MS-E1600
PROBABILITY THEORY

Kalle Kytölä
kalle.kytola@aalto.fi
office Y 241

Lectures Jan 5 – Feb 11, 2016
Tuesdays 10–12 in M3
Thursdays 14–16 in M3

Exercises Wednesdays 14–16 in M3
TA Armando Gutierrez

Exam Tuesday Feb 23 at 16:30 – 19:30

Grading 0 - 5 based on exam.
You may compensate one problem in the exam by solved exercises, with score depending on the number of solved exercises.

Note Wednesday, January 6 is a holiday, so the exercise session is moved to week Jan 11 – 15.

Textbook Jacod & Protter: "Probability Essentials"
or Williams: "Probability with Martingales"

INTRODUCTION

$(\Omega, \mathcal{F}, \mathbb{P})$

Probability theory forms the mathematically precise and powerful foundations for the study of randomness.

Its most basic objects are:

"Outcomes (of a random experiment)"

Sample space Ω consists of all possible outcomes $\omega \in \Omega$.

"Events"

An event E is a subset $E \subset \Omega$: we say the event occurs if the outcome ω belongs to E.

We cannot, however, allow all subsets of the sample space to be events. Rather, we form a suitable collection $\mathcal{F} \subset \mathcal{P}(\Omega)$ of subsets of Ω on which it is possible to have consistent rules of probability.

"Probability (measure)"

To each event E we assign its probability $\mathbb{P}[E] \in [0,1]$.

"Random variable"

A suitable function $X : \Omega \to S$ associating to each outcome $\omega \in \Omega$ a value $X(\omega) \in S$.

"Expected value"

$\mathbb{E}[X] = \int_\Omega X(\omega)\, d\mathbb{P}(\omega)$ integral in the sense of Lebesgue.

Probability theory $\overset{?}{=}$ measure theory

Yes ?

- The basic definitions and results are literally identical.

No ?

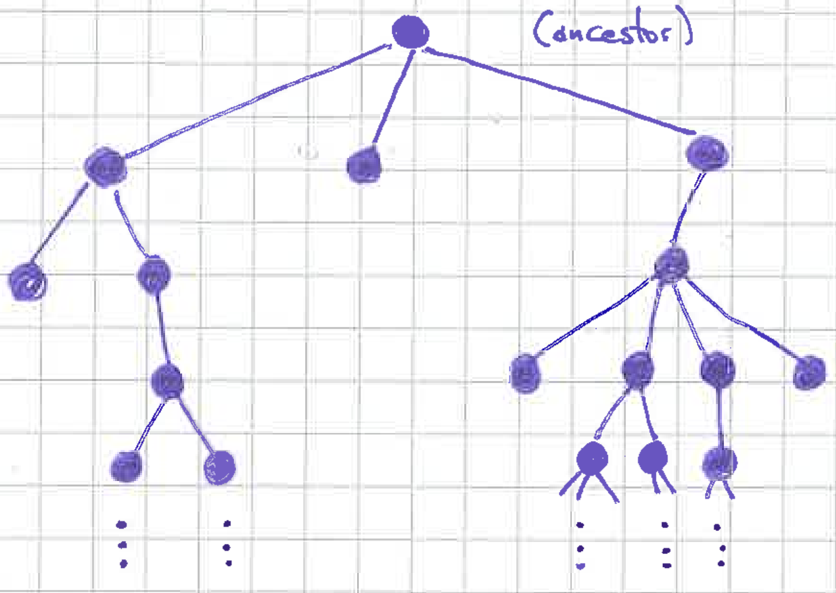
- Probability theory puts emphasis on different notions and provides interesting and useful interpretations.
- Probabilistic thinking offers great intuition.
- There are a number of profound results that are genuinely probabilistic (laws of large numbers, central limit theorem, zero-one laws, martingale convergence theorem, ...)

Although stochastics is concerned with very different questions from measure theory, it is <u>necessary</u> to have a good command of measure and integration in order to do any advanced / mathematically precise probability (or statistics).

INFORMAL MOTIVATING EXAMPLE 1 :

BRANCHING PROCESS

Suppose an ancestor has a random number of descendants, and each descendant has a random number of descendants, and so on and so forth.



The number X of direct descendants of the ancestor is a random variable. Denote, for $k \in \mathbb{Z}_{\geq 0}$,

$$p_k = \mathbb{P}[X = k]$$

and

$$\mu = \mathbb{E}[X] = \sum_{k=0}^{\infty} k \cdot p_k .$$

To define a simple genealogical model assume that all the descendants have random numbers of direct descendants, independently of each other and with the same distribution as X.

Q1 : How do we construct this model mathematically?

(A: A countable product measure)

We might care about whether the lineage of the ancestor ever becomes extinct.

If Z_n denotes the number of descendants in the n:th generation ($Z_0 = 1$, $Z_1 = X$, ...) then extinction happens if $Z_n = 0$ for some n.

Q2: How do we calculate the probability of extinction, and how do we even make sure it is meaningful to talk about the event of extinction?

(A: For the calculation one can use generating functions (characteristic functions), and meaningfulness is guaranteed by properties of σ-algebra.)

It turns out that

if $\mu < 1$ then $\mathbb{P}[\text{extinction}] = 1$,

if $\mu > 1$ then $\mathbb{P}[\text{extinction}] < 1$.

Rather obviously one has $\mathbb{E}[Z_n] = \mu^n$, and consequently for any n the random variable $M_n = Z_n / \mu^n$ has expected value $\mathbb{E}[M_n] = 1$. One can show (martingale convergence theorem) that there exists a limit random variable

$$M_\infty = \lim_{n\to\infty} M_n .$$

Q3: Do we have: $\mathbb{E}[\lim_{n\to\infty} M_n] \overset{?}{=} \lim_{n\to\infty} \underbrace{\mathbb{E}[M_n]}_{=1}$

(A: This is somewhat subtle, it depends. We will learn under which conditions one can interchange the order of limits and expectations / integrals)

EXAMPLE 2

REPEATED COIN TOSSING

The possible outcomes of a single coin toss are "heads" or "tails" abbreviated H and T, respectively.

The possible outcomes of an infinite number of coin tosses are sequences of "heads" and "tails", i.e. functions $\mathbb{N} \to \{H,T\}$.

sample space: $\Omega = \{H,T\}^{\mathbb{N}}$ (uncountable)

Let $X_n = \frac{1}{n} \cdot (\#\text{ heads in first } n \text{ tosses})$ be the frequency of heads in the first n coin tosses.

Compare the following two statements of law of large numbers:

(i) For any $\varepsilon > 0$: $\lim_{n\to\infty} \mathbb{P}[\,|X_n - \frac{1}{2}| > \varepsilon\,] = 0$

(ii) $\mathbb{P}[\lim_{n\to\infty} X_n = \frac{1}{2}] = 1.$

Q: Can we make sense of such statements?
Does one imply the other?
Are the statements true (if $\mathbb{P}$ is the usual probability for fair coin tossing)?

STRUCTURE OF EVENT SPACES

Interpretation of set operations for events

subset	
Ω	sure event (contains all possible outcomes)
$\emptyset$ (empty set)	impossible event (contains no outcomes)
$E_1 \cap E_2$	both E_1 and E_2 occur
$E_1 \cup E_2$	E_1 or E_2 occurs
E^c	E doesn't occur

The collection $\mathcal{F}$ of events should satisfy

- $\emptyset, \Omega \in \mathcal{F}$
- if $E_1, E_2 \in \mathcal{F}$ then $E_1 \cap E_2 \in \mathcal{F}$, $E_1 \cup E_2 \in \mathcal{F}$
- if $E \in \mathcal{F}$ then $E^c = \Omega \setminus E \in \mathcal{F}$

In fact, for a meaningful mathematical theory, we need to be able to form countably infinite intersections and unions.

Def: A collection $\mathcal{F} \subset \mathcal{P}(\Omega)$ of subsets of a set Ω is a *σ-algebra* ("sigma-algebra") on Ω if

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

(2) $E \in \mathcal{F} \Rightarrow E^c \in \mathcal{F}$

(3) $E_1, E_2, \ldots \in \mathcal{F} \Rightarrow \bigcup_{n=1}^{\infty} E_n \in \mathcal{F}$.

Remark (2) and (1) imply: (1') $\emptyset \in \mathcal{F}$

(2) and (3) imply: (3') $E_1, E_2, \ldots \in \mathcal{F} \Rightarrow \bigcap_{n=1}^{\infty} E_n \in \mathcal{F}$.

Examples

(i) $\mathcal{F} = \{\phi, \Omega\}$ is a σ-algebra

(ii) $\mathcal{F} = \mathcal{P}(\Omega)$ is a σ-algebra

(iii) $\mathcal{F} = \mathcal{T}(\mathbb{R}) = \{$ open subsets of $\mathbb{R}\}$ is not a σ-algebra

Example 1 (branching process) Extinction?

Let $E_n = \{Z_n = 0\} = \{$extinction happens by the n:th generation$\}$.

Then $E_1 \subset E_2 \subset E_3 \subset \cdots$ is an increasing sequence of events, $E_n \uparrow E$, where

$$E = \bigcup_{n=1}^{\infty} E_n = \{\text{there exists an } n \text{ such that } Z_n = 0\}$$

is the event of eventual extinction.

Example 2 (coin tossing) Long term frequency of heads?

Let $X_n = \frac{1}{n} \cdot (\#\text{ heads in } n \text{ first tosses})$ be the frequency of heads in n tosses.

We can write equivalent conditions:

$$\lim_{n\to\infty} X_n = \frac{1}{2}$$

$$\iff \forall \varepsilon > 0 \quad \exists k \in \mathbb{Z}_{>0} \text{ s.t. } \forall n \geq k : |X_n - \tfrac{1}{2}| < \varepsilon$$

$$\iff \forall m \in \mathbb{Z}_{>0} \quad \exists k \in \mathbb{Z}_{>0} \text{ s.t. } \forall n \geq k : \tfrac{1}{2} - \tfrac{1}{m} < X_n < \tfrac{1}{2} + \tfrac{1}{m}$$

Therefore, denoting by $E_{m,n}$ the event that $\frac{1}{2} - \frac{1}{m} < X_n < \frac{1}{2} + \frac{1}{m}$, we have that

$$\bigcap_{m \in \mathbb{Z}_{>0}} \bigcup_{k \in \mathbb{Z}_{>0}} \bigcap_{n \geq k} E_{m,n}$$

is the event that the long term frequency of heads is one half.

Lemma If $\mathcal{F}_\alpha$, $\alpha \in I$, are σ-algebras on Ω, then also $\mathcal{F} = \bigcap_{\alpha \in I} \mathcal{F}_\alpha$ is a σ-algebra on Ω.

Proof Note that $E \in \mathcal{F}$ iff $E \in \mathcal{F}_\alpha$ for all $\alpha \in I$.
Since $\phi, \Omega \in \mathcal{F}_\alpha$ for all α, we also have $\phi, \Omega \in \mathcal{F}$, so condition (1) is ok.
Suppose $E \in \mathcal{F}$. Then for all α $E \in \mathcal{F}_\alpha$ so by (2) also $E^c \in \mathcal{F}_\alpha$. This shows that $E^c \in \mathcal{F}$, so condition (2) is ok for $\mathcal{F}$.
Suppose $E_1, E_2, \dots \in \mathcal{F}$. Then for all α $E_1, \dots \in \mathcal{F}_\alpha$, so by (3) also $\bigcup_{n=1}^{\infty} E_n \in \mathcal{F}_\alpha$.
This shows $\bigcup_{n=1}^{\infty} E_n \in \mathcal{F}$, so (3) is ok. $\square$

The reason we care about the previous lemma is that it allows us to "generate" σ-algebras by simpler collections.

Def: Let $\mathcal{C} \subset \mathcal{P}(\Omega)$ be a collection of subsets of Ω. Then we define $\sigma(\mathcal{C}) \subset \mathcal{P}(\Omega)$ as the smallest σ-algebra on Ω which contains the collection $\mathcal{C}$.
This is called the σ-algebra generated by $\mathcal{C}$.

Remark: The smallest σ-algebra containing $\mathcal{C}$ makes sense (it exists and is unique) by the previous lemma: it is the intersection of all σ-algebras which contain $\mathcal{C}$.

It is often difficult to describe explicitly all elements of even very common and reasonable sigma-algebras. Working with generating collections is typically much easier!

<u>Def</u> The Borel σ-algebra $\mathcal{B}$ on $\mathbb{R}$ is the σ-algebra generated by open sets in $\mathbb{R}$. Similarly, the Borel σ-algebra $\mathcal{B}(\mathcal{X})$ of a topological space $\mathcal{X}$ (e.g. $\mathbb{R}^d$, $C[0,1]$, ...) is the σ-algebra gen. by open sets in $\mathcal{X}$.

$\mathcal{B}$ is the most important σ-algebra in all of probability theory!

<u>Proposition</u> The Borel σ-algebra $\mathcal{B}$ on $\mathbb{R}$ is generated by any of the following collections of subsets of $\mathbb{R}$

(i) $\mathcal{C} = \{(-\infty, x] \mid x \in \mathbb{R}\}$

(ii) $\mathcal{C} = \{(x,y) \mid x,y \in \mathbb{R} \quad x<y\}$

(iii) $\mathcal{C} = \{[x,y] \mid x,y \in \mathbb{R} \quad x \leq y\}$

(iv) $\mathcal{C} = \{(x,y] \mid x,y \in \mathbb{R} \quad x \leq y\}$

(etc. etc.)

<u>Proof</u>: We will check (i), others are similar. To show $\sigma(\mathcal{C}) = \mathcal{B}$ we will check separately $\sigma(\mathcal{C}) \subset \mathcal{B}$ and $\sigma(\mathcal{C}) \supset \mathcal{B}$.

"$\subset$": Since $((-\infty, x])^c = \mathbb{R} \setminus (-\infty, x) = (x, +\infty)$ is open, we have $(x, \infty) \in \mathcal{B}$ and thus $(-\infty, x] \in \mathcal{B}$ by (2). Therefore $\mathcal{C} \subset \mathcal{B}$. Since $\sigma(\mathcal{C})$ is the smallest σ-alg. with $\mathcal{C} \subset \sigma(\mathcal{C})$ this implies $\sigma(\mathcal{C}) \subset \mathcal{B}$ as desired.

"$\supset$": To show $\sigma(\mathcal{E}) \supset \mathcal{B}$ it is sufficient to show that $\sigma(\mathcal{E})$ contains all open sets $V \subset \mathbb{R}$ (because $\mathcal{B}$ is the smallest such σ-alg.). Let us show step by step that $\sigma(\mathcal{E})$ contains

(a) intervals $(x,y]$

(b) open intervals (x,z)

(c) open sets $V \subset \mathbb{R}$.

To show (a), note that

$$(x,y] = (-\infty,y] \setminus (-\infty,x]$$
$$= (-\infty,y] \cap \left((-\infty,x]\right)^c$$

so $(x,y]$ is obtained from members of $\mathcal{E}$ by countable (in fact finite) set operations, and thus $(x,y] \in \sigma(\mathcal{E})$.

To show (b), note that

$$(x,z) = \bigcup_{n=1}^{\infty} (x, z-\tfrac{1}{n}]$$

so (x,z) is obtained from intervals $(x,y] \in \sigma(\mathcal{E})$ by countable unions, so $(x,z) \in \sigma(\mathcal{E})$.

Finally, note that any open set $V \subset \mathbb{R}$ is a countable union of open intervals, so V is obtained from $(x,z) \in \sigma(\mathcal{E})$ by countable unions, and thus $V \in \sigma(\mathcal{E})$.

This concludes the proof.

Remark: We used the fact that an open $V \subset \mathbb{R}$ is a countable union of open intervals. Let us remind the argument for this. For any $r \in V$ the open set V contains some ball $(r-\varepsilon_r, r+\varepsilon_r) \subset V$ around r. Thus V contains also an open interval around r with rational endpoints $p_r, q_r \in \mathbb{Q}$, i.e. $r \in (p_r, q_r) \subset V$. Clearly then

$$V = \bigcup_{r \in V} (p_r, q_r).$$

But there are only countably many intervals with rational endpoints, so a countable union suffices.

REMINDERS ON SET THEORY

Let J be some index set and for all $j \in J$ let $A_j \subset \Omega$ be subsets of a set Ω

union: $\bigcup_{j \in J} A_j = \{\omega \in \Omega \mid \omega \in A_j \text{ for some } j \in J\}$
$= \{\omega \in \Omega \mid \exists j \in J: \ \omega \in A_j\}$

intersection: $\bigcap_{j \in J} A_j = \{\omega \in \Omega \mid \omega \in A_j \text{ for all } j \in J\}$
$= \{\omega \in \Omega \mid \forall j \in J: \ \omega \in A_j\}$

complement: $A_j^c = \Omega \setminus A_j = \{\omega \in \Omega \mid \omega \notin A_j\}$

De Morgan's laws:

$$\Big(\bigcup_{j \in J} A_j\Big)^c = \bigcap_{j \in J} A_j^c \quad , \quad \Big(\bigcap_{j \in J} A_j\Big)^c = \bigcup_{j \in J} A_j^c$$

Exercise: Verify De Morgan's laws.

Let $f: \Omega \to S$ be a function and $B \subset S$.

preimage: $f^{-1}(B) = \{\omega \in \Omega \mid f(\omega) \in B\} \subset \Omega$

facts: if $B_j \subset S$ for all $j \in J$, then

$$f^{-1}\Big(\bigcup_{j \in J} B_j\Big) = \bigcup_{j \in J} f^{-1}(B_j) \quad , \quad f^{-1}\Big(\bigcap_{j \in J} B_j\Big) = \bigcap_{j \in J} f^{-1}(B_j)$$

Exercise: Verify the above facts.

disjoint: sets A_j, $j \in J$, are disjoint if
for any $j, k \in J$, $j \neq k$, we have $A_j \cap A_k = \emptyset$.

Some notations for familiar sets

- $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$ = the set of integers
- $\mathbb{N} = \{1, 2, 3, \dots\}$ = the set of natural numbers
- $\mathbb{Z}_{\geq 0} = \{0, 1, 2, \dots\}$ = the set of non-negative integers
- $\mathbb{R}$ = the set of real numbers
- $\mathbb{Q}$ = the set of rational numbers
- $\mathbb{C}$ = the set of complex numbers
- $\mathbb{R}^d = \underbrace{\mathbb{R} \times \cdots \times \mathbb{R}}_{d \text{ times}}$ = the d-dimensional Euclidean vector space
- etc.

<u>countability</u> A set S is countable if there exists an injective mapping $S \to \mathbb{N}$. A countable set can have a finite or infinite number of elements, and is called finite or countably infinite, respectively. A set that is not countable is called uncountable.

Examples :

- the sets $\mathbb{Z}, \mathbb{N}, \mathbb{Q}$ are countable
- the sets $\mathbb{R}, \mathbb{C}, \mathbb{R}^d, [0,1], \{H,T\}^{\mathbb{N}}$ are uncountable

<u>set of functions</u> : if X and Y are sets, we denote by Y^X the set of all functions $f: X \to Y$

<u>power set</u> : $\mathcal{P}(\Omega) = \{A \subset \Omega\}$ = the set of all subsets of Ω

note: $\mathcal{P}(\Omega)$ can be identified with $\{0,1\}^{\Omega}$.

Let $A_1, A_2, A_3, \ldots \subset \Omega$ be a sequence of subsets of Ω.

increasing We say that the sequence is increasing if $A_1 \subset A_2 \subset A_3 \subset \cdots$. We then denote $A_n \uparrow A$, where $A = \bigcup_{n=1}^{\infty} A_n$.

decreasing We say that the sequence is decreasing if $A_1 \supset A_2 \supset A_3 \supset \cdots$. We then denote $A_n \downarrow A$, where $A = \bigcap_{n=1}^{\infty} A_n$.

More generally, we define

$$\limsup A_n = \bigcap_{m=1}^{\infty} \bigcup_{k \geq m} A_k$$
$$= \{ \omega \in \Omega \mid \forall m \ \exists k \geq m : \ \omega \in A_k \}$$
$$= \{ \omega \in \Omega \mid \omega \in A_k \text{ for infinitely many } k \}$$

$$\liminf A_n = \bigcup_{m=1}^{\infty} \bigcap_{k \geq m} A_k$$
$$= \{ \omega \in \Omega \mid \exists m \ \forall k \geq m : \ \omega \in A_k \}$$
$$= \{ \omega \in \Omega \mid \omega \in A_k \text{ for all large enough } k \}$$

Exercise: Make sure you understand why the different descriptions above are equivalent.

MEASURES AND PROBABILITY MEASURES

Recall that the basic objects of probability theory are

Ω : the set of possible outcomes
$\mathcal{F}$: the collection of events
$\mathbb{P}$: probability

We can allow Ω to be any set.
A meaningful theory of probability requires that the collection of events is stable under countable set operations, i.e. $\mathcal{F}$ is a σ-algebra on Ω.
Already this structure is important enough to merit a definition:

Def: A measurable space is a pair (S, Σ) where S is a set and Σ is a σ-algebra on S.

Examples

- $(\mathbb{R}, \mathcal{B})$: the real numbers with Borel σ-algebra
- $(S, \mathcal{P}(S))$: any set S with the collection $\mathcal{P}(S)$ of all of its subsets
- $(\Omega, \mathcal{F})$: any sample space with a collection of events on it.

The topic of this lecture is the requirements we will impose on the last basic object, the probability measure $\mathbb{P}$. We will give definitions for general measures as well, and occasionally specialize to probability measures.

We will also be concerned with questions of how to construct a measure and how to identify a measure. For these two purposes we give two more definitions, where requirements are more relaxed than σ-algebras.

Def: An algebra Σ_0 on a set S is a collection of subsets of S such that

(i) $S \in \Sigma_0$

(ii) $A \in \Sigma_0 \Rightarrow A^c = S \setminus A \in \Sigma_0$

(iii) $A_1, A_2 \in \Sigma_0 \Rightarrow A_1 \cup A_2 \in \Sigma_0$

Remark: It follows that an algebra is stable under finite unions, using (iii) and induction, and under finite intersections, using also (ii).

Def: A π-system $\mathcal{I}$ on a set S is a collection of subsets of S such that if $I_1, I_2 \in \mathcal{I}$ then also $I_1 \cap I_2 \in \mathcal{I}$

Remark: By induction it follows that a π-system is stable under finite intersections.

Remark: Any σ-algebra is an algebra, and any algebra is a π-system. The converse implications are false.

Example: The collection $\mathcal{I} = \{(-\infty, x] \mid x \in \mathbb{R}\}$ of all closed semi-infinite upper bounded intervals in $\mathbb{R}$ is a π-system on $\mathbb{R}$. Recall that the smallest σ-algebra containing $\mathcal{I}$ is the Borel σ-algebra $\mathcal{B}$, i.e. $\sigma(\mathcal{I}) = \mathcal{B}$.

Def: A measure μ on a measurable space (S, Σ) is a function $\mu : \Sigma \longrightarrow [0, +\infty]$ such that

(M1) $\mu[\emptyset] = 0$

(M2) if $A_1, A_2, \ldots \in \Sigma$ are disjoint then $\mu\left[\bigcup_{n=1}^{\infty} A_n\right] = \sum_{n=1}^{\infty} \mu[A_n]$.

Def: A probability measure on $(\Omega, \mathcal{F})$ is a measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ such that $\mathbb{P}[\Omega] = 1$.

__Remark:__ Note that for a general measure μ, some sets can have infinite measure.

__Examples:__

- Counting measure: If S is any set and $\Sigma = \mathcal{P}(S)$ the collection of all its subsets, then $\mu[A] = \# A$ (number of elements of $A \subset S$) defines a measure on $(S, \mathcal{P}(S))$. Note that subsets with infinitely many elements have infinite measure.
- Discrete uniform measure: If Ω is a non-empty finite set and $\mathcal{F} = \mathcal{P}(\Omega)$, then $\mathbb{P}[A] = \frac{\# A}{\# \Omega}$ defines a probability measure on $(\Omega, \mathcal{F})$.
- Lebesgue measure on the real line: The natural notion of "length" on $\mathbb{R}$ corresponds to the Lebesgue measure λ on $(\mathbb{R}, \mathcal{B})$. We have $\lambda[[a,b]] = b-a$, and e.g. $\lambda[\mathbb{R}] = +\infty$. We will later comment on how to construct λ.

__Terminology:__ If μ is a measure on (S, Σ) then we call (S, Σ, μ) a measure space. If $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F})$ then we call $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space.

In particular any probability space is a measure space, and all results for measure spaces can be used for probability spaces.

Probability distributions on countable spaces

Let Ω be a countable set, and $\mathcal{F} = \mathcal{P}(\Omega)$ the collection of all subsets of Ω.

Def: A probability mass function (p.m.f.) on Ω is a function $p : \Omega \to [0,1]$ such that $\sum_{\omega \in \Omega} p(\omega) = 1$.

Fact: If p is a p.m.f. on Ω, then the formula $\mathbb{P}[E] = \sum_{\omega \in E} p(\omega)$ defines a probability measure.

Proof Exercise. $\square$

Fact If $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F})$, then the formula $p(\omega) = \mathbb{P}[\{\omega\}]$ defines a p.m.f.

Proof Exercise. $\square$

Remark If $p : \Omega \to [0, +\infty)$ satisfies $\sum_{\omega \in \Omega} p(\omega) = 1$ then automatically $p(\omega) \leq 1$ for all $\omega \in \Omega$ and p is a p.m.f.

Example (Poisson distribution)

Let $\lambda > 0$ be a parameter.

For $k \in \mathbb{Z}_{\geq 0}$ set $p(k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$.

Then $p(k) \geq 0$ for all $k \in \mathbb{Z}_{\geq 0}$ and $\sum_{k=0}^{\infty} p(k) = 1$, so p is a p.m.f. on $\mathbb{Z}_{\geq 0}$.

The corresponding probability measure is called the Poisson distribution with parameter λ.

Let's see some simple consequences of the definition of measure:

<u>Lemma</u> Let (S, Σ, μ) be a measure space. Then we have:

(a) if $A, B \in \Sigma$ and $A \subset B$ then $\mu[A] \leq \mu[B]$

(b) if $A_1, A_2 \in \Sigma$ then $\mu[A_1 \cup A_2] \leq \mu[A_1] + \mu[A_2]$

(c) if $A_1, A_2, \dots, A_n \in \Sigma$ then $\mu\left[\bigcup_{j=1}^{n} A_j\right] \leq \sum_{j=1}^{n} \mu[A_j]$.

<u>Proof</u>: If $A \subset B$ then $B = A \cup (B \setminus (A \cap B))$ is a disjoint union, so $\mu[B] = \mu[A] + \mu[B \setminus (A \cap B)]$ so (a) follows because $\mu[B \setminus (A \cap B)] \geq 0$.

Write $A_1 \cup A_2 = A_1 \cup (A_2 \setminus (A_1 \cap A_2))$, a disjoint union. Use also $A_2 \setminus (A_1 \cap A_2) \subset A_2$ and (a) to get $\mu[A_1 \cup A_2] = \mu[A_1] + \mu[A_2 \setminus (A_1 \cap A_2)] \leq \mu[A_1] + \mu[A_2]$. This shows (b). For (c), use (b) and induction. $\square$

We also have the following monotone convergence of measures.

<u>Lemma</u> If $A_n \in \Sigma$, and $A_n \uparrow A$ then $\mu[A_n] \uparrow \mu[A]$.

<u>Proof</u>: Write $B_1 = A_1$, $B_2 = A_2 \setminus A_1$, ..., $B_n = A_n \setminus A_{n-1}$, ...
Then the sets $B_1, B_2, \dots$ are disjoint and $A_n = B_1 \cup B_2 \cup \dots \cup B_n$ for any $n \in \mathbb{N}$.
We get $\mu[A_n] = \sum_{k=1}^{n} \mu[B_k]$.
This shows that $\mu[A_n] \uparrow \sum_{k=1}^{\infty} \mu[B_k]$.
But by disjointness,

$$\sum_{k=1}^{\infty} \mu[B_k] = \mu\left[\bigcup_{k=1}^{\infty} B_k\right] = \mu\left[\bigcup_{k=1}^{\infty} A_k\right] = \mu[A]. \quad \square$$

For probability measures, or finite measures more generally, some things are nicer.

Lemma Let (S, Σ, μ) be a measure space with $\mu[S] < \infty$. Then we have

(a) if $A_1, A_2 \in \Sigma$ then
$$\mu[A_1 \cup A_2] = \mu[A_1] + \mu[A_2] - \mu[A_1 \cap A_2]$$

(b) if $A_n \in \Sigma$ and $A_n \downarrow A$ then $\mu[A_n] \downarrow \mu[A]$.

Question: What could go wrong with (a) if we did not assume $\mu[S] < \infty$?

Example Consider $\mathbb{N}$ with the counting measure $\mu[A] = \# A$ for $A \subset \mathbb{N}$.

Set $A_n = \{n, n+1, n+2, \dots\}$ for all $n = 1, 2, 3, \dots$

Then $\mu[A_n] = +\infty$ for all n.

But $A_1 \supset A_2 \supset A_3 \supset \dots$ and $\bigcap_{n=1}^{\infty} A_n = \phi$

so $A_n \downarrow \phi$. Now
$$\mu[\phi] = 0 \neq \lim_{n \to \infty} \underbrace{\mu[A_n]}_{= +\infty} = +\infty .$$

So (b) above does not hold in this case.

Proof of Lemma: (b) is a good exercise.

(a) Write $A_1 \cup A_2 = \big(A_1 \setminus (A_1 \cap A_2)\big) \cup (A_1 \cap A_2) \cup \big(A_2 \setminus (A_1 \cap A_2)\big)$ a disjoint union, to get
$$\mu[A_1 \cup A_2] = \mu[A_1 \setminus (A_1 \cap A_2)] + \mu[A_1 \cap A_2] + \mu[A_2 \setminus (A_1 \cap A_2)]$$
Recall $\mu[A_1] = \mu[A_1 \setminus (A_1 \cap A_2)] + \mu[A_1 \cap A_2]$.

Since everything is finite, we get
$$\mu[A_1 \setminus (A_1 \cap A_2)] = \mu[A_1] - \mu[A_1 \cap A_2].$$
Do the same for A_2 and simplify $\mu[A_1 \cup A_2]$. $\square$

Identification and construction of probability measures

Since σ-algebras are often complicated, we use instead the simpler π-systems for identification and algebras for construction of measures.

Theorem (Dynkin's identification theorem)

Let $\mathcal{I}$ be a π-system on Ω, and $\mathcal{F} = \sigma(\mathcal{I})$ the σ-algebra generated by $\mathcal{I}$. If $\mathbb{P}_1$ and $\mathbb{P}_2$ are two probability measures on $(\Omega, \mathcal{F})$ such that $\mathbb{P}_1[E] = \mathbb{P}_2[E]$ for all $E \in \mathcal{I}$, then we have $\mathbb{P}_1 = \mathbb{P}_2$.

Theorem (Carathéodory's extension theorem)

Let Σ_0 be an algebra on Ω, and $\mathcal{F} = \sigma(\Sigma_0)$ the σ-alg. gen. by Σ_0. Suppose that $\mathbb{P}_0 : \Sigma_0 \to [0,1]$ is such that $\mathbb{P}_0[\phi] = 0$, $\mathbb{P}_0[\Omega] = 1$, and for any $A_1, A_2, \ldots \in \Sigma_0$ disjoint such that $\bigcup_{n=1}^{\infty} A_n \in \Sigma_0$, we have $\mathbb{P}_0[\bigcup_{n=1}^{\infty} A_n] = \sum_{n=1}^{\infty} \mathbb{P}_0[A_n]$. Then there exists a unique probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ such that $\mathbb{P}[A] = \mathbb{P}_0[A]$ for all $A \in \Sigma_0$.

The proof of Carathéodory's extension theorem is slightly too long, so we omit it.

The proof of Dynkin's identification theorem will be given later if there is enough time.

Cumulative distribution functions

Consider probability measures on $(\mathbb{R}, \mathcal{B})$.
Recall that the Borel σ-algebra $\mathcal{B}$ is generated by the collection

$$\mathcal{J} = \{ (-\infty, x] \mid x \in \mathbb{R} \}, \quad \text{i.e.} \quad \mathcal{B} = \sigma(\mathcal{J}),$$

and $\mathcal{J}$ is a π-system.

Def: The cumulative distribution function (c.d.f.) of a probability measure μ on $(\mathbb{R}, \mathcal{B})$ is the function $F : \mathbb{R} \to [0,1]$ given by

$$F(x) = \mu\left[(-\infty, x]\right].$$

Proposition μ is uniquely determined by its c.d.f. F.

Proof Suppose that two measures μ_1, μ_2 on $(\mathbb{R}, \mathcal{B})$ have the same c.d.f. F, i.e.

$$\mu_1\left[(-\infty, x]\right] = F(x) = \mu_2\left[(-\infty, x]\right] \quad \forall x \in \mathbb{R}.$$

This means that μ_1 and μ_2 coincide on the π-system $\mathcal{J} = \{(-\infty, x] \mid x \in \mathbb{R}\}$. By Dynkin's identification thm we have $\mu_1 = \mu_2$ on the whole σ-algebra generated by $\mathcal{J}$, which is $\sigma(\mathcal{J}) = \mathcal{B}$. $\square$

Proposition For a c.d.f. F of a proba meas. μ we have

(a) F is increasing: $F(x) \le F(y)$ $\forall x \le y$

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

(c) if $x_n \downarrow x$ then $F(x_n) \downarrow F(x)$

Proof: (a): If $x \le y$ then $(-\infty, x] \subset (-\infty, y]$

so $F(x) = \mu[(-\infty,x]] \le \mu[(-\infty,y]] = F(y)$.

(b) Suppose $x_n \uparrow +\infty$. Then $(-\infty, x_n] \uparrow \mathbb{R}$.

Therefore $F(x_n) = \mu[(-\infty,x_n]] \uparrow \mu[\mathbb{R}] = 1$.

The case $x_n \downarrow -\infty$ is similar, $(-\infty, x_n] \downarrow \phi$.

(c) Suppose $x_n \downarrow x \in \mathbb{R}$. Then $(-\infty, x_n] \downarrow (-\infty, x]$.

Therefore $F(x_n) = \mu[(-\infty,x_n]] \downarrow \mu[(-\infty,x]] = F(x)$. $\square$

Think: Why don't we necessarily have for $x_n \uparrow x$ that $F(x_n) \uparrow F(x)$?

REMINDERS ABOUT SEQUENCES, SERIES, SUPREMUM AND INFIMUM

For $A \subset \mathbb{R}$, the supremum, $\sup A$, is the smallest number $M \in [-\infty, +\infty]$ such that $x \leq M$ for all $x \in A$. We have $\sup A \in (-\infty, +\infty]$ if $A \neq \emptyset$ and $\sup \emptyset = -\infty$.

Similarly, the infimum, $\inf A$, is the greatest number $m \in [-\infty, \infty]$ such that $x \geq m$ for all $x \in A$.

A sequence $x_1, x_2, x_3, \ldots$ of real numbers is increasing if $x_1 \leq x_2 \leq x_3 \leq \cdots$. Increasing sequences have limits (possibly $+\infty$):

$$\lim_{n\to\infty} x_n = x \quad \text{where} \quad x = \sup\{x_1, x_2, \ldots\}.$$

We then denote $x_n \uparrow x$.

Similarly, the sequence is decreasing if $x_1 \geq x_2 \geq x_3 \geq \cdots$. Decreasing sequences have limits (possibly $-\infty$):

$$\lim_{n\to\infty} x_n = x \quad \text{where} \quad x = \inf\{x_1, x_2, \ldots\}.$$

We then denote $x_n \downarrow x$.

More generally, for a sequence $x_1, x_2, \ldots \in \mathbb{R}$ we define

$$\limsup_{n\to\infty} x_n = \inf_{n\in\mathbb{N}} \left(\sup_{k\geq n} x_k \right)$$

Here we denote, for convenience

$$\sup_{k\geq n} x_k := \sup\{x_k \mid k \geq n\} = \sup\{x_n, x_{n+1}, x_{n+2}, \ldots\}$$

and similarly

$$\inf_n \left(\sup_{k\geq n} x_k \right) = \inf \left\{ \sup_{k\geq n} x_k \;\middle|\; n \in \mathbb{N} \right\}.$$

Note that if $\sup_n x_n < +\infty$ then the sequence $y_n = \sup_{k \geq n} x_k$ is finite and decreasing, so an equivalent definition is

$$\limsup_{n \to \infty} x_n = \lim_{n \to \infty} \left(\sup_{k \geq n} x_k \right).$$

Similarly, we define

$$\liminf_{n \to \infty} x_n = \sup_n \left(\inf_{k \geq n} x_k \right),$$

which can be written equivalently

$$\liminf_{n \to \infty} x_n = \lim_{n \to \infty} \left(\inf_{k \geq n} x_k \right)$$

if $\inf x_n > -\infty$ (limit of increasing sequence).

Fact: For any sequence $x_1, x_2, \dots \in \mathbb{R}$ we have $\liminf_{n \to \infty} x_n \leq \limsup_{n \to \infty} x_n$.

Exercise: Give a proof of this fact.

A very frequently useful observation is:

Fact: A sequence $x_1, x_2, \dots \in \mathbb{R}$ converges to a limit $x \in \mathbb{R}$ if and only if $\liminf_{n \to \infty} x_n = x = \limsup_{n \to \infty} x_n$.

Exercise: Give a proof of this fact.

Let S be any set, and for all $s \in S$ $x_s \in [0, +\infty]$ a non-negative number, possibly infinite. Then we may define

$$\sum_{s \in S} x_s = \sup \left\{ x_{s_1} + x_{s_2} + \dots + x_{s_n} \;\middle|\; \begin{array}{l} s_1, \dots, s_n \in S \\ \text{are distinct} \end{array} \right\}$$

(the series with non-neg. terms is defined as the supremum of all its finite subseries).

We always use the following conventions for calculations with infinities:

- $x + (+\infty) = +\infty$ for any $x \in (-\infty, +\infty]$
- $x + (-\infty) = -\infty$ for any $x \in [-\infty, +\infty)$
- $x \cdot (\pm\infty) = \pm\infty$ for all $x \in (0, \infty)$
- $x \cdot (\pm\infty) = \mp\infty$ for all $x \in (-\infty, 0)$
- $0 \cdot x = 0$ for all $x \in [-\infty, \infty]$.

Exercise: Check that if $x_1, x_2, \dots \geq 0$ then the above definition coincides with the usual one, i.e.

$$\sum_{k \in \mathbb{N}} x_k = \lim_{n \to \infty} \sum_{k=1}^{n} x_k .$$

Exercise*: Check that if $\sum_{s \in S} x_s < +\infty$ then there are only countably many $s \in S$ such that $x_s > 0$.

(*Hint: For each $m \in \mathbb{N}$ show that the set $A_m = \{ s \in S \mid x_s > \frac{1}{m} \}$ is finite, and take the union $\bigcup_{m=1}^{\infty} A_m$.)

RANDOM VARIABLES

$(\Omega, \mathcal{F}, \mathbb{P})$ probability space, i.e.

Ω = set of possible outcomes

$\mathcal{F}$ = collection of events

$\mathbb{P}$ = probability measure

Idea:

- "chance determines the random outcome $\omega \in \Omega$"
- "the random outcome $\omega \in \Omega$ determines various quantities of interest (random variables)"

Therefore, a random variable will be a function X defined on Ω, which to an outcome $\omega \in \Omega$ associates the value $X(\omega)$ of some quantity of interest.

The function $X : \Omega \to S'$ has to be reasonable, so that we can talk about probabilities with which the quantity of interest assumes certain values.

(Here S' is a set of possible values of the quantity of interest)

So whenever $A' \subset S'$ is a reasonable subset of possible values of X, the set of outcomes ω for which $X(\omega)$ belongs to A' should be an event, i.e.,

$$\{\omega \in \Omega \mid X(\omega) \in A'\} \in \mathcal{F}.$$

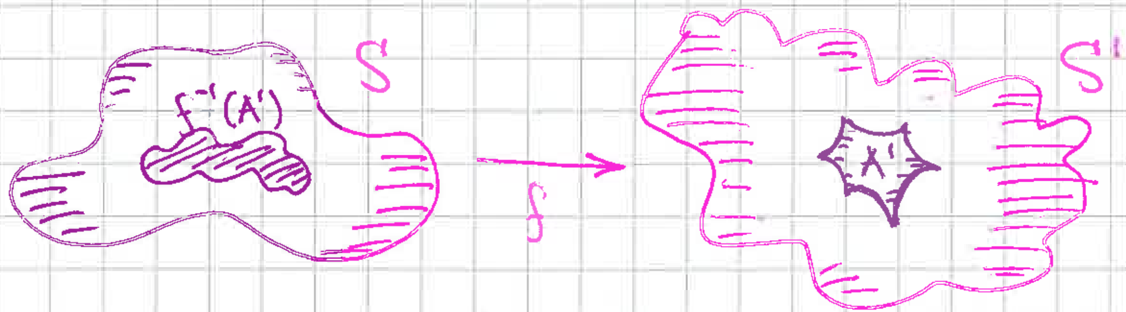
This requirement of reasonability of the function $X : \Omega \to S'$ is what the definition of "measurable function" captures.

Measurable functions and random variables

Let (S, Σ) and (S', Σ') be measurable spaces, i.e. Σ and Σ' are σ-algebras on sets S and S', respectively.

Def: A function $f: S \to S'$ is called Σ/Σ'-measurable (or just measurable, if the σ-algebras are clear from the context) if for all $A' \in \Sigma'$ we have $f^{-1}(A') \in \Sigma$.

Recall: The preimage $f^{-1}(A') := \{ s \in S \mid f(s) \in A' \}$.



$$
\begin{array}{ccccc}
f: & S & \longrightarrow & S' \\
 & \in & & \in \\
 & s & \longmapsto & f(s)
\end{array}
$$

$$S \supset f^{-1}(A') \xleftarrow{\ f^{-1}\ } A' \subset S'$$

Def: If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, and (S', Σ') is a measurable space, then a random variable with values in S' is a $\mathcal{F}/\Sigma'$-measurable function $X: \Omega \to S'$.

Remark: This precisely ensures that for any $A' \in \Sigma'$
$\{X \in A'\} = \{\omega \in \Omega \mid X(\omega) \in A'\}$ is an event.

Examples:

- $S' = \mathbb{R}$, $\Sigma' = \mathcal{B}$ (Borel σ-alg. on $\mathbb{R}$) corresponds to real-valued random variables, i.e. "random numbers".
  When $S' = \mathbb{R}$ we use the Borel σ-algebra without explicit mention, and just call a $\mathcal{F}/\mathcal{B}$-measurable $X: \Omega \to \mathbb{R}$ $\mathcal{F}$-measurable, and denote $X \in m\mathcal{F}$.
- $S' = \mathbb{R}^d$, $\Sigma' = \mathcal{B}(\mathbb{R}^d)$ (Borel σ-alg. on $\mathbb{R}^d$) corresponds to "random vectors"
- $S' = \mathbb{R}^{m \times n}$, $\Sigma' = \mathcal{B}(\mathbb{R}^{m \times n})$ "random matrices"
- S' a set of graphs : "random graph"
- etc.

Usually S and S' are at least topological spaces, so we can (and will) equip them with the Borel σ-algebras $\mathcal{B}(S)$ and $\mathcal{B}(S')$ (generated by open sets in S and S', resp.).
A $\mathcal{B}(S)/\mathcal{B}(S')$-measurable $f: S \to S'$ is called a <u>Borel function</u>.

One more example, slightly more complicated:

- $S' = C([0,1]) = \{$ continuous functions $h: [0,1] \to \mathbb{R} \}$ with the topology induced by the sup-norm $\|h\|_\infty = \sup_{x \in [0,1]} |h(x)|$ and the corresponding Borel σ-alg. $\mathcal{B}(C([0,1]))$ is the relevant space for one-dimensional Brownian motion on unit time interval.

Do random variables exist? How to construct them?

Example
- Ω countable set
- $\mathcal{F} = \mathcal{P}(\Omega) = \{ E \subset \Omega \}$
- probability mass function $p : \Omega \to [0,1]$ and corresp. proba meas. $\mathbb{P}[E] = \sum_{\omega \in E} p(\omega)$

$(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and any function $X : \Omega \to \mathbb{R}$ is a random variable.

(Indeed, for any $B \in \mathcal{B}$
$$X^{-1}(B) = \{ X \in B \} = \{ \omega \in \Omega \mid X(\omega) \in B \} \subset \Omega$$
is a subset, $X^{-1}(B) \in \mathcal{P}(\Omega) = \mathcal{F}$)

To verify measurability it is in fact enough to check the condition $f^{-1}(A') \in \Sigma$ just for A' in some class that generates Σ'.

Lemma: Let $\mathcal{C}' \subset \mathcal{F}(S')$ be such that $\sigma(\mathcal{C}') = \Sigma'$. Then $f : S \to S'$ is Σ / Σ'-measurable iff $f^{-1}(A') \in \Sigma \quad \forall A' \in \mathcal{C}'$.

Proof: The condition is clearly necessary, we only need to prove it is sufficient.

Suppose $f^{-1}(A') \in \Sigma \quad \forall A' \in \mathcal{C}'$.

Define $\mathcal{E} = \{ A' \in \Sigma' \mid f^{-1}(A') \in \Sigma \}$.

Then $\mathcal{C} \subset \mathcal{E}$. But since preimages satisfy
$$f^{-1}(A'^c) = (f^{-1}(A'))^c$$
$$f^{-1}\left(\bigcup_{n=1}^{\infty} A'_n\right) = \bigcup_{n=1}^{\infty} f^{-1}(A'_n)$$
$$f^{-1}(S') = S$$
we see that $\mathcal{E}$ is a σ-algebra on S'.

Therefore $\Sigma' = \sigma(\mathcal{C}') \subset \mathcal{E}$, so f is Σ / Σ' measurable. $\square$

<u>Corollary</u> Any continuous function is Borel-meas.

<u>Proof</u> Let S and S' be topological spaces, and $\mathcal{B}(S)$ and $\mathcal{B}(S')$ the Borel σ-algebras on S and S', resp. Let $f: S \to S'$ be continuous. For any open set $V' \subset S'$ the preimage $f^{-1}(V') \subset S$ is open (definition of continuity). Therefore, with $\mathcal{C}'$ the collection of open sets in S', we have $V' \in \mathcal{C}' \Longrightarrow f^{-1}(V') \in \mathcal{B}(S)$ because any open set in S is Borel. On the other hand $\mathcal{B}(S') = \sigma(\mathcal{C}')$, so $\mathcal{B}(S)/\mathcal{B}(S')$-measurability follows from previous lemma. $\square$

This gives a lot of measurable functions already.

In the case $S' = \mathbb{R}$ we can use an even simpler collection that generates the Borel σ-alg. $\mathcal{B}$.

<u>Corollary</u> A function $f: S \to \mathbb{R}$ is $\Sigma/\mathcal{B}$-measurable iff for all $c \in \mathbb{R}$ we have

$$\{f \leq c\} := \{s \in S \mid f(s) \leq c\} \in \Sigma.$$

<u>Proof</u>: The collection $\mathcal{I} = \{(-\infty, c] \mid c \in \mathbb{R}\}$ generates $\mathcal{B}$, so the result follows from the previous lemma.

Recall notation:

$$m\Sigma = \{f: S \to \mathbb{R} \mid f \text{ is } \Sigma/\mathcal{B}\text{-measurable}\}$$

Proposition: Let (S,Σ) be a measurable space. Then the set $m\Sigma$ of $\Sigma/\mathcal{B}$-measurable functions $S \to \mathbb{R}$ is an algebra:

(i) $f \in m\Sigma,\ \lambda \in \mathbb{R} \implies \lambda f \in m\Sigma$

(ii) $f_1, f_2 \in m\Sigma \implies f_1 + f_2 \in m\Sigma$

(iii) $f_1, f_2 \in m\Sigma \implies f_1 \cdot f_2 \in m\Sigma$

(Here we use the pointwise operations on functions, $(\lambda f)(s) = \lambda \cdot f(s)$, $(f_1+f_2)(s) = f_1(s) + f_2(s)$, $(f_1 \cdot f_2)(s) = f_1(s) \cdot f_2(s)$.)

Proof: Let us prove (ii) and leave (i) and (iii) as exercises.

Suppose $f_1, f_2 \in m\Sigma$. Note that for $c \in \mathbb{R}$,

$$f_1(s) + f_2(s) > c \iff \exists q \in \mathbb{Q} \text{ such that } f_1(s) > q \text{ and } f_2(s) > c-q.$$

Therefore

$$\{f_1+f_2 > c\} = \underbrace{\bigcup_{q \in \mathbb{Q}} \Big(\underbrace{\{f_1 > q\}}_{\substack{\in \Sigma \\ \text{since } f_1 \in m\Sigma}} \cap \underbrace{\{f_2 > c-q\}}_{\substack{\in \Sigma \\ \text{since } f_2 \in m\Sigma}} \Big)}_{\in \Sigma \text{ as a countable union of sets in } \Sigma}.$$

Now $\{f_1 + f_2 \le c\} = \{f_1+f_2 > c\}^c \in \Sigma$ also.

This is sufficient to conclude that $f_1 + f_2 \in m\Sigma$. $\square$

Remark: These operations allow us to construct a lot of new measurable $\mathbb{R}$-valued functions from given measurable functions.

Proposition Let (S, Σ) be a measurable space, and $f_1, f_2, \ldots \in m\Sigma$ measurable functions $S \to \mathbb{R}$. Then we have that

$$\sup_n f_n, \quad \inf_n f_n,$$
$$\limsup_n f_n, \quad \liminf_n f_n$$

are also $\Sigma / \mathcal{B}([-\infty, +\infty])$-measurable.

Remark: We still write $f \in m\Sigma$ for functions $f : S \to [-\infty, +\infty]$ which are $\Sigma / \mathcal{B}([-\infty,+\infty])$-measurable. The topology on $[-\infty, +\infty]$ is the same as on any closed interval, e.g. $[-\frac{\pi}{2}, +\frac{\pi}{2}] \ni x \longmapsto \tan(x) \in [-\infty, +\infty]$ is a homeomorphism.

(Again these are pointwise operations on functions, e.g. $\left(\sup_n f_n\right)(s) = \sup_n f_n(s)$)

Proof: For any $c \in [-\infty, +\infty]$ write for example

$$\left\{ \sup_n f_n \le c \right\} = \bigcap_{n=1}^{\infty} \underbrace{\{ f_n \le c \}}_{\substack{\in \Sigma \\ \text{because } f_n \in m\Sigma}} \in \Sigma .$$

This is sufficient to show that $\sup_n f_n \in m\Sigma$.

Similarly $\inf_n f_n \in m\Sigma$.

Then recall that

$$\limsup_n f_n = \inf_m \left(\sup_{k \ge m} f_k \right).$$

But $f_k \in m\Sigma \quad \forall k \Rightarrow g_m := \sup_{k \ge m} f_k \in m\Sigma$ by what we already proved, and then

$$\limsup_n f_n = \inf_m \left(\sup_{k \ge m} f_k \right) = \inf_m g_m \in m\Sigma \quad \text{also}.$$

Similarly one shows $\liminf_n f_n \in m\Sigma$. $\square$

Corollary If $f_1, f_2, \ldots \in m\Sigma$ and if $\lim_{n\to\infty} f_n$ exists pointwise ($\forall s \in S : \exists \lim_{n\to\infty} f_n(s) \in \mathbb{R}$) then $\lim_{n\to\infty} f_n \in m\Sigma$, too.

Proof: If the limit exists, it coincides with both limsup and liminf, so we can apply the previous proposition. $\square$

Remark: These operations allow us to construct a lot of new measurable $\mathbb{R}$-valued functions from given measurable functions.

Example (Coin tossing)

Let $\Omega = \{H,T\}^{\mathbb{N}}$ be the set of possible sequences of coin tosses (H = "heads", T = "tails"), $\omega = (\omega_1, \omega_2, \ldots) \in \Omega$ with $\omega_j \in \{H,T\}$ $\forall j \in \mathbb{N}$. Let $\mathcal{F}$ be the smallest σ-algebra containing events

$$\text{"}j\text{th coin toss is heads"} = \{\omega \in \Omega \mid \omega_j = H\},$$

i.e. $\mathcal{F} = \sigma\left(\{\{\omega \in \Omega \mid \omega_j = H\} \mid j \in \mathbb{N}\}\right)$.

Then
$$X_j(\omega) = \begin{cases} 1 & \text{if } \omega_j = H \\ 0 & \text{if } \omega_j = T \end{cases}$$
is a random variable (check directly, very easy!).

By an earlier lemma, also
$$M_n = \frac{1}{n}\sum_{j=1}^{n} X_j \quad \text{is a random variable.}$$

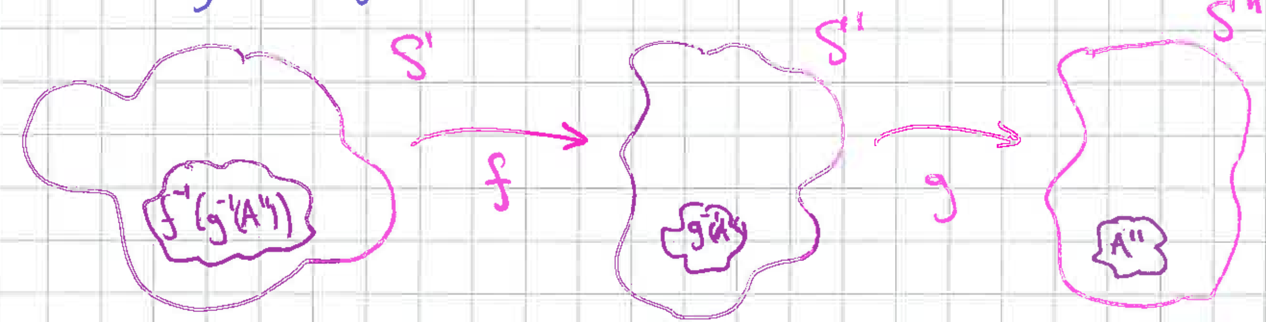
Therefore also $L^+ = \limsup_n \frac{1}{n}\sum_{j=1}^{n} X_j$ and $L^- = \liminf_n \frac{1}{n}\sum_{j=1}^{n} X_j$ are random variables, by proposition above.

For any $r \in [0,1]$ we thus see that
$$\{\omega \in \Omega \mid \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} X_j(\omega) = r\} = \{L^+ = r\} \cap \{L^- = r\}$$
is an event. It is meaningful to consider its probability.

One final, very easy way to construct measurable functions is composition.

<u>Lemma</u>: If (S,Σ), (S',Σ'), and (S'',Σ'') are measurable spaces, and $f: S \to S'$ is Σ/Σ'-measurable, and $g: S' \to S''$ is Σ'/Σ''-measurable, then $g \circ f : S \to S''$ is Σ/Σ''-measurable.

<u>Proof</u> Very easy, see picture:



$$A'' \in \Sigma'' \Rightarrow \underbrace{g^{-1}(A'')}_{=\{s' \in S' \mid g(s') \in A''\}} \in \Sigma'$$

$$\Rightarrow \underbrace{f^{-1}(g^{-1}(A''))}_{} \in \Sigma$$

$$= \{s \in S \mid f(s) \in g^{-1}(A'')\}$$

$$= \{s \in S \mid g(f(s)) \in A''\}$$

$$= (g \circ f)^{-1}(A'') \qquad \square$$

Let (S, Σ) be a measurable space, and consider the set $m\Sigma$ of $\Sigma/\mathcal{B}$-measurable functions $f : S \longrightarrow \mathbb{R}$.

We saw that pointwise limits $\lim_{n\to\infty} f_n$ of measurable functions $f_n \in m\Sigma$ are measurable.

It is very useful to notice that any measurable function is a pointwise limit of simple functions.

Even more powerful is the observation that any non-negative measurable function is a pointwise increasing limit of simple functions.

<u>Def</u>: The indicator (or indicator function) of $A \subset S$ is the function $1_A : S \longrightarrow \mathbb{R}$ given by
$$1_A(s) = \begin{cases} 1 & \text{if } s \in A \\ 0 & \text{if } s \notin A. \end{cases}$$

<u>Exercise</u>: $1_A : S \longrightarrow \mathbb{R}$ is $\Sigma/\mathcal{B}$-measurable if and only if $A \in \Sigma$.

<u>Example</u>: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $E \in \mathcal{F}$ an event. Then 1_E is the random variable which takes the value 1 if the event occurs, and 0 otherwise.

<u>Def</u>: A function $f : S \longrightarrow \mathbb{R}$ is called simple if it is $\Sigma/\mathcal{B}$-measurable, and takes only finitely many values.

<u>Remark</u>: $f : S \longrightarrow \mathbb{R}$ is simple if and only if it is a finite linear combination of indicators of measurable sets,
$$f = \sum_{k=1}^{n} a_k \cdot 1_{A_k} \quad \text{for some } A_k \in \Sigma, \quad k = 1, \dots, n.$$

Indeed, we can for example take $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ be the possible values of f, and $A_k = f^{-1}(\{\alpha_k\})$ for $k = 1, \dots, n$.

Now we prove the remarkably useful approximation result of measurable functions.

<u>Lemma</u> (Approximation of non-negative measurable functions)

Let $f : S \longrightarrow [0, +\infty]$ be a $\Sigma / \mathcal{B}([0, +\infty])$-measurable function. Then there exists a sequence $f_1, f_2, \dots : S \longrightarrow [0, +\infty)$ of simple functions such that $f_n \uparrow f$ pointwise as $n \longrightarrow \infty$.

<u>Proof</u>: We construct the approximating sequence explicitly.

For $n \in \mathbb{N}$ define

$$f_n = \sum_{k=1}^{n2^n} k \cdot 2^{-n} \cdot \mathbb{1}_{A_k^{(n)}}, \quad \text{where}$$

$$A_k^{(n)} = f^{-1}\left([k2^{-n}, (k+1)2^{-n})\right) \quad \text{for } k = 1, \dots, n2^n - 1$$

and $A_{n2^n}^{(n)} = f^{-1}([n, +\infty])$. Clearly each f_n is a simple function $f_n : S \longrightarrow [0, n]$.

We claim that $f_n \uparrow f$ pointwise.

Fix $s \in S$. There are two cases:

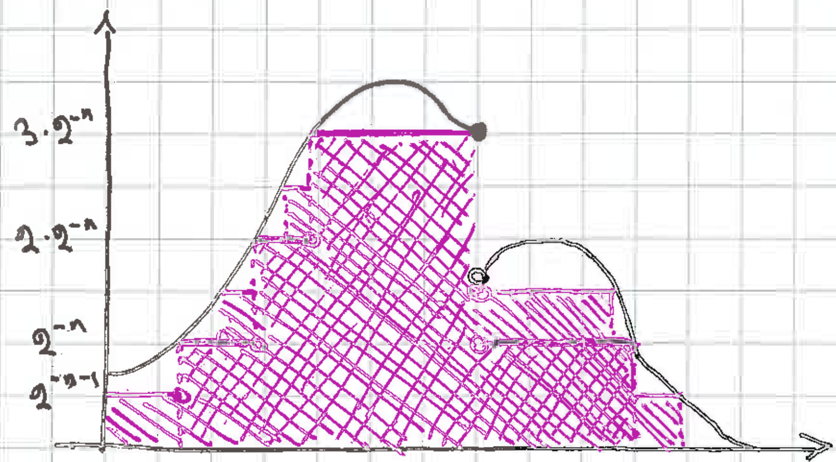
1°) $f(s) = +\infty$: Then $s \in A_{n2^n}^{(n)}$ for any n, so we have $f_n(s) = n \uparrow +\infty$ as $n \to \infty$.

2°) $f(s) \in [0, +\infty)$. Then for $n < f(s)$ we again have $f_n(s) = n$, and for $n \geq f(s)$ we have that $f_n(s) = k2^{-n}$, where $f(s) \in [k2^{-n}, (k+1)2^{-n})$. In particular $f(s) \geq f_n(s)$ and $|f(s) - f_n(s)| \leq 2^{-n} \xrightarrow[n\to\infty]{} 0$, so $f_n(s) \longrightarrow f(s)$. Also $f_n(s) \leq f_{n+1}(s)$ since

$$[k2^{-n}, (k+1)2^{-n}) = [(2k)2^{-(n+1)}, (2k+1)2^{-(n+1)}) \cup [(2k+1)2^{-(n+1)}, (2k+2)2^{-(n+1)})$$

□

Picture of the approximation



Exercise

DRAW A BETTER PICTURE FOR YOURSELF !

INFORMATION GENERATED BY RANDOM VARIABLES

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $(Y_\gamma)_{\gamma \in C}$ a collection of random variables $Y_\gamma : \Omega \to \mathbb{R}$ indexed by $\gamma \in C$. What is the "information generated by these random variables" ?

<u>Idea</u>:

- chance determines the random outcome $\omega \in \Omega$
- the outcome $\omega \in \Omega$ determines the values $Y_\gamma(\omega)$, $\gamma \in C$, of random variables.
- if someone tells you all the values of Y_γ, $\gamma \in C$, then for which events $E \in \mathcal{F}$ can you decide whether E occurred or not ?

Intuitively, the class of events whose occurrence you can decide on the basis of the values of the random variables Y_γ, $\gamma \in C$, is the information generated by $(Y_\gamma)_{\gamma \in C}$.

The formal definition is as follows:

<u>Def</u>: The *σ-algebra generated by $(Y_\gamma)_{\gamma \in C}$* is the smallest σ-algebra $\mathcal{G}$ on Ω such that each $Y_\gamma : \Omega \to \mathbb{R}$ is $\mathcal{G}/\mathcal{B}$-measurable. We denote $\mathcal{G} = \sigma(Y_\gamma : \gamma \in C)$.

An important special case is a collection which contains only one random variable $Y : \Omega \to \mathbb{R}$, in which case we denote the σ-algebra generated by Y by $\sigma(Y)$.

Remark: The definition $\mathcal{Y} = \sigma(Y)$ is clearly equivalent to $\mathcal{Y} = \sigma(\mathcal{C})$, the σ-algebra generated by the collection $\mathcal{C} = \{Y^{-1}(B) \mid B \in \mathcal{B}\}$ of subsets of Ω. In fact, you will show in an exercise that $\mathcal{Y} = \mathcal{C}$.

Remark: Anyway

$$\underbrace{\sigma(Y_\gamma : \gamma \in C)}_{\text{"information generated by random variables } Y_\gamma,\ \gamma \in C\text{"}} \subset \underbrace{\mathcal{F}}_{\text{"full information"}}$$

since all random variables are at least $\mathcal{F}/\mathcal{B}$-measurable.

The slightly abstract definition of $\sigma(Y)$ may be made somewhat more concrete by the following characterization of $\sigma(Y)$-measurable functions.

Theorem (Doob's representation theorem)

A random variable $Z : \Omega \to \mathbb{R}$ is $\sigma(Y)/\mathcal{B}$-measurable if and only if there exists a Borel function $f : \mathbb{R} \to \mathbb{R}$ such that $Z(\omega) = f(Y(\omega))$ $\forall \omega \in \Omega$.

The interpretation is that Z is measurable with respect to the information generated by Y iff the value of Z can be expressed as a function of the value of Y.

As usual, we denote just $Z = f(Y)$ to signify $Z(\omega) = f(Y(\omega))$ $\forall \omega \in \Omega$.
If we were to emphasize that random variables are functions $\Omega \to \mathbb{R}$, we might denote $Z = f \circ Y$.

We will derive Doob's representation theorem from another result, which is important for also other purposes.

Theorem (Monotone class theorem)

Let $\mathcal{H}$ be a class of bounded functions from S to $\mathbb{R}$ such that

(i) $\mathcal{H}$ is a vector space (over $\mathbb{R}$)

(ii) the constant function 1 is in $\mathcal{H}$.

(iii) if $f_1, f_2, \dots \in \mathcal{H}$ is a sequence of non-negative functions in $\mathcal{H}$ such that $f_n \uparrow f$, where f is a bounded function on S, then $f \in \mathcal{H}$.

Then if $\mathcal{H}$ contains the indicator function 1_A of every set $A \in \mathcal{I}$ in a π-system $\mathcal{I}$, then $\mathcal{H}$ contains all bounded $\sigma(\mathcal{I})/\mathcal{B}$-measurable functions.

Proof of Doob's representation theorem:

The "if" part is just composition lemma:

since $Y : \Omega \to \mathbb{R}$ is $\sigma(Y)/\mathcal{B}$-measurable and $f : \mathbb{R} \to \mathbb{R}$ is $\mathcal{B}/\mathcal{B}$-measurable, then $Z_1 = f \circ Y$ is $\sigma(Y)/\mathcal{B}$-measurable.

We first prove the "only if" part assuming that Z_1 is bounded.

So let $\mathcal{H}$ be the class of bounded functions $Z_1 : \Omega \to \mathbb{R}$ such that $Z_1 = f \circ Y$ for some bounded Borel function $f : \mathbb{R} \to \mathbb{R}$.

Note that $\mathcal{J} := \sigma(Y)$ is certainly a π-system, because it is a σ-algebra.
We use the fact that $\mathcal{J} = \sigma(Y) = \{Y^{-1}(B) \mid B \in \mathcal{B}\}$ to see that any $A \in \mathcal{J}$ is $A = Y^{-1}(B)$ for some $B \in \mathcal{B}$, and therefore the indicator $1_A : \Omega \to \mathbb{R}$ can be written as

$$1_A(\omega) = 1_B(Y(\omega)) = \begin{cases} 1 & \text{if } Y(\omega) \in B \\ 0 & \text{otherwise} \end{cases}.$$

But $1_B : \mathbb{R} \to \mathbb{R}$ is a bounded Borel-function, so this shows that $1_A \in \mathcal{H}$.
If we show that $\mathcal{H}$ is a monotone class, i.e. satisfies (i), (ii), (iii), then it follows by Monotone class theorem that $\mathcal{H}$ contains all bounded $\sigma(Y)/\mathcal{B}$-meas. functions.

(i) If $Z_1, Z_2 \in \mathcal{H}$ $(Z_1 = f_1 \circ Y,\ Z_2 = f_2 \circ Y)$ then for $c_1, c_2 \in \mathbb{R}$ $c_1 Z_1 + c_2 Z_2 = f \circ Y$ for a bounded Borel function $f : \mathbb{R} \to \mathbb{R}$ $(f = c_1 f_1 + c_2 f_2)$. Thus $\mathcal{H}$ is a vector space.

(ii) $\mathcal{H}$ contains constant function 1 (take $f \equiv 1$).

(iii) If $Z_n \uparrow Z$ with $Z_n \in \mathcal{H}$ $\forall n \in \mathbb{N}$ and $Z(\omega) \le K$ for some constant K, then $Z_n = f_n \circ Y$ for some $f_n : \mathbb{R} \to \mathbb{R}$ bounded Borel-meas. We can assume $f_n(x) \le K$ $\forall x \in \mathbb{R}$. (replace f_n by $f_n \wedge K = \min(f_n, K)$ if necessary). Define $f = \limsup_n f_n$. Then $f : \mathbb{R} \to \mathbb{R}$ is bounded Borel-meas., and

$$Z = \lim_{n \to \infty} Z_n = \lim_{n \to \infty} f_n \circ Y = \limsup_n f_n \circ Y = f \circ Y.$$

This shows that $Z \in \mathcal{H}$.

Thus every bounded $\sigma(Y)$-meas. $Z : \Omega \to \mathbb{R}$ is $Z = f \circ Y$.
For unbounded Z, define $\tilde{Z} = \arctan \circ Z$ which is bounded, $\tilde{Z} = \tilde{f} \circ Y$, so $Z = f \circ Y$ with $f = \tan \circ \tilde{f}$. $\square$

To prove the Monotone class theorem, we use the following definition and auxiliary results.

Def: A collection $\mathcal{D}$ of subsets of S is called a <u>d-system</u> on S if

(a) $S \in \mathcal{D}$

(b) $A, B \in \mathcal{D}$ and $A \subset B \Rightarrow B \setminus A \in \mathcal{D}$

(c) $A_n \in \mathcal{D}$ and $A_n \uparrow A \Rightarrow A \in \mathcal{D}$.

Proposition: A collection $\Sigma \subset \mathcal{P}(S)$ is a σ-algebra iff it is both a π-system and a d-system

Proof: The "if" part is obvious, so we only prove "only if".

Suppose Σ is both a π-syst. and a d-syst.

Then $S \in \Sigma$ by (a), and $A \in \Sigma \Rightarrow A^c = S \setminus A \in \Sigma$ by (a) and (b). It remains to show that Σ is stable under countable unions.

Suppose first that $A_1, A_2 \in \Sigma$. Now $A_1^c, A_2^c \in \Sigma$ by d-system properties and then $A_1^c \cap A_2^c \in \Sigma$ by π-system property. Therefore

$$A_1 \cup A_2 = S \setminus (A_1^c \cap A_2^c) \in \Sigma \quad \text{again}$$

by d-system properties. By induction, Σ is stable under finite unions.

Let $A_1, A_2, \ldots \in \Sigma$. Then $G_m = A_1 \cup A_2 \cup \cdots \cup A_m \in \Sigma$ as a finite union and $G_m \uparrow \bigcup_{n=1}^{\infty} A_n$.

By (c), $\bigcup_{n=1}^{\infty} A_n \in \Sigma$. $\square$

Def: The d-system generated by a collection $\mathcal{C} \subset \mathcal{P}(S)$ is the smallest d-system $d(\mathcal{C})$ that contains $\mathcal{C}$.

Remark: The definition makes sense again because the intersection of all such d-system is still a d-system (and obviously the smallest).

Lemma (Dynkin's lemma)

If $\mathcal{I}$ is a π-system, then $d(\mathcal{I}) = \sigma(\mathcal{I})$.

Proof: Clearly $d(\mathcal{I}) \subset \sigma(\mathcal{I})$, and by the previous proposition we only need to show that $d(\mathcal{I})$ is also a π-system.

We do this in two steps.

Step 1: Let $\mathcal{D}_1 = \{ B \in d(\mathcal{I}) \mid B \cap C \in d(\mathcal{I}) \ \forall C \in \mathcal{I} \}$.

We show that $\mathcal{D}_1$ is a d-system, and $\mathcal{D}_1 = d(\mathcal{I})$.

For $C \in \mathcal{I}$ we have $S \cap C = C \in \mathcal{I} \subset d(\mathcal{I})$, so $S \in \mathcal{D}_1$. If $A, B \in \mathcal{D}_1$, $A \subset B$, then

$$(B \setminus A) \cap \underbrace{C}_{\in \mathcal{I}} = \underbrace{(B \cap C)}_{\in d(\mathcal{I})} \setminus \underbrace{(A \cap C)}_{\in d(\mathcal{I})} \in d(\mathcal{I}) \text{ by (b) for d-syst. } d(\mathcal{I})$$

If $A_n \in \mathcal{D}_1$, $A_n \uparrow A = \bigcup_{n=1}^{\infty} A_n$, then

$$A \cap \underbrace{C}_{\in \mathcal{I}} = \left(\bigcup_{n=1}^{\infty} A_n \right) \cap C = \bigcup_{n=1}^{\infty} \underbrace{(A_n \cap C)}_{\in d(\mathcal{I})} \uparrow A \cap C \in d(\mathcal{I})$$

Therefore $\mathcal{D}_1$ is a d-system. But clearly $\mathcal{I} \subset \mathcal{D}_1$, so $\mathcal{D}_1 = d(\mathcal{I})$.

Step 2: Let $\mathcal{D}_2 = \{ A \in d(\mathcal{I}) \mid A \cap B \in d(\mathcal{I}) \ \forall B \in d(\mathcal{I}) \}$.

Step 1 showed $\mathcal{I} \subset \mathcal{D}_2$. Just as in Step 1 one can show that $\mathcal{D}_2$ is a d-system. Thus $\mathcal{D}_2 = d(\mathcal{I})$.

From $\mathcal{D}_2 = d(\mathcal{I})$ we conclude that $d(\mathcal{I})$ is a π-system. $\square$

Proof of Monotone class theorem:

Let $\mathcal{D}$ be the class of sets $A \subset S$ s.t. $1_A \in \mathcal{H}$. Properties (i), (ii), (iii) show that $\mathcal{D}$ is a d-system. By assumption $\mathcal{D}$ contains the π-syst. $\mathcal{I}$, so $\mathcal{D} \supset \sigma(\mathcal{I})$. If $f : S \to \mathbb{R}$ is a non-negative bounded $\sigma(\mathcal{I})/\mathcal{B}$-meas. function, we may approximate $f_n \uparrow f$ where each f_n is simple, $f_n = \sum_k a_k 1_{A_k}$. Thus $f_n \in \mathcal{H}$ and $f \in \mathcal{H}$. For general bounded $f : S \to \mathbb{R}$, write $f = f_+ - f_-$ with $f_+ = \max\{f, 0\}$ and $f_- = \max\{-f, 0\}$ to get $f_+, f_- \in \mathcal{H}$ and $f \in \mathcal{H}$. $\square$

The same techniques allow us to prove also Dynkin's identification theorem, which we recall below, and then prove.

Theorem (Dynkin's identification theorem)

Let μ_1 and μ_2 be two probability measures on (S, Σ), and let $\mathcal{J}$ be a π-system on S such that $\sigma(\mathcal{J}) = \Sigma$.
If $\mu_1[A] = \mu_2[A]$ for all $A \in \mathcal{J}$,
then $\mu_1 = \mu_2$.

Proof Let $\mathcal{D} = \{A \in \Sigma \mid \mu_1[A] = \mu_2[A]\}$.
By assumption, $\mathcal{J} \subset \mathcal{D}$. We claim that $\mathcal{D}$ is a d-system, so we need to check

(a): $S \in \mathcal{D}$, since $\mu_1[S] = 1 = \mu_2[S]$

(b): Suppose $A, B \in \mathcal{D}$ and $A \subset B$.
Then $B = A \cup (B \setminus A)$ so by additivity for disjoint sets $\mu_1[B] = \mu_1[A] + \mu_1[B \setminus A]$, and similarly for μ_2. Solve for measure of $B \setminus A$ to get

$$\begin{aligned}\mu_1[B\setminus A] &= \mu_1[B] - \mu_1[A] \\ &= \mu_2[B] - \mu_2[A] \quad \text{since } A, B \in \mathcal{D} \\ &= \mu_2[B \setminus A]\end{aligned}$$

which shows $B \setminus A \in \mathcal{D}$.

(c) Suppose $A_1, A_2, \ldots \in \mathcal{D}$ and $A_n \uparrow A$.
By monotone convergence of measures, $\mu_1[A_n] \uparrow \mu_1[A]$ and similarly for μ_2.
Then

$$\begin{aligned}\mu_1[A] &= \lim_{n \to \infty} \mu_1[A_n] \\ &= \lim_{n\to\infty} \mu_2[A_n] \quad \text{since } A_n \in \mathcal{D}\ \forall n \\ &= \mu_2[A].\end{aligned}$$

This shows $A \in \mathcal{D}$.

Now since $\mathcal{J} \subset \mathcal{D}$ and $\mathcal{D}$ is a d-system, $d(\mathcal{J}) \subset \mathcal{D}$.
But by Dynkin's lemma $d(\mathcal{J}) = \sigma(\mathcal{J}) = \Sigma$. $\square$

INDEPENDENCE

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

The notion of (probabilistic) "independence" is familiar from earlier courses in probability, so we do not elaborate on the interpretation. Rather, we give the general formal definitions, and then show their equivalence with more elementary definitions. Then we start examining some of the more profound consequences of independence.

Def (Independence of σ-algebras)

Let $(\mathcal{G}_j)_{j \in J}$ be a collection of σ-algebras $\mathcal{G}_j \subset \mathcal{F}$. The collection is said to be *independent* if for any distinct $j_1, \ldots, j_n \in J$ and any $A_{j_1} \in \mathcal{G}_{j_1}, \ldots, A_{j_n} \in \mathcal{G}_{j_n}$ we have

$$\mathbb{P}[A_{j_1} \cap \cdots \cap A_{j_n}] = \mathbb{P}[A_{j_1}] \cdot \cdots \cdot \mathbb{P}[A_{j_n}].$$

This is the general definition — independence of random variables and of events reduces to it:

Def (Independence of random variables)

Let $(X_j)_{j \in J}$ be a collection of random variables $X_j : \Omega \to \mathbb{R}$. The collection is said to be *independent* if the σ-algebras $(\sigma(X_j))_{j \in J}$ generated by the random variables are independent.

Def (Independence of events)

Let $(E_j)_{j\in J}$ be a collection of events, $E_j \in \mathcal{F}$. The collection is said to be independent if the indicator functions $(1_{E_j})_{j\in J}$ are independent random variables.

Remark: The σ-algebra generated by the indicator $1_E : \Omega \to \mathbb{R}$ of $E \subset \Omega$ is

$$\sigma(1_E) = \{\phi, E, E^c, \Omega\}.$$

Thus events $(E_j)_{j\in J}$ are independent iff the σ-algebras $(\{\phi, E_j, E_j^c, \Omega\})_{j\in J}$ are independent, i.e. for any distinct $j_1, \dots, j_n \in J$ we have $\mathbb{P}[E_{j_1}^{*_1} \cap \dots \cap E_{j_n}^{*_n}] = \mathbb{P}[E_{j_1}^{*_1}] \cdot \dots \cdot \mathbb{P}[E_{j_n}^{*_n}]$

where each $*_j$ can be a complement or not.

We abbreviate independence by the symbol $\perp\!\!\!\perp$, e.g.

- $(\mathcal{G}_j)_{j\in J}$ $\perp\!\!\!\perp$
- $(X_j)_{j\in J}$ $\perp\!\!\!\perp$
- $(E_j)_{j\in J}$ $\perp\!\!\!\perp$

or in the case of collections with two members

- $\mathcal{G}_1 \perp\!\!\!\perp \mathcal{G}_2$
- $X_1 \perp\!\!\!\perp X_2$
- $E_1 \perp\!\!\!\perp E_2$

or in the case of countably infinite collections (sequences)

- $\mathcal{G}_1, \mathcal{G}_2, \dots$ $\perp\!\!\!\perp$
- $X_1, X_2, \dots$ $\perp\!\!\!\perp$
- $E_1, E_2, \dots$ $\perp\!\!\!\perp$

Verifying independence

To check independence, one does not want to work directly with the general definition, but rather have an easier sufficient condition.

Proposition Let $\mathcal{J}_1, \mathcal{J}_2$ be π-systems on Ω, and $\mathcal{G}_1 = \sigma(\mathcal{J}_1)$, $\mathcal{G}_2 = \sigma(\mathcal{J}_2)$ the σ-algebras generated by them. Then $\mathcal{G}_1 \perp\!\!\!\perp \mathcal{G}_2$ iff
$$\mathbb{P}[A_1 \cap A_2] = \mathbb{P}[A_1]\,\mathbb{P}[A_2] \quad \forall A_1 \in \mathcal{J}_1,\ A_2 \in \mathcal{J}_2.$$

Proof: Fix $A_1 \in \mathcal{J}_1$. If $\mathbb{P}[A_1] = 0$ then $\mathbb{P}[A_1 \cap A_2] = 0 \quad \forall A_2$, so assume $\mathbb{P}[A_1] > 0$.

Define a prob measure μ on $(\Omega, \mathcal{G}_2)$ by
$$\mu[E_2] = \frac{\mathbb{P}[A_1 \cap E_2]}{\mathbb{P}[A_1]} \quad \text{for } E_2 \in \mathcal{G}_2.$$
(Exercise: μ is a measure) By the assumption, μ and $\mathbb{P}$ coincide on the π-system $\mathcal{J}_2$, so by Dynkin's identification theorem they agree on $\mathcal{G}_2 = \sigma(\mathcal{J}_2)$. $\therefore$ $\forall A_1 \in \mathcal{J}_1$, $\forall E_2 \in \mathcal{G}_2$ we have $\mathbb{P}[A_1 \cap E_2] = \mathbb{P}[A_1]\mathbb{P}[E_2]$

Then fix $E_2 \in \mathcal{G}_2$, and assume again $\mathbb{P}[E_2] > 0$. The measure $\tilde{\mu}$ on $(\Omega, \mathcal{G}_1)$ defined by
$$\tilde{\mu}[E_1] = \frac{\mathbb{P}[E_1 \cap E_2]}{\mathbb{P}[E_2]} \quad \text{for } E_1 \in \mathcal{G}_1$$
coincides with $\mathbb{P}$ on $\mathcal{J}_1$ by the first part, so $\tilde{\mu}$ and $\mathbb{P}$ coincide on $\mathcal{G}_1 = \sigma(\mathcal{J}_1)$.

This proves the claim: for all $E_1 \in \mathcal{G}_1$, $E_2 \in \mathcal{G}_2$
$$\mathbb{P}[E_1 \cap E_2] = \mathbb{P}[E_1]\,\mathbb{P}[E_2].$$
$\square$

An example application is the following criterion for the independence of two $\mathbb{R}$-valued random variables.

<u>Corollary</u> Let $X_1, X_2 : \Omega \to \mathbb{R}$ be two random variables. Then $X_1 \perp\!\!\!\perp X_2$ iff

$$\forall x_1, x_2 \in \mathbb{R} : \quad \mathbb{P}[X_1 \le x_1, X_2 \le x_2] = \mathbb{P}[X_1 \le x_1] \cdot \mathbb{P}[X_2 \le x_2].$$

<u>Proof</u> Recall that the π-system $\mathcal{J} = \{(-\infty, x] \mid x \in \mathbb{R}\}$ generates the Borel σ-algebra $\mathcal{B} = \sigma(\mathcal{J})$.

Set $\mathcal{J}_1 = X_1^{-1}(\mathcal{J}) = \{\{\omega \in \Omega \mid X_1(\omega) \le x\} \mid x \in \mathbb{R}\}$.

Then $\mathcal{J}_1$ is a π-system which generates $\sigma(X_1)$ (Exercise). Similarly $\mathcal{J}_2 = X_2^{-1}(\mathcal{J})$ is a π-syst. which generates $\sigma(X_2)$. The result then follows from the previous proposition. $\square$

In a similar way one can prove that random variables $X_1, \dots, X_n : \Omega \to \mathbb{R}$ are independent iff

$$\forall x_1, \dots, x_n \in \mathbb{R} : \quad \mathbb{P}[X_1 \le x_1, \dots, X_n \le x_n] = \prod_{j=1}^{n} \mathbb{P}[X_j \le x_j].$$

To indicate what is done with more than two σ-algebras, consider the following.

<u>Exercise</u>: Let $\mathcal{J}_1, \mathcal{J}_2, \mathcal{J}_3 \subset \mathcal{F}$ be three π-systems such that $\Omega \in \mathcal{J}_k$ for $k = 1, 2, 3$.

Prove that if for all $A_1 \in \mathcal{J}_1$, $A_2 \in \mathcal{J}_2$, $A_3 \in \mathcal{J}_3$

$$\mathbb{P}[A_1 \cap A_2 \cap A_3] = \mathbb{P}[A_1]\, \mathbb{P}[A_2]\, \mathbb{P}[A_3]$$

then $\sigma(\mathcal{J}_1), \sigma(\mathcal{J}_2), \sigma(\mathcal{J}_3)$ are independent.

The Borel – Cantelli lemmas

Recall the following definitions for a sequence $E_1, E_2, \ldots \in \mathcal{F}$ of events :

- "E_n occurs infinitely often" or briefly "E_n i.o." is the event
$$\bigcap_{m\in\mathbb{N}} \bigcup_{k\geq m} E_k = \{\omega\in\Omega \mid \forall m\geq 1\ \exists k\geq m \text{ s.t. } \omega\in E_k\}$$
$$= \{\omega\in\Omega \mid \omega\in E_n \text{ for infinitely many } n\}$$
- "E_n occurs eventually" or briefly "E_n ev." is the event
$$\bigcup_{m\in\mathbb{N}} \bigcap_{k\geq m} E_k = \{\omega\in\Omega \mid \exists m\geq 1 \text{ s.t. } \forall k\geq m\ \omega\in E_k\}$$
$$= \{\omega\in\Omega \mid \omega\in E_n \text{ for all large enough } n\}$$

The first Borel – Cantelli lemma says that if the probabilities of E_n decay fast enough, then E_n does not occur infinitely often.

Lemma (1st Borel - Cantelli lemma)

If $\sum_{n=1}^{\infty} \mathbb{P}[E_n] < \infty$ then $\mathbb{P}[E_n \text{ i.o.}] = 0$.

Proof: Let $G_m = \bigcup_{k=m}^{\infty} E_k$, so $G_m \downarrow G =$ "E_n i.o." and $\mathbb{P}[E_n \text{ i.o.}] = \mathbb{P}[G] = \lim_{m\to\infty} \mathbb{P}[G_m]$ by monotone convergence for finite measures.

Now $0 \leq \mathbb{P}[G] \leq \mathbb{P}[G_m] \leq \sum_{k=m}^{\infty} \mathbb{P}[E_k]$.

But as $m\to\infty$ the left hand side tends to zero, because the series $\sum_{k=1}^{\infty}\mathbb{P}[E_k]$ was assumed convergent. This shows $\mathbb{P}[G]=0$. □

The second Borel-Cantelli lemma is a sort of converse of the first, but requires the assumption of independence in addition.

<u>Lemma</u> (2nd Borel-Cantelli lemma)

If $E_1, E_2, \ldots$ are independent and $\sum_{n=1}^{\infty} \mathbb{P}[E_n] = +\infty$ then $\mathbb{P}[E_n \text{ i.o.}] = 1$.

<u>Proof</u> We will rather show that the complement

$$(\text{"}E_n \text{ i.o."})^c = \Big(\bigcap_{m \geq 1} \bigcup_{k \geq m} E_k\Big)^c = \bigcup_{m \geq 1} \bigcap_{k \geq m} E_k^c$$

$$= (\text{"}E_n^c \text{ ev."})$$

has zero probability.

Denote $p_n = \mathbb{P}[E_n]$. Then

$$\mathbb{P}\Big[\bigcap_{m \leq k \leq m'} E_k^c\Big] \overset{\text{independence}}{=} \prod_{m \leq k \leq m'} \underbrace{(1-p_k)}_{\leq e^{-p_k}} \leq \prod_{m \leq k \leq m'} e^{-p_k}$$

$$= \exp\Big(-\sum_{k=m}^{m'} p_k\Big).$$

Now because $\sum_{n=1}^{\infty} p_n = +\infty$ we have

$$\exp\Big(-\sum_{k=m}^{m'} p_k\Big) \xrightarrow[m' \to \infty]{} 0, \quad \text{for all } m \geq 1.$$

On the other hand, as $m' \to \infty$

$$\bigcap_{m \leq k \leq m'} E_k^c \downarrow \bigcap_{k \geq m} E_k^c \quad \text{so by}$$

monotone convergence of finite measures

$$\mathbb{P}\Big[\bigcap_{k \geq m} E_k^c\Big] = \lim_{m' \to \infty} \mathbb{P}\Big[\bigcap_{m \leq k \leq m'} E_k^c\Big] = 0,$$

for all $m \geq 1$. Then by union bound

$$\mathbb{P}[E_n^c \text{ ev.}] = \mathbb{P}\Big[\bigcup_{m=1}^{\infty} \bigcap_{k \geq m} E_k^c\Big] \leq \sum_{m=1}^{\infty} \mathbb{P}\Big[\bigcap_{k \geq m} E_k^c\Big] = \sum_{m=1}^{\infty} 0 = 0.$$

$\square$

Example Let $X_1, X_2, \ldots$ be i.i.d. ("independent and identically distributed") with distribution $\mathrm{Exp}(1)$:

$$\mathbb{P}[X_n \leq x] = \begin{cases} 1 - e^{-x} & , \text{ for } x \geq 0 \\ 0 & , \text{ for } x < 0 . \end{cases}$$

Q: How big values does the sequence reach in the long run?

E.g. does the sequence ever hit above say 250 000?

≈ "Does some Polonium-214 atom (half life 164 μs) survive one minute without decaying?"

Let $c = 250\,000$, for example.

$$\mathbb{P}[X_n \leq c \ \ \forall n \geq 1] = \mathbb{P}\Big[\bigcap_{n \geq 1} \{X_n \leq c\}\Big]$$

$$= \lim_{m \to \infty} \mathbb{P}\Big[\bigcap_{n=1}^{m} \{X_n \leq c\}\Big] = \lim_{m \to \infty} (1 - e^{-c})^m = 0 .$$

Hence $\mathbb{P}[X_n > c \text{ for some } n \geq 1] = 1$.

Borel-Cantelli lemmas give sharper results.

Note first: for any $\alpha > 0$

$$\mathbb{P}[X_n > \alpha \cdot \log(n)] = e^{-\alpha \cdot \log(n)} = n^{-\alpha} .$$

Hence $$\sum_{n=1}^{\infty} \mathbb{P}[X_n > \alpha \cdot \log(n)] = \sum_{n=1}^{\infty} n^{-\alpha} = \begin{cases} +\infty & , \text{ if } \alpha \leq 1 \\ \text{finite} & , \text{ if } \alpha > 1 . \end{cases}$$

1^{st} BC $\Rightarrow$ $\mathbb{P}[X_n > \alpha \cdot \log(n) \text{ i.o.}] = 0$ for any $\alpha > 1$.

2^{nd} BC $\Rightarrow$ $\mathbb{P}[X_n > \log(n) \text{ i.o.}] = 1$.

A small exercise using these shows that

$$\limsup_{n \to \infty} \frac{X_n}{\log(n)} = 1 \quad \underbrace{\text{almost surely}}_{= \text{ with probability one}}$$

$$\Big(\mathbb{P}\Big[\limsup_{n} \frac{X_n}{\log(n)} = 1\Big] = 1 . \Big)$$

3.2 Infinite-horizon information

$(\Omega, \mathcal{F}, \mathbb{P})$. $(X_1, X_2, \ldots)$ random sequence.

- What is the information content "in the infinite future"?
- Information up to time n: $\sigma(X_1, \ldots, X_n)$.
- Information after n: $\mathcal{T}_n := \sigma(X_{n+1}, X_{n+2}, \ldots)$
- Tail σ-algebra: $\mathcal{T}_\infty := \bigcap_{n \geq 1} \sigma(X_{n+1}, \ldots)$

$$= \bigcap_{n \geq 1} \mathcal{T}_n.$$

Events in $\mathcal{T}_\infty$: (HW)

- $F_1 = (\lim_{n \to \infty} X_n \text{ exists})$
- $F_2 = (\sum_{n \geq 1} X_n \text{ converges})$

$F_3 = (\lim_{k \to \infty} \frac{1}{k}(X_1 + \cdots + X_k) \text{ exists})$

Fact [KO-1, W 4.11]

The tail σ-algebra of a random sequence of indep. terms is a $\mathbb{P}$-trivial:

(i) $F \in \mathcal{T}_\infty \Rightarrow \mathbb{P}(F) = 0$ or $\mathbb{P}(F) = 1$.

(ii) $Z \in \mathcal{T}_\infty \Rightarrow \mathbb{P}(Z = c) = 1$ for some c.

Proof Let $\mathcal{S}_n = \sigma(X_1, \dots, X_n)$
$\mathcal{T}_n = \sigma(X_{n+1}, X_{n+2}, \dots)$

Claim:

1. $\mathcal{S}_n \perp\!\!\!\perp \mathcal{T}_n$

Let $\mathcal{I}_n = \left\{ \bigcap_{j=1}^{n} X_j^{-1}((-\infty, x_j]) : x_j \in \mathbb{R} \cup \{\infty\} \right\}$

$\mathcal{J}_n = \left\{ \bigcap_{j=n+1}^{n+r} X_j^{-1}((-\infty, x_j]) : r \in \mathbb{N},\ x_j \in \mathbb{R} \cup \{\infty\} \right\}$

Then $\mathcal{I}_n, \mathcal{J}_n$ are π-syst., generating $\mathcal{S}_n, \mathcal{T}_n$.

Because $\mathcal{I}_n \perp\!\!\!\perp \mathcal{J}_n$, $\mathcal{S}_n \perp\!\!\!\perp \mathcal{T}_n$ as well.

2. Claim: $\mathcal{S}_n \perp\!\!\!\perp \mathcal{T}_\infty \ \forall n$.

Clear because $\mathcal{T}_\infty = \bigcap \mathcal{T}_k \Rightarrow \mathcal{T}_\infty \subset \mathcal{T}_n$.

Claim 3. $\mathcal{S}_\infty \perp\!\!\!\perp \mathcal{T}_\infty$, ($\infty$-horizon information)

where $\mathcal{S}_\infty = \sigma(X_1, X_2, \dots) = \sigma\left(\bigcup_{n \geq 1} X_n^{-1}(\mathcal{B}) \right)$

information of the random seq

$\mathcal{K}_\infty := \bigcup_{n \geq 1} \mathcal{S}_n$ is a π-system which generates $\mathcal{S}_\infty$. Why?

- $A, B \in \mathcal{K}_\infty \Rightarrow A \in \mathcal{S}_m, B \in \mathcal{S}_n$ for some $m, n \geq 1$.
$\Rightarrow A, B \in \mathcal{S}_{m \vee n}$
$\Rightarrow A \cap B \in \mathcal{S}_{m \vee n} \Rightarrow A \cap B \in \mathcal{K}_\infty$.
- $\mathcal{S}_\infty$ is a σ-alg (by def.), and $\mathcal{S}_\infty$ contains $\bigcup_{n \geq 1} \mathcal{S}_n$. (because $\mathcal{S}_\infty$ contains $\mathcal{S}_n \ \forall n$).
Now $\sigma(\mathcal{K}_\infty) \subset \mathcal{S}_\infty$.
Also $A \in \bigcup_{n \geq 1} X_n^{-1}(\mathcal{B}) \Rightarrow A \in X_n^{-1}(\mathcal{B})$, some n.
$\Rightarrow A \in \mathcal{S}_n$, some $n \Rightarrow A \in \mathcal{K}_\infty$.
$\Rightarrow \bigcup X_n^{-1}(\mathcal{B}) \subset \mathcal{K}_\infty$.
$\Rightarrow \mathcal{S}_\infty = \sigma(\bigcup X_n^{-1}(\mathcal{B})) \subset \sigma(\mathcal{K}_\infty)$.

$\Rightarrow \mathcal{S}_\infty = \sigma(\mathcal{K}_\infty)$.

Step 2 $\Rightarrow \mathcal{K}_\infty \perp\!\!\!\perp \mathcal{T}_\infty \xrightarrow{\text{Fact}} \mathcal{S}_\infty \perp\!\!\!\perp \mathcal{T}_\infty$.

~~Note By~~

∞-horizon information · full information of seq.

Step 4. Note that $\mathcal{T}_\infty \subset \mathcal{J}_\infty$,

and $\mathcal{J}_\infty \perp\!\!\!\perp \mathcal{T}_\infty$.

Hence $\mathcal{T}_\infty \perp\!\!\!\perp \mathcal{T}_\infty \Longrightarrow$??

Thus $\mathbb{P}(F \cap F) = \mathbb{P}(F)\,\mathbb{P}(F) \quad \forall F \in \mathcal{T}_\infty$

$\Longrightarrow \mathbb{P}(F) \in \{0,1\}$. $\square$

(ii) Assume that Z is $\mathcal{T}_\infty$-measurable.

Let $c = \sup\{x : \mathbb{P}(Z \le x) = 0\}$.

Assume $c \in \mathbb{R}$.

Then $\mathbb{P}(Z < c) = \mathbb{P}\Big(\bigcup_{n \ge 1} \{Z \le c - \tfrac{1}{n}\}\Big)$

$= \lim_{n\to\infty} \mathbb{P}(Z \le c - \tfrac{1}{n}) = 0$,

and $\mathbb{P}(Z \le c) = \mathbb{P}\big(\bigcap \{Z \le c + 1/n\}\big)$

$= \lim_{n\to\infty} \mathbb{P}(\underbrace{Z \le c + \tfrac{1}{n}}_{\in \mathcal{T}_\infty})$

$\in \{0,1\} \Longrightarrow = 1$.

$\Longrightarrow \boxed{\mathbb{P}(Z = c) = 1.}$ $= 1$. $\square$

(4

Consequences:

Let $(X_1, \dots)$ be indep.

Then either: $\mathbb{P}(\sum X_n \text{ converges}) = 0$

or $\mathbb{P}(\text{—''—}) = 1$.

Monkey typing Shakespeare

[W 4.9].

Monkey types one ~~letter~~ symbol, ~~unif~~ indep. random, one per time unit.

$\Rightarrow (X_1, X_2, \dots)$: $X_n \in$ alphabet.

WS : ~~....~~ is a sequence of N symbols.

~~Assume $\varepsilon = \min_x \mathbb{P}(X_1 = x) > 0$~~

~~Then~~

What is the pr. that the monkey ~~ever~~ produces ~~...~~ ∞ many copies of WS?

Let H be this event.

~~Let $H(k)$ be the event that~~

Let $H^{(m)}$ be the event that the monkey produces ∞ many copies during $[m+1, \infty)$.

Then $H^{(m)} \in \sigma(X_{m+1}, \dots) = \mathcal{T}_m$

$H = H^{(m)} \; \forall m \geq 1$.

$\Rightarrow H \in \mathcal{T}_m \; \forall m \geq 1$

$\Rightarrow H \in \mathcal{T}_\infty$.

$\Rightarrow \mathbb{P}(H) = 0$ or 1.

INTEGRATION $\int, \Sigma, \mathbb{E}$?!?

Let (S, Σ, μ) be a measure space.

Goal: For all reasonable functions $f : S \to \mathbb{R}$ we want to construct the integral of f with respect to the measure μ,

$$\int f \, d\mu = \int f(s) \, d\mu(s)$$

alternative notations for the same thing.

Special cases:

1°) Summation is the integral with respect to the counting measure ($\mu_{count}[A] = \# A$)

$$\sum_{s \in S} f(s) = \int f(s) \, d\mu_{count}(s)$$

2°) The integral with respect to the Lebesgue measure on $\mathbb{R}$ ($\mu_{Leb}[(a,b)] = b-a$) coincides with the familiar Riemann integral for all Riemann-integrable functions $f : \mathbb{R} \to \mathbb{R}$, i.e.

$$\int_{-\infty}^{\infty} f(x) \, dx = \int f(x) \, d\mu_{Leb}(x).$$

In fact, this Lebesgue integral generalizes the Riemann integral, and enjoys much nicer properties.

3°) Expected value $\mathbb{E}$ of a random variable $X : \Omega \to \mathbb{R}$ is the integral w.r.t. proba meas. $\mathbb{P}$

$$\mathbb{E}[X] = \int X(\omega) \, d\mathbb{P}(\omega).$$

The construction of the integral proceeds by increasing the allowed complexity step by step:

1. Indicator functions
2. Simple functions
3. Positive measurable functions
4. All integrable functions.

Many important results about integration are similarly proved step by step — this proof strategy could be called "the standard machine".

First recall definitions and an approximation result.

<u>Def</u>: The indicator function $1_A : S \to \mathbb{R}$ is

$$1_A(s) = \begin{cases} 1 & \text{if } s \in A \\ 0 & \text{if } s \in S \setminus A. \end{cases}$$

<u>Def</u>: A function $f: S \to \mathbb{R}$ is simple, if

$$f(s) = \sum_{j=1}^{n} a_j \cdot 1_{A_j}$$

for some $a_1, \dots, a_n \in \mathbb{R}$ and $A_1, \dots, A_n \in \Sigma$.

<u>Lemma</u> (Approximation lemma)

Let $f: S \to [0, +\infty]$ be a Σ-measurable function. Then there exists a sequence $f_1, f_2, \dots$ of non-negative simple functions $f_n : S \to [0, \infty)$ such that $f_n(s) \uparrow f(s)$ for all $s \in S$.

<u>Recall proof idea</u>: Let $f_n = \alpha_n \circ f$ where α_n is the n^{th} staircase function $\alpha_n : [0, \infty) \to [0, n]$

$$\alpha_n(x) = \begin{cases} j \cdot 2^{-n} & \text{if } j 2^{-n} \leq x < (j+1) 2^{-n}, \quad j \leq n 2^n \\ n & \text{if } x \geq n. \end{cases}$$

Naive integral for simple functions:

Let us first define a naive integral, denoted by $\int^{(\text{naive})} f \, d\mu$. We want

$$\int^{(\text{naive})} 1_A(s) \, d\mu(s) = \mu[A],$$

and we take this as the definition, which we extend linearly to simple functions $f = \sum_{j=1}^{n} a_j 1_{A_j}$ by

$$\int^{(\text{naive})} f(s) \, d\mu(s) = \sum_{j=1}^{n} a_j \cdot \mu[A_j].$$

Proposition: The naive integral of any simple function is well-defined (it does not depend on the way we choose to express f as a linear combination of indicator functions)

This is rather obvious, and its better if you convince yourself of this. It is just important to note that there is something to check, because the definition of the naive integral uses an expression for f, which is not unique.

Lemma (Properties of the naive integral)

Let $f, g : S \to \mathbb{R}$ be simple functions and $c \in \mathbb{R}$. Then:

(a) If $\mu[\{s \in S \mid f(s) \neq g(s)\}] = 0$ then
$$\int^{(naive)} f \, d\mu = \int^{(naive)} g \, d\mu .$$

(b) $\int^{(naive)} (f+g) \, d\mu = \int^{(naive)} f \, d\mu + \int^{(naive)} g \, d\mu$

(c) $\int^{(naive)} (c \cdot f) \, d\mu = c \cdot \int^{(naive)} f \, d\mu$

(Linearity)

(d) If $f \leq g$ then $\int^{(naive)} f \, d\mu \leq \int^{(naive)} g \, d\mu$ (Monotonicity)

Proof: (These are all very easy, and the details of the proof are left as an exercise.)

Integral for non-negative measurable functions:

Let $f: S \to [0, +\infty]$ is Σ-measurable, then define

$$\int f \, d\mu = \sup \left\{ \int^{(\text{naive})} h \cdot d\mu \;\middle|\; \begin{array}{l} h \text{ simple func.} \\ \text{s.t. } 0 \le h \le f \end{array} \right\}$$

Lemma For $f: S \to [0, +\infty)$ simple function, we have $\int f \, d\mu = \int^{(\text{naive})} f \, d\mu$

This says that for simple functions the actual integral is the same as the naive integral.

Proof: Taking $h = f$ we see that

$$\int f \, d\mu \ge \int^{(\text{naive})} h \, d\mu = \int^{(\text{naive})} f \, d\mu .$$

On the other hand, for any simple h such that $0 \le h \le f$ we have by monotonicity that $\int^{(\text{naive})} h \, d\mu \le \int^{(\text{naive})} f \, d\mu$, so we get

$$\int f \, d\mu = \sup_{\substack{h \text{ simple} \\ 0 \le h \le f}} \int^{(\text{naive})} h \, d\mu \le \int^{(\text{naive})} f \, d\mu . \quad \square$$

Lemma If $f: S \to [0, +\infty]$ is Σ-measurable and $\int f \, d\mu = 0$, then

$$\mu\left[\{ s \in S \mid f(s) > 0 \}\right] = 0$$

(i.e., $f \equiv 0$ μ-almost everywhere)

Proof: For $n \in \mathbb{N}$ let $A_n = \{ s \in S \mid f(s) > \frac{1}{n} \}$. Then $\bigcup_{n=1}^{\infty} A_n = \{ s \in S \mid f(s) > 0 \}$. But by definition of the integral of f we have

$$0 = \int f \, d\mu \ge \int^{(\text{naive})} \frac{1}{n} \mathbb{1}_{A_n} \, d\mu = \frac{1}{n} \cdot \mu[A_n].$$

Therefore $\mu[A_n] = 0$. By subadditivity $\mu\left[\bigcup_{n=1}^{\infty} A_n\right] \le \sum_{n=1}^{\infty} \mu[A_n] = 0$. $\square$

The following theorem is at the same time

- "all there is to integration"
- very important calculation tool in applications
- not too difficult to prove

Theorem (Monotone Convergence Theorem)

If $f_1, f_2, \dots$ are are Σ-measurable functions $S \longrightarrow [0, +\infty]$ and $f_n \uparrow f$, then

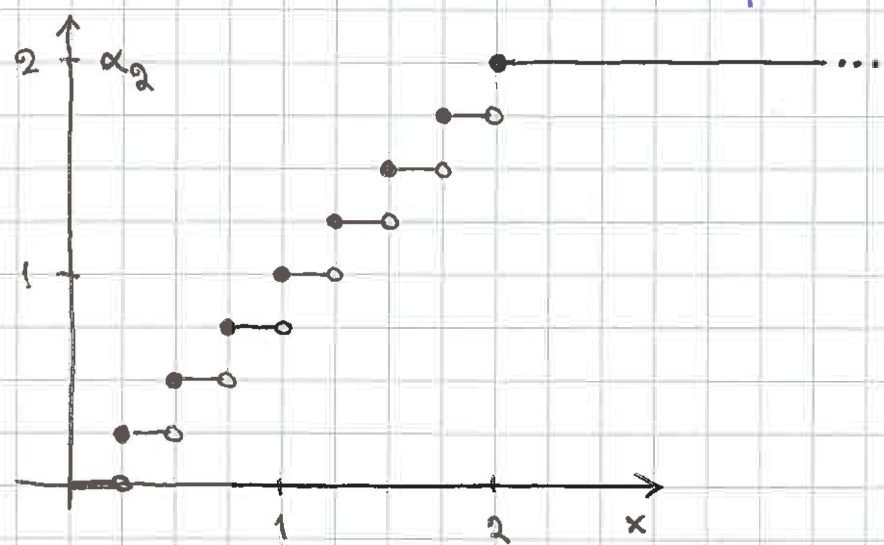
$$\int f_n \, d\mu \uparrow \int f \, d\mu .$$

Before giving the proof, note that since any non-neg. Σ-measurable f can be approximated by simple functions $f_1, f_2, \dots$ in a pointwise increasing way, $f_r \uparrow f$ as $r \to \infty$ we get

$$\int f \, d\mu = \lim_{r \to \infty} \int^{(\text{naive})} f_r \, d\mu$$

(an increasing limit).

We may for example take $f_r = \alpha_r \circ f$ where α_r is the r^{th} step function.



We postpone the proof of Monotone Convergence Theorem to the next lecture.

Integral for integrable functions

For $f \in m\Sigma$ (measurable functions $S \to [-\infty, +\infty]$)

$f^+(s) = \max(f(s), 0)$ and $f^-(s) = \max(-f(s), 0)$.

Then $f^+, f^- : S \to [0, +\infty]$ are Σ-measurable and

$f = f^+ - f^-$ and $|f| = f^+ + f^-$.

Def: We say that f is μ-integrable and write $f \in \mathcal{L}^1(\mu)$ if $\int |f| \, d\mu < +\infty$.

Remark: $\int |f| \, d\mu = \int (f^+ + f^-) \, d\mu = \int f^+ d\mu + \int f^- d\mu$

so $f \in \mathcal{L}^1(\mu) \iff \int f^+ d\mu < +\infty$ and $\int f^- d\mu < +\infty$.

For $f \in \mathcal{L}^1(\mu)$ we define the integral by

$$\int f \, d\mu = \int f^+ d\mu - \int f^- d\mu.$$

(Note: For non-negative $f \geq 0$ we have $f^+ = f$, $f^- = 0$ so obviously the definition coincides with the earlier one)

Lemma We have $\left| \int f \, d\mu \right| \leq \int |f| \, d\mu$ ("Triangle inequality for integrals")

Proof: $\int f \, d\mu = \int f^+ d\mu - \int f^- d\mu \leq \int f^+ d\mu + \int f^- d\mu = \int |f| \, d\mu$

and $-\int f \, d\mu = -\int f^+ d\mu + \int f^- d\mu \leq \int f^+ d\mu + \int f^- d\mu = \int |f| \, d\mu$ $\square$

Proposition For $a, b \in \mathbb{R}$ and $f, g \in \mathcal{L}^1(\mu)$ we have

$$\int (a \cdot f + b \cdot g) \, d\mu = a \cdot \int f \, d\mu + b \cdot \int g \, d\mu$$

Proof: (This is a rather straightforward consequence of a similar lemma for the integral of non-negative functions. We leave the details as an exercise.)

Expectation

The expected value is just the integral with respect to a probability measure.

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

For $X : \Omega \to \mathbb{R}$ a random variable, we set

Just like we defined integral in general!

- if $X = 1_A$ for $A \in \mathcal{F}$, $\mathbb{E}[X] = \mathbb{P}[A]$
- if $X = \sum_{j=1}^{n} a_j 1_{A_j}$ is simple then $\mathbb{E}[X] = \sum_{j=1}^{n} a_j \mathbb{P}[A_j]$
- if X is non-negative then $\mathbb{E}[X] = \sup_{\substack{H \text{ simple} \\ 0 \le H \le X}} \mathbb{E}[H]$
- if $X \in \mathcal{L}^1(\mathbb{P})$, i.e. $\mathbb{E}[X^+] < +\infty$ and $\mathbb{E}[X^-] < +\infty$ then $\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-]$.

Def: The law of X is the probability measure P_X on $(\mathbb{R}, \mathcal{B})$ given by

$$P_X[B] := \mathbb{P}[X \in B] = \mathbb{P}[\{\omega \in \Omega \mid X(\omega) \in B\}]$$

for all $B \in \mathcal{B}$

Exercise Check that P_X is indeed a probability measure on $(\mathbb{R}, \mathcal{B})$.

Proposition For $X : \Omega \to \mathbb{R}$ a random variable with law P_X, and for any $h : \mathbb{R} \to \mathbb{R}$ Borel measurable, we have

$$h(X) \in \mathcal{L}^1(\mathbb{P}) \iff h \in \mathcal{L}^1(P_X),$$

and in this case

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}} h(x) \, dP_X(x).$$

Proof: We put everything to the standard machine.

(a) If $h = 1_B$ for some $B \in \mathcal{B}$ then

$$\mathbb{E}[h(X)] = \mathbb{E}[1_{X^{-1}(B)}] = \mathbb{P}[X^{-1}(B)] = P_X[B].$$

(b) If $h = \sum_{j=1}^{n} a_j \cdot 1_{B_j}$ then

$$\mathbb{E}[h(X)] = \mathbb{E}\Big[\sum_{j=1}^{n} a_j 1_{B_j}\Big] = \sum_{j=1}^{n} a_j \,\mathbb{E}[1_{B_j}]$$

$$\overset{(a)}{=} \sum_{j=1}^{n} a_j \, P_X[B_j] = \int \Big(\sum_{j=1}^{n} a_j 1_{B_j}\Big) dP_X = \int h \, dP_X.$$

(c) If $h_n \uparrow h$ with $h : \mathbb{R} \to [0,+\infty]$ simple, then also $h_n(X) \uparrow h(X)$ and $h_n \circ X : \Omega \to [0,+\infty]$ are simple, so by MCT for $\mathbb{P}$ and P_X we get

$$\mathbb{E}[h(X)] \overset{MCT}{=} \lim_{n\to\infty} \mathbb{E}[h_n(X)]$$

$$\overset{(b)}{=} \lim_{n\to\infty} \int h_n \, dP_X \overset{MCT}{=} \int h \, dP_X.$$

(d) Now for $h : \mathbb{R} \to \mathbb{R}$ general,

$$(h(X))^+ = h^+(X) \quad \text{and} \quad (h(X))^- = h^-(X)$$

and therefore the condition

$$h(X) \in \mathcal{L}^1(\mathbb{P}) \qquad (\mathbb{E}[(h(X))^+] < \infty,\ \mathbb{E}[(h(X))^-] < \infty)$$

is equivalent to the condition

$$h \in \mathcal{L}^1(P_X) \qquad \Big(\int h^+ dP_X < \infty,\ \int h^- dP_X < \infty\Big)$$

by part (b). Also

$$\mathbb{E}[h(X)] := \mathbb{E}[(h(X))^+] - \mathbb{E}[(h(X))^-]$$

is equal to

$$\int h \, dP_X := \int h^+ dP_X - \int h^- dP_X$$

by part (b). $\square$

Recall

(S, Σ, μ) measure space

$$s\Sigma = \{\text{simple functions } S \to \mathbb{R}\}$$
$$= \{ f : S \to \mathbb{R} \mid f = \sum_{j=1}^{n} a_j \cdot 1_{A_j} \text{ for some } a_1, \dots, a_n \in \mathbb{R} \text{ and } A_1, \dots, A_n \in \Sigma \}$$

$$m\Sigma = \{\text{measurable functions } S \to [-\infty, +\infty]\}$$
$$= \{ f : S \to [-\infty, +\infty] \mid \forall B \in \mathcal{B}([-\infty, +\infty]) : f^{-1}(B) \in \Sigma \}$$

By superscript $+$ we denote the corresponding non-negative functions:

$$s\Sigma^{+} = \{ f \in s\Sigma \mid f(s) \geq 0 \quad \forall s \in S \}$$
$$m\Sigma^{+} = \{ f \in m\Sigma \mid f(s) \geq 0 \quad \forall s \in S \}.$$

The integral $\int_S f(s)\, d\mu(s)$ was defined in four stages:

1°) $f = 1_A$, $A \in \Sigma$: $\int f \, d\mu = \mu[A]$

2°) $f \in s\Sigma$, $f = \sum_{j=1}^{n} a_j 1_{A_j}$: $\int f \, d\mu = \sum_{j=1}^{n} a_j \cdot \mu[A_j]$

3°) $f \in m\Sigma^{+}$: $\int f \, d\mu = \sup_{\substack{h \in s\Sigma^{+} \\ h \leq f}} \int h \, d\mu$

4°) $f \in \mathcal{L}^1(S, \Sigma, \mu)$: $\int f \, d\mu = \int f^{+} d\mu - \int f^{-} d\mu$

where $f^{+}(s) = \max(f(s), 0)$
$f^{-}(s) = \max(-f(s), 0)$.

$f \in \mathcal{L}^1(S, \Sigma, \mu) \iff \int f^{+} d\mu < \infty$ and $\int f^{-} d\mu < \infty$

The "standard machine" is the proof technique in which a statement is verified step by step:

1°) $f = 1_A$, 2°) $f \in s\Sigma$, 3°) $f \in m\Sigma^{+}$, 4°) $f \in \mathcal{L}^1(S, \Sigma, \mu)$

PROOF OF THE MONOTONE CONVERGENCE THEOREM

Theorem (Monotone Convergence Theorem)

If $f_1, f_2, \ldots \in m\Sigma^+$ and $f_n \uparrow f$
then $\int f_n \, d\mu \uparrow \int f \, d\mu$.

This is obviously closely related to our earlier result:

Proposition (Monotone Convergence of Measure)

If $A_1, A_2, \ldots \in \Sigma$ and $A_n \uparrow A$,
then $\mu[A_n] \uparrow \mu[A]$.

Our first step is:

Lemma 1: If $A \in \Sigma$ and $h_1, h_2, \ldots \in s\Sigma^+$ are such that $h_n \uparrow \mathbb{1}_A$, then $\int h_n \, d\mu \uparrow \mu[A]$.

Proof: By monotonicity of integral $\int h_n \, d\mu \le \int \mathbb{1}_A \, d\mu = \mu[A]$,
so we only need to prove $\liminf_n \int h_n \, d\mu \ge \mu[A]$.
Let $\varepsilon > 0$ and define $A_n = \{ s \in A \mid h_n(s) > 1 - \varepsilon \}$.
Then $A_n \uparrow A$, so the proposition gives $\mu[A_n] \uparrow \mu[A]$. Note also that
$h_n \ge (1-\varepsilon) \mathbb{1}_{A_n}$ by definition of A_n, so
by monotonicity $\int h_n \, d\mu \ge (1-\varepsilon)\mu[A_n]$.
Taking liminf we get $\liminf_n \int h_n \, d\mu \ge (1-\varepsilon)\mu[A]$
and since $\varepsilon > 0$ was arbitrary
$\liminf_n \int h_n \, d\mu \ge \mu[A]$, which proves the claim. $\square$

The next step is:

Lemma 2: If $h \in s\Sigma^+$ and $h_1, h_2, \ldots \in s\Sigma^+$ are such that $h_n \uparrow h$, then $\int h_n \, d\mu \uparrow \int h \, d\mu$.

Proof: Write $h = \sum_{k=1}^m \alpha_k \mathbb{1}_{A_k}$, where $A_1, \ldots, A_m$ are disjoint, and $\alpha_1, \ldots, \alpha_m > 0$. Then $\frac{1}{\alpha_k} \mathbb{1}_{A_k} h_n \uparrow \mathbb{1}_{A_k}$ as $n \to \infty$
so by Lemma 1 and linearity the assertion follows. $\square$

So far we thus have the MCT when all the functions involved are simple.

The following easy lemma will be used twice.

Lemma A: Let $y_n^{(r)}$, $n \in \mathbb{N}$, $r \in \mathbb{N}$, be an array of numbers in $[0,+\infty]$ which is doubly monotone:

$$\forall r : \quad y_n^{(r)} \uparrow y^{(r)} \quad \text{as } n \to \infty$$

$$\forall n : \quad y_n^{(r)} \uparrow y_n \quad \text{as } r \to \infty .$$

Then $y^{(r)} \uparrow y^{(\infty)}$ as $r \to \infty$ and

$y_n \uparrow y_\infty$ as $n \to \infty$ and $y^{(\infty)} = y_\infty$.

Proof: Assume first that for some $M < \infty$ we have $|y_n^{(r)}| \leq M$ for all r and n ("uniformly bounded").

Let $\varepsilon > 0$. Choose n_0 s.t. $y_{n_0} > y_\infty - \frac{1}{2}\varepsilon$.

Then choose r_0 s.t. $y_{n_0}^{(r_0)} > y_{n_0} - \frac{\varepsilon}{2}$. Then

$y^{(\infty)} \geq y^{(r_0)} \geq y_{n_0}^{(r_0)} > y_\infty - \varepsilon$. Since $\varepsilon > 0$ was arbitrary, we get $y^{(\infty)} \geq y_\infty$.

Similarly $y_\infty \geq y^{(\infty)}$.

If $y_n^{(r)}$ are not uniformly bounded, consider $\tilde{y}_n^{(r)} = \arctan(y_n^{(r)})$ instead. $\square$

We have MCT when all functions are simple, so the next step is to handle general non-negative measurable functions. Note first:

<u>Lemma</u> If $f \in m\Sigma^+$, then there exists a sequence $h_1, h_2, \dots \in s\Sigma^+$ such that $h_n \uparrow f$ and $\int h_n \, d\mu \uparrow \int f \, d\mu$.

<u>Proof</u>: By definition $\int f \, d\mu = \sup \{ \int g \, d\mu \mid g \in s\Sigma^+, g \le f \}$ so there exists some sequence $g_1, g_2, \dots \in s\Sigma^+$ s.t. $g_n \le f$ and $\int g_n \, d\mu \longrightarrow \int f \, d\mu$. Also there exists an approximation $f_1, f_2, \dots \in s\Sigma^+$ such that $f_n \uparrow f$ pointwise. Set

$$h_n = \max(f_n, g_1, g_2, \dots, g_n).$$

Then $h_n \in s\Sigma^+$ and $h_n \le f$. Since $f_n \le h_n \le f$, and $f_n \uparrow f$, we have $h_n \longrightarrow f$ (pointwise). And since $g_n \le h_n \le f$, by monotonicity $\int g_n \, d\mu \le \int h_n \, d\mu \le \int f \, d\mu$ (the latter $\le$ is by def.), and $\int g_n \, d\mu \longrightarrow \int f \, d\mu$, we have $\int h_n \, d\mu \longrightarrow \int f \, d\mu$. It remains to note that by construction $h_1 \le h_2 \le \dots$ and thus in fact $h_n \uparrow f$ and $\int h_n \, d\mu \uparrow \int f \, d\mu$. $\square$

Next note that the choice of the pointwise increasing approximation of f by simple functions does not matter.

<u>Lemma</u> If $f \in m\Sigma^+$ and $g_1, g_2, \dots \in s\Sigma^+$ are such that $g_r \uparrow f$, then $\int g_r \, d\mu \uparrow \int f \, d\mu$.

<u>Proof</u>: Consider $g_1, g_2, \dots$ and use the previous lemma to get also $h_1, h_2, \dots \in s\Sigma^+$ s.t. $h_n \uparrow f$ and $\int h_n \, d\mu \uparrow \int f \, d\mu$.
Let $f_n^{(r)} = \min(g_r, h_n)$, and $y_n^{(r)} = \int f_n^{(r)} \, d\mu$.
Then as $r \to \infty$, $f_n^{(r)} \uparrow h_n$ and as $n \to \infty$, $f_n^{(r)} \uparrow g_r$.
Apply Lemma A: $y_\infty = \int f \, d\mu$ and $\int g_r \, d\mu \uparrow y^{(\infty)} = y_\infty = \int f \, d\mu$. $\square$

We are now ready to prove the MCT.

<u>Proof of Monotone Convergence Theorem</u>:

Let $f_1, f_2, \ldots \in m\Sigma^+$ be such that $f_n \uparrow f$. Then automatically $f \in m\Sigma^+$, and we want to show that $\int f_n \, d\mu \uparrow \int f \, d\mu$.

For each n, let $f_n^{(r)} = \alpha_r \circ f_n$, where α_r is the r^{th} staircase function. Then $f_n^{(r)} \in s\Sigma^+$ and $f_n^{(r)} \uparrow f_n$ as $r \to \infty$. By Lemma 2 we have as $n \to \infty$ $\int f_n^{(r)} d\mu \uparrow \int f^{(r)} d\mu$ where $f^{(r)} = \alpha_r \circ f$. Since $f^{(r)} \in s\Sigma^+$ and $f^{(r)} \uparrow f$ as $r \to \infty$, by the previous lemma we get that $\int f^{(r)} d\mu \uparrow \int f \, d\mu$. If we again denote $y_n^{(r)} := \int f_n^{(r)} d\mu$, then these observations say $y^{(\infty)} = \int f \, d\mu$. Also by the previous lemma $y_n^{(r)} \uparrow y_n = \int f_n \, d\mu$. By Lemma A then $\lim_{n \to \infty} \int f_n \, d\mu = y_\infty = y^{(\infty)} = \int f \, d\mu$. $\square$

We also note that the pointwise increasing approximation $f_n(s) \uparrow f(s) \quad \forall s \in S$ does not have to hold for literally all $s \in S$, but μ-almost all s is enough, i.e.

$$\mu\left[S \setminus \{ s \in S \mid f_n(s) \uparrow f(s) \} \right] = 0 .$$

This is checked as follows:

<u>Lemma</u> If $f, g : S \to [0, +\infty]$ are Σ-measurable, and $\mu[\{ s \in S \mid f(s) \neq g(s) \}]$, then $\int f \, d\mu = \int g \, d\mu$.

"Almost everywhere equal functions have equal integrals"

<u>Proof</u>: Let $f_r = \alpha_r \circ f$, $g_r = \alpha_r \circ g$ for $r \in \mathbb{N}$.
Then $f_r \uparrow f$ and $g_r \uparrow g$ and
$\mu[\{ s \in S \mid f_r(s) \neq g_r(s) \}] = 0$ (i.e., "$f_r = g_r$ μ-a.e.").
Therefore (using Monotone Convergence twice)

$$\int f \, d\mu = \lim_{r \to \infty} \int^{(\text{naive})} f_r \, d\mu = \lim_{r \to \infty} \int^{(\text{naive})} g_r \, d\mu = \int g \, d\mu .$$

$\square$

<u>Corollary</u> (Monotone Convergence Theorem, a.e. version)
If f and $f_1, f_2, \ldots$ are non-negative Σ-measurable functions $S \to [0, +\infty]$, and if $f_n \uparrow f$ μ-almost everywhere, then $\int f_n \, d\mu \uparrow \int f \, d\mu$.

<u>Proof</u>: Let $N = S \setminus \{ s \in S \mid f_n \uparrow f \}$ be the exception set.
By assumption $\mu[N] = 0$. Consider $\tilde{f} = 1_{N^c} \cdot f$ and $\tilde{f}_n = 1_{N^c} \cdot f_n$. We have $f = \tilde{f}$ μ-a.e. and $f_n = \tilde{f}_n$ μ-a.e., and $\tilde{f}_n \uparrow \tilde{f}$. Thus

$$\int f_n \, d\mu = \int \tilde{f}_n \, d\mu \uparrow \int \tilde{f} \, d\mu = \int f \, d\mu .$$

$\square$

MCT

Besides the Monotone convergence theorem, there are also other very practical convergence results that hold for the integrals (now that we have properly constructed the integral).

The following does not even require limits to exist:

<u>Lemma</u> (Fatou's lemma)

For a sequence $f_1, f_2, \ldots \in m\Sigma^+$ we have

$$\int (\liminf_n f_n)\, d\mu \leq \liminf_n \int f_n\, d\mu .$$

<u>Proof</u> Let $g_m = \inf_{k \geq m} f_k$, so $g_m \uparrow g = \liminf_n f_n$ as $m \to \infty$. By MCT: $\int g_m\, d\mu \uparrow \int g\, d\mu$.

Since $f_k \geq g_m$ for any $k \geq m$, we have $\int f_k\, d\mu \geq \int g_m\, d\mu$ and thus $\inf_{k \geq m} \int f_k\, d\mu \geq \int g_m\, d\mu$

Now take limit as $m \to \infty$ to obtain

$$\liminf_n \int f_n\, d\mu \geq \int g\, d\mu . \qquad \square$$

<u>Lemma</u> (Reverse Fatou's lemma)

If $f_1, f_2, \ldots \in m\Sigma^+$ and $f_n \leq g \in m\Sigma^+$ for all n, and $\int g\, d\mu < \infty$, then

$$\int (\limsup f_n)\, d\mu \geq \limsup \int f_n\, d\mu .$$

<u>Proof</u>: Apply Fatou's lemma to $g - f_n$

The following is the most powerful general convergence result:

<u>Theorem</u> (Dominated Convergence Theorem)

Suppose that $f_1, f_2, \ldots \in m\Sigma$ and that for some $g \in m\Sigma^+$ with $\int g \, d\mu < \infty$ we have $|f_n| \leq g$ for all n.
Then if $f_n \longrightarrow f$ (pointwise), we have $\int |f_n - f| \, d\mu \longrightarrow 0$ and $\int f_n \, d\mu \longrightarrow \int f \, d\mu$.

<u>Proof</u>: Note that also $|f| \leq g$, and therefore $|f_n - f| \leq 2g$ (triangle inequality).
Since $\int 2g \, d\mu < \infty$, reverse Fatou's lemma gives
$\limsup\limits_n \int |f_n - f| \, d\mu \leq \int (\limsup\limits_n |f_n - f|) \, d\mu = \int 0 \, d\mu = 0$,
which proves the first claim.
Then use linearity and triangle inequality for integrals to get

$$\left| \int f_n \, d\mu - \int f \, d\mu \right| \leq \left| \int (f_n - f) \, d\mu \right| \leq \int |f_n - f| \, d\mu \longrightarrow 0$$

which proves the second claim. $\square$

MORE ABOUT EXPECTED VALUE

$(\Omega, \mathcal{F}, \mathbb{P})$ proba space

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

(we also allow $X : \Omega \to [-\infty, +\infty]$ i.e. $X \in m\mathcal{F}$)

Expected value: $\mathbb{E}[X] := \int_\Omega X(\omega) \, d\mathbb{P}(\omega)$

Let us summarize basic properties:

(LINEARITY): $\mathbb{E}[aX + bY] = a\,\mathbb{E}[X] + b\,\mathbb{E}[Y]$
for all $a, b \in \mathbb{R}$ and $X, Y \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$

(MONOTONICITY): if $X, Y \in m\mathcal{F}^+$ and
$0 \le X \le Y$, then $0 \le \mathbb{E}[X] \le \mathbb{E}[Y]$

(ALMOST SURE EQUALITY): if $X, Y \in m\mathcal{F}^+$ or $X, Y \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$
and if $X = Y$ almost surely
(i.e. $\mathbb{P}[\{\omega \in \Omega \mid X(\omega) = Y(\omega)\}] = 1$)
then $\mathbb{E}[X] = \mathbb{E}[Y]$

Let us also summarize convergence results:

(MCT): "Monotone Convergence Theorem"
If $X_1, X_2, \ldots \in m\mathcal{F}^+$ and $X_n \uparrow X$
then $\mathbb{E}[X_n] \uparrow \mathbb{E}[X]$

(FATOU): "Fatou's lemma"
If $X_1, X_2, \ldots \in m\mathcal{F}^+$ then $\mathbb{E}[\liminf_n X_n] \le \liminf_n \mathbb{E}[X_n]$

(DCT): "Dominated Convergence Theorem"
If $X_1, X_2, \ldots \in m\mathcal{F}$ and $\exists\, Y \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$
such that $|X_n| \le Y$ for all n, and
if $X_n \to X$ (pointwise), then
$\mathbb{E}[|X_n - X|] \to 0$ and $\mathbb{E}[X_n] \to \mathbb{E}[X]$.

Just a special case of DCT: a constant K is random variable integrable

(BCT): "Bounded Convergence Theorem"
If $X_1, X_2, \ldots \in m\mathcal{F}$ and $\exists\, K < \infty$ s.t. $|X_n| \le K \ \forall n$ and
if $X_n \to X$ then $\mathbb{E}[|X_n - X|] \to 0$ and $\mathbb{E}[X_n] \to \mathbb{E}[X]$.

Some consequences :

- If $X \in m\mathcal{F}^+$ and $\mathbb{E}[X] < \infty$ then $\mathbb{P}[X<\infty] = 1$

Pf: Indeed, let $A = \{\omega \in \Omega \mid X(\omega) = \infty\}$.
Then $X \geq n \cdot 1_A$ for any $n \in \mathbb{N}$ so by monotonicity $\mathbb{E}[X] \geq n \cdot \mathbb{P}[A]$. if $\mathbb{P}[A] > 0$ then letting $n \to \infty$ gives $\mathbb{E}[X] = +\infty$.
Therefore conversely if $\mathbb{E}[X] < \infty$, we must have $\mathbb{P}[A] = 0$, i.e. $\mathbb{P}[A^c] = 1$. $\square$

- If $Z_1, Z_2, \ldots \in m\mathcal{F}^+$ then we have
$$\mathbb{E}\Big[\sum_{k=1}^{\infty} Z_k\Big] = \sum_{k=1}^{\infty} \mathbb{E}[Z_k].$$

Pf: Consider the partial sums $S_n = \sum_{k=1}^{n} Z_k$.
By definition $S_n \uparrow \sum_{k=1}^{\infty} Z_k$ (note $Z_k \geq 0$).
and $\mathbb{E}[S_n] \uparrow \sum_{k=1}^{\infty} \mathbb{E}[Z_k]$ (note $\mathbb{E}[Z_k] \geq 0$ and linearity).
But MCT gives $\mathbb{E}[S_n] \uparrow \mathbb{E}\Big[\sum_{k=1}^{\infty} Z_k\Big]$,
so we must have $\mathbb{E}\Big[\sum_{k=1}^{\infty} Z_k\Big] = \sum_{k=1}^{\infty} \mathbb{E}[Z_k]$. $\square$

- If $Z_1, Z_2, \ldots \in m\mathcal{F}^+$ and $\sum_{k=1}^{\infty} \mathbb{E}[Z_k] < \infty$,
then $\mathbb{P}\Big[\sum_{k=1}^{\infty} Z_k < \infty\Big] = 1$ and $\mathbb{P}[Z_k \to 0] = 1$.

Pf: This follows from the previous two items. $\square$

- The first Borel-Cantelli lemma revisited:
if $E_1, E_2, \ldots \in \mathcal{F}$ are events s.t. $\sum_{k=1}^{\infty} \mathbb{P}[E_k] < \infty$
then $\mathbb{P}[E_k \text{ infinitely often}] = 0$.

Pf: Take $Z_k = 1_{E_k}$ in the previous item.
By assumption $\sum_{k=1}^{\infty} \mathbb{E}[1_{E_k}] = \sum_{k=1}^{\infty} \mathbb{P}[E_k] < \infty$
so $\sum_{k=1}^{\infty} 1_{E_k} < \infty$ almost surely (i.e., with probability 1).
But $\sum_{k=1}^{\infty} 1_{E_k}$ = "number of k such that the event E_k occurs",
so "E_k i.o." $= \{\omega \mid \sum_{i=1}^{\infty} 1_{E_i} = \infty\}$. $\square$

PRODUCTS

Imagine observing two random phenomena simultaneously. The set of possible outcomes is a Cartesian product

$$\Omega = \Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) \mid \omega_1 \in \Omega_1 , \omega_2 \in \Omega_2\}$$

of the sets Ω_1 and Ω_2 of possible outcomes of the two phenomena.

Example: For two coin tosses, the sample space is $\{H,T\} \times \{H,T\} = \{(H,H),(H,T),(T,H),(T,T)\}$.

This part treats measures on spaces that are constructed as Cartesian products. We will in particular study

- the product σ-algebra on a Cartesian product of measurable spaces
- the product measure on a Cartesian product of (σ-finite) measure spaces.

One result of great practical importance is Fubini's theorem: for two measure spaces (S_1, Σ_1, μ_1) and (S_2, Σ_2, μ_2) and a function $f : S_1 \times S_2 \to \mathbb{R}$, under reasonable conditions the change of order of integration formula

$$\int_{S_1} \Big(\int_{S_2} f(s_1,s_2)\, d\mu_2(s_2) \Big) d\mu_1(s_1) = \int_{S_2} \Big(\int_{S_1} f(s_1,s_2)\, d\mu_1(s_1) \Big) d\mu_2(s_2)$$

holds. Recalling that summation and expectation are integrations, this implies (under reasonable conditions) formulas such as

$$E\Big[\sum_{k=1}^{\infty} X_k\Big] = \sum_{k=1}^{\infty} E[X_k], \quad \int \sum_{k=1}^{\infty} f_k(x)\, d\mu(x) = \sum_{k=1}^{\infty} \int f_k(x)\, d\mu(x), \text{ etc.}$$

Product sigma-algebra

Let (S_1, Σ_1) and (S_2, Σ_2) be measurable spaces. We will first have to equip the Cartesian product

$$S_1 \times S_2 = \{ (s_1, s_2) \mid s_1 \in S_1 , \ s_2 \in S_2 \}$$

with a σ-algebra. The coordinate projections

$$\pi_1 : S_1 \times S_2 \longrightarrow S_1 \qquad \pi_1(s_1, s_2) = s_1$$
$$\pi_2 : S_1 \times S_2 \longrightarrow S_2 \qquad \pi_2(s_1, s_2) = s_2$$

are considered fundamental, and we have to require that they are measurable functions to (S_1, Σ_1) and (S_2, Σ_2), respectively.

__Def.__ The product σ-algebra $\Sigma_1 \otimes \Sigma_2$ on $S_1 \times S_2$ is the σ-algebra generated by π_1 and π_2, i.e. the smallest σ-algebra on $S_1 \times S_2$ with respect to which the coordinate projections $\pi_1 : S_1 \times S_2 \to S_1$ and $\pi_2 : S_1 \times S_2 \longrightarrow S_2$ are measurable.

The following lemma says that products of any measurable sets are measurable, and that $\Sigma_1 \otimes \Sigma_2$ is the smallest σ-algebra with this property.

__Lemma__ The collection

$$\mathcal{J} = \{ B_1 \times B_2 \mid B_1 \in \Sigma_1 , \ B_2 \in \Sigma_2 \}$$

is a π-system on $S_1 \times S_2$ and

$$\Sigma_1 \otimes \Sigma_2 = \sigma(\mathcal{J}) .$$

__Proof__ Easy exercise. $\square$

Exercise Show that in $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$, the product σ-algebra $\mathcal{B} \otimes \mathcal{B}$ (of two Borel σ-algebras $\mathcal{B}$ on $\mathbb{R}$) coincides with the Borel σ-algebra $\mathcal{B}(\mathbb{R}^2)$ (the smallest σ-alg. on $\mathbb{R}^2$ which contains all open subsets of the plane).

We will use repeatedly the Monotone Class Theorem, so let us recall its statement:

Theorem (Monotone Class Theorem)

Let S be a set, and $\mathcal{H}$ a class of bounded functions $S \to \mathbb{R}$ such that

(i) $\mathcal{H}$ is a vector space (over $\mathbb{R}$)

(ii) the constant function 1 is in $\mathcal{H}$

(iii) whenever $f_1, f_2, \ldots \in \mathcal{H}$ and $0 \le f_n \uparrow f$ where f is bounded, then $f \in \mathcal{H}$.

Then if $\mathcal{H}$ contains the indicator function 1_A for each set A in some π-system $\mathcal{I}$, then $\mathcal{H}$ contains all bounded $\sigma(\mathcal{I})/\mathcal{B}$-measurable functions

For brevity, let us denote

$$b\Sigma = \{ f : S \to \mathbb{R} \mid f \text{ is bounded } \Sigma/\mathcal{B}\text{-measurable} \}$$

and as before

$$m\Sigma = \{ f : S \to [-\infty, +\infty] \mid f \text{ is } \Sigma/\mathcal{B}([-\infty,+\infty])\text{-meas.} \}$$

and use the superscript $+$ to indicate non-negative func.

$$m\Sigma^+ = \{ f \in m\Sigma \mid f(s) \ge 0 \quad \forall s \in S \}$$

$$b\Sigma^+ = \{ f \in b\Sigma \mid f(s) \ge 0 \quad \forall s \in S \}.$$

Our first use of the Monotone Class Theorem is:

<u>Lemma</u> Let $\mathcal{H}$ denote the class of functions $f: S_1 \times S_2 \to \mathbb{R}$ such that $f \in b(\Sigma_1 \otimes \Sigma_2)$ and

$\forall s_1 \in S_1$: $s_2 \mapsto f(s_1, s_2)$ is Σ_2-meas. $S_2 \to \mathbb{R}$

$\forall s_2 \in S_2$: $s_1 \mapsto f(s_1, s_2)$ is Σ_1-meas. $S_1 \to \mathbb{R}$

Then $\mathcal{H} = b(\Sigma_1 \otimes \Sigma_2)$.

In other words, from any function $f: S_1 \times S_2 \to \mathbb{R}$ which is measurable w.r.t. product σ-algebra, we get measurable functions $S_1 \to \mathbb{R}$ and $S_2 \to \mathbb{R}$ by "freezing" one of the coordinates.

<u>Proof</u>: Clearly (i), (ii), (iii) of the Monotone Class Theorem hold, and $\mathcal{H}$ contains indicator functions of sets in the π-system

$$\mathcal{I} = \{ B_1 \times B_2 \mid B_1 \in \Sigma_1, B_2 \in \Sigma_2 \}$$

which generates $\Sigma_1 \otimes \Sigma_2$. $\square$

Product measure

Suppose that (S_1, Σ_1, μ_1) and (S_2, Σ_2, μ_2) are <u>finite</u> measure spaces, for example probability spaces (finite means $\mu_1[S_1] < +\infty$, $\mu_2[S_2] < +\infty$)

We will define a measure $\mu_1 \otimes \mu_2$ on $(S_1 \times S_2, \Sigma_1 \otimes \Sigma_2)$.

Suppose that $f \in b(\Sigma_1 \otimes \Sigma_2)$.
Then we know that for every $s_1 \in S_1$ the function $s_2 \mapsto f(s_1, s_2)$ is Σ_2-measurable, and it is also integrable since it is bounded in a bounded measure space, so we may define a function $I_1^f : S_1 \to \mathbb{R}$ by

$$I_1^f(s_1) := \int_{S_2} f(s_1, s_2)\, d\mu_2(s_2).$$

Similarly we define a function $I_2^f : S_2 \to \mathbb{R}$ by

$$I_2^f(s_2) := \int_{S_1} f(s_1, s_2)\, d\mu_1(s_1).$$

Lemma For any $f \in b(\Sigma_1 \otimes \Sigma_2)$ the function $I_1 : S_1 \to \mathbb{R}$ is Σ_1-measurable, and the function $I_2 : S_2 \to \mathbb{R}$ is Σ_2-measurable and

$$\int_{S_1} I_1(s_1)\, d\mu_1(s_1) = \int_{S_2} I_2(s_2)\, d\mu_2(s_2).$$

In other words, for any $f : S_1 \times S_2 \to \mathbb{R}$ bounded and measurable (w.r.t. product σ-alg.) we have

$$\int_{S_1} \left(\int_{S_2} f(s_1, s_2)\, d\mu_2(s_2) \right) d\mu_1(s_1) = \int_{S_2} \left(\int_{S_1} f(s_1, s_2)\, d\mu_1(s_1) \right) d\mu_2(s_2)$$

and both sides of the equation make sense.

Proof: Let $\mathcal{H}$ be the class of those $f \in b(\Sigma_1 \otimes \Sigma_2)$ for which $I_1 \in b\Sigma_1$, $I_2 \in b\Sigma_2$, and $\int I_1\, d\mu_1 = \int I_2\, d\mu_2$. Then clearly $\mathcal{H}$ contains the indicator functions of the π-system $\mathcal{I}$: namely for $f(s_1, s_2) = 1_{B_1 \times B_2}(s_1, s_2) = 1_{B_1}(s_1)\, 1_{B_2}(s_2)$ we have for example

$$I_1^f(s_1) = \int 1_{B_1}(s_1)\, 1_{B_2}(s_2)\, d\mu_2(s_2) = 1_{B_1}(s_1) \int 1_{B_2}\, d\mu_2$$
$$= \mu_2[B_2] \cdot 1_{B_1}(s_1),$$

and similarly $I_2^f = \mu_1[B_1] \cdot 1_{B_2}$, so indeed $I_1^f \in b\Sigma_1$ and $I_2^f \in b\Sigma_2$ and

$$\int_{S_2} I_2^f \, d\mu_2 = \mu_1[B_1] \cdot \int_{S_2} 1_{B_2} \, d\mu_2 = \mu_1[B_1] \cdot \mu_2[B_2]$$

and $$\int_{S_1} I_1^f \, d\mu_1 = \mu_2[B_2] \cdot \int_{S_1} 1_{B_1} \, d\mu_1 = \mu_2[B_2] \cdot \mu_1[B_1].$$

We show that $\mathcal{H}$ is a monotone class, which then implies $\mathcal{H} = b(\Sigma_1 \otimes \Sigma_2)$ (because $\mathcal{I}$ generates $\Sigma_1 \otimes \Sigma_2$), and concludes the proof.

Properties (i) and (ii) are obvious for $\mathcal{H}$.

For property (iii), suppose $f_1, f_2, \ldots \in \mathcal{H} \cap b(\Sigma_1 \otimes \Sigma_2)^+$ and $f_n \uparrow f \in b(\Sigma_1 \otimes \Sigma_2)^+$. Then for each $n \in \mathbb{N}$ $I_1^{f_n} \in b\Sigma_1^+$ and by Monotone Convergence Theorem for μ_2 we have

$$I_1^{f_n}(s_1) = \int_{S_2} f_n(s_1, s_2) \, d\mu_2(s_2) \uparrow \int_{S_2} f(s_1, s_2) d\mu_2(s_2) = I_1^f(s_1)$$

and similarly $I_2^{f_n}(s_2) \uparrow I_2^f(s_2)$ by MCT for μ_1.

Thus $I_1^f : S_1 \to \mathbb{R}$ and $I_2^f : S_2 \to \mathbb{R}$ are measurable, as pointwise limits of the measurable functions $I_1^{f_n}$ and $I_2^{f_n}$, respectively.

Both are also bounded, since f is bounded and μ_1 and μ_2 are finite measures.

So $I_1^f \in b\Sigma_1$ and $I_2^f \in b\Sigma_2$. Finally by Monotone Convergence Thm for μ_1 and μ_2

$$\begin{aligned}
\int_{S_1} I_1^f \, d\mu_1 &= \lim_{n \to \infty} \int_{S_1} I_1^{f_n} \, d\mu_1 && (\mu_1\text{-MCT}) \\
&= \lim_{n \to \infty} \int_{S_2} I_2^{f_n} \, d\mu_2 && (f_n \in \mathcal{H}) \\
&= \int I_2^f \, d\mu_2 . && (\mu_2\text{-MCT})
\end{aligned}$$

This establishes property (iii). $\square$

This lemma enables the following unambiguous definition of the product measure:

<u>Def</u>: The product measure $\mu_1 \otimes \mu_2$ is the measure on $(S_1 \times S_2, \Sigma_1 \otimes \Sigma_2)$ given by

$$(\mu_1 \otimes \mu_2)[A] = \int_{S_1} I_1^{1_A} \, d\mu_1 = \int_{S_2} I_2^{1_A} \, d\mu_2 .$$

<u>Lemma</u> $\mu_1 \otimes \mu_2$ is indeed a measure.

<u>Proof</u>: Clearly $(\mu_1 \otimes \mu_2)[A] \geq 0 \quad \forall A \in \Sigma_1 \otimes \Sigma_2$ and $(\mu_1 \otimes \mu_2)[\phi] = 0$. Countable additivity for disjoint sets $A_1, A_2, \dots \in \Sigma_1 \otimes \Sigma_2$ follows by Monotone Convergence and linearity of integrals. Indeed, set $B_n = A_1 \cup A_2 \cup \dots \cup A_n$. Then $B_n \uparrow B = \bigcup_{k \in \mathbb{N}} A_k$. Also $1_{B_n} \uparrow 1_B$

so by MCT for μ_2 we have

$I_1^{1_{B_n}} \uparrow I_1^{1_B}$ and then by MCT for μ_1

$$(\mu_1 \otimes \mu_2)[B_n] = \int_{S_1} I_1^{1_{B_n}} \, d\mu_1 \uparrow \int_{S_1} I_1^{1_B} \, d\mu_1 = (\mu_1 \otimes \mu_2)[B].$$

It remains to note that

$1_{B_n} = 1_{A_1} + \dots + 1_{A_n}$ by disjointness, so linearity gives $I_1^{1_{B_n}} = I_1^{1_{A_1}} + \dots + I_1^{1_{A_n}}$ and again

$$(\mu_1 \otimes \mu_2)[B_n] = \int_{S_1} (I_1^{1_{A_1}} + \dots + I_1^{1_{A_n}}) \, d\mu_1$$

$$= \sum_{k=1}^{n} (\mu_1 \otimes \mu_2)[A_k] \uparrow \sum_{k=1}^{\infty} (\mu_1 \otimes \mu_2)[A_k] .$$

We have obtained two expressions for $\lim_{n \to \infty} (\mu_1 \otimes \mu_2)[B_n]$, and their equality is exactly countable additivity for $\mu_1 \otimes \mu_2$. $\square$

We next show that in many cases the following three expressions are all equal:

1°) $\displaystyle\int_{S_1\times S_2} f \, d(\mu_1\otimes\mu_2)$

2°) $\displaystyle\int_{S_1}\left(\int_{S_2} f(s_1,s_2)\, d\mu_2(s_2)\right) d\mu_1(s_1)$

3°) $\displaystyle\int_{S_2}\left(\int_{S_1} f(s_1,s_2)\, d\mu_1(s_1)\right) d\mu_2(s_2)$

This is very useful in practise, because depending on the situation, one is often much easier to calculate than the other two.

<u>Theorem</u> (Fubini's Theorem) (or Tonelli–Fubini Theorem)

If $f\in m(\Sigma_1\otimes\Sigma_2)^+$, then 1°, 2°, 3° are all equal (either all finite and equal or all equal to $+\infty$). Also if $f\in\mathcal{L}^1(S_1\times S_2, \Sigma_1\otimes\Sigma_2, \mu_1\otimes\mu_2)$ then 1°, 2°, 3° are all equal and finite.

<u>Proof</u>: For $f = 1_{B_1\times B_2}$ with $B_1\in\Sigma_1$, $B_2\in\Sigma_2$ we have 1°, 2°, 3° all equal to $\mu_1[B_1]\mu_2[B_2]$. It is then easy to use the Monotone Class Theorem to show that 1°, 2°, 3° are equal for all $f\in b(\Sigma_1\otimes\Sigma_2)$ (we actually proved 2° = 3° in this case already before). In particular 1° = 2° = 3° for all non-negative simple functions $f\in s(\Sigma_1\otimes\Sigma_2)^+$. But then for $f\in m(\Sigma_1\otimes\Sigma_2)^+$ we can approximate by non-neg. simple functions in a pointwise increasing way. Using Monotone Convergence in 1°, 2°, and 3° shows that they are all equal for $f\in m(\Sigma_1\otimes\Sigma_2)^+$, too.

For the case of $f \in \mathcal{L}^1(S_1 \times S_2, \Sigma_1 \otimes \Sigma_2, \mu_1 \otimes \mu_2)$ just consider the non-negative function $|f| = f^+ + f^-$ and use the first part to conclude that $I_1^f \in \mathcal{L}^1(S_1, \Sigma_1, \mu_1)$ and $I_2^f \in \mathcal{L}^1(S_2, \Sigma_2, \mu_2)$, and then use $f = f^+ - f^-$ and linearity. $\square$

<u>Remark</u>: The product measure satisfies $(\mu_1 \otimes \mu_2)[B_1 \times B_2] = \mu_1[B_1]\, \mu_2[B_2]$. Since the collection $\mathcal{I}$ of sets of this form is a π-system, it immediately follows from Dynkin's identification theorem that this property uniquely determines $\mu_1 \otimes \mu_2$. The work we did above was thus only for existence of the product measure, and for the important Fubini's theorem.

We finally note that the construction above and Fubini's theorem work with milder assumptions than finiteness of μ_1 and μ_2.

<u>Def</u>: A measure space (S, Σ, μ) is <u>σ-finite</u> if there exists a sequence $A_1, A_2, \ldots \in \Sigma$ such that $\mu[A_n] < +\infty$ $\forall n$ and $\bigcup_{n \in \mathbb{N}} A_n = S$.

<u>Remark</u>: By setting $A_n' = A_n \setminus (A_1 \cup A_2 \cup \cdots \cup A_{n-1})$ we get disjoint sets $A_1', A_2', \ldots \in \Sigma$ such that $\mu[A_n'] < +\infty$ and $\bigcup_{n \in \mathbb{N}} A_n' = S$.

<u>Example</u> $(\mathbb{R}, \mathcal{B}, \mathrm{Leb})$ is σ-finite, take e.g. $A_n = [-n, n]$.

Theorem (Fubini's theorem, σ-finite case)

Let (S_1, Σ_1, μ_1) and (S_2, Σ_2, μ_2) be two σ-finite measure spaces. Then there exists a unique measure $\mu_1 \otimes \mu_2$ on $(S_1 \times S_2, \Sigma_1 \otimes \Sigma_2)$ such that

$$(\mu_1 \otimes \mu_2)[B_1 \times B_2] = \mu_1[B_1]\, \mu_2[B_2] \quad \forall B_1 \in \Sigma_1, B_2 \in \Sigma_2 .$$

Moreover, for any $f \in m(\Sigma_1 \otimes \Sigma_2)^+$ we have

$$\int_{S_1 \times S_2} f \, d(\mu_1 \otimes \mu_2) = \int_{S_1} \left(\int_{S_2} f(s_1, s_2)\, d\mu_2(s_2) \right) d\mu_1(s_1)$$

$$= \int_{S_2} \left(\int_{S_1} f(s_1, s_2)\, d\mu_1(s_1) \right) d\mu_2(s_2) .$$

The equalities also hold for any $f \in \mathcal{L}^1(S_1 \times S_2, \Sigma_1 \otimes \Sigma_2, \mu_1 \otimes \mu_2)$.

Proof: Split both S_1 and S_2 to countably many pieces of finite measure, and use the previous result. $\square$

Example application in probability

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $X : \Omega \to [0, +\infty]$ a non-neg. random variable. Consider the set

$$A = \{ (\omega, x) \mid \omega \in \Omega, \ 0 \le x \le X(\omega) \} \subset \Omega \times \mathbb{R}.$$

Remark $A \in \mathcal{F} \otimes \mathcal{B}$, which can be proven as follows.

If $X \in s\mathcal{F}^+$ then $X = \sum_{k=1}^n \alpha_k \cdot 1_{E_k}$ for $E_1, \ldots, E_n$ disjoint. Then we have

$$A = \bigcup_{k=1}^n \left(E_k \times [0, \alpha_k] \right) \quad \text{and} \quad E_k \times [0, \alpha_k] \in \mathcal{J} \subset \mathcal{F} \otimes \mathcal{B}.$$

If $X \in m\mathcal{F}^+$ is general, then by an increasing approximation by simple functions we can write A as the increasing limit (thus a countable union) of sets of the above form.

Consider then the product measure $\mathbb{P} \otimes \mathrm{Leb}$ on $\Omega \times \mathbb{R}$ (Leb denotes the Lebesgue measure on $\mathbb{R}$).

For the function $f = 1_A$ we have

$$\text{(for } x \ge 0\text{)} \quad I_2^f(x) = \int_\Omega 1_A(\omega, x) \, d\mathbb{P}(\omega) = \mathbb{P}[X \ge x] \quad \text{and}$$

$$I_1^f(\omega) = \int_{\mathbb{R}} 1_A(\omega, x) \, dx = \int_0^{X(\omega)} 1 \cdot dx = X(\omega).$$

The second and third expressions in Fubini's theorem are

$$\int_\Omega I_1^f(\omega) \, d\mathbb{P}(\omega) = \int_\Omega X(\omega) \, d\mathbb{P}(\omega) = \mathbb{E}[X] \quad \text{and}$$

$$\int_{\mathbb{R}} I_2^f(x) \, dx = \int_0^\infty \mathbb{P}[X \ge x] \, dx.$$

The equality guaranteed by Fubini's theorem thus gives the occasionally very useful formula

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}[X \ge x] \, dx \quad \text{for non-negative random variables } X.$$

LAWS, JOINT LAWS, AND DENSITIES

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.
Recall that the law of a random variable $X : \Omega \to \mathbb{R}$ is the proba measure P_X on $(\mathbb{R}, \mathcal{B})$ given by $P_X[B] = \mathbb{P}[X \in B]$.

Def: The law of X is said to have a density $f_X : \mathbb{R} \to [0, +\infty]$, if f_X is a Borel function such that for all $B \in \mathcal{B}$ we have

$$\mathbb{P}[X \in B] = P_X[B] = \int_B f_X(x)\, dx$$

dx stands for integration w.r.t. Lebesgue measure short for $d\,\mathrm{Leb}(x)$

We use the notation

$$\int_A f(s)\, d\mu(s) := \int_S 1_A(s) \cdot f(s)\, d\mu(s)$$

for "integration over a subset $A \subset S$"

Recall If $h : \mathbb{R} \to \mathbb{R}$ is Borel, then

$$h \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}, P_X) \iff h(X) \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$$

and if either (then both) of these hold, then

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}} h(x)\, dP_X(x).$$

Exercise If $h \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}, P_X)$ and $X : \Omega \to \mathbb{R}$ has a density $f_X : \mathbb{R} \to [0, +\infty]$, then

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}} h(x) f_X(x)\, dx.$$

Recall: $\mathcal{B}(\mathbb{R}^2) = \mathcal{B} \otimes \mathcal{B}$, π-system $\mathcal{J} = \{B_1 \times B_2 \mid B_1, B_2 \in \mathcal{B}\}$

Def: If X and Y are random variables $\Omega \to \mathbb{R}$, then the joint law of X and Y is the proba measure $P_{X,Y}$ on $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ given by
$$P_{X,Y}[A] = \mathbb{P}[(X,Y) \in A]$$
for all Borel subsets $A \subset \mathbb{R}^2$.

The pair (X,Y) is a "random vector" in $\mathbb{R}^2$, or more precisely $\omega \longmapsto (X(\omega), Y(\omega)) = Z(\omega)$ $Z : \Omega \to \mathbb{R}^2$ is. Note that this is $\mathcal{F}/\mathcal{B}(\mathbb{R}^2)$-measurable, and thus really a random variable with values in $\mathbb{R}^2$. To check measurability note first that for $B_1, B_2 \in \mathcal{B}$ we have
$$\begin{aligned} Z^{-1}(B_1 \times B_2) &= \{\omega \in \Omega \mid Z(\omega) \in B_1 \times B_2\} \\ &= \{\omega \in \Omega \mid X(\omega) \in B_1 \text{ and } Y(\omega) \in B_2\} \\ &= X^{-1}(B_1) \cap Y^{-1}(B_2) \end{aligned}$$
which is an event by the measurability of X and Y. But since the collection $\mathcal{J} = \{B_1 \times B_2 \mid B_1, B_2 \in \mathcal{B}\}$ generates $\mathcal{B} \otimes \mathcal{B} = \mathcal{B}(\mathbb{R}^2)$, this is sufficient for the measurability of $Z : \Omega \to \mathbb{R}^2$.

The Lebesgue measure Leb^2 on $\mathbb{R}^2$ is the product $\mathrm{Leb} \otimes \mathrm{Leb}$ of two Lebesgue measures Leb on $\mathbb{R}$. It corresponds to "area measure": It is determined by e.g. $\mathrm{Leb}^2([a_1,b_1] \times [a_2,b_2]) = (b_1-a_1)(b_2-a_2)$ for all rectangles.

Def: The random variables X and Y have a joint density $f_{X,Y}$ is $f_{X,Y} : \mathbb{R}^2 \to [0,+\infty]$ is a Borel function such that $\forall\, A \in \mathcal{B}(\mathbb{R}^2)$
$$\mathbb{P}[(X,Y) \in A] = P_{X,Y}[A] = \int_A f_{X,Y}(z)\, d(\mathrm{Leb}^2)(z)$$

Proposition If X and Y have a joint density $f_{X,Y} : \mathbb{R}^2 \to [0,+\infty]$, then X has a density $f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y)\, dy$ and Y has a density $f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x,y)\, dx$.

Proof This is a simple consequence of Fubini's theorem. Let $B \in \mathcal{B}$. Write

$$P_X[B] = P[X \in B] = P[X \in B \text{ and } Y \in \mathbb{R}]$$
$$= P[(X,Y) \in B \times \mathbb{R}]$$
$$= \int_{B \times \mathbb{R}} f_{X,Y}(x,y)\, d\,\mathrm{Leb}^2(x,y)$$

Remark: $\mathbb{1}_{B\times\mathbb{R}}(x,y) = \mathbb{1}_B(x)$

$$= \int_{\mathbb{R}^2} \mathbb{1}_{B\times\mathbb{R}}(x,y) f_{X,Y}(x,y)\, d\,\mathrm{Leb}^2(x,y)$$

Fubini for $\mathrm{Leb} \otimes \mathrm{Leb}$

$$= \int_{\mathbb{R}} \underbrace{\left(\int_{\mathbb{R}} \mathbb{1}_B(x) f_{X,Y}(x,y)\, dy \right)}_{= \mathbb{1}_B(x) \cdot \int_{\mathbb{R}} f_{X,Y}(x,y)\, dy \,=\, \mathbb{1}_B(x) f_X(x)} dx$$

$$= \int_{\mathbb{R}} \mathbb{1}_B(x) f_X(x)\, dx = \int_B f_X(x)\, dx$$

The calculation for $P_Y[B] = \int_B f_Y(y)\, dy$ is similar. $\square$

Independence and products

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a proba space and $X, Y : \Omega \to \mathbb{R}$ two random variables. Denote the laws of X and Y by P_X and P_Y, respectively, and the joint law of (X,Y) by $P_{X,Y}$.

Proposition (Independence and laws)

The following conditions are equivalent

(a) $X \perp\!\!\!\perp Y$

(b) $P_{X,Y} = P_X \otimes P_Y$

(c) $\forall x, y \in \mathbb{R}$: $\mathbb{P}[X \leq x, Y \leq y] = \mathbb{P}[X \leq x] \cdot \mathbb{P}[Y \leq y]$.

If moreover (X,Y) has a joint density $f_{X,Y} : \mathbb{R}^2 \to [0, +\infty]$ then (a), (b), (c) are equivalent to

(d) X and Y have densities f_X and f_Y such that $f_{X,Y}(x,y) = f_X(x) f_Y(y)$ for Leb^2 - almost every (x,y).

Proof: "(a) $\Rightarrow$ (b)": Suppose $X \perp\!\!\!\perp Y$. Let $B_1, B_2 \in \mathcal{B}$.

Then $X^{-1}(B_1) \in \sigma(X)$ and $Y^{-1}(B_2) \in \sigma(Y)$.

By independence

$$P_{X,Y}[B_1 \times B_2] = \mathbb{P}[X \in B_1 \text{ and } Y \in B_2]$$
$$= \mathbb{P}[X^{-1}(B_1) \cap Y^{-1}(B_2)]$$
$$\overset{\perp\!\!\!\perp}{=} \mathbb{P}[X^{-1}(B_1)] \cdot \mathbb{P}[Y^{-1}(B_2)] = P_X[B_1] \cdot P_Y[B_2].$$

The product measure is uniquely determined by this condition, so $P_{X,Y} = P_X \otimes P_Y$.

"(b) $\Rightarrow$ (c)": If $P_{X,Y} = P_X \otimes P_Y$ then considering $B_1 = (-\infty, x]$ and $B_2 = (-\infty, y]$ we see that

$$\mathbb{P}[X \leq x, Y \leq y] = P_{X,Y}[B_1 \times B_2] = P_X[B_1] P_Y[B_2] = \mathbb{P}[X \leq x] \mathbb{P}[Y \leq y]$$

"(c) $\Rightarrow$ (a)": We have shown this before: for checking independence, the π-systems $\{X^{-1}((-\infty,x]) \mid x\in\mathbb{R}\}$ and $\{Y^{-1}((-\infty,y]) \mid y\in\mathbb{R}\}$ are enough.

This shows that (a), (b), (c) are equivalent.
If furthermore (X,Y) has joint density $f_{X,Y}$, then we show that (d) $\Rightarrow$ (c) and (b) $\Rightarrow$ (d).

"(d) $\Rightarrow$ (c)": $\mathbb{P}[X\le x_0, Y\le y_0] = \int_{(-\infty, \quad \times(-\infty,} f_{X,Y}(x,y)\, d\,\mathrm{Leb}^2(x,y)$

$$= \int_{\mathbb{R}^2} \underbrace{1_{(-\infty,x_0]\times(-\infty,y_0]}(x,y)}_{=1_{(-\infty,x_0]}(x)\cdot 1_{(-\infty,y_0]}(y)}\, f_{X,Y}(x,y)\, d\,\mathrm{Leb}^2(x,y)$$

$= f_X(x) f_Y(y)$ a.e. by assumption

$$\overset{\text{Fubini}}{=} \int_{\mathbb{R}} \Big(1_{(-\infty,x_0]}(x)\cdot f_X(x)\cdot \underbrace{\int_{\mathbb{R}} 1_{(-\infty,y_0]}(y)\, f_Y(y)\, dy}_{=\mathbb{P}[Y\le y_0]}\Big)\, dx$$

$$= \mathbb{P}[X\le x_0]\cdot \mathbb{P}[Y\le y_0].$$

"(b) $\Rightarrow$ (d)": Suppose $P_{X,Y} = P_X\otimes P_Y$. Recall that
X has density $f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y)\,dy$, and
Y has density $f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x,y)\,dx$.

Define another measure μ on $\mathbb{R}^2$ by
$$\mu[A] = \int_A f_X(x)\cdot f_Y(y)\; d\,\mathrm{Leb}^2(x,y).$$
(It is easy to check that this is a measure)

By Fubini, it is easy to show that
$\mu[B_1\times B_2] = (P_X\otimes P_Y)[B_1\times B_2] = P_{X,Y}[B_1\times B_2]$
for all $B_1, B_2\in\mathcal{B}$. Therefore $\mu = P_{X,Y}$.

Consider the set $E_n = \{(x,y)\in\mathbb{R}^2 \mid f_{X,Y}(x,y) \ge f_X(x) f_Y(y) + \frac{1}{n}\}$.

Then $0 = P_{X,Y}[E_n] - \mu[E_n] = \int_{E_n} (f_{X,Y}(x,y) - f_X(x) f_Y(y))\, d\,\mathrm{Leb}^2(x,y)$
$\ge \frac{1}{n}\cdot \mathrm{Leb}^2[E_n]$ so $\mathrm{Leb}^2[E_n] = 0$.

Thus $\mathrm{Leb}^2\{f_{X,Y} > f_X\cdot f_Y\} = \mathrm{Leb}^2\big(\bigcup_{n\in\mathbb{N}} E_n\big) = 0$. Similarly the other way. $\square$

Theorem Suppose that $X, Y : \Omega \to \mathbb{R}$ are two independent random variables, and $X, Y \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. Then also their product $X \cdot Y \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y].$$

Proof: Assume first that X and Y are non-negative. Let $\alpha_r : [0, +\infty] \to [0, r]$ be the r:th staircase function so that $\alpha_r(X) \uparrow X$ and $\alpha_r(Y) \uparrow Y$ are increasing approximations by non-neg. simple random variables,

$$\alpha_r(X) = \sum_{i=1}^{r \cdot 2^r} c_i^{(r)} \cdot 1_{A_i^{(r)}}, \qquad \alpha_r(Y) = \sum_{j=1}^{r \cdot 2^r} c_j^{(r)} \cdot 1_{B_j^{(r)}}$$

where $A_i^{(r)} \in \sigma(X)$, $B_j^{(r)} \in \sigma(Y)$, and $c_j^{(r)} = j \cdot 2^{-r}$.

For the simple approximations we calculate

$$\begin{aligned}
\mathbb{E}[\alpha_r(X) \cdot \alpha_r(Y)] &= \sum_i \sum_j c_i^{(r)} c_j^{(r)} \, \mathbb{P}[A_i^{(r)} \cap B_j^{(r)}] && \text{(linearity)} \\
&= \sum_i \sum_j c_i^{(r)} c_j^{(r)} \, \mathbb{P}[A_i^{(r)}] \, \mathbb{P}[B_j^{(r)}] && \text{(independence)} \\
&= \mathbb{E}[\alpha_r(X)] \cdot \mathbb{E}[\alpha_r(Y)].
\end{aligned}$$

Letting $r \to \infty$ and using Monotone Convergence Theorem shows $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$.

If X, Y are not necessarily non-negative, write $X = X^+ - X^-$, $Y = Y^+ - Y^-$ and note that e.g. $\mathbb{E}[|X \cdot Y|] = \mathbb{E}[|X| \cdot |Y|] = \mathbb{E}[|X|] \cdot \mathbb{E}[|Y|]$ since $|X| \perp\!\!\!\perp |Y|$, too. Thus $XY \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$.

Finally,

$$\begin{aligned}
\mathbb{E}[XY] &= \mathbb{E}[(X^+ - X^-)(Y^+ - Y^-)] \\
&= \mathbb{E}[X^+ Y^+] + \mathbb{E}[X^+ Y^-] + \mathbb{E}[X^- Y^+] + \mathbb{E}[X^- Y^-] \\
&= \mathbb{E}[X^+]\mathbb{E}[Y^+] + \ldots + \mathbb{E}[X^-] \cdot \mathbb{E}[Y^-] \\
&= (\mathbb{E}[X^+] - \mathbb{E}[X^-]) \cdot (\mathbb{E}[Y^+] - \mathbb{E}[Y^-]) = \mathbb{E}[X]\mathbb{E}[Y]. \qquad \square
\end{aligned}$$

Def: We denote $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ if $X \in m\mathcal{F}$ and $\mathbb{E}[X^2] < +\infty$, and call such X *square integrable*.

Below, denote briefly $\mathcal{L}^1 = \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{L}^2 = \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$.

Theorem (a) If $X, Y \in \mathcal{L}^2$ then $X \cdot Y \in \mathcal{L}^1$ and we have $|\mathbb{E}[XY]| \leq \sqrt{\mathbb{E}[X^2] \cdot \mathbb{E}[Y^2]}$. (Cauchy-Schwarz inequality)

(b) We have $\mathcal{L}^2 \subset \mathcal{L}^1$ and if $X \in \mathcal{L}^2$ then $\mathbb{E}[X]^2 \leq \mathbb{E}[X^2]$

(c) $\mathcal{L}^2$ in a vector space.

Proof (a) We have $|XY| \leq \frac{1}{2}X^2 + \frac{1}{2}Y^2$ so $X, Y \in \mathcal{L}^2 \implies XY \in \mathcal{L}^1$.

$0 \leq (x-y)^2 = x^2 - 2xy + y^2$
$\implies xy \leq \frac{1}{2}x^2 + \frac{1}{2}y^2$
$0 \leq (x+y)^2 = x^2 + 2xy + y^2$
$\implies -xy \leq \frac{1}{2}x^2 + \frac{1}{2}y^2$
$\therefore |xy| \leq \frac{1}{2}x^2 + \frac{1}{2}y^2$

Let $t \in \mathbb{R}$. Then

$$0 \leq \mathbb{E}[(tX + Y)^2] = t^2 \cdot \mathbb{E}[X^2] + 2t \cdot \mathbb{E}[XY] + \mathbb{E}[Y^2].$$

The discriminant of this polynomial of t must therefore be non-positive:

$$4 \cdot \mathbb{E}[XY]^2 - 4 \cdot \mathbb{E}[X^2] \cdot \mathbb{E}[Y^2] \leq 0$$

which gives the Cauchy-Schwarz inequality.

(b) Suppose $X \in \mathcal{L}^2$. Note that the constant function 1 belongs to $\mathcal{L}^2$, $\mathbb{E}[1^2] = 1$. Write $X = X \cdot 1$ and use part (a) to get $X = X \cdot 1 \in \mathcal{L}^1$ and $\mathbb{E}[X \cdot 1]^2 \leq \mathbb{E}[X^2] \cdot \mathbb{E}[1^2]$.

(c) Let $X, Y \in \mathcal{L}^2$ and $a, b \in \mathbb{R}$. Then $(aX + bY)^2 \leq 2a^2X^2 + 2b^2Y^2$ and thus

$0 \geq -(ax-by)^2 = -a^2x^2 + 2abxy - b^2y^2$
Add $2a^2x^2 + 2b^2y^2$ to both sides
$\implies 2a^2x^2 + 2b^2y^2 \geq a^2x^2 + 2abxy + b^2y^2 = (ax+by)^2$

$$\mathbb{E}[(aX + bY)^2] \leq 2a^2\,\mathbb{E}[X^2] + 2b^2\,\mathbb{E}[Y^2] < +\infty,$$

which shows $aX + bY \in \mathcal{L}^2$. $\square$

Def: If $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ then the variance of X is $\mathrm{Var}(X) := \mathbb{E}[(X - \mathbb{E}[X])^2]$.
If $X, Y \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ then the covariance of X and Y is $\mathrm{Cov}(X,Y) := \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]$

Lemma We have $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
and $\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X] \cdot \mathbb{E}[Y]$.

Proof The first statement is a special case of the second because $\mathrm{Var}(X) = \mathrm{Cov}(X,X)$.
Denote $m_X = \mathbb{E}[X]$ and $m_Y = \mathbb{E}[Y]$ and calculate

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathbb{E}[(X-m_X)(Y-m_Y)] \\
&= \mathbb{E}[XY - m_X \cdot Y - X \cdot m_Y + m_X \cdot m_Y] \\
&= \mathbb{E}[XY] - m_X m_Y - m_X m_Y + m_X m_Y \\
&= \mathbb{E}[XY] - m_X m_Y . \qquad \square
\end{aligned}$$

Corollary If $X, Y \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ and $X \perp Y$, then $\mathrm{Cov}(X,Y) = 0$ and $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

Proof Recall that $X \perp Y \Longrightarrow \mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. $\circledast$
Then $\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0$. because of $\circledast$
For the variance of the sum, calculate

$$\begin{aligned}
\mathrm{Var}(X+Y) &= \mathbb{E}[(X+Y)^2] - (\mathbb{E}[X+Y])^2 \\
&= \mathbb{E}[X^2] + 2\,\mathbb{E}[XY] + \mathbb{E}[Y^2] \\
&\quad - (\mathbb{E}[X]^2 + 2\,\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[Y]^2) \\
&= \mathbb{E}[X^2] - \mathbb{E}[X]^2 + \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 \qquad \text{because of } \circledast \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y). \qquad \square
\end{aligned}$$

CONVERGENCE OF RANDOM VARIABLES

Very often in stochastics, one wants to assert that some sequence $X_1, X_2, \dots$ of random variables tends to a limit X in a suitable probabilistic sense.

(You should keep in mind, for comparison, familiar notions such as convergence of sequences of numbers (in calculus) and convergence of functions (especially in functional analysis). Pay attention to similarities, but also to differences.)

Some examples of such contexts might be:

- converge of averages $\frac{X_1 + X_2 + \dots + X_n}{n}$ as $n \to \infty$
- convergence of states X_t of stochastic processes as time t increases
- limits of various random quantities as some parameter of the model tends to an idealized value (e.g. signal-to-noise-ratio $\to \infty$ in communications, size of physical system $\to \infty$ in thermodynamics and statistical mechanics, size of input data $\to \infty$ in randomized algorithms, ...)

In these two remaining lectures we will only consider two types of convergence results in a simple setup of sums of independent identically distributed random variables:

(LLN) "Laws of large numbers": when and in which sense $\frac{1}{n}\sum_{j=1}^{n} X_j \to \mathbb{E}[X]$

(CLT) "Central limit theorems": when and in which sense $\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(X_j - \mathbb{E}[X_j]) \to N(0,\sigma^2)$

($X_1, X_2, \dots$ $\perp\!\!\!\perp$ distributed like X)

Notions of convergence in stochastics

Depending on the situation, we use different notions of convergence of random variables.

Let us first consider an easy example which shows that we should not just define convergence of random variables $X_n : \Omega \to \mathbb{R}$ to a limit $X : \Omega \to \mathbb{R}$ as the convergence of $X_n(\omega) \longrightarrow X(\omega)$ for all outcomes $\omega \in \Omega$.

Example (Coin tossing) ← Interpret as a warning against too naive definitions!

Consider repeated coin tossing,

$\Omega = \{H,T\}^{\mathbb{N}} = \{(\omega_1, \omega_2, \dots) \mid \omega_j \in \{H,T\} \text{ for all } j \in \mathbb{N}\}$

$\mathcal{F}$ = countable product σ-algebra $\bigotimes_{j=1}^{n} \mathcal{P}(\{H,T\})$

$\mathbb{P}$ = countable product measure $\bigotimes_{j=1}^{n} P_{1/2}$ where $P_{1/2}$ is the uniform measure on $\{H,T\}$

Let R_n = relative frequency of heads in the first n coin tosses

$$= \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_{\{\omega_j = H\}}.$$

Then we would of course like to assert that $R_n \longrightarrow \frac{1}{2}$ as $n \longrightarrow \infty$ (in some appropriate sense).

However, for the outcome $\omega = (H,H,H,\dots)$ consisting only of heads we have $R_n(\omega) \longrightarrow 1$, and this outcome is "just as likely as any other. Even worse, for the outcome

$$\omega = (\underbrace{T}_{1}, \underbrace{H,H}_{2}, \underbrace{T,T,T,T}_{4=2^2}, \underbrace{H,\dots,H}_{8=2^3}, \underbrace{T,\dots,T}_{2^4}, \underbrace{H,\dots,H}_{2^5}, \dots)$$

the limit of $R_n(\omega)$ does not even exist although this outcome as well is "just as likely as any other".

In this lecture, we mainly use the following two notions of convergence.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a proba space, and $X_1, X_2, \ldots$ and X real valued random variables on it.

<u>Def</u> (Convergence in probability)

We say that X_n converges to X in probability and denote $X_n \xrightarrow{\mathbb{P}} X$ (as $n \to \infty$) if for any $\varepsilon > 0$ we have

$$\mathbb{P}[\,|X_n - X| \geq \varepsilon\,] \xrightarrow[n\to\infty]{} 0 .$$

<u>Def</u> (Convergence almost surely)

We say that X_n converges to X almost surely (abbreviated a.s.) and denote $X_n \xrightarrow{a.s.} X$ if the event $E = \{\omega \in \Omega \mid X_n(\omega) \to X(\omega)\}$ has full probability, $\mathbb{P}[E] = 1$.

In fact we have briefly encountered already also another notion of convergence, which might be familiar from functional analysis.

<u>Def</u> (L^1 - convergence)

Suppose that $X_1, X_2, \ldots \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. We say that X_n converges to X in L^1, and denote $X_n \xrightarrow{L^1} X$ or $X_n \xrightarrow{\|\cdot\|_1} X$ if $\mathbb{E}[\,|X_n - X|\,] \longrightarrow 0$.

The notions are different, but have relations among them.

<u>Exercise</u>: Show that

- if $X_n \xrightarrow{a.s.} X$ then $X_n \xrightarrow{\mathbb{P}} X$
- if $X_n \xrightarrow{L^1} X$ then $X_n \xrightarrow{\mathbb{P}} X$
- if $X_n \xrightarrow{\mathbb{P}} X$ and for some $Y \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and all n we have $|X_n| \leq Y$ then $X_n \xrightarrow{L^1} X$.

(somewhat harder than the first two)

Exercise: Let $f:\mathbb{R}\to\mathbb{R}$ be a continuous function.

Show that:

- if $X_n \xrightarrow{a.s.} X$ then $f(X_n) \xrightarrow{a.s.} f(X)$
- if $X_n \xrightarrow{\mathbb{P}} X$ then $f(X_n) \xrightarrow{\mathbb{P}} f(X)$.

Weak and strong laws of large numbers

Laws of large numbers assert that averages of i.i.d random variables tend to their expected value (this gives an important interpretation of the meaning of expected value).

One can prove such theorems with various assumptions on the law of X_j, $j\in\mathbb{N}$, but moreover we distinguish between theorems that assert convergence in probability (weak laws of large numbers) and those which assert almost sure convergence (strong laws of large numbers).

Today we will prove one of each type:

Theorem (Weak law of large numbers) (Assuming bounded second moments)

Let $X_1, X_2, \dots \in \mathcal{L}^2(\Omega,\mathcal{F},\mathbb{P})$ be independent random variables such that $\forall j\in\mathbb{N}$: $\mathbb{E}[X_j]=\mu$ and $\mathbb{E}[X_j^2]\le K$.

Then $\frac{1}{n}\sum_{j=1}^{n} X_j \xrightarrow{\mathbb{P}} \mu$ as $n\to\infty$.

Theorem (Strong law of large numbers) (Assuming bounded fourth moments)

Let $X_1, X_2, \dots$ be independent random variables such that for some $K>0$ and all $j\in\mathbb{N}$ $\mathbb{E}[X_j^4]\le K$ and $\mathbb{E}[X_j]=\mu$. Then we have

$$\frac{1}{n}\sum_{j=1}^{n} X_n \xrightarrow{a.s.} \mu \quad \text{as } n\to\infty.$$

Markov and Chebyshev inequalities and weak LLN

The following result is almost obvious:

Proposition (Markov's inequality)

If $X : \Omega \to \mathbb{R}$ is a random variable. Then for any $a > 0$ we have
$$\mathbb{P}[\,|X| \geq a\,] \leq \tfrac{1}{a} \mathbb{E}[|X|].$$

Proof Let $E = \{\omega \in \Omega \mid |X(\omega)| \geq a\}$. Then we have $|X| \geq a \cdot \mathbb{1}_E$, so by monotonicity of expected value (and linearity)
$$\mathbb{E}[\,|X|\,] \geq a \cdot \mathbb{E}[\mathbb{1}_E] = a \cdot \mathbb{P}[E].$$
The result follows by dividing by a. $\square$

Corollary (Chebyshev's inequality)

Suppose $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$, and denote $\mu := \mathbb{E}[X]$ and $\sigma^2 := \mathrm{Var}(X) = \mathbb{E}[(X-\mu)^2]$.
Then for any $c > 0$ we have
$$\mathbb{P}[\,|X-\mu| \geq c\,] \leq \frac{\sigma^2}{c^2}.$$

Proof: Let $Y = (X-\mu)^2$, so $\mathbb{E}[\,|Y|\,] = \sigma^2$.
Note that $\{\omega \in \Omega \mid |X(\omega) - \mu| \geq c\} = \{\omega \in \Omega \mid |Y| \geq c^2\}$.
Now apply Markov's inequality to Y with $a = c^2$. $\square$

Proof of weak LLN: Let $X_1, X_2, \ldots \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ $\perp\!\!\!\perp$ with $\mathbb{E}[X_j] = \mu$ $\forall j$ and $\mathbb{E}[X_j^2] \leq K$ $\forall j$.
Then $\mathrm{Var}(X_j) = \mathbb{E}[X_j]^2 - \mathbb{E}[X_j]^2 \leq K$ $\forall j$.
Let $S_n = \sum_{j=1}^n X_j$ and $Y_n = \frac{1}{n} S_n$. We want to show $Y_n \xrightarrow{\mathbb{P}} \mu$.
By linearity $\mathbb{E}[S_n] = \mathbb{E}[\sum_{j=1}^n X_j] = \sum_{j=1}^n \mathbb{E}[X_j] = n \cdot \mu$
and $\mathbb{E}[Y_n] = \mu$.
By independence (recall exercises)
$$\mathrm{Var}(S_n) = \sum_{j=1}^n \mathrm{Var}(X_j) \leq n \cdot K \quad \text{and} \quad \mathrm{Var}(Y_n) = \mathrm{Var}(\tfrac{1}{n} S_n) \leq \tfrac{1}{n^2} \mathrm{Var}(S_n) \leq K/n.$$

Therefore, applying Chebyshev's inequality to Y_n,

$$\mathbb{P}\left[\,|Y_n - \mu| \geq c\,\right] \leq \frac{\mathrm{Var}(Y_n)}{c^2} \leq \frac{K}{n\,c^2} \xrightarrow[n\to\infty]{} 0 .$$

This proves $Y_n \xrightarrow{\mathbb{P}} \mu$. $\square$

It is remarkable that the easy result above allows one to prove the Weierstrass approximation theorem ("polynomials are dense in the space of continuous functions w.r.t. uniform convergence") with little effort. See [Williams, Chapter 7.4]. This exemplifies the power of probabilistic ideas in other branches of mathematics.

Strong law of large numbers

Recall that we aim to prove

Theorem (Strong LLN)

Let $X_1, X_2, \ldots$ be independent random variables s.t. for some $\mu \in \mathbb{R}$ and $K > 0$ we have $\forall j \in \mathbb{N}$: $\mathbb{E}[X_j] = \mu$ and $\mathbb{E}[X_j^4] \leq K$.

Then $\frac{1}{n}\sum_{j=1}^{n} X_j \xrightarrow{a.s.} \mu$.

We first make a few preliminary observations.

Lemma If $X \in m\mathcal{F}$ and for some $p > 1$ we have $\mathbb{E}[|X|^p] < \infty$ then also for any $r \in [1, p]$ we have $\mathbb{E}[|X|^r] < \infty$.

Proof: When $1 \leq r \leq p$, we have

$$x^r \leq x^p \quad \text{for } x \geq 1$$
$$x^r \leq 1 \quad \text{for } x \in [0,1].$$

Denoting $E = \{\omega \in \Omega \mid |X(\omega)| \leq 1\}$ we therefore have $|X|^r \leq \mathbb{1}_E + |X|^p \cdot \mathbb{1}_{E^c}$, and as a consequence

$$\mathbb{E}[|X|^r] \leq \mathbb{E}[\mathbb{1}_E + |X|^p \mathbb{1}_{E^c}] \leq \mathbb{P}[E] + \mathbb{E}[|X|^p]$$
$$\leq 1 + \mathbb{E}[|X|^p].$$

The claim follows. □

Lemma If $X, Y \in m\mathcal{F}$ and $p \geq 1$ then

$$\mathbb{E}[|X+Y|^p] \leq 2^p \left(\mathbb{E}[|X|^p] + \mathbb{E}[|Y|^p] \right).$$

Proof: For any $x, y \in \mathbb{R}$ we have by triangle ineq.

$|x+y| \leq |x| + |y| \leq 2 \cdot \max(|x|, |y|)$. The mapping $t \mapsto t^p$ is increasing $[0,\infty) \to [0,\infty)$, so

$$|x+y|^p \leq (2 \max(|x|, |y|))^p = 2^p \cdot \max(|x|^p, |y|^p) \leq 2^p (|x|^p + |y|^p).$$

Then we get $|X+Y|^p \leq 2^p (|X|^p + |Y|^p)$ pointwise, so the claim follows by monotonicity and linearity. □

Proof of strong LLN (with bounded fourth moments)

Assume first $\mu = 0$; the general case will follow by considering $X_j' = X_j - \mu$.

Write $S_n = \sum_{j=1}^n X_j$ and $Y_n = \frac{1}{n} S_n$.

We want to prove $Y_n \xrightarrow{a.s.} 0$.

To compute the fourth moment of S_n, expand

$$S_n^4 = (X_1 + X_2 + \dots + X_n)^4$$

$$= \sum_{1 \le i \le n} X_i^4 + \sum_{\substack{1 \le i, j \le n \\ i \ne j}} X_i^3 X_j \cdot \frac{4!}{3!\,1!} + \frac{1}{2!} \sum_{\substack{1 \le i,j \le n \\ i \ne j}} X_i^2 X_j^2 \frac{4!}{2!\,2!}$$

$$+ \frac{1}{2!} \sum_{\substack{1 \le i,j,k \le n \\ i,j,k \text{ different}}} X_i^2 X_j X_k \cdot \frac{4!}{2!\,1!\,1!} + \frac{1}{4!} \sum_{\substack{1 \le i,j,k,l \le n \\ i,j,k,l \text{ different}}} X_i X_j X_k X_l$$

Now note that for $i \ne j$, $X_i \perp X_j$, and thus $\mathbb{E}[X_i^3 X_j] = \mathbb{E}[X_i^3]\, \underbrace{\mathbb{E}[X_j]}_{=0} = 0$

(we used $X_i^3 \in \mathcal{L}^1$, which follows by $\mathbb{E}[|X_i|^3] \le 1 + \mathbb{E}[X_i^4]$).

Similarly, for i,j,k different $\mathbb{E}[X_i^2 X_j X_k] = \mathbb{E}[X_i^2]\, \underbrace{\mathbb{E}[X_j]}_{=0}\, \underbrace{\mathbb{E}[X_k]}_{=0}$

and for i,j,k,l different $\mathbb{E}[X_i X_j X_k X_l] = \underbrace{\mathbb{E}[X_i]}_{=0} \cdots \underbrace{\mathbb{E}[X_l]}_{=0} = 0$.

We also use Cauchy-Scwarz to get, for $i \ne j$,

$$\mathbb{E}[X_i^2 X_j^2] \le \sqrt{\mathbb{E}[X_i^4] \cdot \mathbb{E}[X_j^4]} \le \sqrt{K \cdot K} = K.$$

Thus we can compute the fourth moment of S_n, and bound it

$$\mathbb{E}[S_n^4] = \sum_{1 \le i \le n} \mathbb{E}[X_i^4] + \frac{4!}{3!\,1!} \cdot 0 + \frac{1}{2!} \cdot \frac{4!}{2!\,2!} \sum_{\substack{1 \le i,j \le n \\ i \ne j}} \mathbb{E}[X_i^2 X_j^2]$$

$$+ \frac{1}{2!} \cdot \frac{4!}{2!\,1!\,1!} \cdot 0 + \frac{1}{4!}\, 0$$

$$\le n \cdot K + \frac{n(n-1)}{2} \cdot 6 \cdot K \le 3n^2 K.$$

With this estimate $\mathbb{E}[(\frac{S_n}{n})^4] \le \frac{3K}{n^2}$ we get

$$\mathbb{E}\Big[\sum_{n=1}^{\infty} \Big(\frac{S_n}{n}\Big)^4\Big] \le 3K \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty .$$

But recall that this is possible only if $\sum_{n=1}^{\infty} (\frac{S_n}{n})^4 < \infty$ almost surely. And consequently the terms of this sum must be tending to zero almost surely, i.e. $(\frac{S_n}{n})^4 \xrightarrow[n\to\infty]{a.s.} 0$.

By continuity of the function $t \longmapsto t^{1/4}$ we conclude that $\frac{S_n}{n} \xrightarrow[n\to\infty]{a.s.} 0$, which proves the claim.

In the general case $\mathbb{E}[X_j] = \mu$, consider $\tilde{X}_j = X_j - \mu$. Then $\mathbb{E}[\tilde{X}_j] = 0$ and

$$\mathbb{E}[\tilde{X}_j^4] \le 2^4 \left(\mathbb{E}[X_j^4] + \mu^4\right) \le 2^4 (K + \mu^4) =: \tilde{K}.$$

We can apply the above special case to $\tilde{X}_1, \tilde{X}_2, \ldots$ to get $\frac{1}{n}\sum_{j=1}^{n} \tilde{X}_j \xrightarrow{a.s.} 0$. But then

$$\frac{1}{n}\sum_{j=1}^{n} X_j = \frac{1}{n}\sum_{j=1}^{n} (\tilde{X}_j + \mu) = \frac{1}{n}\sum_{j=1}^{n} \tilde{X}_j + \mu \xrightarrow{a.s.} 0 + \mu = \mu.$$

$\square$

Kolmogorov's strong law of large numbers

We state without proof a strong LLN which has no assumptions on existence of moments other than obviously the expected value.

Theorem (Kolmogorov's strong law of large numbers)

Let $X_1, X_2, \ldots \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ be i.i.d. random variables and denote $\mu = \mathbb{E}[X_j]$. Then we have $\frac{1}{n}\sum_{j=1}^{n} X_j \xrightarrow{a.s.} \mu$ and $\frac{1}{n}\sum_{j=1}^{n} X_j \xrightarrow{L^1} \mu$.

CENTRAL LIMIT THEOREM

Recall that the central limit theorem asserts that for i.i.d. random variables $X_1, X_2, \dots$ with $\mu = E[X_j]$ and $\sigma^2 = Var(X_j)$ (so we need $X_j \in \mathcal{L}^2$) we have

$$\frac{\sum_{j=1}^{n} X_j - n\mu}{\sigma\sqrt{n}} \xrightarrow{?} Z \quad \text{as } n \to \infty$$

where Z has a standard normal distribution, i.e. the law P_Z of Z has density $f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$. In what sense does the convergence hold?

Remarks

- The laws of large numbers say that the sum $S_n = \sum_{j=1}^{n} X_j$ concentrates around $n\mu$ on the scale n, in the sense that $\frac{S_n - n\mu}{n} \to 0$ (almost surely or in probability or in L^1).
- Chebyschev's inequality says that the fluctuations of S_n around $n\mu$ are on a scale not greater than $\sqrt{n}$. Namely $Var(S_n) \overset{\perp}{=} \sum_{j=1}^{n} Var(X_j) = n\sigma^2$ (by independence) so $P[|S_n - n\mu| \geq c\sqrt{n}] \leq \frac{n\sigma^2}{(c\sqrt{n})^2} = \frac{\sigma^2}{c^2}$.

  CLT in particular states that $\sqrt{n}$ is the appropiate scale.
- One can show that $(S_n - n\mu)/\sigma\sqrt{n}$ does not converge to anything almost surely or in probability so the meaning of convergence must be different.

The appropriate notion of convergence is known as "convergence in law" (or "convergence in distribution" or "weak convergence"). We will give the general definition later, but one elementary interpretation is the following convergence of cumulative distribution functions

$$\forall x \in \mathbb{R} \quad P\left[\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right] \xrightarrow[n\to\infty]{} \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt .$$

CHARACTERISTIC FUNCTIONS

The characteristic function φ_X of ~~a~~ random variable $X : \Omega \to \mathbb{R}$ is a very neat and practical way to encode the distribution of X in one well behaved complex-valued function on $\mathbb{R}$.

Def: The characteristic function of $X : \Omega \to \mathbb{R}$ is the function $\varphi_X : \mathbb{R} \to \mathbb{C}$ given by

$$\varphi_X(\theta) = \mathbb{E}\left[e^{i\theta X}\right] = \int_{\mathbb{R}} e^{i\theta x} \, dP_X(x).$$

Here $e^{i\theta x}$ is the complex exponential function $e^z = \sum_{n=0}^{\infty} \frac{1}{n!} z^n$. For $x, \theta \in \mathbb{R}$ we have $e^{i\theta x} = \cos(\theta x) + i \cdot \sin(\theta x)$. The expected value is taken separately for real and imaginary parts, $\mathbb{E}[e^{i\theta X}] = \mathbb{E}[\cos(\theta X)] + i \cdot \mathbb{E}[\sin(\theta X)]$.

Here are some elementary properties of characteristic functions.

Proposition: Let $X : \Omega \to \mathbb{R}$ be a r.v. and $\varphi_X : \mathbb{R} \to \mathbb{C}$ its characteristic function. Then:

(a) $\varphi_X(0) = 1$

(b) $|\varphi_X(\theta)| \leq 1 \quad \forall \, \theta \in \mathbb{R}$

(c) $\varphi_X : \mathbb{R} \to \mathbb{C}$ is continuous

(d) $\varphi_{-X}(\theta) = \overline{\varphi_X(\theta)}$ (complex conjugate)

(e) $\varphi_{aX+b}(\theta) = e^{i\theta b} \cdot \varphi_X(a\theta)$ for any $a, b \in \mathbb{R}$, $\theta \in \mathbb{R}$

Proof: All of these are easy. Let us only check (c). Take $\theta_1, \theta_2, \ldots \in \mathbb{R}$ with $\theta_n \to \theta$. Then by continuity of exp, we have $e^{i\theta_n X} \to e^{i\theta X}$. Moreover $|e^{i\theta_n X}| \leq 1$, so we can use Bounded Convergence Thm:

$$\lim_{n\to\infty} \varphi_X(\theta_n) = \lim_{n\to\infty} \mathbb{E}[e^{i\theta_n X}] = \mathbb{E}[\lim_{n\to\infty} e^{i\theta_n X}] = \mathbb{E}[e^{i\theta X}] = \varphi_X(\theta). \quad \square$$

Lévy's inversion theorem

We now show, very concretely, that the characteristic function $\varphi_X : \mathbb{R} \to \mathbb{C}$ of a random variable $X : \Omega \to \mathbb{R}$ determines the law $\mathbb{P}_X$ of X.

For simplicity of notation, fix $X : \Omega \to \mathbb{R}$, and denote

- $\mu = \mathbb{P}_X$ the law of X
  (proba meas. on $(\mathbb{R}, \mathcal{B})$ s.t. $\mu[B] = \mathbb{P}[X \in B]$)
- $F = F_X$ the cumulative distribution func. of X
  ($F : \mathbb{R} \to [0,1]$ s.t. $F(x) = \mu[(-\infty, x]] = \mathbb{P}[X \leq x]$)
- $\varphi = \varphi_X$ the characteristic function of X
  ($\varphi : \mathbb{R} \to \mathbb{C}$ s.t. $\varphi(\theta) = \mathbb{E}[e^{i\theta X}]$)

__Theorem__ (Lévy's inversion theorem) For any $a < b$ we have

$$(*) \quad \lim_{T \to +\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-i\theta a} - e^{-i\theta b}}{i\theta} \varphi(\theta)\, d\theta = \mu[(a,b)] + \frac{1}{2}\mu[\{a\}] + \frac{1}{2}\mu[\{b\}]$$

$$= \frac{1}{2}\left(F(b) + F(b^-)\right) - \frac{1}{2}\left(F(a) + F(a^-)\right)$$

Moreover, if $\int_{\mathbb{R}} |\varphi(\theta)|\, d\theta < \infty$ then X has a continuous probability density f_X given by

$$f_X(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-i\theta x} \varphi(\theta)\, d\theta.$$

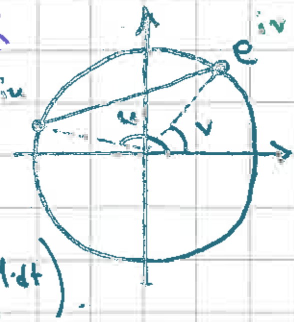
__Remark__: If X has density f_X, then

$$\varphi_X(\theta) = \int_{\mathbb{R}} e^{i\theta x} f_X(x)\, dx.$$

This means that f_X and φ_X are related to each other by Fourier transform.

Proof: Note that for any $u, v \in \mathbb{R}$

$$|e^{iu} - e^{iv}| \le |u-v|$$



(see picture, or calculate $|e^{iu} - e^{iv}| = \left| \int_u^v i e^{it}\,dt \right| \le \int_u^v |i e^{it}|\,dt = \int_u^v 1\,dt$).

Let $a, b \in \mathbb{R}$ with $a < b$. For $T > 0$ use Fubini's theorem to write the left hand side of $\circledast$ as

$$\text{LHS} = \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-i\theta a} - e^{-i\theta b}}{i\theta}\, \varphi(\theta)\, d\theta = \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-i\theta a} - e^{-i\theta b}}{i\theta} \left(\int_{\mathbb{R}} e^{i\theta x}\, d\mu(x) \right) d\theta$$

$$= \frac{1}{2\pi} \int_{\mathbb{R}} \left(\int_{-T}^{T} \frac{e^{i\theta(x-a)} - e^{i\theta(x-b)}}{i\theta}\, d\theta \right) d\mu(x).$$

To justify the use of Fubini, note that

$$\int_{\mathbb{R}} \Bigg(\int_{-T}^{T} \underbrace{\left| \frac{e^{i\theta(x-a)} - e^{i\theta(x-b)}}{i\theta} \right|}_{\le |a-b|} d\theta \Bigg)\, d\mu(x) \le 2T \cdot |a-b| < \infty.$$

The integral over θ is, by evenness of cosine and oddness of sine

$$\int_{-T}^{T} \frac{e^{i\theta(x-a)} - e^{i\theta(x-b)}}{i\theta}\, d\theta = 2\,\mathrm{sgn}(x-a)\, S(|x-a| \cdot T) - 2\,\mathrm{sgn}(x-b)\, S(|x-b| \cdot T)$$

where

$$\mathrm{sgn}(x) := \begin{cases} 1 & \text{for } x > 0 \\ 0 & \text{for } x = 0 \\ -1 & \text{for } x < 0 \end{cases}$$

and

$$S(r) := \int_0^r \frac{\sin(t)}{t}\, dt \qquad (r \ge 0).$$

By a standard exercise in complex analysis (residue calculus) we have

$$S(r) \longrightarrow \frac{\pi}{2} \quad \text{as } r \longrightarrow +\infty$$

(Note, however, that the Lebesgue integral $\int_0^\infty \frac{\sin(t)}{t}\, dt$ does NOT exist.)

Therefore, as $T \to +\infty$, we have

$$\int_{-T}^{T} \frac{e^{i\theta(x-a)} - e^{i\theta(x-b)}}{i\theta}\, d\theta \longrightarrow \begin{cases} 2\pi, & \text{if } a < x < b \\ \pi, & \text{if } x = a \text{ or } x = b \\ 0, & \text{if } x < a \text{ or } x > b \end{cases}$$

and this expression is uniformly bounded in T.

The Bounded Convergence Theorem implies

$$\text{LHS} = \frac{1}{2\pi}\int_{\mathbb{R}}\left(\int_{-T}^{T}\frac{e^{i\theta(x-a)}-e^{i\theta(x-b)}}{i\theta}\,d\theta\right)d\mu(x)$$

$$\xrightarrow[T\to+\infty]{} \mu[(a,b)] + \tfrac{1}{2}\mu[\{a\}] + \tfrac{1}{2}\mu[\{b\}],$$

proving the first part of the theorem.

Suppose now that $\int_{\mathbb{R}}|\varphi(\theta)|\,d\theta < \infty$.
If $a_n \to a$ and $b_n \to b$ then we can use Dominated Convergence Theorem to get

$$\underbrace{\frac{1}{2\pi}\int_{\mathbb{R}}\frac{e^{-i\theta a_n}-e^{-i\theta b_n}}{i\theta}\varphi(\theta)\,d\theta}_{=\frac{1}{2}(F(b_n)+F(b_n^-))-\frac{1}{2}(F(a_n)+F(a_n^-))} \longrightarrow \underbrace{\frac{1}{2\pi}\int_{\mathbb{R}}\frac{e^{-i\theta a}-e^{-i\theta b}}{i\theta}\varphi(\theta)\,d\theta}_{=\frac{1}{2}(F(b)+F(b^-))-\frac{1}{2}(F(a)+F(a^-))}$$

which shows that the c.d.f. $F:\mathbb{R}\to[0,1]$ must be continuous. Moreover, since

$$\left|\frac{e^{-i\theta a}-e^{-i\theta b}}{i\theta(b-a)}\right| \le 1 \quad \text{and} \quad \int_{\mathbb{R}}|\varphi(\theta)|\,d\theta < \infty$$

we can use Dominated Convergence to calculate the derivative of F as

$$F'(a) = \lim_{b\to a}\frac{F(b)-F(a)}{b-a} = \lim_{b\to a}\frac{1}{2\pi}\int_{\mathbb{R}}\frac{e^{-i\theta a}-e^{-i\theta b}}{i\theta(b-a)}\varphi(\theta)\,d\theta$$

$$\overset{DCT}{=} \frac{1}{2\pi}\int_{\mathbb{R}}e^{-i\theta a}\varphi(\theta)\,d\theta.$$

(Actually this calculation gives only the derivative from the right, but derivative from the left is handled similarly)

Obviously $f_X(x) = F'(x)$ is the density of X, since

$$\mu[(a,b)] = F(b)-F(a) = \int_a^b F'(x)\,dx$$

for any interval (a,b).
(And intervals form a π-syst. generating $\mathcal{B}$)
Continuity of $f_X(x) = \frac{1}{2\pi}\int_{\mathbb{R}}e^{-i\theta x}\varphi(\theta)\,d\theta$ is proven by DCT as usually. $\square$

Proposition If $X \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$, then we have

$$\varphi_X(\theta) = 1 + i\theta \cdot \mathbb{E}[X] - \frac{\theta^2}{2}\mathbb{E}[X^2] + o(\theta^2) \quad \text{as } \theta \to 0$$

Notation: Above we use the "Landau little-o notation": $o(\theta^2)$ denotes some function of θ with the property that $\frac{o(\theta^2)}{\theta^2} \longrightarrow 0$ as $\theta \to 0$. This notation is convenient for collecting terms which are of smaller order than the terms that we care about.

Proof The idea is to Taylor expand $e^{i\theta X}$, which is achieved by

$$e^{i\theta x} = 1 + i\theta \cdot \int_0^x e^{i\theta u}\, du \quad \text{(an easy integral)}$$

$$= 1 + i\theta \cdot \int_0^x \left(1 + i\theta \int_0^u e^{i\theta v} dv\right) du \quad \text{(repeat the same)}$$

$$= 1 + i\theta x - \frac{\theta^2 x^2}{2} - \theta^2 \int_0^x du \int_0^u dv\, (e^{i\theta v} - 1).$$

Then we are ready to look at the claim:

$$\varphi_X(\theta) - \left(1 + i\theta \cdot \mathbb{E}[X] - \frac{\theta^2}{2}\mathbb{E}[X^2]\right)$$

$$= \mathbb{E}\left[e^{i\theta X} - 1 - i\theta X + \frac{\theta^2}{2}X^2\right]$$

$$= -\theta^2\, \mathbb{E}\left[\int_0^X du \int_0^u dv\, (e^{i\theta v} - 1)\right].$$

Take absolute values and use triangle ineq. for integrals

$$\left|\varphi_X(\theta) - \left(1 + i\theta\, \mathbb{E}[X] - \frac{\theta^2}{2}\mathbb{E}[X^2]\right)\right|$$

$$\leq |\theta|^2 \cdot \mathbb{E}\left[\int_0^{|X|} du \int_0^u dv\, |e^{i\theta v} - 1|\right] = |\theta|^2 \cdot \mathbb{E}\left[R(\theta, |X|)\right].$$

We denoted $R(\theta, x) = \int_0^x du \int_0^u dv\, \underbrace{|e^{i\theta v} - 1|}_{= 2\cdot \sin\left(\frac{\theta v}{2}\right) \leq \min(2, |\theta v|)} \leq \min\left(x^2, \frac{1}{6}|\theta|\cdot|x|^3\right).$

We have $R(\theta, |X|) \leq |X|^2$ which is integrable, so we can apply Dominated Convergence Theorem to get

$$\lim_{\theta \to 0} \mathbb{E}\left[R(\theta, |X|)\right] \overset{DCT}{=} \mathbb{E}\Big[\underbrace{\lim_{\theta \to 0} R(\theta, |X|)}_{= 0 \text{ since } R(\theta, x) \leq \frac{1}{6}|\theta||x|^3}\Big] = 0.$$

$\square$

WEAK CONVERGENCE / CONVERGENCE IN LAW

The idea of convergence in law (= "convergence in distribution" = "weak convergence") is that anything that can be reliably measured / tested, does converge. The mathematical idealization of a reliably measurable quantity is the expected value of a bounded continuous function of the random variable. So we define:

Def: A sequence of $\mathbb{R}$-valued random variables $X_1, X_2, \ldots$ is said to converge in law to an $\mathbb{R}$-val. r.v. X if for all bounded continuous functions $f: \mathbb{R} \to \mathbb{R}$ we have

$$\mathbb{E}[f(X_n)] \longrightarrow \mathbb{E}[f(X)].$$

(Notation: $X_n \xrightarrow{\text{law}} X$)

Remark: The convergence in law only takes into account the laws P_{X_n} of X_n and P_X of X: indeed, equivalently

$$\int_{\mathbb{R}} f(x)\, dP_{X_n}(x) \longrightarrow \int_{\mathbb{R}} f(x)\, dP_X(x)$$

for all $f: \mathbb{R} \to \mathbb{R}$ cont. bdd.

Same function f integrated against different measures P_{X_n}, P_X on $(\mathbb{R}, \mathcal{B})$. Not a sequence of functions integrated against a fixed measure, as usually!

There are other equivalent conditions for convergence in law.

Theorem Let $X, X_1, X_2, \ldots$ be $\mathbb{R}$-val. random variables with cumulative distribution functions $F, F_1, F_2, \ldots$ and characteristic functions $\varphi, \varphi_1, \varphi_2, \ldots$, respectively. Then the following are equivalent:

(1°) $X_n \xrightarrow{\text{law}} X$

(2°) $\varphi_n(\theta) \longrightarrow \varphi(\theta)$ for all $\theta \in \mathbb{R}$

(3°) $F_n(x) \longrightarrow F(x)$ for all continuity points $x \in \mathbb{R}$ of F

Proof See e.g. the course "Large random systems". □

Theorem (Central Limit Theorem)

Let $X_1, X_2, \ldots \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ be i.i.d.
Denote $\mu = \mathbb{E}[X_j]$ and $\sigma^2 = \mathrm{Var}(X_j)$,
and $S_n = \sum_{j=1}^{n} X_j$. Then we have

$$\frac{S_n - n\mu}{\sqrt{n}} \xrightarrow{\text{law}} Z \sim N(0, \sigma^2)$$

Proof: Without loss of generality assume $\mu = 0$ (otherwise consider $\tilde{X}_j = X_j - \mu$).

Let $\varphi(\theta) = \varphi_{X_j}(\theta) = \mathbb{E}[e^{i\theta X_j}] = 1 - \frac{\theta^2\sigma^2}{2} + o(\theta^2)$ be the characteristic function of the terms X_j (we used the previous Proposition to expand it).

Then calculate the characteristic function of $\frac{S_n}{\sqrt{n}}$ using independence of $X_1, \ldots, X_n$:

$$\varphi_{S_n/\sqrt{n}}(\theta) = \mathbb{E}\left[\exp\left(i\theta \frac{1}{\sqrt{n}} \sum_{j=1}^{n} X_j\right)\right] = \prod_{j=1}^{n} \mathbb{E}\left[e^{i\theta X_j/\sqrt{n}}\right]$$

$$= \left(\varphi\left(\frac{\theta}{\sqrt{n}}\right)\right)^n = \left(1 - \frac{\theta^2\sigma^2}{2n} + o\left(\frac{\theta^2}{n}\right)\right)^n$$

(Exercise 1 below) $\longrightarrow$

$$\longrightarrow e^{-\frac{\sigma^2}{2}\theta^2} \quad \text{as} \quad n \to \infty.$$

We recognize the limit as the characteristic function of $Z \sim N(0, \sigma^2)$

(Exercise 2 below) $\longrightarrow$

$$\mathbb{E}[e^{i\theta Z}] = e^{-\frac{\sigma^2}{2}\theta^2}.$$

By the theorem about convergence in law and characteristic functions, we have thus obtained

$$\frac{S_n}{\sqrt{n}} \xrightarrow{\text{law}} Z. \qquad \square$$

Exercise 1: Show that if $r_1, r_2, \ldots \in \mathbb{C}$ are such that $r_n \to r \in \mathbb{C}$ as $n \to \infty$, then $\left(1 + \frac{r_n}{n}\right)^n \to e^r$ as $n \to \infty$.

Exercise 2: Assume $Z \sim N(\mu, \sigma^2)$, i.e. the law of Z has density $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$. Show $\varphi_Z(\theta) = \exp\left(i\mu\theta - \frac{\sigma^2\theta^2}{2}\right)$.

Further studies in stochastics ?

This semester

- LARGE RANDOM SYSTEMS
  period IV, Feb 22 – Mar 31, Lasse Leskelä & Kalle Kytölä
- RANDOM MATRICES – THEORY AND APPLICATIONS
  period V, Apr 11 – May 20, Christian Webb

Regularly also

- BROWNIAN MOTION AND STOCHASTIC ANALYSIS

Other

- If you're interested in self-study, credits are given if topic and method of completion are agreed on before hand with one of our stochastics / statistics professors (Pauliina Ilmonen, Kalle Kytölä, Lasse Leskelä)
- Topics for B.Sc or M.Sc. or Ph.D theses ? Come talk to the professors and other researchers.