MATHEMATICS OF OPERATIONS RESEARCH
Vol. 21, No. 3, August 1996
Printed in U.S.A.

COMPATIBLE MEASURES AND MERGING

EHUD LEHRER AND RANN SMORODINSKY

Two measures, μ and $\tilde{\mu}$, are updated as more information arrives. If with μ-probability 1, the predictions of future events according to both measures become close, as time passes, we say that $\tilde{\mu}$ merges to μ. Blackwell and Dubins (1962) showed that if μ is absolutely continuous with respect to $\tilde{\mu}$ then $\tilde{\mu}$ merges to μ. Restricting the definition to prediction of near future events and to a full sequence of times yields the new notion of almost weak merging (AWM), presented here. We introduce a necessary and sufficient condition and show many cases with no absolute continuity that exhibit AWM. We show, for instance, that the fact that $\tilde{\mu}$ is diffused around μ implies AWM.

1. Introduction. Two probability measures, μ and $\tilde{\mu}$ are defined on the same space. μ and $\tilde{\mu}$ can be thought of as the true measure and the prior held by an agent, respectively. With time an increasing sequence of information, selected according to μ, becomes available. At any time n an atom, say P_n, of partition $\mathscr{P}_n$, which refines $\mathscr{P}_{n-1}$, is selected. The selection of P_n is done according to the measure $\mu(\cdot|P_{n-1})$, where $P_{n-1} \in \mathscr{P}_{n-1}$. The prior distribution, $\tilde{\mu}$, is updated in light of the information received and, therefore, after time n the real updated measure is $\mu(\cdot|P_n)$ and the assessed one is $\tilde{\mu}(\cdot|P_n)$.

In general, $\tilde{\mu}(\cdot|P_n)$, the posterior distribution, fails to be close to the true one, $\mu(\cdot|P_n)$. In order to get convergence, the belief and the true distribution must be compatible. Various types of compatibility imply different types of convergence. The strongest compatibility assumption is absolute continuity. Blackwell and Dubins (1962) showed that if μ is absolutely continuous with respect to $\tilde{\mu}$ (i.e., for every event A, $\mu(A) > 0$ implies $\tilde{\mu}(A) > 0$), them $\tilde{\mu}$ merges to μ as more information arrives. That is, with μ probability 1 the measures $\mu(\cdot|P_n)$ and $\tilde{\mu}(\cdot|P_n)$ over the future become close as n tends to infinity.

Recently, Kalai and Lehrer (1993) have used this result to show convergence to Nash equilibrium in repeated games. It turns out, however, that a weaker notion of merging is needed. In Blackwell and Dubins (1962), $\mu(\cdot|P_n)$ and $\tilde{\mu}(\cdot|P_n)$ are close to each other on the full range of the whole σ-field, including tail events. For most applications, however, closeness on near future events suffices. This motivated Kalai and Lehrer (1994) to introduce the notion of weak merging. We say that $\tilde{\mu}$ weakly merges to μ if $\tilde{\mu}(A|P_n)$ is close to $\mu(A|P_n)$ whenever A is a short-run event, namely, $A \in \mathscr{P}_{n-1}$. Unfortunately, some natural examples fail to exhibit weak merging.

We propose here a minor modification of the weak merging notion, and we provide a necessary and sufficient condition that accommodates many examples. In the new notion we still require closeness only on near future events but we require it only on a sequence of time periods with density 1. We say that $\tilde{\mu}$ almost weakly merges to μ if $\tilde{\mu}(A|P_n)$ is close to $\mu(A|P_n)$, where $A \in \mathscr{P}_{n-1}$ on all time periods n except, perhaps, of n's in a sparse sequence. The idea is that an agent who observes an increasing

Received June 7, 1994; revised February 24, 1995 and August 15, 1995.
AMS 1991 subject classification. Primary: 60F15; Secondary: 94A17, 90D15.
OR/MS Index 1978 subject classification. Primary: Probability/Limit theorems; Secondary: Games/Stochastic.
Key words. Merging of opinions, almost weak merging, strong law of large numbers.

0364-765X/96/2103/0697/$01.25

number of observations will be able, most of the time, to predict with high precision near future outcomes. Only on sparse set will he be surprised in the sense that the true distribution will not be close to the prediction.

While merging of measures does not depend on the particular filtration (increasing information structure), weak merging and almost weak merging do depend on it. It may occur with one information structure and may not with another. Thus, all the results presented here are relative to one specific filtration. In some contexts dealing with one information structure is natural. In stochastic processes, for instance, the realization of the n first variables naturally provides the information available at time n. In repeated games the histories of length n are the only reasonable information sets that one may deal with.

We present a weaker notion of compatibility than absolute continuity, and show that it implies almost weak merging. Specifically, our main theorem, Theorem 1, states that if with μ probability 1 the lower limit of the sequence $a_n = (\tilde{\mu}(P_n)/\mu(P_n))^{1/n}$ is at least 1, where P_n is the atom selected at time n, then $\tilde{\mu}$ almost weakly merges to μ. Obviously, the main contribution of this paper is to the case where there is a lack of absolute continuity. In this case $\tilde{\mu}(P_n)/\mu(P_n)$ will usually converge to 0 with probability 1. But if it converges to zero slow enough so that a_n converges to 1, then there is almost weak merging. Notice that the condition of absolute continuity should be checked not only on events generated in finite times but also on those events generated by the whole filtration including tail events. In our case, to the contrast, it is enough to restrict attention only to events in $\mathscr{P}_n$, $n = 1, 2, \dots$.

The main interest for applications in game theory, decision science, and economics seems to lie in Corollary 1 and in Example 2. In many instances the true distribution, μ, is not provided in its entirety. What is available is only the stage-transition probabilities (e.g., the probability to choose an action after any given history). Therefore, one may expect that any connection between the assessed distribution, $\tilde{\mu}$, and the true one will be via the stage-transition probabilities.

Similar to what was done in Kalai and Lehrer (1992), we define an ε-perturbation of μ to be a distribution whose stage-transition probabilities are asymptotically close to those of μ up to an ε. Thus, $\tilde{\mu}$ is diffused around μ if every ε-perturbation of μ is assigned a positive probability. In other words, the assessment regarding stage-transition probabilities is partially dispersed around the true one.

It is shown in Corollary 1 that when $\tilde{\mu}$ is diffused around μ then $\tilde{\mu}$ almost weakly merges to μ.

For the convenience of the reader all definitions, main results, and examples are concentrated in §2. The main proofs are given in §3. Section 4 is devoted to a generalization of the main theorem which provides a necessary and sufficient condition for almost weak merging.

2. Definitions and main results.

DEFINITION 1. A *filtration* on a measurable space $(\Omega, \mathscr{E})$ is a sequence of partitions $\{\mathscr{P}_n\}_{n=1}^{\infty}$ of Ω satisfying:

(i) $\forall n \mathscr{P}_n \subset \mathscr{E}$ and $\mathscr{P}_{n+1}$ refines $\mathscr{P}_n$.

(ii) The number of atoms in $\mathscr{P}_n$ is finite or countable.

(iii) Denoting $\mathscr{F}_n$ the field generated by the atoms of $\mathscr{P}_n$ and $\mathscr{F} = \vee_n \mathscr{F}_n$, the σ-field generated by all the fields $\mathscr{F}_n$, then $\mathscr{E} = \mathscr{F}$.

We emphasize the fact that all the assumptions and results apply to a specific filtration and may fail to hold for other ones. Let $\{\mathscr{P}_n\}_{n=1}^{\infty}$ be a fixed filtration throughout the paper. For any $\omega \in \Omega$ we denote by $P_n(\omega)$ the atom of $\mathscr{P}_n$ containing ω.

For any two probability measures $\mu, \tilde{\mu}$ on Ω the notions of merging (Blackwell 1957, Blackwell and Dubins 1962) and of weak merging (Kalai and Lehrer 1994) are defined. Following these definitions we define a weaker notion of merging.

DEFINITION 2. Let $\mathbb{N}$ be the set of integers and let $A \subseteq \mathbb{N}$. $\limsup |A \cap \{1, \dots, n\}|/n$ is the *upper density* of A, denoted $\mathrm{UD}(A)$. We say that A is *sparse* if its upper density is zero. A is *full* if $\mathbb{N} \setminus A$ is sparse.

DEFINITION 3. The probability measure $\tilde{\mu}$ *almost weakly merges* (AWM) to μ (denoted $\mu \xrightarrow{\text{AWM}} \mu$) along the filtration $\{\mathscr{P}_n\}_{n=1}^{\infty}$ if for any natural number l, for all $\varepsilon > 0$ and μ-a.e., $\omega \in \Omega$, there exists a full sequence of indices $\mathbb{N}(\omega, \varepsilon, l)$ such that

$$(1) \quad |\tilde{\mu}(A|P_n(\omega)) - \mu(A|P_n(\omega))| < \varepsilon \quad \forall n \in \mathbb{N}(\omega, \varepsilon, l) \text{ and } \forall A \in \mathscr{F}_{n+l}.$$

In case $\mathbb{N}(\omega, \varepsilon, l)$ is all $\mathbb{N}$ except for a finite number of integers we say that $\tilde{\mu}$ *weakly merges* to μ and when A is not restricted to $\mathscr{F}_{n+l}$ but rather inequality (1) holds for every $A \in \mathscr{F}$ we say that $\tilde{\mu}$ *merges* to μ.

Notice that merging implies weak merging which implies almost weak merging.

REMARK 1. Since $\mu(A|C)\mu(C|B) = \mu(A|B)$ whenever $A \subseteq C \subseteq B$, and since the intersection of a finite number of full sequences is also a full sequence, the previous definition can be rewritten with $l = 1$.

Our main result is the following theorem which provides us with a sufficient condition for $\tilde{\mu}$ to AWM to μ on $\{\mathscr{P}_n\}$:

THEOREM 1. *If for μ-almost every ω there is a full set $\mathbb{N}'$ s.t.*

$$\liminf_{n \in \mathbb{N}'} \left(\frac{\tilde{\mu}(P_n(\omega))}{\mu(P_n(\omega))} \right)^{1/n} \geq 1, \quad \text{then } \tilde{\mu} \overset{\text{AWM}}{\to} \mu.$$

The compatibility assumption of the theorem is that

$$\liminf \left(\frac{\tilde{\mu}(P_n(\omega))}{\mu(P_n(\omega))} \right)^{1/n} \geq 1$$

on a full set μ-a.s. To show that, indeed, it is weaker than absolute continuity, observe that $\tilde{\mu}(P_n(\omega))/\mu(P_n(\omega))$ is a μ-martingale which converges μ-a.s. to a positive number when $\mu \ll \tilde{\mu}$. Therefore, its nth root converges to 1. Thus, when $\mu \ll \tilde{\mu}$ $\liminf \tilde{\mu}(P_n(\omega))/\mu(P_n(\omega))^{1/n} \geq 1$.

In case where μ is not absolutely continuous w.r.t. $\tilde{\mu}$ the likelihood ratio $\tilde{\mu}(P_n(\omega))/\mu(P_n(\omega))$ may go to zero with μ-positive probability. Roughly speaking, the assumption of the theorem actually says that it converges to zero slowly enough to allow the nth root to be at least 1. The theorem refers to the limit inferior. A natural question is what happens to the limit superior. The following lemma tells about the limited superior.

LEMMA 1. *For any two probability measures μ and $\tilde{\mu}$,*

$$\limsup \left(\frac{\tilde{\mu}(P_n(\omega))}{\mu(P_n(\omega))} \right)^{1/n} \leq \quad \mu\text{-a.s.}$$

Lemma 1 states that the hypothesis of Theorem 1 is actually

$$\lim \left(\frac{\tilde{\mu}(P_n(\omega))}{\mu(P_n(\omega))} \right)^{1/n} \underset{n \to \infty}{\to} 1$$

on a full set μ-a.s.

We say that the probability measure λ is a *grain* of $\tilde{\mu}$ if $\tilde{\mu} = \alpha\lambda + (1-\alpha)\lambda'$, where $0 < \alpha \leq 1$ and λ' is a probability measure.

DEFINITION 4. $\tilde{\mu}$ is *diffused* around μ, if for every $\varepsilon > 0$ there exists a probability measure μ_ε satisfying

(i) μ_ε is a grain of $\tilde{\mu}$.

(ii) For μ-a.e. ω there exists a time $N = N(\omega, \varepsilon)$ s.t. for $n > N$,

$$\left| \frac{\mu_\varepsilon(P_n(\omega)|P_{n-1}(\omega))}{\mu(P_n(\omega)|P_{n-1}(\omega))} - 1 \right| < \varepsilon.$$

The measure μ_ε can be thought of as a perturbation of μ since the transition probabilities according to both, μ_ε and μ, are relatively close to each other. All the perturbations μ_ε are assigned positive probability according to $\tilde{\mu}$ and therefore we say that $\tilde{\mu}$ is diffused around μ.

REMARK 2. If $\mu \ll \tilde{\mu}$ then $\tilde{\mu}$ is diffused around μ. This is so because $\tilde{\mu}(P_n(\omega))/\mu(P_n(\omega))$ converges μ-a.s. to a positive number and therefore (ii) of Definition 4 is satisfied with $\mu_\varepsilon = \tilde{\mu}$ for every $\varepsilon > 0$. In other words, $\tilde{\mu}$ is the ε-perturbation of μ for every ε.

COROLLARY 1. *If $\tilde{\mu}$ is diffused around μ, then $\tilde{\mu} \xrightarrow{\text{AWM}} \mu$.*

The following is an example of two measures which neither merge nor weakly merge but they almost weakly merge.

EXAMPLE 1 (see also Kalai and Lehrer 1992). Let Ω be the space $\{0,1\}^{\mathbb{N}}$, and let $\mathscr{P}_n$ be the partition induced by the first n coordinates. Define μ to be the Dirac measure on the point $(1,1,\dots)$. Define $\tilde{\mu}$ as the measure $(1/2)\mu_1 + (1/2)\mu_2$ where μ_1 and μ_2 are defined as follows. μ_1 is the measure induced by a sequence $X_1, X_2, \dots$ of independent Bernoulli random variables, where prob$(X_n = 1)$ is 1 if $n \neq 2^{2^t}$, and it is $1/2$ if $n = 2^{2^t}$. The measure μ_2 is the one induced by the following. Denote by ν_n the measure induced by i.i.d. sequence $X_1, X_2, \dots$ of random variables, where prob$(X_1 = 1) = 1 - 1/n = 1 - \text{prob}(X_1 = 0)$. Set $\mu_2 = \Sigma(1/2^n)\nu_n$. In other words, with probability $(1/2^n)$ $(n = 1, 2, \dots)$ it is defined by a repeated toss of a coin assigning probability $1 - 1/n$ to 1.

One can show that after observing $2^{2^t} - 1$ times the outcome 1 the updated measure of $\tilde{\mu}$ assigns a probability close to $1/2$ to the event that the next outcome will be 1 while the updated measure of μ assigns the same event probability 1. Thus, $\tilde{\mu}$ does not weakly merge to μ, but $\tilde{\mu}$ is diffused around μ and so by Corollary 1, $\tilde{\mu}$ almost weakly merges to μ.

EXAMPLE 2. Let Θ be a set of parameters. For every $\theta \in \Theta$, μ_θ is a measure on Ω. The ε-neighborhood of θ is defined as

$$C(\theta, \varepsilon) = \{\theta' \in \Theta\text{: for } \mu_\theta\text{-almost every } \omega \in \Omega \text{ there is } N \text{ s.t. } n \geq N \text{ implies}$$
$$|\mu_{\theta'}(A|P_n(\omega))/\mu_\theta(A|P_n(\omega)) - 1| < \varepsilon \text{ for every } A \in \mathscr{F}_{n+1}\}.$$

In words, $C(\theta, \varepsilon)$ is the set of parameters θ' s.t. the posteriors of μ_θ and of $\mu_{\theta'}$, restricted to short-run events are close up to an ε.

Let $\mathscr{H}$ be a σ-field on Θ containing every $C(\theta, \varepsilon)$. Suppose that $\tilde{\mu}$ is the measure on Ω induced by a distribution F on $(\Theta, \mathscr{H})$ and by μ_θ. Namely,

$$\tilde{\mu}(A) = \int \mu_\theta(A)\, dF(\theta) \quad \text{for every } A \in \mathscr{F}.$$

If F ascribes a positive probability to every ε-neighborhood of θ then $\tilde{\mu}$ is diffused around μ_θ. By Corollary 1, $\tilde{\mu} \xrightarrow{\text{AWM}} \mu_\theta$. The meaning of this example is that if $\tilde{\mu}$ is diffused so as to assign the neighborhoods of θ positive probability then $\tilde{\mu}$ almost weakly merges to μ_θ.

An obvious special case of this example is a 0, 1 exchangeable process: $\Theta = [0, 1]$, F is a distribution on $[0, 1]$ and for every $\theta \in \Theta$, μ_θ is the measure induced by a sequence of i.i.d. Bernoulli sequence with parameter θ. However, in this case, one may obtain a stronger result.

COROLLARY 2. *Let μ, $\tilde{\mu}$, $\hat{\mu}$ be three probability measures satisfying*:

(i) *$\hat{\mu}$ weakly merges to μ.*

(ii) *$\hat{\mu}$ is a grain of $\tilde{\mu}$.*

Then $\tilde{\mu} \xrightarrow{\text{AWM}} \mu$.

EXAMPLE 3. If $\tilde{\mu}$ is an exchangeable process, as defined in Example 2, and if $F([a, b]) > 0$ for every interval $[a, b]$ containing θ, then $\tilde{\mu}$ weakly merges to μ_θ. Suppose that $\tilde{\mu}$ is a grain of $\tilde{\nu}$, then by Corollary 2, $\tilde{\nu} \xrightarrow{\text{AWM}} \mu_\theta$.

3. Proofs. We begin this section with a well-known lemma (see Smorodinsky 1971, for example).

LEMMA 2. *Let a_i, b_i, $i = 1, 2, \dots$ be nonnegative numbers such that $1 = \Sigma_i a_i \geq \Sigma_i b_i$. Then*

(i) $\Sigma_i a_i \log(b_i/a_i) \leq 0$.

(ii) *Given $\varepsilon > 0$ $\exists \delta = \delta(\varepsilon) > 0$ such that $\Sigma_i |a_i - b_i| > \varepsilon \Rightarrow \Sigma_i a_i \log(b_i/a_i) < -\delta$.*

LEMMA 3. *Let a_i, b_i, $i = 1, 2, \dots$ be nonnegative numbers. $\Sigma a_i = 1$.*

(i) *Given $\eta > 0$ $\exists \phi = \phi(\eta) > 0$ s.t. if $\Sigma_i b_i \leq 1 + \phi$ then $\Sigma a_i \log(b_i/a_i) \leq \eta$.*

(ii) *Given $\varepsilon > 0$ $\exists \phi = \phi(\varepsilon) > 0$ and $\exists \delta(\varepsilon) = \delta > 0$ such that $\Sigma_i b_i \leq 1 + \phi$ and $\Sigma |a_i - b_i| > \varepsilon$ both imply $\Sigma a_i \log(b_i/a_i) < -\delta$.*

PROOF. Take $c_i = b_i/(1 + \phi)$, so $\Sigma c_i \leq 1$. By Lemma 2:

$$0 \geq \sum a_i \log \frac{c_i}{a_i} = \sum a_i \log \frac{b_i/(1+\phi)}{a_i} = \sum a_i \left(\log \frac{b_i}{a_i} + \log\left(\left(\frac{1}{1+\phi} \right) \right) \right)$$

The first part of Lemma 3 is achieved by taking ϕ sufficiently small such that

$$-\log\left(\frac{1}{1+\phi} \right) \leq \eta.$$

For the second part take ϕ such that

$$\sum_i \left| \frac{b_i}{1+\phi} - a_i \right| > \varepsilon/2.$$

By the second part of Lemma 2, $\exists \delta$ such that

$$-2\delta \geq \sum a_i \log \frac{b_i/(1+\phi)}{a_i} = \sum a_i \left(\log \frac{b_i}{a_i} + \log\left(\frac{1}{1+\phi} \right) \right).$$

So:

$$\sum a_i \log \frac{b_i}{a_i} \leq -2\delta - \log \frac{1}{1+\phi}.$$

If necessary, decrease ϕ so that $\log(1/(1 + \phi)) > -\delta$ and the result is obtained. □

PROOF OF THEOREM 1. The hypothesis of the theorem implies that for every $\bar{e} < 1$ the set of those n's satisfying

$$\left(\frac{\tilde{\mu}(P_n(\omega))}{\mu(P_n(\omega))}\right)^{1/n} \geq \bar{e}$$

is a full set μ-a.s. Suppose to the contrary that $\tilde{\mu}$ does not AWM to μ, i.e., there exist $d > 0$ and $B \subset \Omega$ such that $\mu(B) > 0$ and $\forall \omega \in B\ \exists \mathbb{N}(\omega) \subset \mathbb{N}$ with upper density $d(\omega)$ greater than d, satisfying that $\forall n \in \mathbb{N}(\omega)\ \exists A_n(\omega) \in \mathscr{F}_n$ such that

$$|\tilde{\mu}(A_n(\omega)|P_{n-1}(\omega)) - \mu(A_n(\omega)|P_{n-1}(\omega))| > \varepsilon(\omega) > 0.$$

Without loss of generality we may assume that $\varepsilon(\omega) \geq \hat{\varepsilon} > 0\ \forall \omega \in B$. So, $\forall n \in \mathbb{N}(\omega)$,

$$\sum_{P_n \in \mathscr{P}_n} |\tilde{\mu}(P_n|P_{n-1}(\omega)) - \mu(P_n|P_{n-1}(\omega))| > \hat{\varepsilon}.$$

We write $\mu_n(\omega) \equiv \mu(P_n(\omega)|P_{n-1}(\omega))$ and $\tilde{\mu}_n(\omega) \equiv \tilde{\mu}(P_n(\omega)|P_{n-1}(\omega))$.

Define the following random variables:

$$(2) \qquad X_n(\omega) = \log\frac{\tilde{\mu}_n(\omega)}{\mu_n(\omega)} \quad \text{and} \quad Y_n(\omega) = \log\frac{\lambda_n^{\varepsilon_0}(\omega)}{\mu_n(\omega)}, \qquad n = 1, 2, \ldots,$$

where $\lambda_n^{\varepsilon_0}(\omega) = \max\{\tilde{\mu}_n(\omega), \varepsilon_0\, \mu_n(\omega)\}$. We are interested in X_n but since its second moment may fail to exist we modify it to obtain Y_n.

Note that

(i) $X_n(\omega) \leq Y_n(\omega)\ \forall n, \forall \omega$.

(ii) $\forall n\ \sum \lambda_n^{\varepsilon_0}(\omega) \leq \sum \tilde{\mu}_n(\omega) + \varepsilon_0\, \mu_n(\omega) \leq 1 + \varepsilon_0$.

The summations are over all atoms of $\mathscr{P}_n$. Formally, in a given atom of $\mathscr{P}_{n-1}$ we take a representative ω from each atom of $\mathscr{P}_n$. The summation is over all these representatives. Taking $\omega \in B$ and $n \in \mathbb{N}(\omega)$:

$$\sum |\lambda_n^{\varepsilon_0}(\omega) - \mu_n(\omega)| \geq \sum |\tilde{\mu}_n(\omega) - \mu_n(\omega)| - \sum |\lambda_n^{\varepsilon_0}(\omega) - \tilde{\mu}_n(\omega)| \geq \hat{\varepsilon} - \varepsilon_0.$$

So taking $\varepsilon_0 \leq \hat{\varepsilon}/2$ yields

$$(3) \qquad \sum |\lambda_n^{\varepsilon_0}(\omega) - \mu_n(\omega)| \geq \hat{\varepsilon}/2.$$

By the second part of Lemma 3 and (3) we may take ε_0 small enough such that for some positive $\delta = \delta(\hat{\varepsilon}/2)$,

$$(4) \qquad E(Y_n(\omega)|P_{n-1}(\omega)) = \sum \mu_n(\omega) \log\left(\frac{\lambda_n^{\varepsilon_0}(\omega)}{\mu_n(\omega)}\right) < -\delta \quad \forall n \in \mathbb{N}(\omega).$$

For this δ take $\alpha, \beta > 0$ small enough such that

$$-\delta\frac{d}{2} + \left(1 - \frac{d}{2}\right)\alpha \leq -\beta < 0.$$

By the first part of Lemma 3 take ε_0 such that

$$(5)\qquad E(Y_n(\omega)|P_{n-1}(\omega)) = \sum \mu_n(\omega)\log\frac{\lambda_n^{\varepsilon_0}(\omega)}{\mu_n(\omega)} \le \alpha \quad \text{for all } n \in \mathbb{N}.$$

The second moment of $Y_n(\omega)$ given $P_{n-1}(\omega)$ is bounded:

$$\begin{aligned}
&E(Y_n^2(\omega)|P_{n-1}(\omega))\\
&\quad = \sum \mu_n(\omega)\log^2\frac{\lambda_n^{\varepsilon_0}(\omega)}{\mu_n(\omega)}\\
&\quad = \sum_{\tilde{\mu}/\mu<\varepsilon_0} \mu_n(\omega)\log^2\frac{\varepsilon_0\,\mu_n(\omega)}{\mu_n(\omega)} + \sum_{\varepsilon_0\le\tilde{\mu}/\mu\le 1} \mu_n(\omega)\log^2\frac{\tilde{\mu}_n(\omega)}{\mu_n(\omega)}\\
&\qquad + \sum_{\tilde{\mu}/\mu>1} \mu_n(\omega)\log^2\frac{\tilde{\mu}_n(\omega)}{\mu_n(\omega)} \le \log^2\varepsilon_0 + \sum_{\varepsilon_0<\tilde{\mu}/\mu<1} \mu_n(\omega)\log^2\varepsilon_0\\
&\qquad + \sum_{\tilde{\mu}/\mu>1} \mu_n(\omega)\left(\sqrt{\frac{\tilde{\mu}_n(\omega)}{\mu_n(\omega)}}\right)^2 \le \log^2\varepsilon_0 + \log^2\varepsilon_0 + 1.
\end{aligned}$$

So the strong law of large numbers may be applied to the uncorrelated random variables $Y_n(\omega) - E(Y_n(\omega)|P_{n-1}(\omega))$.

For μ-a.e. $\omega \in \Omega$ there exists $n(\omega)$ such that for every $n \ge n(\omega)$,

$$(6)\qquad \frac{1}{n}\sum_{j=1}^{n} Y_j(\omega) \le \frac{1}{n}\sum_{j=1}^{n} E(Y_j(\omega)|P_{j-1}(\omega)) + \frac{\beta}{2}.$$

Take $\omega \in B$ and an infinite sequence $\mathbb{N}_1(\omega) \subset \mathbb{N}(\omega)$ such that $n \in \mathbb{N}_1(\omega)$ implies $n \ge n(\omega)$ and $(\#\{k\colon k \le n \text{ and } k \in \mathbb{N}(\omega)\})/n \ge d/2$. So by (4), (5) and (6) for $n \in \mathbb{N}_1(\omega)$,

$$\frac{1}{n}\sum_{j=1}^{n} Y_n(\omega) \le \frac{1}{n}\left[\left(1-\frac{d}{2}\right)n\alpha + \frac{d}{2}n(-\delta)\right] + \frac{\beta}{2} \le -\beta + \frac{\beta}{2} = -\frac{\beta}{2}$$

$$\Rightarrow \frac{1}{n}\sum_{j=1}^{n}\log\frac{\tilde{\mu}_j(\omega)}{\mu_j(\omega)} = \frac{1}{n}\sum_{j=1}^{n} X_j(\omega) \le \frac{1}{n}\sum_{j=1}^{n} Y_j(\omega) \le -\frac{\beta}{2},$$

which implies

$$\left(\frac{\tilde{\mu}(P_n(\omega))}{\mu(P_n(\omega))}\right)^{1/n} \le e^{-\beta/2} < 1 \quad \text{for all } n \in \mathbb{N}_1(\omega).$$

In order to show contradiction to the hypothesis of the theorem it remains to show that the upper density of $\mathbb{N}_1(\omega)$ is positive. When we delete from $\mathbb{N}(\omega)$ all the numbers in $\mathbb{N}_1(\omega)$ we remain with a set of numbers whose upper density is at most $d/2$. Since the upper density (UD) is subadditive, we obtain $d \le \mathrm{UD}(\mathbb{N}(\omega)) \le \mathrm{UD}(\mathbb{N}(\omega) \setminus \mathbb{N}_1(\omega)) + \mathrm{UD}(\mathbb{N}_1(\omega))$. Thus, $d \le d/2 + \mathrm{UD}(\mathbb{N}_1(\omega))$. Therefore, $\mathrm{UD}(\mathbb{N}_1(\omega)) > 0$ and the proof is complete. □

Recall that $\lambda_n^{\varepsilon_0}(\omega) = \max\{\tilde{\mu}_n(\omega), \varepsilon_0 \mu_n(\omega)\}$. On this basis, we define for every $\varepsilon_0 \geq 0$,

$$\tilde{\phi}_n^{\varepsilon_0}(P_n(\omega)) = \prod_{k=1}^{n} \lambda_k^{\varepsilon_0}(\omega). \tag{7}$$

For $\varepsilon_0 = 0$, $\tilde{\phi}_n^{\varepsilon_0}$ coincides with $\tilde{\mu}$. Thus, $\tilde{\phi}_n^{\varepsilon_0}$ is a modification of $\tilde{\mu}$ in those periods where $\tilde{\mu}(P_n | P_{n-1})$ is small compared to the μ counterpart.

REMARK 3. A careful reading of the proof shows that we have actually shown more than required. Notice that Y_n, defined in (3), depends on ε_0. We have actually shown also that if for every $\varepsilon_0 > 0$, $\liminf(\tilde{\phi}_n^{\varepsilon_0}(P_n(\omega))/\mu(P_n(\omega)))^{1/n} \geq 1$ μ-a.s. (which is a slightly weaker assumption than the one used—with the lim inf over a *full set*), then $\tilde{\mu} \xrightarrow{\text{AWM}} \mu$.

PROOF OF LEMMA 1. Take $Y_n(\omega)$, $X_n(\omega)$ as in the Proof of Theorem 1. By Equation (5) of the proof and using the strong law of large numbers for $Y_n(\omega)$, the following is obtained:

$$\begin{aligned}
\frac{1}{n} \sum_{j=1}^{n} X_j(\omega) &\leq \frac{1}{n} \sum_{j=1}^{n} Y_j(\omega) \\
&\leq \frac{1}{n} \sum_{j=1}^{n} E\big(Y_j(\omega) | P_{j-1}(\omega)\big) + \gamma \leq \alpha + \gamma
\end{aligned}$$

for arbitrarily small α, γ. Thus, $\limsup(1/n)\Sigma_{j=1}^{n} X_j(\omega) \leq 0$, implying

$$\limsup \left(\frac{\tilde{\mu}(P_n(\omega))}{\mu(P_n(\omega))} \right)^{1/n} \leq 1. \quad \square$$

PROOF OF COROLLARY 1.

$$\begin{aligned}
\frac{\tilde{\mu}(P_n(\omega))}{\mu(P_n(\omega))} &= \frac{\alpha_\varepsilon \mu_\varepsilon(P_n(\omega)) + (1 - \alpha_\varepsilon)\hat{\mu}(P_n(\omega))}{\mu(P_n(\omega))} \\
&\geq \alpha_\varepsilon \frac{\mu_\varepsilon(P_n(\omega))}{\mu(P_n(\omega))} = \alpha_\varepsilon \prod_{j=1}^{n} \frac{\mu_\varepsilon\big(P_j(\omega) | P_{j-1}(\omega)\big)}{\mu\big(P_j(\omega) | P_{j-1}(\omega)\big)} \\
&\geq \alpha_\varepsilon \prod_{j=1}^{N(\varepsilon, \omega)} \frac{\mu_\varepsilon\big(P_j(\omega) | P_{j-1}(\omega)\big)}{\mu\big(P_j(\omega) | P_{j-1}(\omega)\big)} \prod_{j=N(\varepsilon, \omega)+1}^{n} \frac{\mu_\varepsilon\big(P_j(\omega) | P_{j-1}(\omega)\big)}{\mu\big(P_j(\omega) | P_{j-1}(\omega)\big)} \\
&\geq \alpha_\varepsilon \prod_{j=1}^{N(\varepsilon, \omega)} \frac{\mu_\varepsilon\big(P_j(\omega) | P_{j-1}(\omega)\big)}{\mu\big(P_j(\omega) | P_{j-1}(\omega)\big)} (1 - \varepsilon)^{n - N(\varepsilon, \omega)}.
\end{aligned}$$

So,

$$\liminf \left(\frac{\tilde{\mu}(P_n(\omega))}{\mu(P_n(\omega))} \right)^{1/n} \geq 1 - \varepsilon,$$

for arbitrary small ε which, by Theorem 1, completes the proof. $\square$

PROOF OF COROLLARY 2. Fix $\varepsilon_0 > 0$ and recall the notation $\tilde{\phi}_n^{\varepsilon_0}(P_n(\omega))$ above. We similarly define $\hat{\phi}_n^{\varepsilon_0}(P_n(\omega))$ corresponding to $\hat{\mu}$. It is clear that since $\hat{\mu}$ merges to μ for μ-a.s. ω and $\forall\varepsilon > 0$ $\exists N$ s.t. $\forall n > N$,

$$\frac{\hat{\phi}_n^{\varepsilon_0}(P_n(\omega))}{\mu(P_n(\omega))} \geq (1-\varepsilon)\frac{\hat{\phi}_{n-1}^{\varepsilon_0}(P_{n-1}(\omega))}{\mu(P_{n-1}(\omega))}.$$

So

$$\frac{\hat{\phi}_n^{\varepsilon_0}(P_n(\omega))}{\mu(P_n(\omega))} \geq (1-\varepsilon)^{n-N}\frac{\hat{\phi}_N^{\varepsilon_0}(P_N(\omega))}{\mu(P_N(\omega))}.$$

Since $\hat{\mu}$ is a grain of $\tilde{\mu}$, $\tilde{\mu}/\hat{\mu} \geq \alpha > 0$. Therefore,

$$\frac{\tilde{\phi}_n^{\varepsilon_0}(P_n(\omega))}{\mu(P_n(\omega))} = \frac{\tilde{\phi}_n^{\varepsilon_0}(P_n(\omega))}{\hat{\phi}_n^{\varepsilon_0}(P_n(\omega))}\cdot\frac{\hat{\phi}_n^{\varepsilon_0}(P_n(\omega))}{\mu(P_n(\omega))} \geq \alpha(1-\varepsilon)^{n-N}\frac{\hat{\phi}_N^{\varepsilon_0}(P_N(\omega))}{\mu(P_N(\omega))}.$$

It follows that

$$\liminf\left(\frac{\tilde{\phi}_n^{\varepsilon_0}(P_n(\omega))}{\mu(P_n(\omega))}\right)^{1/n} \geq 1-\varepsilon.$$

As this is true for arbitrarily small ε, in view of Remark 3 the proof is complete. □

4. A characterization of AWM. The converse of Theorem 1 does not hold. Here is a counterexample.

EXAMPLE 4. We define $\tilde{\mu}$ on the interval $[0,1]$ by defining it on a filtration. The measure μ is the Lebesgue measure. We first define the filtration. We divide the diadic intervals one at a time. Let $\mathscr{P}_1' = \{[0,\frac{1}{2}),[\frac{1}{2},1]\}$. For getting $\mathscr{P}_2'$ we cut $[0,\frac{1}{2})$ into two: $\mathscr{P}_2' = \{[0,\frac{1}{4}),[\frac{1}{4},\frac{1}{2}),[\frac{1}{2},1]\}$. In $\mathscr{P}_3'[\frac{1}{2},1]$ is divided into two: $\mathscr{P}_3' = \{[0,\frac{1}{4}),[\frac{1}{4},\frac{1}{2}),[\frac{1}{2},\frac{3}{4}),[\frac{3}{4},1]\}$, and so forth. Thus, $\{\mathscr{P}_n\}$ generates the Borel σ-algebra. Now we replicate each one of the partitions $\mathscr{P}_n'$ 2^{n-1} times. We get the sequence $\mathscr{P}_1', \mathscr{P}_2', \mathscr{P}_2', \mathscr{P}_3', \mathscr{P}_3', \mathscr{P}_3', \mathscr{P}_3', \dots$ and we call it $\{\mathscr{P}_n\}$. Thus, whatever $\tilde{\mu}$ is, $\tilde{\mu}(P_n(\omega)|P_{n-1}(\omega)) = 1$ on a full sequence of times. The same applies to μ. Therefore, $\tilde{\mu} \xrightarrow{\text{AWM}} \mu$. In order to define $\tilde{\mu}$ we have to define it only in those stages n where one atom is being divided into two. Fix such n and define $\tilde{\mu}$ of the left part, say, A, to be so small compared to its Lebesgue measure that $(\tilde{\mu}(A)/\mu(A))^{1/2n} < 3/4$. Thus, for all time periods m between n and $2n$ we obtain $(\tilde{\mu}(A)/\mu(A))^{1/m} < 3/4$. Moreover, A is an atom of all the partitions between $\mathscr{P}_n$ and $\mathscr{P}_{2n}$. Therefore, for μ-almost all ω (because almost every ω appears infinitely many times in the left part of the divided atom) there exists a sequence of positive (at least $1/2$) upper density s.t. $(\tilde{\mu}(P_m|(\omega))/\mu(P_m(\omega)))^{1/m} < 3/4$. This refutes the hypothesis of Theorem 1.

We use $\phi_n^{\varepsilon_0}$ (see (7)) in order to establish a necessary and sufficient condition for AWM.

PROPOSITION 1. *Suppose that $\tilde{\mu} \to \xrightarrow{\text{AWM}} \mu$. Then for every $\varepsilon_0 > 0$,*

$$\liminf\left(\frac{\tilde{\phi}_n^{\varepsilon_0}(P_n(\omega))}{\mu(P_n(\omega))}\right)^{1/n} \geq 1 \qquad \mu\text{-}a.s.$$

PROOF. Fix $\varepsilon_0 > 0$. Using the random variables Y_n defined with $\lambda_n^{\varepsilon_0}(\omega)$ in (2), one may get, similar to (6), that

$$\frac{1}{n}\sum_{j=1}^{n} E\big(Y_j(\omega)|P_{j-1}(\omega)\big) \le \frac{1}{n}\sum_{j=1}^{n} Y_j(\omega) + \delta, \tag{8}$$

μ-a.s. whenever $n > n(\omega, \delta)$. The assumption of the proposition implies that the left side of (8) converges to 0 and therefore $0 \le \liminf(1/n)\sum_{j=1}^{n} Y_j(\omega)$. Thus,

$$1 \le \liminf\left(\frac{\tilde{\phi}_n^{\varepsilon_0}(P_n(\omega))}{\mu(P_n(\omega))}\right)^{1/n}. \qquad \square$$

We summarize Remark 3 and Proposition 1 in the following characterization of AWM.

THEOREM 2. *$\tilde{\mu} \xrightarrow{\text{AWM}} \mu$ if and only if for every $\varepsilon_0 > 0$,*

$$\liminf\left(\frac{\tilde{\phi}_n^{\varepsilon_0}(P_n(\omega))}{\mu(P_n(\omega))}\right)^{1/n} \ge 1 \qquad \mu\text{-}a.s.$$

With additional assumptions one can obtain a result that resembles the converse of Theorem 1:

COROLLARY 3. *Suppose that $\tilde{\mu} \xrightarrow{\text{AWM}} \mu$ and in addition assume that there is a random variable $c > 0$ s.t. $\liminf \tilde{\mu}(P_n(\omega)|P_{n-1}(\omega))/\mu(P_n(\omega)|P_{n-1}(\omega)) > c$ μ-a.s., then*

$$\left(\frac{\tilde{\mu}(P_n(\omega))}{\mu(P_n(\omega))}\right)^{1/n} \underset{n\to\infty}{\to} 1 \qquad \mu\text{-}a.s.$$

PROOF. The additional assumption assures that for μ-a.e. ω there is $\varepsilon_0 > 0$ s.t. for every n, $\tilde{\mu}(P_n(\omega)) = \tilde{\phi}_n^{\varepsilon_0}(P_n(\omega))$. The proof is complete by Theorem 2 and Lemma 1. $\square$

Acknowledgements. The first author was partially funded by NSF Grant No. SES-9223156. The second author was sponsored by the Roberto Neminovsky Fellowship.

References

Blackwell, D. (1957). On discrete variables whose sum is absolutely continuous. *Ann. Math. Statist.* **28** 520–521.

______, L. Dubins (1962). Merging of opinions with increasing information. *Ann. Math. Statist.* **38** 882–886.

Kalai, I. and E. Lehrer (1993). Rational learning leads to Nash equilibrium. *Econometrica* **61** 1019–1096.

______, ______ (1994). Weak and strong merging of opinions. *J. Math. Econom.* **23** 73–86.

______, ______ (1992). *Bayesian forecasting*, Mimeo.

Smorodinsky, M. (1971). Ergodic theory, entropy, *Lecture Notes in Mathematics*, Springer-Verlag.

E. Lehrer: Department of Managerial Economics and Decision Sciences, J. L. Kellogg Graduate School of Management, and Department of Mathematics, Northwestern University, 2001 Sheridan Road, Evanston, Illinois 60208, and Raymond and Beverly Sackler Faculty of Exact Sciences, School of Mathematical Sciences, Tel Aviv University, Ramat Aviv, Tel Aviv 69978, Israel; e-mail: lehrer@math.tau.ac.il

R. Smorodinsky: Department of Managerial Economics and Decision Sciences, J. L. Kellogg Graduate School of Management, Northwestern University, 2001 Sheridan Road, Evanston, Illinois 60208; e-mail: rann@nwu.edu