Journal of Convex Analysis
Volume 18 (2011), No. 4, 1127–1140

Stationary Stochastic Processes are Mixing of Ergodic Ones: Contingency

Michel Valadier
*Mathématiques, Université Montpellier II, Place Eugène Bataillon,
Case Courier 051, 34095 Montpellier Cedex, France*

Received: December 6, 2010

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 into ergodic classes[1].

Decomposition of more general stationary processes in ergodic classes goes back to von Neumann. This result has been improved and/or rediscovered several times, and it received a lot of different proofs. Its philosophical interpretation as the concept of *contingency* does not seem given in the literature. After some preliminaries we will survey a part of the most basic results.

Added in December 2010. This text was written in November 2000; I keep it unchanged except for small necessary modifications. It is what should be published in place of [65] if there did not happen a misunderstanding (see [66] for essentially a bibliographical supplement).

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 3.2 and the consequence after (Mackey made similar observations: see reference to [43] at the end of Section 3).

When the stationary process is a Markov chain with a finite number of states the classical theory, already in Doob's book [16, Chapter V], shows that the set of states divides into *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 of the process is the one of our life) to the same thing as if the process was ergodic.

[1] My first contact with Markov chains was in [54].

ISSN 0944-6532 / $ 2.50

Decomposition of more general stationary processes in ergodic classes goes back to von Neumann [48, 1932, Theorem 2 p. 617]. This is a several times rediscovered result rarely included in text books: Kallenberg [33, 1997, Theorem 9.12 p. 164] is an exception (see also the more specialized books of Denker, Grillenberger and Sigmund [15, 1976, pp. 73–74] and Kifer [36, 1986, Prop. 2.1 p. 23 and Th. 1.1 pp. 193–194], the survey of Mackey [43, 1974, pp. 192–193] and the course of Aldous [1, 1983, Theorem 12.10 p. 100]). Hopf [29, p. 31] refers to [48] and [38]. From Chersi [7], Jacobs [31] (see in German [30, pp. 85–88]) and Mañé [46] treat the question. All amounts to the following: there is a probability law which I call the *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. Mackey says [44, p. 33]: "there is seldom any loss in generality in restricting attention to those stationary stochastic processes in which the associated measure preserving action is indeed ergodic" (see also [43, p. 202]).

After some preliminaries we will review several statements. I tried gathering a reasonable list of references (excluding works relevant to differential geometry or C^*-algebras). Surely some other papers should be listed. I hope this paper will give taste for the field to many readers.

As references for proofs see Choquet [8, 1956/57], Kryloff-Bogoliouboff [38, 1937] way which was made a bit more precise by Oxtoby [53, 1952] and greatly generalized by Dynkin [18, 1978]. As references for other formulations see Rohlin [58, 1949], Maharam [45, 1950], Farrell [19, 1962], Varadarajan [67–68, 1963], Schmidt [59, 1978], Shimomura [60–61, 1978–90], Kerstan-Wakolbinger [34–35, 1980], Chersi [7, 1987], Zimmermann [69, 1992] and for a generalization to capacities Talagrand [62, 1978].

2. Stationary and ergodic processes

Let $(K, \mathcal{K})$ be a measurable space (for the Choquet and Kryloff-Bogoliouboff points of view, K will be compact metrizable). A *stochastic process* with discrete time taking its values in K is a bilateral sequence $(X_n)_{n\in\mathbb{Z}}$ of random variables (in short r.v.) defined on a probability space $(\Xi, \mathcal{S}, \Pi)$ which take their values in K. The set $\Omega = K^{\mathbb{Z}}$ is more fundamental than Ξ. It is still Borel standard when K is Borel standard and compact metrizable when K is so. Let $\mathcal{F} = \mathcal{K}^{\otimes\mathbb{Z}}$; 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 $\xi \mapsto (X_n(\xi))_{n\in\mathbb{Z}}$. The "canonical" process $(\widetilde{X}_n)_{n\in\mathbb{Z}}$ 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 $(X_n)_{n\in\mathbb{Z}}$.

A point $\omega = (x_n)_{n\in\mathbb{Z}} \in \Omega$ is a *trajectory*. The bijective map T of Ω in itself defined by $T\big((x_n)_{n\in\mathbb{Z}}\big) = (x_{n+1})_{n\in\mathbb{Z}}$ 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 $(X_n)_{n\in\mathbb{Z}}$ is *stationary* if its law is *invariant* i.e. for any $A \in \mathcal{F}$, $P(T^{-1}A) = P(A)$ (that is $T_\#(P) = P$).

Definitions. Let $(X_n)_{n\in\mathbb{Z}}$ 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*[2] 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.*

In ergodic theorems, T^j denotes, when $j \in \mathbb{Z}_+^* = \mathbb{N}^*$, the j-th power of T: $T^j = T \circ \cdots \circ T$, $T^0 = \mathrm{id}_\Omega$ 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 3.1 is elementary. It will be applied, when K is compact, to $\Lambda = \Omega = K^{\mathbb{Z}}$ as well as to $\Lambda = K^d$.

Proposition 3.1. *Let Λ be compact metrizable and D a dense subset*[3] *of $\mathcal{C}(\Lambda)$ (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 $\mathcal{C}(\Lambda)$) 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 3.2. *Assume that K is compact metrizable 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 $P_{(X_{-d+1},\ldots,X_0)}$ of $(X_{-d+1}, \ldots, X_0)$ is the weak limit of*

$$k^{-1} \sum_{j=0}^{k-1} \delta_{(x_{-(j+d-1)},\ldots,x_{-j})}$$

as $k \to +\infty$.

Consequence. Hence P-almost surely, knowing $(x_n)_{n\le 0}$ implies the knowledge of $P_{(X_{-d+1},\ldots,X_0)}$ hence, for $p \le q$ in $\mathbb{Z}$, of $P_{(X_p,X_{p+1},\ldots,X_{q-1},X_q)}$ since stationarity implies $P_{(X_p,\ldots,X_q)} = P_{(X_{p-q},\ldots,X_0)}$. Recall now that P is the projective limit of the measures[4] $P_{(X_p,\ldots,X_q)}$. So, mathematically, P can be identified; from a numerical point of view, this is another story: see the numerous 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 $f \in D_d$ let $\bar{f}$ denote the function on Ω associated to f which is defined by: $\bar{f}\big((x_n)_{n\in\mathbb{Z}}\big) =$

[2]Doob [16, p. 457], and several authors, say "metrically transitive".
[3]It is sufficient that the linear subspace of $\mathcal{C}(\Lambda)$ spanned by D is dense.
[4]More simply, when the $P_{(X_p,\ldots,X_q)}$ are known, the values of P on the algebra of cylindrical sets are known, and this algebra generates $\mathcal{B}(\Omega)$.

$f(x_{-d+1}, \dots, x_0)$. By Birkhoff's theorem (it is in most text books but goes back to the 1931 famous paper [4]), if $(x_n)_{n\in\mathbb{Z}}$ does not belong to a P-negligible set N_f,

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

By Proposition 3.1 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 $(x_n)_{n\in\mathbb{Z}} \notin \bigcup_{f\in D_d} N_f$. So the statement holds for $(x_n)_{n\in\mathbb{Z}}$ not in $\bigcup_{d\in\mathbb{N}^*}\big[\bigcup_{f\in D_d} N_f\big]$. □

Comment. The meaning of Theorem 3.2 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 probability laws on $K^{\mathbb{N}^*}$.

Then knowing $\xi = (x_n)_{n\le 0}$, the future obeys to the conditional law L_ξ on $K^{\mathbb{N}^*}$. I wrote this before reading of [43]: there another way to reconstruct the process law from the complete past of one trajectory is exposed; see specially the bottom of page 204 till the end of Section 5 and the top of page 223. (On the subject of prediction of stationary processes there is the ambitious book of Furstenberg [22, 1960]. It uses systematically C^*-algebras.)

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"[5] family of versions of the $\mathbf{E}^{\mathcal{G}}(1_B)$.

[5]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.*

This has a long story in probability theory: von Neumann [48], Kolmogorov [37], Jirina [32], Hoffmann-Jørgensen [28], Valadier [63–64] for some details[6]. (For textbooks see Bauer [3], Dudley [17].) 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 [17, 10.2.5 p. 272], Doob [16, Th. 9.1 p. 27], Kolmogorov [37, 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[7] $\Omega_1 \times \Omega_2$, $\omega = (\xi, \zeta)$ and $\mathcal{G}$ is generated by the projection on Ω_1 (possibly ξ is the past, ζ is 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 $\Omega = K^{\mathbb{Z}}$ 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 5.1. *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 (I have rediscovered Mackey's point of view), the prediction of stationary processes 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 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 5.1 the roles of time, past and future are essential.

[6] At the time when in France only Bourbaki and Jirina were quoted, I wrote [63] where I gave a result of Hoffmann-Jørgensen [28] in the framework of a product and where I compared several statements of this time. In [64, p. 13] I had the idea, being not aware of [23], of introducing the quotient tribe.
[7] To be more precise $(\Omega_1, \mathcal{F}_1)$ is separated and countably generated and Ω_2 is a "good" topological space: Suslin is a quite general hypothesis (see [14]).

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 5.2. *Assume K is compact metrizable.*

1) *The set $\mathcal{L}_{\text{st}}$ of invariant probabilities on Ω is a non empty convex compact subset of $\mathcal{M}^1_+(\Omega)$.*

2) *The set $\partial\mathcal{L}_{\text{st}}$ of extreme points of $\mathcal{L}_{\text{st}}$ coincide with the set of ergodic laws denoted by $\mathcal{L}_{\text{erg}}$.*

3) *Let $P \in \mathcal{L}_{\text{st}}$. There exists a probability measure λ on $\partial\mathcal{L}_{\text{st}} = \mathcal{L}_{\text{erg}}$ such that*

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

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

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

as

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

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 [5, 1960] and Choquet knew it before; I recommend the proof of Denker-Grillenberger-Sigmund [15, (5.6) p. 24]. Then the conclusion follows from the Choquet integral representation theorem (besides the quoted works of Choquet one can see [6, IV.7.2 Th. 1 p. 219] and Phelps [56, Section 10 pp. 77–85]). □

Remarks. One or two year before Choquet, Hewitt-Savage [27, 1955] 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 [20]; on this subject see Aldous [1]). But in [27] 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-Bogolioubоff's way

We still assume K compact metrizable and denote by $\mathcal{M}^1_+(\Omega)$ the set of probability measures on $(\Omega, \mathcal{B}(\Omega))$ endowed with the weak topology. We will not prove the next statement: see [38], [53] and [15] (the proofs of [18] and [33] contains very good arguments but they do not use the peculiar feature that here T (as T^{-1}) is continuous).

Theorem 5.3. *Assume K is compact metrizable.*

1) *The subset of* Ω

$$\Omega_{\mathrm{erg}} := \left\{ \omega \in \Omega : \frac{1}{n} \sum_{j=0}^{n-1} \delta_{T^j \omega} \text{ weakly converges to an ergodic law} \right\} \tag{2}$$

is Borel and has P*-measure* 1 *for any invariant probability* P.

2) *Let, for* $\omega \in \Omega_{\mathrm{erg}}$, $Q^\omega := \lim \frac{1}{n} \sum_{j=0}^{n-1} \delta_{T^j \omega}$ *and* $\Gamma_\omega := \{\omega' \in \Omega_{\mathrm{erg}} : Q^{\omega'} = Q^\omega\}$. *The set* Γ_ω *is Borel (even* $F_{\sigma\delta}$*), invariant and* Q^ω *is carried by* Γ_ω.

3) *The family* $(Q^\omega)_{\omega \in \Omega_{\mathrm{erg}}}$ *disintegrates any invariant probability* P *relatively to* $\mathcal{I}$, *which means: it is* $\mathcal{I}_0$*-measurable (*$\mathcal{I}_0$ *is the tribe induced by* $\mathcal{I}$ *on* Ω_{erg}*) and*

$$\forall B \in \mathcal{B}(\Omega),\ \forall A \in \mathcal{I}, \quad P(A \cap B) = \int_{A \cap \Omega_{\mathrm{erg}}} Q^\omega(B)\, dP(\omega) \tag{3}$$

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) To be more precise Ω_{erg} and Q^ω must be defined with bilateral limits as in the following formula

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

2) There are two equivalence relations: firstly

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

which makes sense on[8] Ω_{erg} and secondly

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

Let us prove the invariance of Γ_ω, the class of $\omega \in \Omega_{\mathrm{erg}}$ for (R1). For any $f \in \mathcal{C}(\Omega)$ changing ω in $T\omega$ or in $T^{-1}\omega$ does not change the Cesàro limit of the sequence $\big(f(T^j\omega)\big)_j$ that is $\lim_n \int f\, dQ_n^\omega$ where $Q_n^\omega := \frac{1}{n} \sum_{i=0}^{n-1} \delta_{T^i\omega}$. 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 Ω_{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 $P := \varpi^{\otimes \mathbb{Z}}$. Since the X_n are i.i.d. there is a unique class for (R1): the process is ergodic (see for example [16, Th. 1.2 p. 460]) and $Q^\omega \overset{P\text{-a.e.}}{=} P$. For any P-negligible N,

[8] To be more precise Q^ω makes sense on the larger set Ω_{qr} of "quasi-regular" points.

$\Omega \backslash N$ 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 $\Omega \backslash \Omega_{\text{erg}}$ 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 [13].

3) 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 (3), 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 (3) looks as (1) of Theorem 5.2.

For historical works which attacked disintegrating P with respect to $\mathcal{I}$ see Halmos [23] and Ambrose-Halmos-Kakutani [2]. Disintegration is used by Kallenberg in his proofs [33, pp. 162–163] but with rather too much symbolical notations (in my opinion he uses both the idea of using a determining class, as Varadarajan and later Dynkin did, and a disintegration: he has a random variable ξ with values in $(S, \mathcal{S})$ and the tribe of invariant sets $\mathcal{I}$; the kernel $(s, B) \mapsto \mu(s, B)$ he considers disintegrates the image of P on $(S^2, \mathcal{I} \otimes \mathcal{S})$ by $\omega \mapsto (\xi(\omega), \xi(\omega))$). McCutcheon [47, p. 120] discusses quickly ergodic decomposition: he states a disintegration theorem and says that when the sub-σ-algebra under consideration is that of invariant sets one get ergodic measures (see the three lines paragraph after Theorem 4.3.3).

4) Dynkin proves many other results: specially he gets [18, Theorem 3.1] that $\mathcal{L}_{\text{st}}$ is a simplex in the Choquet sense whose extreme points are the ergodic measures.

5.3. The case of measurable spaces

Most processes are unbounded $\mathbb{R}$-valued ones, so the foregoing results do not directly apply. One can consider $\mathbb{R}$ as a subspace of the compact $\overline{\mathbb{R}}$: this "respects" the topology and introduces a compact over-space, but the big drawback is that $\mathbb{R}$ is not closed in $\overline{\mathbb{R}}$.

Several authors obtained an ergodic decomposition under the hypothesis that[9] $(K, \mathcal{K})$ is a *Borel standard* or *Lusin* measurable space, that is a measurable space isomorphic to a Borel subset of a Polish topological space. Any Borel standard space $(K, \mathcal{K})$ is either countable, either has the cardinality of $\mathbb{R}$. In the first case it is isomorphic to $\{1, \ldots, n\}$ or to $\mathbb{N} \cup \{\infty\}$ (the tribe being that of all subsets). In the second case it is isomorphic to $([0, 1], \mathcal{B}([0, 1]))$. As a reference see [14, Appendice au chapitre III, Th. 80 p. 249] (from many authors all properties of Borel standard spaces are proved in Kuratowski's book [39]).

Hence if $(K, \mathcal{K})$ is Borel standard there exists a compact metrizable topology on K whose Borel tribe coincide with $\mathcal{K}$. The Bernoulli shift is still an homeomorphism of $K^{\mathbb{Z}}$. So the conclusion of Theorem 5.3 applies, including existence of the classes Γ_ω, except that the convergence in (2) is with respect to a possibly non natural topology on K.

[9] More precisely, when T is not the Bernoulli shift, the hypotheses concern $(\Omega, \mathcal{F})$.

For example $\mathbb{R}$ is Borel standard. A direct way to check that $\mathbb{R}$ is isomorphic as a measurable space to $[0,1]$ or to $\overline{\mathbb{R}}$ is the following: Let $\varphi : \mathbb{R} \to \overline{\mathbb{R}}$ defined by $\varphi(x) = x$ on $\mathbb{R}\backslash\mathbb{N}$ and any bijection from $\mathbb{N}$ onto $\mathbb{N} \cup \{-\infty, +\infty\}$. Of course when f runs through $\mathcal{C}(\overline{\mathbb{R}})$, $f \circ \varphi$ runs through some subset of the set of all bounded borelian functions on $\mathbb{R}$.

Note that in Ergodic theory maybe three notions of isomorphisms can be used[10]: one-to-one map between sets, one-to-one map between subsets of full measure and isomorphism of the quotient σ-algebras such as $\mathcal{F}/P$ (the set A and B in $\mathcal{F}$ are equivalent if $P(A\Delta B) = 0$). See Petersen [55, pp. 15–17] and (only for the two first notions) [15, pp. 3–5].

Among all works treating Borel standard spaces we will mention two of them.

Maybe only Chersi [7] succeeded proving narrow convergence of the sequence $(Q_n^\omega)_n$; surely this is thanks to the notion he used of Daniell integral.

Dynkin [18], developing his idea of sufficient statistic, covers with a unified approach several other notions: Gibbs states, symmetric laws (de Finetti-Hewitt-Savage), superharmonic functions... Now we will try giving a flavour of Dynkin's paper. He introduced the notion of "B-space" which some times later Ramakrishnan and Rao in [57] proved it coincides with the notion of Borel standard. Let $\widetilde{\mathcal{F}}$ denote the set of all real bounded $\mathcal{F}$-measurable functions on Ω. Dynkin gets the existence of a "determining" set of measurable bounded real functions he denotes by W. In formula (2) of Theorem 5.3, the weak convergence $Q_n^\omega \to Q^\omega$ holds for the $\sigma(\mathcal{M}^b(\Omega), \mathcal{C}(\Omega))$ topology and $\omega \in \Omega_{\mathrm{qr}}$, the set of quasi-regular points. In Dynkin's paper there are more P-negligible sets and equalities P-almost everywhere and he gets only the following two weak convergences:

(i) on a set Ω' of P-measure 1 for all $P \in \mathcal{L}_{\mathrm{st}}$, $\forall f \in W$, $\int_\Omega f\, dQ_n^\omega \to \int_\Omega f\, dQ^\omega$;

(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 $\forall f \in \widetilde{\mathcal{F}}$, $\forall P \in \mathcal{L}_{\mathrm{st}}$, $P(\{\omega \in \Omega' : \int_\Omega f\, dQ_n^\omega \to \int_\Omega f\, dQ^\omega\}) = 1$.

Remark. The work of Lauritzen [40] could have some connections with Dynkin (this author, in a preliminary work, his thesis, does not quote Dynkin's paper. I did not see this book).

6. Further comments about stationary processes

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 knows that the process is Markov and one observes a trajectory living in $\{1,2\}$ obeying to the transition matrix $\begin{pmatrix} 1/4 & 1/2 \\ 3/4 & 1/2 \end{pmatrix}$ then either the whole process is ergodic and obeys to this transition matrix with probabilities of states equal to $\begin{pmatrix} 2/5 \\ 3/5 \end{pmatrix}$ or there exists other ergodic classes about nothing is known.

A remark about a small strange phenomenon: suppose one observes $(x_n)_{n \in \mathbb{Z}_-}$ where

[10] Isomorphism with $[0,1]$ is an essential tool in [58]. See also [45, Th. 6 p. 157].

$x_n = (-1)^n$. One possibility is: there is not any random and this is just a periodic behavior which possibly 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 $(y_n)_{n\in\mathbb{Z}} = \big((-1)^{n+1}\big)_{n\in\mathbb{Z}}$ 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 $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ and probabilities of states equal to $\begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$.

7. More general results

The problem of relaxing the ergodic hypothesis into the stationarity one comes up in the theory of stochastic homogenization (for a few references see Dal Maso-Modica [11–12], Nguyen-Zessin [52] and Licht-Michaille [41–42]). To be more precise a version of the results of Section 5 with respect to the group $\mathbb{R}^N$ in place of $\mathbb{Z}$ is needed.

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 again contingency. But as long as the origin of stones remains unchanged, all amounts as if the ergodic hypothesis was satisfied.

In fact the ergodic decomposition remain valid in the following two extended directions.

7.1. Non bijective transformations

The pointwise ergodic theorem remains true when one consider a measurable transformation T. In this line a countable set of measurable transformations which commute has been considered by Farrell [19], Vadarajan [68] and Dynkin [18, Th. 6.1 p. 717]. More recently Kallenberg [33, Theorem 9.12] proved an ergodic decomposition theorem for a finite number of measurable transformations $T_1, \ldots, T_d$ which commute; he used a mean spatial ergodic theorem he proved before [33, Theorem 9.9].

7.2. Action of a group

One can consider a locally compact group G. The group is not necessarily commutative but it should admit a countable dense subgroup. In 1962 Farrell [19] and Varadarajan [67–68] worked simultaneously and independently in this direction (see also [18, Remark p. 717]). Both use limit theorems about powers of compositions of conditional expectations.

Note that the case of flows (that is $G = \mathbb{R}$) was already treated by von Neumann and Kryloff-Bogoliouboff and that Fomin gave in 1950 some results in this line ([21] is in Russian).

Acknowledgements. The author thanks Gérard Michaille and Charles Castaing for their helpful comments and Baptiste Chapuisat for converting Macintosh files into PC ones (in 2010).

References

[A star points out a reference I could not see myself.]

[1] D. J. Aldous: Exchangeability and Related Topics, Lecture Notes in Math. 1117, Springer, Berlin (1985).

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

[3] H. Bauer: Probability Theory and Elements of Measure Theory, Holt, Rinehart & Winston, New York (1972); transl. of: Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie, De Gruyter, Berlin (1968).

[4] G. D. Birkhoff: Proof of the ergodic theorem, Proc. Natl. Acad. Sci. USA 17 (1931) 656–660.

[5] J. R. Blum, D. L. Hanson: On invariant probability measures. I, Pac. J. Math. 10 (1960), 1125–1129.

[6] N. Bourbaki: Intégration. Chapitres 1 à 4, Hermann, Paris (1965).

[7] F. Chersi: An ergodic decomposition of invariant measures for discrete semiflows on standard Borel spaces, in: Advanced Topics in the Theory of Dynamical Systems (Trento, 1987), G. Fusco et al. (ed.), Notes and Reports in Mathematics in Science and Engineering 6, Academic Press, Boston (1989) 75–87.

[8] G. Choquet: Existence et unicité des représentations intégrales au moyen des points extrémaux dans les cônes convexes, Sémin. Bourbaki 9(139) (1956/57) 15 p.

[9] G. Choquet: Lectures on Analysis. Vol. II: Representation Theory, W. A. Benjamin, New York (1969).

[10] G. Choquet: Représentation intégrale, in: Measure Theory and its Applications (Sherbrooke, 1982), J. M. Belley et al. (ed.), Lecture Notes in Math. 1033, Springer, Berlin (1983) 114–143.

[11] G. Dal Maso, L. Modica: Nonlinear stochastic homogenization, Ann. Mat. Pura Appl., IV. Ser. 144 (1986) 347–389.

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

[13] C. Dellacherie: Nombre au hasard de Borel à Martin-Loef, Gaz. Math., Soc. Math. Fr. 11 (1978) 23–58.

[14] C. Dellacherie, P. A. Meyer: Probabilités et Potentiel. Chapitres I–IV, Herman, Paris (1975); English Ed.: North-Holland, Amsterdam (1978).

[15] M. Denker, C. Grillenberger, K. Sigmund: Ergodic Theory on Compact Spaces, Lecture Notes in Math. 527, Springer, Berlin (1976).

[16] J. L. Doob: Stochastic Processes, Wiley, New York (1953).

[17] R. M. Dudley: Real Analysis and Probability, Wadsworth & Brooks/Cole Mathematics Series, Wadsworth & Brooks/Cole, Pacific Grove (1989).

[18] E. B. Dynkin: Sufficient statistics and extreme points, Ann. Probab. 6 (1978) 705–730.

[19] R. H. Farrell: Representation of invariant measures, Ill. J. Math. 6 (1962) 447–467.

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

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

[22] H. Furstenberg: Stationary Processes and Prediction Theory, Annals of Mathematics Studies 44, Princeton University Press, Princeton (1960).

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

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

[25] P. R. Halmos: Lectures on Ergodic Theory, Mathematical Society of Japan, Tokyo (1956).

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

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

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

[29] E. Hopf: Ergodentheorie, Chelsea, New York (1948); 1st Ed.: Springer, Berlin (1937).

[30] K. Jacobs: Neuere Methoden und Ergebnisse der Ergodentheorie, Springer, Berlin (1960).

[*31] K. Jacobs: Lecture Notes on Ergodic Theory, Matematisk Institut, Aarhus Universitet, Aarhus (1963).

[32] M. Jirina: On regular conditional probabilities, Czech. Math. J. 84 (1959) 445–451.

[33] O. Kallenberg: Foundations of Modern Probability, Springer, New York (1997).

[34] J. Kerstan, A. Wakolbinger: Opérateurs markoviens et l'ensemble de leurs lois de probabilité invariantes, C. R. Acad. Sci., Paris, Sér. A 291 (1980) 163–166.

[35] J. Kerstan, A. Wakolbinger: Ergodic decomposition of probability laws, Z. Wahrsch. Verw. Gebiete 56 (1981) 399–414.

[36] Y. Kifer: Ergodic Theory of Random Transformations, Progress in Probability and Statistics 10, Birkhäuser, Boston (1986).

[37] A. N. Kolmogorov: Foundations of the Theory of Probability, Chelsea, New York (1950); transl. of: Grundbegriffe der Wahrscheinlichkeitsrechnung, Springer, Berlin (1933).

[38] 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. Math. 38 (1937) 65–113.

[39] C. Kuratowski: Topologie. I: Espaces Métrisables, Espaces Complets, 2nd Ed., Monografie Matematyczne 20, Warszawa (1948).

[*40] S. L. Lauritzen: Extremal Families and Systems of Sufficient Statistics, Lecture Notes in Statistics 49, Springer, Berlin (1988).

[41] G. Licht, G. Michaille: Convergence results for subadditive processes, Prépublication 28, Départ. de Math., Université de Montpellier (1994) 15 p.

[42] G. Licht, G. Michaille: Global-local subadditive ergodic theorems and applications to homogenization in elasticity, Ann. Math. Blaise Pascal 9 (2002) 21-62.

[43] G. W. Mackey: Ergodic theory and its significance for statistical mechanics and probability theory, Adv. Math. 12 (1974) 178–268.

[44] G. W. Mackey: Von Neumann and the early days of ergodic theory, in: The Legacy of John von Neumann (Hempstead, 1988), J. Glimm et al. (ed.), Proc. Sympos. Pure Math. 50, Amer. Math. Soc., Providence (1990) 25–38.

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

[*46] R. Mañé: Ergodic Theory and Differentiable Dynamics, Springer, Berlin (1987); transl. from: Introdução à Teoria Ergódica, IMPA, Rio de Janeiro (1983).

[47] R. McCutcheon: Elemental Methods in Ergodic Ramsey Theory, Lecture Notes in Math. 1722, Springer, Berlin (1999).

[48] J. von Neumann: Zur Operatorenmethode in der klassischen Mechanik, Ann. Math. 33 (1932) 587–642, 789–791 (pp. 307–365 in [51]).

[49] J. von Neumann: Proof of the quasi-ergodic hypothesis, Proc. Natl. Acad. Sci. USA 18 (1932) 70–82 (pp. 261–273 in [51]).

[50] J. von Neumann: Physical applications of the ergodic hypothesis, Proc. Natl. Acad. Sci. USA 18 (1932) 263–266 (pp. 274–277 in [51]).

[51] 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).

[52] X. X. Nguyen, H. Zessin: Ergodic theorems for spatial processes, Z. Wahrsch. Verw. Gebiete 48 (1979) 133–158.

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

[54] R. Pallu de La Barrière: Cours d'Automatique Théorique, Dunod, Paris (1966).

[55] K. Petersen: Ergodic Theory, Cambridge Studies in Advanced Mathematics 2, Cambridge University Press, Cambridge (1983).

[56] R. R. Phelps: Lectures on Choquet's Theorem, Van Nostrand, Princeton (1966).

[57] S. Ramakrishnan, B. V. Rao: B-spaces are standard Borel, Ann. Probab. 8 (1980) p. 1191.

[58] V. A. Rohlin: On the fundamental ideas of measure theory, Trans. Amer. Math. Soc. 71 (1952) 1–54; transl. from: Mat. Sb., N. Ser. 67 (1949) 107–150.

[59] K. Schmidt: A probabilistic proof of ergodic decomposition, Sankhyā, Ser. A 40 (1978) 10–18.

[60] H. Shimomura: Ergodic decomposition of quasi-invariant measures, Publ. Res. Inst. Math. Sci. 14 (1978) 359–381.

[61] H. Shimomura: Remark to the ergodic decomposition of measures, Publ. Res. Inst. Math. Sci. 26 (1990) 861–865.

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

[63] 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, Montpellier 2(10) (1972) 21 p.

[64] M. Valadier: Structure des atomes conditionnels d'un espace de probabilité, Sém. Anal. Convexe, Montpellier 10(1) (1980) 17 p.

[65] M. Valadier: What differentiates stationary stochastic processes from ergodic ones: a survey, RIMS Kokyuroku 1215 (2001) 33–52.

[66] M. Valadier: Complement to "What differentiates stationary stochastic processes from ergodic ones: a survey", RIMS Kokyuroku 1264 (2002) 42–44.

[67] V. S. Varadarajan: On a problem in ergodic theory, Bull. Amer. Math. Soc. 68 (1962) 345–350.

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

[69] D. Zimmermann: Uniqueness in ergodic decomposition of invariant probabilities, Ill. J. Math. 36 (1992) 325–344.