Pills of measure theory on Polish spaces

Nicola Gigli *

July 24, 2025

Abstract

These are notes of a short course I gave in SISSA in 2024 and have been written with the intent to give a straight to the point presentation of some fundamental results about measure theory on Polish spaces. Among others, they contain full proofs of some key theorems, such as the disintegration and Kolmogorov's ones, that in the literature are often presented from a probabilistic perspective.

Hopefully, they will convince students that, all in all, the proofs of these crucial statements fit in just a handful of pages, at least if one writes them with a sufficiently small font.

The material collected here is very classical. For more on the topic and references I shall refer to the beautiful textbooks [Bog18] and [Dud02].

Contents

*SISSA, ngigli@sissa.it

1 Basic notions

An algebra on a set X is a collection of subsets containing X and stable by complementation and finite unions and intersections. A σ-algebra $\mathcal{A}$ is also stable by countable unions and intersections. A **measure** on $(\mathrm{X}, \mathcal{A})$ is a map $\mu : \mathcal{A} \to [0, +\infty]$ such that $\mu(\emptyset) = 0$ and

$$\mu(\cup_n A_n) = \sum_n \mu(A_n) \qquad \text{for } (A_n) \subset \mathcal{A} \text{ disjoint.} \tag{1.1}$$

Such measure is finite if $\mu(\mathrm{X}) < +\infty$. A **finite signed measure** on $(\mathrm{X}, \mathcal{A})$ is a map $\mu : \mathcal{A} \to \mathbb{R}$ such that $\mu(\emptyset) = 0$ and (1.1) holds, where now it is part of the requirement the fact that the right hand side is well defined as $\lim_{N\to\infty} \sum_{n\leq N} \mu(A_n)$. Notice that in this case the value of the limit is independent on a reordering of the sequence, because the union $\cup_n A_n$ is so.

Theorem 1.1 (Hahn-Jordan). *Let μ be a finite signed measure on $(\mathrm{X}, \mathcal{A})$. Then we can uniquely write it as $\mu = \mu^+ - \mu^-$ where $\mu^\pm$ are finite measures concentrated on disjoint subsets.*

Proof. We say that $B \in \mathcal{A}$ is *positive* if $\mu(C) \geq 0$ for any $C \subset B$, $C \in \mathcal{A}$. We claim that given $A \in \mathcal{A}$ with $\mu(A) \geq 0$ there is $B \subset A$ positive with $\mu(B) \geq \mu(A)$. To see this let $A_0 := A$, $t_n := \inf_{C\subset A_n} \mu(C)$ and then pick $C_n \subset A_n$ so that $\mu(C_n) \leq \max\{-1, \frac{t_n}{2}\}$ and let $A_{n+1} := A_n \setminus C_n$. Finally put $B := A \setminus \cup_n C_n$ and notice that $\mu(B) = \mu(A) - \sum_n \mu(C_n) \geq \mu(A)$. Suppose B is not positive. Then there is $C \subset B \subset A_n$ for every n with $\mu(C) < 0$. Thus $t_n < \mu(C)$ per ogni n and therefore $\mu(C_n) \leq \max\{-1, \frac{\mu(C)}{2}\}$. Since the C_n's are disjoint we would get $\mu(\cup_n C_n) = \sum_n \mu(C_n) = -\infty$, contradicting the fact that μ takes values in $\mathbb{R}$.

Now define recursively $P_0 := \emptyset$, $p_n := \sup_{A\subset \mathrm{X}\setminus P_{n-1}} \mu(A)$, pick $A_n \subset \mathrm{X} \setminus P_{n-1}$ so that $\mu(A_n) \geq \frac{p_n}{2}$ and use the above to find $P_n \subset A_n$ positive with $\mu(P_n) \geq \min\{\frac{p_n}{2}, 1\}$. Put $P := \cup_n P_n$. We claim that if $A \subset N := \mathrm{X} \setminus P$, then $\mu(A) \leq 0$. Indeed, from $\sum_n \mu(P_n) = \mu(P) < \infty$ we see that $\sum_n p_n < \infty$, and if $\mu(A) > 0$ the above argument produces $P' \subset A$ positive with $\mu(P') \geq \mu(A) > 0$ implying, as above, that $p_n \geq \min\{1, \frac{\mu(P')}{2}\}$ for every $n \in \mathbb{N}$, giving the contradiction.

It is now clear that the formulas $\mu^+(E) := \mu(E \cap P)$ and $\mu^-(E) := -\mu(E \setminus P)$ define finite measures satisfying the claim. For uniqueness, we observe that if $\mu = \nu^+ - \nu^-$ is another decomposition, then the above argument show that ν^- must be concentrated on N and ν^+ on P, giving $\mu(E) = \nu^+(E) - \nu^-(E) = \nu^+(E \cap P) - \nu^-(E \setminus P)$ for any $E \in \mathcal{A}$ and the conclusion. ☐

In particular, the above shows that any finite signed measure must be bounded, i.e. takes value in some compact subinterval of $\mathbb{R}$.

Given a finite signed measure μ we define $|\mu|(E) := \mu^+(E) + \mu^-(E)$ for any $E \in \mathcal{A}$. It is clear that this is a non-negative measure. The **total variation** $\|\mu\|_{\mathsf{TV}}$ of μ is defined as $|\mu|(\mathrm{X})$. It is clear that $\|\cdot\|_{\mathsf{TV}}$ is a norm on the vector space $\mathcal{M}(\mathrm{X}, \mathcal{A})$ of finite signed measures.

Proposition 1.2. *The total variation is a complete norm on $\mathcal{M}(\mathrm{X}, \mathcal{A})$ and probability measures are a closed subset.*

Proof. We have $|\mu(E)| = |\mu^+(E) - \mu^-(E)| \leq \mu^+(E) + \mu^-(E) \leq \mu^+(\mathrm{X}) + \mu^-(\mathrm{X}) = \|\mu\|_{\mathsf{TV}}$ for every E, thus if (μ_n) is $\|\cdot\|_{\mathsf{TV}}$-Cauchy we have $|\mu_n(E) - \mu_m(E)| \leq \|\mu_n - \mu_m\|_{\mathsf{TV}}$ and thus $n \mapsto \mu_n(E)$ is also Cauchy for every E and uniformly so in E, i.e. denoting by $L(E)$ the limit we have that $\lim_n \sup_E |L(E) - \mu_n(E)| = 0$. We claim that L is a measure. Clearly $L(\emptyset) = 0$ and we have finite additivity. For σ-additivity pick (E_n) disjoint and pass to the limit first in n then in i in the bound

$$|L(\cup_n E_n) - L(\cup_{n\leq N} E_n)| = |L(\cup_{n>N} E_n)| \leq |\mu_i|(\cup_{n>N} E_n) + \sup_E |L(E) - \mu_i(E)|$$

to deduce that $\lim_n |L(\cup_n E_n) - L(\cup_{n\leq N} E_n)| = 0$, i.e. that L is a measure. To see that it is the TV-limit of (μ_n) notice that for suitable P_n we have

$$\|L - \mu_n\|_{\mathsf{TV}} = (L - \mu_n)(P_n) - (L - \mu_n)(P_n^c) \leq |(L - \mu_n)(P_n)| + |(L - \mu_n)(P_n^c)| \leq 2 \sup_E |(L - \mu_n)(E)| \to 0.$$

The fact that probability measures are closed is trivial. ☐

Let μ, ν be two measures on the same σ-algebra. We say that μ is **absolutely continuous** with respect to ν, and write $\mu \ll \nu$, provided $\nu(E) = 0$ implies $\mu(E) = 0$. We say that μ is **singular** w.r.t. ν, and write $\mu \perp \nu$, provided there is E with $|\mu|(E) = 0$ and $|\nu|(E) = 0$. A measure is σ-finite provided the underlying space can be decomposed in a countable collection of sets having finite measure.

Theorem 1.3 (Radon–Nikodym). *Let μ, ν be two σ-finite measures on $(\mathrm{X}, \mathcal{A})$ with $\mu \ll \nu$. Then $\mu = f\nu$ for some non-negative $\mathcal{A}$-measurable function f, called Radon–Nikodym derivative and denoted $\frac{\mathrm{d}\mu}{\mathrm{d}\nu}$.*

Proof. Decomposing X in sets where μ, ν are finite we can assume $\mu(\mathrm{X}), \nu(\mathrm{X}) < \infty$. Let $\mathcal{F} := \{g : \mathrm{X} \to [0, +\infty]\ : g \text{ is } \mathcal{A}\text{-meas. and } g\nu \leq \mu\}$, $S := \sup_{g \in \mathcal{F}} \int g \,\mathrm{d}\nu$ and $(g_n) \subset \mathcal{F}$ with $\int g_n \,\mathrm{d}\nu \to S$. It is clear that $f_n := \max\{g_1, \ldots, g_n\} \in \mathcal{F}$ and, by monotone convergence, that $f := \sup_n f_n \in \mathcal{F}$ with $S = \int f \,\mathrm{d}\nu$. We claim that f is the required function, i.e. that $\mu = f\nu$. If not, since we know that $f\nu \leq \mu$ we must have $\mu(\mathrm{X}) - \int_{\mathrm{X}} f \,\mathrm{d}\nu \geq c$ for some $c > 0$ and thus Theorem 1.1 and its proof gives the existence of a positive set $P \in \mathcal{A}$ for $\mu - f\nu$ with $(\mu - f\nu)(P) \geq c$. The positivity of P means that $\mu(A) \geq \int_A f + c \,\mathrm{d}\nu$ for every $A \subset P$, i.e. that $f + c\chi_P \in \mathcal{F}$, contradicting the maximality of f (because $\int f + c\chi_P \,\mathrm{d}\nu > \int f \,\mathrm{d}\nu$). □

Theorem 1.4 (Lebesgue decomposition). *Let μ, ν be two measures on $(\mathrm{X}, \mathcal{A})$ with μ being σ-finite. Then there are unique measures $\mu^{ac}, \mu^{\perp}$ with $\mu^{ac} \ll \nu$, $\mu^{\perp} \perp \nu$ and $\mu = \mu^{ac} + \mu^{\perp}$.*

Proof. As before, we easily reduce to the case $\mu(\mathrm{X}) < \infty$. Let $\mathcal{N} := \{N \in \mathcal{A} : \nu(N) = 0\}$, $m := \sup_{N \in \mathcal{N}} \mu(N)$ and $(N_n) \subset \mathcal{N}$ with $\mu(N_n) \to m$. Put $N := \cup_n N_n$ and notice that $\nu(N) = 0$ and $\mu(N) = m$. Also, the formula $\mu^{\perp}(E) := \mu(E \cap N)$ defines a measure singular w.r.t. ν. We claim that $\mu^{ac} := \mu - \mu^{\perp}$ is absolutely continuous w.r.t. ν. If not, there is $N' \in \mathcal{A}$ with $\nu(N') = 0$ and $0 < \mu^{ac}(N') = \mu(N') - \mu(N' \cap N) = \mu(N' \setminus N)$ contradicting the maximality of N (because $\mu(N \cup N') = \mu(N) \cup \mu(N' \setminus N) > \mu(N)$).

This proves existence. For uniqueness observe that if $\mu = \tilde{\mu}^{ac} + \tilde{\mu}^{\perp}$, then $\mu^{ac} - \tilde{\mu}^{ac} = \tilde{\mu}^{\perp} - \mu^{\perp}$ and the two sides of this equality are mutually singular measures, forcing both to be zero. □

2 Identifying and constructing measures

A π**-system** on a set X is a non-empty collection of subsets stable by finite intersection.

A λ**-system** (or Dyinkin system) on X is a non-empty collection $\mathcal{A}$ of subsets such that:

a) $X \in \mathcal{A}$,

b) if $A, B \in \mathcal{A}$ are so that $B \subset A$, then $A \setminus B \in \mathcal{A}$,

c) if $A_n \in \mathcal{A}$ with and $A_n \subset A_{n+1}$ for every $n \in \mathbb{N}$, then $\cup_n A_n \in \mathcal{A}$.

Theorem 2.1 ($\pi - \lambda$ theorem)**.** *The smallest λ-system containing a π-system P coincides with the σ-algebra generated by P.*

Proof. One inclusion is obvious. For the other one, let D be the smallest λ-system containing P and let $D_1 \subset D$ be the set of those $A \in D$ such that $A \cap C \in D$ for every $C \in P$. It is clear that D_1 is a λ-system (using $(A \setminus B) \cap C = (A \cap C) \setminus (B \cap C)$ and $(\cup_n A_n) \cap C = \cup_n (A_n \cap C)$ and that it contains P (as this is stable by intersection). Thus $D_1 = D$.

Now let $D_2 \subset D$ be the set of those $A \in D$ such that $A \cap C \in D$ for every $C \in D$. Since $D_1 = D$ we have $P \subset D_2$ and the same arguments just used show that D_2 is a λ-system. Hence $D_2 = D$, i.e. D is stable by finite intersection. Since it contains X and is stable by relative complement, it is also stable by finite union. For here the conclusion easily follows using property (c) to improve finite unions to countable ones. □

Corollary 2.2. *Let μ, ν be two finite measures on the same σ-algebra. Assume that they coincide on a π-system generating the σ-algebra. Then they coincide on the whole σ-algebra.*

Proof. The collection of measurable sets E such that $\mu(E) = \nu(E)$ contains the given π-system and is, trivially from the definition of measure, a λ-system. The conclusion follows from the above theorem. □

Let $\mathcal{A}$ be a collection of subsets of X and $\mu : \mathcal{A} \to [0, +\infty]$ be arbitrary. Then we can define an **outer measure** μ^* (i.e. a σ-subadditive function defined on all subsets with $\mu^*(\emptyset) = 0$) by

$$\mu^*(A) := \inf \Big\{ \sum\nolimits_n \mu(A_n) \ : \ (A_n) \subset \mathcal{A} \text{ is an at most countable cover of } A. \Big\} \tag{2.1}$$

It is clear that this is an outer measure and that if $\mu(A) \leq \sum_n \mu(A_n)$ whenever $A, A_n \in \mathcal{A}$, $n \in \mathbb{N}$, are so that $A \subset \cup_n A_n$, then $\mu^*(A) = \mu(A)$ for every $A \in \mathcal{A}$.

Given an outer measure μ^*, a subset $E \subset X$ is called μ^*-measurable provided

$$\mu^*(F) \geq \mu^*(F \cap E) + \mu^*(F \setminus E) \qquad \forall F \subset X.$$

Lemma 2.3 (Carathéodory's criterion)**.** *The set $\mathcal{B}$ of μ^*-measurable sets is a σ-algebra and the restriction of μ^* to it a measure.*

Proof. $\mathcal{B}$ contains the empty set and is stable by complement. Let $(E_n) \subset \mathcal{B}$ and $E := \cup_n E_n$. Given F we have $\mu^*(F) \geq \mu^*(F \setminus E_0) + \mu^*(F \cap E_0)$ and also $\mu^*(F \setminus E_0) \geq \mu^*(F \setminus (E_0 \cup E_1)) + \mu^*((F \setminus E_0) \cap E_1)$ therefore

$$\mu^*(F) \geq \mu^*(F \setminus (E_0 \cup E_1)) + \mu^*(F \cap E_0) + \mu^*((F \setminus E_0) \cap E_1)$$

and thus by induction

$$\mu^*(F) \geq \mu^*(F \setminus (\cup_{i=0}^n E_i)) + \sum_{i=0}^n \mu^*(F \setminus (\cup_{j=0}^{i-1}) E_j) \cap E_i) \geq \mu^*(F \setminus E) + \sum_{i=0}^n \mu^*(F \setminus (\cup_{j=0}^{i-1}) E_j) \cap E_i)$$

having used that $F \setminus E \subset F \setminus \cup_{i=0}^n E_i$ for every n. Letting $n \to \infty$ and using subadditivity we get

$$\mu^*(F) \geq \mu^*(F \setminus E) + \sum_{i \in \mathbb{N}} \mu^*(F \setminus (\cup_{j=0}^{i-1}) E_j) \cap E_i) \geq \mu^*(F \setminus E) + \mu^*(F \cap E),$$

proving that $E \in \mathcal{B}$ and thus that $\mathcal{B}$ is a σ-algebra. To conclude, assume the E_n's are disjoint and pick $F = E = \cup_n E_n$ in the first inequality above to deduce that $\mu^*(E) \geq \sum_i \mu^*(E_i)$. Since the other inequality holds by σ-subadditivity, we are done. $\square$

Proposition 2.4. *Let $\mathcal{A}$ be an algebra on X and $\mu : \mathcal{A} \to [0,1]$ finitely additive. Then μ extends to a measure on the σ-algebra generated by $\mathcal{A}$ if and only if for any $(A_n) \subset \mathcal{A}$ decreasing with $\cap_n A_n = \emptyset$ we have $\mu(A_n) \downarrow 0$.*

Proof. Uniqueness follows from Corollary 2.2. For existence define the outer measure μ^* from μ as in (2.1) and notice that the assumption ensures that $\mu(A) \leq \sum_n \mu(A_n)$ whenever $A, A_n \in \mathcal{A}$ with $A \subset \cup_n A_n$ (because $E_n := A \setminus \cup_{i \leq n} A_i$ is decreasing with $\cap_n E_n = \emptyset$ and thus $\mu(A) - \sum_{i \leq n} \mu(A_i) \leq \mu(E_n) \to 0$). As already noticed, this implies $\mu^*(A) = \mu(A)$ for $A \in \mathcal{A}$. It thus remains to prove that sets in $\mathcal{A}$ are μ^*-measurable. This, however, is a trivial consequence of the additivity of μ. Indeed, for given $A \in \mathcal{A}$, $F \subset \mathrm{X}$ arbitrary and $(A_n) \subset \mathcal{A}$ cover of F we have that $(A_n \cap A), (A_n \setminus A) \subset \mathcal{A}$ are cover of $F \cap A$ and $F \setminus A$ respectively and $\sum_n \mu(A_n) = \sum_n \mu(A_n \cap A) + \mu(A_n \setminus A) \geq \mu^*(F \cap A) + \mu^*(F \setminus A)$. $\square$

Proposition 2.5. *Let μ^* be an outer measure on a metric space X that is additive on distant sets ($A, B \subset \mathrm{X}$ are distant if $\inf_{x \in A, y \in B} \mathsf{d}(x,y) > 0$). Then Borel sets are μ^*-measurable.*

Proof. It suffices to prove that closed sets are measurable. Let thus $C \subset \mathrm{X}$ be closed and $B \subset \mathrm{X}$ arbitrary. We want to prove that

$$\mu^*(B) \geq \mu^*(B \cap E) + \mu^*(B \setminus E)$$

thus we can assume $\mu^*(B) < \infty$. Let $B_n := \{x \in B : \mathsf{d}(x, C) \in [\frac{1}{n+1}, \frac{1}{n}]\}$ and notice that B_n and B_{n+2} are distant, thus by additivity we deduce that $\sum_{i=0}^n \mu^*(B_{2n}) \leq \mu^*(B)$ and $\sum_{i=0}^n \mu^*(B_{2n+1}) \leq \mu^*(B)$ for every n. Hence $\sum \mu^*(B_n) < \infty$ and thus by σ-subadditivity $\mu^*(\cup_{n \geq N} B_n) \to 0$ as $N \to \infty$. Since $B \setminus C = \cup_n B_n$ (here we use that C is closed), from σ-subadditivity we see that

$$\mu^*(B \cap C) + \mu^*(B \setminus C) \leq \mu^*(B \cap C) + \mu^*(\cup_{n<N} B_n) + \mu^*(\cup_{n \geq N} B_n) \leq \mu^*(B) + \mu^*(\cup_{n \geq N} B_n),$$

having used that B and $\cup_{n<N} B_n$ are distant in the second inequality. Letting $N \to \infty$ we conclude. $\square$

Corollary 2.6. *Let X be a metric space and μ be a non-negative map defined on the open sets. Then μ extends to a Borel measure if and only if:*

O1) $\mu(C) \leq \mu(A) + \mu(B)$ whenever $C \subset A \cup B$,

O2) $\mu(A \cup B) \geq \mu(A) + \mu(B)$ if A, B are distant,

O3) $\mu(A) \leq \lim_n \mu(A_n)$ if $A_n \uparrow A$ (i.e. the sequence is non-decreasing and $A = \cup_n A_n$).

Proof. The 'only if' is clear. For the 'if' we notice that $(O1)$ and $(O3)$ imply that $\mu(A) \leq \sum_n \mu(A_n)$ if $A \subset \cup_n A_n$, thus from the initial discussion we see that the outer measure μ^* induced by μ coincides with μ on the open sets. Then, since distant sets are covered by distant open sets (trivially) we see that $(O2)$ and subadditivity imply that μ^* is additive on distant sets, hence the conclusion follows by the above proposition. $\square$

3 Polish spaces

A **Polish space** is a topological space (X, τ) whose topology can be metrized by a complete and separable distance. Fixing a distance, for which there is in general no canonical choice, is not necessary for the results we are going to discuss, that are only linked to topology.

Proposition 3.1. *Open and closed subsets of Polish spaces are Polish and so is the countable product of Polish spaces. If* X *is Polish and* $A_n \subset \mathrm{X}$ *is Polish (with the induced topology) for every* $n \in \mathbb{N}$*, then so is* $\cap_n A_n$*.*

Proof. The claim for closed sets and products are obvious. For the intersection notice that the diagonal $(x, x, \ldots)$ is a closed subset of the product $\Pi_n A_n$. If $U \subsetneq \mathrm{X}$ is open and d a complete and separable distance on X inducing the topology, then $\mathsf{d}_U(x,y) := \mathsf{d}(x,y) + |\mathsf{d}(x, \partial U)^{-1} - \mathsf{d}(y, \partial U)^{-1}|$ is a complete and separable distance on U inducing the subspace topology. □

Proposition 3.2. *A subset* Y *of a Polish space* X *is Polish iff it is a countable intersection of open sets.*

Proof. The 'if' follows from the proposition above. For the 'only if' let $\mathsf{d}, \mathsf{d}_{\mathrm{Y}}$ be complete separable distances on X, Y respectively and $V_n \subset \mathrm{X}$ be the union of all the open subsets W_n of X with d_{X}-diameter $\leq 2^{-n}$ so that $W_n \cap \mathrm{Y}$ has d_{Y}-diameter $\leq 2^{-n}$. We claim that $\mathrm{Y} = \bar{\mathrm{Y}} \cap (\cap_n V_n)$ and notice that this gives the conclusion. The inclusion $\subset$ is obvious. For $\supset$ let $x \in \bar{\mathrm{Y}} \cap (\cap_n V_n)$ and for every n find W_n as above, containing x and contained in V_n. Since d-diameters go to 0, we know that there is at most an element of X belonging to all the X-closures $\bar{W}_n^{\mathrm{X}}$ of the W_n's, and this element is x. Similarly, from the completeness of Y and the fact that the d_{Y}-diameters go to 0 we know that there is an element $y \in \mathrm{Y}$ belonging to the Y-closures $\overline{W_n \cap \mathrm{Y}}^{\mathrm{Y}} \subset \bar{W}_n^{\mathrm{X}}$ of $W_n \cap Y$ for every n. Thus $x = y$. □

For X Polish we define $C_b(\mathrm{X})$ as the space of real valued, bounded and continuous functions, equipped with the (complete) norm $\|f\| := \sup |f|$. The **support** $\mathrm{supp}(f)$ of a function is the closure of $\{f \neq 0\} \subset \mathrm{X}$.

Exercise 3.3. Prove that $C_b(\mathrm{X})$ is separable if and only if X is compact. ■

Proposition 3.4. *Let* X *be Polish. Then there is a countable collection* $\mathcal{D} \subset C_b(\mathrm{X})$ *such that for any* $f \in C_b(\mathrm{X})$ *there is* $(g_n) \subset \mathcal{D}$ *increasing so that* $f(x) = \sup_n g_n(x)$ *for every* $x \in \mathrm{X}$*.*

If d *is a complete and separable distance inducing the topology on* X*, we can choose* $\mathcal{D}$ *so that it only contains* d*-Lipschitz functions.*

Proof. Let $(x_i) \subset \mathrm{X}$ be countable and dense and $\mathcal{D}' \subset C_b(\mathrm{X})$ be the countable collection of functions of the form $(a - b\mathsf{d}(\cdot, x_i)) \vee c$ for $a, b, c \in \mathbb{Q}$ and $i \in \mathbb{N}$. It is easy to see that for any $f \in C_b(\mathrm{X})$, $x \in \mathrm{X}$ and $\varepsilon > 0$ there is $h \in \mathcal{D}'$ so that $h \leq f$ on X and $f(x) \leq h(x) + \varepsilon$. Then $\mathcal{D} := \{h_1 \vee \cdots \vee h_n : n \in \mathbb{N},\ g_i \in \mathcal{D}'\}$ does the job, as for $f \in C_b(\mathrm{X})$ we enumerate as (h_n) the functions in $\mathcal{D}'$ that are $\leq f$ and then consider $g_n := h_1 \vee \cdots \vee h_n$. □

For X Polish we denote by $\mathcal{M}(X)$ the collection of finite signed Borel measures on X and by $\mathcal{P}(\mathrm{X}) \subset \mathcal{M}(\mathrm{X})$ that of Borel probability measures.

Lemma 3.5. *Let* X *be Polish and* $\mu \in \mathcal{P}(\mathrm{X})$*. Then for every* $E \subset \mathrm{X}$ *Borel we have*

$$\mu(E) = \sup\{\mu(C) : C \subset E \ \textit{closed}\} = \inf\{\mu(U) : E \subset U \ \textit{open}\}. \tag{3.1}$$

Proof. Let $\mathcal{A}$ be the collection of Borel sets E for which (3.1) holds. Since open sets in X are countable union of closed sets, $\mathcal{A}$ contains the open sets, so if we prove that it is a σ-algebra we are done. It is clearly stable by complementation, thus it suffices to prove that if $(E_n) \subset \mathcal{A}$ then $E := \cup_n E_n$ is also in $\mathcal{A}$. Let $\varepsilon > 0$ and then $C_n \subset E_n \subset U_n$ with $\mu(E_n \setminus C_n), \mu(U_n \setminus E_n) \leq 2^{-n}\varepsilon$. Then $U := \cup_n U_n$ is open, contains E and $\mu(U \setminus E) \leq \sum_n \mu(U_n \setminus E_n) \leq \varepsilon$. Also, find $N \in \mathbb{N}$ so that $\mu(\cup_n C_n \setminus \cup_{n \leq N} C_n) < \varepsilon$ and put $C := \cup_{n \leq N} C_n \subset E$. Then C is closed and $\mu(E \setminus C) = \mu(E \setminus \cup_n C_n) + \mu(\cup_n C_n \setminus C) \leq 2\varepsilon$. □

4 The Riesz-Daniell-Stone theorem

Lemma 4.1. *Let* X *be Polish and* $\mu \in \mathcal{M}(\mathrm{X})$*. Then*

$$|\mu|(U) = \sup\left\{ \int f \,\mathrm{d}\mu \ : \ f \in C_b(\mathrm{X}),\ \|f\| \leq 1 \quad \mathrm{supp}(f) \subset U \right\} \qquad \forall U \subset \mathrm{X} \text{ open.}$$

and in particular

$$\|\mu\|_{\mathsf{TV}} = \sup\left\{ \int f \,\mathrm{d}\mu \ : \ f \in C_b(\mathrm{X}),\ \|f\| \leq 1 \right\}. \tag{4.1}$$

Proof. $\geq$ is obvious. For $\leq$ let $P \subset \mathrm{X}$ be a positive set of μ with P^c negative as in the proof of Theorem 1.1, $\varepsilon > 0$ and then find closed sets C, D in $U \cap P, U \cap P^c$ respectively with $\mu^+((U \cap P) \setminus C) \leq \varepsilon$ and $\mu^-((U \cap P^c) \setminus D) \leq \varepsilon$. Let $f_n := (1 - n\mathsf{d}(\cdot, C))^+$, $g_n := (1 - n\mathsf{d}(\cdot, D))^+$ and then put $h_n := f_n - g_n$ so that $h_n \to \chi_C - \chi_D$ and $\|h_n\|_C \leq 1$. By dominate convergence we get $\int h_n \,\mathrm{d}\mu \to \mu^+(C) + \mu^-(D) \geq \mu^+(U) + \mu^-(U) - 2\varepsilon = |\mu|(U) - 2\varepsilon$. □

To any $\mu \in \mathcal{M}(\mathrm{X})$ we can associate the linear continuous functional L on $C_b(\mathrm{X})$ given by

$$C_b(\mathrm{X}) \ni f \qquad \mapsto \qquad L(f) := \int f \,\mathrm{d}\mu$$

and the above lemma ensures that the operator norm of L is $\|\mu\|_{\mathsf{TV}}$. The Riesz-Daniell-Stone theorem tells which functionals arise this way. It can be thought of as a regularity theorem, and in fact is the initial step in many structural results, as something that initially only acts on continuous functions turns out to act also on Borel ones[1].

Below we write $\varphi_n \downarrow 0$ to mean that $\varphi_{n+1} \leq \varphi_n$ for every n and $\varphi_n(x) \downarrow 0$ for every $x \in \mathrm{X}$.

Theorem 4.2. *For* $L : C_b(\mathrm{X}) \to \mathbb{R}$ *be linear and continuous the following are equivalent:*

i) L is tight, i.e. for every $f \in C_b(\mathrm{X})$ and $(\varphi_n) \subset C_b(\mathrm{X})$ with $\varphi_n \downarrow 0$ we have: $L(f\varphi_n) \to 0$.

ii) The map $\mathfrak{m}$ defined on open sets as $\mathfrak{m}(U) := \sup\{L(f) : \|f\| \leq 1,\ \mathrm{supp}(f) \subset U\}$ is the restriction to open sets of a finite Borel measure

iii) L is induced by a finite signed Borel measure μ.

If these hold then $\mathfrak{m} = |\mu|$ and the operator norm of L is $\|\mu\|_{\mathsf{TV}}$.

Proof. $(iii) \Rightarrow (i)$ holds by dominated convergence.

$(i) \Rightarrow (ii)$, We check that $\mathfrak{m}$ satisfies $(O1), (O2), (O3)$ of Corollary 2.6. For $(O1)$ we take f with $\|f\| \leq 1$ and support in C and find (see below) $\varphi, \psi \in C_b(\mathrm{X})$ with support in A, B respectively, norm ≤ 1 and with $\varphi + \psi = 1$ on $\mathrm{supp}(f)$. Then $L(f) = L(f\varphi) + L(f\psi) \leq \mathfrak{m}(A) + \mathfrak{m}(B)$ and taking the sup in f we are done. To find φ, ψ notice that there is a continuous function φ' $[0,1]$-valued with support in A and identically 1 on $\mathrm{supp}(f) \setminus B$ (pick $\frac{\mathsf{d}(\cdot, F_1)}{\mathsf{d}(\cdot, F_1) + \mathsf{d}(\cdot, F_2)}$ for $F_1 := \overline{A}$ and $F_2 := \mathrm{supp}(f) \setminus B$). Define symmetrically ψ' and then put $\varphi := \frac{\varphi'}{\varphi' + \psi'}$ and $\psi := \frac{\psi'}{\varphi' + \psi'}$, both set to 0 where $\varphi' + \psi' = 0$.

$(O2)$ is obvious. For $(O3)$ fix f with norm ≤ 1 and support in A, then use the same construction as above to find $\eta : \mathrm{X} \to [0,1]$ with support in A equal to 1 on $\mathrm{supp}(f)$. Then for every n let $(\varphi^n_k) \subset C_b(\mathrm{X})$ so that $\varphi^n_k \uparrow \chi_{A_k}$. Then given an enumeration (n_i, k_i) of $\mathbb{N}^2$ put $\psi_i := \max_{j \leq i} \varphi^{n_j}_{k_j}$ and notice that $\eta - \eta\psi_i \downarrow 0$. By tightness of L we get $L(f) = L(f\eta) = \lim_i L(f\eta\psi_i) = \lim_i L(f\psi_i) \leq \lim_l \mathfrak{m}(A_l)$, hence taking the sup in f we are done.

$(ii) \Rightarrow (iii)$ We have $\mathfrak{m}(\mathrm{X}) = \|L\|_{\mathrm{op}} < +\infty$. We claim that

$$|L(\varphi)| \leq \mathfrak{m}(\{\varphi \neq 0\}) \qquad \text{provided } \varphi \in C_b(\mathrm{X}) \text{ takes values in } [-1,1]. \tag{4.2}$$

[1] This should be taken with a grain of salt: surely we know from Hahn-Banach theorem that any linear continuous functional on $C_b(\mathrm{X})$ can be extended in many ways to a functional on, say, bounded Borel functions. Only one of these extensions will come out from the integral w.r.t. a measure, so we might think at Theorem 4.2 as a 'regular selection' result.

Indeed, given $\varepsilon > 0$ arguing as above we can find $\eta \in C_b(\mathrm{X})$ $[0,1]$-valued identically 1 on $\{|\varphi| > 2\varepsilon\}$ with support in $\{|\varphi| \le \varepsilon\}$. Then $\|\eta\varphi\| \le 1$ and $\|(1-\eta)\varphi\| \le 2\varepsilon$, hence $|L(\varphi)| \le |L((1-\eta)\varphi)| + |L(\eta\varphi)| \le 2\varepsilon\|L\|_{\mathrm{op}} + \mathfrak{m}(\{\varphi \ne 0\})$ and (4.2) follows.

Now let $E \subset \mathrm{X}$ Borel, $\varepsilon > 0$ and find (recall Lemma 3.5) $C \subset E \subset U$ with U open, C closed and $\mathfrak{m}(U \setminus C) < \varepsilon$. If φ_1, φ_2 are 1 on C and 0 outside U, then by (4.2) we get $|L(\varphi_1) - L(\varphi_2)| = |L(\varphi_1 - \varphi_2)| \le \mathfrak{m}(\{\varphi_1 \ne \varphi_2\}) \le 2\varepsilon$. It follows that for $C_n \subset E \subset U_n$ with $\mathfrak{m}(U_n \setminus C_n) \to 0$ and corresponding sequence (φ_n) the limit $\mu(E)$ of $n \mapsto L(\varphi_n)$ exists and thus in particular is independent on the chosen sequence. We claim that

$$|\mu(F) - \mu(E)| \le \mathfrak{m}(F \setminus E) \qquad \forall E \subset F \text{ Borel.} \tag{4.3}$$

To see this fix $C \subset E \subset F \subset U$ and notice that we can take sequences $(\varphi_n), (\psi_n)$ as above defining $\mu(E), \mu(F)$ respectively so that all the functions are 1 on C and 0 outside U. Hence by (4.2) we get $|\mu(F) - \mu(E)| = \lim_n |L(\psi_n - \varphi_n)| \le \mathfrak{m}(U \setminus C)$ and (4.3) follows. Now observe that μ is, as trivial consequence of the definition, additive on sets with disjoint closure. For general disjoint sets E_1, E_2, for $\varepsilon > 0$ we find (by xxx) closed sets $C_i \subset E_i$ with $\mathfrak{m}(E_i \setminus C_i) < \varepsilon$, $i = 1, 2$, and use (4.3) and the arbitrariness of ε to conclude that $\mu(E_1) + \mu(E_2) = \mu(E_1 \cup E_2)$. To conclude that it is a measure it thus suffices to show that if (E_n) is an increasing sequence then $\lim_n \mu(E_n) = \mu(E)$, where $E := \cup_n E_n$: this follows from (4.3) as $|\mu(E) - \mu(\cup_{i \le n} E_i)| \le \mathfrak{m}(E \setminus \cup_{i \le n} E_i) = 0$, having used that $\mathfrak{m}$ is a finite measure.

We now show that $\int f \, \mathrm{d}\mu = L(f)$ for any $f \in C_b(\mathrm{X})$. The key is that for $\varphi : \mathrm{X} \to [0,1]$ continuous we have

$$|L(\varphi) - \mu(\{\varphi > 0\})| \le \mathfrak{m}(\varphi^{-1}(0,1)). \tag{4.4}$$

Indeed, let $\varepsilon > 0$ and then $\psi \in C_b(\mathrm{X})$ $[0,1]$-valued, identically 1 on $\{\varphi \ge \varepsilon\}$ and with support in $\{\varphi > 0\}$. Notice that $L(\psi) \to \mu(\{\varphi > 0\})$ when $\varepsilon \downarrow 0$ (by definition of μ) and that $|L(\varphi) - L(\psi)| = |L(\varphi - \psi)| \le \mathfrak{m}(\{\varphi \ne \psi\}) \le \mathfrak{m}(\varphi^{-1}(0,1))$ (by (4.2)), which proves the claim.

Since $L(1) = \mu(\mathrm{X}) = \int 1 \, \mathrm{d}\mu$, by linearity it suffices to prove the representation formula for $f \ge 0$. In this case we prove that $L(f) = \int_0^\infty \mu(\{f > t\}) \, \mathrm{d}t$, which by Cavalieri's formula is enough to conclude. Let $f_t := f \wedge t$ and $F(t) := L(f_t)$. For $h > 0$ we have

$$\int_0^{\sup f} \frac{F(t+h) - F(t)}{h} \, \mathrm{d}t = \frac{1}{h}\Big(\int_{\sup f}^{\sup f + h} F(t) \, \mathrm{d}t - \int_0^h F(t) \, \mathrm{d}t\Big) \quad \to \quad F(\sup(f)) = L(f) \qquad \text{as } h \downarrow 0,$$

having used that $|F(t)| \le \|f_t\| \|L\|_{op} \le t\|L\|_{op} \to 0$ as $t \downarrow 0$. On the other hand the function $\frac{f_{t+h} - f_t}{h}$ takes values in $[0,1]$, thus (4.4) gives $\Big|L\big(\frac{f_{t+h} - f_t}{h}\big) - \mu(\{f > t\})\Big| \le \mathfrak{m}(\{f \in (t, t+h)\})$ and thus

$$\Big|\int_0^{\sup f} \frac{F(t+h) - F(t)}{h} \, \mathrm{d}t - \int_0^{\sup(f)} \mu(\{f > t\}) \, \mathrm{d}t\Big| \le \int_0^{\sup(f)} \mathfrak{m}(\{f \in (t, t+h)\}) \, \mathrm{d}t.$$

The conclusion follows noticing that the last term on the right goes to zero by dominated convergence.

The last claim now follows from Lemma 4.1 above. □

Lemma 4.3 (Dini). *Let* X *be compact and* $(\varphi_n) \subset C(\mathrm{X})$ *be so that* $\varphi_n \downarrow 0$. *Then* $\|\varphi_n\| \to 0$.

Proof. Say not. Then for some $\varepsilon > 0$ we can find $x_n \in \mathrm{X}$ with $\varphi_n(x_n) \ge \varepsilon$ for all $n \in \mathbb{N}$. By compactness and possibly passing to a subsequence we can assume that $x_n \to x$ for some $x \in \mathrm{X}$. For every $m \in \mathbb{N}$, since $\varphi_m \ge \varphi_n$ for $n \ge m$, we have $\varphi_m(x) = \lim_n \varphi_m(x_n) \ge \overline{\lim}_n \varphi_n(x_n) \ge \varepsilon$ contradicting the assumption $\varphi_m(x) \downarrow 0$. □

Corollary 4.4. *Let* X *be a compact metric space. Then the map* $\mathcal{M}(\mathrm{X}) \ni \mu \mapsto (f \mapsto \int f \, \mathrm{d}\mu)$ *is a bijective isometric isomorphism of* $(\mathcal{M}(\mathrm{X}), \|\cdot\|_{\mathsf{TV}})$ *and the Banach dual* $C(\mathrm{X})^*$ *of* $C(\mathrm{X})$.

Proof. Consequence of Theorem 4.2, as by Dini's lemma every $L \in C(\mathrm{X})^*$ is tight as in (i) of Theorem 4.2. □

Remark 4.5. The Riesz theorem is often stated for positive operators (i.e. sending non-negative functions to non-negative reals): these are automatically bounded as for $f \ge 0$ the monotonicity gives $L(f) \le L(1\|f\|) \le \|f\|L(1)$, and writing a generic f as sum of its positive and negative parts we easily get $|L(f)| \le \|f\|L(1)$. A generic linear bounded operator $L : C_b(\mathrm{X}) \to \mathbb{R}$ can be decomposed in its positive and negative parts via the Riesz-Kantorovich formulas:

$$L^+(f) := \sup\nolimits_{0 \le g \le f} L(g) \qquad \text{and} \qquad L^-(f) := \inf\nolimits_{0 \le g \le f} L(g) \qquad\qquad \forall f \ge 0.$$

It is easy to check that $\|L^\pm\|_{\mathrm{op}} \le \|L\|_{\mathrm{op}}$. It is not clear to me whether tightness of L directly implies that of $L^\pm$ (after Riesz theorem as stated above this is obvious) ■

5 Weak topology

Let X be a Polish space. The **weak** (or narrow, or vague topology) **topology** on $\mathcal{P}(\mathrm{X})$ is the weakest making $\mu \mapsto \int f \,\mathrm{d}\mu$ continuous for every $f \in C_b(\mathrm{X})$.

Notice that for $\mathcal{D} \subset C_b(\mathrm{X})$ as in Proposition 3.4 we have

$$\text{The weak topology is the weakest making } \mu \mapsto \int g \,\mathrm{d}\mu \text{ continuous for every } g \in \mathcal{D}. \tag{5.1}$$

Indeed, from Proposition 3.4 and monotone convergence we see that $\int f \,\mathrm{d}\mu = \sup_{g\in\mathcal{D}, g\leq f} \int g \,\mathrm{d}\mu$. This proves that if $\mu \mapsto \int g \,\mathrm{d}\mu$ is continuous for every $g \in \mathcal{D}$, then $\mu \mapsto \int f \,\mathrm{d}\mu$ is lower semicontinuous for every $f \in \mathcal{D}$ and replacing f with $-f$ we get continuity.

We shall write $\mu_n \rightharpoonup \mu$ to intend that (μ_n) converges to μ in the weak topology.

Fix a complete and separable distance d on X. The **Levy-Prokhorov distance** is:

$$\mathsf{D}_{\mathsf{LP}}(\mu,\nu) := \inf\{r > 0 : \mu(A) \leq \nu(A^r) + r \text{ and } \nu(A) \leq \mu(A^r) + r \text{ for every } A \subset \mathrm{X} \text{ Borel}\},$$

where $A^r := \{x : \mathsf{d}(x, A) < r\}$ is the r-neighbourhood of A. It is easy to check that D_{LP} is indeed a distance. It clearly depends on d, but the topology it induces does not:

Theorem 5.1. *The distance D_{LP} is separable and induces the weak topology on $\mathcal{P}(\mathrm{X})$.*

Proof. Let $f : \mathrm{X} \to \mathbb{R}^+$ be 1-Lipschitz (w.r.t. d) and bounded. Put $S := \sup f$ and notice that for any $c, r > 0$ we have $\{f < c\}^r \subset \{f < c + r\}$. Thus if $\mathsf{D}_{\mathsf{LP}}(\mu, \nu) < r$ we have

$$\int f \,\mathrm{d}\nu = \int_0^S \nu(\{f < c\}) \,\mathrm{d}c \leq \int_0^S \mu(\{f < c\}^r) + r \,\mathrm{d}c \leq \int_0^S \mu(\{f < c + r\}) + r \,\mathrm{d}c \leq r(1 + S) + \int f \,\mathrm{d}\mu,$$

from which it easily follows that $\mathsf{D}_{\mathsf{LP}}(\mu_n, \mu) \to 0$ implies $\int f \,\mathrm{d}\mu_n \to \int f \,d\mu$ and thus, easily from (5.1), that $\mu_n \rightharpoonup \mu$.

Conversely, fix $\mu \in \mathcal{P}(\mathrm{X})$ and $r > 0$: we will prove that there is a weak neighbourhood $\mathcal{U}$ of μ contained in the D_{LP}-ball $\mathcal{B}$ of radius $5r$ centred in μ.

Notice that for any $E \subset \mathrm{X}$ Borel the function $f := \max\{0, 1-2r^{-1}\mathsf{d}(\cdot, \bar{E})\} \in C_b(\mathrm{X})$ satisfies $\chi_E \leq f \leq \chi_{E^r}$, hence there is a neighbourhood $\mathcal{U}(E)$ of μ such that $\mu(E^r)+r \geq \nu(E)$ and $\nu(E^r)+r \geq \mu(E)$ for every $\nu \in \mathcal{U}(E)$. Now let $(x_i) \subset \mathrm{X}$ be countable and dense and notice that since $1 = \mu(\mathrm{X}) = \lim_N \mu(\cup_{n\leq N} B_r(x_i))$ there is $N \in \mathbb{N}$ such that for $F := \cup_{i\leq N} B_r(x_i)$ we have $\mu(\mathrm{X} \setminus F) < r$. For every $J \subset \{1, \ldots, N\}$ let $E_J := \cup_{i\in J} B_{2r}(x_i)$ and then define $\mathcal{U} := \mathcal{U}(\mathrm{X} \setminus F^r) \cap (\cap_J \mathcal{U}(E_J))$. We claim that $\mathcal{U} \subset \mathcal{B}$. To see this, let $A \subset \mathrm{X}$ be arbitrary Borel, let $J := \{i : B_r(x_i) \cap A \neq \emptyset\}$ and notice that $A \subset (\mathrm{X} \setminus F^r) \cup E_J$ and $E_J^r \subset A^{5r}$, thus

$$\mu(A) \leq r + \mu(E_J) \leq 2r + \nu(E_J^r) \leq 2r + \nu(A^{5r}) \qquad \forall \nu \in \mathcal{U}.$$

Noticing that $(\mathrm{X} \setminus F^r)^r \subset \mathrm{X} \setminus F$ we also have

$$\nu(A) \leq \nu(\mathrm{X} \setminus F^r) + \nu(E_J) \leq 2r + \mu(\mathrm{X} \setminus F) + \mu(E_J^r) \leq 3r + \mu(A^{5r}) \qquad \forall \nu \in \mathcal{U},$$

concluding the proof that $\mathcal{U} \subset \mathcal{B}$.

For separability we claim that the collection of probability measures of the form $\sum_{i\leq N} \alpha_i \delta_{x_i}$ with $(x_i) \subset \mathrm{X}$ fixed dense set and $(\alpha_i) \subset \mathbb{Q}$ is D_{LP}-dense. But this is clear, as given $\mu \in \mathcal{P}(\mathrm{X})$ and $\varepsilon > 0$ we can find, as above, $N \in \mathbb{N}$ so that $\mu(\cup_{i\leq N} B_\varepsilon(x_i)) > 1-\varepsilon$. Then picking $\alpha_i \in \mathbb{Q}$ sufficiently close to $\mu(B_\varepsilon(x_i) \setminus \cup_{j<i} B_\varepsilon(x_j))$ and putting $\nu := \sum_{i\leq N} \alpha_i \delta_{x_i}$ it is easy to see that $\mathsf{D}_{\mathsf{LP}}(\mu, \nu) < \varepsilon$. □

We say that $\mathcal{K} \subset \mathcal{P}(\mathrm{X})$ is **tight** if for every $\varepsilon > 0$ there is $K \subset \mathrm{X}$ compact such that $\mu(\mathrm{X} \setminus K) < \varepsilon$ for every $\mu \in \mathcal{K}$.

Theorem 5.2 (Prokhorov)**.** *$\mathcal{K} \subset \mathcal{P}(\mathrm{X})$ is weakly relatively compact if and only if it is tight.*

Proof. We know from the above that the weak topology is metrizable, hence (relative) compactness is equivalent to (relative) sequential compactness. For the 'if' we can thus assume to have a tight sequence (μ_n) and aim to prove that is has a weakly converging subsequence. For every $i \in \mathbb{N}$ let $K_i \subset \mathrm{X}$ be compact with $\mu_n(\mathrm{X}\setminus K_i) < \frac{1}{i}$ for every $n \in \mathbb{N}$. For fixed i, the restrictions $\mu_n|_{K_i}$ of the μ_n's to K_i are bounded in total variation and thus by Corollary 4.4 and Banach-Alaoglu admit a weakly converging subsequence. With a diagonal argument we can pass to a non-relabeled subsequence and assume that for every i we have $\mu_n|_{K_i} \rightharpoonup \nu_i$ (in duality with $C(K_i)$, in particular $\int f\,\mathrm{d}\nu_i = \lim_n \int_{K_i} f\,\mathrm{d}\mu_n$) for some $\nu_i \in \mathcal{M}(K_i) \subset \mathcal{M}(\mathrm{X})$. From the trivial identity $\chi_{K_i} - \chi_{K_j} = \chi_{\mathrm{X}\setminus K_j} - \chi_{\mathrm{X}\setminus K_i}$ we see that $|\int f\,\mathrm{d}(\nu_i - \nu_j)| \leq \underline{\lim}_n \int |f||\chi_{\mathrm{X}\setminus K_j} - \chi_{\mathrm{X}\setminus K_i}|\,\mathrm{d}\mu_n \leq \|f\|(\frac{1}{i} + \frac{1}{j})$ that by (4.1) shows that (ν_i) is $\|\cdot\|_{\mathsf{TV}}$-Cauchy and thus, by Proposition 1.2, admits a limit ν.

We claim that $\mu_n \rightharpoonup \nu$, that would prove relative weak compactness of the μ_n's. Indeed we have

$$\overline{\lim_n}\Big|\int f\,\mathrm{d}(\nu - \mu_n)\Big| \leq \|f\|\big(\|\nu - \nu_i\|_{\mathsf{TV}} + \tfrac{1}{i}\big) + \overline{\lim_n}\Big|\int f\,\mathrm{d}(\nu_i - \mu_n|_{K_i})\Big| = \|f\|\big(\|\nu - \nu_i\|_{\mathsf{TV}} + \tfrac{1}{i}\big) \qquad \forall f \in C_b(\mathrm{X})$$

and letting $i \to \infty$ we conclude.

Conversely, using again that the weak topology is metrizable it suffices to show that if $\mu_n \rightharpoonup \mu$, then (μ_n) is tight. Let $(x_i) \subset \mathrm{X}$ be countable and dense and fix $\varepsilon > 0$. For every $m \in \mathbb{N}$ put $\varepsilon_m := 2^{-m}\varepsilon$ and find $M_m \in \mathbb{N}$ so that for $C_m := \cup_{i \leq M_m} B_{\varepsilon_m}(x_i)$ we have $\mu(C_m) \geq 1 - \varepsilon_m$. Recalling that $\mathsf{D}_{\mathsf{LP}}(\mu_n, \mu) \to 0$ we see that there is $N_m \in \mathbb{N}$ so that $\mu_n(C_m^{\varepsilon_m}) \geq 1 - 2\varepsilon_m$ for every $n \geq N_m$. We now increase M_m to ensure that $\mu_n(C_m) \geq 1 - \varepsilon_m$ holds also for $n \leq N_n$ (thus we have $\mu_n(C_m^{\varepsilon_m}) \geq 1 - 2\varepsilon_m$ for every n, m). Finally we define $K := \cap_m \overline{C_m^{\varepsilon_m}}$ and notice that this is closed and totally bounded, hence compact. Also, for every n we have $\mu_n(\mathrm{X}\setminus K) \leq \sum_m \mu_n(\mathrm{X}\setminus C_m^{\varepsilon_m}) \leq \sum_m 2^{1-m}\varepsilon \leq 4\varepsilon$, proving the desired tightness. □

Corollary 5.3 (Ulam's theorem)**.** *Any $\mu \in \mathcal{P}(\mathrm{X})$ is concentrated on a countable union of compact sets.*

Proof. The singleton $\{\mu\}$ is compact (w.r.t. any topology) hence tight by Prokhorov's theorem. □

Corollary 5.4. *Let $\mu \in \mathcal{P}(\mathrm{X})$. Then for every Borel set $E \subset \mathrm{X}$ we have*

$$\mu(E) = \sup\{\mu(K) \ :\ K \subset E \quad \textit{compact}\} = \inf\{\mu(U) \ :\ E \subset U \quad \textit{open}\}.$$

Proof. The formula with open sets has been proved in Lemma 3.5. For that about compact sets, let $K_n \subset \mathrm{X}$ be compact with $\mu(\mathrm{X}\setminus K_n) \to 0$ (use Ulam's theorem) and, given E, let $C_n \subset E$ be closed with $\mu(E\setminus C_n) \to 0$. Then $C_n \cap K_n \subset E$ is compact for every n and $\mu(E \setminus (C_n \cap K_n)) \leq \mu(E \setminus C_n) + \mu(\mathrm{X}\setminus K_n) \to 0$. □

Proposition 5.5. *The Levy-Prokhorov distance D_{LP} on $\mathcal{P}(\mathrm{X})$ is complete. In particular, $\mathcal{P}(\mathrm{X})$ with the weak topology is Polish.*

Proof. Let $(\mu_n) \subset \mathcal{P}(\mathrm{X})$ be D_{LP}-Cauchy. By Theorems 5.2 and 5.1 it suffices to prove that it is tight. Fix $\varepsilon > 0$ and for every $n \in \mathbb{N}$ let $\varepsilon_n := 2^{-n}\varepsilon$ and find $N_n \in \mathbb{N}$ so that $\sup_{m \geq N_n} \mathsf{D}_{\mathsf{LP}}(\mu_n, \mu_m) < \varepsilon_n$. Also, use the tightness of the finite set $\{\mu_1, \ldots, \mu_{N_n}\}$ to find $K_n \subset \mathrm{X}$ compact so that $\mu_m(K_n) \geq 1 - \varepsilon_n$ for every $m \leq N_n$. By definition of D_{LP} and the choice of N_n we also have $\mu_m(K_n^{\varepsilon_n}) \geq 1 - 2\varepsilon_n$ for every $m > N_n$.

Then, as before, we notice that $K := \cap_n \overline{K_n^{\varepsilon_n}}$ is compact (being closed and totally bounded) and $\mu_m(\mathrm{X}\setminus K) \leq \sum_n \mu_m(\mathrm{X}\setminus K_n^{r_n}) \leq 2\varepsilon\sum_n 2^{-n} = 4\varepsilon$ for every $m \in \mathbb{N}$ and the tightness follows. □

Remark 5.6. One could, and often does, define the weak topology on $\mathcal{M}(\mathrm{X})$ in the very same manner. It is clear from (4.1) that on $\mathcal{M}(\mathrm{X})$ such topology is still Hausdorff and from (5.1) that it has a countable base (exercise).

Still, in general the weak topology on $\mathcal{M}(\mathrm{X})$ is *not* metrizable, not even if we restrict the attention to measures with some given bound on the mass. To see why, consider $\mathrm{X} := \mathbb{N}$ with the discrete, Polish, topology and notice that in this case $\mathcal{M}(\mathrm{X}) \cong \ell^1$ and $C_b(\mathrm{X}) \cong \ell^\infty$, so that the weak topology on $\mathcal{M}(\mathrm{X})$ is in fact the weak topology on ℓ^1. The key fact to notice is that a sequence in ℓ^1 weakly converges if and only if it strongly converges (this is Schur's lemma, see also Lemma 10.1). Thus if the weak topology were metrizable on, say, the close unit ball $\overline{B_1(0)}$ in $\mathcal{M}(\mathrm{X})$, then the weak and strong topologies would agree (because on metric spaces converging sequences characterize the topology). However, the two topologies are not the same, because the strong topology has open sets contained in $B_{1/2}(0)$, while every weakly open set must intersect the unit sphere $\{\|\mu\|_{\mathsf{TV}} = 1\}$. ■

6 Disintegration theorem

Let X, Y be Polish spaces. A map $y \mapsto \mu_y \in \mathcal{P}(X)$ is said to be **weakly Borel** if it is Borel whenever we endow the target space with the weak topology, i.e. if it is measurable w.r.t. the Borel σ-algebra on Y and the σ-algebra generated by the weak topology on $\mathcal{P}(X)$.

Lemma 6.1. *Let* X, Y *be Polish and* $y \mapsto \mu_y \in \mathcal{P}(X)$ *given. Then the following are equivalent:*

i) the map is weakly Borel,

ii) For every $\varphi \in C_b(X)$ *the map* $y \mapsto \int \varphi \, d\mu_y \in \mathbb{R}$ *is Borel,*

iii) For every $E \subset X$ *Borel the map* $y \mapsto \mu_y(E) \in [0, 1]$ *is Borel,*

iv) For every $f : X \to [0, +\infty]$ *Borel the map* $y \mapsto \int f \, d\mu_y \in [0, +\infty]$ *is Borel.*

Proof. $(iii) \Leftrightarrow (iv)$ is trivial and so are $(i) \Rightarrow (ii)$ and $(iv) \Rightarrow (ii)$. For $(ii) \Rightarrow (iii)$ we notice that for $U \subset X$ open the functions $\varphi_n := (n\mathsf{d}(\cdot, U^c)) \wedge 1 \in C_b(X)$ monotonically converge to χ_U, thus by monotone convergence it follows that $\mu_y(U) = \sup_n \int \varphi_n \, d\mu_y$ for every y, showing that $y \mapsto \mu_y(U)$ is Borel. The class of Borel sets $E \subset X$ such that $y \mapsto \mu_y(E)$ is Borel is easily seen to be a λ-system, and since it contains the π-system of open sets, by the $\pi - \lambda$ theorem it coincides with the Borel σ-algebra.

For $(ii) \Rightarrow (i)$ we notice that the open sets generated by the maps $y \mapsto \int \varphi \, d\mu_y$ for $\varphi \in C_b(X)$ are a subbase of the weak topology and since such topology is separable, every weakly open set is the countable union of finite intersections of open sets in this subbase (alternatively, use a countable collection of φ's as in (5.1)). This is enough to conclude. □

Theorem 6.2. *Let* X, Y *be Polish,* $\mu \in \mathcal{P}(X)$*,* $T : X \to Y$ *Borel and let* $\nu := T_*\mu$ *(i.e.* $\nu(E) = \mu(T^{-1}(F))$ *for every* $F \subset Y$ *Borel). Then there is a weakly Borel map* $y \mapsto \mu_y \in \mathcal{P}(X)$ *such that:*

1) μ_y *is concentrated on* $T^{-1}(y)$ *for* ν*-a.e.* y*,*

2) For any $\varphi \in C_b(X)$ *we have*

$$\int \varphi \, d\mu = \int \Big(\int \varphi \, d\mu_y \Big) \, d\nu(y) \tag{6.1}$$

If $y \mapsto \tilde{\mu}_y$ *has the same properties, then* $\mu_y = \tilde{\mu}_y$ *for* ν*-a.e.* y*. Also, for any such* (μ_y) *formula (6.1) also holds for any* $\varphi : X \to [0, +\infty]$ *Borel.*

Proof.
Final statement. By monotone convergence we see that the class of φ's for which (6.1) holds is closed by the operation of pointwise supremum of an increasing sequence. Thus, by the approximation also used in the above lemma, such class contains the characteristic functions of open sets. Then, again as above, we notice that the collection of Borel sets E such that for $\varphi := \chi_E$ formula (6.1) holds is a λ-system: since it contains the open sets it coincides with the whole Borel σ-algebra. Then by linearity we see that if φ Borel only attains a finite number of values, then (6.1) holds. A final monotone approximation yields the claim for general non-negative Borel functions.
Uniqueness. Let $E \subset X$ and $F \subset Y$ be Borel. Then by what just proved we can pick $\varphi := \chi_E \chi_{T^{-1}(F)}$ in (6.1) written with both (μ_t) and $(\tilde{\mu}_t)$ to deduce, taking into account also property (1), that $\int_F \mu_y(E) \, d\nu(y) = \int_F \tilde{\mu}_y(E) \, d\nu(y)$. Being this true for any Borel $F \subset Y$ we can deduce that $\mu_y(E) = \tilde{\mu}_y(E)$ for every $y \in Y \setminus N(E)$, with $N(E) \subset Y$ Borel and ν-negligible. Let then A be a countable algebra on X generating the Borel σ-algebra (e.g. the algebra generated by a countable base of open sets) and notice that $N := \cup_{E \in A} N(E)$ is Borel and ν-negligible and that for every $y \in Y \setminus N$ we have $\mu_y(E) = \tilde{\mu}_y(E)$ for every $E \in A$. By Corollary 2.2 this suffices to conclude that $\mu_y = \tilde{\mu}_y$ for every $y \in Y \setminus N$.

Existence As before, let A be a countable algebra on X generating the Borel σ-algebra. For each $E \in A$ choose an increasing sequence of compact sets $K_n \subset E$ with $\mu(K_n) \uparrow \mu(E)$ and then let $\tilde{A}$ be the algebra, still countable, generated by A and these compact sets. For each $E \in \tilde{A}$ we obviously have $0 \leq \chi_E \mu \leq \mu$ and thus the Radon-Nikodym derivative $\frac{\mathrm{d}(T_*(\chi_E\mu))}{\mathrm{d}\nu}$ admits a Borel representative $\rho(E)(\cdot) : \mathrm{Y} \to [0, 1]$. By the very definition of Borel representative and the fact that $\tilde{A}$ is countable we see that there is $N \subset \mathrm{Y}$ Borel and ν-negligible such that for every $y \in \mathrm{Y} \setminus N$ we have:

a) $\rho(\emptyset)(y) = 0$ and $\rho(\mathrm{X})(y) = 1$ (because $\chi_\emptyset\mu = 0$ and $\chi_\mathrm{X}\mu = \mu$)

b) $\rho(E)(y) + \rho(F)(y) = \rho(E + F)(y)$ if $E \cap F = \emptyset$ (because $\chi_E\mu + \chi_F\mu = \chi_{E\cup F}\mu$ if $E \cap F = \emptyset$)

c) For any $E \in A$ and the previously chosen sequence $(K_n) \subset \tilde{A}$ associated to E we have $\rho(K_n)(y) \uparrow \rho(E)(y)$ (because $\chi_{K_n}\mu \leq \chi_E\mu$ for every n and $\mu(K_n) \uparrow \mu(E)$).

We claim that for any $y \in \mathrm{Y} \setminus N$ the map $A \ni\mapsto \rho_E(y) \in [0, 1]$ is the restriction to A of a Borel probability measure. According to Proposition 2.4, given $(a), (b)$ above to prove this it suffices to prove that if $(E_n) \subset A$ is decreasing with $\cap_n E_n = \emptyset$, then $\rho_{E_n}(y) \downarrow 0$. Say not. Then using (c) we can find compact sets $K_n \subset E_n$ with $\sum_n \rho_{E_n\setminus K_n}(y) < \varepsilon$. It follows that for every $n \in \mathbb{N}$ we have

$$\rho(\cap_{i\leq n}K_i)(y) = \rho(E_n)(y) - \rho(\cup_{i\leq n}E_n \setminus K_i)(y) > \varepsilon - \rho(\cup_{i\leq n}E_i \setminus K_i)(y) > 0$$

and in particular (by (a)) we have $\cap_{i\leq n}K_i \neq \emptyset$ for any $n \in \mathbb{N}$. Being the K_n's compact, this implies that $\cap_{i\in\mathbb{N}}K_i \neq \emptyset$, contradicting the fact that $K_n \subset E_n$ and $\cap_n E_n = \emptyset$.

Thus for $y \in \mathrm{Y} \setminus N$ the map $A \ni\mapsto \rho_E(y) \in [0, 1]$ is indeed the restriction to A of a Borel probability (by (a)) measure, that we shall denote μ_y. For $y \in N$ we set $\mu_y := \delta_{\bar{x}}$, where $\bar{x} \in \mathrm{X}$ is some arbitrarily chosen point. Since for any $E \in A$ Borel the map $y \mapsto \rho(E)(y)$ is Borel, it is easy to see that also $y \mapsto \mu_y(E)$ is Borel. Since the class of Borel sets E for which $y \mapsto \mu_y(E)$ is Borel is, quite clearly, a λ-system, and A generates the Borel σ-algebra, we see that $y \mapsto \mu_y(E)$ is Borel for any $E \subset \mathrm{X}$ Borel, which by the above lemma proves that $y \mapsto \mu_y$ is weakly Borel.

For any $E \in A$ we have $\int \mu_y(E)\,\mathrm{d}\nu(y) = \int \rho(E)(y)\,\mathrm{d}\nu(y) = \int \frac{\mathrm{d}T_*(\chi_E\mu)}{\mathrm{d}\nu}\,\mathrm{d}\nu = \int 1\,\mathrm{d}T_*(\chi_E\mu) = \int \chi_E\,\mathrm{d}\mu$, proving that formula (6.1) holds for $\varphi := \chi_E$. Then again an approximation argument, based first on the $\pi - \lambda$-theorem to extend the above to all Borel sets, then linearity and uniform density of simple functions, shows that (6.1) holds for any bounded and Borel functions, and in particular for $\varphi \in C_b(\mathrm{X})$.

It remains to prove that μ_y is concentrated on $T^{-1}(y)$ for ν-a.e. y. To see this recall that for $E \in A$ the identity $\mu_y(E) = \frac{\mathrm{d}T_*(\chi_E\mu)}{\mathrm{d}\nu}(y)$ holds for ν-a.e. $y \in \mathrm{Y}$, and since the collection of sets with this property is clearly a λ-system, by the $\pi - \lambda$-theorem we see that this holds for general Borel sets E. In particular, this is true for $E := T^{-1}(F)$ for $F \subset \mathrm{Y}$ Borel and recalling the identity $T_*(f \circ T\mu) = fT_*\mu$ (that can be checked directly from the definition) we see that $\mu_y(T^{-1}(F)) = \chi_F(y)$ holds for ν-a.e. y for all $F \subset \mathrm{Y}$ Borel. Let B be a countable base of the topology of Y. Then there is a ν-negligible Borel set $N \subset \mathrm{Y}$ such that for any $y \in \mathrm{Y}\setminus N$ we have $\mu_y(T^{-1}(F)) = \chi_F(y)$ for any $F \in B$. Fix $y \in \mathrm{Y} \setminus N$ and find a decreasing sequence $(F_n) \subset B$ with $\cap_n F_n = \{y\}$: then we have $\mu_y(T^{-1}(y)) = \lim_n \mu_y(T^{-1}(F_n)) = 1$, proving the claim. □

Remark 6.3 (Variants)**.**

- The statements holds, with trivial modifications to the proof, even if the measure μ is only σ-finite, provided $\nu = T_*\mu$ is σ-finite as well.
- The Polish structure on Y played no role: we only used that its σ-algebra is countably generated and contains the singletons.
- In probability theory the disintegration theorem is often used to speak about *conditional expectations*. In this case, rather than having a map between two spaces one typically has two different σ-algebras on the same space, one included in the other, with a measure given on the biggest one. The smallest one, that 'does not distinguish all the events' so to say, is the one on which we want to condition our expectation. In the language of the theorem as we stated it, the biggest σ-algebra would be the Borel σ-algebra on X and the smallest one that generated by sets of the form $T^{-1}(E)$ for $E \subset \mathrm{Y}$ Borel.
- In some instances, rather than giving a target set Y and a map, one is only given an equivalence relation on X (or, which is the same, a partition of X) and takes Y as the quotient space and T as the quotient map. In this scenario the required regularity properties of Y and T are read through properties of the equivalence relation, such as, for instance, the existence of a Borel subset of X meeting each equivalence class precisely once. ■

Exercise 6.4. Let X be Polish and $\mu \in \mathcal{P}(\mathrm{X})$. Prove that $\{\nu \in \mathcal{P}(\mathrm{X}) : \nu \ll \mu\}$ is weakly Borel. ■

7 Kolmogorov's theorem

Let I be an infinite set of indexes (not necessarily countable) and for each $i \in I$ let X_i be a Polish space. Consider the product space $\mathrm{X} := \Pi_{i\in I}\mathrm{X}_i$ equipped with the product topology (always Polish if I is countable, not necessarily so otherwise). For $F \subset I$ not empty put $\mathrm{X}_F := \Pi_{i\in F}\mathrm{X}_i$ and for $F_1 \subset F_2 \subset I$ let $\mathrm{Pr}^{F_2}_{F_1} : \mathrm{X}_{F_2} \to \mathrm{X}_{F_1}$ be the natural projection. Let $\mu \in \mathcal{P}(\mathrm{X})$ and define $\mu_F := (\mathrm{Pr}^I_F)_*\mu$ for any $\emptyset \neq F \subset I$. Since $\mathrm{Pr}^{F_2}_{F_1} \circ \mathrm{Pr}^I_{F_2} = \mathrm{Pr}^I_{F_1}$ for $F_1 \subset F_2$ we see that

$$(\mathrm{Pr}^{F_2}_{F_1})_*\mu_{F_2} = \mu_{F_1} \qquad \forall \emptyset \neq F_1 \subset F_2 \subset I. \tag{7.1}$$

Now assume that for any non-empty finite set $F \subset I$ we are given $\mu_F \in \mathcal{P}(\mathrm{X}_F)$ and we ask whether there is $\mu \in \mathcal{P}(\mathrm{X})$ such that $\mu_F = (\mathrm{Pr}^I_F)_*\mu$ for any such F. Clearly, the above relation (7.1) is a necessary condition. Kolmogorov's theorem ensures it is sufficient as well:

Theorem 7.1. *With the above notation and assumption, assume that for every non-empty $\emptyset \neq F \subset I$ finite we are given $\mu_F \in \mathcal{P}(\mathrm{X}_F)$ so that the compatibility relation (7.1) holds for F_1, F_2 finite.*

Then there is a unique $\mu \in \mathcal{P}(\mathrm{X})$ such that $\mu_F = (\mathrm{Pr}^I_F)_*\mu$ holds for any such F.

Proof. Let $\mathcal{A}$ be the collection of subsets of X of the form $(\mathrm{Pr}^I_F)^{-1}(B)$ for some $F \subset I$ finite and $B \subset \mathrm{X}_F$ Borel. Then, clearly, $\mathcal{A}$ is an algebra that generates the Borel σ-algebra on X. The requirement $\mu_F = (\mathrm{Pr}_F)_*\mu$ forces $\mu((\mathrm{Pr}^I_F)^{-1}(B))$ to be equal to $\mu_F(B)$, so that uniqueness follows from Proposition 2.4. For existence we start defining μ on $\mathcal{A}$ as above. Then to conclude again by Proposition 2.4 it suffices to prove that if $(A_n) \subset \mathcal{A}$ is a decreasing sequence with $\cap_n A_n = \emptyset$, then $\mu(A_n) \to 0$.

Say not i.e. that $\mu(A_n) > \varepsilon > 0$ for every n and some ε. Then there are finite subsets F_n of I and $B_n \subset \mathrm{X}_{F_n}$ Borel such that $A_n = (\mathrm{Pr}^I_{F_n})^{-1}(B_n)$ for every n. Replacing F_n with $\cup_{i\le n}F_i$ we can assume (F_n) to be increasing. Then by inner regularity we can find compact sets $K_n \subset B_n$ such that $\sum_n \mu_{F_n}(B_n \setminus K_n) < \varepsilon$. Put for brevity $\mathrm{P}_{nm} := \mathrm{Pr}^{F_n}_{F_m}$ for every $m \le n$ and let $C_n := \cap_{i\le n}\mathrm{P}^{-1}_{ni}(K_i)$. Since K_n is compact and $\mathrm{P}^{-1}_{ni}(K_i)$ closed, we see that $C_n \subset \mathrm{X}_{F_n}$ is compact. Since $B_n \subset \mathrm{P}^{-1}_{ni}(B_i)$ for $i \le n$ (because $A_n \subset A_i$) by (7.1) we get

$$\mu_{F_n}(C_n) = \mu_{F_n}(B_n) - \mu_{F_n}(B_n \setminus C_n) > \varepsilon - \mu_{F_n}\big(\bigcup_{i\le n} B_n \setminus \mathrm{P}^{-1}_{ni}(K_i)\big) \ge \varepsilon - \sum_{i\le n}\mu_{F_i}(B_i \setminus K_i) > 0$$

proving that $C_n \neq \emptyset$. We claim that

$$\mathrm{P}_{kn}(C_k) \subset \mathrm{P}_{mn}(C_m) \qquad \forall k \ge m \ge n \tag{7.2}$$

and notice that to prove this it suffices to show that $\mathrm{P}_{km}(C_k) \subset C_m$. In turn, this follows from $\mathrm{P}^{-1}_{km}(C_m) = \cap_{i\le m}\mathrm{P}^{-1}_{km}(\mathrm{P}^{-1}_{mi}(K_i)) = \cap_{i\le m}\mathrm{P}^{-1}_{ki}(K_i)) \supset C_k$. We then define $D_n := \cap_{k\ge n}\mathrm{P}_{kn}(C_k) \subset \mathrm{X}_{F_n}$ and notice that being this the intersection of a decreasing sequence of non empty compact sets it is also non empty. We claim that

$$\mathrm{P}_{mn}(D_m) = D_n \qquad \forall m \ge n \tag{7.3}$$

and notice that since 'image of the intersection $\subset$ intersection of the images', the inclusion $\mathrm{P}_{mn}D_m \subset D_n$ is obvious. For the converse, let $x \in D_n$ so that in particular $x \in \mathrm{P}_{kn}(C_k) = \mathrm{P}_{mn}\mathrm{P}_{km}(C_k)$ for every $k \ge m \ge n$. Then the sets $E_k := \mathrm{P}_{km}(C_k) \cap \mathrm{P}^{-1}_{mn}(x)$ are non empty (by what just said), compact (as intersection of a compact and a closed), and decreasing in k (by (7.2)). Hence $\cap_k E_k \neq \emptyset$ and any element y of this intersection belongs to D_m and satisfies $\mathrm{P}_{mn}(y) = x$, proving our claim (7.3).

Pick $x_0 \in D_0$ arbitrary, use (7.3) to find $x_1 \in D_1$ with $\mathrm{P}_{10}(x_1) = x_0$ and iterate to produce a sequence $n \mapsto x_n \in D_n$ with $\mathrm{P}_{mn}(x_m) = x_n$ for every $m \ge n$. In other words, for $i \in F_n \subset I$ and $m \ge n$ the i-th coordinates of x_n and x_m agree. Hence we can define $x \in \mathrm{X}$ so that for every $n \in \mathbb{N}$ and $i \in F_n$ the i-th coordinate of x is that of x_n (what just proved shows that this is a good definition, meaning it only depends on i and not on n) and chosen arbitrarily for $i \notin \cup_n F_n$.

Then we have $\mathrm{P}^I_{F_n}(x) = x_n \in D_n \subset C_n \subset K_n$ for every $n \in \mathbb{N}$, proving that $\cap_n C_n = \cap_n(\mathrm{Pr}^I_{F_n})^{-1}(K_n) \subset \cap_n(\mathrm{Pr}^I_{F_n})^{-1}(B_n) = \cap_n A_n$ is not empty, giving the desired contradiction. □

8 Borel selection

A **multivalued map** $F : \mathrm{X} \to \mathrm{Y}$ is a map that assigns to every $x \in \mathrm{X}$ a subset of Y. Its **graph** is the set $\{(x, y) : y \in F(x)\} \subset \mathrm{X} \times \mathrm{Y}$. A **selector** for F is a (ordinary) map $f : \mathrm{X} \to \mathrm{Y}$ such that $f(x) \in F(x)$ for every $x \in \mathrm{X}$. A necessary condition for a selector to exist (also sufficient, if one believes in the Axiom of Choice) is that $F(x) \neq \emptyset$ for every $x \in \mathrm{X}$. In general, if F has some regularity/structural properties, one would like to build a selector sharing these properties (and here the Axiom of Choice is of no help). Here we discuss measurability.

Let $\mathcal{A}$ be a σ-algebra on X and Y a topological space. We say that F is $\mathcal{A}$-measurable if

$$\text{for any } U \subset \mathrm{Y} \text{ open the set } \{x \in \mathrm{X} : F(x) \cap U \neq \emptyset\} \text{ is in } \mathcal{A}. \tag{8.1}$$

Notice that if F is single valued this reduces to ordinary $\mathcal{A}$-measurability.

Theorem 8.1 (Kuratowski–Ryll-Nardzewski)**.** *Let* $(\mathrm{X}, \mathcal{A})$ *be a set with a* σ*-algebra,* Y *Polish and* $F : \mathrm{X} \to \mathrm{Y}$ *multivalued,* $\mathcal{A}$*-measurable so that* $F(x)$ *is a non-empty* closed *set for every* $x \in \mathrm{X}$*. Then* F *admits an* $\mathcal{A}$*-measurable selector.*

Proof. Let $\mathsf{d} < 1$ be a complete and separable distance inducing the topology of Y (replace a given d with $\frac{1}{2} \wedge \mathsf{d}$ if necessary). We shall define a sequence of $\mathcal{A}$-measurable maps $f_n : \mathrm{X} \to \mathrm{Y}$ so that $\mathsf{d}(f_n(x), F(x)) < 2^{-n}$ and $\mathsf{d}(f_n(x), f_{n+1}(x)) < 2^{-(n-1)}$ for every $n \in \mathbb{N}$. If these are built, the desired selector f can be found as pointwise limit of the f_n's.

Let f_0 be constant. Given f_n and given $y \in \mathrm{Y}$ and $r > 0$ consider the sets $\{x \in \mathrm{X} : \mathsf{d}(f_n(x), y) < r\}$ and $\{x \in \mathrm{X} : \mathsf{d}(F(x), y) < r\}$ (here $\mathsf{d}(A, y) := \inf_{z \in A} \mathsf{d}(z, y)$). These are $\mathcal{A}$-measurable: the first by the measurability of f_n, the second by that of F. Let now $(y_i) \subset \mathrm{Y}$ be countable and dense and put

$$A_i := \{x \in \mathrm{X} : \mathsf{d}(F(x), y_i) < 2^{-n},\ \mathsf{d}(f_n(x), y_i) < 2^{-(n-1)}\}.$$

The density of (y_i) and the inductive assumption give that $\cup_i A_i = \mathrm{X}$ and by what previously said we know that the A_i's are measurable. Let then $B_0 := A_0$, $B_i := A_i \setminus \cup_{j<i} A_j$ and define f_{n+1} so that, for every i, it is constantly equal to y_i on B_i. $\square$

9 Measurable projection and selection

Let X be Polish. A **Souslin subset** $S \subset \mathrm{X}$ is, by definition, a set of the form $f(\mathrm{Y})$ for some Polish Y and $f : \mathrm{Y} \to \mathrm{X}$ continuous (here we allow Y to be empty, so that $\emptyset \subset \mathrm{X}$ is Souslin).

Exercise 9.1. Prove that any non-empty Polish space is a continuous image of $\mathbb{N}^{\mathbb{N}}$. ■

Lemma 9.2. *The class of Souslin subsets of* X *is closed by countable unions and intersections*[2].

Proof. Let $S_n = f_n(\mathrm{Y}_n)$, put $\mathrm{Y} := \mathbb{N} \times \Pi_n \mathrm{Y}_n$ and define $f : \mathrm{Y} \to \mathrm{X}$ as $f(n, (y_i)) := f_n(y_n)$. Then f is continuous with $f(\mathrm{Y}) = \cup_n S_n$. Also, the subset $\mathrm{Z} \subset \mathrm{Y}$ of $(n, (y_i))$ so that $f_1(y_1) = f_2(y_2) = ...$ is closed, hence Polish, and $f(\mathrm{Z}) = \cap_n S_n$. □

Proposition 9.3. *Borel subsets of* X *are Souslin.*

Proof. Let $\mathcal{A}$ be the smallest collection of subsets of X containing the closed sets and stable by countable unions and intersections and notice that by the above lemma it easily follows that elements of $\mathcal{A}$ are Souslin. Put $\mathcal{A}' := \{E \in A : E^c \in A\} \subset A$ and observe that $\mathcal{A}'$ is also stable by countable unions and intersections. Also, since every open subset of X is the countable union of closed sets, we see that closed sets are in $\mathcal{A}'$ and thus that $\mathcal{A}' \supset \mathcal{A}$ and therefore $\mathcal{A} = \mathcal{A}'$. This shows that $\mathcal{A}$ is a σ-algebra, hence the Borel σ-algebra. □

Lemma 9.4. *Let μ be a finite non-negative measure, μ^* the outer measure it induces (as in* (2.1)*) and $A_n \uparrow A$. Then $\mu^*(A_n) \uparrow \mu^*(A)$.*

Proof. Given $\varepsilon > 0$ we have $\mu^*(A_n) + \varepsilon \geq \mu(B_n)$ for some measurable set $B_n \supset A_n$. Put $B := \cup_n \cap_{i \geq n} B_i \supset A$ and notice that

$$\mu^*(A) \leq \mu(B) = \lim_n \mu(\cap_{i \geq n} B_i) \leq \lim_n \mu(B_n) \leq \lim_n \mu^*(A_n) + \varepsilon. \qquad \square$$

Theorem 9.5. *Let μ be a finite Borel measure on* X*. Then Souslin sets are μ-measurable.*

Proof. Let $S \subset \mathrm{X}$ be Souslin and $\varepsilon > 0$. It suffices to find $K \subset S$ compact with $\mu(K) \geq \mu^*(S) - \varepsilon$. Let $S = f(\mathrm{Y})$ and $C_1 \subset \mathrm{Y}$ the union of a finite number of closed balls of radius ≤ 1 with $\mu^*(f(C_1)) > \mu^*(S) - \varepsilon$ (since Y is the union of a countable number of such balls, the existence of C_1 follows by the Lemma above). Then by induction find $C_n \subset C_{n-1}$ finite union of closed balls of radius $\leq \frac{1}{n}$ such that $\mu^*(f(C_n)) > \mu^*(S) - \varepsilon$. Notice that $\cap_n C_n$ is compact, as is closed and totally bounded, and then let $K := f(\cap_n C_n)$, that is compact as the continuous image of a compact set. We claim that $K = \cap_n \overline{f(C_n)}$ and notice that if we prove this we are done, as then $\mu(K) = \lim_n \mu(\overline{f(C_n)}) \geq \overline{\lim}_n \mu^*(f(C_n)) > \mu^*(S) - \varepsilon$.

The inclusion $K \subset \cap_n \overline{f(C_n)}$ is trivial (and irrelevant for our argument). For $\supset$ we take $x \in \cap_n \overline{f(C_n)}$ and then $y_n \in C_n$ such that $\mathsf{d}(f(y_n), x) \leq \frac{1}{n}$. The key observation is that (y_n) is relatively compact, because is totally bounded, hence has a subsequence admitting a limit $y \in \cap_n C_n$. The continuity of f gives $f(y) = x$. □

Theorem 9.6. *Let* X, Y *be Polish, $\mathcal{A}$ a σ-algebra on* X *containing Souslin sets and $F : \mathrm{X} \to \mathrm{Y}$ a multivalued map with Souslin graph and $F(x) \neq \emptyset$ for all $x \in \mathrm{X}$. Then there is a selection $\mathcal{A}$-measurable.*

Proof. Since the graph is Souslin, there are a Polish space Z and a continuous map $g : \mathrm{Z} \to \mathrm{X} \times \mathrm{Y}$ with $g(\mathrm{Z}) = \mathrm{Graph}(F)$. Let $\tilde{F} : \mathrm{X} \to \mathrm{Z}$ multivalued as "$g^{-1} \circ F$", i.e. (x, z) is in the graph of $\tilde{F}$ iff $x = \mathrm{Pr}_{\mathrm{X}}(g(z))$. The continuity of g grants that $\tilde{F}(x) \subset \mathrm{Z}$ is closed and the fact that $F(x) \neq \emptyset$ that $\tilde{F}(x) \neq \emptyset$ as well. Given $U \subset \mathrm{Z}$ arbitrary we have $\{x \in \mathrm{X} : \tilde{F}(x) \cap U \neq \emptyset\} = \mathrm{Pr}_{\mathrm{X}}(g(U))$. In particular this holds if U is open, which shows that $\tilde{F}$ is $\mathcal{A}$-measurable (recall that $\mathcal{A}$ contains Souslin sets) as in (8.1). Hence $\tilde{F}$ admits a selector $\mathcal{A}$-measurable $\tilde{f} : \mathrm{X} \to \mathrm{Z}$. It follows that $f(x) := \mathrm{Pr}_{\mathrm{Y}}(g(\tilde{f}(x)))$ is an $\mathcal{A}$-measurable selector for F. □

Exercise 9.7. Let X, Y be Polish, μ a σ-finite Borel measure on X and $f : \mathrm{X} \to \mathrm{Y}$ so that preimages of open sets are in the σ-algebra generated by Souslin sets. Prove that there is $\tilde{f} : \mathrm{X} \to \mathrm{Y}$ Borel equal to f outside a μ-negligible set. ■

---

[2] but not complementation!

10 A tightness criterion

Lemma 10.1. *(Steinhaus-Schur) Let $a_n = (\ldots, a_n^i, \ldots) \in \ell^1$ for every $n \in \mathbb{N}$ be so that for any bounded sequence $\lambda = (\ldots, \lambda^i, \ldots) \in \ell^\infty$ the sequence $n \mapsto \sum_i a_n^i \lambda^i \in \mathbb{R}$ is Cauchy.*

Then (a_n) is is Cauchy w.r.t. the ℓ^1-distance.

Proof. It is immediate to verify that it is sufficient to prove that 'tails are uniformly small', i.e. that for every $\varepsilon > 0$ there is $N \in \mathbb{N}$ such that $\sum_{i>N} |a_n^i| < \varepsilon$ for every $n \in \mathbb{N}$. Say not, i.e. assume that

$$\exists \varepsilon > 0 \text{ such that } \forall N, M \in \mathbb{N} \text{ there is } n = n(N, M) > M \in \mathbb{N} \text{ such that } \sum\nolimits_{i>N} |a_n^i| > \varepsilon. \tag{10.1}$$

Choosing $\lambda = e^i := (\ldots, 0, 1, 0, \ldots)$ with the 1 in the i-th position, from the fact that $n \mapsto \sum_i a_n^i \lambda^i \in \mathbb{R}$ is Cauchy we see that for every $i \in \mathbb{N}$ there exists the limit a^i of (a_n^i). Let $a := (\ldots, a^i, \ldots)$, notice that our assumption and the uniform boundedness principle grant that (a_n) is a bounded sequence in ℓ^1 and then use Fatou's lemma to deduce that $\sum_i |a^i| \le \underline{\lim}_n \sum_i |a_n^i| < \infty$. Thus $a \in \ell^1$ and therefore

$$\text{there exists } \bar{N} \text{ such that } \sum\nolimits_{i>\bar{N}} |a^i| < \frac{\varepsilon}{3}. \tag{10.2}$$

Let $N_1 := \bar{N}$ and apply our assumption (10.1) to find $n_1 = n(N_1, 0)$ and N_2 such that $\sum_{i=N_1}^{N_2} |a_{n_1}^i| > \varepsilon$ and $\sum_{i>N_2} |a_{n_1}^i| < \frac{\varepsilon}{3}$. Then, using the convergence $a_n^i \to a^i$ and (10.2), find $M_1 \in \mathbb{N}$ such that $\sum_{i=N_1}^{N_2} |a_n^i| < \frac{\varepsilon}{3}$ for every $n > M_1$. Continue recursively: assume N_j, M_{j-1} are given, then define $n_j := n(N_j, M_{j-1})$, let N_{j+1} be such that $\sum_{i=N_j}^{N_{j+1}} |a_{n_j}^i| > \varepsilon$ and $\sum_{i>N_{j+1}} |a_{n_j}^i| < \frac{\varepsilon}{3}$ and finally let M_j be such that $\sum_{i=\bar{N}}^{N_{j+1}} |a_n^i| < \frac{\varepsilon}{3}$ for every $n > M_j$. Define the bounded sequence $\lambda = (\ldots, \lambda^i, \ldots)$ as

$$\lambda^i := \begin{cases} 0, & \forall i < \bar{N}, \\ \operatorname{sign}(a_{n_j}^i), & \forall i \in \{N_j, \ldots, N_{j+1} - 1\} \quad \text{if } j \text{ is odd}, \\ -\operatorname{sign}(a_{n_j}^i), & \forall i \in \{N_j, \ldots, N_{j+1} - 1\} \quad \text{if } j \text{ is even} \end{cases}$$

and notice that for every $j \in \mathbb{N}$ we have $\sum_i a_{n_j}^i \lambda^i = \sum_{i=\bar{N}}^{N_j - 1} a_{n_j}^i \lambda^i + \sum_{i=N_j}^{N_{j+1}-1} a_{n_j}^i \lambda^i + \sum_{i \ge N_{j+1}} a_{n_j}^i \lambda^i$. By construction the first and last term on the right hand side are bounded in modulus by $\frac{\varepsilon}{3}$, while the middle one is $> \varepsilon$ or $< -\varepsilon$ depending on the parity of j. Hence $n \mapsto \sum_i a_n^i \lambda^i$ is not Cauchy, giving the desired contradiction. □

Theorem 10.2. *(Alexandrov) Let X be Polish, d a complete and separable distance inducing the topology and $(\mu_n) \subset \mathcal{P}(\mathrm{X})$ so that for every $f : \mathrm{X} \to \mathbb{R}$ bounded and Lipschitz the sequence $n \mapsto \int f \, \mathrm{d}\mu_n$ is Cauchy.*

Then (μ_n) is tight (and thus by Prokhorov's theorem and (5.1) has a weak limit).

Proof. It suffices to prove that for every $\varepsilon, r > 0$ there is a compact set $K = K(\varepsilon, r) \subset \mathrm{X}$ such that for the r-neighborhood $K^r := \{x \in \mathrm{X} : \mathsf{d}(x, K) < r\}$ of K we have $\sup_n \mu_n(\mathrm{X} \setminus K^r) \le \varepsilon$. Indeed, if this is the case, for any $\varepsilon > 0$ and chosen sequence $r_n \downarrow 0$, for $K_n := K(\varepsilon 2^{-n}, r_n)$ we have that the set $K := \cap_n \overline{K_n^{r_n}}$ is compact (being closed and totally bounded) and such that $\mu_n(\mathrm{X} \setminus K) \le 2\varepsilon$ for every $n \in \mathbb{N}$.

We shall argue by contradiction and assume instead that

$$\exists \varepsilon, r > 0 \text{ such that } \forall K \subset \mathrm{X} \text{ compact there is } n = n(K) \in \mathbb{N} \text{ such that } \mu_n(\mathrm{X} \setminus K^r) > \varepsilon.$$

Let $K_0 := \emptyset$ and define recursively indexes n_i and compact sets K_i, H_i for $i > 0$ as follows. We put $n_i := n(K_{i-1})$, the by inner regularity we find $H_i \subset \mathrm{X} \setminus K_{i-1}^r$ such that $\mu_{n_i}(H_i) > \varepsilon$ and then we put $K_i := K_{i-1} \cup H_i$. Also, for every $i > 0$ let $f^i \in \mathrm{Lip_{bs}}(\mathsf{X})$ be defined as $f^i := \max\{0, 1 - (3r)^{-1} \mathsf{d}(\cdot, H_i)\}$ and notice that these functions are uniformly bounded and Lipschitz and with disjoint supports. In particular, for any $\lambda = (\ldots, \lambda^i, \ldots) \in \ell^\infty$ the function $f^\lambda := \sum_i \lambda^i f^i$ is bounded and Lipschitz and thus putting $a_j^i := \int f^i \, \mathrm{d}\mu_{n_j}$ we have that $j \mapsto \sum_{i>0} a_j^i \lambda^i = \int f^\lambda \, \mathrm{d}\mu_{n_j}$ is a Cauchy sequence.

By Proposition 10.1 we deduce that the sequence $j \mapsto a_j := (\ldots, a_j^i, \ldots) \in \ell^1$ is Cauchy w.r.t. the ℓ^1-norm, thus letting $a = (\ldots, a^i, \ldots) \in \ell^1$ its limit we see that

$$\varlimsup_{i \to \infty} |a_{n_i}^i| \le \varlimsup_{i \to \infty} |a^i| + \varlimsup_{i \to \infty} \sum_j |a_{n_i}^j - a^j| = 0.$$

This contradicts the fact that, by construction, we have $a_{n_i}^i > \varepsilon$ for every $i > 0$. The conclusion follows. □

References

[Bog18] Vladimir I. Bogachev. *Weak convergence of measures*, volume 234 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2018.

[Dud02] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.