Measure theoretic basics of Probability

Imre Péter Tóth

September 24, 2013

Abstract

For the "Probability 1" course at CEU, Fall 2013. Occasionally changing draft, with many typos, mistakes, errors and inconsistencies. These notes are mostly copied from my earlier lecture notes on Mathematical Statistical Physics.

Contents

1 Probability warming-up **1**
1.1 Measure, measure space, probability, probability space 1
1.2 Measurable functions, random variables and their distributions 7
1.3 Integral and expectation 9
1.3.1 Integral, integrability 9
1.3.2 Exchanging limits with integrals 12

Introduction

Nothing of the material presented is my own result. In fact, it is by now considered "classical" or "standard knowledge" in Mathematics, and I will not attempt to give references to the original papers of the true authors. Instead, I will (when I get to it) cite some textbooks and lecture notes where the material is easily accessible, and which I myself use to reduce the number of false statements written.

1 Probability warming-up

1.1 Measure, measure space, probability, probability space

Definition 1.1 (σ-algebra)**.** *For a nonempty set Ω, a family $\mathcal{F}$ of subsets of ω (i.e. $\mathcal{F} \subset 2^\Omega$, where $2^\Omega := \{A : A \subset \Omega\}$ is the power set of Ω) is called a σ-algebra over Ω if*

- $\emptyset \in \mathcal{F}$
- *if $A \in \mathcal{F}$, then $A^C := \Omega \setminus A \in \mathcal{F}$ (that is, $\mathcal{F}$ is closed under complement taking)*
- *if $A_1, A_2, \dots \in \mathcal{F}$, then $(\cup_{i=1}^\infty A_i) \in \mathcal{F}$ (that is, $\mathcal{F}$ is closed under countable union).*

Definition 1.2 (measurable space, measurable set)**.** *If Ω is a nonempty set and $\mathcal{F}$ is a σ-algebra over Ω, then the pair $(\Omega, \mathcal{F})$ is called a measurable space. The elements of $\mathcal{F}$ arer called measurable subsets of Ω.*

Lemma 1.3. *A σ-algerba is closed under finite intersection, countable union and finite union.*

Proof. Homework. □

Note that a σ-algebra is in general not closed under arbitrary intersection and union. For example, the Borel σ-algebra on the set $\mathbb{R}$ of real numbers (see later) contains every 1-element subset of $\mathbb{R}$, but it does not contain every subset (a fact we will not prove).

Two trivial examples of σ-algebra:

Definition 1.4 (indiscrete σ-algebra)**.** *For a nonempty set Ω, the family of subsets $\mathcal{F}_{ind} = \{\emptyset, \Omega\}$ is called the indiscrete or trivial σ-algebra over Ω.*

Definition 1.5 (discrete σ-algebra)**.** *For a nonempty set Ω, the family of subsets $\mathcal{F}_{discr} = 2^\Omega$ (the entire power set) is called the discrete σ-algebra over Ω.*

It is immediate from the definition that these are indeed σ-algebras over Ω.

Lemma 1.6. *The intersection of any (nonempty) family of σ-algebras over the same Ω is also a σ-algerba over Ω. That is, if Ω is a nonempty set and $\mathcal{F}_i$ is a σ-algerba over Ω for every $i \in I$ where I is* any *nonempty index set, then $\mathcal{F} := \cap_{i \in I} \mathcal{F}_i$ is also a σ-algebra over Ω.*

Proof. (trivial set algebra) By definition, $\emptyset \in \mathcal{F}_i$ for every $i \in I$, and I is nonempty, so $\emptyset \in \cap_{i\in I}\mathcal{F}_i = \mathcal{F}$. Similarly, if $A \in \mathcal{F} = \cap_{i\in I}\mathcal{F}_i$, then $A \in \mathcal{F}_i$ for every $i \in I$, so by definition $\Omega \setminus A \in \mathcal{F}_i$ for every $i \in I$, so $\Omega \setminus A \in \mathcal{F}$. Finally, if $A_1, A_2, \cdots \in \mathcal{F}$, then $A_1, A_2, \cdots \in \mathcal{F}_i$ for every $i \in I$, so by definition $(\cup_{k=1}^{\infty} A_k) \in \mathcal{F}_i$ for every $i \in I$, which means that $(\cup_{k=1}^{\infty} A_k) \in \mathcal{F}$. □

It is important to note that I being *any* nonempty set means in particular that it can well be a large set, having infinitely many, or even uncountably many, or possibly much more elements.

Corollary 1.7. *If Ω is a nonempty set and $H \subset 2^{\Omega}$ is any family of subsets, then there exists a unique σ-algebra $\sigma(H)$ over Ω, which is the smallest σ-algebra containig H in the following sense:*

- *$H \subset \sigma(H)$*
- *If $\mathcal{F}$ is any σ-algebra over Ω with $H \subset \mathcal{F}$, then $\mathcal{F} \subset \sigma(H)$.*

Proof. The family

$$\{\mathcal{F} : \ \mathcal{F} \text{ is a } \sigma\text{-algebra over } \Omega \text{ and } H \subset \mathcal{F}\}$$

is nonempty, since it contains at least the discrete σ-algebra 2^{Ω}. Thus by the above lemma,

$$\sigma(H) := \cap\{\mathcal{F} : \ \mathcal{F} \text{ is a } \sigma\text{-algebra over } \Omega \text{ and } H \subset \mathcal{F}\}$$

will do. Uniqueness also follows from the lemma: if there were two diferent such minimal σ-algebras, their intersection would also be a σ-algebra, but it would not contain them – a contradiction. □

Definition 1.8 (σ-algebra generated by a family of sets)**.** *The above $\sigma(H)$ is called the σ-algebra generated by H.*

Definition 1.9 (Borel σ-algebra)**.** *If $(\Omega, \mathcal{T})$ is a topological space (which means that it makes sense to talk about open subsets of Ω, and $\mathcal{T}$ is the set of these open subsets), then $\mathcal{B}(\Omega) := \sigma(\mathcal{T})$ is called the Borel σ-algebra on Ω.*

Remark 1.10. *The set $\mathcal{T}$ is called the* topology*, so the Borel σ-algebra is the σ-algebra generated by the topology. For those, who haven't heard, but are interested: A collection $\mathcal{T} \subset 2^{\Omega}$ of subsets of Ω is called (by definition) a* topology *over Ω if it contains $\emptyset$ and Ω, and it is closed under finite intersection and* arbitrary *union. Then the elements of $\mathcal{T}$ are called open sets, and the pair $(\Omega, \mathcal{T})$ is called a topological space. (So the definition only says that $\emptyset$ and Ω are open, the intersection of finitely many open set is open, and that the union of* any *family of open sets is open.)*

When we talk about the Borel sets on $\mathbb{R}$ or $\mathbb{R}^n$, we always think of the usual notion of open sets on these spaces.

Remark 1.11. *Not every subset of* $[0, 1]$ *is Borel. In fact, a non-Borel subset can be constructed (and not only the existence can be proven). We don't go into that.*

Notation 1.12. *We denote by* $\mathbb{R}^+$ *the set of nonnegative real numbers – that is,* $\mathbb{R}^+ = [0, \infty)$*. In particular,* $\mathbb{R}^+$ *includes zero.*

Definition 1.13 (measure space, measure)**.** *Let* $(\Omega, \mathcal{F})$ *be a measurable space. The nonnegative extended real valued function* μ *on* $\mathcal{F}$ *(that is,* $\mu : \mathcal{F} \to \mathbb{R}^+ \cup \{\infty\}$*) is called a measure on* Ω*, if*

- $\mu(\emptyset) = 0$,
- μ *is* σ*-additive, meaning that if* $\{A_i\}_{i\in I}$ *is a countable family of pairwise disjoint measureable sets (with formulas:* $A_i \in \mathcal{F}$ *for every* $i \in I$ *where* I *is a* countable *index set, and* $A_i \cap A_j = \emptyset$ *for every* $i \neq j$*,* $i, j \in I$*), then*
$$\mu\left(\cup_{i\in I} A_i\right) = \sum_{i\in I} \mu(A_i).$$

Then the triple $(\Omega, \mathcal{F}, \mu)$ *is called a measure space and* $\mu(A)$ *is called the measure of the set* A*.*

Remark 1.14. *It is absolutely important that in the definition of* σ*-additivity* I *is countably infinite:* σ*-additivity is more than finite additivity and less than arbitrary additivity. For example, if* μ *is the Lebesgue measure on* $\mathbb{R}$ *(see later), every* 1*-element set* $\{x\}$ *has* $\mu(\{x\}) = 0$*, which implies that every countable set has zero measure, but of course*

$$1 = \mu([0, 1]) = \mu(\cap_{x\in[0,1]}\{x\}) \neq \sum_{x\in[0,1]} \mu(\{0\}) = \sum_{x\in[0,1]} 0 = 0.$$

(Whatever the sum of uncountably many real numbers could mean.)

Once it came up, we mention the following, *absolutely non-important* definition:

Definition 1.15 (sum of many nonnegative extended real numbers)**.** *If* $a_i \in \mathbb{R}^+ \cup \{\infty\}$ *for every* $i \in I$ *where* I *is* any *index set, then we define the sum of all* a_i *as*

$$\sum_{i\in I} a_i := \sup\left\{\sum_{i\in J} a_i \ : \ J \subset I \text{ and } J \text{ is finite}\right\}.$$

Note that it is important that the a_i are nonnegative. This definition obviously coincides with the usual sum of (an arbitrarily reorderable) series if I is countable. This new notion of an infinite sum is no serious extention of the well known notion of a countable series: it is easy to see that if the sum is finite, then at most countably many terms can be nonzero.

Remark 1.16. *In the definition of the measure, the first requirement $\mu(\emptyset) = 0$ is almost automatic from σ-additivity: it's only there to rule out the trivial nonsense $\mu(\emptyset) = \infty$. In fact it would be enough to require that at least one measurable set A has finite measure: σ-additivity implies*

$$\mu(A) = \mu(A \cup \emptyset \cup \emptyset \cup \emptyset \cup \dots) = \mu(A) + \mu(\emptyset) + \mu(\emptyset) + \mu(\emptyset) + \dots = \mu(A) + \sum_{i=1}^{\infty} \mu(\emptyset).$$

If $\mu(A) < \infty$, then this implies $0 = \sum_{i=1}^{\infty} \mu(\emptyset)$, so $\mu(\emptyset) = 0$.

Definition 1.17. *The measure χ on a nonempty set Ω equipped with the discrete σ-algebra 2^Ω defined as*

$$\chi(A) := \sharp A := \begin{cases} \textit{number of elements in } A, & \textit{if } A \textit{ is finite} \\ \infty, & \textit{if } A \textit{ is infinite} \end{cases}$$

is called the counting measure on Ω. The restriction of χ to any σ-algebra $\mathcal{F} \subset 2^\Omega$ is still called a counting measure (on $\mathcal{F}$).

One of the most important examples of a measure is the Lebesgue measure on $\mathbb{R}$ or on $\mathbb{R}^d$

Definition 1.18 (Lebesgue measure vaguely)**.** *Consider the set $\mathbb{R}$ with the Borel σ-algebra $\mathcal{B}$. The measure $\mathrm{Leb} : \mathcal{B} \to \mathbb{R}^+ \cup \{\infty\}$ is called the Lebesgue measure on $\mathbb{R}$, if it assigns to every interval its length – that is, for every $a, b \in \mathbb{R}$, $a \le b$ we have*

$$\mathrm{Leb}((a,b)) = \mathrm{Leb}((a,b]) = \mathrm{Leb}([a,b)) = \mathrm{Leb}([a,b]) = b - a.$$

The restriction of Leb to a Borel subset of $\mathbb{R}$ (e.g. an interval $[c,d]$ or (c,∞)) is still called Lebesgue measure and is still denoted by Leb. (More precisely, if $(\mathbb{R}, \mathcal{B}, \mathrm{Leb})$ is the Lebesgue measure space on $\mathbb{R}$, and $I \in \mathcal{B}$, than one can define the "restriction of Leb to I" as the measure space $(I, \mathcal{B}_I, \mathrm{Leb}_I)$ where $\mathcal{B}_I := \{A \cap I \,:\, A \in \mathcal{B}\} = \{B \,:\, B \in \mathcal{B}, B \subset I\} \subset \mathcal{B}$ and $\mathrm{Leb}_I := \mathrm{Leb}|_{\mathcal{B}_I}$ is the restriction of Leb to $\mathcal{B}_I$.)

Similarly, the "Lebesgue measure on $\mathbb{R}^d$" is the measure on Borel subsets of $\mathbb{R}^d$ which assigns to every box its d-dimensional volume, i.e. for every $a_1 \le b_1, a_2 \le b_2, \dots a_d \le b_d \in \mathbb{R}$ we have

$$\mathrm{Leb}_d([a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_d, b_d]) = \prod_{i=1}^{d} (b_i - a_i).$$

Restrictions to Borel subsets of $\mathbb{R}^d$ are still called Lebesgue measure, and denoted by Leb_d *or just* Leb.

Remark 1.19. *The above "definition" of the Lebesgue measure is far from being complete, and is not the usual definition – it's actually a characterization of the Lebesgue measure which shows its essence. It can be (and needs to be) shown that such a measure indeed exists, since in the "definition" we only gave the value of* Leb *for a few very special sets, and not every Borel set. Also uniqueness can and needs to be shown. These questions lead to the construction of measures based on their pre-known values on certain pre-chosen "to-be-measurable" sets, which can sometimes be of crucial importance, but we don't go into that here.*

Remark 1.20. *In the measure theory literature, Lebesgue measure is defined on a σ-algerba $\mathcal{F}$ which is* larger *than the Borel σ-algebra (i.e. $\mathcal{B} \subsetneqq \mathcal{F}$), called the "set of Lebesgue measurable sets". In particular, $\mathcal{F}$ has the property that if $B \in \mathcal{F}$, $\mathrm{Leb}(B) = 0$ and $A \subset B$, then $A \in \mathcal{F}$, which is not true for Borel sets. However, in probability theory it is usual to consider Lebesgue measure restricted to Borel sets only (as in the above definition).*

The following definition shows that the basic object of probability theory, called "the probability" is in fact a measure.

Definition 1.21 (Kolmogorov probability space)**.** *The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a Kolmogorov probability space (or just probability space) if it is a measure space and $\mathbb{P}(\Omega) = 1$. Then $\mathbb{P}$ is called the* probability *or a "probability measure", elements of $\mathcal{F}$ are called* events, *and elements of Ω are the* elementary events. *For $A \in \mathcal{F}$, $\mathbb{P}(A) \in [0, 1]$ is called the probability of the event A.*

Picture $\omega \in \Omega$ as possible outcomes of an experiment, so an event $A \in \mathcal{F}$ often consists of many possible outcomes of that experiment, which have some common property that we are interested in. By definition, an "event" is something which has a probability.

1.2 Measurable functions, random variables and their distributions

Notation 1.22. *For a function $f : \Omega \to \Omega'$ and a set $A' \subset \Omega'$, let $f^{-1}(A')$ denote, as usual, the complete inverse image of A' defined as $f^{-1}(A') := \{\omega \in \Omega \,:\, f(\omega) \in A'\}$. Note that this makes sense for any function and any A' – in particular, f need not be invertable.*

Definition 1.23 (measureable function, observable, random variable)**.** *Let $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be measurable spaces. The function $X : \Omega \to \Omega'$ is called* measurable *or $(\mathcal{F}, \mathcal{F}')$-measurable, if for every $A' \in \mathcal{F}'$ we have $X^{-1}(A') \in \mathcal{F}$. (That is, if the inverse image of any measurable set is also measurable.) In physical applications, when Ω is the (possibly complicated) phase space of a system and Ω' is a (usually simple) set of possible measurement results (e.g. $\Omega' = \mathbb{R}$), the same X is called an* observable*. In the context of probability theory, when $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, X is called a (Ω'-valued)* random variable*.*

Note that the notion of measurability of a function depends on the choice of the σ-algebras $\mathcal{F}$ and $\mathcal{F}'$. However, in many cases when this choice is clear from the context, we simply say "measurable" instead of "$(\mathcal{F}, \mathcal{F}')$-measurable". When we talk about a random variable, and do not specify the range, usually $(\Omega', \mathcal{F}') = (\mathbb{R}, \mathcal{B})$ is understood.

Remark 1.24. *If $X : \Omega \to \Omega'$ is a random variable and $x \in \Omega'$, then we denote by $\{X = x\}$ the set of elementary events where X takes the value x – that is,*

$$\{X = x\} := \{\omega \in \Omega \,:\, X(\omega) = x\} = X^{-1}(\{x\}).$$

Similarly, if $A' \subset \Omega'$, then $\{X \in A'\}$ denotes the set of elementary events where X takes values in A':

$$\{X \in A'\} := \{\omega \in \Omega \,:\, X(\omega) \in A'\} = X^{-1}(A).$$

With this in mind, the definition of a random variable as a measurable *function is very natural. The definition says exactly that is A' is a measurable subset of the range Ω', then the set $\{X \in A\}$ is indeed an* event *– that is, it makes sense to talk about its probability.*

Example 1.25. *coordinate, number rolled, sum of these*

Random variables are the central objects of study in probability theory. In a typical situation they extract fairly little information (e.g. a single number) from a big probability space containing many complicated possible outcomes of an experiment. So to "understand" a random variable $X : \Omega \to \Omega'$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ well, we need less information than the what $\mathbb{P}$ and X contain. From another point of view: when we consider a random variable, Ω is often not needed, or not even known. All we need to know is the possible values (in Ω') X can take, and the probability of these being taken. This information is contained exactly in a measure on Ω', as the following definition shows.

Definition 1.26 (distribution of a random variable)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $(\Omega', \mathcal{F}')$ a measurable space and $X : \Omega \to \Omega'$ a random variable. Then the distribution of X is the measure μ on $(\Omega', \mathcal{F}')$ which is defined as*

$$\mu(A) := \mathbb{P}(\{X \in A\}) = \mathbb{P}(X^{-1}(A)) \quad \textit{for every } A \in \mathcal{F}'.$$

This can be written in short as

$$\mu := \mathbb{P} \circ X^{-1}.$$

μ is nothing else than the "push-forward" of the probability $\mathbb{P}$ by X to Ω'.

In the special case when $\Omega' = \mathbb{R}$ and X is discrete (meaning that it can take finitely many, or at most countably many values), there is a convenient alternative way to "desribe the distribution of X", by simply listing the possible values x_k and their probabilities $p_k := \mathbb{P}(X = x_k) := \mathbb{P}(\{X = x_k\})$. Then the information contained in the sequence of pairs $\{(x_k, p_k)\}_{k=1}^N$ (with possibly $N = \infty$) is called the *discrete probability distribution.* Having this information, one can calculate probabilities of events by summation:

$$\mu(A) = \mathbb{P}(\{X \in A\}) = \sum_{k : x_k \in A} p_k.$$

Similarly, in the special case when $\Omega' = \mathbb{R}$ and X is absolutely continuous (see later), there is convenient alternative "description of the distribution" by a density function $f : \mathbb{R} \to \mathbb{R}^+$ from which one can calculate probabilities of events by integration:

$$\mu(A) = \mathbb{P}(\{X \in A\}) = \int_A f(x)\, \mathrm{d}x.$$

The above notion of a probability distribution is a far-reaching generalization of both notions.

Example 1.27. *sum of two rolled numbers real number generated by a sequence of fair coin tosses real number generated by a sequence of biased coin tosses*

1.3 Integral and expectation

1.3.1 Integral, integrability

For an (extended) *real-valued* measurable functions $X : \Omega \to \mathbb{R}$ on a measure space $(\Omega, \mathcal{F}, \mu)$ it makes sense to talk about the integral $\int_\Omega X \, \mathrm{d}\mu$. This is an essential tool, and also an important object of study both in measure theory and in probability theory. We don't go deep into the definition and properties of the integral here – we don't want to, and we can't substitute a measure theory course now. I just give very briefly one of the shortest possible definitions, and point out a few main feaures.

Since we don't want to exclude the case when either a function or a measure takes the value ∞, we work with extended real numbers, and use the convention

$$0 \cdot \infty := \infty \cdot 0 := 0.$$

We start by defining the integral of nonnegative functions.

Definition 1.28 (integral of non-negative extended real valued functions)**.** *If $(\Omega, \mathcal{F}, \mu)$ is a measure space and $X : \Omega \to \mathbb{R}^+ \cup \{\infty\}$ is measurable, we introduce a sequence $X_n : \Omega \to \mathbb{R}^+$ of* simple *functions (i.e. taking only finitely many values) which approximate X (from below) as*

$$X_n(\omega) := \max\{x \in \{0, \frac{1}{2^n}, \frac{2}{2^n}, \dots, 2^n - \frac{1}{2^n}, 2^n\} \,:\, x \le X(\omega)\}.$$

Than we define the n-th integral-approximating sum as

$$I_n := \sum_{x \in \{0, \frac{1}{2^n}, \frac{2}{2^n}, \dots, 2^n - \frac{1}{2^n}, 2^n\}} x \mu(X_n^{-1}(\{x\})),$$

and the integral of X as

$$\int_\Omega X \, \mathrm{d}\mu := \lim_{n \to \infty} I_n.$$

The sets $X_n^{-1}(\{x\}) \subset \Omega$ in the definition of I_n are ensured to be $\mathcal{F}$-measurable by the assumption that X is measurable (and the fact that $\{x\}$ is Borel-measurable in $\mathbb{R}$), so $\mu(X_n^{-1}(\{x\}))$ makes sense. The sequence X_n of functions is cleverly defined to be increasing, and so is the sequence I_n, so the limit in the above definition exists, but is possibly infinite.

We can now go on to the general definition of the integral for extended real valued functions:

Definition 1.29 (integral)**.** *If* $(\Omega, \mathcal{F}, \mu)$ *is a measure space and* $X : \Omega \to \mathbb{R} \cup \{-\infty, \infty\}$ *is measurable, we introduce the positive part* X_+ *and the negative part* X_- *of* X *as*

$$X_+(\omega) := \begin{cases} X(\omega), & \text{if } X(\omega) > 0, \\ 0, & \text{if not} \end{cases}, \quad X_-(\omega) := \begin{cases} -X(\omega), & \text{if } X(\omega) < 0, \\ 0, & \text{if not} \end{cases}.$$

Note that both X_+ *and* X_- *are nonnegative and* $X = X_+ - X_-$.

Now

- *If either* $\int_\Omega X_+ \,\mathrm{d}\mu < \infty$ *or* $\int_\Omega X_- \,\mathrm{d}\mu < \infty$*, then we define the integral of* X *as*
$$\int_\Omega X \,\mathrm{d}\mu := \int_\Omega X_+ \,\mathrm{d}\mu - \int_\Omega X_- \,\mathrm{d}\mu,$$
which can possibly be ∞ *or* $-\infty$.

- *If both* $\int_\Omega X_+ \,\mathrm{d}\mu = \infty$ *and* $\int_\Omega X_- \,\mathrm{d}\mu = \infty$*, then we say that the integral of* X *doesn't exist (or that it is undefined).*

In the usual mathematical language, there is an important distinction between the existence of an integral and the integrability of a function. We emphasize this in the following definition:

Definition 1.30 (integrability)**.** *Let* $(\Omega, \mathcal{F}, \mu)$ *be a measure space and* $X : \Omega \to \mathbb{R} \cup \{-\infty, \infty\}$ *measurable. If the integral* $\int_\Omega X \,\mathrm{d}\mu$ *exists and it is* finite*, then we say that* X *is integrable (with respect to* μ*).*

So integrability of X means $-\infty < \int_\Omega X \,\mathrm{d}\mu < \infty$, which is equivalent to both $\int_\Omega X_+ \,\mathrm{d}\mu$ and $\int_\Omega X_- \,\mathrm{d}\mu$ being finite.

Remark 1.31 (comparison to the Riemann integral)**.** *The above definition of the integral is similar to the construction of the good old Riemann integral: in both cases the domain of integration,* Ω*, is chopped up into pieces, on each of which the function* X *takes nearly constant values. Than the "size" of each piece is multiplied by the approximate value of the function there, and these products are added up to obtain an integral approximating sum. The crucial difference is that in the case of Riemann integral, these small pieces of* Ω *had to be* intervals*, while here they can be any measurable subset of* Ω*. In particular, the points in* $X_n^{-1}(\{x\})$ *don't need to be "close" to each other in any sense, thus* Ω *absolutely doesn't need to be the real line or anything similar. It doesn't need to have any additional structure that would give sense to the words "distance" or "being close". Really, any measure space will do.*

A trivial but important example of integrable functions:

Lemma 1.32 (bounded functions on finite measure spaces are integrable)**.** *If* $(\Omega, \mathcal{F}, \mu)$ *is a finite measure space (meaning that* $\mu(\Omega) < \infty$*) and* $X : \Omega \to \mathbb{R}$ *is measurable and bounded – meaning that there exists an* $M \in \mathbb{R}$ *such that* $-M \leq X(\omega) \leq M$ *for every* $\omega \in \Omega$*, then* X *is integrable w.r.t.* μ*.*

Proof. X being bounded by M implies that $0 \leq X_+(\omega), X_-(\omega) \leq M$. when calculating the integral of, say, X_+, we have $X_n(\omega) \leq X_+(\omega) \leq M$, and thus

$$\begin{aligned} I_n &:= \sum_{x\in\{0,\frac{1}{2^n},\frac{2}{2^n},\dots,2^n-\frac{1}{2^n},2^n\}} x\mu(X_n^{-1}(\{x\})) \\ &\leq M \sum_{x\in\{0,\frac{1}{2^n},\frac{2}{2^n},\dots,2^n-\frac{1}{2^n},2^n\}} \mu(X_n^{-1}(\{x\})) \\ &= M\mu(\cup_{x\in\{0,\frac{1}{2^n},\frac{2}{2^n},\dots,2^n-\frac{1}{2^n},2^n\}} X_n^{-1}(\{x\})) \\ &= M\mu(\Omega). \end{aligned}$$

So we get $\int_\Omega X_+ \,\mathrm{d}\mu < M\mu(\Omega)$. Similarly, $\int_\Omega X_- \,\mathrm{d}\mu < M\mu(\Omega)$, so

$$-\infty < -M\mu(\Omega) \leq \int_\Omega X \,\mathrm{d}\mu \leq M\mu(\Omega) < \infty.$$

□

A basic property of the integral with a huge importance is linearity in the function integrated (the integrand):

Theorem 1.33 (linearity of integrability and the integral)**.** *Let* X_1 *and* X_2 *be real valued measurable functions on the same probability space* $(\Omega, \mathcal{F}, \mu)$*, and let* $\alpha_1, \alpha_2 \in \mathbb{R}$*. If both* $I_1 := \int_\Omega X_1 \,\mathrm{d}\mu$ *and* $I_2 := \int_\Omega X_2 \,\mathrm{d}\mu$ *exist and* $\alpha_1 I_1 + \alpha_2 I_2$ *is not of the form* $\infty - \infty$*, then* $\int_\Omega (\alpha_1 X_1 + \alpha_2 X_2) \,\mathrm{d}\mu$ *exists and*

$$\int_\Omega (\alpha_1 X_1 + \alpha_2 X_2) \,\mathrm{d}\mu = \alpha_1 \int_\Omega X_1 \,\mathrm{d}\mu + \alpha_2 \int_\Omega X_2 \,\mathrm{d}\mu.$$

As a consequence, if X_1 *and* X_2 *are both integrable, then so is* $\alpha_1 X_1 + \alpha_2 X_2$*.*

The proof is easy from the definition, but we don't discuss it here. See any measure theory book. It is useful to note that the integral is linear not only in the integral, but also in the measure:

Theorem 1.34 (linearity of the integral II.). *Let $(\Omega, \mathcal{F})$ be a measurable space, μ_1 and μ_2 measures on it, X a real valued measurable function and $0 \leq \alpha_1, \alpha_2 \in \mathbb{R}$. If both $I_1 := \int_\Omega X \,\mathrm{d}\mu_1$ and $I_2 := \int_\Omega X \,\mathrm{d}\mu_2$ exist and $\alpha_1 I_1 + \alpha_2 I_2$ is not of the form $\infty - \infty$, then $\int_\Omega X \,\mathrm{d}(\alpha_1\mu_1 + \alpha_2\mu_2)$ exists and*

$$\int_\Omega X \,\mathrm{d}(\alpha_1\mu_1 + \alpha_2\mu_2) = \alpha_1 \int_\Omega X \,\mathrm{d}\mu_1 + \alpha_2 \int_\Omega X \,\mathrm{d}\mu_2.$$

As a consequence, if X is integrable w.r.t. both μ_1 and μ_2, then so it is w.r.t. $\alpha_1\mu_1 + \alpha_2\mu_2$.

Remark 1.35 (bilinearity of the integral). *In the last theorem, we required that α_1 and α_2 be nonnegative – otherwise $\alpha_1\mu_1 + \alpha_2\mu_2$ may not be a measure, since in our definition a measure has to be nonnegative. For the same reason, the measures on a measurable space do not form a linear space. In functional analysis, to overcome that limitation, it is common to introduce the notion of "signed measures" (say, as a difference of two measures), which already form a linear space (with the usual notion of addition and multiplication). Then the two-variable real-valued mapping*

$$\langle \mu, X \rangle := \int X \,\mathrm{d}\mu$$

can be defined on suitably chosen linear spaces of measures and functions (e.g. $\mu \in \{$signed finite measures$\}$, $X \in \{$bounded measurable functions$\}$.) The above two theorems show that this mapping is bilinear, which is the property where functional analysis starts.

1.3.2 Exchanging limits with integrals

It is utmost important that Theorem 1.33 is about the linear combination of *two* integrable functions. Of course, it immediately implies linearity of the integral for *finite* linear combinations, but *does not say anything about infinite sums.* Indeed, linearity of the integral for infinite sums is not at all true in general. In fact, it is an important issue, in which cases exchanging an integral with a limit is possible – one has to be at least always careful. In the following we state (without proof) three theorems, which are the most frequently (and almost exclusively) used tools in checking exchangeability. In a situation where none of them works, exchangeability is usually hard to prove, and may very well not be true.

The first and most used tool is the Lebesgue dominated convergence theorem:

Theorem 1.36 (dominated convergence). *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f_1, f_2, \ldots$ measurable real valued functions on Ω which converge to the limit function pointwise, μ-almost everywehere. (That is, $\lim_{n\to\infty} f_n(x) = f(x)$ for every $x \in \Omega$, except possibly for a set of x-es with μ-measure zero.) Assume furthermore that the f_n admit a common* integrable dominating function*: there exists a $g : \Omega \to \mathbb{R}$ such that $|f_n(x)| \leq g(x)$ for every $x \in \Omega$ and $n \in \mathbb{N}$, and $\int_\Omega g \,\mathrm{d}\mu < \infty$. Then (all the f_n and also f are integrable and)*

$$\lim_{n\to\infty} \int_\Omega f_n \,\mathrm{d}\mu = \int_\Omega f \,\mathrm{d}\mu.$$

As you see, it is enough to require "almost everywhere" convergence, which is no surprise, since changing a function on a set of measure zero doesn't change the integral. In fact, it would be enough to require that g dominates the f_n almost everywhere – moreover, it would be enough to require that the f_n and g be extended real valued and defined almost everywhere. This is not a serious generalization, so I decided to rather keep the formulation simple. In the literature usually the most general form is given.

The second and easiest tool is Beppo Levi's monotone convergence theorem:

Theorem 1.37 (monotone convergence). *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f_1, f_2, \ldots$ a sequence of measurable* nonnegative *real valued functions on Ω which is pointwise increasing. (That is, $0 \leq f_n(x) \leq f_{n+1}(x)$ for every $n \in \mathbb{N}$ and $x \in \Omega$.) Then the pointwise limit function f defined by $f(x) := \lim_{n\to\infty} f_n(x)$ is also measurable and*

$$\lim_{n\to\infty} \int_\Omega f_n \,\mathrm{d}\mu = \int_\Omega f \,\mathrm{d}\mu.$$

The third and trickiest tool is the Fatou lemma:

Theorem 1.38 (Fatou lemma). *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and f_1, f_2, $\ldots$ a sequence of measureabale functions $f_n : \Omega \to \mathbb{R}$, which are nonneagtive, e.g. $f_n(x) \geq 0$ for every $n = 1, 2, \ldots$ and every $x \in \Omega$. Then*

$$\int_\Omega \liminf_{n\to\infty} f_n(x) \,\mathrm{d}\mu(x) \leq \liminf_{n\to\infty} \int_\Omega f_n(x) \,\mathrm{d}\mu(x)$$

(and both sides make sense).

Definition 1.39 (Absolute continuity of measures, singularity of measures). *Let μ and ν be two measures on the same measurable space $\Omega, \mathcal{F}$. We say that ν is absolute continuous with respect to μ (notation: $\nu \ll \mu$) if for every*

$A \in \mathcal{F}$ for which $\mu(A) = 0$, also $\nu(A) = 0$. We say that ν and μ are singular (with respect to each other) (notation: $\nu \perp \mu$) if there exists and $A \in \mathcal{F}$ for which $\nu(A) = 0$ and $\mu(\Omega \setminus A) = 0$.

The best known probability distributions are all examples of either one or the other of these: the "discrete" probability distributions are singular w.r.t. Lebesgue measurer on $\mathbb{R}$, while the ones that are loosely called "continuous" are actually absolutely continuous w.r.t. Lebesgue measurer on $\mathbb{R}$.

Theorem 1.40 (Radon-Nykodim)**.** *If μ and ν are two measures on the same measurable space $\Omega, \mathcal{F}$ and $\nu \ll \mu$, then there exists a measurable $f : \Omega \to \mathbb{R}$, called the **density** of ν w.r.t. μ, which satisfies $\nu(A) = \int_A f \, \mathrm{d}\mu$.*

absolute continuity w.r.t. counting measure
absolute continuity w.r.t. Lebesgue measure
integration w.r.t counting measure
integration w.r.t. Lebesgue measure
integration w.r.t. absolutely continouos measures
density function, distribution function
continuity of measures
expectation if it exists - real-valued - complex or R^d -valued - remark about more complicated spaces
integration by substitution

Theorem 1.41. *Let $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be measurable spaces, μ a measure on Ω, $X : \Omega \to \Omega'$ measurable and $\nu := \mu \circ X^{-1}$ the push-forward of μ by X (a measure on Ω') (the distribution of X, if μ is a probability). Then for any measurable $g : \Omega' \to \mathbb{R}$*

$$\int_\Omega g(X) \, \mathrm{d}\mu = \int_{\Omega'} g \, \mathrm{d}\nu.$$

expectation of a distribution
linearity of expectation
variance, standard deviation
moments, centered moments, moment-generating function, characteristic function
product measure space, product measure - be careful with infinite products
pairwise independence, marginal distributions
mutual independence
variance of linear combination of independent random variables
Markov inequality

weak law of large numbers
central limit theorem
weak convergence of measures
weak convergence of random variables
continuity of the characteristic function