)$$

since $\mathbb{E}Y_n \to \mathbb{E}Y$. $\square$

We can produce a similar inequality for limit superior if we impose an additional condition.

Proposition 25. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $(X_n)_{n=1}^\infty$ be a sequence of non-negative random variables, where $X_n \leq X$ for some random variable X satisfying $\mathbb{E}|X| < \infty$. Then,

$$\limsup_{n\to\infty} \mathbb{E}X_n \leq \mathbb{E}\big(\limsup_{n\to\infty} X_n\big).$$

Proof. Since $-X_n \geq -X$, we can apply Fatou's lemma to the random variables $Y_n := -X_n + X$ to yield

$$\liminf_{n\to\infty} \mathbb{E}(-X_n) \geq \mathbb{E}\big(\liminf_{n\to\infty}(-X_n)\big).$$

This implies,

$$-\liminf_{n\to\infty} \mathbb{E}(-X_n) \leq \mathbb{E}\big(-\liminf_{n\to\infty}(-X_n)\big),$$

or equivalently

$$\limsup_{n\to\infty} \mathbb{E}X_n \leq \mathbb{E}\big(\limsup_{n\to\infty} X_n\big).$$

$\square$

Observe that the dominating condition is required on the X_n, for the limit superior version of Fatou's lemma. The following example shows why this condition is necessary.

Example 18. Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), \lambda)$ and define the random variables $X_n := n^2\mathbf{1}_{(0,\frac{1}{n})}$ for each $n \in \mathbb{N}$. Fix $\omega \in [0,1]$ and choose $N \in \mathbb{N}$ so that $\omega < \frac{1}{N}$. Then it follows that $X_n(\omega) = 0$ for $n > N$ showing that $X_n \to 0$. However, it is clear that $\mathbb{E}X_n = n^2\frac{1}{n} = n$ for all $n \in \mathbb{N}$. Hence $\mathbb{E}\big(\limsup_{n\to\infty} X_n\big) = 0$ while $\limsup_{n\to\infty} \mathbb{E}X_n = \infty$.

Theorem 14. (Dominated Convergence Theorem) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $(X_n)_{n=1}^{\infty}$ be a sequence of random variables. Suppose there exists random variables X and Y such that $X_n \overset{\text{a.s.}}{\to} X$, $\mathbb{E}|Y| < \infty$ and $|X_n| \leq Y$ for each $n \in \mathbb{N}$. Then $\mathbb{E}|X| < \infty$ and $\mathbb{E}X_n \to \mathbb{E}X$.

Proof. Since $(|X_n|)_{n=1}^{\infty}$ is a sequence of non-negative random variables, by Fatou's lemma we have

$$\mathbb{E}|X| = \mathbb{E}\big(\liminf_{n\to\infty} |X_n|\big) = \liminf_{n\to\infty} \mathbb{E}|X_n| \leq \mathbb{E}Y < \infty.$$

Since $|X_n| \leq Y$, we have that $X_n + Y \geq 0$, $-X_n + Y \geq 0$ and $\mathbb{E}|X_n| \leq \mathbb{E}Y < \infty$ for each $n \in \mathbb{N}$. Hence by Fatou's lemma we get

$$\begin{aligned}\mathbb{E}Y + \liminf_{n\to\infty} \mathbb{E}X_n &= \liminf_{n\to\infty} (\mathbb{E}X_n + \mathbb{E}Y) \\ &= \liminf_{n\to\infty} \mathbb{E}(X_n + Y) \\ &\geq \mathbb{E}\big(\liminf_{n\to\infty} X_n + Y\big) \\ &= \mathbb{E}(X + Y) = \mathbb{E}X + \mathbb{E}Y.\end{aligned}$$

As $\mathbb{E}Y < \infty$, we have

$$\liminf_{n\to\infty} \mathbb{E}X_n \geq \mathbb{E}X.$$

Repeating the same argument but with $-X_n$ instead of X_n, we have

$$\liminf_{n\to\infty} \mathbb{E}(-X_n) \geq \mathbb{E}(-X),$$

or equivalently

$$\limsup_{n\to\infty} \mathbb{E}X_n \leq \mathbb{E}X.$$

Thus,

$$\mathbb{E}X \leq \liminf_{n\to\infty} \mathbb{E}X_n \leq \limsup_{n\to\infty} \mathbb{E}X_n \leq \mathbb{E}X$$

completing the proof. □

Example 19. (Moment Generating Function) The *moment-generating function* (m.g.f) of a random variable X is defined by

$$M(t) = \mathbb{E}(e^{tX}),$$

for any $t \in \mathbb{R}$ such that $\mathbb{E}(e^{tX}) < \infty$. Suppose $M(t) < \infty$ for all $t \in (-\delta, \delta)$ for some $\delta > 0$, then $\frac{d^n}{dt^n} M(t)|_{t=0} = \mathbb{E}X^n$, and $\mathbb{E}|X^n| < \infty$ for all $n \in \mathbb{N}$.

Proof. Let $t \in (0, \delta)$. Using the inequality

$$e^{t|x|} \leq e^{tx} + e^{-tx},$$

we have

$$\mathbb{E}(e^{t|X|}) \leq M(t) + M(-t) < \infty.$$

By a Taylor series expansion of the exponential function, we have

$$e^{t|X|} = \sum_{n=0}^{\infty} \frac{t^n |X|^n}{n!}.$$

Let $Y_m := \sum_{n=0}^{m} \frac{t^n |X|^n}{n!}$, and observe that $(Y_m)_{m=1}^{\infty}$ is an increasing sequence of non-negative random variables with $Y_m \overset{\text{a.s.}}{\to} e^{t|X|}$. Since $|Y_m| = Y_m \leq e^{t|X|}$, by dominated convergence we have that

$$\mathbb{E}(e^{t|X|}) = \lim_{m\to\infty} \mathbb{E}(Y_m) = \lim_{m\to\infty} \mathbb{E} \sum_{n=0}^{m} \frac{t^n |X|^n}{n!} = \sum_{n=0}^{\infty} \frac{t^n \mathbb{E}|X|^n}{n!} < \infty.$$

This means that $\mathbb{E}|X^n| < \infty$ for all n. Furthermore, since

$$\mathbb{E}X^n \leq |\mathbb{E}X^n| \leq \mathbb{E}|X^n|,$$

the series $\sum_{n=0}^{\infty} \frac{t^n \mathbb{E}X^n}{n!}$ is absolutely convergent for any $t \in (-\delta, \delta)$. Let $Z_m = \sum_{n=0}^{m} \frac{t^n \mathbb{E}X^n}{n!}$ and we have

$$|M(t) - \mathbb{E}Z_m| \leq \sum_{n=m+1}^{\infty} \frac{t^n \mathbb{E}|X|^n}{n!} \to 0, \text{ as } m \to \infty.$$

Thus,

$$M(t) = \sum_{n=0}^{\infty} \frac{t^n \mathbb{E}X^n}{n!}.$$

Take the n-th derivative of $M(t)$ using the previous expansion for $t \in (-\delta, \delta)$ and set $t = 0$, we have

$$\frac{d^n}{dt^n} M(t)|_{t=0} = \mathbb{E}X^n.$$

□

Example 20. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\{A_n\}_{n=1}^{\infty}$ be a disjoint collection of events. If $\mathbb{E}|X| < \infty$ then

$$\mathbb{E}(X\mathbf{1}_A) = \sum_{n=1}^{\infty} \mathbb{E}(X\mathbf{1}_{A_n})$$

where $A := \bigcup_{n=1}^{\infty} A_n$. By definition, we have

$$\mathbb{E}(X\mathbf{1}_A) = \mathbb{E}\Big(\sum_{n=1}^{\infty} X\mathbf{1}_{A_n}\Big).$$

Define the sequence of random variables $X_n := \sum_{i=1}^{n} X\mathbf{1}_{A_i}$ for each $n \in \mathbb{N}$. We have that $X_n \to X\mathbf{1}_A$ and $|X_n| \leq |X|$. Hence by dominated convergence we have

$$\begin{aligned}
\mathbb{E}(X\mathbf{1}_A) &= \lim_{n\to\infty} \mathbb{E}X_n \\
&= \lim_{n\to\infty} \mathbb{E}\Big(\sum_{i=1}^{n} X\mathbf{1}_{A_i}\Big) \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E}(X\mathbf{1}_{A_i}) \\
&= \sum_{n=1}^{\infty} \mathbb{E}(X\mathbf{1}_{A_n}).
\end{aligned}$$

4.3 Expectation Inequalities

The expected value of a random variable is an important quantity and as such, it is of interest to construct specific bounds that relate the expected value and the probability of specific events pertaining to a random variable.

Theorem 15. (Markov's Inequality). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be probability space and let $X \geq 0$ be a random variable on Ω. Then for all $t > 0$ we have

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}X}{t}$$

for all $t > 0$.

Proof. Fix $t > 0$ and observe that

$$t\mathbb{P}(X \geq t) = t\mathbb{E}(\mathbf{1}_{\{X\geq t\}}) = \mathbb{E}(t\mathbf{1}_{\{X\geq t\}}) \leq \mathbb{E}(X\mathbf{1}_{\{X\geq t\}}) \leq \mathbb{E}X.$$

□

We can generalize the previous inequality, to functions of random variables and on specific events of the sample space.

Theorem 16. (Chebyshev's Inequality) Suppose $f : \mathbb{R} \to \mathbb{R}$ is a non-negative Borel measurable function and fix $B \in \mathcal{B}(\mathbb{R})$. Then it follows that

$$\inf_{x \in B} f(x)\mathbb{P}(X \in B) \le \mathbb{E}(f(X)\mathbf{1}_B).$$

Proof. Fix $\omega \in \Omega$, if $X(\omega) \notin B$ then $\mathbf{1}_{\{X \in B\}}(\omega) = 0$ and thus

$$\inf_{x \in B} f(x)\mathbf{1}_{\{X \in B\}}(\omega) \le f(X(\omega))\mathbf{1}_{\{X \in B\}}(\omega) \le f(X(\omega)).$$

On the other hand if $X(\omega) \in B$, then $f(\omega) \ge \inf_{x \in B} f(x)$ and thus

$$\inf_{x \in B} f(x)\mathbf{1}_{\{X \in B\}}(\omega) \le f(X(\omega))\mathbf{1}_{\{X \in B\}}(\omega) \le f(X(\omega)).$$

Since this holds for all $\omega \in \Omega$, by monotonicity of expectation we have that

$$\inf_{x \in B} f(x)\mathbb{P}(X \in B) \le \mathbb{E}(f(X)\mathbf{1}_B).$$

$\square$

We can recover the Markov inequality by choosing $B := [t, \infty)$ and $f(x) := x$. The classic Chebyshev's inequality that we are familiar with is also recovered by choosing $f(x) := x^2$ and considering the random variable $(X - \mathbb{E}X)^2$ and the set $B := [t, \infty)$. Hence the Chebyshev's inequality becomes

$$\mathbb{P}(|X - \mathbb{E}X| \ge t) \le \mathbb{P}((X - \mathbb{E}X)^2 \ge t^2) \le \frac{\mathrm{Var}\,(X)}{t^2},$$

where $\mathrm{Var}\,(X) := \mathbb{E}(X - \mathbb{E}X)^2$ is the *variance* of a random variable X.

Similarly, choosing $f(x) := e^{tx}$ for a fixed $t > 0$, we can produce another useful bound that relates the probability of an event to the moment generating function of a random variable. The proof of which is immediate from the Markov inequality.

Theorem 17. (Chernoff's Bound) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X be a random variable with $\mathbb{E}X = 0$. Then for all $x \ge 0$ we have

$$\mathbb{P}(X \ge x) \le \inf_{t>0} e^{-tx}M(t) \quad \text{and} \quad \mathbb{P}(X \le x) \le \inf_{t<0} e^{-tx}M(t)$$

where M is the moment generating function of X.

A consequence of the Chernoff bound applied to sum's of independent random variables is the Hoeffding's inequality.

Theorem 18. (Hoeffding's Inequality) Suppose $X_1, \dots, X_n$ are independent random variables, and $a_i \le X_i \le b_i$ for $i = 1, \dots, n$. Then, for $\varepsilon \ge 0$,

$$\mathbb{P}\big(|\sum_{i=1}^{n}(X_i - \mathbb{E}X_i) \ge \varepsilon\big) \le 2\exp\left(-\frac{2\varepsilon^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right).$$

Proof. The case of $\varepsilon = 0$ is obvious. Let $\varepsilon > 0$. We first apply Chernoff's inequality to yield

$$\mathbb{P}\big(\sum_{i=1}^{n}(X_i - \mathbb{E}X_i) \ge \varepsilon\big) \le e^{-t\varepsilon}\mathbb{E}\big(e^{t\sum_{i=1}^{n}(X_i - \mathbb{E}X_i)}\big) \le e^{-t\varepsilon}\prod_{i=1}^{n}\mathbb{E}\big(e^{t(X_i - \mathbb{E}X_i)}\big),$$

for all $t > 0$. We need to estimate $\mathbb{E}\big(e^{t(X_i - \mathbb{E}X_i)}\big)$. The estimation is provided at the end of the proof, and thus

$$\mathbb{P}\big(\sum_{i=1}^{n}(X_i - \mathbb{E}X_i) \ge \varepsilon\big) \le e^{-t\varepsilon}\prod_{i=1}^{n} e^{t^2(b_i - a_i)^2/8} = \exp\big(\frac{t^2}{8}\sum_{i=1}^{\infty}(b_i - a_i)^2 - t\varepsilon\big),$$

for $t > 0$. Since t is arbitrary, we choose $t = 4\varepsilon/\sum_{i=1}^{\infty}(b_i - a_i)^2$ and yield

$$\mathbb{P}\big(\sum_{i=1}^{n}(X_i - \mathbb{E}X_i) \geq \varepsilon\big) \leq \exp\Big(-\frac{2\varepsilon^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\Big).$$

Applying the similar reasoning to $-X_1, \ldots, -X_n$, we yield

$$\mathbb{P}\Big(\sum_{i=1}^{n}(X_i - \mathbb{E}X_i) \leq -\varepsilon\Big) \leq \exp\Big(-\frac{2\varepsilon^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\Big).$$

Combing these two inequalities completes the proof of the Hoeffding inequality. It remains to provide the estimation on $\mathbb{E}\big(e^{t(X_i - \mathbb{E}X_i)}\big)$. We will prove that for a random variable Y with $\mathbb{E}Y = 0$ and $a \leq Y \leq b$, we have

$$\mathbb{E}e^{tY} \leq e^{t^2(b-a)^2/8}$$

for all t. The statement is clearly true for $a = b$. Suppose $a < b$ and define $Z := (b - Y)/(b - a)$. Then, $Y = Za + (1 - Z)b$ and

$$e^{tY} = e^{Zta+(1-Z)tb} \leq Ze^{ta} + (1 - Z)e^{tb} = \frac{b - Y}{b - a}e^{ta} + \frac{Y - a}{b - a}e^{tb}.$$

Taking expectation we yield,

$$\begin{aligned}\mathbb{E}e^{tY} &\leq \frac{b}{b-a}e^{ta} - \frac{a}{b-a}e^{tb} = \exp\left(\log\left(\frac{be^{ta} - ae^{tb}}{b-a}\right)\right) = \exp\left(ta + \log\left(\frac{b - ae^{t(b-a)}}{b-a}\right)\right)\\ &= \exp\left(u(p-1) + \log(p - (1-p)e^u)\right), \qquad (1)\end{aligned}$$

where $p := b/(b - a)$ and $u := t(b - a)$. Let $\varphi(u) := u(p - 1) + \log(p + (1 - p)e^u)$ for $u \geq 0$ (note $b \geq 0$ and thus $p \geq 0$). By Taylor's expansion, we have

$$\varphi(u) = \varphi(0) + \varphi'(0)u + \frac{1}{2}\varphi''(s)u^2$$

for some $0 \leq s \leq u$. Note $\varphi(0) = 0$, and $\varphi'(u) = (p-1) + \frac{(1-p)e^u}{p+(1-p)e^u}$, i.e., $\varphi'(0) = 0$. Regarding φ'',

$$\varphi''(u) = \frac{(1-p)e^u}{p + (1-p)e^u} - \frac{(1-p)^2e^{2u}}{(p + (1-p)e^u)^2} \leq \frac{1}{4}, \quad u \geq 0.$$

It follows that $\varphi(u) \leq u^2/8$. This together with the bound on $\mathbb{E}e^{tY}$ concludes the proof.

□

With restricting the class of functions that we use, we can produce another useful inequality. We begin with a definition.

Definition 13. A function $f : \mathbb{R} \to \mathbb{R}$, defined on an open $U \subset \mathbb{R}$ is called *convex* on U if and only if

$$f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y)$$

for all $t \in [0, 1]$ and all $x, y \in U$.

Theorem 19. (Jensen's Inequality) Let $(\Omega, \mathcal{F}, \mathbb{P})$ denote a probability space and X be a random variable with $\mathbb{E}|X| < \infty$. If $f : \mathbb{R} \to \mathbb{R}$ is a convex function with $\mathbb{E}|f(X)| < \infty$, then

$$f(\mathbb{E}X) \leq \mathbb{E}f(X).$$

Proof. Since f is convex, there exists $a, b \in \mathbb{R}$ such that $ax + b \leq f(x)$ for all $x \in \mathbb{R}$, and $f(\mathbb{E}X) = a\mathbb{E}X + b$. Hence for all $\omega \in \Omega$, we have that

$$f(\mathbb{E}X) = a\mathbb{E}X + b = \mathbb{E}(aX + b) \leq \mathbb{E}f(X).$$

□

Example 21. (Lyapunov's Inequality) For $0 < p < q < \infty$ we have that $(\mathbb{E}|X|^p)^{\frac{1}{p}} \leq (\mathbb{E}|X|^q)^{\frac{1}{q}}$. If $q > p > 0$, then $f(x) := x^{\frac{q}{p}}$ is a convex function. Then by the previous exercise we have

$$E|X|^q = \mathbb{E}|X|^{p\frac{q}{p}} \leq (\mathbb{E}|X|^p)^{\frac{q}{p}}.$$

Taking the q-root of both sides yields the desired result.

This example can be generalized to another inequality that relates to products of random variables.

Theorem 20. (Hölder's Inequality) Let X and Y denote two random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If $1 < p, q < \infty)$ with $\frac{1}{p} + \frac{1}{q} = 1$ then

$$\mathbb{E}|XY| \leq (\mathbb{E}|X|^p)^{\frac{1}{p}} (\mathbb{E}|Y|^q)^{\frac{1}{q}}.$$

Proof. If $\mathbb{E}|X|^p = 0$, then $|X|^p = 0$ a.s., implying that $\mathbb{E}|XY| = 0$ and the inequality holds trivially. Similar case holds when $\mathbb{E}|Y|^q = 0$. Thus we assume $\mathbb{E}|X|^p, \mathbb{E}|Y|^q > 0$. If either $\mathbb{E}|X|^p = \infty$ or $\mathbb{E}|Y|^q = \infty$ once again the inequality holds trivially. Suppose $0 < \mathbb{E}|X|^p, \mathbb{E}|Y|^q < \infty$, and set

$$X' := \frac{X}{(\mathbb{E}|X|^p)^{\frac{1}{p}}} \quad \text{and} \quad Y' := \frac{Y}{(\mathbb{E}|Y|^q)^{\frac{1}{q}}}.$$

We have that

$$\mathbb{E}|X'Y'| = \frac{\mathbb{E}|XY|}{(\mathbb{E}|X|^p)^{\frac{1}{p}} (\mathbb{E}|Y|^q)^{\frac{1}{q}}}$$

Choose $x := |X'|^p$ and $y := |Y'|^q$ and observe that

$$|X'Y'| = x^{\frac{1}{p}} y^{\frac{1}{q}} \leq \frac{1}{p}a + \frac{1}{q}b = \frac{1}{p}|X'|^p + \frac{1}{q}|Y'|^q.$$

Taking expectations and using the above expression for $\mathbb{E}|X'Y'|$ we have

$$\frac{\mathbb{E}|XY|}{(\mathbb{E}|X|^p)^{\frac{1}{p}} (\mathbb{E}|Y|^q)^{\frac{1}{q}}} = \mathbb{E}|X'Y'| \leq \frac{1}{p}\mathbb{E}|X'|^p + \frac{1}{q}\mathbb{E}|Y'|^q = \frac{1}{p} + \frac{1}{q} = 1.$$

$\square$

A special case of Hölder's inequality, when $p = q = 2$ is the famous Cauchy-Schwartz inequality that is used ubiquitously in probability and statistics.

Example 22. (Cauchy-Schwarz Inequality) Let X and Y denote two random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then,

$$\mathbb{E}|XY| \leq \sqrt{\mathbb{E}|X|^2 \mathbb{E}|Y|^2}.$$

With this, we can immediately deduce a useful tail-bound for non-negative random variables in relation to their first and second moments. Let X be a non-negative random variable. Applying the Cauchy-Schwarz to X and $\mathbf{1}_{\{X>0\}}$

$$\mathbb{E}X = \mathbb{E}|X\mathbf{1}_{X>0}| \leq \sqrt{E|X|^2 \mathbb{E}|\mathbf{1}_{\{X>0\}}|^2} = \sqrt{\mathbb{E}X^2 \mathbb{P}(X > 0)}$$

and re-arranging,

$$\mathbb{P}(X > 0) \geq \frac{(\mathbb{E}X)^2}{\mathbb{E}X^2}.$$

Theorem 21. (Minkowski's Inequality) Let X and Y be two random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If $p \geq 1$, then

$$(\mathbb{E}|X + Y|^p)^{\frac{1}{p}} \leq (\mathbb{E}|X|^p)^{\frac{1}{p}} + (\mathbb{E}|Y|^p)^{\frac{1}{p}}.$$

Proof. If either $\mathbb{E}|X|^p = \infty$ or $\mathbb{E}|Y|^p < \infty$, the result is immediate, so assume the both quantities are finite. If $p = 1$, then the result follows trivially from the triangle inequality and monotonicity of expectation. Fix $1 < p < \infty$ and choose $q = \frac{p}{p-1}$. Then by an application of Hölder's inequality

$$\begin{aligned}
\mathbb{E}|X+Y|^p &= \mathbb{E}(|X+Y||X+Y|^{p-1}) \\
&\leq \mathbb{E}((|X|+|Y|)|X+Y|^{p-1}) \\
&= \mathbb{E}(|X||X+Y|^{p-1}) + \mathbb{E}(|Y||X+Y|^{p-1}) \\
&= (\mathbb{E}|X|^p)^{\frac{1}{p}}(\mathbb{E}|X+Y|^{(p-1)q})^{\frac{1}{q}} + (\mathbb{E}|Y|^p)^{\frac{1}{p}}(\mathbb{E}|X+Y|^{(p-1)q})^{\frac{1}{q}} \\
&= (\mathbb{E}|X|^p)^{\frac{1}{p}}(\mathbb{E}|X+Y|^p)^{\frac{1}{q}} + (\mathbb{E}|Y|^p)^{\frac{1}{p}}(\mathbb{E}|X+Y|^p)^{\frac{1}{q}}.
\end{aligned}$$

If $0 < (\mathbb{E}|X+Y|^p)^{1/q} < \infty$, then dividing out this quantity completes the proof. If $(\mathbb{E}|X+Y|^p)^{\frac{1}{q}} = 0$ the result is immediate, so suppose $(\mathbb{E}|X+Y|^p)^{\frac{1}{q}} = \infty$. Observe that

$$|X+Y|^p \leq (|X|+|Y|)^p \leq 2^{p-1}(|X|^p + |Y|^p),$$

and thus

$$\mathbb{E}|X+Y|^p \leq 2^{p-1}(\mathbb{E}|X|^p + \mathbb{E}|Y|^p)$$

showing that $\mathbb{E}|X|^p = \mathbb{E}|Y|^p = \infty$, and thus $(\mathbb{E}|X|^p)^{\frac{1}{p}} + (\mathbb{E}|Y|^p)^{\frac{1}{p}} = \infty$. □

As we have seen before, the expectation operator is interchangeable with summation. It can also be shown that expectation is interchangeable with differentiation and integration under suitable conditions.

Theorem 22. Let X be a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and suppose $f : \mathbb{R}^2 \to \mathbb{R}$ is a differentiable function in it's first coordinate, satisfying the property that $\mathbb{E}f(t,X) < \infty$ and $\sup_{t\in\mathbb{R}} |f'(t,X)| \leq M$ for some $M > 0$. It follows that

$$\frac{\partial}{\partial t}\mathbb{E}f(t,X) = \mathbb{E}f'(t,X).$$

Proof. Observe that,

$$f'(t,X) = \lim_{n\to\infty} \frac{f(t+1/n,X) - f(t,X)}{1/n},$$

and

$$\left| \frac{f(t+1/n,X(\omega)) - f(t,X(\omega))}{1/n} \right| \leq M$$

for sufficiently large n, and all $\omega \in A$ for some $A \in \mathcal{F}$ with $\mathbb{P}(A) = 1$. Hence by dominated convergence we have that

$$\begin{aligned}
\frac{\partial}{\partial t}\mathbb{E}f(t,X) &= \lim_{n\to\infty} \frac{\mathbb{E}f(t+1/n,X) - \mathbb{E}f(t,X)}{1/n} \\
&= \lim_{n\to\infty} \mathbb{E}\frac{f(t+1/n,X) - f(t,X)}{1/n} \\
&= \mathbb{E}\lim_{n\to\infty} \frac{f(t+1/n,X) - f(t,X)}{1/n} \\
&= \mathbb{E}f'(t,X).
\end{aligned}$$

□

We now present the result that allows use to interchange the expectation operator with integration. The proof of which is outside the scope of this course.

Theorem 23. (Fubini-Tonelli Theorem) Let X be a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a function, and suppose that either $f(t,X)$ is non-negative or that $\int_{-\infty}^{\infty} \mathbb{E}|f(t,X)| < \infty$. Then we have that

$$\int_{-\infty}^{\infty} \mathbb{E}f(t,X)dt = \mathbb{E}\Big(\int_{-\infty}^{\infty} f(t,X)dt\Big).$$

With this, we can prove a very useful tail-integral formula that provides an alternative way of computing expectations.

Theorem 24. (Tail-Integral Formula) If X is a non-negative random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then

$$\mathbb{E}X^n = \int_0^\infty nx^{n-1}\mathbb{P}(X > x)dx.$$

Proof. We have that

$$\begin{aligned}\mathbb{E}X^n &= \mathbb{E}\Big(\int_0^X nx^{n-1}dx\Big)\\ &= \mathbb{E}\Big(\int_0^\infty nx^{n-1}\mathbf{1}_{\{X>x\}}dx\Big)\\ &= \int_0^\infty nx^{n-1}\mathbb{E}\mathbf{1}_{\{X>x\}}dx\\ &= \int_0^\infty nx^{n-1}\mathbb{P}(X > x)dx.\end{aligned}$$

□

4.4 L^p Convergence

In this section we explore another mode of convergence, called *L^p-convergence* or *convergence in p-th mean*.

Definition 14. Let $(X_n)_{n=1}^\infty$ be a sequence of random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let X be a random variable. We say that $(X_n)_{n=1}^\infty$ *converges in L^p* or *converges in p-th mean* (denoted $X_n \overset{L^p}{\to} X$) if $\lim\limits_{n\to\infty} \mathbb{E}|X_n - X|^p = 0$.

By the triangle-inequality and Jensen's inequality, it is clear that if $X_n \overset{L^p}{\to} X$ then $\mathbb{E}X_n \to \mathbb{E}X$. The following two examples show that convergence almost surely and convergence in p-th mean do not imply one another.

Example 23. Consider the probability space $([0, 1], \mathcal{B}([0, 1]), \lambda)$ and define the sequence of random variables $X_n := n\mathbf{1}_{(0,\frac{1}{n})}$ for each $n \in \mathbb{N}$. It is clear that $X_n \overset{\text{a.s.}}{\to} 0$. However on the other hand, $\mathbb{E}|X_n| = 1$ for $n \in \mathbb{N}$ showing that X_n does not converge to 0 in mean.

Example 24. Consider the probability space $([0, 1], \mathcal{B}([0, 1]), \lambda)$ and define the sequence of random variables as follows: $X_1 := \mathbf{1}_{[0,1/2]}$, $X_2 := \mathbf{1}_{[1/2,1]}, X_3 := \mathbf{1}_{[0,1/4]}, X_4 := \mathbf{1}_{[1/4,1/2]}, X_5 := \mathbf{1}_{[1/2,3/4]}, X_6 := \mathbf{1}_{[3/4,1]}, X_7 := \mathbf{1}_{[0,1/8]}, \ldots$, and so on. It is clear that $\mathbb{E}|X_n| \to 0$, however $\lim\limits_{n\to\infty} X_n(\omega)$ doesn't exist for any $n \in \mathbb{N}$.

4.5 Problems

Problem 31. (Essential Supremum) Let X be a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We define the *essential supremum* of X to be,

$$\|X\|_\infty := \inf\{x \in \mathbb{R} : \mathbb{P}(|X| > x) = 0\}.$$

Prove that $\|X\|_\infty = \lim_{p\to\infty}(\mathbb{E}|X|^p)^{\frac{1}{p}}$.

Problem 32. (Chebyshev's Inequality Sharpness) Fix $0 < \sigma \le t$. Construct a random variable X such that $\mathbb{E}X^2 = \sigma^2$ for which $\mathbb{P}(|X| \ge t) = \frac{\sigma^2}{t^2}$. On the other hand, if there exists a random variable X such that $0 < \mathbb{E}X^2 < \infty$, prove that $\lim_{n\to\infty} \frac{n^2}{\mathbb{E}X^2}\mathbb{P}(|X| \ge n) = 0$.

Problem 33. Let X be a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Prove the following:

1. $\inf\{\mathbb{P}(|X| > \varepsilon) : \mathbb{E}X = 0, \mathbb{E}X^2 = 1\} = 0$ for all $\varepsilon > 0$.
2. $\inf\{\mathbb{P}(|X| > x) : \mathbb{E}X = 1, \mathrm{Var}\,(X) = \sigma^2\} = 0$ for $y \geq 1$ and $0 < \sigma^2 < \infty$.

Problem 34. Let X be a non-negative random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, satisfying $\mathbb{E}\frac{1}{X} = \infty$. Prove that:

$$\lim_{n\to\infty} \mathbb{E}\big(\frac{n}{X}\mathbf{1}_{\{X>n\}}\big) = 0 \quad \text{and} \quad \lim_{n\to\infty} \mathbb{E}\big(\frac{1}{nX}\mathbf{1}_{\{X>n\}}\big) = 0.$$

Problem 35. Fix $1 \leq p \leq 2$, and let X and Y be independent random variables where Y has *symmetric* distribution (that is, Y and $-Y$ have the same distribution). Prove that

$$\mathbb{E}|X+Y|^p \leq \mathbb{E}|X|^p + \mathbb{E}|Y|^p.$$

Hint: For $a, b \geq 0$ and $1 \leq p \leq 2$, we have $(a^2+b^2)^p \leq (a^2+b^2)^p$.