Probability Theory I

MAT 5170

Hanan Ather

Fall 2022

Contents

⋆ These notes were created during my review process to aid my own understanding and not written for the purpose of instruction. I originally wrote them only for myself, and they may contain typos and errors [a]. *No professor has verified or confirmed the accuracy of these notes.* With that said, I've decided to share these notes on the off chance they are helpful to anyone else.

[a]Any corrections are greatly appreciated.

§1 September 7, 2022

§1.1 Basic definitions

Any experiment involving randomness can be modelled as a probability space. In this course we define probability space $(\Omega, \mathcal{F}, \mathbb{P})$ using the terminology of measure theory. Measure theory was initially created to generalize the notion of length, area and volume of subsets of Euclidean space. In this course we use Kolmogorov's framework of probability exclusively, however, there are certain *non-commutative* versions that are used in quantum mechanics, which are generalizations of the Kolmogorov model.[1]

Its impossible to assign length to all subsets of $\mathbb{R}$ while preserving additivity and invariance.

- The *sample space* Ω is a set of all possible outcomes $\omega \in \Omega$ of a random experiment (e.g. $\Omega = [0, 1]$ for the uniform distribution).
- The *event space* $\mathcal{F}$ is a set of subsets of Ω
- and $\mathbb{P} : \mathcal{F} \to [0, 1]$, probabilities are assigned $A \mapsto \mathbb{P}(A)$ through *probability measure* $\mathbb{P}$

Definition 1.1 (Field) A **set** is a collection of *elements*. A **class** is a set of sets. All the elements of a class are themselves sets. A nonempty class $\mathcal{A}$ of subsets of Ω is called a **field** if

(i) $A \in \mathcal{A} \implies A^c \in \mathcal{A}$ (closed under complementation)

(ii) $A_1 \in \mathcal{A}, A_2 \in \mathcal{A} \implies A_1 \cup A_2 \in \mathcal{F}$ (Closed under finite union)

The smallest field we can have is $\{\varnothing, \Omega\}$.We are using the word "small" in the sense that the number of sets in the field.[2] Note that we can actually recover the property $\Omega \in \mathcal{A}, \varnothing \in \mathcal{A}$ by applying conditions (i) and (ii). If $A \in \mathcal{A}$ by (i) $A^c \in \mathcal{A}$. By (ii) $A \cup A^c \in \mathcal{A} \implies \Omega \in \mathcal{A}$. Since $\Omega \in \mathcal{A}$, by (i) $\Omega^c \in \mathcal{A} \implies \varnothing \in \mathcal{A}$.

Example 1.2 (Finite and cofinite fields)**.** $\mathcal{A} = \{A : A \subset \Omega \text{ with A is finite or cofinite}\}$; A is called **cofinite** if A^c is finite. Let $\Omega = \{x_1, ..., x_n, ...\}$; $A_n = \{x_n\}$: singleton set $\implies$ A_n^c is cofinite. Let $A, B \in \mathcal{A}$. To show that $A \cup B \in \mathcal{A}$:

1. if A and B are finite $\implies$ $A \cup B =$ finite and hence $A \cup B \in \mathcal{A}$;
2. If at least one is cofinite $\implies$ $A \cup B =$ cofinite

Therefore, $\mathcal{A}$ is a field.

Example 1.3 (Subsets of $\mathbb{R}$)**.** Let $\Omega = \mathbb{R}, \mathcal{A} = \{(-\infty, x], x \in \mathbb{R}\}$. Let $A \in \mathcal{A}, B \in \mathcal{A}$. Then we have $A \cup B = (\infty, max(x_1, x_2)] \in \mathcal{A} \implies A$ is closed under finite union.
However, $A^c = (x_1, \infty) \notin \mathcal{A} \implies$ not closed under complementation, hence not a field.

[1]Parthasarathy, K. R. An introduction to quantum stochastic calculus.
[2]We are using the word "small" in the sense that the number of sets in the field.

Definition 1.4 ($\sigma-$field) A nonempty class $\mathcal{F}$ of subsets of of Ω is called a **$\sigma-$field** if

(i) if for every $A \in \mathcal{F}, A^c \in \mathcal{F}$ (closed under complementation);

(ii) for every sequence of sets $\{A_n\}$, $A_n \in \mathcal{F}$, $\bigcup A_n \in \mathcal{F}$ (closed under countable union)

If $\mathcal{F}$ is a σ-field it implies $\mathcal{F}$ is a field. Is the converse true? NO! Lets go back to example (1.2), we already proved $\mathcal{A}$ is a field. Consider $A_n = \{x_{2n}\}$, $A_n \in \mathcal{F}$ but $\bigcup_{n=1}^{\infty} A_n = \{x_2, x_4, ...\} \notin \mathcal{F}$. Why? since the complement of $\bigcup_{n=1}^{\infty} A_n$ is a set of all odd elements, which is also infinite, $\bigcup_{n=1}^{\infty} A_n$ is neither finite or cofinite.

§1.2 The σ-field generated by a given class $\mathcal{C}$

We now turn our attention to how sigma fields are constructed, and how to guarantee a desired sigma field exists.
Lets say we have an indexed family of subsets of Ω: $\{C_t : t \in T\}$, T is an index set, such that for each $t \in T$, C_t is closed under countable unions. [3] Then

$$\mathcal{C} = \bigcap_{t \in T} C_t \quad \text{is closed under countable unions.}$$

⋆ If we have a sequence of nested sets in Ω. Suppose that $\{\mathcal{F}_n\}$ are all σ-fields. We can show that $\bigcup_{n\in\mathbb{N}} \mathcal{F}_n$ is a *field*. However, suppose that $\{\mathcal{F}_n\}$ are all σ-fields. $\bigcup_{n\in\mathbb{N}} \mathcal{F}_n$ may not be a σ-field.

Important for midterm!

Proof. Complete the proof. □

Proposition 1.5 (Intersection of σ-algebras)**.** If $\mathcal{F}, \mathcal{G} \subseteq 2^{\Omega}$ are σ-algebras on Ω, then their intersection $\mathcal{F} \cap \mathcal{G}$ is also a σ-algebra. This also holds for infinite intersections.

Proof. Straightforward, verify the σ-algebra properties directly. □

Definition 1.6 ($\sigma-$field generated) Given a collection of subsets $\mathcal{G} \subseteq 2^{\Omega}$, we define σ-algebra generated by $\mathcal{G}$ to be the smallest σ-algebra containing $\mathcal{G}$, i.e.,

$$\sigma(\mathcal{G}) = \bigcap_{\substack{\mathcal{G} \subseteq \mathcal{F} \subseteq 2^{\Omega} \\ \mathcal{F} \text{ is a } \sigma\text{-algebra}}} \mathcal{F}$$

Here is a collection of useful and somewhat related facts. Let $\mathcal{C}$ be a collection of subsets of Ω:

⋆ $\sigma(\mathcal{C}) \supset \mathcal{C}$

⋆ If $\mathcal{B}$ is some other σ-field containing $\mathcal{C}$, then $\mathcal{B} \supset \sigma(\mathcal{C})$

⋆ If A is a subset; then $\sigma(A) = \{\varnothing, A, A^c, \Omega\}$

⋆ If $\Omega = \mathbb{R}$ (or more generally if Ω is a space with a topology), the **Borel σ-algebra** is the σ-algebra generated by the open sets (or by closed sets, which is equivalent).

Example 1.7 (Borel sets)**.** Let $\mathcal{C}$ be the class of sub-intervals of $\Omega = (0, 1]$, and define $\mathcal{B} = \sigma(\mathcal{C})$. The elements of $\mathcal{B}$ are called *Borel sets* of the unit interval. $\mathcal{B}$ contains the open sets in $(0, 1]$.

[3] This is NOT true for $\bigcup_{t\in T} C_t$

§1.3 Probability Measure

Definition 1.8 A *set function* is a function defined on some class of subsets of Ω. A set function $\mathbb{P}$ is a **probability measure** if it satisfies the following conditions

(i) $0 \leq \mathbb{P}(A) \leq 1$ for $A \in \mathcal{F}$

(ii) $\mathbb{P}(\Omega) = 1; \mathbb{P}(\varnothing) = 0$

(iii) $\mathbb{P}$ is **additive**: if $A_1, A_2, ...$ is a sequence of disjoint events then

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) \tag{1}$$

§2 September 12, 2022

§2.1 Properties of probability measure

Proposition 2.1 (Countable sub-additivity property of probability measures)**.** *For any sequence $A_1, A_2, ... \in \mathcal{F}$ (whether disjoint or not), we have by*

$$\mathbb{P}\left(\bigcup_j A_j\right) \leq \sum_j \mathbb{P}(A_j)$$

Proof. Let $B_1 = A_1$, $B_j = A_j / \cup_{k=1}^{j-1} A_k$. This implies that $\mathbb{P}(\cup_j A_i) = \mathbb{P}(\cup_j B_j) = \sum_j \mathbb{P}(B_j) \leq \sum_j \mathbb{P}(A_j)$. □

Given a probability triple $(\Omega, \mathcal{F}, \mathbb{P})$, and sequence of events, $A, A_1, A_2, ... \in \mathcal{F}$, the notation $A_i \uparrow A$ means that $A_1 \subseteq A_2...$ and $\bigcup_i A_i = A$. In words, the events A_i *increase* to A. Likewise, we use the notation $A_i \downarrow A$ to mean that $\{A_i^C\} \uparrow A^C$, or equivalently that $A_1 \supseteq A_2 \supseteq A_3 \supseteq ...$, and $\bigcap_i A_i = A$. In words, the events A_i decrease to A.

Proposition 2.2 (Continuity of probabilities)**.** If $A_i \uparrow A$ or $A_i \downarrow A$, then $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$.

Proof. Suppose that $A_i \uparrow A$. Let $B_1 = A_1$, $B_i = A_i \cap A_{i-1}^C$ for $i > 1$. It follows that $\{B_i\}$ are disjoint, with $\bigcup_i B_i = \bigcup_i A_i = A$. Thus,

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

Note: $=$ holds from the fact that $\{A_i\}$ are nested sequence. □

From the proof we can see that if $\{A_i\}$ are *not* nested, then $=$ may not hold, and thus we many not have $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$.

Theorem 2.3 (Part of Borel-Cantelli Lemma) — Let $A_1, A_2, ... \in \mathcal{F}$. If $\sum_j \mathbb{P}(A_j) < \infty$, then

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m\right) = 0$$

§2.2 Borel sets

Proposition 2.4. Let $\Omega_0 \subset \Omega$.

1. If $\mathcal{B}$ is a $\sigma-$field of subsets of Ω, then $\mathcal{B}_0 := \{A \cap \Omega_0 : A \in \mathcal{B}\}$ is a $\sigma-$field of subsets of Ω_0

§2.3 Extension theorem

Motivation for Extension theorem. How can we formally define a probability triple $(\Omega, \mathcal{F}, \mathbb{P})$ which corresponds to the uniform distribution? Clearly we should choose $\Omega = [0, 1]$. But what about $\mathcal{F}$? We know that $\mathcal{F}$ cannot contain *all* interval of $[0, 1]$, but it should certainly contain intervals $[a, b], [a, b), etc.$ The construction of $\mathcal{F}$ and $\mathbb{P}$ is a challenge. To deal with them we prove a very general theorem about constructing probability triples.

Before leaping into technical details of the Extension Theorem, its useful to think about what the theorem is trying to accomplish. In many cases its difficult to define a probability function $\mathbb{P}(\cdot)$ on all sets of a $\sigma-$algebras.

The Extension theorem is a method of constructing complicated probability triples on full σ-algebras, using only probabilities defined on much simpler subsets.
Let $\Omega = [0, 1]$, and $I = \{(a, b] : 0 \leq a < b \leq 1\}$. Let $\mathcal{B}_0$ be a field generated by I. Let

$\lambda([a,b]) = b - a$ extend to $\mathcal{B}_0$ by finite additivity [4]. Then we make the following two claims:
Claim 1: λ is *finitely additive* probability on B_0 called Lebesgue probability
Claim 2: $\exists!$ unique extension to probability on $\mathcal{B} = \sigma(I)$

Theorem 2.5 (Extension theorem) — Suppose that $\mathbb{P}$ is a probability measure on a field $\mathcal{F}_0$ of subsets of Ω, and let $\mathcal{F} = \sigma(\mathcal{F}_0)$ Then there exists a probability measure $Q(\cdot)$ on $\mathcal{F}$ such that $Q(A) = \mathbb{P}(A)$ for $A \in \mathcal{F}_0$ [a]. Further, if Q^* is another probability measure on $\mathcal{F}$ such that $Q^*(A) = \mathbb{P}(A)$ for $A \in F_0$, then $Q^*(A) = Q(A)$ for $A \in \mathcal{F}$.

[a] Although the measure extended to $\mathcal{F}$ is usually denoted by the same letter as the original measure on $\mathcal{F}_0$, they are really different set functions, since they have different domains of definition.

§2.4 Outer Measure

Definition 2.6 (Outer Measure) Let $\mathbb{P}$ be a probability function on field $\mathcal{F}$. For any $A \in \Omega$, let

$$\mathbb{P}^*(A) = \underset{A \subseteq \bigcup_i A_i}{inf} \sum_i \mathbb{P}(A_n)$$

An alternative notion we could use is

$$\mathbb{P}^*(A) = inf \left\{ \sum_i \mathbb{P}(A_n) \mid A \subseteq \bigcup_i A_i \right\}$$

In other words, we are choose a collection of $A_1, A_2, ... \in \mathcal{B}_0$ which covers A and minimizes the sum $\sum_i \mathbb{P}(A_n)$. What do we get from doing this? For any set $A \subseteq \Omega$, we can define a probability through using $\mathbb{P}$.

Proposition 2.7. The outer probability measure function $\mathbb{P}^*$ satisfies the following properties.

(i) Empty set. $\mathbb{P}^*(\varnothing) = 0$.

(ii) Non-Negativity. $\mathbb{P}(A) \geq 0$ for all $A \subseteq \Omega$.

(iii) Monotonicity. $A \subset B$ implies $\mathbb{P}^*(A) \leq \mathbb{P}(B)$.

(iv) Countable Sub-additivity. $\mathbb{P}\left(\bigcup_{i \in \mathbb{N}} A_i\right) \leq \sum_{i \in \mathbb{N}} \mathbb{P}^*(A_i)$

Proof. (i) holds since $\varnothing$ is covered by itself, and has probability measure $\mathbb{P}(\varnothing) = 0$. (ii) follows from the fact that $\mathbb{P}(\cdot)$ is non-negative. (iii) $\mathbb{P}^*(\cdot)$ is monotone, if $A \subset B$, then the infimum of $\mathbb{P}^*(A)$ includes choices of $\{A_i\}$ which work for $\mathbb{P}(B)^*$ plus many more besides that, so $\mathbb{P}^*(A) \leq \mathbb{P}^*(B)$. (iv) We prove the fourth property below. For any ε and any n we construct a sequence of sets $B_{nk}, k \in \mathbb{N}$ such that $A_n \subset \bigcup_k B_{nk}$ and that $\sum_k \mathbb{P}(B_{nk}) \leq \mathbb{P}^*(A_n) + \varepsilon 2^{-n}$. Such a construction is possible since $\mathbb{P}^*$ is defined as the infimum of all possible coverings. □

[4] A set function μ is **finitely additive**, if given any finite disjoint collection of sets $\{A\}_{i=1}^n$ on which μ is defined, $\mu\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mu(A_i)$

§3 September 14, 2022

§3.1 Construction of the Extension

We set

$$\mathcal{M} = \{A \subseteq \Omega \mid \mathbb{P}^*(A \cap E) + \mathbb{P}^*(A^c \cap E) = \mathbb{P}^*(E), \forall E \subseteq \Omega\} \tag{2}$$

That is, $\mathcal{M}$ is set of all subsets A with the property that $\mathbb{P}^*$ is additive on the union of $(A \cap E)$ and $(A^c \cap E)$ for *all* subsets E, i.e.,

$$\mathbb{P}\left((A \cap E) \cup (A^c \cap E)\right) = \mathbb{P}^*(A \cap E) + \mathbb{P}^*(A^c \cap E) = \mathbb{P}(E).$$

Since $E = (A \cap E) \cup (A^c \cap E)$. Note that by sub-additivity, we always have $\mathbb{P}(E) \leq \mathbb{P}^*(A \cap E) + \mathbb{P}^*(A^c \cap E)$, thus (2) is equivalent to

$$SM = \{A \subseteq \Omega \mid \mathbb{P}^*(A \cap E) + \mathbb{P}^*(A^c \cap E) \leq \mathbb{P}^*(E), \forall E \subseteq \Omega\} \tag{3}$$

Proposition 3.1 ($\mathbb{P}^*$ is countably additive on $\mathcal{M}$)**.** If $A_1, A_2, ... \in \mathcal{M}$ are disjoint, then $\mathbb{P}^*\left(\bigcup_n A_n\right) = \sum_n \mathbb{P}^*(A_n)$

Proposition 3.2. Set set $\mathcal{M}$ is a σ-algebra

§3.2 Uniqueness and $\pi - \lambda$ system

Definition 3.3 (π-system) A set of sets $\mathcal{C}$ is called a $\pi-$system if the following holds:

$$A, B \in \mathcal{C}, \quad \implies A \cap B \in \mathcal{C}$$

It follows from induction that for a $\pi-$system $\mathcal{C}$, if $A_1, ..., A_n \in \mathcal{C}$, then we have $A_1 \cap ... \cap A_n \in \mathcal{C}$.

Definition 3.4 (λ-system) A set of sets $\mathcal{C}$ is a $\pi-$system if it satisfies

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

(ii) $A \in \mathcal{C} \implies A^c \in \mathcal{C}$

(iii) $A_n \in \mathcal{C}$, $n \in \mathbb{N}$ and $A_n \cap A_n = \varnothing$ for $n \neq m \implies \cap_n A_n \in \mathcal{C}$

Proposition 3.5. A set of sets that is both a $\pi-$system and a λ is a σ-algebra.

Proposition 3.6 (Dynkin's Theorem)**.** If $\mathcal{C}$ is a π-system, and $\mathcal{D}$ is a $\lambda-$system. Then $\mathcal{C} \subseteq \mathcal{D} \implies \sigma(\mathcal{C}) \subset \mathcal{D}$

§3.3 Lebesgue Measure

Theorem 2.5 allows us to automatically construct valid probability triples. We can use this to construct the **Uniform**$[0, 1]$ distribution. The *Lebesgue measure* is a measure function λ on the interval $\Omega = (0, 1]$ that agrees with our intuitive notion of length. [5] Consider the unit interval $(0, 1]$ together with the field $\mathcal{B}_0$ of finite disjoint unions of sub-intervals and the $\sigma-$field $\mathcal{B} = \sigma(\mathcal{B}_0)$ of *Borel sets* in $(0, 1]$.

$$\lambda(A) = \sum_{i=1}^{k} |b_i - a_i| \quad \text{where } A = \bigcup_{i=1}^{k} (a_i, b_i)$$

By Theorem 2.5, λ extends to $\mathcal{B}$, the extended λ being **Lebesgue measure**. The probability space $((0, 1], \mathcal{B}, \lambda)$ will be the basis for much of the probability theory.

[5] The Lebesgue measure can be defined on different types of intervals (a, b), $[a, b]$, or $[a, b)$ in accordance with the definition above.

§4 September 19, 2022

§4.1 Limit Events

Definition 4.1 (Limit) For a given sequence of subsets $A_1, A_2, ... \subseteq \Omega$, we define

$$\limsup_n A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \{\omega : \omega \in A_n \text{ infinitely often}\}$$
$$= \{\omega : \omega \in A_{n_k}, k = 1, 2, ..\}$$

for some subsequence n_k depending on ω , and

$$\liminf_n A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \{\omega : \omega \in A_n \text{ all but finitely many } n\text{'s}\}$$
$$= \{\omega : \omega \in A_n, \forall n \geq n_0(\omega)\}$$

The **limit** of a sequence of sets is defined as follows:

$$\lim_{n\to\infty} A_n := \limsup_n A_n = \liminf_n A_n$$

Note that ω lies in 4.1 if and only if for each n there is some $k \geq n$ for which $\omega \in A_k$; in other words, ω lies in $\limsup_n A_n$ if and only if it lies in *infinitely many* of the A_n. In the same way ω lines in $\liminf_n A_n$ if and only if there exists some n such that $\omega \in A_k$ for all $k \geq n$; in words, ω lies in $\liminf_n A_n$ (4.1) if and only if it lies in *all but finitely many of the* A_n[6].

- $(\limsup_n A_n)^C = \liminf_n (A_n^C)$ therefore $\mathbb{P}(A_n \ i.o) = 1 - \mathbb{P}(A_n^C \ a.a)$
- Note that $\bigcup_{k=n}^{\infty} A_n \downarrow \limsup_n A_n$;
- and $\bigcap_{k=n}^{\infty} A_n \uparrow \liminf_n A_n$

For example, suppose want to model infinite coin tossing, let $(\Omega, \mathcal{F}, \mathbb{P})$ is infinite fair coin tossing. Then $\Omega = \{(r_1, r_2, \dots) : r_i = 0 \text{ or } 1\}$, is the collection of all binary sequences, and let H_n be the event the n^{th} coin is heads.

Proposition 4.2. The relationship between lim sup and lim inf is

$$\liminf_{n\to\infty} A_n \subset \limsup_{n\to\infty} A_n$$

Proof. (Obvious fact!) If $\omega \in A_n$, for all but finitely many A_n, then certainly, $\omega \in A_n$ infinitely many times. □

Theorem 4.3 (Fatou's) — For each sequence $\{A_n\}$, we always have that

(i) $\mathbb{P}(\liminf_n A_n) \leq \liminf_n \mathbb{P}(A_n) \leq \limsup_n \mathbb{P}(A_n) \leq \mathbb{P}(\limsup_n A_n)$

(ii) If $A_n \to A$, then $\mathbb{P}(A_n) \to \mathbb{P}(A)$

Proof. The middle inequality is obvious. *Prove this!* □

[6] Note that facts about limits inferior and superior can usually be deduced from the logic they involve more easily than by formal set-theoretic manipulations

Corollary 4.4
If $\{A_n, n \geq 1\}$ is sequence of events, then

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\lim_{n\to\infty} A_n\right) \quad \text{(when both exist)}$$

Proof. *Prove this!* □

Example 4.5. Let the following two conditions be true:

$$\begin{cases} \mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = \mathbb{P}\left(\liminf_{n\to\infty} B_n\right) = 1 \\ \mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = \mathbb{P}\left(\limsup_{n\to\infty} B_n\right) = 1 \end{cases}$$

what is $\mathbb{P}\left[\limsup(A_n \cap B_n)\right]$

Proof. The first equation is the stronger,

$$\mathbb{P}\left(\liminf_{n\to\infty} B_n\right) = 1 \implies \mathbb{P}\left(\limsup_{n\to\infty} B_n\right) = 1$$

If $\omega \in \liminf B_n \implies \exists N$ s.t, $\forall n > N$, $\omega \in B_n$. Now if $\omega \in \limsup A_n \implies \exists$ an infinite sequence $\{n(k)\}_{k\in\mathbb{N}}$ such that $\omega \in A_{n(k)}$
Let $\omega \in \{A_{n(k)} \cap B_{n(k)}\}$ for $n(k) > N$, this implies that $\omega \in \limsup(A_n \cap B_n)$ □

§4.2 Independence

- Two events A and B are *independent* if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$
- Two random variables X and Y are *independent* if for all $C, D \in \Re$,

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \cdot \mathbb{P}(Y \in B)$$

- Two $\sigma-$fields are independent if for all $A \in \mathcal{F}$ and $B \in \mathcal{G}$ the events A and B are independent

Definition 4.6 (Independence) More generally, a finite collection $A_1, ..., A_n$ of events is independent if

$$\mathbb{P}(A_{k_1} \cap \cdots \cap A_{k_j}) = \mathbb{P}(A_{k_1}) \cdots \mathbb{P}(A_{k_j})$$

for $2 \leq j \leq n$ and $1 \leq k_1 < \cdots < k_j \leq n$. An infinite collection of events is defined to be independent in each of its finite sub-collections is independent.

Note that the above equation actually represents $\sum_{k=2}^{n} \binom{n}{k} = 2^n - n - 1$ equations.

⋆ It is not enough to assume that $P(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$ for all $i \neq j$. A sequence with this property is called **pairwise independent**. Its obvious that independent events are pairwise independent. But the converse is not true!

In analogy with independence of events we can define the independence of two random vectors and more generally, that of two σ-algebras

Definition 4.7 (Independent collection) Let $S_i \subseteq \mathcal{F}$ for $i = 1, 2, \ldots, n$. The classes S_i are independent, if for any choice $A_1, A_2, \ldots, A_n$ with $A_i \in S_i$ the events $A_1, A_2, \ldots, A_n$ are independent.

This is useful theorem for proving if two sigma fields are independent.

Theorem 4.8 — Let S_1, S_2 be independent fields. Then $\sigma(S_1), \sigma(S_2)$ are independent algebras. Additionally, if S_i for each $i = 1, 2, \ldots, n$ are just *classes* of events satisfying

1. S_i is a $\pi-$system
2. $S_i, i = 1, 2, \ldots, n$ are independent

then, $\sigma(C_1), \cdots, \sigma(C_n)$ are independent.

Proof. The sketch for the proof: We start with fixing $A_2 \in S_2$ for $n = 2$, and show

$$\mathcal{L} = \{A \in \Omega : \mathbb{P}(A \cap A_2) = \mathbb{P}(A) \cdot \mathbb{P}(A_2)\}$$

is a $\lambda-$system. And apply Dynkin's theorem which implies that $\mathcal{L} \supset S_1 \implies \mathcal{L} \supset \sigma(S_1)$ □

§4.3 Borel-Centelli

The Borel-Cantelli lemma is a theorem about sequences of events. The lemma basically says that under certain conditions, an event will have probability either 0 or 1. It belongs to a class of theorems known as *Zero-One laws*[7] .

Lemma 4.9 (Borel-Cantelli Lemma) — Let $A_1, A_2, \cdots \in \mathcal{F}$

(i) If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, then, $\mathbb{P}(\limsup_n A_n) = 0$

(ii) If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, and $\{A_n\}$ are independent, then $\mathbb{P}(\limsup_n A_n) = 1$

Proof. Prove this! □

The theorem asserts that if the sum of the probabilities of events is a finite number, then the set of all outcomes that occur infinitely often must be 0.

[7] Kolmogorov's zero–one law and the Hewitt–Savage zero–one law

§5 September 21, 2022

§5.1 Tail σ-fields

Definition 5.1 (Tail $\sigma-$fields) Given a sequence of events $A_1, A_2, ...$ in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we define their **tail field** by

$$\tau = \bigcap_{n=1}^{\infty} \sigma(A_n, A_{n+1}, A_{n+2}, ...).$$

Or equivalently,

$$\tau = \mathcal{F}\left(\{A_j\}_{j\in\mathbb{N}}\right) \equiv \bigcap_{n=1}^{\infty} \sigma(\{A_j\}_{j\geq n})$$

This is the tail $\sigma-$ field associated with the sequence $\{A_n\}$, and its elements are called *tail events*. In words, an event $B \in \tau$ must have the property that for any n, it depends only on the events $A_n, A_{n+1}, ...$; it does not care about any finite number of events A_n.
Consider tossing a fair coin infinitely many times, and let H_n be the event that n^{th} coin comes up heads. Then $\{\limsup_n H_n\}$ is the event that we obtain infinitely many heads. [8] On the other hand, $\{\liminf_n H_n\}$ is the event that we obtain only finitely many tails.

Example 5.2. Since $\limsup_n A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$, and $\liminf_n \bigcup_{n=1}^{\infty} \bigcap_{k\geq n} A_n$ are both in $\sigma(A_n, A_{n+1}, A_{n+2}, ...)$, limit superior and limit inferior are *tail events* for the sequence $\{A_n\}$.

Proof. For all $N \in \mathbb{N}$, $\limsup_n A_n = \bigcap_{n=1}^{\infty} \bigcup_{k\geq n} A_k = \bigcap_{n=N}^{\infty} \bigcup_{k\geq n} A_k \in \tau$ □

§5.2 Kolmogorov's Zero-One Law

Incomplete!!!

Theorem 5.3 (Kolmogorov's Zero-One Law) — For a sequence $A_1, A_2, \ldots$ of events in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a tail-field τ, and if $A \in \tau$, then $\mathbb{P}(A) = 0$ or 1.

§5.3 Random Variables

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be an arbitrary probability space, and let X be a real-value function on Ω; X is a **simple random variable** if it takes finitely many values S, and

$$[\omega : X(\omega) = s] = X^{-1}(s) \in \mathcal{F}, \quad \forall s \in S$$

Whether or not X satisfies this condition only depends on $\mathcal{F}$, not on $\mathbb{P}$. A finite sum

$$X = \sum_i x_i \mathbb{I}_{A_i}$$

is a random variable if the A_i form a finite partition of Ω into $\mathcal{F}$-sets. If $\mathcal{G}$ is a sub-σ-field of $\mathcal{F}$

Proposition 5.4. Let $X_1, ..., X_n$ be simple random variables.

1. The σ-field $\sigma(X_1, ..., X_n)$ consists of the sets

$$\{\omega : (X_1(\omega), ..., X_n(\omega)) \in H\}$$

[8] Billingsley's defines this much more rigorously

Definition 5.5 (Random variable) Let $(\Omega, \mathcal{F})$ and $(\mathbb{S}, \mathcal{S})$ be two measurable spaces. A function $X : \Omega \mapsto \mathbb{S}$ between is **Random variable** (R.V) if

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

for every $B \in \mathcal{S}$. This is also the definition of a **measurable mapping** [a]. A *real-valued* random variable X is a function from Ω to $\mathbb{R}$ such that

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}, \quad x \in \mathbb{R}$$

which can be written as $\{X \leq x\} \in \mathcal{F}$ or $X^{-1}((\infty, x]) \in \mathcal{F}$.

[a] If Ω is a topological space and $\mathcal{F} = \sigma(\{\mathcal{O} \subseteq \Omega \text{ open}\})$ is the corresponding Borel $\sigma-$algebra, we say that $X : \Omega \mapsto \mathbb{R}$ is a **Borel function**

⋆ Requirement of wanting pre-images of a measurable function to be in $\mathcal{F}$ comes from how we will soon define the Lebesgue Integral. Lets say we have an indicator function $\mathbb{I}_A$. The $\int_\Omega \mathbb{I}_A(\omega) d\mathbb{P}(\omega) = \mathbb{P}(f^{-1}(\{1\}))$ should equal to the measure of the set A. Hence the pre-image of any set, $f^{-1}(A) \in \mathcal{F}$ (i.e., shoudl be measurable)

As we can see random variables are just measurable functions that get special notation. Lets say we have an random variable Y, and we would like to know the probability it lies between the interval a and b. The rigorous way to express this would be

$$\mathbb{P}(Y^{-1}([a, b])) = \mathbb{P}(\{\omega \in \Omega \mid Y(\omega) \in [a, b]\})$$

This is a bit cumbersome, which is why the notation "$Y \in B$" gets used to express the same thing as $Y^{-1}(B)$. Therefore, we have

$$\mathbb{P}(Y^{-1}([a, b])) = \mathbb{P}(Y \in [a, b])$$

Proposition 5.6. Let X and Y be independent random variables. Let $f, g : \mathbb{R} \to \mathbb{R}$ be *Borel-measurable functions.* Then the random variables $f(X)$ and $g(X)$ are independent.

Proof. *Complete this* We want to show that

$$\mathbb{P}(f(X) \in S_1, g(Y) \in S_2) = \mathbb{P}(f(X) \in S_1) \cdot \mathbb{P}(g(Y) \in S_2)$$

□

Proposition 5.7. Let X and Y be two random variables, defined jointly on some probability triple $(\Omega, \mathcal{F}, \mathbb{P})$. Then X and Y are independent if and only if $\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y)$ for all $x, y \in \mathbb{R}$.

§5.4 σ-algebras induced by Inverse Maps

Definition 5.8 Given a function $X : \Omega \to \mathbb{R}$ we denote $\sigma(X)$ as the smallest $\sigma-$algebra $\mathcal{F}$ such that $X(\omega)$ is a measurable mapping from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B})$. Alternatively,

$$\begin{aligned} \sigma(X) &= \sigma(\{\omega \in \Omega : X(\omega) \in B\}, B \in \mathcal{B}) \\ \Leftrightarrow \sigma(X) &= \{\underbrace{\{X \in B\}}_{=X^{-1}(B)} : B \in \mathcal{B}\} \\ \Leftrightarrow \sigma(X) &= X^{-1}(\mathcal{B}) \end{aligned}$$

Extreme Example: Let $X(\omega) = 17$ for all ω. This mean that

$$\begin{aligned} \sigma(X) &= \{\{X \in B\} : B \in \mathcal{B}\} \\ &= \sigma(\varnothing, \Omega) \end{aligned}$$

Less Extreme Example: Suppose that $X = \mathbb{I}_A$ for some $A \in \mathcal{F}$, then X has range of $\{0,1\}$. Then

$$X^{-1}(\{0\}) = A^C, \quad X^{-1}(\{1\}) = A$$

and therefore,

$$\sigma(X) = \{\varnothing, \Omega, A, A^C\}$$

The results below states that a map between spaces where one of the spaces is measurable implicitly define a σ-algebra on the other one

Theorem 5.9 (σ-algebras Induced by Inverse Maps) — Let $f : \mathbb{X} \to \mathbb{Y}$ be a map, and $\mathbb{X}$ and $\mathbb{Y}$ are non-empty sets.

1. If $\mathcal{B}$ is a σ-algebra on Y, then

$$\sigma(f) = f^{-1}(\mathcal{B}) = \{f^{-1}(B) : B \in \mathcal{B}\}$$

is a σ-algebra on $\mathbb{X}$

2. If $\mathcal{F}$ is a σ-algebra on $\mathbb{X}$, then

$$\{B \subseteq Y : f^{-1}(B) \in \mathcal{F}\}$$

is a σ-algebra on $\mathbb{Y}$.

3. $\mathcal{A}$ is a colleciton of sets in $\mathbb{Y}$, then

$$\sigma(f^{-1}(\mathcal{A})) = f^{-1}(\sigma(\mathcal{A}))$$

The σ-algebra $\mathcal{S}$ can be huge, so its useful to know that we can verify that a given mapping is measurable without the need to check that the pre-image $X^{-1}(B)$ is in $\mathcal{F}$ for every $B \in \mathcal{S}$. It suffices to do this for a collection (of our choice) of generators of $\mathcal{S}$.

Theorem 5.10 — Suppose that $(\Omega, \mathcal{F})$ and $(\mathbb{S}, \mathcal{S})$ are two measurable spaces, and $\mathcal{S} = \sigma(\mathcal{A})$ is generated by the collection of sets $\mathcal{A}$ in $\mathbb{S}$. Then $X : \Omega \to \mathbb{S}$ is a measurable if and only if

$$X^{-1}(A) \in \mathcal{F} \quad \textit{for every } A \in \mathcal{A}$$

Proof. Based on the following facts:

1. $X^{-1}(\mathbb{S} \setminus B) = \Omega \setminus X^{-1}(B)$

2. $X^{-1}\left(\bigcup_{i=1}^{\infty} B_i\right) = \bigcup_{i=1}^{\infty} X^{-1}(B_i)$

It follows that

$$\mathcal{M} = \{B \subseteq \mathbb{S} : X^{-1}(B) \in \mathcal{F}\}$$

σ-algebra on $\mathbb{S}$. By assumption $\mathcal{M} \supset \mathcal{A}$ and therefore, $\mathcal{M} \supset \sigma(\mathcal{A}) = \mathcal{S}$, which implies that X is measurable □

Alternatively, A function f defined on a measurable subset E of $\mathbb{R}^d$ is **measurable** if for all $a \in \mathbb{R}$

$$f^{-1}([-\infty, a)) = \{x \in E : f(x) < a\}$$

is measurable.

Definition 5.11 If $(\Omega, \mathcal{F})$ is a measurable space, then

- $X : \Omega \to \mathbb{R}$ is **measurable** if $X^{-1}(B) \in \mathcal{F}$ for every Borel set $B \in \mathcal{B}(\mathbb{R})$.
- A function $X : \mathbb{R}^n \to \mathbb{R}$ is **Lebesgue measurable** if $X^{-1}(B)$ is a *Lebesgue measurable* subset of $\mathbb{R}^n$ for every Borel subset B of $\mathbb{R}$
- A function $X : \mathbb{R}^n \to \mathbb{R}$ is **Borel measurable** if $X^{-1}(B)$ is a *Borel measurable* subset of $\mathbb{R}^n$ for every Borel subset B of $\mathbb{R}$.

Proposition 5.12. If $(\Omega, \mathcal{F})$ is a measurable space, then $X : \Omega \to \mathbb{R}$ is a measurable function if and only if one of the following conditions holds:

$$\{\omega \in \Omega : X(\omega) < x\} \in \mathcal{F} \quad \textit{for every } x \in \mathbb{R}$$
$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F} \quad \textit{for every } x \in \mathbb{R}$$
$$\{\omega \in \Omega : X(\omega) > x\} \in \mathcal{F} \quad \textit{for every } x \in \mathbb{R}$$
$$\{\omega \in \Omega : X(\omega) \geq x\} \in \mathcal{F} \quad \textit{for every } x \in \mathbb{R}$$

§6 September 26, 2022

§6.1 Existence of i.i.d Sequences

We started this lecture the remark that proving the existence of i.i.d sequences is tricky, and usually done towards the end of most Probability textbooks.

Theorem 6.1 — Assume $X_n \sim F_n$ (independently distributed), assume F_n^{-1} exists. We fix a sequence of random variables $\{F_n\}_{n\in\mathbb{N}}$. There exists independent sequence of random variables $\{X_n\}_{n\in\mathbb{N}}$ where $X_n \sim F_n$. [a]

[a] We fixing the the sequence of distributions

§6.2 Kolmogorov's Extension Theorem

Let

$$\begin{aligned} \mathcal{B}^n &= \sigma\left(\{A_1 \times A_2 \times \cdots \times A_n\}_{A_i \in \mathcal{B}}\right) \\ &= \sigma\left(\{[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]\}\right) \quad \text{where } a_i < b_i \in \mathbb{Q} \end{aligned}$$

Then we have

$$\mathcal{B}^{\mathbb{N}} = \sigma\left(\{A_1 \times A_2 \times \cdots \times A_n \times \mathbb{R} \times \cdots\}\right)$$

Theorem 6.2 (Kolmogorov's Extension Theorem) — Suppose we are given sequence of probability measures $\{\mu\}_{n\in\mathbb{N}}$ on $(\mathbb{R}^n, \mathcal{B}^N)$, on a "growing probability space" [a].

$$\mu_{n+1}(\{A_1 \times A_2 \dots A_n \times \mathbb{R}\}) = \mu_{n+1}(\{A_1 \times A_2 \dots A_n\})$$

There exists a (*unique?*) measure μ on $(\mathcal{B}^{\mathbb{N}}, \mathbb{R}^{\mathbb{N}})$ such that

$$\mu_n(\{A_1 \times A_2 \dots A_n \times \mathbb{R} \times \mathbb{R} \times \cdots\}) = \mu_n(\{A_1 \times A_2 \dots A_n\})$$

[a] verify what this means

Example 6.3 (Example from)**.** Let $F_1, F_2, \dots$ be distribution functions and let μ_n be the measure on $\mathbb{R}^n$ with

$$\mu\left(\{[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]\}\right) = \prod_{m=1}^{n} \left(F_m(b_m) - F_m(a_m)\right)$$

§6.3 Measurable maps and R.V's

Theorem 6.4 — If $X : (\Omega_1, \mathcal{F}_1) \to (\Omega_2, \mathcal{F}_2)$ and $f : (\Omega_2, \mathcal{F}_2) \to (\Omega_3, \mathcal{F}_3)$ are all **measurable maps** then $f(X)$ is a measure map from $(\Omega, \mathcal{F}) \to (\Omega_3, \mathcal{F}_3)$.

Proof. $(X \circ f)^{-1}(A_3) = X^{-1}\underbrace{(f^{-1}(A_3))}_{\in \mathcal{F}_2} \implies$ Since X is a measurable map,

$$X^{-1}(f^{-1}(A_3) \in \mathcal{F}_1$$

□

Theorem 6.5 — If f is a continuous function, it is measurable.

Proof. *prove this* □

Given collection of real-valued functions $\{f_i\}_{i=1}^{\infty}$, we define **new functions** by specifying their values at each $x \in \mathbb{R}$.

$$\inf_i f_i(x) = \inf f_i(x) = \inf_i \{f_i(x)\}$$
$$\sup_i f_i(x) = \sup f_i(x) = \sup_i \{f_i(x)\}$$
$$\limsup_i f_i(x) = \limsup f_i(x) = \inf_{j>1} \sup_{i>j} \{f_i(x)\}$$
$$\liminf_i f_i(x) = \liminf f_i(x) = \sup_{j>1} \inf_{i>j} \{f_i(x)\}$$

Theorem 6.6 — If $X_1, X_2, \dots$ are random variables then so are

$$\inf_n X_n \quad \sup_n X_n \quad \limsup_n X_n \quad \liminf_n X_n$$

Proof. Let $g = \sup_n X_n$ so for each i, $X_i(\omega) \le g(\omega)$ for all $\omega \in \Omega$. Fix $a \in \mathbb{R}^+$ and observe that

$$\{\omega : max\{X_1(\omega), X_2(\omega)\} > a\} = \{\omega : X_1(\omega) \ge a\} \bigcup \{\omega : X_2(\omega) \ge a\}$$

The set of ω's that $\sup_n X_n > a$ is equal to the union of all the sets of ω's where $X_i > a$ for all $i \in \mathbb{N}$ □

§6.4 Almost Sure Convergence

Point-wise convergence of R.V. i.e. $(X_n(\omega) \to X(\omega)$ for every $\omega \in \Omega)$ is often too strong of a requirement, since the R.V. may not be defined on sets of zero measure.

Definition 6.7 A sequence of measurable functions or random variables $\{X_n\}$ on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ converge **almost surely** to X if

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

Another way to define **almost sure convergence** is:
Two R.V's X and Y are almost surely the same if $\mathbb{P}(\{\omega : X(\omega) \ne Y(\omega)\}) = 0$

Lemma 6.8 — If $\mathbb{P}[\{|X_n - X| \ge \varepsilon\}\ i.o] = 0$, then $X_n \xrightarrow{a.s} X$.

Proof. Let $A = \{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) \ne X\}$. Then for all $\omega \in A$, $\exists \delta > 0, n(k)$ such that $|X_{n(k)}(\omega) - X(\omega)| > \delta$ for all $k \in \mathbb{N}$. But then $|X_n(\omega) - X(\omega)| > \frac{1}{n}$ i.o, so

$$\omega \in B \equiv \{|X_n - X| \ge \frac{1}{n}\ i.o\} \implies \mathbb{P}(A) \le \mathbb{P}(B) = 0$$

□

§7 September 28, 2022

§7.1 Convergence in Probability

Recall: "$X_n \underset{a.s}{\to} X$" $\Leftrightarrow \mathbb{P}(\{\omega : X_n(\omega) \to X(\omega)\}) = 1$

Definition 7.1 (Convergence in Probability) Suppose that $\{X_n\}_{n\geq 1}$ and X are random variables. Then $\{X_n\}$ converges in probability to X, written $X_n \overset{p}{\to} X$, if for any $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}[|X_n - X| > \varepsilon] = 0$$

- Almost sure convergence of $\{X_n\}$ demands that for "almost all" ω, $X_n(\omega) - X(\omega)$ gets small
- $X_n(\omega)$ fails to converge to $X(\omega)$ if and only if there is some ε such that for no m does $|X_n(\omega) - X(\omega)|$ below ε for all n exceeding m
- Convergence in probability is weaker and merely requires that all the probability of the difference $X_n(\omega) - X(\omega)$ becomes small

Theorem 7.2 (Convergence a.s implies convergence in probability) — Suppose that $\{X_n\}_{n\geq 1}$ and X are random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If

$$X_n \to X, \quad \text{almost surely (a.s)}$$

then

$$X_n \overset{p}{\to} X$$

Proof. If $X_n \overset{a.s}{\to} X$, it means the set of points $A = \{\omega : \lim X_n(\omega) \neq X(\omega)\}$ has probability 0. Lets fix $\varepsilon > 0$, define

$$A_n = \bigcup_{m\geq n} \{\omega : |X_m(\omega) - X_n(\omega)| > \varepsilon\}$$

Clearly A_n's are monotonically decreasing, i.e., $A_n \supset A_{n+1} \supset A_{n+2} \cdots$, and they are decreasing towards the set

$$A_\infty = \bigcap_{n\geq 1} A_n$$

For these decreasing sequence of sets, their probabilities are also a decreasing sequence, and the probabilities are decreasing to $\mathbb{P}(A_\infty)$; we will show that $\mathbb{P}(A_\infty) = 0$. Now for any point ω in the complement of A is such that $\lim X_n(\omega) = X(\omega)$, which implies $|X_n(\omega) - X(\omega)| < \varepsilon$ for all n greater than a number N. Therefore, for all $n \geq N$, ω will not belong to A_n, and consequently will not belong to A_∞. This implies that A^c and A_∞ are *disjoint* events, which in turn implies that A_∞ is a subset of A which implies that $\mathbb{P}(A_\infty) = 0$. By continuity from above, we have $\lim \mathbb{P}(A_n) = \mathbb{P}(A_\infty) = 0$, and

$$\mathbb{P}(|X_n - X)| > \varepsilon) \leq \mathbb{P}(A_n) \to 0, \quad n \to \infty$$

Second Proof:
If $X_n \to X$ a.s. then for any $\varepsilon > 0$,

$$0 = \mathbb{P}([|X_n - X| > \varepsilon] i.o) =$$

□

Example 7.3. Let $X_n \sim Bern(\frac{1}{n})$. Fix $\delta > 0$.

$$\mathbb{P}(|X_n| > \delta) \leq \mathbb{P}(X_n = 1) = \frac{1}{n} \to 0, \quad n \to \infty$$

This implies that $X_n \xrightarrow{p} 0$. Now we have

$$\mathbb{P}\left(|X_n| > \frac{1}{2}\right) = \frac{1}{n} \implies \sum_{n=1}^{\infty}\left(|X_n| > \frac{1}{2}\right) = \infty$$

By Borel-Cantelli lemma

$$\mathbb{P}\left(|X_n| > \frac{1}{2} \quad \text{i.o}\right) = 1$$

therefore, this sequence of variables does not converge almost surely.

§7.2 Summary of convergences

Let $X_1, \cdots, X_n$ be random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say:

(a) $X_n \to X$ **almost surely**, written $X_n \xrightarrow{a.s} X$, if $\{\omega \in \Omega : X_n(\omega) \to X(\omega) \text{ as } n \to \infty\}$ is an event with probability 1.[9]

(b) $X_n \to X$ **in rth mean**, where $r \geq 1$, written $X_n \xrightarrow{r} X$, if $\mathbb{E}[X_n^r] < \infty$ for all n and

$$\mathbb{E}(|X_n - X|^r) \to 0 \quad \text{as} \quad n \to \infty,$$

(c) $X_n \to X$ **in probability**, written $X_n \xrightarrow{p} X$, if

$$\mathbb{P}(|X_n - X| > \varepsilon) \to 0 \quad \text{as} \quad n \to \infty, \quad \text{for all } \varepsilon > 0,$$

(d) $X_n \to X$ **in distribution**, written $X_n \xrightarrow{D} X$, if

$$\mathbb{P}(X_n \leq x) \to \mathbb{P}(X \leq x) \quad \text{as} \quad n \to \infty$$

There are several notations for $X_n \xrightarrow{a.s} X$. They include $X_n \to X$ *almost everywhere*, or $X_n \xrightarrow{a.e} X$, or $X_n \to X$ *with probability 1*. We can check using the Minkowski's inequality that

$$\| Y \|_r = (\mathbb{E}|Y^r|)^{\frac{1}{r}} = \left(\int |y|^r dF_Y\right)^{\frac{1}{r}}$$

define a norm on the collection of random variables with a finite rth moment, for any any value of $r \geq 1$. Its obvious that these four modes of convergence are not equivalent to each other. Convergence in distribution is the weakest, sine its a condition only on *distribution functions* of X_n, that is, it contains no reference to to the sample space Ω, and no information regarding the independence or dependence of X_n.

Example 7.4. Let X be a Bernoulli variable taking values 0 and 1 with equal probability $\frac{1}{2}$. Let $X_1, X_2 \cdots$ be identical random variables given by $X_n = X$ for all n. The X_n are clearly not independent, but $X_n \xrightarrow{D} X$. Let $Y = 1 - X$. Certainly $X_n \xrightarrow{D} Y$ also, since X and Y have the same distribution. However, X_n cannot converge to Y is any other mode because $|X_n - Y| = 1$.

§7.3 Lebesgue Integral

We define expectation for all simple positive random variables and then for all $(0, \infty)-$*valued* random variables. Let

$$SF_+ := \{X \geq 0 : X : (\Omega, \mathcal{F}) \mapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))\}$$

be non-negative, measurable functions with domain Ω. We fix measure space $(\Omega, \mathcal{F}, \mathbb{P})$ and we define the $\mathbb{E}(X)$ by the following four step procedure.

1. For $A \in \mathcal{F}$, define $\mu_0(\mathbb{I}_A) := \mathbb{P}(A)$

[9] we do not require the whole space Ω, but rather that its complement, is a null set

2. For any $f \in SF_+$, has a representation $f = \sum_{j=1}^{n} a_j \mathbb{I}_{A_j}$; for some finite $n < \infty$, and $c_j > 0$, which gives us the definition of the integral

$$\mu_0(f) = \sum_{j=1}^{n} a_j \mathbb{P}(A_j)$$

3. for all random variables $X \geq 0$, we define

$$\mathbb{E}[X] := \sup\{\mu_0(f) : f \in SF_+, f(\omega) \leq X(\omega)\}$$

$$\Leftrightarrow \mathbb{E}[X] = \sup\{\mathbb{E}[Y] : Y \text{ is simple }, Y(\omega) \leq X(\omega)\}$$

4. For X, let $X^+ = max(X, 0)$ and $X^- = max(-X, 0)$, then $\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-]$

$\star$ Expectation $\mathbb{E}[X]$ of a random variable X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is just the Lebesgue integral $\int X(\omega) d\mathbb{P}(\omega)$ of X with respect to $\mathbb{P}$

The expected value of a simple random variable is defined as

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_i x_i \mathbb{I}_{A_i}\right] = \sum_i x_i \mathbb{P}(A_i)$$

or in the alternative form

$$\mathbb{E}[X] = \sum_x x_i \mathbb{P}[X = x]$$

where the sum is extending over the range of X; We derive some important properties of μ_0 from our definition. In particular want to to prove μ_0 is invariant, linear and monotone.

Lemma 7.5 — μ_0 is "well-defined"

(a) $\mu_0(\varphi) = \mu_0(\psi)$ if $\varphi, \psi \in SF_+$ and $\mu_0(\{\omega : \varphi(\omega) = \psi(\omega)\}) = 0$

(b) μ_0 is linear, that is $\mu_0(\varphi + \psi) = \mu_0(\varphi) + \mu_0(\psi)$

(c) μ_0 is monotone that is, for all $\omega \in \Omega \varphi(\omega) \leq \psi(\omega) \Leftrightarrow \mu_0(\varphi) \leq \mu_0(\psi)$

Theorem 7.6 (Monotone Convergence Theorem) — Suppose that $X_1, X_2, ...$ are random variables and $0 \leq \{X_n(\omega)\}_{n=1}^{\infty} \uparrow X(\omega)$ for all $\omega \in \Omega$, then X is a random variable

$$\mathbb{E}[X_n] \uparrow \mathbb{E}[X] \Leftrightarrow \lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$$

Lemma 7.7 — Now we prove a similar lemma for expectations:

(a) The mathematical expectation is is well-defined $\mathbb{E}(\varphi) = \mathbb{E}(\psi)$ if $\varphi, \psi \in SF_+$ and $\mathbb{E}(\{\omega : \varphi(\omega) = \psi(\omega)\}) = 0$

(b) Expectation is a linear operation, that is $\mathbb{E}(\varphi + \psi) = \mathbb{E}(\varphi) + \mathbb{E}(\psi)$

(c) $\mathbb{E}$ is monotone that is, for all $\omega \in \Omega \quad \varphi(\omega) \leq \psi(\omega) \Leftrightarrow \mathbb{E}(\varphi) \leq \mathbb{E}(\psi)$

Relation to Riemann Integral. Let $f : [a, b] \to \mathbb{R}$ be a Riemann integrable function, where $[a, b]$ is equipped with Borel measure. Then f is also Lebesgue integrable, and the integrals agree:

$$\int_a^b f(x) dx = \int_{[a,b]} f d\mu$$

Thus in practice we do all theory with Lebesgue integrals, since they are nicer, but when we need to actually compute anything, we revert back to the fundamental theorem of calculus.

There are three "big" theorem for exchanging limits with Lebesgue integrals:

1. Monotone convergence theorem

2. Fatou's Lemma: most general statement since it can be applied to any non-negative measurable function

3. Dominated Convergence theorem [10]: bounded by some absolutely integrable functions (i.e, limits are not too big)

[10] Bounded convergence theorem is a special case of dominated convergence

§8 October 3, 2022

§8.1 Important properties of Lebesgue integrals

Theorem 8.1 (Beppo-Levi) — Let $\{X_n\}$ be a sequence of monotone (*increasing?*) random variables, then

$$\mathbb{E}[\sup_{n\in\mathbb{N}} X_n] = \sup_{n\in\mathbb{N}} \mathbb{E}[X_n]$$

$$\int \sup_{n\in\mathbb{N}} X_n = \sup_{n\in\mathbb{N}} \int X_n$$

Lemma 8.2 — For every $\mathbb{R}-$valued R.V. $X(\omega)$ there exists a sequence of simple functions $X_n(\omega)$ such that $X_n(\omega) \to X(\omega)$ (*converge point wise*).

Proof. Let

$$X_n(\omega) \equiv n\mathbb{I}_{x>n} + \sum_{k=0}^{n2^n-1} \frac{k}{2^n} \mathbb{I}_{X(\omega)\in(\frac{k}{2^n},\frac{k+1}{2^n})}$$

Note that:

1. $X_n(\omega) \leq X_{n+1}(\omega) \leq X(\omega)$ for all $n \in \mathbb{N}$
2. $X(\omega) - X_n(\omega) = 2^{-n}$ for all $n \in \mathbb{N}$, it follows that $X_n(\omega) \to X(\omega)$.

□

Theorem 8.3 (Bounded Convergence) — If $\{X_n\}$ is uniformly bounded, and $X_n \overset{a.s}{\to} X$, then

$$\mathbb{E}[X_n] \to \mathbb{E}[X], \quad n \to \infty$$

Proof. Let $C > \sup_n |X_n|$, then $|X - X_n| \leq 2C$. Fix $\delta > 0$, and let

$$A_n\{\omega : |X(\omega) - X_n(\omega)| > \delta\}$$

. Then

$$|X(\omega) - X_n(\omega)| \leq \delta\mathbb{I}_{A^c}(\omega) + 2C\mathbb{I}_{A_n}(\omega) \leq \delta + 2C\mathbb{I}_{A_n}(\omega), \quad \text{for all } \omega$$

This implies that

$$\mathbb{E}[|X(\omega) - X_n(\omega)|] \leq \delta + 2C\mathbb{E}[\mathbb{I}_{A_n}(\omega)] = \delta + 2C\mathbb{P}[X(\omega) - X_n(\omega)| > \delta]$$

Since $X_n \overset{p}{\to} X$, this implies that $\mathbb{P}(A_n) \to 0$, and since δ is arbitrary,

$$\mathbb{E}[|X(\omega) - X_n(\omega)|] \to 0 \Leftrightarrow \mathbb{E}[X_n] = \mathbb{E}[X]$$

□

§8.2 Inequalities

Markov's inequality. Let X be an arbitrary non-negative random variable. Then for all $\alpha > 0$,

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

Proof. We define a new random variable Z by

$$Z(\omega) = \begin{cases} \alpha & \text{if } X(\omega) \geq \alpha \\ 0 & \text{if } X(\omega) < \alpha \end{cases}$$

Clearly $Z \leq X \implies \mathbb{E}(Z) \leq \mathbb{E}(X) \Leftrightarrow \alpha\mathbb{P}(X \geq \alpha) \leq \mathbb{E}(X)$. □

Chebychev's inequality Let X be an arbitrary random variable with $\mathbb{E}(X) < \infty$. Then for all $\alpha > 0$,

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \alpha) \leq \frac{\mathbb{V}ar(X)}{\alpha^2}$$

The probability that X differs from its mean by more than α is bounded above by its variance divided by α^2.

Proof. To prove the Chebychev's inequality we apply Markov's:

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \alpha) = \mathbb{P}(|X - \mathbb{E}(X)|^2 \geq \alpha^2) \leq \frac{\mathbb{E}(X - \mathbb{E}(X))^2}{\alpha^2} = \frac{\mathbb{V}ar(X)}{\alpha^2}$$

□

Jensen's inequality
Hölder's Inequality

$$\mathbb{E}[XY] \leq \mathbb{E}[X^P]^{\frac{1}{p}} \cdot \mathbb{E}[Y^q]$$

§9 October 5, 2022

§9.1 Inequalities Continued

Minkowski's Inequality.

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

§9.2 Law of Large Numbers (LLN)

Let $X_1, \dots, X_n$ be a sequence of random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. They are *identically distributed* if their distributions are all the same. We define $S_n = X_1 + \cdots + X_n$. We are interested in the asymptotic behaviour of S_n as $n \to \infty$[11]. The general problem can be described as under what conditions does the following convergence occur:

$$\frac{S_n}{b_n} - a_n \to S \quad \text{as} \quad n \to \infty$$

where a The **weak law of large numbers** is

> **Theorem 9.1** (Weak Law of large numbers) —
>
> $$\lim_{n\to\infty} \mathbb{P}\left(| \frac{1}{n}(X_1 + \cdots + X_n) - \mu | \geq \varepsilon \right) = 0$$

In words, the partial averages $\frac{1}{n}(X_1 + \cdots + X_n)$ converge in probability to μ. The **strong law of large numbers** is:

> **Theorem 9.2** (Strong Law of Large Numbers) — Let $X_1, \dots, X_n$ be independent and identically distributed (i.i.d) and $\mathbb{E}(X_i) = m < \infty$, then,
>
> $$\mathbb{P}\left(\lim_{n\to\infty} \frac{1}{n}(X_1 + \cdots + X_n) = \mu \right) = 1$$
>
> Or in other words,
>
> $$\frac{S_n}{n} \overset{a.s}{\to} m \qquad \text{as } n \to \infty$$

In words the partial averages $\frac{1}{n}(X_1 + \cdots + X_n)$ converge *almost surely* to μ.

Proof. Suppose that X_i are non-negtive random variables with $\mathbb{E}|X| = \mathbb{E}[X_1] < \infty$, and let $\mu = \mathbb{E}[X_i]$. We 'truncate' the X_n to obtain a new sequence $\{Y_n\}$ given by

$$Y_n = X_n I_{\{X_n < n\}} = \begin{cases} X_n & \text{if } X_n < n, \\ 0 & \text{if } X_n \geq n. \end{cases}$$

□

[11] this long-term behaviour depends crucially upon the sequence $\{X_n\}$

§10 October 17, 2022

§10.1 Bernstein Approximation Theorem

Theorem 10.1 (Bernstein Polynomial) — Let $f : [0, 1] \to \mathbb{R}$ be a continuous function on the interval $[0, 1]$. The *Bernstien polynomial* of degree n associated with f is

$$B_n(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{x} x^k (1-x)^{n-k}$$

If f is continuous, then $B_n(x)$ converges to $f(x)$ uniformly on $[0, 1]$. i.e,

$$\sup_x |f(x) - f(y)| \to 0$$

Proof. Let $M = \sup_x |f(x)|$, $\delta(\varepsilon) = \sup\{|f(x) - f(y)| : |x - y| < \varepsilon\}$. Our objective will be to show that

$$\sup_x |f(x) - B_n(x)| \leq \delta(\varepsilon) + \frac{2M}{n\varepsilon^2}$$

By uniform continuity of f, we have $\lim_{\varepsilon \to 0} \sup\{|f(x) - f(y)| : |x - y| < \varepsilon\} = 0$
Fix: $x \in [0, 1]$, and let $X_1, ..., X_n \sim Bern(x)$, $S_n = X_1 + ...X_n$, then

$$|B_n(x) - f(x)| \leq \mathbb{E}\left[|f\left(\frac{S_n}{n}\right) - f(x)|\right]$$

□

§10.2 Connections between a.s and i.p convergence

Theorem 10.2 — If

$$\sum_n \mathbb{P}(A_n) = 0, \quad \text{and} \quad \liminf_{n\to\infty}\left(\frac{\sum_{j,k\leq n} \mathbb{P}(A_j \cap A_k)}{\sum_{j\leq n} \mathbb{P}(A_j)^2}\right) \leq 1,$$

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

Theorem 10.3 — If $X_n \xrightarrow{p} X$ then there exists a subsequence $\{X_{n(k)}\}$ such that $X_{n(k)} \xrightarrow{a.s} X$.

Lemma 10.4 — If $\mathbb{P}[\{|X_n - X| \geq \varepsilon\}\ i.o] = 0$, then $X_n \xrightarrow{a.s} X$.

Proof. Let $A = \{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) \neq X\}$. Then for all $\omega \in A$, $\exists \delta > 0, n(k)$ such that $|X_{n(k)}(\omega) - X(\omega)| > \delta$ for all $k \in \mathbb{N}$. But then $|X_n(\omega) - X(\omega)| > \frac{1}{n}$ i.o, so

$$\omega \in B \equiv \{|X_n - X| \geq \frac{1}{n}\ i.o\} \implies \mathbb{P}(A) \leq \mathbb{P}(B) = 0$$

□

§11 October 19, 2022

§11.1 Applications of Strong Law of Large Numbers

Example 11.1 (The Glivenko-Cantelli Theorem)**.** As $n \to \infty$, $\sup |F_n(x) - F(x)| \to 0$ a.s.

$$1(X_m \leq x) \tag{4}$$

The next result shows that F_n converges uniformly to F as $n \to \infty$.

Proof: Fix x and let $Y_n = 1(X \leq x)$. Since the Y_n are i.i.d. with $EY_n = P(X_n \leq x) = F(x)$, the strong law of large numbers implies that $F_n(x) = \frac{1}{n}\sum_{m=1}^{n} Y_m \to F(x)$ a.s. In general, if F_n is a sequence of nondecreasing functions that converges pointwise to a bounded and continuous limit F, then $\sup_x |F_n(x) - F(x)| \to 0$. However, the distribution function $F(x)$ may have jumps, so we have to work a little harder.

Again, fix x and let $Z_n = 1(X_n < x)$. Since the Z_n are i.i.d. with $EZ_n = P(X_n < x) = F(x^-) = \lim_{y \uparrow x} F(y)$, the strong law of large numbers implies that $F_n(x^-) = \frac{1}{n}\sum_{m=1}^{n} Z_m \to F(x^-)$ a.s. For $1 \leq j \leq k-1$, let $x_{j,k} = \inf\{y : F(y) \geq j/k\}$. The pointwise convergence of $F_n(x)$ and $F_n(x^-)$ imply that we can pick $N_k(\omega)$ so that if $n \geq N_k(\omega)$, then

$$|F_n(x_{j,k}) - F(x_{j,k})| < k^{-1} \quad \text{and} \quad |F_n(x_{j,k^-}) - F(x_{j,k^-})| < k^{-1} \tag{5}$$

for $1 \leq j \leq k-1$. If we let $x_{0,k} = -\infty$ and $x_{k,k} = \infty$, then the last two inequalities hold for $j = 0$ or k. If $x \in (x_{j-1,k}, x_{j,k})$ with $1 \leq j \leq k$ and $n \geq N_k(\omega)$, then using the monotonicity of F_n and F, and $F(x_{j,k^-}) - F(x_{j-1,k}) \leq k^{-1}$, we have

$$F_n(x) \leq F_n(x_{j,k^-}) \leq F(x_{j,k^-}) + k^{-1} \leq F(x_{j-1,k}) + 2k^{-1} \leq F(x) + 2k^{-1} \tag{6}$$

$$\tag{7}$$

§11.2 Examples

§12 November 2, 2022

The following are equivalent:

- $F : \mathbb{R} \to [0, 1]$

§13 November 14, 2022

We want to be able to describe the distribution of likelihoods of possible values of X. We can do this through a distribution function. Recall from Probability 101 that distribution of X evaluated at x is just the probability that X will taken on a value less than or equal to x. More formally, we say the **distribution** or *law* of the random variable X is the probability measure μ on $\mathbb{R}, \mathcal{B}(\mathbb{R})$ defined by

$$\mu(A) = \mathbb{P}[X \in A], \quad A \in \mathcal{B}(\mathbb{R})$$

For simple R.V's distribution is defined for every subset of the line, however, this is clearly not the case anymore. From now on μ will be defined only for Borel subsets. The **distribution function** of X is the function $F : \mathbb{R} \to [0, 1]$ defined as

$$F(x) = \mu(-\infty, x) = \mathbb{P}(X^{-1}(-\infty, x)) = \mathbb{P}(X \leq x), \quad x \in \mathbb{R}$$

⋆ What is the key difference between a distribution and a distribution function?

Lemma 13.1 — A distribution function F has the following properties:

(a) $\lim_{x\to-\infty} F(x) = 0, \quad \lim_{x\to\infty} F(x) = 1$

(b) if $x < y$, then $F(x) \leq F(y)$,

(c) F is right-continuous, that is $F(x + h) \to F(x)$ as $\downarrow 0$

Proof. (a) Let $B_n = \{\omega \in \Omega : X(\omega) \leq -n\} = \{X \leq -n\}$. The sequence $B_1, B_2, \cdots$ is decreasing with the empty set as limit. Thus, $\mathbb{P}(B_n) \to \mathbb{P}(\varnothing) = 0$.

(b) Let $A_x = \{\omega \in \Omega : X(\omega) \leq x\} = \{X \leq x\} \subseteq \Omega$. Similarly, let $A_{x,y} = \{x < X \leq y\}$. It follows that $A_y = A_x \cup A_{x,y}$, which is a disjoint union, and by definition of $\mathbb{P}$,

$$\mathbb{P}(A_y) = \mathbb{P}(A_x) + \mathbb{P}(A_{x,y}) \implies F(y) = F(x) + \mathbb{P}(A_{x,y}) \geq F(x)$$

□

Theorem 13.2 — If F is a non-decreasing, right continuous function satisfying

$$\lim_{x\to-\infty} F(x) = 0 \text{ and } \lim_{x\to\infty} F(x) = 1,$$

then there exists on some probability space a random variable X for which $F(x) = P(X \leq x)$

Proof. First proof

- F is nondecreasing + right-continuous $\implies$ measure μ on $\mathbb{R}, \mathcal{B}(\mathbb{R})$ where $\mu(a, b] = F(b) - F(a)$
- define $\varphi(u) = \inf\{x : u \leq F(x)\}$ and show that its a R.V

□

Definition 13.3 (Weak Convergence) If F_n and F are distribution functions, F_n converges weakly to F if

$$\lim_n F_n(x) = F(x)$$

for each x which F if continuous

Proposition 13.4. Let $X_n \sim F_n$ and $X \sim F$, assume that $X_n \xrightarrow{p} X$. We claim that $F_n \Rightarrow F$.

Proof. $\forall \delta > 0\ \exists N(\delta)$ such that for all $n \geq N$, $\mathbb{P}(|X_n - X| > \delta) \leq \delta$.

$$\begin{aligned} F(a) = \mathbb{P}(X \leq a) &\leq \mathbb{P}\left(\{X_n \leq a + \delta\} \cup \{|X_n - X| > \delta\}\right) \\ &\leq \mathbb{P}(X_n \leq a + \delta) + \mathbb{P}(|X_n - X| > \delta) \\ &\leq F_n(a + \delta) + \delta \end{aligned}$$

Therefore, we conclude that

$$F_n \Rightarrow F$$

□

Lemma 13.5 — If $F_n \Rightarrow F$, $a_n \to a$, and $b_n \to b$, then $F_n(a_n x + b_n) \Rightarrow F(ax + b)$

Proof. □

§13.1 Properties of Integral

Theorem 13.6 (Fatou's Lemma) — For sequence of measurable functions or R.V.'s $\{X_n\} > 0$,

$$\int \liminf_n X_n d\mu \leq \liminf \int X_n d\mu \Leftrightarrow \mathbb{E}\left(\liminf_n X_n\right) \leq \liminf_n \mathbb{E}(X_n)$$

Theorem 13.7 (Dominated convergence theorem (DCT)) — If $|X_n| \leq Y$ almost everywhere, and Y is integrable, and $X_n \overset{a.s}{\to} X$, then (i) f and f_n are integrable and $\int X_n d\mu = \int X d\mu$

Note in DCT we are not assuming that X_n or X is integrable. Can a non-integrable be dominated by an integrable function?

The following is is proposition concerning the convergence of sequences of integrable functions.

Theorem 13.8 (Scheffé's lemma) — Let μ_n and μ have densities f_n and f, respectively, such that $\mu_n(\Omega) = \mu(\Omega) < \infty$. Let

$$\sup_{A \in \mathcal{F}} \mid \mu_n(A) - \mu(A) \mid \leq \int_\Omega |f - f_n \mid d\lambda \to 0$$

§13.2 Uniform Integrability

Uniform integrability is a property of a family of random variables which says that: (1) first absolute moments are uniformly bounded (2) distribution tails of the R.Vs in the family converge to 0 at a unifrom rate. If X is integrable, then $|X|\mathbb{I}_{\{|X_n| \geq a\}} \to 0$ almost everywhere as $a \to \infty$ and is dominated by $|X|$ and hence

$$\lim_{a \to \infty} \int_{\{|X_n| \geq a\}} |X| d\mu = 0$$

Definition 13.9 A sequence $\{X_n\}$ of random variables is said to be **uniformly integrable** if

$$\lim_{a \to \infty} \sup_n \int_{\{|X_n| \geq a\}} |X_n| d\mu = 0$$

or equivalently,

$$\sup_n \mathbb{E}(|X_n| \mathbb{I}_{\{|X_n| \geq a\}}) \to 0 \quad \text{as} \quad a \to \infty$$

Theorem 13.10 — Suppose that $\mu(\Omega) < \infty$ and $X_n \overset{a.s}{\to} X$, then

(i) If the f_n are uniformly integrable, then f is integrable and

$$\int f_n d\mu \to \int f d\mu$$

(ii) if f and f_n are uniformly integrable, it follows by

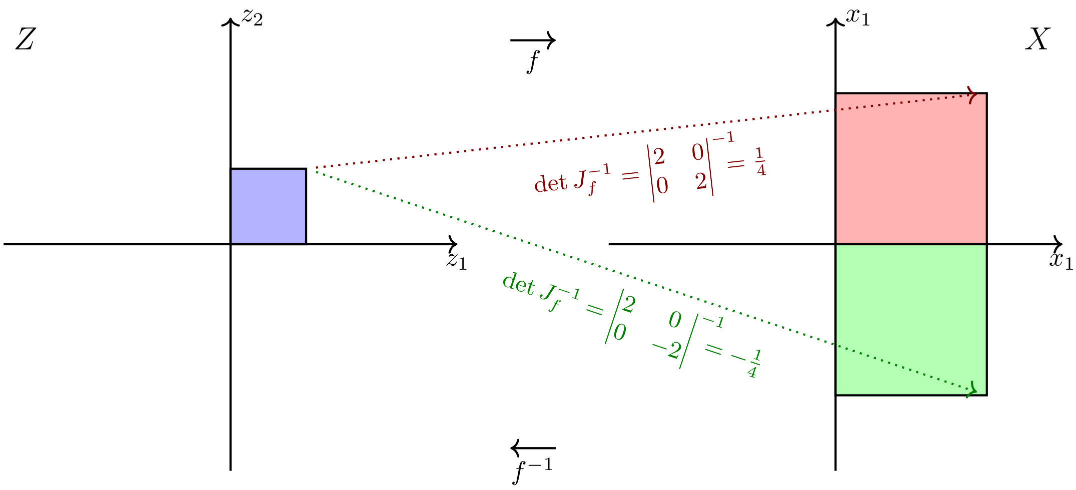
§13.3 Change of variable

Let $(\Omega^{(1)}, \mathcal{F}^{(1)})$ and $(\Omega^{(2)}, \mathcal{F}^{(2)})$ be mesurable spaces, and suppose that the mapping $T : \Omega^{(1)} \to \Omega^{(2)}$ is measurable $\mathcal{F}^{(1)}/\mathcal{F}^{(2)}$. For a measure μ on $\mathcal{F}^{(1)}$, define a measure

$$\mu T^{-1}(A) := \mu(T^{-1}(A)), \quad A \in \mathcal{F}^{(2)}$$

Theorem 13.11 — If $f \geq 0$, then

$$\int_{\Omega^{(1)}} f(T\omega)\mu(d\omega) = \int_{\Omega^{(2)}} f(\omega)\mu T^{-1} d\omega$$



§14 November 21, 2022

§14.1 Egorov's theorem

- Pointwise convergence does not imply uniform convergence
- Pointwise a.e convergence does not imply L^∞
- L^∞ is a function space, its elements are *essentially* bounded measurable functions [12]

Egorov's theorem: if we allow our self to reduce the domain of functions, we can find a subset on which we have uniform convergence [13].

Theorem 14.1 (Egorov's theorem) — *Let $(X, \mathcal{F}, \mu)$ be a finite measure space. Let $\{X_n\}$ be a sequence of R.V's.* $X_n \overset{a.s}{\to} X$, $\forall \delta > 0$, *there exists a measurable set* $A \subseteq X \equiv A(\delta)$ such that

1. $X_n \to X$ uniformly on A, i.e.,

$$\lim_{n\to\infty}\left(\sup_{a\in A}|X(a) - X_n(a)|\right) = 0$$

2. $\mu(A^c) < \delta$

Also $\forall A \in \mathcal{B}$, $\forall \delta > 0$, $\exists \{I_n\}_{n=1}^N$ disjoint such that $\lambda(A \triangle \cup_n I_n) < \delta$

Proof. $\forall f > 0$, $\exists g \leq f$ such that □

Proposition 14.2. Similar result to Egorov's theorem

1. $\forall f \in L^1$, $\forall \delta > 0$, $\exists g = \sum_{n=1}^N c_n \mathbb{I}_{I_n}$ such that $\int_\Omega |f - g| < \delta$

§14.2 Product measures

Let $(X, \mathcal{A}, \mu_1)$ and $(Y, \mathcal{B}, \mu_2)$ be two σ-finite measure spaces. Let $\Omega = X \times Y = \{(x, y) : x \in X, y \in Y\}$ and $\mathcal{S} = \{A \times B : A \in \mathcal{A}, B \in \mathcal{B}\}$. Sets in $\mathcal{S}$ are called rectangles. It is easy to see that $\mathcal{S}$ is a semi-algebra:

$$(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D),$$
$$(A \times B)^c = (A^c \times B) \cup (A \times B^c) \cup (A^c \times B^c).$$

Let $\mathcal{F} = \mathcal{A} \times \mathcal{B}$ be the σ-algebra generated by $\mathcal{S}$.

Theorem 1.7.1 There is a unique measure μ on $\mathcal{F}$ with $\mu(A \times B) = \mu_1(A)\mu_2(B)$. Notation μ is often denoted by $\mu_1 \times \mu_2$.

Proof By Theorem 1.1.9, it is enough to show that if $A \times B = \bigcup_i (A_i \times B_i)$ is a finite or countable disjoint union, then

$$\mu(A \times B) = \sum_i \mu_1(A_i)\mu_2(B_i).$$

For each $x \in A$, let $I(x) = \{i : x \in A_i\}$. $B = \bigcup_{i\in I(x)} B_i$, so

$$\mu_1(A)\mu_2(B) = \int_X 1_A(x)\mu_2(B)\,\mathrm{d}\mu_1(x).$$

[12] Essentially bound functions are bounded except on a set of zero measure

[13] Uniform convergence meas that there is an overall speed to the convergence. Pointwise convergence means at every point the sequence of functions has its own speed of convergence (can be fast at at some points and very very slow at other points)

Integrating with respect to μ_1 and using Exercise 1.5.6 gives

$$\mu(A \times B) = \sum_i \mu_1(A_i)\mu_2(B_i),$$

which proves the result.

Let $(X, \mathcal{F}, \mu)$ and $(Y\mathcal{G}, \nu)$ be measurable spaces. Our aim is to construct on the Cartesian Product $X \times Y$ a **product measure** π such that $\pi(A \times B) = \mu(A)\nu(B)$ for $A \subset X$ and $B \subset Y$. [14]

Theorem 14.3 — (i) If $E \in \mathcal{F} \times \mathcal{G}$, then for each x the set $\{y : (x, y) \in E\} \in \mathcal{G}$

(ii) If f is measurable $\mathcal{F} \times \mathcal{G}$, then for each fixed x the function $f(x, \cdot)$ is measurable $\mathcal{G}$ and $f(\cdot, y)$ is measurable $\mathcal{F}$

Let us define;

$$\pi_1(E) = \int_X \nu(\{y : (x, y) \in E\})\mu dx$$

$$\pi_2(E) = \int_Y \nu(\{x : (x, y) \in E\})\nu dy$$

Theorem 14.4 (Fubini's theorem) —

$$\int_{X \times Y} f(x, y)(\mu \times \nu)d(x, y) = \int_X \left[\int_Y f(x, y)d\nu(y)\right] d\mu(x)$$

for $f \in L^1$

Example 14.5. Let $X = (0, 1)$, $Y = (1, \infty)$, both equipped with the Borel sets and Lebesgue measure. Let $f(x, y) = e^{-xy} - 2e^{-2xy}$.

$$\int_0^1 \int_1^\infty f(x, y)dydx = \int_0^1 x^{-1}(e^{-x} - e^{-2x})dx > 0$$

$$\int_1^\infty \int_0^1 f(x, y)dxdy = \int_1^\infty y^{-1}(e^{-2y} - e^{-y})dy < 0$$

Example 14.6. Let $X = (0, 1)$ with $A =$ the Borel sets and $\mu_1 =$ Lebesgue measure. Let $Y = (0, 1)$ with $B =$ all subsets and $\mu_2 =$ counting measure. Let $f(x, y) = 1$ if $x = y$ and 0 otherwise.

$$\int f(x, y)\mu_2(dy) = 1 \text{ for all } x \text{ so } \int\int f(x, y)\mu_2(dy)\mu_1(dx) = 1$$

$$\int f(x, y)\mu_1(dx) = 0 \text{ for all } y \text{ so } \int\int f(x, y)\mu_1(dy)\mu_2(dx) = 0$$

Our last example shows that measurability is important or maybe that some of the axioms of set theory are not as innocent as they seem.

[14] In the case mu and ν are Lebesgue measure on the line, π will be Lebesgue measure on the plane.