On Convergence of Public Beliefs in Observational Learning

Xuanye Wang (UT Austin)

July 5, 2018

Abstract

In an observational learning environment, the public belief λ_t forms a martingale and converges almost surely to a limit belief λ_∞. If $\{\lambda_t\}_{t\in\mathbb{N}}$ is not uniformly bounded, we invent a method to show that λ_t doesn't converge to λ_∞ in L^1 or in mean.

1 Introduction

Learning through observation of actions plays a crucial role in daily life and economics activities, and is studied by a large literature. A widely used argument in this literature is the martingale argument introduced in Smith and Sørensen (2000). Let λ_t be the posterior belief formed by the society after observing $t-1$ actions. As more and more actions being observed, the sequence of posterior beliefs forms a positive martingale. Positive martingale converges almost surely to a limit variable with finite support. Then one can conclude that the posterior belief must converge after observing enough actions, and learning is obtained in this sense.

An unanswered question in the literature is: could we prove λ_t converges in other sense, like in L^1 or in mean? This question has both a theoretical and a practical motivation. Theoretically, since λ_t is a martingale, $E[\lambda_t] = E[\lambda_1]$ for any finite t. It is natural to ask whether $E[\lambda_\infty] = E[\lambda_1]$. Practically, in solving an observational learning model, one often finds that the public belief converges to two sets. For example, the public belief can either converge to the point where everyone completely learns, or to the point where the action provides no more information (confounded learning point). Convergence in mean will enable one to compute the probability of posterior belief converges to each set. Also, convergence in L^1 restricts the behavior of λ_t, by forcing it not leaving λ_∞ too far and too fast before converges back.

In this paper, we provide a negative answer to above question. Based on a simplified version of the classic model in Smith and Sørensen (2000), we show that the public belief process λ_t doesn't converge to λ_∞ in L^1 and in mean provided that $\{\lambda_t\}_{\mathbb{N}}$ is not uniformly bounded. Our proof explores the power of Dunford-Pettis theorem to prove that non-uniformly bounded $\{\lambda_t\}_{t\in\mathbb{N}}$ is not uniformly integrable, and is general enough to be applied in other models. We also provide a result which can establish L^1 convergence from convergence in probability, see Proposition 3.

In the rest of the introduction, we briefly describe the model, explain the intuition, and briefly summarize the proof strategy.

We consider a standard observational learning model with two underlying states A and B. With probability p, the player at period t would like to take an action matches the underlying state; with the rest probability, player t would like to mismatch the state. Let $\lambda_t = \frac{\Pr(B|\alpha_1,\ldots,\alpha_{t-1})}{\Pr(A|\alpha_1,\ldots,\alpha_{t-1})}$ be the posterior belief conditioned on first $t-1$ actions. Using the standard argument we can show λ_t is a martingale and converges to λ_∞ almost surely. We assume the private signal is of unbounded strength. The support of λ_∞ consists of the full learning point $\{0\}$ and the confounding learning points $\lambda^* \in K$, where $\lambda^* \in K$ solves $\Pr(a|B,\lambda^*) = \Pr(a|A,\lambda^*)$.

If $\{\lambda_t\}_{t\in\mathbb{N}}$ is not uniformly bounded, λ_t increases on a sequence of shrinking sets. For example, let α be a sequence of actions such that $\lambda_t(\alpha)$ increase to infinity. Let E_t be the set such that first t actions agrees with first t actions in α, then λ_t increases to infinity along the decreasing sequence of sets E_t. This reminds us a classic example where a sequence of functions converges a.s. but not in L^1 to a limit:

$$f_n(x) = \begin{cases} 0, & \text{if } x \in [\frac{1}{n}, 1); \\ n, & \text{if } x \in (0, \frac{1}{n}). \end{cases}$$

Here f_n converges to 0 point-wisely but not in L^1 and in mean. The failure of L^1 convergence and mean convergence is due to that f_n increases rapidly on a sequence of shrinking sets. The public belief martingale λ_t meets the same problem. It also increases fast enough on a shrinking sequence of sets which shrinks comparatively slow.

To prove the theorem, We use the Dunford-Pettis theorem to show λ_t is not uniformly integrable. See Theorem 14. Then we explore the equivalence of L^1 convergence, convergence in mean, and uniformly integrability of a martingale, see Theorem 15.

To our best of knowledge, there is no other literature discussing the convergence property of public beliefs in observational learning model.

The paper is organized as following: in section 2 we give a detailed description of the model; in section 3 we solve the model using standard martingale argument and state our main result; in section 4 we prove our main result.

2 Model

We work with a simplified version of the classic model with unbounded private signal as in Smith and Sørensen (2000).

We assume two underlying states A and B, and without loss of generality further assume the true underlying state is A. Time is discrete, that is, $t \in \mathbb{N}$. In each period t, there is a player t who chooses between actions a and b. With probability p, player t is of type M, whose payoff is given as:

	state A	state B
action a	u	0
action b	0	1

A type M player is a player who receives positive payoff from choosing the action that matches the state. With probability $1-p$, player t is of type DM, whose payoff is given as:

	state A	state B
action a	0	1
action b	v	0

A type DM player is a player who receives positive payoff from choosing the action that mismatches the state. We assume $u \neq v \neq 1$.

Below we provide a detailed description of the model with the primary purpose of introducing notations. Readers who are familiar with standard herding model can omit it.

As standard in the literature, player t observes actions of player 1 through player $t-1$. Such a sequence of actions is denoted by $\{\alpha_1, \dots, \alpha_{t-1}\}$, and is referred as the public history h_t at period t. Player t also observes a private signal s_t which is generated from a state-dependent distribution. So player t's information set before taking an action is given by $\{\alpha_1, \dots, \alpha_{t-1}, s_t\}$. Before observing the public history and the private signal, each player t holds a flat prior of underlying states, that is, $\Pr(A|\emptyset) = \Pr(B|\emptyset) = \frac{1}{2}$. Following the literature, a player's private signal is identified with one's posterior belief of state being A only after observing the private signal. With a slight abuse of notation, the above assumption is $s_t = \Pr(A|s_t)$. With underlying state being $state \in \{A, B\}$, s_t is i.i.d drawn from distribution $F^{state}(s)$. We assume the private signal strength is unbounded, that is, $supp(F^{state}(s)) = (0, 1)$. To guarantee no private signal fully reveals the underlying state, we assume that $F^A(s)$ and $F^B(s)$ are mutually absolutely continuous.

3 Description of Public Belief Process λ_t

In this section we define the public belief process λ_t and establish that λ_t converges almost surely to a finite limit variable. The argument used is the standard martingale argument first introduced in Smith and Sørensen (2000). Readers who are familiar with the standard martingale argument can jump to Theorem 1, which is the main theorem of this paper.

The public belief at period t

$$\lambda_t = \frac{\Pr(B|\alpha_1, \ldots, \alpha_{t-1})}{\Pr(A|\alpha_1, \ldots, \alpha_{t-1})}, \tag{1}$$

is the posterior likelihood ratio after observing the public history up to period t. Such public belief is the same across any player acts in or after period t.

Using Bayes' formula, we see

$$\begin{aligned} \lambda_t &= \frac{\Pr(B|\alpha_1, \ldots, \alpha_{t-2}) \Pr(\alpha_{t-1}|B, \alpha_1, \ldots, \alpha_{t-2})}{\Pr(A|\alpha_1, \ldots, \alpha_{t-2}) \Pr(\alpha_{t-1}|A, \alpha_1, \ldots, \alpha_{t-2})} \\ &= \lambda_{t-1} \frac{\Pr(\alpha_{t-1}|B, \alpha_1, \ldots, \alpha_{t-2})}{\Pr(\alpha_{t-1}|A, \alpha_1, \ldots, \alpha_{t-2})}. \end{aligned} \tag{2}$$

Therefore, λ_t can be inductively defined once we know $\Pr(\alpha_{t-1}|state, \alpha_1, \ldots, \alpha_{t-2})$. From now on, we write h_{t-1} for $\{\alpha_1, \ldots, \alpha_{t-2}\}$ in order to simplify notations.

By total probability formula,

$$\Pr(\alpha_{t-1}|state, h_{t-1}) = \sum_{i \in M, DM} \Pr(\alpha_{t-1}|state, h_{t-1}, type_t = i) \Pr(type_t = i|state, h_{t-1}). \tag{3}$$

By assumption we have $\Pr(type_t = M|state, h_{t-1}) = p$.

Player $t-1$'s posterior belief

$$\frac{\Pr(B|h_{t-1}, s_{t-1})}{\Pr(A|h_{t-1}, s_{t-1})} = \frac{\Pr(B|h_{t-1})}{\Pr(A|h_{t-1})}\frac{\Pr(s_{t-1}|B)}{\Pr(s_{t-1}|A)} = \lambda_{t-1}\frac{1-s_{t-1}}{s_{t-1}}.$$

Therefore, if player $t-1$ is of type M, he/she will choose action a if and only if $s_{t-1} > \frac{\lambda_{t-1}}{\lambda_{t-1}+u}$. If player $t-1$ is of type DM, he/she will choose action a if and only if $s_{t-1} < \frac{\lambda_{t-1}}{\lambda_{t-1}+v}$. Substitute into 4, we have

$$\Pr(\alpha_{t-1} = a|state, h_{t-1}) = p[1 - F^{state}(\frac{\lambda_{t-1}}{\lambda_{t-1}+u})] + (1-p)F^{state}(\frac{\lambda_{t-1}}{\lambda_{t-1}+v}). \tag{4}$$

By the flat prior assumption, $\lambda_0 = 1$. Therefore, public belief λ_t can be computed inductively using formulas 2 and 4.

It is not hard to observe that

$$\begin{aligned} E[\lambda_t|h_t] &= E[\lambda_{t-1}\frac{\Pr(\alpha_t|B, h_t)}{\Pr(\alpha_t|A, h_t)}|h_t] \\ &= \lambda_{t-1}[\sum_{\alpha_t\in\{a,b\}}\frac{\Pr(\alpha_t|B, h_t)}{\Pr(\alpha_t|A, h_t)}\Pr(\alpha|A, h_t)] \\ &= \lambda_{t-1}. \end{aligned} \tag{5}$$

Therefore, we conclude that public belief process $\{\lambda_t\}_{t\in\mathbb{N}}$ is a martingale.

Let us pause here to describe the underlying probability space on which $\{\lambda_t\}_{t\in\mathbb{N}}$ is defined. The probability space can be thought as either a space of actions $(R^{\mathbb{N}}, \mathcal{R}^{\mathbb{N}}, \mu^A)$, or a space of signals and types $(\Omega, \Sigma, \mathbb{P})$. A generic element in $R^{\mathbb{N}}$ is denoted as $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_t, \ldots)$, where α_t is action taken at period t. A generic element in Ω is denotes as $\omega = (\omega_1, \omega_2, \ldots, \omega_t, \ldots)$. Here $\omega_t = (s_t, type_t)$ specifies the private signal and the type of player t. In short,

$$\Omega = \Pi_{t\in\mathbb{N}}(0,1) \times \{M, DM\}.$$

The σ-field Σ and probability $\mathbb{P}$ is defined through the standard procedure of taking product of infinite copies of probability spaces. For any $S \in \mathcal{R}^{\mathbb{N}}$, $\mu^A(S) = \mathbb{P}(\{\omega \in \Omega|\alpha(\omega) \in S\})$. Here $\alpha(\omega)$ is the realized sequence of actions along ω.

The martingale convergence theorem (see Theorem 11.5 in Williams (1991)) asserts that a non-negative martingale converges almost surely to a limit random variable with a finite support. Hence

$$\lambda_t \to \lambda_\infty \ a.s.,$$

where $supp(\lambda_\infty) < +\infty$.

The primary conclusion of this paper is that λ_t doesn't converge to λ_∞ in mean and in L^1 generically. We state the result here, and give the intuition and proof in next section.

Theorem 1 *With λ_t defined as in 1, $\lambda_t \to \lambda_\infty$ almost surely. But if $up - v(1-p) \neq 0$, $f^A(1) > 0$, and $\{\lambda_t\}_{t\in\mathbb{N}}$ is not uniformly bounded, then λ_t doesn't converge in mean or in L^1. That is,*

$$\lim_{t\to\infty} E[\lambda_t] \neq E[\lim_{r\to\infty} \lambda_t],$$
$$\lim_{t\to\infty} E[|\lambda_t - \lambda_\infty|] \neq 0.$$

We conclude this section by proving that $supp(\lambda_\infty)$ is generically a bounded set in $\mathbb{R}$, which will be used in the proof of the main theorem. Following theorem B.2 in Smith and Sørensen (2000), $\lambda \in supp(\lambda_\infty)$ if and only if $\lambda = \lambda\frac{\Pr(\alpha|B,\lambda)}{\Pr(\alpha|A,\lambda)}$. Therefore, $supp(\lambda_\infty)$ consists of $\{0, K\}$ where $\lambda^* \in K$ solves

$$\Pr(\alpha|B,\lambda^*) = \Pr(\alpha|A,\lambda^*).$$

Substitute 4 into above equation, we obtain that λ^* solves

$$p[1 - F^A(\frac{\lambda^*}{\lambda^*+u})] + (1-p)F^A(\frac{\lambda^*}{\lambda^*+v}) - p[1 - F^B(\frac{\lambda^*}{\lambda^*+u})] - (1-p)F^B(\frac{\lambda^*}{\lambda^*+v}) = 0. \quad (6)$$

We further compute that

$$\begin{aligned} & \frac{d}{d\lambda^*}(\Pr(a|A,\lambda^*) - \Pr(a|B,\lambda^*)) \\ = \quad & -pf^A(\frac{\lambda^*}{\lambda^*+u})\frac{u}{(\lambda^*+u)^2} + (1-p)f^A(\frac{\lambda^*}{\lambda^*+v})\frac{v}{(\lambda^*+v)^2} \\ & +pf^B(\frac{\lambda^*}{\lambda^*+u})\frac{u}{(\lambda^*+u)^2} - (1-p)f^B(\frac{\lambda^*}{\lambda^*+v})\frac{v}{(\lambda^*+v)^2}. \end{aligned}$$

We have

$$\lim_{\lambda^*\to+\infty}(\lambda^*)^2\frac{d}{d\lambda^*}(\Pr(a|A,\lambda^*) - \Pr(a|B,\lambda^*)) = (up - v(1-p))(f^B(1) - f^A(1)).$$

If $f^A(1) > 0$, by lemma A.1 in Smith and Sørensen (2000) we have $f^B(1) - f^A(1) < 0$. Hence, for generic (p, u, v) satisfying $up - v(1-p) \neq 0$, $\lim_{\lambda^*\to+\infty}(\lambda^*)^2\frac{d}{d\lambda^*}(\Pr(a|A,\lambda^*) - \Pr(a|B,\lambda^*))$ is either strictly positive or strictly negative. By continuity, $\frac{d}{d\lambda^*}(\Pr(a|A,\lambda^*) - \Pr(a|B,\lambda^*))$ is strictly positive or strictly negative, for any λ^* bigger than a constant $M(p, u, v)$. Also,

$\Pr(a|B,+\infty) = \Pr(a|A,+\infty)$. Thus no $\lambda^* > M(p,u,v)$ can solve $\Pr(a|B,\lambda^*) = \Pr(a|A,\lambda^*)$. In other words, $supp(\lambda_\infty) < M(p,u,v)$.

To summarize, we have

Lemma 2 *If $up - v(1-p) \neq 0$ and $f^A(1) > 0$, then $supp(\lambda_\infty)$ is a bounded set in $\mathbb{R}$.*

Lastly, although the confounded learning point λ^* may not exist for all parameters (p,u,v). On the other hand, it is not hard to find examples where confounded learning point exists. Let $F^B(s) = 2s - s^2, F^A(s) = s^2$, then for any (p,u,v) satisfying

$$\min\{\frac{u}{v},\frac{v}{u}\} < \frac{p}{1-p} < \max\{\frac{u}{v},\frac{v}{u}\},$$

we have λ^* exists.

4 Main Result

A classic example of a sequence of functions which converge almost surely but fail to converge in L^1 and in mean on a probability space is:

$$f_n(x) = \begin{cases} 0, & \text{if } x \in [\frac{1}{n}, 1); \\ n, & \text{if } x \in (0, \frac{1}{n}). \end{cases} \tag{7}$$

The underlying probability space is $(0,1)$ equipped with ordinary Borel field and Lebesgue measure. We check that $f_n \to 0$ but

$$\lim_{n\to\infty} E[f_n] = 1 \neq E[\lim_{n\to\infty} f_n] = 0.$$

From this example we can clearly see the problem of being almost surely converge but fail to converge in mean. Despite that f_n converges to 0 on a larger and large portion of $(0,1)$, it rapidly increases on a shrinking sequence of sets. When the function values increases too fast, convergence in mean fails.

The problem described in example 7 explains the intuition why λ_t fails to converge to λ_∞ in mean. After sketching the evolution of λ_t, we do see that λ_t increases on a shrinking sequence of sets. Whether λ_t converges in mean, depends on how fast the sequence of sets shrinks, and how fast λ_t increases.

We first gives out a result which can be used to establish convergence in L^1 from convergence in probability.

Proposition 3 *Let $(B, \mathcal{B}, \nu)$ be a probability space, let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of r.v. defined on B with $E[|X_n|] < +\infty$ for every n. If there exists a bounded r.v. X_∞ ($|X_\infty| < M$ for some $M > 0$) such that X_n converges to X_∞ in probability, then X_n converges to X_∞ in L^1 provided that*

$$\exists C > M, s.t. \lim_{n\to+\infty} \int_{\{b\in B||X_n|>C\}} |X_n| d\nu = 0.$$

To prove theorem 1, we combine two powerful results in mathematics. The first result (see Theorem 15) establishes the equivalence of convergence in mean, in L^1 and uniform integrability of a martingale. The second result (see Theorem 3 in Diestel (1991)) establishes that a sequence of random variables (not necessarily martingales) being uniformly integrable is equivalent to a certain induced finitely additive signed measure being countably additive and positive. Then we show the induced measure is not countably additive.

To start the proof, we first give the necessary definitions and notations.

Let $(B, \mathcal{B}, \nu)$ be a generic probability space. Let $L^1(B)$ be the normed space consists of functions $f : B \to \mathbb{R}$ with L^1 norm $\|f\|_1 = \int_B |f| d\nu$. Let $L^\infty(B)$ be the normed space consists of functions $f : B \to \mathbb{R}$ with L^∞ norm $\|f\|_\infty = \inf\{x|\nu(\{b \in B|f(b) \geq x\}) = 0\}$. Let $ba(B)$ be the normed space with generic element m be a bounded signed finitely additive measure on B, with norm $\|m\| = sup_{E\in\mathcal{B}} \int_E |m(E)| d\nu$.

Definition 4 *Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables on B, $\{X_n\}_{n\in\mathbb{N}}$ is uniformly integrable if*

$$\lim_{M\to\infty} \int_{|X_n|>M} X_n d\nu \to 0 \tag{8}$$

uniformly in n.

Let $L^1(B)^*$ denotes the dual space of $L^1(B)$, let $L^\infty(B)^*$ denotes the dual space of $L^\infty(B)$. Following the standard theory in functional analysis, we have

Lemma 5 *$L^1(B)^* = L^\infty(B)$ and $L^\infty(B)^* = ba(B)$. For a generic $g \in L^\infty(B)$,*

$$g(f) = \int_B g f d\nu, \forall f \in L^1(B).$$

For a generic $m \in ba(B)$,

$$m(g) = \int_B g dm, \forall g \in L^\infty(B).$$

For details about lemma 5, and spaces $L^1(B), L^\infty(B), ba(B)$, see III.7 and IV.8 in Dunford and Schwartz (1957).

An immediate corollary of lemma 5 is

Corollary 6 $L^1(B)k$ *is isometrically embedded in* $ba(B)$.

The following Dunford-Pettis theorem states a sufficient and necessary condition for a sequence of r.v. $\{X_n\}_{n\in\mathbb{N}}$ being uniformly integrable.

Theorem 7 *(Dunford-Pettis) Let* $\{X_n\}_{n\in\mathbb{N}} \subset L^1(B) \subset ba(B)$, *then* $\{X_n\}_{n\in\mathbb{N}}$ *is uniformly integrable if and only if*

$$\mu \in \overline{\{X_n\}_{n\in\mathbb{N}}}^{wk*} - \{X_n\}_{n\in\mathbb{N}} \Rightarrow \mu \in L^1(B). \tag{9}$$

Here the weak-star closure in 9 is taken in $ba(B)$. For statement and proof of theorem 7, see pg45-pg50 in Diestel (1991).

Let us view $\{\lambda_t\}_{t\in\mathbb{N}}$ as a martingale defined on $(\Omega, \Sigma, \mathbb{P})$. Obviously $\{\lambda_t\}_{t\in\mathbb{N}} \subset L^1(\Omega)$. From now on, we reserve μ for elements in $\overline{\{\lambda_t\}_{t\in\mathbb{N}}}^{wk*} - \{\lambda_t\}_{t\in\mathbb{N}}$.

By Dunford-Pettis theorem 7, $\{\lambda_t\}_{t\in\mathbb{N}}$ is uniformly integrable iff there exists $g \in L^1(\Omega)$ such that $\mu(E) = \int_E g d\mathbb{P}$, $\forall E \in \Sigma$. Then μ must be countably additive on $(\Omega, \Sigma, \mathbb{P})$. Following results from lemma 8 to theorem 14 show that μ cannot be countably additive.

To start, we need a result to describe the μ.

Lemma 8 *If* $\mu \in \overline{\{\lambda_t\}_{t\in\mathbb{N}}}^{wk*} - \{\lambda_t\}_{t\in\mathbb{N}} \subset ba(\Omega)$, *then* $\forall E \in \Sigma$, $\forall \epsilon > 0$, *there exists a subsequence* $t_k(E, \epsilon)$ *such that*

$$|\mu(E) - \int_E \lambda_{t_k} d\mathbb{P}| < \epsilon.$$

Proof. Let χ_E be the indicator function of $E \in \Sigma$. By definition of weak-star closure, $\forall \epsilon > 0$, there exists a subsequence t_k such that

$$|\int_\Omega \chi_E d\mu - \int_\Omega \lambda_{t_k} d\mathbb{P}| < \epsilon.$$

■

Following two lemmas can be proved directly.

Lemma 9 *μ is positive, that is, $\forall E \in \Sigma$, $\mu(E) \geq 0$.*

Proof. Using lemma 8, $\forall E \in \Sigma$, $\forall \epsilon > 0$, $\exists t_k$ such that

$$\mu(E) > \int_E \lambda_{t_k} d\mathbb{P} - \epsilon.$$

The result follows because $\lambda_{t_k} \geq 0$. ■

Lemma 10 *μ is $\mathbb{P}$-continuous, that is, if $\forall E \in \Sigma$ such that $\mathbb{P}(E) = 0$, then $\mu(E) = 0$.*

Proof. Using lemma 8, $\forall E \in \Sigma$,$\forall \epsilon > 0$, $\exists t_k$ such that

$$\int_E \lambda_{t_k} d\mathbb{P} - \epsilon < \mu(E) < \int_E \lambda_{t_k} d\mathbb{P} + \epsilon.$$

If $\mathbb{P}(E) = 0$, then $\int_E \lambda_{t_k} d\mathbb{P} = 0$. Then the result follows. ■

It is well known that a finitely additive measure μ is countably additive if and only if μ is continuous at empty set. See Exercise 2.8 in Royden and Fitzpatrick (2010). Following lemma rigorously states this fact.

Lemma 11 *Let ν be a finitely additive measure on a measurable space $(B, \mathcal{B})$. Then ν is countably additive if and only if $\forall B_n \in \mathcal{B}$ satisfying*

- $B_1 \supset B_2 \ldots B_n \supset B_{n+1} \supset \ldots$;
- $\bigcap_{n \in \mathbb{N}} B_n = \emptyset$;

We have $\lim_{n\to\infty} \nu(B_n) = 0$.

For notation abbreviation, we will write $B_n \downarrow \emptyset$ for the sequence of sets as described in lemma 11.

Proof. Let $\{E_n\}_{n\in\mathbb{N}}$ be a disjoint countable sequence of sets. Define $B_n = \bigcup_{k \geq n} E_k$. Then $B_n \downarrow \emptyset$. Then we have

$$\nu(\dot{\cup}_{n\in\mathbb{N}} E_n) \;\; = \;\; \sum_{k=1}^{n-1} \nu(E_n) + \nu(B_n)$$

for any n due to finite additivity of ν. Then

$$\nu(\dot{\cup}_{n\in\mathbb{N}} E_n) = \sum_{k=1}^{\infty} \nu(E_n)$$

if and only if $\lim_{n\to\infty} \nu(B_n) = 0$. ■

Using lemma 3, we can easily obtain a proposition which describe an if and only if condition for μ being countably additive.

Proposition 12 *μ is countably additive if and only if $\forall B_n \downarrow B$ with $\mathbb{P}(B) = 0$, we have $\lim_{n\to\infty} \mu(B_n) = 0$.*

Proof. For any $B_n \downarrow B$ with $\mathbb{P}(B) = 0$, there is associated sequence $\tilde{B}_n \doteq B_n - B$ such that $\tilde{B}_n \downarrow \emptyset$. That μ is countably additive if and only if $\lim_{n\to} \mu(\tilde{B}_n) = 0$, which is equivalent to $\lim_{n\to}(\mu(B_n) - \mu(B)) = 0$. But $\mu(B) = 0$ due to lemma 10. ■

However, we can construct a sequence of sets S_n decreasing to a null set S but with $\lim_{n\to\infty} \mu(S_n) > 0$.

Let M be an upper bound of $supp(\lambda_\infty)$. If $\{\lambda_t\}_{t\in\mathbb{N}}$ is not uniformly bounded, $\exists T$ such that $\{\omega \in \Omega | \lambda_T(\omega) > M\} \neq \emptyset$. Define

$$S_n = \{\omega \in \Omega | \lambda_{T+k}(\omega) > M \text{ for } k = 1, \ldots, n\}.$$

It is obvious that $S_n \supset S_{n+1}$. Also,

$$\bigcap_{n\in\mathbb{N}} S_n = \{\omega \in \Omega | \lambda_t(\omega) > M,\ \forall t \geq T + 1\}.$$

This intersection is obviously a null set since $\lim_{t\to\infty} \lambda_t = \lambda_\infty$ almost surely, and $supp(\lambda_\infty) < M$.

Following lemma is a key observation.

Lemma 13 $\lim_{n\to\infty} \lim_{t\to\infty} \int_{S_n} \lambda_t d\mathbb{P} > 0$.

Proof. First we note that

$$\lim_{t\to\infty} \int_{S_n} \lambda_t d\mathbb{P} = \int_{S_n} \lambda_{T+n} d\mathbb{P} = \int_{A(S_n)} \lambda_{T+n} d\mu^A.$$

This is because $\forall t \geq T + n$, we have $E[\lambda_t | S_n] = E[\lambda_{T+n} | S_n]$.

Let α_{T+n} be a sequence of actions up to period $T + n$. By definition,

$$\lambda_{T+n}(\alpha_{T+n}) = \frac{\Pr(B|\alpha_{T+n})}{\Pr(A|\alpha_{T+n})} = \frac{\Pr(a_{T+n}|B)}{\Pr(a_{T+n}|A)}.$$

Here the last equality uses the assumption of flat prior. Then

$$\int_{A(S_n)} \lambda_{T+n} d\mu^A = \sum_{\alpha_{T+n} \in A(S_n)} \frac{\Pr(a_{T+n}|B)}{\Pr(a_{T+n}|A)} \mu^A(\alpha_{T+n}) = \Pr(A(S_n)|B).$$

Here $A : \Omega \to R^{\mathbb{N}}$ is defined as $A(\omega) = \alpha(\omega)$, that is, A maps a sequence of realized private signals and types ω into the sequence of actions happens along ω. Then

$$\lim_{n\to\infty} \lim_{t\to\infty} \int_{S_n} \lambda_t d\mathbb{P} = \lim_{n\to\infty} \Pr(A(S_n)|B) > 0.$$

Here we conclude the limit is strictly positive because $\lambda_t \to \infty$ with positive probability if the true underlying state is B. ■

Using lemma 13, we obtain our first theorem.

Theorem 14 *If $\mu \in \overline{\{\lambda_t\}_{t\in\mathbb{N}}}^{wk*} - \{\lambda_t\}_{t\in\mathbb{N}} \subset ba(\Omega)$, then μ is not countably additive for the reason that $\lim_{n\to\infty} \mu(S_n) \neq 0$. Hence $\{\lambda_t\}_{t\in\mathbb{N}}$ is not uniformly integrable.*

Proof. For any $\epsilon > 0$, there exists subsequence t_k such that

$$\mu(S_n) > \int_{S_n} \lambda_{t_k} d\mathbb{P} - \epsilon.$$

Therefore

$$\lim_{n\to\infty} \mu(S_n) \geq \lim_{n\to\infty} \lim_{k\to\infty} \int_{S_n} \lambda_{t_k} d\mathbb{P} - \epsilon > 0.$$

By Dunford-Pettis theorem, $\{\lambda_t\}_{t\in\mathbb{N}}$ is not uniformly integrable. ■

The following well-known truth establishes that λ_t cannot converge to λ_∞ in L^1 and in mean.

Theorem 15 *Let $\{M_t\}_{t\in\mathbb{N}}$ be a non-negative martingale, then $M_t \to M_\infty$ a.s. The following are equivalent:*

- *$\{M_t\}_{t\in\mathbb{N}}$ is uniformly integrable.*
- *$\lim_{t\to\infty} \int_\Omega |M_t - M_\infty| d\mathbb{P} = 0$.*
- *$\lim_{t\to\infty} E[M_t] = E[M_\infty]$.*

See Theorem 13.7 in Williams (1991) for a proof.

A Proof of Proposition 3

In this section, all notations possess the same meaning as in Proposition 3.

Lemma 16 *$\exists C > M$ such that $\lim_{n\to+\infty}\int_{\{b\in B||X_n|>C\}}|X_n|d\nu = 0$ implies*

$$\forall\epsilon < C - M, \lim_{n\to+\infty}\int_{\{b\in B||X_n-X_\infty|>\epsilon\}}|X_n|d\nu = 0.$$

Proof. Arbitrarily choosing an $\epsilon < \delta$, we have

$$\begin{aligned}
& \lim_{n\to+\infty}\int_{\{b\in B||X_n-X_\infty|>\epsilon\}}|X_n|d\nu \\
= & \lim_{n\to+\infty}[\int_{\{b\in B||X_n-X_\infty|>\epsilon\}\cap\{b\in B||X_n|>C\}}|X_n|d\nu + \int_{\{b\in B||X_n-X_\infty|>\epsilon\}\cap\{b\in B||X_n|\le C\}}|X_n|d\nu] \\
\le & \lim_{n\to+\infty}[\int_{\{b\in B||X_n|>C\}}|X_n|d\nu + C\nu(\{b\in B||X_n-X_\infty|>\epsilon\}\cap\{b\in B||X_n|\le C\})] \\
= & \lim_{n\to+\infty}\int_{\{b\in B||X_n|>C\}}|X_n|d\nu.
\end{aligned}$$

Here the last equation uses that X_n converges to X_∞ in probability. ■

Lemma 17 *If $\mu \in \overline{\{X_n\}_{n\in\mathbb{N}}}^{wk*} - \{X_n\}_{n\in\mathbb{N}}$, then $\forall E\in\mathcal{B}$, we have $\mu(E) = \int_E X_\infty d\nu$.*

Proof. Arbitrarily choosing $\epsilon < C - M$ and $E \in \mathcal{B}$, by lemma 8, there exists subsequence n_k such that

$$\int_E X_{n_k}d\nu - \epsilon < \mu(E) < \int_E X_{n_k}d\nu + \epsilon.$$

We have

$$\begin{aligned}
\mu(E) &< \int_E X_{n_k} d\nu + \epsilon \\
&= \int_{E\cap\{b\in B||X_{n_k}-X_\infty|>\epsilon\}} X_{n_k} d\nu + \int_{E\cap\{b\in B||X_{n_k}-X_\infty|\leq\epsilon\}} X_{n_k} d\nu + \epsilon \\
&\leq \int_{E\cap\{b\in B||X_{n_k}-X_\infty|>\epsilon\}} X_{n_k} d\nu + \int_{E\cap\{b\in B||X_{n_k}-X_\infty|\leq\epsilon\}} X_\infty d\nu \\
&\quad +\epsilon(1+\nu(E\cap\{b\in B||X_{n_k}-X_\infty|\leq\epsilon\})) \\
&\leq \limsup_{k\to+\infty} \int_{E\cap\{b\in B||X_{n_k}-X_\infty|>\epsilon\}} X_{n_k} d\nu + \int_E X_\infty d\nu \\
&\quad -\limsup_{k\to+\infty} \int_{E\cap\{b\in B||X_{n_k}-X_\infty|>\epsilon\}} X_\infty d\nu + \epsilon(1+\nu(E)))
\end{aligned}$$

But

$$\begin{aligned}
& \limsup_{k\to+\infty} \int_{E\cap\{b\in B||X_{n_k}-X_\infty|>\epsilon\}} X_{n_k} d\nu \\
\leq\ & \limsup_{k\to+\infty} \left|\int_{E\cap\{b\in B||X_{n_k}-X_\infty|>\epsilon\}} X_{n_k} d\nu\right| \\
\leq\ & \limsup_{k\to+\infty} \int_{E\cap\{b\in B||X_{n_k}-X_\infty|>\epsilon\}} |X_{n_k}| d\nu \\
\leq\ & 0,
\end{aligned}$$

and

$$\begin{aligned}
& \limsup_{k\to+\infty} \int_{E\cap\{b\in B||X_{n_k}-X_\infty|>\epsilon\}} X_\infty d\nu \\
\geq\ & -M \limsup_{k\to+\infty} \nu(E\cap\{b\in B||X_{n_k}-X_\infty|>\epsilon\}) \\
=\ & 0.
\end{aligned}$$

Therefore, we have $\mu(E) < \int_E X_\infty d\nu + \epsilon(1+\nu(E))$ for all $0<\epsilon<\delta$. Similarly we can prove that $\mu(E) > \int_E X_\infty d\nu + \epsilon(1+\nu(E))$ for all $0<\epsilon<\delta$. Hence $\mu(E) = \int_E X_\infty d\nu, \forall E \in \mathcal{B}$. ■

Using Dunford-Pettis theorem, lemma 17 implies that $\{X_n\}_{n\in\mathbb{N}}$ is uniformly integrable. Then using Theorem 13.7 in Williams (1991), we have X_n converges to X_∞ in L^1.

References

DIESTEL, J. (1991): "Uniform Integrability: An Introduction," *Rendiconti dellIstituto di Matematica dellUniversit di Trieste. An International Journal of Mathematics*, 23, 41–80.

DUNFORD, N. AND J. T. SCHWARTZ (1957): *Linear Operators, Part 1: General Theory.*

ROYDEN, H. AND P. FITZPATRICK (2010): *Real Analysis*, Prentice Hall, 4 ed.

SMITH, L. AND P. SØRENSEN (2000): "Pathological Outcomes of Observational Learning," *Econometrica*, 68, 371–398.

WILLIAMS, D. (1991): *Probability with Martingales*, Cambridge University Press.