Hindawi Publishing Corporation
International Journal of Stochastic Analysis
Volume 2011, Article ID 801303, 23 pages
doi:10.1155/2011/801303

Research Article

Blackwell Spaces and ϵ-Approximations of Markov Chains

Giacomo Aletti[1] and Diane Saada[2]

[1] *ADAMSS Centre (ex MIRIAM) and Department of Mathematics, University of Milan, Via Saldini 50, 20133 Milan, Italy*
[2] *Department of Statistics, The Hebrew University of Jerusalem, Jerusalem 91905, Israel*

Correspondence should be addressed to Giacomo Aletti, giacomo.aletti@unimi.it

Received 31 December 2010; Revised 11 April 2011; Accepted 24 May 2011

Academic Editor: Michel Benaïm

On a weakly Blackwell space we show how to define a Markov chain approximating problem, for the target problem. The approximating problem is proved to converge to the optimal reduced problem under different pseudometrics.

1. Introduction

A target problem (TP) is defined as a homogeneous Markov chain stopped once it reaches a given absorbing class, the target. Our purpose is to only use the necessary information relevant with respect to the target and in consequence to reduce the available information. A new Markov chain, associated with a new equivalent but reduced matrix is defined. In the (large) finite case, the problem has been solved for TPs: in [1–3], it has been proved that any TP on a finite set of states has its "best target" equivalent Markov chain. Moreover, this chain is unique and there exists a polynomial time algorithm to reach this optimum.

The question is now to find, in generality, an ϵ-approximation of the Markov problem when the state space is measurable. The idea is to merge into one group the points that ϵ-behave the same with respect to the objective and, at the same time, to keep an almost equivalent Markov chain, with respect to the other "groups". The construction of these groups is done through equivalence relations. Each group will correspond to a class of equivalence. In fact, there are many other mathematical fields where approximation problems are faced by equivalences. For instance, in integration theory, we use simple functions, in functional analysis, we use the density of countable generated subspaces, and, in numerical analysis, we use the finite elements method.

In this paper, the approximation is made by means of discrete equivalences, which will be defined in the following. We prove that the sequence of approximations tends to the optimal exact equivalence relation defined in [1–3], when we refine the groups. Finer equivalence will imply better approximation, and accordingly the limit will be defined as a countably generated equivalence.

Under a very general Blackwell-type hypothesis on the measurable space, we show that it is equivalent to speak on countably generated equivalence relationships or on measurable real functions on the measurable space of states. If we do not work under this framework of Blackwell spaces, we can be faced to paradoxes, as it is explained by [4], of enlarging σ-algebras, while decreasing the information available to a decision-maker. The ϵ-approximation of the Markov chain depends always upon the kind of objective. In [5], Jerrum deals with ergodic Markov chains. His objective is to approximate the stationary distribution by means of a discrete approximating Markov chain, whose limit distribution is close in a certain sense to the original one. However, unlike our following work, his purpose is not the explicit and unified construction of the approximating process. In this paper, we focus on the target problem. We solve extensively the TP, where the objective is connected with the conditional probability of reaching the target T, namely $\mathbb{P}(X_n \in T \mid X_0 = x)$, for any n, x. This part extends the work in [1–3].

2. Main Results

Let $(X, \mathcal{X})$ be a measurable space, equipped with a assumption (A0) that will be explained when necessary. Let P be any transition probability on $(X, \mathcal{X})$. A homogeneous Markov process $(X_n)_{n\geq 0}$ is naturally associated to $(X, \mathcal{X}, P)$. In the target problem, we are interested in the probabilities of reaching the target class T within n steps, namely in

$$\mathbb{P}(\{X_n \in T\} \mid X_0 = x) \quad \text{for any } n \text{ and } x. \tag{2.1}$$

The set T is a priori given, and does not change through the computations. T is supposed to be an absorbing set lying in $\mathcal{X}$.

Definition 2.1. Let $(X, \mathcal{X})$ be a measurable space and let $T \in \mathcal{X}$. Let $\mathcal{F} \subseteq \mathcal{X}$ be a sub σ-algebra of $\mathcal{X}$ such that $T \in \mathcal{F}$. A function $P : X \times \mathcal{F} \to [0,1]$ is a *transition probability on* $(X, \mathcal{F})$ if

(i) $P(x, \cdot)$ is a probability measure on $\mathcal{F}$, for any $x \in X$,

(ii) $P(\cdot, F)$ is $\mathcal{F}$-measurable, for any $F \in \mathcal{F}$.

Given a transition probability P on $(X, \mathcal{F})$, we denote by P^n the transition probability on $(X, \mathcal{F})$ given inductively by

$$P^1 = P; \qquad P^{n+1}(x, F) = \int_X P(x, dy) P^n(y, F). \tag{2.2}$$

We denote by $\mathrm{TrP}(X, \mathcal{F})$ the set of the transition probabilities on $(X, \mathcal{F})$. We denote by $\mathbb{TP}_X = \cup_{\mathcal{F} \subseteq \mathcal{X}} \mathrm{TrP}(X, \mathcal{F})$ the set of all transition probabilities on X that we equip with a suitable pseudometric d:

$$d(P_1, P_2) = \sup_x \sum_n \beta^n \left| P_1^n(x, T) - P_2^n(x, T) \right|, \quad \text{with } \beta \in (0,1). \tag{2.3}$$

It is such that

$$d(P_1, P_2) = 0 \Longleftrightarrow P_1^n(x, T) = P_2^n(x, T), \quad \text{for any } n \text{ and } x. \tag{2.4}$$

This pseudometric, which is compatible with T, allows to approximate P by simpler kernels.

A target problem is defined through a transition probability $P \in (\mathbb{TP}_X, d)$. More precisely, we have the following definition.

Definition 2.2. A *target problem* is a quadruple $(X, \mathcal{F}, T, P)$, where $P \in \mathrm{TrP}(X, \mathcal{F})$ and $T \in \mathcal{F}$. A *simple target problem* is a target problem where $\mathcal{F}$ is generated by an at most countable partition of X.

The main purpose of this paper is to approximate any target problem by a sequence of simple target problems in the spirit of the construction of the Lebesgue integral, where the integral of a function f is approximated by the integral of simple functions $f_n = \sum_{i \in I} c_i I_{C_i}$. The strategies will play the role of the approximating subdivisions $(C_i)_{i \in I}$.

Definition 2.3. We call *strategy* Str a sequence of maps $(\mathrm{Str}_n)_{n \geq 0}$ from the set of the target problems to the set of the simple target problems. A strategy is a *target algorithm* if it is built as in Section 3.

In the "Lebesgue example" given above, the strategy is related to the "objective" of the problem (the integral) and the pseudometric $d(f, f_n) = \int |f - f_n| dx$ is required to go to 0 as n goes to infinity. Here also, a strategy is meaningful if $d(P, \mathrm{Str}_n(P))$ tends to 0 as n goes to infinity. Moreover, for what concerns applications, given a target problem $(X, \mathcal{X}, T, P)$ a good strategy should not need the computation of P^n, $n > 1$. The first main result of this paper states that the target algorithms are always good strategies.

Theorem 2.4. *For any target problem $(X, \mathcal{F}, T, P)$ and any target algorithm* Str*,*

$$\lim_{n \to \infty} d(P, P_n) = 0, \tag{2.5}$$

where $(X, \mathcal{F}_n, T, P_n) = \mathrm{Str}_n(X, \mathcal{F}, T, P)$.

Two questions immediately arise: does the sequence $(\mathrm{Str}_n(P))_{n \geq 0}$ have a limit (and in which sense)? Moreover, since d is defined as a pseudometric, does this limit depend on the choice of Str?

The extension of the concept of compatible projection given in [1–3] to our framework will enable us to understand better the answer to these questions. A measurable set $A \neq \emptyset$ of a measurable space $(X, \mathcal{X})$ is an $\mathcal{X}$-atom if it has no nonempty measurable proper subset. No two distinct atoms intersect. If the σ-field is countably generated, say by the sequence $\{A_n\}$ then the atoms of $\mathcal{X}$ are of the form $\cap_n C_n$ where each C_n is either A_n or $X \setminus A_n$.

Definition 2.5. An equivalence relationship π on a measurable space $(X, \mathcal{X})$ is *measurable* (*discrete*) if there exists a (discrete) random variable $f : (X, \mathcal{X}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ ($\mathcal{B}_{\mathbb{R}}$ denotes the Borel σ-algebra), such that

$$x \pi y \Longleftrightarrow f(x) = f(y), \tag{2.6}$$

and we denote it by $\pi = \pi_f$. Let $(X, \mathcal{F}, T, P)$ be a target problem. A *compatible projection* is a measurable equivalency π_f such that $T \in \sigma(f)$ and

$$P(x, F) = P(y, F), \quad \forall x \pi_f y, \ \forall F \in \sigma(f). \tag{2.7}$$

We say that $\pi \supseteq \pi'$ if π corresponds to partitions finer than π'. Finally, a compatible projection π is said to be *optimal* if $\pi \supseteq \pi'$, for any other compatible projection π'.

Remark 2.6. This definition is well posed if

$$\pi_f = \pi_g \Longleftrightarrow \sigma(f) = \sigma(g). \tag{2.8}$$

Assumption (A0) ensures that the definition of measurable equivalency is indeed well posed. This assumption will be stated and discussed in Section 4.

Theorem 2.7. *If $\pi = \pi_f$ is a compatible projection for the target problem $(X, \mathcal{F}, T, P)$, then there exists a target problem $(X, \sigma(f), T, P_\pi)$. such that $P_\pi(x, F) = P(x, F)$ for any $F \in \sigma(f)$.*

It is not said "a priori" that an optimal compatible projection must exist. If it is the case, then this equivalence is obviously unique.

Theorem 2.8. *For any target problem $(X, \mathcal{F}, T, P)$, there exists a (unique) optimal compatible projection π.*

To conclude the main results, let us first go back to the Lebesgue example. The simple functions $f_n = \sum_{i \in I} c_i I_{C_i}$ are chosen so that $\sigma(C_i, i \in I)$ increases to $\sigma(f)$ and $f_n(x) \to f(x)$. The following theorem guarantees these two similar facts by showing the "convergence" of any strategy to the optimal problem.

Theorem 2.9. *Let* $\mathrm{Str}_n(X, \mathcal{F}, T, P) = (X, \mathcal{F}_n, T, P_n)$, *with* Str *target algorithm and let π be the optimal compatible projection associated to the target problem $(X, \mathcal{F}, T, P)$. Then*

(i) *$\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$ for any n, and $\vee_n \mathcal{F}_{n \geq 0} = \mathcal{F}_\pi$;*

(ii) *$\lim_n P_n(x, F) = P_\pi(x, F)$, for any $(x, F) \in (X \times \cup_m \mathcal{F}_m)$.*

Remark 2.10 (The topology top). In Theorems 2.4–2.9, we have proved the convergence of $(P_n)_n$ to P_π with respect to the pseudometric d. The pseudometric topology top is the topology induced by the open balls $B_r(P) = \{Q \in \mathbb{TP}_X : d(P, Q) < r\}$, which form a basis for the topology. Accordingly, the previous theorems may be reread in terms of convergence of P_n to P on the topological space $(\mathbb{TP}_X, \mathrm{Top})$.

2.1. Connection with Weak Convergence

Given a strategy $(X, \mathcal{F}_n, T, P_n)_{n \geq 0}$, if we want to show a sort of weak convergence of $P_n(x, \cdot)$ to $P(x, \cdot)$, for any x, we face the two following problems:

(i) each $P_n(x, \cdot)$ is defined on a different domain (namely, on $\mathcal{F}_n$),

(ii) we did not have required a topology on X.

We overcome the first restriction by introducing a new definition of probability convergence. The idea is given in the following example.

Example 2.11. Let $\mathcal{F}_n = \sigma(\{(i2^{-n}, (i+1)2^{-n}], \ i = 0, \dots, 2^n - 1\})$ be the σ-algebra on $(0,1]$ generated by the dyadic subdivision. Suppose we know that $\nu_n : \mathcal{F}_n \to [0,1]$ is the unique probability on $\mathcal{F}_n$ s.t. for any i, $\nu_n((i2^{-n}, (i+1)2^{-n}]) = 2^{-n}$. Even if ν_n is not defined on the Borel sets of $(0,1]$, it is clear that in "some" sense, it must happen that $\nu_n \to \nu_*$, where ν_* is the Lebesgue measure on the Borel sets of $(0,1]$. Note that the cumulative function of ν_n is not defined, and therefore a standard weak convergence cannot be verified.

In fact, we know that

$$\nu_n\left(\left(-\infty, \frac{i}{2^n}\right]\right) = \nu_n\left(\left(0, \frac{i}{2^n}\right]\right) = \frac{i}{2^n}, \tag{2.9}$$

that is, in this case, as $n \to \infty$, we can determine the cumulative function on a dense subset. This fact allows to hope that $\nu_n \to \nu_*$ in a particular sense.

Definition 2.12. Let $(X, \mathcal{X}, (\mathcal{X}_n)_{n\geq 0})$ be a filtered space, and set $\mathcal{X}_\infty = \vee_{n\geq 0}\mathcal{X}_n$. Let $\nu_n : \mathcal{X}_n \to [0,1]$, $n \geq 1$ and $\nu_\infty : \mathcal{X}_\infty \to [0,1]$ be probability measures. One says that ν_n *converges totally to* ν_∞ *on the topological space* (X, τ) as n tends to infinity if $\overline{\nu}_n \xrightarrow[\tau]{w} \nu_\infty$ (converges in weak sense on (X, τ)), for any $\overline{\nu}_n : \mathcal{X}_\infty \to [0,1]$, such that $\overline{\nu}_{n|\mathcal{X}_n} = \nu_n$. One writes $\nu_n \xrightarrow[\tau]{\text{tot}} \nu_\infty$.

Going back to the example, it is simple to check that $\nu_n \xrightarrow[\tau(0,1]]{\text{tot}} \nu_*$, where ν_n, ν_* are given in Example 2.11 and $\tau(0,1]$ is the standard topology on $(0,1]$. In fact, let $(\overline{\nu}_n)_{n\geq 1}$ be any extension of $(\nu_n)_{n\geq 1}$ to the Borel sets of $(0,1]$. For any $t \in (0,1)$, we have by (2.9) that

$$t - \frac{1}{2^n} \leq F_{\overline{\nu}_n}(t) \leq t + \frac{1}{2^n}, \tag{2.10}$$

where $F_{\overline{\nu}_n}$ is the cumulative function of $\overline{\nu}_n$, which implies the weak convergence of $\overline{\nu}_n$ to ν_* and, therefore, $\nu_n \xrightarrow[\tau(0,1]]{\text{tot}} \nu_*$.

For what concerns the topology on X, we will define the topological space (X, ϱ_P) induced by the pseudometric d_P associated to the target problem $(X, \mathcal{F}, T, P)$, and the pseudometric d. In this way ϱ_P is only defined with the data of the problem. One may ask: is this topology too poor? The answer is no, since it is defined by the pseudometric d_P. In fact, $d_P(x,y) < \epsilon$ means that x and y play "almost the same role" with respect to T. A direct algorithm which takes d_P into account needs the computation of P^n at each step. In any case, even if d_P may not be computable, it defines a nontrivial interesting topology ϱ_P on X. As expected, we have the following theorem.

Theorem 2.13. *Let* $\mathrm{Str}_n(X, \mathcal{F}, T, P) = (X, \mathcal{F}_n, T, P_n)$*, with* Str *target algorithm. Then, for any given* x*,*

$$P_n(x, \cdot) \xrightarrow[\varrho_P]{\text{tot}} P(x, \cdot). \tag{2.11}$$

3. The Target Algorithm

In this section, we introduce the core of the approximating target problem, namely a set of strategies Str which solves the target problem.

Given a measurable space $(X, \mathcal{X})$ and a target problem $(X, \mathcal{F}, T, P)$, the target algorithm is built in the spirit of the exact one given in [1, 2], which starts from the largest classes T and $X \setminus T$ and then reaches the optimal classes according to a backward construction.

The target algorithm defines a strategy $\mathrm{Str} = (\mathrm{Str}_n)_{n\geq 0}$, where

$$\mathrm{Str}_n(X, \mathcal{F}, T, P) = (X, \mathcal{F}_n, T, P_n), \tag{3.1}$$

and it consists of three steps:

(1) the choice of a sequence $(\sim_{\epsilon_n})_{n\geq 1}$ of equivalences on the simplex defined on the unit ball of ℓ_1 with $\epsilon_n \to 0$;

(2) the definition of a filtration $(\mathcal{F}_n)_n$ based on $(\sim_{\epsilon_n})_{n\geq 1}$ where each $\mathcal{F}_n$ is generated by a countable partition of X;

(3) the choice of a suitable measure μ and the definition of $(P_n)_{n\geq 0}$.

3.1. Preliminary Results on Measurability and Equivalency and the Choice of $(\sim_{\epsilon_n})_{n\geq 1}$

Associated to each countably generated sub σ-algebra $\mathcal{A} \subseteq \mathcal{X}$, we define the equivalence relationship $\pi_{\mathcal{A}}$ induced by the atoms of $\mathcal{A}$:

$$x\pi_{\mathcal{A}}y \Longleftrightarrow [x]_{\mathcal{A}} := \cap\{A \in \mathcal{A} : x \in A\} = \cap\{A \in \mathcal{A} : y \in A\} =: [y]_{\mathcal{A}}. \tag{3.2}$$

Thus, if $(\mathcal{A}_n)_n$ is a sequence of countably generated σ-algebras, then

$$\pi_{\vee_n \mathcal{A}_n} = \cap_n \pi_{\mathcal{A}_n}. \tag{3.3}$$

Now, the atoms of the σ-algebra $\mathcal{F}$ of each simple target problem $(X, \mathcal{F}, T, Q)$ are at most countable, by definition. Then Q may be represented as a transition matrix on the state set $\mathbb{N}$. Each row of Q is a distribution probability on $\mathbb{N}$ (i.e., a sequence $(p_n)_{n\geq 1}$ in the simplex S of ℓ_1). The first step of the target algorithm is to equip S with the ℓ_1-norm and then to define an ϵ-equivalence on S.

We will alternatively use both the discrete equivalencies and the countable measurable partitions, as a consequence of the following result, whose proof is left to the appendices.

Lemma 3.1. *Given a measurable space $(X, \mathcal{X})$, there exists a natural bijection between the set of discrete equivalencies on X and the set of the countable measurable partitions of it.*

Let S_{ℓ_1} be the unit sphere in ℓ_1 and $S = \{x \geq 0\} \cap S_{\ell_1}$ be the simplex on ℓ_1. Let $\Omega_n = [0, 1]$, for any n, and τ be the standard topology on $[0, 1]$. Denote by $\mathcal{B}_{[0,1]}$ the Borel σ-algebra on $[0, 1]$ generated by τ. We look at S as a subset of $\Pi_{n=1}^{\infty}\Omega_n$ so that the Borel σ-algebra $\mathcal{B}_S$ induced on S is $\bigotimes_{n=1}^{\infty}\mathcal{B}_{[0,1]} \cap S$.

Definition 3.2. $\sim_\epsilon$ is an *ϵ-equivalence on S* if it is the product of finite equivalences on each $(\Omega_n, \mathcal{B}_{[0,1]})$, and $\|p - q\|_1 < \epsilon$ whenever $p \sim_\epsilon q$.

Remark 3.3. The choice of the ℓ_1-norm on S is linked to the total variation distance between probability measures. This distance between two probability measures P and Q is defined by

$d_{TV}(P,Q) = \sup_{A\in\Omega}|P(A) - Q(A)|$. On the other hand, the total variation of a measure μ is $\|\mu\|(\Omega) = \sup\sum_i |\mu(A_i)|$, where the supremum is taken over all the possible partitions of Ω. As $(P-Q)(\Omega) = 0$, we have that $d_{TV}(P,Q) = (1/2)\|P-Q\|$; see [6]. To each $p \in S$ corresponds the probability measure P on $\mathbb{N}$ with $P(i) = p_i$ (and vice versa). In fact, $p \in S$ implies $p_i \geq 0$ and $\sum_i p_i = 1$. Therefore, since $\|p-q\|_1 = \|P-Q\| = 2d_{TV}(P,Q)$, we have

$$p \sim_\epsilon q \Longrightarrow d_{TV}(P,Q) < \frac{\epsilon}{2}. \tag{3.4}$$

Example 3.4. Define the ϵ-cut as follows: $p \sim_\epsilon q \Leftrightarrow \lfloor p_n/\epsilon 2^{-n}\rfloor = \lfloor q_n/\epsilon 2^{-n}\rfloor$, for all n, where $\lfloor x\rfloor$ denotes the entire part of x. $\sim_\epsilon$ is an ϵ-equivalence on S. Indeed,

(i) for each n, define $p_n \sim_n q_n \Leftrightarrow \lfloor p_n/\epsilon 2^{-n}\rfloor = \lfloor q_n/\epsilon 2^{-n}\rfloor$. Then $\sim_n$ is a finite equivalence on Ω_n and $p \sim_\epsilon q \Leftrightarrow (p_n \sim_n q_n)$, for all n,

(ii) for any $p \in S$

$$[p] = \{q \in S : \pi_{\sim_\epsilon}(q) = \pi_{\sim_\epsilon}(p)\} = \prod_n \left[\frac{\lfloor 2^n p_n/\epsilon\rfloor \epsilon}{2^n}, \frac{(\lfloor 2^n p_n/\epsilon\rfloor + 1)\epsilon}{2^n}\right) \bigcap S \tag{3.5}$$

is measurable with respect to $\mathcal{B}_S$,

(iii) for all $p \sim_\epsilon q$,

$$\|p-q\|_1 \leq \sum_n \epsilon 2^{-n} = \epsilon. \tag{3.6}$$

3.2. The Choice of $(\mathcal{F}_n)_{n\geq 0}$

The following algorithm is a good candidate to be a strategy for the approximating problem we are facing. Given a sequence $(\sim_{\epsilon_n})_{n\geq 1}$ of ϵ-equivalences on S, we define $(\mathcal{F}_n)_{n\geq 0}$ inductively. Consider the equivalence classes given by $\mathcal{F}_{n-1}$ and divide them again according to the next rule. Starting from any two points in the same class, we check whether the probabilities to attain any other $\mathcal{F}_{n-1}$-classes are ϵ-the same. Mathematically speaking: we have the following steps.

Step 0. $\mathcal{F}_0 = \sigma(T) = \{\emptyset, T, X \setminus T, X\}$.

Step n. $\mathcal{F}_n$ is based on the equivalence $\mathcal{F}_{n-1}$ and on $\sim_{\epsilon_n}$, inductively. $\mathcal{F}_{n-1}$ is generated by a countable partition of X, say $(A_i^{(n-1)})_i$. We define, for any couple $(x,y) \in X^2$,

$$(x\pi_n y) \Longleftrightarrow (x\pi_{n-1}y) \wedge \left(\left(P\left(x, A_i^{(n-1)}\right)_i\right) \sim_{\epsilon_n} \left(P\left(y, A_i^{(n-1)}\right)_i\right)\right). \tag{3.7}$$

Lemma 3.6 shows that π_n is a discrete equivalency on $(X,\mathcal{X})$, and therefore it defines $\mathcal{F}_n = \sigma(X/\pi_n)$ as generated by a countable partition of X.

Remark 3.5. The choice of the "optimal" sequence $(\sim_{\epsilon_n})_{n\geq 1}$ is not the scope of this work. We only note that the definition of $\sim_\epsilon$ can be relaxed and the choice of the sequence $(\sim_{\epsilon_n})_{n\geq 1}$ may be done interactively, obtaining a fewer number of classes $(A_i^{(n)})_i$ at each step.

Lemma 3.6. *Let $(X, \mathcal{F}, T, P)$ be a target problem. $(\mathcal{F}_n)_{n\geq 0}$ defined as above, is a filtration on $(X, \mathcal{F})$ and for any $n \in \mathbb{N}$, π_n is a finite (and hence discrete) equivalency on $(X, \mathcal{X})$.*

Proof. The monotonicity of $(\mathcal{F}_n)_{n\geq 0}$ is a simple consequence of (3.7).

The statement is true for $n = 0$, since $T \in \mathcal{X}$. For the induction step, let $\{A_1^{(n-1)}, A_2^{(n-1)}, \ldots, A_{k_n}^{(n-1)}\} \in \mathcal{X}$ be the measurable partition of X given by X/π_{n-1}. The map $h : (X, \mathcal{X}) \to (S, \mathcal{B}(S))$ given by $x \mapsto (P(x, A_i^{(n-1)}))_i$ is therefore measurable. As $\sim_{\epsilon_n}$ is a finite equivalency on $\prod_1^{k_n}(\Omega_n, \mathcal{B}_{[0,1]})$, the map $\pi_{\sim_{\epsilon_n}} \circ h : (X, \mathcal{X}) \to (S/\sim_{\epsilon_n}, 2^{S/\sim_{\epsilon_n}})$ is also measurable, where $\pi_{\sim_{\epsilon_n}}$ is the natural projection associated with $\sim_{\epsilon_n}$. Thus, two points $x, y \in X$ are such that

$$\left(\left(P\left(x, A_i^{(n-1)}\right)_{i=1}^{k_n}\right) \sim_{\epsilon_n} \left(P\left(y, A_i^{(n-1)}\right)_{i=1}^{k_n}\right)\right) \tag{3.8}$$

if and only if their image by $\pi_{\sim_{\epsilon_n}} \circ h$ is the same point of $S/\sim_{\epsilon_n}$. It results that the new partition of X built by π_n is obtained as an intersection of the sets $A_i^{(n-1)}, 1 \leq i \leq k_n$—which formed the π_{n-1}-partition- with the counter-images of $(\prod_{i=1}^{k_n} \Omega_i)/\sim_{\epsilon_n}$ by $\pi_{\sim_{\epsilon_n}} \circ h$. Intersections between two measurable finite partitions of X being a measurable finite partition of X, we are done. □

3.3. *The Choice of μ and the Definition of $(P_n)_{n\geq 0}$*

Before defining $(P_n)_{n\geq 0}$, we need the following result, which will be proved in Section 5.

Theorem 3.7. *Let $(\pi_n)_{n\geq 0}$ be defined as in the previous section and let $\pi_\infty = \cap_n \pi_n$. Then π_∞ is a compatible projection.*

As a consequence of Theorems 2.7 and 3.7, a target problem $(X, \vee_n \mathcal{F}_n, T, P_\infty)$ is well defined. We intend to define P_n as the μ-weighted mean average of P_∞ given the information carried by $\mathcal{F}_n$.

More precisely, let μ be a probability measure on $(X, \vee_n \mathcal{F}_n)$ such that $\mu(F) > 0$, for any $F \in \mathcal{F}_n, F \neq \emptyset$ (the existence of such a measure is shown in Example 3.8).

For any $F \in \mathcal{F}_n$, let Y^F be the $\vee_n \mathcal{F}_n$-random variable such that $Y^F(\omega) = P_\infty(\omega, F)$. Define

$$P_n(x, F) = \mathbb{E}_\mu\left[Y^F \mid \mathcal{F}_n\right](x), \quad \forall x \in X, \ \forall F \in \mathcal{F}_n. \tag{3.9}$$

P_n is uniquely defined on $(X \times \mathcal{F}_n)$, the only μ-null set of $\mathcal{F}_n$ being the empty set. We claim that $P_n(x, \cdot)$ is a probability measure, for any $x \in X$.

We give in the following an example of the measure μ that has been used in (3.9) which justifies its existence.

Example 3.8. Let $(Y_n)_{n\geq 0}$ be a sequence of independent and identically distributed geometric random variables, with $\mathbb{P}_{Y_i}(j) = 1/2^j$, $j \in \mathbb{N}$. Let $\mathcal{A}_n = \sigma(Y_0, \ldots, Y_n)$ and set $\mathcal{A} = \vee \mathcal{A}_n$. There exists a probability measure $\mathbb{P}$ on $\mathcal{A}$ such that

$$\mathbb{P}(\cap_{i=0}^n \{Y_{l_i} = y_i\}) = \mathbb{P}_{Y_{l_1}}(y_1) \otimes \cdots \otimes \mathbb{P}_{Y_{l_n}}(y_n) = \frac{1}{2^{\sum_{i=0}^n y_i}}, \tag{3.10}$$

and thus, $\mathbb{P}(A) > 0$, for all $A \in \mathcal{A}_n$, $A \neq \emptyset$. Moreover, it follows that for any n,

$$A_1 \in \mathcal{A}_n, \quad A_2 \in \sigma(Y_{n+1}), \quad A_1 \neq \emptyset, \quad A_2 \neq \emptyset, \Longrightarrow \mathbb{P}(A_1 \cap A_2) > 0. \tag{3.11}$$

We check by induction that we can embed $\mathcal{F}_n$ into $\mathcal{A}_n$, for any $n \geq 0$. The required measure μ will be the trace of $\mathbb{P}$ on the embedded σ-field $\vee_n \mathcal{F}_n$.

For $n = 0$, define $T \mapsto \{Y_0 = 1\}$, $X \setminus T \mapsto \{Y_0 \geq 2\}$. The embedding forms a nontrivial partition, and therefore the restriction of $\mathbb{P}$ to the embedding of $\mathcal{F}_0$ defines a probability measure on $\mathcal{F}_0$ with $\mu_0(F) > 0$ if $F \neq \emptyset$.

For the induction step, suppose it is true for n. Given $F_i^{(n)} \in \mathcal{F}_n$, we have $F_i^{(n)} \mapsto A_i^{(n)}$, where $(A_i^{(n)})_i$ is a nontrivial partition in $\mathcal{A}_n$ and therefore the restriction of $\mathbb{P}$ to the embedding of $\mathcal{F}_n$ defines a probability measure μ_n on $\mathcal{F}_n$ with $\mu_n(F) > 0$ if $F \neq \emptyset$.

Given $F_i^{(n)}$, let $H_i^{(n+1)} := \{F_j^{(n+1)} : F_j^{(n+1)} \subseteq F_i^{(n)}\}$. The monotonicity of π_n ensures that each $F_j^{(n+1)}$ will belong to one and only one $H_i^{(n+1)}$. Moreover, by definition of $F_j^{(n+1)}$, we have that

$$F_i^{(n)} = \cup \left\{ F_j^{(n+1)} : F_j^{(n+1)} \in H_i^{(n+1)} \right\}. \tag{3.12}$$

Since X/π_{n+1} is at most countable, we may order $H_i^{(n+1)}$ for any i. We have accordingly defined an injective map $X/\pi_{n+1} \to \mathbb{N}^2$, where

$$F_j^{(n+1)} \longmapsto (i,k) \Longleftrightarrow F_j^{(n+1)} \text{ is the } k\text{th element in } H_i^{(n+1)}. \tag{3.13}$$

According to the cardinality of $H_i^{(n+1)}$, define the $n+1$-embedding

$$F_j^{(n+1)} \longmapsto (i,k) \longmapsto A_j^{(n+1)} := A_i^{(n)} \cap \begin{cases} \{Y_{n+1} = k\} & \text{if } k < \#\left\{ H_i^{(n+1)} \right\}, \\ \{Y_{n+1} \geq k\} & \text{if } k = \#\left\{ H_i^{(n+1)} \right\}. \end{cases} \tag{3.14}$$

By definition of $A_j^{(n+1)}$ and (3.12), it follows that we have mapped $\mathcal{F}_{n+1}$ into a partition in $\mathcal{A}_{n+1}$. Moreover, $\mathbb{P}(A_j^{(n+1)}) > 0$ as a consequence of (3.11). The restriction of $\mathbb{P}$ to the embedding of $\mathcal{F}_{n+1}$ defines a probability measure on $\mathcal{F}_{n+1}$ with $\mu_{n+1}(F) > 0$ if $F \neq \emptyset$. Note that μ_{n+1} is by construction an extension of μ_n to $\mathcal{F}_{n+1}$ since by (3.14),

$$\mu_n\left(F_i^{(n)}\right) = \sum_{F_j^{(n+1)} \in H_i^{(n+1)}} \mu_{n+1}\left(F_j^{(n+1)}\right). \tag{3.15}$$

Finally, the Carathéodory's extension theorem ensures the existence of the required μ, as $\mu(F) = \mu_n(F)$, for any $F \in \mathcal{F}_n$. Note that μ is just mapped to the trace of $\mathbb{P}$ on the embedded $\mathcal{F}_\infty$.

4. Blackwell

The problem of approximation is mathematically different if we start from a Markov process with a countable set of states or with an uncountable one. Let us consider, for the moment, the countable case: X is an at most countable set of the states and $\mathcal{X} = 2^X$ is the power set. Each function on X is measurable. If we take any equivalence relation on X, it is both measurable and identified by the σ-algebra it induces (see Theorem 4.6). This is not in general the case when we deal with a measurable space $(X, \mathcal{X})$, with X uncountable. In this section, we want to connect the process of approximation with the upgrading information. More precisely, a measurable equivalence $\pi = \pi_f$ defines both the partition X/π and the sigma algebra $\sigma(f)$. One wishes these two objects to be related, in the sense that ordering should be preserved. Example 4.5 shows a paradox concerning π_f and $\sigma(f)$ when X is uncountable. In fact, we have the following lemma.

Lemma 4.1. *Let $\mathcal{A}_1 \subseteq \mathcal{A}_2$ be countably generated sub σ-algebras of a measurable space $(X, \mathcal{X})$. Then $[x]_{\mathcal{A}_1} \supseteq [x]_{\mathcal{A}_2}$.*

In particular, let f, g be random variables. If $\sigma(f) \supseteq \sigma(g)$, then $\pi_f \subseteq \pi_g$.

Proof. See Appendix A. □

The problem is that even if a partition is more informative than another one, it is not true that it generates a finer σ-algebra, that is, the following implication is not always true for any couple of random variables f and g:

$$\pi_f \subseteq \pi_g \Longrightarrow \sigma(f) \supseteq \sigma(g). \tag{A0}$$

Then Lemma 4.1 is not invertible, if we do not require the further assumption (A0) on the measurable space $(X, \mathcal{X})$. This last fact connects the space $(X, \mathcal{X})$ with the theory of Blackwell spaces (see Lemma 4.3). We will assume the sole assumption (A0).

Example 4.2 ($\pi_f = \pi_g \nRightarrow \sigma(f) = \sigma(g)$). We give here a counterexample to assumption (A0), where two random variables f, g generate two different sigma algebras $\sigma(f) \neq \sigma(g)$ with the same set of atoms. Obviously, assumption (A0) does not hold. Let $(X, \mathcal{B}_X)$ be a Polish space and suppose $\mathcal{B}_X \subsetneq \mathcal{X}$. Let $A \in \mathcal{X} \setminus \mathcal{B}_X$ and consider the sequence $\{A_n, n \in \mathbb{N}\}$ that determines $\mathcal{B}_X$, that is, $\mathcal{B}_X = \sigma(A_n, n \in \mathbb{N})$. Let $\mathcal{A} = \sigma(A, A_n, n \in \mathbb{N})$. $\mathcal{B}_X \subsetneq \mathcal{A}$. As a consequence of Lemma A.3, there exist two random variables f, g such that $\mathcal{B}_X = \sigma(f)$ and $\mathcal{A} = \sigma(g)$. The atoms of $\mathcal{B}_X$ are the points of X, and then the atoms of $\mathcal{A}$ are also the points of X, since $\mathcal{B}_X \subseteq \mathcal{A}$.

We recall here the definition of Blackwell spaces. A measurable space $(X, \mathcal{X})$ is said *Blackwell* if $\mathcal{X}$ is a countably generated σ-algebra of X and $\mathcal{A} = \mathcal{X}$ whenever $\mathcal{A}$ is another countably generated σ-algebra of X such that $\mathcal{A} \subseteq \mathcal{X}$, and $\mathcal{A}$ has the same atoms as $\mathcal{X}$. A metric space X is Blackwell if, when endowed with its Borel σ-algebra, it is Blackwell. The measurable space $(X, \mathcal{X})$ is said to be a *strongly Blackwell space* if $\mathcal{X}$ is a countably generated σ-algebra of X and

(A1) $\mathcal{A}_1 = \mathcal{A}_2$ if and only if the sets of their atoms coincide, where $\mathcal{A}_1$ and $\mathcal{A}_2$ are countably generated σ-algebras with $\mathcal{A}_i \subseteq \mathcal{X}$, $i = 1, 2$.

For what concerns Blackwell spaces, the literature is quite extensive. Blackwell proved that every analytic subset of a Polish space is, with respect to its relative Borel σ-field, a strongly Blackwell space (see [7]). Therefore, if $(X, \mathcal{B}_X)$ is (an analytic subset of) a Polish space and $\mathcal{B}_X \subsetneq \mathcal{X}$, then $(X, \mathcal{X})$ cannot be a weakly Blackwell space (see Example 4.2). Moreover, as any (at most) countable set equipped with any σ-algebra may be seen as an analytic subset of a Polish space, then it is a strongly Blackwell space. More connections and examples involving Blackwell spaces, measurable sets and analytical sets in connection with continuum hypothesis (CH) may be found in [8–11]. Finally, note that assumption (A0) and assumption (A1) coincide, as the following lemma states.

Lemma 4.3. *Let* $(X, \mathcal{X})$ *be a measurable space. Then* (A0) *holds if and only if (A1) holds.*

Proof. Lemma A.3 in Appendix A asserts that $\mathcal{A} \subseteq \mathcal{X}$ is countably generated if and only if there exists a random variable f such that $\mathcal{A} = \sigma(f)$. In addition, as a consequence of Lemma 4.1, we have only to prove that (A1) implies (A0). By contradiction, assume (A1), $\pi_f \subseteq \pi_g$, but $\sigma(g) \not\subseteq \sigma(f)$. We have $\sigma(f,g) \neq \sigma(f)$, and then $\pi_{\sigma(f,g)} \neq \pi_f$ by (A1) and Lemma A.3. On the other hand, from (3.3), we have that $\pi_{\sigma(f,g)} = \pi_{\sigma(f)\vee\sigma(g)} = \pi_f \cap \pi_g = \pi_f$. □

We call *weakly Blackwell space* a measurable space $(X, \mathcal{X})$ such that assumption (A0) holds. If $(X, \mathcal{X})$ is a weakly Blackwell space, then $(X, \mathcal{F})$ is a weakly Blackwell space, for any $\mathcal{F} \subseteq \mathcal{X}$. Moreover, every strong Blackwell space is both a Blackwell space and a weakly Blackwell space whilst the other inclusions are not true in general. In [12, 13], examples are provided of Blackwell spaces which may be shown not to be weakly Blackwell. The following example shows that a weakly Blackwell space need not be Blackwell.

Example 4.4 (weakly Blackwell $\not\Rightarrow$ Blackwell). Let X be an uncountable set and $\mathcal{X}$ be the countable-cocountable σ-algebra on X. $\mathcal{X}$ is easily shown to be not countably generated, and therefore $(X, \mathcal{X})$ is not a Blackwell space. Take any countably generated σ-field $\mathcal{A} \subseteq \mathcal{X}$, that is, $\mathcal{A} = \sigma(\{A_i, i \in \mathbb{N}\})$.

(i) Since each set (or its complementary) of $\mathcal{X}$ is countable, then, without loss of generality, we can assume the cardinality of $X \setminus A_i$ to be countable.

(ii) Each atom B of $\sigma(A_i, i \in \mathbb{N})$ is of the form

$$B = \bigcap_{i=1,2,\ldots} C_i, \quad \text{where } C_i = A_i \text{ or } C_i = X \setminus A_i, \text{ for any } i. \tag{4.1}$$

Note that the cardinality of the set $A := \cup_i (X \setminus A_i)$ is countable, as it is a countable union of countable sets. As a consequence of (4.1), we face two types of atoms:

(1) for any i, $C_i = A_i$. This is the atom made by the intersections of all the uncountable generators. This is an uncountable atom, as it is equal to $X \setminus A$.

(2) exists i such that $C_i = X \setminus A_i$. This implies that this atom is a subset of the countable set A. Therefore, all the atoms (except $X \setminus A$) are disjoint subsets of the countable set A and hence they are countable.

It follows that the number of atoms of $\mathcal{A}$ is at most countable. Thus, $(X, \mathcal{A})$ is a strongly Blackwell space, that is, $(X, \mathcal{X})$ is a weakly Blackwell space.

Example 4.5 (Information and σ-algebra (see [4])). Suppose $X = [0,1]$, $\mathcal{X} = \sigma(\mathcal{Y}, A)$ where $\mathcal{Y}$ is the countable-cocountable σ-algebra on X and $A = [0, 1/2)$. Consider a decisionmaker who chooses action 1 if $x < 1/2$ and action 2 if $x \geq 1/2$. Suppose now that the information is modeled either as the partition of all elements of X, $\tau = \{x, x \in X\}$ and in this case the decisionmaker is perfectly informed, or as the partition $\tau' = \{A, X \setminus A\}$. If we deal with σ-algebras as a model of information then $\sigma(\tau) = \mathcal{Y}$ and $\sigma(\tau') = \sigma(A)$. The partition τ is more informative than τ', whereas $\sigma(\tau)$ is not finer than $\sigma(\tau')$. In fact $A \notin \mathcal{Y}$ and therefore if the decisionmaker uses $\sigma(\tau)$ as its structure of information, believing it more detailed than $\sigma(\tau')$, he will never know whether or not the event A has occurred and can be led to take the wrong decision. In this case, σ-algebras do not preserve information because they are not closed under arbitrary unions. However, if we deal with Blackwell spaces, any countably generated σ-algebra is identified by its atoms and therefore will possess an informational content (see, e.g., [14]).

The following theorem, whose proof is in Appendix B, connects the measurability of any relation to the cardinality of the space and assumption (A0). It shows the main difference between the uncountable case and the countable one.

Theorem 4.6. *Assume (CH). Let $(X, \mathcal{X})$ be a measurable space. The following properties are equivalent:*

(1) *any equivalence relation π on X is measurable and assumption* (A0) *holds,*

(2) $(X, 2^X)$ *is a weakly Blackwell space,*

(3) *X is countable and $\mathcal{X} = 2^X$.*

5. Proofs

The following theorem mathematically motivates our approximation problem: any limit of a monotone sequence of discrete equivalence relationships is a measurable equivalence.

Theorem 5.1. *For all $n \in \mathbb{N}$, let π_n be a discrete equivalency. Then $\pi_\infty = \cap_n \pi_n$ is a measurable equivalency. Conversely, for any measurable equivalency π, there exists a sequence $(\pi_n)_{n\geq 0}$ of discrete equivalencies such that $\pi_\infty = \cap_n \pi_n$.*

Proof. See Appendix A. □

Proof of Theorem 2.7. Let $\pi = \pi_f$ be a compatible projection. We define

$$P_\pi(x, F) := P(x, F), \quad \forall x \in X,\ \forall F \in \sigma(f). \tag{5.1}$$

What remains to prove is that $P_\pi \in \mathrm{TrP}(X, \mathcal{X}, \sigma(f))$. More precisely, we have to show that $P_\pi(\cdot, F)$ is $\sigma(f)$-measurable, for all $F \in \sigma(f)$. By contradiction, there exists $F \in \sigma(f)$ such that the random variable $Y^F(\omega) = P_\pi(\omega, F)$ is not $\sigma(f)$-measurable. Then $\sigma(Y^F) \not\subseteq \sigma(f)$, and hence $\pi_{Y^F} \not\supseteq \pi_f$ by assumption (A0), which contradicts (2.7). □

Proof of Theorem 3.7. As a consequence of Theorem 5.1, $\pi_\infty = \pi_f$, where $\sigma(f) = \vee_n \mathcal{F}_n$. Define

$$P_\infty(x, F) := P(x, F), \quad \forall x \in X,\ \forall F \in \sigma(f). \tag{5.2}$$

We will prove that, for any $F \in \sigma(f)$, $P_\infty(\cdot, F)$ is $\sigma(f)$-measurable and consequently π_∞ will be a compatible projection. This implies that there exists a measurable function $h_F : (\mathbb{R}, \mathcal{B}_R) \to (\mathbb{R}, \mathcal{B}_R)$ so that $P_\infty(\omega, F) = h_F(f(\omega))$. Therefore, if $x \pi_f y$, then $P_\infty(x, F) = P_\infty(y, F)$, which is the thesis.

We need to show that for any $F \in \sigma(f)$ and $t \in \mathbb{R}$, we have

$$H := \{x : P(x, F) \leq t\} \in \sigma(f). \tag{5.3}$$

We first show that it is true when $F \in \mathcal{F}_n$ by proving that

$$H = \bigcap_{m>n} \pi_m^{-1} \pi_m(H), \tag{5.4}$$

which implies that $H \in \sigma(f)$. The inclusion $H \subseteq \cap_m \pi_m^{-1} \pi_m(H)$ is always true. For the other inclusion, let $y \in \cap_{m>n} \pi_m^{-1} \pi_m(H)$. Let $m > n$; there exists $x_m \in H$ such that $y \pi_m x_m$. Therefore, (3.7) and the definition of $\sim_{\epsilon_n}$ imply $P(y, F) \leq P(x_m, F) + \epsilon_m \leq t + \epsilon_m$, for any $m > n$. As $\epsilon_m \searrow 0$, we obtain that $y \in H$. Then (5.3) is true on the algebra $\mathrm{Alg} := \cup_n \mathcal{F}_n$.

Actually, let $F_n \in \mathrm{Alg}$ such that $F_n \nearrow F$. We prove that (5.3) holds for F by showing that

$$H = \{x : P(x, F) \leq t\} = \bigcap_n \{x : P(x, F_n) \leq t\} =: \bigcap_n H_n. \tag{5.5}$$

Again, since $F_n \subseteq F$, then $P(x, F_n) \leq P(x, F)$ and therefore $H \subseteq \bigcap_n H_n$. Conversely, the set $\cap_n H_n \setminus H$ is empty since the sequence of $\mathcal{X}$-measurable maps $P(\cdot, F) - P(\cdot, F_n)$ converges to 0:

$$P(\cdot, F) - P(\cdot, F_n) = P(\cdot, F \setminus F_n) \longrightarrow P(\cdot, \emptyset) = 0. \tag{5.6}$$

Then (5.3) is true on the monotone class generated by the algebra $\mathrm{Alg} = \cup_n \mathcal{F}_n$, that is, on $\sigma(f)$. □

Proof of Theorem 2.8. Given a target algorithm $(X, \mathcal{F}_n, T, P_n)_n$, let $\pi_\infty = \pi_f$ be defined as in Theorem 3.7. We claim that π_∞ is optimal. Let ψ_g be another compatible projection and let $(X, \sigma(g), T, P_g)$ be the target problem given by Theorem 2.7. We are going to prove by induction on n that

$$\forall n \in \mathbb{N}, \quad \mathcal{F}_n \subseteq \sigma(g). \tag{5.7}$$

In fact, for $n = 0$ it is sufficient to note that $\mathcal{F}_0 = \sigma(\{T\}) \subseteq \sigma(g)$.

Equation (3.7) states that $\mathcal{F}_n = \sigma(\mathcal{F}_{n-1}, h_n)$, where h_n is the discrete random variable, given by Lemma 3.1, s.t.

$$\begin{gathered} x \pi_{h_n} y \\ \Updownarrow \\ \left(\left(P\left(x, A_i^{(n-1)}\right)\right)_i\right) \sim_{\epsilon_n} \left(P\left(y, A_i^{(n-1)}\right)_i\right)\right). \end{gathered} \tag{5.8}$$

Let $k_i^{(n-1)} : \mathrm{X} \to [0,1]$ be defined as $k_i^{(n-1)}(x) = P(x, A_i^{(n-1)})$. Then

$$\begin{tikzcd} X \arrow[d, "(k_i^{(n-1)})_{i\in\mathbb{N}}"'] \arrow[dr, "h_n"] & \\ S \arrow[r, "\epsilon_n"'] & S/\epsilon_n \end{tikzcd} \tag{5.9}$$

Obviously, $\sigma(h_n) \subseteq \sigma(k_1^{(n-1)}, k_2^{(n-1)}, \ldots)$. For the induction step, as $A_i^{(n-1)} \in \mathcal{F}_{n-1} \subseteq \sigma(g)$, we have that $P_g(\cdot, A_i^{(n-1)})$ is $\sigma(g)$-measurable, and therefore $\sigma(k_i^{(n-1)}) \subseteq \sigma(g)$. Then $\mathcal{F}_n = \sigma(\mathcal{F}_{n-1}, h_n) \subseteq \sigma(\mathcal{F}_{n-1}, k_1^{(n-1)}, k_2^{(n-1)}, \ldots) \subseteq \sigma(g)$. Therefore $\sigma(f) = \vee_n \mathcal{F}_n \subseteq \sigma(g)$, which implies $\pi_\infty \supseteq \psi_g$ by Lemma 4.1, and hence π_∞ is optimal. □

Corollary 5.2. *π_∞ does not depend on the choice of* Str.

Proof. $\pi_\infty = \cap_n \pi_n$ is optimal, for all $(\pi_n)_{n\geq 0} = \mathrm{Str}(P)$. The optimal projection being unique, we are done. □

Proof of Theorem 2.4. Let $\pi_\infty = \pi_f$ be defined as in Theorem 3.7 and $(X, \sigma(f), T, P_\infty)$ be given by Theorem 2.7 so that $P(x, F) = P_\infty(x, F)$ for any $F \in \sigma(f)$. Then each P_n of (3.9) can be rewritten as

$$P_n(x, F) = \frac{\int_{[x]_n} P_\infty(x, F)\mu(dz)}{\mu([x]_n)}, \quad \forall x \in X,\ \forall F \in \mathcal{F}_n, \tag{5.10}$$

where $[x]_n$ is the π_n-class of equivalence of x and $\mu([x]_n) > 0$ since $[x]_n \neq \emptyset$.

Note that $d(P, P_m) \leq 2\sum_n \beta^n$. Then, for any $\epsilon > 0$, there exists an N so that $\sum_{n>N} \beta^n \leq \epsilon/2$. Therefore we are going to prove by induction on n that

$$\sup_x |P_m^n(x, T) - P^n(x, T)| \longrightarrow 0 \quad \text{as } m \text{ tends to infinity}, \tag{5.11}$$

which completes the proof. If $n = 1$, then by definition of ϵ_m, since $T \in \mathcal{F}_{m-1}$, we have that

$$\begin{aligned} |P_m(x, T) - P(x, T)| &\leq \frac{\int_{[x]_m} |P_\infty(z, T) - P(x, T)|\mu(dz)}{\mu([x]_m)} \\ &= \frac{\int_{[x]_m} |P(z, T) - P(x, T)|\mu(dz)}{\mu([x]_m)} \\ &\leq \epsilon_m \frac{\int_{[x]_m} \mu(dz)}{\mu([x]_m)} = \epsilon_m. \end{aligned} \tag{5.12}$$

For the induction step, we note that

$$\left|P_m^{n+1}(x, T) - P^{n+1}(x, T)\right| \leq \sum_i \left| P_m\left(x, A_i^{(m)}\right) P_m^n\left(A_i^{(m)}, T\right) - \int_{A_i^{(m)}} P(x, dz) P^n(z, T) \right|, \tag{5.13}$$

where $(A_i^{(m)})_i$ is the partition of X given by π_m. By induction hypothesis, for any $\tilde{\epsilon} > 0$,

$$|P_m^n(z, T) - P^n(z, T)| \leq \tilde{\epsilon} \tag{5.14}$$

for $m \geq m_0$ large enough. Since $[z]_m = A_i^{(m)}$ if $z \in A_i^{(m)}$, it follows that

$$\int_{A_i^{(m)}} P(x,dz)\left|P_m^n\left(A_i^{(m)},T\right) - P^n(z,T)\right| \leq \tilde{\epsilon}\int_{A_i^{(m)}} P(x,dz). \tag{5.15}$$

Equation (5.13) becomes

$$\begin{aligned}\left|P_m^{n+1}(x,T) - P^{n+1}(x,T)\right| &\leq \tilde{\epsilon} + \sum_i P_m^n\left(A_i^{(m)},T\right)\left|P_m\left(x,A_i^{(m)}\right) - P\left(x,A_i^{(m)}\right)\right| \\ &\leq \tilde{\epsilon} + \sum_i \left|P_m\left(x,A_i^{(m)}\right) - P\left(x,A_i^{(m)}\right)\right|.\end{aligned} \tag{5.16}$$

On the other hand, by (5.10),

$$P_m\left(x,A_i^{(m)}\right) - P\left(x,A_i^{(m)}\right) = \int_{[x]_m} \frac{P_\infty\left(z,A_i^{(m)}\right) - P\left(x,A_i^{(m)}\right)}{\mu([x]_m)}\mu(dz). \tag{5.17}$$

The definition of $\sim_{\epsilon_{m+1}}$ states that

$$\sum_i \left|P_\infty\left(z,A_i^{(m)}\right) - P\left(x,A_i^{(m)}\right)\right| \leq \epsilon_{m+1} \tag{5.18}$$

whenever $z \in [x]_m$ and therefore

$$\left|P_m^{n+1}(x,T) - P^{n+1}(x,T)\right| \leq \tilde{\epsilon} + \int_{[x]_m}\sum_i \left|P_\infty\left(z,A_i^{(m)}\right) - P\left(x,A_i^{(m)}\right)\right|\frac{\mu(dz)}{\mu([x]_m)} \leq \tilde{\epsilon} + \epsilon_{m+1}. \tag{5.19}$$

Since $\epsilon_m \to 0$ as m tends to infinity, we get the result. □

Proof of Theorem 2.9. By (3.9) and Lemma 3.6, $(P_n(\cdot,F))_{n\geq m}$ is a martingale with respect to the filtration $(\mathcal{F}_n)_{n\geq m}$, for any $F \in \mathcal{F}_m$. Then, if $Y^F(x) = P(x,F)$ as in (3.9), we have that

$$P_n(x,F) \underset{n\to\infty}{\longrightarrow} \mathbb{E}_\mu\left[Y^F \mid \vee_n \mathcal{F}_n\right](x) = Y^F(x), \quad \text{for } \mu\text{-a.e. } x \in X,\ \forall F \in \mathcal{F}_m. \tag{5.20}$$

Let $\pi_\infty = \pi_f$ be defined as in Theorem 3.7 and $(X,\sigma(f),T,P_\infty)$ given by Theorem 2.7 so that $P(x,F) = P_\infty(x,F)$ for any $F \in \sigma(f)$. Unfortunately, (5.20) is not enough to state that

$$P_n(x,F) \underset{n\to\infty}{\longrightarrow} P_\infty(x,F), \quad \text{for } x \in X,\ \forall F \in \bigcup_m \mathcal{F}_m, \tag{5.21}$$

even if $\sigma(f)$ is countably generated (see, e.g., [15], for counterexample). In fact, Polish assumption is assumed in [15] to guarantee (5.21).

Here, we will deal with the specific properties of P_n and P_∞ to deduce (5.21). Take $F \in \mathcal{F}_m$ and $n > m$. By (5.10) and the definition of ϵ_n, since $F \in \mathcal{F}_{n-1}$, we have, for any $x \in X$, that

$$\begin{aligned} |P_n(x,F) - P(x,F)| &\le \frac{\int_{[x]_n} |P_\infty(z,F) - P(x,F)|\mu(dz)}{\mu([x]_n)} \\ &= \frac{\int_{[x]_n} |P(z,F) - P(x,F)|\mu(dz)}{\mu([x]_n)} \\ &\le \epsilon_n \frac{\int_{[x]_n} \mu(dz)}{\mu([x]_n)} = \epsilon_n \end{aligned} \tag{5.22}$$

since the only μ-null set in $\mathcal{F}_n$ is the empty set. Then

$$P_n(x,F) \underset{n\to\infty}{\longrightarrow} P_\infty(x,F) \tag{5.23}$$

for any $x \in X$ and $F \in \cup_m \mathcal{F}_m$. □

5.1. Weak Convergence of Conditional Probabilities

Let the target problem $(X,\mathcal{F},T,P)$ be given and let $\mathrm{Str} = (\mathrm{Str}_n)_n$, where $\mathrm{Str}_n(X,\mathcal{F},T,P) = (X,\widetilde{\mathcal{F}}_n,T,\widetilde{P}_n)$ be a target algorithm. In order to prove Theorem 2.13, which states the total convergence of the probability measure $P_n(x,\cdot)$ towards $P(x,\cdot)$, we proceed as follows:

(i) first, we define the topology ϱ_P on X;

(ii) then, we define a "natural" topology τ_{Str} on X associated to any target algorithm $(\mathrm{Str}_n)_n$. We prove in Theorem 5.4 the total convergence of $(P_n)_{n\ge 0}$ to P_∞, under this topology;

(iii) then, we define the topology τ_P on X as the intersection of all the topologies τ_{Str};

(iv) finally, we show Theorem 2.13 by proving that $\varrho_P \subseteq \tau_{\mathrm{Str}}$. The nontriviality of ϱ_P will imply that of τ_P.

We introduce the pseudometric d_P on X as follows:

$$d_P(x,y) = \sum_n \beta^n \left| P^n(x,T) - P^n(y,T) \right|. \tag{5.24}$$

Now, let τ_{Str} be the topology generated by $\cup_n \mathcal{F}_n$. C is a closed set if and only if $C = \cap_n C_n$, $C_n \in \mathcal{F}_n$. In fact, if $C \in \mathcal{F}_n$, for a given n, then $C \in \mathcal{F}_{n+p}$, for any p and therefore C is closed. (X,τ_{Str}) is a topological space.

Remark 5.3. Let us go back to Example 2.11. The topology defined by asking that the sets in each $\mathcal{F}_n$ are closed is strictly finer than the standard topology. On the other hand, the same example may be explained with left closed-right opened dyadic subdivisions, which leads to a different topology that also contains the natural one. Any other "reasonable" choice of subdivision will lead to the same point: the topologies are different, and all contain the standard one. In the same manner, we are going to show that all the topologies τ_{Str} contain the standard one, ϱ_P.

Theorem 5.4. *Let the target problem* $(X, \mathcal{F}, T, P)$ *and the target algorithm* $(X, \mathcal{F}_n, T, P_n)_n$ *be given. For any target algorithm* Str,

$$P_n(x,\cdot) \xrightarrow[\tau_{\mathrm{Str}}]{\mathrm{tot}} P(x,\cdot), \quad \forall x \in X. \tag{5.25}$$

Proof. Denote by $\overline{P}_n$ any extension of P_n to $\vee_n \mathcal{F}_n$. We have to check that $\limsup_n \overline{P}_n(x,C) \le P(x,C)$, for any closed set C of Str and $x \in X$ (see, e.g., [6]).

Let $\{C_n \in \mathcal{F}_n\}$, with $C_n \supseteq C_{n+1}$ and $C = \cap_n C_n$ (take, e.g., C_n as the closure of C in $\mathcal{F}_n$).

Note that, since $C \in \vee_n \mathcal{F}_n$, we have $P(x,C) = P_\infty(x,C)$. But, $\overline{P}_n(x,C) - P_\infty(x,C) \le \overline{P}_n(x,C_{n-1}) - P_\infty(x,C) = P_n(x,C_{n-1}) - P_\infty(x,C)$. Actually,

$$P_n(x,C_{n-1}) - P_\infty(x,C) = \left(\underbrace{P_n(x,C_{n-1}) - P_\infty(x,C_{n-1})}_{I}\right) + \left(\underbrace{P_\infty(x,C_{n-1}) - P_\infty(x,C)}_{II}\right). \tag{5.26}$$

$I \to 0$ as n tends to infinity, from the target algorithm and $II \to 0$ as n tends to infinity, from the continuity of the measure. □

An example of a natural extension of P_n to $\overline{P}_n$ is given by

$$\overline{P}_n(x,F) = \mathbb{E}_\mu\left[Y^F \mid \mathcal{F}_n\right](x), \quad \forall x \in X,\ \forall F \in \vee_n \mathcal{F}_n, \tag{5.27}$$

where, for any $F \in \vee_n \mathcal{F}_n$, Y^F is the $\vee_n \mathcal{F}_n$-random variable such that $Y^F(\omega) = P_\infty(\omega, F)$. As mentioned for P_n, $\overline{P}_n(x,\cdot)$ is a probability measure, for any $x \in X$.

Corollary 5.5. *For any fixed strategy* $\mathrm{Str}(P)$, *let* P_n *be as in Theorem 2.4. We have*

$$P_n(x,\cdot) \xrightarrow[\tau_P]{\mathrm{tot}} P(x,\cdot), \tag{5.28}$$

for any given x.

In order to describe the topology τ_P, we will denote by $[[F]]_*$ the closure of a set $F \subseteq X$ in a given topology $*$. Note that the monotonicity of π_n implies

$$[[F]]_{\tau_{\mathrm{Str}}} = \bigcap_n [[F]]_{\tau_{\mathrm{Str}_n}} \tag{5.29}$$

where τ_{Str_n} is the (discrete) topology on X generated by $\mathcal{F}_n$. Since τ_P is the intersection of all the topologies τ_{Str}, we have

$$[[F]]_{\tau_P} \supseteq [[F]]_{\tau_{\mathrm{Str}}} = \bigcap_n [[F]]_{\tau_{\mathrm{Str}_n}}, \quad \forall F \in 2^X,\ \forall \mathrm{Str}. \tag{5.30}$$

Proof of Theorem 2.13. Let F be the closed set in ϱ_P so defined

$$F := \{y \in X : d_P(y, x) \geq r\}, \tag{5.31}$$

that is, F is the complementary of an open ball in (X, d_P) with center x and radius r. If we show that $F \in \tau_P$, then we are done, as the arbitrary choice of x and r spans a base for the topology ϱ_P.

We are going to prove

$$F = [[F]]_{\tau_{\mathrm{Str}}} = \bigcap_m [[F]]_{\tau_{\mathrm{Str}_m}}, \quad \forall \mathrm{Str}, \tag{5.32}$$

which implies $[[F]]_{\tau_P} = F$. It is always true that $F \subseteq [[F]]_*$; we prove the nontrivial inclusion $F \supseteq \cap_m [[F]]_{\tau_{\mathrm{Str}_m}}$. Assume that $y \in [[F]]_{\tau_{\mathrm{Str}}}$. Now, $y \in [[F]]_{\tau_{\mathrm{Str}_m}}$, for any m, and then there exists a sequence $(y_m)_{n \geq 0}$ with $y_m \in F$ such that $y \pi_m y_m$, for any m. Thus, $y \in \cap_m [y_m]_m$, where $[x]_m$ is the π_m-class of equivalence of x. Thus

$$P_m^n(y_m, T) = P_m^n(y, T), \quad \forall m, n \tag{5.33}$$

since $P_m(\cdot, T)$ is $\mathcal{F}_m$-measurable. By Theorem 2.4, for any $n \in \mathbb{N}$,

$$\left|P^n(y, T) - P_m^n(y, T)\right| + \left|P_m^n(y_m, T) - P^n(y_m, T)\right| \underset{m \to \infty}{\longrightarrow} 0. \tag{5.34}$$

Now, let N be such that $\sum_{n=N}^{\infty} \beta^n \leq \epsilon/4$ and take n_0 sufficiently large s.t.

$$\sum_{n=0}^{N} \left|P^n(y, T) - P_{n_0}^n(y, T)\right| + \left|P_{n_0}^n(y_{n_0}, T) - P^n(y_{n_0}, T)\right| \leq \frac{\epsilon}{2}. \tag{5.35}$$

We have

$$\begin{aligned}
d_P(y_{n_0}, y) &= \sum_n \beta^n \left|P^n(y, T) - P^n(y_{n_0}, T)\right| \\
&\leq \sum_{n=0}^{N} \left|P^n(y, T) - P^n(y_{n_0}, T)\right| + 2\sum_{n=N}^{\infty} \beta^n \\
&\leq \sum_{n=0}^{N} \Big(\left|P^n(y, T) - P_{n_0}^n(y, T)\right| \\
&\qquad + \left|P_{n_0}^n(y, T) - P_{n_0}^n(y_{n_0}, T)\right| + \left|P_{n_0}^n(y_{n_0}, T) - P^n(y_{n_0}, T)\right| \Big) \\
&\quad + 2\frac{\epsilon}{4} \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,
\end{aligned} \tag{5.36}$$

and therefore

$$d_P(x, y) \geq d_P(x, y_{n_0}) - d_P(y_{n_0}, y) \geq r - \epsilon. \tag{5.37}$$

The arbitrary choice of ϵ implies $y \in F$, which is the thesis. □

Appendices

A. Results on Equivalence Relations

In this appendix we give the proof of auxiliary results that connect equivalency with measurability.

Proof of Lemma 3.1. Let $\pi = \pi_f$ be a discrete equivalency on X. Then X/π defines a countable measurable partition of X. Conversely, let $\{A_1, A_2, \ldots\}$ be a countable measurable partition on X. Define $f : X \to \mathbb{N}$ s.t. $f(x) = n \Leftrightarrow x \in A_n$. Therefore f is measurable and $\pi = \pi_f$ is a discrete equivalency on X. □

Lemma A.1. *Let f, g be two random variables such that $g(x) < g(y) \Rightarrow f(x) < f(y)$. Then $\sigma(g) \subseteq \sigma(f)$.*

Proof. Let $t \in \mathbb{R}$ be fixed. We must prove that $\{g \le t\} \in \sigma(f)$. If $\{g \le t\}$ or $\{g > t\}$ are empty, then we are done. Assume then that $\{g \le t\}, \{g > t\} \ne \emptyset$ and define $t^* = \sup(f(\{g \le t\}))$. We have the following two cases.

Case 1 ($t^* \in f(\{g \le t\}) : \exists x^* \in \{g \le t\}$ such that $t^* = f(x^*)$). By definition of t^*, $\{g \le t\} \subseteq \{f \le t^*\}$. Conversely, let $y \in \{g > t\}$. Since $g(x^*) \le t < g(y)$, then $f(x^*) = t^* < f(y)$, that is, $\{g > t\} \subseteq \{f > t^*\}$. Then $\{g \le t\} = \{f \le t^*\} \in \sigma(f)$.

Case 2 ($t^* \notin f(\{g \le t\}) : \forall x \in \{g \le t\}$ we have that $f(x) < t^*$). Then $\{g \le t\} \subseteq \{f < t^*\}$. Conversely, let $y \in \{g > t\}$. Since $\forall x \in \{g \le t\} g(y) > g(x)$, then $f(y) > f(x)$, which implies $f(y) \ge \sup f(\{g \le t\}) = t^*$, that is, $\{g > t\} \subseteq \{f \ge t^*\}$. Then $\{g \le t\} = \{f < t^*\} \in \sigma(f)$. □

The next lemma plays a central rôle. Its proof is common in set theory.

Lemma A.2. *For all $n \in \mathbb{N}$, let π_n be a discrete measurable equivalency. Then there exists a random variable f such that $\sigma(f) = \vee_n \sigma(X/\pi_n)$.*

Proof. Before proving the core of the Lemma, we build a sequence $(g_n)_{n\in\mathbb{N}}$ of functions $g_n : \mathbb{N}^n \to \mathbb{R}$ that will be used to define the function f.

Take $h : \mathbb{N} \cup \{0\} \to [0, 1)$ to be the increasing function $h(m) = 1 - 2^{-m}$ and let $(g_n)_{n\in\mathbb{N}}$ the sequence of function $g_n : \mathbb{N}^n \to \mathbb{R}$ so defined:

$$\begin{aligned} g_1(m_1) &= h(m_1 - 1), \\ g_2(m_1, m_2) &= g_1(m_1) + h(m_2 - 1)\Delta g_1(m_1), \\ &\vdots \\ g_{n+1}(\mathbf{m}_n, m_{n+1}) &= g_n(\mathbf{m}_n) + h(m_{n+1})\Delta g_n(\mathbf{m}_n), \\ &\vdots \end{aligned} \tag{A.1}$$

where, for all n, $\mathbf{m}_n = (m_1, \ldots, m_n)$ and

$$\Delta g_n(\mathbf{m}_n) = g_n(\mathbf{m}_{n-1}, m_n + 1) - g_n(\mathbf{m}_{n-1}, m_n). \tag{A.2}$$

As a first consequence of the definition, note that for any choice of n and $\mathbf{m}_{n+1}$, it holds that

$$g_n(\mathbf{m}_{n-1}, m_n) \leq g_{n+1}(\mathbf{m}_{n+1}) < g_n(\mathbf{m}_{n-1}, m_n + 1) \tag{A.3}$$

since $h \in [0,1)$. We now prove by induction on n_1+n_2 that for any choice of $n_1 \in \mathbb{N}$, $n_2 \in \mathbb{N}\cup\{0\}$ and $\mathbf{m}_{n_1+n_2}$, we have

$$g_{n_1}(\mathbf{m}_{n_1-1}, m_{n_1}) \leq g_{n_1+n_2}(\mathbf{m}_{n_1+n_2}) < g_{n_1}(\mathbf{m}_{n_1-1}, m_{n_1} + 1). \tag{A.4}$$

Equation (A.4) is clearly true for $n_1 + n_2 = 1$, since h is strictly monotone. The same argument shows that (A.4) is always true for $n_2 = 0$ and therefore we check it only for $n_2 > 0$. We assume by induction that (A.4) is true for $n_1 + n_2 \leq n$ and we prove it for $n_1 + n_2 = n + 1$. By using twice the induction hypothesis, as $n_2 - 1 \geq 0$, we obtain

$$\begin{aligned} g_{n_1}(\mathbf{m}_{n_1-1}, m_{n_1}) &\leq g_{n_1+n_2-1}(\mathbf{m}_{n_1+n_2-2}, m_{n_1+n_2-1}) \\ &< g_{n_1+n_2-1}(\mathbf{m}_{n_1+n_2-2}, m_{n_1+n_2-1} + 1) \\ &\leq g_{n_1}(\mathbf{m}_{n_1-1}, m_{n_1} + 1). \end{aligned} \tag{A.5}$$

Equation (A.4) is now a consequence of (A.3).

Now, we come back to the proof of the lemma. First note that, without loss of generality, we can (and we do) require the sequence $(\pi_n)_{n\in\mathbb{N}}$ to be monotone, by taking the sequence $\pi'_n = \cap_{i=1}^n \pi_i$ instead of π_n. π'_n is again a countable measurable equivalency on X. In fact, by Lemma 3.1 we can read this statement in trivial terms of partitions: an at most countable intersection of countable measurable partitions is still a countable measurable partition. Moreover, by definition, $\vee_{i=1} 1^n \sigma(X/\pi_i) = \vee_{i=1} 1^n \sigma(X/\pi'_i)$.

Let $\tau_n = X/\pi_n$ be the increasing sequence of countable measurable dissections of X. We are going to give a consistent inductive method of numbering the set of atoms of τ_n to build the functions f_n. Let $\tau_1 = \{A_1^{(1)}, A_2^{(1)}, \ldots\}$ be any ordering of τ_1. By induction, let $\{A_{\mathbf{m}_n,1}^{(n+1)}, A_{\mathbf{m}_n,2}^{(n+1)}, \ldots\}$ be the partition of the atom $A_{\mathbf{m}_n}^{(n)} \in \tau_n$ given by τ_{n+1}. Define, for any $n \in \mathbb{N}$,

$$f_n(x) = g_n(\mathbf{m}_n) \Longleftrightarrow x \in A_{\mathbf{m}_n}^{(n)}. \tag{A.6}$$

To complete the proof, we first show that $\sigma(f_n) = \sigma(X/\pi_n)$, $\forall n$, and then we prove $\sigma(f) = \sigma(f_1, f_2, \ldots)$ by proving that $f_n \to f$ pointwise.

To prove that $\sigma(f_n) = \sigma(X/\pi_n)$ we show that $f_n(x) = f_n(y) \Leftrightarrow \exists \mathbf{m}_n : x, y \in A_{\mathbf{m}_n}^{(n)}$. One implication is a consequence of the fact that f_n is defined on the partition of X given by $X/\pi_n = \tau_n$. For the converse, assume that $x \in A_{\mathbf{m}_n}^{(n)} \neq A_{\mathbf{m}'_n}^{(n)} \ni y$ and consider $n_1 := \min\{j \leq n : m_j \neq m'_j\}$. Thus $\mathbf{m}_{n_1-1} = \mathbf{m}'_{n_1-1}$ and, without loss of generalities, $m_{n_1} < m'_{n_1}$. By (A.4), we have

$$f_n(x) = g_n(\mathbf{m}_n) < g_{n_1}(\mathbf{m}_{n_1-1}, m_{n_1} + 1) \leq g_{n_1}\left(\mathbf{m}'_{n_1-1}, m'_{n_1}\right) \leq g_n(\mathbf{m}'_n) = f_n(y). \tag{A.7}$$

We are going to prove that $\sigma(f) = \sigma(f_1, f_2, \ldots)$.

Proof of $\subseteq$. The sequence $(f_n)_n$ is monotone by definition and bounded by (A.4). Then $\exists f$: $f_n \uparrow f$ and thus $\sigma(f) \subseteq \sigma(f_1, f_2, \ldots)$. $\square$

Proof of $\supseteq$. Let n be fixed, and take $x, y \in X$ with $f_n(x) < f_n(y)$. Then, for any $h \geq 0$, $\tau_n \subseteq \tau_{n+h}$ implies $x \in A^{(n+h)}_{\mathbf{m}_{n+h}} \neq A^{(n+h)}_{\mathbf{m}'_{n+h}} \ni y$. As above, consider $n_1 := \min\{j \leq n : m_j \neq m'_j\}$. As $f_n(x) < f_n(y)$, we have $\mathbf{m}_{n_1-1} = \mathbf{m}'_{n_1-1}$ and $m_{n_1} < m'_{n_1}$. Again, by (A.4), for $h > n_1 + 1 - n$,

$$\begin{aligned} f_{n+h}(x) &= g_{n+h}(\mathbf{m}_{n+h}) \\ &< g_{n_1+1}(\mathbf{m}_{n_1}, m_{n_1+1} + 1) = \alpha \\ &< g_{n_1}(\mathbf{m}_{n_1-1}, m_{n_1} + 1) \\ &\leq g_n(\mathbf{m}'_n) = f_n(y), \end{aligned} \tag{A.8}$$

that is, $\forall h$, $f_{n+h}(x) < \alpha < f_n(y)$. As $f_l \uparrow f$, $f(x) < f(y)$. Apply Lemma A.1 with $g = f_n$ to conclude that $\sigma(f_n) \subseteq \sigma(f)$. $\square$

$\square$

As a consequence of Lemma A.2, any countably generated sub-σ-algebra is generated by a measurable equivalence π, as the following lemma states.

Lemma A.3. *$\mathcal{A} \subseteq \mathcal{X}$ is countably generated if and only if there exists a random variable f such that $\mathcal{A} = \sigma(f)$.*

Proof. $\Rightarrow$ Let $\mathcal{A} = \sigma(A_1, A_2, \ldots)$. Apply Lemma A.2 with $X/\pi_n = \{A_n, X \setminus A_n\}$.

$\Leftarrow$ Take a countable base $B_1, B_2, \ldots$ of $\mathcal{B}_{\mathbb{R}}$ and simply note that $\sigma(f) = \sigma(\{f^{-1}(B_1), f^{-1}(B_2), \ldots\})$. $\square$

Proof of Lemma 4.1. Let $x \in X$ be fixed. By hypothesis, $\mathcal{A}_1 \subseteq \mathcal{A}_2$. If $\mathcal{A}_1 = \sigma(A^1_1, A^1_2, \ldots)$ then $\mathcal{A}_2$ will be of the form $\mathcal{A}_2 = \sigma(A^1_1, A^2_1, A^1_2, A^2_2, \ldots)$. Without loss of generality (if needed, by choosing $X \setminus A^j_n$ instead of A^j_n) we can require $x \in A^j_n$, for any $n \in \mathbb{N}$ and $j = 1, 2$. Then $[x]_{\mathcal{A}_2} = \cap_n (A^1_n \cap A^2_n) \subseteq \cap_n A^1_n = [x]_{\mathcal{A}_1}$.

The last part of the proof is a consequence of Lemma A.3 and of the first point, since

$$f^{-1}(\{f(x)\}) = [x]_{\pi_f} \subseteq [x]_{\pi_g} = g^{-1}(\{g(x)\}), \tag{A.9}$$

or, equivalently, $f(x) = f(y) \Rightarrow g(x) = g(y)$ which is the thesis. $\square$

Proof of Theorem 5.1. Note that $X/\pi_\infty \subseteq \mathcal{X}$ is countable and generated by $\cup_n X/\pi_n$. Then π_∞ is a measurable equivalency by Lemma A.3.

Conversely, we can use the standard approximation technique: if $\pi = \pi_f$ is measurable, let $f_n = 2^{-n}\lfloor 2^n f \rfloor$ for any n. Since f_n are discrete random variables, π_n are defined through Lemma 3.1. By Lemma 4.1 and (3.3), the thesis $\pi_f = \cap_n \pi_n$ will be a consequence of the fact that $\sigma(f) = \vee_n \sigma(f_n)$.

$\sigma(f_n) \subseteq \sigma(f)$ by definition, which implies $\sigma(f_1, f_2, \ldots) \subseteq \sigma(f)$. Finally, as $f_n \to f$, we have $\sigma(f) \subseteq \sigma(f_1, f_2, \ldots)$, which completes the proof. $\square$

B. Proof of Theorem 4.6

Before proving the theorem, we state the following lemma.

Lemma B.1. *Let $(X, \mathcal{X})$ be a measurable space.*

(1) *If any equivalence relationship π on X is measurable, then $\mathcal{X} = 2^X$ and card$(X) \leq$ card$(\mathbb{R})$.*

(2) *The converse is true under the axiom of choice.*

Proof. (1) $\Rightarrow$ (2). Let π_I be the identity relation: $x\pi_I y \Leftrightarrow x = y$. By hypothesis, there exists f such that $\pi_I = \pi_f$, and thus f is injective. Then $\text{card}(X) \leq \text{card}(\mathbb{R})$. Now, take $A \subsetneq X$ and let π_A be the relation so defined:

$$x\pi_A y \Longleftrightarrow \{x, y\} \subseteq A \quad \text{or} \quad \{x, y\} \subseteq X \setminus A. \tag{B.1}$$

Since any equivalency is measurable, then there exists $f : (X, \mathcal{X}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that $\pi_A = \pi_f$. But $\sigma(f) = \sigma(A)$, which shows that $A \subsetneq X \Rightarrow A \in \mathcal{X}$, that is, $\mathcal{X} = 2^X$.

(2) $\Rightarrow$ (1). Since $\text{card}(X) \leq \text{card}(\mathbb{R})$, there exists an injective function $h : X \to \mathbb{R}$. Let π be an equivalence relationship on X, and define the following equivalence on $\mathbb{R}$:

$$r_1 R r_2 \Longleftrightarrow \left(\{r_1, r_2\} \subseteq h(X),\ h^{-1}(r_1)\pi h^{-1}(r_2)\right) \quad \text{or} \quad \{r_1, r_2\} \subseteq \mathbb{R} \setminus h(X). \tag{B.2}$$

By definition of R, if we denote by π_R the canonical projection of $\mathbb{R}$ on $\mathbb{R}/R$, then $\pi_R \circ h : X \to \mathbb{R}/R$ is such that

$$\pi_R \circ h(x) = \pi_R \circ h(y) \Longleftrightarrow x\pi y. \tag{B.3}$$

The axiom of choice ensures the existence of a injective map $g : \mathbb{R}/R \to \mathbb{R}$. Then $f := g \circ \pi_R \circ h : X \to \mathbb{R}$ is such that $\pi = \pi_f$. f is measurable since $\mathcal{X} = 2^X$. □

Proof of Theorem 4.6. (1) $\Rightarrow$ (2). By Lemma B.1 and assumption (A0), $(X, 2^X)$ is weakly Blackwell.

(2) $\Rightarrow$ (3). Assume X is uncountable. By CH, exists $Y \subseteq X$ s.t. $Y \overset{g_1}{\leftrightarrow} \mathbb{R}$ (i.e., Y is in bijection with $\mathbb{R}$ via g_1). Take a bijection $\mathbb{R} \overset{g_2}{\leftrightarrow} \mathbb{R} \setminus \{0\}$. Then the map

$$g(x) = \begin{cases} g_2(g_1(x)) & \text{if } x \in Y, \\ 0 & \text{if } x \in X \setminus Y, \end{cases} \tag{B.4}$$

is a bijective map from $\{Y, \{X \setminus Y\}\}$ to $\mathbb{R}$. Equip $\mathbb{R}$ with the Borel σ-algebra $\mathcal{B}_{\mathbb{R}}$ and let $\mathcal{A}_1 = g^{-1}(\mathcal{B}_{\mathbb{R}})$. $\mathcal{A}_1$ is countably generated and its atoms are all the points in Y and the set $X \setminus Y$. Now, take a nonBorel set N of the real line. $\mathcal{A}_2 = g^{-1}(\sigma(\mathcal{B}_{\mathbb{R}}, N))$ is also countably generated, $\mathcal{A}_1 \subsetneq \mathcal{A}_2$ and its atoms are all the points in Y and the set $X \setminus Y$, too. Since $\mathcal{A}_1 \subseteq 2^X$ and $\mathcal{A}_2 \subseteq 2^X$, $(X, 2^X)$ is not a weakly Blackwell space by Lemma 4.3.

(3) $\Rightarrow$ (1). Since X is countable, then X/π is. Therefore, Lemma 3.1 ensures that any equivalence π is measurable, since $\mathcal{X} = 2^X$. Finally, just note that each countable set is strongly Blackwell. And thus Lemma 4.3 concludes the proof. □

Acknowledgments

The authors are very grateful to the anonymous referees whose incisive comments improved the paper. This work was made while the first author was hosted in Jerusalem. He wishes to thank for the warm hospitality.

References

[1] G. Aletti, "Laplace transformation and weak convergence with an application to fluorescence resonance energy transfer (FRET)," *Applied Mathematics Letters*, vol. 19, no. 10, pp. 1057–1061, 2006.

[2] G Aletti, "The behavior of a Markov network with respect to an absorbing class: the target algorithm," *RAIRO Operations Research*, vol. 43, no. 3, pp. 231–245, 2009.

[3] G Aletti and E Merzbach, "Stopping Markov processes and first path on graphs," *Journal of the European Mathematical Society*, vol. 8, no. 1, pp. 49–75, 2006.

[4] J. Dubra and F. Echenique, "Information is not about measurability," *Mathematical Social Sciences*, vol. 47, no. 2, pp. 177–185, 2004.

[5] M. Jerrum, "On the approximation of one Markov chain by another," *Probability Theory and Related Fields*, vol. 135, no. 1, pp. 1–14, 2006.

[6] P. Billingsley, *Probability and Measure*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, New York, NY, USA, 3rd edition, 1995.

[7] D. Blackwell, "On a class of probability spaces," in *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability, 1954-1955, vol. II*, pp. 1–6, University of California Press, Berkeley, Calif, USA.

[8] J. Jasiński, "On the Blackwell property of Luzin sets," *Proceedings of the American Mathematical Society*, vol. 95, no. 2, pp. 303–306, 1985.

[9] A. Maitra, "Coanalytic sets that are not Blackwell spaces," *Fundamenta Mathematicae*, vol. 67, pp. 251–254, 1970.

[10] M. Orkin, "A Blackwell space which is not analytic," *Bulletin de l'Académie Polonaise des Sciences. Série des Sciences Mathématiques, Astronomiques et Physiques*, vol. 20, pp. 437–438, 1972.

[11] R. M. Shortt, "Sets with no uncountable Blackwell subsets," *Czechoslovak Mathematical Journal*, vol. 37(112), no. 2, pp. 320–322, 1987.

[12] J. Jasiński, "On the combinatorial properties of Blackwell spaces," *Proceedings of the American Mathematical Society*, vol. 93, no. 4, pp. 657–660, 1985.

[13] R. M. Shortt, "Combinatorial properties for Blackwell sets," *Proceedings of the American Mathematical Society*, vol. 101, no. 4, pp. 738–742, 1987.

[14] M. B. Stinchcombe, "Bayesian information topologies," *Journal of Mathematical Economics*, vol. 19, no. 3, pp. 233–253, 1990.

[15] P. Berti, L. Pratelli, and P. Rigo, "Almost sure weak convergence of random probability measures," *Stochastics*, vol. 78, no. 2, pp. 91–97, 2006.