ALMOST EVERYBODY DISAGREES ALMOST ALL THE TIME:

The Genericity of Weakly-Merging Nowhere

August 1997 (Minor Revisions: September 1997)

Ronald I. Miller
Chris William Sanchirico

PRELIMINARY DRAFT OF LONG VERSION
Please do not cite without authors' permission.

Department of Economics
Columbia University
420 West 118th Street, 10th Floor
New York, New York 10027
rm170@columbia.edu
cs282@columbia.edu

Suppose we randomly pull two agents from a population and ask them to observe an unfolding, infinite sequence of zeros and ones. If each agent starts with a prior belief about the true sequence and updates this belief on revelation of successive observations, what is the chance that the two agents will come to agree on the likelihood that the next draw is a one? In this paper we show that there is no chance. More formally, we show that under a very unrestrictive definition of what it means to draw priors "randomly," the probability that two priors have any chance of *weakly merging* is zero. Indeed, almost surely, the two measures will be *singular*—one prior will think certain to occur a set of sequences that the other thinks impossible, and vice versa.

Our result is meant as a critique of the "rational learning" literature, which seeks positive convergence results on infinite product spaces by augmenting the process of Bayesian updating with seeming regularity conditions, variously labeled "consistency" or "compatibility" assumptions. Our object is to investigate just how regular these assumption and results are when considered in the space of all possible prior distributions. Our results on the genericity of nowhere weak merging and singularity speak not just to the specific assumptions and results that appear in the literature, but to the "rational learning" approach generally. We call instead for a different approach to learning, one that recognizes the necessity of genuine, substantive restrictions on beliefs and proposes "extra rational" restrictions that are explicitly grounded in our best understanding of human behavior, ideally gleaned from experimental data.

Printed: 09/26/97: 5:29 PM

Suppose we randomly pull two agents from a population and ask them to observe an unfolding, infinite sequence of zeros and ones. If each agent starts with a prior belief about the true sequence and updates this belief on revelation of successive coordinates, what are the chances that the two agents will eventually agree on the likelihood that the next draw is a one? Alternatively, if we fix a particular objective measure and then pick a single agent from the population and ask her to observe a sequence of zeros and ones generated by that measure, what are the chances that her belief will eventually conform to the true likelihood? In this paper we show that there is no chance in either case. Formally, we show, in a more general context, under a very unrestrictive definition of what it means to draw priors "randomly," that the probability that two priors have any chance of weakly merging is zero. Indeed, almost surely, the two measures will be *singular*—one prior will think certain a set of sequences that the other thinks impossible (and vice versa for the complement of this set).

Our result is meant as a critique of the "rational learning" literature, which seeks positive convergence results on infinite product spaces by augmenting the process of Bayesian updating with seeming regularity conditions, variously labeled "consistency" or "compatibility" assumptions. This literature is grounded in a famous theorem due to Blackwell and Dubins (1962): if Q and P are measures on an infinite product space and Q is *absolutely continuous* with respect to P (i.e., all zero measure events under P are also zero measure under Q) then with Q probability one, Q and P will *merge,* where merging means that the distance (appropriately defined) between their conditionals on the full continuation of the sequence will limit to zero as the sequence unfolds. Kalai and Lehrer (1993) rely on this result to claim, as in their title, that "Rational Learning Leads to Nash Equilibrium" in the context of an infinitely repeated game. The same authors (1994) subsequently introduce *weak merging*: convergence of the conditionals on only the next period's draw as opposed to the infinite future. And Lehrer and Smorodinsky (1995) show that a "compatibility assumption" weaker than absolute continuity implies "almost weak merging."

Importantly, the assumptions generating these convergence results are not offered as attractive characterizations of how individuals' current beliefs are affected by past observations, but rather as regularity conditions on priors that are proffered as if they rule out the occasional perverse case. It then seems fair to ask just how regular these assumption and results are when considered in the space of all possible prior distributions.

Our answer is that it is the assumptions and results themselves that are perverse. Indeed, we show that not only probability one weak merging, but any chance of weak merging is "non-generic:" with probability one the two measures will almost surely fail to merge. As a corollary, we prove that with probability one, not only will the measures lack absolute continuity, but they will be mutually *singular* (meaning that we can find a zero measure event for P to which Q assigns not just positive measure, but full measure 1).

In general terms, we conclude that a more fruitful approach to learning would start by recognizing the inherent vacuity of Bayesian updating on infinite product spaces and the corresponding necessity of genuine, substantive restrictions on beliefs. Such an approach would then propose explicitly "extra rational" restrictions grounded in our best understanding of human behavior, as in literature on probabilistic learning: Hurkens (1995), Sanchirico (1996) and Sonsino (1997).

This raises the question of whether absolute continuity itself can be recast as a useful behavioral assumption. In an earlier paper (Miller and Sanchirico [1996]), we suggest that this is not the case. There we provide an alternative proof of Blackwell and Dubins (1962) Merging of

Opinions Theorem that makes clear the sense in which assuming absolute continuity is in fact the same as assuming merging.

Mathematically, an exercise such as the one conducted in this paper poses several challenges. Primary among these is the difficulty of working with and conceiving of probability measures over a space of probability measures that in turn pertain to an infinite product space. We address this difficulty by viewing measures from the equivalent "local perspective," as discussed below. As the reader will see, this reduces the measurable space of measures to a more familiar object to which the usual tools may be applied.

A related difficulty is that of generating a notion of picking measures on infinite product spaces "at random." The problem is that there is no proper uniform distribution on infinite dimensional spaces. One simple notion of random choice, for binary sequences, is to assign each "branch probability" (defined below) by drawing from an i.i.d. uniform distribution. This case was worked out by Freedman (1966) (to whom we owe our application of Fubini's Theorem in Lemma 1). Whether Freedman's is the proper notion of randomness is open to debate. In this paper, however, we sidestep the issue by showing that no merging holds for a range of measures over priors that are broad enough to include any reasonable notion of what it means to draw priors at random. In particular, we offer two sets of restrictions, both of which include the i.i.d. uniform case, and each of which (alone or in path-wise combination with the other) is sufficient to guarantee that the probability of weak merging anywhere is zero.[1] Naturally, some restriction is necessary: if, for example, our measure on priors puts all weight on one particular prior, then there would always be immediate merging.

[1] We also generalize Freedman (1966) by drawing sequences from arbitrary countable factor spaces.

In Section 1 we lay out the general framework and identify the "local" perspective on priors over infinite sequences. In Section 2 we imagine that the true sequence is drawn according to some fixed "objective" measure and that we randomly draw an individual/prior. The question is whether observation of the past will enable the individual to "learn" the true measure generating the rest of the sequence. In Section 3 we imagine that two individuals are drawn randomly and question whether their opinions about the future will "merge" as they jointly observe the sequence unfold.

1. Notation and Set-up: The "Local" Perspective

First we define the underlying measurable space of sequences. We write $X = \prod_{i=1}^{\infty} X_i$, where the factors X_i are assumed to be countable. Each X_i is endowed with the discrete σ-algebra and X has the usual infinite product σ-algebra. We call elements of *X* *paths*. Let $B = \{h_\varnothing\} \cup \bigcup_{n=1}^{\infty} \prod_{i=1}^{n} X_i$ be the set of all *partial histories* or *nodes*. ($h_\varnothing$ is the null history.) We say that partial history *b* is of *length* $n \geq 1$ if it is an element of $\prod_{i=1}^{n} X_i$. The null history has length 0. Let $h_n(x)$ be the *n* length partial history of path *x*: the projection of *x* onto its first *n* coordinates. Given any measure *Q* on *X*, any path *x*, let $q_n(x) = Q([h_n(x') = h_n(x)])$, i.e. the probability that the true path agrees with *x* on the first *n* coordinates.

Next, we define the measurable space of *positive* probability measures on *X*, i.e., probability measures that put positive weight on every partial history.[2] Here we exploit a simple but crucial equivalence between the "local" and "global" form of measures on *X*. Given any positive

[2] Our restriction to such measures is for convenience only and is a typical simplification in this literature.

measure Q and any partial history b of length n, we can easily derive, via Bayes rule, the probability measure on X conditional on the event of reaching b. We can then find the marginal of this conditional measure on the $n+1^{\text{th}}$ factor of X. Formally, for any positive measure Q, any partial history b of length n and any element y of X_{n+1}, we may define $t(b)(y)$ according to

$$Q\left(\overline{by}\right) = t(b)(y)Q\left(\overline{b}\right), \tag{1}$$

where for any b', $\overline{b'}$ denotes the set of all paths with partial history b' and *by* is the *n+1* length partial history consisting of b followed by y. The resulting *local measure* on X_{n+1} will be an element of the (interior of the) infinite dimensional simplex, $S \equiv \left\{ y \in (0,1)^{Z} \middle| \sum_{j=1}^{\infty} y_j = 1 \right\}$. Applying this operation to each partial history, we see that every positive measure Q induces one and only one *system of local measures:* an element t of the infinite product of simplices S^B.

We may think of the elements of X as paths through an infinite tree, with each b a node and each X_{n+1}, the list of branches emanating from nodes of length n. Then $q_n(x)$ is the probability of reaching the n^{th} node along path x. And the coordinate $t(b)$ is the measure over branches emanating from node b, while $t(b)(y)$ is the probability that the particular branch $y \in X_{n+1}$ is taken at node b. Thus, associated with each positive measure Q is a comprehensive list of nodal probability measures.

Somewhat less obvious is the fact that the map just described—from the set of positive measures on X to the (interior of the) set of systems of local measures, S^B—is actually one to one and onto. That is, under the map defined by (1), every system of local measures $t \in S^B$ is induced by one and only one positive measure on X. This would be trivial were X a <u>finite</u> product. We could then obtain the measure on any of the finite number of full paths simply by multiplying up the local measures associated with each of the branches along that path. As it turns out, the same intuition carries over to the case of countably infinite products. (See Ash (1972) Theorem 2.7.2, p. 109 for a more general case than that considered here, that of countably many factors of arbitrary cardinality.) *Thus a measure on the countably infinite product X is nothing more, nothing less, than a system of branch probabilities.*[3] This simple but often overlooked insight is crucial not only to an understanding of the results in this paper, but also, we feel, to a balanced view of the "rational learning" approach.

Given the equivalence just discussed, we will use the notation t for both the system of local measures and the corresponding positive measure on X. Moreover, we will endow the space of positive probability measures on X with the natural σ-algebra for S^B. In particular, we endow the factor S with the usual σ-algebra: namely, the restriction to S of the product σ-algebra generated by the Borel σ-algebra on $(0,1)$. Then we endow S^B with the product σ-algebra. We let μ denote a probability measure on this measurable space. Occasionally, we refer to μ as the *meta-measure*.

2. ONE MEASURE FIXED, ONE MEASURE "RANDOMLY" DRAWN

In this section we imagine that the sequence $x \in X$ is drawn according to some "objective" measure Q on X and that we randomly draw a prior t from S^B according to the measure μ. The question is whether t will "learn" Q—formally, whether t will Q-weakly merge with Q (as

[3] This is not in general true for factor spaces of arbitrary cardinality. But the literature on merging of opinions, whenever it does consider a more general product space, restricts attention to measures for which this is true. Such measures are called "predictive." (See, e.g., Blackwell and Dubins (1962).)

defined in the main theorem). The statement of the theorem requires one non-standard definition, and one standard definition that may not be familiar to all readers.

2.1 No Asymptotic Point Mass

The new definition is the notion of no asymptotic point mass in a sequence of random variables. An explanation follows the formal definition.

Definition 1: A sequence of real valued random variables $\{z_n\}$ with support on an interval is said to have *no asymptotic point mass* if $\lim_{\varepsilon\to 0}\left(\underline{\lim}_{n\to\infty} \sup_{q\in\Re} P\left(|z_n - q| \le \varepsilon\right)\right) = 0$.

In order to help interpret this definition it is helpful to break it into pieces. The innermost part of the expression, $\sup_{q\in\Re}\{P(|z_n - q| \le \varepsilon)\}$, is the most probability mass contained in any 2ε interval, which might be called the "ε-concentration of z_n." (Clearly, any single z_n has no point mass if and only if its ε concentration goes to zero with ε.) The limit infimum of the sequence of these ε concentrations is the smallest persistent degree of ε-concentration. We require that this go to zero with ε.

Some examples may help clarify this property. Clearly a sequence of uniform i.i.d. random variables satisfies the condition: for every n, the ε-concentration is 2ε and so the limit infimum of the ε-concentrations is also 2ε, which obviously goes to zero in ε. However, it is important to note that *the no asymptotic point mass condition is on the marginal distributions of the random variables*, so that independence is not needed: a sequence of perfectly correlated identical uniform distributions also has the property. Homogeneity too may be relaxed. It would suffice for infinitely many of the z_n to be uniform; the rest could be anything at all including degenerate distributions. And it goes without saying that there is nothing special about the uniform distribution: similar logic applies to any distribution not having point mass. Even more, a sequence can have a mass point in the distribution for every single random variable while still having no asymptotic mass point. An example is a sequence whose n^{th} element has mass 1/n on ½ and is uniform everywhere else. Since the mass point disappears asymptotically it meets the criterion. Another special case of no asymptotic point mass is one in which infinitely many of the z_n have densities and the modes of the densities do not go to infinity.

So what sorts of sequences *do* have asymptotic point mass? One simple example is any sequence which converges in distribution to a random variable with some point mass. Another is a sequence of random variables, possibly non-convergent, such that for some $\delta > 0$, each has some point with at least δ mass.

2.2 Mixing

The following standard definitions, commonly used in the statistics and econometrics literatures, are provided for the reader's convenience.

Definition 2: Given two σ-algebras, $\Im_1$ and $\Im_2$ on an arbitrary set, define:

$$\alpha(\Im_1, \Im_2) = \sup_{A \in \Im_1,\, C \in \Im_2} \{ |P(A \cap C) - P(A)\,P(C)| \}$$

For a conditioned stochastic sequence $\{z_n, \Im_n\}$ define the *α-mixing coefficients*:

$$\alpha(m) = \sup_n \alpha(\Im_n, \Im_{n+m}^{\infty}),$$

where $\Im_{n+m}^{\infty}$ is the continuation Borel σ-algebra induced by the random variables $\{z_{n+m}, z_{n+m+1}, \ldots\}$. Then the stochastic sequence is said to be *α-mixing* if $\lim_{m\to\infty} \alpha(m) = 0$.

This is one way of defining the notion of “asymptotic independence.” There are other mixing conditions such as ϕ- and ρ-mixing, but α-mixing is implied by the others. Further discussion of α-mixing and other mixing conditions may be found in White (1984).

2.3 Statement and Discussion of Main Theorem

If we draw *t* according to μ, then for any fixed partial history *b* and next branch *y*, $t(b)(y)$ is a random variable on the open unit interval, while $t(b)$ is a random vector with values on the infinite dimensional simplex. Further, for any fixed path *x*, $\left\{t(h_n(x))(x_{n+1})\right\}_{n=1}^{\infty}$ is a sequence of random variables and $\left\{t(h_n(x))\right\}_{n=1}^{\infty}$ a sequence of random vectors. The assumptions placed on μ to capture the notion of randomly drawing priors will be stated in terms of the sequences $\left\{t(h_n(x))(x_{n+1})\right\}_{n=1}^{\infty}$.

Theorem 1: Consider any measure μ *on the space* S^B *of positive probability measures on* X *(those that assign zero to no partial history). For any fixed path,* $x \in X$ *consider the sequence* $t(h_n(x))(x_{n+1})$ *of* μ*-random variables, representing for each* n*, the randomly drawn n+1*th *branch probability along path x.*

Suppose that for all paths x the sequence $t(h_n(x))(x_{n+1})$ *has at least one of the following properties: a) it has no asymptotic point mass; b) it is* α*-mixing and its variance does not go to zero (i.e.,* $\overline{\lim}_{n\to\infty}$ *var [t(*h_n*(x))(*x_{n+1}*)] > 0).*

Then whatever the "true measure" Q, there is no chance that a measure drawn according to μ *will Q-anywhere merge to Q. That is, for all positive measures Q on X, for* μ *a.e.* $t \in S^B$*, the event* $\left[\lim_{n\to\infty}\left(t(h_n(x))(x_{n+1}) - q_{n+1}(x)/q_n(x)\right) = 0\right]$ *has Q-measure zero.*

Further, the randomly drawn measure will almost surely be singular with respect to Q. That is, $\mu(t \perp Q) = 1$.

Two examples on the binary tree may help clarify the result and the relation between the two possible conditions. Suppose the population of potential priors only contains individuals who believe that the sequence is generated by i.i.d. Bernoulli random draws. Thus each individual may be associated with her Bernoulli probability. Suppose also that the population distribution is uniform across these Bernoulli probabilities. Then the implied measure satisfies no asymptotic point mass, since along each path the branch probabilities are perfectly correlated uniforms, which was one of the examples used in Section 2.1 above. To gain some intuition as to why there is also no chance of any weak merging, consider a fixed path. In order for *Q* and any individual's beliefs to merge, it must be that *Q*'s branch probabilities converge to a fixed number, say one quarter. But the chance of drawing the individual with a one quarter Bernoulli probability is zero.

Note that the example in the preceding paragraph does not satisfy α-mixing, since each individual's branch probabilities are perfectly correlated over time. We can also construct an example that fails to satisfy no asymptotic point mass, but does satisfy the mixing conditions and thus exhibits no weak merging. Suppose that the population is made up of individuals who, for each branch, believe that the probability is either 1/3 or 2/3. Let the measure, μ, by which we draw individuals from the population be equivalent to drawing either 1/3 or 2/3 with equal probability independently across all nodes. Obviously this measure fails the no asymptotic point mass condition along every path. However, since the branch probabilities are drawn independently along each path, α-mixing is satisfied. Further, since the variance is constant and finite over time, the second half of the mixing condition is also satisfied. To get a sense of why there can be no weak merging here, consider a *Q* that has only 2/3 and 1/3 branch probabilities, to give merging its best chance. At every node there is a 50% chance that the individual's branch probability does not match *Q*'s, thus almost surely the individual's branch probability will infinitely often be 1/3 away from *Q*'s.

Note that the condition in the theorem is that each path individually satisfy either condition a or condition b, not that all paths together satisfy the same condition. Lastly, as the reader will note from the proof, for any given *Q*, it suffices that the conditions on $t(h_n(x))(x_{n+1})$ hold *Q*-a.e. We impose the conditions on all paths, since we seek conditions on μ that prevent merging no matter what the true *Q*. (Compare this to imposition of the same conditions in Theorem 2 below.)

2.4 Proof of Main Theorem

We will develop the proof of Theorem 1 through a series of lemmas. The first lemma reduces the problem to one of showing that on any path the probability of merging to any particular fixed sequence is zero. It does this by first showing that if on each path there is no μ-chance of weak merging then there is no μ-chance of weak merging on Q-any path. This is non-trivial because the fact that an event has zero probability on each of the paths does not obviously imply it has zero measure over any continuum of paths. Secondly, the lemma shows that with probability one, the prior chosen will be singular with respect to *Q*.

Lemma 1: Let Q be any positive measure on X. Suppose that along all paths x, the S^B*-event* $\left[\lim_{n\to\infty}\left(t\left(h_n(x)\right)\left(x_{n+1}\right)-q_{n+1}(x)/q_n(x)\right) = 0\right]$ *has* μ*–probability 0. Then with* μ *probability 1, the X-event* $\left[\lim_{n\to\infty}\left(t\left(h_n(x)\right)\left(x_{n+1}\right)-q_{n+1}(x)/q_n(x)\right) = 0\right]$ *has Q probability zero. Further,* $\mu(t\perp Q)=1$.

Proof: Consider $E=\overline{\left[\lim_{n\to\infty}\left(t\left(h_n(x)\right)\left(x_{n+1}\right)-q_{n+1}(x)/q_n(x)\right) = 0\right]}$, the event that there is <u>no</u> weak merging. (Here the over-bar indicates the complement of the event). We will regard this as an event in the product space $S^B\times X$ with the usual product σ-algebra and the product measure induced by the coordinate probabilities *Q* and μ. Take I_E to be the indicator function for the set E. By hypothesis $\int_X\left(\int_{S^B} I_E(x,t)\,d\mu\right)dQ=1$. By Fubini's Theorem $\int_{S^B}\left(\int_X I_E(x,t)\,dQ\right)d\mu=1$. Since $I_E(x,t)\le 1$, μ a.s., *Q* a.s., $I_E(x,t)=1$ and we obtain the conclusion for part a.

For part b, first define $m_n(x,t) = t(\overline{h_n(x)})$ to be the *t* probability of observing the partial history of *x* up to time n. Thus $t(h_n(x))(x_{n+1}) = m_{n+1}(x,t)/m_n(x,t)$. We claim that if $m_{n+1}(x,t)/m_n(x,t) - q_{n+1}(x)/q_n(x)$ does not converge to zero, then $m_n(x,t)/q_n(x)$ does not have a strictly positive limit. This can be seen by noting that:

$$\left|\frac{m_{n+1}}{m_n} - \frac{q_{n+1}}{q_n}\right| \leq \left|\frac{m_{n+1}}{m_n} - \frac{q_{n+1}}{q_n}\right| / \frac{q_{n+1}}{q_n} = \left|\frac{m_{n+1}}{q_{n+1}} - \frac{m_n}{q_n}\right| / \frac{m_n}{q_n}$$

which tells us that $\lim_{n\to\infty} m_n / q_n = a > 0$ implies $\lim_{n\to\infty}\left(m_{n+1}(x,t)/m_n(x,t) - q_{n+1}(x)/q_n(x)\right) = 0$, which is the contrapositive of the statement above. Thus, for μ almost all *t*, for *Q* almost no *x*, does $m_n(x,t) / q_n(x)$ converge to a finite strictly positive limit.

Now fix a *t* in the μ -measure 1 set wherein $m_n(x,t) / q_n(x)$ converges to a finite strictly positive limit with *Q*-probability zero. For the usual reasons (see, e.g., Shiryaev, p. 525) $m_n(x,t) / q_n(x)$ is a *Q*-martingale with constant expectation so that it must converge *Q* almost everywhere. But by choice of *t* there is no *Q*-chance that $m_n(x,t) / q_n(x)$ converges to a strictly positive limit. Thus since $m_n(x,t) / q_n(x)$ is non-negative, there exists an event $A \subset X$, with $Q(A) = 1$, on which $m_n(x,t) / q_n(x)$ converges to <u>zero</u>. However, on *A*, the inverse ratio, $q_n(x) / m_n(x,t)$, must go to infinity. But since this random variable is a *t*-martingale with constant expectation, it converges *t* almost everywhere. Thus $t(A) = 0$ and the conclusion follows.//

The next two lemmas establish that either of the two conditions in the theorem are sufficient to guarantee that, as in the Lemma 1's hypothesis, the μ-probability of weak merging along any given path is zero. The first uses the no asymptotic point mass condition, while the second uses the mixing condition.

Lemma 2a: Consider a sequence of real numbers {q_n} on (0,1) and a stochastic sequence {z_n} with support on (0,1). If {z_n} has no asymptotic point mass, then almost surely z_n does not converge to q_n,, i.e. $z_n - q_n$ does not converge to zero.

Proof: First define the event $A_n^\varepsilon = \{|z_n - q_n| > \varepsilon\}$. Note that showing almost sure non-convergence is equivalent to showing that $P\left(\bigcup_{k=1}^{\infty} P\left(\overline{\lim_{n\to\infty}} A_n^{1/k}\right)\right) = 1$. Now $\underline{\lim}_{n\to\infty} \sup_{q\in\Re} P(|z_n - q| \le \varepsilon) \ge \underline{\lim}_{n\to\infty} P(|z_n - q_n| \le \varepsilon) = 1 - \overline{\lim}_{n\to\infty} P(A_n^\varepsilon)$. Thus, taking the limit as ε goes to zero: $0 = \lim_{\varepsilon\to 0} \underline{\lim}_{n\to\infty} \sup_{q\in\Re} P(|z_n - q| \le \varepsilon) \ge \lim_{\varepsilon\to 0}\left(1 - \overline{\lim}_{n\to\infty} P(A_n^\varepsilon)\right) = 1 - \lim_{\varepsilon\to 0}\left(\overline{\lim}_{n\to\infty} P(A_n^\varepsilon)\right)$, where the first equality is by hypothesis. Thus $1 = \lim_{\varepsilon\to 0}\left(\overline{\lim}_{n\to\infty} P(A_n^\varepsilon)\right) \le \lim_{\varepsilon\to 0}\left(P\left(\overline{\lim}_{n\to\infty} A_n^\varepsilon\right)\right)$ in which the inequality is a standard result from probability theory (see, e.g. Billingsley [1995], p 53). Now for each *n*, $A_n^{\frac{1}{k}}$ increases in *k*. Thus $\overline{\lim}_{n\to\infty} A_n^{\frac{1}{k}}$ increases as well and so by the continuity of measures,

$$1 = \lim_{\varepsilon\to 0}\left(P\left(\overline{\lim}_{n\to\infty} A_n^\varepsilon\right)\right) = \lim_{k\to\infty}\left(P\left(\overline{\lim}_{n\to\infty} A_n^{\frac{1}{k}}\right)\right) = P\left(\bigcup_{k=1}^{\infty} P\left(\overline{\lim}_{n\to\infty} A_n^{\frac{1}{k}}\right)\right).//$$

Lemma 2b: Consider a sequence of real numbers {q_n} on (0,1) and a stochastic sequence {z_n} with support on (0,1). If {z_n} is a) α-mixing and b) its variance does not go to zero (i.e. $\overline{\lim}_{n\to\infty} \mathrm{var}(z_n) > 0$), then almost surely z_n does not converge to q_n,, i.e. $z_n - q_n$ does not converge to zero.

Proof: By property b) there exists a subsequence $\{ z_{n_i} \}$ whose variance converges to some $\gamma > 0$.

Thus $\exists$ N s.t. $\forall\ n_i > N$, $\text{var}(z_{n_i}) > \gamma / 2$. Then, we claim, it must be that $\forall$ ε sufficiently small, $\exists$ δ>0 s.t. $P(|z_{n_i} - q_{n_i}| > \varepsilon) > \delta$. To prove this claim we will show that for any random variable *z*, with values on [0,1]: $\forall q \in [0,1], \forall \varepsilon \in (0,1), P(|z - q| > \varepsilon) \geq \text{var}(z) - \varepsilon^2$. First, note that

$\text{var}(z) = E[(z - E(z))^2] \leq E[(z - q)^2] = \int_0^1 (z - q)^2 \, d\mu_z$, where μ_z is the distribution function for *z*.

Second,

$$
\begin{aligned}
\int_0^1 (z-q)^2 \, d\mu_z &= \int_{(|z-q| \leq \varepsilon)} (z-q)^2 \, d\mu_z + \int_{(|z-q| > \varepsilon)} (z-q)^2 \, d\mu_z \\
&\leq \varepsilon^2 P(|z-q| \leq \varepsilon) + P(|z-q| > \varepsilon) \\
&\leq \varepsilon^2 + P(|z-q| > \varepsilon)
\end{aligned}
$$

Thus $\text{var}(z) \leq \varepsilon^2 + P(|z - q| > \varepsilon)$.

Returning to our chosen subsequence, pick such a particular small ε and consider only the portion of the sequence beyond N. Then showing that the event $|z_{n_i} - q_{n_i}| > \varepsilon$, which we shall call A_{n_i}, occurs infinitely often is sufficient for showing that $z_n - q_n$ does not converge to zero. First note that $\{A_{n_i}$ i.o.$\}$ is equivalent to the condition that $P(\bigcap_{k=b}^{\infty} \overline{A}_{n_k}) = 0, \forall b$. Now, for any ϕ, α-mixing implies that there exists m_ϕ such that if $n_j - n_i \geq m_\phi$ then $\forall\ B_{n_j} \in \mathfrak{I}_{n_j}^{\infty}$, $\forall\ B_{n_i} \in \mathfrak{I}_{n_i}$, $|P(B_{n_i} \cap B_{n_j}) - P(B_{n_i})P(B_{n_j})| < \phi$ where the $\mathfrak{I}$'s are defined as in the definition of α-mixing above. Now, consider any subsequence of $\overline{A}_{n_k}$ such that the n_k are separated by at least m_ϕ, and call it $\overline{A}_{n_k}^{\phi}$. Then:

$$P(\bigcap_{k=b}^{K}\overline{A}_{n_k}^{\phi}) \le P(\bigcap_{k=b}^{K-1}\overline{A}_{n_k}^{\phi})P(\overline{A}_{n_K}^{\phi}) + \phi$$
$$\le \left[P(\bigcap_{k=b}^{K-2}\overline{A}_{n_k}^{\phi})P(\overline{A}_{n_{K-1}}^{\phi}) + \phi \right] \cdot P\left(\overline{A}_{n_K}^{\phi}\right) + \phi .$$

Continuing in this fashion yields:

$$P(\bigcap_{k=b}^{K}\overline{A}_{n_k}^{\phi}) \le \prod_{k=b}^{K} P(\overline{A}_{n_k}^{\phi}) + \phi + \phi \sum_{j=1}^{K}\prod_{k=0}^{j} P(\overline{A}_{n_k}^{\phi})$$

Recall that $P(|z_{n_i} - q_{n_i}| > \varepsilon) > \delta$, so that $P(\overline{A}_{n_k}) \le 1-\delta,\ \forall\ k$. Thus, following from the line above:

$$P(\bigcap_{k=b}^{K}\overline{A}_{n_k}^{\phi}) \le (1-\delta)^{K-b+1} + \phi\sum_{j=0}^{K}(1-\delta)^{j} \le (1-\delta)^{K-b+1} + \phi/\delta$$

Taking the limit as K goes to infinity: $P(\bigcap_{k=b}^{\infty}\overline{A}_{n_k}^{\phi}) \le \phi/\delta$. Note that $P(\bigcap_{k=b}^{\infty}\overline{A}_{n_k})$ is less than or equal to the probability of the same intersection taken over any subsequence. Thus:

$$P(\bigcap_{k=b}^{\infty}\overline{A}_{n_k}) \le P(\bigcap_{k=b}^{\infty}\overline{A}_{n_k}^{\phi}) \le \phi/\delta,\ \forall \phi > 0$$

so that we can conclude that $P(\bigcap_{k=b}^{\infty}\overline{A}_{n_k}) = 0$, $\forall k$ and thus $\{A_{n_i}$ i.o.$\}$ so that $z_n - q_n$ does not converge to zero.//

Proof of Theorem 1: Let Q be any fixed positive probability measure on X. Consider the ratio $m_n(x,t)/q_n(x)$. Fix a particular x, which makes $q_n(x)$ a sequence of reals on the open unit interval and $m_n(x,t)$ a sequence of random variables (defined over the σ-algebra on S^B). Property a) in the statement of the theorem means that the sequence $m_{n+1}(x,t)/m_n(x,t)$ satisfies the no asymptotic point mass assumption in Lemma 2a. Similarly, properties (b) mean satisfaction of the corresponding assumptions in Lemma 2b. Thus, applying either version of Lemma 2 to the sequences $m_{n+1}(x,t) / m_n(x,t)$ and $q_{n+1}(x) / q_n(x)$, we conclude that for all x, μ-almost surely $m_{n+1}(x,t)/m_n(x,t) - q_{n+1}(x)/q_n(x)$ does not converge to zero. Lemma 1 then implies the conclusion.//

3. EXTENSION TO RANDOMLY DRAWING TWO MEASURES INDEPENDENTLY

Thus far we have fixed one measure Q and considered the likelihood that a randomly drawn second measure would Q-merge to it. This corresponds to the economic question of how likely it is that any given individual will learn the objective measure that is generating the infinite sequence unfolding before her. In this section we consider the related problem of whether any <u>two</u> individuals will come to agree in their subjective probabilistic evaluations of the future course of the sequence. Formally, the experiment now is to draw two priors independently, each according to the measure μ, and ask whether the two will merge with respect to either. We show that the conclusion analogous to that of Theorem 1 holds (indeed with slightly weaker conditions).

Theorem 2: Consider any measure μ on the space S^B of positive probability measures on X. For any fixed path $x \in X$, consider the sequence $t(h_n(x))(x_{n+1})$ of μ-random variables, representing for each n, the randomly drawn $n+1^{th}$ branch probability along path x.

Suppose that there exists an event $A \subset X$ with μ-almost sure measure 1 (i.e., $\mu[\ t(A) = 1] = 1$) such that for all paths x in A, the sequence $t(h_n(x))(x_{n+1})$ of μ-random *variables has at least one of the following properties: a) it has no asymptotic point mass; b) it is α-mixing and its variance does not go to zero (i.e., $\overline{\lim}$ var $[t(h_n(x))(x_{n+1})] > 0$).*

Then there is no chance that two measures drawn independently according to μ will merge anywhere with respect to either measure. That is, for $\mu \times \mu$ almost all $(t_1, t_2) \in S^B \times S^B$, the event $[\lim_{n\to\infty}(t_1(h_n(x))(x_{n+1}) - t_2(h_n(x))(x_{n+1})) = 0] \subset X$ has measure zero under both t_1 and t_2.

Further, the two measures will almost surely be singular. That is, $\mu x \mu$ $(t_1 \perp t_2) = 1$.

The following Lemma 3 is similar in purpose and argument to the part of Lemma 1 that relies on Fubini's theorem. It differs only in that the assumptions on the population of measures are assumed to hold for not all, but **almost** all measures, on not all paths, but **almost** all paths—and correspondingly, the conclusion of the lemma applies to not all, but **almost** all draws of the first measure t_1. After establishing this result, we prove Theorem 2 with another application of Fubini's theorem—here with respect to the product $\mu x \mu$ rather than the product $t_1 \times \mu$ or "$Q \times \mu$" as in Lemma 1.

Note that the binary relation "t_1 merges to t_2 with t_1-probability zero" is not symmetric—even though for fixed x the relation "t_1 does not merge to t_2 on x" is. It is simple to construct measures that never merge with respect to one, but sometimes merge with respect to the other.

For any fixed measure $t_1 \in S^B$, write

$E_{t_1} = \overline{\left\{(t_2, x) \in S^B \times X \,\middle|\, \lim_{n\to\infty}\left(t_1\left(h_n(x)\right)\left(x_{n+1}\right) - t_2\left(h_n(x)\right)\left(x_{n+1}\right)\right) = 0\right\}}$ for the event in $S^B \times X$ that merging does <u>not</u> occur. Let $I_{E_{t_1}}(t_2, x)$ be the corresponding indicator function. Thus $\left[I_{E_{t_1}}(t_2, x) = 0\right]$ signifies that t_1 <u>does</u> merge to t_2 along path x.

Lemma 3: Suppose that μ satisfies the assumptions recited in paragraph 2 of Theorem 2. Then for μ-almost every measure $t_1 \in S^B$, there is no chance under μ of drawing a second measure t_2 that merges to t_1 t_1-anywhere. That is, $\exists$ an event H in S^B with $\mu(H) = 1$ s.t. $\forall t_1 \in H$, $\exists$ an event G_{t_1} in S^B with $\mu\left(G_{t_1}\right) = 1$ s.t. $\forall t_2 \in G_{t_1}$, $t_1\left(I_{E_{t_1}}(t_2, x) = 0\right) = 0$.

Proof: Let $H \subset S^B$ be the μ-measure 1 event on which $A \subset X$ has t_1-measure 1. Take any $t_1 \in H$. By Lemmas 2a and 2b, $\forall x \in A$, $\mu\left(I_{E_{t_1}}(\mathbf{t}_2, x) = 1\right) = 1$. Given $t_1 A = 1$, this means $\int_X \left(\int_{S^B} I_{E_{t_1}}(t_2, x)\, d\mu\right) dt_1 = 1$ and so by Fubini's Theorem $\int_{S^B}\left(\int_X I_{E_{t_1}}(t_2, x)\, dt_1\right) d\mu = 1$. Now, $\forall t_2$, $\int_X I_{E_{t_1}}(t_2, \mathbf{x})\, dt_1 \le 1$. Thus, $\exists$ an event G_{t_1} with $\mu\left(G_{t_1}\right) = 1$ s.t. $\forall t_2 \in G_{t_1}$, $\int_X I_{E_{t_1}}(t_2, \mathbf{x})\, dt_1 = 1$, implying that $\forall t_2 \in G_{t_1}$, $t_1\left(I_{E_{t_1}}(t_2, \mathbf{x}) = 0\right) = 0$.//

Notice that the μ-measure 1 event G_{t_2} of t_2's is not necessarily uniformly chosen across the potentially uncountable infinity of t_1's in H.

Proof of Theorem 2: Write $F_i = \left\{(t_1, t_2) \in S^B \times S^B \,\middle|\, t_i\left(I_{E_{t_1}}(t_2, x) = 0\right) = 0\right\}$. Lemma 3 implies $\int_{S^B}\left(\int_{S^B} I_{F_1}(t_1, t_2)\, d\underset{t_2}{\mu}\right) d\underset{t_1}{\mu} = 1$. By Fubini's Theorem $\int_{S^B \times S^B} I_{F_1}(t_1, t_2)\; d(\mu \times \mu) = 1$, so $\mu \times \mu\left(I_{F_1}(t_1, t_2) = 1\right) = 1$. The same argument shows $\mu \times \mu\left(I_{F_2}(t_1, t_2) = 1\right) = 1$. Thus,

$\mu \times \mu\left(I_{F_1}(t_1, t_2) = I_{F_2}(t_1, t_2) = 1\right) = 1$, which is a restatement of the result on merging that we seek.

Almost sure singularity follows as in Theorem 1.//

4. REFERENCES

Ash R., (1972), *Real Analysis and Probability*, Boston, Academic Press.

Billingsley, P. (1995), *Probability and Measure (3rd ed.)*, New York, John Wiley & Sons.

Blackwell, D and L. Dubins (1962), "Merging of Opinions with Increasing Information," *Annals of Mathematical Statistics*, 32:882-886.

Freedman, D (1966), "On Two Equivalence Relations Between Measures," *Annals of Mathematical Statistics* 37:686-689.

Hurkens, S. (1995), "Learning by Forgetful Players," *Games and Economic Behavior*, 11:304-29.

Kalai, E. and E. Lehrer (1993), "Rational Learning Leads to Nash Equilibrium," *Econometrica*, 61:1019-1045.

_____ and _____ (1994), "Weak and Strong Merging of Opinions," *Journal of Mathematical Economics*, 23:73-86.

Lehrer E. and R. Smorodinsky (1995), "Compatible Measures and Merging," *forthcoming Mathematics of Operations Research*.

Miller, R. and C. Sanchirico (1996), "The Role of Absolute Continuity in 'Merging of Opinions' and 'Rational Learning,'" mimeo Columbia Economics Department.

Sanchirico, C. (1996), "A Probabilistic Model of Learning in Games," *Econometrica*, 64: 1375-1393.

Shiryaev, A. (1996), *Probability (2nd Edition)*, New York, Springer.

Sonsino, D. (1994), "Learning to Learn, Pattern Recognition, and Nash Equilibrium," *Games and Economic Behavior*, 18:286-331.

White, H. (1984), *Asymptotic Theory for Econometricians*, Academic Press, New York.