Statistics, Probability and Game Theory
IMS Lecture Notes – Monograph Series (1996) Volume 30

MERGING AND LEARNING

EHUD LEHRER AND RANN SMORODINSKY
Tel Aviv University and Northwestern University

Abstract.

This review presents the well-known notion of merging, introduced by Blackwell and Dubins, and its later generalizations. While the original concept of merging refers to all future events, the two new concepts, of weak and of almost weak merging, are concerned only with forecasting near-future events. Necessary and sufficient conditions for almost weak merging and necessary conditions for weak merging are given.

1. Introduction. The subject of converging to an equilibrium in game theoretical models has captured a lot of attention in the last years. The literature is roughly divided into three main branches: rational learning, fictitious-play type processes and evolutionary models. In this review we focus solely on rational learning, where players consider their future utility as well as their present one. In a sense the foundations were laid in a seminal paper of Blackwell and Dubins (1962).

Consider a discrete time stochastic process attaining only countably many values. Suppose that there are two distributions over the underlying probability space: the true distribution according to which this process evolves and a subjective distribution held by the decision maker who observes the process realizations. We say that an agent learns if his ability to forecast forthcoming events improves with time. Blackwell and Dubins (1962) introduced the learning notion of "Merging", and they entitled their paper "Merging of Opinions with Increasing Information". Our main concern, in this review, is to introduce the original concept of merging and some later extensions of it.

Naturally it may so happen that at an initial situation, when an agent has only his belief and has not yet seen any observation from the process, he may not be able to forecast correctly. However, it may happen that as time passes and the agent observes more outcomes of the stochastic process he updates his belief, in a Bayesian manner, and learns to forecast with increasing precision. Blackwell and Dubins focused on the ability to forecast long-term events, including tail events. Later developments, inspired by models where agents discount their future payoffs, emphasize forecasting of short-term events.

Obviously, not all beliefs will guarantee learning. We are about to introduce various initial conditions on the true measure and on the initial belief that ensure different types of asymptotic learning. Throughout this article, we refer to these initial conditions as "compatibility conditions".

It is important to distinguish between the kind of Bayesian learning treated here and other kinds of Bayesian Learning. In many general questions where learning is concerned one may have a model in which, initially, some parameter is chosen at random and according to this parameter the stochastic process evolves. The agent knows how the parameter is chosen but does not know its realization. When one says that the agent learns, one may refer to one of two kinds of learning : learning the parameter and learning to forecast future outcomes. These two types of learning are distinct and need not occur simultaneously. Any of these types may happen without the other. Motivated by game theoretical models we are mainly concerned with the second kind of learning. In order to clarify the distinction between the two kinds we provide some examples in Section 2. In all these examples, as well as all examples throughout this review, we shall look into stochastic processes which take only two values at any stage, 0 and 1.

In Section 3 we present the two models which we will work with throughout this paper. The first is referred to as the subjective model and the second is the Bayesian model, which is frequently used in Bayesian Statistics. Section 4 is devoted to the different notions of compatibility between two given probability measures, defined on the same measurable space. Section 5 consists of the definitions of various notions of merging. Section 6 and 7 tie compatibility conditions with merging and Section 8 is devoted to examples.

2. Examples.

Example 1 - Learning the parameter without merging. Suppose the parameter space is the interval $[0,1]$. Given a parameter, the process on $\{0,1\}^{\mathbf{N}}$ is deterministic, and it is simply the binary expansion of the parameter. Namely at time t the process will give the $t-th$ digit of the binary expansion of the parameter. Note that as time goes by the agent, not knowing the parameter has increasing knowledge about it. Nevertheless the agent's forecast for the next outcome will always be 1 with probability $\frac{1}{2}$ and 0 with probability $\frac{1}{2}$, as opposed to the forecast of someone who is knowledgeable of the parameter, which is either 1 with probability 1 or 0 with probability 1.

Example 2 - Merging without learning the parameter. Now suppose the parameter space has two components, a and b, chosen by flipping a fair coin. If a is chosen, then the process is such that the first outcome is determined by a $(\frac{1}{3}, \frac{2}{3})$ coin, and from the second outcome on a fair coin is used. On the other hand if the parameter b is chosen then the outcomes from the first stage are chosen according to a fair coin. In this example it is obvious that an agent knows from stage 2 on to forecast as if he knew the true parameter, but he never learns anything about the parameter, other than what he learned in stage 1.

Example 3 - Merging and learning the parameter. Once again suppose the parameter space has 2 components, a and b chosen by flipping a fair coin. If a is chosen then the process is deterministically 1, and if b is chosen then the process is deterministically 0. In this example, independent of the realization of the parameter, the agent knows the parameter after the first stage, and can forecast accurately the future.

Example 4 - No merging and no learning of the parameter. This example is a slight modification of Example 1. Let the parameter space be the interval $[0, 1]$ and let the outcome of the process at time t be the digit in the 2^t coordinate of the binary expansion of the parameter. Once again no merging occurs as an agent who does not know the parameter will forever assign probability $\frac{1}{2}$ to the next out-come being 0 or 1, where as the next outcome is deterministic. As opposed to Example 1 no complete learning of the parameter occurs. The agent does collect some information on the parameter but the accuracy by which he guesses the parameter does not diminish to zero.

As previously mentioned we study here the notion of merging, while the issue of learning the parameter is either ignored or considered as a tool for merging. That is, in some cases learning the parameter will suffice for merging to occur (which is not a general phenomenon as clarified by Example 1).

3. The Model. We consider here only the simple case in which the set of outcomes of the stochastic process is either finite or countable. The results can be easily extended to any Borel set of outcomes.

Let $(\Omega, \mathcal{B})$ be a measurable space.

Definition 1. A *filtration* on $(\Omega, \mathcal{B})$ is a sequence of partitions of Ω, denoted $\{\mathcal{P}_n\}_{n=1}^{\infty}$, satisfying:

i) $\forall n \; \mathcal{P}_n \subset \mathcal{B}$.

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

iii) Denoting by $\mathcal{F}_n$ the field generated by the atoms of $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_n$, and $\mathcal{F} = \bigvee_n \mathcal{F}_n$, the σ-field generated by all the fields $\mathcal{F}_n$, then $\mathcal{B} = \mathcal{F}$.

The results presented throughout, unless specifically mentioned otherwise, are for a given, fixed filtration. For any $w \in \Omega$ we denote by $P_n(w)$ the atom of $\mathcal{F}_n$ containing w.

In what follows we shall state definitions and results in two models. The first model, called the *subjective model*, is a model where two measures $\mu, \tilde{\mu}$, on $(\Omega, \mathcal{B})$, referred to as the "true" measure and the "belief" respectively, are the primitives. In the second model, a measurable set $(\Theta, \mathcal{C})$ of parameters and a distribution F on it, are given as primitives and $\forall \theta \in \Theta \; \mu_\theta$ is a measure on $(\Omega, \mathcal{B})$. For obvious reasons we assume that the map $\theta \longrightarrow \mu_\theta(S)$ is Borel measurable $\forall S \in \mathcal{B}$. The second model is referred to as the *Bayesian model*. In this model denote by θ_0 the parameter initially chosen according to the distribution F, and think of μ_{θ_0} as the "truth" and of $\mu_\Theta = \int \mu_\theta dF(\theta)$ as the "belief". For any measurable $\Theta' \subset \Theta$ we shall use the notation $\mu_{\Theta'} = \frac{1}{F(\Theta')} \int_{\theta \in \Theta'} \mu_\theta dF(\theta)$.

4. Different Notions of Compatibility. This section primarily consists of the definitions of the various notions of compatibility. The section begins with the stronger notions, which will later yield strong notions of learning.

Some of the definitions will be stated in two versions. Version (a) will comply with the subjective model and version (b) will comply with the Bayesian model. It is easy to check that in all of these definitions, given version (b), version (a) of the definition is satisfied, with the proper interpretation of the true measure (μ) and the belief ($\tilde{\mu}$).

Definition 2(a) (Kalai & Lehrer). Given two probability measures μ and $\tilde{\mu}$ on $(\Omega, \mathcal{B})$ we say that μ is a *grain* of $\tilde{\mu}$ if there exists $\alpha \in (0,1]$ s.t $\tilde{\mu} = \alpha\mu + (1-\alpha)\widehat{\mu}$ where $\widehat{\mu}$ is an arbitrary probability measure on $(\Omega, \mathcal{B})$. In other words we say that $\tilde{\mu}$ holds a *grain of truth* (w.r.t. the "truth", μ).

Definition 2(b). The measurable set of parameters $(\Theta, \mathcal{C}, F)$ is said to be a *set of grains* if F is purely atomic.

Definition 3(a). μ is *absolutely continuous* with respect to (w.r.t) $\tilde{\mu}$ (denoted $\mu \ll \tilde{\mu}$) if for any $S \in \mathcal{B}$ $\mu(S) > 0$ implies $\tilde{\mu}(S) > 0$.

Remark 1. One can interpret definition 3(a) as follows: μ is absolutely continuous w.r.t $\tilde{\mu}$ means that an agent with a subjective belief can never be surprised by an event which can actually happen (i.e. an event which is "truly" assigned positive probability). Note that among all such events are included events which are not in any of the fields $\mathcal{F}_n$. That is given the filtration $\{\mathcal{P}_n\}_{n=1}^{\infty}$ definition $3(a)$ includes sets which are in the tail σ-field. These can be interpreted, in turn, as events which cannot be verified in any finite time, that is events which belong to the infinite horizon future.

Definition 3(b). The set of parameters Θ is *absolutely continuous within* if $\forall\theta \in \Theta$ μ_θ is absolutely continuous w.r.t $\tilde{\mu} = \int \mu_\theta dF(\theta)$.

Remark 2.

i) If $\tilde{\mu}$ is a grain of truth w.r.t. μ then μ is absolutely continuous w.r.t $\tilde{\mu}$.
ii) If the set of parameters Θ is a set of grains then it is also absolutely continuous within.

Note that, as opposed to the following definition, Definitions $2(a), (b)$ and $3(a), (b)$ are independent of the filtration.

In what follows 0/0 is understood as 0 and $\mu(A \mid B) = 0$ when $\mu(B) = 0$ for any probability measure μ. .

Definition 4. A measure μ_2 is said to be ϵ - *asymptotically near* the measure μ_1 if μ_1-almost everywhere (a.e.) the accumulation points of the sequence $\frac{\mu_2\left(P_n(w)|P_{n-1}(w)\right)}{\mu_1\left(P_n(w)|P_{n-1}(w)\right)}$ for $n = 1, 2, \ldots$ lie in the interval $(1-\epsilon, 1+\epsilon)$.

Definition 5(a). The belief $\tilde{\mu}$ is *diffused* around μ if $\forall\epsilon > 0$ there exists a probability measure μ_ϵ on $(\Omega, \mathcal{B})$ such that:

i) μ_ϵ is a grain of $\tilde{\mu}$;
ii) μ_ϵ is ϵ - asymptotically near μ.

What definition 5(a) states is that the belief $\tilde{\mu}$ holds a grain, which is not necessarily a grain of truth but is, in some sense an arbitrarily small neighborhood of the truth.

Definition 5(b). The parameter set Θ is *diffused* if for every $\theta \in \Theta$ and $\epsilon > 0$ the ϵ-neighborhood of θ, denoted $C(\theta, \epsilon)$, and defined by:

$$C(\theta,\epsilon) = \Big\{\theta' \in \Theta;\ \text{for } \mu_\theta\text{-almost every } w \in \Omega, \exists N \text{ s.t. } n \geq N \text{ implies} \\ \Big|\mu_{\theta'}\big(A \mid P_n(w)\big) \Big/ \mu_\theta\big(A \mid P_n(w)\big) - 1\Big| < \epsilon \text{ for every } A \in \mathcal{F}_{n+1}\Big\}.$$

satisfies $F(C(\theta, \epsilon)) > 0$.

The next pair of definitions are concerned with the non-compatibility of two measures.

Definition 6(a). A measure λ has *the separation property* w.r.t μ if there exists $d > 0$ such that for sufficiently many stages, λ is not d-close to μ. That is, the following set of stages:

$$\{s \in \mathbb{N}; \exists A \in \mathcal{F}_{s+1} \quad s.t. \quad |\lambda(A|P_s(\omega)) - \mu(A|P_s(\omega))| > d\}$$

has positive lower density[1], μ-a.e.

Definition 6(b). A set $\Theta' \subset \Theta$ has *the separation property* w.r.t a given $\theta_0 \in \Theta$ if $\mu_{\Theta'}$ has the separation property w.r.t μ_θ, where $\mu_{\Theta'} = \int_{\theta\in\Theta'} \mu_\theta dF(\theta)$.

We now turn to the last pair of definitions related to compatibility.

Definition 7(a). The belief $\tilde{\mu}$ *accommodates* μ if $\forall\epsilon > 0$ there exists a probability measure μ_ϵ on $(\Omega, \mathcal{B})$ such that:

i) μ_ϵ is a grain of $\tilde{\mu}$;
ii) μ_ϵ is ϵ - asymptotically near μ.
iii) λ_ϵ is a convex combination of probability measures $\lambda_\epsilon^j (1 \leq j \leq M)$, $\lambda_\epsilon = \sum \beta_j \lambda_\epsilon^j$, where λ_ϵ^j has the separation property w.r.t μ.

Definition 7(b). The set Θ is *accommodating* if $\forall\theta \in \Theta$ and $\forall\epsilon > 0$ the ϵ-neighborhood of θ, $C(\theta, \epsilon)$ (see definition 5(b)) satisfies:

i) $F(C(\theta, \epsilon)) > 0$;
ii) There exists a finite partition of $C(\theta, \epsilon)^c$, i.e., the complementary of $C(\theta, \epsilon)$, $C(\theta, \epsilon)^c = \cup_{j=1}^M \Theta_j$ s.t. $\forall j = 1, 2, \ldots, M$, Θ_j has the separation property w.r.t θ_0.

[1] Let $\mathbb{N}$ be the set of integers and let $A \subseteq \mathbb{N}$. $\liminf \big|A \cap \{1, \ldots, n\}\big|/n$ is the *lower density* of A.

Remark 3.

i) If $\tilde{\mu}$ accommodates μ then, trivially, $\tilde{\mu}$ is diffused around μ.

ii) If Θ is accommodating then, trivially, Θ is diffused.

Remark 4.

i) If μ is absolutely continuous w.r.t $\tilde{\mu}$ then $\tilde{\mu}$ accommodates μ.

ii) If Θ is absolutely continuous within then Θ is accommodating.

Proof

i) Simply take for every $\epsilon > 0$ $\mu_\epsilon = \tilde{\mu}$ and $\alpha = 1$. An application of the Martingale Convergence Theorem gives that the following quotient, $\tilde{\mu}(P_{s+1}(\omega)|P_s(\omega))/\mu(P_{s+1}(\omega)|P_s(\omega))$ converges to 1.

ii) Follows immediately from part i). ■

We have therefore defined four notions of compatibility. By Remarks 1, 2 and 3 we note that these notions have the following hierarchy: grain $\Rightarrow$ absolute continuity $\Rightarrow$ accommodation $\Rightarrow$ diffusion.

5. Notions of Merging. This section is devoted to the definition of the different notions of learning to forecast. We begin with the strongest definition due to Blackwell & Dubins (1962):

Definition 8 (Blackwell & Dubins). The probability measure $\tilde{\mu}$ *merges* to μ (denoted $\tilde{\mu} \xrightarrow{\text{M}} \mu$) along the filtration $\{\mathcal{P}_n\}_{n=1}^{\infty}$ if for all $\epsilon > 0$ and $\mu - a.e\ w \in \Omega$ there exists $N = N(\epsilon, \omega) \in \mathbb{N}$ s.t for all $n > N$

$$(1) \qquad \left|\tilde{\mu}(A \mid P_n(w)) - \mu(A \mid P_n(w))\right| < \epsilon \quad \text{and } \forall A \in \mathcal{B}.$$

What definition 8 states is that an agent who has an initial belief $\tilde{\mu}$ and who updates his belief in a Bayesian manner will asymptotically assign the same probabilities to any event. Note that this forecasting power of the agent is very strong as the agent is eventually able to forecast correctly even events in the infinite horizon (i.e., events in the tail σ-field), the occurrence of which can never be verified.

In decision problems when the near future is more payoff-relevant than the far future, forecasting short-term events is more important than forecasting long-term events. Motivated by such problems, the following definition refers solely to finite horizon events.

Definition 9 (Kalai & Lehrer). The probability measure $\tilde{\mu}$ *weakly merges* (WM) to μ (denoted $\tilde{\mu} \xrightarrow{\text{WM}} \mu$) along the filtration $\{\mathcal{P}_n\}_{n=1}^{\infty}$ if for any natural number ℓ, for all $\epsilon > 0$ and $\mu - a.e\ w \in \Omega$, there exists $N = N(\epsilon, \omega, \ell) \in \mathbb{N}$ s.t. for all $n > N$

$$(2) \qquad \left|\tilde{\mu}(A \mid P_n(w)) - \mu(A \mid P_n(w))\right| < \epsilon \qquad \forall A \in \mathcal{F}_{n+\ell}\,.$$

Remark 5. The definition of weak merging can be rewritten with $\ell = 1$, instead of an arbitrary ℓ.

Proof: The general argument for extending the definition from ℓ to $\ell+1$ is a mere repetition of the following argument which extends the definition from $\ell = 1$ to $\ell = 2$.

Define $g_t(w)$ to be the indicator function of the set

$$B_t(\epsilon) = \{w; \forall A \in \mathcal{F}_{t+1} \quad |\mu(A|P_t(w)) - \tilde{\mu}(A|P_t(w))| < \epsilon\}.$$

Assume definition 9 holds with $\ell = 1$. Then μ-a.e. $g_t(w) \to 1$. By Lemma 1 of Kalai & Lehrer (1994) the set

$$\begin{aligned} C_t(\epsilon) &= \{w; g_s(w) \neq 1 \quad \forall s > t\} \\ &= \{w; |\frac{g_s(w)}{1} - 1| > 1 - \epsilon \quad \forall s > t\}, \end{aligned}$$

satisfies that for some $t(\epsilon)$ if $t > t(\epsilon)$ then

$$\mu(\{w; \mu(C_t(\epsilon)|P_s(w)) < \epsilon, \forall s > t\}) > 1 - \epsilon.$$

By the assumption, for μ-a.e $w_0 \in \Omega$ there is $N = N(\epsilon, w_0)$ such that for all $n > N$

$$\Big|\tilde{\mu}(A \mid P_n(w_0)) - \mu(A \mid P_n(w_0))\Big| < \epsilon \quad \forall A \in \mathcal{F}_{n+1}.$$

Let $n \geq \max(N, t(\epsilon))$ and set $C = C_t(\epsilon)$. Now let $A \in \mathcal{F}_{n+2}$ and define, $D = \cup_{P_{n+1} \not\subset C} P_{n+1}$. Notice that $\mu(A|D) = \sum_{P_{n+1} \subset D} \big(\frac{\mu(A|P_{n+1})\mu(P_{n+1}|P_n(w_0))}{\mu(D|P_n(w_0))}\big)$. Thus, $|\mu(A|D) - \tilde{\mu}(A|D)| \leq 5\epsilon$. We therefore obtain,

$$|\mu(A|P_n(w_0)) - \tilde{\mu}(A|P_n(w_0))| \leq$$

$$\leq |\mu(A|D)\mu(D|P_n(w_0)) - \tilde{\mu}(A|D)\tilde{\mu}(D|P_n(w_0))| +$$

$$|\mu(A|D^c)\mu(D^c|P_n(w_0)) - \tilde{\mu}(A|D^c)\tilde{\mu}(D^c|P_n(w_0))|$$

$$\leq \mu(D|P_n(w_0))|\mu(A|D) - \tilde{\mu}(A|D)| + \tilde{\mu}(A|D)|\mu(D|P_n(w_0)) - \tilde{\mu}(D|P_n(w_0))| +$$

$$\mu(D^c|P_n(w_0))|\mu(A|D^c) - \tilde{\mu}(A|D^c)| + \tilde{\mu}(A|D^c)|\mu(D^c|P_n(w_0)) - \tilde{\mu}(D^c|P_n(w_0))|$$

$$\leq \mu(D|P_n(w_0))5\epsilon + \tilde{\mu}(A|D)\epsilon + \epsilon|\mu(A|D^c) - \tilde{\mu}(A|D^c)| + \tilde{\mu}(A|D^c)\epsilon \leq 9\epsilon. \quad \blacksquare$$

Looking at specific examples of measures μ and $\tilde{\mu}$ (see Examples 7 and 9 in the last part of this review) the next definition seems to be quite natural. Definition 9 requires that forecasting near-future events with ϵ accuracy will take place from a certain point in time on. Instead, Definition 10 requires it for "almost all" periods. That is, for having "almost weak merging", forecasting near-future events with ϵ accuracy must occur only over a set of periods having density 1.

Definition 10 (Lehrer & Smorodinsky). The probability measure $\tilde{\mu}$ *almost weakly merges* (AWM) to μ (denoted $\tilde{\mu} \xrightarrow{\text{AWM}} \mu$) along the filtration $\{\mathcal{P}_n\}_{n=1}^{\infty}$ if for any natural number ℓ, for all $\epsilon > 0$ and $\mu - a.e\ w \in \Omega$, there exists a full[2] sequence of indices $\mathbb{N}(w, \epsilon, \ell) \subset \mathbb{N}$ such that

$$(3)\quad \Big|\tilde{\mu}\big(A \mid P_n(w)\big) - \mu\big(A \mid P_n(w)\big)\Big| < \epsilon \quad \forall n \in \mathbb{N}(w, \epsilon, \ell) \quad \text{and } \forall A \in \mathcal{F}_{n+\ell}\,.$$

Remark 6. Due to similar arguments as in Remark 5 and due to the fact that the intersection of a finite number of full sequences is also a full sequence, the definition of almost weak merging can be rewritten with $\ell = 1$, instead of an arbitrary ℓ. In other words, one can restrict the definition to the next stage events.

6. Relating Compatibility to Merging. We now make the connections between the different notions of compatibility and merging. We begin with a well-known result of Blackwell & Dubins (1962):

Theorem 1. If μ is absolutely continuous w.r.t $\tilde{\mu}$ then $\tilde{\mu} \xrightarrow{\text{M}} \mu$.

We shall only mention that the proof is an application of the Radon-Nikodym and the Martingale Convergence theorems. Note that Theorem 1 is actually independent of the specific filtration. This is because the definition of absolute continuity is independent of the filtration.

An immediate Corollary of Theorem 1 is obtained by applying the theorem to a set of parameters which is absolutely continuous within.

Corollary 1. If Θ is a set of parameters which is absolutely continuous within, then $\forall \theta_0 \in \Theta$, μ_Θ merges to μ_{θ_0}.

Corollary 2.

i) If μ is a grain of $\tilde{\mu}$ then $\tilde{\mu} \xrightarrow{\text{M}} \mu$.

ii) If Θ is a set of grains then, $\forall \theta_0 \in \Theta$, μ_Θ merges to μ_{θ_0}.

Proof: Immediate from Remark 2, Theorem 1 (part (i)) and Corollary 1 (part (ii)). ∎

Theorem 1 relates two notions, both involve tail σ-field events. The main interest in this review is to model learning by agents who discount the future. In such situations the agents are almost indifferent among events in the far horizon, not to mention events in the infinite horizon. We therefore turn to study the case where weak and almost weak merging occur.

Theorem 2. If $\tilde{\mu}$ accommodates μ then $\tilde{\mu}$ weakly merges to μ.

The proof of Theorem 2 relies on the following three lemmas (stated without the proofs. The proofs can be found in Lehrer & Smorodinsky (1994)).

Lemma 1. Let $\{p_i\}$ and $\{q_i\}$ be two sequences of nonnegative numbers that sum to at most 1. Then, $\sum_i |p_i - q_i| \geq d$ implies $\sum_i \sqrt{p_i \cdot q_i} \leq 1 - \frac{d^2}{8}$.

[2] A sequence of indices in $\mathbb{N}$ is full if its lower density is 1.

Lemma 2. Suppose that μ_ϵ is ϵ - asymptotic near μ. Then for μ-almost every ω there is a time S s.t. $s \geq S$ implies

$$(4). \qquad |\mu_\epsilon(A|P_s(\omega)) - \mu(A|P_s(\omega))| < \epsilon \quad \text{for every} \quad A \in \mathcal{P}_{s+1}$$

Lemma 3. Suppose $\tilde{\mu}$ is diffused around μ then μ-a.e $\liminf \left(\frac{\tilde{\mu}\left(P_s(w)\right)}{\mu\left(P_s(w)\right)} \right)^{1/s}$ ≥ 1.

We now turn to the proof of Theorem 2.

Proof of Theorem 2. Fix an $\epsilon > 0$ and let μ_ϵ and λ_ϵ^j be as in definition 7(a), for some $1 \leq j \leq M$. Define $X_s(\omega) = \left[\frac{\lambda_\epsilon^j\left(P_s(\omega)|P_{s-1}(\omega)\right)}{\mu\left(P_s(\omega)|P_{s-1}(\omega)\right)} \right]^{1/2}$, and $Y_s = X_s - E\left(X_s|\mathcal{P}_{s-1}\right)$. Notice that the sequence $\{Y_s\}$ satisfies:

i) The second moment is uniformly bounded;

ii) $\forall s \neq t$ Y_s and Y_t are uncorrelated. We show this for the case $t = s+1$ (showing this for all s and t such that $s \neq t$ is done in the same way, using more cumbersome notation): $E(Y_s Y_{s+1}) = E(E(Y_s Y_{s+1}|\mathcal{F}_s)) = E(Y_s E(Y_{s+1}|\mathcal{F}_s)) = E(Y_s E(X_{s+1} - E(X_{s+1}|\mathcal{F}_s)|\mathcal{F}_s)) = 0$. On the other hand $E(Y_s)E(Y_{s+1}) = 0$ as well, and therefore $Cov(Y_s, Y_{s+1}) = 0$.

We can now apply the strong law of large numbers for uncorrelated random variables and conclude that there exists a sequence δ_s diminishing to zero such that:

$$(5) \qquad (1/S)\sum_{s=1}^{S} X_s \leq (1/S)\sum_{s=1}^{S} E\left(X_s|\mathcal{F}_{s-1}\right) + \delta_S \; .$$

For a period s when there is an $A \in \mathcal{F}_s$ s.t. $|\lambda_\epsilon^j\left(A|P_{s-1}(\omega)\right) - \mu\left(A|P_{s-1}(\omega)\right)| > d$ (see part (iii) of Definition 7(a) and Definition 6(a)) denote by p_i the measures of atoms of $\mathcal{F}_s$ according to $\lambda_\epsilon^j\left(\bullet|P_{s-1}(w)\right)$, and by q_i the measures of these atoms according to $\mu\left(\bullet\,|P_{s-1}(w)\right)$ (this follows from the inequality of averages). Thus, $E\left(X_s|P_{s-1}(w)\right) = \sum_i q_i\sqrt{\frac{p_i}{q_i}} = \sum_i \sqrt{p_i q_i}$, where the expectation is according to μ.

By Lemma 1, $\sum\sqrt{p_i q_i} \leq (1 - d^2/8)$. Certainly, $E\left(X_s|P_{s-1}(w)\right)$ is at most 1 for all s. However, on a sequence of times with a lower density, say, $\eta > 0$, it is at most $(1 - d^2/8)$. Thus, from the right side of (3),

$$(6) \quad (1/S)\sum_{s=1}^{S} E\left(X_s|P_{s-1}(w)\right) + \delta_S \leq (1 - d^2/8)(\eta/2) + (1 - \eta/2) + \delta_S \; ,$$

for sufficiently large S. (The factor is $\eta/2$ because the density η is achieved only in the limit, but for sufficiently large S the density of those times where λ_ϵ^j and μ are remote by at least d is greater than $\eta/2$.)

If δ_S is smaller than $\eta/2$ we can bound the right side of (6) by $(1-\beta)$, where $\beta > 0$. Combine (5) and (6) and the inequality of averages to get:

$$\Big(\prod_{s=1}^{S} X_s\Big)^{1/S} \leq (1/S)\sum_{s=1}^{S} X_s \leq (1-\beta)\,.$$

Hence, $\prod_{s=1}^{S} X_s \leq (1-\beta)^S$. But

$$\prod_{s=1}^{S} X_s = \left(\frac{\lambda_\epsilon^j(P_S(\omega))}{\mu(P_S(\omega))}\right)^{1/2}.$$

We conclude that $\frac{\lambda_\epsilon(P_s(\omega))}{\mu(P_s(\omega))} \leq (1-\beta)^{2s}$, for some $\beta > 0$.

By Remark 3 $\tilde{\mu}$ is diffused around μ, so by Lemma 3 we can conclude that $\left(\frac{\tilde{\mu}(P_s(\omega))}{\mu(P_s(\omega))}\right)^{1/s} \to 1\ \mu - a.s.$

We now obtain

(7)
$$\frac{\lambda_\epsilon^j((P_s(\omega))}{\tilde{\mu}(P_s(\omega))} = \frac{\lambda_\epsilon^j(P_s(\omega))}{\mu(P_s(\omega))}\,\frac{\mu(P_s(\omega))}{\tilde{\mu}(P_s(\omega))} \leq (1-\beta)^{2s}\left[\frac{\mu(P_s(\omega))}{\tilde{\mu}(P_s(\omega))}\right] =$$
$$= \left[(1-\beta)^2\left(\frac{\mu(P_s(\omega))}{\tilde{\mu}(P_s(\omega))}\right)^{1/s}\right]^s \sim (1-\beta)^{2s} \to 0 \quad \mu - a.s.$$

Since $\tilde{\mu} = \alpha\mu_\epsilon + (1-\alpha)\lambda_\epsilon$ for some $0 < \alpha \leq 1$ and since (7) holds for every j (see Definition 7) we obtain that $\frac{\mu_\epsilon(P_s(\omega))}{\tilde{\mu}(P_s(\omega))} \to \frac{1}{\alpha}\ \mu - a.e.$, which in turn implies $\frac{\tilde{\mu}(P_{s+1}(\omega)|P_s(\omega))}{\mu_\epsilon(P_{s+1}(\omega)|P_s(\omega))} \to 1\ \mu - a.e.$.

Combining this with $1-\epsilon < \liminf \frac{\mu_\epsilon(P_{s+1}|P_s)}{\mu(P_{s+1}|P_s)} \leq \limsup \frac{\mu_\epsilon(P_{s+1}|P_s)}{\mu(P_{s+1}|P_s)} < 1+\epsilon, \mu - a.e.$, we get that for an arbitrary $\epsilon > 0$ $\quad 1-\epsilon < \liminf \frac{\tilde{\mu}(P_{s+1}|P_s)}{\mu(P_{s+1}|P_s)} \leq \limsup \frac{\tilde{\mu}(P_{s+1}|P_s)}{\mu(P_{s+1}|P_s)} < 1+\epsilon\ \mu - a.e.$ Now conclude that $\tilde{\mu}$ is ϵ-asymptotic near μ for any $\epsilon > 0$, which by Lemma 2 gives the desired result. ■

The next theorem is the parametric version of Theorem 2 and its proof closely follows that of Theorem 2.

Theorem 3. If Θ is an accommodating set of parameters then $\forall\theta_0 \in \Theta$, μ_Θ weakly merges to μ_{θ_0}.

Proof: Take $\theta_0 \in \Theta$. For any $\epsilon > 0$ we look at the neighborhood, $C(\theta_0, \epsilon)$, and at $C(\theta_0, \epsilon)^c = \cup_{j=1}^{K} \Theta_j$ (see definition 7(b)). It is straightforward that $\mu_{C(\theta_0,\epsilon)}$ is a grain of μ_Θ and that, furthermore, $\mu_{C(\theta_0,\epsilon)}$ is ϵ-asymptotically near μ_{θ_0}.

Using similar arguments as in the proof of Theorem 2 one can show that for any $j = 1, 2, \ldots, M$ $\frac{\mu_{\Theta_j}\left((P_s(\omega)\right)}{\mu_\Theta\left(P_s(\omega)\right)} \to 1$ $\mu_{\theta_0} - a.e.$ With this in hand it can be concluded that (again, by a mere repetition of the final argument in the proof of Theorem 2) that $\frac{\mu_\Theta\left(P_{s+1}|P_s\right)}{\mu_{\theta_0}\left(P_{s+1}|P_s\right)} \to 1$ $\mu_{\theta_0} - a.e.$

It is concluded that μ_Θ is ϵ-asymptotically near μ_{θ_0} for any $\epsilon > 0$. Now apply Lemma 2 for the desired result. ■

We now turn to treat the even weaker notion of merging, the Almost Weak Merging. The next theorem presents sufficient conditions for AWM to occur. We shall later see (Theorems 6 and 7) that these conditions are very close to necessary ones.

Although, mathematically, the key result regarding AWM is stated in the next theorem, we believe that the more important results, from a "practical" point of view (when implementing AWM in economic/game-theoretic situations), are Corollaries 4 and 5 which follow.

Theorem 4. If $\liminf \left(\frac{\tilde{\mu}\left(P_n(w)\right)}{\mu\left(P_n(w)\right)}\right)^{1/n} \geq 1$ μ-almost surely then $\tilde{\mu} \overset{\text{AWM}}{\longrightarrow} \mu$.

Before giving the proof we emphasize that a similar proof to the proof of Theorem 1 can achieve the result. The proof given here is different and will be useful for obtaining some more results, such as necessary conditions for AWM.

The proof of theorem 4 needs the following lemma, stated without the proof (the proof appears in Lehrer & Smorodinsky (1993)).

Lemma 4. Let a_i, b_i $i = 1, 2, \ldots$ be non-negative numbers. $\sum a_i = 1$.

i) Given $\eta > 0$ $\exists \varphi = \varphi(\eta) > 0$ s.t. if $\sum_i b_i \leq 1 + \varphi$ then $\sum a_i \log \frac{b_i}{a_i} \leq \eta$.

ii) Given $\epsilon > 0$ $\exists \varphi = \varphi(\epsilon) > 0$ and $\exists \delta(\epsilon) = \delta > 0$ such that if $\sum_i b_i \leq 1 + \varphi$ and $\sum |a_i - b_i| > \epsilon$ then $\sum a_i \log \frac{b_i}{a_i} < -\delta$.

Proof of Theorem 4: Suppose $\liminf \left(\frac{\widetilde{\mu}\left(P_n(w)\right)}{\mu\left(P_n(w)\right)}\right)^{1/n} \geq 1$ $\mu - a.s.$ and that $\widetilde{\mu}$ does not AWM to μ, i.e., there exist $d > 0$ and $B \subset \Omega$ such that $\mu(B) > 0$ and $\forall w \in B$ $\exists \mathbb{N}(w) \subset \mathbb{N}$ with upper density $d(w)$ greater than d, satisfying that $\forall n \in \mathbb{N}(w)$ $\exists A_n(w) \in \mathcal{F}_n$ such that

$$\left|\widetilde{\mu}\left(A_n(w) \mid P_{n-1}(w)\right) - \mu\left(A_n(w) \mid P_{n-1}(w)\right)\right| > \epsilon(w) > 0 \ .$$

Without loss of generality we may assume that $\epsilon(w) \geq \widehat{\epsilon} > 0$ $\forall w \in B$.

So:

$$\forall n \in \mathbb{N}(w) \sum_{P_n \in \mathcal{F}_n(w)} \Big|\widetilde{\mu}(P_n \mid P_{n-1}(w)) - \mu(P_n \mid P_{n-1}(w))\Big| > \widehat{\epsilon}.$$

We write $\mu_n(w) \equiv \mu(P_n(w) \mid P_{n-1}(w))$ and $\widetilde{\mu}_n(w) \equiv \widetilde{\mu}(P_n(w) \mid P_{n-1}(w))$.

Define the following random variables:

$$\text{(8)} \qquad \forall n \in N \quad X_n(w) = \log \frac{\tilde{\mu}_n(w)}{\mu_n(w)} \quad \text{and} \quad Y_n(w) = \log \frac{\lambda_n^{\epsilon_0}(w)}{\mu_n(w)} ,$$

where $\lambda_n^{\epsilon_0}(w) = \max\{\tilde{\mu}_n(w), \epsilon_0 \cdot \mu_n(w)\}$.

Note that

i) $X_n(w) \le Y_n(w) \quad \forall n, \forall w$.

ii) $\forall n \sum \lambda_n^{\epsilon_0}(w) \le \sum \tilde{\mu}_n(w) + \epsilon_0 \mu_n(w) \le 1 + \epsilon_0$. (The summations are over all atoms of $\mathcal{P}_n$.)

Taking $w \in B$ and $n \in \mathbb{N}(w)$:

$$\sum |\lambda_n^{\epsilon_0}(w) - \mu_n(w)| \ge \sum |\widetilde{\mu}_n(w) - \mu_n(w)| - \sum |\lambda_n^{\epsilon_0}(w) - \tilde{\mu}_n(w)| \ge \widehat{\epsilon} - \epsilon_0 .$$

And so taking $\epsilon_0 \le \widehat{\frac{\epsilon}{2}}$ yields

$$\text{(9)} \qquad \sum |\lambda_n^{\epsilon_0}(w) - \mu_n(w)| \ge \frac{\widehat{\epsilon}}{2}.$$

By the second part of Lemma 4 and (9) we may take ϵ_0 small enough such that for some positive $\delta = \delta\Big(\widehat{\frac{\epsilon}{2}}\Big)$

$$\text{(10)} \qquad E\Big(Y_n(w) \mid P_{n-1}(w)\Big) = \sum \mu_n(w) \log \frac{\lambda_n^{\epsilon_0}(w)}{\mu_n(w)} < -\delta \quad \forall n \in \mathbb{N}(w) .$$

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

$$-\delta \cdot \frac{d}{2} + \Big(1 - \frac{d}{2}\Big) \cdot \alpha \le -\beta < 0 .$$

By the first part of Lemma 4 take ϵ_0 such that

$$\text{(11)} \qquad E(Y_n(w) \mid P_{n-1}(w)) = \sum \mu_n(w) \log \frac{\lambda_n^{\epsilon_0}(w)}{\mu_n(w)} \le \alpha \quad \forall n \in \mathbb{N} .$$

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

$$E(Y_n^2(w) \mid P_{n-1}(w)) = \sum \mu_n(w) \cdot \log^2 \frac{\lambda_n^{\epsilon_0}(w)}{\mu_n(w)} =$$

$$= \sum_{\frac{\tilde{\mu}}{\mu} < \epsilon_0} \mu_n(w) \cdot \log^2 \frac{\epsilon_0 \cdot \mu_n(w)}{\mu_n(w)} + \sum_{\epsilon_0 \le \frac{\tilde{\mu}}{\mu} \le 1} \mu_n(w) \cdot \log^2 \frac{\tilde{\mu}_n(w)}{\mu_n(w)}$$

$$+ \sum_{\frac{\tilde{\mu}}{\mu} > 1} \mu_n(w) \log^2 \frac{\tilde{\mu}_n(w)}{\mu_n(w)} \le$$

$$\leq \log^2 \epsilon_0 + \sum_{\epsilon_0 < \frac{\tilde{\mu}}{\mu} < 1} \mu_n(w) \cdot \log^2 \epsilon_0 + \sum_{\frac{\tilde{\mu}}{\mu} > 1} \mu_n(w) \cdot \left(\sqrt{\frac{\tilde{\mu}_n(w)}{\mu_n(w)}} \right)^2$$

$$\leq \log^2 \epsilon_0 + \log^2 \epsilon_0 + 1 \ .$$

So the strong law of large numbers may be applied to the uncorrelated random variables (see Shiryayev, 1984) $Y_n(w) - E\big(Y_n(w) \mid P_{n-1}(w)\big)$.

For μ -a.e. $w \in \Omega$ exists $m(w)$ such that for every $n \geq m(w)$,

$$\text{(12)} \qquad \frac{1}{n} \sum_{j=1}^{n} Y_j(w) \leq \frac{1}{n} \cdot \sum_{j=1}^{n} E\big(Y_j(w) \mid P_{j-1}(w)\big) + \frac{\beta}{2} \ .$$

Take $w \in B$ and an infinite sequence $\mathbb{N}_1(w) \subset \mathbb{N}$ such that $n \in \mathbb{N}_1(w)$ implies $n \geq m(w)$ and $\frac{\#\{k | k \leq n \text{ and } k \in \mathbb{N}(w)\}}{n} \geq \frac{d}{2}$. So by (10), (11) and (12) for $n \in \mathbb{N}_1(w)$

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

$$\leq -\beta + \frac{\beta}{2} = -\frac{\beta}{2}$$

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

which implies

$$\left(\frac{\tilde{\mu}\big(P_n(w)\big)}{\mu\big(P_n(w)\big)} \right)^{1/n} \leq e^{-\frac{\beta}{2}} < 1 \quad \forall n \in \mathbb{N}_1(w)$$

thus contradicting the assumption over μ and $\tilde{\mu}$. ■

Remark 7. A careful reading of the proof shows that we have actually shown more than required. Notice that Y_n, defined in (8), depends on ϵ_0. Define for every ϵ_0

$$\tilde{\varphi}_n^{\epsilon_0}\big(P_n(w)\big) = \prod_{k=1}^{n} \lambda_k^{\epsilon_0}(w) \ .$$

We have actually shown that if $\liminf \Big(\tilde{\varphi}_n^{\epsilon_0}\big(P_n(w)\big) \Big/ \mu\big(P_n(w)\big) \Big)^{1/n} \geq 1$ $\mu - a.s$ for every $\epsilon_0 > 0$, then $\tilde{\mu} \overset{\text{AWM}}{\longrightarrow} \mu$.

Using this remark the condition in Theorem 4 can be slightly relaxed:

Corollary 3. If μ-a.e $\omega \in \Omega$ there exists a full sequence of indices $N(\omega) \subset \mathbb{N}$ such that $\liminf_{n \in N(\omega)} \left(\frac{\tilde{\mu}\left(P_n(w)\right)}{\mu\left(P_n(w)\right)} \right)^{1/n} \geq 1$ then $\tilde{\mu} \overset{\text{AWM}}{\longrightarrow} \mu$.

Proof: Fix $\delta > 0$. $\exists M = M(w,\delta) \in N(w)$ s.t. if $n \geq M$ and $n \in N(w)$ then $\left(\frac{\tilde{\mu}(P_n(w))}{\mu(P_n(w))}\right)^{\frac{1}{n}} > 1-\delta$ and $\frac{|N(w)\cap\{1,\ldots,n\}|}{n} > 1-\delta$

$$\text{(13)} \quad \forall k \in N(w),\ k \geq m: \quad \left(\frac{\hat{\varphi}_n^{\epsilon}(P_n(w))}{\mu(P_n(w))}\right)^{\frac{1}{n}} \geq \left(\frac{\tilde{\mu}(P_n(w))}{\mu(P_n(w))}\right)^{\frac{1}{n}} \geq 1-\delta\,.$$

$$\text{(14)} \quad \begin{array}{l} \forall k \notin N(w), k \geq m: \exists M \leq k(w) < k \quad \text{such that} \\ \qquad k(w) \in N(w),\ \{k(w)+1,\ldots,k\} \cap N(w) = \emptyset \end{array}$$

i.e. $k(w)$ is the last number which is in $N(w)$ and is less than k. It is easy to verify that $\lim_{k\to\infty} \frac{k(w)}{k} = 1$ μ-a.e.

$$\left(\frac{\hat{\varphi}_k^{\epsilon}(P_k(w))}{\mu(P_k(w))}\right)^{\frac{1}{k}} = \left(\frac{\hat{\varphi}_{k(w)}^{\epsilon}(P_{k(w)}(w))}{\mu(P_{k(w)}(w))}\right)^{\frac{1}{k(w)}\cdot\frac{k(w)}{k}} \cdot \left(\frac{\prod_{j=k(w)+1}^{k} \lambda_j^{\epsilon_j}(w)}{\prod_{j=k(w)+1}^{k} \mu_j(w)}\right)^{\frac{1}{k}} \geq$$
$$(1-\delta)^{\frac{k(w)}{k}} \cdot \left(\epsilon_0^{k-k(w)}\right)^{\frac{1}{k}} \underset{\substack{k\to 0 \\ \frac{k(w)}{k}\to 1}}{\longrightarrow} (1-\delta)\cdot 1$$

As δ is arbitrary the result follows from (13) and (14). ■

It will be shown the the condition stated in the above corollary is not equivalent to AWM. In Example 10 AWM is satisfied, while the condition from Corollary 3 is violated.

Corollary 4. If $\tilde{\mu}$ is diffused around μ then $\tilde{\mu} \overset{\text{AWM}}{\longrightarrow} \mu$.

Proof: Apply Lemma 3 to get the condition of Theorem 4. ■

Corollary 5. If Θ is a diffused set of parameters then $\forall \theta_0 \in \Theta$, μ_Θ AWM to μ_{θ_0}.

Corollary 6. Let $\mu, \tilde{\mu}, \widehat{\mu}$ be three probability measures satisfying:
 i) $\widehat{\mu}$ weakly merges to μ.
 ii) $\widehat{\mu}$ is a grain of $\tilde{\mu}$, i.e. $\tilde{\mu} = \alpha\cdot\widehat{\mu} + (1-\alpha)\lambda$, where $0 < \alpha \leq 1$ and λ is a probability measure.

Then $\widetilde{\mu} \overset{\text{AWM}}{\longrightarrow} \mu$.

Proof: Fix $\epsilon_0 > 0$ and recall the notation $\widetilde{\varphi}_n^{\epsilon_0}(P_n(w))$ in Remark 3. We similarly define $\widehat{\varphi}_n^{\epsilon_0}(P_n(w))$ corresponding to $\widehat{\mu}$. It is clear that since $\widehat{\mu}$ merges to μ for $\mu - a.e\ w$ and $\forall\epsilon > 0\ \ \exists N$ s.t. $\forall n > N$

$$\frac{\widehat{\varphi}_n^{\epsilon_0}(P_n(w))}{\mu(P_n(w))} \geq (1-\epsilon)\cdot\frac{\widehat{\varphi}_{n-1}^{\epsilon_0}(P_{n-1}(w))}{\mu(P_{n-1}(w))}\,.$$

So
$$\frac{\widehat{\varphi}_n^{\epsilon_0}(P_n(w))}{\mu(P_n(w))} \geq (1-\epsilon)^{n-N} \cdot \frac{\widehat{\varphi}_N^{\epsilon_0}(P_N(w))}{\mu(P_N(w))} .$$

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

$$\frac{\tilde{\varphi}_n^{\epsilon_0}(P_n(w))}{\mu(P_n(w))} = \frac{\tilde{\varphi}_n^{\epsilon_0}(P_n(w))}{\widehat{\varphi}_n^{\epsilon_0}(P_n(w))} \cdot \frac{\widehat{\varphi}_n^{\epsilon_0}(P_n(w))}{\mu(P_n(w))} \geq \alpha \cdot (1-\epsilon)^{n-N} \frac{\widehat{\varphi}_N^{\epsilon_0}(P_N(w))}{\mu(P_N(w))} .$$

It follows that
$$\liminf \left(\frac{\tilde{\varphi}_n^{\epsilon_0}(P_n(w))}{\mu(P_n(w))} \right)^{1/n} \geq 1 - \epsilon .$$

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

Another result which is achieved with the technique of the previous proof is the following:

Remark 8. For any two probability measures μ and $\tilde{\mu}$, we have μ-almost surely $\limsup \left(\frac{\tilde{\mu}(P_n(w))}{\mu(P_n(w))} \right)^{1/n} \leq 1$. In other words, the hypothesis of Theorem 4 is actually $\left(\frac{\tilde{\mu}(P_n(w))}{\mu(P_n(w))} \right)^{1/n} \underset{n\to\infty}{\longrightarrow} 1$ μ-almost surely.

7. Necessary Conditions for Merging. The previous section dealt with sufficient conditions for all kinds of merging. It related the different notions of compatibility to those of merging. These conditions are certainly not necessary. We refer the reader to Section 8 for examples which clarify this claim.

In order to form necessary conditions for merging we must first define properly the environments in which the belief, $\tilde{\mu}$, can be properly updated using Bayes rule. Therefore we need the following definitions:

Definition 11(a). μ is *locally absolutely continuous* (abbreviated LAC) w.r.t $\tilde{\mu}$ if for any $n \in \mathbb{N}$ and any $S \in \mathcal{F}_n$ $\mu(S) > 0$ implies $\tilde{\mu}(S) > 0$.

Definition 11(b). The set of parameters, Θ, is *locally absolutely continuous within* (abbreviated LAC within) if for any $\theta_0 \in \Theta$ μ_{θ_0} is LAC w.r.t μ_Θ.

Throughout Section 6 we shall assume that the the truth is LAC w.r.t the belief and that the set of parameters is LAC within.

The next theorem, stated without a proof, is the converse of Theorem 1 under local absolute continuity (for the proof see Kalai and Lehrer(1994).)

Theorem 5. If μ is LAC w.r.t $\tilde{\mu}$ and $\tilde{\mu}$ merges to μ then μ is absolutely continuous w.r.t $\tilde{\mu}$.

An example can be provided (see Example 10) to show that the converse of Theorem 4 is generally incorrect. We therefore, turn to specify necessary conditions for AWM. We will use the notations $\lambda_n^{\epsilon_0}(w)$ (introduced in the proof of Theorem 4) and $\tilde{\varphi}_n^{\epsilon_0}\big(P_n(w)\big)$ (see Remark 7).

Theorem 6. Suppose μ is LAC w.r.t $\tilde{\mu}$ and $\tilde{\mu}$ AWM to μ. Then for every $\epsilon_0 > 0$:

$$\liminf \left(\frac{\tilde{\varphi}_n^{\epsilon_0}\big(P_n(w)\big)}{\mu\big(P_n(w)\big)} \right)^{1/n} \geq 1 \quad \mu - a.s\ .$$

Proof: We use the proof of Theorem 4. Fix $\epsilon_0 > 0$. Using the random variables Y_n defined with $\lambda_n^{\epsilon_0}(w)$ in (8), one may get, similar to (12) that

$$\frac{1}{n}\sum_{j=1}^{n} E\big(Y_j(w) \mid P_{j-1}(w)\big) \leq \frac{1}{n}\sum_{j=1}^{n} Y_j(w) + \delta \tag{15}$$

$\mu - a.s$ whenever $n > n(w,\delta)$. The assumption of the proposition implies that the left side of (15) converges to 0 and therefore $0 \leq \liminf \frac{1}{n}\sum_{j=1}^{n} Y_j(w)$. Thus, $1 \leq \liminf \left(\frac{\tilde{\varphi}_n^{\epsilon_0}\left(P_n(w)\right)}{\mu\left(P_n(w)\right)} \right)^{1/n}$. ∎

Now one can summarize this last theorem and Remark 5 to get a characterization of AWM:

Theorem 7. $\tilde{\mu} \overset{\text{AWM}}{\longrightarrow} \mu$ if and only if $\mu - a.e$ and for every $\epsilon_0 > 0$

$$\liminf \left(\frac{\tilde{\varphi}_n^{\epsilon_0}\big(P_n(w)\big)}{\mu\big(P_n(w)\big)} \right)^{1/n} \geq 1 \quad .$$

A necessary condition which is slightly more elegant is stated in the following corollary:

Corollary 7. Suppose μ is LAC w.r.t $\tilde{\mu}$ and $\tilde{\mu}$ AWM to μ. If there is a random variable $c > 0$ s.t. $\liminf \tilde{\mu}\big(P_n(w) \mid P_{n-1}(w)\big)/\mu\big(P_n(w) \mid P_{n-1}(w)\big) > c \quad \mu - a.e$, then $\left(\frac{\tilde{\mu}\left(P_n(w)\right)}{\mu\left(P_n(w)\right)} \right)^{1/n} \underset{n\to\infty}{\longrightarrow} 1$ μ-almost surely.

8. Examples. This section is devoted to various examples which demonstrate the different theorems presented here. Other examples are provided to show that natural relaxations of the conditions stated in some of the theorems will yield conditions which are insufficient for maintaining the results.

In all of the following examples let $\Omega = \{0,1\}^{\mathbf{N}}$, let $\mathcal{P}_n = \{$all the cylinder sets determined by the first n coordinates of ω $\}$ and let $\mathcal{B}$ be the σ-field generated by $\cup_{n=1}^{\infty}\mathcal{P}_n$.

Example 5 - Merging. Suppose a stochastic process on $\{0,1\}$ is a determined by a "rational" coin, i.e. at each stage the probability for the next outcome to be 1 is a fixed number p, where p is rational. Suppose this is all our agent knows. By Corollary 2, no matter what the true p is, as long as it is rational, merging will occur. This is so because the parameter set $\{p \in [0,1];\ \ p \text{ is rational }\}$, together with any measure, which has full support, is a set of grains.

Example 6 - Weak Merging. This example is quite close to the previous one, yet in this example merging does not occur, but rather weak merging.

Suppose a stochastic process on $\{0,1\}$ is a determined by a coin, i.e. at each stage the probability for the next outcome to be 1 is a fixed number p, where $0 < p < 1$ is the outcome of a uniform distribution on $(0,1)$. Denote by $C(p,\epsilon)$ the interval $(p - p\epsilon, p + p\epsilon)$ and decompose the complementary of $C(p,\epsilon)$ as follows $C(p,\epsilon)^c = (0, p - p\epsilon] \cup [p + p\epsilon, 1)$. It is easy to verify that $\mu_{C(p,\epsilon)}$ is ϵ-asymptotically near μ_p and that both $(0, p - p\epsilon]$ and $[p + p\epsilon, 1)$ have the separation property w.r.t. μ_p. Conclude that $(0,1)$, along with the uniform distribution on it, is an accommodating set of parameters. By Theorem 3 weak merging occurs. independently of p the agent, after sufficiently many periods, will be able to forecast the next outcome, or the next ℓ outcomes, almost as good as if he knew the true p.

Nevertheless, merging does not occur here. The conditions of Theorem 1 are not satisfied, and as local absolute continuity is satisfied, we may conclude by Theorem 5 that merging does not occur.

To see why the condition of Theorem 1 is not satisfied look at the following set:

$$B = \{\omega \in \{0,1\}^{\mathbf{N}} |\ \text{density of 1's is exactly p }\}$$

Note that $\mu_p(B) = 1$ and simultaneously $\mu_{(0,1)}(B) = 0$, i.e., the absolute continuity condition is violated. Also note that the event B is a tail event.

Example 7 - Almost weak merging. Let Ω be the space $\{0,1\}^{\mathbf{N}}$, and let $\mathcal{P}_n$ be the partition induced by the first n coordinates. Define μ to be the Dirac measure on the point $(1,1,\ldots)$. Define $\tilde{\mu}$ to be the measure $\frac{1}{2}\mu_1 + \frac{1}{2}\mu_2$ where μ_1 and μ_2 are defined as follows. μ_1 is the measure induced by a sequence $X_1, X_2, \ldots$ of independent Bernoulli random variable, where $\text{prob}(X_n = 1)$ is 1 if $n \neq 2^{2^t}$ and it is $\frac{1}{2}$ if $n = 2^{2^t}$. The measure μ_2 is the one induced by the following. Denote by ν_n the measure induced by the i.i.d. sequence $X_1, X_2, \ldots$ of random variables, where prob $(X_1 = 1) = 1 - \frac{1}{n} = 1 -$ prob $(X_1 = 0)$. Set $\mu_2 = \sum \frac{1}{2^n}\nu_n$. In other words, with probability $\frac{1}{2^n}$ $(n = 1, 2, \ldots)$ it is defined by a repeated toss of a coin assigning probability $1 - \frac{1}{n}$ 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 $\frac{1}{2}$ to the event that the next outcome will be 1 while the updated measure of μ assigns the same event the measure 1. Thus, $\tilde{\mu}$ does not weakly merge to μ, which in turn yields that $\tilde{\mu}$ does not merge to μ. But as $\tilde{\mu}$ is diffused around μ it is concluded by Corollary 3 that $\tilde{\mu}$ almost weakly merges to μ.

Looking at the definitions of "accommodating" (Definition 7(a)) and "diffused" (Definition 5(a)) a natural question is whether one can relax the condition on the residual measure λ_ϵ and still maintain the property of weak (as opposed to almost weak) merging. We now show a pair of examples in which the conditions on the residual measure λ_ϵ are indeed relaxed, yet in the first example weak merging occurs, while in the second only almost weak merging can be obtained.

Example 8 - Relaxing the definition of accommodation without violating the weak merging property. Modify the definition of $\tilde{\mu}$ from the previous example, such that for any n of the form $n = 2^t$ (instead of $n = 2^{2^t}$) prob $(X_n = 1)$ is $\frac{1}{2}$ and is 1 otherwise. For any $\epsilon > 0$ let $\mu_\epsilon = \mu_2$ and $\lambda_\epsilon = \mu_1$. Obviously the set of indices for which the conditional of λ_ϵ is not equal to the conditional of μ has density zero, yet λ_ϵ is eventually assumed away and WM actually occurs.

In the following example we have a measure $\tilde{\mu}$ which almost weakly merges to another measure μ but does not weakly merge. In this example $\tilde{\mu}$ is diffused around μ (thus, AWM is obtained) and furthermore $\tilde{\mu}$ can be decomposed as follows:

(i) $\forall \epsilon > 0 \quad \tilde{\mu} = \frac{1}{2}\mu_\epsilon + \frac{1}{2}\lambda_\epsilon$

(ii) $\forall t \in \mathbb{N}\ |\mu_\epsilon(P_n|P_{n-1}) - \mu(P_n|P_{n-1})| < |\lambda_\epsilon(P_n|P_{n-1}) - \mu(P_n|P_{n-1})|$

(iii) On a sequence of periods with upper density 1 and lower density zero the following occurs:

$$|\mu_\epsilon(P_n|P_{n-1}) - \mu(P_n|P_{n-1})| + \frac{1}{8} < |\lambda_\epsilon(P_n|P_{n-1}) - \mu(P_n|P_{n-1})|$$

i.e. the differences are bounded away.

For any measure $\alpha(\cdot)$ we denote $\alpha_n = \alpha_n(\omega) = \alpha(P_n(\omega)|P_{n-1}(\omega))$.

Example 9 - Relaxing the definition of Accommodation and violating the weak merging property. Let $\Omega = \{\text{"}H\text{"}, \text{"}T\text{"}\}^{\mathbf{N}}$ and let $\mathcal{B}$ be the Borel σ-field on Ω. Let μ be the measure generated by tosses of a fair coin. Let $P_n(\omega)$ be the first n outcomes of the coin.

Defining $\tilde{\mu}$. In order to define $\tilde{\mu}$ we begin with a decreasing sequence $\{\epsilon_n\}_{n=1}^\infty$ $0 < \epsilon_n \to 0$, which will be determined later. Using it, we define the following sequence of triples of random variables. For $n = 1$:

* $T_1(\omega)$ is the first time where the difference between the number of Heads and the number of Tails is greater than 1.
* $S_1(\omega) = 1 \cdot T_1(\omega)$
* $K_1(\omega) = \min\left\{ k \middle| \left(\frac{\frac{1}{2}+\epsilon_1}{\frac{1}{2}-\epsilon_1}\right)^k \cdot \left(\frac{\frac{1}{2}}{\frac{1}{2}+\epsilon_1}\right)^{S_1(\omega)} \geq 2 \cdot 1 \right\}$.

We continue the definitions inductively:

* $T_n(\omega)$ is the first time where the difference between the number of Heads and the number of Tails since $S_{n-1}(\omega)$ is greater than $K_{n-1}(\omega)$.

* $S_n(\omega) = n \cdot T_n(\omega)$
* $K_n(\omega) = \min\left\{k \,\middle|\, \left(\frac{\frac{1}{2}+\epsilon_k}{\frac{1}{2}-\epsilon_k}\right)^k \cdot \left(\frac{\frac{1}{4}}{\frac{1}{2}+\epsilon_k}\right)^{S_n(\omega)} \geq 2 \cdot n\right\}$.

Remark a) T_n is a stopping time and thus S_n and K_n, are $\mathcal{F}_{T_n}$ measurable.

Remark b) Since $T_n(\omega) < \infty$ $\forall n$ $\mu - a.e.$, $S_n(\omega) < \infty$ and $K_n(\omega) < \infty$ $\forall n$ $\mu - a.e.$. Having defined the sequence $\{(T_n, S_n, K_n)\}_{n=1}^{\infty}$ we proceed by defining $\tilde{\mu} = \frac{1}{2}\nu + \frac{1}{2}\lambda$ where ν and λ are defined as follows. (ν will serve as μ_ϵ $\forall \epsilon > 0$ and λ will be λ_ϵ $\forall \epsilon > 0$).

Defining ν. ν is derived from a sequence of Bernoulli random variables:

$$\forall t \leq T_1(\omega) \quad Prob(\text{“}H\text{”}) = \frac{1}{2} - \epsilon_0 = 1 - Prob(\text{“}T\text{”})$$

$$\forall T_{n-1}(\omega) < t \leq T_n(\omega) \quad Prob(\text{“}H\text{”}) = \frac{1}{2} - \epsilon_{n-1} = 1 - Prob(\text{“}T\text{”})$$

Remark c) As $T_n < \infty$ $\mu - a.e.$ and $\epsilon_n \to 0$, $lim\left|\frac{\nu_n}{\mu_n} - 1\right| = 0 \quad \mu - a.e.$

Since ν is equal to μ_ϵ $\forall \epsilon > 0$ (see Definition 5(a)) one obtains that $\tilde{\mu}$ is diffused around μ and by Corollary 4 $\tilde{\mu} \overset{AWM}{\longrightarrow} \mu$.

Defining λ We first need to define, $\forall n \in \mathbb{N}$, a decreasing sequence of positive number $\{\eta_n(i)\}_{i=1}^{\infty}$ s.t. $\prod_{i=1}^{\infty} \frac{\frac{1}{2}-\epsilon_n-\eta_n(i)}{\frac{1}{2}+\epsilon_n} > \frac{1}{2}$. Having these sequences, we define λ by the following sequence of Bernoulli random variables:

$$\forall t \leq T_1 \quad Prob(\text{“}H\text{”}) = \frac{1}{2} + \epsilon_0 + \eta_0(t) = 1 - Prob(\text{“}T\text{”})$$

$$\forall T_1 < t \leq S_1 \quad Prob(\text{“}H\text{”}) = \frac{3}{4} = 1 - Prob(\text{“}T\text{”})$$

$$\forall S_{n-1} < t \leq T_n \quad Prob(\text{“}H\text{”}) = \frac{1}{2} + \epsilon_{n-1} + \eta_{n-1}(t) = 1 - Prob(\text{“}T\text{”})$$

$$\forall T_n < t \leq S_n \quad Prob(\text{“}H\text{”}) = \frac{3}{4} = 1 - Prob(\text{”}T\text{”}) .$$

Remark d) For every time period t

$$|\nu(\cdot|P_t) - \mu(\cdot|P_t)| < |\lambda(\cdot|P_t) - \mu(\cdot|P_t)| .$$

Remark e) For any time t satisfying $T_n < t \leq S_n$

(*i*) $$|\lambda_n - \mu_n| = |\frac{1}{2} - \frac{1}{4}| = \frac{1}{4}$$

(*ii*) $$|\nu_n - \mu_n| = \epsilon_n \longrightarrow 0 .$$

Remark f) The set of periods satisfying Remark e) has upper density 1. (Because $S_n = n \cdot T_n$). Finally we show that in spite of Remarks c,d and e there is no WM (although AWM occurs – Remark b). We show that infinitely often $|\tilde{\mu}_n - \mu_n| = \frac{1}{4}\ \mu - a.e.$

Claim: $\tilde{\mu}$ does not WM to μ.

Proof: Look at the infinite sequence of periods $\{T_n(\omega)\}_{n=1}^{\infty}$. We shall show that $\frac{\lambda(P_{T_n})}{\nu(P_{T_n})} \underset{n\to\infty}{\longrightarrow} \infty$. I.e., λ is much more likely to produce the atom P_{T_n} then ν and thus a-posteriori $\tilde{\mu}$ gives substantially more weight to λ than to ν on the times $\{T_n\}_{n=1}^{\infty}$. Thus, $|\tilde{\mu}_{T_n} - \lambda_{T_n}| \to 0$, yielding $lim|\mu_{T_n} - \tilde{\mu}_{T_n}| = lim|\mu_{T_n} - \lambda_{T_n}| = \frac{1}{4}\ \forall n$.

$$\frac{\lambda(P_{T_{n+1}}|P_{T_n})}{\nu(P_{T_{n+1}}|P_{T_n})} = \prod_{j=T_n+1}^{T_{n+1}} \frac{\lambda(P_{j+1}|P_j)}{\nu(P_{j+1}|P_j)} = \prod_{j=T_n+1}^{T_{n+1}} \frac{\lambda_j}{\nu_j} =$$

$$= \prod_{j=T_n+1}^{S_n} \frac{\lambda_j}{\nu_j} \cdot \prod_{j=S_n+1}^{T_{n+1}} \frac{\lambda_j}{\nu_j} \geq \left(\frac{\frac{1}{4}}{\frac{1}{2}+\epsilon_n}\right)^{S_n-(T_n+1)} \cdot \prod_{j=S_n+1}^{T_{n+1}} \frac{\lambda_j}{\nu_j} \geq$$

$$\geq \left(\frac{\frac{1}{4}}{\frac{1}{2}+\epsilon_n}\right)^{S_n} \cdot \prod_{j=S_n+1}^{T_{n+1}} \frac{\lambda_j}{\nu_j} = (*)$$

Define $A = A(\omega) =$ the set of times at which "H" occurred and $B = B(\omega) =$ the set of times at which "T" occurred. We obtain,

$$(*) = \left(\frac{\frac{1}{4}}{\frac{1}{2}+\epsilon_n}\right)^{S_n} \prod_{\substack{S_n<j\leq T_{n+1}\\ j\in A}} \frac{\lambda_j}{\nu_j} \cdot \prod_{\substack{S_n<j\leq T_{n+1}\\ j\in B}} \frac{\lambda_j}{\nu_j} =$$

$$= \left(\frac{\frac{1}{4}}{\frac{1}{2}+\epsilon_n}\right)^{S_n} \prod_{\substack{S_n<j\leq T_{n+1}\\ j\in A}} \left(\frac{\frac{1}{2}+\epsilon_n+\eta_n(j)}{\frac{1}{2}-\epsilon_n}\right) \cdot \prod_{\substack{S_n<j\leq T_{n+1}\\ j\in B}} \left(\frac{\frac{1}{2}-\epsilon_n-\eta_n(j)}{\frac{1}{2}+\epsilon_n}\right)$$

$$= (**)$$

Recall that between S_n and T_{n+1} the number of outcomes of "H" exceeds that of "T" by K_n. So:

$$(**) \geq \left(\frac{\frac{1}{4}}{\frac{1}{2}+\epsilon_n}\right)^{S_n} \cdot \left(\frac{\frac{1}{2}+\epsilon_n}{\frac{1}{2}-\epsilon_n}\right)^{K_n} \cdot \prod_{S_n<j\leq T_{n+1}} \frac{\frac{1}{2}-\epsilon_n-\eta_n(j)}{\frac{1}{2}+\epsilon_n} = (***)$$

Using the definitions of $\{\eta_n(j)\}$ and of K_n:

$$(***) \geq 2n \cdot \frac{1}{2} = n\ .$$

And so

$$\frac{\lambda(P_{T_{n+1}})}{\nu(P_{T_{n+1}})} = \frac{\lambda(P_{T_n})}{\nu(P_{T_n})} \cdot \frac{\lambda(P_{T_{n+1}}|P_{T_n})}{\nu(P_{T_{n+1}}|P_{T_n})} \geq \frac{\lambda(P_{T_n})}{\nu(P_{T_n})} \cdot n \underset{n\to\infty}{\longrightarrow} \infty.$$

■

In the following example we show that the converse of Theorem 4 is generally incorrect. In particular we show that $\tilde{\mu} \overset{WM}{\longrightarrow} \mu$, and therefore $\tilde{\mu} \overset{AWM}{\longrightarrow} \mu$ and yet

$$\sqrt[n]{\frac{\tilde{\mu}(P_n(w))}{\mu(P_n(w))}} \longrightarrow 0 \quad \mu - a.e.$$

Example 10. As usual let $\Omega = \{0,1\}^{\mathbf{N}}$. Take a sequence $\{d_n\}_{n=1}^{\infty}$ such that $d_n \to 0$ and $\sum_{n=1}^{\infty} d_n = \infty$. Let μ be the measure defined by an independent sequence of tosses of coins, where at stage n the coin has probability d_n to be 1 and $1-d_n$ to be 0. For any $\omega \in \Omega$ let $N(\omega)$ be the set of indices for which $w_n = 1$. As $\sum_{n=1}^{\infty} d_n = \infty$ we deduce by the Borel-Cantelli Lemma that μ-a.e. the set $N(\omega)$ is an infinite set of indices. For any ω let $\{n_j(\omega)\}_{j=1}^{\infty}$ be the set $N(\omega)$. Now take k_j such that the set

$$B_j = \{\omega | n_{j+1}(\omega) - n_j(\omega) < k_j\}$$

satisfies $\mu(B_j) > 1 - \frac{1}{2}^j$.

Now define a sequence $\{c_n\}_{n=1}^{\infty}$ with the following properties:

i) $\forall n \ \frac{1-c_n}{1-d_n} < 2$

ii) $\left\{ \left(\frac{c_n}{d_n} \right)^{\frac{1}{n+k_n}} \right\}_{n=1}^{\infty}$ is a decreasing sequence which converges to zero.

We define $\tilde{\mu}$ in a similar way to μ. At stage n a coin with the probability c_n of getting the outcome 1 is used. As $c_n \to 0$ and $d_n \to 0$ it is clear that $\tilde{\mu} \overset{WM}{\longrightarrow} \mu$, and therefore $\tilde{\mu} \overset{AWM}{\longrightarrow} \mu$. On the other hand, we shall show that

$$\sqrt[n]{\frac{\tilde{\mu}(P_n(w))}{\mu(P_n(w))}} \longrightarrow 0 \quad \mu - a.e.$$

By the construction of B_j^c we deduce that μ-a.e $\exists j_0(w)$ s.t. $w \in \bigcap_{j \geq j_0(w)} B_j$.

Denote $j_0 = j_0(w)$ and $n_{j_0} = n_{j_0(w)}(w)$ and take $m > n_{j_0}$. There exists $j(m) \geq j_0$ which satisfies $n_{j(m)} \leq m < n_{j(m)+1}$. So:

$$\left(\frac{\tilde{\mu}(P_m(w))}{\mu(P_m(w))}\right)^{\frac{1}{m}} \leq \left[\prod_{j=1}^{n_{j(m)}-1} \left(\frac{1-c_j}{1-d_j}\right) \cdot \frac{c_{n_{j(m)}}}{d_{n_{j(m)}}} \cdot \prod_{j=n_{j(m)}+1}^{m} \left(\frac{1-c_j}{1-d_j}\right) \right]^{\frac{1}{m}} \leq$$

$$\leq \left(\frac{c_{n_{j(m)}}}{d_{n_{j(m)}}} \cdot 2^m\right)^{\frac{1}{m}} \leq \left(\frac{c_{n_{j(m)}}}{d_{n_{j(m)}}}\right)^{\frac{1}{m}} \cdot 2 \leq$$

$$\leq \left(\frac{c_{n_{j(m)}}}{d_{n_{j(m)}}}\right)^{\frac{1}{n_{j(m)+1}}} \cdot 2 \leq 2 \cdot \left(\frac{c_{n_{j(m)}}}{d_{n_{j(m)}}}\right)^{\frac{1}{n_{j(m)}+k_{n_{j(m)}}}} \underset{m\to\infty}{\longrightarrow} 0\,.$$

REFERENCES

BLACKWELL, D. AND DUBINS, L. (1962). Merging of Opinions with Increasing Information. *Ann. Math. Statist.* **38** 882-886.

BLUME, L. AND EASLEY, D. (1992). Rational Expectations and Rational Learning. Notes from a talk given at the Economic Theory Workshop in Honor of Roy Radner, Cornell University.

JORDAN, J. S. (1991). Bayesian Learning in Normal Form Games. *Games and Economic Behavior* **3** 60-81.

KALAI, E. AND LEHRER, E. (1993). Rational Learning Lead to Nash Equilibrium. *Econometrica* **61** 1019-1045.

KALAI, E. AND LEHRER, E. (1994). Weak and Strong Merging of Opinions. *Journal of Mathematical Economics* **23** 73-86.

LEHRER, E. AND SMORODINSKY, R. (1993). Compatible Measures and Merging. To appear in *Mathematics of Operations Research.*

LEHRER, E. AND SMORODINSKY, R. (1994). Repeated Large Games with Incomplete Information. To appear in *Games and Economic Behavior.*

EHUD LEHRER
RAYMOND AND BEVERLY SACKLER FACULTY OF EXACT SCIENCES
SCHOOL OF MATHEMATICAL SCIENCES, TEL AVIV UNIVERSITY
RAMAT AVIV 69978, ISRAEL
lehrer@math.tau.ac.il

RANN SMORODINSKY
DEPARTMENT OF MANAGERIAL ECONOMICS AND DECISION SCIENCES
KELLOGG GRADUATE SCHOOL OF MANAGEMENT
NORTHWESTERN UNIVERSITY, EVANSTON, IL 60208, USA
rsm722@lulu.acns.nwu.edu