The Annals of Probability
1996, Vol. 24, No. 2, 882–904

DECREASING SEQUENCES OF σ-FIELDS AND A MEASURE CHANGE FOR BROWNIAN MOTION

By Lester Dubins, Jacob Feldman,[1] Meir Smorodinsky and Boris Tsirelson[2]

University of California at Berkeley, University of California at Berkeley, Tel Aviv University and Tel Aviv University

Dedicated to Anatoly Vershik on the occasion of his 60th birthday

Let $(\mathscr{F}_t)_{t\geq 0}$ be the filtration of a Brownian motion $(B(t))_{t\geq 0}$ on $(\Omega, \mathscr{F}, P)$. An example is given of a measure $Q \sim P$ (in the sense of absolute continuity) for which $(\mathscr{F}_t)_{t\geq 0}$ is *not* the filtration of any Brownian motion on $(\Omega, \mathscr{F}, Q)$. This settles a 15-year-old question.

1. Introduction. Let Ω be the Borel space of real continuous functions on $[0,\infty)$ which are 0 at 0, and P the Wiener measure, that is, the probability measure which makes the coordinate process $(B(t))_{t\geq 0}$ a standard Brownian motion. Let Q be another probability measure equivalent to P.

Problem 0. Given such a Q, is it always possible to define on Ω a process $(B'(t))_{t\geq 0}$ such that under Q, B' is a standard Brownian motion and $B'(t)$ is $\mathscr{F}_t$-measurable for each t, where $\mathscr{F}_t$ is the σ-field generated by $\{B(s)\colon 0 < s \leq t\}$?

If we denote by $\mathscr{F}_t'$ the corresponding σ-field of B', then this amounts to the requirement that $\mathscr{F}_t' \subset \mathscr{F}_t$ for all $t > 0$. (All σ-fields are supposed complete, that is, containing all sets of zero measure.)

The system $(\Omega, \mathscr{F}, P, (\mathscr{F}_t)_{t\geq 0})$ is frequently called a *Brownian filtration*. Of course, one may speak more generally about the filtration of a stochastic process, or a filtered probability space.

Problem 1. Given such a Q, is it always possible to define B' as required in Problem 0 and such that $\mathscr{F}_t' = \mathscr{F}_t$ for all $t > 0$? That is, is $(\Omega, \mathscr{F}, Q, (\mathscr{F}_t)_{t\geq 0})$ a Brownian filtration?

Received October 1994; revised June 1995.

[1]Supported in part by NSF Grant DMS-91-13642.
[2]Supported by the Rashi Foundation.
AMS 1991 *subject classifications*. Primary 60J65; secondary 28C20, 60G07, 60H10.
Key words and phrases. Brownian filtration, equivalent measure, decreasing sequence of measurable partitions.

Problem 0 has an elegant solution given by Girsanov's formula [see Revuz and Yor (1991), 8.1, or Protter (1990), 3.6]:

$$B'(t) = B(t) - \int_0^t \Phi(s)\, ds,$$

where $(\Phi(t))_{t \geq 0}$ is the process, adapted to the Brownian filtration, which arises in the Cameron–Martin representation of the Radon–Nikodym derivative:

$$\frac{dQ}{dP} = \exp\left(\int_0^\infty \Phi(t)\, dB(t) - \frac{1}{2}\int_0^\infty \Phi^2(t)\, dt\right).$$

(The Girsanov process is a martingale even with respect to the filtration of B.)

However, Tsirelson (1975) constructed an example of a measure Q such that the process defined by the Girsanov formula does *not* solve Problem 1.

If such B' as required in Problem 1 existed, then the Girsanov process would have a representation

$$dB - \Phi\, ds = \Psi\, dB',$$

with a process Ψ adapted to $(\mathscr{F}_t)_{t \geq 0}$ and assuming only the values ± 1. So Problem 1 may be viewed as the problem of finding, for given Φ, a process Ψ adapted to $(\mathscr{F}_t)_{t \geq 0}$, taking on only the values ± 1, and such that the process

$$B'(t) = \int_0^t \Psi(s)\, dB(s) - \int_0^t \Psi(s)\Phi(s)\, ds$$

has the property $\mathscr{F}_t' = \mathscr{F}_t$.

Problem 1 was pointed out in Stroock and Yor (1980), Question (G), page 161, and repeated in Revuz and Yor (1991), Question 1, page 336. We are grateful to Marc Yor for bringing it to our attention.

In the present paper we show that Problem 1 has a negative solution.

It should be mentioned that in Theorem 7 of Skorokhod (1986) a general assertion is made about the structure of filtrations which, taking what we regard as a reasonable interpretation of that assertion, would imply a *positive* solution to Problem 1. Thus the present paper contradicts this interpretation of the assertion. Although it was not our original intention to elaborate on this, the editors and referees have strongly urged that we do so; therefore, such a discussion is appended as Section 6.

2. Tools. Our main tool is borrowed from the theory of decreasing sequences of measurable partitions (in other words, σ-fields) developed by Vershik in a sequence of works [Vershik (1968, 1970, 1971, 1973, 1994)]. Here, for brevity, we will call them "reverse filtrations." Throughout, either by assumption or by construction, all measure spaces will be Lebesgue spaces. The reader may recall that all "reasonable" constructions with Lebesgue spaces, such as quotients and products, produce Lebesgue spaces; see Rokhlin (1949).

DEFINITION 2.1. A reverse filtration $(\mathscr{X}, \mathbf{F}, m)$ is a probability space $(\mathscr{X}, \mathscr{F}, m)$ equipped with a decreasing sequence $\mathbf{F} = (\mathscr{F}_n)_{n=0}^{\infty}$ of σ-fields: $\mathscr{F} = \mathscr{F}_0 \supset \mathscr{F}_1 \supset \mathscr{F}_2 \supset \cdots$ (all σ-fields are taken to be complete for m) such that $\bigcap_n \mathscr{F}_n$ is the trivial σ-field consisting of the sets of measure 0 and 1.

The natural notion of isomorphism in the class of reverse filtrations is as follows.

DEFINITION 2.2. A reverse filtration $(\mathscr{X}, \mathbf{F}, m)$ is *isomorphic* to $(\mathscr{X}', \mathbf{F}', m')$, if there exists an almost one-to-one map $\varphi: \mathscr{X} \to \mathscr{X}'$ sending m to m', and such that $\varphi^{-1}\mathscr{F}_n' = \mathscr{F}_n$.

Any sequence of random variables $X_1, X_2, \dots$ on $(\mathscr{X}, \mathscr{F}, m)$ determines a decreasing sequence of σ-fields: $\mathscr{F}_n$ is the σ-field generated by the variables $X_{n+1}, X_{n+2}, \dots$. If $\mathscr{F}_0 = \mathscr{F}$ and the intersection of the $\mathscr{F}_n$ is trivial, then we call $(\mathscr{X}, (\mathscr{F}_n)_{n=0}^{\infty}, m)$ the reverse filtration *generated* by $(X_n)_{n=1}^{\infty}$.

The following definition is analogous to one for filtrations which will be given in Section 6.

DEFINITION 2.3. An *extension* of a reverse filtration $(\mathscr{X}, \mathbf{F}, m)$ consists of a reverse filtration $(\tilde{\mathscr{X}}, \tilde{\mathbf{F}}, \tilde{m})$ and a measure-preserving map π from $\tilde{\mathscr{X}}$ to $\mathscr{X}$ with $\pi^{-1}(\mathscr{F}_n) \subset \tilde{\mathscr{F}}_n$, $n = 0, 1, \dots,$ such that the equality

$$\tilde{E}\big(X \circ \pi | \tilde{\mathscr{F}}_n\big) = \big(E(X|\mathscr{F}_n)\big) \circ \pi$$

holds almost everywhere for any bounded measurable function $X: \mathscr{X} \to \mathbb{R}$.

LEMMA 2.4. *Let a reverse filtration* $(\mathscr{X}, \mathbf{F}, \lambda)$ *be given, and an extension* $(\tilde{\mathscr{X}}, \tilde{\mathbf{F}}, \tilde{\lambda})$, $\pi: \tilde{\mathscr{X}} \to \mathscr{X}$. *Let* m *be a measure on* $\mathscr{X}$ *equivalent to* λ. *Define a measure* $\tilde{m}$ *on* $\tilde{\mathscr{X}}$, *equivalent to* $\tilde{\lambda}$, *by the equality*

$$\frac{d\tilde{m}}{d\tilde{\lambda}} = \left(\frac{dm}{d\lambda}\right) \circ \pi.$$

Then $(\tilde{\mathscr{X}}, \tilde{\mathbf{F}}, \tilde{m})$, π *form an extension of* $(\mathscr{X}, \mathbf{F}, m)$.

The proof is left to the reader.

Vershik called one of his decreasing sequences of σ-fields standard if it was generated by a sequence of independent random variables, each variable being either nonatomic or purely atomic with all its atoms having the same probability. But we prefer to focus on the nonatomic case.

DEFINITION 2.5. A reverse filtration will be called *standard* if it is generated by a sequence of independent nonatomic random variables.

Clearly, any reverse filtration isomorphic to a standard one is also standard, and all standard reverse filtrations are isomorphic. Nonstandard reverse filtrations exist trivially, since, in general, conditional measures may contain atoms. A reverse filtration will be called *nonatomic* if for each $n \in \mathbb{N}$ the restriction of the measure m to $\mathscr{F}_{n-1}$, when conditioned by $\mathscr{F}_n$, is a.e. nonatomic; in this case the measures obtained by conditioning m on $\mathscr{F}_n$ are likewise a.e. nonatomic.

For any two reverse filtrations $(\mathscr{X}, \mathbf{F}, m)$ and $(\mathscr{X}', \mathbf{F}', m')$, their product $(\mathscr{X} \times \mathscr{X}', \mathbf{F} \otimes \mathbf{F}', m \otimes m')$ is a reverse filtration; here $\mathbf{F} \otimes \mathbf{F}' = (\mathscr{F}_n \otimes \mathscr{F}_n')_{n=0}^{\infty}$. If at least one of the given reverse filtrations is nonatomic, then the product is also nonatomic.

The product $(\mathscr{X} \times \mathscr{X}', \mathbf{F} \otimes \mathbf{F}', m \otimes m')$ together with the map $\pi: \mathscr{X} \times \mathscr{X}' \to \mathscr{X}$, $\pi(\omega, \omega') = \omega$ may be treated as a special case of an extension as in Definition 2.3. Thus every reverse filtration admits a nonatomic extension.

Does every reverse filtration have a *standard* extension? The answer is no, but this is far from evident. Vershik was first to construct examples of reverse filtrations admitting no standard extensions. Here is a description of one.

EXAMPLE [Vershik (1973)]. It is possible to construct a Markov chain $X_1, X_2, \dots$ which generates a reverse filtration with no standard extension. Each X_n is distributed uniformly on $[0, 1]$. When $X_n, X_{n+1}, \dots$ are given, the conditional distribution of X_n depends on X_{n+1} only, and in the following manner: if X_{n+1} takes the value x, which we write in binary as $.x_1 x_2 x_3 \dots$, then X_n is equal to either $.x_1 x_3 x_5 \dots$ or $.x_2 x_4 x_6 \dots$, each with probability $1/2$. A proof that this indeed generates a reverse filtration with no standard extension may be found in Smorodinsky (1995).

A standard reverse filtration obviously has a standard extension, namely itself. We note that conversely any *nonatomic* reverse filtration with a standard extension is itself standard; but we will not make use of this fact, and therefore will not prove it.

Denote by λ the Bernoulli measure on the space $\mathscr{X} = \{0, 1\}^{\mathbb{N}}$, which makes the coordinate functions $X_1, X_2, \dots$ independent random variables taking only the values 0 and 1, each with probability $1/2$, that is, a *Bernoulli sequence*. The "Bernoulli reverse filtration" $(\mathscr{X}, \mathbf{F}, \lambda)$ generated by $(X_n)_{n=1}^{\infty}$ is of course atomic. Its nonatomic extension, constructed by taking its direct product, in the obvious sense, with a standard reverse filtration, is easily seen to be standard.

THEOREM 2.6. *There is a measure m equivalent to λ for which the reverse filtration $(\mathscr{X}, \mathbf{F}, m)$ admits no standard extension.*

This theorem will provide a negative solution to Problem 1 as follows.

Fix a sequence $t_1 > t_2 > \cdots \to 0$, consider the independent increments $B(t_n) - B(t_{n+1})$ of Brownian motion and let

$$X_n = \begin{cases} 1, & \text{when } B(t_n) - B(t_{n+1}) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

The X_n form a Bernoulli sequence, so the map $\pi\colon \omega \mapsto (X_1(\omega), X_2(\omega), \dots)$ is a measure-preserving map from (Ω, P) to the $(\mathscr{X}, \lambda)$ of Theorem 2.6. Moreover, the reverse filtration $(\Omega, (\mathscr{F}_t)_{n=0}^{\infty}, P)$ and the map π form an extension of $(\mathscr{X}, \mathbf{F}, \lambda)$. According to Lemma 2.4, $(\Omega, (\mathscr{F}_{t_n})_{n=0}^{\infty}, Q)$ and π form an extension of $(\mathscr{X}, \mathbf{F}, m)$, where Q is the measure equivalent to P defined by the equality

$$\frac{dQ}{dP} = \left(\frac{dm}{d\lambda}\right) \circ \pi.$$

Since $(\mathscr{X}, \mathbf{F}, m)$ admits no standard extension, we conclude that the reverse filtration $(\Omega, (\mathscr{F}_{t_n})_{n=0}^{\infty}, Q)$ is nonstandard. But if the filtered probability space $(\Omega, \mathscr{F}, Q, (\mathscr{F}_t)_{t \ge 0})$ were generated by a Brownian motion B', then one could choose independent random variables $Y_1, Y_2, \dots$ on $(\Omega, \mathscr{F}, Q)$ with each Y_n generating the σ-field generated by $\{B'(t) - B'(s)\colon t_n > t > s > t_{n+1}\}$, which would imply standardness of $(\Omega, (\mathscr{F}_{t_n})_{n=1}^{\infty}, Q)$. Thus there can be no such B'.

Our main tool for proving Theorem 2.6 is Vershik's criterion for standardness [Vershik (1970)]. Here we will state it in terms of multistep Kantorovich–Rubinstein metrics, to be defined below.

Let $(\mathscr{X}, \mathscr{F}, m)$ be a probability space and f a measurable function from $\mathscr{X}$ to a compact metric space with metric ρ. This determines a pseudometric on $\mathscr{X}$: $\rho_f(x, y) = \rho(f(x), f(y))$. The class of all such ρ_f is exactly the class of all measurable precompact pseudometrics on $\mathscr{X}$. *From here on, the term "pseudometric" will always mean pseudometric of this kind.* The pseudometric itself is then measurable on $(\mathscr{X} \times \mathscr{X}, \mathscr{F} \otimes \mathscr{F})$. We identify two such pseudometrics if there is a subset of $\mathscr{X}$ of full measure such that they agree on all pairs of points which both lie in this subset. A pseudometric ρ determines the set $\operatorname{Lip}(\rho)$ of all real-valued functions l on $\mathscr{X}$ satisfying the Lipschitz condition

$$|l(x) - l(y)| \le \rho(x, y)$$

for all $x, y \in \mathscr{X}$. Such functions are necessarily measurable. It is more convenient to deal with equivalence classes rather than individual functions, so we weaken the Lipschitz condition to hold only on a set of measure 1; this makes it insensitive to a change of l within its equivalence class, as well as a change of ρ within its equivalence class (defined above). Then

$$\operatorname{Lip}(\rho) \subset L_\infty(m).$$

Furthermore, $\operatorname{Lip}(\rho)$ determines ρ uniquely (up to the stipulated equivalence).

If ρ is $\mathscr{F}' \otimes \mathscr{F}'$-measurable, $\mathscr{F}'$ being a sub-σ-field of $\mathscr{F}$, then $\operatorname{Lip}(\rho) \subset L_\infty(\mathscr{X}, \mathscr{F}', m)$. Moreover, ρ may be transferred to the quotient space $\mathscr{X}/\mathscr{F}'$, and $\operatorname{Lip}(\rho)$ may be calculated on the quotient space, giving essentially the same result.

Let ρ be a pseudometric on $\mathscr{X}$. Given two probability measures μ, ν on $\mathscr{X}$, the *Kantorovich–Rubinstein distance* [Kantorovich and Rubinstein (1958); see also Dudley (1989)] is defined as

$$\rho_{\mathrm{KR}}(\mu, \nu) = \sup\left\{\left|\int l\, d\mu - \int l\, d\nu\right| : l \in \operatorname{Lip}(\rho)\right\}.$$

Another (equivalent) definition will be mentioned and used later; but in the present section only the above definition will be used.

If ρ is $\mathscr{F}' \otimes \mathscr{F}'$-measurable, $\mathscr{F}' \subset \mathscr{F}$, then $\rho_{\mathrm{KR}}(\mu, \nu)$ depends only on the restrictions $\mu|\mathscr{F}'$, $\nu|\mathscr{F}'$. Moreover, $\rho_{\mathrm{KR}}(\mu, \nu)$ may equally well be calculated on the quotient space $\mathscr{X}/\mathscr{F}'$.

The following multistep counterpart of the KR construction is crucial. For any reverse filtration $(\mathscr{X}, \mathbf{F}, m)$ and any given pseudometric ρ on $\mathscr{X}$, we define recursively a sequence of pseudometrics (ρ_n), ρ_n being $\mathscr{F}_n \otimes \mathscr{F}_n$-measurable: $\rho_0 = \rho$, and recursively $\rho_n(x, y)$ is the $(\rho_{n-1})_{\mathrm{KR}}$ distance between the conditional measures, given $\mathscr{F}_n$, corresponding to x and y; these conditional measures exist, since we have assumed that $(\mathscr{X}, \mathscr{F}, m)$ is a Lebesgue space. Due to the $\mathscr{F}_{n-1} \otimes \mathscr{F}_{n-1}$-measurability of ρ_{n-1}, the result will be the same if the conditioning is done for the pair $\mathscr{F}_{n-1}$ and $\mathscr{F}_n$ rather than $\mathscr{F}$ and $\mathscr{F}_n$. Restating the definition symbolically, $\rho_0 = \rho$ and

$$\rho_n(x, y) = \sup\{|E(l|\mathscr{F}_n)(x) - E(l|\mathscr{F}_n)(y)| : l \in \mathrm{Lip}(\rho_{n-1})\}.$$

The conditional measures may be changed on a set of measure 0; but because of the identification we have made between pseudometrics, such a change has no significance. The usual caution is required when taking a supremum of uncountably many equivalence classes: either a dense countable subset should be selected, or the supremum should not be treated pointwise. It is easily seen that precompactness and measurability continue to hold.

Next we define a sequence of numbers α_n associated with ρ and $(\mathscr{X}, \mathbf{F}, m)$:

$$\alpha_n(\rho) = \iint \rho_n(x, y)\, dm(x)\, dm(y);$$

then $\alpha_0 \ge \alpha_1 \ge \alpha_2 \ge \cdots \ge 0$. Their limit $\alpha_\infty(\rho)$ is used in Vershik's criterion of standardness [Vershik (1970)]. His sufficiency statement, specialized to the nonatomic case, is reproduced below as Theorem 2.7(a). However, what we actually need is a necessity statement, given for arbitrary reverse filtrations as part (b) of the same theorem.

THEOREM 2.7. (a) *If a nonatomic reverse filtration $(\mathscr{X}, \mathbf{F}, m)$ satisfies the condition $\alpha_\infty(\rho) = 0$ for every pseudometric ρ on $\mathscr{X}$, then the reverse filtration is standard.*

(b) *If a reverse filtration $(\mathscr{X}, \mathbf{F}, m)$ admits a standard extension, then $\alpha_\infty(\rho) = 0$ for every pseudometric ρ on $\mathscr{X}$.*

In order to make our presentation self-contained, a proof of (b) follows.

LEMMA 2.8. *For every standard reverse filtration $(\mathscr{X}, \mathbf{F}, m)$ and every pseudometric ρ on $\mathscr{X}$, the corresponding $\alpha_\infty(\rho) = 0$.*

PROOF. A standard reverse filtration is generated by independent random variables. Without loss of generality, we suppose that they are distributed uniformly on $[0, 1]$, and $\mathscr{X}$ is the space of sequences. Consider metrics

$\rho_0, \rho_1, \dots$ constructed from a given $\rho_0 = \rho$. The first n coordinates actually do not affect the value of ρ_n, due to its $\mathscr{F}_n$-measurability.

REMARKS ON NOTATION. Here and henceforth we use the notation $x_1^{n-1} = (x_1, \dots, x_{n-1})$; $y_n^\infty = (y_n, y_{n+1}, \dots)$. Also, here and in other places we will deliberately abuse notation by sometimes writing $\rho_n(x, y)$ and sometimes $\rho_n(x_{n+1}^\infty, y_{n+1}^\infty)$. It seemed preferable to do this rather than introduce more new notation.

Returning to the argument, it will be shown by induction on n that, for all $n \in \mathbb{N}$,

$$\rho_{n-1}(y_n^\infty, z_n^\infty) \le \int \rho\big((x_1^{n-1}, y_n^\infty), (x_1^{n-1}, z_n^\infty)\big)\, dx_1^{n-1}$$

for any y_n^∞, z_n^∞, all scalar variables running over $[0, 1]$.

For $n = 1$ there is nothing to check. Suppose, for some $n \ge 1$, that the inequality holds. It suffices to prove that

$$\rho_n(y_{n+1}^\infty, z_{n+1}^\infty) \le \int \rho_{n-1}((x_n, y_{n+1}^\infty), (x_n, z_{n+1}^\infty))\, dx_n.$$

Let a function l be ρ_{n-1}-Lipschitz. Then

$$\begin{aligned}
&|E(l|\mathscr{F}_n)(y_{n+1}^\infty) - E(l|\mathscr{F}_n)(z_{n+1}^\infty)| \\
&\quad = \left| \int l(x_n, y_{n+1}^\infty)\, dx_n - \int l(x_n, z_{n+1}^\infty)\, dx_n \right| \\
&\quad \le \int |l(x_n, y_{n+1}^\infty) - l(x_n, z_{n+1}^\infty)|\, dx_n \\
&\quad \le \int \rho_{n-1}((x_n, y_{n+1}^\infty), (x_n, z_{n+1}^\infty))\, dx_n.
\end{aligned}$$

It follows that

$$\begin{aligned}
\alpha_n(\rho) &= \int \rho_n(y_{n+1}^\infty, z_{n+1}^\infty)\, dy_{n+1}^\infty\, dz_{n+1}^\infty \\
&\le \int \rho((x_1^n, y_{n+1}^\infty), (x_1^n, z_{n+1}^\infty))\, dx_1^n\, dy_{n+1}^\infty\, dz_{n+1}^\infty.
\end{aligned}$$

Denote the right-hand side by $\beta_n(\rho)$.

For any two measurable functions f, g from $\mathscr{X}$ to a compact metric space with metric ρ, we have

$$\begin{aligned}
&|\rho(f(x_1^n, y_{n+1}^\infty), f(x_1^n, z_{n+1}^\infty)) - \rho(g(x_1^n, y_{n+1}^\infty), g(x_1^n, z_{n+1}^\infty))| \\
&\qquad \le \rho(f(x_1^n, y_{n+1}^\infty), g(x_1^n, y_{n+1}^\infty)) + \rho(f(x_1^n, z_{n+1}^\infty), g(x_1^n, z_{n+1}^\infty));
\end{aligned}$$

hence

$$|\beta_n(\rho_f) - \beta_n(\rho_g)| \le 2 \int \rho(f(x_1^\infty), g(x_1^\infty))\, dx_1^\infty$$

and

$$\left| \beta_\infty(\rho_f) - \beta_\infty(\rho_g) \right| \le 2\int_{\mathscr{X}} \rho(f(x), g(x))\, dx.$$

So it will suffice to show that $\beta_\infty(\rho_f) = 0$ for a dense set of functions f. If f depends on only finitely many coordinates, then there is some n such that $f(x_1^\infty) = f(x_1^n)$ and $\beta_n(\rho_f) = 0$. Any measurable function from $\mathscr{X}$ to a compact metric space may be approximated uniformly by a measurable function taking only finitely many values, and such a function may be approximated in probability by a function depending on only finitely many coordinates. This completes the proof. □

LEMMA 2.9. *Let* $(\mathscr{X}, \mathbf{F}, m)$ *be a reverse filtration and* $(\tilde{\mathscr{X}}, \tilde{\mathbf{F}}, \tilde{m}), \pi$ *an extension. For any pseudometric* ρ *on* $\mathscr{X}$, *define* $\tilde{\rho}$ *by* $\tilde{\rho}(\tilde{x}, \tilde{y}) = \rho(\pi\tilde{x}, \pi\tilde{y})$. *Then* $\alpha_\infty(\rho) = \alpha_\infty(\tilde{\rho})$.

PROOF. It will be shown, by induction on n, that $\tilde{\rho}_n(\tilde{x}, \tilde{y}) = \rho_n(\pi\tilde{x}, \pi\tilde{y})$ for any $\tilde{x}, \tilde{y} \in \tilde{\mathscr{X}}$; here $(\tilde{\rho}_n)$ is the sequence of metrics on $\tilde{\mathscr{X}}$ obtained from $\tilde{\rho}_0 = \tilde{\rho}$ by the multistep Kantorovich–Rubinstein construction on $(\tilde{\mathscr{X}}, \tilde{\mathbf{F}}, \tilde{m})$. For $n = 0$, it is true by definition. Suppose that it is true for some $n \ge 1$. Then any $\tilde{\rho}_n$-Lipschitz function $\tilde{l}$ on $\tilde{\mathscr{X}}$ is of the form $l \circ \pi$ with a ρ_n-Lipschitz function l on $\mathscr{X}$. Hence (see Definition 2.3)

$$\tilde{E}\left(\tilde{l} | \tilde{\mathscr{F}}_{n+1}\right) = E(l | \mathscr{F}_{n+1}) \circ \pi.$$

It follows that

$$\left| \tilde{E}\left(\tilde{l} | \tilde{\mathscr{F}}_{n+1}\right)(\tilde{x}) - \tilde{E}\left(\tilde{l} | \tilde{\mathscr{F}}_{n+1}\right)(\tilde{y}) \right| = \left| E(l | \mathscr{F}_{n+1})(x) - E(l | \mathscr{F}_{n+1})(y) \right|,$$

where $x = \pi\tilde{x}$, $y = \pi\tilde{y}$; the supremum in l gives

$$\tilde{\rho}_{n+1}(\tilde{x}, \tilde{y}) = \rho_{n+1}(x, y).$$

Integration in $\tilde{x}, \tilde{y}$ gives

$$\alpha_n(\tilde{\rho}) = \alpha_n(\rho);$$

hence $\alpha_\infty(\tilde{\rho}) = \alpha_\infty(\rho)$, completing the argument. □

Statement (b) of Theorem 2.7 follows directly from Lemmas 2.8 and 2.9.

3. Constructing the measure on sequences.

Here and henceforth $X_1, X_2, \dots$ *denote the coordinate functions on* $\mathscr{X} = \{0, 1\}^{\mathbb{N}}$, *treated as random variables on* $(\mathscr{X}, m)$.

The measure m on $\mathscr{X} = \{0, 1\}^{\mathbb{N}}$, mentioned in Theorem 2.6, will be described by its conditional probabilities $m(X_i = 1 | X_{i+1}^\infty)$ for each $i = 1, 2, \dots$. We restrict ourselves to the case

$$m\left(X_i = 1 | X_{i+1}^\infty\right) = \frac{1 \pm \delta_i}{2},$$

where $\delta_1, \delta_2, \dots$ are numbers in $(0,1)$. Alternately,

$$m(X_i = 1 | X_{i+1}^\infty) = \tfrac{1}{2}(1 + \delta_i \tau_i(X_{i+1}^\infty)),$$

where each τ_i is a Borel function, taking only two values ± 1, defined on the space of $0, 1$-sequences.

In the next lemma, although we prove existence and uniqueness, it is actually only the existence part which will be needed.

LEMMA 3.1. *If $\delta_1, \delta_2, \dots$ are in $(0,1)$, and $\sum \delta_i^2 < \infty$, then for any functions $\tau_1, \tau_2, \dots$ as above there exists one and only one measure m equivalent to the Bernoulli-$1/2$ measure λ and such that*

$$m(X_i = 1 | X_{i+1}^\infty) = \tfrac{1}{2}(1 + \delta_i \tau_i(X_{i+1}^\infty))$$

almost surely for each i.

PROOF. It is more convenient to deal with the values $-1, 1$ instead of $0, 1$. Let m be a measure equivalent to λ. Take $D = dm/d\lambda$, $D_i = \mathbb{E}_\lambda(D|\mathscr{F}_i)$. Then (D_i) is a reverse martingale, and $D_i \to 1$ λ-almost everywhere. Since

$$m(X_i = 1 | X_{i+1}^\infty) = \frac{D_i(1, X_{i+1}^\infty)}{D_i(-1, X_{i+1}^\infty) + D_i(1, X_{i+1}^\infty)} = \frac{1}{2} \frac{D_i(1, X_{i+1}^\infty)}{D_{i+1}(X_{i+1}^\infty)},$$

the needed equality takes the form

$$\frac{D_i(x_i, x_{i+1}^\infty)}{D_{i+1}(x_{i+1}^\infty)} = 1 + \delta_i x_i \tau_i(x_{i+1}^\infty),$$

which is equivalent to

$$D(x_1^\infty) = \prod_{i=1}^\infty (1 + \delta_i x_i \tau_i(x_{i+1}^\infty))$$

(λ-almost everywhere). The uniqueness of m is thus proved; a proof of its existence follows.

Each sequence (x_i) of ± 1's determines a new sequence (y_i):

$$y_i = x_i \tau_i(x_{i+1}^\infty);$$

thus a map $\varphi\colon \mathscr{X} \to \mathscr{X}$ is defined, $\mathscr{X} = \{-1, 1\}^{\mathbb{N}}$, $\varphi(x) = y$. The map φ is perhaps not one-to-one, but nevertheless φ sends λ to λ. Indeed, for any $i_1 < \cdots < i_n$,

$$\int y_{i_1} \cdots y_{i_n}\, d(\varphi(\lambda)) = \int x_{i_1} \tau_{i_1}(x_{i_1+1}^\infty) \cdots x_{i_n} \tau_{i_n}(x_{i_n+1}^\infty)\, d\lambda$$

$$= \int x_{i_1} \times \left(\text{some function of } x_{i_1+1}^\infty\right) d\lambda = 0.$$

The infinite product

$$\prod_{i=1}^\infty (1 + \delta_i y_i)$$

converges λ-almost everywhere to the density $d\mu/d\lambda$ of a product measure μ equivalent to λ; see Kakutani (1948), Sections 8 and 10. Hence the infinite product

$$\prod_{i=1}^{\infty}\left(1+\delta_i x_i \tau_i(x_{i+1}^{\infty})\right)$$

converges λ-almost everywhere to the density of a measure m equivalent to λ, completing the argument. □

From now on, both in this section and the two to follow, we will always have *everywhere defined* conditional probabilities: there will be no ambiguity on sets of measure 0. Then, similarly, all pseudometrics and their corresponding Kantorovich–Rubinstein sequences will be unambiguously defined everywhere. To this end, we impose the two following conditions on the measure m on $\mathscr{X}=\{0,1\}^{\mathbb{N}}$ with which we work:

1. Nondegeneracy: each cylinder set has positive measure; that is,

$$m(X_1^n = x_1^n) > 0$$

   for all n and $x_1^n \in \{0,1\}^n$.
2. Finite range dependence: for any $n\in\mathbb{N}$ there is $N\in\mathbb{N}$ such that X_1^n and X_N^∞ are conditionally independent, given X_{n+1}^{N-1}.

These two conditions will be satisfied in the examples which will be constructed. They could actually be eliminated from any general statement we make, provided everything were stated with enough reservations. However, this will not be needed, so the matter will not be pursued.

We now specialize to the "block-Markov" case with exponentially large blocks. That is,

$$\frac{dm}{d\lambda}(x)=\prod_{k=0}^{\infty}\frac{p_k(x^{(k)}|x^{(k+1)})}{(1/2)^{2^k}};$$

here $x=(x_1,x_2,\dots)\in\mathscr{X}=\{0,1\}^{\mathbb{N}}$; for $k=0,1,\dots$ we denote by $x^{(k)}$ the following piece of the sequence x:

$$x^{(k)}=(x_{2^k},x_{2^k+1},\dots,x_{2^{k+1}-1})\in\mathscr{X}^{(k)}=\{0,1\}^{2^k}$$

and each p_k is a Markovian transition probability from $\mathscr{X}^{(k+1)}$ to $\mathscr{X}^{(k)}$:

$$\forall\, z\in\mathscr{X}^{(k+1)},\qquad \sum_{y\in\mathscr{X}^{(k)}} p_k(y|z)=1.$$

To avoid writing nested exponents we will often write n instead of 2^k. *The reader is asked to keep in mind this implicit dependence of n on k*. Each p_k may be considered as a family of 2^{2n} probability measures on $\{0,1\}^n$ indexed by points of $\{0,1\}^{2n}$.

Choose a sequence of numbers $\varepsilon_0, \varepsilon_1, \dots$ in $(0,1)$. Set $\delta_i = \varepsilon_k$ for $2^k \le i < 2^{k+1}$. For all measures μ arising as one of the $p(\cdot|z)$, we will so arrange matters that the only values taken by $\mu(X_i = 1|X_{i+1}^n)$ are of the form

$$\mu(X_i = 1|X_{i+1}^n) = \frac{1 \pm \varepsilon_k}{2}$$

everywhere for $i = 1, \dots, n$. Due to the block-Markov nature of m, we then also have

$$m(X_i = 1|X_{i+1}^\infty) = \frac{1 \pm \varepsilon_k}{2}$$

for $2^k \le i < 2^{k+1}$. Now Lemma 3.1 may be applied, with $\delta_i = \varepsilon_k$ for $i = 2^k, \dots, 2^{k+1} - 1$: if the condition

$$\text{(i*)} \qquad \sum_k 2^k \varepsilon_k^2 < \infty$$

is satisfied, then the infinite product for $dm/d\lambda$ converges λ-almost everywhere to the density of a measure m equivalent to λ.

In the next two sections we will show that if the additional condition

$$\text{(ii*)} \qquad \sum_k \frac{1}{2^k \varepsilon_k} < \infty$$

is satisfied, then the transition probabilities p_k can be so chosen that furthermore $(\mathscr{X}, \mathbf{F}, m)$ has no standard extension.

4. The pseudometric KR^n and the fundamental lemma. Reverse filtrations of the form $(\mathscr{X}, \mathbf{F}, m)$ are again considered, where $\mathscr{X} = \{0,1\}^{\mathbb{N}}$, $\mathbf{F} = (\mathscr{F}_n)_{n=0}^\infty$ is the sequence of σ-fields corresponding to the coordinates and m satisfies the conditions of nondegeneracy and finite range dependence stipulated in Section 3.

We write $m(\cdot|x_{n+1}^\infty)$ for the conditional measure $m(\cdot|X_{n+1}^\infty = x_{n+1}^\infty)$. If Z is a measurable function from $\mathscr{X}$ to some measurable space $\mathscr{Z}$, then we write $m(Z|x_{n+1}^\infty)$ for the image of $m(\cdot|x_{n+1}^\infty)$ under Z, so $m(Z|x_{n+1}^\infty)$ is a measure on $\mathscr{Z}$. The quotient space $\mathscr{X}/\mathscr{F}_n$ may be identified with $\{0,1\}^{\{n+1, n+2, \dots\}}$, the projection being $x_1^\infty \mapsto x_{n+1}^\infty$. Thus, for a pseudometric ρ on $\mathscr{X}$, the Kantorovich–Rubinstein sequence is defined by $\rho_0 = \rho$ and, for $n \ge 1$,

$$\rho_n(x_{n+1}^\infty, y_{n+1}^\infty) = (\rho_{n-1})_{\mathrm{KR}}\big(m(X_n^\infty|x_{n+1}^\infty), m(X_n^\infty|y_{n+1}^\infty)\big).$$

Let $\mathscr{X}^n$ be $\{0,1\}^n$, $\mathscr{P}_n$ be the set of all probability measures on $\{0,1\}^n$ for which each point has positive measure and $\mathscr{R}_n$ the set of functions on $\mathscr{X}^n \times \mathscr{X}^n$ of the form $(x_1^n, y_1^n) \mapsto \sigma((x_1^n, 0), (y_1^n, 1))$, where σ is a pseudometric on $\mathscr{X}^{n+1}$. One easily sees that a pseudometric on $\mathscr{X}^n$ is in $\mathscr{R}_n$; however, a member of $\mathscr{R}_n$ need not be a pseudometric; for example, it may not be symmetric in its two arguments.

Note that each $m(X_1^n|x_{n+1}^\infty)$ is in $\mathscr{P}_n$. If ρ is a pseudometric on $\mathscr{X}$ and we define $\rho(\cdot,\cdot \,\|\, x_{n+1}^\infty, y_{n+1}^\infty)$ by $\rho(x_1^n, y_1^n \,\|\, x_{n+1}^\infty, y_{n+1}^\infty) = \rho(x_1^\infty, y_1^\infty)$, then also every $\rho(\cdot, \cdot \| x_{n+1}^{\infty}, y_{n+1}^{\infty})$ is in $\mathscr{R}_n$. In fact, if we map $\mathscr{X}^{n+1}$ into $\mathscr{X}$ with $(x_1^n, 0) \mapsto x_1^{\infty}$ and $(y_1^n, 1) \mapsto y_1^{\infty}$, then the pullback of ρ with respect to this map is a pseudometric σ on $\mathscr{X}^{n+1}$, and $\rho(\cdot, \cdot \| x_{n+1}^{\infty}, y_{n+1}^{\infty})$ is precisely $\sigma((\cdot, 0), (\cdot, 1))$. Furthermore, letting m and ρ vary, everything in $\mathscr{P}_n$ and $\mathscr{R}_n$ may be obtained in this way.

It will also be convenient to consider $\mathscr{X}^n$, $\mathscr{P}_n$ and $\mathscr{R}_n$ when $n = 0$: $\mathscr{X}^0$ is the set of all functions from the empty set to $\{0, 1\}$, that is, the one-point set whose only member is the empty set; $\mathscr{P}_0$ contains only one member, a point mass at the one point of $\mathscr{X}^0$; and $\mathscr{R}_0$ may be identified with $[0, \infty)$.

We will need an alternate definition of the Kantorovich–Rubinstein metric [see Kantorovich and Rubinstein (1958) or Dudley (1989), Section 11.8]; namely for any measurable pseudometric σ on a measurable space $(\mathscr{Y}, \mathscr{G})$ and probability measures μ, ν on $\mathscr{G}$,

$$\sigma_{\mathrm{KR}}(\mu, \nu) = \inf\left\{\int \sigma(x, y)\, d\lambda(x, y) \colon \lambda \in \mathscr{T}(\mu, \nu)\right\},$$

where $\mathscr{T}(\mu, \nu)$ is the set of joinings of μ with ν. We remind the reader that a *joining* of an ordered pair of probability measures is a measure on the product σ-field which has the original measures as marginals.

Lemma 4.1. *Given $n \geq 0$, there is a function Φ_n on $\mathscr{P}_n \times \mathscr{P}_n \times \mathscr{R}_n$ so that, for all reverse filtrations $(\mathscr{X}, \mathbf{F}, m)$ as above and all pseudometrics ρ on $\mathscr{X}$, we have*

$$\rho_n(x_{n+1}^{\infty}, y_{n+1}^{\infty}) = \Phi_n\Big(m(X_1^n | x_{n+1}^{\infty}), m(X_1^n | y_{n+1}^{\infty}), \rho(\cdot, \cdot \| x_{n+1}^{\infty}, y_{n+1}^{\infty})\Big).$$

(The Φ_n are, in fact, unique and continuous in the obvious sense, but this will not be needed, so the matter will not be pursued.)

Proof of Lemma 4.1. The argument goes by induction on n. For $n = 0$, it is obvious: Φ_0 is essentially the identity function on $[0, \infty)$.

Suppose for some $n \geq 1$ that Φ_{n-1} exists and works for all m and ρ. For μ and ν in $\mathscr{P}_n$ and φ in $\mathscr{R}_n$, set

$$\Phi_n(\mu, \nu, \varphi) = \inf\left\{\int \Phi_{n-1}\big(\mu(X_1^{n-1} | x_n), \nu(X_1^{n-1} | y_n), \varphi(\cdot, \cdot \| x_n, y_n)\big)\, d\lambda(x_n, y_n)\right\},$$

where λ varies over $\mathscr{T}(\mu \circ (X_n)^{-1}, \nu \circ (X_n)^{-1})$. It is easy to check that $\varphi(\cdot, \cdot \| x_n, y_n)$ is in $\mathscr{R}_{n-1}$ by a similar argument to the one which showed that $\rho(\cdot, \cdot \| x_{n+1}^{\infty}, y_{n+1}^{\infty})$ is in $\mathscr{R}_n$. Then, using the induction hypothesis, we have

$$\Phi_n\big(m(X_1^n | x_{n+1}^{\infty}), m(X_1^n | y_{n+1}^{\infty}), \rho(\cdot, \cdot \| x_{n+1}^{\infty}, y_{n+1}^{\infty})\big) = \inf\left\{\int \rho_{n-1}(x_n^{\infty}, y_n^{\infty})\, d\lambda(x_n, y_n)\right\},$$

where λ varies over $\mathscr{T}(m(X_n | x_{n+1}^{\infty}), m(X_n | y_{n+1}^{\infty}))$.

The map $x_n \mapsto (x_n, x_{n+1}^\infty) = x_n^\infty$ sends $m(X_n|x_{n+1}^\infty)$ to $m(X_n^\infty|x_{n+1}^\infty)$. Thus the map $(x_n, y_n) \mapsto (x_n^\infty, y_n^\infty)$ sends $\mathscr{T}(m(X_1^n|x_{n+1}^\infty)\circ X_n^{-1}, m(X_1^n|y_{n+1}^\infty)\circ X_n^{-1})$ to $\mathscr{T}(m(X_n^\infty|x_{n+1}^\infty), m(X_n^\infty|y_{n+1}^\infty))$. Then the measures λ in the integral may be interpreted as measures varying over $\mathscr{T}(m(X_n^\infty|x_{n+1}^\infty), m(X_n^\infty|y_{n+1}^\infty))$. Under this interpretation the infimum is seen to be exactly $(\rho_{n-1})_{\mathrm{KR}}(x_{n+1}^\infty, y_{n+1}^\infty)$, which equals $\rho_n(x_{n+1}^\infty, y_{n+1}^\infty)$. $\square$

The whole construction is shift invariant in the following sense: for $n \ge r$,

$$\rho_n(x_{n+1}^\infty, y_{n+1}^\infty) = \Phi_{n-r}\big(m(X_{r+1}^n|x_{n+1}^\infty), m(X_{r+1}^n|y_{n+1}^\infty), \rho_r(\cdot,\cdot\,\|x_{n+1}^\infty, y_{n+1}^\infty)\big),$$

where $\rho_r(x_{r+1}^n, y_{r+1}^n\|x_{n+1}^\infty, y_{n+1}^\infty) = \rho_r(x_{r+1}^\infty, y_{r+1}^\infty)$.

If ρ depends only on a finite number of coordinates, $\rho(x,y) = \bar\rho(x_1^n, y_1^n)$, then $\rho(\cdot,\cdot\,\|x_{n+1}^\infty, y_{n+1}^\infty) = \bar\rho$; hence

$$\rho_n(x_{n+1}^\infty, y_{n+1}^\infty) = \Phi_n\big(m(X_1^n|x_{n+1}^\infty), m(X_1^n|y_{n+1}^\infty), \bar\rho\big).$$

In particular, set $d^n(x,y) = 0$ when $x_1^n = y_1^n$ and 1 otherwise, and define

$$\mathrm{KR}^n(\mu,\nu) = \Phi_n(\mu,\nu,\bar d^n).$$

Then

$$(d^n)_n(x_{n+1}^\infty, y_{n+1}^\infty) = \mathrm{KR}^n\big(m(X_1^n|x_{n+1}^\infty), m(X_1^n|y_{n+1}^\infty)\big).$$

Now, for ε in $(0,1)$, we define $\mathscr{P}_{n,\varepsilon} \subset \mathscr{P}_n$ to be those probability measures μ on $\{0,1\}^n$ for which $\mu(X_i = 1|X_{i+1},\dots,X_n) = (1\pm\varepsilon)/2$ for $i = 1,\dots,n$.

FUNDAMENTAL LEMMA. *There exists an absolute constant C such that, for any $\varepsilon \in (0,1/2)$ and $n = 3,4,5,\dots,$ there exists a subset $\mathscr{M}$ of $\mathscr{P}_{n,\varepsilon}$ of cardinality 2^{2n} such that $\mathrm{KR}^n(\mu,\nu) \ge 1 - C/(n\varepsilon)$ for any $\mu \ne \nu$ in $\mathscr{M}$.*

The remainder of this section will be devoted to proving Theorem 2.6 on the basis of the fundamental lemma.

First, suppose that, for a certain $n \in \mathbb{N}$ and some $\gamma > 0$, the pseudometric ρ on $\mathscr{X} = \{0,1\}^{\mathbb{N}}$ satisfies $\rho(x,y) \ge \gamma$ whenever $x_1^n \ne y_1^n$. Then $\rho(x,y) \ge \gamma d^n(x,y)$, which implies $\rho_n(x_{n+1}^\infty, y_{n+1}^\infty) \ge \gamma (d^n)_n(x_{n+1}^\infty, y_{n+1}^\infty)$; that is,

$$\rho_n(x_{n+1}^\infty, y_{n+1}^\infty) \ge \gamma\,\mathrm{KR}^n\big(m(X_1^n|x_{n+1}^\infty), m(X_1^n|y_{n+1}^\infty)\big).$$

Similarly, if $r \le n$ and $\rho_r(x_{r+1}^\infty, y_{r+1}^\infty) \ge \gamma$ whenever $x_{r+1}^n \ne y_{r+1}^n$, then

$$\rho_n(x_{n+1}^\infty, y_{n+1}^\infty) \ge \gamma\,\mathrm{KR}^{n-r}\big(m(X_{r+1}^n|x_{n+1}^\infty), m(X_{r+1}^n|y_{n+1}^\infty)\big).$$

Now we return to the block-Markov case of Section 3. When n is of the form 2^k, $k = 0,1,\dots,$ we see that $m(X_1^{2n-1}|x_{2n}^\infty)$ depends only on $x_{2n}^{4n-1} = x^{(k+1)}$. Restricting this measure to the coordinates $n,\dots,2n-1$, we get $m(X_n^{2n-1} = x^{(k)}|x_{2n}^\infty) = p_k(x^{(k)}|x^{(k+1)})$.

If ρ is a pseudometric on $\mathscr{X}$ depending only on finitely many coordinates, then we have $\rho_{n-1}(x_n^\infty, y_n^\infty) = \Phi_{n-1}(m(X_1^{n-1}|x_n^\infty), m(X_1^{n-1}|y_n^\infty), \bar\rho)$ for all sufficiently large n. For n of the form 2^k, this depends only on $(x^{(k)}, y^{(k)})$, as seen in the previous paragraph. If $\rho_{n-1}(x_n^\infty, y_n^\infty) \ge \gamma$ whenever $x_n^{2n-1} \ne y_n^{2n-1}$, then $\rho_{2n-1}(x_{2n}^\infty, y_{2n}^\infty) \ge \gamma\,\mathrm{KR}^n(m(X_n^{2n-1}|x_{2n}^\infty), m(X_n^{2n-1}|y_{2n}^\infty))$. For $n = 2^k$ with k

large enough, this means the following. If $\rho_{n-1}(x^{(k)}, y^{(k)}) \geq \gamma$ whenever $x^{(k)} \neq y^{(k)}$, then

$$\rho_{2n-1}\big(x^{(k+1)}, y^{(k+1)}\big) \geq \gamma \,\mathrm{KR}^n\big(p_k(\cdot|x^{(k+1)}), p_k(\cdot|y^{(k+1)})\big).$$

For $k = 0, 1, \dots,$ choose $\varepsilon_k \in (0,1)$ satisfying the two conditions given at the end of Section 3; for example, $\varepsilon_k = \theta^k$ with $\theta \in (1/2, 1/\sqrt{2})$. Then choose for each $k = 0, 1, \dots$ a subset $\mathscr{M}_k$ of $\mathscr{P}_{n, \varepsilon_k}$ containing 2^{2n} members so that:

(ii) for any $\mu \neq \nu \in \mathscr{M}_k$,

$$\mathrm{KR}^n(\mu, \nu) \geq 1 - \frac{C}{n\varepsilon_k}.$$

This may be done by the fundamental lemma. [Note: the condition is called (ii) because of its relation to (ii*); the reader should not be disturbed by the absence of any condition (i).] Finally, make transition probabilities p_k by choosing $z \mapsto p_k(\cdot|z)$ to be any bijection from $\mathscr{X}^{(k+1)}$ onto $\mathscr{M}_k$. Then Section 3 tells us how to construct a probability measure m on $\mathscr{X}$ equivalent to the Bernoulli product measure, giving a reverse filtration $(\mathscr{X}, \mathbf{F}, m)$.

For any ρ depending on only finitely many coordinates, we have, from the above remarks,

$$\begin{aligned}\min\big\{\rho_{2n-1}\big(x^{(k+1)}, y^{(k+1)}\big)&: x^{(k+1)} \neq y^{(k+1)}\big\}\\ &\geq \left(1 - \frac{C}{n\varepsilon_k}\right)\cdot \min\big\{\rho_{n-1}\big(x^{(k)}, y^{(k)}\big): x^{(k)} \neq y^{(k)}\big\}\end{aligned}$$

for all k large enough. Take k_0 so that

$$1 - \frac{C}{2^k\varepsilon_k} > 0 \quad \text{for } k \geq k_0.$$

Then

$$\prod_{k=k_0}^{\infty}\left(1 - \frac{C}{2^k\varepsilon_k}\right) > 0.$$

Choose ρ to be d^{2n_0-1}, $n_0 = 2^{k_0}$. Then

$$\inf\big\{\rho_{2^k}\big(x^{(k)}, y^{(k)}\big): x^{(k)} \neq y^{(k)}\big\} \geq \prod_{l=k_0}^{k-1}\left(1 - \frac{C}{2^l\varepsilon_l}\right)$$

for $k \geq k_0$. Now, $m \otimes m(\{(x, y): x^{(k)} \neq y^{(k)}\}) \to 1$ as $k \to \infty$. It follows that the numbers $\alpha_n(\rho) = \int\int \rho_n(x, y)\, dm(x)\, dm(y)$, with $n = 2^{k_0} - 1, 2^{k_0+1} - 1, \dots,$ are bounded away from 0. So $\alpha_\infty(\rho) > 0$, and by Theorem 2.7(b) the reverse filtration $(\mathscr{X}, \mathbf{F}, m)$ admits no standard extension. □

5. Proof of the fundamental lemma. The proof will be given after some auxiliary lemmas. Return to the "discrete" metric d^n giving

$$\mathrm{KR}^n(\mu, \nu) = \Phi_n(\mu, \nu, \bar{d}^n).$$

In the next three lemmas we describe how KR^n may be estimated from below in terms of KR^{n-1} by working with a certain KR metric in a four-point space.

LEMMA 5.1. *If* $0 \le i \le n$, $(d^n)_i(x_{i+1}^\infty, y_{i+1}^\infty)$ *is equal to*

$$\mathrm{KR}^i\big(m\big(X_1^i | x_{i+1}^\infty\big), m\big(X_1^i | y_{i+1}^\infty\big)\big)$$

when $x_{i+1}^n = y_{i+1}^n$, *and is identically* 1 *otherwise.*

PROOF. If $x_{i+1}^n = y_{i+1}^n$ then, by Lemma 4.1,

$$(d^n)_i(x_{i+1}^\infty, y_{i+1}^\infty) = \Phi_i\big(m\big(X_1^i | x_{i+1}^\infty\big), m\big(X_1^i | y_{i+1}^\infty\big), d^n(\cdot, \cdot \,\|\, x_{i+1}^\infty, y_{i+1}^\infty)\big).$$

Since $d^n(\cdot, \cdot \,\|\, x_{i+1}^\infty, y_{i+1}^\infty) = \bar{d}^i$ when $x_{i+1}^n = y_{i+1}^n$, this equals

$$\Phi_i\big(m\big(X_1^i | x_{i+1}^\infty\big), m\big(X_1^i | y_{i+1}^\infty\big), \bar{d}^i\big) = \mathrm{KR}^i\big(m\big(X_1^i | x_{i+1}^\infty\big), m\big(X_1^i | y_{i+1}^\infty\big)\big).$$

The case $x_{i+1}^n \ne y_{i+1}^n$ is proved by a finite induction on i. For $i = 0$ the claim is true by definition. So assume that $n \ge i \ge 1$ and that the assertion is true for $i - 1$.

Because we are in the case $x_{1+1}^n \ne y_{i+1}^\infty$, we have $(d^n)_i(x_{i+1}^\infty, y_{i+1}^\infty) = ((d^n)_{i-1})_{\mathrm{KR}}(m(X_i^\infty | x_{i+1}^\infty), m(X_i^\infty | y_{i+1}^\infty))$. Any joining λ of the two measures in the argument is supported on points (x, y) with tail $(x_{i+1}^\infty, y_{i+1}^\infty)$. Since $x_{i+1}^n \ne y_{i+1}^n$, also $x_i^n \ne y_i^n$, by the induction hypothesis the values of $(d^n)_{i-1}$ in the support of λ are all 1, and the integral obtained using any such λ is therefore also 1. □

The values $\rho_n(x_{n+1}^\infty, y_{n+1}^\infty)$ may be calculated within the four-point structure formed by the two-point sets $\{0, 1\} \times \{x_{n+1}^\infty\}$ and $\{0, 1\} \times \{y_{n+1}^\infty\}$ equipped with measures $\bar{\mu}, \bar{\nu}$ and a pseudometric $\bar{\rho}$, the first measure $\bar{\mu} = m(X_n^\infty | x_{n+1}^\infty) = m(X_n | x_{n+1}^\infty) \times \delta(x_{n+1}^\infty)$ supported on the first two-point set, the second one $\bar{\nu} = m(X_n^\infty | y_{n+1}^\infty) = m(X_n | y_{n+1}^\infty) \times \delta(y_{n+1}^\infty)$ on the second set. $\bar{\rho}$ is defined on the product of these two-point sets to be the appropriate restriction of ρ_{n-1}; that is, $\bar{\rho}((x_n, x_{n+1}^\infty), (y_n, y_{n+1}^\infty)) = \rho_{n-1}(x_n^\infty, y_n^\infty)$. Actually, only the restriction of $\bar{\rho}$ to certain pairs of points is needed, but this need not concern us. Notice that $\rho_n(x_{n+1}^\infty, y_{n+1}^\infty)$ is the infimum of $\int \bar{\rho}\, d\lambda$ over all joinings λ of $\bar{\mu}$ with $\bar{\nu}$, which is exactly $\bar{\rho}_{\mathrm{KR}}(\bar{\mu}, \bar{\nu})$.

Let us simplify the notation:

$$a = (0, x_{n+1}^\infty), \qquad c = (0, y_{n+1}^\infty),$$

$$b = (1, x_{n+1}^\infty), \qquad d = (1, y_{n+1}^\infty);$$

$$\bar{\mu}\{a\} = m(X_n = 0 | x_{n+1}^\infty), \qquad \bar{\nu}\{c\} = m(X_n = 0 | y_{n+1}^\infty),$$

$$\bar{\mu}\{b\} = m(X_n = 1 | x_{n+1}^\infty), \qquad \bar{\nu}\{d\} = m(X_n = 1 | y_{n+1}^\infty),$$

$$\bar{\rho}(a, c) = \rho_{n-1}(a, c) = \rho_{n-1}((0, x_{n+1}^\infty), (0, y_{n+1}^\infty)),$$

and similarly for $\bar{\rho}(a, d)$, $\bar{\rho}(b, c)$ and $\bar{\rho}(b, d)$.

Suppose, in addition, that $\rho = d^N$ for some $N \geq n$. Then Lemma 5.1 gives

$$\begin{aligned}\rho_n(x_{n+1}^\infty, y_{n+1}^\infty) &= \mathrm{KR}^n\big(m(X_1^n|x_{n+1}^\infty), m(X_1^n|y_{n+1}^\infty)\big),\\ \bar\rho(a,c) &= \mathrm{KR}^{n-1}\big(m(X_1^{n-1}|0, x_{n+1}^\infty), m(X_1^{n-1}|0, y_{n+1}^\infty)\big),\\ \bar\rho(b,d) &= \mathrm{KR}^{n-1}\big(m(X_1^{n-1}|1, x_{n+1}^\infty), m(X_1^{n-1}|1, y_{n+1}^\infty)\big),\\ \bar\rho(a,d) &= 1, \qquad \bar\rho(b,c) = 1.\end{aligned}$$

A recurrence relation for KR^n, implicit in the above considerations, will become explicit in the following lemma, after carrying out a "four-point" computation.

LEMMA 5.2. *Let each of two probability measures $\bar\mu, \bar\nu$ be concentrated on a two-point set, $\bar\mu$ on $\{a, b\}$ and $\bar\nu$ on $\{c, d\}$, in a space with metric $\bar\rho$. Suppose that*

$$\begin{aligned}\bar\rho(a,c) &\leq 1, \qquad \bar\rho(b,d) \leq 1,\\ \bar\rho(a,d) &= 1, \qquad \bar\rho(b,c) = 1.\end{aligned}$$

Then

$$\begin{aligned}1 - \bar\rho_{\mathrm{KR}}(\bar\mu,\bar\nu) \leq{}& (1-\bar\rho(a,c))\min(\bar\mu\{a\},\bar\nu\{c\})\\ &+(1-\bar\rho(b,d))\min(\bar\mu\{b\},\bar\nu\{d\}).\end{aligned}$$

(In fact, equality holds, but we do not need it here.)

PROOF OF LEMMA 5.2. Without loss of generality, we may suppose that $\bar\mu\{a\} \geq \bar\nu\{c\}$; then $\bar\mu\{b\} \leq \bar\nu\{d\}$. Take a function $l\colon \{a, b, c, d\} \to \mathbb{R}$ such that

$$\begin{aligned}l(c) - l(a) &= \bar\rho(a,c),\\ l(d) - l(b) &= \bar\rho(b,d),\\ l(d) - l(a) &= \bar\rho(a,d) = 1.\end{aligned}$$

This l is a Lipschitz function. Indeed, $|l(a) - l(b)| = |\bar\rho(a, d) - \bar\rho(b, d)| \leq \bar\rho(a, b)$; the same for c, d; and $|l(b) - l(c)| = |\bar\rho(a, c) + \bar\rho(b, d) - \bar\rho(a, d)| \leq \bar\rho(b, c)$. Hence

$$\begin{aligned}1 &- \bar\rho_{\mathrm{KR}}(\bar\mu,\bar\nu)\\ &\leq 1 - \left(\int l\,d\bar\nu - \int l\,d\bar\mu\right)\\ &= 1 + l(a)(1-\bar\mu\{b\}) + l(b)\bar\mu\{b\} - l(c)\bar\nu\{c\} - l(d)(1-\bar\nu\{c\})\\ &= 1 + (l(a) - l(d)) + (l(b) - l(a))\bar\mu\{b\} + (l(d) - l(c))\bar\nu\{c\}.\end{aligned}$$

Taking into account that $l(d) - l(a) = 1$, we obtain

$$\begin{aligned}1 - \bar\rho_{\mathrm{KR}}(\bar\mu,\bar\nu) &\leq (1 + l(b) - l(d))\bar\mu\{b\} + (1 + l(a) - l(c))\bar\nu\{c\}\\ &= (1-\bar\rho(b,d))\bar\mu\{b\} + (1-\bar\rho(a,c))\bar\nu\{c\},\end{aligned}$$

which completes the proof. □

For $\mu \in \mathscr{P}_{n,\varepsilon}$ write μ_x for the conditional measure $\mu(X_1^{n-1}|X_n = x)$, and define $\tau(\mu)$ by the equation $\mu(X_n = 1) = (1 + \tau(\mu)\varepsilon)/2$. Note that $\tau(\mu)$ plays a similar role to that of the τ_i in the beginning of Section 3.

PROPOSITION 5.3. *For every μ and ν in $\mathscr{P}_{n,\varepsilon}$*

$$1 - \mathrm{KR}^n(\mu, \nu) \le \big(1 - \mathrm{KR}^{n-1}(\mu_0, \nu_0)\big) \times \big(1 - \varepsilon \max\{\tau(\mu), \tau(\nu)\}\big)/2 + \big(1 - \mathrm{KR}^{n-1}(\mu_1, \nu_1)\big) \times \big(1 + \varepsilon \min\{\tau(\mu), \tau(\nu)\}\big)/2.$$

LEMMA 5.4. *Define*

$$F_n(\lambda, \varepsilon) = \frac{1}{|\mathscr{P}_{n,\varepsilon}|^2} \sum_{\mu, \nu \in \mathscr{P}_{n,\varepsilon}} \exp(\lambda(1 - \mathrm{KR}^n(\mu, \nu)))$$

for $\lambda \ge 0$. Then

$$F_n(\lambda, \varepsilon) \le F_{n-1}\left(\frac{1-\varepsilon}{2}\lambda, \varepsilon\right) \times \frac{1}{2}\left(F_{n-1}\left(\frac{1-\varepsilon}{2}\lambda, \varepsilon\right) + F_{n-1}\left(\frac{1+\varepsilon}{2}\lambda, \varepsilon\right)\right).$$

(In fact, equality holds, but we do not need it here.)

PROOF OF LEMMA 5.4. If M and N are independent random variables, each taking on all values in $\mathscr{P}_{n,\varepsilon}$ with equal probability, then the expression to be estimated is precisely $\mathbb{E}(\exp(\lambda(1 - \mathrm{KR}^n(M, N))))$. Recalling $\tau(\mu)$ of Corollary 5.3, we then get independent random variables $\tau(M)$ and $\tau(N)$. Because of the uniformity of the distributions of M and N, the random variables $\tau(M)$ and $\tau(N)$ take on ± 1 with equal probability. Also, recalling μ_0 and μ_1 of Corollary 5.3, we get independent random variables M_0, M_1, N_0, N_1 taking on all values in $\mathscr{P}_{n-1,\varepsilon}$ with equal probability; these values are just the values of M and N conditioned by the event in $\{0,1\}^n$ described by $X_n = 0$ or $X_n = 1$, as indicated by the subscript. Then Corollary 5.3 gives

$$1 - \mathrm{KR}^n(M, N) \le \big(1 - \mathrm{KR}^{n-1}(M_0, N_0)\big) \times \big(1 - \varepsilon \max\{\tau(M), \tau(N)\}\big)/2 + \big(1 - \mathrm{KR}^{n-1}(M_1, N_1)\big) \times \big(1 + \varepsilon \min\{\tau(M), \tau(N)\}\big)/2.$$

Average, using independence. First condition on $\tau(M) = \tau(N) = 1$, getting:

$$\begin{aligned}
&\mathbb{E}\big(\exp(\lambda(1 - \mathrm{KR}^n(M, N)))\,|\,\tau(M) = \tau(N) = +1\big) \\
&\quad \le \mathbb{E}\exp\left(\frac{\lambda(1-\varepsilon)}{2}\big(1 - \mathrm{KR}^{n-1}(M_0, N_0)\big) + \frac{\lambda(1+\varepsilon)}{2}\big(1 - \mathrm{KR}^{n-1}(M_1, N_1)\big)\right) \\
&\quad = F_{n-1}\left(\frac{1-\varepsilon}{2}\lambda, \varepsilon\right) \times F_{n-1}\left(\frac{1+\varepsilon}{2}\lambda, \varepsilon\right).
\end{aligned}$$

Similar reasoning holds for $\tau(M) = \tau(N) = -1$. The two other cases give $F^2_{n-1}([(1-\varepsilon)/2]\lambda, \varepsilon)$, completing the argument. □

LEMMA 5.5. *The function* F_n *introduced in Lemma* 5.4 *satisfies the inequality*

$$F_n(\lambda, \varepsilon) \le \exp\left(\left(1 - \frac{\varepsilon}{3}\right)^n \lambda\right)$$

for $0 \le \lambda \le \varepsilon^{-1}(1-\varepsilon/3)^{-n+1}$ *and* $0 < \varepsilon \le 1/2$.

PROOF. It is natural to put $F_0(\lambda, \varepsilon) = e^{\lambda}$; then Lemma 5.4 holds also for $n = 1$, and we may prove our inequality by induction, the case $n = 0$ being trivial. Due to Lemma 5.4, it suffices to check that

$$\begin{aligned}\exp&\left(\left(1-\frac{\varepsilon}{3}\right)^{n-1}\frac{1-\varepsilon}{2}\lambda\right)\\ &\times\frac{1}{2}\left(\exp\left(\left(1-\frac{\varepsilon}{3}\right)^{n-1}\frac{1-\varepsilon}{2}\lambda\right)+\exp\left(\left(1-\frac{\varepsilon}{3}\right)^{n-1}\frac{1+\varepsilon}{2}\lambda\right)\right)\\ &\le \exp\left(\left(1-\frac{\varepsilon}{3}\right)^{n}\lambda\right)\end{aligned}$$

for $n = 1, 2, \dots,$ $0 \le \lambda \le \varepsilon^{-1}(1-\varepsilon/3)^{-n+1}$ and $0 < \varepsilon \le 1/2$; the requirement

$$\frac{1 \pm \varepsilon}{2}\lambda \le \varepsilon^{-1}\left(1 - \frac{\varepsilon}{3}\right)^{-(n-1)+1}$$

is fulfilled, since $(1+\varepsilon)/2 \le 1 - \varepsilon/3$. Simplify the inequality:

$$\frac{1}{2}\exp\left(-\left(1-\frac{\varepsilon}{3}\right)^{n-1}\times\frac{2\varepsilon}{3}\lambda\right)+\frac{1}{2}\exp\left(\left(1-\frac{\varepsilon}{3}\right)^{n-1}\times\frac{\varepsilon}{3}\lambda\right)\le 1.$$

The left-hand side is convex in λ; hence it suffices to check the inequality for $\lambda = 0$ and $\lambda = \varepsilon^{-1}(1-\varepsilon/3)^{-n+1}$. The former case is trivial. The latter reduces to checking whether $\frac{1}{2}\exp(-\frac{2}{3}) + \frac{1}{2}\exp(\frac{1}{3}) \le 1$. This is indeed the case, which completes the proof. □

PROOF OF FUNDAMENTAL LEMMA. Let $L = 2^{2n}$ and $M^{(1)}, \dots, M^{(L)}$ be independent uniformly distributed random variables taking on all values in $\mathscr{P}_{n,\varepsilon}$ with equal probability. Denote by q the probability of violating the inequality $\mathrm{KR}^n(M^{(i)}, M^{(j)}) \ge 1 - C/(n\varepsilon)$ at least for one pair $i \ne j$. Lemma 5.5 gives

$$q \le \frac{L(L-1)}{2} \times \frac{F_n(\lambda, \varepsilon)}{\exp(\lambda \times C/n\varepsilon)} \le 2^{4n}\exp\left(\left(1-\frac{\varepsilon}{3}\right)^n\lambda - \lambda\cdot\frac{C}{n\varepsilon}\right).$$

Take the maximal λ, that is, $\lambda = \varepsilon^{-1}(1-\varepsilon/3)^{-n+1}$. Then

$$\begin{aligned}
\log q &\le 4n\log 2 + \frac{1}{\varepsilon}\left(1-\frac{\varepsilon}{3}\right) - \frac{C}{n\varepsilon^2}\left(1-\frac{\varepsilon}{3}\right)^{-n+1}\\
&= -\frac{C}{\varepsilon}\left(\frac{1}{n\varepsilon}\left(1-\frac{\varepsilon}{3}\right)^{-n+1} - \frac{1}{C}\left(1-\frac{\varepsilon}{3}\right) - \frac{4\log 2}{C}n\varepsilon\right)\\
&\le -\frac{C}{\varepsilon}\left(\frac{1}{n\varepsilon}\left(1-\frac{\varepsilon}{3}\right)\exp\left(\frac{n\varepsilon}{3}\right) - \frac{1}{C}\left(1-\frac{\varepsilon}{3}\right) - \frac{4\log 2}{C}n\varepsilon\right)\\
&= -\frac{C}{\varepsilon}\cdot\frac{1}{n\varepsilon}\exp\left(\frac{n\varepsilon}{3}\right)\cdot\left(1-\frac{\varepsilon}{3} - \frac{1}{C}n\varepsilon\exp\left(-\frac{n\varepsilon}{3}\right)\left(1-\frac{\varepsilon}{3} + 4\log 2\cdot n\varepsilon\right)\right).
\end{aligned}$$

If we choose the absolute constant C large enough, then the last term is always greater than or equal to $1/2 > \varepsilon$, so:

$$\log q \le -C\cdot\frac{1}{n\varepsilon}\exp\left(\frac{n\varepsilon}{3}\right).$$

When C is large enough we have $q < 1$, so that some value $(\mu^{(1)},\dots,\mu^{(L)})$ of $(M^{(1)},\dots,M^{(L)})$ will provide the needed set $\mathscr{M}$ for the nontrivial case when $C/(n\varepsilon) < 1$ [the inequality $\mathrm{KR}^n(\mu_i,\mu_j) \ge 1 - C/(n\varepsilon)$ also ensures that all μ_i are pairwise different]. For the trivial case when $C/(n\varepsilon) \ge 1$, it suffices to note that $|\mathscr{P}_{n,\varepsilon}| \ge 2^{2n} = L$ for $n \ge 3$. So the fundamental lemma is proved. □

REMARK. For each $k = 0, 1, \dots$ let $n = 2^k$ and let a subset $\mathscr{M}_k$ of cardinality 2^{2n} be chosen *randomly* from $\mathscr{P}_{n,\varepsilon_k}$ and independently for different k. As seen in the proof of the fundamental lemma, for each k the number $q = q_k$, the probability that $\mathscr{M}_k$ violates the KR^n condition of that lemma, satisfies the inequality

$$\log q_k \le -C\cdot\frac{1}{2^k\varepsilon_k}\exp\left(\frac{2^k\varepsilon_k}{3}\right).$$

Since we have assumed $\sum_k (2^k\varepsilon_k)^{-1} < \infty$, it follows that $\sum_k q_k < \infty$. Thus, with probability 1, $\mathscr{M}_k$ satisfies the KR^n condition of the fundamental lemma for all sufficiently large k. Then, if the transition probabilities p_k are chosen to be arbitrary one-to-one maps $z \mapsto p_k(\cdot|z)$, $\{0,1\}^n \to \mathscr{M}_k$ and m is the corresponding measure on $\mathscr{X}$, the reverse filtration $(\mathscr{X}, \mathbf{F}, m)$ has no standard extension.

6. Discussion of a statement of Skorokhod. The purpose of this section is to give a plausible precise interpretation of Theorem 7 of Skorokhod (1986), alluded to in the introduction, and to show that this interpretation is contradicted by our results.

Skorokhod's statement proposes an exhaustive constructive description of filtrations [i.e., increasing families $(\mathscr{F}_t)_{t \ge 0}$ of σ-fields $\mathscr{F}_t \subset \mathscr{F}$] on probability spaces $(\Omega, \mathscr{F}, P)$, provided the following condition holds:

(S_1) For any $(\mathscr{F}_t)_{t \ge 0}$-adapted square integrable martingale X on $(\Omega, \mathscr{F}, P)$, the process $(\langle X \rangle_t)_{t \ge 0}$ [the dual previsible projection of the quadratic variation $([X]_t)_{t \ge 0}$; see Rogers and Williams (1987), page 377, or Protter (1990), page 98] is almost surely absolutely continuous in t.

The constructive description is given in terms of Wiener and Poisson processes. Since we are not interested in the Poisson component, we restrict ourselves to filtrations satisfying one more condition:

(S_2) Every $(\mathscr{F}_t)_{t \ge 0}$-adapted square integrable martingale X on $(\Omega, \mathscr{F}, P)$ has a modification which is almost surely continuous in t. (So $\langle X \rangle_t = [X]_t$.)

As usual, the constructive description in terms of auxiliary Brownian motions requires an extension of the probability space; as seen in Section 2, this has an analog for *reverse* filtrations, which was used there. The notion of filtrations may be found in Getoor and Sharpe (1972) under the name "lifting," or in Ikeda and Watanabe [(1989), Chapter 2, Definition 7.1]. Skorokhod does not explicitly mention extensions, but we believe that their use must have been intended, for otherwise his Theorem 7 would be false for trivial reasons. Extensions for filtrations are defined as follows.

DEFINITION 6.1. Let $(\Omega, \mathscr{F}, P)$ be a probability space equipped with a filtration $(\mathscr{F}_t)_{t \ge 0}$. An extension of $(\Omega, \mathscr{F}, P, (\mathscr{F}_t)_{t \ge 0})$ consists of another probability space $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{P})$ equipped with a filtration $(\tilde{\mathscr{F}}_t)_{t \ge 0}$ and a measure-preserving map $\pi: \tilde{\Omega} \to \Omega$ so that $\pi^{-1}\mathscr{F}_t \subset \tilde{\mathscr{F}}_t$ for each t, and the equality

$$\tilde{E}\big(X \circ \pi | \tilde{\mathscr{F}}_t\big) = E(X|\mathscr{F}_t) \circ \pi$$

holds almost everywhere for any bounded measurable function $X: \Omega \to \mathbb{R}$.

Given $(\Omega, \mathscr{F}, P)$ and $(\mathscr{F}_t)_{t \ge 0}$ satisfying (S_1) and (S_2), we define a *Skorokhod representation* for $(\mathscr{F}_t)_{t \ge 0}$ as the following objects (a)–(c) satisfying condition (d):

(a) an extension $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{P}, (\tilde{\mathscr{F}}_t)_{t \ge 0}), \pi$ of $(\Omega, \mathscr{F}, P, (\mathscr{F}_t)_{t \ge 0})$;

(b) a sequence $(w_i)_{i=1,2,\dots}$ of independent $(\tilde{\mathscr{F}}_t)$-Brownian motion processes on $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{P})$, [The independence of w_i was not explicitly assumed in Skorokhod (1986), but here again we believe that it must have been intended.] where as usual, $w_i(t) - w_i(s)$ is supposed to be independent of $\tilde{\mathscr{F}}_s$ when $s < t$;

(c) a random process $(r_t)_{t \ge 0}$ on $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{P})$, predictable with respect to $(\tilde{\mathscr{F}}_t)_{t \ge 0}$ with values in $\{0, 1, 2, \dots, +\infty\}$. [An even stronger predictability condition, with respect to $(w_i)_{i=1,2,\dots}$ or even with respect to $(\pi^{-1}\mathscr{F}_t)_{t \ge 0}$, seems to be imposed in Skorokhod (1986).]

(d) For any $t \ge 0$, the σ-field $\pi^{-1}\mathscr{F}_t$ coincides mod 0 with the σ-field generated by $\{u_i(s): 0 \le s \le t,\ i = 1, 2, \dots\}$, where

$$u_i(t) = \int_0^t I_i(s)\, dw_i(s),$$

$$I_i(s) = \begin{cases} 1, & \text{when } i \le r(s), \\ 0, & \text{otherwise.} \end{cases}$$

The relevant consequence of our interpretation of Theorem 7 of Skorokhod (1986) may now be formulated as follows:

> A Skorokhod representation exists for any filtration which satisfies conditions (S_1) and (S_2).

But this is simply not the case. In the remaining discussion we will show that it is contradicted by our negative solution to Problem 1 of the Introduction. Specifically, let Q be the measure equivalent to P constructed in the discussion after Theorem 2.6. We will show that $(\Omega, \mathscr{F}, Q, (\mathscr{F}_t)_{t \ge 0})$ satisfies (S_1) and (S_2) but admits no Skorokhod representation.

Any local P-martingale [i.e., local martingale on the filtered probability space $(\Omega, \mathscr{F}, P, (\mathscr{F}_t)_{t \ge 0})$] is continuous, by Revuz and Yor (1991), Theorem 3.4, page 187. There exists a (continuous) positive P-martingale $(D_t)_{t \ge 0}$ such that D_∞ exists and is equal to the Radon-Nikodym derivative dQ/dP. In fact

$$D_t = \exp\left(\int_0^t \Phi(s)\, dB(s) - \frac{1}{2} \int_0^t \Phi^2(s)\, ds \right),$$

but this will play no role here. The two measures P and Q [or rather filtered probability spaces $(\Omega, \mathscr{F}, P, (\mathscr{F}_t)_{t \ge 0})$ and $(\Omega, \mathscr{F}, Q, (\mathscr{F}_t)_{t \ge 0})$] form a Girsanov pair, as defined in Revuz and Yor (1991), Definition 1.8, page 305. Hence the class of continuous semimartingales is the same for P and Q [Revuz and Yor (1991), page 305], and the notions of quadratic variation $\langle X \rangle_t$ and quadratic covariation $\langle X, Y \rangle_t$ for such semimartingales are the same for P and Q [Revuz and Yor (1991), page 302], where a continuous semimartingale is defined as a sum of a continuous local martingale and a continuous adapted process of (locally) finite variation [Revuz and Yor (1991), Definition 1.17, page 121].

An adapted process $(X_t)_{t \ge 0}$ is a local Q-martingale if and only if the process $(X_t D_t)_{t \ge 0}$ is a local P-martingale [Protter (1990), the lemma on page 109]. Hence any local Q-martingale is continuous, too. Thus condition (S_2) holds for Q. Condition (S_1) holds for Q, since it holds for P by the representation theorem [Revuz and Yor (1991), Theorem 3.4, page 187].

The same representation theorem ensures that the following statement holds for P (and hence for Q): for any two continuous semimartingales

X_1, X_2 the determinant

$$\det\begin{vmatrix} \frac{d}{dt}\langle X_1\rangle_t & \frac{d}{dt}\langle X_1, X_2\rangle_t \\ \frac{d}{dt}\langle X_2, X_1\rangle_t & \frac{d}{dt}\langle X_2\rangle_t \end{vmatrix}$$

vanishes (almost everywhere in t, almost surely). Indeed, for $i = 1, 2$ we have $X_i = \int \Psi_i\, dB +$ (a finite variation process), which implies $(d/dt)\langle X_i, X_j\rangle_t = \Psi_i(t)\Psi_j(t)$.

Suppose $(\Omega, \mathscr{F}, Q, (\mathscr{F}_t)_{t\ge 0})$ admitted a Skorokhod representation. Extension of the probability space as in Definition 6.1 preserves the notions of martingale, continuous semimartingale and quadratic (co-)variation [Ikeda and Watanabe (1989), page 89]. That is, if $(X_t)_{t\ge 0}$ is a continuous semimartingale on $(\Omega, \mathscr{F}, Q, (\mathscr{F}_t)_{t\ge 0})$, then $(X_t \circ \pi)_{t\ge 0}$ is a continuous semimartingale on $(\tilde\Omega, \tilde{\mathscr{F}}, \tilde Q, (\tilde{\mathscr{F}}_t)_{t\ge 0})$, and $\langle X\rangle_t \circ \pi = \langle X \circ \pi\rangle_t$. On the other hand, a process $(\tilde X_t)_{t\ge 0}$ on $(\tilde\Omega, \tilde{\mathscr{F}}, \tilde Q, (\tilde{\mathscr{F}}_t)_{t\ge 0})$ is of the form $X \circ \pi$ if and only if it is adapted to $(\pi^{-1}\mathscr{F}_t)_{t\ge 0}$. Hence the above statement about the determinant holds for all $(\pi^{-1}\mathscr{F}_t)_{t\ge 0}$-adapted continuous semimartingales on $(\tilde\Omega, \tilde{\mathscr{F}}, \tilde Q, (\tilde{\mathscr{F}}_t)_{t\ge 0})$. Therefore the Skorokhod rank $r_t \le 1$ (almost everywhere in t, almost surely). But r_t could not vanish, since there exists a continuous $(\pi^{-1}\mathscr{F}_t)_{t\ge 0}$-adapted semimartingale X on $(\tilde\Omega, \tilde{\mathscr{F}}, \tilde Q, (\tilde{\mathscr{F}}_t)_{t\ge 0})$ such that $(d/dt)\langle X\rangle_t$ does not vanish (this follows from the corresponding fact for Q, which follows in turn from the same for P, the latter being evident).

So $r_t = 1$. This means that $(\pi^{-1}\mathscr{F}_t)_{t\ge 0}$ is generated by a single Brownian motion $\tilde w$ on $(\tilde\Omega, \tilde{\mathscr{F}}, \tilde Q, (\tilde{\mathscr{F}}_t)_{t\ge 0})$. Being adapted to $(\pi^{-1}\mathscr{F}_t)_{t\ge 0}$, this $\tilde w$ is of the form $\tilde w = w \circ \pi$ for some adapted process w on $(\Omega, \mathscr{F}, Q, (\mathscr{F}_t)_{t\ge 0})$. Then w generates σ-fields $\mathscr{G}_t(w)$ such that $\pi^{-1}\mathscr{G}_t(w)$ coincides mod 0 with $\pi^{-1}\mathscr{F}_t$; hence $\mathscr{G}_t(w)$ coincides mod 0 with $\mathscr{F}_t$. Finally, the finite-dimensional distributions of w coincide with those of $\tilde w$; hence w is a Brownian motion on $(\Omega, \mathscr{F}, Q, (\mathscr{F}_t)_{t\ge 0})$ generating $(\mathscr{F}_t)_{t\ge 0}$, contradicting the choice of Q.

REFERENCES

DUDLEY, R. M. (1989). *Real Analysis and Probability*. Wadsworth and Brooks/Cole, Pacific Grove, CA.

GETOOR, R. and SHARPE, M. (1972). Conformal martingales. *Invent. Math.* **16** 271–308.

IKEDA, N. and WATANABE, S. (1989). *Stochastic Differential Equations and Diffusion Processes*, 2nd ed. North-Holland, Amsterdam/Kodansha, Tokyo.

KAKUTANI, S. (1948). On equivalence of infinite product measures. *Ann. of Math. (2)* **49** 214–224.

KANTOROVICH, L. V. and RUBINSTEIN, G. S. (1958). On space of completely additive functions. *Vestnik Leningrad Univ.* **7** 52–59. (In Russian.)

PROTTER, P. (1990). *Stochastic Integration and Differential Equations*. Springer, Berlin.

REVUZ, D. and YOR, M. (1991). *Continuous Martingales and Brownian Motion*. Springer, Berlin.

ROGERS, L. C. G. and WILLIAMS, D. (1987). *Diffusions, Markov Processes, and Martingales* **2**. *Ito Calculus*. Wiley, New York.

ROKHLIN, V. A. (1949). Main notions of measure theory. *Mat. Sb.* **67** 107–150. (In Russian.)

SKOROKHOD, A. V. (1986). Random processes in infinite dimensional spaces. *Proc. Internat. Cong. Math.* (A. M. Gleason, ed.) 163–171. Amer. Math. Soc., Providence, RI. (In Russian.)

SMORODINSKY, M. (1995). An example of a nonstandard inverse filtration. Preprint.

STROOCK, D. W. and YOR, M. (1980). On extremal solutions of martingale problems. *Ann. Sci. École Norm. Sup. (4)* **13** 95–164.

TSIRELSON, B. S. (1975). An example of a stochastic differential equation having no strong solution. *Theory Probab. Appl.* **20** 416–418.

VERSHIK, A. M. (1968). A theorem on lacunary isomorphisms. *Functional Anal. Appl.* **2** 200–203.

VERSHIK, A. M. (1970). Decreasing sequences of measurable partitions, and their applications. *Soviet Math. Dokl.* **11** 1007–1011.

VERSHIK, A. M. (1971). A continuum of pairwise nonisomorphic dyadic sequences. *Functional Anal. Appl.* **5** 16–18.

VERSHIK, A. M. (1973). Approximation in measure theory. Dissertation, Leningrad Univ. (In Russian.)

VERSHIK, A. M. (1994). The theory of decreasing sequences of measurable partitions. *St. Petersburg Math. J.* **6** 705–761.

L. DUBINS
J. FELDMAN
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720
E-MAIL: dubins@stat.berkeley.edu
feldman@math.berkeley.edu

M. SMORODINSKY
B. TSIRELSON
SCHOOL OF MATHEMATICS
TEL AVIV UNIVERSITY
TEL AVIV 69978
ISRAEL