Department of Mathematical Sciences
UNIVERSITY OF COPENHAGEN

June 11, 2021

Weak Convergence in Metric Spaces

Milos Mathias Koch
Bacherlor's thesis in mathematics

Advisor: Ernst Hansen

Abstract

The theory of weak convergence in Euclidean spaces is large and eclectic, pulling on knowledge from functional analysis, Fourier analysis and much more. What is even more broad is the application of the theory of weak convergence, ranging from statistics to number theory. While the theory of weak convergence in Euclidean spaces is beautiful, there does exists more abstract spaces than the Euclidean ones.

The goal of this thesis is to introduce a theory of weak convergence in general metric spaces. The first three sections are dedicated to this theory and will conclude with a proof of Prohorov's compactness theorem. The last two sections are more concerned with surrounding subjects and applications.

Contents

Introduction

This bachelor's thesis aims to introduce a theory of weak convergence in metric spaces, which is a notion of convergence of probability measures defined on some σ-algebra on a general metric space. Other notions of weak convergence exists in various forms. What sets this particular theory apart is our exclusion of the Borel-σ-algebra from much of the theory.

This thesis is divided into 5 sections. Some preliminary results not depending on weak convergence theory and will be established in section 1. Section 2 introduces our notion of weak convergence in metric spaces, as well as other modes of convergence. As is customary in probability theory, measurability problems are abound! Section 3 is entirely concerned with the proof of two specific theorems, due to a highly influential Russian probabilists: Yuri Prohorov and Anatoliy Skorokhod. The proof of Prohorov's theorem is heavily inspired by the proof of the same theorem in David Pollard's book *Convergence of Stochastic Processes* and discussions with my advisor Ernst Hansen who introduced me to the proof strategy used. A more topological approach to weak convergence is developed in section 4 through an attempt to metrize weak convergence. While it fails in the general case, it turns out to be possible if we only consider sequences that converge to amenable limits. In the fifth and last section we will consider a specific, though important, example of a metric space: The cádlág space $D[0,1]$ equipped with the Skorokhod metric. The Skorokhod metric is usually defined using the group of strictly increasing homeomorphism from $[0,1]$ to itself. However, in this thesis it will be defined using the group of strictly increasing diffeomorphism from $[0,1]$ to itself in an attempt to elucidate certain aspects of the metric.

Acknowledgements

I want to express my sincerest gratitude to the at most countably many people who helped me in the process of writing this thesis. Especially, I would like to thank my friend Victor for helping me with making figures in LaTex, and I would like to thank my little sister Solvej for helping me find the area of trapezia and triangles. I would like to profoundly thank my advisor Ernst who proved to be a crucial source of meetings in a time where speaking with other people at all was a rarity. Their many contributions cannot be overstated, and this thesis would not be possible without them.

1 Preliminaries

We will review concepts that are not considered standard knowledge in the theory on measure theory, metric spaces and stochastic processes.

1.1 Metric spaces and measure theory

Example 1.1. Consider the mapping $\phi : \mathbb{R} \to L^\infty(\mathbb{R})$, given by $x \longmapsto \phi_x = 1_{[x,\infty)}$. We recognize ϕ_x as the distribution function for the degenerate distribution concentrated on x. We immediately note that the for distinct real numbers x_1 and x_2, $\phi(x_1)$ and $\phi(x_2)$ are far apart in uniform norm. Indeed, we have

$$\left\|\phi(x_1) - \phi(x_2)\right\|_\infty = \left\|1_{[x_1,\infty)} - 1_{[x_2,\infty)}\right\|_\infty = \left\|1_{[\min(x_1,x_2),\max(x_1,x_2))}\right\|_\infty = 1.$$

Hence any open ball with radius strictly less than $\frac{1}{2}$ contains at most one ϕ_x.
Let A be a (possibly non-Borel-measurable) subset of $\mathbb{R}$. then

$$\bigcup_{x\in A} B\left(\phi_x, \frac{1}{2}\right),$$

is an open set in $L^\infty(\mathbb{R})$, hence it is Borel-measurable. But as $\phi^{-1}(\bigcup_{x\in A} B\left(\phi_x, \frac{1}{2}\right)) = A$, we see ϕ cannot be measurable, as not all subsets of $\mathbb{R}$ are measurable.

If ϕ is not measurable, we might not be working in the most appropriate framework. The Borel σ-algebra simply contains too many sets. To remedy this, we introduce a new σ-algebra on our metric space.

Definition 1.2. Let (M, d) be a metric space. Then the σ-algebra generated by all open balls in (M, d),

$$\sigma(B(x, r) \mid x \in M, r > 0),$$

is denoted by $\mathcal{B}_0(M)$.

Let $B[x, r] = \{z \in M \mid d(x, z) \leq r\}$ be in (M, d) with center $x \in M$ and radius $r > 0$. As

$$B(x, r) = \bigcup_{n\in N} B\left[x, r - \frac{1}{n}\right],$$

and

$$B[x, r] = \bigcap_{n\in N} B\left(x, r + \frac{1}{n}\right),$$

the ball σ-algebra is equivalently characterised as the σ-algebra generated by the closed balls. In most texts on elementary measure theory, this definition is superfluous, by the following proposition.

Proposition 1.3. Let (M, d) be a separable metric space. Then ball-algebra $\mathcal{B}_0(M)$ and Borel-algebra $\mathcal{B}(M)$ coincides.

Proof. We will proof the proposition by showing that $\tau_d \subseteq \mathcal{B}_0$, where τ_d is the metric topology on (M, d). This will imply that $\mathcal{B}(M) \subseteq \mathcal{B}_0(M)$. As the reverse inclusion is trivial, we will have shown the proposition.
Let $U \in \tau_d$. As (M, d) is separable, so is U (Let C be a countable, dense subset of M, then $C \cap U$ is countable, dense subset of U.), hence for U satisfies the Lindelöf property, i.e for every open covering of U, there exist a countable subcovering. Since U is open, for each $x \in U$, there exist $\varepsilon_x > 0$ such that $B(x, \varepsilon_x) \subseteq U$. The family $\{B(x, \varepsilon_x)\}_{x \in U}$ is clearly an open cover of U with $\bigcup_{x \in U} B(x, \varepsilon_x) = U$. Therefore union of any subcover would also be equal to U. In particular, the union of any countable cover will be equal to U. As U is the union of countably many open balls, it is in $\mathcal{B}_0(M)$. □

Corallary 1.4. Let (M, d) be a metric space. Let $C \subseteq M$ be separable. Then the trace algebra of $\mathcal{B}(M)$ with respect to C is equal to the trace algebra of $\mathcal{B}_0(M)$ with respect to C.

Proposition 1.5. Let (M, d) be a metric space, any closed and separable subset of M is $\mathcal{B}_0(M)$-measurable.

Proof. Let $A \subseteq M$ be closed and separable, and let C be dense subset of A. By denseness, for every $k \in \mathbb{N}$, it holds that

$$A \subseteq \bigcup_{x \in C} B\left(x, \frac{1}{k}\right).$$

The right hand side is clearly $\mathcal{B}_0(M)$-measurable. Let $y \in \bigcap_{k \in \mathbb{N}} \bigcup_{x \in C} B(x, \frac{1}{k})$. For all $k \in \mathbb{N}$, we can then find $x_k \in A$ such that $y \in B\left(x_k, \frac{1}{k}\right)$. It is evident that $(x_k)_{k \in \mathbb{N}} \subseteq A$ and $x_k \to y$ as $k \to \infty$. As A is closed, we conclude $y \in A$. □

The above result in particular implies that compact sets are ball-measurable

A last measure-theoretic concept, we will be using extensively is tight measures.

Definition 1.6. A probability measure, μ, on the ball-σ-algebra of a metric space (M, d) is said to be tight, if for every $\varepsilon > 0$ there exists a compact set $K \subseteq \Omega$ such that

$$\mu(K) > 1 - \varepsilon.$$

We similarly say a collection of probability measures, $(\mu_i)_{i \in I}$, is uniformly tight, if for every $\varepsilon > 0$, there exists a compact set, K, such that

$$\sup_{i \in I} \mu_i(K^c) < \varepsilon.$$

Remark. A family of probability measures being uniformly tight clearly implies that every probability measure in the family is tight. The reverse implication is not true, unless the family is finite. Another immediate consequence is that any subfamily of an uniformly tight family of probability measures is also uniformly tight.

It turns out that tightness of limit measures is just what we need, when working with weak convergence in general metric spaces. Recall that a metric space is compact if and only if it is totally bounded and complete.

Lemma 1.7. Let (M,d) be a complete metric space. Sets of the form

$$\bigcap_{\varepsilon\in\mathbb{Q}_+}\bigcup_{n=1}^{n_\varepsilon} B(x_n,\varepsilon),$$

where $n_\varepsilon \in \mathbb{N}$ for all $\varepsilon \in \mathbb{Q}$, have compact closure.

Proof. We will proof the lemma by showing that sets of the given form are totally bounded. As the closure of a totally bounded set is totally bounded, the closure will be compact, as closed subsets of complete metric spaces are again complete. For any $\delta > 0$, choose $\varepsilon' \in \mathbb{Q}_+$ such that $\varepsilon' < \delta$, then

$$\bigcap_{\varepsilon\in\mathbb{Q}_+}\bigcup_{n=1}^{n_\varepsilon} B\left(x_n,\varepsilon\right) \subseteq \bigcup_{n=1}^{n_{\varepsilon'}} B\left(x_n,\varepsilon'\right) \subseteq \bigcup_{n=1}^{n_{\varepsilon'}} B\left(x_n,\delta\right),$$

hence $\bigcap_{\varepsilon\in\mathbb{Q}_+}\bigcup_{n=1}^{n_\varepsilon} B\left(x_n,\varepsilon\right)$ is totally bounded, and so it is relatively compact. $\square$

Remark. $\varepsilon \in \mathbb{Q}_+$ can without any further work be replaced with $\frac{1}{k}$ with $k \in \mathbb{N}$

Proposition 1.8. A probability measure, μ, on the ball-σ-algebra of a metric space (M,d) is tight only if there exists a closed, separable subset $C \subseteq M$ such that $\mu(C) = 1$. If the metric space is complete, the reverse conclusion holds.

Proof. Suppose μ is tight. Then there exists a sequence of compacta, $(K_n)_{n\in\mathbb{N}}$, such that $\mu(K_n) > 1 - \frac{1}{n}$. As K_n is compact for all $n \in \mathbb{N}$, they are separable. The countable union,

$$\bigcup_{n\in\mathbb{N}} K_n,$$

has probability 1, as for all $m \in \mathbb{N}$ $\mu(\bigcup_{n\in\mathbb{N}} K_n) \geq \mu(K_m) > 1 - \frac{1}{m}$, and separable, as it is the countable union of separable spaces. The closure of $\bigcup_{n\in\mathbb{N}} K_n$ satisfies the desired criteria.
Next, suppose that M is complete and there exists a closed, separable subset $C \subseteq M$ such that $\mu(C) = 1$. Let $(x_n)_{n\in\mathbb{N}}$ be a dense sequence of C. By denseness, for all $k \in \mathbb{N}$ the following holds

$$C \subseteq \bigcup_{n\in\mathbb{N}} B\left(x_n, \frac{1}{k}\right).$$

By downwards continuity of μ, for every $\varepsilon > 0$, we can find $n_k \in \mathbb{N}$ such that

$$\mu\left(\left(\bigcup_{n=1}^{n_k} B\left(x_n,\frac{1}{k}\right)\right)^c\right) \leq \frac{\varepsilon}{2^k}.$$

Consider the relatively compact set

$$\bigcap_{k\in\mathbb{N}}\bigcup_{n=1}^{n_k} B\left(x_n,\frac{1}{k}\right).$$

Boole's inequality implies

$$\mu\left(\left(\bigcap_{k\in\mathbb{N}}\bigcup_{n=1}^{n_k} B\left(x_n,\frac{1}{k}\right)\right)^c\right) \leq \sum_{n=1}^{\infty}\mu\left(\left(\bigcup_{n=1}^{n_k} B\left(x_n,\frac{1}{k}\right)\right)^c\right) \leq \varepsilon.$$

Thus, as the closure of $\bigcap_{k\in\mathbb{N}}\bigcup_{n=1}^{n_k} B\left(x_n,\frac{1}{k}\right)$ clearly has larger measure than $\bigcap_{k\in\mathbb{N}}\bigcup_{n=1}^{n_k} B\left(x_n,\frac{1}{k}\right)$ it self. Hence

$$\mu\left(\overline{\bigcap_{k\in\mathbb{N}}\bigcup_{n=1}^{n_k} B\left(x_n,\frac{1}{k}\right)}\right) > 1-\varepsilon.$$

As this holds for all $\varepsilon > 0$, we conclude that μ is tight $\square$

Corallary 1.9. Every probability measure on a separable and complete metric space is tight.

In the following proof, we will be using the fact that the family of finite intersections of closed balls in (M,d) is a $\cap$-stable generator of the ball-σ-algebra on (M,d).

Lemma 1.10. Let $(M,d,\mathcal{B}_0(M))$ be a metric space endowed with its associated ball-σ-algebra. Let μ be a tight measure on $\mathcal{B}_0(M)$. if

$$\int f\ d\mu = \int f\ d\nu$$

for all ball-measurable, bounded, Lipschitz functions, f, then $\mu = \nu$.

Proof. Let C be a closed and separable subset of M, and let $(c_n)_{n\in\mathbb{N}}$ be a countable dense subset of C. For fixed $x \in M$, it then holds that

$$\inf_{c\in C} d(x,c) = \inf_{n\in\mathbb{N}} d(x,c_n).$$

Consider the family of intervals $\{(-\infty,r) \mid r\in\mathbb{R}\}$. It is well known that this family generates the Borel algebra $\mathcal{B}$. Note that for fixed $x\in M$ the mapping $y \longmapsto d(y,x)$ satisfy

$$(d(y,x)\in(-\infty,r)) = \begin{cases} B(x,r) \text{ for } r>0 \\ \emptyset \text{ for } r\leq 0 \end{cases},$$

and

$$\left|d(z,x)-d(y,x)\right| \leq d(z,y).$$

This implies that $x \longmapsto d(y,x)$ is Lipschitz and $\mathcal{B}_0(M)-\mathcal{B}$-measurable where $\mathcal{B}$ denotes the Borel-algebra on $\mathbb{R}$. As $x \longmapsto d(x,C) = \inf_{n\in\mathbb{N}} d(x,c_n)$ is the countable infimum of ball-measurable mappings, it itself is ball-measurable. Note that, for $x,y\in M$,

$$d(x,C) \leq d(x,y) + d(y,C),$$

hence $d(x,C)-d(y,C)\leq d(x,y)$. Similarly we have, for $z\in C$

$$d(y,C) \leq d(y,z) \leq d(y,x) + d(x,z),$$

hence $d(y, C) \leq d(y, x) + d(x, C)$, or equivalently $d(y, C) - d(x, C) \leq d(x, y)$. These two inequalities imply that $x \longmapsto d(x, C)$ is Lipschitz. Now, for $n \in \mathbb{N}$, let $g_n : \mathbb{R} \to \mathbb{R}$ be given by

$$g_n(t) = \begin{cases} 1 & \text{for } t \leq 0 \\ 1 - nt & \text{for } t \in [0, \frac{1}{n}] \\ 0 & \text{for } t \geq \frac{1}{n}, \end{cases}.$$

We see that g_n is piecewise C^1 with bounded derivative, and hence Lipschitz and Borel-measurable, for all $n \in \mathbb{N}$. For $n \in \mathbb{N}$, let $f_n = g_n \circ d(\cdot, C)$. It is easy to see that f_n is ball-measurable and Lipschitz. Hence

$$\int f_n \, d\mu = \int f_n \, d\nu \text{ for } n \in \mathbb{N}.$$

Note that for $x \in C$, it holds that $f_n(x) = g_n(d(x, C)) = 1$. If $x \notin C$ and as C is closed, we have $d(x, C) > 0$, and hence for large enough $n \in \mathbb{N}$, it holds that $f_n(x) = 0$. This implies that f_n converges pointwise to 1_C. As $f_n \leq 1$ for all $n \in \mathbb{N}$, Lebesgue's dominated convergence theorem implies that

$$\mu(C) = \lim_{n\to\infty} \int f_n \, d\mu = \lim_{n\to\infty} \int f_n \, d\nu = \nu(C),$$

and so μ and ν agree on closed, separable subsets of M.
Now let μ be tight and let C' be a closed separable subset of M such that $\mu(C') = 1$. As μ and ν agree on closed, separable sets, we see that $\nu(C') = 1$. Let F be a finite intersection of closed balls. Then $F \cap C'$ is a closed, separable subset of M and therefore μ and ν will agree on $F \cap C'$. As both measures are concentrated on C', it holds that

$$\mu(F) = \mu(F \cap C') = \nu(F \cap C') = \nu(F).$$

As noted before, the family of finite intersections of closed balls is a $\cap$-stable generator of the ball-σ-algebra and hence $\mu = \nu$.

□

1.2 Stochastic processes

One of the most commonly occuring classes of spaces in probability theory is the class of cádlág space. The cádlág space of interest in this thesis is the space $D[0, 1]$, which consists of functions $x : [0, 1] \to \mathbb{R}$ that are right-continuous with left limits.

Proposition 1.11. The càdlàg space $D[0, 1]$ is a closed subspace of $L^\infty([0, 1])$ under the uniform norm.

Proof. The fact that $D[0, 1]$ is a vector space under pointwise addition and scaling, is a consequence of linearity of limits, demonstrated by the following computations: For sequences $(t_n)_{n\in\mathbb{N}}$ and $(u_n)_{n\in\mathbb{N}}$ with $t_n \searrow t$ and $u_n \nearrow t$ for $n \to \infty$ for some $x, y \in \mathbb{R}$, and $f, g \in D[0, 1]$.

$$(xf + yg)(t_n) = xf(t_n) + yg(t_n) \to xf(t) + yg(t) = (xf + yg)(t)$$
$$(xf + yg)(u_n) = xf(u_n) + yg(u_n) \to xf(u-) + yg(u-) = (xf + yg)(u-).$$

Let $x \in D[0,1]$. Assume for a contradiction, that $x \notin L^\infty([0,1])$, then $\sup_{t\in[0,1]}(\big|x(t)\big|) = \infty$. There exists a sequence $(t_n)_{n\in\mathbb{N}} \subseteq [0,1]$ such that $\big|x(t_n)\big| \to \infty$. As $[0,1]$ is compact, there exist a convergent subsequence $(t_{n_j})_{j\in\mathbb{N}}$, converging to some $t \in [0,1]$. We can without loss of generality assume that $(t_{n_j})_{j\in\mathbb{N}}$ is converging from either the left or the right. If it is not, let $(t_{n_{j_k}})_{k\in\mathbb{N}}$ be the subsequence of $(t_{n_j})_{j\in\mathbb{N}}$ with all elements either to the left or right of t. This subsequence is clearly also convergent. By definition of $D[0,1]$, if $(t_{n_j})_{j\in\mathbb{N}}$ is converging from the right, right continuity of x implies that $\lim_{j\to\infty} x(t_{n_j}) = x(t)$ is finite. If $(t_{n_j})_{j\in\mathbb{N}}$ is converging from the left, the existence of left limits of x implies that $\lim_{j\to\infty}\big|x(t_{n_j})\big|$ is finite. This is a contradiction, and hence we conclude that x is bounded. Hence $D[0,1] \subseteq L^\infty([0,1])$. We will now prove that $D[0,1]$ is closed in $L^\infty([0,1])$. Let $(x_n)_{n\in\mathbb{N}} \subseteq D[0,1]$ be a sequence converging uniformly to $x \in L^\infty([0,1])$ and let $\varepsilon > 0$. As the sequence is uniformly convergent, we can find $N \in \mathbb{N}$ such that $\big\|x_j - x\big\|_\infty < \varepsilon$ for $j \geq N$, and hence $\|x_n - x_m\|_\infty \leq \|x_n - x\|_\infty + \|x_m - x\|_\infty < 2\varepsilon$ for $n, m \geq N$. Let $u_n \nearrow u$ be an increasing sequence of numbers in $[0,1]$. As $(x_n)_{n\in\mathbb{N}}$ is càdlàg, there exists a sequence, $(c_n)_{n\in\mathbb{N}}$ of real numbers, such that, for each $i \in \mathbb{N}$, there exists $N_i \in \mathbb{N}$ such that $\big|x_i(u_n) - c_i\big| < \varepsilon$ for $n \geq N_i$, i.e. $c_n = x_n(u-)$ for all $n \in \mathbb{N}$. Thus, by the triangle inequality, we have for $n, m \geq N$

$$\begin{aligned}
|c_n - c_m| &\leq \big|c_n - x_n(u_i)\big| + \big|x_n(u_i) - x_m(u_i)\big| + \big|c_m - x_m(u_i)\big| \\
&\leq \big|c_n - x_n(u_i)\big| + \|x_n - x_m\|_\infty + \big|c_m - x_m(u_i)\big| \\
&< 4\varepsilon
\end{aligned}$$

for any $i > \max(N_n, N_m)$. Hence $(c_n)_{n\in\mathbb{N}}$ is Cauchy, therefore there exist $c \in \mathbb{R}$ and $N_c \in \mathbb{N}$, such that $|c_n - c| < \varepsilon$ for $n \geq N_c$, i.e. $(c_n)_{n\in\mathbb{N}}$ is convergent to $c \in \mathbb{R}$. Once again by the triangle inequality for $k \geq \max(N, N_c)$

$$\begin{aligned}
\big|x(u_i) - c\big| &\leq \big|x(u_i) - x_k(u_i)\big| + \big|x_k(u_i) - c_k\big| + |c_k - c| \\
&\leq \|x - x_k\|_\infty + \big|x_k(u_i) - c_k\big| + |c_k - c| \\
&< 3\varepsilon,
\end{aligned}$$

for any $i > \max(N_c, N_k, N)$. Hence x has left limits. Letting $u_n \searrow u$ be a decreasing sequence in $[0,1]$, and $c_m = x_m(u)$ for all $m \in \mathbb{N}$, similar computations shows that x is right continuous, hence $x \in D[0,1]$. □

An immediate consequence of this, is of course that $D[0,1]$ is complete with respect to the uniform norm, i.e. it is a Banach space. However a small modification of example 1.1 shows that $D[0,1]$ is not a separable Banach space with respect to the uniform norm.
The space $D[0,1]$ usually arises when discussing continuous time stochastic processes. Such processes are projection measurable. In the $D[0,1]$ case this will imply that they are also ball measurable and vice versa. In the following we will denote the projection onto $t \in [0,1]$ by π_t.

Proposition 1.12. On $D[0,1]$, the projection algebra and ball algebra coincides.

Proof. For fixed $t \in [0,1]$, consider the sets of the form

$$\{x \in D[0,1] \mid \pi_t(x) > \alpha\},$$

with $\alpha \in \mathbb{R}$. If this set is ball-measurable, then π_t is ball-measurable. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence of cádlág functions, given by

$$x_n(s) = \alpha + 1_{[t,t+\frac{1}{n})}(s)(n + \frac{1}{n}),$$

where $1_{[t,t+\frac{1}{n})}$ is the indicator function of $[t, t + \frac{1}{n})$. We obviously let n be large enough for $t + \frac{1}{n} \leq 1$. For $t = 1$, consider the functions $x_n = \alpha + 1_{\{t\}}(n + \frac{1}{n})$. Let $x \in B(x_n, n)$ for some $n \in \mathbb{N}$. Then

$$\begin{aligned} x(t) &= x(t) + x_n(t) - x_n(t) \\ &\geq x_n(t) - \left|x(t) - x_n(t)\right| \\ &> x_n(t) - n \\ &\geq \alpha + \frac{1}{n} \\ &> \alpha \end{aligned}$$

Now, let $x(t) > \alpha$ and let $t \in [0, 1)$. By right continuiuty, there exists an $n_1 \in \mathbb{N}$ such that $n \geq n_1$ implies $x(s) > \alpha + \frac{1}{n}$ for $s \in [t, t + \frac{1}{n})$. As $D[0, 1] \subseteq L_\infty([0, 1])$, there exists $n_2 \in \mathbb{N}$ such that $\|x\|_\infty + |\alpha| < n < 2n$ for $n \geq n_2$. Let $n > \max(n_1, n_2)$, then for $s \in [t, t + \frac{1}{n})$, we have

$$\begin{aligned} x(s) - x_n(s) &= x(s) - \alpha - n - \frac{1}{n} \\ &\leq \|x\|_\infty + |\alpha| - n - \frac{1}{n} \\ &< n - \frac{1}{n} \\ &< n \end{aligned}$$

and

$$\begin{aligned} x(s) - x_n(s) &= x(s) - \alpha - n - \frac{1}{n} \\ &> -n \end{aligned}$$

Thus $\left|x(s) - x_n(s)\right| < n$. If $s \notin [t, t + \frac{1}{n})$. Then

$$\left|x(s) - x_n(s)\right| \leq \|x\|_\infty + |\alpha| < n.$$

Hence

$$\{x \in D[0, 1] \mid \pi_t(x) > \alpha\} = \bigcup_{n\in\mathbb{N}} B(x_n, n),$$

for all $t \in [0, 1]$ and $\alpha \in \mathbb{R}$. As sets of the form of the left hand side generates the projection algebra, we conclude that the projection is contained in the ball algebra.
For any cádlág function, $x \in D[0, 1]$, it holds that

$$\sup_{t\in[0,1]} x(t) = \sup_{t\in[0,1]\cap\mathbb{Q}} x(t).$$

As $[0,1] \cap \mathbb{Q}$ is dense in $[0,1]$ and x is right continuous. Let $B(z,r)$ denote the open ball in $D[0,1]$ with center $z \in D[0,1]$ and radius $r > 0$. We see

$$\begin{aligned}\overline{B}(z,r) &= \{x \in D[0,1] \mid \|z-x\|_\infty < r\} \\ &= \bigcap_{t\in[0,1]} \{x \in D[0,1] \mid \big|z(t)-x(t)\big| < r\} \\ &= \bigcap_{t\in[0,1]\cap\mathbb{Q}} \{x \in D[0,1] \mid \big|z(t)-x(t)\big| < r\}.\end{aligned}$$

As $\{x \in D[0,1] \mid \big|z(t)-x(t)\big| < r\} = (\pi_t, \pi_t)^{-1}(\{(x_1,x_2) \in \mathbb{R}^2 \mid |x_1 - x_2| < r\})$ is projection measurable, we conclude that the ball algebra is contained in the projection algebra. $\square$

Example 1.13. If we revisit 1.1, we see that the specific problem raised there is actually solved by the introduction of the ball-σ-algebra. For and $t \in [0,1]$ and $B \in \mathcal{B}$ it holds

$$\phi^{-1}(\pi^{-1}(B)) = \{x \in \mathbb{R} \mid \phi_x(t) \in B\} = \{x \in \mathbb{R} \mid 1_{[0,t]}(x) \in B\},$$

and therefore it is evident that ϕ is Borel-projection-measurable. As the projection algebra and ball algebra coincides on $D[0,1]$, we conclude ϕ is Borel-ball-measurable.

2 Weak convergence in metric spaces

2.1 Definition and first results

In this section, we will establish the basics of a theory of weak convergence in metric spaces. This, of course, begins with a definition:

Definition 2.1. Let (M, d) be a metric space equipped with the associated ball-σ-algebra $\mathcal{B}_0(M)$. Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence of probability measures on M. $(\mu_n)_{n\in\mathbb{N}}$ is said to converge weakly to some probability measure μ, if for all bounded, continuous, $\mathcal{B}_0(M)$-measurable functions, $f : M \to \mathbb{R}$, it holds that

$$\int f \ d\mu_n \to \int f \ d\mu.$$

We say that a sequence of M-valued stochastic variables, $(X_n)_{n\in\mathbb{N}}$ on some probability space $(\Omega, \mathbb{F}, P)$ converges in distribution to some M-valued stochastic variable X if for all bounded, continuous, $\mathcal{B}_0(M)$-measurable functions, f, it hold that

$$\int f(X_n) \ dP \to \int f(X) \ dP.$$

We will write $X_n \overset{D}{\to} X$

A lot of open sets in (M, d) are not ball-measurable, so we can not guarantee that every continuous function $f : M \to \mathbb{R}$ is automatially ball-measurable. However, there are still plenty of continuous functions that are ball-measurable, and even better, there are plenty of Lipschitz functions that are ball-measurable.

Proposition 2.2. Let $(M, d, \mathcal{B}_0(M))$ be a metric space endowed with the associated ball-σ-algebra. For every point $x \in M$ and $0 < s < r$, there exists a bounded, Lipschitz, ball-measurable function $f : M \to \mathbb{R}$ such that

$$1_{B(x,s)} \leq f \leq 1_{B(x,r)}.$$

Proof. We have already seen that $y \longmapsto d(y, x)$ is Lipschitz and $\mathcal{B}_0(M) - \mathcal{B}$-measurable for fixed $x \in M$. Now consider the function $g : \mathbb{R} \to \mathbb{R}$, given by

$$g(t) = \begin{cases} 1 \text{ for } t \leq s \\ \frac{r-x}{r-s} \text{ for } t \in [s, r] \\ 0 \text{ for } t \geq r \end{cases}.$$

Clearly g is bounded, $\mathcal{B} - \mathcal{B}$-measurable and piecewise C^1 with bounded derivative, and hence g is Lipschitz and Borel-measurable. Thus the function $f = g \circ d(\cdot, x)$ satisfy the desired properties. □

Lemma 2.3. Let $(M, d, \mathcal{B}_0(M))$ be a metric space endowed with the associated ball-σ-algebra and let $C \in \mathcal{B}_0(M)$ be separable. Suppose that $h : M \to \mathbb{R}$ is a ball-measurable function, such that

1. $h(x) \geq 0$ for all $x \in M$,
2. h is continuous on C.

Then there exists a sequence, $(f_n)_{n\in\mathbb{N}}$, of bounded, Lipschitz and ball-measurable functions such that $0 \leq f_1 \leq f_2 \leq \ldots$ and $f_n(x) \nearrow h(x)$ for all $x \in C$ for all $n \in \mathbb{N}$.

Proof. Let $(c_n)_{n\in\mathbb{N}}$ be a countable dense subset of C. for each $c \in (c_n)_{n\in\mathbb{N}}$ and $s, r \in \mathbb{Q}_+$ with $s < r$, let $f_{c,s,r}$ denote a bounded, Lipschitz ball-measurable function, such that

$$1_{B(c,s)} \leq f_{c,s,r} \leq 1_{B(c,r)}.$$

Let $\mathcal{C}$ be the class of functions given by

$$\mathcal{C} = \{q f_{c,s,r} \mid c \in (c_n)_{n\in\mathbb{N}},\ s, r \in \mathbb{Q}_+\ s < r,\ q \in \mathbb{Q}\}.$$

It is easy to see that $\mathcal{C}$ is countable, hence the subset $\mathcal{C}^h = \{f \in \mathcal{C} \mid f \leq h\}$, is also countable. Let $(g_n)_{n\in\mathbb{N}}$ be an enumeration of $\mathcal{C}^h$. Let $(f_n)_{n\in\mathbb{N}}$ be a sequence given by $f_n = \max(g_1, g_2 \ldots g_n)$. As the maximum of Lipschitz functions are Lipschitz, $(f_n)_{n\in\mathbb{N}}$ immediatly satisfy every desired property, except pointwise convergence to h.
Let $y \in C$. If $h(y) = 0$, then pointwise convergence is clear, $f_n(y) = 0$ for all $n \in \mathbb{N}$. If $h(y) > 0$ then let $p \in \mathbb{Q}_+$ be such that $h(y) > p$. By continuity, there exists $t \in \mathbb{Q}_+$, such that

$$h(z) > p \text{ for } z \in B(z, t).$$

Let $x \in (c_n)_{n\in\mathbb{N}} \cap B(y, \frac{t}{3})$. By the triangle inequality, we have

$$y \in B\left(x, \frac{t}{3}\right) \subseteq B\left(x, \frac{2t}{3}\right) \subseteq B\left(y, t\right).$$

These two statements together imply that $p f_{x,\frac{t}{3},\frac{2t}{3}} \in \mathcal{C}^h$. Hence, there exists n_0 such that $p f_{x,\frac{t}{3},\frac{2t}{3}} = g_{n_0}$. This implies that

$$f_n(y) \geq g_{n_0}(y) = p \text{ for } n \geq n_0.$$

As this hold for all points $y \in C$, we conclude that f_n converges pointwise to h on C. $\square$

Lemma 2.4 (Convergence lemma)**.** Let $(M, d, \mathcal{B}_0(M))$ be a metric space endowed with the associated ball-σ-algebra and let $C \in \mathcal{B}_0(M)$ be separable and ball-measurable. Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence of probability measures on $\mathcal{B}_0(M)$, and let μ be a tight probability measure on $\mathcal{B}_0(M)$ concentrated on C. Suppose that, for all $f : M \to \mathbb{R}$, bounded, Lipschitz and ball-measurable, it holds

$$\int f \, d\mu_n \to \int f \, d\mu.$$

then for all bounded, continuous on C, ball-measurable functions, $h : M \to \mathbb{R}$, it holds that

$$\int h \, d\mu_n \to \int h \, d\mu.$$

In particular, $\mu_n \to \mu$ weakly.

Proof. Without loss of generality, assume that $h \geq 0$. If h is not non-negative, consider instead $h' = h + \|h\|_\infty$. If the lemma holds for h', then

$$\int h \, d\mu_n = \int h' \, d\mu_n - \|h\|_\infty \to \int h' \, d\mu - \|h\|_\infty = \int h \, d\mu.$$

Let $(f_n)_{n \in \mathbb{N}}$ be an non-negative increasing sequence of bounded, Lipschitz and ball-measurable functions dominated by h. By monotonocity of integrals, we have, for any $m \in \mathbb{N}$

$$\int f_m \, d\mu_n \leq \int h \, d\mu_n.$$

As $(f_n)_{n \in \mathbb{N}}$ is a sequence of bounded, Lipschitz and ball-measurable functions, we have by assumption that $\int f_m \, d\mu_n \to \int f_m \, d\mu$ for all $m \in \mathbb{N}$. Hence

$$\int f_m \, d\mu \leq \liminf_{n \to \infty} \int h \, d\mu_n.$$

Note, since $f_n \nearrow h$ μ-almost surely, the monotone convergene theorem implies $\int f_n \, d\mu \to \int h \, d\mu$. Hence

$$\int h \, d\mu \leq \liminf_{n \to \infty} \int h \, d\mu_n.$$

Repeat the argument above with $-h$ to obtain

$$-\int h \, d\mu \leq -\limsup_{n \to \infty} \int h \, d\mu_n,$$

or equivalently

$$\int h \, d\mu \geq \limsup_{n \to \infty} \int h \, d\mu_n.$$

Collecting all this information yields

$$\int h \, d\mu \leq \liminf_{n \to \infty} \int h \, d\mu_n \leq \limsup_{n \to \infty} \int h \, d\mu_n \leq \int h \, d\mu,$$

and so $\int h \, d\mu_n \to \int h \, d\mu$. As any bounded, continuous ball measurable function, g, from M to $\mathbb{R}$ will be continuous on C, we can conclude that $\mu_n \to \mu$ weakly. □

Corallary 2.5 (Continuous mapping theorem)**.** Let $(M, d, \mathcal{B}_0(M))$ and $(M', d', \mathcal{B}_0(M'))$ be metric spaces endowed with their associated ball-σ-algebras. $(\mu_n)_{n \in \mathbb{N}}$ be a sequence of probability measures on $\mathcal{B}_0(M)$ such that $\mu_n \to \mu$ weakly to some tight probability measure on $\mathcal{B}_0(M)$ concentrated on some separable $\mathcal{B}_0(M)$-measurable $C \subseteq M$. Suppose $H : M \to M$ is bounded, $\mathcal{B}_0(M) - \mathcal{B}_0(M')$-measurable and continuous on C, then $H(\mu_n) \to H(\mu)$ weakly.

Proof. Let f be bounded, continuous and $\mathcal{B}_0(M')$-measurable. Then $f \circ H$ is bounded, $\mathcal{B}_0(M)$-measurable and continuous on C. Hence, by abstract change of variables, we have

$$\int f \, dH(\mu_n) = \int f \circ H \, d\mu_n \to \int f \circ H \, d\mu = \int f \, dH(\mu).$$

□

2.2 Other modes of convergence in metric spaces

Just like the classical theory of weak convergence in $\mathbb{R}$, there is a set of stronger notions of convergence which might be useful in proving weak convergence. However, they are quite a bit more complicated to define in non-separable metric spaces than they are in $\mathbb{R}$ or even Euclidean spaces. This difficulty stems from the fact that the distance, $(X, Y) \longmapsto d(X, Y)$, between two metric space valued random variables, X and Y, is not in general measurable with respect to the ball-algebra. If we assume some regularity on at least one of the variables, this problem disappears.

Definition 2.6. We say a stochastic variable, $X : (\Omega, \mathbb{F}, P) \to (M, d, \mathcal{B}_0)$ is separable if there exists a closed, separable set, C, such that

$$\forall \omega \in \Omega : X(\omega) \in C.$$

Remark. The above definition is a bit stronger than X having tight distribution. If X simply has tight distribution, we will modify X as such: Let C be a closed and separable set such that $P(X \in C) = 1$ and let $c_0 \in C$. Then define

$$X'(\omega) = \begin{cases} X(\omega) & \text{if } X(\omega) \in C \\ c_0 & \text{if } X(\omega) \notin C \end{cases}.$$

Then X' is separable and $X \sim X'$.

Example 2.7. Consider the space of continuous time stochastic processes on $[0, 1]$ with cádlág path and let B be the Brownian motion restricted to $[0, 1]$. As $C[0, 1]$ is seperable, it follows that B is a seperable stochastic variable as the brownian motion can be assumed to be surely in $C[0, 1]$.

Lemma 2.8. Let $(M, d, \mathcal{B}_0)$ be a metric space, endowed with it's associated ball σ-algebra. Let X and Y be M-valued ball-measurable stochastic variables on some probability space $(\Omega, \mathbb{F}, P)$. If at least one of X and Y are separable, then $d(X, Y)$ is a real-valued stochastic variable.

Proof. Let Y be seperable. Let C be a closed separable set such that Y is surely in C and let $(c_n)_{n\in\mathbb{N}}$ be a dense subset of C. Let $r > 0$. If we can show that

$$\{(x, y) \mid d(x, y) < r\} \cap M \times C \tag{1}$$

$$= \left(\bigcup_{n=1}^{\infty} \bigcup_{s,t\in\mathbb{Q}_+ : s+t<r} B(c_n, s) \times B(c_n, t) \right) \cap M \times C, \tag{2}$$

then we are done. Indeed, since (X, Y) is $\mathbb{F} - \mathcal{B}_0(M) \otimes \mathcal{B}_0(M)$-measurable this will imply

$$(d(X, Y) < r) = (X, Y)^{-1}(\{(x, y) \mid d(x, y) < r\}) = (X, Y)^{-1}(\{(x, y) \mid d(x, y) < r\} \cap M \times C).$$

(2) above is clearly in $\mathcal{B}_0(M) \otimes \mathcal{B}_0(M)$, we will have shown the proposition.
Let $(u, v) \in M \times C$ be in (2), then there in particular exists $n \in \mathbb{N}$ and $s, t \in \mathbb{Q}_+$ with $s+t < r$, such that $u \in B(c_n, s)$ and $v \in B(c_n, t)$. The triangle inequality implies that $d(u, v) < r$.

If $(u,v) \in M \times C$ is in the lefthand side, then $d(u,v) < r$. Let $t \in \mathbb{Q}_+$ such that $d(u,v)+3t < r$. By denseness of $(c_n)_{n\in\mathbb{N}}$, we can find $n \in \mathbb{N}$ such that $d(v,c_n) < t$. By the triangle inequality, we now have

$$d(u,c_n) \leq d(u,v) + d(v,c_n) < d(u,v) + t.$$

If $s \in (d(u,v)+t, d(u,v)+2t) \cap \mathbb{Q}_+$, then $s+t < r$ and

$$(u,v) \in B(c_n,s) \times B(c_n,t).$$

□

Definition 2.9. Let $(M,d,\mathcal{B}_0)$ be a metric space, endowed with it's associated ball σ-algebra. Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of stochastic variables on some probability space $(\Omega,\mathbb{F},P)$ and let X be separable. If

$$P\left(\bigcap_{\varepsilon\in\mathbb{Q}_+}\bigcup_{N=1}^{\infty}\bigcap_{n\geq N}^{\infty}(d(X_n,X)<\varepsilon)\right) = 1,$$

we say that $X_n \to X$ almost surely and write $X_n \overset{as}{\to} X$. We will use the shorthand

$$(X_n \to X) = \bigcap_{\varepsilon\in\mathbb{Q}_+}\bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty}(d(X_n,X)<\varepsilon).$$

If

$$P(d(X_n,X)<\varepsilon) \to 1, \text{ for } n\to\infty \text{ for all } \varepsilon > 0,$$

we say that $X_n \to X$ in probability and write $X_n \overset{P}{\to} X$.

Proposition 2.10. Let $(M,d,\mathcal{B}_0)$ be a metric space, endowed with it's associated ball σ-algebra. Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of stochastic variables on some probability space $(\Omega,\mathbb{F},P)$ and let X be separable. Then we have the following sequence of implication:

$$X_n \overset{as}{\to} X \implies X_n \overset{P}{\to} X \implies X_n \overset{D}{\to} X.$$

Proof. Assume $X_n \overset{as}{\to} X$. Then $d(X_n,X) \to 0$ almost surely (in the classical $\mathbb{R}$-sense). Hence $1_{(d(X_n,X)\geq\varepsilon)}$ also converges to 0 almost surely. By Lebesgue's dominated convergence theorem $X_n \overset{P}{\to} X$.
Now assume $X_n \overset{P}{\to} X$. In view of the convergence lemma, let f be bounded, non-constant, Lipschitz and ball-measurable. As

$$P(\left|f(X_n)-f(X)\right| \geq \varepsilon) \leq P\left(d(X_n,X) \geq \frac{\varepsilon}{\|f\|_{Lip}}\right),$$

where $\|f\|_{Lip}$ is the infimum of all Lipschitz constants for f, we see $f(X_n) \overset{P}{\to} f(X)$ (in the classical $\mathbb{R}$-sense). By the extended Lebesgue dominated convergence theorem[1]

$$\int f(X_n)dP \to \int f(X)dP,$$

hence $X_n \overset{D}{\to} X$. □

[1] See [Hansen], problem 2.14.

3 Two theorems

We will state two major theorems in the theory of weak convergence in metric spaces: Skorokhod's representation theorem and Prohorov's compactness theorem. For the sake of brevity we will only prove Prohorov's compactness theorem. In the proof of this remarkable theorem we will argue on stochastic variables with certain amenable properties. Therefore, we will need procedures to generate such variables. We will start by stating a general result on the existence of variables with given distribution.

3.1 Constructions of stochastic variables

Theorem 3.1. Let $((\mathcal{X}_n, \mathbb{K}_n, \mu_n))_{n\in\mathbb{N}}$ be a sequence if probability spaces. Then there exists a probability space $(\Omega, \mathbb{F}, P)$ and a sequence $(X_n)_{n\in\mathbb{N}}$ of independent stochastic variables, such that $X_n : (\Omega, \mathbb{F}, P) \to (\mathcal{X}_n, \mathbb{K}_n, \mu_n)$ and $X_n(P) = \mu_n$ for all $n \in \mathbb{N}$.

Proof. See theorem B in [Halmos], page 157. □

For a sequence of probability measures, $(\mu_n)_{n\in\mathbb{N}}$, on some metric space (M, d), we will say that a sequence of M-valued stochastic variables $(X_n)_{n\in\mathbb{N}}$ on some probability space $(\Omega, \mathbb{F}, P)$, represents $(\mu_n)_{n\in\mathbb{N}}$ if $X_n(P) = \mu_n$ for all $n \in \mathbb{N}$.
In the following, we will say a partition of a metric space M is ball-measurable if every set in the partition is ball-measurable.

Theorem 3.2 (Construction 1)**.** Let $(\Omega, \mathbb{F}, P)$ be a probability space, and let $(M, d, \mathcal{B}_0(M))$ be a metric space, endowed with its associated ball-σ-algebra. Let X be a M-valued, separable stochastic variable. Suppose for some $\varepsilon > 0$, there exists a finite ball-measurable partition, $(B_i)_{i=0}^k$, of M such that

1. $diam(B_i) = \sup(d(x, y) \mid x, y \in B_i) < 2\varepsilon$ for $1 \le i \le k$
2. $P(X \in B_0) < \varepsilon$.

If there exists a sequence $(X_n)_{n\in\mathbb{N}}$ of M-valued stochastic variables on $(\Omega, \mathbb{F}, P)$, such that

$$P(X_n \in B_i) \ge (1 - \varepsilon)P(X \in B_i) \text{ for } 0 \le i \le k \text{ and } n \in \mathbb{N},$$

then there exist another sequence $(X'_n)_{n\in\mathbb{N}}$ such that $X_n \sim X'_n$ for all $n \in \mathbb{N}$ and such that

$$P\left(\bigcup_{n=1}^{\infty} \left(d(X'_n, X) \ge 2\varepsilon\right)\right) < 2\varepsilon$$

Proof. Without loss of generality, assume that $P(X \in B_i) > 0$ for $i = 1, \ldots, k$. If this is not the case, simply define $B'_0 = B_0 \cup \bigcup_{i\in N} B_i$, where N is the set of indices of all the B_i's with $P(X \in B_i) = 0$, and recount the remaining B_i's. Hence, we can safely assume that $P(X_n \in B_i) > 0$ for $1 \le i \le k$ and $n \in \mathbb{N}$. We will not lose much sleep if it happens that $P(X \in B_0) = 0$. For $0 \le i \le k$, and $n \in \mathbb{N}$, define stochastic variables, Z_i^n, with distributions

$$P(Z_i^n \in A) = P(X_n \in A \mid X_n \in B_i) = \frac{P(X_n \in A \cap B_i)}{P(X_n \in B_i)}.$$

If it happens that $P(X_n \in B_0) = 0$ for some $n \in \mathbb{N}$, we simply let Z_0^n have whatever distribution it so pleases, it will not matter. In the following, we let V and $(W_n)_{n \in \mathbb{N}}$ be stochastic variables, independent of the Z variables, X and each other. Let V be uniformly distributed on $[0, 1]$. For every $n \in \mathbb{N}$, define a $\{1, \dots k\}$-valued stochastic variable, W_n, with point probabilities yet to be specified.
For $n \in \mathbb{N}$, define X'_n as

$$X'_n = \begin{cases} Z_0^n & \text{on } (V \leq 1 - \varepsilon, X \in B_0) \cup (V > 1 - \varepsilon, W_n = 0) \\ Z_1^n & \text{on } (V \leq 1 - \varepsilon, X \in B_1) \cup (V > 1 - \varepsilon, W_n = 1) \\ \vdots & \\ Z_k^n & \text{on } (V \leq 1 - \varepsilon, X \in B_k) \cup (V > 1 - \varepsilon, W_n = k). \end{cases}$$

We note that on $(V \leq 1 - \varepsilon, X \notin B_0)$ X'_n and X are both in the same set in the partition, in particular $d(X'_n, X) < 2\varepsilon$. Hence

$$P(d(X'_n, X) \geq 2\varepsilon) \leq P(V > 1 - \varepsilon) + P(X \in B_0) < 2\varepsilon.$$

As $(d(X'_n, X) \geq 2\varepsilon) \subseteq (V > 1 - \varepsilon, X \in B_0)$ for all $n \in \mathbb{N}$, it holds that

$$P\left(\bigcup_{n=1}^{\infty} \big(d(X'_n, X) \geq 2\varepsilon\big)\right) < 2\varepsilon.$$

It is elementary to calculate the distribution of X'_n:

$$\begin{aligned} P(X'_n \in A) &= P(X'_n \in A, V \leq 1 - \varepsilon) + P(X'_n \in A, V > 1 - \varepsilon) \\ &= \sum_{i=0}^{k} P(X'_n \in A, V \leq 1 - \varepsilon, X \in B_i) + P(X'_n \in A, V > 1 - \varepsilon, W_n = i) \\ &= \sum_{i=0}^{k} P(Z_i^n \in A, V \leq 1 - \varepsilon, X \in B_i) + P(Z_i^n \in A, V > 1 - \varepsilon, W_n = i) \\ &= \sum_{i=0}^{k} P(Z_i^n \in A) P(V \leq 1 - \varepsilon) P(X \in B_i) + P(Z_i^n \in A) P(V > 1 - \varepsilon) P(W_n = i) \\ &= \sum_{i=0}^{k} P(Z_i^n \in A)((1 - \varepsilon) P(X \in B_i) + \varepsilon P(W_n = i)). \end{aligned}$$

The above is equal to $P(X_n \in A)$ if

$$(1 - \varepsilon) P(X \in B_i) + \varepsilon P(W_n = i) = P(X_n \in B_i) \text{ for all } i \in \{0, 1, \dots k\},$$

which is the case if

$$P(W_n = i) = \frac{P(X_n \in B_i) - (1 - \varepsilon) P(X \in B_i)}{\varepsilon}.$$

As $(B_i)_{i=0}^k$ is a partition and $P(X_n \in B_i) \geq (1 - \varepsilon) P(X \in B_i)$ for $0 \leq i \leq k$ and $n \in \mathbb{N}$,, the above specification of point probabilities gives rise to a valid probability distribution, whence the result follows. □

Before we state our second construction, we need a definition.

Definition 3.3. Let $(M, d, \mathcal{B}_0(M))$ be a metric space, endowed with its associated ball-σ-algebra. Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of separable stochastic variables on some probability space $(\Omega, \mathbb{F}, P)$. If

$$P\left(\bigcap_{\varepsilon\in\mathbb{Q}_+}\bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty}\left(d(X_n, X_N) < \varepsilon\right)\right) = 1,$$

we say that $(X_n)_{n\in\mathbb{N}}$ is almost surely Cauchy

If (M, d) is complete, it is easy to see, that a process, which is almost surely Cauchy, is almost surely convergent: Every sample sequence, which is Cauchy, is by completeness convergent.

Theorem 3.4 (Construction 2)**.** Let $(\Omega, \mathbb{F}, P)$ be a probability space, and let $(M, d, \mathcal{B}_0)$ be a metric space, endowed with its associated ball-σ-algebra. Let $(\varepsilon_j)_{j\in\mathbb{N}}$ be a summable sequence of non-increasing positive numbers and suppose for $j \in \mathbb{N}$. Assume there exists a sequence of finite ball-meaurable partitions $((B_i^j)_{i=0}^{k_j})_{j\in\mathbb{N}}$ of M such that for $p \leq q$ we have $(B_i^q)_{i=0}^{k_q}$ is a refinement of $(B_i^p)_{i=0}^{k_p}$, and let $\lambda : \mathcal{A} \to [0, 1]$ be a finitely additive set function such that $\lambda(M) = 1$, where $\mathcal{A}$ is an sub-algebra of the ball-σ-algebra that contains $\bigcup_{j=1}^{\infty}(B_i^j)_{i=0}^{k_j}$. Furthermore, assume that $\bigcup_{i=1}^{k_j} B_i^j$ is compact for all $j \in \mathbb{N}$ and

1. $diam(B_i^j) = \sup(d(x, y) \mid x, y \in B_i) < 2\varepsilon_j$ for $1 \leq i \leq k_j$ and $j \in \mathbb{N}$
2. $\lambda(B_0^j) < \varepsilon_j$ for all $j \in \mathbb{N}$.

Then there exists a sequence of discrete M-valued stochastic variables $(Y_n)_n$, where $P(Y_j \in B_i^j) = \lambda(B_i^j)$ for $j \in \mathbb{N}$ and $i = 0, \ldots, k_j$, converging almost surely to a separable stochastic variable X.

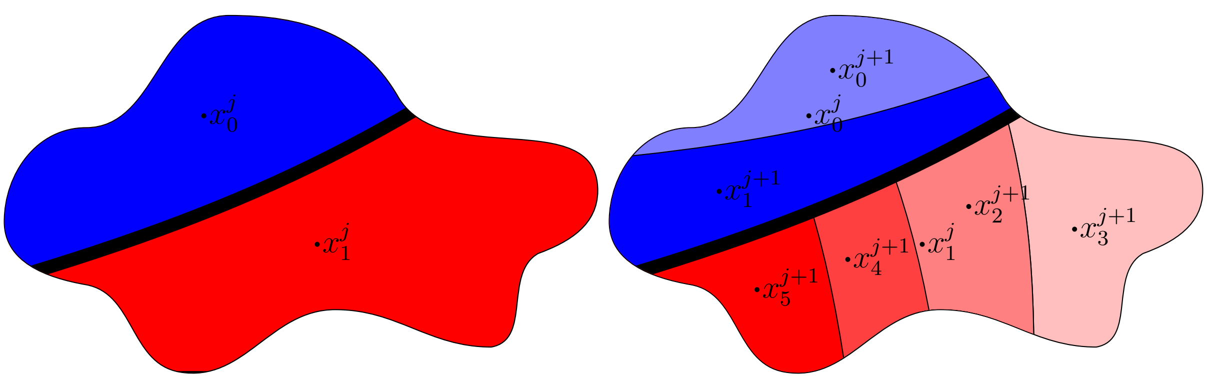

Figure 1: Example of a refinement of a partition.

Proof. Similarly to before, assume $\lambda(B_i^j) > 0$ for $1 \leq i \leq k_j$ and $j \in \mathbb{N}$. For $j \in \mathbb{N}$, let $\{x_0^j, \ldots, x_{k_j}^j\}$ be a finite subset of M, such that $x_i^j \in B_i^j$ for $0 \leq i \leq k_j$. Define the $(L_i^j)_{j\in\mathbb{N}}^{i=0,\ldots k_{j-1}}$ to be a triangular array of independent M-valued stochastic variable with point probabilities

$$P(L_0^1 = x_i^1) = \lambda(B_i^1) \text{ for } i = 0, \ldots k_1,$$

and

$$P(L_i^{j+1} = x_\ell^{j+1}) = \frac{\lambda(B_\ell^{j+1} \cap B_i^j)}{\lambda(B_i^j)} \text{ for } i = 0, \ldots, k_j \text{ and } \ell = 0, \ldots k_{j+1}$$

and $P(L_0^{j+1} = x_\ell^{j+1}) = 0$ if $\lambda(B_0^j) = 0$.
Finite additivity implies that for all $j \in \mathbb{N}$ it holds L_i^{j+1} is almost surely in B_i^j for $i = 0, \ldots, k_j$.
Now let $Y_1 = L_0^1$, and inductively define $(Y_j)_{j \in \mathbb{N}}$ as

$$Y_{j+1} = \begin{cases} L_0^{j+1} & \text{on } (Y_j = x_0^j) \\ L_1^{j+1} & \text{on } (Y_j = x_1^j) \\ \vdots & \\ L_{k_j}^{j+1} & \text{on } (Y_j = x_{k_j}^j) \end{cases}, \tag{3}$$

Note that $Y_j \in B_i^j \implies Y_{j+1} \in B_i^j$ for $i = 1, \ldots, k_j$. We see that

$$(Y_{j+1} = x_\ell^{j+1}) = (L_\ell^{j+1} = x_\ell^{j+1}, Y_j = x_i^j) \text{ where } B_\ell^{j+1} \subseteq B_i^j.$$

We wish to prove that $P(Y_j = x_i^j) = \lambda(B_i^j)$ for $j \in \mathbb{N}$ and $i = 1, \ldots, k_j$. For $j = 1$, it is part of the definition of the Y-variables. For $j+1$, assume the claim holds for j. Since L_i^{j+1} is independent of Y_j, as Y_j is completely determined by L_i^l-variables for $l \leq j$, we get

$$\begin{aligned} P(Y_{j+1} = x_\ell^{j+1}) &= P(L_i^{j+1} = x_\ell^{j+1}, Y_j = x_i^j) \\ &= P(L_i^{j+1} = x_\ell^{j+1}) P(Y_j = x_i^j) \\ &= \frac{\lambda(B_\ell^{j+1} \cap B_i^j)}{\lambda(B_i^j)} \lambda(B_i^j) \\ &= \lambda(B_\ell^{j+1}), \end{aligned}$$

for $B_\ell^{j+1} \subseteq B_i^j$. By induction, the claim holds for all $j \in \mathbb{N}$. Note, as the Y-variables are discrete and therefore can safely be assumed to be separable. On $(Y_n \neq x_0^n)$ for any $n \in \mathbb{N}$, it holds $d(Y_n, Y_m) < 2\varepsilon_n$ for all $m \geq n$. Hence, for all $j \in \mathbb{N}$, it holds

$$\bigcup_{m=j}^{\infty} (d(Y_j, Y_m) \geq 2\varepsilon_j) \subseteq (Y_j = x_0^j).$$

Thus, the Borel-Cantelli lemma implies

$$P\left(\left(\bigcup_{m=j}^{\infty} d(Y_j, Y_m) \geq 2\varepsilon_j\right) \text{ for infinitely many values of } j\right) = 0.$$

In particular $(Y_n)_{n \in \mathbb{N}}$ is almost surely Cauchy. If $Y_n \neq x_0^n$, which happens with probability at least $1 - \varepsilon_n$, then $Y_m \in \bigcup_{i=1}^{k_n} B_i^n$ almost surely for all $m \geq n$. As $\bigcup_{i=1}^{k_n} B_i^n$ is compact, hence complete, $(Y_n)_{n \in \mathbb{N}}$ is, with probability at least $1 - \varepsilon_n$, convergent in $\bigcup_{i=1}^{k_n} B_i^n$. This implies that $(Y_n)_{n \in \mathbb{N}}$ is almost surely convergent in $\bigcup_{n=1}^{\infty} \bigcup_{i=1}^{k_n} B_i^n$ to some variable X, which can be assumed to be seperable as $\bigcup_{n=1}^{\infty} \bigcup_{i=1}^{k_n} B_i^n$ is separable. $\square$

Corallary 3.5. If the assumptions of 3.4 holds and if there for fixed $j \in \mathbb{N}$ exists a sequence $(X_n)_{n\in\mathbb{N}}$ of separable M-valued stochastic variables a strictly increasing sequence of natural numbers $(N_j)_{j\in\mathbb{N}}$ such that

$$P(X_n \in B_i^j) \geq (1-\varepsilon_j)\lambda(B_i^j) \text{ for } 0 \leq i \leq k_j \text{ and } n \geq N_j,$$

then there exist another sequence $(X'_n)_{n\in\mathbb{N}}$ such that $X_n \sim X'_n$ for all $n \in \mathbb{N}$ and such that $(X'_n)_{n\in\mathbb{N}}$ is almost surely convergent.

Proof. Let $(Y_n)_{n\in\mathbb{N}}$ be the sequence constructed using 3.4 and let X be the separable stochastic variable with $Y_n \overset{as}{\to} X$. For every $j \in \mathbb{N}$, we can, by construction 1, construct a new sequence $(\hat{X}_n^j)_{n=N_j}^\infty$ using Y_j as the target variable, such that for all $j \in \mathbb{N}$, we have $\hat{X}_n^j \sim X_n$ for $n \in \mathbb{N}$. Let $(X'_n)_{n\in\mathbb{N}}$ be the sequence given by

$$X'_n = \begin{cases} \hat{X}_n^1 \text{ if } n = N_1, \dots, N_2 - 1 \\ \hat{X}_n^2 \text{ if } n = N_2, \dots, N_3 - 1 \\ \vdots \end{cases}.$$

This constructions can be visualised by the following diagram.

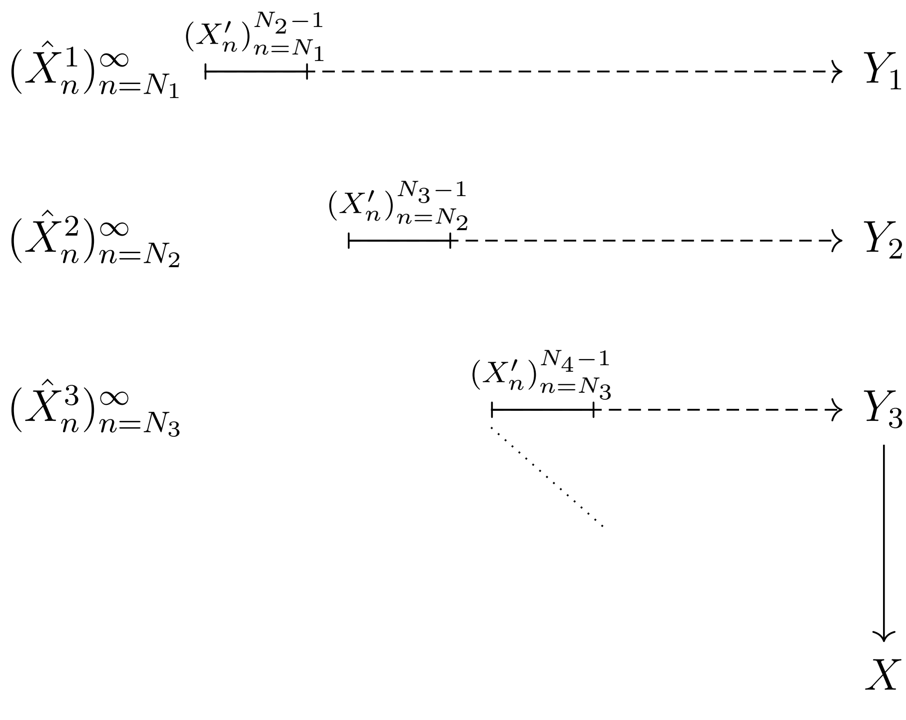

Figure 2: The construction of $(X')_{n\in\mathbb{N}}$. The dashed lines indicate that $(\hat{X}_n^i)_{n\in\mathbb{N}}$ does not necessarily converge to Y_i, but simply approaches with suitably high probability.

Note

$$P\left(\bigcup_{n=N_j}^{N_{j+1}-1} (d(X'_n, Y_j) > 2\varepsilon_j)\right) < 2\varepsilon_j \text{ for all } j \in \mathbb{N},$$

Hence the Borel-Cantelli lemma implies that with probability 1 it holds $d(X'_n, Y_j) \leq 2\varepsilon_j$ eventually, given $n = N_j, \dots, N_{j+1} - 1$. Let

$$\mathcal{E} = \left(\bigcap_{n=N_j}^{N_{j+1}-1} (d(X'_n, Y_j) < 2\varepsilon_j) \text{ eventually as } j \to \infty \right) \cap (Y_j \to X).$$

For $\varepsilon > 0$ and $\omega \in \mathcal{E}$ let $k_\omega \in \mathbb{N}$ be such that $2\varepsilon_{k_\omega} < \varepsilon$ and $d(Y_m(\omega), X(\omega)) < \varepsilon$ for $m \geq k_\omega$. If $n \geq N_{k_\omega}$ then there exists $\ell \geq k_\omega$, such that $N_\ell \leq n \leq N_{\ell+1} - 1$. It then holds

$$d(X'_n(\omega), X(\omega)) \leq d(X'_n(\omega), Y_\ell(\omega)) + d(Y_\ell(\omega), X(\omega)) < 2\varepsilon_\ell + \varepsilon < 2\varepsilon,$$

for $\ell \geq k$. As $P(\mathcal{E}) = 1$ we conclude that $X'_n \overset{as}{\to} X$.

□

3.2 The Russian theorems

We can now prove Prohorov's compactness theorem. The proof given here is heavily inspired by the proof of the same theorem in [Pollard].

Theorem 3.6 (Prohorov's compactness theorem)**.** Let $(M, d, \mathcal{B}_0(M))$ be a metric space, endowed with the associated ball-σ-algebra, and let $(\mu_n)_{n\in\mathbb{N}}$ be an uniformly tight sequence of probability measures on $(M, d, \mathcal{B}_0)$. Then there exists a subsequence $(\mu_{n_k})_{k\in\mathbb{N}}$ and a probability measure μ on $(M, d, \mathcal{B}_0(M))$, such that $(\mu_{n_k})_{k\in\mathbb{N}}$ converges weakly to μ.

Proof. Let $(\varepsilon_j)_{j\in\mathbb{N}}$ be a summable sequence. For each $j \in \mathbb{N}$, let K_j be a compact set, such that $\sup_{n\in\mathbb{N}}(\mu_n(K_j^c)) < \varepsilon_j$. Without loss of generality, we can assume $(K_j)_{j\in\mathbb{N}}$ to be an increasing sequence of sets. Indeed, if $(K_j)_{j\in\mathbb{N}}$ is not increasing, consider the sequence of sets

$$(K'_j)_{j\in\mathbb{N}} = \left(\bigcup_{i=1}^{j} K_i \right)_{j\in\mathbb{N}},$$

then $(K'_j)_{j\in\mathbb{N}}$ clearly has all of the desired properties of $(K_j)_{j\in\mathbb{N}}$, with the added property of being increasing. Let $\mathcal{U}_j$ be an open cover of K_j, consisting of the open ball-measurable sets with diameter at most ε_j. By compactness, we can extract a finite subcover, $\mathcal{F}_j = (F_i^j)_{i=1}^{k_j}$, of $\mathcal{U}_j$. Let $\mathbb{F}_k = \sigma\left(\bigcup_{n=1}^{k} \left(\mathcal{F}_n \cup \{K_n\}\right)\right)$ and let $\mathbb{A}_\infty = \bigcup_{n=1}^{\infty} \mathbb{F}_n$. Note that $\mathbb{F}_m$ is finite for all $m \in \mathbb{N}$ and $\mathbb{A}_\infty$ is countable. As $[0,1]^{\mathbb{N}}$ is compact, there exists a subsequence $(\mu_{n_k})_{k\in\mathbb{N}}$, such that $\mu_{n_k}(A)$ converges in $[0,1]$ for all $A \in \mathbb{A}_\infty$. Let $(X_{n_k})_{k\in\mathbb{N}}$ be a representing sequence of $(\mu_{n_k})_{k\in\mathbb{N}}$.
Define $\lambda : \mathbb{A}_\infty \to [0,1]$, given by

$$\lambda(A) = \lim_{k\to\infty} \mu_{n_k}(A) \text{ for all } A \in \mathbb{A}_\infty.$$

To avoid cluttering our notation, we assume that $n_k = n$.
Linearity of limits implies that λ is finitely additive. By the definition of λ, for every $j \in \mathbb{N}$ and for $A \in \mathbb{F}_\infty$ with $\lambda(A) \neq 0$, there exists $N_j \in \mathbb{N}$ such that

$$\varepsilon_j \lambda(A) \geq \left|\lambda(A) - \mu_n(A)\right| \text{ for all } n \geq N_j \text{ and } A \in \mathbb{F}_j,$$

this in turn implies that

$$\mu_n(A) \geq (1-\varepsilon_j)\lambda(A) \text{ for all } n \geq N_j \text{ and } A \in \mathbb{F}_j. \tag{4}$$

If $\lambda(A) = 0$, then 4 is trivially satisfied. For each finite subcover, $\mathcal{F}_j$, let $(B_i^j)_{i=0}^{k_j}$ be the partition given by $B_0^j = K_j^c$ and

$$B_i^j = \left(F_i^j \setminus \bigcup_{\ell=1}^{i-1} F_\ell^j \right) \cap K_j \text{ for } i \leq k_j.$$

We can assume that $(B_l^{j+1})_{l=0}^{k_{j+1}}$ is a refinement of $(B_i^j)_{i=0}^{k_j}$. If it is not, simply intersect every set in $(B_l^{j+1})_{l=0}^{k_{j+1}}$ with every set in $(B_i^j)_{i=0}^{k_j}$ (this is the maximum or join of the two partitions in the lattice of finite partitions of M) and replace $(B_l^{j+1})_{l=0}^{k_{j+1}}$ with the resulting partition. As $(K_j)_{j\in\mathbb{N}}$ is a an increasing sequence of sets, then no sets of $(B_i^j)_{i=1}^{k_j}$ will intersect K_{j+1}^c, and hence this recounting will not produce a set of diameter more than $2\varepsilon_{j+1}$. By construction 2, we can now generate a sequence of M-valued stochastic variables, $(X_n')_{n\in\mathbb{N}}$, converging almost surely to some separable stochastic variable X, such that $X_n \sim X_n'$ for all $n \in \mathbb{N}$. As almost sure convergence imply convergence in probability, which in turn implies convergence in distribution, we have that $X_n'(P) \to X(P)$ weakly. As $\mu_n = X_n'(P)$ for all $n \in \mathbb{N}$, we conclude that $\mu_n \to X(P)$ weakly. □

Lemma 3.7. Let $(M, d, \mathcal{B}_0)$ be a metric space, endowed with the associated ball-σ-algebra. Let μ be a tight probability measure on $\mathcal{B}_0(M)$. For any $\varepsilon > 0$, there exists a finite ball-measurable partition, $(B_i)_{i=0}^k$, of M such that

1. $diam(B_i) = \sup(d(x,y) \mid x, y \in B_i) \leq 2\varepsilon$ for $1 \leq i \leq k$,
2. ∂B_i is a null set for $1 \leq i \leq k$,
3. $\mu(B_0) < \varepsilon$.

Proof. Let C be a closed and separable set such that $\mu(C) = 1$. For $\varepsilon > 0$ and $x \in C$, let $f_x : M \to \mathbb{R}$ be a Lipschitz and ball-measurable function, such that

$$1_{B(x,\frac{\varepsilon}{2})} \leq f_x \leq 1_{B(x,\varepsilon)}.$$

Let s_x be the survival function of f_x , that is $s_x(t) = \mu(f_x > t)$. Let $y \in (0,1)$ be a point where s_x is continuous, and consider the set

$$A_x = (f_x > y).$$

As f_x is continuous and ball-measurable, A_x is open and ball-measurable. Note that

$$A_x \subseteq (f_x > 0) \subseteq B(x, \varepsilon),$$

and so $diam(A_x) \leq 2\varepsilon$. We also note that $A_x \subseteq (f_x \geq y)$. As $(f_x \geq y)$ is, by continuity, closed, this implies that the closure of A_x is also contained in $(f_x \geq y)$. And so

$$\partial A_x = \overline{A_x} \setminus A_x \subseteq (f_x \geq y) \setminus (f_x > y) = (f_x = y).$$

As y was chosen as a point of continuity of s, $\mu(f_x = y) = 0$, hence ∂A_x is a μ-nullset. As $f_x(x) = 1$, the family $(A_x)_{x\in C}$ is an open cover of C. By separability of C, we can extract a countable subcover $(A_{x_n})_{n\in\mathbb{N}}$. Define the sequence of sets $(B_n)_{n\in\mathbb{N}}$ as

$$B_n = A_{x_n} \setminus \bigcup_{i=1}^{x_{n-1}} A_{x_i}.$$

That $diam(B_n) \leq 2\varepsilon$ is a consequece of the natural disjointification and that ∂B_n is a nullset, for all $n \in \mathbb{N}$ is an consequence of the general topological fact that $\partial B_n \subseteq \bigcup_{i=1}^{n} \partial A_i$. Another consequence of the disjointification of is that $\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n$, and hence

$$\sum_{n=1}^{\infty} \mu(B_n) = 1.$$

Therefore, we can find $k \in \mathbb{N}$ such that

$$\sum_{n=1}^{k} \mu(B_n) > 1 - \varepsilon.$$

If we let $B_0 = \left(\bigcup_{k+1}^{\infty} B_n\right)^c$, then we are done. $\square$

Theorem 3.8. (Skorokhod's representation theorem) Let $(M, d, \mathcal{B}_0(M))$ be a metric space, endowed with the associated ball-σ-algebra, and let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence of probability measures on $(M, d, \mathcal{B}_0)$. Suppose that $\mu_n \to \mu$ for some tight probability probability measure. Then there exists stochastic variables, X and $(X_n)_{n\in\mathbb{N}}$, on some probability space $(\Omega, \mathbb{F}, P)$ such that $(X_n)_{n\in\mathbb{N}}$ represents $(\mu_n)_{n\in\mathbb{N}}$, $X(P) = \mu$ and $X_n \overset{as}{\to} X$, where X is separable.

Proof. Let Let $(\varepsilon_j)_{j\in\mathbb{N}}$ be the sequence given by $\varepsilon_j = 2^{-j}$ and $(B_i)_{i=0}^{k}$ be the partition given in the previous lemma, generated by μ. For $i = 1, \ldots, k$, let D_i be a ball-measurable of μ-measure zero such that $\partial B_i \subseteq D_i$. Note that μ is concentrated on separable set $C \setminus D_i$. It is easy to see that 1_{B_i} is continuous on $C \setminus D_i$ for all $i = 1, \ldots, k$, hence, by the convergence lemma, we have that $P(X_n \in B_i) = \mu_n(B_i) \to \mu(B_i) = P(X \in B_i)$ for $i = 1, \ldots, k$. This allows us to conclude that, for $j \in \mathbb{N}$ there exists N_j, such that

$$P(X_n \in B_i) \geq (1 - \varepsilon_j) P(X_n \in B_i),$$

for $n \geq N_j$ and $i = 1, \ldots k$. Hence, for every $j \in \mathbb{N}$, we can, for $n = N_j, \ldots N_{j+1} - 1$ construct variables X'_n, such that $X'_n \sim X_n$ and

$$P\left(\bigcup_{n=N_j}^{N_{j+1}-1} (d(X'_n, X)) > 2\varepsilon_j\right) \leq 2\varepsilon_j.$$

Thus, by the Borel-Cantelli lemma,

$$P\left(\bigcup_{n=N_j}^{N_{j+1}-1} (d(X'_n, X)) > 2\varepsilon_j) \text{ for infinitely many values of } j\right) = 0,$$

Hence $X'_n \overset{as}{\to} X$. $\square$

4 Metrizability of weak convergence

The topologically inclined reader might find the name ”Prohorov’s compactness theorem” a bit cocky. Even if we take for granted that weak convergence induces a topology, what Prohorov’s compactness theorem shows is not a criterion for compactness in the space of probability measures on some metric space, it is at best a criterion for relative sequential compactness. In this section we will investigate this problem by introducing a metric on (a subset of) the space of probability measures on a metric space.

4.1 The bounded-Lipschitz (pseudo)metric

Let $BL_0(M)$ denote the set of bounded, ball-measurable Lipschitz function on some metric space (M, d). Consider the mapping $\|\cdot\|_{BL} : BL_0(M) \to [0, \infty)$ given by

$$\|f\|_{BL} = \|f\|_\infty + \|f\|_{Lip} \;\; f \in BL_0(M),$$

where $\|f\|_{Lip}$ is the infimum of all Lipschitz constants for f. Let $Pr(M)$ be the set of probabibilty measures on the metric space $(M, d, \mathcal{B}_0)$ and let $Pt(M)$ be set of tight probability measures on the metric space $(M, d, \mathcal{B}_0)$. We define the bounded-Lipschitz metric as the mapping $p : Pr(M) \times Pr(M) \to \mathbb{R}$, given by

$$p(\mu, \nu) = \sup \left(\left| \int f \; d\mu - \int f \; d\nu \right| \;\mid\; \|f\|_{BL} \leq 1 \right).$$

Proposition 4.1. Let $(M, d, \mathcal{B}_0(M))$ be a metric space endowed with its associated ball-σ-algebra. Then the bounded-Lipschitz metric is a pseudometric on $Pr(M)$.

Proof. Symmetry and non-negativity follows trivially from symmetry and non-negativity of $(x, y) \longmapsto |x - y|$ and so does $p(\mu, \mu) = 0$ for all $\mu \in Pr(M)$. Now, let $\mu, \nu, \gamma \in Pr(M)$ and $f \in BL_0(M)$ with $\|f\|_{BL} \leq 1$, then it holds that

$$\begin{aligned} \left| \int f \; d\mu - \int f \; d\nu \right| &\leq \left| \int f \; d\mu - \int f \; d\gamma \right| + \left| \int f \; d\gamma - \int f \; d\nu \right| \\ &\leq p(\mu, \gamma) + p(\gamma, \nu) \end{aligned}$$

which implies that $p(\mu, \nu) \leq p(\mu, \gamma) + p(\gamma, \nu)$. □

If we restrict our scope to $Pt(M)$, then p is actually a proper metric.

Proposition 4.2. Let $(M, d, \mathcal{B}_0(M))$ be a metric space endowed with its associated ball-σ-algebra. Then the bounded-Lipschitz metric is a metric on $Pt(M)$.

Proof. Symmetry, non-negativity and the triangle inequality all follow from the previous proposition.

Let μ and ν be tight probability measures on $\mathcal{B}_0(M)$ with $p(\mu, \nu) = 0$. Let f be a non-zero function in $BL_0(M)$, then we see

$$\begin{aligned}\left|\int f \, d\mu - \int f \, d\nu\right| &= \|f\|_{BL}\left|\int \frac{f}{\|f\|_{BL}} \, d\mu - \int \frac{f}{\|f\|_{BL}} \, d\nu\right| \\ &\le \|f\|_{BL} \, p(\mu, \nu) \\ &= 0\end{aligned}$$

and hence $\mu = \nu$ by 1.10. For $f = 0$, the above inequality holds trivially. □

Theorem 4.3. Let $(M, d, \mathcal{B}_0(M))$ be a metric space endowed with its associated ball-σ-algebra. Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence of probability measures on $\mathcal{B}_0(M)$ and let μ be a tight probability on $\mathcal{B}_0(M)$. Then

$$\mu_n \to \mu \text{ weakly if and only if } p(\mu_n, \mu) \to 0.$$

Proof. Assume that $p(\mu_n, \mu) \to 0$. Then for non-zero $f \in BL_0(M)$ it holds

$$\begin{aligned}\left|\int f \, d\mu_n - \int f \, d\mu\right| &= \|f\|_{BL}\left|\int \frac{f}{\|f\|_{BL}} \, d\mu_n - \int \frac{f}{\|f\|_{BL}} \, d\mu\right| \\ &\le \|f\|_{BL} \, p(\mu_n, \mu) \to 0,\end{aligned}$$

and hence $\mu_n \to \mu$ weakly by the convergence lemma.
Now assume $\mu_n \to \mu$ weakly. By Skorokhod's representation theorem, there exists representing stochastic variables $(X_n)_{n\in\mathbb{N}}$ of $(\mu_n)_{n\in\mathbb{N}}$ and X of μ respectively, on some probabilty space $(\Omega, \mathbb{F}, P)$, such that $X_n \overset{as}{\to} X$. We see that, for $f \in BL_0(M)$ with $\|f\|_{BL} \le 1$ and for $\varepsilon > 0$, it holds

$$\begin{aligned}\left|\int f \, d\mu_n - \int f \, d\mu\right| &= \left|\int f(X_n) - f(X) \, dP\right| \\ &\le \int \big|f(X_n) - f(X)\big| \, dP \\ &= \int_{(d(X_n,X)<\varepsilon)} \big|f(X_n) - f(X)\big| \, dP + \int_{(d(X_n,X)\ge\varepsilon)} \big|f(X_n) - f(X)\big| \, dP \\ &\le \int_{(d(X_n,X)<\varepsilon)} \|f\|_{Lip} \, d(X_n, X) \, dP + 2\int_{(d(X_n,X)\ge\varepsilon)} \|f\|_\infty \, dP \\ &\le \|f\|_{Lip} \, \varepsilon + 2\|f\|_\infty \, P(d(X_n, X) \ge \varepsilon) \\ &\le (\|f\|_\infty + \|f\|_{Lip})\varepsilon + 2(\|f\|_\infty + \|f\|_{Lip})P(d(X_n, X) \ge \varepsilon) \\ &\le \varepsilon + 2P(d(X_n, X) \ge \varepsilon).\end{aligned}$$

Hence, it holds that

$$p(\mu_n, \mu) \le \varepsilon + 2P(d(X_n, X) \ge \varepsilon)$$

As $X_n \overset{as}{\to} X \implies X_n \overset{P}{\to} X$, it holds that $P(d(X_n, X) \ge \varepsilon) < \varepsilon$ eventually, and therefore $p(\mu_n, \mu) < 3\varepsilon$ eventually. As this holds for all $\varepsilon > 0$, we see that $p(\mu_n, \mu) \to 0$. □

If (M, d) is separable and complete then $Pr(M) = Pt(M)$ by 1.9, and we see that $Pr(M)$ equipped with topology of weak convergence of measures is metrizable. If (M, d) is not separable, this seems to not be the case. If we, however, restrict ourselves to the weakly convergent sequences with tight limit, then weak convergence is metrizable.

4.2 Applications of metrizability of weak convergence

This knowledge of the topological properties of weak convergence allows us to restate Prohorov's compactness theorem in topological terms.

Theorem 4.4. Let $(M, d, \mathcal{B}_0(M))$ be a metric space endowed with its associated ball-σ-algebra. If a subset of $Pt(M)$ is uniformly tight, it is relatively compact.

We will now prove a "converse" of the Prohorov theorem. While we did not require additional structure on our metric in Prohorov's theorem, the "converse" does seem to require both completeness and separability, hence the quotation marks. Firstly, we need a lemma.

Lemma 4.5 (Portmanteau lemma)**.** Let $(M, d, \mathcal{B}_0(M))$ be a metric space endowed with its associated ball-σ-algebra. Let $(\mu_n)_{n\in\mathbb{N}}$ be sequence of probability measures, converging weakly to some tight measure μ. Then for all open, ball-measurable subsets, U, of M, it holds

$$\liminf_{n\to\infty} \mu_n(U) \geq \mu(U).$$

Proof. In view of Skorokhod's representation theorem, the proof of corollary 6.21 in [Hansen] is perfectly adaptable to this result. □

Theorem 4.6 (2nd half of Prohorov's compactness theorem)**.** Let $(M, d, \mathcal{B}(M))$ be a complete, separable metric space, with its associated Borel-σ-algebra. If $\Gamma \subseteq Pr(M)$ is relatively compact, it is uniformly tight.

Proof. We have already noted that if M is separable and complete, then $(Pr(M), p)$ is a metric space. Hence, if a subset Γ is relatively compact, it is relatively sequentially compact. Let $\{x_n\}_{n\in\mathbb{N}}$ be a countable, dense subset of M and let $\varepsilon > 0$. Let $(U_n)_{n\in\mathbb{N}}$ be an increasing sequence of open ball measurable sets, with $\lim_{n\to\infty} U_n = M$. We wish to show that for every $\varepsilon > 0$, there exists $n \in \mathbb{N}$, such that $\mu(U_n) > 1 - \varepsilon$ for all $\mu \in \Gamma$. Assume for a contradiction that for each $n \in \mathbb{N}$, there exists a probability measure, $\mu_n \in \Gamma$, such that $\mu_n(U_n) \leq 1 - \varepsilon$. By relatively sequentially compactness, there exists a subsqeunce $(\mu_{n_k})_{k\in\mathbb{N}}$ and a measure μ (not necessarily in Γ), such that $\mu_{n_k} \to \mu$ weakly. By the Portmanteau lemma, upwards continuity and the fact that for any fixed n we have $U_n \subseteq U_{n_k}$ eventually, it holds that

$$\begin{aligned}
1 &= \lim_{n\to\infty} \mu(U_n) \\
&\leq \limsup_{n\to\infty} \liminf_{n_k\to\infty} \mu_{n_k}(U_n) \\
&\leq \limsup_{n\to\infty} \liminf_{n_k\to\infty} \mu_{n_k}(U_{n_k}) \\
&\leq 1 - \varepsilon,
\end{aligned}$$

a clear contradiction. Now let $(B_n^i)_{n\in\mathbb{N}}$ be the open cover of M consisting of the sets of the form $B_n^i = B(x_n, \frac{1}{i})$. As we noted before, for all $k \in \mathbb{N}$ there exists n_k, such that $\mu\left(\bigcup_{n=1}^{n_k} B_n^k\right) > 1 - \frac{\varepsilon}{2^k}$ for all $\mu \in \Gamma$. We see

$$\overline{\bigcap_{k\in\mathbb{N}} \bigcup_{n=1}^{n_k} B_n^k}$$

is compact by 1.7. We see

$$\mu\left(\overline{\bigcap_{k\in\mathbb{N}} \bigcup_{n=1}^{n_k} B_n^k}\right) > 1 - \varepsilon.$$

As this holds for all $\mu \in \Gamma$ we conclude Γ is uniformly tight. $\square$

For $A \in \mathcal{B}(M)$ of a separable metric space, M, and $\varepsilon > 0$ we define the set A^ε as $A^\varepsilon = (x \in M \mid d(x, A) < \varepsilon) = \bigcup_{x\in A} B(x, \varepsilon)$. It is easy to see that A^ε is open, hence Borel.

Theorem 4.7. Let (M, d) be a separable and complete metric space. Then $(Pr(M), p)$ is separable and complete.

Proof. Consider the function $b : Pr(M) \times Pr(M) \to \mathbb{R}_+$ given by

$$b(\mu, \nu) = \inf\left(\varepsilon > 0 \mid \forall A \in \mathcal{B}(M) : \mu(A) \leq \nu(A^\varepsilon) + \varepsilon\right).$$

(It can be shown, though we will not do it here, that b is a metric equivalent to the bounded-Lipschitz metric. b is commonly referred to as the Prohorov-Levy metric). Let A be a Borel subset of M, let $\varepsilon > 0$ and let $f : M \to \mathbb{R}$ be given by $f(x) = \max\left(0, 1 - \frac{d(x,A)}{\varepsilon}\right)$. By the usual stability results for bounded and Lipschitz functions respectively, f is in $BL(M)$ and $1_A \leq f \leq 1_{A^\varepsilon}$. and $\|f\|_{BL} \leq 1 + \frac{1}{\varepsilon}$, hence for probability measures μ and ν, it holds

$$\begin{aligned}
\mu(A) &\leq \int f \, d\mu \\
&= \int f \, d\nu + \int f \, d\mu - \int f \, d\nu \\
&\leq \int f \, d\nu + \left(1 + \frac{1}{\varepsilon}\right) p(\mu, \nu) \\
&\leq \nu(A^\varepsilon) + \left(1 + \frac{1}{\varepsilon}\right) p(\mu, \nu)
\end{aligned}$$

hence $b(\mu, \nu) \leq \max(\varepsilon, (1 + \frac{1}{\varepsilon})p(\mu, \nu))$. Therefore, if $p(\mu, \nu) \leq \varepsilon^2$, then $b(\mu, \nu) \leq \varepsilon^2 + \varepsilon$. As b clearly is bounded by 1, it is easy to see that $b(\mu, \nu) \leq 2\sqrt{p(\mu, \nu)}$; Let $\varepsilon = \sqrt{p(\mu, \nu)}$. As $p(\mu, \nu) \leq \sqrt{p(\mu, \nu)}^2$, if $p(\mu, \nu) \leq 1$ we have that $b(\mu, \nu) \leq p(\mu, \nu) + \sqrt{p(\mu, \nu)} \leq 2\sqrt{p(\mu, \nu)}$, and if $p(\mu, \nu) > 1$, the inequality is a triviality.

Completeness:

We will show completeness by showing that any Cauchy sequence, $(\mu_n)_{n\in\mathbb{N}}$, in (M, d) is uniformly tight, Prohorov's theorem will provide a convergent subsequence. Now let $(\mu_n)_{n\in\mathbb{N}}$ be a p-Cauchy sequence and let $(x_n)_{n\in\mathbb{N}}$ be a countable dense subset of M. We want to show that for any rational $\eta > 0$, we can find a finite number of balls with centres in $(x_n)_{n\in\mathbb{N}}$, and radii

of η, whose union have probability at least $1-2\eta$.
By the above considerations, if $p(\mu_m,\mu_n) < \left(\frac{\eta}{2}\right)^2$ for $n,m \geq N$, then $b(\mu_m,\mu_n) < \eta$ for $n,m \geq N$. Let $(A_n)_{n\in\mathbb{N}}$ be the open cover consisting of open balls given by $A_n = B(x_n,\eta)$. Choose $m \in \mathbb{N}$ such that $\mu_n\left(\bigcup_{n=1}^m A_n\right) > 1-\eta$ for $n \leq N_\eta$. Now let $(B_n)_{n\in\mathbb{N}}$ be the open cover consisting of open balls given by $B_n = B(x_n, 2\eta)$. If $x \in \left(\bigcup_{j=1}^m A_j\right)^\eta$, then $d\left(x, \bigcup_{n=1}^m A_n\right) < \eta$. This implies that there exists $i \leq m$, such that $d\left(x, A_i\right) < \eta$, hence $x \in B(x_i, 2\eta)$. Hence for $k \geq N_\eta$, it holds

$$\begin{aligned}\mu_k\left(\bigcup_{n=1}^m B_n\right) &= \mu_k\left(\left(\bigcup_{n=1}^m A_n\right)^\eta\right)\\ &\geq \mu_{N_\eta}\left(\bigcup_{n=1}^m A_n\right) - \eta\\ &> 1-2\eta.\end{aligned}$$

For $k \leq N_\eta$, it holds

$$\mu_k\left(\bigcup_{n=1}^m B_n\right) \geq \mu_k\left(\bigcup_{n=1}^m A_n\right) > 1-\eta > 1-2\eta.$$

Hence we can find a finite number of balls with union having arbitrarily large probability for each measure in $(\mu_n)_{n\in\mathbb{N}}$. Therefore, we can choose relatively compact sets of the form

$$\bigcap_{k\in\mathbb{N}}\bigcup_{n=1}^m B\left(x_n, \frac{1}{k}\right),$$

to have arbitrarily large probability, and we see $(\mu_n)_{n\in\mathbb{N}}$ is uniformly tight, as those sets are relatively compact by 1.7. Prohorov's compactness theorem then implies that there exists a convergent subsequence of $(\mu_n)_{n\in\mathbb{N}}$, and as $(\mu_n)_{n\in\mathbb{N}}$ is p-Cauchy, it is convergent in the p-metric.
Separability:
Let $\mu \in Pr(M)$ and let Y be a M-valued stochastic variable representing μ. Let $(\varepsilon_j)_{j\in\mathbb{N}}$ be a summable sequence, and for each j, let $(K_j)_{j\in\mathbb{N}}$ be a sequence of compacta such that $\mu(K_j) = P(Y \in K_j) \geq 1-\epsilon_j$. By construction 2 we can construct a sequence of discrete stochastic variables $(Y_n)_{n\in\mathbb{N}}$. Using the notation of the proof of construction 2, the Borel-Cantelli lemma implies $P(Y_n \notin B_0^n \text{ eventually}) = 1$, and hence we conclude that $Y_n \to Y$ almost surely, hence in distribution. This implies that the set

$$\mathcal{R} = \left\{\sum_{i=1}^n r_i\delta_{y_i} \,|r_i \in \mathbb{R}_+, \sum_{i=1}^n r_i = 1, y_1, y_2, \ldots y_n \in M\right\}$$

is dense in $Pr(M)$. Let $\mathcal{C}$ be a countable dense set in M. To see that the countable set

$$\mathcal{Q} = \left\{\sum_{i=1}^n q_i\delta_{z_i} \,|q_i \in \mathbb{Q}_+, \sum_{i=1}^n q_i = 1, z_1, z_2, \ldots z_n \in \mathcal{C}\right\}$$

is dense in $\mathcal{R}$, let $\sum_{i=1}^n r_i \delta_{y_i} \in \mathcal{R}$ and let $(q_i^m)_{m\in\mathbb{N}}^{i=1,\dots n}$ be sequences of positive rational numbers converging to r_i respectively and $(z_i^n)_{n\in\mathbb{N}}^{i=1,\dots n}$ be sequences of points in $\mathcal{C}$ converging to y_i respectively.
Let f be a bounded, continuous function. By linearity of existing limits and integrals and continuity of f, we get

$$\begin{aligned}\int f \ d\sum_{i=1}^n \frac{q_i^m}{\sum_{i=1}^n q_i^m}\delta_{z_i^m} &= \sum_{i=1}^n \frac{q_i^m}{\sum_{i=1}^n q_i^m}\int f \ d\delta_{z_i^m} \\ &= \sum_{i=1}^n \frac{q_i^m}{\sum_{i=1}^n q_i^m} f(z_i^m) \\ &\to \sum_{i=1}^n r_i f(y_i) \\ &= \sum_{i=1}^n r_i \int f \ d\delta_{y_i} \\ &= \int f \ d\sum_{i=1}^n r_i \delta_{y_i},\end{aligned}$$

as $m \to \infty$. Hence $\sum_{i=1}^n q_i^m \delta_{z_i^m} \to \sum_{i=1}^n r_i \delta_{y_i}$ weakly. Hence $\mathcal{Q}$ is dense $\mathcal{R}$, and thus $\mathcal{Q}$ is dense in $Pr(M)$. □

Remark. While both completeness and separablity of M is needed to prove completeness of $Pr(M)$, only separablity of M is needed to prove separablity of $Pr(M)$.

5 Cádlág spaces

In this section we will consider an important example of a metric space in the context of probability theory, the cádlág space $D[0,1]$. Most stochastic processes arising in applications will almost surely have cádlág sample paths, and hence such a stochastic process can be viewed as a $D[0,1]$-valued stochastic variable. We have already seen that under the uniform metric $D[0,1]$ is complete, but not separable. This is all well and good if the weak limit of the sequence in question is tightly distributed, or at least concentrates on a closed and separable subset of $D[0,1]$, such as the if the limit is the Brownian motion. However, in order to utilise the full scope of tools we have seen so far, we need a metric which is both separable and complete, or at least, making $D[0,1]$ a Polish space.

5.1 The Skorokhod topology

Uniform closeness between two continuous functions f and g on $[0,1]$ indicates that, for all $t \in [0,1]$, $f(t)$ can be uniformly moved onto $g(t)$, i.e. the graph f can be uniformly "wiggled" onto the graph of g, without changing the time variable. In order to study $D[0,1]$ as a Polish space, we will need to allow uniformly "wiggles" in space as well as in time. In this section, we will consider a metric, which formalises this idea.
In the following we will be studying the group of increasing C^1 diffeomorphism from $[0,1]$ to itself. We note that if $f : [0,1] \to [0,1]$ is continuous and increasing and $x : [0,1] \to \mathbb{R}$ is cádlág, then $x \circ f$ is also cádlág.

Definition 5.1 (Skorokhod metric)**.** Let Λ be the group of increasing C^1 diffeomorphisms $\lambda : [0,1] \to [0,1]$. Let $I \in \Lambda$ be the identity mapping on $[0,1]$. The Skorokhod metric $d : D[0,1] \times D[0,1] \to \mathbb{R}_+$ is given by

$$d(x,y) = \inf_{\lambda \in \Lambda} \left(\max(\|\lambda - I\|_\infty, \|x - y \circ \lambda\|_\infty) \right).$$

In the above $\|\lambda - I\|_\infty$ can be viewed as a uniform measurement of the deformation of time and $\|x - y \circ \lambda\|_\infty$ can be viewed as the uniform distance between x and y after the time variable has been deformed.
Many of the technical arguments in this sections depends on the existence of deformations with specific qualities. In order to make such arguments, we need a construction of one of those deformations.

Lemma 5.2. Let $\theta : 0 = t_0 < \ldots < t_n = 1$ and $\sigma : 0 = s_0 < \ldots < s_n = 1$ be finite subdivisions of $[0,1]$ and for $i = 1, \ldots, n$, define

$$c_i = \frac{t_i - t_{i-1}}{s_i - s_{i-1}}.$$

Let $\varepsilon > 0$, then there exists an increasing C^1-function $\varphi : [0,1] \to [0,1]$ with non-vanishing derivative, such that

1. $\varphi(s_i) = t_i$ for $i = 1, \ldots, n$
2. $|\varphi'(\xi) - c_i| \leq |\min_{i=1,\ldots n} c_i - \max_{i=1,\ldots,n} c_i| + \varepsilon$ and $\xi \in [s_i - s_{i-1}]$ and $i = 1, \ldots, n$
3. $\|\phi - I\|_\infty < \max_{i=0,\ldots,n}(|t_i - s_i|) + \varepsilon$.

Proof. Let $c_0 = \min_{i=1,\dots,n} c_i$ Note $c_i > 0$ for $i = 1,\dots,n$, and thus $c_0 > 0$. Let $\delta \in \left(0, \frac{\min_{i=1,\dots,n}(s_i - s_{i-1})}{2}\right)$ and let $g : [0,1] \to \mathbb{R}$ be the piecewise affine function that linearly interpolates from the point (s_{i-1}, c_0) to $(s_{i-1}+\delta, h_i)$ to $(s_i - \delta, h_i)$ to (s_i, h_i), for $i = 1,\dots n$ and where $(h_i)_{i=1}^n$ is yet to be defined. We note that the graph of g forms a finite sequence of (possibly flat) trapezia with height $(h_i - c_0)$, upper width $(s_i - s_{i-1} - 2\delta)$ and lower width $(s_i - s_{i-1})$, elevated from the time axis by c_0.

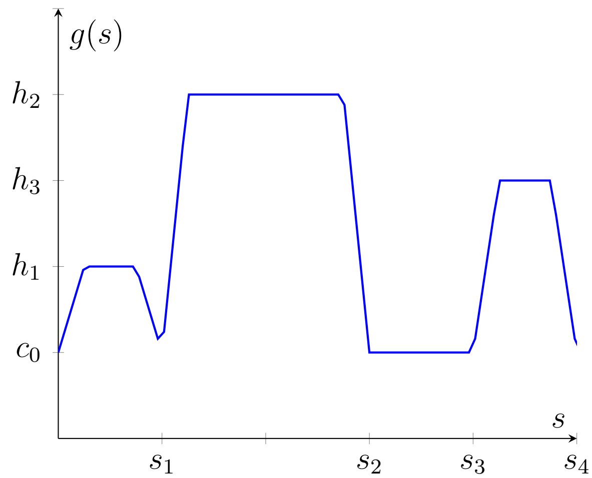

Figure 3: An example of how g could look.

Now let $\varphi : [0,1] \to \mathbb{R}$ be given by $\varphi(s) = \int_0^s g(x)\, dx$. For $i = 1,\dots,n$ we adjust h_i such that

$$\int_{s_{i-1}}^{s_i} g(x)\, dx = (s_i - s_{i-1})c_0 + (h_i - c_0)(s_i - s_{i-1} - \delta) = t_i - t_{i-1} \text{ for } i = 1,\dots n. \tag{5}$$

It is elementary to see that the above equation implies that $h_i = c_0$ or $h_i = \frac{t_i - t_{i-1} - \delta c_0}{s_i - s_{i-1} - \delta}$ for $i = 1,\dots n$. If $h_i \neq c_0$, then (5) implies that $h_i > c_i$. We also note that for $i = 1,\dots n$ it holds that $h_i \to c_i$ for $\delta \to 0$. This implies for $i = 1,\dots n$ we can bound $\varphi' = g$ on $[s_i, s_{i-1}]$ as in the statement of the lemma. By the fundamental theorem of calculus, $\varphi(s_i) - \varphi(s_{i-1}) = \int_{s_{i-1}}^{s_i} g(x)\, dx$ and we note $\varphi(0) = 0$, thus it holds $\varphi(s_i) = t_i$ for $i = 1,\dots,n$.
It remains to be shown that

$$\|\varphi - I\|_\infty < \max_{i=0,\dots,n}(|t_i - s_i|) + \left|\min_{i=1,\dots,n} c_1 - 1\right|\delta,$$

If $h_i \le 1$ for $i = 1,\dots n$. Then $(\varphi - I)'$ is non-positive on $[s_{i-1}, s_i]$, and hence

$$\sup_{s\in[s_{i-1},s_i]}\left(\left|(\varphi - I)'(s)\right|\right) = \max(\left|\varphi(s_i) - s_i\right|, \left|\varphi(s_{i-1}) - s_{i-1}\right|) = \max(|t_i - s_i|, |t_{i-1} - s_{i-1}|).$$

If $h_i > 1$ then $(\varphi - I)' = 0$ has two solutions on $[s_{i-1}, s_i)$ for every $i = 1,\dots,n$, and therefore $\varphi - I$ has two stationary points in $[s_{i-1} - s_i)$, respectively $s^* \in [s_{i-1}, s_{i-1}+\delta)$ and $s^{**} \in [s_i - \delta, s_i)$. By the mean value theorem, there exists $\xi \in (s^{**}, s_i)$ such that

$$\left|(\varphi(s^{**}) - s^{**}) - (\varphi(s_i) - s_i)\right| = \left|(\varphi'(\xi) - 1)(s_i - s^{**})\right|.$$

As $c_0-1 \leq \varphi'(\xi)-1 < 0$, it holds $\left|\varphi'(\xi) - 1\right| \leq |c_0 - 1|$ and thus $\left|(\varphi(s^{**}) - s^{**}) - (\varphi(s_i) - s_i)\right| \leq |c_0 - 1|\,\delta$. Hence it holds

$$\begin{aligned}\left|\varphi(s^{**}) - s^{**}\right| \leq& \left|(\varphi(s^{**}) - s^{**}) - (\varphi(s_i) - s_i)\right| + \left|(\varphi(s_i) - s_i)\right| \\ \leq& |c_0 - 1|\,\delta + |t_i - s_i|\,.\end{aligned}$$

Similar computations show $\left|\varphi(s^*) - s^*\right| \leq |c_0 - 1|\,\delta + |t_{i-1} - s_{i-1}|$. If $\delta < \frac{\varepsilon}{|c_0-1|}$, we conclude

$$\|\varphi - I\| \leq \max_{i=1,\ldots,n}(|t_i - s_i|) + \varepsilon.$$

□

We will refer to φ, cosntructed above, as the C^1 interpolation of the subdivisions $\theta : 0 = t_0 < \ldots < t_n = 1$ and $\sigma : 0 = s_0 < \ldots < s_n = 1$.
Before we continue, we need a technical lemma regarding time deformations.

Lemma 5.3. Let $\lambda \in \Lambda$ and $x, y \in D[0,1]$, then

$$\|x \circ \lambda\|_\infty = \|x\|_\infty\,.$$

Proof. For all $t \in [0,1]$, it holds that $\lambda(t) \in [0,1]$. Hence

$$\left|x(\lambda(t))\right| \leq \|x\|_\infty \text{ for all } t \in [0,1],$$

from which we surmise $\|x \circ \lambda\|_\infty \leq \|x\|_\infty$. As $x \circ \lambda$ is cádlág and $\lambda^{-1} \in \Lambda$, we note that the above implies

$$\|x\|_\infty = \left\|x \circ \lambda \circ \lambda^{-1}\right\|_\infty \leq \|x \circ \lambda\|_\infty\,,$$

and from this we conclude the desired result. □

From the lemma above, we note these immediate consequences for $\lambda \in \Lambda$ and $x, y \in D[0,1]$

$$\begin{aligned}\|\lambda - I\|_\infty =& \left\|\lambda^{-1} - I\right\|_\infty \text{ and} \\ \|x - y \circ \lambda\|_\infty =& \left\|x \circ \lambda^{-1} - y\right\|_\infty\,.\end{aligned}$$

Note that the second claim also holds for deformations in Λ.

Proposition 5.4. The Skorokhod metric is a well-defined metric on $D[0,1]$.

Proof. As $\max(\|\lambda - I\|_\infty, \|x - y \circ \lambda\|_\infty)$ is finite when $\lambda = I$ and $I \in \Lambda$, the mapping is well-defined. Non-negativity and the fact that $d(x,x) = 0$ are obvious from the definition. If $d(x,y) = 0$, then there exists a sequence $(\lambda_n)_{n\in\mathbb{N}} \subseteq \Lambda$ such that

$$\max(\|\lambda_n - I\|_\infty, \|x - y \circ \lambda_n\|_\infty) \to 0$$

as $n \to \infty$. This is the case if and only if $\|\lambda_n - I\|_\infty \to 0$ and $\|x - y \circ \lambda_n\|_\infty \to 0$ as $n \to \infty$. In other words: $\lambda_n \to I$ and $y \circ \lambda_n \to x$ uniformly as $n \to \infty$. Hence, for all $t \in [0,1]$, it holds that $\lambda_n(t) \to t$ and $y \circ \lambda_n(t) \to x(t)$. If t is a continuity point of y, then $y(\lambda(t_n)) \to y(t)$, and so $y(t) = x(t)$. As right continuous functions at most can have countably many discontinuities,

it holds that the complement of the set of discontinuities is dense, hence x and y agree on a dense subset of $[0,1]$. It easy to see if $t<1$, then this implies that $x(t)=y(t)$ (just consider a sequence of continuity points converging to t from the right and use right continuity). If $t=1$, then note that $\lambda_n(1)=1$ for all $n\in\mathbb{N}$. The results follows easily from there.
By 5.3, it holds that $\|x-y\circ\lambda\|_\infty=\left\|y-x\circ\lambda^{-1}\right\|_\infty$ and $\|\lambda-I\|_\infty=\left\|\lambda^{-1}-I\right\|_\infty$ for all $x,y\in D[0,1]$ and $\lambda\in\Lambda$. Hence, if $(\lambda_n)_{n\in\mathbb{N}}$ is a sequence of deformations such that

$$\max(\|\lambda_n-I\|_\infty,\|x-y\circ\lambda_n\|_\infty)\to d(x,y),$$

it then holds that $\max(\left\|I-\lambda_n^{-1}\right\|_\infty,\left\|y-x\circ\lambda_n^{-1}\right\|_\infty)\to d(x,y)$, and by the group structure of Λ we conclude $d(x,y)\geq d(y,x)$. The reverse inequality is obtained thorugh similar arguments. The triangle inequality remains to be shown. By 5.3, we have

$$\begin{aligned}\|\lambda_2\circ\lambda_1-I\|_\infty&=\left\|\lambda_2-\lambda_1^{-1}\right\|_\infty\\&\leq\|\lambda_2-I\|_\infty+\left\|I-\lambda_1^{-1}\right\|_\infty\\&=\|\lambda_2-I\|_\infty+\|\lambda_1-I\|_\infty\end{aligned}$$

and

$$\begin{aligned}\|x-z\circ\lambda_2\circ\lambda_1\|_\infty&\leq\|x-y\circ\lambda_1\|_\infty+\|y\circ\lambda_1-z\circ\lambda_2\circ\lambda_1\|_\infty\\&=\|x-y\circ\lambda_1\|_\infty+\|y-z\circ\lambda_2\|_\infty.\end{aligned}$$

Thus, if $(\xi_n)_{n\in\mathbb{N}}$ and $(\zeta_n)_{n\in\mathbb{N}}$ are sequences such that

$$\begin{aligned}&\lim_{n\to\infty}\max(\|\lambda_n-I\|_\infty,\|x-y\circ\lambda_n\|)=d(x,y)\text{ and}\\&\lim_{n\to\infty}\max(\|\xi_n-I\|_\infty,\|y-z\circ\xi_n\|)=d(y,z).\end{aligned}$$

As $\liminf_{n\to\infty}\max(\|\lambda_n\circ\xi_n-I\|_\infty,\|x-z\circ\lambda_n\circ\xi_n\|_\infty)\leq d(x,y)+d(y,z)$, we conclude that the triangle inequality holds. □

Recall that $d(x_n,x)\to0$ for $n\to\infty$ if and only there exists a sequence of deformations $(\lambda_n)_{n\in\mathbb{N}}$ such that $\|\lambda_n-I\|_\infty\to0$ and $\|x-x_n\circ\lambda_n\|_\infty\to0$ for $n\to\infty$. By the triangle inequality, for $t\in[0,1]$, it holds

$$\left|x_n(t)-x(t)\right|\leq\left|x_n(t)-x(\lambda_n(t))\right|+\left|x(\lambda_n(t))-x(t)\right|.$$

Therefore, if $d(x_n,x)\to0$ and x is continuous at t, then $\left|x_n(t)-x(t)\right|\to0$.

Example 5.5. $(D[0,1],d)$ is not complete.

Proof. Consider the mapping $\phi:[0,1]\to D[0,1]$ given by $\phi(x)=1_{[x,1]}$. Let $(a_n)_{n\in\mathbb{N}}\subseteq[0,1]$ be the sequence given by $a_n=2^{-n}$. Consider the sequence $(x_n)_{n\in\mathbb{N}}=(\phi(a_n))_{n\in\mathbb{N}}$. Let $(\lambda_n)_{n\in\mathbb{N}}\subseteq\Lambda$ be a sequence of C^1-interpolations such that $\lambda(2^{-n})=2^{-(n+1)}$ for $n\in\mathbb{N}$ and $\|\lambda_n-I\|_\infty\leq\left|2^{-n}-2^{-(n+1)}\right|+2^{-(n+1)}=2^{-n}$. As $\|x_{n+1}\circ\lambda_n-x_n\|_\infty=0$. This implies that $d(x_n,x_{n+1})\leq2^{-n}$ and we conclude that $(x_n)_{n\in\mathbb{N}}$ is d-Cauchy. However, for $t>0$ it holds that $x_n(t)\to1$ for $n\to\infty$, and so the only potential limit function is the constant $t\longmapsto1$. But as $\left|x_n(0)-1\right|=1$ for all $n\in\mathbb{N}$ and $t=0$ is a continuity point for all cádlág functions, we conclude that $(x_n)_{n\in\mathbb{N}}$ is not convergent. □

Even though d is not a complete metric, it is equivalent to a complete metric, that is $(D[0,1],d)$ is a Polish space. Before we can introduce this metric, let us discuss how we will change d to make $D[0,1]$ complete. If $\lambda \in \Lambda$, then we can discuss the uniform distance between λ and the identity function on $[0,1]$. In the following, we will discuss the distance between the slopes of all of the secant lines of λ and 1. The mean value theorem implies that instead of slopes of the secant line of Λ between $s,t \in [0,1]$ with $s < t$, we can simply compute the distance between the derivative of λ, evaluated at some point $\xi \in (s,t)$ and 1. For purely computational reason, we will discuss the distance between the logarithm of the slopes of the secant lines and 0. As every $\lambda \in \Lambda$ has finite, non-vanishing derivative, it follows that the logarithm is uniformly continuous on image of any $\lambda \in \Lambda$. Thus it holds

$$\lambda_n' \to 1 \text{ uniformly as } n \to \infty \text{ if and only if } \log\big(\lambda_n'\big) \to \log(1) = 0 \text{ uniformly as } n \to \infty.$$

We introduce the notation

$$\overline{\|\lambda\|} = \sup_{\xi \in (0,1)} \Big|\log\big(\lambda'(\xi)\big)\Big| = \sup_{\xi \in [0,1]} \Big|\log\big(\lambda'(\xi)\big)\Big|.$$

Lemma 5.6. For $\lambda, \lambda_1, \lambda_2 \in \Lambda$ it holds

$$\overline{\|\lambda^{-1}\|} = \overline{\|\lambda\|},$$

and

$$\overline{\|\lambda_2 \circ \lambda_1\|} \leq \overline{\|\lambda_2\|} + \overline{\|\lambda_1\|}.$$

Proof. By the inverse function differentiation rule, it holds

$$\Big|\log\big((\lambda^{-1})'(\zeta)\big)\Big| = \Big|\log\big(\lambda'(\lambda^{-1}(\zeta))\big)\Big|,$$

for $\zeta \in (0,1)$. For all $\zeta \in [0,1]$, there exists one $\xi \in [0,1]$, such that $\lambda(\xi) = \zeta$, or equivalently, $\lambda^{-1}(\zeta) = \xi$. If $(\zeta_n)_{n\in\mathbb{N}}$ is a sequence such that

$$\Big|\log\big(\lambda'(\lambda^{-1}(\zeta_n))\big)\Big| \to \overline{\|\lambda^{-1}\|} \text{ as } n \to \infty,$$

then there exists a sequence $(\xi_n)_{n\in\mathbb{N}}$ such that

$$\Big|\log\big(\lambda'(\xi_n)\big)\Big| \to \overline{\|\lambda^{-1}\|} \text{ as } n \to \infty.$$

This implies that $\overline{\|\lambda\|} \geq \overline{\|\lambda^{-1}\|}$. The reverse inequality is obtained similarly.
By the chain rule, we see

$$\begin{aligned}\Big|\log\big(\lambda_2(\lambda_1(\xi))'\big)\Big| &= \Big|\log\big(\lambda_2'(\lambda_1(\xi))\lambda_1'(\xi)\big)\Big| \\ &\leq \Big|\log\big(\lambda_2'(\lambda_1(\xi))\big)\Big| + \Big|\log\big(\lambda_1'(\xi)\big)\Big|,\end{aligned}$$

whence the claim follows. ∎

We define the mapping $\hat{d} : D[0,1] \times D[0,1] \to \mathbb{R}$ to be given by

$$\overline{d}(x,y) = \inf_{\lambda \in \Lambda}(\max(\overline{\|\lambda\|}, \|x - y \circ \lambda\|_\infty)).$$

In view of the lemma above, the proof of the fact that $\overline{d}$ is a pseudo-metric proceeds in much the same way as 5.4. That $\overline{d}$ is indeed a metric is an immediate consequence of the following result.

Proposition 5.7. For all $x, y \in D[0,1]$ it holds

$$d(x,y) \leq 4\overline{d}(x,y).$$

Proof. Assume $\overline{d}(x,y) < \frac{1}{4}$ and let $\lambda \in \Lambda$ be such that $\overline{\|\lambda\|} < \frac{1}{4}$ and $\|x - y \circ \lambda\| < \frac{1}{4}$. Let $t \in (0,1]$ and consider the slope of the secant line from 0 to t,

$$\frac{\lambda(t) - \lambda(0)}{t - 0} = \frac{\lambda(t)}{t} = \lambda'(\xi) \text{ for some } \xi \in (0,t).$$

It then holds

$$\log\Big(1 - 4 \cdot \overline{\|\lambda\|}\Big) \leq \log\big(\lambda'(\xi)\big) \leq \log\Big(1 + 4 \cdot \overline{\|\lambda\|}\Big),$$

which implies

$$1 - 4 \cdot \overline{\|\lambda\|} \leq \lambda'(\xi) \leq 1 + 4 \cdot \overline{\|\lambda\|}.$$

Thus $\big|\lambda'(\xi) - 1\big| = \Big|\frac{\lambda(t)}{t} - 1\Big| \leq 4 \cdot \overline{\|\lambda\|}$. As $t \in (0,1]$, it holds $\big|\lambda(t) - t\big| = t\Big|\frac{\lambda(t)}{t} - 1\Big| \leq 4 \cdot \overline{\|\lambda\|}$, and so the result follows. If $\overline{d}(x,y) \geq \frac{1}{4}$ the result is trivial as $\|\lambda - I\|_\infty \leq 1$. $\square$

Hence it holds that convergence in the $\overline{d}$-metric implies convergence in the d-metric. To see the reverse implication, we will need the concept of modulus of continuity for cádlág functions. For $x \in D[0,1]$ and $A \subseteq [0,1]$, let $\psi_A(x) = \sup_{s,t \in I}(\big|x(t) - x(s)\big|)$.

Lemma 5.8. For $x \in D[0,1]$ and $\varepsilon > 0$, there exists a finite subdivision of $[0,1]$, $\sigma : 0 = s_0 < \ldots < s_n \leq 1$ such that

$$\max_{i=1,\ldots,n} \psi_{[s_{i-1}, s_i)}(x) < \varepsilon.$$

Proof. Let T be the supremum of $t \in [0,1] \cap \mathbb{Q}$ such that $[0,t)$ can be decomposed as in the statement. As $x(0) = x(0+)$, we have that T exists. As x has left limits, it holds that $[0,T]$ can be decomposed. For all $t \in [0,1]$, we have that x is right continuous in t, and hence T cannot be strictly smaller than any $t \in [0,1]$. We conclude that $T = 1$. $\square$

We now define modulus of continuity for cádlág functions. In the following, let Π be set of all finite subdivisions of $[0,1]$ and for $0 < \delta \leq 1$ let Π_δ be the set of finite partitions such that the smallest interval in any subdivision in Π_δ is at least δ in length.

Definition 5.9. For $x \in D[0,1]$ and $\delta \in (0,1]$, we define

$$\psi'_\delta(x) = \inf_{\Pi_\delta}(\max_{1 \leq i \leq n}(\psi_{[t_{i-1}, t_i)}(x)))$$

We note that 5.8 implies that $\psi'_\delta(x) \to 0$ as $\delta \to 0$.
For a finite subdivision $\sigma : 0 = s_0 < \ldots < s_n = 1 \in \Pi$, let $A_\sigma : D[0,1] \to D[0,1]$ be given by

$$A_\sigma(x)(s) = \begin{cases} x(s_{i-1}) \text{ for } s \in [s_{i-1}, s_i) \text{ for some } i = 1, \ldots, n \\ x(1) \text{ for } s = 1 \end{cases}.$$

Lemma 5.10. Let $\sigma : 0 = s_0 < \ldots < s_n = 1 \in \Pi$ with $\max_{i=1,\ldots,n}(s_i - s_{i-1}) < \delta$, then $d(x, A_\sigma(x)) \leq \max(\delta, \psi'_\delta(x))$

Proof. Let $\mu : [0,1] \to [0,1]$ be given by

$$\mu(s) = \begin{cases} s_{i-1} \text{ for } s \in [s_{i-1}, s_i) \text{ for some } i = 1, \ldots, n \\ 1 \text{ for } s = 1 \end{cases}.$$

Note that $A_\sigma(x) = x \circ \mu$. Let $\theta : 0 = t_0 < \ldots < t_k \in \Pi_\delta$ such that $\max_{j=1,\ldots,k} \psi_{[t_{j-1},t_j)}(x) < \psi'_\delta(x) + \varepsilon$ for some $\varepsilon > 0$. Let $\upsilon : (t_j)_{j=1}^k \to [0,1]$ be given by

$$\upsilon(t) \begin{cases} s_i \text{ for } t \in (s_{i-1}, s_i] \text{ for some } i = 1, \ldots n \\ 0 \text{ for } t = 0 \end{cases}.$$

As $\theta \in \Pi_\delta$ and $\sigma \in \Pi \setminus \Pi_\delta$, no two endpoints of θ lie in the same interval of σ. We note that υ is strictly increasing, and hence the image of υ is a subdivision. Let λ be the C^1 interpolation of θ and the image of υ. As $t \in [0,1]$ and $\lambda(t)$ are always in the interval in σ, it clearly holds that $\|\lambda - I\|_\infty \leq \delta$.
We wish to show that $\left|x - A_\sigma(x)(\lambda(t))\right| = \left|x(\lambda^{-1}(t)) - A_\sigma(x)(t)\right| \leq \psi'_\delta(x) + \varepsilon$ for all $t \in [0,1]$. For $t = 0$ or $t = 1$, it is clear. For $t \in (0,1)$, it is enough to show that $\mu(t)$ and $\lambda^{-1}(t)$ is always in the same interval in θ. This is because $\max_{j=1,\ldots,k} \psi_{[t_{j-1},t_j)}(x) < \psi'_\delta(x) + \varepsilon$. Let $j \neq 0$, then $t_j \in (s_{i-1}, s_i]$ for some $i = 1, \ldots, n$. As $\mu(t) \in (s_i)_{i=1}^n$, it holds $t_j \leq \mu(t)$ is equivalent to $s_i \leq \mu(t)$, which is, by definiton of μ, equivalent to $s_i \leq t$. As $t_j \in (s_{i-1}, s_i]$ for some $i = 1, \ldots, n$, it holds that $\lambda(t_j) = s_i \leq t$, which, by definition of λ, is equivalent to $t_j \leq \lambda^{-1}(t)$. This implies that $t_j \leq \mu(t)$ if and only if $t_j \leq \lambda^{-1}(t)$, or in other words, μ and λ^{-1} are always in the same interval of θ.
Hence $d(x, A_\sigma(x)) < \max(\delta, \psi'_\delta(x) + \varepsilon)$. As this holds for all ε, this concludes the proof. $\square$

Proposition 5.11. Let $\delta < \frac{1}{100}$ and $x, y \in D[0,1]$. If $d(x,y) \leq \delta^2$, then $\overline{d}(x,y) \leq \psi'_\delta(x) + 4\delta$.

Proof. Let $\sigma : 0 = s_0 < \ldots < s_n = 1 \in \Pi_\delta$, be such that $\max_{1 \leq i \leq n}(\psi_{[s_{i-1},s_i)}(x)) < \psi'_\delta(x) + \delta$. Let $\hat{\lambda} \in \Lambda$, be such that $\left\|\hat{\lambda} - I\right\|_\infty < \delta^2$ and $\left\|x - y \circ \hat{\lambda}\right\|_\infty = \left\|x \circ \hat{\lambda}^{-1} - y\right\|_\infty < \delta^2$. Let $\lambda \in \Lambda$ be the C^1 interpolation, such that $\hat{\lambda}(s_i) = \lambda(s_i)$ for $i = 0, \ldots n$. As $\hat{\lambda}^{-1}(\lambda(s_i)) = s_i$ for $i = 0, \ldots n$ and $\hat{\lambda}^{-1} \circ \lambda$ is strictly increasing and continuous, it holds that for all $s \in [0,1]$, both s and

$\hat{\lambda}^{-1}(\lambda(s))$ are in the same subinterval of σ.

$$\begin{aligned}\left|x(t)-y(\lambda(t))\right| \leq& \left|x(t)-x(\hat{\lambda}^{-1}(\lambda(t)))\right| + \left|x(\hat{\lambda}^{-1}(\lambda(t)))-y(\lambda(t))\right| \\ \leq& \max_{1\leq i\leq n}(\psi_{[t_{i-1},t_i)}(x)) + \delta^2 \\ <& \psi'_\delta(x)+\delta+\delta^2 \\ <& \psi'_\delta(x)+4\delta.\end{aligned}$$

For $i=1,\dots,n$, it holds

$$\begin{aligned}\left|\lambda(s_i)-\lambda(s_{i-1})-(s_i-s_{i-1})\right| =& \left|\hat{\lambda}(s_i)-\hat{\lambda}(s_{i-1})-(s_i-s_{i-1})\right| \\ \leq& \left\|\hat{\lambda}-I\right\|_\infty + \left\|\hat{\lambda}-I\right\|_\infty \\ <& 2\delta^2 \\ \leq& 2\delta(s_i-s_{i-1}).\end{aligned}$$

We wish to show

$$\left|\frac{\lambda(t)-\lambda(s)}{t-s}-1\right| \leq 3\delta \tag{6}$$

for $s,t\in[0,1]$. Let $\gamma:[0,1]\to[0,1]$ be the piecewise affine function, such that $\gamma(s_i)=\lambda(s_i)$ and affine on each of the intervals in σ. For $i=1,\dots,n$, γ is affine on $[t_{i-1},t_i]$, therefore exist a_i and b_i, such that $\gamma(s)=a_is+b_i$ for $s\in[t_{i-1},t_i]$. In particular it holds

$$\left|\gamma(s_i)-\gamma(s_{i-1})-(s_i-s_{i-1})\right| = \left|a_is_i+b_i-(a_is_{i-1}+b_i)-(s_i-s_{i-1})\right| = |a_i-1|\,(s_i-s_{i-1})$$

As we noted above, it holds

$$\left|\gamma(s_i)-\gamma(s_{i-1})-(s_i-s_{i-1})\right| = \left|\lambda(s_i)-\lambda(s_{i-1})-(s_i-s_{i-1})\right| < 2\delta(s_i-s_{i-1}).$$

Combining these observations, we conclude $|a_i-1|<2\delta$ for all $i=1,\dots,n$. Hence, for $s,t\in[t_{i-1},t_i]$ with $s<t$ for some $i=1,\dots,n$, it holds $\left|\gamma(t)-\gamma(s)-(t-s)\right|\leq 2\delta(t-s)$ or equivalently

$$\left|\frac{\gamma(t)-\gamma(s)}{t-s}-1\right| \leq 2\delta.$$

Let $s,t\in[t_{i-1},t_i]$ with $s<t$, by 5.2, we can choose λ such that

$$\begin{aligned}\left|\frac{\lambda(t)-\lambda(s)}{t-s}-\frac{\gamma(t)-\gamma(s)}{t-s}\right| \leq& \left|\min_{i=,\dots n}a_i-\max_{i=,\dots n}a_i\right|+\varepsilon \\ \leq& \left|\min_{i=,\dots n}a_i-1\right|+\left|\max_{i=,\dots n}a_i-1\right|+\varepsilon \\ <& 5\delta,\end{aligned}$$

for any $\varepsilon>0$. Let $\varepsilon<\delta$. Note

$$\begin{aligned}\left|\frac{\lambda(t)-\lambda(s)}{t-s}-1\right| \leq& \left|\frac{\lambda(t)-\lambda(s)}{t-s}-\frac{\gamma(t)-\gamma(s)}{t-s}\right|+\left|\frac{\gamma(t)-\gamma(s)}{t-s}-1\right| \\ <& 7\delta.\end{aligned}$$

If $s, t \in [0,1]$, then (6) follows from the triangle inequality and the considerations above.
As (6) holds, it also holds that

$$\log(1-7\delta) < \log\big(\lambda'(\xi)\big) < \log(1+7\delta),$$

as $\delta < \frac{1}{100}$, it holds that $\overline{\|\lambda\|} \le 4\delta$. Hence, we conclude that $\overline{d}(x,y) \le \psi'_\delta(x) + 4\delta$. $\square$

We state an immediate consequence of 5.7 and 5.11.

Theorem 5.12. The metrics d and $\overline{d}$ are equivalent.

But why even look at these metrics. As teased before, under these metrics $D[0,1]$ is a polish space. We have already seen that d is not complete, but $\overline{d}$ is. Both d and $\overline{d}$ are separable, but it is quite a bit easier to prove it for d.

Proposition 5.13. The metric space $(D[0,1], d)$ is separable.

Proof. Let $\varepsilon > 0$ and $x \in D[0,1]$. For $n \in \mathbb{N}$, let $B_n \subseteq D[0,1]$ be sets of functions such that they attain constant, rational value on $[\frac{i-1}{n}, \frac{i}{n})$ for $i = i, \dots, n$ and on $\{1\}$. Note that $\bigcup_{n=1}^\infty B_n$ is countable. Choose $k \in \mathbb{N}$ such that $k^{-1} < \frac{\varepsilon}{2}$ and $\psi'_{k^{-1}}(x) < \frac{\varepsilon}{2}$. Let $\tau : 0 < \frac{1}{k} < \dots < \frac{k-1}{k} < 1$. As $\max_{i=1,\dots,k}(\frac{i}{k} - \frac{i-1}{k}) = \frac{1}{k} < \frac{\varepsilon}{2}$, it holds that $d(x, A_\tau(x)) < \frac{\varepsilon}{2}$. Let $y \in B_n$ be such that $\left\|A_\tau(x) - y\right\|_\infty < \frac{\varepsilon}{2}$, then $d(A_\tau(x), y) < \frac{\varepsilon}{2}$, and $d(x,y) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$. Hence $\bigcup_{n=1}^\infty B_n$ is dense in $(D[0,1], d)$. $\square$

It turns out that $\overline{d}$ is a complete metric. In the following theorem we will be utilising the fact that $C^1[0,1]$ is complete with respect to the C^1-norm given by $\|f\|_{C^1} = \|f\|_\infty + \|f'\|_\infty$. The proof of this is a rote application of completeness as the uniform norm and korollar 3.18 of [Christiandl].

Theorem 5.14. The metric space $(D[0,1], \overline{d})$ is complete.

Proof. Let $(x_n)_{n\in\mathbb{N}}$ be a $\overline{d}$-Cauchy sequence in $D[0,1]$. Let $(x_{n_k})_{k\in\mathbb{N}}$ be a subsequence such that $\overline{d}(x_{n_k}, x_{n_{k+1}}) < 2^{-k}$. Then there exists $(\lambda_k)_{k\in\mathbb{N}}$ such

$$\left|f_{n_k}(t) - f_{n_k}(\lambda_k(t))\right| \le 2^{-k} \text{ and } \overline{\|\lambda_k\|} \le 2^{-k}.$$

Note, by 5.11, the above implies that $\|\lambda_k - I\|_\infty \le 2^{-k+2}$ Let $(\mu^n_m)_{m\in\mathbb{N}}$ be the sequence in Λ given by $\mu^n_m = \lambda_{n+m} \circ \dots \lambda_n$.
We wish to show that $(\mu^n_m)_{m\in\mathbb{N}}$ is a Cauchy sequence in the C^1-norm for all $n \in \mathbb{N}$. This is the case if both $(\mu^n_m)_{m\in\mathbb{N}}$ and $((\mu^n_m)')_{m\in\mathbb{N}}$ are uniformly Cauchy for all $n \in \mathbb{N}$. By 5.3, it holds

$$\begin{aligned}\left\|\mu^n_{m+1} - \mu^n\right\|_\infty =& \|\lambda_{n+m+1} \circ \dots \circ \lambda_n - \lambda_{n+m} \circ \dots \circ \lambda_n\| \\ =& \|\lambda_{n+m+1} - I\| \\ &\le 2^{-(m+n+1)},\end{aligned}$$

whence we see that $(\mu^n_m)_{m\in\mathbb{N}}$ is uniformly Cauchy. By the chain rule, we see

$$\begin{aligned}\left\|(\mu^n_{m+1})' - (\mu^n_m)'\right\|_\infty =& \left\|(\lambda_{n+m+1} \circ \dots \circ \lambda_n)' - (\lambda_{n+m} \circ \dots \circ \lambda_n)'\right\|_\infty \\ \le& \left\|(\lambda'_{n+m+1} \circ \dots \circ \lambda_n) \cdot (\lambda_{n+m} \circ \dots \circ \lambda_n)' - (\lambda_{n+m} \circ \dots \circ \lambda_n)'\right\|_\infty \\ \le& \left\|(\lambda_{n+m} \circ \dots \circ \lambda_n)'\right\|_\infty \left\|\lambda'_{n+m+1} \circ \dots \circ \lambda_n - 1\right\|_\infty.\end{aligned}$$

As $\log\big(\lambda'_{n+m+1}\big)$ converges uniformly fast to 0 for $m \to \infty$, it holds that λ'_{n+m+1} converges uniformly fast to 1 as $m \to \infty$. By the chain rule, it holds that $(\lambda_{n+m} \circ \ldots \circ \lambda_n)' = (\lambda'_{n+m} \circ \ldots \circ \lambda_n) \cdot (\lambda_{n+m-1} \circ \ldots \circ \lambda_n)'$, and hence $\big\|(\mu^n_m)'\big\|_\infty \leq \big\|\lambda'_{n+m}\big\|_\infty \big\|(\mu^n_{m-1})'\big\|_\infty$. By induction it therefore holds that

$$\big\|(\mu^n_m)'\big\|_\infty \leq \prod_{j=0}^{m} \Big\|\lambda'_{n+j}\Big\|_\infty .$$

As $(\overline{\|\lambda_k\|})_{k\in\mathbb{N}}$ is summable, it holds that $(\log\big(\lambda'_k(\xi_k)\big))_{k\in\mathbb{N}}$ is summable for any sequence $(\xi_k)_{k\in\mathbb{N}} \subseteq [0,1]$. Therefore it holds that $\Big(\prod_{i=1}^{k} \lambda'_i(\xi_i)\Big)_{k\in\mathbb{N}}$ is convergent. In particular, it holds that $\prod_{j=0}^{m}\Big\|\lambda'_{n+j}\Big\|_\infty$ converges as $m \to \infty$.
Thus it holds that $\big\|(\mu^n_{m+1})' - (\mu^n_m)'\big\|_\infty$ is summable. Thus $(\mu^n_m)_{m\in\mathbb{N}}$ is a Cauchy sequence in the C^1-norm. As the C^1-norm is complete, $(\mu^n_m)_{m\in\mathbb{N}}$ has a limit, we will call it μ_n, for all $n \in \mathbb{N}$. As μ^n_m is strictly increasing for all $m \in \mathbb{N}$, it holds that μ_n is weakly increasing. Furthermore, as

$$\overline{\|\mu^n_m\|} \leq \sum_{j=0}^{m} \overline{\big\|\lambda_{n+j}\big\|} \leq 2^{-(n-1)},$$

we conclude that μ_n is strictly increasing with non-vanishing derivative, hence $\mu_n \in \Lambda$ for all $n \in \mathbb{N}$.
We note that $\mu_n = \mu_{n+1} \circ \lambda_n$. Therefore

$$\Big\|f_{n_j} \circ \mu_j^{-1} - f_{n_{j+1}} \circ \mu_{j+1}^{-1}\Big\|_\infty = \Big\|f_{n_j}\mu_j^{-1} - f_{n_{j+1}} \circ \mu_{j+1}^{-1}\Big\|_\infty \leq 2^{-j}.$$

Hence $(f_{n_j} \circ \mu_j^{-1})_{j\in\mathbb{N}}$ is uniformly Cauchy, hence there exists a limit function f. As $D[0,1]$ is closed in the uniform norm, it holds that $f \in D[0,1]$. As

$$\big\|f_{n_j} - f\big\| \to 0 \text{ and } \overline{\big\|\mu_j\big\|} \to 0 \text{ as } j \to \infty,$$

we conclude that $\overline{d}(f_{n_j}, f) \to 0$. Hence every $\overline{d}$-Cauchy sequence is is convergent. □

As separability is a topological property, we conclude that $(D[0,1], \overline{d})$ is complete and separable.

5.2 Projections

While it is convenient that we can find a metric that makes $D(0,1)$ a Polish space, there is an addtional reason for it being the metric of choice when doing probability theory in $D[0,1]$: The Borel-σ-algebra induced by the Skorokhod metric coincides with the projection algebra on $D[0,1]$, and hence the ball algebra induced by the uniform norm.
In the following we will need one lemma.

Lemma 5.15. If $x : [0,1] \to \mathbb{R}$ is right continuous, then for $t \in [0,1)$

$$m \int_t^{t+\frac{1}{m}} x(s)\; ds \to x(t) \text{ for } m \to \infty.$$

Proof. As x is right continuous, it holds that for all $\varepsilon > 0$ there exists $m \in \mathbb{N}$ such that $s \in [t, t + \frac{1}{m})$ implies $\left|x(t) - x(s)\right| < \varepsilon$. Hence

$$\begin{aligned}\left|x(t) - m \int_t^{t+\frac{1}{m}} x(s)\ ds\right| =& \left|m \int_t^{t+\frac{1}{m}} x(t)\ ds - m \int_t^{t+\frac{1}{m}} x(s)\ ds\right| \\ \leq& m \int_t^{t+\frac{1}{m}} \left|x(t) - x(s)\right|\ ds \\ \leq& m \int_t^{t+\frac{1}{m}} \varepsilon\ ds \\ =& \varepsilon.\end{aligned}$$

□

We note that the projection at $t = 1$ is continuous; as $\lambda(1) = 1$ for all $\lambda \in \Lambda$, it holds that $d(x_n, x) \to 0$ implies $\left|x(1) - x_n(\lambda_n(1))\right| = \left|x(1) - x_n(1)\right| \to 0$. In the following, let $\mathcal{D}$ denote the Borel-σ-algebra on $D[0,1]$ induced by d (or $\overline{d}$).

Theorem 5.16. The projection algebra on $D[0,1]$ coincides with the Borel-σ-algebra on $(D[0,1], d)$.

Proof. We prove $\sigma(\pi_t \mid t \in [0,1]) \subseteq \mathcal{D}$ by proving that every projection is $\mathcal{D} - \mathcal{B}(\mathbb{R})$ measurable. As π_1 is continuous, it is measurable. For $0 \leq t < 1$ and $m \in \mathbb{N}$ such that $t + \frac{1}{m} \leq 1$, let $h_m^t : D[0,1] \to \mathbb{R}$ given by

$$h_m^t(x) = m \int_t^{t+\frac{1}{m}} x(s)\ ds.$$

As a cádlag function has at most countably many discontinuities it holds that a sequence of cádlág functions converging in the Skorokhod topology to some cádlág function is almost surely convergent to that limit. As convergence in the Skorokhod topology implies that $x_n \circ \lambda_n$ converges uniformly, it holds that a Skorokhod convergent sequence is uniformly bounded. Lebesgue's dominated convergence theorem then yields that $d(x_n, x) \to 0$ implies $h_m^t(x_n) \to h_m^t(x)$. In other words, h_m^t is continuous, hence measureable. By the previous lemma $h_m^t(x) \to \pi_t(x)$, hence π_t is the pointwise limit of measurable function, and hence measurable itself.
We wish to show that the identity mapping on $D[0,1]$ is $\sigma(\pi_t \mid t \in [0,1]) - \mathcal{D}$-measurable. This will imply that $\mathcal{D} \subseteq \sigma(\pi_t \mid t \in [0,1])$.
For $m \in \mathbb{N}$, let $\theta_m : 0 = s_0 < \ldots < s_{k(m)} = 1 \in \Pi \setminus \Pi_{\frac{1}{m}}$. Then the projection $\pi_{\theta_m} = \pi_{s_0 \ldots, s_{k(m)}} : D[0,1] \to \mathbb{R}^{k(m)+1}$ is $\sigma(\pi_t \mid t \in [0,1]) - \mathcal{B}(\mathbb{R}^{k(m)+1})$-measurable[2]. Consider the mapping $V_{\theta_m} : \mathbb{R}^{k(m)+1} \to D[0,1]$ given by

$$V_{\theta_m}(a_1, a_2 \ldots a_{k(m)})(t) = \begin{cases} a_{i-1} \text{ for } t \in [s_{i-1}, s_i), \text{ for } i = 1, \ldots, k(m) - 1 \\ a_{k(m)} \text{ for } t = 1 \end{cases}.$$

It is easy to see that V_{θ_m} is Skorokhod continuous, it is in fact continuous in the uniform norm on $D[0,1]$. Therefore V_{θ_m} is $\mathcal{B}(\mathbb{R}^{k(m)+1}) - \mathcal{D}$-measurable. Note that $V_{\theta_m} \circ \pi_{\theta_m} = A_{\theta_m}$. This

[2] $k(m)$ indicates that the number of endpoints in θ_m depends on m.

composition is then $\sigma(\pi_t \mid t \in T) - \mathcal{D}$-measurable. By 5.10, it holds that $d(x, A_{\theta_m}(x)) \to 0$ as $m \to \infty$, and hence the identity mapping on $D[0,1]$ is $\sigma(\pi_t \mid t \in T) - \mathcal{D}$-measurable.
Thus we conclude that $\mathcal{D} = \sigma(\pi_t \mid t \in [0,1])$. $\square$

This last result shows just how well the metric space $(D[0,1], d)$ suits our theory of weak convergence in general metric spaces. The natural σ-algebra to equip $D[0,1]$ with is the projection algebra, if we consider continuous time stochastic processes as $D[0,1]$-valued stochastic variables. Now we know that both the ball-σ-algebra of the uniform norm and the Borel algebra induced by the Skorokhod metric coincides with this canonical algebra. This implies that the theory of weak convergence in complete and separable metric spaces can, with some care, be used in the theory of weak convergence in $D[0,1]$, even if one decides to equip $D[0,1]$ with the uniform norm. Furthermore, it also allows us to connect our theory to functional analysis through the Riesz representation theorem, a connection that disappeared once we dropped the Borel algebra.

References

[Hansen] Ernst Hansen, *Stochastic Processes*, Department of mathematical sciences, University of Copenhagen, 1. edition, 2020.

[Pollard] David Pollard, *Convergence of Stochastic Processes*, Springer Verlag New York, 1. edition, 1984.

[Billingsley] Patrick Billingsley, *Convergence of Probability Measures*, John Wiley & Sons, Inc, 2. edition, 1999.

[Halmos] Paul R. Halmos, *Measure theory*, Springer Verlag New York, 2. edition, 1974.

[Dudley] R. M. Dudley, *Real Analysis and Probability*, Cambridge University Press, 1. edition, 2002.

[Christiandl] Matthias Christiandl, *Analyse 1*, Department of mathematical sciences, University of Copenhagen, 1. edition, 2019.