Advanced Probability Theory

Davi Sales Barreira

December 6, 2021

Contents

List of Definitions

List of Theorems

Notes mostly based on Teixeira and Lacoin [5], Sokol and Rønn-Nielsen [4] and the course Probability II at IMPA ministered in 2020 by professor Augusto Teixeira. Most introductions to Probability focuse on $\mathbb{R}$. Thus, I call this "Advanced Probability", because we'll present the theory in Metric Spaces.

1 Probability Basics

1.1 Defining a Probability Measure

Definition 1.1 (σ-Algebra). We say that $\mathcal{F} \subset \mathcal{P}(\Omega)$ is a σ-algebra if

(i) $\Omega \in \mathcal{F}$,

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

(iii) if $A_n \in \mathcal{F}$ for all $n \in \mathbb{N} \implies \cup_{n\in\mathbb{N}} A_n \in \mathcal{F}$.

Definition 1.2 (Borel σ-algebra). The Borel σ-algebra of $\mathbb{R}$ is the one generated by the family of all open sets on $\mathbb{R}$. Let U be the set of all open sets, $\Lambda := \{\mathcal{G} : U \subset \mathcal{G}\}$ is the family of all σ-algebras that contains all open sets, hence

$$\mathcal{B} := \bigcap_{\mathcal{F}_i \in \Lambda} \mathcal{F}_i \tag{1}$$

Definition 1.3 (Measurable space). We call $(\Omega, \mathcal{F})$ a measurable space, where $\mathcal{F}$ is a σ-algebra of Ω.

Definition 1.4 (Measure). We call μ a (positive) measure in the measurable space $(\Omega, \mathcal{F})$ if $\mu : \mathcal{F} \to [0, +\infty]$ and μ is *countably-additive*, i.e. for any disjoint countable collection of sets $A_n \in \mathcal{F}$, we have

$$\mu\left(\bigcup_{n=1}^{+\infty} A_n\right) = \mu\left(\sum_{n=1}^{+\infty} A_n\right) = \sum_{n\in\mathbb{N}} \mu(A_n). \tag{2}$$

Definition 1.5 (Probability Measure). A measure μ defined on $(\Omega, \mathcal{F})$ is a probability measure if

$$\mu(\Omega) = 1. \tag{3}$$

Definition 1.6 (Probability Space). We call the triple $(\Omega, \mathcal{F}, P)$ a probability space if P is a probability measure $P : \mathcal{F} \to [0,1]$, and $\mathcal{F}$ is a σ-algebra on Ω.

Definition 1.7 (Space of Probabilities). We call $\mathcal{P}(E)$ the space of probability measures defined on $(E, \mathcal{F})$, where the σ-algebra $\mathcal{F}$. Let (E, d) be a metric space. Whenever we write $\mathcal{P}(E)$ without specifying more about the underlying space E, then we are assuming that we are using the topology induced by d and that we are working on the Borel σ-algebra of such induced topology.

Proposition 1.8 (Basic properties of probability measures). Let $(\Omega, \mathcal{F}, P)$ be a probability space. The following affirmatives are true:

(i) (**Complement**). $P(A) = 1 - P(A^c)$;

(ii) (**Monotonicity**). If $A \subset B$, then $P(A) \leq P(B)$;

(iii) (**Subadditivity**). $P(\cup_{n=1}^{+\infty} A_n) \leq \sum_{n=1}^{+\infty} P(A_n)$;

(iv) (**Continuity from below**). If $A_n \uparrow A$, then $\lim_n P(A_n) = P(A)$;

(v) (**Continuity from above**). If $A_n \downarrow A$, then $\lim_n P(A_n) = P(A)$;

(vi) (**Continuity**). If $A_n \to A$, then $\lim_n P(A_n) = P(A)$.

Proof.
(i)
$$1 = P(\Omega) = P(A \cup A^c) = P(A + A^c) = P(A) + P(A^c).$$

(ii)

$$B \cap A = A, \; B = (B \backslash A) + (B \cap A) = (B \backslash A) + A \implies P(B) = P(B \backslash A) + P(A) \implies P(B) \geq P(A).$$

(iii) Let $B_1 = A_1, B_2 = A_2 \setminus A_1, B_3 = A_3 \setminus (A_1 \cup A_2)...$ which implies that $B_1 + B_2 + ... = A$. Hence,

$$P(A) = P(\sum_{n=1}^{+\infty} B_n) = \sum_{n=1}^{+\infty} P(B_n) = \lim_{k \to +\infty} \sum_{n=1}^{k} P(B_n) = \lim_{k \to +\infty} P(A_k).$$

(iv) Note that $A_1 \supset A_2 \implies A_1^c \subset A_2^c$, thus $A_n \downarrow A$ implies $A_n^c \uparrow A^c$. Now use the same argument as (iii) to show that $\lim P(A_n^c) = P(A^c) = 1 - P(A) = \lim 1 - P(A_n) \implies P(A) = \lim P(A_n)$.

(v) If $A_n \to A$, then for $B_k = \cup_{n\geq k} A_n$, we have $B_k \supset B_{k+1}$ and

$$\lim A_n = \limsup A_n = \lim_{k\to+\infty} \bigcup_{n\geq k} A_n = \bigcap_{k=1}^{+\infty} \bigcup_{n\geq k} A_n = \bigcap_{k=1}^{+\infty} B_k \implies B_k \downarrow A.$$

Hence, $\lim_k P(B_k) = P(A)$ and $P(B_k) \geq P(A_n)$ for every $n \geq k$. Thus, $\sup_{n\geq k} P(A_n) \leq P(B_k) \implies \limsup_k P(A_k) \leq \lim P(B_k) = P(\limsup A_n) = P(A)$. Now, make $C_k = \bigcap_{n\geq k} A_n$, then

$$\lim A_n = \liminf A_n = \lim_{k\to+\infty} \bigcap_{n\geq k} A_n = \bigcup_{k=1}^{+\infty} \bigcap_{n\geq k} A_n = \bigcup_{k=1}^{+\infty} C_k \implies C_k \uparrow A.$$

Hence, $\lim_k P(C_k) = P(A)$ and $P(C_k) \leq P(A_n)$ for every $n \geq k$. Thus, $\inf_{n\geq k} P(A_n) \geq P(C_k) \implies \liminf_k P(A_k) \geq \lim P(C_k) = P(A)$.

Finally, $\liminf P(A_n) \geq P(A) \geq \limsup P(A_n) \implies P(A_n) \to P(A)$.

$\square$

It might seem odd why most books state the properties of continuity from below and from above instead of going straight to continuity (vi). The reason for this is that one can actually show that σ-additivity is actually equivalent to finite-additivity with continuity from either below or above.

Proposition 1.9. Let P be a measure that is finitely additive, i.e. for disjoint sets $A_1, ..., A_n$,

$$P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i).$$

And P is continuous from below (above). This implies that P is σ-additive.

Proof. Let $A = \cup_{i=1}^\infty A_i$ where $A_1, ...$ are disjoint. We want to show that indeed $P(A) = \sum_{i=1}^\infty P(A_i)$. In order to do this, consider the sets $B_n := A \setminus \cup_{i=1}^n A_i$. It's clear that $B_n \downarrow \varnothing$ and $A_i \cap B_i = \varnothing$ for every $i \in \mathbb{N}$. Thus,

$$P(A) = P(\cup_{i=1}^n A_i \cup B_n) = \sum_{i=1}^n P(A_i) + P(B_n)$$
$$\implies P(A) = \lim_{n\to\infty} \sum_{i=1}^n P(A_i) + P(B_n).$$

Since everything is positive, the limit of the sum is the sum of the limits, i.e. $\lim_{n\to\infty}\sum_{i=1}^{n} P(A_i) + P(B_n) = \lim_{n\to\infty}\sum_{i=1}^{n} P(A_i) + \lim_{n\to\infty} P(B_n)$. Since $\lim_{n\to\infty} P(B_n) = 0$, we conclude the proof.

For the case of above continuity, just make $B_n = \cup_{i=1}^{n} A_i$ and hence $P(B_n) = \sum_{i=1}^{n} P(A_i) \uparrow P(A)$.

$\square$

1.2 Random Variables

Definition 1.10 (Random Variable). Let $(\Omega, \mathcal{F}, P)$ be a probability space, and $(E, \mathcal{E})$ be a measurable space (i.e. $\mathcal{B}$ is a σ-algebra of E). A function $X : \Omega \to E$ is called a random variable if X is measurable, i.e.

$$X^{-1}(B) = \{\omega : X(\omega) \in B\} \in \mathcal{F}, \quad \forall B \in \mathcal{E}. \tag{4}$$

Note that the notion of measurability is related to the spaces on which the function is acting.

It's common to define random variables from $(\Omega, \mathcal{F}, P)$ to $(\mathbb{R}, \mathcal{B})$, where $\mathcal{B}$ is the Borel σ-algebra.

Once we have a random variable (which is nothing more than a function), we can construct an image probability measure, which is also sometimes called the random variable's distribution. This is done via the pushforward operation which is formalized below.

Definition 1.11 (Pushforward). Let $(E, \mathcal{F})$ and $(E', \mathcal{G})$ be measurable spaces, $h : E \to E'$ a measurable map and $P \in \mathcal{P}(E)$. We call $h_{\#}P$ the pushforward of h by P, where:

$$h_{\#}P(B) = P(h^{-1}(B)), \quad \forall B \in \mathcal{G}. \tag{5}$$

Theorem 1.12 (Change of Variable). Let $T : E \to E'$ be a measurable map between $(E, \mathcal{F}, P)$ and $(E', \mathcal{G})$. Then, $T_{\#}P$ is a probability measure on $(E', \mathcal{G})$ and $\forall f$ measurable and integrable with respect to $T_{\#}P$ one has:

$$\int_{E'} f dT_{\#}P = \int_{E} f \circ T dP \tag{6}$$

Proof. Let f_n be a simple positive measurable function. Hence

$$f_n(y) = \sum_{i=0}^{N} a_i \mathbb{1}_{A_i}(y) \ \therefore \int_{E'} f_n \ dT_{\#}P = \sum_{i=0}^{N} a_i T_{\#}P(A_i) = \sum_{i=0}^{N} a_i P(T^{-1}(A_i))$$

$$(f_n \circ T)(x) = \sum_{i=0}^{N} a_i \mathbb{1}_{A_i}(T(x)) = \sum_{i=0}^{N} a_i \mathbb{1}_{T^{-1}(A_i)}(x)$$

$$\therefore$$

$$\int_E f_n \circ T \ dP = \sum_{i=0}^{N} a_i P(T^{-1}(A_i))$$

Hence, $\int_E f_n \circ T \ dP = \int_{E'} f_n \ dT_{\#}P$.

Now, for a positive integrable measurable function f, there exists a sequence of positive simple functions such that $f_n \uparrow f$. Then, by the Monotone Convergence Theorem,

$$\int_{E'} f \ dT_{\#}P = \int_{E'} \lim_{n\to+\infty} f_n \ dT_{\#}P = \lim_{n\to+\infty} \int_{E'} f_n \ dT_{\#}P =$$
$$= \lim_{n\to+\infty} \int_E f_n \circ T \ dP = \int_{E'} f \ dT_{\#}P$$

If f is non-positive, just use the same argument by splitting the negative and positive portions of the function.

□

Definition 1.13 (Regular Probability). Let $(E, \mathcal{B}(E), P)$ be a metric space with the Borel σ-algebra and a probability measure P. We say that such probability measure is regular if for every $\varepsilon > 0$ and $A \in \mathcal{B}$, there is a closed set F and an open set G such that $F \subset A \subset G$ and $P(G \setminus F) < \varepsilon$.

Theorem 1.14 (Regular Metric Spaces). Every metric space $(E, \mathcal{B}(E), P)$ is regular, where $\mathcal{B}(E)$ is the Borel σ-algebra for the metric space E. Remember that every metric space has an induced topology, which we'll use to characterize the Borel σ-algebra.

Proof. For a closed set A, make $F = A$. Let

$$G^{\delta} = B(A, \delta) = \{x \in E : d(x, A) < \delta\}.$$

Note that G^δ is open, since $G^\delta = \cup_{x \in A} B(x, \delta)$ is the union of open sets. Since $G^\delta \downarrow A$, then $P(G^\delta) \to P(A)$.

Now, let's prove for the case hat A is not close. Define

$$\mathcal{G} := \{A \in \mathcal{B}(E) : \forall \varepsilon > 0 \exists F \subset A \subset G \ : G \text{ is open and } F \text{ is closed}\}$$

We just proved above that every closed set is in $\mathcal{G}$. Note that $\mathcal{G} \subset \mathcal{B}(E)$, since every element of $\mathcal{G}$ is by definition in $\mathcal{B}(E)$. But, if we can prove that $\mathcal{G}$ is a σ-algebra, then since $\mathcal{G}$ contains all the closed sets, then $\mathcal{B} \subset \mathcal{G}$, which would imply that $\mathcal{B} = \mathcal{G}$ [1].

Let's show that $\mathcal{G}$ is a σ-algebra.

(i) $\Omega \in \mathcal{G}$ since it's closed;

(ii) For $A \in \mathcal{G}$, for every ε we know that $F \subset A \subset G$ and $G^c \subset A^c \subset F^c$ and $P(F^c \setminus G^c) = P(F^c \cap G) = P(G \setminus F) < \varepsilon$, hence, $A^c \in \mathcal{G}$;

(iii) $A_1, A_2 \ldots \in \mathcal{G}$. For every A_n, there is F_n, G_n, such that $P(G_n \setminus F_n) < \frac{2^{-n}\varepsilon}{2}$. Since $\cup_{i=1}^n F_i \uparrow \cup_{i=1}^\infty F_i$, there is $\cup_{i=1}^N F_i$ closed, such that $P(\cup_{i=1}^\infty F_i \setminus \cup_{i=1}^N F_i) < \varepsilon/2$. $\cup_{i=1}^N F_i \subset \cup A_i \subset \cup_{i=1}^\infty G_i = G$. Finally, $P(G \setminus \cup_{i=1}^N F_i) \leq P(G \cup_{i=1}^\infty F_i) + P(\cup_{i=1}^\infty F_i \setminus \cup_{i=1}^N F_i) \leq \sum_{i=1}^\infty P(G_i \setminus F_i) + \varepsilon/2 \leq \sum_{i=1}^\infty \frac{2^{-i}\varepsilon}{2} + \varepsilon/2 = \varepsilon$.

Thus we conclude the proof.

□

Corollary 1.15. Let $(E, \mathcal{B}(E), P)$ and $(E, \mathcal{B}(E), Q)$ be probability spaces such that $P(F) = Q(F), \forall F$ closed. Then, $P(A) = Q(A), \forall A \in \mathcal{B}(E)$, i.e. $P = Q$.

Proof. We can approximate any measurable set A by a closed set F. And since probability measures are continuous, we have that

$$F_n \uparrow A \implies \lim_{n \to \infty} P(F_n) = \lim_{n \to \infty} Q(F_n) = P(A) = Q(A).$$

□

Theorem 1.16. Let P, Q be probability measures defined on $(E, \mathcal{B}(E))$ and $\int f dP = \int f dQ, \forall f : E \to \mathbb{R}$ continuous and bounded (i.e $f \in C_b(E)$). Then, $P = Q$.

[1] This is a useful technique in Measure Theory, where we construct a set implicitly with the property we want to prove.

Proof. Let F be a closed set. Next, define

$$f(x) := \left(1 - \frac{d(x,F)}{\varepsilon}\right)_+ = \max\left(0, 1 - \frac{d(x,F)}{\varepsilon}\right).$$

This function is continuous and bounded by $[0,1]$.

But,$\mathbb{1}_F(x) \leq f(x) \leq \mathbb{1}_{B_\varepsilon(F)}(x)$, thus

$$P(F) \leq \int f dP = \int f dQ \leq Q(B_\varepsilon(F)).$$

Since $B_\varepsilon(F) \downarrow F$, then $\lim_{\varepsilon \to 0} Q(B_\varepsilon(F)) = Q(F)$. We can make the same argument for P instead of Q, thus, we conclude that $P = Q$.

□

1.3 Weak Convergence of Probability Measures

Definition 1.17 (Weak Convergence for Probability). Let $(E, \mathcal{B}(E))$. We say that $P_n \rightharpoonup P$ (i.e. P_n converges weakly to P) if $\forall f \in C_b(E)$ we have

$$\int f dP_n \to_n \int f dP.$$

Definition 1.18 (Continuity Set). $A \in \mathcal{B}(E)$ is called a continuity set of E if $P(\partial A) = 0$ (remember that ∂A is the boundary of A, i.e. $\partial A := \bar{A} \setminus \mathring{A}$).

Theorem 1.19 (Theorem of Portmanteau). Consider $(E, \mathcal{B}(E))$. The following statements are equivalent:

(i) $P_n \rightharpoonup P$;

(ii) $\forall f$ uniformly continuous and bounded, then $\int f dP_n \to \int f dP$;

(iii) $\forall F$ closed, then $\limsup_n P_n(F) \leq P(F)$;

(iv) $\forall G$ open, then $\liminf P_n(G) \geq P(G)$;

(v) $P_n(A) \to P(A), \forall A$ in the continuity set of E.

Proof. [**(i)** $\implies$ **(ii)**] This is clear, since by definition of weak convergence we have that $\int f dP_n \to \int f dP$ for every $f \in C_b$, thus, the same is true if f is C_b and uniformly continuous.

[**(ii)** $\implies$ **(iii)**] Let F be a closed set, and define

$$f(x) := \left(1 - \frac{d(x, F)}{\varepsilon}\right)_+.$$

This function is uniformly continuous and bounded. Hence,

$$\limsup_n P_n(F) \leq \limsup_n \int f dP_n = \int f dP \leq P(B_\varepsilon(F)).$$

Take the limit for $\varepsilon \to 0$, and conclude that $\limsup_n P_n(F) \leq P(F)$.

[**(iii)** $\implies$ **(iv)**] Just use F^c and

$$\liminf_n P_n(F^c) = \liminf_n 1 - P_n(F) = 1 - \limsup_n P(F) \geq 1 - P(F) = P(F^c).$$

[**(iii)** & **(iv)** $\implies$ **(v)**] Note that if A is in the continuity set, then $P(\bar{A}) = P(\mathring{A})$, but using (iii) and (iv)

$$\begin{aligned} P(\bar{A}) \geq \limsup_n P_n(\bar{A}) \geq \limsup P_n(A) &\geq \liminf_n P_n(A) \\ &\geq \liminf_n P_n(\mathring{A}) \\ &\geq P(\mathring{A}) = P(\bar{A}). \end{aligned}$$

[**(v)** $\implies$ **(i)**] We'll use an identity often proved in introductory courses of Probability, but which will only prove later. The identity is for $f \geq 0$, then $E[f] = \int f dP = \int_0^\infty P(f > x) dx$.

First, take a function $f : E \to \mathbb{R}$ that is continuous and bounded. Let's suppose that $f : E \to [0, 1]$. Note that the following argument can be adapted easily for the case that $f : E \to [a, b]$ for any $a, b \in \mathbb{R}$.

We begin by using the fact that:

$$\int f dP = \int_0^1 P(f > x) dx.$$

Note that $\partial\{y \in E : f(y) > x\} \subset \{y \in E : f(y) = x\}$. Also,

$$P(E) = P(\cup_{x\in[0,1]}\{y \in E : f(y) = x\}) = 1.$$

Since $\{y \in E : f(y) = x\}$ are disjoint set (i.e if $y \in \{f = x_1\}$, then $y \notin \{f = x_2\}$ for $x_1 \neq x_2$), then there are at most an enumerable number of x_i such that $P(\{f = x_i\}) > 0$.

Therefore,

$$\int f dP_n = \int_{[0,1]} P_n(f > x)dx = \int_{[0,1]\setminus\{x_1,x_2,...\}} P_n(f > x)dx + \int_{\{x_1,x_2,...\}} P_n(f > x)dx.$$

But the set $\{x_1, x_2, ...\}$ has null Lebesgue measure. Hence,

$$\int f dP_n = \int_{[0,1]\setminus\{x_1,x_2,...\}} P_n(f > x)dx$$

Also, note that $P(\partial\cup_{x\in[0,1]\setminus\{x_1,...\}}\{f > x\}) = 0$. Hence, by (v), we know that if a set is a continuity set, then $P_n(A) \to P(A)$. We can then use the Dominated Convergence Theorem to conclude that

$$\int f dP_n = \int_{[0,1]\setminus\{x_1,x_2,...\}} P_n(f > x)dx \to \int_{[0,1]\setminus\{x_1,x_2,...\}} P(f > x)dx = \int f dP.$$

□

Proposition 1.20. Let $(E, \mathcal{B}(E))$ and $(E', \mathcal{B}(E'))$ be metric spaces and $h : E \to E'$ is a continuous function. Suppose that P_n and P are probability measures defined on E, and $P_n \rightharpoonup P$. Then,

$$h_\# P_n = P_n \circ h^{-1} \rightharpoonup h_\# P = P \circ h^{-1}. \tag{7}$$

Proof. Note that $h_\# P_n$ and $h_\# P$ are probability measures defined on E'. So, to prove that $h_\# P_n$ converges weakly to $h_\# P$, we need to prove that for any $f \in C_b(E')$, the integrals converge. Which is true, since

$$\int f dP_n \circ h^{-1} = \int \underbrace{f \circ h}_{\in C_b(E')} dP_n \to \int f dP.$$

□

Theorem 1.21 (Continuity of Pushforward). Let $(E, \mathcal{B}(E))$ and $(E', \mathcal{B}(E'))$ be metric spaces and $h : E \to E'$ a function such that it's set of points of descontinuity D_h has probability zero (i.e. $P(D_h = 0)$). Then,

$$h_{\#}P_n = P_n \circ h^{-1} \rightharpoonup h_{\#}P = P \circ h^{-1}. \tag{8}$$

Note that this is just a stronger version of Propostion 1.20.

Proof. Take $F \in E'$ a closed set. Then,

$$\begin{aligned}\limsup_n (P_n \circ h^{-1})(F) &= \limsup_n P_n(h^{-1}(F)) \\ &\leq \limsup_n P_n(\overline{h^{-1}(F)}) \leq P(\overline{h^{-1}(F)}),\end{aligned}$$

where we used (iii) of Portmanteau's Theorem 1.19 in the last inequality.

Next, note that $\overline{h^{-1}(F)} \subset D_h \cup h^{-1}(F)$. This is true since, for $x \in \overline{h^{-1}(F)}$, there exists $x_n \in h^{-1}(F)$ such that $x_n \to x$. If $x \in D_h$, then we are done. Now, if $x \notin D_h$, then x is a point of continuity of h, hence $h(x_n) \to h(x)$. But since $x_n \in h^{-1}(F)$, then $h(x_n) \in F$ closed, meaning that $h(x) \in F$, thus $x \in h^{-1}(F)$.

Finally, since $\overline{h^{-1}(F)} \subset D_h \cup h^{-1}(F)$, then

$$P\left(\overline{h^{-1}(F)}\right) \leq P(D_h \cup h^{-1}(F)) \leq \underbrace{P(D_h)}_{=0} + P(h^{-1}(F)) = P(h^{-1}(F)).$$

Thus, for any closed set $F \in E'$ we have $\limsup_n (P_n \circ h^{-1})(F) \leq P(h^{-1}(F))$, which again by Portmanteau 1.19 implies that $P_n \circ h^{-1} \rightharpoonup P \circ h^{-1}$.

□

Lemma 1.22. If P_n and P are such that for every subsequence $\forall n_k$ there exists n_{k_j} such that $P_{n_{k_j}} \rightharpoonup P$, then $P_n \rightharpoonup P$.

Proof. Suppose that P_n does not converges weakly to P. Hence, there exists an $f \in C_b$ and $\varepsilon_0 > 0$ such that for every $N \in \mathbb{N}$ and there exists $n \geq N$ that

$$\left| \int f dP_n - \int f dP \right| \geq \varepsilon_0.$$

But if this is true, then take a subsequence n_k such that for every n_k we have

$$\left| \int f dP_{n_k} - \int f dP \right| \geq \varepsilon_0.$$

Next, take a sub-subsequence n_{k_j}. By the hypothesis, we have that $P_{n_{k_j}} \rightharpoonup P$, which is a contradiction, since

$$\left| \int f dP_{n_{k_j}} - \int f dP \right| \geq \varepsilon_0.$$

□

1.4 Tightness of Probability Families

The main goal of this section is to prove the celebrated Prokhorov's Theorem, which relates the notion of compactness with the notion of tightness. This can be quite useful, since in many situations we can easily see that a family of distributions is tight, but it's not clear that they are (pre)compact.

Definition 1.23 (Tightness). We say that a family $\{P_\lambda\}_{\lambda\in\Lambda}$ of probability distributions in E is *tight* if $\forall \varepsilon > 0$, there exists $K \subset E$ with K compact, such that

$$\inf_{\lambda\in\Lambda} P_\lambda(K) \geq 1 - \varepsilon. \tag{9}$$

Note 1.1 (Pre-Compactness). We say that $A \subset X$ is pre-compact if $\bar{A} = X$. This means that A is dense in X. And we say that A is compact if every open cover of A has a finite sub-cover. In metric spaces, this notion of compactness is equivalent to what we call sequential compactness.

For a metric space (X, d), we say that $A \subset X$ is sequentially compact if every $(x_n) \subset A$ has convergent subsequence $x_{n_k} \to x \in A$. We say that it is sequentially pre-compact if instead $x \in \bar{A}$.

Consider now a family of probability measures $\Pi = \{P_\lambda\}_{\lambda\in\Lambda}$ We say that Π is sequentially pre-compact with respect to weak-convergence if every $(P_n) \in \Pi$ has a subsequence $P_{n_k} \rightharpoonup P \in \overline{\Pi}$.

A last point to remember is the idea of totally boundedness. In $\mathbb{R}$, we know that compactness is equivalent to being closed and bounded. But this is not enough in more general spaces. Instead of boundedness, we need the stronger notion of totally boundedness.

We say that a set A in a metric space (X, d) is totally bounded if for every $r > 0$, there is a collection of finite balls that cover A, i.e. for $x \in A$

there is $B_r(y)$ such that $x \in B_r(y)$. This is another way of saying that there exists an r-net that covers

Theorem 1.24 (Prokhorov). Prokhorov's theorem contains two results that are almost one the converse of the other. The first result is the main (and most useful) one:

(i) If a family $\Pi = \{P_\lambda\}_{\lambda \in \Lambda}$ is tight, then Π is sequentially pre-compact with respect to the weak convergence.

(ii) Also, if E is a Polish space (i.e. separable and complete metric space) and Π is sequentially pre-compact with respect to weak convergence, then Π is tight.[2]

Proof. For the proof, we'll start proving (ii) and then (i).

(ii) This is the easier one in the theorem. Note that if E is Polish, then there exists a countable dense set that covers E. Also, we can fix a diameter $\frac{1}{k}$, and define a family of open subsets with $\operatorname{diam}(A_n^k) \leq 1/k$, such that $\cup_{i=1}^n A_i^k \uparrow E$.

Let's start proving that since $\cup_{i=1}^n A_i^k = B_n^k$ grows to E and our family Π is pre-compact, then for any $\varepsilon > 0$, there exists n big enough such that $\inf_{\lambda \in \Lambda} P_\lambda(B_n^k) \geq 1 - \varepsilon$. If this was not true, then there would exists $\varepsilon_0 > 0$, such that $\inf P_\lambda(B_n^k) < 1 - \varepsilon$ for every n, in other words, there would exist a sequence $P_n(B_n^k) \leq 1 - \varepsilon$.

Since the family is pre-compact, there exists a convergent subsequence $P_{n_j} \rightharpoonup Q$. But by Portmanteau 1.19,

$$Q(B_{n_j}^k) \leq \liminf_j P_{n_j}(B_{n_j}^k) \leq 1 - \varepsilon_0,$$

which is a contradiction, since $B_{n_j}^k \uparrow E \implies Q(B_{n_j}^k) \uparrow 1$.

Note that the proof above works for every $k \in \mathbb{N}$. So we can use a diagonal argument. Since $\forall k, B_n^k \uparrow E$, then, for each k there exists a n_k such that

$$\inf_{\lambda \in \Lambda} P_\lambda(B_{n_k}^k) \geq 1 - \frac{\varepsilon}{2^k} \geq 1 - \varepsilon.$$

[2]Note that in Prokhorov, the second statement is not the converse of the first, since it needs stronger hypothesis.

Now, define $B := \cap_{k\in\mathbb{N}} B^k_{n_k}$. Note that B is pre-compact, since for any $r > 0$, there exists a finite r-net, just note that for $k \in \mathbb{N} : 1/k < r$, the set $B^k_{n_k} = \cup_{i=1}^{n_k} A^k_i$ is covered by a finite union of balls with diameter r. This implies that B is totally limited, thus $\overline{B}$ is closed and totally limited, which implies that $\overline{B}$ is compact and B is pre-compact.

Finally, for $B^c = \cup_{k\in\mathbb{N}} (B^k_{n_k})^c$. We have

$$P_\lambda(B^k_{n_k}) \geq 1 - \frac{\varepsilon}{2^k} \iff 1 - P_\lambda(B^k_{n_k}) \leq \frac{\varepsilon}{2^k} \iff P_\lambda((B^k_{n_k})^c) \leq \frac{\varepsilon}{2^k}.$$

Therefore,

$$P_\lambda(B^c) \leq \sum_{k=1}^{\infty} P_\lambda((B^k_{n_k})^c) \leq \sum_{k=1}^{\infty} \frac{\varepsilon}{2^k} = \varepsilon \implies P_\lambda(B) \geq 1 - \varepsilon.$$

Since this is true for any $P_\lambda \in \Pi$, we conclude that

$$\inf_{\lambda\in\Lambda} P_\lambda(B) \geq 1 - \varepsilon.$$

(i) To prove that a tight family $\{P_\lambda\}$ is sequentially pre-compact, we need to show that each sequence $(P_n) \subset \{P_\lambda\}$ there is a subsequence that converges weakly to P. The hardest aspect of the proof is to find a candidate for the limiting measure P.

Without loss of generality, we can assume that E is separable. The reason for this is that since $\{P_\lambda\}$ is tight, then for $i \in \mathbb{N}$ we have a sequence of compact sets K_i such that

$$\inf_\lambda P_\lambda(K_i) \geq 1 - \frac{1}{i} \implies \inf_\lambda P_\lambda(\cup_{i\in\mathbb{N}} K_i) \geq \inf_\lambda P_\lambda(\sup_i K_i) \geq 1.$$

Knowing that this enumerable sequence covers all the space in measure, we can cover each K_i and obtain a finite sub-cover of open sets, and thus, construct a sequence of enumerable open sets that covers E.

The rest of this proof is from Chin [1].

$\square$

The following proof is from Teixeira and Lacoin [5].

Theorem 1.25 (Kolmogorov's Extension Theorem - Simpler Version). For $n \in \mathbb{N}$, let $P_n \in \mathcal{P}(\mathbb{R}^n)$ such that

$$P_{n+1}(A \times \mathbb{R}) = P_n(A), \quad \forall\, A \in \mathcal{B}(\mathbb{R}^n).$$

Therefore, there exists a unique probability measure P in $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$ such that $P(A \times \mathbb{R} \times ...) = P_n(A)$ for every n and every $A \in \mathbb{R}^n$.

2 Probability in $C[0, 1]$

Definition 2.1 (Cylinder).

$$\mathcal{A} := \{X\}$$

Note 2.1 (Ascoli-Arzela). $D \subset C[0, 1]$ is compact if and only if

(i) $\exists a \in \mathbb{R}$ such that $|x_0| < a \forall x \in D$

(ii) D is equicontinuous, i.e. $\forall \varepsilon > 0 \exists \delta > 0$ such that $|t - s| < \delta \implies |x_t - x_s| < \varepsilon$.

Definition 2.2 (Modulus of Continuity). The modulus of continuity $\omega_x(\delta)$ is

$$\sup_{s,t \in [0,1], |s-t| < \delta} |x_t - x_s|.$$

Note that D is equicontinuous $\iff \lim_{\delta \to 0} \sup_{x \in D} \omega_x(\delta) = 0$.

3 Gaussians

This section has some results from Le Gall [2]. It consists of a collection of results related to Gaussian Random Variables.

3.1 1D Gaussian

Definition 3.1 (Standard Gaussian). A probability measure $\mu \in \mathcal{P}(\mathbb{R})$ is said to be a *standard Gaussian* (or standard normal) if it's density with respect to Lebesgue measure is

$$p(x) = \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} \tag{10}$$

To prove that this indeed defines a probability distribution, note that

$$\left(\int_{-\infty}^{\infty} e^{-x^2/2} dx \right)^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)/2} dx dy = \int_0^{2\pi} \int_0^{\infty} e^{-(r^2)/2} r d\theta dr d\theta = \sqrt{(2\pi)}.$$

Definition 3.2 (Gaussian). A probability measure $\mu \in \mathcal{P}(\mathbb{R})$ is said to be a *Gaussian* if it's density with respect to Lebesgue measure is

$$p(x) = \frac{e^{\frac{-(x-a)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} \tag{11}$$

3.2 Vector of Gaussians

3.3 Gaussian Process

4 Expected Value

Lemma 4.1 (Layered Cake Representation). Note that for a r.v X on $(\Omega, \mathcal{F}, P)$, then

$$X(\omega) = \int_0^{X(\omega)} 1 dx = \int_{\mathbb{R}} \mathbb{1}_{[0,X(\omega)]}(x) dx. \tag{12}$$

Theorem 4.2 (Tail Probability Expectation Identity). Let X be a random variable with finite expectation. Then

$$EX = \int_0^{\infty} P(X > x) dx - \int_{-\infty}^{0} P(X < x) dx. \tag{13}$$

If X is a positive discrete r.v, then

$$EX = \sum_{k=0}^{\infty} P(X > k) = \sum_{k=1}^{\infty} P(X \geq k). \tag{14}$$

Proof. Let's start by proving for the case where $X \geq 0$.

$$\mathbb{1}_{[0,X(\omega)]}(x) = \mathbb{1}_{\{X(\omega)>x\}\cap\{X(\omega)\geq 0\}}(\omega)$$

$$\begin{aligned} EX &= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_{\{X(\omega)>x\}\cap\{X(\omega)\geq 0\}}(\omega) dx \; dP(\omega) \\ \text{(Fubini)} &= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_{\{X(\omega)>x\}\cap\{X(\omega)\geq 0\}}(\omega) dP(\omega) \; dx \\ &= \int_{\mathbb{R}} P(\{X(\omega) > x\} \cap \{X(\omega) \geq 0\}) dt \end{aligned}$$

Note that $P(\{X(\omega) > x\} \cap \{X(\omega) \geq 0\}) = 0$ if $x \leq 0$. Hence,

$$EX = \int_0^{\infty} P(X > x) dx.$$

For the general case where X is not necessarily positive, repeat the same proof for $X^+ = \max 0, X$ and $X^- = -\min 0, X$. Thus, obtaining the desired result.

Finally, for the case of positive discrete random variable, just note that

$$\int_0^\infty P(X > x)dx = \int_0^1 P(X > x)dx + \int_1^2 P(X > x)dx + ... \int_n^{n+1} P(X > x)dx + ...$$
$$= \sum_{k=0}^{\infty} P(X > k).$$

□

Corollary 4.3 (Tail Probability Expectation Identity for Functions). Let X be a *positive* r.v and $f : \mathbb{R}_+ \to \mathbb{R}$ such that $f \in C^1$. Then

$$E[f(X)] = f(0) + \int_0^\infty f'(t)P(X \geq t)dt \tag{15}$$

Proof.

$$Ef(X) = \int_0^\infty f(x)dP_X, \quad f(x) - f(0) = \int_0^x f'(t)dt$$

$$\begin{aligned}
\int_0^\infty f(x)dP_X &= \int_0^\infty f(0) + \left(\int_0^x f(x)dt\right) dP_X \\
&= f(0) + \int_0^\infty \int_0^x f(x)dtdP_X \\
&= f(0) + \int_{\mathbb{R}} \mathbb{1}_{(0,\infty)}(x) \int_{\mathbb{R}} \mathbb{1}_{(0,x)}(t)dt\ dP_X
\end{aligned}$$

Note that $\mathbb{1}_{(0,\infty)}(x) = \mathbb{1}_{(t,\infty)}(x)$, hence

$$\begin{aligned}
Ef(X) &= f(0) + \int_{\mathbb{R}} \mathbb{1}_{(0,\infty)}(x) \int_{\mathbb{R}} \mathbb{1}_{(t,\infty)}(x)dt\ dP_X \\
\text{(Fubini)} &= f(0) + \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_{(0,\infty)}(x)\mathbb{1}_{(t,\infty)}(x)dP_X\ dt \\
&= f(0) + \int_{\mathbb{R}} P_X(\{x \in (0,\infty)\} \cap \{x \in (t,\infty)\})dt \\
&= f(0) + \int_{\mathbb{R}} P(X > t)dt.
\end{aligned}$$

□

5 Extra

This section is a collection of interesting results that are not conventional. Some of them might be specially useful for Optimal Transport.

5.1 Probabilities and Lipschitz Functions

Here we point out the interesting relation between probability measures with finite first moment and Lipschitz functions. This is based on Perrone [3].

Theorem 5.1. Let (X, d) be a metric space

References

[1] Wooyoung Chin. New simple proofs of the kolmogorov extension theorem and prokhorov's theorem. *arXiv preprint arXiv:1911.12979*, 2019.

[2] Jean-François Le Gall. *Brownian motion, martingales, and stochastic calculus*. Springer, 2016.

[3] Paolo Perrone. Categorical probability and stochastic dominance in metric spaces, 2018.

[4] Alexander Sokol and Anders Rønn-Nielsen. Advanced probability. *University of Copenhagen: Department of Mathematical Sciences*, 2013.

[5] Augusto Teixeira and Hubert Lacoin. *Notas de aula: Probabilidade I (Remix)*. 2018.