Centro de Investigación en Matemáticas

Weak convergence in measure

T H E S I S

To obtain the degree of

Master of Science

P R E S E N T S:

Guillermo Elías Martínez Dibene

Thesis advisor:

Ph.D. Martin Barlow

University of British Columbia

GUANAJUATO, GTO AUGUST 2015

✠ Acknowledgements.

I would like to thank the Mexican National Council for Science and Technology (CONACYT) for the scholarship (with registration number 370909), which I received during my master. In the same matter, I thank the Mathematics Research Center (CIMAT) for the several financial supports given during my studies there.

Also, I thank Martin Barlow for the financial support he gave me to make an academic stay in the University of British Columbia (UBC) during the Winter term of 2014. This thesis is written with the research made in this stay at UBC.

Finally, it is not possible to mention everyone who have been with me during the last years but I will mention my father, José Guillermo Eugenio Martínez Arias, who has alway been there for me and supported me.

To everyone of you, I appreciate your company in the good and the bad times.

Sincerely,
Guillermo Elías Martínez Dibene

✠ Introduction.

The present thesis has as objective to be a brief exposition of the topic “weak convergence in measure”. It has been a while since the very first articles, those in which weak convergence in measure first appeared, were published. Recently Barlow, Burzdy and Timar have proved some type of Central Limit theorems in which the convergence is of this type, “weakly in measure”.

To understand the relevance of this research, it is necessary to give a brief historical note about this type of convergence. First of all, in 1986, Kipnis and Varadham published the article [5]. There, they proved a Central Limit type theorem for additive functionals of stationary reversible ergodic Markov chains. Their proof let them conclude that some random measures actually converge in measure (with respect to the measure of the space on where they were defined). So, they did not give a name to this type of convergence and they simply mentioned it as a remark (see Remark 1.10 of [5]). Even when they informally gave the definition of weak convergence in measure, they did not explicitly mention it and never investigated its properties. A few years later, in 1989, De Masi et al., in [1] explicitly gave the definition of weak convergence in measure. As a matter of fact, they stated the following as their definition

> “In applications, the state ξ of our reversible Markov process will represent the environment seen from a ‘tagged’ particle. Since we wish to investigate asymptotic behavior for a fixed initial environment, as well as the behavior arising from averaging (with respect to μ) over the initial envirionment, we employ the following notion of convergence. Let $X^\varepsilon = (X^\varepsilon_t)_{t\geq 0}$, $\varepsilon > 0$, be a family of $\mathbb{R}^d$-valued processes on (Ω, P_μ). Then we say that X^ε converges *weakly in μ-measure* (or probability) to the $\mathbb{R}^d$-valued process Y, and we write $X^\varepsilon \to Y$, if for all bounded, continuous functions F on $D \equiv D([0,\infty), \mathbb{R}^d)$ (equipped with the Skorohod topology)
>
> $$E_\mu[F(X^\varepsilon)|\xi_0 = \xi) \to E(F(Y))$$
>
> as $\varepsilon \to 0$ in μ-probability. Note that $X^\varepsilon \to Y$ in μ-probability implies that X^ε tends to Y in distribution.”

Again, they proved a Central Limit type theorem with this convergence but, like Kipnis and Varadham, they did not investigate properties of this convergence. Finally, in a paper of Barlow, Burzdy and Timar, to appear [3], there is defined “weak convergence in measure” in a more general context (moreover, this definition is the same as the one given here (2.1.2)), and also prove a Central Limit type theorem but do not research what properties hold in this case.

After this historical development, it seemed necessary to investigate which of the theorems that hold for the already well known topic of weak convergence hold equally well for weak convergence in measure, that is, which theorems could be generalized in a natural way? In fact, as Prohorov’s theorem seems to be one of the most important theorem for weak convergence, the question that arose was: does Prohorov’s theorem hold or have a natural analogue in the context of weak convergence in measure? The whole point of this thesis is to answer this question.

In Chapter 1, the definition of “classical” weak convergence will be given and the details will be developed. It has to be mentioned here that, the familiar reader with the topic can skip this chapter without losing the thread of the second chapter. Now, the definition of weak convergence will be given so that it will coincide with the same as the standard definition of the “vague topology” (same as “weak topology”) used in functional analysis. It seems convenient to start from an analytical point of view: in one hand, Billingsley have already developed the topic starting from a probability point of view (see [4]) but at the cost of not giving the connection between the two topics; secondly, this view will include the topic as it is, in other words, there is a metric space in which is to further study of the convergence and, of course, study its “nice” subsets, namely, those which are relatively compact. This chapter is thus written from this point of view. It is assumed that the reader is familiar with general analisys and general topology (the books of the bibliography are recommended for references, see [6] and [8]) but not with advanced probability (the reference for probabily topics will be [2]).

In Chapter 2, the development of the theory of weak convergence in measure is given. All definitions are being proposed and in this point it is necessary that the reader be familiar with Lebesgue’s integration. Firstly, some questions about mesurability are answered. Then, one proceeds to propose the definition of weak convergence in measure and will always be stated that *with arbitrarily large probability, the is a uniformity that happens with high chance*; this will be clearer once starting the reading (see the remarks of this chapter). After that, the steps are to prove Prohorov’s theorem in measure and in order to do that, definition of tightness in measure (2.1.2) is given and a analogous of portmanteau theorem is proved (see (1.5.3) and (2.3.1)). Finally, one half (the “direct half”) of Prohorov’s theorem for this context is stated and proved and a construction (with a construction of multiple explicit counterexamples) is given to prove *falssity* of the sufficiency of this theorem.

✠ Some notation employed in the thesis.

During the thesis all notations will be consistent inside this writing. Also, I will explicitly used the italic letters to emphasize an important point in a sentence and I will reserve the quotation marks (the symbols “ and ”) when I write a definition, given by a word or by a sentence. Something else must be mention also before I start giving the notation; I have followed several advices given in the book *How to write Mathematics* by Steenrod et al. To understand the cross referencing inside this text the following convention was taken. Every property (Definition, Lemma, Remark, Theorem) stated here will have asociated an *unique* number of the form (k.n.m); here, **k** denotes the number of the chapter, **n** the number of the section and **m** will denote the number of the property in that section. To make a reference to such a property, *only* its unique number will be given.

Each of the following sets are subsets of $\mathbb{R}^S$, the vector space of functions $S \to \mathbb{R}$, and the specific property is expressed (it turns out that all of these spaces are subspaces of $\mathbb{R}^S$):

1. $\mathscr{C}_{\mathbb{R}}(S)$ the continuous elements;
2. $\mathscr{C}_{\mathbb{R}}^{\infty}(S)$ the bounded continuous elements;
3. $\mathscr{C}_{\mathbb{R}}^{0}(S)$ the continuous elements that vanish at ∞ (this is, for every $\varepsilon > 0$ there is a compact set $K \subset S$ such that $\|f(x)\| \leq \varepsilon$ for every point x in the *complement* of K);
4. $\mathscr{K}_{\mathbb{R}}(S; K)$ the continuous elements with compact support contained in K elements;
5. $\mathscr{B}_{\mathbb{R}}(S)$ the bounded elements;
6. $\mathscr{M}_{\mathbb{R}}(S)$ the measurable elements;
7. $\mathscr{M}_{\mathbb{R}}^{\infty}(S)$ the bounded measurable elements;
8. $\mathscr{L}_{\mathbb{R}}^{p}(S, \mu)$ the p-integrable elements (with respect to μ).

Remark. In this thesis the space of k times differentiable functions with continuity (also called as continuously differentiable functions) will never be used or mentioned. It is then clear that it will cause no confusion the previous notations.

Also, the following general notation is used:

9. $\mathrm{B}_S(x; r)$ is the ball in S with centre x and radius $r > 0$; when S is clear from the context, it will be written simply as $\mathrm{B}(x; r)$.
10. ∂A denotes the frontier of the set A;
11. $\overline{A}$ denotes the closure of the set A;
12. $\mathfrak{P}(S)$ denotes the power set of S, this is, the set of all subsets of S;
13. $\mathcal{B}_S$ denotes the Borel sets of S;

14. $\mathcal{C}_S$ denotes the closed sets of S;

15. $\mathcal{O}_S$ denotes the topology of S;

16. If x belongs to a normed vector space, $\|x\|$ denotes the norm (the same symbol for all norms that will be considered).

17. For a given set $A \subset S$, the function $\mathbb{1}_A : S \to \{0,1\}$ given by $\mathbb{1}_A(x) = 1$ if $x \in A$ and $\mathbb{1}_A(x) = 0$ if $x \notin A$ is the "indicator function" of the set A (or as is called in analysis, the "characteristic function" of the set A; this term will not be used in this thesis).

It is convenient to bear in mind the following conventions:

18. $\mathcal{M}_{\mathbb{R}}(S)$ is the set of all real-valued signed measures defined on $\mathcal{B}_S$;

19. $\mathcal{P}_S$ is the set of all probability measures defined on $\mathcal{B}_S$.

Also, accordingly with the previous notations, the following alphabet is reserved for sets of subsets of a given set S (in other word, subsets of $\mathfrak{P}(S)$ but with no special structure):

$$\mathfrak{A},\ \mathfrak{B},\ \mathfrak{C},\ \mathfrak{D},\ \mathfrak{E},\ \mathfrak{F},\ \mathfrak{G},\ \mathfrak{H},\ \mathfrak{I},\ \mathfrak{J},\ \mathfrak{K},\ \mathfrak{L},\ \mathfrak{M},\ \mathfrak{N},\ \mathfrak{O},\ \mathfrak{P},\ \mathfrak{Q},\ \mathfrak{R},\ \mathfrak{S},\ \mathfrak{T},\ \mathfrak{U},\ \mathfrak{V},\ \mathfrak{W},\ \mathfrak{X},\ \mathfrak{Y}, \mathfrak{Z}.$$

In the same way, the following alphabet is often used for sets whose elements are functions $S \to \mathbb{R}$ (in other words, subsets of $\mathbb{R}^S$):

$$\mathscr{A},\ \mathscr{B},\ \mathscr{C},\ \mathscr{D},\ \mathscr{E},\ \mathscr{F},\ \mathscr{G},\ \mathscr{H},\ \mathscr{I},\ \mathscr{J},\ \mathscr{K},\ \mathscr{L},\ \mathscr{M},\ \mathscr{N},\ \mathscr{O},\ \mathscr{P},\ \mathscr{Q},\ \mathscr{R},\ \mathscr{S},\ \mathscr{T},\ \mathscr{U},\ \mathscr{V},\ \mathscr{W},\ \mathscr{X},\ \mathscr{Y}, \mathscr{Z}.$$

It has to be said, however, that Ω will denote a "generic sample space" with its "generic σ-algebra" $\mathscr{F}$; μ and ν "generic measures", $\mathbb{P}$ a "generic probability measure". Finally, the next alphabet will mostly be used to designate subset of "signed measures", "special laws" or subsets of $\mathcal{B}_S$ (such as in the case of $\mathcal{O}_S$ and $\mathcal{C}_S$):

$$\mathcal{A},\ \mathcal{B},\ \mathcal{C},\ \mathcal{D},\ \mathcal{E},\ \mathcal{F},\ \mathcal{G},\ \mathcal{H},\ \mathcal{I},\ \mathcal{J},\ \mathcal{K},\ \mathcal{L},\ \mathcal{M},\ \mathcal{N},\ \mathcal{O},\ \mathcal{P},\ \mathcal{Q},\ \mathcal{R},\ \mathcal{S},\ \mathcal{T},\ \mathcal{U},\ \mathcal{V},\ \mathcal{W},\ \mathcal{X},\ \mathcal{Y}, \mathcal{Z}.$$

In chapter 2, a probability space will be given $(\Omega, \mathscr{F}, \mathbb{P})$ and two metric spaces will be on consideration, namely (S, d) and $(\mathcal{P}_S, \rho)$. The subsets of S will be denoted with no special emphasis, in other hand, the letter to specific subsets of $\mathcal{P}_S$ are already stated (since these subset are sets of probability measures). Hence, to avoid possible misunderstandings with the alphabets, this last will be used to denote event (sets in $\mathscr{F}$) of Ω :

$$\mathsf{A},\ \mathsf{B},\ \mathsf{C},\ \mathsf{D},\ \mathsf{E},\ \mathsf{F},\ \mathsf{G},\ \mathsf{H},\ \mathsf{I},\ \mathsf{J},\ \mathsf{K},\ \mathsf{L},\ \mathsf{M},\ \mathsf{N},\ \mathsf{O},\ \mathsf{P},\ \mathsf{Q},\ \mathsf{R},\ \mathsf{S},\ \mathsf{T},\ \mathsf{U},\ \mathsf{V},\ \mathsf{W},\ \mathsf{X},\ \mathsf{Y}, \mathsf{Z}.$$

Chapter 1

✠ Weak convergence of probability measures.

As a brief introduction, consider the following situation. Suppose that a family $(S_\lambda, \tau_\lambda)_{\lambda \in L}$ of topological spaces is given (that is, S_λ is assumed to be a nonempty set and τ_λ a topological structure over S_λ). Define $S = \prod_{\lambda \in L} S_\lambda$ the Cartesian product of the sets S_λ and τ to be the product topology on S, that is, τ is the finest topology on S which makes all projections $\mathrm{pr}_\mu : (s_\lambda)_{\lambda \in L} \mapsto s_\mu$ continuous from S to S_μ. If it happens that all of the S_λ coincide with a *topological vector space* V, then S is a vector space, which is denoted by V^L, and the projection are *continuous, linear* and *open* functions. In this case, in order for a sequence (f_n), defined on the space V^L, to converge to a point $f \in V^L$, it is necessary and sufficient that the sequence $(\mathrm{pr}_\mu(f_n) = f_n(\mu))$, defined on V, converges, in V, to the point $f(\mu)$. As this is true, it is usual to call such a product topology the *topology of simple convergence* or the *topology of pointwise convergence.* A special case is when V is the *field* of *real numbers* $\mathbb{R}$ or the *complex numbers* $\mathbb{C}$ and L is also a *linear space.* Here, it is obtained $\mathbb{R}^L$ or $\mathbb{C}^L$. In the subspace of *linear forms* on $\mathbb{R}^L$ or $\mathbb{C}^L$, the topology of simple convergence is called the *weak topology*; it is readily seen that this topology is defined by the *seminorms*

$$x \mapsto |f(x)|, \quad x \in L.$$

As was mentioned earlier, in this topology, a sequence of linear forms (f_n) will converge to a linear form f if $f_n(x)$ converges to $f(x)$ in $\mathbb{R}$ or $\mathbb{C}$. Finally, consider a metric space (E, d) and the linear space $\mathscr{C}^\infty_{\mathbb{R}}(E)$, of functions defined on E with values in $\mathbb{R}$ which are *continuous* and *bounded*; a linear form $\mu \in \mathbb{R}^{\mathscr{C}^\infty_{\mathbb{R}}(E)}$ for which each restriction $\mu\Big|_{\mathscr{K}_{\mathbb{R}}(E;K)}$ is continuous, where $K \subset E$ a compact set, is a measure on E (this is known as a version of "Daniell's theorem"); let $\mathcal{M}_{\mathbb{R}}(E)$ to be the subspace of $\mathbb{R}^{\mathscr{C}^\infty_{\mathbb{R}}(E)}$ whose elements are measures. In this space the weak topology is called *vague topology* (in french and spanish literature, while in english literature the term "weak topology" is usually fixed); saying that a sequence of measures (μ_n) converges to the measure μ is equivalent to saying that for each $f \in \mathscr{C}^\infty_{\mathbb{R}}(E)$, the sequence of real numbers $\mu_n(f)$ will converge to the real number $\mu(f)$; this last notion is what is called μ_n *converges vaguely to the measure* μ (again, in the english literature, this is not used and they use the teminology μ_n *converges weakly to the measure* μ). In this chapter we consider the special case in which we take all the measures μ_n to be *unitary*, or equivalently *probability measures*, and the interest lies in further studying the subset (which is not a subspace) of $\mathcal{M}_{\mathbb{R}}(E)$ of probability measures; it will be shown that this set is metrizable (whereas, in general, $\mathcal{M}_{\mathbb{R}}(E)$ is not).

§1.1. The Prohorov distance.

During this chapter, the next convention will be taken. Always S will denote a metric space (sometimes it will happen that the theorem to be proved here will still be valid when S is assumed to be *only* a topological space); d the distance over S; $\mathcal{B}_S$ will denote the smallest σ-algebra that contains the topology of S (the "Borel σ-algebra") and its elements will be called "Borel sets" of S; note that inside $\mathcal{B}_S$ is the set $\mathcal{C}_S$ of the closed subsets of S and $\mathcal{O}_S$ of the open ones; $\mathcal{P}_S$ will denote the set of all probability measures[1] defined on the σ-algebra $\mathcal{B}_S$. For a given set F of S, and a positive number $\varepsilon > 0$, the symbol F^ε will denote de set of points x in S whose *distance from* F will be smaller[2] than ε, where the distance from F is defined to be the non-negative (and necessarily finite) number

$$d(x,F) = \inf_{y \in F} d(x,y).$$

Theorem (1.1.1) The real function $\rho_S = \rho$ defined on the set $\mathcal{P}_S \times \mathcal{P}_S$ given by the formula

$$\rho_S(P,Q) = \inf \left\{ \varepsilon > 0 \middle| \forall F \in \mathcal{C}_S,\ P(F) \leq Q\left(F^\varepsilon\right) + \varepsilon \right\}$$

is a metric on $\mathcal{P}_S$; the "Prohorov distance" and the space $(\mathcal{P}_S, \rho)$ will be refered as the "Prohorov space" of S.

Proof: it is evident that ρ is non-negative; the other properties must be proven. To prove that $\rho(P,Q) = \rho(Q,P)$; it will be first proven that, for α and β two positive numbers, the relation

$$\forall F \in \mathcal{C}_S,\ P(F) \leq Q\left(F^\alpha\right) + \beta$$

implies the relation

$$\forall F \in \mathcal{C}_S,\ Q(F) \leq P\left(F^\alpha\right) + \beta;$$

for if this is proved, it will follows immediately that $\rho(P,Q) = \rho(Q,P)$. Now, assume the first relation holds and take $F_1 \in \mathcal{C}_S$; let F_2 be the set $\complement_S F_1^\alpha$, where $\complement_S A$ is the set "complement of A with respect to S" defined by $S - A$. As F_1^α is open[3], F_2 is closed, thus, it belongs to $\mathcal{C}_S$. As the relation $A \subset B$ is equivalent to the relation $\complement_S B \subset \complement_S A$, it follows that $F_1 \subset \complement_S F_2^\alpha$. Then,

$$P\left(F_1^\alpha\right) = 1 - P(F_2) \geq 1 - Q\left(F_2^\alpha\right) - \beta \geq Q(F_1) - \beta,$$

and the second relation is proved.

Assume now that $\rho(P,Q) = 0$. Then, if $F \in \mathcal{C}_S$ is arbitrary, $P(F) \leq Q\left(F^\varepsilon\right) + \varepsilon$ for all $\varepsilon > 0$; hence, taking the infimum over the set of $\varepsilon > 0$, it is possible to conclude that[4]

$$\forall F \in \mathcal{C}_S,\ P(F) \leq Q(F).$$

As d is symmetric, it can be concluded that $P(F) = Q(F)$ for all the closed sets in S and, taking complements, for all of the open sets in S. By hypothesis, P and Q are defined on $\mathcal{B}_S$ and therefore, they must be equal since they concide on the generators of $\mathcal{B}_S$.

Now, let $P, Q, R \in \mathcal{P}_S$ and suppose $\rho(P,Q) < \delta$ and $\rho(Q,R) < \varepsilon$. Then,

$$P(F) \leq Q\left(F^\delta\right) + \delta \leq R\left(\left(\overline{F^\delta}\right)^\varepsilon\right) + +\varepsilon + \delta \leq R\left(F^{\delta+\varepsilon}\right) + \delta + \varepsilon$$

---

[1] Throughout this text, "measure" will always mean a "positive measure"; however, a "charge" or "signed measure" will mean the general concept of measure.

[2] The terms "smaller" and "bigger" will always mean $<$ and $>$, respectively; for the symbols $\leq$ and $\geq$, it shall be used the phrases "at most" and "at least". Similarly, "positive" will mean > 0 and "negative" will mean < 0; the term "non-negative" is reserved to mean ≥ 0.

[3] In fact, if F is an *arbitrary* set of S, then F^α (called as "augmentation" by $\alpha > 0$ of F) is open: indeed, for if $x \in F^\alpha$, then there is some $y \in F$ such that $d(x,y) < \alpha$; the ball $\mathrm{B}\left(x; \alpha - d(x,y)\right)$ is contained in F^α as follows directly from use once the triangle inequality. In what follows, it will be used that such sets F^α are open without further notice.

[4] This follows since F is closed: for in this case, $F = \bigcap_{\varepsilon > 0} F^\varepsilon = \bigcap_{n \in \mathbb{N}} F^{\frac{1}{n}}$ (this last property shows that *every* closed set in a metric space is "G_δ" -the letter G is for the German word *gebiet*, which mean "area"; the δ is supposed to mean "intersection" from the German *durchschnitt*- that is, every closed set is the intersection of a sequence of open sets) and the fundamental property of monotone limit for probability measures.

because if $d\left(x, \overline{F^\delta}\right) < \varepsilon$, then for some $y \in \overline{F^\delta}$ it is true that $d(x,y) < \varepsilon$, and for such y there exists a sequence (y_n) defined on F^δ such that (y_n) converges to y, therefore for sufficiently large n it will happen that $d(x, y_n) < \varepsilon$ and, with this, the last inequality follows[5]. Finally, note that by definition of ρ,

$$\rho(P,R) \leq \delta + \varepsilon,$$

hence $\rho(P,R) - \delta$ is a lower bound for the set of $\varepsilon > 0$ such that if $F \in \mathcal{C}_S$ then $Q(F) < R(F^\varepsilon) + \varepsilon$, and therefore it cannot be bigger than the infimum, namely $\rho(Q,R)$; that is,

$$\rho(P,R) \leq \delta + \rho(Q,R);$$

the same argument shows that

$$\rho(P,R) \leq \rho(P,Q) + \rho(Q,R),$$

which is the triangle inequality. ■

§1.2. Some technical lemmas.

In this section, several technical lemmas will be proved with the aid of giving a proof for Skorohod's representation theorem and Prohorov's theorem.

Lemma (1.2.1) Assume that μ is a finite measure over the σ-algebra $\mathcal{B}_S$, that $(p_i)_{i=1,\ldots,n}$ is a family of non-negative real numbers, and that $(A_i)_{i=1,\ldots,n}$ is a finite family of Borels of S. If

$$\sum_{i\in I} p_i \leq \mu\left(\bigcup_{i\in I} A_i\right)$$

whatever the subset $I \subset \{1,\ldots,n\}$ is, then there exist finite measures $\lambda_1,\ldots,\lambda_n$ on $\mathcal{B}_S$ such that $\lambda_i(A_i) = \lambda_i(S) = p_i$ $(i = 1,\ldots,n)$ and $\sum_{i=1}^n \lambda_i(A) \leq \mu(A)$ for every $A \in \mathcal{B}_S$.

Proof: if some $p_i = 0$, let $\lambda_i = 0$; then, it may be supposed, without loss of generality, that each $p_i > 0$. First note that for each i, $\mu(A_i) \geq p_i$, otherwise, the hypothesis would be violated when $I = \{i\}$, therefore $\mu(A_i) > 0$. Now, it will be proceed inductively. For the case $n = 1$, let $\lambda_1 : \mathcal{B}_S \to \mathbb{R}_+$ be defined by $\lambda_1(A) = p_1 \dfrac{\mu(A\cap A_1)}{\mu(A_1)}$. It is then clear that λ_1 satisfies the conclusion and, hence, the case $n = 1$ is proved. Now, suppose that (1.2.1) has already been proved for all cases $1,\ldots,n-1$ and that μ, $(p_i)_{i=1,\ldots,n}$ and $(A_i)_{i=1,\ldots,n}$ satisfies the hypothesis. Define the measure η on $\mathcal{B}_S$ by

$$\eta(A) = \frac{\mu(A\cap A_n)}{\mu(A_n)}$$

and let ε_0 be the largest real number $\varepsilon \geq 0$ such that

$$\sum_{i\in I} p_i \leq (\mu - \varepsilon\eta)\left(\bigcup_{i\in I} A_i\right),$$

for all subsets I of $\{1,\ldots,n-1\}$ (note that ε_0 may be 0). To continue, first suppose that $\varepsilon_0 \geq p_n$. Let $\lambda_n = p_n\eta$ and $\mu' = \mu - \lambda_n$. As $\mu(A_n) \geq p_n$, it follows that μ' is never negative, thus, it is a measure on $\mathcal{B}_S$. By definition of ε_0,

$$\sum_{i\in I} p_i \leq \mu'\left(\bigcup_{i\in I} A_i\right), \qquad \text{for every } I \subset \{1,\ldots,n-1\}.$$

[5]Note that $\left(F^\delta\right)^\varepsilon \subset F^{\delta+\varepsilon}$ for if x belongs to the first set, take y in F^δ such that $d(x,y) < \varepsilon$ and take $z \in F$ such that $d(x,z) < \delta$, the triangle inequality shows that $d(x,z) < \delta + \varepsilon$ and $z \in F$.

By induction, it is possible to conclude there exist measures $\lambda_1, \dots, \lambda_{n-1}$ defined on $\mathcal{B}_S$ such that $\lambda_i(A_i) = \lambda_i(S) = p_i$ for $i = 1, \dots, n-1$ and $\sum_{i=1}^{n-1} \lambda_i(A) \leq \mu'(A) = \mu(A) - \lambda_n(A)$, for every $A \in \mathcal{B}_S$. As $\lambda_n(A_n) = p_n$, it follows that $\lambda_1, \dots, \lambda_n$ satisfies the conclusion and, therefore, this case is concluded.

To finish the cases, assume now that $\varepsilon_0 < p_n$. Let $\mu' = \mu - \varepsilon_0 \eta$. Then μ' is a measure on $\mathcal{B}_S$ as for any $A \in \mathcal{B}_S$,

$$\mu'(A) = \mu(A) - \varepsilon_0 \eta(A) = \frac{\mu(A_n)\mu(A) - \varepsilon_0 \mu(A \cap A_n)}{\mu(A_n)} \geq \frac{p_n \mu(A) - \varepsilon_0 \mu(A \cap A_n)}{\mu(A_n)} \geq 0,$$

where the last was concluded since $p_n \geq \varepsilon_0$ and $\mu(A) \geq \mu(A \cap A_n)$. The definition of ε_0 allows to conclude that there is a non-empty set $I_0 \subset \{1, \dots, n-1\}$ such that

$$\sum_{i \in I} p_i \leq \mu' \left(\bigcup_{i \in I} A_i \right), \quad \text{for every } I \subset I_0$$

with equality when $I = I_0$ (if there were no such I_0, then the first inequality would be strict for all subsets I and this would violate the definition of ε_0). By induction, there are measures $(\lambda_i)_{i \in I_0}$ defined on $\mathcal{B}_S$ such that $\lambda_i(A_i) = \lambda_i(S) = p_i$ for $i \in I_0$ and

$$\sum_{i \in I_0} \lambda_i(A) \leq \mu'(A),$$

for every $A \in \mathcal{B}_S$. Let $p_i' = p_i$ for $i = 1, \dots, n-1$, $p_n' = p_n - \varepsilon_0$, put

$$B_0 = \bigcup_{i \in I_0} A_i$$

and define the measure μ'' on $\mathcal{B}_S$ by

$$\mu''(A) = \mu'(A) - \mu'(A \cap B_0), \quad A \in \mathcal{B}_S.$$

Let $I_1 = \{1, \dots, n\} - I_0$. Then, for $I \subset I_1$,

$$\begin{aligned} \sum_{i \in I} p_i' + \mu'(B_0) &= \sum_{i \in I \cup I_0} p_i' \leq \mu' \left(\bigcup_{i \in I \cup I_0} A_i \right) \\ &= \mu' \left(\bigcup_{i \in I} A_i \right) + \mu'(B_0) - \mu' \left(\bigcup_{i \in I} A_i \cap B_0 \right) \\ &= \mu'' \left(\bigcup_{i \in I} A_i \right) + \mu'(B_0), \end{aligned}$$

where the first equality is the definition of I_0, the first inequality is a consequence of the definition of ε_0 when $n \notin I$; when $n \in I$, by the hypothesis and definition of p_i' and μ',

$$\sum_{i \in I \cup I_0} p_i' \leq \sum_{\substack{i \in I \cup I_0 \\ i \neq n}} p_i' + p_n' = \sum_{i \in I \cup I_0} p_i - \varepsilon_0 \leq \mu \left(\bigcup_{i \in I \cup I_0} A_i \right) - \varepsilon_0 = \mu' \left(\bigcup_{i \in I \cup I_0} A_i \right).$$

Consequently,

$$\sum_{i \in I} p_i' \leq \mu'' \left(\bigcup_{i \in I} A_i \right), \quad \text{for every } I \subset I_1.$$

By induction, there exists λ'_i $(i \in I_1)$ defined on $\mathcal{B}_S$ such that $\lambda'_i(A_i) = \lambda'_i(S) = p_i$ and $\sum_{i\in I_1} \lambda'_i(A) \leq \mu''(A)$ for every $A \in \mathcal{B}_S$. To conclude, let $\lambda_i = \lambda'_i$ for $i \in I_1 - \{n\}$ and $\lambda_n = \lambda'_n + \varepsilon_0\eta$. Then, all λ_i are measures on $\mathcal{B}_S$ and

$$\lambda_i(A_i) = \lambda_i(S) = p_i, \quad \forall i \in I_1 - \{n\}$$

and

$$\lambda_n(A_n) = \lambda'_n(A_n) + \varepsilon_0\eta(A_n) = p_n.$$

Finally, for $A \in \mathcal{B}_S$,

$$\begin{aligned}\sum_{i=1}^{n} \lambda_i(A) &= \sum_{i\in I_0} \lambda_i(A) + \sum_{i\in I_1} \lambda_i(A) = \sum_{i\in I_0} \lambda_i(A \cap B_0) + \sum_{i\in I_1} \lambda'_i(A) + \varepsilon_0\eta(A) \\ &\leq \mu'(A \cap B_0) + \mu''(A) + \varepsilon_0\eta(A) = \mu'(A) + \varepsilon_0\eta(A) = \mu(A).\end{aligned}$$

This concludes the proof of (1.2.1). ∎

Lemma (1.2.2) Suppose that μ is a finite measure defined on $\mathcal{B}_S$, that $(p_i)_{i=1,\dots,n}$ is a finite family of non-negative real numbers and $(A_i)_{i=1,\dots,n}$ is a finite family of Borels of S. If $\varepsilon > 0$ is such that

$$\sum_{i\in I} p_i \leq \mu\left(\bigcup_{i\in I} A_i\right) + \varepsilon$$

for all subsets $I \subset \{1, \dots, n\}$, then there exist finite measures $\lambda_1, \dots, \lambda_n$ defined on $\mathcal{B}_S$ such that $\lambda_i(A_i) = \lambda_i(S) \leq p_i$ for $i = 1, \dots, n$, $\sum_{i=1}^{n} \lambda_i(S) \geq \sum_{i=1}^{n} p_i - \varepsilon$ and $\sum_{i=1}^{n} \lambda_i(A) \leq \mu(A)$ for every $A \in \mathcal{B}_S$.

Proof: consider Alexandroff's construction[6] $S^* = S \cup \{\infty\}$ and note that (1.2.1) will be valid for any topological space (S, τ), where, of course, in this case $\mathcal{B}_S$ is the σ-algebra generated by the open sets of S. If μ is a given measure defined on $\mathcal{B}_S$, then it easy to see that if μ^* is defined on $\mathcal{B}_{S^*}$, and given by $\mu^*(A) = \mu(A)$ if $A \in \mathcal{B}_S$ and $\mu^*(\{\infty\}) = \varepsilon$, so that $\mu^*(A) = \mu(A \setminus \{\infty\}) + \varepsilon$ if $\infty \in A$, then μ^* is a measure on $\mathcal{B}_{S^*}$.

As it is true that, for $A_i^* = A_i \cup \{\infty\}$,

$$\sum_{i\in I} p_i \leq \mu\left(\bigcup_{i\in I} A_i^*\right)$$

for every subset $I \subset \{1, \dots, n\}$, (1.2.1) implies the existence of finite measures $\lambda_1^*, \dots, \lambda_n^*$ on $\mathcal{B}_{S^*}$ such that $\lambda_i^*(A_i^*) = \lambda_i^*(S^*) = p_i$ and $\sum_{i=1}^{n} \lambda_i^*(A) \leq \mu^*(A)$ for every $A \in \mathcal{B}_{S^*}$. For each i, λ_i be the restriction of λ_i^* to $\mathcal{B}_S$. Then $\lambda_i(A_i) = \lambda^*(A_i) \leq \lambda_i^*(A_i^*) = p_i$ and $\lambda_i(S - A_i) = \lambda_i^*(S^* - A_i^*) = 0$, for each i. Also,

$$\sum_{i=1}^{n} \lambda_i(S) = \sum_{i=1}^{n} (p_i - \lambda_i^*(\{\infty\})) \geq \sum_{i=1}^{n} p_i - \mu^*(\{\infty\}) = \sum_{i=1}^{n} p_i - \varepsilon$$

and

$$\sum_{i=1}^{n} \lambda_i(A) = \sum_{i=1}^{n} \lambda_i^*(A) \leq \mu(A),$$

for every $A \in \mathcal{B}_S$. ∎

[6]In this construction, $S \subset S^*$ is an *open* subset of the enlarged space. Therefore, $\mathcal{B}_S \subset \mathcal{B}_{S^*}$.

Lemma (1.2.3) Assume that (S, d) is a separable metric space and that P is a element of $\mathcal{P}_S$. For every $\varepsilon > 0$ and $\delta > 0$ there exists a finite "measurable partition" $B_0, \ldots, B_N$ of S (that is, there is a finite partition[7] of Borels of S) such that the diameter[8] of each of $B_1, \ldots, B_N$ is at most ε and $P(B_0) \leq \delta$.

Proof: by metrizability and separability, there exists a sequence $(x_n)_{n\in\mathbb{N}}$ that is dense in S; for each of the terms in the sequence, consider the ball $V_n = \mathrm{B}\left(x_n; \frac{\varepsilon}{2}\right)$, so, by the triangle inequality, it is true that $\delta(V_n) \leq \varepsilon$ no matter what the $n \in \mathbb{N}$ is. It is then clear that $S = \bigcup_{n=1}^{\infty} V_n$, so there exists an index N such that $P\left(\bigcup_{n=N+1}^{\infty} V_n\right) \leq \delta$; the sets $B_1 = V_1$ and $B_k = V_k - \bigcup_{j=1}^{k-1} B_j$, for $k = 2, \ldots, N$ must have a diameter at most ε (as they are contained in sets of diameter at most ε), and define then $B_0 = S - \bigcup_{k=1}^{N} B_k$; it is important to note that by construction, $\bigcup_{k=1}^{N} V_k = \bigcup_{k=1}^{N} B_k$, so $B_0 \subset \bigcup_{n=N+1}^{\infty} V_n$ and hence, by monotonicity, $P(B_0) \leq \delta$. Finally, the finite family $(B_0, \ldots, B_N)$ has the required properties. ∎

Lemma (1.2.4) Assume the hypotheses of (1.2.3). Let ε, δ be two positive real numbers and $Q \in \mathcal{P}_S$ such that $\rho(P, Q) < \varepsilon$. Let $(B_0, \ldots, B_N)$ be a measurable partition for which $P(B_0) \leq \delta$ and the sets $B_1, \ldots, B_N$ have diameter at most ε (which exists by (1.2.3)). Then, there exist real numbers $c_1, \ldots, c_N$ in $[0, 1]$ and independent random objects[9] $X, Y_0, \ldots, Y_N, U$ all of them defined over a probability space $(\Omega, \mathscr{F}, \nu)$ such that $X, Y_0, \ldots, Y_N$ take their values in the metric space S and U takes its values in $[0, 1]$, the distribution of X is P, the distribution of U is uniform on $[0, 1]$ and the random object Y, given by

(1.2.4.1)
$$Y = \begin{cases} Y_i & \text{on the set} \quad \{X \in B_i, U \geq c_i\} \text{ for } i = 1, \ldots, N, \\ Y_0 & \text{on the set} \quad \{X \in B_0\} \cup \bigcup_{i=1}^{N} \{X \in B_i, U < c_i\} \end{cases}$$

has distribution Q, the inclusion

(1.2.4.2)
$$\{d(X, Y) \geq \delta + \varepsilon\} \subset \{X \in B_0\} \cup \left\{U < \max_{i:P(B_i)>0} \frac{\varepsilon}{P(B_i)}\right\}$$

holds and, finally,

(1.2.4.3)
$$\nu(d(X, Y) \geq \delta + \varepsilon) \leq \delta + \varepsilon.$$

Proof: let $p_i = P(B_i)$ and $A_i = B_i^{\varepsilon}$ for $i = 1, \ldots, N$. Then, since $\rho(P, Q) < \varepsilon$,

$$\sum_{i\in I} p_i \leq \sum_{i\in I} P\left(\overline{\bigcup_{i\in I} B_i}\right) \leq Q\left(\bigcup_{i\in I} A_i\right) + \varepsilon,$$

[7] A "partition" of a set is a family of pairwise disjoint sets whose union is the whole set.

[8] For *any* set A in the metric space (S, d), the "diameter" of A is defined as the non-negative number, possibly $+\infty$, $\delta(A) = \sup_{x,y\in A} d(x, y)$.

[9] A "random object" is a map between two measurable spaces that is measurable with respect to the given σ-algebras. If such a map takes values in $\mathbb{R}$ it is called "random variable", if the values are taken in $\mathbb{R}^d$ it is called a "random vector" and the adjective "complex" is used when $\mathbb{R}$ (resp. $\mathbb{R}^d$) is replaced by $\mathbb{C}$ (resp. $\mathbb{C}^p$). A family of random objects is "independent" if the corresponding family of generated σ-algebras is independent (a family $(\mathscr{F}_\alpha)_{\alpha\in I}$ of σ-algebras is "independent with respect to the probability measure $\mathbb{P}$" if for every finite number of *distinct indices* $\alpha_1, \ldots, \alpha_k$ and sets $N_1, \ldots, N_k$ such that $N_i \in \mathscr{F}_{\alpha_i}$ -for $i = 1, \ldots, N$- one has $\mathbb{P}(N_1 \cap \ldots \cap N_k) = \mathbb{P}(N_1) \cdots \mathbb{P}(N_k)$).

since

$$\left(\overline{\bigcup_{i\in I} B_i}\right)^{\varepsilon} \subset \bigcup_{i\in I} A_i.$$

By (1.2.2), there exist measures $\lambda_1, \ldots, \lambda_N$ defined on $\mathcal{B}_S$ such that $\lambda_i(A_i) = \lambda_i(S) \leq p_i$ for $i = 1, \ldots, N$,

$$\sum_{i=1}^{N} \lambda_i(S) \leq \sum_{i=1}^{N} p_i - \varepsilon$$

and $\sum_{i=1}^{N} \lambda_i(A) \leq Q(A)$ for all the Borels A of S. Define $c_1, \ldots, c_N \in [0, 1]$ by

$$c_i = \frac{p_i - \lambda_i(S)}{p_i}$$

if $p_i \neq 0$ and $c_i = 0$ if $p_i = 0$. By the definitions,

$$(1 - c_i)P(B_i) = \lambda_i(S), \quad i = 1, \ldots, N$$

and

$$P(B_0) + \sum_{i=1}^{N} c_i P(B_i) = 1 - \sum_{i=1}^{N} \lambda_i(S).$$

Define *probability* measures $Q_0, \ldots, Q_N$ on $\mathcal{B}_S$ according to the following cases:

1. for $i = 1, \ldots, N$,
   a) if $(1 - c_i)P(B_i) = 0$, take Q_i to be *any* probability measure[10]; and
   b) if $(1 - c_i)P(B_i) \neq 0$, take $Q_i = \dfrac{\lambda_i}{(1 - c_i)P(B_i)}$;

   then, for this case,

   $$Q_i(B)(1 - c_i)P(B_i) = \lambda_i(B)$$

   for every Borel B of S.

2. For $i = 0$,
   a) if $P(B_0) + \sum_{i=1}^{N} c_i P(B_i) = 0$, proceed as before, take Q_0 to be any probability measure; and
   b) if $P(B_0) + \sum_{i=1}^{N} c_i P(B_i) \neq 0$, take $Q_0 = \dfrac{Q - \sum_{i=1}^{N} \lambda_i}{P(B_0) + \sum_{i=1}^{N} c_i P(B_i)}$;

   then, for every Borel B of S,

   $$Q_0(B)\left(P(B_0) + \sum_{i=1}^{N} c_i P(B_i)\right) = Q(B) - \sum_{i=1}^{N} \lambda_i(B).$$

[10] For example, take $x \in A_i$ and Q_i to be the "Dirac's measure concentrated at x" ε_x given by $\varepsilon_x(A) = 1$ if $x \in A$ and $\varepsilon_x(A) = 0$ if $x \notin A$; with this, Q_i is "concentrated" on A_i (that is, $Q_i(S - A_i) = 0$).

Now the space $(\Omega, \mathscr{F}, \nu)$ will be constructed. The procedure goes as follows: take $\Omega = S^{N+1} \times [0,1]$ and $\mathscr{F}$ the product σ-algebra $\mathcal{B}_S \otimes \ldots \otimes \mathcal{B}_S \otimes \mathcal{B}_{[0,1]}$ and take ν to be the product measure $P \otimes Q_0 \otimes \ldots \otimes Q_N \otimes \mathcal{U}$, where $\mathcal{U}$ denotes Lebesgue measure on $[0,1]$. It follows immediately that the random object $I_\Omega : \omega \to \omega$ (the identity function on Ω) has law[11] ν; henceforth, if it is written $I_\Omega = (X, Y_0, \ldots, Y_N, U)$, then the random objects $X, Y_0, \ldots, Y_N$ and U satisfy the required conditions, as shall be shown. For $i = 1, \ldots, N$, the probability measure Q_i is concentrated on A_i, hence Y_i takes values in A_i with probability one, hence, it can be redefined so that Y_i takes *all* its values in A_i. Now, let Y to be the random object defined in (1.2.4.1). Then, for any Borel B of S,

$$\begin{aligned}
\nu(Y \in B) &= \sum_{i=1}^{N} \nu(Y \in B, X \in B_i, U \geq c_i) + \nu\left(\{Y \in B\} \cap \left[\{X \in B_0\} \cup \bigcup_{i=1}^{N} \{X \in B_i, U < c_i\}\right]\right) \\
&= \sum_{i=1}^{N} Q_i(B) P(B_i)(1 - c_i) + Q_0(B)\left(P(B_0) + \sum_{i=1}^{N} c_i P(B_i)\right) \\
&= \sum_{i=1}^{N} \lambda_i(B) + Q(B) - \sum_{i=1}^{N} \lambda_i(B) = Q(B).
\end{aligned}$$

Now note that

$$\{X \in B_i, U \geq c_i\} \subset \{X \in B_i, Y \in A_i\}$$

since Y_i takes its values in A_i. Since the B_i have diameters at most δ and $A_i = B_i^\varepsilon$,

$$\{X \in B_i, Y \in A_i\} \subset \{d(X,Y) < \delta + \varepsilon\}.$$

Starting with this inclusion, which is valid for all $i = 1, \ldots, N$, it is then true that

$$\bigcup_{i=1}^{N} \{X \in B_i, U \geq c_i\} \subset \{d(X,Y) < \delta + \varepsilon\}.$$

Take the complement of these sets to obtain that,

$$\{d(X,Y) \geq \delta + \varepsilon\} \subset \bigcap_{i=1}^{N} \big(\{X \notin B_i\} \cup \{U < c_i\}\big);$$

it is claimed that the right hand set is included in the set

$$\{X \in B_0\} \cup \bigcup_{i=1}^{N} \{X \in B_i, U < c_i\};$$

for if $\omega \in \bigcap_{i=1}^{N} \big(\{X \notin B_i\} \cup \{U < c_i\}\big)$ but $\omega \notin \{X \in B_0\}$, then since the family $(B_0, \ldots, B_N)$ is a partition, there exists an index j such that $\omega \in \{X \in B_j\} \cap \bigcap_{\substack{i=1 \\ i \neq j}}^{N} \{X \notin B_i\}$, but for each i it must happen that $\omega \in \{X \notin B_i\}$ or

[11] The term "law" is used for the probability measure asociated with a random object, while "distribution" is used, almost exclusively, for random variables or random vectors (real or complex); explicitly, given a random object $Z : (\Omega, \mathscr{F}) \to (\Omega', \mathscr{F}')$ and a measure (typically, a probability measure) μ, the "law of Z with respect to μ" is the measure $\mathcal{L}_Z$, defined on $\mathscr{F}'$, by $\mathcal{L}_Z(B') = \mu(Z \in B')$; this is also called the "image measure of μ by Z" and denoted by $Z(\mu)$.

$\omega \in \{U < c_i\}$, hence for the index j it is readily seen that $\omega \in \{X \in B_j, U < c_j\}$ and the inclusion follows. By now, it has been established that

$$\{d(X,Y) \geq \delta + \varepsilon\} \subset \{X \in B_0\} \cup \bigcup_{i=1}^{N} \{X \in B_i, U < c_i\};$$

the last set it is clearly a subset of $\{X \in B_0\} \cup \left\{U < \max_{i=1,\ldots,N} c_i\right\}$; noting that $p_i \geq \lambda_i(S)$ and that $\varepsilon \geq \sum_{i=1}^{N}(p_i - \lambda_i(S))$, it follows that $\varepsilon \geq p_i - \lambda_i(S)$, hence, for all i such that $P(B_i) > 0$, one has $\frac{\varepsilon}{P(B_i)} \geq c_i$ and, therefore,

$$\left\{U < \max_{i=1,\ldots,N} c_i\right\} \subset \left\{U < \max_{i:P(B_i)>0} \frac{\varepsilon}{P(B_i)}\right\}.$$

Putting all the inclusions together, yields the inclusion (1.2.4.2). Finally, from subadditivity,

$$\nu(d(X,Y) \geq \delta + \varepsilon) \leq \nu(X \in B_0) + \sum_{i=1}^{N} \nu(X \in B_i, U < c_i) = P(B_0) + \sum_{i=1}^{N} c_i P(B_i),$$

by hypothesis $P(B_0) \leq \delta$ and $P(B_i) = p_i$, so that $\sum_{i=1}^{N} c_i P(B_i) = \sum_{i=1}^{N}(p_i - \lambda_i(S)) \leq \varepsilon$ and the last inequality (1.2.4.3) follows. ■

Lemma (1.2.5) Assume that (S,d) is a separable metric space and that $P, Q \in \mathcal{P}_S$. Define $\mathcal{M}(P,Q)$ to be the set of probabilities $\mu \in \mathcal{P}_{S\times S}$ whose first marginal[12] is P and second is Q. Then,

$$\rho(P,Q) = \inf_{\mu \in \mathcal{M}(P,Q)} \inf_{\varepsilon > 0} \Big(\mu(\{(x,y) \in S \times S | d(x,y) > \varepsilon\}) < \varepsilon\Big).$$

Proof: first, assume that $\varepsilon > 0$ and that $\mu \in \mathcal{M}(P,Q)$ satisfy that

$$\mu(\{(x,y) \in S \times S | d(x,y) < \varepsilon\}) < \varepsilon.$$

Then,

$$\begin{aligned} P(A) &= \mu(A \times S) \leq \mu((A \times S) \cap \{(x,y) | d(x,y) < \varepsilon\}) + \varepsilon \\ &\leq \mu\left(A \times A^{\varepsilon}\right) + \varepsilon \leq \mu\left(S \times A^{\varepsilon}\right) + \varepsilon \\ &= Q\left(A^{\varepsilon}\right) + \varepsilon. \end{aligned}$$

It follows that

$$\rho(P,Q) \leq \inf_{\mu \in \mathcal{M}(P,Q)} \inf_{\varepsilon > 0} \Big(\mu(\{(x,y) \in S \times S | d(x,y) > \varepsilon\}) < \varepsilon\Big).$$

Now, the other inequality will be proven. Let $\delta > 0$ and choose a measurable partition $(B_0, \ldots, B_N)$ of S such that the $B_1, \ldots, B_N$ have diameters at most $\delta' < \delta$ and $P(B_0) \leq \delta'$ (which exist by (1.2.3)). By (1.2.4), there exists a $\mu \in \mathcal{M}(P,Q)$ such that

$$\mu(\{(x,y) | d(x,y) \geq \delta + \varepsilon\}) \leq \varepsilon + \delta.$$

Therefore,

$$\inf_{\eta \in \mathcal{M}(P,Q)} \inf_{r > 0} \Big(\eta(\{(x,y) | d(x,y) \geq r\}) \leq r\Big) \leq \varepsilon + \delta.$$

[12]The "marginal measures" are defined for product measures; these are the measures on $\mathcal{B}_S$ given by $\mu_1(A) = \mu(A \times S)$ and $\mu_2(B) = \mu(S \times B)$; they will be called "first marginal" and "second marginal", respectively.

Taking the infimum over $\varepsilon > 0$, it is seen that

$$\rho(P,Q) + \delta \geq \inf_{\mu \in \mathcal{M}(P,Q)} \inf_{\varepsilon > 0} \Big(\mu(\{(x,y) \in S \times S | d(x,y) > \varepsilon\}) < \varepsilon\Big);$$

as this is true for every $\delta > 0$, the conclusion follows. ■

§1.3. Harvesting the consequences.

Now it is possible to obtain nice results in what follows. Specifically, Skorohod's representation theorem which allow, in several important cases, changing convergence of elements of $\mathcal{P}_S$ to convergence of random objects. Also, it will follow a theorem which allows concluding convergence of image measures of elements of $\mathcal{P}_S$ (see (1.3.4)).

Theorem (1.3.1) Assume that (S,d) is a separable metric space. Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of random objects defined on $(\Omega, \mathscr{F}, \nu)$ with values in S and let $(P_n)_{n \in \mathbb{N}}$ be the family of their laws. Let X be another random object on $(\Omega, \mathscr{F}, \nu)$ with values in S and law P. If the sequence of random variables $(d(X_n, X))_{n \in \mathbb{N}}$ converges to 0 in ν-probability[13], then $\lim_{n \to \infty} \rho(P_n, P) = 0$.

Proof: for every $n \in \mathbb{N}$, let μ_n be the law of the random object (X_n, X). The hypothesis implies that, for a given $\varepsilon > 0$,

$$\lim_{n \to \infty} \mu_n(\{(x,y) \in S \times S | d(x,y) > \varepsilon\}) = 0.$$

The result then follows immediately from (1.2.5). ■

Theorem (1.3.2) Let (S,d) be a separable metric space. Then, the metric space $(\mathcal{P}_S, \rho)$ is separable. If (S,d) is also complete, the space of probabilities has the same property.

Proof: this is a solution of problem 3 of chapter 3 of [9]. Let $(x_n)_{n \in \mathbb{N}}$ be a dense sequence in S. Consider the set $\mathcal{D}$ of probability measures that assume the form $\sum_{i=1}^{N} a_i \varepsilon_{x_i}$, where N runs over $\mathbb{N}$, the $a_1, \ldots, a_N$ run over the non-negative rationals, and ε_x denotes the Dirac measure concentrated at x. The set $\mathcal{D}$ is everwhere dense in $\mathcal{P}_S$, as will be shown below. Let $P \in \mathcal{P}_S$ and $\varepsilon > 0$. Take the measurable partition $(B_0, \ldots, B_N)$ constructed in (1.2.3) such that the diameters of $B_1, \ldots, B_N$ are at most $\frac{\varepsilon}{4}$ and $P(B_0) < \frac{\varepsilon}{2}$ (note that $B_i \subset \mathrm{B}\left(x_i; \frac{\varepsilon}{4}\right)$, so that $x_i \in B_i^{\frac{\varepsilon}{4}}$). Choose $a_i \in [0, P(B_i)] \cap \mathbb{Q}$ for $i = 1, \ldots, N$ such that

$$P(B_i) \leq a_i + \frac{\varepsilon}{2N}.$$

Given $A \in \mathcal{B}_S$,

$$P(A) = \sum_{i=1}^{N} P(A \cap B_i) + P(B_0);$$

note that the relation $B_i \cap A \neq \varnothing$ implies the relation $x_i \in B_i^{\frac{\varepsilon}{4}} \subset A^{\varepsilon}$, hence, if $P(A \cap B_i)$ is not zero, it is bounded above by $a_i \varepsilon_{x_i}(A^{\varepsilon}) + \frac{\varepsilon}{2N}$, and therefore

$$P(A) \leq \sum_{i=1}^{N} a_i \varepsilon_{x_i}(A^{\varepsilon}) + \varepsilon,$$

[13] A sequence of measurable functions (f_n) defined over a measure space $(\Omega, \mathscr{F}, \mu)$ and with values in the metric space (S,d), converges "in μ-measure" to the measurable function $f : \Omega \to S$ if, for every $\varepsilon > 0$, the following limit holds: $\lim_{n \to \infty} \mu(d(f_n, f) > \varepsilon) = 0$; this is equivalent to saying that the sequence of random variables $(d(f_n, f))$ converges in μ-measure to 0.

and this shows that $\rho\left(P, \sum_{i=1}^{N} a_i \varepsilon_{x_i}\right) < \varepsilon$, proving that $\mathcal{D}$ is dense.

To prove that $\mathcal{P}_S$ is a complete metric space with respect to ρ, it suffices to consider fundamental sequences[14] (P_n) for which $\rho(P_n, P_{n+1}) < 2^{-n}$ for $n \geq 2$. For each such n, choose measurable partitions $\left(B_j^{(n)}\right)_{j=0,\ldots,N(n)}$ such that $B_1^{(n)}, \ldots, B_{N(n)}^{(n)}$ have diameters at most 2^{-n} and $P\left(B_0^{(n)}\right) \leq 2^{-n}$. Using the Kolmogorov Extension Theorem[15], there exists a probability space $(\Omega, \mathscr{F}, \nu)$ supporting random objects $\left(Y_j^{(n)}\right)_{j=0,\ldots,N(n)}$, for $n \geq 2$ and with values in S, random variables $U^{(n)}$, for $n \geq 2$ and with values in $[0,1]$, a random object X_1 with values in S and with law P_1, all of them independent, and $c_j^{(n)} \in [0,1]$ for $j = 0, \ldots, N(n)$ and $n \geq 2$ such that

$$X_n = \begin{cases} Y_i^{(n)} & \text{on the set} \quad \left\{X_{n-1} \in B_i^{(n)}, U^{(n)} \geq c_i^{(n)}\right\} \text{ for } i = 1, \ldots, N(n), \\ Y_0^{(n)} & \text{on the set} \quad \left\{X_{n-1} \in B_0^{(n)}\right\} \cup \bigcup_{i=1}^{N(n)} \left\{X_{n-1} \in B_i^{(n)}, U^{(n)} < c_i^{(n)}\right\} \end{cases}$$

has law P_n and that

$$\nu(d(X_{n-1}, X_n) \geq 2^{-n+1}) \leq 2^{-n+1}, \quad n \geq 2.$$

By the Borel-Cantelli lemma

$$\nu\left(\sum_{n=2}^{\infty} d(X_{n-1}, X_n) < \infty\right) = 1$$

and, by completeness, $\lim_{n\to\infty} X_n$ exists for almost every point with respect to ν. Let X be any random object defined on *all* Ω and such that $X(\omega)$ is equal to be the previous limit when exists. Let P be the law of X. Then, by (1.3.1), $\lim_{n\to\infty} \rho(P_n, P) = 0$ and the theorem is proved. ■

Theorem (1.3.3) Suppose that (S, d) is a separable metric space and the sequence $(P_n)_{n\in\mathbb{N}}$ converges to P in $\mathcal{P}_S$. Then there exists a probability space $(\Omega, \mathscr{F}, \nu)$ which supports random objects X, X_n ($n \in \mathbb{N}$) with values in S and whose laws are, respectively, P and P_n ($n \in \mathbb{N}$) and such that X_n converges to X for almost every point with respect to ν; this is known as "Skorohod's representation theorem".

Proof: by virtue of (1.2.3), for each $k \in \mathbb{N}$, it is possible to define a measurable partition sets

$$B_0^{(k)}, \ldots, B_{N(k)}^{(k)}$$

for which the diameters of the sets $B_1^{(k)}, \ldots, B_{N(k)}^{(k)}$ are at most 2^{-k} and with $P\left(B_0^{(k)}\right) \leq 2^{-k}$. Without loss of generality, it can be supposed that $\varepsilon_k = \min_{1\leq i\leq N(k)} P\left(B_i^{(k)}\right) > 0$. Define

$$k(n) = \max\{1\} \cup \left\{k \in \mathbb{N} \middle| \rho(P_n, P) < \frac{\varepsilon_k}{k}\right\}.$$

Apply (1.2.4) with $Q = P_n$, $\varepsilon = \dfrac{\varepsilon_{k(n)}}{k(n)}$ if $k(n) > 1$ or $\varepsilon = \rho(P_n, P) + \dfrac{1}{n}$ if $k(n) = 1$, $\delta = 2^{-k(n)}$, and the partition being $B_0^{(k(n))}, \ldots, B_{N(k(n))}^{(k(n))}$. Kolmogorov's Extension Theorem gives a probability space $(\Omega, \mathscr{F}, \nu)$ with random objects $Y_0^{(n)}, \ldots, Y_{N(k(n))}^{(n)}$ (with values in S), a random variable U (uniformly distributed in $[0,1]$), a random object

[14] In *any* metric space (S, d), a sequence of point $(x_n)_{n\in\mathbb{N}}$ is "fundamental" if it satisfies the "fundamental condition": for every $\varepsilon > 0$, there exists an index n_0 such that the relation $n \geq n_0$ implies the relation $d(x_n, x_{n_0}) < \varepsilon$.

[15] See, for example, Ash, *Real Analysis and Probability* [2].

X (with values in S) whose law is P, all of them independent and $c_1^{(n)}, \ldots, c_{N(k(n))}^{(n)} \in [0,1]$ such that for every $n \in \mathbb{N}$,

$$X_n = \begin{cases} Y_i^{(n)} & \text{on the set} \quad \left\{X \in B_i^{(k(n))}, U \geq c_i^{(n)}\right\} \text{ for } i = 1, \ldots, N(k(n)), \\ Y_0^{(n)} & \text{on the set} \quad \left\{X \in B_0^{(k(n))}\right\} \cup \bigcup_{i=1}^{N(k(n))} \left\{X \in B_i^{(k(n))}, U < c_i^{(n)}\right\} \end{cases}$$

is a random object with law P_n and the following inclusion holds for $k(n) > 1$

$$\left\{d(X_n, X) \geq 2^{-k(n)} + \frac{\varepsilon_{k(n)}}{k(n)}\right\} \subset \left\{X \in B_0^{(k(n))}\right\} \cup \left\{U < \frac{1}{k(n)}\right\}.$$

If $K_n = \min_{m \geq n} k(m) > 1$, then

$$\nu\left(\bigcup_{m=n}^{\infty} \left\{d(X_m, X) \geq 2^{-k(m)} + \frac{\varepsilon_{k(m)}}{k(m)}\right\}\right) \leq \sum_{k=K_n}^{\infty} \nu\left(X \in B_0^{(k)}\right) + \nu\left(U < \frac{1}{K_n}\right) \leq 2^{-K_n+1} + \frac{1}{K_n},$$

and, since $\rho(P_n, P) \to 0$, it follows $\lim_{n\to\infty} K_n = \infty$, and it turns out that the limit $\lim_{n\to\infty} X_n$ exists for almost every point with respect to ν. Now take X to be any random object whose values coincide with the previous limit when that limit exists. Then the sequence X_n converges to X for almost every point with respect to ν and Skorohod's Theorem is proved. ■

Now the Continuity Theorem for Borel functions is stated and proved.

Theorem (1.3.4) Let (S, d) and (S', d') be two separable metric spaces and suppose that $h : S \to S'$ is a Borel function[16]. Let (P_n) be a sequence in $\mathcal{P}_S$ which converges to P and define $Q_n = h(P_n)$ and $Q = h(P)$ the corresponding image measures of P_n and P by h. Let C_h be the set of points where h is continuous[17]. If $P(C_h) = 1$, then

$$\lim_{n\to\infty} \rho'(Q_n, Q) = 0,$$

where ρ' is the Prohorov distance on $\mathcal{P}_{S'}$; the "Continuity theorem" for Borel functions.

Proof: by (1.3.3), there exists a probability space $(\Omega, \mathscr{F}, \nu)$ and random objects X, X_n $(n \in \mathbb{N})$ so that X has law P and X_n has law P_n, all of them taking values in S, and such that $X_n \to X$ for almost every point with respect to ν. As $P(C_h) = \nu(X \in C_h) = 1$, it follows that $h(X_n) \to h(X)$ for almost every point with respect to ν. By (1.3.1), $\rho'(h(P_n), h(P)) = \rho'(Q_n, Q)$ converges to 0 and the thesis follows. ■

§1.4. Prohorov's Theorem.

By now, the metric space $(\mathcal{P}_S, \rho)$ has been defined. It is clear that the most important sets in a metric space are those which are compact and, in fact, the whole "Theory of Weak Topologies" appears motivated to find coarser topologies with more compact sets. Now, the compact sets in this particular metric space will be characterized.

Definition (1.4.1) Let $(P_\lambda)_{\lambda \in L}$ be a family of elements of $\mathcal{P}_S$. Such family is called "tight" if, for every $\varepsilon > 0$, there exists a compact set $K \subset S$ such that

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

[16]A function is "Borel" if it is measurable with respect to the corresponding Borel σ-algebras.
[17]It is known that $C_h \in \mathcal{B}_S$.

The definition of a "tight" set is analogous, explicitly, a set $\mathcal{T}$ of probability measures is "tight" if, for every $\varepsilon > 0$, there exists a compact set $K \subset S$ such that

$$\inf\{P(K) | P \in \mathcal{T}\} \geq 1 - \varepsilon.$$

When the phrase "assume $P_1, \ldots, P_n$ is tight" is used, it will mean that the set $\{P_1, \ldots, P_n\}$ is tight.

Lemma (1.4.2) If (S, d) is a separable and complete metric space, then each $P \in \mathcal{P}_S$ is tight.

Proof: let $(x_n)_{n \in \mathbb{N}}$ be a dense sequence in S and, for each $n \in \mathbb{N}$, choose integers $N_n \in \mathbb{N}$ (as in (1.2.3)) such that

$$P\left(\bigcup_{k=1}^{N_n} \mathrm{B}\left(x_k; \frac{1}{n}\right)\right) \geq 1 - \frac{\varepsilon}{2^n}.$$

Let K be the *closure* of the set $\bigcap_{n=1}^{\infty} \bigcup_{k=1}^{N_n} \mathrm{B}\left(x_k; \frac{1}{n}\right)$, so that K is totally bounded[18]. Now, K is closed in the complete space (S, d), hence (K, d) is complete and therefore (K, d) is compact[19]. Finally,

$$\begin{aligned} P(K) &\geq P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=1}^{N_n} \mathrm{B}\left(x_k; \frac{1}{n}\right)\right) = 1 - P\left(\bigcup_{n=1}^{\infty} \bigcap_{k=1}^{N_n} \complement \mathrm{B}\left(x_k; \frac{1}{n}\right)\right) \\ &\geq 1 - \sum_{n=1}^{\infty} P\left(\bigcap_{k=1}^{N_n} \mathrm{B}\left(x_k; \frac{1}{n}\right)\right) \geq 1 - \sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} = 1 - \varepsilon, \end{aligned}$$

which concludes the proof. ■

Theorem (1.4.3) Let (S, d) be a complete and separable metric space and suppose that $\mathcal{K} \subset \mathcal{P}_S$. The following conditions are equivalent:

(I) $\mathcal{K}$ is tight;

(II) for every $\varepsilon > 0$, there exists a compact set $K \subset S$ such that

$$\inf_{P \in \mathcal{K}} P\left(K^{\varepsilon}\right) \geq 1 - \varepsilon;$$

(III) $\mathcal{K}$ is relatively compact[20];

this is "Prohorov's Theorem".

Proof: note that condition (I) implies condition (II) trivially. So, the proof consist in showing that (II) implies (III) and (III) implies (I).

(II) implies (III) By Theorem (1.3.2), the metric space $(\mathcal{P}_S, \rho_S)$ is a complete metric space, hence, $\overline{\mathcal{K}}$ is also complete; then, to conclude that (III) holds, it suffices to show that $\mathcal{K}$ is totally bounded. Let $\delta > 0$ and $0 < \varepsilon < \frac{\delta}{2}$. Choose a compact set $K \subset S$ that satisfies hypothesis (II). As K is compact, there are finite points $x_1, \ldots, x_N \in K$ such that

$$K^{\varepsilon} \subset \bigcup_{i=1}^{N} \mathrm{B}\left(x_i; 2\varepsilon\right),$$

[18] In *any* metric space (S, d) a set A is "totally bounded" if for every $\varepsilon > 0$ there exists a finite set $F \subset S$ such that $A \subset F^{\varepsilon}$. To see that K is totally bounded, take $\varepsilon > 0$ and choose an $n \in \mathbb{N}$ so that $\frac{1}{n} < \varepsilon$, then take F to be the set $\{x_1, \ldots, x_{N_n}\}$.

[19] See (3.16.1) of [6].

[20] In *any* metric space (S, d), a set A is called "relatively compact" if $\overline{A}$ is compact.

put $B_i = \mathrm{B}(x_i; 2\varepsilon)$. Let $x_0 \in S$ and $m \geq \frac{N}{\varepsilon}$ be an integer number. Define $\mathcal{N}$ as the set of probability measures that can be written in the form

$$P = \sum_{i=0}^{N} \frac{k_i}{m} \varepsilon_{x_i},$$

where $0 \leq k_i \leq m$ is an integer number and $\sum_{i=0}^{N} k_i = m$. Given an element $Q \in \mathcal{K}$, put k_i to be the integer part of $mQ(E_i)$ for $i = 1, \ldots, N$, where the sets E_i are defined via $E_i = B_i - \bigcup_{j=1}^{i-1} B_j$ and, for the k_i to add up to m, set $k_0 = m - \sum_{i=1}^{N} k_i$. Let P be the element generated by these constants, that is,

$$P = \sum_{i=0}^{N} \frac{k_i}{m} \varepsilon_{x_i}.$$

Then,

$$Q(F) = Q\left(F \cap \bigcup_{i=1}^{N} E_i\right) + Q\left(F \cap \complement\left[\bigcup_{i=1}^{N} E_i\right]\right) \leq Q\left(\bigcup_{F \cap E_i \neq \varnothing}^{N} E_i\right) + Q\left(\complement K^{\varepsilon}\right),$$

the last inequality is due to $K^{\varepsilon} \subset \bigcup_{i=1}^{N} E_i$. Observe that

$$Q(E_i) \leq \frac{k_i + 1}{m},$$

so

$$Q(F) \leq \sum_{F \cap E_i \neq \varnothing} \frac{k_i}{m} + \frac{N}{m} + \varepsilon.$$

Now, m was selected in such a way that $\frac{N}{m} \leq \varepsilon$; using this inequality in the previous one,

$$Q(F) \leq P\left(F^{2\varepsilon}\right) + 2\varepsilon.$$

This has shown that $Q \in \mathcal{N}^{2\varepsilon} \subset \mathcal{N}^{\delta}$ as $2\varepsilon \leq \delta$. Therefore, condition (III) holds.

(III) implies (I) Let $\varepsilon > 0$. As the set $\mathcal{K}$ is relatively compact, it is totally bounded, so there exist finite sets $\mathcal{N}_n \subset \mathcal{K}$ ($n \in \mathbb{N}$) such that $\mathcal{K}$ is a subset of the augmentation by $\frac{\varepsilon}{2^{n+1}}$ of $\mathcal{N}_n$. By (1.4.2), for each n, there can be found a compact set[21] $K_n \subset S$ such that $P(K_n) \geq 1 - \frac{\varepsilon}{2^{n+1}}$ for every $P \in \mathcal{N}_n$ ($n \in \mathbb{N}$). Let $Q \in \mathcal{K}$. Then, for each $n \in \mathbb{N}$, there is a $P_n \in \mathcal{N}_n$ such that $\rho(Q_n, P_n) \leq \frac{\varepsilon}{2^{n+1}}$ or, as the definition implies,

$$Q\left(K^{\frac{\varepsilon}{2^{n+1}}}\right) \geq P_n(K_n) - \frac{\varepsilon}{2^{n+1}} \geq 1 - \frac{\varepsilon}{2^n}.$$

[21] The lemma proved that *one* probability measure is tight, but an easy inductive argument can show that this actually holds for sets of finite probability measures.

Define K to be the *closure* of the set $\bigcap_{n=1}^{\infty} K_n^{\frac{\varepsilon}{2^{n+1}}}$; such a K is compact as it is closed and totally bounded. Finally,

$$Q(K) \geq 1 - \sum_{n=1}^{\infty} P\left(\complement K_n^{\frac{\varepsilon}{2^{n+1}}}\right) \geq 1 - \sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} = 1 - \varepsilon,$$

and this demonstrates that, since Q was any element of $\mathcal{K}$, the set $\mathcal{K}$ is tight. ■

The following Scholium is here to show the powerfullness of the previous theorems. It can be skipped without interfering with the rest of the thesis.

Scholium. The previous theorem can be greatly generalized since it will *not* assume separability of completeness of S. In fact, if (S, d) is *any* metric space and $\mathcal{K} \subset \mathcal{P}_S$ is tight, then $\mathcal{K}$ is relatively compact. The proof goes as follows. For each $m \in \mathbb{N}$, the tightness of $\mathcal{K}$ implies the existence of a compact set $K_m \subset S$ such that

$$\inf_{P \in \mathcal{K}} P(K_m) \geq 1 - \frac{1}{m},$$

and it can be supposed that $K_m \subset K_{m+1}$ ($m \in \mathbb{N}$). Given $P \in \mathcal{K}$ and $m \in \mathbb{N}$, define $P^{(m)} \in \mathcal{P}_S$ by

$$P^{(m)}(A) = \frac{P(A \cap K_m)}{P(K_m)};$$

note that it is possible to restrict $P^{(m)}$ to $\mathcal{B}_{K_m}$ and then assume that $P^{(m)} \in \mathcal{P}_{K_m}$. As K_m is complete and separable (since *all* compact metric spaces are separable), the set

$$\mathcal{K}^{(m)} = \left\{ P^{(m)} \middle| P \in \mathcal{K} \right\}$$

is relatively compact in $\mathcal{P}_{K_m}$ ($m \in \mathbb{N}$). Also, for $P \in \mathcal{K}$, $A \in \mathcal{B}_S$ and $m, p, q \in \mathbb{N}$, the following *first* inequality is obtained:

$$P(A) \leq P(A \cap K_m) + \frac{1}{m} \leq P^{(m)}(A) + \frac{1}{m},$$

and a *second* inequality also holds:

$$P^{(q)}(A) = \frac{P(A \cap K_q)}{P(K_q)} \leq \frac{P(A \cap K_p) + \frac{1}{p}}{P(K_p)} \leq P^{(p)}(A) + \frac{1}{p-1} \leq P^{(p)}(A) + \frac{2}{p},$$

for $q \geq p \geq 2$. In the same way (and with the same notation), a *third* inequality is obtained:

$$P(A) \geq P(K_m) P^{(m)} P(A) \geq \left(1 - \frac{1}{m}\right) P^{(m)}(A)$$

and a *fourth* one also arises:

$$P^{(q)}(A) \geq \frac{P(K_p) P^{(p)}(A)}{P(K_q)} \geq \left(1 - \frac{1}{p}\right) P^{(p)}(A),$$

for $q \geq p$. The first inequality above shows that

$$\rho\left(P, P^{(m)}\right) \leq \frac{1}{m}$$

no matter what the $P \in \mathcal{K}$ and $m \in \mathbb{N}$ are. Now, let $(A_n)_{n \in \mathbb{N}}$ be a sequence of disjoint Borels of S. The second and fourth inequalities show that (for $P \in \mathcal{K}$ and $q \geq p \geq 2$) $\sum_{n=1}^{\infty} \left| P^{(q)}(A_n) - P^{(p)}(A_n) \right| \leq \sum_{n=1}^{\infty} \left\{ \left| P^{(q)}(A_n) - \left(1 - \frac{1}{p}\right) P^{(p)}(A_n) \right| + \frac{1}{p} P^{(p)}(A_n) \right\}$, which follows from the triangle inequality; the term on the rigth is equal to $\sum_{n=1}^{\infty} \left[P^{(q)}(A_n) - \left(1 - \frac{1}{p}\right) P^{(p)}(A_n) + \frac{1}{p} P^{(p)}(A_n) \right]$, which is deduced immediately from the fourth inequality; some algebra reduces this to $\sum_{n=1}^{\infty} \left[P^{(q)}(A_n) - P^{(p)}(A_n) + \frac{2}{p} P^{(p)}(A_n) \right]$; reordering terms gives $\sum_{n=1}^{\infty} \left| P^{(q)}(A_n) - P^{(p)}(A_n) \right| = \left[P^{(q)}\left(\bigcup_{n=1}^{\infty} (A_n) \right) - P^{(p)}\left(\bigcup_{n=1}^{\infty} A_n \right) \right] + \frac{2}{p} P^{(p)}\left(\bigcup_{n=1}^{\infty} A_n \right)$. Note now that the term on the rigth above is at most $\frac{4}{p}$, as the second inequality shows at once. Now let $(P_n)_{n \in \mathbb{N}}$ be a sequence in $\mathcal{K}$. By the relative compactness of $\mathcal{K}$, there exists a subsequence $\left(P_{\varphi(n)}\right)_{n \in \mathbb{N}}$ of (P_n) and $Q^{(m)} \in \mathcal{P}_{K_m}$ ($m \in \mathbb{N}$) such that

$$\lim_{n \to \infty} \rho\left(P_{\varphi(n)}^{(m)}, Q^{(m)}\right) = 0$$

for every $m \in \mathbb{N}$; in fact, for every $m \in \mathbb{N}$, the sequence $\left(P_n^{(m)}\right)_{n\in\mathbb{N}}$ belongs to the compact metric space $\mathcal{P}_{K_m}$ (this follows at once from Prohorov's theorem (1.4.3)) and, consequently, as the product of the compact and metric spaces $\prod_{m=1}^{\infty} \mathcal{P}_{K_m}$ is metric and compact[22], hence there exists a subsequence which converges. This means that there exists a $\left(P_{\varphi(n)}\right)$ such that, for every $m \in \mathbb{N}$, the coordinate sequence $\left(P_{\varphi(n)}^{(m)}\right)_{n\in\mathbb{N}}$ converges in $\mathcal{P}_{K_m}$ to some element $Q^{(m)}$, as stated earlier. It follows that

$$Q^{(m)}(F) = \lim_{\varepsilon\downarrow 0}\left(\limsup_{n\to\infty} P_{\varphi(n)}^{(m)}\left(F^{\varepsilon}\right)\right), \quad \forall F \in \mathcal{C}_S;$$

for if $F \in \mathcal{C}_S$ and $\varepsilon > 0$, there exists some $n_0 \in \mathbb{N}$ such that the relation $n \geq n_0$ implies that $\rho\left(P_{\varphi(n)}^{(m)}, Q^{(m)}\right) \leq \varepsilon$ and, hence,

$$Q^{(m)}(F) \leq P_{\varphi(n)}^{(m)}\left(F^{\varepsilon}\right) + \varepsilon,$$

taking $\limsup n \to \infty$ and then letting $\varepsilon \downarrow 0$, there follows the inequality $\leq$; now, to prove the reverse inequality, the symmetry of ρ shall be used, for then, there is an $n_1 \in \mathbb{N}$ for which if $n \geq n_1$, then $P_{\varphi(n)}\left(F^{\varepsilon}\right) \leq Q^{(m)}\left(\overline{F^{\varepsilon}}^{\varepsilon}\right) + \varepsilon \leq Q^{(m)}\left(F^{2\varepsilon}\right) + \varepsilon$, and taking $\limsup_{n\to\infty}$ it is seen that $\limsup_{n\to\infty} P_{\varphi(n)}^{(m)}\left(F^{\varepsilon}\right) \leq Q^{(m)}\left(F^{2\varepsilon}\right) + \varepsilon$, if $\varepsilon \downarrow 0$, the right hand side converges to $Q^{(m)}(F)$ and this proves the inequality $\geq$. Previously, it was proven that

$$P^{(q)}(A) \leq P^{(p)}(A) + \frac{2}{p} \quad \text{and} \quad P^{(q)}(A) \geq \left(1 - \frac{1}{p}\right) P^{(p)}(A),$$

for any elements $P \in \mathcal{K}$ and integers $q \geq p \geq 2$. Using these inequalities and the previous formula for $Q^{(m)}(F)$, it is then clear that

$$Q^{(q)}(F) \leq Q^{(p)}(F) + \frac{2}{p} \quad \text{and} \quad Q^{(q)}(F) \geq \left(1 - \frac{1}{p}\right) Q^{(p)}(F),$$

for any $F \in \mathcal{C}_S$ and integers $q \geq p \geq 2$. Now, since Q is a regular measure[23] and, hence, the two last inequalities are valid for all $F \in \mathcal{B}_S$. As proved earlier for $P^{(q)}$ and $P^{(p)}$, it follows that

$$\sum_{n=1}^{\infty}\left|Q^{(q)}(A_n) - Q^{(p)}(A_n)\right| \leq \frac{4}{p}$$

for integers $q \geq p \geq 2$. Hence, for $A \in \mathcal{B}_S$,

$$Q(A) = \lim_{m\to\infty} Q^{(m)}(A)$$

exists for every $A \in \mathcal{B}_S$. Up to now, there have been proved the existence of a subsequence $P_{\varphi(n)}$ for which $P_{\varphi(n)}^{(m)}$ converges to $Q^{(m)}$ and a *function* Q, defined on $\mathcal{B}_S$, such that $Q(A)$ is the limit of $Q^{(m)}(A)$. It seem natural to use an argument of a "$\frac{\varepsilon}{3}$-triangle inequality" type to conclude that $P_{\varphi(n)}$ converges to Q. But first, is must be shown that Q belongs to $\mathcal{P}_S$. It is clear that Q is a non-negative function such that $Q(S) = 1$, so, the complete additivity, also known as σ-additivity, needs to be proved to conclude that $Q \in \mathcal{P}_S$. So, assume that $(A_n)_{n\in\mathbb{N}}$ is a sequence of pairwise disjoint Borel sets and let A be their union. Then,

$$Q(A) - \sum_{n=1}^{N} Q(A_n) = \lim_{m\to\infty}\left[Q^{(m)}(A) - \sum_{n=1}^{N} Q^{(m)}(A_n)\right] = \lim_{m\to\infty}\sum_{n=N+1}^{\infty} Q^{(m)}(A_n) \geq 0.$$

Now, as

$$\left|\sum_{n=N+1}^{\infty}\left[Q^{(q)}(A_n) - Q^{(p)}(A_n)\right]\right| \leq \sum_{n=N+1}^{\infty}\left|Q^{(q)}(A_n) - Q^{(p)}(A_n)\right| \leq \frac{4}{p},$$

[22]For *any* countable family of metric spaces $(S_n, d_n)_{n\in\mathbb{N}}$, the product space $S = \prod_{n=1}^{\infty} S_n$ is metrizable (via the metric $d(x,y) = \sum_{n=1}^{\infty} 2^{-n}\min\{d_n(x_n,y_n),1\}$) and if each of the factors is separable (resp. totally bounded; resp. complete; resp. a sequence in S is fundamental; resp. a sequence in S is convergent), the product space is separable (resp. totally bounded; resp. complete; resp. the coordinate sequences in the corresponding S_n are fundamental; resp. the coordinate sequences in the corresponding S_n are convergent). Similarly, Tychonoff's theorem (see theorem 1.4 (4), chap. XI, of [8] shows that the Cartesian product of *any* family of *compact* spaces is also compact.

[23]A "regular measure" is a measure μ defined on a topological space such that $\mu(A) = \sup\{\mu(K) | K \subset A \text{ is compact}\} = \inf\{\mu(O) | A \subset O \text{ is open}\}$. As probability measures are tight, they are also regular and, in fact, for a probability measure P it follows that $P(A) = \sup\{P(C) | C \subset A \text{ is closed}\}$ since compact sets are closed.

the following inequality holds

$$\sum_{n=N+1}^{\infty} Q^{(m)}(A_n) \leq \frac{4}{p} + \sum_{n=N+1}^{\infty} Q^{(p)}(A_n),$$

for all $m \geq p$. Thus,

$$0 \leq Q(A) - \sum_{n=1}^{N} Q(A_n) \leq \frac{4}{p} + \sum_{n=N+1}^{\infty} Q^{(p)}(A_n).$$

Letting $N \to \infty$, the second term on the rigth hand side disappears (since $Q^{(p)}$ is a probability in S) and letting $p \to \infty$, it follows that $Q(A) = \sum_{n=1}^{\infty} Q(A_n)$, proving the complete additivity of Q. Finally,

$$\rho\left(P_{\varphi(n)}, Q\right) \leq \rho\left(P_{\varphi(n)}, P_{\varphi(n)}^{(m)}\right) + \rho\left(P_{\varphi(n)}^{(m)}, Q^{(m)}\right) + \rho\left(Q^{(m)}, Q\right).$$

Letting $n \to \infty$, it follows that

$$0 \leq \liminf_{n\to\infty} \rho\left(P_{\varphi(n)}, Q\right) \leq \limsup_{n\to\infty} \rho\left(P_{\varphi(n)}, Q\right) \leq \frac{2}{m}.$$

Letting $m \to 0$, it can be seen that $\lim_{n\to\infty} \rho\left(P_{\varphi(n)}, Q\right) = 0$, therefore $\mathcal{K}$ is relatively compact. ■

§1.5. Weak Convergence.

In the begining of this chapter some concepts for the study of weak convergence were mentioned. Now, those terms will be explained in a more detailed way.

‡ Motivation for the definition of weak convergence.

First of all, assume that $(E_\lambda, \tau_\lambda)_{\lambda\in L}$ is a family of topological spaces and let E be *any* non empty set. If for each $\lambda \in L$, a function $f_\lambda : E \to E_\lambda$ is given, a natural question arises: which topologies make the family $(f_\lambda)_{\lambda\in L}$ continuous? That is, what condition should a topology τ on E satisfy in order to make each of the functions f_λ continuous? Several important points will now be made.

1º Recall that a function $f : (E, \tau) \to (E', \tau')$ is continuous is equivalent to saying that for every open set O' in E', the set $f^{-1}(O')$ (defined as the subset of points $x \in E$ such that $f(x) \in O'$) is open in E.

2º Since $f^{-1}\left(\bigcup_{\alpha\in I} O'_\alpha\right) = \bigcup_{\alpha\in I} f^{-1}(O'_\alpha)$ and $f^{-1}\left(\bigcap_{i=1}^{n} O'_i\right) = \bigcup_{i=1}^{n} f^{-1}(O'_i)$, the set $f^{-1}(\tau')$ of subsets of E that assume the form $f^{-1}(O')$, where O' run over τ', is a *topology* on E.

3º From the previous two points, in order for a function $f : (E, \tau) \to (E', \tau')$ to be continuous, it is necessary and sufficient that the topology $f^{-1}(\tau')$ be contained in the topology τ; in other words, τ is "finer" than $f^{-1}(\tau)$.

4º This proves that there exists "coarsest" topology that makes f continuous, namely $f^{-1}(\tau')$.

5º The more elements a topology has, the fewer the compact sets there are; and, as an extreme case, if the "discrete" topology (that is, the topology where *all subset are open*) is given, only the *finite* sets are compact. Since compactness plays the same role in topology as finiteness plays in set theory, this is a *bad thing for the topology to do*. Therefore, *if a topology makes f continuous and has fewer open sets, it is a better topology.*

6º Returning to the case of the $(f_\lambda)_{\lambda\in L}$ above, the same argument shows that the *best* topology for which all the functions f_λ are *continuous* is the *coarsest* topology that contains all the topologies $f_\lambda^{-1}(\tau_\lambda)$. If such a topology exists, it will be called "weak topology induced by the family (f_λ)".

7º For a given set $\mathfrak{G} \subset \mathfrak{P}(E)$, there is always a coarsest topology $\tau_{\mathfrak{G}}$ that contains $\mathfrak{G}$ and, in fact, $\bigcap_{\tau} \tau$ is such topology where τ runs over the set of all topologies of E that contain all the sets of $\mathfrak{G}$ as elements (there is always at least one such τ, namely, $\tau = \mathfrak{P}(E)$). Hence, weak topologies induced by families of functions *always* exist.

8º It can be shown easily that if $\mathfrak{S} = \bigcup_{\lambda \in L} f_{\lambda}^{-1}(\tau_{\lambda})$ (the letter $\mathfrak{S}$ -which is an 'S'- stands for "subbasic"), then the set $\mathfrak{B}$ of subsets of E that take the form $\bigcap_{\lambda \in F} f_{\lambda}^{-1}(O'_{\lambda})$, where $F \subset L$ is *finite* and, for each O'_{λ} is open in E_{λ}, (the letter $\mathfrak{B}$ -which is a 'B'- stands for "basic") has the property *each* element of the weak topology induced by the family $(f_{\lambda})_{\lambda \in L}$ is of the form $\bigcup_{\alpha \in I} B_{\alpha}$, where $(B_{\alpha})_{\alpha \in I}$ is *some* family of elements of $\mathfrak{B}$.

Now, more specifications will be given. Take E to be the Cartesian product $\prod_{\lambda \in L} E_{\lambda}$ and f_{λ} to be the "projection function" pr_{λ} given by $\mathrm{pr}_{\lambda}(x_{\alpha})_{\alpha \in L} = x_{\lambda}$. The weak topology on E defined by the family of projection is the "product topology" and, by construction, *all the projections are continuous*. But more, all these projections are *open functions*[24].

Note that with these definitions, a sequence $\left(x^{(n)} = \left(x_{\lambda}^{(n)}\right)_{\lambda \in L}\right)_{n \in \mathbb{N}}$ will converge to the point $x = (x_{\lambda})_{\lambda \in L}$ if and only if each coordinate sequence $\left(x_{\lambda}^{(n)}\right)_{n \in \mathbb{N}}$ converges to x_{λ} (for every $\lambda \in L$)[25]. For this reason, the product topology is also called "topology of simple convergence" or "topology of pointwise convergence".

Now, observe that if all the E_{λ} are vector spaces, then so is E (the vector space structure is defined coordinate-wise) and the functions pr_{λ} must be *linear* by the definition of the vector space operations on E. Therefore, if the spaces E_{λ} are linear, the projections are *continuous*, *linear* and *open* functions.

The indexing set L is also not arbitrary: it will be a topological linear space (and, in this case, a normed space) and all the E_{λ} will be a vector space V. Hence, $E = V^L$ is the vector space of *all* functions $L \to V$, so it contains the subspace $\mathscr{L}(L;V)$ of all *continuous linear* functions $L \to V$. If both L and V have norms, $\mathscr{L}(L;V)$ will alway be considered with the "natural norm" induced by those of L and V; explicitly, if $\mu \in \mathscr{L}(L;V)$, then

$$\|\mu\| = \sup_{f \in L: \|f\|_L \leq 1} \|\mu(f)\|_V, \quad \forall \mu \in \mathscr{L}(L;V)$$

defines a norm (as can be verified by direct computation). Taking into consideration that for $f = 0$, $\|\mu(f)\|_V = 0 = \|\mu\| \, \|f\|_L$, and if $f \neq 0$, then $\hat{f} = \dfrac{f}{\|f\|_L}$ satisfies $\left\|\hat{f}\right\|_L \leq 1$, thus, $\|\mu(f)\|_V = \|f\|_L \left\|\mu\left(\hat{f}\right)\right\| \leq \|\mu\| \, \|f\|_L$. So, for every $f \in L$,

$$\|\mu(f)\| \leq \|\mu\| \, \|f\|_L .$$

It must be stated that the projections, for which in this case will always be called "evaluation maps", are continuous: for if $f \in L$ and e_f denotes the evaluation map at f, then

$$\|e_f(\mu + \nu) - e_f(\mu)\|_V = \|(\mu + \nu)(f) - \mu(f)\|_V = \|\nu(f)\|_V \leq \|\nu\| \, \|f\|_L ,$$

[24] This means that they take open sets to open sets; indeed, for if a set $O \subset S$ is open and $x_{\lambda} \in \mathrm{pr}_{\lambda}(O)$, take $y = (y_{\alpha})_{\alpha \in L}$ a point in O, where $y_{\lambda} = x_{\lambda}$; there is a basic $B = \mathrm{pr}_{\lambda_1}^{-1}\left(O'_{\lambda_1}\right) \cap \ldots \cap \mathrm{pr}_{\lambda_k}^{-1}\left(O'_{\lambda_k}\right)$ (the λ_i are different and O'_{λ_i} is open in E_{λ_i}) such that $y \in B \subset O$, and hence $x_{\lambda} \in \mathrm{pr}_{\lambda}(B) \subset \mathrm{pr}_{\lambda}(O)$; note that $\mathrm{pr}_{\lambda}(B) = S$ if λ is different from all of the λ_i and $\mathrm{pr}_{\lambda}(B) = O_{\lambda_i}$ for $\lambda = \lambda_i$, in any case, x_{λ} is an interior point of $\mathrm{pr}_{\lambda}(O)$, which means that O is open.

[25] To prove this, observe first that if $x^{(n)}$ converges to x, then by continuity, $\mathrm{pr}_{\lambda}\left(x^{(n)}\right)$ will converge to $\mathrm{pr}_{\lambda}(x)$; conversely, if all coordinate sequences $\mathrm{pr}_{\lambda}\left(x^{(n)}\right)$ converge to respective coordinate $\mathrm{pr}_{\lambda}(x)$ and O is an open neighbourhood of x, there is a basic neighbourhood $B = \mathrm{pr}_{\lambda_1}^{-1}\left(O'_{\lambda_1}\right) \cap \ldots \cap \mathrm{pr}_{\lambda_k}^{-1}\left(O'_{\lambda_k}\right)$ (the λ_i are different and O'_{λ_i} is open in S_{λ_i}) of x contained in O; as the indices $\lambda_1, \ldots, \lambda_k$ are finite, there is an index $n_0 \in \mathbb{N}$ such that if $n \geq n_0$ then, for $i = 1, \ldots, k$, $s_{\lambda_i}^{(n)} \in O'_{\lambda_i}$, which implies that $x^{(n)}$ belong to O for all $n \geq n_0$, and this proves that $x^{(n)}$ converges to x.

therefore, if $\|\nu\| \to 0$, then $e_f(\mu+\nu) \to e_f(\mu)$, which proves the continuity of the projection with respect to the topology generated by the norm in $\mathscr{L}(L;V)$. So, this norm generates a *finer* topology than the weak topology generated by the projection, and, as mentioned earlier, this is not the best behaviour possible.

The case of interest is when L is the normed vector space of bounded[26] *continuous* functions $S \to \mathbb{R}$ (recall that (S,d) is a metric space), which will be denoted by $\mathscr{C}^{\infty}_{\mathbb{R}}(S)$. The norm on $\mathscr{C}^{\infty}_{\mathbb{R}}(S)$ is given by the "supremum norm" (usually called "sup-norm") defined to be

$$\|f\| = \sup_{x\in S} |f(x)|, \quad \forall f \in \mathscr{C}^{\infty}_{\mathbb{R}}(S).$$

The space V is taken to be the set of real numbers $\mathbb{R}$. So, from now on, instead of taking $E = \mathbb{R}^{\mathscr{C}^{\infty}_{\mathbb{R}}(S)}$, the set E shall be taken to be $\mathscr{L}(\mathscr{C}^{\infty}_{\mathbb{R}}(S);\mathbb{R})$. In general, for a given linear normed space L, the space $\mathscr{L}(L;\mathbb{R})$ is called the "dual space" of L and will be denoted by L'; the elements in L' are called "linear forms" or "linear functionals" (however, the last term is prefered when L is a "function space"). In L' the topology will not be the topology generated by the natural norm, it will be the topology of simple convergence and this topology is called the *weak topology*.

According to Daniell's theorem, if a linear form $\mu \in \mathscr{C}^{\infty}_{\mathbb{R}}(S)'$ satisfies that:

1. for every $f \in \mathscr{C}^{\infty}_{\mathbb{R}}(S)$ which is non-negative ($f(x) \geq 0$ for *every* $x \in S$), the linear form μ assumes a positive value, that is $\mu(f) \geq 0$ (this is will be expressed as "the linear form μ is positive");
2. for every sequence (f_n) defined on $\mathscr{C}^{\infty}_{\mathbb{R}}(S)$ which converges monotonically to the constant function $0 \in \mathscr{C}^{\infty}_{\mathbb{R}}(S)$, the sequence of numbers $(\mu(f_n))$ converges to $0 \in \mathbb{R}$;

then, such linear form μ can be extended to a space $\mathscr{L}^1_{\mathbb{R}}(S,\mu)$ (the extension will also be denoted by μ) in a unique way such that the space $\mathscr{L}^1_{\mathbb{R}}(S,\mu)$ contains the set of functions $\mathbb{1}_A$, where A runs through the Borel sets of S, and the function $A \mapsto \mu(\mathbb{1}_A)$ is a positive measure on $\mathcal{B}_S$ (the value $\mu(\mathbb{1}_A)$ will, of course, be denoted by $\mu(A)$). A linear form $\mu \in \mathscr{C}^{\infty}_{\mathbb{R}}(S)'$ that have these two hypothesis will be called a "positive integral" or a "measure" (here, the term measure is restricted to "Borel positive measures" while $\mu(A) = \mu(\mathbb{1}_A)$ might be defined for sets $A \notin \mathcal{B}_S$). The subset of linear combinations of elements of $\mathscr{C}^{\infty}_{\mathbb{R}}(S)'$ that are positive integrals is a subspace and will be denoted by $\mathcal{M}_+(S)$.

In this last case, the *weak topology* on $\mathcal{M}_+(S)$ is called the *vague topology* and this topology is induced by the evalation maps: saying that "a sequence $(\mu_n)_{n\in\mathbb{N}}$ will converge *vaguely* to μ" is equivalent to saying that for each $f \in \mathscr{C}^{\infty}_{\mathbb{R}}(S)$, the sequence of real numbers $\mu_n(f)$ converges to $\mu(f)$. Finally, it is necessary to observe that the space $\mathcal{P}_S$ can be "naturally identified" with the subset $\widetilde{\mathcal{P}}_S \subset \mathcal{M}_+(S)$ of positive integrals μ such that $\mu(S) = 1$ in the following way:

1. given $P \in \mathcal{P}_S$, use Daniell's theorem to obtain an extension $P \in \mathscr{P}_S$;
2. given $P \in \widetilde{\mathcal{P}}_S$, take P to be the positive Borel measure $A \mapsto P(A)$.

‡ Definition and portmanteau theorem.

Definition (1.5.1) Let (S,d) be an arbitrary metric space and $(P_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{P}_S$. It shall be said that the sequence (P_n) "converges weakly" to the probability measure $P \in \mathcal{P}_S$ if

$$\lim_{n\to\infty} \int_S f dP_n = \int_S f dP, \quad \forall f \in \mathscr{C}^{\infty}_{\mathbb{R}}(S).$$

If random objects $X : (\Omega,\mathscr{F}) \to S$, $X_n : (\Omega_n,\mathscr{F}_n) \to S$ ($n \in \mathbb{N}$) are given, the expresion "the sequence $(X_n)_{n\in\mathbb{N}}$ converges in law to X" if $\mathcal{L}_{X_n}$ converges weakly to $\mathcal{L}_X$; the law of a random object Y will also be denoted by μ_Y, where, of course, Y is defined on the measure space $(\Omega,\mathscr{F},\mu)$. This is equivalent to saying that

$$\lim_{n\to\infty} \mathbb{E}\left(f(X_n)\right) = \mathbb{E}\left(f(X)\right), \quad \forall f \in \mathscr{C}^{\infty}_{\mathbb{R}}(S),$$

[26] A function f with values in a normed space in "bounded" if there is some $M > 0$ such that $\|f(x)\| \leq M$ for every $x \in S$.

where the expectation is taken in the corresponding space. Convergence in measure will be denoted $P_n \xrightarrow{\text{w}} P$ and convergence in distribution will be denoted by $\mathcal{L}_{X_n} \xrightarrow{\text{w}} \mathcal{L}$, $\mu_{X_n} \xrightarrow{\text{w}} \mu_X$ or by $X_n \xrightarrow{\text{D}} X$.

Remark (1.5.2) It must be observed that every *weak limit* is unique in the sense that if $P_n \xrightarrow{\text{w}} P$ and $P_n \xrightarrow{\text{w}} Q$ then $P(A) = Q(A)$ for every $A \in \mathcal{B}_S$.

Indeed, the weak limit implies at once $\displaystyle\int f dP = \int f dQ$ for every bounded continuous function $f : S \to \mathbb{R}$. But if $A \in \mathcal{O}_S$, then $\mathbb{1}_A$ is the simply and incresing limit of a sequence of functions $0 \leq f_n \leq 1$ and such that $f_n \in \mathscr{C}^\infty_{\mathbb{R}}(S)$ (see Theorem (A6.6) of Ash: [2]). Therefore, the monotone convergence theorem shows that $P(A) = Q(A)$ for every $A \in \mathcal{O}_S$. As $\mathcal{O}_S$ is a set of generators of $\mathcal{B}_S$, the two measures P and Q coincide on $\mathcal{B}_S$.

The following theorem will give several different characterizations of weak convergence and, of course, some terminology facilitates its reading.

1. The "frontier" of a set $A \subset S$ is the closed (hence, Borel) set $\partial A = \overline{A} \cap \overline{\complement_S A}$; for a probability $P \in \mathcal{P}_S$, a set A is a "continuity set with respect to P" if $P(\partial A) = 0$; for short, the phrase "P-continuity set" will often be used.

2. As in the previous discussion, given any measure μ and any function $f \in \mathscr{L}^1_{\mathbb{R}}(S, \mu)$, the number $\displaystyle\int f d\mu$ will be denoted by $\mu(f)$ or even by μf.

Theorem (1.5.3) Let (S, d) be any metric space and assume that P is an element of $\mathcal{P}_S$ and (P_n) is a sequence in $\mathcal{P}_S$. Of the following conditions, *the first implies all the other* and the second to the sixth are equivalent; additionally, *if S is separable, all conditions are equivalent.*

1. $\lim\limits_{n\to\infty} P_n = P$ with respect to ρ;
2. $P_n \xrightarrow{\text{w}} P$;
3. for every $f \in \mathscr{C}^\infty_{\mathbb{R}}(S)$ which is *uniformly continuous*, $\lim\limits_{n\to\infty} P_n(f) = P(f)$;
4. for every $F \in \mathcal{C}_S$, a *closed* set of S, $\limsup\limits_{n\to\infty} P_n(F) \leq P(F)$;
5. for every $G \in \mathcal{O}_S$, an *open* set of S, $\liminf\limits_{n\to\infty} P_n(G) \geq P(G)$;
6. for every $A \in \mathcal{B}_S$, a P-continuity set, $\lim\limits_{n\to\infty} P_n(A) = P(A)$.

Proof: note that (2. $\Rightarrow$ 3.), (4. $\Leftrightarrow$ 5.) and (4. & 5. $\Rightarrow$ 6.) follows immediately from the definitions[27]. Therefore, it suffices to show that (1. $\Rightarrow$ 2.), (3. $\Rightarrow$ 4.), (6. $\Rightarrow$ 2.) and, with the extra hypothesis that S is separable, (5. $\Rightarrow$ 1.). Each case will be done separately.

(1. $\Rightarrow$ 2.) Given $n \in \mathbb{N}$, let $\varepsilon_n = \rho(P_n, P) + \dfrac{1}{n}$; take $f \in \mathscr{C}^\infty_{\mathbb{R}}(S)$ non-negative and note that the function

$$t \mapsto P_n(f \geq t), \quad t \in \mathbb{R}$$

[27] A sketch of the proof goes as follows: 2. $\Rightarrow$ 3. is true since all f from case 3. also belong to those of case 2.; the condition 4. $\Leftrightarrow$ 5. follows since a set is closed if and only if its complement is open, then the formula $P(F) = 1 - P(G)$, if F and G are complementary, is used; if both 4. and 5. hold, note that it is true $P(A) = P\left(\mathring{A}\right) = P\left(\overline{A}\right)$ as A is a P-continuity set, and hence, the two cases 4. and 5. are used to conclude that

$$P(A) = P\left(\mathring{A}\right) \leq \liminf_{n\to\infty} P_n\left(\mathring{A}\right) \leq \limsup_{n\to\infty} P_n\left(\overline{A}\right) \leq P\left(\overline{A}\right) = P(A).$$

is non-increasing and with support[28] in $[0, \infty)$; hence, such function is Borel. Now, use the Lebesgue-Fubini theorem to note that

$$\int\limits_{[0,\infty)} P_n(f \geq t)dt = \int\limits_{[0,\infty)} \int\limits_{\{f\geq t\}} dP_n dt = \int\limits_{\{f\geq 0\}} \int\limits_{[0,f(\omega))} dt dP_n(\omega) = \int\limits_{\{f\geq 0\}} f(\omega) dP_n(\omega) = P_n(f)$$

since $f \geq 0$ at every point of S. The first expresion for $P_n(f)$ gives,

$$\int\limits_{[0,\infty)} P_n(f \geq t)dt = \int\limits_0^{\|f\|} P_n(f \geq t)dt \leq \int\limits_0^{\|f\|} P\left(\{f \geq t\}^{\varepsilon_n}\right) dt + \varepsilon_n \|f\|.$$

Let $n \to \infty$; use that $P\left(\{f \geq t\}^{\varepsilon_n}\right) \to P(f \geq t)$ (by the monotonicity of P) and the closureness of $\{f \geq t\}$ (recall that f is continuous) to obtain

$$\limsup_{n\to\infty} \int\limits_S f dP_n \leq \int\limits_S f dP.$$

This last inequality is true for every $f \in \mathscr{C}^\infty_{\mathbb{R}_+}(S)$; so, given $f \in \mathscr{C}^\infty_{\mathbb{R}}(S)$, the functions $\|f\| + f, \|f\| - f \geq 0$ and, hence,

$$\limsup_{n\to\infty} \int\limits_S (\|f\| + f) dP_n \leq \int\limits_S (\|f\| + f) dP$$

and

$$\limsup_{n\to\infty} \int\limits_S (\|f\| - f) dP_n \leq \int\limits_S (\|f\| - f) dP.$$

Observe now that

$$\int\limits_S (c + g) dP_n = c + P_n(g),$$

for every constant $c \in \mathbb{R}$, and, in particular for $c = \|f\|$; this observation gives

$$\|f\| + \limsup_{n\to\infty} \int\limits_S f dP_n \leq \|f\| + \int\limits_S dP$$

and

$$\|f\| - \liminf_{n\to\infty} \int\limits_S f dP_n \leq \|f\| - \int\limits_S f dP.$$

Therefore, $P_n \xrightarrow{\text{w}} P$ as was to be shown.

(3. ⇒ 4.) Given any closed set $F \in \mathcal{C}_S$ and any $\varepsilon > 0$, the function

$$f_\varepsilon(x) = \max\left(1 - \frac{d(x, F)}{\varepsilon}, 0\right)$$

is uniformly continuous: indeed, if φ and ψ are uniformly continuous, the same is true for $\max(\varphi, \psi)$ and for $\min(\varphi, \psi)$ and, hence, it suffices to show that $x \mapsto 1 - \dfrac{d(x, F)}{\varepsilon}$ is uniformly continuous but this is obvious since

[28] The "support" of a function is the *complement* of the biggest open set where the function is identically zero; that is, the support is the *complement* of the set of $x \in S$ such that there is a ball centred at x and for which f restricted to such ball is identically zero.

the funtion $x \mapsto d(x,F)$ is uniformly continuous[29]. The function f_ε is uniformly continuous and $0 \le f_\varepsilon \le 1$, so the hypothesis allows concluding that

$$\lim_{n\to\infty} \int_S f_\varepsilon dP_n = \int_S f_\varepsilon dP.$$

Note that if $x \in F$, then $f_\varepsilon(x) = 1$ so $\mathbb{1}_F \le f_\varepsilon$ and, hence

$$\limsup_{n\to\infty} P_n(F) \le \lim_{n\to\infty} \int_S f_\varepsilon dP_n = \int_S f_\varepsilon dP.$$

If $\varepsilon \downarrow 0$, then $f_\varepsilon(x) \downarrow \mathbb{1}_F$ as F is closed; Lebesgue's convergence theorem then yields

$$\limsup_{n\to\infty} P_n(F) \le P(F).$$

(6. ⇒ 2.) Let $f \in \mathscr{C}^\infty_{\mathbb{R}_+}(S)$ (so, f is non-negative). Then $\partial\{f \ge t\} \subset \{f = t\}$, and therefore the set $\{f \ge t\}$ is a P-continuity for every t, with the exception, perhaps, of a countable set. Therefore,

$$\lim_{n\to\infty} \int_S f dP_n = \lim_{n\to\infty} \int_0^{\|f\|} P_n(f \ge t)dt = \int_0^{\|f\|} P(f \ge t)dt = \int_S f dP.$$

Now, for $f \in \mathscr{C}^\infty_{\mathbb{R}}(S)$, not necessarily non-negative, write $f = f^+ - f^-$, where $f^+ = \frac{1}{2}(f + |f|)$ and $f^- = \frac{1}{2}(-f + |f|)$ belong to $\mathscr{C}^\infty_{\mathbb{R}_+}(S)$. The previous case shows that

$$\lim_{n\to\infty} P_n(f) = \lim_{n\to\infty} \left(P_n(f^+) - P_n(f^-)\right) = P(f^+) - P(f^-) = P(f),$$

which is 2.

(5. ⇒ 1.) For this implication, we are to assume that S is separable. As in the proof of (1.2.3), for given $\varepsilon > 0$ there exists a measurable partition (E_n) of Borels of S such that the diameter of each of them is $< \frac{\varepsilon}{2}$. Let N be the minimum integer such that $P\left(\bigcup_{i=1}^N E_i\right) > 1 - \frac{\varepsilon}{2}$ and let $\mathcal{G}$ be the finite set of subsets of S of the form

$$\left(\bigcup_{i\in I} E_i\right)^{\frac{\varepsilon}{2}}, \quad I \subset \{1,\dots,N\}.$$

Note that all the sets in $\mathcal{G}$ are open and, as there are only finitely many of them, the assumption allows obtaining an integer number $n_0 \in \mathbb{N}$ such that

$$P(G) \le P_n(G) + \frac{\varepsilon}{2}$$

for every $G \in \mathcal{G}$ and every $n \ge n_0$. Now, assume that $F \in \mathcal{C}_S$ and let F_0 be the union of the sets E_i such that $E_i \cap F \neq \varnothing$ for $i = 1,\dots,N$. It is then clear that $F^{\frac{\varepsilon}{2}} \in \mathcal{G}$ and

$$P(F) \le P(F_0) + \frac{\varepsilon}{2} \le P_n\left(F_0^{\frac{\varepsilon}{2}}\right) + \varepsilon \le P_n\left(F_0^{\varepsilon}\right) + \varepsilon,$$

no matter what $n \ge n_0$ is. This last statement is exactly that $\lim_{n\to\infty} P_n = P$ with respect to ρ. ■

[29] To see this uniform continuity, it suffices to show that $|d(x,F) - d(y,F)| \le d(x,y)$ for every pair $x, y \in S$; for any such pair, note that $d(x,F) \le d(x,z) \le d(x,y) + d(y,z)$ for any $z \in F$. Now, in the last term take the infimum over $z \in F$ to obtain $d(x,F) - d(y,F) \le d(x,y)$; similarly, $d(y,F) - d(x,F) \le d(y,x) (= d(x,y))$ and the conclusion follows.

Chapter 2

✠ Weak convergence in measure.

For the rest of this thesis, let (S, d) be a metric space and let $(\mathcal{P}_S, \rho)$ be the Prohorov space associated to it (see (1.1.1)).

The objective of this thesis is to propose a definition of "weak convergence in measure" in the sense of (1.5.1) but taking into account that now the measures P_n are random, so, some type of random convergence must be employed, this will be done in this chapter. Of course, such a convergence will be the one described in the Introduction. The author must clarify that *the sufficiency of Prohorov's theorem in measure does not hold.*

§2.1. Basic definitions and properties.

Recall the notation mentioned at the begin of this thesis. The spaces:

$$\mathscr{K}_{\mathbb{R}}(S;K), \quad \mathscr{B}_{\mathbb{R}}(S), \quad \mathscr{C}_{\mathbb{R}}(S) \quad \text{and} \quad \mathscr{L}^p_{\mathbb{R}}(S,\mu)$$

are normed and complete[1]; the following inclusions hold

$$\mathscr{K}_{\mathbb{R}}(S;K) \subset \mathscr{C}^0_{\mathbb{R}}(S) \subset \mathscr{C}^\infty_{\mathbb{R}}(S) \subset \mathscr{C}_{\mathbb{R}}(S) \subset \mathscr{B}_{\mathbb{R}}(S)$$

and

$$\mathscr{M}^\infty_{\mathbb{R}}(S) \subset \mathscr{M}_{\mathbb{R}}(S), \quad \mathscr{M}^\infty_{\mathbb{R}}(S) \subset \mathscr{L}^\infty_{\mathbb{R}}(S,\mathbb{R});$$

if, aditionally, μ is finite,

$$\mathscr{L}^q_{\mathbb{R}}(S,\mathbb{R}) \subset \mathscr{L}^p_{\mathbb{R}}(S,\mathbb{R})$$

for any $1 \leq p \leq q \leq \infty$. In everywhere that preceded, the symbol $\mathbb{R}$ might be replaced by $\mathbb{C}$ if the functions takes values in $\mathbb{C}$ or, more generally, by F, where F is *any* subset of an arbitrary *normed vector space*[2] or a subset of the *extended real line* $\overline{\mathbb{R}} = [-\infty, +\infty]$.

This is a first result that seems to be very useful hereinafter since it will answer questions about the mesurability of the functions to be considered.

Theorem (2.1.1) For every $f \in \mathscr{C}^\infty_{\mathbb{R}}(S)$, define the "evaluation function" $e_f : \mathcal{P}_S \to \mathbb{R}$ "associated with" or "based at" f to be the function e_f given by

$$e_f(P) = Pf = \int_S f dP.$$

Then,

[1] In fact, the space $\mathscr{L}^p_{\mathbb{R}}(S,\mu)$ is not a normed space, but it is always identified with one via the equivalence relation: "$f \sim g$ if and only if $\mu(|f-g|>0)=0$"; the space obtained with this relation is a quotien space normed complete vector space, denoted by $\mathbf{L}^p_{\mathbb{R}}(S,\mu)$

[2] A function $f : S \to V$, where V is a normed vector space, is p-integrable with respect to μ if f is measurable and $\mu(\|f\|^p)$ is finite.

1. every evaluation map based at some $f \in \mathscr{C}_{\mathbb{R}}^{\infty}(S)$ is continuous;
2. for every $f \in \mathscr{M}_{\mathbb{R}}^{\infty}(S)$, the evaluation function associated with f is measurable;
3. if $\omega \to P^{\omega}$ defines a function $P : (\Omega, \mathscr{F}) \to (\mathcal{P}_S, \mathcal{B}_{\mathcal{P}_S})$, then, for every $f \in \mathscr{M}_{\mathbb{R}}^{\infty}(S)$, the function $Pf : \Omega \to \mathbb{R}$, given by
$$(Pf)(\omega) = P^{\omega} f = \int_S f(s) P^{\omega}(ds),$$
is a random variable.

Proof: the proof is divided into several steps.

(a) As $\mathcal{P}_S$ is a metric space, it suffices to show that if P_n converges to P, then $e_f(P_n)$ converges to $e_f(P)$. But if P_n converges to P, then $P_n \xrightarrow{w} P$ (see (1.5.3)) and then $e_f(P_n) = P_n f$ converges to $Pf = e_f(P)$ by definition of weak convergence. So, e_f is continuous.

(b) Now, the following claim will be proved: if $f_n \in \mathscr{M}_{\mathbb{R}}^{\infty}(S)$, for $n \in \mathbb{N}$, and the f_n converge simply to f and are uniformly bounded, then e_{f_n} converges simply to e_f. Indeed, let $P \in \mathcal{P}_S$ and note that, by the uniform boundedness of the f_n, the bounded convergence theorem applies th $e_{f_n}(P) = P(f_n)$. The claim follows.

(c) The following result is Theorem (A6.6) of *Real Analysis and Probability* by Ash [2]: if $G \subset S$ is open, then $\mathbb{1}_G$ is the limit of an increasing sequence of continuous functions $(f_n)_{n\in\mathbb{N}}$ bounded by 1. Hence, the previous claim shows that every evaluation map associated with the indicator of a open set is measurable.

(d) Let $\mathscr{H}$ the set of functions in $\mathscr{M}_{\mathbb{R}}^{\infty}(S)$ such that the evaluation maps associated are measurable. Clearly, $\mathscr{H}$ is a vector space since the following is inmediate $e_f + a e_g = e_{f+ag}$. By the previous steps, $e_{\mathbb{1}_G} \in \mathscr{H}$ for all the open sets G and $\mathscr{H}$ is closed with respect to increasing limits of uniformly bounded sequences of functions. Therefore, the Theorem of Dynkin classes[3] shows that $\mathbb{1}_G \in \mathscr{H}$ for all measurable sets $G \subset S$.

(e) If $f \in \mathscr{M}_{\mathbb{R}}^{\infty}(S)$ is positive, then e_f is measurable. Such a function is the pointwise limit of an increasing sequence of simple functions. The previous steps prove the claim.

(f) All $f \in \mathscr{M}_{\mathbb{R}}^{\infty}(S)$ are such that e_f is measurable. For such an f, it can be written $f = f^+ - f^-$ and f^+, f^- are positive and measurable, the claim follows from the previous step.

(g) Finally, if P is a random object as in the hypothesis, then the function Pf is the composition of the two measurable maps e_f and P and, therefore, is a random variable (in fact, if P is continuous, so is Pf).

This concludes the proof. ■

Definition (2.1.2) Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and (S, d) be a metric space and, as in the previous chapter, let $(\mathcal{P}_S, \rho)$ the space of probability measures on $\mathcal{B}_S$. Any *measurable function* $P : (\Omega, \mathscr{F}) \to (\mathcal{P}_S, \mathcal{B}_{\mathcal{P}_S})$ will receive the name "random probability measure"; hence, random probability measures are random objects with values in the space of probability measures. Assume that $(P_n)_{n\in\mathbb{N}}$ is a sequence of random probability measures $(\Omega, \mathscr{F}) \to (\mathcal{P}_S, \mathcal{B}_{\mathcal{P}_S})$ and let P be another such random measure. It will be said that "the sequence (P_n) converges weakly in measure" to P if for all $f \in \mathscr{C}_{\mathbb{R}}^{\infty}(S)$, the sequence of random variables $(P_n f)_{n\in\mathbb{N}}$ converges in $\mathbb{P}$-probability to Pf. The notation to be employed is $P_n \xrightarrow{\text{wim}} P$. Instead of writing $\omega \mapsto P_n(\omega)$, here it will be written $\omega \mapsto P_n^{\omega}$ and similarly without the subscript n. Therefore, $P_n(f)$ is the random variable

$$\omega \mapsto \int f(t) dP^{\omega}(t)$$

[3]See Theorem 4.1.2 of Ash [2]

The sequence $(P_n)_{n\in\mathbb{N}}$ will be called "tight in measure" if for every pair of positive numbers ε and δ, there exists a *compact* subset K of S such that if $n \in \mathbb{N}$, then there exists an event $\mathsf{F}_n \in \mathscr{F}$ that satisfies $\mathbb{P}(\mathsf{F}_n) \geq 1-\varepsilon$ and $\omega \in \mathsf{F}_n \Rightarrow P_n^\omega(K) \geq 1-\delta$.

As a consequence of (2.1.1), the following remark can be made.

Remark (2.1.3) The definition of tightness in measure is equivalent to the following: for all $\delta, \varepsilon > 0$ there exists a compact $K \subset S$ such that $n \in \mathbb{N} \Rightarrow \mathbb{P}(P_n \mathbb{1}_K > 1-\delta) > 1-\varepsilon$. That is, *with arbitrarily large probability, there is a uniform compact K which has a high chance of happening for all the P_n.*

The equivalence is since $\{P_n\mathbb{1}_K > 1-\delta\} \in \mathscr{F}$ as $P_n\mathbb{1}_K$ is a random variable, so one may take this set to be F_n.

The following result gives hope for Prohorov's theorem in measure (see (1.4.3)) in this context to be true. It says that all random probability measures on a complete and separable metric space are tight in measure.

Theorem (2.1.4) If P is a random probability measure $(\Omega, \mathscr{F}) \to (\mathcal{P}_S, \mathcal{B}_{\mathcal{P}_S})$ and S is complete and separable then P is tight in measure.

Proof: the idea of Theorem 1.3 in Billingsley's book will be followed (see [4]). As S is a separable metric space, there exists a sequence $(x_n)_{n\in\mathbb{N}}$ whose range is dense in S. For $k \in \mathbb{N}$, the family $\left(\mathrm{B}\left(x_n; \frac{1}{k}\right)\right)_{n\in\mathbb{N}}$ is an open cover of S. By the monotone convergence theorem,

$$\lim_{N\to\infty} P^\omega\left(\bigcup_{j=1}^{N} \mathrm{B}\left(x_j; \frac{1}{k}\right)\right) = P^\omega(S) = 1 \quad \text{for } \textit{all} \quad \omega \in \Omega.$$

Note also that the sets

$$\mathsf{E}_N = \left\{\omega \in \Omega \middle| P^\omega\left(\bigcup_{j=1}^{N} \mathrm{B}\left(x_j; \frac{1}{k}\right)\right) > 1 - \frac{\delta}{2^k}\right\}$$

are increasing in N. Hence, by monotonicity of $\mathbb{P}$, if $\delta, \varepsilon > 0$ are given and $k \in \mathbb{N}$, it possible to choose (take the minimum) an integer N_k in such a way that

$$n \geq N_k \Rightarrow \mathbb{P}(\mathsf{E}_n) > 1 - \frac{\varepsilon}{2^k}.$$

Let $K \subset S$ be the *closure* of the set $\displaystyle\bigcap_{k=1}^{\infty}\bigcup_{j=1}^{N_k} \mathrm{B}\left(x_j; \frac{1}{k}\right)$. Obviously, by completeness of S, K is a complete metric space. But K is also totally bounded since if $\eta > 0$, then it is possible to choose an integer k such that $\frac{1}{k} < \eta$ and K is contained in N_k balls of radii less than η. So, K is a compact set. It is also clear that $P^\omega(K) \geq P^\omega\left(\bigcap_{k=1}^{\infty}\bigcup_{j=1}^{N_k} \mathrm{B}\left(x_j; \frac{1}{k}\right)\right)$ for *all* $\omega \in \Omega$, so, if $\mathsf{A} = \left\{\omega \in \Omega \middle| P^\omega\left(\bigcap_{k=1}^{\infty}\bigcup_{j=1}^{N_k} \mathrm{B}\left(x_j; \frac{1}{k}\right) > 1-\delta\right)\right\}$ and

$$\mathbb{P}(\mathsf{A}) > 1-\varepsilon$$

then

$$\mathbb{P}(P\mathbb{1}_K > 1-\delta) > 1-\varepsilon,$$

which is the wanted tightness in measure.

Let $\mathsf{D}_k = \left\{\omega \in \Omega \middle| P^\omega\left(\bigcap_{m=1}^{\infty}\bigcup_{j=1}^{N_m} \mathrm{B}\left(x_j; \frac{1}{m}\right) > 1 - \frac{\delta}{2^k}\right)\right\}$. Then, $\mathbb{P}(\mathsf{D}_k) > 1 - \frac{\varepsilon}{2^k}$ and, hence,

$$\mathbb{P}\left(\bigcap_{k=1}^{\infty} \mathsf{D}_k\right) > 1-\varepsilon.$$

So, it suffices to show that

$$\omega \in \bigcap_{k=1}^{\infty} \mathsf{D}_k \Rightarrow P^{\omega}\left(\bigcap_{m=1}^{\infty}\bigcup_{j=1}^{N_m} \mathrm{B}\left(x_j; \frac{1}{m}\right)\right) > 1-\delta,$$

which is immediate from the definition of the D_k. ■

§2.2. Some results obtained for weak convergence in measure.

The following seems to be the correct definition of a "continuity set in measure" in the case of random probability measures.

Definition (2.2.1) It will be said that a set $A \in \mathcal{B}_S$ is a "continuity set in measure" for the random probability measure $P : (\Omega, \mathscr{F}, \mathbb{P}) \to (\mathcal{P}_S, \mathcal{B}_{\mathcal{P}_S})$ (in short form, a "P-continuity set in measure"), for ever pair $\delta > 0$ and $\varepsilon > 0$,

$$\mathbb{P}\left(P\mathbb{1}_{\mathring{A}} \geq P\mathbb{1}_A - \delta, P\mathbb{1}_{\overline{A}} \leq P\mathbb{1}_A + \delta\right) \geq 1-\varepsilon.$$

Remark (2.2.2) Compare this definition with the one given just *after* (1.5.1); observe that in the non-random case, a continuity set for P is a Borel A such that $P(\partial A) = 0$ so, in this case *with arbitrarily large probability, the frontier of A has small chance to happen*. Returning to the random case; assume A is a P-continuity set in measure and let δ and ε be as in (2.2.1). One notes that, with probability at least $1-\varepsilon$,

$$P\mathbb{1}_A - P\mathbb{1}_{\mathring{A}} = P\mathbb{1}_{\partial A \cap A} \leq \delta$$

and that

$$P\mathbb{1}_{\overline{A}} - P\mathbb{1}_A = P\mathbb{1}_{\partial A \cap \complement A} \leq \delta,$$

so, both inequalities give $P\mathbb{1}_{\partial A} \leq 2\delta$ with probability at least $1-\varepsilon$. Similarly, assume $A \in \mathcal{B}_S$ is such that for every pair $\varepsilon, \delta > 0$, one has $\mathbb{P}\left(P\mathbb{1}_{\partial A} \leq \delta\right) \geq 1-\varepsilon$. Then

$$P(\mathbb{1}_{\partial A}) = P\mathbb{1}_{\partial A \cap \complement A} + P\mathbb{1}_{\partial A \cap \complement A} \leq \delta,$$

and, hence, as they are non-negative valued functions,

$$P\mathbb{1}_A - P\mathbb{1}_{\mathring{A}} = P\mathbb{1}_{\partial A \cap A} \leq \delta \quad \text{and} \quad P\mathbb{1}_{\overline{A}} - P\mathbb{1}_A = P\mathbb{1}_{\partial A \cap \complement A} \leq \delta.$$

Therefore, what was proven here is that the definition of continuity set in measure is equivalent to: *for every pair $\delta > 0$ and $\varepsilon > 0$,*

$$\mathbb{P}\left(P\mathbb{1}_{\partial A} \leq \delta\right) \geq 1-\varepsilon;$$

and as with the previous definitions "in measure": *a set A is a P-continuity set in measure if it is approximately a continuity set uniformly for all the P^{ω} with high probability.*

The following result was derived in the search for a proof of Prohorov's theorem in the case of weak convergence in measure. The author can say that this result is refined, as it generalizes a classic result of measure theory: for every finite measure and any bounded continuous function, the image measure can only have countably many points of positive mass. This result generalizes as the previous definitions: for a given random probability measure and a continuous and bounded function, the random image measure have an *uniform* and *countable* set of points that have positive mass. The formal idea follows.

Theorem (2.2.3) Let P be a random probability measure $(\Omega, \mathscr{F}) \to (\mathcal{P}_S, \mathcal{B}_{\mathcal{P}_S})$ and $g \in \mathscr{C}_{\mathbb{R}}^{\infty}(S)$. Then, there exist a countable set $D \subset \mathbb{R}$ such that

$$t \notin D \Rightarrow \mathbb{P}\left(P\mathbb{1}_{\{g=t\}} = 0\right) = 1.$$

In other (more meaningful) terms, if gP denotes the random probability measure $(\omega, A) \mapsto P^{\omega}\mathbb{1}_{\{g \in A\}}$, then, for almost every point ω with respect to $\mathbb{P}$, the (non-random) measure gP^{ω} has no atoms inside D (so, D contains all the atoms of all the gP^{ω}, which is surprising since ω might run over an immense set Ω).

Proof: let $\varepsilon > 0$ and let D_ε be the set of $t \in \mathbb{R}$ such that

$$\mathbb{P}\left(P\mathbb{1}_{\{g=t\}} > \varepsilon\right) > \varepsilon$$

It is claimed that D_ε is finite. Otherwise, there would be a sequence $(t_n)_{n\in\mathbb{N}}$ of distinct elements of D_ε. Consider the set $\mathsf{D} = \{P\mathbb{1}_{\{g=t_n\}} > \varepsilon,\ \text{i.o.}\} \in \mathscr{F}$, which has probability at least[4] ε. As D is not empty, there is an $\omega \in \mathsf{D}$ and for this ω there is the probability measure gP^ω, which is the image of P^ω by g, such that it has an infinite number of distinct points that have mass of at least ε, which is absurd. Therefore, D_ε is finite.

Define $D = \bigcup_{n=1}^{\infty} D_{\frac{1}{n}}$. Note that D is countable and that if $t \notin D$, then

$$\mathbb{P}\left(P\mathbb{1}_{\{g=t\}} > \frac{1}{n}\right) \leq \frac{1}{n} \quad \text{for every } n \in \mathbb{N}.$$

Let $n \to \infty$ to obtain[5] $t \notin D \Rightarrow \mathbb{P}\left(P\mathbb{1}_{\{g=t\}} > 0\right) = 0.$ ■

Remark (2.2.4) Recall the basic result of Lebesgue: *for any function $f \in \mathscr{M}_{\mathbb{R}_+}(S)$, there is a sequence of simple measurable and bounded functions (f_n) such that f_n converges increasing and simply to f.* In fact, the proof will be relevant shortly and as it is rather short it is sketched immediately. Define $h_n = \frac{1}{2^n}\sum_{k=1}^{n2^n} \mathbb{1}_{\{f > k2^{-n}\}}$; the 2^{-n} is so as to ensure that the next partition is contained in the previous one and, hence, the sequence will be increasing. Note that $\frac{k-1}{2^n} \leq h_n \leq \frac{k}{2^n}$ if $\frac{k}{2^n} \leq f \leq \frac{k+1}{2^n}$, so the h_n increase to f.

Now, assume a probability measure μ is given and note that the extremities of the intervals $\left[\frac{k-1}{2^n}, \frac{k}{2^n}\right)$ might be modified *just a little* to make them points of continuity for μ; hence, Lebesgue's result can be improved to: *for any function $f \in \mathscr{M}_{\mathbb{R}_+}(S)$, there is a sequence of simple measurable and bounded functions (f_n) such that f_n converges increasing and simply to f and if $f_n = \sum_{k=1}^{m} c_k \mathbb{1}_{A_k}$, then A_k is a continuity set for μ.*

The next result states that the previous example can also be constructed for continuity set in measures.

Theorem (2.2.5) Let P be a random probability measure $(\Omega, \mathscr{F}) \to (\mathcal{P}_S, \mathcal{B}_{\mathcal{P}_S})$ and $g \in \mathscr{C}^{\infty}_{\mathbb{R}_+}(S)$ (note that g is assumed non-negative valued). Then, there exist a sequence $(g_m)_{m\in\mathbb{N}}$ of *simple functions* such that the following properties are verified:

1. for all $x \in S$, $\lim_{m\to\infty} g_m(x) = \frac{g(x)}{\|g\|}$;

---

[4] By definition $\{A_n,\ \text{i.o.}\} = \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k$, so its probability is at least $\limsup_{n\to\infty} \mathbb{P}(A_n)$. In the context of the theorem, each A_n has probability at least ε.

[5] There is some trickery in this step: take $\varepsilon > 0$ and let n_0 be so big as for to let $\frac{1}{n_0} < \varepsilon$; then for $n \geq n_0$ it can be deduced that

$$\mathbb{P}\left(P\mathbb{1}_{\{g=t\}} > \frac{1}{n}\right) \leq \varepsilon;$$

now, the sets $\left\{P\mathbb{1}_{\{g=t\}} > \frac{1}{n}\right\}$ are increasing in n, so the monotonicity of $\mathbb{P}$ gives that

$$\mathbb{P}\left(P\mathbb{1}_{\{g=t\}} > 0\right) = \lim_{n\to\infty} \mathbb{P}\left(P\mathbb{1}_{\{g=t\}} > \frac{1}{n}\right) \leq \varepsilon;$$

letting $\varepsilon \downarrow 0$ gives the desired claim.

2. for all $m \in \mathbb{N}$, one has $0 \le g_m \le \frac{g}{\|g\|} \le g_m + \frac{3}{2^m}$;

3. if $g_m = \sum_{k=1}^{n} c_k \mathbb{1}_{A_k}$, then each A_k is a P-continuity set in measure.

Proof: let $B_{k,m} = \left\{ \frac{k-1}{2^m} \le \frac{g}{\|g\|} < \frac{k}{2^m} \right\} \in \mathcal{B}_S$ for $k = 1, \dots, 2^m + 1$. Then,

$$\partial B_{k,m} \subset \left\{ \frac{g}{\|g\|} = \frac{k-1}{2^m} \right\} \cup \left\{ \frac{g}{\|g\|} = \frac{k}{2^m} \right\}$$

for $k = 1, \dots, 2^m + 1$. By (2.2.3), it is possible to choose points $s_{k,m}$ for $k = 0, \dots, 2^m + 1$, such that:

1. $s_{0,m} < 0 < s_{k,m} < s_{k+1,m}$ for $k \ge 1$ and $s_{2^m+1,m} > 1$;
2. $\left\{ \frac{g}{\|g\|} = s_{k,m} \right\}$ is a P-continuity set in measure for all k;
3. $s_{0,m} \uparrow 0$ and $s_{2^m+1,m} \downarrow 1$; and
4. $\left| \frac{k}{2^m} - s_{k,m} \right| < \frac{1}{2^m}$ for all k.

Define $c_{k,m} = s_{k,m}$ for all $k \ge 1$ and m, $c_{0,m} = 0$ and let

$$A_{k,m} = \left\{ s_{k-1,m} \le \frac{g}{\|g\|} < s_{k,m} \right\} \in \mathcal{B}_S.$$

Let $g_m = \sum_{k=1}^{2^m+1} c_{k-1,m} \mathbb{1}_{A_{k,m}}$. Obviously, for all $m \in \mathbb{N}$, it is true that $0 \le g_m \le \frac{g}{\|g\|} \le g_m + \frac{3}{2^m}$ and that each of the sets $A_{k,m}$ is a continuity set in measure for P. If $x \in S$, then there exists a sequence $(k(x,m))_{m \in \mathbb{N}}$ such that

$$\frac{k(x,m) - 1}{2^m} \uparrow \frac{g(x)}{\|g\|} \quad \text{and} \quad \frac{k(x,m)}{2^m} \downarrow \frac{g(x)}{\|g\|}.$$

Then,

$$\begin{aligned} \frac{k(x,m) - 3}{2^m} &< s_{k(x,m)-1,m} < \frac{k(x,m) - 1}{2^m} \\ &< s_{k(x,m),m} < \frac{k(x,m) + 1}{2^m} \\ &< s_{k(x,m)+1,m} < \frac{k(x,m) + 3}{2^m}, \end{aligned}$$

and all these numbers converge (as $m \to \infty$) to $\frac{g(x)}{\|g\|}$. This proves that $g_m(x)$ converges to $\frac{g(x)}{\|g\|}$. As $x \in S$ was arbitrary, the convergence is on all of S. ■

§2.3. Portmanteau theorem in measure.

Now it is possible to state and give a proof of the portmanteau theorem in weak convergence in measure context.

Theorem (2.3.1) Let $P_n : (\Omega, \mathscr{F}, \mathbb{P}) \to (\mathcal{P}_S, \mathcal{B}_{\mathcal{P}_S})$ be a sequence of random probability measures and P another such. The following five conditions are equivalent.

(1) $P_n \xrightarrow{\text{wim}} P$;

(2) $P_n f \xrightarrow{\mathbb{P}} Pf$ for every $f \in \mathscr{C}_{\mathbb{R}}^{\infty}(S)$ that are uniformly continuous;

(3) for every closed $F \subset S$ and $\varepsilon, \delta > 0$ there exists an $N \in \mathbb{N}$ such that

$$n \geq N \Rightarrow \mathbb{P}\left(P_n \mathbb{1}_F > P\mathbb{1}_F + \delta\right) < \varepsilon;$$

(4) for every open $G \subset S$ and every pair $\varepsilon, \delta > 0$ there exists an $N \in \mathbb{N}$ such that

$$n \geq N \Rightarrow \mathbb{P}\left(P_n \mathbb{1}_G < P\mathbb{1}_G - \delta\right) < \varepsilon;$$

(5) for every continuity set in measure $A \in \mathcal{B}_S$ of P and every pair $\varepsilon, \delta > 0$ there exists an $N \in \mathbb{N}$ such that

$$n \geq N \Rightarrow \mathbb{P}\left(|P_n \mathbb{1}_A - P\mathbb{1}_A| > \delta\right) < \varepsilon,$$

that is, for every continuity set in measure A of P,

$$P_n \mathbb{1}_A \xrightarrow{\mathbb{P}} P\mathbb{1}_A.$$

Proof: obviously $(1) \Rightarrow (2)$.

It will be proved that $(2) \Rightarrow (3)$. Let $F \subset S$ be closed and $\eta > 0$. Set $f_\eta \in \mathscr{C}_{\mathbb{R}}^{\infty}(S)$ by

$$f_\eta(x) = \max\left(1 - \frac{d(x,F)}{\eta}, 0\right).$$

Then, f_η is uniformly continuous (see the *proof* of (1.5.3)) and, note that for every $x \in S$, $f_\eta(x)$, which is non-negative and bounded by 1, will *decrease* to $\mathbb{1}_F(x)$. The hypothesis (2) yields, $P_n f_\eta \xrightarrow[n\to\infty]{\mathbb{P}} Pf_\eta$. It is also true that

$$P_n \mathbb{1}_F \leq P_n f_\eta \quad \text{and} \quad \lim_{\eta \downarrow 0} Pf_\eta = P\mathbb{1}_F \quad \text{simply on } \textit{all} \quad \Omega,$$

where the convergence is obtained by the bounded convergence theorem applied to each of the P^ω for $\omega \in \Omega$. Let $\varepsilon > 0$ and $\delta > 0$. By the convergence in probability above, there exists an $N \in \mathbb{N}$ such that

$$n \geq N \Rightarrow \mathbb{P}\left(|P_n f_\eta - Pf_\eta| > \delta\right) < \varepsilon$$

and, since the sets $\{Pf_\eta - P\mathbb{1}_F > \delta\} = \{|Pf_\eta - P\mathbb{1}_F| > \delta\}$ are decreasing with η, there is an $\eta_0 > 0$ such that

$$0 < \eta \leq \eta_0 \Rightarrow \mathbb{P}\left(|Pf_\eta - P\mathbb{1}_F| > \delta\right) < \varepsilon.$$

For ω in a set of probability at least $1 - 2\varepsilon$,

$$|P_n^\omega f_\eta - P^\omega f_\eta| \leq \delta \quad \text{and} \quad |P^\omega f_\eta - P^\omega \mathbb{1}_F| \leq \delta,$$

so, for such ω,

$$P_n^\omega \mathbb{1}_F - P^\omega \mathbb{1}_F \leq P_n^\omega f_\eta - P^\omega f_\eta + \delta \leq 2\delta,$$

which gives

$$\mathbb{P}\left(P_n \mathbb{1}_F > P\mathbb{1}_F + 2\delta\right) < 2\varepsilon.$$

This proves (3).

Now it will be shown that (3) $\Leftrightarrow$ (4). Let F be a closed set of S and $G = S - F$, which is open. Then, for any $\delta > 0$,

$$\mathbb{P}\left(P_n \mathbb{1}_G < P\mathbb{1}_G - \delta\right) = \mathbb{P}\left(P_n \mathbb{1}_F > P\mathbb{1}_F + \delta\right),$$

which proves the equivalence.

Together, (3) and (4) imply (5). Let $A \in \mathcal{B}_S$ be a continuity set in measure for P. Then, there exists an $N \in \mathbb{N}$ such that, setting $\mathsf{R}_n = \left\{\omega \in \Omega \middle| P_n^\omega\left(\overline{A}\right) > P^\omega\left(\overline{A}\right) + \frac{\delta}{2}\right\}$,

$$n \geq N \Rightarrow \mathbb{P}\left(\mathsf{R}_n\right) \leq \frac{\varepsilon}{4}$$

and setting $\mathsf{S}_n = \left\{\omega \in \Omega \middle| P_n^\omega\left(\mathring{A}\right) < P^\omega\left(\mathring{A}\right) - \frac{\delta}{2}\right\}$

$$n \geq N \Rightarrow \mathbb{P}\left(\mathsf{S}_n\right) \leq \frac{\varepsilon}{4}.$$

Therefore

$$n \geq N \Rightarrow \mathbb{P}\left(\complement\mathsf{R}_n \cap \complement\mathsf{S}_n\right) \geq 1 - \frac{\varepsilon}{2}.$$

Thus, using (2.2.1) it is readily seen that, setting $\complement T_N = \left\{\omega \in \Omega \middle| P^\omega\left(\mathring{A}\right) \geq P(A) - \frac{\delta}{2}, P^\omega\left(\overline{A}\right) \leq P^\omega(A) + \frac{\delta}{2}\right\}$, one has $\mathbb{P}\left(\complement\mathsf{T}_n\right) \geq 1 - \frac{\varepsilon}{2}$. Thus, $\mathbb{P}\left(\complement\mathsf{R}_n \cap \complement\mathsf{S}_n \cap \complement\mathsf{T}_n\right) \geq 1 - \varepsilon$; that is, for $n \geq N$ and $\omega \in \complement\mathsf{R}_n \cap \complement\mathsf{S}_n \cap \complement\mathsf{T}_n$, which is a set of probability at least $1 - \varepsilon$,

$$P_n^\omega(A) \geq P_n^\omega\left(\mathring{A}\right) \geq P^\omega\left(\mathring{A}\right) - \frac{\delta}{2} \geq P^\omega(A) - \delta$$

and

$$P_n^\omega(A) \leq P_n^\omega\left(\overline{A}\right) \leq P^\omega\left(\overline{A}\right) + \frac{\delta}{2} \leq P^\omega(A) + \delta.$$

Therefore, for such $n \geq N$ and $\omega \in \complement\mathsf{R}_n \cap \complement\mathsf{S}_n \cap \complement\mathsf{T}_n$, $|P_n^\omega(A) - P^\omega(A)| \leq \delta$, that is

$$n \geq N \Rightarrow \mathbb{P}\left(\left|P_n \mathbb{1}_A - P\mathbb{1}_A\right| > \delta\right) \leq \varepsilon,$$

which is the wanted result.

Finally, the last step is to prove (5) $\Rightarrow$ (1). It suffices to show that for every $g \in \mathscr{C}_{\mathbb{R}_+}^\infty(S)$, then $P_n g \xrightarrow{\mathbb{P}} Pg$. Let $g \in \mathscr{C}_{\mathbb{R}_+}^\infty(S)$. By (2.2.5), there exists a sequence of simple functions $(g_m)_{m\in\mathbb{N}}$ such that $0 \leq g_m \leq g \leq g_m + \frac{3\|g\|}{2^m}$. If $g_m = \sum_{k=1}^n c_k \mathbb{1}_{A_k}$, then each A_k is a continuity set in measure of P and $g_m \to g$ on all S. By hypothesis, we may conclude that

$$P_n g_m \xrightarrow[n\to\infty]{\mathbb{P}} P g_m.$$

Note that $0 \leq g - g_m \leq \frac{3\|g\|}{2^m}$, so, if Q is a random probability measure, then $Q(g - g_m) \leq \frac{3\|g\|}{2^m}$. By the bounded convergence theorem, $Pg_m \xrightarrow[m\to\infty]{} Pg$ on *all* Ω. Then, for given $\delta > 0$ and any pair $m, n \in \mathbb{N}$,

$$\mathbb{P}\left(|Pg - P_n g| > \delta\right) \leq \mathbb{P}\left(|Pg - Pg_m| > \frac{\delta}{3}\right) + \mathbb{P}\left(|Pg_m - P_n g_m| > \frac{\delta}{3}\right) + \mathbb{P}\left(|P_n g_m - P_n g| > \frac{\delta}{3}\right).$$

Proceed as follows. Given $\varepsilon > 0$, take m_0 to be the minimum of the $k \in \mathbb{N}$ such that $\|g\| < \dfrac{\delta 2^k}{9}$. Then, for $m \geq m_0$, the set $\left\{|P_n g_m - P_n g| > \dfrac{\delta}{3}\right\}$ is empty. Now, given $m \geq m_0$, by convergence in probability of the sequence $(P_n g_m)_{n\in\mathbb{N}}$ to the random variable Pg_m, it follows

$$\lim_{n\to\infty} \mathbb{P}\left(|Pg_m - P_n g_m| > \frac{\delta}{3}\right) = 0.$$

Henceforth, for such $m \geq m_0$,

$$\limsup_{n\to\infty} \mathbb{P}\left(|Pg - P_n g| > \delta\right) \leq \mathbb{P}\left(|Pg - Pg_m| > \frac{\delta}{3}\right).$$

Let $m \to \infty$ to obtain that

$$\lim_{n\to\infty} \mathbb{P}\left(|Pg - P_n g| > \delta\right) = 0.$$

This implies the relation $P_n g \xrightarrow[n\to\infty]{\mathbb{P}} Pg$, which proves (1) and concludes the theorem. ■

§2.4. The necessity of a Prohorov's theorem in measure.

With this theorem, the direct half of what it is presumed to be Prohorov's theorem in measure claims can be stated and proved. First, the terminology to be employed is needed.

Definition (2.4.1) Suppose that (P_n) is a sequence of random probability measures $(\Omega, \mathscr{F}) \to (\mathcal{P}_S, \mathcal{B}_{\mathcal{P}_S})$ and that $\mathbb{P}$ is a probability measure on Ω. The phrase "the sequence (P_n) is relatively compact in measure" will be used to mean that there is a subsequence $\left(P_{\varphi(n)}\right)$ and, for this subsequence, a random probability measure $P : (\Omega, \mathscr{F}) \to (\mathcal{P}_S, \mathcal{B}_{\mathcal{P}_S})$ such that $P_{\varphi(n)} \xrightarrow{\text{wim}} P$.

Theorem (2.4.2) If (P_n) is relatively compact in measure then it is also tight in measure.

Proof: the following claim will be proved first. If $(G_m)_{m\in\mathbb{N}}$ is a sequence of open sets in S such that $G_m \uparrow S$ then, for every pair δ and ε of positive numbers, there exist a positive integer number M such that

$$m \geq M \Rightarrow \inf_{n\in\mathbb{N}} \mathbb{P}\left(P_n \mathbb{1}_{G_m} > 1 - \delta\right) \geq 1 - \varepsilon.$$

Assume it is *false*. Then, it is possible to choose a pair δ and ε of positive numbers such that for all $m \in \mathbb{N}$ there is a $P_{\sigma(m)}$ such that $\mathbb{P}\left(P_{\sigma(m)}\mathbb{1}_{G_m} > 1 - \delta\right) < 1 - \varepsilon$. Define $\mathsf{K} = \left\{\omega \in \Omega \middle| P^{\omega}_{\sigma(m)}(G_m) \leq 1 - \delta, \text{ i.o.}\right\}$, so that $\mathbb{P}\left(\mathsf{K}\right) \geq \varepsilon$. By relatively compactness in measure, and passing through another subsequence if necessary, there is a P such that $P_{\sigma(m)} \xrightarrow{\text{wim}} P$. By (2.3.1), part (4), for all $m \in \mathbb{N}$ there exists an $N_m \in \mathbb{N}$ such that, if

$$\mathsf{K}_{n,m} = \left\{\omega \in \Omega \middle| P^{\omega}_{\sigma(n)}(G_m) < P^{\omega}(G_m) - \frac{\delta}{2}\right\},$$

then

$$n \geq N_m \Rightarrow \mathbb{P}\left(\mathsf{K}_{n,m}\right) < \frac{\varepsilon}{2^{m+1}}.$$

Let $\mathsf{H}_m = \left\{\omega \in \Omega \middle| \liminf_{n\to\infty} P^{\omega}_{\sigma(n)}(G_m) < P^{\omega}(G_m) - \dfrac{\delta}{2}\right\}$. Then, $\mathsf{H}_m \subset \liminf_{n\to\infty} \mathsf{K}_{n,m}$ for the last set is

$$\liminf_{n\to\infty} \mathsf{K}_{n,m} = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} \mathsf{K}_{n,m},$$

and if $\omega \in \mathsf{H}_m$, the $\liminf$ allows concluding the existence of a number $k \in \mathbb{N}$ such that the relation $n \geq k$ implies the relation $P^{\omega}_{\sigma(n)}(G_m) < P^{\omega}(G_m) - \frac{\delta}{2}$, which says that $\omega \in \bigcap_{n=k}^{\infty} \mathsf{K}_{n,m}$, showing the desired inclusion. Therefore,

$$\mathbb{P}\left(\mathsf{H}_m\right) \leq \liminf_{n\to\infty} \mathbb{P}\left(\mathsf{K}_{n,m}\right) < \frac{\varepsilon}{2^{m+1}}.$$

Define $\mathsf{H} = \bigcup_{m=1}^{\infty} \mathsf{H}_m$. Then $\mathbb{P}\left(\mathsf{H}\right) \leq \frac{\varepsilon}{2}$. Look now at $\mathsf{K} \cap \complement\mathsf{H}$ and note that

$$\mathbb{P}\left(\mathsf{K} \cap \complement\mathsf{H}\right) = \mathbb{P}\left(\mathsf{K}\right) + \mathbb{P}\left(\complement\mathsf{H}\right) - \mathbb{P}\left(\mathsf{K} \cup \complement\mathsf{H}\right) \geq \varepsilon + 1 - \frac{\varepsilon}{2} - \mathbb{P}\left(\mathsf{K} \cup \complement\mathsf{H}\right) \geq \frac{\varepsilon}{2}.$$

Therefore, the set $\mathsf{K} \cap \complement\mathsf{H}$ is not empty. Let $\omega \in \mathsf{K} \cap \complement\mathsf{H}$. Then, as $\omega \in \mathsf{K}$, there exists an increasing function $\psi : \mathbb{N} \to \mathbb{N}$ such that $P^{\omega}_{\psi\sigma(m)}\left(G_{\psi\sigma(m)}\right) \leq 1 - \delta$, no matter which $m \in \mathbb{N}$. Also, as $\omega \in \complement\mathsf{H}$, it can be observed that for every $m \in \mathbb{N}$,

$$\liminf_{n\to\infty} P^{\omega}_{\sigma(n)}(G_m) \geq P^{\omega}(G_m) - \frac{\delta}{2}.$$

Therefore,

$$\begin{aligned}
P^{\omega}(G_m) &\leq \liminf_{n\to\infty} P^{\omega}_{\sigma(n)}(G_m) + \frac{\delta}{2} \\
&\leq \liminf_{n\to\infty} P^{\omega}_{\psi\sigma(n)}(G_m) + \frac{\delta}{2}, \quad \text{since } \psi \text{ is increasing} \\
&\leq \liminf_{n\to\infty} P^{\omega}_{\psi\sigma(n)}(G_{\psi\sigma(n)}) + \frac{\delta}{2}, \quad \text{since } G_m \subset G_{\psi\sigma(n)} \text{ for } \textit{every } n \textit{ large enough} \\
&\leq 1 - \frac{\delta}{2}.
\end{aligned}$$

But, if $m \uparrow \infty$ then $G_m \uparrow S$, so the left hand side will go to 1, but is bounded by something strictly less than one. This contradiction proves the claim.

The rest of the proof goes exactly as in the *proof* of (2.1.4). Indeed, let δ and ε two positive numbers. Using the same notation for the balls and dense sequence, the previous claim shows that, for every $k \in \mathbb{N}$, one can choose an $N_k \in \mathbb{N}$ such that for

$$\mathsf{J}_{n,m} = \left\{\omega \in \Omega \middle| P^{\omega}_n\left(\bigcup_{j=1}^{m} \mathrm{B}\left(x_j; \frac{1}{k}\right)\right) > 1 - \frac{\delta}{2^k}\right\}$$

$$m \geq N_k \Rightarrow \inf_{n\in\mathbb{N}} \mathbb{P}\left(\mathsf{J}_{n,m}\right) > 1 - \frac{\varepsilon}{2^k}$$

Define K as the *closure* of the set $\bigcap_{k=1}^{\infty}\bigcup_{j=1}^{N_k} \mathrm{B}\left(x_j; \frac{1}{k}\right)$. Then, for every P_n, the *proof* of (2.1.4) shows that

$$\mathbb{P}\left(P_n \mathbb{1}_K > 1 - \delta\right) > 1 - \varepsilon.$$

As K does not depend on P_n, the proof has been completed. ■

§2.5. Falsity of the sufficiency of Prohorov's theorem in measure.

Before giving the counterexample it is necessary to recall some concepts of probability theory.

Giving a sequence of random variables $(X_n)_{n\in\mathbb{N}}$ on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$; the set $\Sigma(X_n|n \in \mathbb{N})$ is defined to be the minimum σ-algebra on Ω such that it makes all the random variables (X_n) continuous; such set always exists since it is the intersection of all σ-algebras that makes all the X_n continuous (one such σ-algebra is $\mathfrak{P}(\Omega)$). Now, assume a sequence of random variables is given and set $\mathscr{F}_n = \Sigma(X_k|k \geq n)$. Then $\mathscr{F}_\infty = \bigcap_{n=1}^{\infty} \mathscr{F}_n$ is a σ-algebra, the "tail σ-algebra". The events inside $\mathscr{F}_\infty$ are called "tail events" and, as a matter of fact, the Kolmogorov zero-one law states that if the sequence is independent then every tail event has probability zero or one (such sets are called "trivial events"). A function $f : (\Omega, \mathscr{F}_\infty) \to (\mathbb{R}, \mathcal{B}_\mathbb{R})$ is called a "tail function" and, with the hypothesis of independence, all tail functions are constants for almost every point with respect to $\mathbb{P}$. All this results are stated and proved in *Real Analysis and Probability* by Ash: [2] (see 7.2.6 and 7.2.7). Also, 7.2.6 proves that the event $\mathsf{C} = \{X_n \text{ converges}\}$ is a tail event and, therefore, for sequences of independent random variables, such C has probability zero or one. Assume it is one and let X to be any random variable whose value at ω is this limit when it exists and zero everywhere else. Both 7.2.6 and 7.2.7 proves that X is almost surely a constant c and, therefore, the sequence $Y_n = X_n - c$ is also a sequence of independent random variables but now this sequence converges to zero. Impose the additional hypothesis that the X_n are independent and identically distributed, then so are the Y_n. If $\mathbb{P}(|Y_n| > \varepsilon) > 0$ for *some* $\varepsilon > 0$, then, as they are identically distributed, $\sum_{n=1}^{\infty} \mathbb{P}(|Y_n| > \varepsilon) = \infty$ and the second Borel-Cantelli lemma (see 7.1.5 of Ash: [2]) implies $\mathbb{P}(|Y_n| > \varepsilon, \text{ i.o}) = 1$; this last relation implies that for almost every ω with respect to $\mathbb{P}$, the sequence $Y_n(\omega)$ assumes for infinitely many n a value such that $|Y_n(\omega)| > \varepsilon$ and, therefore, it cannot converge to zero, which is of course a contradiction. Therefore, $\mathbb{P}(|Y_n| = 0) = 1$ and this implies that $\mathbb{P}(X_n = c) = 1$. If $\mathsf{M}_n = \{X_n = c\}$ then $\mathbb{P}(\mathsf{M}_n) = 1$ and $\mathbb{P}\left(\bigcap_{n=1}^{\infty} \mathsf{M}_n\right) \geq 1 - \sum_{n=1}^{\infty} \mathbb{P}\left(\complement \mathsf{M}_n\right) = 1$. Hence for every $\omega \in \Omega$, with the exception of a set of probability zero, all the $X_n(\omega) = c$. What it was proved here is that if X_n defines a sequence of independent and identically distributed random variables which converges with the exception of a set of probability zero, then the sequence is a constant; the converse is obvious.

Now, weak convergence in measure ask about convergence in probability and not for almost every point, so the previous argument must be modified a little. Assume then that (X_n) is a sequence of independent, identically distributed random variables which converges in probability to *some* random variable X. There is a subsequence $(X_{n'})$ of the sequence which converges for almost every point; by the previous paragraph, there exists a constant c such that the subsequence is equal to it almost surely. Observe then that $(X_{n'})$ converges in probability to *both* X and c, and therefore $X = c$ for almost every point with respect to $\mathbb{P}$. To clarify, we have proved that if a sequence of independent, identically distributed random variables converges in probability to some random variable, then the whole sequence is equal to a constant (and, hence, the limit also).

With the preliminarities of the previous paragraphs, let $S = \mathbb{R}$ and assume that (P_n) is a sequence of random probability measures that are identically distributed, independent and that is not a constant sequence (take, for instance, a sequence of random variables (X_n) with these properties and let P_n to be the random Dirac measure: ε_{X_n}). For every $f \in \mathscr{C}_\mathbb{R}^\infty(\mathbb{R})$, note that $e_f(P_n)$ defines a sequence of independent, identically distributed random variables; unless, the sequence is a constant, it cannot converge in probability. So, P_n cannot converge weakly in measure. As the P_n are identically distributed and one alone is tight (see (2.1.4)), for every $\varepsilon > 0$ and $\delta > 0$, there is a $K \subset S$ which is compact and

$$\mathbb{P}(P_1 \mathbb{1}_K \geq 1 - \delta) \geq 1 - \varepsilon,$$

but then, for *every* $n \in \mathbb{N}$,

$$\mathbb{P}(P_n \mathbb{1}_K \geq 1 - \delta) \geq 1 - \varepsilon,$$

and so for the infimum. Therefore, the sequence (P_n) is tight in measure and does not have a subsequence which converges weakly in measure. Finally, this is exactly what the *negation* of what was assumed to be Prohorov's theorem in measure.

✠ Bibliography

[1] A. De Masi, et al. "An Invariance Principle for Reversible Markov Processes. Applications to Random Motions in Random Environments." *Journal of Statistical Physics*, 55, Dec. 23, 1988: 787-855.

[2] Ash, Robert B. *Real Analysis and Probability.* Academic Press, 1972.

[3] Barlow, Burdzy, Timar. "Comparison of quenched and annealed invariance principles for random conductance mode". *To appear in Probability Theory and Related Fields.*

[4] Billingsley, Patrick. *Convergence of Probability Measures.* Wiley, Second Ed., 1999.

[5] Kipnis, C., S. R. S. Varadham. "Central Limit Theorem for Additive Functionals of Reversible Markov Processes and Applications to Simple Exclusions." *Communications in Mathematical Phyisics 104*, 104, 1-19, 1986.

[6] Dieudonné, Jean. *Foundations of Modern Analysis.* Academic Press, 1969.

[7] Dieudonné, Jean. *Treatise on Analysis.* Vol. II, Academic Press, 1976.

[8] Dugundji, James. *Topology.* Allyn and Bacon Inc., 1967.

[9] Ethier, Kurtz. *Markov Processes: Characterization and Convergence.* Wiley, 2005.