

MEASURE THEORY AND LEBESGUE INTEGRATION

In this lecture we will abstract the notion of length, area or volume of a set and with that, we will be able to build a notion of integration over arbitrary spaces.

5.1 Measure Theory

Why do we say that the area of a unit circle is π ? By definition, π is the length of half its perimeter, but why is then the interior's area π ? It has no easy answer, unless we go back to elementary school and recall what area means. We define that a rectangle's area is the product of its base times height. To know the area of any other figure we resort to that: we either split a rectangle into some parts and rearrange them to build a new figure (say, to know the area of a triangle) or, if possible, we split the figure into rectangles and then sum their areas. If none of them is possible, as is the case for the circle, we resort to the ancient Greeks. We approximate the desired shape (say, the circle) by a union of small squares and compute its area, then with even smaller squares and compute its area, repeating this, we get a sequence of areas of better and better approximations of the objective shape. We define the area of the shape to be the limit of that sequence. Alternatively, we could try to fill the shape with countably infinitely many rectangles (which are smaller and smaller as we approach the “non-straight” edge of the shape), then we add the areas of all of them. As surprising as it may sound, there is really no other way to know the area of a circle.

Thus, the notion of area (and similarly, that of volume and length) is not a mere attribution of non-negative real numbers to shapes, namely, to subsets of $\mathbb{R}^2$ (or rather of Euclidean 2-space). It is an intrinsic part of the concept that given a finite or countable partition of a shape, the sum of the areas of the parts is the area of the shape. That is why, the obvious abstraction of area on an arbitrary set X —called, *measure* on X — is a map that assigns non-negative numbers to subsets of X (which might be zero —like the area of a point— or infinite —like the area of a half-plane), such that given finite or countably many disjoint subsets, the measure attributed to their union equals the sum of the subsets' measures.

When making this abstraction however, mathematicians found a problem: there exist subsets of $\mathbb{R}^2$ which can be given no area (not even zero or infinity), i.e., sets which would yield a mathematical contradiction if they had a “Greek-style area”. Much the same for lengths and volumes. In a future lecture, we will see an explicit example of such a reason-defying shape for the case of length.

This tells us that we can only hope to define areas on a subfamily of the power set. Hence, when we generalize it, we will usually only define measures on certain subfamilies of the power set. Which ones? Well, we know for areas that if a subset of $\mathbb{R}^2$ has a well-defined area, its complement has a well-defined area too. Moreover, if a countable family of sets has an area, their union also has an area. Thus, it is natural to consider the following type of family.

Definition 69. Given a set X , a family of subsets $\mathcal{A} \subseteq \mathcal{P}(X)$ is called σ -algebra when

- (i) $X \in \mathcal{A}$
- (ii) $A \in \mathcal{A} \implies X \setminus A \in \mathcal{A}$
- (iii) $A_n \in \mathcal{A} \ \forall n \in \mathbb{N} \implies \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$

The elements of $\mathcal{A}$ are called the $\mathcal{A}$ -measurable sets and $(X, \mathcal{A})$ is called a *measurable space*. $\blacklozenge$

Trivially, both $\{\emptyset, X\}$ and $\mathcal{P}(X)$ are σ -algebras, but the most natural choice is often built using the following:

Proposition 35. Given a set X , an index set I and a family of σ -algebras on X , $\{\mathcal{A}_j\}_{j \in I}$, their intersection $\bigcap_{j \in I} \mathcal{A}_j$ is still a σ -algebra on X . Hence, given any $B \subseteq \mathcal{P}(X)$, it is possible to define:

$$\sigma(B) := \bigcap_{\substack{\mathcal{A} \subseteq \mathcal{P}(X): \\ \mathcal{A} \text{ is a } \sigma\text{-algebra on } X \\ B \subseteq \mathcal{A}}} \mathcal{A},$$

which is the *smallest σ -algebra containing all sets in B* , called *σ -algebra generated by B* . $\blacklozenge$

Then, given a topological space, there is already a distinguished choice of sets, namely, the open and closed sets. In particular, in $\mathbb{R}, \mathbb{R}^2, \mathbb{R}^3$, every closed and open set has a length, area and volume, so the σ -algebra they generate is a canonical choice of measurable sets:

Definition 70. Given a topological space $(X, \mathcal{T})$, we denote the σ -algebra generated by the topology $\mathfrak{B}(X) := \sigma(\mathcal{T})$ and call it the *Borel σ -algebra of X* . $\blacklozenge$

That said, one defines the generalization of area as we anticipated:

Definition 71. Let $(X, \mathcal{A})$ be a measurable space. A map $\mu : \mathcal{A} \rightarrow [0, \infty]$ is called a *measure* when

- (i) $\mu(\emptyset) = 0$,
- (ii) if $(A_n)_{n \in \mathbb{N}} \subseteq \mathcal{A}$ are pairwise disjoint, then

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n) \quad (\sigma\text{-additivity}). \quad (5.1)$$

The triple $(X, \mathcal{A}, \mu)$ is called a *measure space*. Furthermore,

- A set $B \in \mathcal{A}$ is called a *null set* when it has measure 0 (i.e., $\mu(B) = 0$).
- In the context of integration, one denotes indistinguishably $d\mu := \mu$.
- We say that μ is a *finite* measure when $\mu(X) < \infty$.
- We say that μ is a *σ -finite* measure, when there exists $(A_n)_{n \in \mathbb{N}} \subseteq \mathcal{A}$ such that $X = \bigcup_{n=1}^{\infty} A_n$ and $\mu(A_n) < \infty \ \forall n \in \mathbb{N}$. $\blacklozenge$

The prime example of measure is the generalization of length, area and volume to arbitrary $\mathbb{R}^n$ and it is uniquely characterized by the obvious fact that a given shape has the same length/area/volume irrespective of its “location”:

Theorem 17. There *exists a unique* measure λ^n on $(\mathbb{R}^n, \mathfrak{B}(\mathbb{R}^n))$ satisfying:

1. that the unit cube has measure 1, i.e., $\lambda^n([0, 1]^n) = 1$.
2. It is *translation invariant*, i.e.,

$$\lambda^n(A + v) = \lambda^n(A) \quad \forall A \in \mathcal{B} \quad \forall v \in \mathbb{R}^n.$$

We call it the *Lebesgue(-Borel) measure* of n dimensions or n -volume and also denote it $d^n x := \lambda^n$ (especially in the context of integration theory). It is a σ -finite measure, since one can partition $\mathbb{R}^n$ into countably many hypercubes of side 1, each of which has n -volume 1. λ^1 is called *length*, λ^2 *area* and λ^3 *volume*. $\diamond$

However, there are many other measures as well:

Example 30. Given an arbitrary set X ,

$$\begin{aligned} v : \mathcal{P}(X) &\longrightarrow [0, \infty], \\ A &\longmapsto v(A) = \begin{cases} |A| & \text{if } A \text{ is finite} \\ \infty & \text{otherwise} \end{cases} \end{aligned} \quad (5.2)$$

is a measure on $(X, \mathcal{P}(X))$ and we call it the *counting measure*. Also, for each $x_0 \in X$,

$$\begin{aligned} \delta_{x_0} : \mathcal{P}(X) &\longrightarrow [0, \infty], \\ A &\longmapsto \delta_{x_0}(A) = \begin{cases} 1 & \text{if } x_0 \in A \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (5.3)$$

is a measure on $(X, \mathcal{P}(X))$, and we call it the *Dirac measure at x_0* . $\diamond$

Here are some basic properties of σ -algebras:

Lemma 28. Let $\mathcal{A}$ be a σ -algebra on X . Then

(i) $X \in \mathcal{A}$

(ii) $A_k \in \mathcal{A}$ for $k \in \mathbb{N} \implies \bigcap_{k=1}^{\infty} A_k \in \mathcal{A}$

(iii) $A, B \in \mathcal{A} \implies A \setminus B \in \mathcal{A}$. $\diamond$

And some basic properties of measures:

Proposition 36. Let μ be a measure on $(X, \mathcal{A})$ and $A, B \in \mathcal{A}$. Then

(i) $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$.

(ii) if $A \subset B$, then $\mu(A) \leq \mu(B)$.

(iii) For $A_j \in \mathcal{A}$, $j \in \mathbb{N}$,

$$\mu\left(\bigcup_{j=1}^{\infty} A_j\right) \leq \sum_{j=1}^{\infty} \mu(A_j). \quad (5.4)$$

(iv) If $A_j \subset A_{j+1}$ then

$$\lim_{j \rightarrow \infty} \mu(A_j) = \mu\left(\bigcup_{j=1}^{\infty} A_j\right). \quad \diamond$$

5.1.1 Build New Measures Out of Known Ones: Product and Push-Forward

One defines the area of the rectangles as the product of the length of two of its orthogonal edges, and similarly, the volume of a parallelepiped is the product of the lengths of three of its orthogonal edges (or the product of the area of a face and an orthogonal edge). Hence, it feels like the notion of area is the “square” of the notion of length, while the volume is the “cube” of length. Similarly, given two measure spaces $(X_1, \mathcal{A}_1, \mu_1)$, $(X_2, \mathcal{A}_2, \mu_2)$ and two sets $B_1 \in \mathcal{A}_1$, $B_2 \in \mathcal{A}_2$, to the rectangle $B_1 \times B_2$, one would like to attribute measure $\mu_1(B_1) \cdot \mu_2(B_2)$.

Proposition 37. Given a measure space $(X_j, \mathcal{A}_j, \mu_j)$ for each $j \in \{1, \dots, n\}$ and some $n \in \mathbb{N}$, we call the σ -algebra on $X_1 \times \dots \times X_n$,

$$\mathcal{A}_1 \odot \dots \odot \mathcal{A}_n := \sigma \left(\left\{ B_1 \times \dots \times B_n \mid B_j \in \mathcal{A}_j \ \forall j \in \{1, \dots, n\} \right\} \right),$$

the *product σ -algebra* of $\mathcal{A}_1, \dots, \mathcal{A}_n$. There exists a measure $\mu_1 \odot \dots \odot \mu_n$ on $(X_1 \times \dots \times X_n, \mathcal{A}_1 \odot \dots \odot \mathcal{A}_n)$ such that

$$\mu_1 \odot \dots \odot \mu_n(B_1 \times \dots \times B_n) := \mu_1(B_1) \dots \mu_n(B_n) \quad \forall B_j \in \mathcal{A}_j, \ \forall j \in \{1, \dots, n\}.$$

If all μ_j are σ -finite, then there is a unique such measure. We then call it *the product measure of $\mu_1, \dots, \mu_n$* .

Example 31. For every $n \in \mathbb{N}$, $\mathfrak{B}(\mathbb{R}^n) = \mathfrak{B}(\mathbb{R}) \odot \dots \odot \mathfrak{B}(\mathbb{R})$ and $\lambda^n = \lambda^1 \odot \dots \odot \lambda^1$. ♦

The following is another way to build new measures, which finds many applications in probability theory.

Definition 72. Let $(X, \mathcal{A})$ and $(Y, \mathcal{C})$ be measurable spaces. A map $f : X \rightarrow Y$ is called *measurable* when the pre-image of every $\mathcal{C}$ -measurable set is $\mathcal{A}$ -measurable, i.e.,

$$f^{-1}(C) \in \mathcal{A} \quad \forall C \in \mathcal{C}. \tag{5.5} \quad \blacklozenge$$

Then, if the departure space X has a measure, one trivially gets a measure on Y by defining the measure of a set C in Y to be that of the set of X points mapping there.

Proposition 38. Given a measure space $(X, \mathcal{A}, \mu)$, for any measurable space $(Y, \mathcal{C})$ and measurable function $f : X \rightarrow Y$, the map

$$\begin{aligned} f_*\mu : \mathcal{C} &\longrightarrow [0, +\infty] \\ C &\longmapsto \mu(f^{-1}(C)) \end{aligned}$$

is a measure on Y and we call it the *push-forward of μ by f* . ♦

5.2 Lebesgue Integration

For this section it will be convenient to introduce:

Definition 73. Given a set X and $f : X \rightarrow \overline{\mathbb{R}}$, we will write the positive and negative parts of f as

$$f_+(x) := \max(0, f(x)) \quad \text{and} \quad f_-(x) := \max(0, -f(x)) \quad \forall x \in X,$$

such that $f = f_+ - f_-$. ♦

Now, let us define a function that is 0 everywhere except for a given set:

Definition 74. Given a set X and a subset $A \subseteq X$, we define the *indicator function of A* to be:

$$\begin{aligned} \mathbb{1} : X &\longrightarrow \{0, 1\} \\ x &\longmapsto \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases} . \end{aligned}$$

It is also often denoted as $\chi_A := \mathbb{1}_A$. ♦

This allows to define arbitrary “bar chart”/“stair function”-looking functions:

Definition 75. Given a set X , we call a function $g : X \rightarrow \overline{\mathbb{R}}$ for which there exists $n \in \mathbb{N}$, disjoint sets $A_1, \dots, A_n \subseteq X$ and different values $\alpha_1, \dots, \alpha_n \in \overline{\mathbb{R}}$ such that

$$g := \sum_{j=1}^n \alpha_j \mathbb{1}_{A_j},$$

a *simple function*. Equivalently, a simple function is a map $g : X \rightarrow \overline{\mathbb{R}}$ with finite range —just define $n := |\text{range}(g)|$, enumerate $\text{range}(g) =: \{\alpha_1, \dots, \alpha_n\}$ and take $A_j := g^{-1}(\alpha_j) \forall j \in \{1, \dots, n\}$. We call this last choice of α_j and A_j the *canonical representation of g* . Trivially, g_+ and g_- are also simple functions.

If $\mathcal{A}$ is a σ -algebra on X , note that demanding that the canonical sets A_j are measurable (i.e., $g^{-1}(\alpha_j) \in \mathcal{A}$) is exactly to demand that g is a measurable function. $\diamond$

For non-negative simple functions g , there is an obvious way to define the “volume under the bars of g ” or integral of g as the sum of the “heights times base’s area”. Then, for arbitrary simple functions, we just need to compute the integral of the positive part and negative part separately and then subtract the results —unless they are not both infinity, since such a subtraction would be ill-defined.

Definition 76. Given a measure space $(X, \mathcal{A}, \mu)$ and a measurable simple function $g : X \rightarrow [0, +\infty]$ in canonical form $g = \sum_{j=1}^n \alpha_j \mathbb{1}_{A_j}$, we define

$$\int_{x \in X} g(x) d\mu := \sum_{j=1}^n \alpha_j \mu(A_j)$$

where $+\infty \cdot (+\infty) := +\infty$ and $0 \cdot \infty := 0$ (this always yields a number in $[0, +\infty]$). Then, for a general simple function $h : X \rightarrow \overline{\mathbb{R}}$, if

$$P := \int_{x \in X} h_+(x) d\mu < +\infty \quad \text{or} \quad N := \int_{x \in X} h_-(x) d\mu < +\infty,$$

we define

$$\int_{x \in X} h(x) d\mu := P - N,$$

where $\pm\infty + a := a \pm \infty := \pm\infty \forall a \in \mathbb{R}$. $\diamond$

5.2.1 The Difficulty with the Riemann Integral

The notion of integration for functions $f : [a, b] \rightarrow \mathbb{R}$ usually taught in calculus courses is called *Riemann integration*. It essentially consists of partitioning the domain $[a, b]$, choosing a point in each part and evaluating f over it, then consider the sum of areas of the bars in the resulting “bar-chart” approximating f and take the limit as we make the partition finer and finer:

Definition 77. Given $a, b \in \mathbb{R} : a < b$, we define the set of *tagged partitions* of $[a, b]$ as

$$P[a, b] := \left\{ \left((I_1, \dots, I_n), (t_1, \dots, t_n) \right) \mid \begin{array}{l} n \in \mathbb{N}, \quad I_n := [x_{n-1}, x_n], \quad I_j := [x_{j-1}, x_j] \quad \forall j \in \{1, \dots, n-1\} \\ a = x_0 < x_1 < \dots < x_{n-1} < x_n = b, \quad t_j \in I_j \end{array} \right\}.$$

Each element is a choice of partition for $[a, b]$ and a choice of *tag* t_j in each part. Given a function $f : [a, b] \rightarrow \mathbb{R}$ and some $J := ((I_1, \dots, I_n), (t_1, \dots, t_n)) \in P[a, b]$, we define their *Riemann sum* $S(f, J)$ as the signed area between the x axis and the simple function

$$g(f, J) := \sum_{j=1}^n f(t_j) \mathbb{1}_{I_j}$$

that approximates f , i.e.,

$$S(f, J) := \int_{x \in [a, b]} g(f, J)(x) dx = \sum_{j=1}^n f(t_j) \lambda^1(I_j) = \sum_{j=1}^n f(t_j)(x_j - x_{j-1}). \quad \diamond$$

Lemma 29. Define the *refinement* relation $\leq$ in $P[a, b]$ such that

$$T := \left((I_j)_{j=1}^n, (t_j)_{j=1}^n \right), R := \left((K_j)_{j=1}^m, (r_j)_{j=1}^m \right) \in P[a, b],$$

$$T \leq R \quad :\Longleftrightarrow \quad \forall \ell \in \{1, \dots, m\} \exists j \in \{1, \dots, n\} : I_j \supseteq K_\ell, \quad (5.6)$$

i.e., when the parts of R 's partition are subsets of T 's —it is said that R is a *finer* partition than T , or a *refinement*. This relation makes $(P[a, b], \leq)$ a *directed set*. $\blacklozenge$

Definition 78. We call $f : [a, b] \rightarrow \mathbb{R}$ *Riemann integrable* when the net of Riemann sums $(S(f, J))_{J \in P[a, b]}$ converges as partitions become finer. If so, we call the limit point the *Riemann integral* of f and denote

$$\int_{x \in [a, b]}^R f(x) dx := \lim_{J \in P[a, b]} S(f, J).$$

(Thus, it is the limit of step function approximations over finer and finer partitions of the domain.) $\blacklozenge$

Imagine we want to generalize this to now integrate $f : \mathbb{R} \rightarrow \mathbb{R}$ over any subset B of $\mathbb{R}$ with well-defined length (say, any set in $\mathfrak{B}(\mathbb{R})$). For that, we would now define the family of tagged partitions $P[B]$ as the finite partitions of B by measurable sets (so their length is defined) and a choice of tag per each such set. Then, we could define the integral of f over B exactly as above. An issue would be that if B has infinite length, a finite partition will always have at least one part with infinite measure and thus, the Riemann sums would be generically infinite —unless the function is zero in that infinite set. Hence, there would be essentially no function with finite integral over sets with infinite length. Even worse, we could face a problem of needing to add $+\infty$ and $-\infty$ if there are two infinite sets where the function is positive and negative respectively.

Still, as long as $\mu(X) < +\infty$ in a measure space $(X, \mathcal{A}, \mu)$, we could easily generalize the prescription to measurable $f : X \rightarrow \mathbb{R}$ avoiding all such problems. Take $P[X]$ to be the family of finite partitions of X by measurable sets with tags,

$$P[X] := \left\{ \left((I_j)_{j=1}^n, (t_j)_{j=1}^n \right) \mid n \in \mathbb{N}, (I_j)_{j=1}^n \subseteq \mathcal{A} \text{ partition of } X, t_j \in I_j \right\},$$

direct them by refinement (exactly as in Eq. (5.6)) and define the integral as the limit of the resulting net of generalized Riemann sums (if it exists):

$$\int_{x \in X}^R f(x) d\mu := \lim_{J \in P[X]} \int_{x \in X} g(f, J)(x) d\mu.$$

The problem with this notion of integrability is that the net has finite limit *if and only if* f is what we will call *essentially bounded* (i.e., an element of $L^\infty(X, \mathbb{R}, \mu)$ —to be defined).^[a] Consequently, the integrable functions $f : X \rightarrow \mathbb{R}$ for which

$$\|f\|_{L^p}^R := \left(\int_{x \in X}^R |f(x)|^p d\mu \right)^{1/p} < +\infty$$

(which is a generalization of the p -norms) do not generally make a complete normed vector space when equipped with $\|\cdot\|_{L^p}^R$.^{[b][c]} In particular, there are $\|\cdot\|_{L^p}^R$ -Cauchy sequences that do not converge in $\|\cdot\|_{L^p}^R$ (despite the point-wise limit might be well-defined). In particular, there will be no monotone or dominated convergence theorems.

^[a] And if so, the limit will equal the Lebesgue integral of f in X . Thus, in $L^\infty(X, \mathbb{R}, d\mu)$, the generalized Riemann integral and the Lebesgue integral agree.

^[b] Not even after identifying “almost everywhere equal” functions —to be explained below.

^[c] Because $\mu(X) < \infty$, $L^\infty(X, \mathbb{R}, d\mu) \subseteq L^p(X, \mathbb{R}, d\mu)$, but in general, not the converse. However, $L^\infty(X, \mathbb{R}, d\mu)$ is dense in $L^p(X, \mathbb{R}, d\mu)$ in the $\|\cdot\|_{L^p}$ -norm, so, instead of using the Lebesgue integral, one can alternatively define $L^p(X, \mathbb{R}, d\mu)$ as the completion of $L^\infty(X, \mathbb{R}, d\mu)$ in $\|\cdot\|_{L^p}^R$ norm. It is in this sense that the notion of Lebesgue integral is the “completion” of that of Riemann integral.

The problem with this is that those L^p -norms, particularly the $p = 1$ and 2 cases are the norms that matter for statistical and quantum mechanics, so we would get a bad mathematical structure for physics.

Instead, mathematicians noticed that one gets a more robust notion of integral if we consider a construction first for non-negative functions and then for arbitrary and importantly, because that allows to define the integral using a supremum rather than a net limit. For $f : [a, b] \rightarrow [0, +\infty)$ one could define its integral as the supremum of the integrals of simple functions that “live” under the curve:

$$\int_{x \in [a, b]} f(x) dx := \sup \left\{ \int_{x \in [a, b]} g(x) dx \mid g : [a, b] \rightarrow [0, +\infty) \text{ simple, measurable and } g(x) \leq f(x) \forall x \right\}.$$

Since this set is non-empty (the zero function is a simple function), it is well-defined and we saw that in the extended real line, every set has a supremum, so it always gives a definite number in $[0, +\infty]$. Then, for arbitrary measurable $f : [a, b] \rightarrow \mathbb{R}$ one may split f as a sum of the positive and negative parts and integrate them separately. If they both yield infinity, then we would have an indetermination, but otherwise a clear result. It turns out that doing that

Proposition 39. If $f : [a, b] \rightarrow \mathbb{R}$ is Riemann integrable, then $\int_{x \in [a, b]} f_{\pm}(x) d\mu < +\infty$ and

$$\int_{x \in [a, b]}^R f(x) dx = \int_{x \in [a, b]} f_+(x) dx - \int_{x \in [a, b]} f_-(x) dx. \quad \blacklozenge$$

Interestingly the converse is not true, there are more functions that have well-defined integral in the supremum sense than Riemann sense (for instance $\mathbb{1}_{\mathbb{Q} \cap [0, 1]}$ is not Riemann integrable, but is integrable in the supremum sense). Precisely enough so that we get that all “ L^p norm spaces” are complete.

5.2.2 Understanding the Lebesgue Integral

The generalization of integration using the supremum definition is what we call Lebesgue integration. Since we will allow integrals to yield infinity, we will allow that functions are valued infinity (so that we can integrate also functions with singularities). In particular, we will use $+\infty \cdot (+\infty) := +\infty$ and $0 \cdot (+\infty) = 0$.

Definition 79. Let $(X, \mathcal{A}, \mu)$ be a measure space and let $f : X \rightarrow [0, +\infty]$ be measurable. Then, we define

$$\int_{x \in X} f(x) d\mu = \sup \left\{ \int_{x \in X} g(x) d\mu \mid g : X \rightarrow [0, \infty] \text{ simple, measurable and } g(x) \leq f(x) \forall x \right\}. \quad (5.7)$$

Then, for a measurable $g : X \rightarrow \overline{\mathbb{R}}$ we call it *integrable* when

$$\int_{x \in X} g_+(x) d\mu < +\infty \quad \text{and} \quad \int_{x \in X} g_-(x) d\mu < +\infty, \quad (5.8)$$

and for such functions we define

$$\int_{x \in X} g(x) d\mu := \int_{x \in X} g_+(x) d\mu - \int_{x \in X} g_-(x) d\mu.$$

Finally, we call a measurable $\psi : X \rightarrow \mathbb{C}$ *integrable* when its real and imaginary parts are integrable $\mathbb{R}$ -valued functions and if so, we define

$$\int_{x \in X} \psi(x) d\mu := \int_{x \in X} \operatorname{Re}\{\psi(x)\} d\mu + i \int_{x \in X} \operatorname{Im}\{\psi(x)\} d\mu. \quad \blacklozenge$$

You will also see the integrals denoted alternative but equivalently, as

$$\int_{x \in X} \psi(x) d\mu \equiv \int_X \psi(x) d\mu(x) \equiv \int_X \psi d\mu.$$

Lemma 30. In a measure space $(X, \mathcal{A}, \mu)$, a measurable function $\psi : X \rightarrow \mathbb{C}$ is *integrable* (both as a complex valued or $\mathbb{R}$ -valued function) *if and only if*

$$\|\psi\|_{L^1} := \int_{x \in X} |\psi(x)| d\mu < +\infty. \quad \blacklozenge$$

Now, a supremum is hard to be used in many partial calculations and hence, not only to understand better what Lebesgue integration is, but for actual proofs, the following is very practical:

Proposition 40. Given a measure space $(X, \mathcal{A}, \mu)$, every measurable function $f : X \rightarrow [0, +\infty]$ is the point-wise limit of a monotonically increasing sequence of non-negative simple functions, i.e. $\exists (s_n)_{n \in \mathbb{N}}$ with each $s_n : X \rightarrow [0, +\infty]$ simple and measurable, such that for all $x \in X$,

$$(i) \quad s_n(x) \leq s_{n+1}(x) \leq f(x) \quad \forall n \in \mathbb{N} \quad \text{and} \quad (ii) \quad s_n(x) \xrightarrow{n \rightarrow \infty} f(x).$$

Explicitly, such a sequence is given by

$$s_n := \sum_{k=1}^{2^{2n}+1} \frac{k-1}{2^n} \mathbb{1}_{A_{n,k}} \quad \begin{cases} I_{n,k} := [\frac{k-1}{2^n}, \frac{k}{2^n}) & \forall k \in \{1, 2, \dots, 2^{2n}\} \\ I_{n, 2^{2n}+1} := [2^n, +\infty] \\ A_{n,k} := f^{-1}(I_{n,k}), \end{cases}$$

which is a partition of the co-domain that becomes finer as n increases and looking at the pre-images of each part, gives an each time better point-wise approximation of f from below. $\blacklozenge$

Then, the monotone convergence theorem (to be seen in a moment), allows to prove that:

Corollary 9. Given a measure space $(X, \mathcal{A}, \mu)$ and $f : X \rightarrow [0, +\infty]$ measurable,

$$\int_{x \in X} f(x) d\mu = \lim_{n \rightarrow \infty} \int_{x \in X} s_n(x) d\mu,$$

for one (and hence any) of the simple function sequences of Prop. 40. $\blacklozenge$

Therefore, one can equivalently define Lebesgue integration instead of using the supremum, as the limit of the integrals of monotonic simple function sequences that approximate f point-wise from below (then, for $\mathbb{R}$ - and $\mathbb{C}$ -valued functions we proceed as before). This provides a more graphical understanding of Lebesgue integration. In this sense, it is said that while Riemann integration is defined by taking finer and finer partitions of the *domain* of a function (then taking the integrals of simple functions over them), Lebesgue integration is based on taking instead refinements of partitions of the *co-domain*.

Finally, let us give some important properties of the Lebesgue integral.

Proposition 41. Let $f, g : X \rightarrow \mathbb{C}$ be integrable and $\alpha \in \mathbb{C}$ (equivalently, if f, g $\mathbb{R}$ -valued). Then, the integral is linear:

$$(i) \quad \int_X \alpha f d\mu = \alpha \int_X f d\mu$$

$$(ii) \quad \int_X (f + g) d\mu = \int_X f d\mu + \int_X g d\mu.$$

Furthermore, it satisfies the following triangle-inequality and monotonicity-like properties:

$$(iii) \quad \left| \int_X f d\mu \right| \leq \int_X |f| d\mu$$

$$(iv) \quad \text{If } f, g \text{ are } \mathbb{R}\text{-valued and } f(x) \leq g(x) \quad \forall x \in X \implies \int f d\mu \leq \int g d\mu. \quad \blacklozenge$$

5.2.3 Almost Everywhere-ness and Sampling

Definition 80. Given a measurable space $(X, \mathcal{A})$, we say that with respect to a measure μ a property holds *almost everywhere* (a.e.), or holds for *almost every point* $x \in X$ (a.e. $x \in X$), when the set of exceptions has measure zero, i.e., when it only fails to hold in a null set. More precisely, when there is $A \in \mathcal{A}$ where the property holds for all $x \in A$ and $\mu(X \setminus A) = 0$. ♦

Each measure μ gives a different notion of *almost everywhere-ness*, so one often uses “ μ -almost every” to clarify the employed measure in the claim.

This is the sense in which one makes mathematical the notion of *typicality*: we say that a property for points in a measure space $(X, \mathcal{A}, \mu)$ is *typical* when it holds for μ -almost every point. For example, since $\lambda^1(\mathbb{Q}) = 0$ (as you will prove in the exercises), $\lambda^1([0, 1]) = \lambda^1([0, 1] \setminus \mathbb{Q})$ and thus, with respect to length, almost-every real number in $[0, 1]$ is irrational. We say that a “typical” number in $[0, 1]$ is irrational. In this line, one finds how to use measure theory to formalize probability theory after the following:

Informally, given a measure space $(X, \mathcal{A}, \mu)$ with $\mu(X) = 1$, a countable set of “samples” $(x_n)_{n \in \mathbb{N}} \subseteq X$ is said to be (*randomly*) *sampled following the (probability) distribution* of μ when the relative number of them in each relevant set tends to the measure given by μ , i.e., when

$$\mu(B) = \lim_{N \rightarrow \infty} \frac{|\{n \in \{1, \dots, N\} : x_n \in B\}|}{N}$$

for all relevant $B \in \mathcal{A}$ (although not necessarily for all $B \in \mathcal{A}$).

It is in this sense that the Lebesgue measure gives a “uniform distribution”. By translation invariance it gives the same measure to a set translated around the space. Hence, a sequence of points $(x_n)_{n \in \mathbb{N}}$ in say, $[0, 1]$ is sampled according to the Lebesgue measure exactly when the relative frequency of them falling in a set $(0, 0.1)$ is the same as for $(0.2, 0.3)$ or $(0.3, 0.4)$ etc. It is “equally likely” that they lie in any interval of the same length.

On the other hand, “almost-everywhere”-ness allows us to identify functions up to differences in null sets, and this allows us to get rid of potential singularities.

Proposition 42. Let $(X, \mathcal{A}, \mu)$ be a measure space and $f : X \rightarrow [0, +\infty]$ measurable. Then,

$$\int_{x \in X} f(x) d\mu = 0 \iff f(x) = 0 \quad \mu\text{-a.e. } x \in X.$$

Thus, when $h, g : X \rightarrow \mathbb{C}$ (or to $\mathbb{R}$) are integrable,

$$\int_{x \in X} |g(x) - h(x)| d\mu = 0 \iff g(x) = h(x) \quad \text{-a.e. } x \in X$$

If so, we say that g, h *agree almost everywhere* and write $g = h$ a.e. In particular, this defines an *equivalence relation* (both within $\mathbb{R}$ - and $\mathbb{C}$ -valued functions). We will denote the equivalence class of g as $[g]_{a.e.}$. Note that in particular, it implies

$$\int_{x \in X} g(x) d\mu = \int_{x \in X} h(x) d\mu. \quad \blacklozenge$$

This tells us in particular, that if we change the value of a function $f : X \rightarrow [0, +\infty]$ in a null set, the resulting function will have exactly the same integral. Now, since by Corollary 10,^[d] an integrable function can be infinite at most in a null set, *each equivalence class of a.e. equal functions has at least one representative that is everywhere finite.*

^[d]Take $A_1 = |f|^{-1}(\{\pm\infty\})$ and $A_2 = X \setminus A_1$.

5.2.4 Density Functions and Radon-Nikodym Derivative

Integrals give us a way to obtain new measures by interpreting a function as a “density”:

Definition 81. Given a measurable space $(X, \mathcal{A})$, a measurable function $\rho : X \rightarrow [0, +\infty]$ and a measure^[e] $d\mu$, we denote by $\rho d\mu$ the mapping $\mathcal{A} \rightarrow [0, +\infty]$ such that

$$\rho d\mu(B) := \int_{x \in B} \rho(x) d\mu, \quad B \in \mathcal{A}. \quad \blacklozenge$$

Lemma 31. In the above situation, whichever the measurable function $\rho : X \rightarrow [0, +\infty]$ is, $d\nu := \rho d\mu$ is *still* a measure on $(X, \mathcal{A})$. We call ρ the *density* of the measure $d\nu$ with respect to $d\mu$ and call it the Radon-Nikodym derivative of ν by μ , denoted $\frac{d\nu}{d\mu} := \rho$. $\blacklozenge$

Lemma 31 proves by definition of measure that:

Corollary 10. Given a measure space $(X, \mathcal{A}, \mu)$, a measurable function $f : X \rightarrow [0, +\infty]$ (also an integrable function to either $\overline{\mathbb{R}}$ or $\mathbb{C}$) satisfies that given pairwise disjoint $A_n \in \mathcal{A}$ for $n \in \mathbb{N}$,

$$\int_{x \in \bigcup_{n \in \mathbb{N}} A_n} f(x) d\mu = \sum_{n=1}^{\infty} \int_{x \in A_n} f(x) d\mu.$$

In general, each result of Prop. 36 for measures gives a useful property for integrals. $\blacklozenge$

Example 32. Here is an example showing that not all Lebesgue integrable functions are Riemann integrable (despite the all Riemann are Lebesgue integrable). The *Dirichlet function* $f : [0, 1] \rightarrow \mathbb{R}$, $f(x) := \chi_{\mathbb{Q} \cap [0, 1]}(x)$ is famously non-Riemann integrable, but it is Lebesgue integrable and

$$\int_{x \in [0, 1]} f(x) dx \stackrel{Cor. 10}{=} 1 \cdot \lambda(\mathbb{Q} \cap [0, 1]) + 0 \cdot \lambda([0, 1] \setminus \mathbb{Q}) = 0. \quad \blacklozenge$$

Now, Lemma 31 proves that there are measures (like $d\nu$) that can be written as a density times another measure, which we may call a *background measure*. We may then wonder, when such a thing is possible, and for σ -finite measures there is an exact characterization called the *Radon-Nikodym Theorem*.

Definition 82. Given a measurable space $(X, \mathcal{A})$ and two measures $d\mu_1, d\mu_2$:

- $d\mu_1$ is *absolutely continuous* (a.c.) with respect to $d\mu_2$, denoted $d\mu_1 \ll d\mu_2$, when

$$d\mu_2(B) = 0 \implies d\mu_1(B) = 0 \quad \forall B \in \mathcal{A}.$$

Equivalently

$$d\mu_1(B) > 0 \implies d\mu_2(B) > 0 \quad \forall B \in \mathcal{A}.$$

- $d\mu_1, d\mu_2$ are *mutually absolutely continuous*, denoted $d\mu_1 \sim d\mu_2$, when $d\mu_1 \ll d\mu_2$ and $d\mu_2 \ll d\mu_1$, i.e., when they agree on which sets have zero and non-zero measure.
- $d\mu_1, d\mu_2$ are *mutually singular*, denoted $d\mu_1 \perp d\mu_2$ when $\exists B \in \mathcal{A}$ such that $d\mu_1(B) = 0$ and $d\mu_2(X \setminus B) = 0$ (i.e., they have disjoint “supports”). $\blacklozenge$

Theorem 18. Given a measurable space $(X, \mathcal{A})$ and two σ -finite measures $d\mu_1, d\mu_2$:

$$d\mu_1 \ll d\mu_2 \iff \exists \rho : X \rightarrow [0, +\infty] \text{ measurable s.th. } d\mu_1 = \rho d\mu_2.$$

Moreover, ρ is *unique* (up to changes over $d\mu_2$ -null sets). We denote $\rho =: \frac{d\mu_1}{d\mu_2}$. $\blacklozenge$

^[e]Remember that the “d” in front of a measure has no mathematical meaning, it is just nota ggf v gftionally convenient in the context of Lebesgue integration, i.e., $d\mu \equiv \mu$.

5.2.5 Convergence Theorems and L^p -spaces

One of the big advantages of Lebesgue integration with respect to other generalizations is the following theorem, called the *monotone convergence* or *Levi theorem*. It gives a criterion of when one can exchange taking a limit and taking the integral.

Theorem 19. Given a measure space $(X, \mathcal{A}, \mu)$ and a sequence $(f_n)_{n \in \mathbb{N}}$ with $f_n : X \rightarrow [0, +\infty]$ measurable, if they are monotonously increasing almost everywhere, i.e.,

$$f_n(x) \leq f_{n+1}(x) \quad \text{a.e } x \in X \quad \forall n \in \mathbb{N},$$

then, there is a measurable $f : X \rightarrow [0, +\infty]$ that is a.e the point-wise limit of f_n , i.e.,

$$f(x) := \lim_{n \rightarrow \infty} f_n(x) \quad \text{a.e } x \in X$$

and it satisfies

$$\lim_{n \rightarrow +\infty} \int_{x \in X} f_n(x) d\mu = \int_{x \in X} f(x) d\mu \quad \diamond$$

As a corollary, for general sequences one gets the following result, called *Fatou's lemma*:

Definition 83. Given a sequence $(a_n)_{n \in \mathbb{N}} \subseteq \mathbb{R}$, one defines

$$\liminf_{n \rightarrow +\infty} a_n := \lim_{n \rightarrow +\infty} \left(\inf_{m \in \{n, n+1, \dots\}} a_m \right), \quad \limsup_{n \rightarrow +\infty} a_n := \lim_{n \rightarrow +\infty} \left(\sup_{m \in \{n, n+1, \dots\}} a_m \right)$$

(They always exist as numbers in $\overline{\mathbb{R}}$.) ♦

Corollary 11. Given a measure space $(X, \mathcal{A}, \mu)$ and a sequence $f_n : X \rightarrow [0, \infty]$ ($n \in \mathbb{N}$) of measurable functions, defining

$$f^{inf}(x) := \liminf_{n \rightarrow \infty} f_n(x) \quad \forall x \in X,$$

then,

$$\int_{x \in X} f^{inf}(x) d\mu \leq \liminf_{n \rightarrow \infty} \int f_n d\mu. \quad \diamond$$

Now, the main reason why we use Lebesgue integration is the following:

Proposition 43. Let $(X, \mathcal{A}, \mu)$ be a measure space and $p \in [1, \infty)$. Define for $f : X \rightarrow \mathbb{C}$ measurable

$$\|f\|_{L^p} := \left(\| |f|^p \|_{L^1} \right)^{1/p} = \left(\int_{x \in X} |f(x)|^p d\mu \right)^{1/p}.$$

Then, the set

$$\begin{aligned} \mathcal{L}^p(X, \mu) &:= \left\{ f : X \rightarrow \mathbb{C} \mid f \text{ measurable and } |f|^p \text{ integrable} \right\} = \\ &= \left\{ f : X \rightarrow \mathbb{C} \mid f \text{ measurable and } \|f\|_{L^p} < +\infty \right\}, \end{aligned} \quad (5.9)$$

is a vector space. $\|\cdot\|_{L^p}$ satisfies homogeneity and triangle inequality but not definiteness because both $f(x) := 0 \quad \forall x \in X$ and any other $g \in [f]_{a.e}$ yields $\|f\|_{L^p} = 0 = \|g\|_{L^p}$. Yet, in

$$L^p(X, \mu) := \left\{ [f]_{a.e} \mid f \in \mathcal{L}^p(X, \mu) \right\}$$

the operations for $\mu, \lambda \in \mathbb{C}$, $[f]_{a.e}, [g]_{a.e} \in L^p(X, \mu)$ given by

$$\mu[g]_{a.e} + \lambda[f]_{a.e} := [\mu g + \lambda f]_{a.e}, \quad \text{and} \quad \|[f]\|_{L^p} := \|f\|_{L^p}$$

are well-defined (independent of the representative) and they make $L^p(X, \mu)$ a normed vector space. ♦

Proposition 44. Given $(X, \mathcal{A}, \mu)$ a measure space and $f : X \rightarrow \mathbb{C}$ measurable, we define its *essential supremum* as

$$\|f\|_{L^\infty} := \inf \left\{ \lambda \in [0, +\infty] : \mu(|f|^{-1}((\lambda, \infty])) = 0 \right\} =: \text{ess sup } |f|. \quad (5.10)$$

It represents the supremum of f up to null sets. Indeed, there is always some $g \in [f]_{a.e}$ for which $\|g\|_\infty = \|f\|_{L^\infty}$. As in the last, proposition, we then define

$$\mathcal{L}^\infty(X, \mu) := \left\{ f : X \rightarrow \mathbb{C} \mid f \text{ measurable, } \|f\|_{L^\infty} < +\infty \right\} \quad (5.11)$$

and then,

$$L^\infty(X, \mu) := \left\{ [f]_{a.e} \mid f \in \mathcal{L}^\infty(X, \mu) \right\} \quad (5.12)$$

makes a normed vector space with $\|[f]_{a.e}\|_{L^\infty} := \|f\|_{L^\infty}$. $\blacklozenge$

Theorem 20. For any measure space $(X, \mathcal{A}, \mu)$ and $p \in [1, +\infty]$, $(L^p(X, \mu), \|\cdot\|_{L^p})$ is a Banach space. $\blacklozenge$

Very important abuse of notation: We write an elements of an $L^p(X, \mu)$ space always as the representative of a class (i.e., instead of $[f]_{a.e}$ we use f) and just keep in the back of our minds that technically, it is an equivalence class and not a function.

Now, here is the other big theorem telling us when we can exchange limit and integral signs: the *dominated convergence theorem*.

Theorem 21. Given a measure space $(X, \mathcal{A}, \mu)$ and a sequence $(f_n)_{n \in \mathbb{N}}$ with $f_n : X \rightarrow \mathbb{C}$ measurable, if $\exists g \in L^p(X, \mu)$ for some $p \in [1, +\infty)$ dominating the sequence, i.e.,

$$|f_n(x)| \leq |g(x)| \quad \text{a.e. } x \in X \quad \forall n \in \mathbb{N},$$

and if $\exists f : X \rightarrow \mathbb{C}$ such that

$$f(x) = \lim_{n \rightarrow \infty} f_n(x) \quad \text{a.e. } x \in X,$$

then,

(i) $f_n, f \in L^p(X, \mu)$

(ii) $\lim_{n \rightarrow \infty} f_n = f$ in the $\|\cdot\|_{L^p}$ -norm topology, i.e.,

$$\lim_{n \rightarrow \infty} \|f_n - f\|_{L^p} = 0 \iff \lim_{n \rightarrow \infty} \int_{x \in X} |f_n(x) - f(x)|^p d\mu = 0.$$

(iii) By the reverse triangle inequality, (ii) implies that

$$\lim_{n \rightarrow \infty} \int_{x \in X} |f_n(x)|^p d\mu = \int_{x \in X} |f(x)|^p d\mu,$$

(iv) In the case $p = 1$,

$$\lim_{n \rightarrow \infty} \int_{x \in X} f_n(x) d\mu = \int_{x \in X} f(x) d\mu. \quad \blacklozenge$$

Remark 2. If μ is a finite measure ($\mu(X) < \infty$), then, trivially, any constant function is in all L^p 's. Hence, in such a case, it is enough that the sequence of functions is bounded uniformly a.e in X to be able to use the dominated convergence theorem. $\blacklozenge$

Lastly, the following result is called the Hölder inequality (Hölder was a mathematician here in Tübingen!). It is important because it encodes the inter-play between the different L^p spaces, which in general are not contained in each other (unless μ is a finite measure —fact you will prove in the exercises).

Theorem 22. Let $p, q, r \in [1, \infty]$ be such that $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$, where $\frac{1}{\infty} := 0$. If $f \in L^p(X, \mu)$ and $g \in L^q(X, \mu)$, then,

$$\|fg\|_{L^r} \leq \|f\|_{L^p} \cdot \|g\|_{L^q},$$

and thus, $fg \in L^r(X, \mu)$. $\diamond$

Remark 3. The $p = q = 2, r = 1$ of the Hölder inequality proves the so-called *Cauchy-Schwarz inequality* on the space $L^2(X, \mu)$:

$$\left| \int_{x \in X} \overline{f(x)} g(x) d\mu \right| \leq \int_{x \in X} |\overline{f(x)} g(x)| d\mu \leq \|f\|_{L^2} \cdot \|g\|_{L^2}. \quad (5.13)$$

The relevance of this will become clear in the chapter on Hilbert spaces. It also shows that if $f, g \in L^2(X, \mu)$, then, $\overline{f}g \in L^1(X, \mu)$. $\diamond$

5.2.6 Extra: Relation Between Function Convergence Types

In this extra section we will relate different notions of convergence that are used in the context of measure theory, just for you to have a reference. You may skip the section.

Definition 84. Let (X, μ) be a measure space. $(f_n)_{n \in \mathbb{N}}$ with $f_n : X \rightarrow \mathbb{C}$ measurable is said to

- (a) converge to $f : X \rightarrow \mathbb{C}$ (*pointwise almost everywhere* (AE)) when $f_n(x) \xrightarrow[n \rightarrow \infty]{} f(x)$ a.e. $x \in X$.
- (b) converge to $f : X \rightarrow \mathbb{C}$ (*almost uniformly* (AU)) when for every $\varepsilon > 0$, there exists $E_\varepsilon \in \mathcal{A}$ such that $\mu(E_\varepsilon) \leq \varepsilon$ and

$$f_n|_{X \setminus E_\varepsilon} \longrightarrow f|_{X \setminus E_\varepsilon} \quad \text{uniformly.}$$

- (c) converge to f (*in measure* (M)) when

$$\forall \varepsilon > 0, \quad \mu\left(\{x \in X : |f(x) - f_n(x)| \geq \varepsilon\}\right) \xrightarrow[n \rightarrow \infty]{} 0. \quad \diamond$$

In the following diagrams we summarize the interplay between these convergence notions. *Solid arrow* means that if the convergence at the tail of the arrow holds, automatically the convergence at the tip of the arrow holds too. *Dashed arrow* means that if convergence at the tail of the arrow holds, then there exists a subsequence that convergence in the way given by the arrow's tip. The absence of arrow means there exists a counter example for such an arrow. Whenever L^p convergence is involved (which we denote by L^p), it is assumed that the functions are in $L^p(X, \mu)$ and we mean $p \in [1, \infty)$.

Theorem 23. For a general measure space $(X, \mathcal{A}, \mu)$ and a sequence $(f_n)_{n \in \mathbb{N}}$, $f_n : X \rightarrow \mathbb{C}$ measurable,

$$\begin{array}{ccc} AE & \longleftarrow & AU \\ \uparrow & \nwarrow & \uparrow \\ L^p & \longrightarrow & M \end{array} \quad (5.14)$$

- If the measure is finite, i.e., if $\mu(X) < +\infty$, then,

$$\begin{array}{ccc} AE & \longleftrightarrow & AU \\ \uparrow & \nwarrow & \uparrow \\ L^p & \longrightarrow & M \end{array} \quad (5.15)$$

- Under L^p domination, i.e., if $\exists g \in L^p(X, \mu) : |f_n(x)| \leq |g(x)|$ a.e. $x \in X$, then,

$$\begin{array}{ccc} AE & \longleftrightarrow & AU \\ \uparrow & \nwarrow & \uparrow \\ L^p & \longrightarrow & M \end{array} \quad (5.16) \quad \diamond$$

5.3 The Order of Iterated Integration

Recall that for σ -finite measures there is a unique product measure. In such a case, under very moderate conditions, one can choose to integrate first in one variable and then in the other one and also to exchange the order of integration. For non-negative functions for instance, it is automatically allowed: it is called *Tonelli's theorem*.

Theorem 24. Let $(X_1, \mathcal{A}_1, \mu_1)$, $(X_2, \mathcal{A}_2, \mu_2)$ be σ -finite measure spaces. Then, for arbitrary $\mathcal{A}_1 \odot \mathcal{A}_2$ -measurable $f : X_1 \times X_2 \rightarrow [0, +\infty]$, the partially evaluated function $f(x_1, \cdot)$ is $\mathcal{A}_2$ -measurable $\forall x_1 \in X_1$ and $f(\cdot, x_2)$ is $\mathcal{A}_1$ -measurable $\forall x_2 \in X_2$. Furthermore,

$$\begin{aligned} & \int_{(x_1, x_2) \in X_1 \times X_2} f(x_1, x_2) d(\mu_1 \odot \mu_2) = \\ &= \int_{x_1 \in X_1} \left(\int_{x_2 \in X_2} f(x_1, x_2) d\mu_2 \right) d\mu_1 = \int_{x_2 \in X_2} \left(\int_{x_1 \in X_1} f(x_1, x_2) d\mu_1 \right) d\mu_2, \end{aligned}$$

where the partially integrated functions are also measurable. $\diamond$

For the rest of functions, we only need to check that the function is integrable. In fact, it is enough that the partially evaluated functions are iteratedly integrable: it is called *Fubini's theorem*.

Theorem 25. Let $(X_1, \mathcal{A}_1, \mu_1)$, $(X_2, \mathcal{A}_2, \mu_2)$ be σ -finite measure spaces. Then, for arbitrary $\mathcal{A}_1 \odot \mathcal{A}_2$ -measurable $f : X_1 \times X_2 \rightarrow \mathbb{C}$, f is integrable, i.e., $f \in L^1(X_1 \times X_2, \mu_1 \odot \mu_2)$, if and only if at least one holds:

$$\int_{x_1 \in X_1} \left(\int_{x_2 \in X_2} |f(x_1, x_2)| d\mu_2 \right) d\mu_1 < +\infty \quad \text{or} \quad \int_{x_2 \in X_2} \left(\int_{x_1 \in X_1} |f(x_1, x_2)| d\mu_1 \right) d\mu_2 < +\infty.$$

If so, then,

$$\begin{aligned} & \int_{(x_1, x_2) \in X_1 \times X_2} f(x_1, x_2) d(\mu_1 \odot \mu_2) = \\ &= \int_{x_1 \in X_1} \left(\int_{x_2 \in X_2} f(x_1, x_2) d\mu_2 \right) d\mu_1 = \int_{x_2 \in X_2} \left(\int_{x_1 \in X_1} f(x_1, x_2) d\mu_1 \right) d\mu_2. \end{aligned} \quad \diamond$$

5.4 Series are also Integrals

Proposition 45. Given a set I , recall that we denote $z : I \rightarrow \mathbb{C}$ also as the tuple $(z_j)_{j \in I} \subseteq \mathbb{C}$. Then, for the measure space $(I, \mathcal{P}(I), \nu)$, where ν is the counting measure,

(i) if $x : I \rightarrow [0, +\infty]$,

$$\int_{j \in I} x(j) d\nu = \sum_{j \in I} x_j,$$

so, a general $z : I \rightarrow \mathbb{C}$ satisfies

$$\left(\int_{j \in I} |z(j)| d\nu < +\infty \right) \iff \left(\sum_{j \in I} |z_j| < +\infty \right),$$

i.e., z is integrable if and only if it is absolutely summable. Thus, in general, for $p \in [1, \infty)$,

$$\|z\|_{L^p} = \left(\int_{j \in I} |z_j|^p d\nu \right)^{1/p} = \left(\sum_{j \in I} |z_j|^p \right)^{1/p}.$$

(ii) For the counting measure the only null set is $\emptyset$, so, equality almost-everywhere is the same as equality everywhere, implying that $[z]_{a.e} = z$. Thus,

$$\|z\|_{L^\infty} = \text{ess sup } |z| = \sup_{j \in I} |z_j| = \|z\|_\infty.$$

We denote for each $p \in [1, \infty]$,

$$\ell^p(I) := L^p(I, \nu).$$

which by the above, are equivalently given by:

$$\ell^p(I) = \left\{ (z_j)_{j \in I} \in \prod_{j \in I} \mathbb{C} \mid \sum_{j \in I} |z_j|^p < +\infty \right\} \quad \forall p \in [1, \infty)$$

$$\ell^\infty(I) = \left\{ (z_j)_{j \in I} \in \prod_{j \in I} \mathbb{C} \mid \sup_{j \in I} |z_j| < +\infty \right\}. \quad \blacklozenge$$

Remark 4. This shows that we can treat absolutely convergent series as Lebesgue integrals. Therefore, in order to justify exchanging an infinite sum and another limit, one can use the monotone and dominated convergence theorems (when their hypotheses apply). Similarly, when we have two iterated infinite sums, we can exchange the order in which we take them by using the Fubini and Tonelli theorems (when applicable). This is highly useful because it gives us “high level” tricks to exchange the order of two iterated limits. Keep this fact in mind. $\blacklozenge$

5.5 How all this is used in Probability Theory

Measure theory turns out to be the right mathematical structure to make probability theory precise. However, when it is used for that purpose, one employs ad hoc names for its symbols. In this section we will give an informal overview of what changes.

- Given a measurable space (Ω, Σ) , instead of “sets” $B \in \Sigma$ one says *events*, and their elements are called *elementary events* —as such, (Ω, Σ) is called the *event space*. One thinks of Ω as the set of possible configurations that a system may adopt, while Σ are the possible “questions” about the system’s configurations that one may ask. Importantly, we are usually representing either

- (i) a process that (at least hypothetically) occurs countably many times and adopts a definite configuration of Ω at each realization —also called each *run*. Namely, some $\omega : \mathbb{N} \rightarrow \Omega$, where the parameter $t \in \mathbb{N}$ labels the *run*.
- (ii) Similarly, a process that occurs along a continuous parameter $t \in I \subseteq \mathbb{R}$ (bounded interval) called the *instance* of realization, adopting a definite configuration of Ω at each such instance. Namely, some measurable $\omega : I \subseteq \mathbb{R} \rightarrow \Omega$ —if Ω has a topology, ω is often assumed to be continuous.

Then, one defines for some $B \in \Sigma$ “the probability of B ” as

- (i) the asymptotic relative number of runs when the realization of the process adopts a configuration in B (assumed to be convergent).
- (ii) the fraction of I that ω spends in B , i.e., $\lambda^1(\omega^{-1}(B))/\lambda^1(I)$.

Then, given a measure μ with $\mu(\Omega) = 1$, one interprets μ as a given assignment of probabilities for each event $B \in \Sigma$, calling $\mu(B)$ the *probability that an elementary event in B occurs*, or plainly, *the probability of B* —instead of the “measure of B ”. Hence, we call such a μ , a *probability measure* and one uses often the symbol $\mathbb{P} := \mu$. Importantly, one says *almost surely* instead of *almost everywhere*.

Example 33. 1. If we are describing the possible outcomes after rolling two dice of different colors (say, red and blue), each with six faces: $\Omega = \{1, \dots, 6\} \times \{1, \dots, 6\}$, where an elementary event $(r, b) \in \Omega$ represents the outcome of red coming out a and blue coming out b . Then, if $\Sigma = \mathcal{P}(\Omega)$, an example composite events is say, $A := \{(a, b) \in \Omega \mid a, b \text{ are even}\}$ (the event that both dice turn even).

2. To describe the time in days between the last drop of a storm over a spot in Tübingen until the first drop of the next storm, we would use $\Omega = [0, +\infty)$. There, an elementary event $\Delta t \in \Omega$ represents the time coming out to be Δt days. Then, given $\Sigma = \mathfrak{B}([0, +\infty))$, an example composite event is $A := [0, 7)$, representing that it takes less than a week before the following storm.

3. Given a lecture room of height h , occupying a Cartesian subset $L \subseteq \mathbb{R}^3$ with the origin at the bottom left corner, if we model the $N \in \mathbb{N}$ molecules in the air as point-like particles, the space of their possible positions is $\Omega := L^N$: each elementary event $((x_1, y_1, z_1), \dots, (x_N, y_N, z_N)) \in \Omega$ is a tuple of N points in $\mathbb{R}^3$, giving the x, y, z -coordinates of each particle. Under $\Sigma = \mathfrak{B}(\Omega)$, a composite event is for instance $A := \{((x_1, y_1, z_1), \dots, (x_N, y_N, z_N)) \in \Omega \mid z_j > h/2 \forall j\}$, representing a configuration where all there is no molecule in the bottom half of the room. ♦

- One calls any measurable function $Y : \Omega \rightarrow \mathbb{R}$ a *random variable* and calls the push-forwarded measure on $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$, $\mathbb{P}_Y := Y_*\mathbb{P} = \mathbb{P} \circ Y^{-1}$, the *probability distribution of Y* . One writes for $B \in \mathfrak{B}(\mathbb{R})$, $\mathbb{P}(Y \in B) := \mathbb{P}(Y^{-1}(B))$ saying that this is “the probability for the variable Y to take a value in B ”.

Random variables are used to describe other processes that are *exactly determined* by the configuration of Ω resulting at a particular run or instance of ω . That is, when the “randomness” (i.e., our lack of knowledge) of a process comes exclusively from the “randomness” of ω . Importantly, we also assume that the configuration of this derived process is encode-able as a real numbers (otherwise, we talk about random vectors, sequences etc).

Example 34. For the example of the dice, $Y : (r, b) \mapsto r + b$, represents the sum of the two results and obviously, its probability measure is the push-forward: after all, the probability to obtain a sum of dice equal to say, 5, is the probability of the event $\{(a, b) \in \Omega : a + b = 5\} = Y^{-1}(\{5\})$. ♦

- The *expectation value* of Y is defined to be $\mathbb{E}(Y) := \int_{x \in \Omega} Y(x) d\mathbb{P}$ (so, it is the L^1 -norm if Y is non-negative) —it can equally be computed as $\int_{y \in \mathbb{R}} y d\mathbb{P}_Y$. Going down to the characterization of the integral in terms of simple functions and using the definition of probability as above, one can prove that it is the asymptotic average Y -configuration along the runs or instances of ω . Similarly, one defines the *variance of Y* as $\text{Var}(Y) := \mathbb{E}((Y - \mathbb{E}(Y))^2)$. It is the asymptotic average deviation of Y from its expectation value along the runs or instances of ω . As a short computation shows, the variance equals $\mathbb{E}(Y^2) - \mathbb{E}(Y)^2$. In general, one calls $\int_{x \in \Omega} Y(x)^p d\mathbb{P}$ the *p -th moment of Y* and the norm $\|Y\|_{L^p(\Omega, \mu)}^p$ the *absolute p -th moment of Y* .

- If given a set of random variables $\{Y_j\}_{j=1}^n$, one can consider an $\mathbb{R}^n$ -valued measurable function $Y : \Omega \rightarrow \mathbb{R}^n, w \mapsto (Y_1(w), \dots, Y_n(w))$ that is called *random vector*, of *probability distribution* $\mu \circ Y^{-1}$ on $\mathfrak{B}(\mathbb{R}^n)$. It’s interpretation is obvious. In this context, the distributions of each component, $\mu \circ Y_k^{-1}$, is called the *k -th marginal distribution*.

- Each random variable Y induces in Ω a σ -algebra by pulling back the σ -algebra on $\mathbb{R}$: $\sigma(Y) := Y^{-1}(\mathfrak{B}(\mathbb{R}))$ —note that it is a subset of Σ . These are the Ω -events that matter if one is solely interested on Y ’s outcomes. Then, given a set of random variables $(Y_j)_{j=1}^n$, one says that they are *independent* when $\mu(\bigcap_{j=1}^n B_j) = \prod_{j=1}^n \mu(B_j)$ for all choices of $B_j \in \sigma(Y_j)$. That is, when

$$\mathbb{P}\left(\bigcap_{j=1}^n \{Y_j \in A_j\}\right) = \prod_{j=1}^n \mathbb{P}(Y_j \in A_j) \quad \forall A_j \in \mathfrak{B}(\mathbb{R}).$$

In words, when the probabilities of outcomes for each random variable do not depend on —are independent of— what the other variable’s outcomes are. One can prove that $(Y_j)_{j=1}^n$ are independent *if and only if* the distribution of the random vector they form is exactly the product measure of their marginal distributions:

$$\mu \circ (Y_1, \dots, Y_n)^{-1} = \mu \circ Y_1^{-1} \odot \dots \odot \mu \circ Y_n^{-1}.$$

In a similar fashion, one says that a *sequence* of random variables $(Y_j)_{j \in \mathbb{N}}$ (or *random sequence*) is *independent* if any finite subfamily of the sequence is independent. Finally, one says $(Y_j)_{j \in \mathbb{N}}$ are *equally distributed* if all marginal distributions yield the same $\mathfrak{B}(\mathbb{R})$ -measure: $\mu \circ Y_k^{-1} = \mu \circ Y_j^{-1} \forall j, k \in \mathbb{N}$.

Proposition 46 (*Kolmogorov's Strong Law of Large Numbers*). Given a probability space $(\Omega, \Sigma, d\mu)$ and a sequence of independent random variables $(Y_j)_{j \in \mathbb{N}} \subseteq L^2(\Omega, d\mu)$,

$$\sum_{j=1}^{\infty} \frac{\text{Var}(Y_j)}{j^2} < +\infty \quad \implies \quad \lim_{N \rightarrow \infty} \sum_{j=1}^N \frac{Y_j - \mathbb{E}(Y_j)}{N} = 0 \quad \text{almost surely.}$$

♦

Exercise Sheet 5

1. Prove Lemma 28 and Proposition 36.
2. Prove Propositions 41 and 42.
3. Prove Lemma 31.
4. Given $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ denotes the Lebesgue measure, prove that $\lambda(\mathbb{Q}) = 0$.
5. Let μ be a finite measure on $(X, \mathcal{A})$. Show that

$$L^p(X, \mu) \subseteq L^q(X, \mu)$$

for all $1 \leq q \leq p \leq \infty$. (Note: when μ is a general measure there is no inclusion relation between the L^p spaces.)