数理解析研究所講究録 1215 巻 2001 年 33-52

WHAT DIFFERENTIATES STATIONARY STOCHASTIC PROCESSES FROM ERGODIC ONES: A SURVEY

MICHEL VALADIER

ABSTRACT. Stationarity of a stochastic process seems connected to the idea of constancy. But ergodicity is needed for the property that almost surely the observation of a trajectory from time $-\infty$ to 0 makes possible the identification of the law of the whole process, including the future. When the stationary process is a Markov chain with a finite number of states it is well known that the set of states divides in ergodic classes. Decomposition of more general stationary processes in ergodic classes goes back to von Neumann. This seems a hidden result rarely developed in text books. After some preliminaries we will expose Choquet way and the Kryloff-Bogoliouboff way which was made a bit more precise by Oxtoby and greatly generalized by Dynkin.

1. Introduction

Very often in contemporary papers the authors assume that a stochastic process is ergodic because under the weaker hypothesis of stationarity their proofs no longer hold. One aim of this paper is to show that this difficulty may be immaterial.

A specially interesting problem is prediction. Stationarity (of a stochastic process — for a precise definition see Section 2) seems connected to the idea of constancy. But ergodicity is needed for the property that (almost surely of course) the observation of a trajectory from time $-\infty$ to 0 makes possible the identification of the law of the whole process, including the future: this is made precise in Theorem 2 and the consequence after.

When the stationary process is a Markov chain with a finite number of states the classical theory, already in [Do, Chapter V], shows that the set of states divides

1991 *Mathematics Subject Classification.* 60–02, 60 G 10, 60 G 25.
Key words and phrases. stationarity, ergodicity, Birkhoff's theorem, prediction.

in *ergodic classes*. And when one observes a trajectory, this trajectory lives in one ergodic class and it amounts (if only one observation is done which is the only possibility if the time is the one of our life) to the same thing as if the process was ergodic. Decomposition of more general stationary processes in ergodic classes goes back to von Neumann [N1]. This seems a hidden result rarely developed in text books: Denker, Grillenberger and Sigmund's book [DGS] is an exception but it does not emphasize the consequences. All amounts to the following: there is a probability law which I call it contingency law, λ, on the set of ergodic laws; firstly an ergodic law Q is chosen according to λ and then the trajectory is chosen according to Q.

After some preliminaries we will give the main result. It was proved by different methods. The elegant Choquet way will be summarized quickly. The Kryloff-Bogoliouboff way which was made a bit more precise by Oxtoby and greatly generalized by Dynkin will be more detailed. For some other works see Maharam [Mah], Varadarajan [Var] and for a generalization to capacities Talagrand [T].

I have discovered the existence of Chersi's paper [Che] in May 2000. I don't know more than the title and the very short Math. Reviews analysis. Surely it is also on the subject.

The author thanks Gérard MICHAILLE and Charles CASTAING for their helpful comments.

2. Stationary and ergodic processes

Let $(K, \mathcal{K})$ be a *Borel standard* measurable space, that is a measurable space isomorphic to a Borel subset of a Polish topological space (for the Choquet point of view, K will be compact metrizable). A *stochastic process* with discrete time taking its values in K is a bilateral sequence $(X_n)_{n\in\mathbf{Z}}$ of random variables (in short r.v.) defined on a probability space $(\Xi, \mathcal{S}, \Pi)$ which take their values in K. The set $\Omega = K^{\mathbf{Z}}$ is more fundamental than Ξ. It is still Borel standard and is compact metrizable when K is compact. Let $\mathcal{F} = \mathcal{K}^{\otimes \mathbf{Z}}$; when K is compact and $\mathcal{K} = \mathcal{B}(K)$, $\mathcal{F} = \mathcal{B}(\Omega)$, that is the product of the Borel tribes coincides with the Borel tribe of the product topology. The *law* of the process, always denoted by P in this paper, is the probability measure on $(\Omega, \mathcal{F})$ image of Π by $\xi \mapsto (X_n(\xi))_{n\in\mathbf{Z}}$. The "canonical" process $(\widetilde{X}_n)_{n\in\mathbf{Z}}$ is defined on Ω by the property that $\widetilde{X}_n$ is the n-th coordinate function. In the sequel we do not use $(\Xi, \mathcal{S}, \Pi)$ and the canonical process is simply denoted by $(X_n)_{n\in\mathbf{Z}}$.

A point $\omega = (x_n)_{n\in\mathbf{Z}} \in \Omega$ is a *trajectory*. The bijective map T of Ω in itself defined by $T\big((x_n)_{n\in\mathbf{Z}}\big) = (x_{n+1})_{n\in\mathbf{Z}}$ is the *Bernoulli shift*. It is an homeomorphism when K is compact. The image of P by T is denoted by $T_{\#}(P)$.

Definition. The process $(X_n)_{n\in\mathbf{Z}}$ is *stationary* if its law is *invariant* i.e. for any $A \in \mathcal{B}(\Omega)$, $P(T^{-1}A) = P(A)$ (that is $T_{\#}(P) = P$). We write $P \in \mathcal{L}_{\mathrm{st}}$.

Definitions. Let $(X_n)_{n\in\mathbf{Z}}$ be a stationary process. The *invariant* events are the $A \in \mathcal{F}$ satisfying $T^{-1}A = A$ (or $TA = A$). The set they constitute is a tribe denoted by $\mathcal{I}$. The process of law P is *ergodic*[1] if $\mathcal{I}$ is coarse (up to P-negligible sets), that is if $A \in \mathcal{I} \Rightarrow P(A) = 0$ or 1. One also says that P is an *ergodic law*. We write $P \in \mathcal{L}_{\mathrm{erg}}$.

In ergodic theorems, T^j denotes, when $j \in \mathbb{Z}_+^* = \mathbb{N}^*$, the j-th power of T :

[1] Doob [Do, p.457], and several authors, say "metrically transitive".

$T^j = T \circ T \circ \cdots \circ T$, $T^0 = \mathrm{id}_\Omega$ and, when $j \in \mathbb{Z}_-^*$, T^j denotes $(T^{-1})^{|j|}$.

The notation δ_x denotes the Dirac measure at x.

3. Identification of the law of an ergodic process from the observation of its past

Proposition 1 is elementary. It will be applied, when K is compact, to $\Lambda = \Omega = K^{\mathbb{Z}}$ as well as to $\Lambda = K^d$. For the Borel standard case see Proposition 5.

Proposition 1. *Let Λ be compact metrizable and D a dense subset*[2] *of $C(\Lambda)$ (usually D will be countable). Let $(P_k)_{k\in\mathbb{N}}$ be a sequence of probability measures on Λ. It weakly converges (i.e. for the weak topology relative to the duality with $C(\Lambda)$) iff $\forall f \in D$, the sequence $(\int_\Lambda f\, dP_k)_{k\in\mathbb{N}}$ converges in $\mathbb{R}$.*

Proof. The "only if" part is obvious. For the converse assume that $\forall f \in D$, the sequence $(\int_\Lambda f\, dP_k)_{k\in\mathbb{N}}$ converges in $\mathbb{R}$. The space $\mathcal{M}_+^1(\Lambda)$ of all probability measures on Λ is weakly compact metrizable. The sequence $(P_k)_{k\in\mathbb{N}}$ has a unique limit point. Indeed if Q_1 and Q_2 are two limit points, one has, for $i = 1$ and 2, $\forall f \in D$, $\int_\Lambda f\, dQ_i = \lim_k \int_\Lambda f\, dP_k$ hence $Q_1 = Q_2$. □

Theorem 2. *Assume that K is compact and that the process $(X_n)_{n\in\mathbb{Z}}$ with values in K is ergodic. Then almost surely, for all $d \in \mathbb{N}^*$, the law $P_{(X_{-d+1},\dots,X_0)}$ of $(X_{-d+1},\dots,X_0)$ is the weak limit of $k^{-1}\sum_{j=0}^{k-1} \delta_{(x_{-(j+d-1)},\dots,x_{-j})}$ as $k \to +\infty$.*

Consequence. Hence P-almost surely, knowing $(x_n)_{n\le 0}$ implies the knowledge of $P_{(X_{-d+1},\dots,X_0)}$ hence of $P_{(X_p,X_{p+1},\dots,X_{q-1},X_q)}$ (this writing supposes $(p,q) \in \mathbb{Z}^2$ and $p \le q$) since stationarity implies $P_{(X_p,\dots,X_q)} = P_{(X_{p-q},\dots,X_0)}$. Recall now that P

[2] It is sufficient that the linear subspace of $C(\Lambda)$ spanned by D is dense.

is the projective limit of the measures[3] $P_{(X_p,\dots,X_q)}$. So, mathematically, P can be identified; from a numerical point of view, this is another story: see all the concepts defined and studied in Statistical Theory.

Proof. Let $d \in \mathbb{N}^*$ and D_d be a countable dense subset of $\mathcal{C}(K^d)$. For any $f_0 \in D_d$ let f denote the function on Ω associated to f_0 which is defined by: $f\big((x_n)_{n\in\mathbb{Z}}\big) = f_0(x_{-d+1},\dots,x_0)$. By Birkhoff's theorem if $(x_n)_{n\in\mathbb{Z}}$ does not belong to a P-negligible set N_{f_0}:

$$\begin{aligned}\int_\Omega f_0\, d\Big[k^{-1}\sum_{j=0}^{k-1}\delta_{(x_{-(j+d-1)},\dots,x_{-j})}\Big] &= k^{-1}\sum_{j=0}^{k-1} f_0(x_{-(j+d-1)},\dots,x_{-j})\\ &= k^{-1}\sum_{j=0}^{k-1} f\big[T^{-j}\big((x_n)_n\big)\big]\\ &\xrightarrow{(k\to+\infty)} \int_\Omega f\, dP\\ &= \int_{K^d} f_0\, dP_{(X_{-d+1},\dots,X_0)}\,.\end{aligned}$$

By Proposition 1 this proves the convergence of $k^{-1}\sum_{j=0}^{k-1}\delta_{(x_{-(j+d-1)},\dots,x_{-j})}$ to $P_{(X_{-d+1},\dots,X_0)}$ if $(x_n)_{n\in\mathbb{Z}} \notin \bigcup_{f_0\in D_d} N_{f_0}$. So the statement holds for $(x_n)_{n\in\mathbb{Z}}$ not in $\bigcup_{d\in\mathbb{N}^*}\big[\bigcup_{f_0\in D_d} N_{f_0}\big]$. $\square$

Comment. The meaning of Theorem 2 is that almost surely the mere observation of the past (from $-\infty$) of an ergodic process allows to identify the law of the whole process (including the future). From a theoretical point of view this is a perfect situation for *prediction*. Indeed when Ω is written $\Omega = K^{\mathbb{Z}_-}\times K^{\mathbb{N}^*}$ and $\omega = (\xi,\zeta)$ that is $\xi = (x_n)_{n\le 0} = \omega_{|\mathbb{Z}_-}$ and $\zeta = (x_n)_{n>0} = \omega_{|\mathbb{N}^*}$, there is a disintegration (see next Section) of P unique up to equality a.e. which is a family $(L_\xi)_{\xi\in K^{\mathbb{Z}_-}}$ of

[3] More simply when the $P_{(X_p,\dots,X_q)}$ are known, the values of P on the algebra of cylindrical sets are known, and this algebra generates $\mathcal{B}(\Omega)$.

probability laws on $K^{\mathbf{N}^*}$. Then knowing $\xi = \omega_{|\mathbf{Z}_-} = (x_n)_{n\leq 0}$, the future obeys to the conditional law L_ξ on $K^{\mathbf{N}^*}$.

4. Basic ideas of disintegration

When a sub-tribe $\mathcal{G}$ of $\mathcal{F}$ is given, there exists under very general topological hypotheses concerning Ω, a disintegration with respect to $\mathcal{G}$, that is a family of probability measures $(Q^\omega)_{\omega\in\Omega}$ on $(\Omega,\mathcal{F})$ which is $\mathcal{G}$-measurable in ω and which satisfies

$$\forall B \in \mathcal{F},\ \forall A \in \mathcal{G}, \quad P(A\cap B) = \int_A Q^\omega(B)\,dP(\omega).$$

Let us consider, as it always should be, the conditional expectation $\mathbf{E}^{\mathcal{G}}(1_B)$ as a class of random variables up to equality a.s. The functions $\omega \mapsto Q^\omega(B)$ (B running through $\mathcal{F}$) constitute a "consistent"[4] family of versions of the $\mathbf{E}^{\mathcal{G}}(1_B)$.

This has a long story in probability theory: von Neumann [N1], Kolmogorov [Ko], Jirina [Ji], Hoffmann-Jørgensen [HJ], Valadier [V1–2] for some details[5]. (For textbooks see Bauer [Ba], Dudley [Du].) But disintegration is unduly considered as a hard concept reserved to experts and, in my opinion, too rarely used.

Classically for any real integrable r.v. Y (see for example Dudley [Du, 10.2.5 p.272], Doob [Do, Th.9.1 p.27], Kolmogorov [Ko, ch.V (12) and (14)]) :

$$\int_\Omega Y(\omega')\,dQ^\omega(\omega') \overset{\text{a.s.}}{=} (\mathbf{E}^{\mathcal{G}}Y)(\omega).$$

An important particular case is the following. Suppose Ω is a product[6] $\Omega_1\times\Omega_2$, $\omega = (\xi,\zeta)$ and $\mathcal{G}$ is generated by the projection on Ω_1 (possibly ξ is the past, ζ is

[4] The problem if one chose anyhow versions of $\mathbf{E}^{\mathcal{G}}(1_B)$ would lie in the σ-additivity with respect to B. A classical expression for disintegration is *regular conditional probabilities.*

[5] At the time when in France only Bourbaki and Jirina were quoted, I wrote [V1] where I gave a result of Hoffmann-Jørgensen [HJ] in the framework of a product and where I compared several statements of this time. In [V2, p.13] I had the idea, being not aware of [Hl1], of introducing the quotient tribe.

[6] To be more precise $(\Omega_1,\mathcal{F}_1)$ is separated and countably generated and Ω_2 is a "good" topological space, that is Polish or Suslin...

the future). Then Q^ω depends only on ξ and has the form $\delta_\xi \otimes L_\xi$ where δ_ξ is the Dirac mass at ξ and L_ξ is the *conditional law of ζ given ξ*.

5. Decomposition of a stationary process. The contingency law

In the following P always denote a probability measure on $\Omega = K^{\mathbf{Z}}$ and we will say equivalently that P is invariant or stationary. This refers to the stationarity of the "canonical" process defined in Section 2. And P is said ergodic if the process is ergodic. Although the decomposition theorem admits several non trivial proofs and some variants in its formulation, it roughly says at least the following:

Theorem 3. *Any stationary law P on Ω is a mixing of ergodic laws.*

Comments. 1) All amounts to the following: there is a probability law which I call *contingency law*, λ, on the set of ergodic laws; firstly an ergodic law Q is chosen according to λ and then the trajectory is chosen according to Q. So, if only one observation is done, one observes a trajectory of an ergodic process. And in my opinion, the prediction of stationary processes[7] is not a problem different from the prediction of ergodic ones.

2) For example imagine the set of meteorological phenomena appearing during one year is the value of a stationary process with time in $\mathbb{Z}$, and imagine that this process has been always observed. Then it could be treated as an ergodic process: two moons or another rotational velocity of the planet Earth could have occurred if the world has been created differently. This is contingency.

3) Any probability is a mixing of Dirac measures: if $P \in \mathcal{M}_+^1(\mathbb{R})$ it is the mixing

[7] On the subject of prediction of stationary processes there is a very ambitious book: Fursten-

of the measures δ_r according to the image $\overline{P}$ of P on $\mathcal{M}^1_+(\mathbb{R})$ by $r \mapsto \delta_r$. This has not any interest. In the interpretation of Theorem 3 the roles of time, past and future are essential.

Theorem 3 was originally proved by von Neumann in 1932 [N1]. It received several interesting proofs: the one of Choquet [Cho1, 1956/1957], the method of Kryloff-Bogoliouboff [KB, 1937]; see also Maharam [Mah, 1950], Farrell [Fa, 1962], and specially Dynkin [Dy, 1978].

5.1 Choquet's way.

Assume K is compact metrizable. Let $\mathcal{M}^1_+(\Omega)$ denote the set of probability measures on $(\Omega, \mathcal{B}(\Omega))$ endowed with the weak topology.

Theorem 4. *Assume K is compact metrizable.* 1) *The set $\mathcal{L}_{\mathrm{st}}$ of invariant probabilities on Ω is a non empty convex compact subset of $\mathcal{M}^1_+(\Omega)$.* 2) *The set $\partial\mathcal{L}_{\mathrm{st}}$ of extreme points of $\mathcal{L}_{\mathrm{st}}$ coincide with the set of ergodic laws $\mathcal{L}_{\mathrm{erg}}$.* 3) *Let $P \in \mathcal{L}_{\mathrm{st}}$. There exists a probability measure λ on $\partial\mathcal{L}_{\mathrm{st}} = \mathcal{L}_{\mathrm{erg}}$ such that*

$$P = \int_{\mathcal{L}_{\mathrm{erg}}} Q \, d\lambda(Q) \,. \tag{1}$$

(In (1) the right-hand side is a weak integral of measures whose meaning is as well

$$\forall \varphi \in \mathcal{C}(\Omega), \quad \int_\Omega \varphi \, dP = \int_{\mathcal{L}_{\mathrm{erg}}} \Big[\int_\Omega \varphi \, dQ \Big] \, d\lambda(Q)$$

as

$$\forall B \in \mathcal{B}(\Omega), \quad P(B) = \int_{\mathcal{L}_{\mathrm{erg}}} Q(B) \, d\lambda(Q) \,.)$$

Some ideas of the proof. The first assertion has an easy proof. The second is well known of specialists and a rather old result: see Blum-Hanson [BH, 1960] and Choquet knew it before; I recommend the proof of Denker-Grillenberger-Sigmund [DGS, (5.6) p.24]. Then the conclusion follows from the Choquet integral representation theorem (besides the quoted works of Choquet one can see [Bo, IV.7.2 Th.1 p.219] and Phelps [Ph]).

Remarks. One year before Choquet, Hewitt-Savage [HS] used the same argument with the set of laws on $\mathbb{R}^I$ invariant by permutation of coordinates whose extreme points are the laws of families of i.i.d. random variables (they continued a famous work of de Finetti [Fi]). But in [HS] the set of extreme points is closed which makes the integral representation elementary while here $\mathcal{L}_{\text{erg}}$ is not closed (think of stationary Markov chains with matrices $\begin{pmatrix} 1-\frac{1}{n} & \frac{1}{n} \\ \frac{1}{n} & 1-\frac{1}{n} \end{pmatrix}$, for $n \in \mathbb{N}^*$, $n \to +\infty$).

5.2 Kryloff-Bogoliouboff's way (generalized by Dynkin).

Let us summarize the main result of Kryloff-Bogoliouboff as it was generalized by Dynkin [Dy].

Proposition 5. *Let Ω be Borel standard. There exists a countable set W of bounded measurable functions satisfying:*

BS1. *Let $(P_k)_{k\in\mathbb{N}}$ be a sequence of probability measures on Ω such that $\forall f \in W$, the sequence $(\int_\Omega f\, dP_k)_{k\in\mathbb{N}}$ converges in $\mathbb{R}$. Then there exists a probability P such that $\forall f \in W$, $\int_\Omega f\, dP_k \to \int_\Omega f\, dP$.*

BS2. *If $\mathcal{H}$ is a linear space of real functions on Ω containing W and stable with respect to the "bounded pointwise convergence[8]", then $\mathcal{H}$ contains all bounded mea-*

[8] That is: if $f_n \in \mathcal{H}$, if the sequence $(f_n)_n$ is uniformly bounded and converges pointwisely to

surable functions on Ω.

Thanks to **BS2** there is a unique P satisfying **BS1**.

Remark. A short time after Dynkin's paper [Dy, 1978], Ramakrishnan and Rao proved in [RR, 1980] that the notion of B-space introduced by Dynkin coincides with the notion of Borel standard.

Proof. See Dynkin [Dy] Section 4.2 page 714. □

Notations. Let $\widetilde{\mathcal{F}}$ denote the set of all real bounded $\mathcal{F}$-measurable functions on Ω. Let $\widehat{\mathcal{F}}$ denote the intersection

$$\widehat{\mathcal{F}} = \bigcap_{P \in \mathcal{L}_{st}} \widehat{\mathcal{F}}_P$$

of the P-completions of $\mathcal{F}$.

Theorem 6. (1) *Let Q_n^ω denote $\frac{1}{n}\sum_{i=0}^{n-1} \delta_{T^i\omega}$ and W given by Proposition 5. The two following subsets of Ω, Ω' and Ω_{erg} defined by:*

$$\Omega' := \{\omega \in \Omega : \forall f \in W, \left(\int_\Omega f\, dQ_n^\omega\right)_n \text{ converges in } \mathbb{R}\}$$

and, if Q^ω denotes the probability — whose existence follows from **BS1** *— satisfying $\forall f \in W$, $\int_\Omega f\, dQ_n^\omega \to \int_\Omega f\, dQ^\omega$,*

$$\Omega_{\mathrm{erg}} := \{\omega \in \Omega' : Q^\omega \text{ is an ergodic law}\}$$

belong to $\widehat{\mathcal{F}}$ and have P-measure 1 for any invariant probability P. Moreover $\forall P \in \mathcal{L}_{\mathrm{st}}$, $\forall f \in \widetilde{\mathcal{F}}$, $\int_\Omega f\, dQ_n^\omega \xrightarrow{P\text{-a.s.}} \int_\Omega f\, dQ^\omega$.

---

f, then $f \in \mathcal{H}$.

(2) *The family* $(Q^\omega)_{\omega\in\Omega_{\mathrm{erg}}}$ *disintegrates any invariant probability* P *relatively to* $\mathcal{I}$, *which means: for any* $B \in \mathcal{F}$, $\big(\mathbf{E}_P^{\mathcal{I}}(1_B)\big)(\omega) \overset{P\text{-a.s.}}{=} Q^\omega(B)$ *or*

$$\forall B \in \mathcal{F},\ \forall A \in \mathcal{I}, \quad P(A\cap B) = \int_{A\cap\Omega_{\mathrm{erg}}} Q^\omega(B)\, dP(\omega) \tag{2}$$

as well as, for any $Y \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ *(or* $Y \geq 0$ *and* $\mathcal{F}$*-measurable),*

$$(\mathbf{E}_P^{\mathcal{I}} Y)(\omega) \overset{P\text{-a.s.}}{=} \int_\Omega Y\, dQ^\omega\,.$$

Remarks. 1) When K is compact metrizable as in [KB], [Ox] and [DGS], one gets on a set denoted by Ω_{qr}, the weak convergence $Q_n^\omega \to Q^\omega$ for the $\sigma(\mathcal{M}^b(\Omega), \mathcal{C}(\Omega))$ topology. There are less P-negligible sets and equalities P-almost everywhere than here (in the [Dy] framework). And here one gets only the following two weak convergences: (i) on the set Ω', $\forall f \in W$, $\int_\Omega f\, dQ_n^\omega \to \int_\Omega f\, dQ^\omega$;

(ii) $\forall P \in \mathcal{L}_{\mathrm{st}}$, $\forall f \in \widetilde{\mathcal{F}}$, $\int_\Omega f\, dQ_n^\omega \xrightarrow{P\text{-a.s.}} \int_\Omega f\, dQ^\omega$, that is $\forall f \in \widetilde{\mathcal{F}}$, $\forall P \in \mathcal{L}_{\mathrm{st}}$, $P(\{\omega \in \Omega' : \int_\Omega f\, dQ_n^\omega \to \int_\Omega f\, dQ^\omega\}) = 1$.

2) In the Dynkin framework Ω is not necessarily a product and T is not assumed to be bijective. When $\Omega = K^{\otimes\mathbb{Z}}$ and T is the Bernoulli shift, as in [KB], [Ox] and [DGS], Q^ω can be treated as a bilateral limit as written in the following formula

$$Q^\omega := \lim_{\substack{|n|\to+\infty \\ n\in\mathbb{Z}^*}} \Bigg[|n|^{-1} \sum_{\substack{|j|\leq|n|-1 \\ \operatorname{sgn} j = \operatorname{sgn} n}} \delta_{T^j(\omega)} \Bigg].$$

Proof. All arguments are those of [Dy] with some simplifications possible thanks to the framework and some more explanations when I feel them useful. Part (A) comes from Lemmas 4.1 and 6.1 of [Dy], (B) comes from Lemma 3.1 and Theorem 3.4 and (C) is a part of Theorem 3.1.

(A) It is easy to prove $\Omega' \in \mathcal{F}$. By Birkhoff's theorem

$$\forall P \in \mathcal{L}_{\mathrm{st}},\ \forall f \in \widetilde{\mathcal{F}},\quad \int_\Omega f\, dQ_n^\omega \xrightarrow{P\text{-a.s.}} (\mathbf{E}_P^{\mathcal{I}} f)(\omega)\,. \tag{3}$$

Since W is countable, (3) implies $\forall P \in \mathcal{L}_{\mathrm{st}}$, $P(\Omega') = 1$. By the definition of Q^ω and (3), $\forall P \in \mathcal{L}_{\mathrm{st}}$, $\forall f \in W$,

$$(\mathbf{E}_P^{\mathcal{I}} f)(\omega) \overset{P\text{-a.s.}}{=} \int_\Omega f\, dQ^\omega\,. \tag{4}$$

We postpone the discussion of the measurability properties of $\omega \mapsto Q^\omega$ to the end of Part (A). Now let for $P \in \mathcal{L}_{\mathrm{st}}$,

$$\mathcal{H}_P := \{f \in \widetilde{\mathcal{F}} : (\mathbf{E}_P^{\mathcal{I}} f)(\omega) \overset{P\text{-a.s.}}{=} \int_\Omega f\, dQ^\omega\}\,.$$

Property **BS2** applies to $\mathcal{H}_P$ so $\mathcal{H}_P = \widetilde{\mathcal{F}}$. As consequences, (4) holds firstly if $f \in \widetilde{\mathcal{F}}$; then (4) still holds if f is $\mathcal{F}$-measurable and $[0, +\infty]$-valued and it holds too for $f \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ (use $f = f^+ - f^-$). This proves Part 2 of the statement (with letter f in place of Y). Now, again by Birkhoff's theorem, for $f \in \widetilde{\mathcal{F}}$ and $P \in \mathcal{L}_{\mathrm{st}}$,

$$P(\{\omega \in \Omega' : \int_\Omega f\, dQ_n^\omega \to \int_\Omega f\, dQ^\omega\}) = 1\,.$$

Now we discuss the measurability properties of $\omega \mapsto Q^\omega$. Firstly for any $f \in \widetilde{\mathcal{F}}$, $\omega \mapsto \int_\Omega f\, dQ^\omega$ is $\mathcal{F}_{|\Omega'}$-measurable because it is the limit of a sequence of measurable functions. Thanks to **BS2**

$$\mathcal{H} := \{f \in \widetilde{\mathcal{F}} : \omega \mapsto \int_\Omega f\, dQ^\omega \text{ is } \mathcal{F}_{|\Omega'}\text{-measurable}\}$$

equals $\widetilde{\mathcal{F}}$. Then (4) extended to $\widetilde{\mathcal{F}}$ implies: for any $B \in \mathcal{F}$, $\omega \mapsto Q^\omega$ is $\widehat{\mathcal{I}}$-measurable, hence there exists $\varphi_B \colon \Omega \to \mathbb{R}$ which is $\mathcal{I}$-measurable and satisfies $\forall P \in \mathcal{L}_{\mathrm{st}}$,

$P(N_B) = 0$ where $N_B := \{\omega \in \Omega' : Q^\omega(B) \neq \varphi_B(\omega)\}$. Let $\mathcal{A}$ be a countable algebra which generates $\mathcal{F}$ and $N := \bigcup_{B \in \mathcal{A}} N_B$. Then $\forall P \in \mathcal{L}_{\mathrm{st}}$, $P(N) = 0$ and $\omega \mapsto Q^\omega(B)$ is $\mathcal{I}$-measurable on $\Omega' \backslash N$, firstly for $B \in \mathcal{A}$, and then, thanks to the monotone class lemma, for $B \in \mathcal{F}$.

(B) Now we prove that for any $P \in \mathcal{L}_{\mathrm{st}}$, for P-almost every ω in Ω', Q^ω is T-invariant which writes also $Q^\omega \in \mathcal{L}_{\mathrm{st}}$. Firstly let us prove

$$\forall P \in \mathcal{L}_{\mathrm{st}},\ \forall f \in \widetilde{\mathcal{F}},\quad \mathbf{E}_P^{\mathcal{I}}(f) \overset{P\text{-a.s.}}{=} \mathbf{E}_P^{\mathcal{I}}(f \circ T). \tag{5}$$

Indeed, for any $f \in \widetilde{\mathcal{F}}$ and any $A \in \mathcal{I}$,

$$\int_\Omega 1_A (f \circ T)\, dP = \int_\Omega (1_A \circ T)(f \circ T)\, dP = \int_\Omega 1_A f\, d[T_\# P] \overset{*}{=} \int_\Omega 1_A f\, dP.$$

Formula (5) can be written

$$\int_\Omega f\, dQ^\omega \overset{P\text{-a.s.}}{=} \int_\Omega (f \circ T)\, dQ^\omega. \tag{6}$$

Now let $\mathcal{A}$ be a countable algebra which generates $\mathcal{F}$. Taking, for any $B \in \mathcal{A}$, $f = 1_B$, one gets from (6) $Q^\omega(B) \overset{P\text{-a.s.}}{=} Q^\omega(T^{-1}B)$. So there exists a P-negligible set N such that $\forall \omega \in \Omega' \backslash N$, one has $[\forall B \in \mathcal{A},\ Q^\omega(B) = Q^\omega(T^{-1}B)]$. Hence $P(\{\omega \in \Omega' : Q^\omega \in \mathcal{L}_{\mathrm{st}}\}) = 1$.

(C) Now we have to prove that $\Omega_{\mathrm{erg}} := \{\omega \in \Omega' : Q^\omega \text{ is ergodic}\}$ belongs to $\widehat{\mathcal{F}}$ and has P-measure 1 for any $P \in \mathcal{L}_{\mathrm{st}}$. Let for $B \in \mathcal{F}$ and $\nu \in \mathcal{L}_{\mathrm{st}}$,

$$f_B(\nu) := \int_{\Omega'} [Q^\omega(B) - \nu(B)]^2\, d\nu(\omega).$$

Since $\nu(B)$ is the ν-mean of $\omega \mapsto Q^\omega(B)$,

$$f_B(\nu) := \int_{\Omega'} Q^\omega(B)^2\, d\nu(\omega) - \nu(B)^2.$$

Now let us consider $Y_B(\omega) := Q^\omega(B)^2$ which is $\mathcal{I}$-measurable (more precisely $\widehat{\mathcal{I}}$-measurable) in ω. Then, for $P \in \mathcal{L}_{\mathrm{st}}$,

$$f_B(Q^\omega) = \int_\Omega Y_B\, dQ^\omega - Y_B(\omega) \overset{P\text{-a.s.}}{=} [\mathbf{E}_P^{\mathcal{I}}(Y_B)](\omega) - Y_B(\omega) \overset{P\text{-a.s.}}{=} 0\,.$$

So there exists N a P-negligible set such that $\forall B \in \mathcal{A}$, $\forall \omega \in \Omega' \backslash N$, $f_B(Q^\omega) = 0$.

Now let us prove that if $\nu \in \mathcal{L}_{\mathrm{st}}$, [$\nu$ is ergodic] $\Longleftrightarrow$ [$\forall B \in \mathcal{A}$, $f_B(\nu) = 0$]. Firstly [$\forall B \in \mathcal{A}$, $f_B(\nu) = 0$] implies $\nu(\{\omega \in \Omega' : \forall B \in \mathcal{A},\ Q^\omega(B) = \nu(B)\}) = 1$, hence $\nu(\{\omega \in \Omega' : Q^\omega = \nu\}) = 1$. This ensures $\forall B \in \mathcal{F}$, $\nu(\{(\mathbf{E}_\nu^{\mathcal{I}}(1_B))(.) = \nu(B)\}) = 1$ and, taking $B \in \mathcal{I}$, $1_B \overset{\nu\text{-a.s.}}{=} \mathbf{E}_\nu^{\mathcal{I}}(1_B) \overset{\nu\text{-a.s.}}{=} \nu(B)$ hence $\nu(B) = 0$ or 1. This proves the implication $\Longleftarrow$. The converse is easily checked.

The functions f_B are measurable functions with respect to the tribe on $\mathcal{L}_{\mathrm{st}}$ generated by the maps $P \mapsto P(C)$ (C running through $\mathcal{F}$). Hence

$$\mathcal{L}_{\mathrm{erg}} = \bigcap_{B \in \mathcal{A}} \{\nu \in \mathcal{L}_{\mathrm{st}} : f_B(\nu) = 0\}$$

is a measurable subset of $\mathcal{L}_{\mathrm{st}}$ and $\Omega_{\mathrm{erg}} \in \widehat{\mathcal{F}}$. The property $P(\{\omega \in \Omega' : Q^\omega \in \Omega_{\mathrm{erg}}\}) = 1$ follows from the foregoing observations. $\square$

Remarks. 1) There are two equivalence relations: firstly

$$Q^\omega = Q^{\omega'} \tag{R1}$$

which makes sense on Ω' and secondly

$$\forall A \in \mathcal{I}, \quad 1_A(\omega) = 1_A(\omega')\,. \tag{R2}$$

Let Γ_ω denote the class of $\omega \in \Omega'$ for (R1) and let us prove its invariance. For any $f \in W$ changing ω in $T\omega$ or in $T^{-1}\omega$ does not change the Cesàro limit of the sequence $\left(f(T^j\omega)\right)_j$ that is $\lim_n \int f\, dQ_n^\omega$. Hence $Q^{T\omega} = Q^{T^{-1}\omega} = Q^\omega$, so $T\omega$ and $T^{-1}\omega$ both belong to Γ_ω, and since T is bijective, $T\Gamma_\omega = \Gamma_\omega$.

The class of ω for (R2) is $\dot{\omega} = \{T^j\omega : j \in \mathbb{Z}\}$ because this is the smallest Borel invariant set containing ω.

Thanks to the invariance of Γ_ω relation (R2) is always finer than (R1). But in general they do not coincide neither on $\Omega_{\mathbf{erg}}$ nor on $\Omega \backslash N$ where N is any negligible set. Let us give an example: let ϖ be a probability measure on K not reduced to a Dirac mass and $P := \varpi^{\otimes \mathbf{Z}}$. Since the X_n are i.i.d. $Q^\omega \overset{P\text{-a.e.}}{=} P$ (and as well known — [Do] Th.1.2 page 460 — the process is ergodic). So there is a unique class for (R1). For any P-negligible N, $\Omega \backslash N$ has the cardinal of $\mathbb{R}$ (because P is isomorphic to the Lebesgue measure). The existence of only one class for (R2) would lead to a contradiction. Indeed, suppose there is only the class $\dot{\omega} = \{T^j\omega : j \in \mathbb{Z}\}$. For any ω', $\exists j \in \mathbb{Z}$ such that $\omega' = T^j\omega$ and the set $\Omega \backslash N$ of trajectories under consideration would be countable.

The set $\Omega \backslash \Omega_{\mathbf{erg}}$ is not the biggest P-negligible set possible. For a discussion of negligible sets when $K = \{0, 1\}$ in connection with the notion of random numbers see Dellacherie [De].

2) Since (R2) is finer than (R1), $\omega \mapsto Q^\omega$ is constant on each class $\dot{\omega}$; let $\Theta^{\dot{\omega}}$ denotes its value on $\dot{\omega}$ and $\dot{\Omega}$ the set of all classes. Then as a consequence of (2), P is the mixing of the ergodic laws $\Theta^{\dot{\omega}}$ $(\dot{\omega} \in \dot{\Omega})$ according to the image of P on $\dot{\Omega}$ by $\omega \mapsto \dot{\omega}$. Thus (2) looks as (1) of Theorem 4. For historical works which attacked disintegrating P with respect to $\mathcal{I}$ see Halmos [Hl1] and Ambrose-Halmos-Kakutani [AHK].

3) Dynkin proves many other results: specially he gets (Theorem 3.1) that $\mathcal{L}_{\mathrm{st}}$

is a simplex in the Choquet sense whose extreme points are the ergodic measures.

6. Further comments

When one observes only one trajectory of a process which is assumed to be stationary, one can only identify the law Q^ω corresponding to the observed trajectory. For example if one observes a Markov trajectory living in $\{1,2\}$ obeying to the transition matrix $\begin{pmatrix} \frac{1}{4} & \frac{1}{2} \\ \frac{3}{4} & \frac{1}{2} \end{pmatrix}$ then either the whole process is ergodic and obeys to this transition matrix with probabilities of states equal to $\begin{pmatrix} \frac{2}{5} \\ \frac{3}{5} \end{pmatrix}$ or there exists other ergodic classes about nothing is known.

A remark about a small strange phenomenon: suppose one observes $(x_n)_{n\in\mathbf{Z}_-}$ where $x_n = (-1)^n$. One possibility is: there is not any random and this is just a periodic behavior which may continue with $x_n = (-1)^n$ for $n \geq 1$. If we are sure that there is behind a stationary stochastic process then the observed trajectory continues in this way and the trajectory $(y_n)_{n\in\mathbf{Z}} = \left((-1)^{n+1}\right)_{n\in\mathbf{Z}}$ is another (hidden) possibility. So, if the process is ergodic, these two trajectories are the only ones and have probability 1/2. This is the Markov chain with states $\{-1,1\}$, matrix of transitions $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and probabilities of states equal to $\begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$.

7. Prospects

Let us consider manufacturing of concrete. The dimensions and shapes of the stones are random variables with stochastic characteristics which are the same as long as the stones come from the same origin. This origin could change when building a new work. This is contingency. But as long as the origin of stones remains unchanged, all amounts as if the ergodic hypothesis was satisfied. To be more precise the results of Section 5 would have to be extended to the group $\mathbb{R}^N$ in place of $\mathbb{Z}$. Fomin [Fo] gives in Russian some results in this line; see also Remark 6.1 in [Dy, p.717]. More generally the problem of relaxing the ergodic hypothesis into the stationarity one comes up in stochastic homogenization. For a few references see Dal Maso-Modica [DMM1–2], Nguyen-Zessin and Licht-Michaille [LM1–2].

REFERENCES

[AHK] W. Ambrose, P.R. Halmos & S. Kakutani, *The decomposition of measures, II*, Duke Math. J. **9** (1942), 43–47.

[Ba] H. Bauer, *Probability theory and elements of measure theory*, Holt, Rinehart & Winston, New York, 1972 (translation of *Wahrscheinlichkeitstheorie und Grundzüge der Masstheorie*, De Gruyter, Berlin, 1968).

[Bir] G.D. Birkhoff, *Proof of the ergodic theorem*, Proc. Nat. Acad. Sci. U.S.A. **17** (1931), 656–660.

[BH] J.R. Blum & D.L. Hanson, *On invariant probability measures I*, Pacific J. Math. **10** (1960), 1125–1129.

[Bo] N. Bourbaki, *Intégration, Chapitres 1 à 4*, Hermann, Paris, 1965.

[Che] F. Chersi, *An ergodic decomposition of invariant measures for discrete semiflows on standard Borel spaces*, Advanced topics in the theory of dynamical systems. Proceedings of the International Conference held in Trento, June 1–6, 1987 (G. Fusco, M. Iannelli & L. Salvadori, eds.), Notes and Reports in Mathematics in Science and Engineering, vol. 6, 1989, pp. 75–87.

[Cho1] G. Choquet, *Existence et unicité des représentations intégrales au moyen des points extrémaux dans les cônes convexes*, Séminaire Bourbaki, 9ème année (1956/1957), Paris, 1959, pp. 139.1–139.15.

[Cho2] G. Choquet, *Lectures on Analysis, vol.II Representation Theory*, W.A. Benjamin, Inc., New York, 1969.

[Cho3] G. Choquet, *Représentation intégrale*, Measure Theory and its Applications (J.M. Belley, J. Dubois & P. Morales, eds.), Lecture Notes in Mathematics, vol. 1033, Springer Verlag, Berlin, 1983 pp. 114–143.

[DMM1] G. Dal Maso & L. Modica, *Nonlinear stochastic homogenization*, Annali Mat. Pura ed Appl. **144** (1986), 347–389.

[DMM2] G. Dal Maso & L. Modica, *Nonlinear stochastic homogenization and ergodic theory*, J. Reine Angew. Math. **368** (1986), 28–42.

[De] C. Dellacherie, *Nombre au hasard de Borel à Martin-Loef*, Gazette des Mathématiciens (Société Mathématique de France) **11** (1978), 23–58.

[DGS] M. Denker, C. Grillenberger & K. Sigmund, *Ergodic theory on compact spaces*, Lecture Notes in Math., vol. 527, Springer-Verlag, Berlin, 1976.

[Do] J.L. Doob, *Stochastic processes*, Wiley, New York, 1965 (sixth printing. First printing, 1953).

[Du] R.M. Dudley, *Real analysis and probability*, Wadsworth & Brooks/Cole Mathematics Series, California, 1989.

[Dy] E.B. Dynkin, *Sufficient statistics and extreme points*, Ann. Probability **6** (1978), 705–730.

[Fa] R.H. Farrell, *Representation of invariant measures*, Illinois J. Math. **6** (1962), 447–467.

[Fi] B. de Finetti, *La prévision : ses lois logiques, ses sources subjectives*, Ann. Inst. H. Poincaré **7** (1937), 1–68.

[Fo] S.V. Fomin, *On measures invariant under certain groups of transformations*, Izvestiya Akad. Nauk SSSR Ser. Mat. **14** (1950), 261–274.

[Fu] H. Furstenberg, *Stationnary processes and prediction theory*, Annals of Mathematics Studies n°44, Princeton University Press, Princeton, 1960 (MR (1963) 25 # 3573).

[Hl1] P.R. Halmos, *The decomposition of measures*, Duke Math. J. **8** (1941), 386–392.

[Hl2] P.R. Halmos, *Measurable transformations*, Bull. Amer. Math. Soc. **55** (1949), 1015–1034.

[Hl3] P.R. Halmos, *Lectures on Ergodic Theory*, The Mathematical Society of Japan, Tokyo, 1956.

[HN] P.R. Halmos & J. von Neumann, *Operator methods in classical mechanics II*, Ann. of Math. **43** (1942), 332–350.

[HS] E. Hewitt & L.J. Savage, *Symmetric measures on cartesian products*, Trans. Amer. Math. Soc. **80** (1955), 470–501.

[HJ] J. Hoffmann-Jørgensen, *Existence of conditional probabilities*, Math. Scand. **28** (1971), 257–264.

[Ji] M. Jirina, *On regular conditional probabilities*, Czechoslovak Math. J. **84** (1959), 445–451.

[Ko] A.N. Kolmogorov, *Foundations of the theory of probability*, Chelsea, New York, 1950 (translation of *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Springer, Berlin, 1933).

[KB] N. Kryloff & N. Bogoliouboff, *La théorie générale de la mesure dans son application à l'étude des systèmes dynamiques de la mécanique non linéaire*, Ann. of Math. **38** (1937), 65–113.

[LM1] G. Licht & G. Michaille, *Convergence results for subadditive processes*, Prépublication n° 28 (15 pages) (1994), Départ. de Math., Université de Montpellier, France.

[LM2] G. Licht & G. Michaille, *Global-local subadditive ergodic theorems and applications to homogenization in elasticity*, submitted (1999).

[Mah] D. Maharam, *Decompositions of measure algebras and spaces*, Trans. Amer. Math. Soc. **69** (1950), 142–160.

[N1] J. von Neumann, *Zur Operatorenmethode in der klassischen Mechanik*, Ann. of Math. **33** (1932), 587–642, 789–791 (pages 307–365 in [N4]).

[N2] J. von Neumann, *Proof of the quasi-ergodic hypothesis*, Proc. Nat. Acad. Sci. U.S.A. **18** (1932), 70–82 (pages 261–273 in [N4]).

[N3] J. von Neumann, *Physical applications of the ergodic hypothesis*, Proc. Nat. Acad. Sci. U.S.A. **18** (1932), 263–266 (pages 274–277 in [N4]).

[N4] J. von Neumann, *Collected works*, vol. II, Operators, Ergodic theory and almost periodic functions in a group (A.H. Taub, ed.), Pergamon Press, Oxford, 1961.

[NZ] X.X. Nguyen & H. Zessin, *Ergodic theorems for spatial processes*, Z. Wahrscheinlichkeitstheorie verw. Gebiete **48** (1979), 133–158.

[Ox] J.C. Oxtoby, *Ergodic sets*, Bull. Amer. Math. Soc. **58** (1952), 116–136.

[Ph] R.R. Phelps, *Lectures on Choquet's theorem*, Van Nostrand, Princeton, 1966.

[RR] S. Ramakrishnan & B.V. Rao, *B-spaces are standard Borel*, The Annals of Probability **8** (1980), 1191.

[T] M. Talagrand, *Capacités invariantes extrémales*, Ann. Inst. Fourier **28** (1978), 79–146.

[V1] M. Valadier, *Désintégration d'une mesure sur un produit*, C. R. Acad. Sci. Paris Sér. A **276** (1973), 33–35 (summary of *Comparaison de trois théorèmes de désintégration*, Sém. Anal. Convexe **2** (1972) 10.1–10.21).

[V2] M. Valadier, *Structure des atomes conditionnels d'un espace de probabilité*, Sém. Anal. Convexe **10** (1980), 1.1–1.17.

[Var] V.S. Varadarajan, *Groups of automorphisms of Borel spaces*, Trans. Amer. Math. Soc. **109** (1963), 191–220.

Département de Mathématiques — case 051, Université Montpellier II, place Eugène Bataillon, 34 095 MONTPELLIER CEDEX 5, France

E-mail address: valadier@math.univ-montp2.fr