Séminaire de probabilités (Strasbourg)

Michel Émery
Walter Schachermayer

On Vershik's standardness criterion and Tsirelson's notion of cosiness

Séminaire de probabilités (Strasbourg), tome 35 (2001), p. 265-305
<http://www.numdam.org/item?id=SPS_2001__35__265_0>

ON VERSHIK'S STANDARDNESS CRITERION AND TSIRELSON'S NOTION OF COSINESS

M. Émery and W. Schachermayer

Abstract. — Building on work done by A. Vershik some thirty years ago, the insight into different types of filtrations has recently seen important progress, due in particular to B. Tsirelson, and L. Dubins, J. Feldman, M. Smorodinsky, B. Tsirelson. Key concepts are the notions of a standard filtration (due to A. Vershik) and of a cosy filtration (due to B. Tsirelson). We investigate the relation between these two concepts and try to provide a comprehensive and self-contained presentation of the topic.

Part of this work is expository, and consists in translating into a probabilist's language Vershik's necessary and sufficient condition for standardness, and his theorem on lacunary isomorphism. There are also original results: Theorem 2 proves that standardness is in fact equivalent to a certain variant of the notion of cosiness, which we call *I-cosiness;* an example borrowed from Vershik and Smorodinsky then shows that I-cosiness is strictly stronger than another variant, D-cosiness, used in earlier works. Another new result is a (negative) answer to a question of H. von Weizsäcker: the last section gives an example of a filtration $(\mathcal{F}_n)_{n\in-\mathbb{N}}$ and a σ-field $\mathcal{B}$ such that $\mathcal{F}_0$ and $\mathcal{B}$ are "almost independent", but nevertheless

$$\bigcap_{n\in-\mathbb{N}} (\mathcal{F}_n \vee \mathcal{B}) \supsetneq \Big(\bigcap_{n\in-\mathbb{N}} \mathcal{F}_n \Big) \vee \mathcal{B} \, .$$

Many thanks to A. M. Vershik for enjoyable and fruitful conversations about his theory of filtrations. We are also grateful to the Schrödinger Institut in Vienna, where part of this work was done during the Mini-Symposium on the Classification of Filtrations held in December 1998, and to C. Leuridan for his remarks.

Introduction

The objects of this study are filtrations. We shall not be interested in their set-theoretical properties, but in their probabilistic ones: we shall only consider filtrations on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and the notions relevant for our analysis, e.g., that of independence, are not invariant under changes of the measure $\mathbb{P}$. (In full rigor, we should speak of filtered probability spaces rather than filtrations.) We refer to the next section for a precise definition of an isomorphism between two filtrations in the present context.

In the late sixties and early seventies, A. Vershik [17] initiated a classification of filtrations. Consider filtrations $(\mathcal{F}_n)_{n\in-\mathbb{N}}$, where time is a negative integer, such that $\bigcap_n \mathcal{F}_n$ is degenerate and each $\mathcal{F}_n$ is generated by $\mathcal{F}_{n-1}$ and a random variable independent of $\mathcal{F}_{n-1}$, uniformly distributed on $[0,1]$. A typical example of this situation is, of course, the filtration generated by an i.i.d. sequence $(X_n)_{n\in-\mathbb{N}}$ of random variables with uniform law on $[0,1]$; a natural (and innocent-looking) question is whether this example already covers all cases of filtrations verifying the above properties. One of Vershik's results is the following, highly non-trivial, fact:

these filtrations are not all isomorphic to each other. More precisely, calling *standard* a filtration generated by an independent sequence of uniformly distributed random variables, he exhibited non standard filtrations satisfying the above conditions, and obtained his *standardness criterion,* a necessary and sufficient condition for a filtration to be standard.

Written in the language of ergodic theory, these ideas did not find their way into the probabilistic culture until 25 years later, when they were used by L. Dubins, J. Feldman, M. Smorodinsky and B. Tsirelson [5] to show that standard filtrations are not stable under equivalent changes of probability. They deduced therefrom that Brownian filtrations are not stable either under equivalent changes of probability.

A further step in the search of filtration invariants was made by B. Tsirelson, who showed in [15] that a Walsh process is not immersible into a Brownian filtration, that is, into a filtration generated by some (finite- or infinite-dimensional) Brownian motion. The strategy of his proof involves introducing a new property, *cosiness,* possessed by all Brownian filtrations (but more general than mere "Brownianness"); he then shows that a Walsh process is not immersible into a cosy filtration.

This strategy, establishing non-cosiness to deduce non-Brownianness, has been adapted to other situations: J. Warren proves in [18] that the filtration generated by sticky Brownian motion is not cosy; the non-Brownian change of probability constructed on Wiener space by Dubins, Feldman, Smorodinsky and Tsirelson in [5] is shown in [3] to be non-cosy; a non-cosy change of time on Wiener space is constructed in [7]. In the latter two articles, [3] and [7], Tsirelson's definition of cosiness is slightly modified (weakened, and adapted to discrete time); what is used there is the variant of cosiness which we call *D-cosiness* below.

These two tools, the standardness criterion on the one hand, and cosiness and its variants on the other hand, are very efficient means of establishing that some given filtration is not standard (or not Brownian). The present article aims at bridging the gap between them, by establishing that standardness is equivalent to yet another variant of cosiness (we call it *I-cosiness*).

We shall first copy the proof of Vershik's criterion in the language of stochastic processes, and graft thereupon the equivalence between standardness and I-cosiness (Theorem 2 and Corollary 5).

Then we shall test the efficiency of this new criterion on one of Vershik's non standard, hence also non I-cosy, examples; somewhat unexpectedly, this particular non I-cosy example turns out to be D-cosy (Proposition 9).

Last, we shall show in Proposition 10 that the same example answers negatively a question raised by H. von Weizsäcker: if a filtration $(\mathcal{F}_n)_{n\leqslant 0}$ and a σ-field $\mathcal{B}$ are almost independent (this will be explained), does the germ σ-field $\bigcap_n (\mathcal{F}_n \vee \mathcal{B})$ always equal $\big(\bigcap_n \mathcal{F}_n\big) \vee \mathcal{B}$?

Notation and definitions

All probability spaces $(\Omega, \mathcal{A}, \mathbb{P})$ will be $\mathbb{P}$-complete; by a sub-σ-field of $\mathcal{A}$, we always mean an $(\mathcal{A}, \mathbb{P})$-complete sub-$\sigma$-field of $\mathcal{A}$. For instance, $\sigma(X)$ denotes the σ-field generated by the r.v. X and the null events; and when we consider a product $(\Omega' \times \Omega'', \mathcal{A}' \otimes \mathcal{A}'', \mathbb{P}' \times \mathbb{P}'')$ of probability spaces, $\mathcal{A}' \otimes \mathcal{A}''$ is completed for $\mathbb{P}' \times \mathbb{P}''$. Also, all filtered probability spaces $(\Omega, \mathcal{A}, \mathbb{P}, \mathcal{F})$ will satisfy the usual hypotheses of the general theory of processes: each $\mathcal{F}_t$ is $(\mathcal{A}, \mathbb{P})$-complete and, if the time-parameter is continuous, the filtration is right-continuous.

Recall that the σ-field $\mathcal{A}$ is essentially separable (resp. essentially finite) if it is generated by countably (resp. finitely) many events (and the null events; this is implicit). Equivalently, $\mathcal{A}$ is generated by some random variable (resp. simple random variable). This is tantamount to saying that $\mathrm{L}^1(\Omega, \mathcal{A}, \mathbb{P})$ is a separable Banach space (resp. is finite-dimensional), and entails that every sub-σ-field of $\mathcal{A}$ is also essentially separable (resp. essentially finite).

DEFINITION. — *An* embedding *of a probability space* $(\Omega, \mathcal{A}, \mathbb{P})$ *into another one* $(\bar{\Omega}, \bar{\mathcal{A}}, \bar{\mathbb{P}})$ *is a mapping* Ψ *from* $\mathrm{L}^0(\Omega, \mathcal{A}, \mathbb{P})$ *to* $\mathrm{L}^0(\bar{\Omega}, \bar{\mathcal{A}}, \bar{\mathbb{P}})$ *that commutes with Borel operations on r.v.'s:*

$$\Psi\big(f \circ (X_1, \dots, X_n)\big) = f \circ \big(\Psi(X_1), \dots, \Psi(X_n)\big) \text{ for every Borel } f : \mathbb{R}^n \to \mathbb{R}$$

and preserves probability laws:

$$\bar{\mathbb{P}}\big[\Psi(X) \in E\big] = \mathbb{P}\big[X \in E\big] \qquad \text{for every Borel } E \subset \mathbb{R}.$$

An embedding is always injective and transfers not only random variables, but also sub-σ-fields, filtrations, processes, etc. It is called an *isomorphism* if it is surjective; it then has an inverse. An embedding Ψ of $(\Omega, \mathcal{A}, \mathbb{P})$ into $(\bar{\Omega}, \bar{\mathcal{A}}, \bar{\mathbb{P}})$ is always an isomorphism between $(\Omega, \mathcal{A}, \mathbb{P})$ and $\big(\bar{\Omega}, \Psi(\mathcal{A}), \bar{\mathbb{P}}\big)$. See for instance [2] for more details.

DEFINITION. — *Given two filtered probability spaces* $(\Omega, \mathcal{A}, \mathbb{P}, \mathcal{F})$ *and* $(\bar{\Omega}, \bar{\mathcal{A}}, \bar{\mathbb{P}}, \bar{\mathcal{F}})$, *the filtrations* $\mathcal{F}$ *and* $\bar{\mathcal{F}}$ *are* isomorphic *if there exists an isomorphism* Ψ *from* $(\Omega, \mathcal{F}_\infty, \mathbb{P})$ *to* $(\bar{\Omega}, \bar{\mathcal{F}}_\infty, \bar{\mathbb{P}})$ *such that* $\Psi(\mathcal{F}) = \bar{\mathcal{F}}$.

It would be more rigorous to say that the filtered probability spaces (and not only the filtrations) are isomorphic; this precision is necessary when there may be an ambiguity on the probabilities $\mathbb{P}$ and $\bar{\mathbb{P}}$. In the sequel we shall never change probabilities, so we shall allow ourselves this abuse of language.

DEFINITIONS. — *Let* $\mathcal{F}$ *and* $\mathcal{G}$ *be two filtrations on a probability space* $(\Omega, \mathcal{A}, \mathbb{P})$.

The filtration $\mathcal{F}$ *is* included *in* $\mathcal{G}$ *if* $\mathcal{F}_t \subset \mathcal{G}_t$ *for each time* t.

The filtration $\mathcal{F}$ *is* immersed *in* $\mathcal{G}$ *if every* $\mathcal{F}$*-martingale is a* $\mathcal{G}$*-martingale* (this is stronger than mere inclusion).

The filtrations $\mathcal{F}$ *and* $\mathcal{G}$ *are* jointly immersed[1] *if each of them is immersed in* $\mathcal{F} \vee \mathcal{G}$ (the smallest filtration where $\mathcal{F}$ and $\mathcal{G}$ are included; it can be defined by $(\mathcal{F} \vee \mathcal{G})_t = \bigcap_{\varepsilon > 0} \big(\mathcal{F}_{t+\varepsilon} \vee \mathcal{G}_{t+\varepsilon}\big)$ for each time t).

As with the previous definition, the role played by $\mathbb{P}$ should be stressed: The fact that $\mathcal{F}$ is immersed in $\mathcal{G}$ is in general not stable under a change of probability.

Note also that, by a density argument, $\mathcal{F}$ is immersed in $\mathcal{G}$ if and only if every bounded $\mathcal{F}$-martingale is a (bounded) $\mathcal{G}$-martingale; and, by stopping, if and only if every local martingale for $\mathcal{F}$ is a local martingale for $\mathcal{G}$.

1. Or, more precisely, *jointly immersed in* $\mathcal{F} \vee \mathcal{G}$. But we shall see in Lemma 4 b) that this holds as soon as there exists a filtration $\mathcal{H}$ such that both $\mathcal{F}$ and $\mathcal{G}$ are immersed in $\mathcal{H}$.

The notion of an immersion is fundamental in many aspects of stochastic calculus; for instance, it is hidden inside the definition of a Brownian motion for a filtration, or of the Markov property with respect to a filtration. It has been used by many authors, sometimes implicitly, without giving it a name, sometimes explicitly, under various names; see [3] for more details and for some references. (The works [1] by D. Aldous and M. Barlow and [10] by D. Hoover should be added to those references.)

LEMMA 1. — *Let* $\mathcal{F}$ *and* $\mathcal{G}$ *be two independent filtrations* (that is, $\mathcal{F}_t$ and $\mathcal{G}_t$ are independent for each t). *Then* $\mathcal{F}$ *and* $\mathcal{G}$ *are jointly immersed in* $\mathcal{F}\vee\mathcal{G}$.

PROOF. — An independent enlarging of a filtration preserves its martingales. ∎

LEMMA 2. — *Let* $\mathcal{F}$, $\mathcal{G}$, $\mathcal{H}$ *and* $\mathcal{K}$ *be four filtrations on the same sample space, such that* $\mathcal{F}$ *is immersed in* $\mathcal{H}$ *and* $\mathcal{G}$ *in* $\mathcal{K}$. *If* $\mathcal{H}$ *and* $\mathcal{K}$ *are independent,* $\mathcal{F}\vee\mathcal{G}$ *is immersed in* $\mathcal{H}\vee\mathcal{K}$.

PROOF. — It suffices to show that the product FG of a bounded $\mathcal{F}$-martingale F and a bounded $\mathcal{G}$-martingale G is an $\mathcal{H}\vee\mathcal{K}$-martingale. This is obtained by taking any bounded, $\mathcal{H}_t$-measurable (respectively $\mathcal{K}_t$-measurable) random variable H_t (respectively K_t) and writing

$$\mathbb{E}[F_\infty G_\infty H_t K_t] = \mathbb{E}[F_\infty H_t]\,\mathbb{E}[G_\infty K_t] = \mathbb{E}[F_t H_t]\,\mathbb{E}[G_t K_t] = \mathbb{E}[F_t G_t H_t K_t]\ . \quad ∎$$

REMARK. — If $\mathcal{F}$ and $\mathcal{G}$ are immersed in $\mathcal{H}$, it is not always true that $\mathcal{F}\vee\mathcal{G}$ is immersed in $\mathcal{H}$, even when $\mathcal{F}$ and $\mathcal{G}$ are independent. A very simple counter-example can be built from two independent random variables U and V with uniform law on $\{-1,1\}$. Put $M_t = U\,\mathbb{1}_{\{t\geqslant 1\}}$ and $N_t = V\,\mathbb{1}_{\{t\geqslant 1\}}$. The filtrations $\mathcal{F}$ and $\mathcal{G}$ respectively generated by M and N are independent and immersed (by Lemma 1) in the filtration $\mathcal{H}$ given by $\mathcal{H}_t = \sigma(UV)$ if $t<1$ and $\mathcal{H}_t = \sigma(U,V)$ if $t\geqslant 1$; but the process $MN = UV\,\mathbb{1}_{[\![1,\infty[\![}$ is not an $\mathcal{H}$-martingale, though it is of course an $\mathcal{F}\vee\mathcal{G}$-martingale.

In other words, if $\mathcal{F}$ and $\mathcal{G}$ are two independent filtrations, the product of an $\mathcal{F}$-martingale and a $\mathcal{G}$-martingale is always an $\mathcal{F}\vee\mathcal{G}$-martingale; but if M and N are two independent martingales for a filtration $\mathcal{H}$, the product MN is not necessarily an $\mathcal{H}$-martingale. (Note, however, that the product of two independent *continuous* $\mathcal{H}$-martingales is always an $\mathcal{H}$-martingale, for in this case $\langle M,N\rangle = 0$.)

A sufficient condition for the product of two martingales to be a local martingale is that their covariation process is constant; this suggests the following statement: *Let two filtrations* $\mathcal{F}$ *and* $\mathcal{G}$ *be immersed in* $\mathcal{H}$. *Suppose there exists an* $\mathcal{H}$*-optional subset* A *of* $\mathbb{R}_+\times\Omega$ *such that, for all* $\mathcal{F}$*-martingales* M *and* $\mathcal{G}$*-martingales* N, *the processes* $\int \mathbb{1}_A\, d[M,M]$ *and* $\int \mathbb{1}_{A^c}\, d[N,N]$ *are constant. Then the filtration* $\mathcal{F}\vee\mathcal{G}$ *is immersed in* $\mathcal{H}$. The simple proof of this statement is left to the reader; we shall only need the particular instance when A is the deterministic interval $[\![0,t]\!]$:

LEMMA 3. — *Let two filtrations* $\mathcal{F}$ *and* $\mathcal{G}$ *be immersed in some filtration* $\mathcal{H}$. *Suppose that for some time* t, $\mathcal{F}_\infty$ *is included in* $\mathcal{H}_t$ *and* $\mathcal{G}_t$ *is degenerate. The filtration* $\mathcal{F}\vee\mathcal{G}$ *is immersed in* $\mathcal{H}$.

PROOF. — It suffices to show that the product FG of a bounded $\mathcal{F}$-martingale F and a bounded $\mathcal{G}$-martingale G is an $\mathcal{H}$-martingale. The martingale equality may be checked separately on the intervals $[0,t]$ and $[t,\infty)$, since they have t in common.

On $[t,\infty)$, this equality holds because $F = F_t$ is constant and $\mathcal{H}_t$-measurable; on $[0,t]$, it holds because $G = \mathbb{E}[G_\infty]$ is constant and deterministic. ∎

LEMMA 4. — *Let $\mathcal{F}$, $\mathcal{G}$ and $\mathcal{H}$ be three filtrations on some sample space $(\Omega,\mathcal{A},\mathbb{P})$.*

a) *If $\mathcal{F}$ is included in $\mathcal{G}$ and $\mathcal{G}$ in $\mathcal{H}$, and if $\mathcal{F}$ is immersed in $\mathcal{H}$, then $\mathcal{F}$ is immersed in $\mathcal{G}$.*

b) *if $\mathcal{F}$ and $\mathcal{G}$ are immersed in $\mathcal{H}$, they are jointly immersed (in $\mathcal{F}\vee\mathcal{G}$).*

REMARK. — Lemma 4 b) can be rewritten as follows: *two filtrations $\mathcal{F}$ and $\mathcal{G}$ on the same sample space are jointly immersed if and only if there exists on that space a filtration $\mathcal{H}$ such that $\mathcal{F}$ and $\mathcal{G}$ are immersed in $\mathcal{H}$.* This equivalence explains a posteriori the choice of the name for this property.

PROOF. — a) Any $\mathcal{F}$-martingale is an $\mathcal{H}$-martingale adapted to $\mathcal{G}$, whence a $\mathcal{G}$-martingale.

b) If both $\mathcal{F}$ and $\mathcal{G}$ are immersed in $\mathcal{H}$, applying a) to $\mathcal{F}$, $\mathcal{F}\vee\mathcal{G}$ and $\mathcal{H}$ shows that $\mathcal{F}$ is immersed in $\mathcal{F}\vee\mathcal{G}$; similarly for $\mathcal{G}$. ∎

If a filtration $\mathcal{F}$ is included in a filtration $\mathcal{G}$, each of the following three statements is a necessary and sufficient condition for $\mathcal{F}$ to be immersed in $\mathcal{G}$:
for each t, the σ-fields $\mathcal{F}_\infty$ and $\mathcal{G}_t$ are conditionally independent given $\mathcal{F}_t$;
for each t, the operators of conditional expectation verify $\mathbb{E}^{\mathcal{F}_\infty}\mathbb{E}^{\mathcal{G}_t} = \mathbb{E}^{\mathcal{F}_t}$;
for each t, the operators of conditional expectation verify $\mathbb{E}^{\mathcal{G}_t}\mathbb{E}^{\mathcal{F}_\infty} = \mathbb{E}^{\mathcal{F}_t}$.
These three characterizations of immersion can be found in Exercise V.4.16.1° of Revuz-Yor [12]. We shall not use them directly, but in a disguised form: Lemma 5 will rephrase them in terms of $\mathcal{F}$-saturation.

DEFINITION. — *Let $(\Omega,\mathcal{A},\mathbb{P},\mathcal{F})$ be a filtered probability space. A sub-σ-field $\mathcal{B}$ of $\mathcal{A}$ is* $\mathcal{F}$-saturated *if $\mathcal{B}\subset\mathcal{F}_\infty$ and if $\mathbb{P}[B|\mathcal{F}_t]$ $(=\mathbb{E}[\mathbb{1}_B|\mathcal{F}_t])$ is $\mathcal{B}$-measurable for each $B\in\mathcal{B}$ and each time t.*

LEMMA 5. — *Let $(\Omega,\mathcal{A},\mathbb{P},\mathcal{F})$ be a filtered probability space. The map $\mathcal{E}\mapsto\mathcal{E}_\infty$ is a bijection between all filtrations immersed in $\mathcal{F}$ and all $\mathcal{F}$-saturated sub-σ-fields of $\mathcal{A}$. Its inverse is the map $\mathcal{B}\mapsto\mathcal{E}$ defined by $\mathcal{E}_t = \mathcal{B}\cap\mathcal{F}_t$.*

Consequently, the filtrations $\mathcal{E}$ immersed in $\mathcal{F}$ are characterized by their end σ-fields $\mathcal{E}_\infty$, and verify $\mathcal{E}_t = \mathcal{E}_\infty\cap\mathcal{F}_t$ for every t.

PROOF. — Let $\mathcal{E}$ be immersed in $\mathcal{F}$ and set $\mathcal{B} = \mathcal{E}_\infty$. Pick any $B\in\mathcal{B}$ and consider the $\mathcal{E}$-martingale $M_t = \mathbb{P}[B|\mathcal{E}_t]$; it is also an $\mathcal{F}$-martingale, and $\mathbb{P}[B|\mathcal{F}_t]$ is equal to $\mathbb{P}[B|\mathcal{E}_t]$, whence $\mathcal{B}$-measurable; so $\mathcal{B}$ is $\mathcal{F}$-saturated. This equality also shows that if $B\in\mathcal{B}\cap\mathcal{F}_t$, $\mathbb{P}[B|\mathcal{E}_t] = \mathbb{P}[B|\mathcal{F}_t] = \mathbb{1}_B$, whence $B\in\mathcal{E}_t$, and $\mathcal{E}_t = \mathcal{B}\cap\mathcal{F}_t$.

Conversely, starting with any $\mathcal{F}$-saturated sub-σ-field $\mathcal{B}$ of $\mathcal{A}$, define a filtration $\mathcal{E}$ included in $\mathcal{F}$ by $\mathcal{E}_t = \mathcal{B}\cap\mathcal{F}_t$. Clearly, $\mathcal{E}_\infty\subset\mathcal{B}$. For each $X\in\mathrm{L}^\infty(\mathcal{B})$, the $\mathcal{F}$-martingale $M_t = \mathbb{E}[X|\mathcal{F}_t]$ is adapted to the smaller filtration $\mathcal{E}$, so it is an $\mathcal{E}$-martingale. Consequently, noticing that $\mathbb{E}[X|\mathcal{E}_\infty] = M_\infty = \mathbb{E}[X|\mathcal{F}_\infty] = X$ (because $\mathcal{B}\subset\mathcal{F}_\infty$), one sees that $\mathcal{B}$ is included in $\mathcal{E}_\infty$, whence $\mathcal{B} = \mathcal{E}_\infty$. So M is the most general bounded $\mathcal{E}$-martingale, and $\mathcal{E}$ is immersed in $\mathcal{F}$. ∎

It is obvious that the intersection of two $\mathcal{F}$-saturated σ-fields is $\mathcal{F}$-saturated too; Lemma 5 translates this into a statement on immersed filtrations:

LEMMA 6. — *If two filtrations $\mathcal{F}$ and $\mathcal{G}$ are jointly immersed in $\mathcal{F}\vee\mathcal{G}$, their intersection $\mathcal{F}\cap\mathcal{G}$* (the filtration consisting of the intersections $\mathcal{F}_t\cap\mathcal{G}_t$) *is immersed in each of them.*

PROOF. — Put $\mathcal{H} = \mathcal{F}\vee\mathcal{G}$; by hypothesis, both $\mathcal{F}$ and $\mathcal{G}$ are immersed in $\mathcal{H}$. It suffices to show that $\mathcal{F}\cap\mathcal{G}$ is immersed in $\mathcal{H}$, and the result will follow by applying Lemma 4 a) to $\mathcal{F}\cap\mathcal{G}$, $\mathcal{F}$ and $\mathcal{H}$. By Lemma 5, the σ-fields $\mathcal{F}_\infty$ and $\mathcal{G}_\infty$ are $\mathcal{H}$-saturated; consequently, so is also $\mathcal{F}_\infty\cap\mathcal{G}_\infty$. Still by Lemma 5, the filtration $\mathcal{I}$ defined by $\mathcal{I}_t = \mathcal{F}_\infty\cap\mathcal{G}_\infty\cap\mathcal{H}_t$ is immersed in $\mathcal{H}$. Applying Lemma 5 again gives $\mathcal{F}_\infty\cap\mathcal{H}_t = \mathcal{F}_t$ and $\mathcal{G}_\infty\cap\mathcal{H}_t = \mathcal{G}_t$, wherefrom $\mathcal{I}_t = \mathcal{F}_t\cap\mathcal{G}_t$. ∎

LEMMA 7. — *Let $(\Omega,\mathcal{A},\mathbb{P},\mathcal{F})$ be a filtered probability space and $\mathcal{B}$ an $\mathcal{F}$-saturated sub-σ-field of $\mathcal{A}$. For a given t, $\mathcal{B}$ is independent of $\mathcal{F}_t$ if and only if $\mathcal{B}\cap\mathcal{F}_t$ is degenerate.*

If this holds, and if $\mathcal{C}$ is any $\mathcal{F}$-saturated σ-field included in $\mathcal{F}_t$, then $\mathcal{B}\vee\mathcal{C}$ is $\mathcal{F}$-saturated.

PROOF. — Supposing $\mathcal{B}\cap\mathcal{F}_t$ is degenerate, take $B \in \mathcal{B}$. The random variable $\mathbb{P}[B|\mathcal{F}_t]$ is measurable with respect to $\mathcal{B}$ (by saturation) and to $\mathcal{F}_t$, hence a.s. constant. Consequently, B is independent of $\mathcal{F}_t$; so $\mathcal{B}$ and $\mathcal{F}_t$ are independent. Conversely, if $\mathcal{B}$ and $\mathcal{F}_t$ are independent, $\mathcal{B}\cap\mathcal{F}_t$ is independent of itself, that is, degenerate.

The second part of the lemma is a corollary of Lemmas 3 and 5. ∎

From now on, the discussion will be restricted to filtrations indexed by the time-axis $-\mathbb{N} = \{\dots,-2,-1,0\}$: the instants of time are negative integers. (In fact, only a neighbourhood of $-\infty$ is interesting; at the cost of a few minor changes, everything extends to the case when the time-axis is $\mathbb{Z}$.) All statements seen so far on immersion and saturation are still valid, with naturally $\mathcal{F}_\infty$ being replaced by $\mathcal{F}_0$. In this situation (and more generally whenever time is discrete), there is a very simple and useful instance of immersion:

LEMMA 8. — *Let $(\Omega,\mathcal{A},\mathbb{P},\mathcal{F})$ be a filtered probability space and, for each $n \leqslant 0$, let $\mathcal{C}_n$ be a sub-σ-field of $\mathcal{F}_n$, independent of $\mathcal{F}_{n-1}$. The filtration $\mathcal{E}$ defined by $\mathcal{E}_n = \sigma(\mathcal{C}_m, m\leqslant n)$ is immersed in $\mathcal{F}$, and the σ-field $\sigma(\mathcal{C}_n, n\leqslant 0)$ is $\mathcal{F}$-saturated.*

PROOF. — Every bounded, $\mathcal{E}_0$-measurable r.v. E has the form $\phi(\dots,C_{-1},C_0)$ where each C_n is a $\mathcal{C}_n$-measurable r.v. For fixed $n \leqslant 0$, the r.v.'s $\dots,C_{n-1},C_n$ are $\mathcal{F}_n$-measurable and $(C_{n+1},\dots,C_0)$ is independent of $\mathcal{F}_n$; hence $\mathbb{E}[E|\mathcal{F}_n]$ equals $\int \phi(\dots,C_{n-1},C_n,c_{n+1},\dots,c_0)\,d\gamma_{n+1}(c_{n+1})\dots d\gamma_0(c_0)$, where γ_m is the law of C_m. As this is $\mathcal{E}_n$-measurable, $\mathbb{E}[E|\mathcal{F}_n]$ equals $\mathbb{E}[E|\mathcal{E}_n]$, showing immersion of $\mathcal{E}$ in $\mathcal{F}$ and $\mathcal{F}$-saturation of $\mathcal{E}_0$. ∎

DEFINITION. — *Two filtrations $\mathcal{F} = (\mathcal{F}_n)_{n\in-\mathbb{N}}$ and $\mathcal{G} = (\mathcal{G}_n)_{n\in-\mathbb{N}}$, defined on the same sample space $(\Omega,\mathcal{A},\mathbb{P})$, are* I-separate *if there exists an $n \in -\mathbb{N}$ such that the σ-fields $\mathcal{F}_n$ and $\mathcal{G}_n$ are independent.*

The letter I in this name stands for Independence. Later on, we shall meet other separation conditions (D-separation, H-separation, ...).

DEFINITION. — *A filtered probability space* $(\Omega, \mathcal{A}, \mathbb{P}, \mathcal{F})$, *where* $\mathcal{F} = (\mathcal{F}_n)_{n \in -\mathbb{N}}$, *is* I-cosy *if for every* $\mathcal{F}_0$*-measurable r.v.* R *and every* $\delta > 0$, *there exists a probability space* $(\bar{\Omega}, \bar{\mathcal{A}}, \bar{\mathbb{P}})$ *with two filtrations* $\mathcal{F}'$ *and* $\mathcal{F}''$ *such that*

(i) *each of* $\mathcal{F}'$ *and* $\mathcal{F}''$ *is isomorphic to* $\mathcal{F}$*;*

(ii) $\mathcal{F}'$ *and* $\mathcal{F}''$ *are jointly immersed;*

(iii) $\mathcal{F}'$ *and* $\mathcal{F}''$ *are I-separate;*

(iv) *the copies* $R' \in \mathrm{L}^0(\bar{\Omega}, \mathcal{F}'_0, \bar{\mathbb{P}})$ *and* $R'' \in \mathrm{L}^0(\bar{\Omega}, \mathcal{F}''_0, \bar{\mathbb{P}})$ *of* R *by the isomorphisms in* (i) *are* δ*-close in probability:* $\bar{\mathbb{P}}\big[|R' - R''| \geqslant \delta\big] < \delta$.

When there is no ambiguity on the underlying space $(\Omega, \mathcal{A}, \mathbb{P})$, we shall simply say that the filtration $\mathcal{F}$ is I-cosy.

If the filtration $\mathcal{F}$ were indexed by $\mathbb{Z}$ instead of $-\mathbb{N}$, the σ-fields $\mathcal{F}_0$, $\mathcal{F}'_0$ and $\mathcal{F}''_0$ in the above definition should be replaced with $\mathcal{F}_\infty$, $\mathcal{F}'_\infty$ and $\mathcal{F}''_\infty$.

The definition of I-cosiness is inspired from two sources. The first one is Tsirelson's definition of cosiness in [15]; it is the same as I-cosiness, save the separation condition (iii) (Tsirelson works in continuous time and assumes that all martingales are continuous; in this framework, his separation condition is the existence of a constant $\rho < 1$ such that, for each $\mathcal{F}'$-martingale M' and each $\mathcal{F}''$-martingale M'', $[M', M'']_t \leqslant \rho\, [M', M']_t^{1/2}\, [M'', M'']_t^{1/2}$). The other source is a proof of non-standardness by Smorodinsky [14], who implicitly uses I-cosiness, without giving it an explicit name.

Observe that I-cosiness is invariant by isomorphisms: two isomorphic filtrations are either both I-cosy, or both non I-cosy.

PROPOSITION 1. — *A filtration immersed in an I-cosy filtration is itself I-cosy.*

PROOF. — If $(\Omega, \mathcal{A}, \mathbb{P})$ is endowed with two filtrations $\mathcal{F}$ and $\mathcal{G}$, if $\mathcal{F}$ is immersed in $\mathcal{G}$ and if Ψ is an embedding of $(\Omega, \mathcal{G}_0, \mathbb{P})$ into some probability space, then the filtration $\Psi(\mathcal{F})$ is immersed in $\Psi(\mathcal{G})$. The proposition follows immediately from this remark, the definition of I-cosiness and the transitivity of immersions. ∎

DEFINITION. — *Let* $\mathcal{F}$ *and* $\mathcal{G}$ *be two filtrations, not necessarily on the same probability space. The filtration* $\mathcal{F}$ *is* immersible *into* $\mathcal{G}$ *if there exists a filtration immersed in* $\mathcal{G}$ *and isomorphic to* $\mathcal{F}$.

Using the invariance of I-cosiness under isomorphisms, Proposition 1 immediately bootstraps into a stronger result:

COROLLARY 1. — *A filtration immersible into an I-cosy filtration is itself I-cosy.*

PROPOSITION 2. — *I-cosiness is stable by taking subsequences: Let* $\mathcal{F} = (\mathcal{F}_n)_{n \in -\mathbb{N}}$ *be an I-cosy filtration and* $\sigma : -\mathbb{N} \to -\mathbb{N}$ *a strictly increasing map. The filtration* $\mathcal{G} = (\mathcal{G}_n)_{n \in -\mathbb{N}}$ *defined by* $\mathcal{G}_n = \mathcal{F}_{\sigma(n)}$ *is I-cosy too.*

PROOF. — Let R be a $\mathcal{G}_0$-measurable r.v.; it is also $\mathcal{F}_0$-measurable, so for $\delta > 0$ we have two isomorphic copies $\mathcal{F}'$ and $\mathcal{F}''$ of $\mathcal{F}$, jointly immersed, I-separate, and verifying condition (iv). Put $\mathcal{G}'_n = \mathcal{F}'_{\sigma(n)}$, $\mathcal{G}''_n = \mathcal{F}''_{\sigma(n)}$ and $\mathcal{H} = \mathcal{G}' \vee \mathcal{G}''$. Plainly, the filtrations $\mathcal{G}'$ and $\mathcal{G}''$ are isomorphic to $\mathcal{G}$, immersed in $\mathcal{H}$, and I-separate. ∎

LEMMA 9. — *Let (E,ρ) be a separable metric space, and $\mathcal{F}$ an I-cosy filtration. The property defining I-cosiness still holds when the random variables R are taken E-valued (with $\rho(R',R'')$ replacing $|R'-R''|$).*

PROOF. — Approximating in probability R by a simple r.v., we may suppose that R takes only finitely many values $x_1,\dots,x_p$ in E. It then suffices to apply the definition of I-cosiness to the $\{1,\dots,p\}$-valued r.v. S defined by $R=x_S$ and to $\delta'=\delta\wedge\frac{1}{2}\inf_{i\neq j}\rho(x_i,x_j)$. ∎

DEFINITION. — *A filtration $\mathcal{F}=(\mathcal{F}_n)_{n\in-\mathbb{N}}$ is of* product type *if there exists an independent sequence $(\mathcal{C}_n)_{n\in-\mathbb{N}}$ of sub-σ-fields of $\mathcal{A}$ such that $\mathcal{F}_n=\sigma(\mathcal{C}_m,\ m\leqslant n)$ for each $n\leqslant 0$.*

This definition is borrowed from Feldman [8] and Feldman-Smorodinsky [9], who define a *filtered probability space* $(\Omega,\mathcal{A},\mathbb{P},\mathcal{F})$ to be of product type when it satisfies the above property, thus stressing the role of the measure $\mathbb{P}$. But, as we already do for isomorphisms and immersions, we shall simply speak of *filtrations* of product type, keeping in mind that this notion is not invariant under changes of measure.

Notice that if $\mathcal{F}$ is of product type, the σ-fields $\mathcal{C}_n$ in the preceding definition are in general not uniquely determined. Consider for instance the natural filtration of a process $(\varepsilon_n)_{n\leqslant 0}$ made of i.i.d. r.v.'s uniform on $\{-1,1\}$. This filtration is of product type, with $\mathcal{C}_n=\sigma(\varepsilon_n)$; replacing $\mathcal{C}_0$ by $\sigma(\varepsilon_{-1}\varepsilon_0)$ yields another family of σ-fields with the same property.

Of course, if $(Y_n)_{n\in-\mathbb{N}}$ is an independent sequence of random variables, the filtration generated by the process Y is of product type. Conversely, every filtration $\mathcal{F}$ of product type and such that $\mathcal{F}_0$ is essentially separable, is the natural filtration of such an independent process.

PROPOSITION 3. — *Every filtration of product type is I-cosy.*

PROOF. — Let $\mathcal{F}$ be of product type: there exists an independent sequence $(\mathcal{C}_n)_{n\in-\mathbb{N}}$ of sub-σ-fields of $\mathcal{A}$ such that $\mathcal{F}_n=\sigma(\dots,\mathcal{C}_{n-1},\mathcal{C}_n)$. Fix $R\in\mathrm{L}^0(\mathcal{F}_0)$ and $\delta>0$. Remark that the σ-fields $\mathcal{B}_n=\sigma(\mathcal{C}_{n+1},\mathcal{C}_{n+2},\dots,\mathcal{C}_0)$ form a monotone sequence $(\mathcal{B}_n)_{n\in-\mathbb{N}}$ of sub-σ-fields of $\mathcal{A}$, with limit $\bigvee_n\mathcal{B}_n=\mathcal{F}_0$ when $n\to-\infty$. By Doob's direct martingale convergence theorem, there exist an $n<0$ and a r.v. $S\in\mathrm{L}^0(\mathcal{B}_n)$ such that S is δ-close to R in probability. Fix these n and S.

On a suitable sample space $(\bar\Omega,\bar{\mathcal{A}},\bar{\mathbb{P}})$, for instance the product $(\Omega\times\Omega,\mathcal{A}\otimes\mathcal{A},\mathbb{P}\times\mathbb{P})$, there exist two independent sub-σ-fields $\mathcal{A}^1$ and $\mathcal{A}^2$ of $\bar{\mathcal{A}}$ such that both $(\bar\Omega,\mathcal{A}^1,\bar{\mathbb{P}})$ and $(\bar\Omega,\mathcal{A}^2,\bar{\mathbb{P}})$ are isomorphic to $(\Omega,\mathcal{A},\mathbb{P})$; call Ψ^1 and Ψ^2 the isomorphisms: $\mathcal{A}^i=\Psi^i(\mathcal{A})$ for $i\in\{1,2\}$, and put $\mathcal{C}^i_m=\Psi^i(\mathcal{C}_m)$. Define three filtrations $\mathcal{F}'$, $\mathcal{F}''$ and $\mathcal{G}$ on $(\bar\Omega,\bar{\mathcal{A}},\bar{\mathbb{P}})$ as follows:

$$
\begin{aligned}
\text{for every } m\leqslant 0,\quad & \mathcal{F}'_m=\sigma(\dots,\mathcal{C}^1_{m-1},\mathcal{C}^1_m)\ ;\\
\text{for every } m\leqslant 0,\quad & \mathcal{G}_m=\sigma(\dots,\mathcal{C}^1_{m-1},\mathcal{C}^2_{m-1},\mathcal{C}^1_m,\mathcal{C}^2_m)\ ;\\
\text{for } m\leqslant n,\quad & \mathcal{F}''_m=\sigma(\dots,\mathcal{C}^2_{m-1},\mathcal{C}^2_m)\\
\text{and for } n<m\leqslant 0,\quad & \mathcal{F}''_m=\sigma(\dots,\mathcal{C}^2_{n-1},\mathcal{C}^2_n,\mathcal{C}^1_{n+1},\dots,\mathcal{C}^1_{m-1},\mathcal{C}^1_m)\ .
\end{aligned}
$$

To show that $\mathcal{F}$ is I-cosy, we shall check that $\mathcal{F}'$ and $\mathcal{F}''$ verify the four conditions in the definition of I-cosiness.

(i) The restriction Ψ' of Ψ^1 to $\mathcal{F}_0$ is an isomorphism from $\mathcal{F}$ to $\mathcal{F}'$. An isomorphism Ψ'' between $\mathcal{F}$ and $\mathcal{F}''$ is given by the following algorithm: If C_m are $\mathcal{C}_m$-measurable r.v.'s and if ϕ is a Borel function, put

$$\begin{aligned}\Psi''\big(\phi(\ldots,C_{n-1},C_n,C_{n+1},\ldots,C_0)\big)\\ = \phi\big(\ldots,\Psi^2(C_{n-1}),\Psi^2(C_n),\Psi^1(C_{n+1}),\ldots,\Psi^1(C_0)\big)\ .\end{aligned}$$

(ii) $\mathcal{F}'$ and $\mathcal{F}''$ are immersed in $\mathcal{G}$ by Lemma 1, because $\mathcal{G}$ is an independent enlargement of each of them.

(iii) The σ-fields $\mathcal{F}'_n$ and $\mathcal{F}''_n$ are independent because they are respectively included in $\mathcal{A}^1$ and $\mathcal{A}^2$.

(iv) Put $R' = \Psi'(R)$, $R'' = \Psi''(R)$, $S' = \Psi'(S)$ and $S'' = \Psi''(S)$. By isomorphic transfer, R' and S' (respectively R'' and S'') are δ-close in probability. Owing to the definitions of Ψ' and Ψ'', these isomorphisms have the same restriction to $\mathcal{B}_n$; so $S' = \Psi'(S) = \Psi''(S) = S''$. Consequently, R' and R'' are 2δ-close in probability. ∎

COROLLARY 2. — *Any filtration immersible into a filtration of product type is I-cosy.*

PROOF. — Immediate from Proposition 3 and Corollary 1. ∎

As we shall now see, the converse is also true. We shall get it as a straightforward consequence of Vershik's criterion, more precisely of condition 3 in Theorem 3.2 of Vershik [17] (it becomes condition (vii) in our Theorem 2 below). The next three sections are devoted to this topic; the reader already familiar with Vershik's theory can skip these sections and jump directly to the (short and easy) Proposition 5.

Vershik's standardness criterion: Preliminary notions

Vershik's work on filtrations is written in Rohlin's language [13], where the idea of conditioning with respect to a sub-σ-field is expressed by quotienting the probability space. Sticking to a vocabulary more familiar to probabilists (at least, to us), the next proposition recalls what happens when the factor space is diffuse (that is, all equivalence classes in the quotient are isomorphic to the Lebesgue space $[0,1]$).

PROPOSITION 4 AND DEFINITION. — *Given a sample space $(\Omega,\mathcal{A},\mathbb{P})$, let $\mathcal{B}$ be an essentially separable sub-σ-field of $\mathcal{A}$ and $\mathcal{C}$ a sub-σ-field of $\mathcal{B}$. The following four conditions are equivalent:*

(i) *there exists a $\mathcal{B}$-measurable random variable Y such that, for every $\mathcal{C}$-measurable random variable Z, $\mathbb{P}[Y=Z]=0$;*

(ii) *there exists a $\mathcal{B}$-measurable random variable, independent of $\mathcal{C}$ and having a diffuse law;*

(iii) *there exists a $\mathcal{B}$-measurable random variable X, independent of $\mathcal{C}$, with uniform law on $[0,1]$, and such that $\mathcal{C}\vee\sigma(X)=\mathcal{B}$;*

(iv) *every random variable V that verifies $\mathcal{C}\vee\sigma(V)=\mathcal{B}$, has a diffuse law.*

When these hypotheses and conditions are met, we shall say that $\mathcal{B}$ is conditionally non-atomic given $\mathcal{C}$, *and every X satisfying condition* (iii) *will be called a* complement to $\mathcal{C}$ in $\mathcal{B}$.

PROOF. — Trivially, (iii) $\Rightarrow$ (ii).

(ii) $\Rightarrow$ (i). Let Y be $\mathcal{B}$-measurable, independent of $\mathcal{C}$, and with diffuse law η. For any $\mathcal{C}$-measurable Z, calling ζ the law of Z and using the fact that η is diffuse, one has $\mathbb{P}[Y=Z] = \iint \mathbb{1}_{\{y=z\}}\,\eta(dy)\,\zeta(dz) = \int \eta(\{z\})\,\zeta(dz) = 0$.

(i) $\Rightarrow$ (iii). Let Y be a $\mathcal{B}$-measurable random variable such that $\mathbb{P}[Y=Z] = 0$ for every $\mathcal{C}$-measurable Z. The essentially separable σ-field $\mathcal{B}$ is generated by some random variable B; its sub-σ-field $\mathcal{C}$ is essentially separable too and is generated by some C. Call γ the law of C, and $(\beta_c)_{c\in\mathbb{R}}$ a regular version of the law of B given C: β_c is a probability well-defined for γ-almost every c, depending measurably on c, and the joint law of (C,B) is $\gamma(dc)\,\beta_c(db)$.

Remark that Y has the form $y{\circ}B$ for some measurable function y; consequently, for any $\mathcal{C}$-measurable Z, $\mathbb{P}[B=Z] \leqslant \mathbb{P}[Y=y{\circ}Z] = 0$ by the choice of Y. This implies that almost all probabilities β_c are diffuse; indeed, for $\varepsilon > 0$, call $a(c)$ the smallest (i.e., leftmost) atom of β_c with mass at least ε, or $+\infty$ if there is no such atom; as $\mathbb{P}[B=a{\circ}C] = 0$, for almost every c, β_c has no atom weighing at least ε.

For each c, call F_c the distribution function of β_c, given by $F_c(x) = \beta_c\big((-\infty,x)\big)$; as β_c has no atoms, for every $t\in[0,1]$, β_c gives mass t to the interval $F_c^{-1}\big([0,t]\big)$. Define a $\mathcal{B}$-measurable, $[0,1]$-valued random variable X by $X = F_C{\circ}B$.

For every bounded, measurable function ϕ and every $t\in[0,1]$, one has

$$\begin{aligned}\mathbb{E}[\phi{\circ}C\ \mathbb{1}_{\{X\leqslant t\}}] = \mathbb{E}[\phi{\circ}C\ \mathbb{1}_{\{F_C\circ B\leqslant t\}}] &= \iint \phi(c)\ \mathbb{1}_{\{F_c(b)\leqslant t\}}\ \gamma(dc)\,\beta_c(db)\\ &= t\int \phi(c)\,\gamma(dc) = t\,\mathbb{E}[\phi{\circ}C]\ ;\end{aligned}$$

thus X is independent of C (of $\mathcal{C}$) and uniformly distributed on $[0,1]$. Calling G_c the right-continuous inverse of F_c, defined on $[0,1]$, one almost surely has $B = G_C{\circ}X$, so B is $\mathcal{C}{\vee}\sigma(X)$-measurable and $\mathcal{B} = \mathcal{C}\vee\sigma(X)$.

(i) $\Rightarrow$ (iv). Suppose Y verifies (i) and C generates $\mathcal{C}$; let V be such that $\mathcal{C}\vee\sigma(V) = \mathcal{B}$. For some Borel f, one has $Y = f(C,V)$, and for each $v\in\mathbb{R}$, one can write $\mathbb{P}[V{=}v] \leqslant \mathbb{P}[Y=f(C,v)] = 0$.

(iv) $\Rightarrow$ (i). Assuming (iv), let B be a r.v. generating $\mathcal{B}$ and such that $B\neq 0$ a.s. If Z is any $\mathcal{C}$-measurable r.v., $V = B\,\mathbb{1}_{\{B\neq Z\}}$ is $\mathcal{B}$-measurable and one has $\{V=0\} = \{B=Z\}$, whence $B = V + Z\,\mathbb{1}_{\{V=0\}}$; so $\mathcal{C}\vee\sigma(V) = \mathcal{B}$. By (iv), V must be diffuse, wherefrom $\mathbb{P}[B=Z] = \mathbb{P}[V=0] = 0$, and B fulfills (i). ∎

COROLLARY 3. — *Let $\mathcal{B}$, $\mathcal{C}$ and $\mathcal{D}$ be three sub-σ-fields in an essentially separable sample space $(\Omega,\mathcal{A},\mathbb{P})$; assume $\mathcal{C}$ is included in $\mathcal{B}$.*

a) *If $\mathcal{B}{\vee}\mathcal{D}$ is conditionally non-atomic given $\mathcal{C}{\vee}\mathcal{D}$, then $\mathcal{B}$ is conditionally non-atomic given $\mathcal{C}$.*

b) *If $\mathcal{B}$ and $\mathcal{D}$ are independent, and if $\mathcal{B}$ is conditionally non-atomic given $\mathcal{C}$, then $\mathcal{B}{\vee}\mathcal{D}$ is conditionally non-atomic given $\mathcal{C}{\vee}\mathcal{D}$; moreover, every complement of $\mathcal{C}$ in $\mathcal{B}$ is also a complement of $\mathcal{C}{\vee}\mathcal{D}$ in $\mathcal{B}{\vee}\mathcal{D}$.*

PROOF. — a) If $\mathcal{B}{\vee}\mathcal{D}$ is conditionally non-atomic given $\mathcal{C}{\vee}\mathcal{D}$, for any V such that $\mathcal{C}{\vee}\sigma(V) = \mathcal{B}$, one also has $\mathcal{C}{\vee}\mathcal{D}{\vee}\sigma(V) = \mathcal{B}{\vee}\mathcal{D}$, so V is diffuse by condition (iv), and $\mathcal{B}$ is conditionally non-atomic given $\mathcal{C}$ by the same condition.

b) If $\mathcal{B}$ and $\mathcal{D}$ are independent and if $\mathcal{B}$ is conditionally non-atomic given $\mathcal{C}$, every complement X to $\mathcal{C}$ in $\mathcal{B}$ is independent of $\mathcal{C}\vee\mathcal{D}$ (because X is independent of $\mathcal{C}$, and $\mathcal{D}$ is independent of $\mathcal{C}\vee\sigma(X)$); as $\mathcal{C}\vee\sigma(X)\vee\mathcal{D} = \mathcal{B}\vee\mathcal{D}$, X is also a complement to $\mathcal{C}\vee\mathcal{D}$ in $\mathcal{B}\vee\mathcal{D}$. ∎

REMARK. — At this stage, it may be useful to warn the reader against two pitfalls:

a) *If $\mathcal{B}$ is conditionally non-atomic given $\mathcal{C}$ and if Y is $\mathcal{B}$-measurable, with diffuse law, and independent of $\mathcal{C}$, there may exist no complement X to $\mathcal{C}$ in $\mathcal{B}$ such that $\sigma(X) \supset \sigma(Y)$.* Consider for instance three independent random variables C, Y and Z with uniform law on $[0,1]$. Call Γ the event $\{C<\frac{1}{2}\}$, define $\mathcal{C}$ as the σ-field $\sigma(C)$ and $\mathcal{B}$ as the σ-field $\sigma(C,Y,Z\mathbb{1}_\Gamma)$. Since Y satisfies condition (ii) of Proposition 4, $\mathcal{B}$ is conditionally non-atomic given $\mathcal{C}$. Let X be any $\mathcal{B}$-measurable r.v. independent of $\mathcal{C}$ and such that $\sigma(X) \supset \sigma(Y)$. We shall show $\sigma(X) = \sigma(Y)$, thus preventing $\mathcal{C}\vee\sigma(X)$ to equal $\mathcal{B}$ (for clearly, $Z\mathbb{1}_\Gamma$ is not $\mathcal{C}\vee\sigma(Y)$-measurable). There are two Borel functions f and g such that $Y = f(X)$ and $X = g(C,Y,Z\mathbb{1}_\Gamma)$ a.s. Putting $h(c,y) = g(c,y,0)$, one has $X = h\big(C,f(X)\big)$ a.s. on Γ^c. Consequently, for almost every $c \geqslant \frac{1}{2}$, one has $x = h\big(c,f(x)\big)$ for almost all x. Fix such a c and put $k(y) = h(c,y)$; equality $x = k\circ f(x)$ holds for almost all x, whence $X = k\circ f(X)$ a.s., giving $X = k\circ Y$ a.s., and $\sigma(X) = \sigma(Y)$.

b) *If $\mathcal{B}$ is conditionally non-atomic given $\mathcal{C}$ and if $\mathcal{D}$ is such that $\mathcal{C}\vee\mathcal{D} = \mathcal{B}$, there may exist no $\mathcal{D}$-measurable complement to $\mathcal{C}$ in $\mathcal{B}$.* Take for instance Ω equal to the union of the three rectangles $[0,\frac{1}{2}]\times[0,1]$, $[\frac{1}{2},1]\times[0,\frac{1}{2}]$ and $[1,\frac{3}{2}]\times[\frac{1}{2},1]$, endowed with the restriction of the Lebesgue measure. Call C the first coordinate (it is $[0,\frac{3}{2}]$-valued) and D the second one (it is $[0,1]$-valued); define $\mathcal{C}$ as $\sigma(C)$, $\mathcal{D}$ as $\sigma(D)$ and $\mathcal{B}$ as $\sigma(C,D)$. Since $|D-\frac{1}{2}|$ is independent of C, $\mathcal{B}$ is conditionally non-atomic given $\mathcal{C}$. Let X be any $\mathcal{D}$-measurable random variable such that $\mathcal{C}\vee\sigma(X) = \mathcal{B}$. We shall show $\sigma(X) = \mathcal{D}$, thus preventing X to be independent of $\mathcal{C}$. There are two Borel functions f and g such that $X = f(D)$ and $D = g(C,X)$ a.s. Consequently, $D = g\big(C,f(D)\big)$ a.s. and for almost every $c < \frac{1}{2}$, one has $d = g\big(c,f(d)\big)$ for almost all d in $[0,1]$. Fix such a c and put $h(x) = g(c,x)$; equality $d = h\circ f(d)$ holds for almost all $d \in [0,1]$, whence $D = h\circ f(D)$ a.s., giving $D = h\circ X$ a.s., and $\sigma(X) = \mathcal{D}$.

In counter-example a) above, Y had a diffuse law. If, on the opposite, Y is discrete, a complement X always exists, as can easily be seen:

COROLLARY 4. — *Let $\mathcal{B}$ and $\mathcal{C}$ be two sub-σ-fields of $\mathcal{A}$, with $\mathcal{C} \subset \mathcal{B}$ and $\mathcal{B}$ conditionally non-atomic given $\mathcal{C}$.*

If Y is a $\mathcal{B}$-measurable random variable taking finitely or countably many values, $\mathcal{B}$ is conditionally non-atomic given $\mathcal{C}\vee\sigma(Y)$.

If furthermore Y is independent of $\mathcal{C}$, there exists a complement X to $\mathcal{C}$ in $\mathcal{B}$ such that $\sigma(X) \supset \sigma(Y)$.

PROOF. — For every random variable V such that $\mathcal{C}\vee\sigma(Y)\vee\sigma(V) = B$, the pair (Y,V) has a diffuse law by Proposition 4 (iv). Since Y is discrete, this implies $\mathbb{P}[V{=}v] = \sum_y \mathbb{P}[V{=}v, Y{=}y] = 0$, so V is diffuse, and $\mathcal{B}$ is conditionally non-atomic given $\mathcal{C}\vee\sigma(Y)$ by the same condition (iv).

If Y is independent of $\mathcal{C}$, let X' denote any complement to $\mathcal{C}\vee\sigma(Y)$ in $\mathcal{B}$, and X a random variable generating $\sigma(Y,X')$; notice that $\sigma(X) \supset \sigma(Y)$ and that X is diffuse because X' is. So we may choose X uniform on $[0,1]$. Since $\mathcal{C}$, Y and X' are independent and $\mathcal{C}\vee\sigma(Y)\vee\sigma(X') = B$, X is a complement to $\mathcal{C}$ in $\mathcal{B}$. ∎

Vershik's standardness criterion: First level

Vershik's theory of standard filtrations is a two-storied building. The first floor, which we now shall enter, gives a necessary and sufficient condition for a filtration to be standard, in terms of $\mathcal{F}$-saturation. Theorem 1 below translates into our language the equivalence between 1 and 2 in Vershik's Theorem 3.2 of [17]. (His condition 3 uses a more sophisticated tool; this is postponed to the next section.) Besides changing the language, another difference is that, in this section, we deal with the non-atomic case only, whereas his statement covers the atomic case as well. Indeed, for the sake of simplification, in [17] some statements are given in the atomic case only; the general case has recently been written in full details by J. Feldman [8]; see also J. Feldman and M. Smorodinsky [9].

DEFINITIONS. — *A filtration $(\mathcal{F}_n)_{n\leqslant 0}$ is* non-atomic *if $\mathcal{F}_0$ is essentially separable and if, for each $n \leqslant 0$, $\mathcal{F}_n$ is conditionally non-atomic given $\mathcal{F}_{n-1}$.*

A filtration $(\mathcal{F}_n)_{n\leqslant 0}$ is standard non-atomic *if it is generated by a process $(X_n)_{n\leqslant 0}$, where X_n are independent random variables with uniform law on $[0,1]$.*

These definitions are compatible: A standard non-atomic filtration is non-atomic.

The important point in the latter definition is that the r.v.'s X_n are independent, with diffuse laws; that these laws can be chosen uniform on $[0,1]$ is irrelevant, but shows that all standard non-atomic filtrations are isomorphic to each other. Clearly, a filtration isomorphic to a standard non-atomic filtration is standard non-atomic too.

THEOREM 1 (Vershik [17]). — *Let $\mathcal{F} = (\mathcal{F}_n)_{n\leqslant 0}$ be a non-atomic filtration on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The following four statements are equivalent:*

(i) *$\mathcal{F}$ is standard non-atomic;*

(ii) *$\mathcal{F}$ is of product type;*

(iii) *the tail σ-field $\mathcal{F}_{-\infty}$ is degenerate; and, for every $\mathcal{F}_0$-measurable random variable R and every $\delta > 0$, there exist an essentially finite, $\mathcal{F}$-saturated σ-field $\mathcal{B}$ and a $\mathcal{B}$-measurable random variable S, such that $\mathbb{P}\big[|R-S| \geqslant \delta\big] < \delta$;*

(iv) *For every $\delta > 0$, $n \leqslant 0$ and every simple, $\mathcal{F}_n$-measurable r.v. R, there exist $m < n$, an $\mathcal{F}$-saturated $\mathcal{B}$ independent of $\mathcal{F}_m$ verifying $\mathcal{F}_m\vee\mathcal{B} = \mathcal{F}_0$, and a $\mathcal{B}\cap\mathcal{F}_n$-measurable r.v. S such that $\mathbb{P}[S \neq R] < \delta$.*

The gist of the equivalence (i) $\Leftrightarrow$ (iv) in this theorem can be better understood with the help of the following lemma:

LEMMA. — *Assume $\mathcal{F}$ is non-atomic and fix $m \leqslant 0$. If an $\mathcal{F}$-saturated σ-field $\mathcal{B}$ is independent of $\mathcal{F}_m$ and if $\mathcal{F}_m\vee\mathcal{B} = \mathcal{F}_0$, there exist $X_{m+1}, \ldots, X_0$ such that each X_ℓ is a complement to $\mathcal{F}_{\ell-1}$ in $\mathcal{F}_\ell$ and $\sigma(X_{m+1}, \ldots, X_0) = \mathcal{B}$.*

We mention this lemma only for the light it sheds on the theorem; we shall neither use it nor prove it. In the second paragraph of (iv) $\Rightarrow$ (iv$'$) in the proof of the theorem, we shall need and establish a slightly stronger property. (That paragraph contains a parameter n; when $n = 0$ it reduces to a proof of the lemma.)

PROOF OF THEOREM 1. — (i) $\Rightarrow$ (ii) is trivial: calling $\mathcal{C}_n$ the σ-field generated by X_n, one has $\mathcal{F}_n = \sigma(\ldots, \mathcal{C}_{n-1}, \mathcal{C}_n)$.

(ii) $\Rightarrow$ (iii). Hypothesis (ii) says that $\mathcal{F}_n = \sigma(\ldots, \mathcal{C}_{n-1}, \mathcal{C}_n)$, where the $\mathcal{C}_n$ are independent; the degeneracy of $\mathcal{F}_{-\infty}$ follows by Kolmogorov's zero-one law. As $\mathcal{F}_0$ is essentially separable, so is also each $\mathcal{C}_n$; hence there exists for each n an increasing sequence $(\mathcal{C}_n^j)_{j\in\mathbb{N}}$ of essentially finite sub-σ-fields such that $\mathcal{C}_n = \bigvee_{j\geqslant 0} \mathcal{C}_n^j$. For each $j \geqslant 0$, put

$$\mathcal{B}^j = \mathcal{C}_{-j}^j \vee \mathcal{C}_{-j+1}^j \vee \ldots \vee \mathcal{C}_0^j \;;$$

this is an $\mathcal{F}$-saturated σ-field by Lemmas 8 and 5. The σ-fields $\mathcal{B}^j$ form an increasing sequence whose limit $\sigma(\mathcal{B}^j, j \geqslant 0)$ contains every $\mathcal{C}_n$; consequently $\sigma(\mathcal{B}^j, j \geqslant 0) = \mathcal{F}_0$, and $\bigcup_{j\geqslant 0} \mathrm{L}^0(\mathcal{B}^j)$ is dense in $\mathrm{L}^0(\mathcal{F}_0)$.

The proof of (iii) $\Rightarrow$ (iv) will be made clearer by breaking it into two smaller steps. We shall introduce a new condition (iii′), and establish (iii) $\Rightarrow$ (iii′) $\Rightarrow$ (iv). Here is this intermediate statement (the letters SC stand for 'saturated complement'):

(iii′) *Call* $\mathrm{SC}(m, \mathcal{F})$ *the set of all* $\mathcal{F}$*-saturated sub-*σ*-fields* $\mathcal{B}$ *of* $\mathcal{F}_0$ *that are independent of* $\mathcal{F}_m$ *and verify* $\mathcal{F}_m \vee \mathcal{B} = \mathcal{F}_0$*. For every* $\mathcal{F}_0$*-measurable* R *and every* $n \leqslant 0$*, there exist an* $m < n$*, a* $\mathcal{B} \in \mathrm{SC}(m, \mathcal{F})$ *and a* $\mathcal{B}$*-measurable* S *such that* $\mathbb{P}\big[|R-S| \geqslant \delta\big] < \delta$.

(iii) $\Rightarrow$ (iii′). This is a straightforward consequence of the following fact: *If* $\mathcal{F}$ *is non-atomic and if* $\mathcal{F}_{-\infty}$ *is degenerate, for every* $n \leqslant 0$ *and every essentially finite,* $\mathcal{F}$*-saturated* $\mathcal{B}$*, there exist an* $m < n$ *and a* $\mathcal{C} \in \mathrm{CS}(m, \mathcal{F})$ *such that* $\mathcal{C} \supset \mathcal{B}$. To establish this claim, remark first that if B is an event such that $0 < \mathbb{P}[B] < 1$, the degeneracy hypothesis implies $B \notin \mathcal{F}_m$ for m small enough. Since $\mathcal{B}$ contains only finitely many events (modulo negligibility), there is an m (fixed in the sequel) smaller than n, such that $B \notin \mathcal{F}_m$ for every $B \in \mathcal{B}$ verifying $0 < \mathbb{P}[B] < 1$. So $\mathcal{B} \cap \mathcal{F}_m$ is degenerate, and, by Lemma 7, $\mathcal{B}$ is independent of $\mathcal{F}_m$.

By Lemma 5, the filtration defined by $\mathcal{D}_\ell = \mathcal{B} \cap \mathcal{F}_\ell$ is immersed in $\mathcal{F}$; $\mathcal{D}_\ell$ is degenerate for $\ell \leqslant m$. Noticing that each $\mathcal{D}_\ell$ is essentially finite, Corollary 4 asserts that $\mathcal{F}_\ell$ is conditionally non-atomic given $\mathcal{F}_{\ell-1} \vee \mathcal{D}_\ell$; call Z_ℓ a complement to $\mathcal{F}_{\ell-1} \vee \mathcal{D}_\ell$ in $\mathcal{F}_\ell$, and define a filtration $\mathcal{E}$ by taking $\mathcal{E}_\ell$ degenerate if $\ell \leqslant m$ and $\mathcal{E}_\ell = \mathcal{D}_\ell \vee \sigma(Z_{m+1}, \ldots, Z_\ell)$ if $\ell > m$. The claim will be shown with $\mathcal{C} = \mathcal{E}_0$.

First, $\mathcal{F}_m \vee \mathcal{E}_\ell \supset \mathcal{F}_\ell$ for each $\ell \geqslant m$: This inclusion is trivial for $\ell = m$, and if it holds for ℓ, then $\mathcal{F}_m \vee \mathcal{E}_{\ell+1} \supset \mathcal{F}_m \vee \mathcal{E}_\ell \vee \mathcal{D}_{\ell+1} \vee \sigma(Z_{\ell+1}) \supset \mathcal{F}_\ell \vee \mathcal{D}_{\ell+1} \vee \sigma(Z_{\ell+1}) = \mathcal{F}_{\ell+1}$. So $\mathcal{F}_m \vee \mathcal{C} = \mathcal{F}_m \vee \mathcal{E}_0 = \mathcal{F}_0$.

Second, $\mathcal{C} = \mathcal{E}_0$ is $\mathcal{F}$-saturated. By Lemma 5, it suffices to show that $\mathcal{E}$ is immersed in $\mathcal{F}$. For every $\ell \leqslant 0$ and every bounded, $\mathcal{E}_\ell$-measurable U, we have to show that $\mathbb{E}[U|\mathcal{F}_{\ell-1}]$ is $\mathcal{E}_{\ell-1}$-measurable. When $\ell \leqslant m$, U is deterministic; so we may suppose $\ell > m$. Without loss of generality, we may also suppose that U is a product $WV_\ell D_\ell$, where $W \in \mathrm{L}^\infty\big(\sigma(Z_{m+1}, \ldots, Z_{\ell-1})\big)$, $V_\ell \in \mathrm{L}^\infty\big(\sigma(Z_\ell)\big)$ and $D_\ell \in \mathrm{L}^\infty(\mathcal{D}_\ell)$. Taking W out of the conditional expectation, it remains to show that $\mathbb{E}[V_\ell D_\ell|\mathcal{F}_{\ell-1}]$ is $\mathcal{E}_{\ell-1}$-measurable. We may replace V_ℓ by $\mathbb{E}[V_\ell|\mathcal{F}_{\ell-1} \vee \mathcal{D}_\ell]$; but V_ℓ is independent of $\mathcal{F}_{\ell-1} \vee \mathcal{D}_\ell$ by definition of Z_ℓ; so $\mathbb{E}[V_\ell|\mathcal{F}_{\ell-1} \vee \mathcal{D}_\ell]$ is a constant, and we are left with $\mathbb{E}[D_\ell|\mathcal{F}_{\ell-1}]$. This is $\mathcal{D}_{\ell-1}$-measurable because $\mathcal{D}$ is immersed in $\mathcal{F}$.

Last, by Lemma 5, $\mathcal{C}\cap\mathcal{F}_m = \mathcal{E}_0\cap\mathcal{F}_m = \mathcal{E}_m$ is degenerate, and $\mathcal{C}$ is independent of $\mathcal{F}_m$ by Lemma 7.

So $\mathcal{C}$ is in $\mathrm{SC}(m,\mathcal{F})$; as $\mathcal{C} = \mathcal{E}_0 \supset \mathcal{D}_0 = \mathcal{B}$, the proof of (iii) $\Rightarrow$ (iii′) is complete.

(iii′) $\Rightarrow$ (iv). Assuming (iii′), fix $n \leqslant 0$, $\delta > 0$ and R measurable for $\mathcal{F}_n$ and F-valued, where F is a finite subset of $\mathbb{R}$; without loss of generality, we shall take $F = \{1,\dots,p\}$. Put $\delta' = \delta/p$. Hypothesis (iii′) provides us with an $m < n$, a $\mathcal{B} \in \mathrm{SC}(m,\mathcal{F})$ and a $\mathcal{B}$-measurable r.v. T such that $\mathbb{P}\big[|T-R|\geqslant\delta'\big] < \delta'$; by replacing if necessary T with $1 \vee T \wedge p$, we may further suppose $|T-R| \leqslant p-1$. This implies $\mathbb{E}\big[|T-R|\big] \leqslant \delta' + (p-1)\,\mathbb{P}\big[|T-R|\geqslant\delta'\big] \leqslant \delta' + (p-1)\delta' = \delta$. By L^1-contractivity of conditional expectations, $T' = \mathbb{E}[T|\mathcal{F}_n]$ is also δ-close to R in L^1. Since T is $\mathcal{B}$-measurable and $\mathcal{B}$ is $\mathcal{F}$-saturated, T' is $\mathcal{B}\cap\mathcal{F}_n$-measurable.

For $x \in \mathbb{R}$, call $\psi(x)$ the point in F closest to x (take the smallest such point if there are two of them). Among all F-valued r.v.'s, $S = \psi\circ T'$ is closest to T' in L^1, whence $\mathbb{E}\big[|S-T'|\big] \leqslant \mathbb{E}\big[|T'-R|\big] \leqslant \delta$, and $\mathbb{E}\big[|R-S|\big] \leqslant 2\delta$. Since R and S are F-valued, $\mathbb{P}[R\neq S] = \mathbb{P}\big[|R-S| \geqslant 1\big] \leqslant \mathbb{E}\big[|R-S|\big] \leqslant 2\delta$; as S is $\mathcal{B}\cap\mathcal{F}_n$-measurable, (iv) is established.

The proof of (iv) $\Rightarrow$ (i) will also be sliced into two smaller steps, by introducing a new statement (iv′) and establishing (iv) $\Rightarrow$ (iv′) $\Rightarrow$ (i). This intermediate step is:

(iv′) *Suppose given $n < 0$ and $X_{n+1},\dots,X_0$ such that each X_ℓ is a complement to $\mathcal{F}_{\ell-1}$ in $\mathcal{F}_\ell$. For every $R \in \mathrm{L}^0(\mathcal{F}_0)$ and $\delta > 0$, there exist some $m < n$, some $X_{m+1},\dots,X_n$ with the same property (each X_ℓ is a complement to $\mathcal{F}_{\ell-1}$ in $\mathcal{F}_\ell$) and some r.v. $S \in \mathrm{L}^0\big(\sigma(X_{m+1},\dots,X_0)\big)$ verifying $\mathbb{P}\big[|R-S|\geqslant\delta\big] < \delta$.*

(iv) $\Rightarrow$ (iv′). Take R, δ, n and $X_{n+1},\dots,X_0$ as in (iv′); by the assumption on the X_ℓ, $\mathcal{F}_0$ is equal to $\mathcal{F}_n \vee \sigma(X_{n+1},\dots,X_0)$. Writing $\mathcal{F}_n$ as the limit of an increasing sequence of essentially finite sub-σ-fields, one can δ-approximate R by a r.v. of the form $\phi(T,X_{n+1},\dots,X_0)$, where ϕ is Borel and T is $\mathcal{F}_n$-measurable and simple. Applying (iv) to T, we obtain an $m < n$, a $\mathcal{B} \in \mathrm{SC}(m,\mathcal{F})$ and a $\mathcal{B}\cap\mathcal{F}_n$-measurable S verifying $\mathbb{P}[S\neq T] < \delta$. This gives $\mathbb{P}\big[\phi(S,X_{n+1},\dots,X_0)\neq\phi(T,X_{n+1},\dots,X_0)\big] < \delta$, and R is 2δ-close in probability to $\phi(S,X_{n+1},\dots,X_0)$.

The filtration $\mathcal{E}_\ell = \mathcal{B}\cap\mathcal{F}_\ell$ associated to $\mathcal{B}$ by Lemma 5 has the following properties: $\mathcal{E}$ is immersed in $\mathcal{F}$, S is $\mathcal{E}_n$-measurable, and $\mathcal{F}_m\vee\mathcal{E}_0 = \mathcal{F}_0$; moreover, by definition of $\mathcal{B}$, $\mathcal{E}$ is independent of the σ-field $\mathcal{F}_m$. According to Lemma 3, the filtration $\mathcal{F}' = \mathcal{F}^{m]}\vee\mathcal{E}$, equal to $\mathcal{F}$ up to time m and to $\mathcal{F}_m\vee\mathcal{E}$ from m on, is immersed in $\mathcal{F}$. Its end σ-field is $\mathcal{F}_m\vee\mathcal{E}_0 = \mathcal{F}_0$, so Lemma 5 gives $\mathcal{F}' = \mathcal{F}$ and one has $\mathcal{F}_m\vee\mathcal{E}_\ell = \mathcal{F}_\ell$ for all $\ell \in [m,0]$. For $m < \ell \leqslant 0$, $\mathcal{E}_\ell$ is conditionally non-atomic given $\mathcal{E}_{\ell-1}$, for this property is inherited from $\mathcal{F}_\ell$ and $\mathcal{F}_{\ell-1}$ by Corollary 3 a). For $m < \ell \leqslant n$, choose a complement X_ℓ to $\mathcal{E}_{\ell-1}$ in $\mathcal{E}_\ell$. By Corollary 3 b), X_ℓ is also a complement to $\mathcal{F}_{\ell-1}$ in $\mathcal{F}_\ell$. As $\mathcal{E}_m$ is degenerate, $\sigma(X_{m+1},\dots,X_n) = \mathcal{E}_n$, and S is $\sigma(X_{m+1},\dots,X_n)$-measurable; consequently, R is 2δ-close to some $\psi(X_{m+1},\dots,X_n,X_{n+1},\dots,X_0)$.

(iv′) $\Rightarrow$ (i). Choose any r.v. R generating $\mathcal{F}_0$, and a sequence $(\delta_j)_{j\in\mathbb{N}}$ tending to 0. Starting for instance with $n = -1$ and an arbitrary complement X_0 to $\mathcal{F}_{-1}$ in $\mathcal{F}_0$, and using repeatedly (iv′) for each δ_j in turn, construct a sequence $(X_\ell)_{\ell\leqslant 0}$, a strictly decreasing sequence $(n_j)_{j\in\mathbb{N}}$ in $-\mathbb{N}$, and random variables S_j, respectively $\sigma(X_{n_j+1},\dots,X_0)$-measurable and δ_j-close to R in probability. Being the limit in probability of S_j, R is $\sigma(X_\ell, \ell \leqslant 0)$-measurable, whence $\sigma(X_\ell, \ell \leqslant 0) = \mathcal{F}_0$. The filtration generated by the process X is immersed in $\mathcal{F}$ by Lemma 8, and its value at time 0 is $\mathcal{F}_0$, so it is equal to $\mathcal{F}$ by Lemma 5. ∎

Vershik's standardness criterion: Second level

Given a filtration $(\mathcal{F}_n)_{n\leqslant 0}$ (with $\mathcal{F}_0$ essentially separable) and an $\mathcal{F}_0$-measurable r.v. R, all the information on R available at time -1 is contained in the conditional law $\pi_{-1}R = \mathcal{L}[R|\mathcal{F}_{-1}]$ of R given $\mathcal{F}_{-1}$. Similarly, all the information available at time -2 on *the values of* R is carried by the conditional law $\mathcal{L}[R|\mathcal{F}_{-2}]$. But this does not encapture everything that can be said about R at time -2: it may miss some possible prediction at time -2 of *how the values of* R *will progressively be revealed in the future.* Specifically, the conditional law $\pi_{-2}R = \mathcal{L}[\pi_{-1}R|\mathcal{F}_{-2}]$ may contain more information than $\mathcal{L}[R|\mathcal{F}_{-2}]$. For an example, take a triple (N, X_{-1}, X_0) of independent and non-degenerate r.v.'s, such that N takes values -1 and 0, and X_{-1} and X_0 have the same law; take $\mathcal{F}_{-2} = \sigma(N)$, $\mathcal{F}_{-1} = \sigma(N, X_{-1})$ and $\mathcal{F}_0 = \sigma(N, X_{-1}, X_0)$, and choose $R = X_N$. The r.v. $\mathcal{L}[R|\mathcal{F}_{-2}]$ only tells you that R is independent of $\mathcal{F}_{-2}$, with the same law as X_{-1} and X_0; whereas the r.v. $\pi_{-2}R$ generates a bigger σ-field: it further tells you that if $N = -1$, R will actually be known at time -1, and if $N = 0$, R will still be completely unknown at time -1.

A key idea in Vershik's theory is to repeat this operation by putting $\pi_{-3}R = \mathcal{L}[\pi_{-2}R|\mathcal{F}_{-3}]$, and so on; he uses the full sequence $(\pi_n R)_{n\in-\mathbb{N}}$ of such iterated conditional laws. These $\pi_n R$ will be rigorously defined before Lemma 13, that characterizes the information they contain. They play the central role in the second part of Vershik's criterion, which says that $\mathcal{F}$ is standard non-atomic if and only if for each R the iterated prediction $\pi_n R$ becomes closer and closer to being deterministic when n tends to $-\infty$. This should be compared to the well-known, much easier fact, that $\mathcal{F}_{-\infty}$ is degenerate if and only if for each R the conditional law of R given $\mathcal{F}_n$ becomes closer and closer to being deterministic when n tends to $-\infty$.

As the successive $\pi_n R$ do not live in the same space, Vershik introduces for each n a distance ρ_n on the corresponding space, these distances being related to each other in a precise way; only then can the asymptotic condition be rigorously stated (condition (vii) in Theorem 2 below). It is also possible to give an equivalent statement that does not involve the distances ρ_n, namely I-cosiness; as we shall see, equivalence between I-cosiness and Vershik's second-level condition is easily established. We feel that I-cosiness may prove handier in some instances, because the ρ_n no longer appear; but it is essentially the same thing. (Vershik also gives another restatement, condition 4 in his theorem 3.2, in terms of his "tower of measures". This is a space where all the $\pi_n R$ can be made to live together; but the ρ_n are still implicitly there, in the very definition of the tower. We shall not elaborate further on this topic.)

Instead of working with real random variables, we shall take them K-valued, where (K, ρ) is a non-empty compact metric space; this will make some iterations easier, because the space K' of all probability measures on K is also compact and metrizable (for the topology of weak convergence). The set of all a.s. defined, K-valued random variables will be called $\mathrm{L}(K)$ (or $\mathrm{L}(\mathcal{A}, K)$ to specify the σ-field), and endowed with the distance $\mathbb{E}[\rho(R,S)]$. Given an essentially separable sub-σ-field $\mathcal{C}$ in a sample space $(\Omega,\mathcal{A},\mathbb{P})$, the conditional law $\mathcal{L}[R|\mathcal{C}]$ of an $R\in \mathrm{L}(K)$ given $\mathcal{C}$ is a random variable belonging to $\mathrm{L}(\mathcal{C},K')$; it is almost surely well defined, for instance by disintegrating the joint law of (C,R) where C is any r.v. generating $\mathcal{C}$.

On the set K' of all probabilities on K, the Kantorovich-Rubinshtein distance ρ' is defined as

$$\rho'(\mu,\nu)=\inf_{\substack{\lambda \text{ has margins}\\ \mu \text{ and } \nu}}\int_{K\times K}\rho(r,s)\,\lambda(dr,ds)\;,$$

where the infimum is taken over all probabilities λ on the product $K\times K$ with margins μ and ν. References on this definition are given by Vershik [17] and Dubins, Feldman, Smorodinsky and Tsirelson [5]; see also the survey by Belili [4], or the recent book [11] by Rachev and Rüschendorf. All we need to know about ρ' is recalled in the next lemma.

Lemma 10. — *With these notations, ρ' is indeed a distance on K'. Moreover, the topology generated by ρ' is the topology of weak convergence; in particular, (K',ρ') is compact.*

Let $g : K \to \mathbb{R}$ be a c-Lipschitz function, that is, $|g(r)-g(s)|\leqslant c\,\rho(r,s)$. The function g' defined on K' by $g'(\mu)=\int_K g(r)\,\mu(dr)$ is c-Lipschitz too: $|g'(\mu)-g'(\nu)|\leqslant c\,\rho'(\mu,\nu)$.

Proof. — First, to establish the triangle inequality, assume $\rho'(\mu_1,\mu_2)<\alpha$ and $\rho'(\mu_2,\mu_3)<\beta$. There exist a probability λ' with marginals μ_1 and μ_2, such that $\int\rho(r,s)\,\lambda'(dr,ds)<\alpha$, and a λ'' with marginals μ_2 and μ_3, such that $\int\rho(s,t)\,\lambda''(ds,dt)<\beta$. Disintegrating λ' and λ'' gives probabilities ν'_s and ν''_s, defined for μ_2-almost all s, such that $\lambda'(dr,ds)=\mu_2(ds)\,\nu'_s(dr)$ and $\lambda''(ds,dt)=\mu_2(ds)\,\nu''_s(dt)$. Putting $\lambda(dr,dt)=\int_{s\in K}\nu'_s(dr)\nu''_s(dt)\,\mu_2(ds)$, it is a child's play to verify that λ is a probability with marginals μ_1 and μ_3 such that $\int\rho(r,t)\,\lambda(dr,dt)<\alpha+\beta$.

We now verify that a sequence $(\mu_j)_{j\in\mathbb{N}}$ in K' converges weakly to a limit $\nu\in K'$ if and only if $\rho(\mu_j,\nu)$ tends to 0; taking a constant sequence $\mu_j=\mu$ will by the same token give the separation condition and show that ρ is a distance.

If μ_j converges weakly to ν, there exist, on a suitable sample space, random variables R_j and S with these laws and such that $\rho(R_j,S)$ tends in probability to 0. As $\rho(R_j,S)$ is bounded by the diameter of K, convergence also holds in L^1 and $\mathbb{E}[\rho(R_j,S)]\to 0$. But, calling λ_j the joint law of R_j and S, one has $\rho'(\mu_j,\nu)\leqslant\int\rho(r,s)\,\lambda_j(dr,ds)=\mathbb{E}[\rho(R_j,S)]$, wherefrom $\rho'(\mu_j,\nu)\to 0$.

Conversely, supposing $\rho'(\mu_j,\nu)\to 0$, we have to verify that $\mu_j(f)\to\nu(f)$ for every continuous function f on K. By compactness, f is uniformly continuous, and given any $\varepsilon>0$ there exists $\delta>0$ such that

$$|f(r)-f(s)|\leqslant\left\{\begin{array}{ll}\varepsilon & \text{if } \rho(r,s)<\delta\\ 2\sup|f| & \text{if } \rho(r,s)\geqslant\delta\end{array}\right\}\leqslant\varepsilon+\frac{2\sup|f|}{\delta}\,\rho(r,s)\;.$$

Now choose a probability λ_j with marginals μ_j and ν, such that

$$\int\rho(r,s)\,\lambda_j(dr,ds)\leqslant\rho'(\mu_j,\nu)+\frac{\delta\varepsilon}{2\sup|f|}$$

and write

$$\Big|\int f\,d\mu_j - \int f\,d\nu\Big| = \Big|\int f(r)\,\lambda_j(dr,ds) - \int f(s)\,\lambda_j(dr,ds)\Big|$$
$$\leqslant \int |f(r)-f(s)|\,\lambda_j(dr,ds)$$
$$\leqslant \varepsilon + \frac{2\sup|f|}{\delta}\int \rho(r,s)\,\lambda_j(dr,ds)$$
$$\leqslant \varepsilon + \frac{2\sup|f|}{\delta}\,\rho'(\mu_j,\nu)+\varepsilon\;.$$

This implies $\limsup_j |\int f\,d\mu_j - \int f\,d\nu| \leqslant 2\varepsilon$, and $\int f\,d\mu_j$ tends to $\int f\,d\nu$.

Last, if g is a real, c-Lipschitz function on K, given μ and ν in K' a similar computation yields for every λ with marginals μ and ν

$$|g'(\mu)-g'(\nu)| \leqslant \int |g(r)-g(s)|\,\lambda(dr,ds) \leqslant c\int \rho(r,s)\,\lambda(dr,ds)\;;$$

taking the infimum over all such λ yields $|g'(\mu)-g'(\nu)| \leqslant c\,\rho'(\mu,\nu)$. ∎

The next two lemmas show how the definition of ρ' is taylor-made to transfer distance estimates all the way down or up a filtration. Going down the ladder is immediate:

LEMMA 11. — *Let R and S be two K-valued random variables. If $\mathcal{C}$ is an essentially separable sub-σ-field, one has $\rho'\big(\mathcal{L}[R|\mathcal{C}],\mathcal{L}[S|\mathcal{C}]\big) \leqslant \mathbb{E}[\rho(R,S)|\mathcal{C}]$ a.s.*

PROOF. — For almost all ω, the conditional law $\mathcal{L}[(R,S)|\mathcal{C}](\omega)$ is a probability on $K\times K$ with marginals $\mathcal{L}[R|\mathcal{C}](\omega)$ and $\mathcal{L}[S|\mathcal{C}](\omega)$. Calling it $\lambda(\omega)$ and inserting it into the definition of ρ' yields the almost sure inequality

$$\rho'\big(\mathcal{L}[R|\mathcal{C}](\omega),\mathcal{L}[S|\mathcal{C}](\omega)\big) \leqslant \int \rho(r,s)\,\lambda(\omega)(dr,ds) = \mathbb{E}[\rho(R,S)|\mathcal{C}](\omega)\;. \qquad ∎$$

Climbing up the ladder is a little more arduous; this is done in the next lemma. To make things technically easier, we shall deal with simple random variables, that is, random variables that take only finitely many values.

LEMMA 12. — *On some $(\Omega,\mathcal{A},\mathbb{P})$, let $\mathcal{B}$ and $\mathcal{C}$ be sub-σ-fields such that $\mathcal{C}\subset\mathcal{B}$ and $\mathcal{B}$ is conditionally non-atomic given $\mathcal{C}$; suppose $R\in \mathrm{L}(\mathcal{B},K)$ and $L\in\mathrm{L}(\mathcal{C},K')$ are simple and the values (in K') taken by the random variable L are probability measures on K with finite support. There exists a r.v. $S\in\mathrm{L}(\mathcal{B},K)$ verifying $\mathcal{L}[S|\mathcal{C}]=L$ and $\mathbb{E}[\rho(R,S)] = \mathbb{E}\big[\rho'\big(\mathcal{L}[R|\mathcal{C}],L\big)\big]$.*

PROOF. — There exists a finite set $F\subset K$ such that the values of R are points of F and the (finitely many) values taken by L are probabilities with supports in F. By weak compactness of the set of all probabilities on $K\times K$, the infimum in the definition of ρ' is reached for some λ; writing this for $\mu=\mathcal{L}[R|\mathcal{C}](\omega)$ and $\nu=L(\omega)$ shows for almost every ω the existence of a probability $\lambda(\omega)$ on $F\times F$, verifying

$$\sum_{s\in F}\lambda(\omega,\,.\,,s) = \mathcal{L}[R|\mathcal{C}](\omega) \qquad \sum_{r\in F}\lambda(\omega,r,\,.\,) = L(\omega)$$

and
$$\sum_{(r,s)\in F\times F}\rho(r,s)\,\lambda(\omega,r,s) = \rho'\big(\mathcal{L}[R|\mathcal{C}](\omega),L(\omega)\big)\;;$$

moreover, $\lambda(\omega)$ can be taken $\mathcal{C}$-measurable, for instance by the measurable section theorem.

Put $\Sigma(\omega) = \sum_{s\in F} \lambda(\omega, R(\omega), s) = \sum_{r\in F} \mathbb{1}_{\{R=r\}}\, \mathbb{P}[R=r|\mathcal{C}]$ and deduce from

$$\mathbb{P}[\Sigma=0 \text{ and } R=r|\mathcal{C}] = \mathbb{P}[R=r|\mathcal{C}]\; \mathbb{1}_{\{\mathbb{P}[R=r|\mathcal{C}]=0\}} = 0$$

that $\mathbb{P}[\Sigma=0] = 0$ and $\Sigma > 0$ a.s. Call $s_1, \ldots, s_p$ the points of F and for $0 \leqslant j \leqslant p$ set

$$Q_j(\omega) = \frac{\lambda(\omega, R(\omega), s_1) + \ldots + \lambda(\omega, R(\omega), s_j)}{\Sigma(\omega)} ;$$

these r.v.'s are $\mathcal{C}\vee\sigma(R)$-measurable and verify $0 = Q_0 \leqslant Q_1 \leqslant \ldots \leqslant Q_p = 1$. Now R is simple, so, by Corollary 4, $\mathcal{B}$ is conditionally non-atomic given $\mathcal{C}\vee\sigma(R)$, and by Proposition 4 there exists a complement X to $\mathcal{C}\vee\sigma(R)$ in $\mathcal{B}$. The F-valued r.v. S defined by

$$S(\omega) = s_j \qquad \Longleftrightarrow \qquad Q_{j-1}(\omega) < X(\omega) \leqslant Q_j(\omega)$$

verifies

$$\mathbb{P}[S=s|\mathcal{C}\vee\sigma(R)] = \frac{\lambda(\omega, R(\omega), s)}{\Sigma(\omega)} ;$$

$$\mathbb{P}[R=r \text{ and } S=s|\mathcal{C}\vee\sigma(R)] = \mathbb{1}_{\{R=r\}} \frac{\lambda(\omega, r, s)}{\mathbb{P}[R=r|\mathcal{C}]} ;$$

$$\mathbb{P}[R=r \text{ and } S=s|\mathcal{C}] = \mathbb{P}[R=r|\mathcal{C}] \frac{\lambda(\omega, r, s)}{\mathbb{P}[R=r|\mathcal{C}]} = \lambda(\omega, r, s)$$

(everything vanishes if the denominator is null). This shows that the conditional law of (R, S) given $\mathcal{C}$ is λ. It implies on the one hand $\mathcal{L}[S|\mathcal{C}] = L$ and on the other hand

$$\mathbb{E}[\rho(R, S)|\mathcal{C}] = \sum_{(r,s)\in F\times F} \rho(r, s)\, \lambda(r, s) = \rho'\big(\mathcal{L}[R|\mathcal{C}], L\big) ,$$

whence $\mathbb{E}[\rho(R, S)] = \mathbb{E}\big[\rho'\big(\mathcal{L}[R|\mathcal{C}], L\big)\big]$. ∎

Fixing a filtered space $(\Omega, \mathcal{A}, \mathbb{P}, (\mathcal{F}_n)_{n\leqslant 0})$ such that $\mathcal{F}_0$ is essentially separable, the (a.s. well defined) conditional law $\mathcal{L}[R|\mathcal{F}_{-1}]$ of an $R \in \mathrm{L}(K)$ given $\mathcal{F}_{-1}$ is a random variable belonging to $\mathrm{L}(\mathcal{F}_{-1}, K')$. As mentioned at the beginning of this section, this will be iterated by considering the conditional law of this r.v. given $\mathcal{F}_{-2}$, and so on. To do so, we shall use the following notation: $(K_0, \rho_0) = (K, \rho)$, $\pi_0 R = R$; and for each $n \leqslant 0$, $(K_{n-1}, \rho_{n-1}) = (K_n{}', \rho_n{}')$ and $\pi_{n-1}R$ is the conditional law of $\pi_n R$ given $\mathcal{F}_{n-1}$. Notice that each (K_n, ρ_n) is a compact metric space, and the random variable $\pi_n R$ belongs to $\mathrm{L}(\mathcal{F}_n, K_n)$.

What information is conveyed by the r.v. $\pi_n R$? It contains the conditional law of R given $\mathcal{F}_n$, but also (see the example at the beginning of this section) predictions about how these conditional laws may evolve from the present time n to the future time 0 (call this *second-order information*). And it also contains predictions about how this second-order information itself may evolve in the future (this is third order), and so on, up to order $|n|$ (no more information is added to $\pi_n R$ by iterating beyond order $|n|$). This procedure is an informal description from inside; the next lemma gives a (simpler) characterization from outside.

LEMMA 13. — *Assume $\mathcal{F}_0$ is essentially separable and fix $R \in \mathrm{L}(\mathcal{F}_0, K)$. The filtration generated by the process $(\pi_n R)_{n \leqslant 0}$ is the smallest filtration $\mathcal{E}$ immersed in $\mathcal{F}$ such that R is $\mathcal{E}_0$-measurable. The σ-field $\sigma(\pi_n R, n \leqslant 0)$ is the smallest $\mathcal{F}$-saturated σ-field making R measurable.*

PARTIAL PROOF. — We shall only show the immersion and saturation properties. Minimality will not be used in the sequel; its proof is left to the reader.

To see that the filtration generated by the process $(\pi_n R)_{n \leqslant 0}$ is immersed in $\mathcal{F}$, it suffices to verify that $\mathbb{E}[\phi(\pi_m R, m \leqslant 0) | \mathcal{F}_n]$ is $\sigma(\pi_m R, m \leqslant n)$-measurable for each $n \leqslant 0$. Writing $\mathbb{E}[\phi(\ldots, \pi_{-1}R, \pi_0 R) | \mathcal{F}_{-1}]$ as $\int \phi(\ldots, \pi_{-1}R, r)\, d(\pi_{-1}R)(r)$ shows it for $n = -1$; the general case follows by iteration.

Consequently, by Lemma 5, the σ-field $\sigma(\pi_n R, n \leqslant 0)$ is $\mathcal{F}$-saturated. ∎

LEMMA 14. — *Suppose $\mathcal{F}_0$ to be essentially separable. If R and S are two K-valued, $\mathcal{F}_0$-measurable random variables, the process $\rho_n(\pi_n R, \pi_n S)$ is a submartingale. In particular,*

$$\mathbb{E}[\rho_n(\pi_n R, \pi_n S)] \leqslant \mathbb{E}[\rho(R, S)] \ .$$

PROOF. — Immediate by Lemma 11. ∎

LEMMA 15. — *Let $\mathcal{F}$ be non-atomic and R be an $\mathcal{F}_0$-measurable r.v. with values in a finite set $F \subset K$. For every $\delta > 0$, there exists an $\mathcal{F}_0$-measurable, F-valued r.v. S such that $\pi_{-1}S$ is simple and $\mathbb{E}[\rho(R, S)] < \delta$.*

PROOF. — Call F' the set of all probabilities on F; it is compact, so there exist a finite $G \subset F'$ and a Borel $f : F' \to G$ such that $\rho_{-1}\big(\mu, f(\mu)\big) < \delta$ for every $\mu \in F'$. Lemma 12 with $L = f(\pi_{-1}R)$ gives a K-valued S such that $\pi_{-1}S = L$ (hence S is F-valued) and that $\mathbb{E}[\rho(R, S)] = \mathbb{E}[\rho_{-1}(\pi_{-1}R, L)] = \mathbb{E}\big[\rho_{-1}\big(\pi_{-1}R, f(\pi_{-1}R)\big)\big] < \delta$. ∎

DEFINITION. — *The filtration $(\mathcal{F}_m)_{m \leqslant 0}$ and $n \leqslant 0$ being fixed, a r.v. $R \in \mathrm{L}(\mathcal{F}_0, K)$ is* n-simple *if the $|n|+1$ r.v.'s $\pi_n R, \pi_{n+1} R, \ldots, \pi_0 R = R$ are simple.*

The next lemma says that n-simplicity of R is in fact a property of $\pi_n R$.

NOTATION. — Set $K_0^{\mathrm{S}} = K$ and for $n < 0$ call K_n^{S} the set of all probabilities carried by finitely many points of K_{n+1}^{S}. Observe that K_n^{S} is included in K_n, but K_n^{S} is not compact (unless $n = 0$ or $|K| = 1$).

LEMMA 16. — *Assume $\mathcal{F}_0$ is essentially separable. For $n \leqslant 0$, a random variable $R \in \mathrm{L}(\mathcal{F}_0, K)$ is n-simple if and only if $\pi_n R$ is simple and takes its values in K_n^{S}.*

PROOF. — For $n = 0$, this is just the definition of a simple r.v. Assuming the lemma holds for some n, we shall prove it for $n - 1$.

If R is $(n-1)$-simple, it is also n-simple, and, by induction hypothesis, $\pi_n R$ is simple and takes its values in K_n^{S}; these values belong to some finite $F \subset K_n^{\mathrm{S}}$. The set F' of all probabilities on F is included in K_{n-1}^{S}; $\pi_{n-1}R$ has values in F' and a fortiori in K_{n-1}^{S}. It is also simple, by definition of $(n-1)$-simplicity.

Conversely, if $\pi_{n-1}R$ is simple and takes its values in K_{n-1}^{S}, these values are finitely many probabilities finitely supported in K_n^{S}, so $\pi_n R$ is simple and K_n^{S}-valued, and, by induction hypothesis, R is n-simple. As $\pi_{n-1}R$ is simple too, R is $(n-1)$-simple. ∎

LEMMA 17. — *Suppose $\mathcal{F}$ to be non-atomic; for some $n \leqslant 0$, let $R \in \mathrm{L}(\mathcal{F}_0, K)$ and L be a simple, $\mathcal{F}_n$-measurable r.v. with values in K_n^{S}. There exists an n-simple $S \in \mathrm{L}(\mathcal{F}_0, K)$ such that $\pi_n S = L$ and $\mathbb{E}[\rho(R,S)] = \mathbb{E}[\rho_n(\pi_n R, L)]$.*

PROOF. — If $n = 0$, take $S = L$. If $n < 0$, writing L_n instead of L and working by induction, it suffices to show that our hypothesis implies the existence of a simple, K_{n+1}^{S}-valued $L_{n+1} \in \mathrm{L}(\mathcal{F}_{n+1}, K_{n+1})$ verifying $\mathcal{L}[L_{n+1}|\mathcal{F}_n] = L_n$ and $\mathbb{E}[\rho_{n+1}(\pi_{n+1}R, L_{n+1})] = \mathbb{E}[\rho_n(\pi_n R, L_n)]$. But this is just Lemma 12 with K_{n+1} instead of K, $\mathcal{B} = \mathcal{F}_{n+1}$ and $\mathcal{C} = \mathcal{F}_n$. ∎

LEMMA 18. — *Assume $\mathcal{F}$ is non-atomic. For fixed $n \leqslant 0$, the set of all n-simple r.v.'s is dense in $\mathrm{L}(\mathcal{F}_0, K)$.*

PROOF. — For $n = 0$, this just recalls that simple r.v.'s are dense. To prove the lemma, it suffices to show that the set of all $(n-1)$-simple r.v.'s is dense in the set of all n-simple ones. So let R be n-simple; $\pi_n R$ is simple and K_n^{S}-valued (Lemma 16), its values belong to a finite subset F of K_n^{S}. Applying Lemma 15 to $\pi_n R$ yields an F-valued L_n such that $\mathcal{L}[L_n|\mathcal{F}_{n-1}]$ is simple and $\mathbb{E}[\rho_n(\pi_n R, L_n)]$ arbitrarily small; then Lemma 17 gives a n-simple S such that $\pi_n S = L_n$ and $\mathbb{E}[\rho(R,S)]$ is small; as $\pi_{n-1}S = \mathcal{L}[L_n|\mathcal{F}_{n-1}]$ is simple, S is $(n-1)$-simple. ∎

With Lemmas 17 and 18 at our disposal, the equivalence between standardness and Vershik's second-level criterion is within hand reach.

DEFINITION. — *Let (K,ρ) denote the unit interval $[0,1]$ with the usual distance. A filtration $\mathcal{F}$* satisfies Vershik's criterion *if $\mathcal{F}_0$ is essentially separable and, for every $\delta > 0$ and every $\mathcal{F}_0$-measurable, $[0,1]$-valued R, there exist an $n \leqslant 0$ and a $\mu \in K_n$ such that $\mathbb{E}[\rho_n(\pi_n R, \mu)] < \delta$,* where (K_n, ρ_n) and $\pi_n R$ are inductively defined as above, starting with (K,ρ).

REMARK. — This definition is not changed if the interval $[0,1]$ is replaced with an arbitrary infinite compact metric space (K,ρ). To see it, observe first that the property holds with (K,ρ) if and only if it holds for every (F,ρ), where F ranges over all finite subsets of K (approximate $R \in \mathrm{L}(\mathcal{F}_0, K)$ by simple random variables and use Lemma 14). Then notice that if it holds for (F,ρ), it also holds for any other metric space (F^1, ρ^1) with the same cardinality as F (since, after identifying F and F^1 by an arbitrary bijection, the ratio ρ^1/ρ is bounded above and below).

PROPOSITION 5. — *If a filtration $\mathcal{F} = (\mathcal{F}_n)_{n \leqslant 0}$ is I-cosy and if $\mathcal{F}_0$ is essentially separable, $\mathcal{F}$ satisfies Vershik's criterion.*

PROOF. — Given an $\mathcal{F}_0$-measurable, $[0,1]$-valued R and a $\delta > 0$, the cosiness hypothesis provides us with an n, an $(\overline{\Omega}, \overline{\mathcal{A}}, \overline{\mathbb{P}})$ and two isomorphic copies $\mathcal{F}'$ and $\mathcal{F}''$ of $\mathcal{F}$ on $\overline{\Omega}$, jointly immersed (in $\overline{\mathcal{F}} = \mathcal{F}' \vee \mathcal{F}''$), independent at time n, and such that $\overline{\mathbb{P}}[|R' - R''| \geqslant \delta] < \delta$. As $|R' - R''| \leqslant 1$, $\overline{\mathbb{E}}[|R' - R''|] < 2\delta$, whence $\overline{\mathbb{E}}[\rho_n(\pi_n R', \pi_n R'')] < 2\delta$ by Lemma 14. In this formula, the K_n-valued r.v. $\pi_n R'$ is defined in the filtration $\overline{\mathcal{F}}$; by immersion, it remains the same when computed in $\mathcal{F}'$. By isomorphic transfer, $\pi_n R'$ and $\pi_n R''$ have the same law as $\pi_n R$; call q this law. As $\pi_n R'$ and $\pi_n R''$ are independent, the measurable function ϕ defined on K_n by

$\phi(\lambda) = \int \rho_n(\lambda', \lambda)\, q(d\lambda') = \mathbb{E}[\rho_n(\pi_n R, \lambda)]$ verifies

$$\mathbb{E}[\phi(\pi_n R)] = \int \phi(\lambda'')\, q(d\lambda'') = \iint \rho_n(\lambda', \lambda'')\, q(d\lambda')\, q(d\lambda'') = \bar{\mathbb{E}}[\rho_n(\pi_n R', \pi_n R'')] < 2\delta\ ;$$

hence $\mathbb{P}[\phi(\pi_n R) \geqslant 3\delta] \leqslant \frac{2}{3}$, and there exists an $\omega_0 \in \Omega$ such that $\phi\big((\pi_n R)(\omega_0)\big) < 3\delta$. Taking $\mu = (\pi_n R)(\omega_0)$, one has $\mathbb{E}[\rho_n(\pi_n R, \mu)] = \phi(\mu) < 3\delta$. ∎

THEOREM 2 (Vershik [17]). — *Let $\mathcal{F} = (\mathcal{F}_n)_{n\leqslant 0}$ be a non-atomic filtration on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The following three statements are equivalent to each other and to statements* (i), (ii), (iii) *and* (iv) *of Theorem 1:*

(v) *$\mathcal{F}$ is immersible into a filtration of product type;*

(vi) *$\mathcal{F}$ is I-cosy;*

(vii) *$\mathcal{F}$ satisfies Vershik's criterion.*

PROOF. — (ii) $\Rightarrow$ (v) is trivial, (v) $\Rightarrow$ (vi) has already been seen in Corollary 2, and (vi) $\Rightarrow$ (vii) repeats Proposition 5.

(vii) $\Rightarrow$ (iii). Define a sequence of functions $g_n : K_n \to [0,1]$ by $g_0 = \mathrm{Id}$ and $g_{n-1}(\mu) = \int_{K_n} g_n(\lambda)\, \mu(d\lambda)$. Given any $\mathcal{F}_0$-measurable, $[0,1]$-valued r.v. R, equality $\mathbb{E}[R|\mathcal{F}_n] = g_n(\pi_n R)$ holds for $n = 0$, and if it holds for some n, the definition of $\pi_{n-1}R$ implies

$$\mathbb{E}[R|\mathcal{F}_{n-1}] = \mathbb{E}[g_n(\pi_n R)|\mathcal{F}_{n-1}] = \int_{K_n} g_n(\lambda)\, (\pi_{n-1}R)(d\lambda) = g_{n-1}(\pi_{n-1}R)\ ;$$

so $\mathbb{E}[R|\mathcal{F}_n] = g_n(\pi_n R)$ for all n. Now, by Lemma 10, each g_n is 1-Lipschitz. Consequently, $\big|\mathbb{E}[R|\mathcal{F}_n] - g_n(\mu)\big| \leqslant \rho_n(\pi_n R, \mu)$ for any $\mu \in K_n$. Combining this estimate with hypothesis (vii), we obtain for every $\delta > 0$ an n and a constant c such that $\mathbb{E}[R|\mathcal{F}_n]$ is δ-close to c in L^1. As this holds for some n, it also holds for all n small enough, so there are constants c_n such that $\mathbb{E}[R|\mathcal{F}_n] - c_n$ tends to 0 in L^1 when $n \to -\infty$. This implies that $\mathbb{E}[R|\mathcal{F}_n] \to \mathbb{E}[R]$, which in turn shows that $\mathcal{F}_{-\infty}$ is degenerate.

It remains to show that every $R \in \mathrm{L}^0(\mathcal{F}_0)$ can be approximated to any given accuracy by a r.v. belonging to some essentially finite, $\mathcal{F}$-saturated σ-field. By truncation, we may suppose that R is bounded, and by an affine transformation, that $0 \leqslant R \leqslant 1$. Take $K = [0,1]$ and fix $\delta > 0$. Hypothesis (vii) gives an n (fixed in the sequel) and a $\mu \in K_n$ such that $\mathbb{E}[\rho_n(\pi_n R, \mu)] < \delta$. According to Lemma 18, there exists an n-simple r.v. S such that $\mathbb{E}[\rho(R,S)] < \delta$. Lemma 14 gives $\mathbb{E}[\rho_n(\pi_n R, \pi_n S)] < \delta$, whence $\mathbb{E}[\rho_n(\pi_n S, \mu)] < 2\delta$, and $\mathbb{P}[\rho_n(\pi_n S, \mu) \geqslant 3\delta] < \frac{2}{3}$. This estimate, and the fact that $\pi_n S$ is a.s. K_n^{S}-valued (apply Lemma 16 to S), imply that for some ω_0 one has $\rho_n\big((\pi_n S)(\omega_0), \mu\big) < 3\delta$ and $(\pi_n S)(\omega_0) \in K_n^{\mathrm{S}}$. Putting $\nu = (\pi_n S)(\omega_0)$, one gets $\rho_n(\mu, \nu) < 3\delta$ and $\mathbb{E}[\rho_n(\pi_n S, \nu)] < 5\delta$. Taking now L deterministic and equal to ν in Lemma 17, we get an n-simple T such that $\mathbb{E}[\rho(S,T)] < 5\delta$ and $\pi_n T = \nu$.

The σ-field $\mathcal{B} = \sigma(\pi_m T, m \leqslant 0)$ is $\mathcal{F}$-saturated by Lemma 13. For each $m \leqslant n$, $\pi_m T$ is deterministic (easy induction, starting from $\pi_n T = \nu$); and for $m \geqslant n$, $\pi_m T$ is simple because T is n-simple; so $\mathcal{B}$ is essentially finite. Since $\mathbb{E}[\rho(R,T)] < 6\delta$ and T is $\mathcal{B}$-measurable, (iii) is established. ∎

Another proof of (v) $\Rightarrow$ (i) is given by Feldman and Smorodinsky [9]; see also part 1 of Tsirelson's preprint [16].

LEMMA 19. — *If two filtered sample spaces* $(\Omega^1, \mathcal{A}^1, \mathbb{P}^1, \mathcal{F}^1)$ *and* $(\Omega^2, \mathcal{A}^2, \mathbb{P}^2, \mathcal{F}^2)$ *satisfy Vershik's criterion, their product* $(\Omega^1{\times}\Omega^2, \mathcal{A}^1{\otimes}\mathcal{A}^2, \mathbb{P}^1{\otimes}\mathbb{P}^2, \mathcal{F}^1{\otimes}\mathcal{F}^2)$ *satisfies it too.*

PROOF. — The proof is elementary and involves no new idea; but it is made tedious and lengthy by the need of verifying inequalities in such spaces as $(K{\times}K)_n$. For more clarity, we shall denote by $\star$ the push-forward of a measure by a function: if σ is a probability on E and if $f : E \to F$ is measurable, $f{\star}\sigma$ is the probability $\sigma{\circ}f^{-1}$ on F.

Starting from a compact metric space (K, ρ), we have defined a sequence (K_n, ρ_n) of such spaces; a product $(\tilde{K}, \tilde{\rho})$ can also be defined by $\tilde{K} = K{\times}K$ and $\tilde{\rho}\big((x^1, x^2), (y^1, y^2)\big) = \rho(x^1, y^1) + \rho(x^2, y^2)$. Then, $(\tilde{K}_n, \tilde{\rho}_n)$ is defined by $\tilde{K}_n = (\tilde{K})_n$ and $\tilde{\rho}_n = (\tilde{\rho})_n$, and $(\widetilde{K_n}, \widetilde{\rho_n})$ by $\widetilde{K_n} = \widetilde{(K_n)}$ $(= K_n{\times}K_n)$ and $\widetilde{\rho_n} = \widetilde{(\rho_n)}$ as above.

There is for each $n \leqslant 0$ a natural map $i_n : \widetilde{K_n} \to \tilde{K}_n$, defined by induction: i_0 is the identity on $\tilde{K}$, and, for $n < 0$, an arbitrary element $u = (\mu^1, \mu^2)$ of $\widetilde{K_n} = K_n{\times}K_n$ is mapped to $i_n u = i_n(\mu^1, \mu^2) = i_{n+1} \star (\mu^1{\otimes}\mu^2)$; this is meaningful because μ^1 and μ^2 are probabilities on K_{n+1}.

This map i_n is useful in our setting because if R^1 (respectively R^2) is K-valued, $\mathcal{F}^1_0$-measurable (respectively $\mathcal{F}^2_0$-measurable), $R = (R^1, R^2)$ is $\tilde{K}$-valued and $\mathcal{F}_0$-measurable (we denote by $(\Omega, \mathcal{A}, \mathbb{P}, \mathcal{F})$ the product space), and one has, with obvious notations, $\pi_n R = i_n(\pi^1_n R^1, \pi^2_n R^2)$. Indeed, if this holds at level n, then

$$\begin{aligned}
\pi_{n-1} R &= \mathcal{L}[\pi_n R | \mathcal{F}_{n-1}] = \mathcal{L}\big[i_n(\pi^1_n R^1, \pi^2_n R^2) \mid \mathcal{F}_{n-1}\big] \\
&= i_n \star \mathcal{L}\big[(\pi^1_n R^1, \pi^2_n R^2) \mid \mathcal{F}_{n-1}\big] = i_n \star \big(\mathcal{L}[(\pi^1_n R^1)|\mathcal{F}^1_{n-1}]{\otimes}\mathcal{L}[(\pi^2_n R^2)|\mathcal{F}^2_{n-1}]\big) \\
&= i_n \star (\pi^1_{n-1} R^1{\otimes}\pi^2_{n-1} R^2) = i_{n-1}(\pi^1_{n-1} R^1, \pi^2_{n-1} R^2)\ .
\end{aligned}$$

Now, it is a fact that the distances $\tilde{\rho}_n$ on $\tilde{K}_n$ and $\widetilde{\rho_n}$ on $\widetilde{K_n}$ are related by

$$\text{for all } u \text{ and } v \text{ in } \widetilde{K_n}, \qquad \tilde{\rho}_n(i_n u, i_n v) \leqslant \widetilde{\rho_n}(u, v)\ .$$

[This formula is trivial for $n = 0$, so to prove it by induction, we may suppose it to hold for n and check it for $n-1$. We are given $u = (\mu^1, \mu^2)$ and $v = (\nu^1, \nu^2)$ in $\widetilde{K_{n-1}} = K_{n-1}{\times}K_{n-1}$; μ^1, μ^2, ν^1, and ν^2 are elements of K_{n-1}, that is, probabilities on K_n; we want to majorize $\tilde{\rho}_{n-1}(i_{n-1}u, i_{n-1}v)$. If $\lambda^1(ds^1, dt^1)$ is any probability on $K_n{\times}K_n$ with marginals μ^1 and ν^1, and $\lambda^2(ds^2, dt^2)$ any probability on $K_n{\times}K_n$ with marginals μ^2 and ν^2, then $\lambda\big(d(s^1, s^2), d(t^1, t^2)\big) = \lambda^1(ds^1, dt^1)\,\lambda^2(ds^2, dt^2)$ is a probability on $\widetilde{K_n}{\times}\widetilde{K_n}$ with marginals $\mu^1{\otimes}\mu^2$ and $\nu^1{\otimes}\nu^2$; and, calling j_n the map from $\widetilde{K_n}{\times}\widetilde{K_n}$ to $\tilde{K}_n{\times}\tilde{K}_n$ defined by $j_n(s,t) = (i_n s, i_n t)$, the image $j_n \star \lambda$ is a probability on $\tilde{K}_n{\times}\tilde{K}_n$, with marginals $i_n \star (\mu^1{\otimes}\mu^2)$ and $i_n \star (\nu^1{\otimes}\nu^2)$, that is, with marginals $i_{n-1}u$ and $i_{n-1}v$. As a consequence, $\tilde{\rho}_{n-1}(i_{n-1}u, i_{n-1}v)$ is majorized by the infimum, over all such choices of λ^1 and λ^2, of the integral

$$\begin{aligned}
\int_{\widetilde{K_n}\times\widetilde{K_n}} \tilde{\rho}_n(\sigma, \tau)\, \ell(d\sigma, d\tau) &= \int_{\widetilde{K_n}\times\widetilde{K_n}} \tilde{\rho}_n(\sigma, \tau)\, (j_n{\star}\lambda)(d\sigma, d\tau) \\
&= \int_{\tilde{K}_n\times\tilde{K}_n} \tilde{\rho}_n(i_n s, i_n t)\, \lambda(ds, dt)\ .
\end{aligned}$$

By induction hypothesis, and by the definition of $\widetilde{\rho_n}$ as a sum, this is majorized by

$$\int_{\tilde{K}_n\times\tilde{K}_n} \widetilde{\rho_n}(s,t)\,\lambda(ds,dt) = \int_{K_n^4} \left(\rho_n(s^1,t^1)+\rho_n(s^2,t^2)\right)\,\lambda^1(ds^1,dt^1)\,\lambda^2(ds^2,dt^2)\ ;$$

taking now the infimum over all λ^1 with marginals μ^1 and ν^1, and all λ^2 with marginals μ^2 and ν^2, we get $\rho_{n-1}(\mu^1,\nu^1)+\rho_{n-1}(\mu^2,\nu^2)$, that is, $\widetilde{\rho_{n-1}}(u,v)$. The inequality is established.]

To prove the lemma, we have to show that if R is $\mathcal{F}_0$-measurable and $[0,1]$-valued, $\pi_n R$ is well approximated by some constant μ for some n. But R itself is approximated by a finite sum $R_1^1R_1^2+\ldots+R_p^1R_p^2$, where R_k^1 (respectively R_k^2) are $\mathcal{F}_0^1$-measurable (respectively $\mathcal{F}_0^2$-measurable) and bounded. By Lemma 14, it suffices to verify the property when R is such a sum, that is, a scalar product $R^1{\cdot}R^2$, where R^1 and R^2 have their values in some compact $K=[-M,M]^p$. Endow K with the Euclidean distance ρ. As $\mathcal{F}^1$ and $\mathcal{F}^2$ satisfy Vershik's criterion, the remark following the definition of the criterion gives $\mathbb{E}^1[\rho_n(\pi_n^1R^1,\mu^1)]<\delta$ and $\mathbb{E}^2[\rho_n(\pi_n^2R^2,\mu^2)]<\delta$ for some n, μ^1 and μ^2. The inequality seen above yields

$$\tilde{\rho}_n\big(\pi_n(R^1,R^2),i_n(\mu^1,\mu^2)\big) \leqslant \rho_n(\pi_n^1R^1,\mu^1)+\rho_n(\pi_n^2R^2,\mu^2)\ .$$

Now if $f:(\hat{K},\hat{\rho})\to(\bar{K},\bar{\rho})$ is a c-Lipschitz map between two compact metric spaces, $f^{(n)}:(\hat{K}_n,\hat{\rho}_n)\to(\bar{K}_n,\bar{\rho}_n)$ can be inductively defined by $f^0=f$ and $f^{(n)}\mu=f^{n+1}\star\mu$ for $n<0$; it is elementary to check that $\pi_n(f{\circ}S)=f^{(n)}(\pi_nS)$ and that $f^{(n)}$ is c-Lipschitz too. Applying this to the scalar product $f:K{\times}K\to\bar{K}$ where $\bar{K}=[-pM^2,pM^2]$ gives

$$\bar{\rho}_n\big(\pi_n(f(R^1,R^2)),f^{(n)}(i_n(\mu^1,\mu^2))\big) \leqslant c\,\tilde{\rho}_n\big(\pi_n(R^1,R^2),i_n(\mu^1,\mu^2)\big)\ ,$$

where c is a Lipschitz constant for f (this constant depends on p and M, but these parameters are fixed). So the left-hand side has expectation less than $2c\delta$, and the lemma is proved. ∎

To lift the assumptions on $\mathcal{F}$ in Theorems 1 and 2, we need a definition of standardness suitable for the general case (atomic or not).

DEFINITION — *A filtration $\mathcal{F}$ is* standard[2] *if it is immersible into a standard non-atomic filtration.*

This name is not misleading, because a non-atomic filtration is standard if and only if it is standard non-atomic. Indeed, a standard non-atomic filtration is standard because it is immersed in itself; conversely, if a non-atomic filtration is standard, it is standard non-atomic by (v) $\Rightarrow$ (i) in Theorem 2. In other words, a filtration is standard non-atomic if and only if it is both standard and non-atomic.

2. We follow Vershik's terminology. Feldman [8] and Feldman-Smorodinsky [9] say *substandard;* and they call *prestandard* a filtration satisfying condition (iii) of Corollary 5.

COROLLARY 5. — *Let* $(\Omega, \mathcal{A}, \mathbb{P}, \mathcal{F})$ *be a filtered probability space. The following five conditions are equivalent:*

(i) $\mathcal{F}$ *is standard;*

(ii) $\mathcal{F}$ *is immersible into a standard filtration;*

(iii) *if* $(\bar{\Omega}, \bar{\mathcal{A}}, \bar{\mathbb{P}}, \bar{\mathcal{F}})$ *is a standard non-atomic filtered sample space, the independent product of* $\mathcal{F}$ *and* $\bar{\mathcal{F}}$ *is standard non-atomic too;*

(iv) $\mathcal{F}_0$ *is essentially separable and* $\mathcal{F}$ *is I-cosy;*

(v) $\mathcal{F}$ *satisfies Vershik's criterion.*

The independent product in condition (iii) just means a filtration of the form $\mathcal{F}' \vee \bar{\mathcal{F}}'$, where $\mathcal{F}'$ and $\bar{\mathcal{F}}'$ are defined on the same sample space, independent and respectively isomorphic to $\mathcal{F}$ and $\bar{\mathcal{F}}$. The product of $\mathcal{F}$ with a standard non-atomic filtration is well-defined, up to isomorphism; so if (iii) holds for some standard non-atomic $(\bar{\Omega}, \bar{\mathcal{A}}, \bar{\mathbb{P}}, \bar{\mathcal{F}})$, it holds for every standard non-atomic $(\bar{\Omega}, \bar{\mathcal{A}}, \bar{\mathbb{P}}, \bar{\mathcal{F}})$.

PROOF OF COROLLARY 5. — (i) $\Rightarrow$ (iv). A standard non-atomic filtration satisfies (iv) by Theorem 2; and a filtration immersible into a filtration satisfying (iv) also satisfies (iv) by Corollary 1.

(iv) $\Rightarrow$ (v) has already been seen in Proposition 5.

(v) $\Rightarrow$ (iii). As $\bar{\mathcal{F}}$ is standard non-atomic, it satisfies Vershik's criterion by Theorem 2. By Lemma 19, its product with $\mathcal{F}$ satisfies Vershik's criterion too. But this product is non-atomic; consequently, by Theorem 2 again, it is standard non-atomic.

(iii) $\Rightarrow$ (ii). By Lemma 1, a filtration is always immersible into its independent product with any other filtration.

(ii) $\Rightarrow$ (i). If $\mathcal{F}$ is immersible into a filtration $\mathcal{G}$ which is in turn immersible into a standard non-atomic filtration $\mathcal{H}$, then $\mathcal{F}$ is immersible into $\mathcal{H}$, whence standard. ∎

Among the reasons that make standardness an interesting notion stands the following fact. *Let* $(\mathcal{F}_n)_{n\leqslant 0}$ *be a filtration, and* $(E_n)_{n\leqslant 0}$ *be a sequence of sets, each of which is either a finite set or the interval* $[0,1]$. *Suppose there exists for each* n *an* $\mathcal{F}_n$*-measurable r.v.* X_n*, uniformly distributed on* E_n*, independent of* $\mathcal{F}_{n-1}$*, and such that* $\mathcal{F}_{n-1} \vee \sigma(X_n) = \mathcal{F}_n$*. The filtration* $\mathcal{F}$ *is standard (if and) only if it is of product type.*

We shall not prove this statement. When each X_n has a diffuse law, it just repeats equivalence (i) $\Leftrightarrow$ (v) in Theorems 1 and 2. In the general case, it is an immediate corollary of Theorem 3.2 of Vershik [17]. Vershik gives a complete proof only in the case when each E_n is finite, but the indices n such that E_n is infinite are easy to deal with, in the same way as above: approximate random variables by simple ones and use Corollary 4. Another proof is provided by Feldman [8]; a key step in his method consists in showing that, for some special K, ρ, μ and ν, it is possible to find a probability λ on $K \times K$, carried by a graph, and arbitrarily close to being optimal in the definition of the Kantorovich-Rubinshtein distance $\rho'(\mu, \nu)$. Still another proof is given by Feldman and Smorodinsky [9]; instead of using the distances ρ_n on K_n, they use (non-separating) distances on the quotients $(\Omega, \mathcal{F}, \mathbb{P})/\mathcal{F}_n$.

The restriction that each X_n is uniformly distributed is essential. A very simple counter-example is attributed to Vinokurov by Vershik ([17] page 756; see also Feldman [8]): the natural filtration $\mathcal{F}$ of the stationary Markov chain $(M_n)_{n\leqslant 0}$ with two states and transition matrix $\begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}$, where $0 < p < 1$ and $p \neq \frac{1}{2}$. The following coupling argument shows that $\mathcal{F}$ is I-cosy: Consider two independent copies M^1 and M^2 of M, put $T_m = \inf\{n \geqslant m\ M_n^1 = M_n^2\}$ and define another copy M^3 of the process M by $M_n^3 = M_n^1$ if $n \leqslant T_m$ and $M_n^3 = M_n^2$ if $n \geqslant T_m$. The filtrations generated by M^1 and M^3 are isomorphic to that of M and jointly immersed in that of (M^1, M^2); by taking m small enough, the processes M^1 and M^3 have a large probability of being equal on a given interval $\{n, \dots, 0\}$. So $\mathcal{F}$ is I-cosy, hence standard (Corollary 5). On the other hand, every $\mathcal{F}_n$-event A has the form $A = \big(B \cap \{M_n = M_{n-1}\}\big) \cup \big(C \cap \{M_n \neq M_{n-1}\}\big)$ with B and C in $\mathcal{F}_{n-1}$; so $\mathbb{P}[A|\mathcal{F}_{n-1}] = p\mathbb{1}_B + (1-p)\mathbb{1}_C$, and, as $p \neq 1-p$, there are (modulo negligibility) only two non-degenerate $\mathcal{F}_n$-events independent of $\mathcal{F}_{n-1}$, namely $\{M_n = M_{n-1}\}$ and $\{M_n \neq M_{n-1}\}$. Consequently, if $\mathcal{F}$ were of product type, it would be generated by the process $Y_n = \mathbb{1}_{\{M_n = M_{n-1}\}}$; but this process determines M only up to one bit of information (interchange both states), a contradiction.

Vershik's theorem on lacunary isomorphism

The same tools that were needed to establish Vershik's criterion will now be used to prove his theorem on lacunary isomorphism, a phemenon that we find still much more mind–boggling than the existence of non standard filtrations. It says that a non standard filtration can always be made standard by a deterministic time-change, that is, by replacing $-\mathbb{N}$ with one of its subsequences. We keep following closely Vershik [17].

THEOREM 3 (lacunary isomorphism). — *Let $\mathcal{F}$ be a filtration such that $\mathcal{F}_0$ is essentially separable and $\mathcal{F}_{-\infty}$ is degenerate. There exists a strictly increasing map $\sigma : -\mathbb{N} \to -\mathbb{N}$ such that the filtration $\mathcal{G}$ defined by $\mathcal{G}_n = \mathcal{F}_{\sigma(n)}$ is standard.*

The argument will be split into several statements.

LEMMA 20. — *Let $\mathcal{B}$ and $\mathcal{C}$ be two sub-σ-fields of $\mathcal{A}$, with $\mathcal{C}$ included in $\mathcal{B}$ and $\mathcal{B}$ conditionally non-atomic given $\mathcal{C}$. If R is a $\mathcal{B}$-measurable r.v. taking values in some finite set F, there exists a $\mathcal{B}$-measurable, F-valued r.v. S independent of $\mathcal{C}$ and such that $\mathbb{P}[S \neq R] = \frac{1}{2} \sum_{r\in F} \mathbb{E}\big[|\mathbb{P}[R{=}r|\mathcal{C}] - \mathbb{P}[R{=}r]|\big]$.*

PROOF. — Remark first that if K is a finite set endowed with the distance $\rho(r,s) = \mathbb{1}_{\{r\neq s\}}$, then the Kantorovich-Rubinshtein distance ρ' on the compact K' is explicitly given by the formula $\rho'(\mu,\nu) = \frac{1}{2}\sum_{t\in K} |\mu(t) - \nu(t)|$. To see this, define

$$p(r) = \mu(r) - \mu(r)\wedge\nu(r) = \big(\mu(r) - \nu(r)\big)^+ ,$$

$$q(s) = \nu(s) - \mu(s)\wedge\nu(s) = \big(\mu(s) - \nu(s)\big)^- ,$$

$$C = 1 - \sum_t \mu(t)\wedge\nu(t) = \sum_r p(r) = \sum_s q(s) = \tfrac{1}{2}\sum_t |\mu(t) - \nu(t)|$$

and observe that $\lambda(r,s) = \mu(r)\wedge\nu(s)\,\mathbb{1}_{\{r=s\}} + p(r)q(s)/C$ is a probability on $K\times K$ with marginals μ and ν, and verifying $\lambda(t,t) = \mu(t)\wedge\nu(t)$, thus achieving the infimum in the definition of $\rho'(\mu,\nu)$.

Now apply Lemma 12 to $K = F$ endowed with this ρ, to R, and to the constant r.v. $L = \ell$ where ℓ is the law of R. This gives an S with values in F such that $\mathcal{L}[S|\mathcal{C}] = \ell$ (so S is independent of $\mathcal{C}$) and $\mathbb{P}[S\neq R] = \frac{1}{2}\sum_{r\in F}\mathbb{E}\big[|\mathcal{L}[R|\mathcal{C}](r)-\ell(r)|\big]$. ∎

LEMMA 21. — *Let $\mathcal{F}$ be a non-atomic filtration such that $\mathcal{F}_{-\infty}$ is degenerate. For every $\delta > 0$ and every $\mathcal{F}_0$-measurable, simple r.v. R, there exist an $m < 0$ and an $\mathcal{F}_0$-measurable r.v. S, independent of $\mathcal{F}_m$ and verifying $\mathbb{P}[S\neq R] < \delta$.*

PROOF. — Call F a finite set where R takes its values. For each $r \in F$, the degeneracy of $\mathcal{F}_{-\infty}$ implies that $\mathbb{P}[R{=}r|\mathcal{F}_m] \to \mathbb{P}[R{=}r]$ in L^1 when $m \to -\infty$. So there is an $m < 0$ such that $\sum_{r\in F}\mathbb{E}\big[|\mathbb{P}[R{=}r|\mathcal{F}_m] - \mathbb{P}[R{=}r]|\big] < \delta$; and Lemma 20 gives the S sought for. ∎

LEMMA 22. — *Let $\mathcal{F}$ be a filtration verifying the same hypotheses as in the previous lemma. Suppose given an $n \leqslant 0$ and $|n|$ random variables $X_{n+1},\ldots,X_0$ such that each X_ℓ is a complement to $\mathcal{F}_{\ell-1}$ in $\mathcal{F}_\ell$. For every $\delta > 0$ and every $\mathcal{F}_0$-measurable r.v. R, there exist an $m < n$, a complement X' to $\mathcal{F}_m$ in $\mathcal{F}_n$, and a r.v. S, $\sigma(X', X_{n+1},\ldots,X_0)$-measurable and δ-close to R in probability.*

PROOF. — Writing $\mathcal{F}_0 = \mathcal{F}_n\vee\sigma(X_{n+1},\ldots,X_0)$ and approximating $\mathcal{F}_n$ by essentially finite σ-fields, we may suppose R to be of the form $\phi(T, X_{n+1},\ldots,X_0)$ where ϕ is Borel and T is $\mathcal{F}_n$-measurable and simple. Lemma 21 applied to T and to the shifted filtration $(\ldots,\mathcal{F}_{n-1},\mathcal{F}_n)$ gives an $m < n$ and a simple, $\mathcal{F}_n$-measurable S, independent of $\mathcal{F}_m$ and such that $\mathbb{P}[S\neq T] < \delta$; a fortiori, $\mathbb{P}[\phi(S, X_{n+1},\ldots,X_0)\neq R] < \delta$. According to Corollary 4, there exists a complement X' to $\mathcal{F}_m$ in $\mathcal{F}_n$ such that $\sigma(X') \supset \sigma(S)$; so $S = \psi(X')$ and $\phi\big(\psi(X'), X_{n+1},\ldots,X_0\big)$ is δ-close to R in probability. ∎

PROPOSITION 6. — *Let $\mathcal{F}$ be a non-atomic filtration such that $\mathcal{F}_{-\infty}$ is degenerate. There exists a strictly increasing map $\sigma : -\mathbb{N} \to -\mathbb{N}$ such that the filtration $\mathcal{G}$ defined by $\mathcal{G}_n = \mathcal{F}_{\sigma(n)}$ is standard non-atomic.*

PROOF. — Choose a r.v. R generating $\mathcal{F}_0$ and a sequence $(\delta_\ell)_{\ell\leqslant 0}$ such that $\delta_0 = 1$, $\delta_\ell > 0$ and $\delta_\ell \to 0$ when $\ell \to -\infty$. We shall first show the existence of a strictly increasing mapping $\sigma : -\mathbb{N} \to -\mathbb{N}$, a sequence $(X'_\ell)_{\ell\leqslant 0}$ such that each X'_ℓ is a complement to $\mathcal{F}_{\sigma(\ell-1)}$ in $\mathcal{F}_{\sigma(\ell)}$, and and a sequence $(S_\ell)_{\ell\leqslant 0}$ of r.v.'s such that S_ℓ is $\sigma(X'_\ell,\ldots,X'_0)$-measurable and δ_ℓ-close to R in probability. Indeed, supposing $\sigma(\ell),\ldots,\sigma(0)$, $X'_{\ell+1},\ldots,X'_0$ and $S_{\ell+1},\ldots,S_0$ have already been constructed, $\sigma(\ell{-}1)$, X'_ℓ and S_ℓ are obtained simply by applying Lemma 22 to the filtration $(\ldots,\mathcal{F}_{\sigma(\ell+1)-2},\mathcal{F}_{\sigma(\ell+1)-1},\mathcal{F}_{\sigma(\ell+1)},\mathcal{F}_{\sigma(\ell+2)},\ldots,\mathcal{F}_{\sigma(-1)},\mathcal{F}_{\sigma(0)})$; this gives σ, X' and S by induction (the first step just starts with $\sigma(0) = 0$, no X' and no S; applying Lemma 22 then yields $\sigma(-1)$, X'_0 and S_0, and so on).

Now, Lemma 8 says that the (standard non-atomic) filtration $\mathcal{G}'$ generated by the process X' is immersed in the filtration $\mathcal{G}$ defined by $\mathcal{G}_n = \mathcal{F}_{\sigma(n)}$. But R is the limit in probability of S_ℓ, so it is $\mathcal{G}'_0$-measurable, whence $\mathcal{G}'_0 = \mathcal{F}_0$ by the choice of R. As $\mathcal{G}_0$ is sandwiched between $\mathcal{G}'_0$ and $\mathcal{F}_0$, one also has $\mathcal{G}'_0 = \mathcal{G}_0$; consequently, by Lemma 5, $\mathcal{G}' = \mathcal{G}$ (both these filtrations are immersed in $\mathcal{G}$), and $\mathcal{G}$ is standard non-atomic. ∎

PROOF OF THEOREM 3. — Let $\mathcal{F}$ be any filtration such that $\mathcal{F}_0$ is essentially separable and $\mathcal{F}_{-\infty}$ is degenerate. By enlarging the sample space if necessary, we may suppose the existence of a standard non-atomic filtration $\mathcal{H}$ independent of $\mathcal{F}$. The filtration $\mathcal{F}\vee\mathcal{H}$ satisfies the sames hypotheses as $\mathcal{F}$ and is non-atomic. So Proposition 6 can be applied to this filtration, giving a sub-sequence σ such that the filtration $\mathcal{K}_n = \mathcal{F}_{\sigma(n)}\vee\mathcal{H}_{\sigma(n)}$ is standard non-atomic. Set $\mathcal{G}_n = \mathcal{F}_{\sigma(n)}$. Being immersed in $\mathcal{K}$ by Lemma 1, the filtration $\mathcal{G}$ is standard. ∎

Study of an example

To illustrate the notion of I-cosiness and how it can be used, we now turn to Vershik's Example 3 ([17], page 744). This will be done in two steps: we start with a modified, easier version of the example, due to Smorodinsky [14]; there the state spaces are finite. Then we come back to Vershik's version, which is slightly less simple (the state space is $[0,1]$), but stationary.

DEFINITIONS. — *Given a filtered probability space* $\big(\Omega,\mathcal{A},\mathbb{P},(\mathcal{F}_n)_{n\leqslant 0}\big)$, *a process* $(\varepsilon_n)_{n\leqslant 0}$ *is an* $\mathcal{F}$-coin-tossing *if, for each* $n \leqslant 0$, ε_n *is* $\mathcal{F}_n$*-measurable, independent of* $\mathcal{F}_{n-1}$, *and uniformly distributed on* $\{-1,1\}$.

A filtration $\mathcal{F}$ *has the* predictable representation property *with respect to an* $\mathcal{F}$*-coin-tossing* ε *if* $\mathcal{F}_{-\infty}$ *is degenerate and, for each* n, $\mathcal{F}_n$ *is generated by* $\mathcal{F}_{n-1}$ *and* ε_n.

The definition of an $\mathcal{F}$-coin-tossing is equivalent to demanding that the law of the process ε is that of a fair coin-tossing, and that the filtration it generates is immersed in $\mathcal{F}$.

The predictable representation property with respect to ε amounts to saying that every $\mathcal{F}$-martingale M has the form $M_n = \mathbb{E}[M_0] + \sum\limits_{m\leqslant n} H_m\varepsilon_m$ for some $\mathcal{F}$-predictable process H.

THEOREM 4 (Vershik [17], Smorodinsky [14]). — *There exists a filtered probability space* $(\Omega,\mathcal{A},\mathbb{P})$ *with a filtration* $\mathcal{F} = (\mathcal{F}_n)_{n\leqslant 0}$ *and an* $\mathcal{F}$*-coin-tossing* $\varepsilon = (\varepsilon_n)_{n\leqslant 0}$ *such that* $\mathcal{F}$ *has the predictable representation property with respect to* ε, *but* $\mathcal{F}$ *is not standard.*

In particular, this $\mathcal{F}$ is not of product type (this would imply I-cosiness by Proposition 3), so it is not generated by any coin-tossing whatsoever; according to Corollary 2, is it not even immersible into a filtration of product type (for instance the filtration generated by some coin-tossing process).

We now describe Smorodinsky's construction of this paradoxical filtration.

Let A be a finite set, called the alphabet, with k elements, called letters ($k \geqslant 2$). An ordered sequence of m letters is called an m-word; the set of all m-words is A^m, its cardinality is k^m. An m-word w and an m'-word w' can be concatenated to form the $(m{+}m')$-word $w{\cdot}w'$. For $n \leqslant 0$, put $B_n = A^{2^{|n|}}$.

Consider a Markov process $(X_n, \varepsilon_n)_{n\leqslant 0}$ with the following law: for each $n \leqslant 0$, (X_n, ε_n) is uniformly distributed on $B_n\times\{-1,1\}$ (that is, the random $2^{|n|}$-word X_n and the random sign ε_n are uniform and independent); the transition from $n-1$ to n is obtained by taking ε_n independent of $(X_{n-1}, \varepsilon_{n-1})$ and choosing X_n as the first half of the word X_{n-1} if $\varepsilon_n = 1$, and as the second half of X_{n-1} if $\varepsilon_n = -1$.

Clearly, this transition probability, when applied to a uniformly distributed $(X_{n-1}, \varepsilon_{n-1})$, yields a uniformly distributed (X_n, ε_n); this compatibility implies existence and uniqueness in law of the process (X, ε). From now on, we suppose (X, ε) is realized on a sample space $(\Omega, \mathcal{A}, \mathbb{P})$, and we call $\mathcal{F}$ the filtration it generates.

The Markov property and the independence of ε_n and $(X_{n-1}, \varepsilon_{n-1})$ imply that ε_n is independent of $\mathcal{F}_{n-1}$; in other words, the process ε is an $\mathcal{F}$-coin-tossing.

THEOREM 4′ (Smorodinsky [14]). — *With the above definitions, the filtration $\mathcal{F}$ has the predictable representation property with respect to ε, but it is not standard.*

The proof is copied from Smorodinsky; the only difference is that we use the language of I-cosiness, but the mathematics are exactly the same.

PROOF OF THE PREDICTABLE REPRESENTATION PROPERTY. — As X_n is the first or second half of X_{n-1} according to the value of ε_n, X_n is a function of X_{n-1} and ε_n, and $\mathcal{F}_n$ is generated by $\mathcal{F}_{n-1}$ and ε_n. To establish the predictable representation property, it remains to verify that $\mathcal{F}_{-\infty}$ is degenerate, or, equivalently, that for each $Z \in \mathrm{L}^1(\mathcal{F}_0)$, the conditional expectation $\mathbb{E}[Z|\mathcal{F}_n]$ tends to $\mathbb{E}[Z]$ when n tends to $-\infty$.

We know that X_n is one half of the word X_{n-1}; call W_n the other half (which may happen to be equal to X_n). Knowing X_0, the sequence (ε_n) and the sequence (W_n), it is easy to recover all the X_n by backward induction; so $\mathcal{F}_0$ is generated by X_0, the ε and the W. Consequently, we may suppose that Z is a function of X_0, $(\varepsilon_{\ell+1}, \varepsilon_{\ell+2}, \ldots, \varepsilon_0)$ and $(W_{\ell+1}, W_{\ell+2}, \ldots, W_0)$ for some $\ell < 0$. The vector $(W_{\ell+1}, W_{\ell+2}, \ldots, W_0, X_0)$ is itself a function of X_ℓ and $(\varepsilon_{\ell+1}, \varepsilon_{\ell+2}, \ldots, \varepsilon_0)$; so $Z = f(X_\ell, \varepsilon_{\ell+1}, \ldots, \varepsilon_0)$. Now, for $n \leqslant \ell$, to compute the conditional expectation $\mathbb{E}[Z|\mathcal{F}_n]$, we may replace Z by $\mathbb{E}[Z|\mathcal{F}_\ell]$; the latter is but $g(X_\ell)$, with $g(x) = \mathbb{E}[f(x, \varepsilon_{\ell+1}, \ldots, \varepsilon_0)]$, for $(\varepsilon_{\ell+1}, \ldots, \varepsilon_0)$ is independent of $\mathcal{F}_\ell$. So it suffices to check that the conditional law of X_ℓ given $\mathcal{F}_n$ tends to the uniform law on $A^{2^{|\ell|}}$. This conditional law can easily be described: the random $2^{|n|}$-word X_n can be sliced into $2^{|n|-|\ell|}$ smaller words, each with length $2^{|\ell|}$; conditionally on $\mathcal{F}_n$, X_ℓ is uniformly chosen among those $2^{|n|-|\ell|}$ words. As they are independent and uniformly distributed on $A^{2^{|\ell|}}$, they constitute a $2^{|n|-|\ell|}$-sample of the uniform law on $A^{2^{|\ell|}}$, and the conditional law of X_ℓ given $\mathcal{F}_n$ is the empirical measure associated to this sample. When $n \to -\infty$, the size of the sample tends to infinity (it doubles at each step), and the empirical measure converges almost surely to the uniform law on $A^{2^{|\ell|}}$ by the law of large numbers. ∎

Before we turn to non-standardness, some definitions and estimates will be useful. In the next lemma, the symmetric group $\mathfrak{S}_m$ is identified with the group of all permutations of the set $I_m = \{1, 2, \ldots, m\}$.

LEMMA 23. — *There exists a sequence* $(G_n)_{n\leqslant 0}$ *such that*

(i) *for each* $n \leqslant 0$, G_n *is a sub-group of the symmetric group* $\mathfrak{S}_{2^{|n|}}$, *having* $2^{2^{|n|}-1}$ *elements;*

(ii) *the permutation of* $I_{2^{|n|+1}}$ *defined as the translation by* $2^{|n|}$ *modulo* $2^{|n|+1}$ *is in* G_{n-1} *(it globally exchanges both halves of the interval* $I_{2^{|n|+1}}$ *without modifying the order inside them);*

(iii) *if* g_1 *and* g_2 *are in* G_n, *then the permutation* g *of* $I_{2^{|n|+1}}$ *acting as* g_1 *on the first half* $I_{2^{|n|}}$ *of* $I_{2^{|n|+1}}$, *and as* g_2 *on its second half (identified with* $I_{2^{|n|}}$ *by a translation) is in* G_{n-1}.

The G_n are unique; this is easy to check but we shall not need it. They are the groups of automorphisms of the dyadic tree; Vershik calls them $\mathcal{D}_n$.

PROOF OF LEMMA 23. — By induction: define G_0 as $\{\mathrm{Id}\}$ and G_{n-1} as the set of all permutations g of $I_{2^{|n|+1}}$ such that

either both halves of the interval $I_{2^{|n|+1}}$ are globally stable by g, and the restriction of g to each half is in G_n;

or both halves of the interval are globally exchanged by g, and g is the product of the translation defined in condition (ii) by an element of the previous type.

It is easy to see that this set is indeed a group, whose order $|G_{n-1}|$ verifies $|G_{n-1}| = 2\,|G_n|^2$; this gives (i). ∎

A word $w \in B_n$ can be written as $w_1 w_2 \dots w_{2^{|n|}}$. The group G_n acts on B_n in the obvious way, by permuting the $2^{|n|}$ letters: $g(w_1 w_2 \dots w_{2^{|n|}}) = w_{g(1)} w_{g(2)} \dots w_{g(2^{|n|})}$. Notice the following property: if g_1 and g_2 are in G_n, there exists $g \in G_{n-1}$ such that, for all words u and v in B_n, $g(u{\cdot}v) = g_1(u){\cdot}g_2(v)$ and there exists $h \in G_{n-1}$ such that, for all words u and v in B_n, $h(v{\cdot}u) = g_1(u){\cdot}g_2(v)$.

Endow the set B_n with the distance

$$\delta_n(w', w'') = \frac{1}{2^{|n|}} \sum_{i=1}^{2^{|n|}} \mathbb{1}_{\{w'_i \neq w''_i\}} \,.$$

Remark that the distance δ_n is invariant under the action of G_n (more generally, of the whole symmetric group) and define on $B_n \times B_n$ the symmetric function

$$d_n(w', w'') = \inf_{g \in G_n} \delta_n\bigl(w', g(w'')\bigr) \,.$$

LEMMA 24. — *If* u', v', u'' *and* v'' *are four words in* B_n,

$$\tfrac{1}{2}\,\bigl[\bigl(d_n(u',u'')+d_n(v',v'')\bigr) \wedge \bigl(d_n(u',v'')+d_n(v',u'')\bigr)\bigr] \geqslant d_{n-1}(u'{\cdot}v', u''{\cdot}v'') \,.$$

(Actually, Smorodinsky establishes equality; only this inequality will be needed.)

PROOF OF LEMMA 24. — We first show $\frac{1}{2}\bigl(d_n(u',u'')+d_n(v',v'')\bigr) \geqslant d_{n-1}(w',w'')$, with $w' = u'{\cdot}v'$ and $w'' = u''{\cdot}v''$. The left side is $\frac{1}{2}\bigl[\delta_n\bigl(u',g_1(u'')\bigr)+\delta_n\bigl(v',g_2(v'')\bigr)\bigr]$, where g_1 and g_2 are in G_n; this is also $\delta_{n-1}\bigl(u'{\cdot}v', g_1(u''){\cdot}g_2(v'')\bigr)$. But $g_1(u''){\cdot}g_2(v'')$ is equal to $g(u''{\cdot}v'')$, where g is the element of G_{n-1} acting as g_1 on the first half of the word and as g_2 on the second half. So we have on the left $\delta_{n-1}\bigl(w', g(w'')\bigr)$, and this majorizes $d_{n-1}(w',w'')$ by definition of d_{n-1}.

The other minoration is similar, with an $h \in G_{n-1}$ such that $g_1(u''){\cdot}g_2(v'')$ equals $h(v''{\cdot}u'')$. ∎

LEMMA 25. — *Let $\mathcal{F}'$ and $\mathcal{F}''$ be two isomorphic copies of $\mathcal{F}$, both immersed in some filtered probability space $(\bar{\Omega}, \bar{\mathcal{A}}, \bar{\mathbb{P}}, (\bar{\mathcal{F}}_n)_{n\leqslant 0})$; call X' and X'' the copies of X in $\mathcal{F}'$ and $\mathcal{F}''$. The process $d_n(X'_n, X''_n)$ is an $\bar{\mathcal{F}}$-submartingale.*

PROOF OF LEMMA 25. — Fix $n \leqslant 0$ and write $X'_{n-1} = U' \cdot V'$ and $X''_{n-1} = U'' \cdot V''$, where U' is the first half of the random word X'_{n-1} and similarly for X''. By immersion, the copies ε' and ε'' of ε in $\mathcal{F}'$ and $\mathcal{F}''$ verify

$$\bar{\mathbb{P}}[\varepsilon'_n{=}1|\bar{\mathcal{F}}_{n-1}] = \bar{\mathbb{P}}[\varepsilon'_n{=}-1|\bar{\mathcal{F}}_{n-1}] = \bar{\mathbb{P}}[\varepsilon''_n{=}1|\bar{\mathcal{F}}_{n-1}] = \bar{\mathbb{P}}[\varepsilon''_n{=}-1|\bar{\mathcal{F}}_{n-1}] = \tfrac{1}{2}\,.$$

This implies

$$\begin{aligned}\bar{\mathbb{P}}[\varepsilon'_n{=}\varepsilon''_n{=}1|\bar{\mathcal{F}}_{n-1}] = \bar{\mathbb{P}}[\varepsilon'_n{=}\varepsilon''_n{=}-1|\bar{\mathcal{F}}_{n-1}] &= \tfrac{1}{2}\,\bar{\mathbb{P}}[\varepsilon'_n{=}\varepsilon''_n|\bar{\mathcal{F}}_{n-1}]\,;\\ \bar{\mathbb{P}}[\varepsilon'_n{=}1, \varepsilon''_n{=}-1|\bar{\mathcal{F}}_{n-1}] = \bar{\mathbb{P}}[\varepsilon'_n{=}-1, \varepsilon''_n{=}1|\bar{\mathcal{F}}_{n-1}] &= \tfrac{1}{2}\,\bar{\mathbb{P}}[\varepsilon'_n{\neq}\varepsilon''_n|\bar{\mathcal{F}}_{n-1}]\,.\end{aligned}$$

By construction of X,

$$X'_n = \begin{cases} U' & \text{if } \varepsilon'_n = 1\\ V' & \text{if } \varepsilon'_n = -1\end{cases} \qquad\qquad X''_n = \begin{cases} U'' & \text{if } \varepsilon''_n = 1\\ V'' & \text{if } \varepsilon''_n = -1;\end{cases}$$

so

$$\begin{aligned}\bar{\mathbb{E}}[d_n(X'_n, X''_n)|\bar{\mathcal{F}}_{n-1}] = {} & d_n(U', U'')\,\bar{\mathbb{P}}[\varepsilon'_n{=}\varepsilon''_n{=}1|\bar{\mathcal{F}}_{n-1}]\\ & + d_n(U', V'')\,\bar{\mathbb{P}}[\varepsilon'_n{=}1, \varepsilon''_n{=}-1|\bar{\mathcal{F}}_{n-1}]\\ & + d_n(V', U'')\,\bar{\mathbb{P}}[\varepsilon'_n{=}-1, \varepsilon''_n{=}1|\bar{\mathcal{F}}_{n-1}]\\ & + d_n(V', V'')\,\bar{\mathbb{P}}[\varepsilon'_n{=}\varepsilon''_n{=}-1|\bar{\mathcal{F}}_{n-1}]\\ = {} & \tfrac{1}{2}\left[d_n(U', U'') {+} d_n(V', V'')\right]\bar{\mathbb{P}}[\varepsilon'_n{=}\varepsilon''_n|\bar{\mathcal{F}}_{n-1}]\\ & + \tfrac{1}{2}\left[d_n(U', V'') {+} d_n(V', U'')\right]\bar{\mathbb{P}}[\varepsilon'_n{\neq}\varepsilon''_n|\bar{\mathcal{F}}_{n-1}]\end{aligned}$$

and

$$\bar{\mathbb{E}}[d_n(X'_n, X''_n)|\bar{\mathcal{F}}_{n-1}] \geqslant \tfrac{1}{2}\left[\left(d_n(U', U''){+}d_n(V', V'')\right) \wedge \left(d_n(U', V''){+}d_n(V', U'')\right)\right].$$

It now suffices to appeal to Lemma 24 to get

$$\bar{\mathbb{E}}[d_n(X'_n, X''_n)|\bar{\mathcal{F}}_{n-1}] \geqslant d_{n-1}(X'_{n-1}, X''_{n-1})\,. \qquad ∎$$

LEMMA 26. — *Recall that k denotes the cardinality of A. Fix $n < 0$ and $m = 2^{|n|}$. If W' and W'' are two independent random m-words with uniform law on B_n,*

$$\mathbb{P}\big[d_n(W', W'') \leqslant \tfrac{1}{2}\big] < \frac{1}{2}\left(\frac{4}{\sqrt{k}}\right)^m.$$

In particular, if $k \geqslant 16$, $\mathbb{E}[d_n(W', W'')] \geqslant \frac{1}{4}$.

PROOF OF LEMMA 26. — We start with the distance δ_n instead of d_n:

$$\begin{aligned}\mathbb{P}\big[\delta_n(W', W'') \leqslant \tfrac{1}{2}\big] &= \mathbb{P}\big[W' \text{ and } W'' \text{ have at least } m/2 \text{ matching letters}\big]\\ &= \mathbb{P}\big[\exists J \subset \{1, \dots, m\}\ \ |J| = \tfrac{1}{2}m\ \ \forall j \in J\ \ W'(j) = W''(j)\big]\\ &\leqslant \sum_{\substack{J\subset\{1,\dots,m\}\\ |J| = \frac{m}{2}}} \mathbb{P}[\forall j \in J\ \ W'(j) = W''(j)]\\ &= \sum_{\substack{J\subset\{1,\dots,m\}\\ |J| = \frac{m}{2}}} \left(\frac{1}{k}\right)^{\frac{m}{2}} \leqslant \sum_{J\subset\{1,\dots,m\}} \left(\frac{1}{k}\right)^{\frac{m}{2}} = \left(\frac{2}{\sqrt{k}}\right)^m.\end{aligned}$$

Now, since for any g in G_n (or, for that matter, in $\mathfrak{S}_m$), gW'' is uniformly distributed and independent of W',

$$\mathbb{P}\big[d_n(W',W'')\leqslant\tfrac{1}{2}\big] = \mathbb{P}\big[\exists g\in G_n\ \delta_n(W',gW'')\leqslant\tfrac{1}{2}\big] \leqslant \sum_{g\in G_n}\mathbb{P}\big[\delta_n(W',gW'')\leqslant\tfrac{1}{2}\big]$$
$$= 2^{m-1}\,\mathbb{P}\big[\delta_n(W',W'')\leqslant\tfrac{1}{2}\big] \leqslant \frac{1}{2}\Big(\frac{4}{\sqrt{k}}\Big)^m .$$

When $k\geqslant 16$, the m-th power is at most 1, one has $\mathbb{P}[d_n(W',W'')\leqslant\frac{1}{2}]\leqslant\frac{1}{2}$, whence $\mathbb{E}[d_n(W',W'')]\geqslant\mathbb{E}\big[\frac{1}{2}\,\mathbb{1}_{\{d_n(W',W'')>\frac{1}{2}\}}\big]=\frac{1}{2}\,\mathbb{P}[d_n(W',W'')>\frac{1}{2}]>\frac{1}{4}$. ∎

End of the proof of Theorem 4′. To establish that $\mathcal{F}$ is not standard, we shall show it is not I-cosy.

First case : $k\geqslant 16$. — On some sample space $(\bar{\Omega},\bar{\mathcal{A}},\bar{\mathbb{P}})$, let $\mathcal{F}'$ and $\mathcal{F}''$ be any two filtrations isomorphic to $\mathcal{F}$, jointly immersed and I-separated. By triviality of G_0 and by Lemma 25, one has for every $n\leqslant 0$

$$\bar{\mathbb{P}}[X_0'\neq X_0''] = \bar{\mathbb{E}}[\delta_0(X_0',X_0'')] = \bar{\mathbb{E}}[d_0(X_0',X_0'')] \geqslant \bar{\mathbb{E}}[d_n(X_n',X_n'')] .$$

By I-separation, there is an $n<0$ such that X_n' and X_n'' are independent. For this n, Lemma 26 gives $\bar{\mathbb{E}}[d_n(X_n',X_n'')]>\frac{1}{4}$, wherefrom $\bar{\mathbb{P}}[X_0'\neq X_0'']>\frac{1}{4}$. This minoration shows that condition (iv) in the definition of I-cosiness cannot be satisfied for $R=X_0$, and $\mathcal{F}$ is not I-cosy.

Second case : $k<16$. — The new alphabet $\widehat{A}=A^4$ has at least 16 letters; it consists of "new letters", which are blocks of 4 old letters. Calling $\widehat{B}_n$ the space $\widehat{A}^{2^{|n|}}$, there is a natural identification between B_{n-2} and $\widehat{B}_n$, obtained by considering a word of $2^{|n|+2}=4\times 2^{|n|}$ old letters as a word of $2^{|n|}$ new letters. Putting $\widehat{X}_n=X_{n-2}$ and $\widehat{\varepsilon}_n=\varepsilon_{n-2}$ for $n\leqslant 0$, the natural filtration $\widehat{\mathcal{F}}$ of the process $(\widehat{X},\widehat{\varepsilon})$ is given by $\widehat{\mathcal{F}}_n=\mathcal{F}_{n-2}$. The first case applies to the new, hatted, process and shows that $\widehat{\mathcal{F}}$ is not I-cosy; according to Proposition 2, neither is $\mathcal{F}$. ∎

Other forms of cosiness

As recalled in the introduction, two other definitions of cosiness can be found in the literature; all three definitions are identical but for the separation condition (iii). The genuine one, hereafter called II-cosiness, was introduced by Tsirelson [15] in a framework where all martingales are continuous; the separation condition it uses is a reinforcement of the Kunita-Watanabe inequality: Two jointly immersed filtrations $\mathcal{F}$ and $\mathcal{G}$ (such that all martingales are continuous) are II-separate if for some $r<1$ and for all $\mathcal{F}$-martingales M and $\mathcal{G}$-martingales N started at 0, one has

$$[M,N]^2\leqslant r\,[M,M]\,[N,N] .$$

When rewritten in discrete time, this is equivalent to a conditional correlation inequality: For $\mathcal{F}$ and $\mathcal{G}$ jointly immersed in $\mathcal{H}$ and for all $F\in\mathrm{L}^2(\mathcal{F}_n)$ and $G\in\mathrm{L}^2(\mathcal{G}_n)$,

$$\mathrm{Cov}\,[F,G\mid\mathcal{H}_{n-1}]^2\leqslant r\ \mathrm{Var}\,[F\mid\mathcal{H}_{n-1}]\ \mathrm{Var}\,[G\mid\mathcal{H}_{n-1}] ;$$

with this separation condition, we do not know if the non I-cosy filtration of the above example is II-cosy or not.

The separation condition used in [2], [3] and [7] is slightly different: it does not involve time, but only the end σ-fields $\mathcal{F}_\infty$ and $\mathcal{G}_\infty$ of the filtrations (or $\mathcal{F}_0$ and $\mathcal{G}_0$ when time is $-\mathbb{N}$); this makes it in some sense a coarse tool when compared to II-separation or I-separation. We shall call it *D-separation* (D for diffuse). Its definition is quite simple: Two filtrations $\mathcal{F}$ and $\mathcal{G}$ are *D-separate* if for all random variables $F \in \mathrm{L}^0(\mathcal{F}_\infty)$ and $G \in \mathrm{L}^0(\mathcal{G}_\infty)$ with diffuse laws, one has $\mathbb{P}[F=G] = 0$. And *D-cosiness* is defined exactly as I-cosiness, but with D-separation instead of I-separation in condition (iii).

At the risk of adding a little more confusion to this admittedly already rather messy situation, we shall introduce yet another variant of cosiness. Not only by sheer pleasure of losing the reader in a maze of definitions, but for a logical reason too: we shall establish that the non I-cosy filtration of the previous section is cosy in this new sense, so it is worth stating this result with a definition of cosiness as strong as possible. (By contradistinction, D-cosiness was used in [2], [3] and [7] to show that some filtrations are *not* D-cosy, that is why its definition was made as weak as possible.)

This new definition formalizes an idea introduced and brilliantly used, whithout giving it a name, by Tsirelson in [15]; it consists in "rotating" a Gaussian processes and all associated filtrations. If $G = (G_\lambda)_{\lambda\in\Lambda}$ is a centered Gaussian process and G' an independent copy of G, define a new centered Gaussian process $(G^\alpha_\lambda)_{\alpha\in\mathbb{R},\lambda\in\Lambda}$ by $G^\alpha_\lambda = G_\lambda \cos\alpha + G'_\lambda \sin\alpha$. Notice that for any fixed α, the process $G^\alpha = (G^\alpha_\lambda)_{\lambda\in\Lambda}$ has the same law as G; so the σ-fields $\sigma(G)$ and $\sigma(G^\alpha)$ are isomorphic, with an isomorphism Ψ^α such that $\Psi^\alpha(G_\lambda) = G^\alpha_\lambda$.

DEFINITION. — *A filtration $\mathcal{F}$* (not necessarily indexed by $-\mathbb{N}$) *is* G-cosy *if there exist two independent centered Gaussian processes G and G' with the same law, defined on some sample space $(\bar\Omega, \bar{\mathcal{A}}, \bar{\mathbb{P}})$, and an embedding Φ of $(\Omega, \mathcal{F}_\infty, \mathbb{P})$ into $(\bar\Omega, \sigma(G), \bar{\mathbb{P}})$, such that, with the above notation, for each α the filtrations $\Phi(\mathcal{F})$ and $\Psi^\alpha{\circ}\Phi(\mathcal{F})$ are jointly immersed.*

This simply means the existence (possibly on an extension of Ω) of a Gaussian process G verifying $\mathcal{F}_\infty \subset \sigma(G)$ and such that $\mathcal{F}$ and any copy $\mathcal{F}^\alpha$ of $\mathcal{F}$ obtained by rotating G are jointly immersed.

The simplest example of a G-cosy filtration is any filtration generated by a Gaussian process, for instance a Brownian filtration. This is the case Tsirelson considered when introducing cosiness; his proof in [15] that such filtrations are cosy is copied below, in Proposition 8 and in the remark following it. But we cannot restrict ourselves to this case: in the proof of Proposition 9, we shall need the definition of G-cosiness in its full extent.

LEMMA 27. — *A filtration immersible into a G-cosy filtration is itself G-cosy.*

PROOF. — Suppose that $\mathcal{F}$ satisfies the above definition and $\mathcal{E}$ is immersed in $\mathcal{F}$. One has $\Phi(\mathcal{E}_\infty) \subset \Phi(\mathcal{F}_\infty) \subset \sigma(G)$; and $\Phi(\mathcal{E})$ is immersed in $\Phi(\mathcal{F})$. By hypothesis, the filtrations $\Psi^\alpha{\circ}\Phi(\mathcal{F})$ and $\Phi(\mathcal{F})$ on $\bar\Omega$ are immersed in their supremum $\mathcal{H}$, so $\Phi(\mathcal{E})$ and $\Psi^\alpha{\circ}\Phi(\mathcal{E})$ are also immersed in $\mathcal{H}$, and jointly immersed by Lemma 4 b).

This shows that a filtration immersed in a G-cosy filtration is itself G-cosy, and the lemma follows by isomorphism. ∎

PROPOSITION 7. — *Every standard filtration* $(\mathcal{F}_n)_{n\leqslant 0}$ *is G-cosy.*

PROOF. — A standard filtration is immersible into a standard, non-atomic filtration; so by the preceding lemma it suffices to verify that a (or "the") standard non-atomic filtration is G-cosy. Now a standard non-atomic filtration is generated by an i.i.d. sequence $(G_n)_{n\leqslant 0}$ of standard Gaussian random variables; it suffices to enlarge Ω to accomodate both G and an independent copy G' of G, and the definition of G-cosiness is readily verified. ∎

PROPOSITION 8. — *Every G-cosy filtration is D-cosy.*

PROOF. — Suppose a filtration $\mathcal{F}$ is G-cosy. By replacing $\mathcal{F}$ with the isomorphic filtration $\Phi(\mathcal{F})$, we may suppose that $\overline{\Omega} = \Omega$ and $\Phi = \mathrm{Id}$ in the definition of G-cosiness.

Fix $\delta > 0$ and R in $\mathrm{L}^0(\mathcal{F}_\infty)$; notice that $R \in \mathrm{L}^0(\sigma(G))$. By Slutsky's lemma (see Théorème 1 of [2]), the map $\alpha \mapsto \Psi^\alpha(R)$ is continuous for the topology of convergence in probability; so, for α close enough to 0, the distance between $\Psi^\alpha(R)$ and $\Psi^0(R)$ is less than δ. For any α, the isomorphic copies $\mathcal{F}^\alpha = \Psi^\alpha(\mathcal{F})$ and $\mathcal{F}$ of the filtration $\mathcal{F}$ are jointly immersed by hypothesis. Last, for $\alpha \neq 0 \pmod{\pi}$, the filtrations $\mathcal{F}^\alpha$ and $\mathcal{F}$ are D-separate; see for instance Proposition 2 of [3].

Consequently, for α close enough to 0 but not null, the filtrations $\mathcal{F}^\alpha$ and $\mathcal{F}^0$ fulfill all four conditions in the definition of D-cosiness. ∎

REMARK. — Define two σ-fields $\mathcal{B}$ and $\mathcal{C}$ to be *H-separate* if, for some $p \in [1,2)$,

$$\forall B \in \mathrm{L}^p(\mathcal{B}) \quad \forall C \in \mathrm{L}^p(\mathcal{C}) \qquad \mathbb{E}[BC] \leqslant \|B\|_{\mathrm{L}^p}\, \|C\|_{\mathrm{L}^p}$$

(the letter H stands for Hypercontractivity, or for Hypo-independence). Define two filtrations $\mathcal{F}$ and $\mathcal{G}$ to be H-separate if their end σ-fields $\mathcal{F}_\infty$ and $\mathcal{G}_\infty$ are, and a filtration to be H-cosy if it satisfies the definition of D-cosiness (or I-cosiness, or II-cosiness) with H-separation instead of D-separation (or I-separation, or II-separation). It is not difficult to see that H-separation implies D-separation (see [2], or Proposition 1 of [3]); therefore H-cosiness implies D-cosiness. And the proof of Proposition 8 (or the proof of Proposition 2 of [3]) shows in fact a stronger result: G-cosiness $\Rightarrow$ H-cosiness $\Rightarrow$ D-cosiness.

PROPOSITION 9. — *The* (non standard, hence also non I-cosy) *filtration* $\mathcal{F}$ *of Theorem 4′ is G-cosy* (hence also H-cosy and D-cosy).

The proof of this proposition will use some properties of a very small, abelian subgroup of the group G_n described in Lemma 23. Tsirelson calls it the group of cube automorphisms (as opposed to the whole group G_n of tree automorphisms).

LEMMA 28. — *For* $n \leqslant 0$ *and* $(\eta_1, \dots, \eta_{|n|}) \in \{-1,1\}^{|n|}$, *define an operation* $\gamma^n_{\eta_1,\dots,\eta_{|n|}}$ *on all* $2^{|n|}$*-words as follows:* $\gamma^0_\varnothing$ *is the identity on* A^1, *and for* $n \leqslant -1$, *writing every* $2^{|n|}$*-word of* $A^{2^{|n|}}$ *as the concatenation* $u{\cdot}v$ *of two* $2^{|n|-1}$*-words,*

$$\gamma^n_{\eta_1,\dots,\eta_{|n|}}(u{\cdot}v) = \begin{cases} \gamma^{n+1}_{\eta_2,\dots,\eta_{|n|}}(u) \cdot \gamma^{n+1}_{\eta_2,\dots,\eta_{|n|}}(v) & \text{if } \eta_1 = 1, \\ \gamma^{n+1}_{\eta_2,\dots,\eta_{|n|}}(v) \cdot \gamma^{n+1}_{\eta_2,\dots,\eta_{|n|}}(u) & \text{if } \eta_1 = -1. \end{cases}$$

These operations form a commutative group, called H_n, *with* $2^{|n|}$ *elements. The map* $(\eta_1, \dots, \eta_{|n|}) \mapsto \gamma^n_{\eta_1,\dots,\eta_{|n|}}$ *is a group isomorphism between* $\{-1,1\}^{|n|}$ *and* H_n.

For $\gamma \in H_n \setminus \{\mathrm{Id}\}$, there is a partition of $\{1, \dots, 2^{|n|}\}$ into $2^{|n|-1}$ pairs such that γ (considered as an element of $\mathfrak{S}_n$) acts by swapping both elements of each pair.

PROOF OF LEMMA 28. — Both properties

$$\gamma^n_{\eta_1,\dots,\eta_{|n|}} = \mathrm{Id} \qquad \Leftrightarrow \qquad (\eta_1, \dots, \eta_{|n|}) = (1, \dots, 1)$$

$$\gamma^n_{\eta'_1\eta''_1,\dots,\eta'_{|n|}\eta''_{|n|}} = \gamma^n_{\eta'_1,\dots,\eta'_{|n|}} \gamma^n_{\eta''_1,\dots,\eta''_{|n|}}$$

are readily verified by induction on n.

For $1 \leqslant k \leqslant |n|$, divide the interval $\{1, \dots, 2^{|n|}\}$ into 2^k blocks (i.e., subintervals) of length $2^{|n|-k}$. If $\eta_1 = \dots = \eta_{k-1} = 1$ and $\eta_k = -1$, then $\gamma^n_{\eta_1,\dots,\eta_{|n|}}$ sends every element of the $(2j+1)$st block into the $(2j+2)$nd block and conversely (induction on n and k, with $|n| - k$ fixed, starting from $k = 1$ and $|n| \geqslant 1$). Consequently, if $(\eta_1, \dots, \eta_{|n|}) \neq (1, \dots, 1)$, the permutation $\gamma^n_{\eta_1,\dots,\eta_{|n|}}$ has no fixed point; and as its square is Id (isomorphism with $\{-1, 1\}^{|n|}$), its orbits form a partition of $\{1, \dots, 2^{|n|}\}$ into $2^{|n|-1}$ classes of 2 elements each. ∎

PROOF OF PROPOSITION 9. — Consider a quadruply infinite supply $(E_n)_{n\leqslant 0}$, $(C_p)_{p\geqslant 1}$, $(E'_n)_{n\leqslant 0}$, $(C'_p)_{p\geqslant 1}$ of independent standard normal random variables; let f be a Borel function from $\mathbb{R}$ to A transforming the standard normal law into the uniform law on the alphabet A. For $n \leqslant 0$, define random variables ε_n in $\{-1, 1\}$ and X_n in $B_n = A^{2^{|n|}}$ by

$$\varepsilon_n = \operatorname{sgn} E_n \; ; \qquad X_n = \gamma^n_{\varepsilon_{n+1},\dots,\varepsilon_0}\big(f(C_1) \dots f(C_{2^{|n|}})\big)$$

where γ^n is the same as in Lemma 28.

The first step of the proof consists in justifying this notation by showing that the process (X, ε) has the same law as in the example studied in Theorem 4$'$. It is easy to see that for each n the law of (X_n, ε_n) is uniform on $B_n \times \{-1, 1\}$: just notice that the random word $W = f(C_1) \dots f(C_{2^{|n|}})$ is uniform on B_n and that X_n is obtained from W by a random permutation independent of W. It is also straigtforward to get X_{n+1} as the first or second half of X_n according to the value 1 or -1 of ε_{n+1}: just replace $\gamma^n_{\varepsilon_{n+1},\dots,\varepsilon_0}$ by its definition. To complete the first step, it only remains to verify that the r.v. ε_n is independent of $(X_k, \varepsilon_k)_{k<n}$. This will be done later; meanwhile, the first step is left uncompleted.

Call G (respectively G') the centered Gaussian process (E, C) (respectively (E', C')); notice that G' is an independent copy of G and that (X, ε) can be written as $\psi(G)$ for some Borel functional ψ. Define $G^\alpha = G \cos\alpha + G' \sin\alpha$, put $(X^\alpha, \varepsilon^\alpha) = \psi(G^\alpha)$, call $\mathcal{F}^\alpha$ the filtration generated by $(X^\alpha, \varepsilon^\alpha)$, and $\mathcal{H}$ the filtration generated by all processes $(X^\beta, \varepsilon^\beta)$ when β takes all possible values. With the notation used in the definition of G-cosiness, $\mathcal{F}^\alpha$ is but $\Psi^\alpha(\mathcal{F})$. The second step is to establish that each $\mathcal{F}^\alpha$ is immersed in $\mathcal{H}$; when both steps are done, the proposition will be proved. Trivially, $\mathcal{F}^\alpha$ is included in $\mathcal{H}$. As $\mathcal{F}^\alpha$ has the predictable representation property with respect to ε^α, immersion amounts to saying that ε^α_n is independent of $\mathcal{H}_{n-1}$. Showing it will a fortiori establish independence of ε^α_n and $\mathcal{F}^\alpha_{n-1}$ and, taking $\alpha = 0$, of ε_n and $\sigma\big((X_k, \varepsilon_k)_{k<n}\big)$; so by the same token step 1 will also be completed.

Fixing n from now on, it only remains to prove independence of ε^α_n and $\mathcal{H}_{n-1}$; this is equivalent to $\mathbb{E}[\varepsilon^\alpha_n | \mathcal{H}_{n-1}] = 0$, or to

$$\mathbb{E}\big[\varepsilon^\alpha_n \, h\big((X^{\beta_1}_k)_{k<n}, (\varepsilon^{\beta_1}_k)_{k<n}, \dots, (X^{\beta_q}_k)_{k<n}, (\varepsilon^{\beta_q}_k)_{k<n}\big)\big] = 0$$

for all $q > 0$, $\beta_1, \dots, \beta_q \in \mathbb{R}$ and for all bounded, Borel h.

The random variable inside the expectation is measurable for $\sigma(E, E', C, C')$; therefore it is a functional

$$\Phi(E, E', C, C') = \Phi\big((E_k)_{k\leqslant 0}, (E'_k)_{k\leqslant 0}, (C_p)_{p\geqslant 1}, (C'_p)_{p\geqslant 1}\big)$$

of infinitely many independent $\mathrm{N}(0,1)$ random variables. To prove the claim, we shall exhibit another Gaussian family $(\mathbf{E}, \mathbf{E}', \mathbf{C}, \mathbf{C}')$ of random variables, with the same law as (E, E', C, C'), but such that $\Phi(\mathbf{E}, \mathbf{E}', \mathbf{C}, \mathbf{C}') \equiv -\Phi(E, E', C, C')$; this will imply $\mathbb{E}[\Phi(E, E', C, C')] = \mathbb{E}[\Phi(\mathbf{E}, \mathbf{E}', \mathbf{C}, \mathbf{C}')] = -\mathbb{E}[\Phi(E, E', C, C')]$, whence $\mathbb{E}[\Phi(E, E', C, C')] = 0$.

Define $\mathbf{E}$ and $\mathbf{E}'$ by

$$\mathbf{E}_k = \begin{cases} E_k & \text{if } k < n \\ -E_k & \text{if } k \geqslant n \end{cases} \qquad \mathbf{E}'_k = \begin{cases} E'_k & \text{if } k < n \\ -E'_k & \text{if } k \geqslant n \end{cases}$$

and put $\mathbf{C}_p = C_{s(p)}$ and $\mathbf{C}'_p = C'_{s(p)}$, where s is the *deterministic* permutation of $\{1, 2, \dots\}$ which globally preserves each interval $\{i\,2^{|n|+1}+1, \dots, (i+1)\,2^{|n|+1}\}$, but completely reverses the order inside this interval. For instance, if $n = -2$, s is the permutation

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 & \dots \\ 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 & 16 & 15 & 14 & 13 & 12 & 11 & 10 & 9 & 24 & 23 & \dots \end{pmatrix}.$$

Observe the effect of replacing (E, E', C, C') by $(\mathbf{E}, \mathbf{E}', \mathbf{C}, \mathbf{C}')$ in the argument of Φ. For $k < n$, $\varepsilon^\alpha_k = \operatorname{sgn}(E_k \cos\alpha + E'_k \sin\alpha)$ does not change, whereas for $k \geqslant n$, $\varepsilon^\beta_k = \operatorname{sgn}(E_k \cos\beta + E'_k \sin\beta)$ is replaced by $\operatorname{sgn}(\mathbf{E}_k \cos\beta + \mathbf{E}'_k \sin\beta) = -\varepsilon^\beta_k$. And for $k < n$,

$$X^\beta_k = \gamma^k_{\varepsilon^\beta_{k+1}, \dots, \varepsilon^\beta_{n-1}, \varepsilon^\beta_n, \dots, \varepsilon^\beta_0}\big(f(C_1 \cos\beta + C'_1 \sin\beta) \dots f(C_{2^{|k|}} \cos\beta + C'_{2^{|k|}} \sin\beta)\big)$$

is replaced by

$$\gamma^k_{\varepsilon^\beta_{k+1}, \dots, \varepsilon^\beta_{n-1}, -\varepsilon^\beta_n, \dots, -\varepsilon^\beta_0}\big(f(C_{s(1)} \cos\beta + C'_{s(1)} \sin\beta) \dots f(C_{s(2^{|k|})} \cos\beta + C'_{s(2^{|k|})} \sin\beta)\big) ;$$

the claim will be established and the proposition proved if we show that the latter is equal to X^β_k itself. This is a particular instance of the general identity

$$\gamma^k_{\eta_{k+1}, \dots, \eta_{n-1}, -\eta_n, \dots, -\eta_0}(w_{s(1)} \dots w_{s(2^{|k|})}) = \gamma^k_{\eta_{k+1}, \dots, \eta_{n-1}, \eta_n, \dots, \eta_0}(w_1 \dots w_{2^{|k|}}) ,$$

valid for all $k < n \leqslant 0$. To show this identity, notice first that, since

$$\gamma^k_{\eta_{k+1}, \dots, \eta_{n-1}, -\eta_n, \dots, -\eta_0} = \gamma^k_{\eta_{k+1}, \dots, \eta_{n-1}, \eta_n, \dots, \eta_0} \gamma^k_{1, \dots, 1, -1, \dots, -1}$$

as a consequence of the group homomorphism (Lemma 28), the identity reduces to

$$\gamma^k_{1, \dots, 1, -1, \dots, -1}(w_{s(1)} \dots w_{s(2^{|k|})}) = w_1 \dots w_{2^{|k|}} .$$

For $k = n - 1$, this is just the formula

$$\gamma^{n-1}_{-1, \dots, -1}(w_{2^{|n|+1}} \dots w_1) = w_1 \dots w_{2^{|n|+1}} ,$$

which can be verified by induction on n, using the definition of γ^{n-1}; for fixed n and general $k < n$, it is proved by induction on $n-k$ using the definition of γ^k and the definition of s. The proof is now complete. ∎

REMARK. — Proposition 9 and its proof remain true if the finite alphabet A is replaced with any separable probability space $(A, \mathcal{A}, \pi)$: it suffices to replace f by a Borel function transforming the normal law into the probability π. This remark is not interesting per se (the simpler the model, the better), but will be technically useful in the next section.

Vershik's Example 3

When showing that his example is not immersible into a standard filtration, Smorodinsky's aim was to explain the same property for Vershik's example; indeed, this property carries over immediately from the former to the latter. As we shall see, G-cosiness also transfers easily.

Vershik's example is a Markov process indexed by the signed integers $\mathbb{Z}$, with state space the set $A^{\mathbb{N}}$ made of all infinite words $v = v^0 v^1 v^2 \ldots$ and endowed with the product measure μ (the letters are independent and uniform). The transition probability is

$$v^0 v^1 v^2 v^3 v^4 \ldots \longmapsto \begin{cases} v^0 v^2 v^4 \ldots & \text{with probability } \frac{1}{2} \\ v^1 v^3 v^5 \ldots & \text{with probability } \frac{1}{2}; \end{cases}$$

plainly, μ is invariant, so we have a stationary process $V = (V_n)_{n \in \mathbb{Z}}$ with this probability transition and with stationary law μ. Associated to it is the coin-tossing $(\varepsilon_n)_{n \in \mathbb{Z}}$ such that V_n is the "even half" of V_{n-1} if $\varepsilon_n = 1$ and the "odd half" if $\varepsilon_n = -1$. The process (V, ε) has the same natural filtration $\mathcal{G}$ as V, since both halves of V_{n-1} are almost surely different.

For $n \leqslant 0$, consider the word $W_n = V_n^0 V_n^1 \ldots V_n^{2^{|n|}}$ obtained from the infinite word V_n by keeping only the first $2^{|n|}$ letters. The process (W, ε) is very similar to Smorodinsky's (X, ε)-process: at each step, one half of the word is lost and one half is retained, but these halves are not the left half and right half; they are the halves consisting of letters in even or odd position. There exists for each n a (deterministic) permutation σ_n of $\{0, 1, \ldots, 2^{|n|} - 1\}$ such that the words $X_n = W_n \circ \sigma_n$ form a process with the same law as in Smorodinsky's example. The existence of those σ_n is obvious by induction; an explicit description is also possible: $\sigma_n(i)$ and i, when written as binary numbers with $|n|$ digits, are mirror-image from each other. For instance, $X_{-3} = W_{-3}^0 W_{-3}^4 W_{-3}^2 W_{-3}^6 W_{-3}^1 W_{-3}^5 W_{-3}^3 W_{-3}^7$. In this way, the process X is immersed in the process V. This really is an immersion, in the sense that the filtration $\mathcal{F}$ generated by (X, ε) becomes a sub-filtration of $\mathcal{G}$, immersed in $\mathcal{G}$; consequently, non-standardness of $\mathcal{G}$ is a corollary of non-standardness of $\mathcal{F}$.

It is almost as easy to get G-cosiness of $\mathcal{G}$ from Proposition 9. Given V, define for $n \leqslant 0$ and $0 \leqslant k < 2^{|n|}$ a random infinite word $Z_n^k \in A^{\mathbb{N}}$ by

$$Z_n^k = V_n^{\sigma_n(k)} V_n^{\sigma_n(k)+2^{|n|}} V_n^{\sigma_n(k)+2.2^{|n|}} V_n^{\sigma_n(k)+3.2^{|n|}} \ldots$$

and put $Y_n = Z_n^0 \ldots Z_n^{2^{|n|}-1}$. This Y_n can be considered as a $2^{|n|}$-word with letters Z_n^k in the new alphabet $A^{\mathbb{N}}$. This defines a process $(Y_n)_{n \leqslant 0}$ with two properties. First, as the finite word Y_n is just a rewriting of the infinite word V_n, the processes $(Y_n)_{n \leqslant 0}$ and $(V_n)_{n \leqslant 0}$ generate the same filtration. Second, by the choice of σ_n, Y_n is the first or second half of Y_{n-1} according to the value ± 1 of ε_n; in other terms, Y is an instance of Smorodinsky's process, but with the infinite alphabet $A^{\mathbb{N}}$ instead of A. Consequently, by the remark following the proof of Proposition 9, the filtration $(\mathcal{G}_n)_{n\leqslant 0}$ is G-cosy. And so is also the whole filtration $(\mathcal{G}_n)_{n\in\mathbb{Z}}$, as can be seen by inserting another Gaussian sequence $(E_n)_{n\geqslant 1}$ in the proof of Proposition 9, and putting $\varepsilon_n = \operatorname{sgn} E_n$ for $n \geqslant 1$ as well.

On a question by von Weizsäcker

The proof of Proposition 9 rests on a construction of the process X via certain random permutations of a sequence of random letters. The same construction will now be used to answer a question raised by H. von Weizsäcker during the Mini-Symposium on the Classification of Filtrations at the ESI, Vienna, December 1998.

It is well known that if $\mathcal{F} = (\mathcal{F}_n)_{n\leqslant 0}$ is a filtration and $\mathcal{B}$ a σ-field, the equality $\bigcap_n(\mathcal{F}_n\vee\mathcal{B}) = \mathcal{F}_{-\infty}\vee\mathcal{B}$ does not always hold true, but it sometimes does, for instance when $\mathcal{B}$ and $\mathcal{F}_0$ are independent; see von Weizsäcker [19] for a necessary and sufficient condition and for references. The question is, is equality always obtained when $\mathcal{F}_0$ and $\mathcal{B}$ are "almost independent"? The hope for an affirmative answer relied on the fact that all previously known examples of situations where the above formula goes wrong, had a big intersection between the σ-fields $\mathcal{F}_0$ and $\mathcal{B}$, which therefore were far from independent.

By almost independence of $\mathcal{F}_0$ and $\mathcal{B}$, we mean either a hypercontractivity inequality:

$$\exists p\in(1,2)\quad \forall B\in \mathrm{L}^p(\mathcal{B})\quad \forall F\in \mathrm{L}^p(\mathcal{F}_0)\qquad \mathbb{E}[BF]\leqslant \|B\|_{\mathrm{L}^p}\|F\|_{\mathrm{L}^p}$$

with p close to 1, or a correlation inequality:

$$\exists r\in(0,1)\quad \forall B\in \mathrm{L}^2(\mathcal{B})\quad \forall F\in \mathrm{L}^2(\mathcal{F}_0)\qquad \big|\mathrm{Cov}\,[B,F]\big|\leqslant r\,\sqrt{\mathrm{Var}\,[B]}\,\sqrt{\mathrm{Var}\,[F]}$$

with r close to 0. These inequalities are always satisfied with $p=2$ and $r=1$; they become stronger when p and r decrease, and for $p=1$ and $r=0$ each of them is equivalent to independence. The former was called *H-separation* in the remark after Proposition 8; the latter could be called C-separation, and C-cosiness could then be defined—but enough is enough!

Rotating a Gaussian process is an efficient way to generate σ-fields satisfying such inequalities. (We also find it efficient in providing some intuition about "almost independence".) If $G = (G_\lambda)_{\lambda\in\Lambda}$ is a centered Gaussian process and if G' is an independent copy of G, recall that the rotated process is $G^\alpha = G\cos\alpha + G'\sin\alpha$. The σ-fields $\sigma(G)$ and $\sigma(G^\alpha)$ satisfy the above inequalities with exponent $p = 1+|\cos\alpha|$ and correlation constant $r = |\cos\alpha|$; so G and G^α are almost independent when α is close to $\frac{\pi}{2}$. (On hypercontractivity, see for instance the references in and after the proof of Proposition 2 of [3]. On the correlation inequality, called Gebelein's inequality in this case, see Exercise V.3.13 of Revuz-Yor [12]; or Dym-McKean [6] page 66.)

The answer to von Weizsäcker's question is negative:

PROPOSITION 10. — *For all $\delta \in (0,1)$, there exist a filtration $(\mathcal{F}_n)_{n\leqslant 0}$ and a σ-field $\mathcal{B}$ such that the σ-fields $\mathcal{F}_0$ and $\mathcal{B}$ satisfy the hypercontractivity and correlation inequalities recalled above with $p = 1+\delta$ and $r = \delta$, and such that the germ σ-field $\bigcap_n(\mathcal{F}_n\vee\mathcal{B})$ contains a non-constant r.v. independent of $\mathcal{F}_{-\infty}\vee\mathcal{B}$.*

PROOF. — It is inspired from that of Proposition 9. As in that proof, start with a a quadruply infinite supply $(E_n)_{n\leqslant 0}$, $(C_q)_{q\geqslant 1}$, $(E'_n)_{n\leqslant 0}$, $(C'_q)_{q\geqslant 1}$ of independent standard normal random variables; our alphabet will have two letters: $A = \{-1,1\}$. For $n \leqslant 0$, define random variables ε_n in $\{-1,1\}$ and X_n in $B_n = A^{2^{|n|}}$ by

$$\varepsilon_n = \operatorname{sgn} E_n \ ; \qquad X_n = \gamma^n_{\varepsilon_{n+1},\dots,\varepsilon_0}\big(\operatorname{sgn}(C_1)\dots\operatorname{sgn}(C_{2^{|n|}})\big) \ ,$$

where $\gamma^{|n|}_{\varepsilon_{n+1},\dots,\varepsilon_0}$ is the (random) cube automorphism associated to the (random) signs $\varepsilon_{n+1},\dots,\varepsilon_0$ by Lemma 28. As was seen in the proof of Proposition 9, the process (X,ε) is a realization of the example studied in Theorem 4′; call $\mathcal{F} = (\mathcal{F}_n)_{n\leqslant 0}$ its natural filtration. Now rotate the Gaussian process (E,C) by putting

$$E^\alpha_n = E_n\cos\alpha + E'_n\sin\alpha \qquad \text{and} \qquad C^\alpha_q = C_q\cos\alpha + C'_q\sin\alpha$$

and fix $\alpha\in(0,\frac{\pi}{2})$ such that $\cos\alpha = \delta$. The properties of rotated Gaussian processes recalled earlier imply that, for α close to $\frac{\pi}{2}$, the σ-fields $\sigma(E,C)$ and $\sigma(E^\alpha,C^\alpha)$ are almost independent (hyercontractivity inequality with $p = 1+\cos\alpha$ and correlation inequality with $r = \cos\alpha$). This property is immediately inherited by the sub-σ-field $\mathcal{B} = \sigma\big(\operatorname{sgn}(C^\alpha_q), q\geqslant 1\big)$ of $\sigma(E^\alpha,C^\alpha)$ and the sub-σ-field $\mathcal{F}_0$ of $\sigma(E,C)$.

From now on, the parameter α is fixed in $(0,\frac{\pi}{2})$. It should be considered as close to $\frac{\pi}{2}$ (to make p close to 1 and r close to 0), but the sequel does not depend on that.

We have seen in Theorem 4′ that $\mathcal{F}_{-\infty}$ is degenerate; to prove the proposition, it remains to exhibit in the germ σ-field $\bigcap_n(\mathcal{F}_n\vee\mathcal{B})$ a non degenerate r.v. independent of $\mathcal{B}$. We shall show that the process $(\varepsilon_n)_{n\leqslant 0}$ is measurable with respect to the germ σ-field; this will prove the proposition, for this process is independent of (C,C'), and a fortiori of $\mathcal{B}$. (Notice that E' plays no rôle: all these σ-fields are included in $\sigma(E,C,C')$. The sequence E' was introduced only as a phantom, to lubricate the proof of hypercontractivity and correlation.)

So the rest of the proof will consist in showing that the knowledge of the sequence $\operatorname{sgn}(C^\alpha_1),\operatorname{sgn}(C^\alpha_2),\dots$ and of the remote past of (X,ε) is sufficient to recover the whole process ε.

Set $u_q = \operatorname{sgn}(C_q)$ and $v_q = \operatorname{sgn}(C^\alpha_q)$. The joint law of the random words $U_n = u_1\dots u_{2^{|n|}}$ and $V_n = v_1\dots v_{2^{|n|}}$ is easily described: u_q and v_q are positively correlated bits, with $\mathbb{P}[u_q\neq v_q] = \frac{\alpha}{\pi} < \frac{1}{2}$, and all pairs (u_q,v_q) are independent of each other. Observe that the word X_n is obtained from U_n by a random permutation $\gamma^n_{\varepsilon_{n+1},\dots,\varepsilon_0}$ independent of (U_n,V_n). The idea of the proof is that it is possible to recover this permutation $\gamma^n_{\varepsilon_{n+1},\dots,\varepsilon_0}$ from the observation of the words X_n and V_n with a probability of error that tends to 0 when $n\to-\infty$. The recipe to get $\gamma^n_{\varepsilon_{n+1},\dots,\varepsilon_0}$ is as follows: construct all the words γV_n, where γ ranges in H_n, match these words against X_n, and retain the γ giving the best fit, that is, the largest number of matching letters. This works well for large $|n|$ because, with high probability, U_n and V_n have more than half their letters in common, and the group

H_n is not too large. The main quantitative estimate is the following lemma, that uses the distance δ_n defined above Lemma 24. (But all what follows remains valid if the factor 2^n in the definition of δ_n is dropped, that is, if $\delta_n(w', w'')$ is defined as the number of non-matching letters in the words w' and w''.)

LEMMA 29. — *Given $\alpha \in (0, \frac{\pi}{2})$, there exists a constant $c_\alpha \in (0,1)$ such that, for all $n < 0$ and all $\gamma \in H_n \setminus \{\mathrm{Id}\}$, one has*

$$\mathbb{P}\big[\delta_n(U_n, \gamma V_n) \leqslant \delta_n(U_n, V_n)\big] \leqslant c_\alpha^{2^{|n|}} .$$

PROOF OF LEMMA 29. — Put $m = 2^{|n|}$. The event to be evaluated is $\{S \geqslant 0\}$, where

$$S = \sum_{i=1}^{m} \mathbb{1}_{\{u_i \neq v_i\}} - \sum_{i=1}^{m} \mathbb{1}_{\{u_i \neq v_{\gamma^{-1}(i)}\}} .$$

Now, since γ is in $H_n \setminus \{\mathrm{Id}\}$, Lemma 28 gives a partition of $\{1, \ldots, m\}$ into $\dfrac{m}{2}$ pairs $\{i, j\}$ such that $\gamma(i) = j$ and $\gamma(j) = i$. For such a pair $\{i, j\}$, set

$$D_{ij} = \mathbb{1}_{\{u_i \neq v_i\}} + \mathbb{1}_{\{u_j \neq v_j\}} - \mathbb{1}_{\{u_i \neq v_j\}} - \mathbb{1}_{\{u_j \neq v_i\}} .$$

The sum S is also the sum of the $m/2$ r.v.'s D_{ij} when $\{i, j\}$ ranges over all these pairs. These r.v.'s are independent and identically distributed, their law depends on α only. This law could easily be computed; we shall not need it, we only retain that $\mathbb{E}[D_{ij}] = \frac{\alpha}{\pi} + \frac{\alpha}{\pi} - \frac{1}{2} - \frac{1}{2} < 0$. The rest of the proof is Chernoff's classical argument of large deviations: Since the derivative at $t = 0$ of the Laplace transform $\phi_\alpha(t) = \mathbb{E}[\exp(tD_{ij})]$ is $\mathbb{E}[D_{ij}] < 0$, there exists $t_\alpha > 0$ such that $0 < \phi_\alpha(t_\alpha) < 1$, and it suffices to write

$$\mathbb{P}[S \geqslant 0] = \mathbb{P}[\exp(t_\alpha S) \geqslant 1] \leqslant \mathbb{E}[\exp(t_\alpha S)] = \big(\phi_\alpha(t_\alpha)\big)^{m/2} . \qquad \blacksquare$$

END OF THE PROOF OF PROPOSITION 10. — If x and v are two $2^{|n|}$-words, define $k_n(x, v)$ as the element $(\eta_1, \ldots, \eta_{|n|}) \in \{-1, 1\}^{|n|}$ that minimizes the distance $\delta_n(x, \gamma^n_{\eta_1, \ldots, \eta_{|n|}} v)$ if the minimum is uniquely achieved, and put for instance $k_n(x, v) = (1, \ldots, 1)$ if the minimum is not unique.

Consider the event $\Omega_n = \big\{k_n(X_n, V_n) = (\varepsilon_{n+1}, \ldots, \varepsilon_0)\big\}$ that the recipe described before Lemma 29 gives the correct answer. On Ω_n^{c}, there exists a (random) $\Gamma \in H_n$ such that $\Gamma \neq \gamma^n_{\varepsilon_{n+1}, \ldots, \varepsilon_0}$ and $\delta_n(X_n, \Gamma V_n) \leqslant \delta_n(X_n, \gamma^n_{\varepsilon_{n+1}, \ldots, \varepsilon_0} V_n)$. As the distance between two words is not modified when the same permutation is applied to both of them, we may let $\gamma^n_{\varepsilon_{n+1}, \ldots, \varepsilon_0}$ act on all four words in this inequality; the word X_n becomes $\gamma^n_{\varepsilon_{n+1}, \ldots, \varepsilon_0} X_n = U_n$, the word $\gamma^n_{\varepsilon_{n+1}, \ldots, \varepsilon_0} V_n$ becomes V_n, and ΓV_n becomes $\Gamma' V_n$ with a (random) $\Gamma' = \gamma^n_{\varepsilon_{n+1}, \ldots, \varepsilon_0} \Gamma \neq \mathrm{Id}$. Consequently,

$$\Omega_n^{\mathrm{c}} \subset \big\{\exists \gamma \in H_n \setminus \{\mathrm{Id}\} \quad \delta_n(U_n, \gamma V_n) \leqslant \delta_n(U_n, V_n)\big\} .$$

So, remembering that H_n has $2^{|n|}$ elements and using Lemma 29,

$$\mathbb{P}[\Omega_n^{\mathrm{c}}] \leqslant \sum_{\gamma \in H_n \setminus \{\mathrm{Id}\}} \mathbb{P}\big[\delta_n(U_n, \gamma V_n) \leqslant \delta_n(U_n, V_n)\big] \leqslant 2^{|n|} c_\alpha^{2^{|n|}} ;$$

hence the series $\sum_n \mathbb{P}[\Omega_n^{\mathrm{c}}]$ is convergent, and equality $(\varepsilon_{n+1}, \ldots, \varepsilon_0) = k_n(X_n, V_n)$ holds for all n less than some $N(\omega)$. Consequently, for each fixed $n < 0$, the whole process ε is a function of the sequence $\big(k_\ell(X_\ell, V_\ell), \ell \leqslant n\big)$; it is therefore measurable with respect to $\mathcal{F}_n \vee \mathcal{B}$. $\blacksquare$

REMARKS. — a) The filtration $\mathcal{F}$ in this proof can be replaced, with exactly the same proof, by the smaller filtration $\mathcal{X}$ generated by X. The σ-field $\mathcal{X}_{-\infty}$ is degenerate because it is included in $\mathcal{F}_{-\infty}$; and since $(\varepsilon_{n+1},\dots,\varepsilon_0) = k_n(X_n,V_n)$ for all but finitely many n, the process ε is measurable for each of the σ-fields $\mathcal{X}_n \vee \mathcal{B}$.

b) Since $\mathcal{F}_n \vee \mathcal{B}$ contains X_n and the whole process ε, it also contains $\mathcal{F}_0$, so $\mathcal{F}_n \vee \mathcal{B} = \mathcal{F}_0 \vee \mathcal{B}$ for each n and the germ σ-field $\bigcap_n (\mathcal{F}_n \vee \mathcal{B})$ too is equal to $\mathcal{F}_0 \vee \mathcal{B}$.

c) The filtration $\mathcal{F}$ in the proof of Proposition 10 is not standard (Theorem 4′); in fact, it can be considered as an arch-example of a non standard filtration. One might ask whether the situation changes when one imposes the restriction that $\mathcal{F}$ must be standard. By Vershik's lacunary isomorphism theorem (Theorem 3), the above construction easily carries over to this case, simply by taking a subsequence. And once this has been done, one can also get an example where $\mathcal{F}$ is standard non-atomic, simply by replacing $\mathcal{F}$ by its product with a standard non-atomic filtration.

d) The construction in the proof of Proposition 10 is a small modification of that of Proposition 9, which features a whole family of processes $(X^\alpha,\varepsilon^\alpha)$ isomorphic to (X,ε), each of them immersed in the filtration $\mathcal{H}$ generated by all of them. To be in a situation similar to Proposition 10, consider the setting of Proposition 9, but only in the case when the alphabet has two letters -1 and 1, and $f(x) = \operatorname{sgn} x$. Call $\mathcal{F}^\alpha$ the natural filtation of $(X^\alpha,\varepsilon^\alpha)$ and put $\mathcal{F}^{\alpha\beta}_n = \mathcal{F}^\alpha_n \vee \mathcal{F}^\beta_n$ and $\varepsilon^{\alpha\beta}_n = \varepsilon^\alpha_n \varepsilon^\beta_n$. If $\alpha-\beta$ is not a multiple of $\frac{\pi}{2}$, the same argument as above yields $(\varepsilon^{\alpha\beta}_{n+1},\dots,\varepsilon^{\alpha\beta}_0) = k_n(X^\alpha_n, X^\beta_n)$ with overwhelming probability when n is small, so for each n, the whole process $\varepsilon^{\alpha\beta}$ is measurable with respect to the σ-field $\mathcal{F}^{\alpha\beta}_n$; in other words, *this process $\varepsilon^{\alpha\beta}$ is measurable with respect to the germ σ-field $\mathcal{F}^{\alpha\beta}_{-\infty}$ of the filtration $\mathcal{F}^{\alpha\beta}$*. Consequently, still for $\alpha-\beta$ not a multiple of $\frac{\pi}{2}$, the filtration $\mathcal{F}^{\alpha\beta}$ is dyadic:

$$\mathcal{F}^{\alpha\beta}_n = \mathcal{F}^{\alpha\beta}_{n-1} \vee \sigma(\varepsilon^\alpha_n) = \mathcal{F}^{\alpha\beta}_{n-1} \vee \sigma(\varepsilon^\beta_n)\ ,$$

and one has for all $m \leqslant n \leqslant 0$ the identity $\mathcal{F}^\alpha_m \vee \mathcal{F}^\beta_n = \mathcal{F}^\beta_m \vee \mathcal{F}^\alpha_n = \mathcal{F}^\alpha_n \vee \mathcal{F}^\beta_n$. But $\mathcal{F}^{\alpha\beta}$ does not have the predictable representation property with respect to ε^α, nor to ε^β, nor to any other coin-tossing, simply because $\mathcal{F}^{\alpha\beta}_{-\infty}$ is not degenerate: we just saw that it contains $\varepsilon^{\alpha\beta}$.

Thus, for $\alpha \neq \beta \bmod \frac{\pi}{2}$, we have an example of two filtrations $\mathcal{F}^\alpha$ and $\mathcal{F}^\beta$ immersed in $\mathcal{F}^\alpha \vee \mathcal{F}^\beta$, arbitrarily close to independence as in Proposition 10, such that the three filtrations $\mathcal{F}^\alpha$, $\mathcal{F}^\beta$ and $\mathcal{F}^\alpha \vee \mathcal{F}^\beta$ are dyadic, $\mathcal{F}^\alpha_{-\infty}$ and $\mathcal{F}^\beta_{-\infty}$ are trivial, and $\bigcap_n (\mathcal{F}^\alpha_n \vee \mathcal{F}^\beta_n)$ is not trivial.

(When $|\alpha-\beta| = \frac{\pi}{2}$, the filtrations $\mathcal{F}^\alpha$ and $\mathcal{F}^\beta$ are independent, $\mathcal{F}^{\alpha\beta}_{-\infty}$ is degenerate, and $\mathcal{F}^{\alpha\beta}_n = \mathcal{F}^{\alpha\beta}_{n-1} \vee \sigma(\varepsilon^\alpha_n, \varepsilon^\beta_n)$; this is just the product situation.)

Last, observe that the filtration $\mathcal{H}$ generated by all the $(X^\alpha,\varepsilon^\alpha)$ is dyadic too, because $\mathcal{H}_n = \mathcal{H}_{n-1} \vee \sigma(\varepsilon^\alpha_n)$ for any α. Indeed, $\mathcal{H}_{-\infty}$ contains the products $\varepsilon^\alpha_n \varepsilon^\beta_n$ for all α, β and n (if $\alpha-\beta = k\frac{\pi}{2}$, write $\varepsilon^{\alpha\beta} = \varepsilon^{\alpha\gamma}\varepsilon^{\gamma\beta}$); the reason why $\mathcal{H}$ can be dyadic and at the same time contain two independent coin-tossing processes ε^0 and $\varepsilon^{\pi/2}$, is simply that their product $\varepsilon^0\varepsilon^{\pi/2}$ is $\mathcal{H}_{-\infty}$-measurable.

Conversely, it looks likely that $\mathcal{H}_{-\infty} = \sigma(\varepsilon^{0\alpha}, \alpha \in \mathbb{R})$ and, for $\alpha-\beta \notin \mathbb{Z}\frac{\pi}{2}$, $\bigcap_n (\mathcal{F}^\alpha_n \vee \mathcal{F}^\beta_n) = \sigma(\varepsilon^{\alpha\beta})$; we have not investigated this question.

References

[1] D. Aldous & M. Barlow. On countable dense random sets. *Séminaire de Probabilités XV*, Lecture Notes in Mathematics 850, Springer 1981.

[2] M.T. Barlow, M. Émery, F.B. Knight, S. Song & M. Yor. Autour d'un théorème de Tsirelson sur des filtrations browniennes et non browniennes. *Séminaire de Probabilités XXXII*, Lecture Notes in Mathematics 1686, Springer 1998.

[3] S. Beghdadi-Sakrani & M. Émery. On certain probabilities equivalent to coin-tossing, d'après Schachermayer. *Séminaire de Probabilités XXXIII*, Lecture Notes in Mathematics 1709, Springer 1999.

[4] N. Belili. Dualité du problème des marges et ses applications. *Séminaire de Probabilités XXXIII*, Lecture Notes in Mathematics 1709, Springer 1999.

[5] L. Dubins, J. Feldman, M. Smorodinsky & B. Tsirelson. Decreasing sequences of σ-fields and a measure change for Brownian motion. *Ann. Prob. 24*, 882–904, 1996.

[6] H. Dym & H.P. McKean. Gaussian Processes, Function Theory, and the Inverse Spectral Problem. Academic Press, 1976.

[7] M. Émery & W. Schachermayer. Brownian filtrations are not stable under equivalent time-changes. *Séminaire de Probabilités XXXIII*, Lecture Notes in Mathematics 1709, Springer 1999.

[8] J. Feldman. Decreasing sequences of σ-fields: product type, standard, and prestandard. ESI preprint, 1998.

[9] J. Feldman & M. Smorodinsky. Decreasing sequences of measurable partitions: product type, standard, and prestandard. *Ergodic Theory and Dynamical Systems 20,* 1079–1090, 2000.

[10] D.N. Hoover. Extending probability spaces and adapted distributions. *Séminaire de Probabilités XXVI*, Lecture Notes in Mathematics 1526, Springer 1992.

[11] S. Rachev & L. Rüschendorf. Mass Transportation Problems. Springer, 1998 (2 volumes).

[12] D. Revuz & M. Yor. Continuous Martingales and Brownian Motion. Springer 1991, 1994, 1999.

[13] V. A. Rohlin. On the fundamental ideas of measure theory. *Mat. Sb. 25,* 107–150, 1949. English translation: *A.M.S. Transl. 71,* 3–55, 1952.

[14] M. Smorodinsky. Processes with no standard extension. *Israel J. Math. 107,* 327–331, 1998.

[15] B. Tsirelson. Triple points: From non-Brownian filtrations to harmonic measures. *GAFA, Geom. funct. anal. 7,* 1096–1142, 1997.

[16] B. Tsirelson. About Yor's problem. Unpublished preprint, Tel Aviv University.

[17] A. M. Vershik. Approximation in measure theory. *Doctor Thesis,* Leningrad 1973. Expanded and updated english version: The theory of decreasing sequences of measurable partitions. *St. Petersburg Math. J. 6,* 705–761, 1995.

[18] J. Warren. On the joining of sticky Brownian motion. *Séminaire de Probabilités XXXIII*, Lecture Notes in Mathematics 1709, Springer 1999.

[19] H. von Weizsäcker. Exchanging the order of taking suprema and countable intersections of σ-algebras. *Ann. Institut H. Poincaré Section B, 19,* 91–100, 1983.

Michel Émery
Université Louis Pasteur et C.N.R.S.
I.R.M.A.
7 rue René Descartes
F 67 084 Strasbourg Cedex
`emery@math.u-strasbg.fr`

Walter Schachermayer
Technische Universität Wien
Institut für Statistik, Wahrscheinlichkeitstheorie und Versicherungsmathematik
Wiedner Hauptstraße 8–10 / E 107–5
A 1040 Wien
`walter.schachermayer@fam.tuwien.ac.at`