DIVISION OF HUMANITIES AND SOCIAL SCIENCES
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA 91125

BAYESIAN ECONOMISTS ... BAYESIAN AGENTS I: AN ALTERNATIVE APPROACH TO OPTIMAL LEARNING

Mahmoud A. El-Gamal
California Institute of Technology

Rangarajan K. Sundaram
University of Rochester

SOCIAL SCIENCE WORKING PAPER 705

August 1989

BAYESIAN ECONOMISTS ... BAYESIAN AGENTS I:*
AN ALTERNATIVE APPROACH TO OPTIMAL LEARNING

Mahmoud A. El-Gamal
California Institute of Technology

Rangarajan K. Sundaram
University of Rochester

August 16, 1989

Abstract

We study the framework of optimal decision making under uncertainty where the agents do not know the full structure of the model and try to learn it optimally. We generalize the results on Bayesian learning based on the martingale convergence theorem to the sequential framework instead of the repeated framework for which results are currently available. We also show that the variability introduced by the sequential framework is sufficient under very mild identifiability conditions to circumvent the incomplete learning results that characterize the literature. We then question the type of convergence so achieved, and give an alternative Bayesian approach whereby we let the economist himself be a Bayesian with a prior on the priors that his agents may have. We prove that such an economist cannot justify endowing all his agents with the same (much less the true) prior on the basis that the model has been running long enough that we can almost surely approximate any agent's beliefs by any other's. We then examine a possibly weaker justification based on the convergence of the economist's measure on beliefs, and fully characterize it by the Harris ergodicity of the relevant Markov kernel. By means of very simple examples, we then show that learning, partial learning, and non-learning may all occur under the weak conditions that we impose. For complicated models where the Harris ergodicity of the Markov kernel in question can neither be proved nor disproved, the mathematical/statistical test of Domowitz and El-Gamal (1989) can be utilized.

Keywords: Rational Expectations, bayesian learning, controlled Markov processes, sequential statistical decision framework.

---

* We wish to thank Larry Blume, David Easley, and Nick Kiefer and participants in the theory workshops of the University of Rochester and Penn State for valuable suggestions and comments. All remaining errors are of course our own. Send communications to Division of Humanities and Social Sciences 228-77, Caltech, Pasadena, CA 91125.

1 Introduction

In recent years, the issue of how rational optimizing agents collect and process data relevant to their activity has received a considerable degree of attention. The underlying structure is quite familiar; a single agent plays a game against nature where the agent is maximizing expected discounted lifetime payoff while contemporaneously learning about the underlying structure of the economy. The simplest example of this framework are bandit problems a survey of which can be found in Berry and Fristedt [2]. We shall be concerned with a reasonably general framework where the agent knows the entire structure of the economy up to a set of parameters. The agent has some prior belief about those parameter values and gets to use his/her observations to update those beliefs. It is no surprise, therefore, that a dominant theme in the relevant literature has been *optimal learning* which has a Bayesian flavor varying in strength according to the authors' tastes. The attention has been twofold: first, what properties does the agent's sequence of beliefs display? and second, if this sequence of beliefs converges, do limit beliefs necessarily place point mass at the true parameter value (in other words, does full learning occur)? The answer to the first question has relied almost exclusively on the use of the Martingale convergence theorem. It is argued that a rational agent cannot expect his/her beliefs to change in any systematic (predictable) way based on his/her chosen action. Consequently, the agent's best guess of tomorrow's belief is today's, and the belief process follows a martingale. An appeal to the martingale convergence theorem, while imposing appropriate conditions immdeiately furnishes the existence of limit beliefs. The focus is then shifted to a characterization of limit beliefs in various environments.[1] It has been known for some time that full learning need not occur even in an infinite horizon problem. The agent may decide to play an uninformative action forever, never getting a chance to learn the truth. Many examples of such situations

[1]The references cited in this paper do not purport to be exhaustive; they merely attempt to give an idea of the work done in studying single agent decision problems under uncertainty. It is worth mentioning here that a number of papers have also studied the problem from the point of view of a (general or partial) equilibrium perspective (see, e.g., Blume and Easley [5], Feldman [18, 19], or the survey in Blume, Bray, and Easley [6]). To overcome problems arising from strategic considerations in these cases, agents are assumed to be atomistic, and therefore myopic maximizers. No single agent believes that his actions will unilaterally matter, so indulges in single period maximization with respect to prior beliefs. Collective actions, however, generate information, which all agents use to update their beliefs. This form of *passive* learning, where tradeoffs between information acquisition and current payoffs do not enter the agent's decision problem, is not our focus in this paper. Neither do we consider ad hoc learning procedures such as least squares learning (see e.g. Jordan [24], Marcet and Sargent [30], or Brock, Marcet, Rust and Sargent [7]), not because we believe that these are unreasonable mechanisms (quite the contrary), but because we wish to examine the validity of the rational expectations hypothesis under the conditions most favorable to it. In economic contexts, Rothschild [36] has invoked the Bandit framework to study pricing by a monopolist. McLennan [31] constructs an example of a monopolist who fails, with non-zero probability, under optimal conditions to identify parameters of the demand curve that he faces. In a general framework, Kiefer and Nyarko [26] examine the related problem of learning the parameters of a linear process and arrive at similar conclusions.

are available in the literature, and we note that they do not seem to require any special structure.[2]

In this paper, we present an alternative Bayesian approach to the issue of optimal learning. The objective of this paper being rather modest, we confine ourselves to describing the new framework, obtaining preliminary theoretical results and examining some other conjectures by means of numerical analysis. The crucial point behind our approach is that beliefs need not, and indeed in general will not, follow a martingale process from the point of view of an objective outside observer informed of the true parameter value.[3] It must be noted here that in general, different initial beliefs can have very different effects on the behavior of the system. From that point of view, we have to consider all possible initial beliefs and track down their evolution according to the true stochastic process. Indeed this is one of the crucial points that we shall discuss in a later section of the paper when the necessary technical toolbox has been developed.[4]

Since the economist knows that different agent priors may result in different behavior, and since there is no a priori reason (save for ones built into the model and/or circular ones) to choose any particular prior, the modelling economist himself becomes a Bayesian in our framework. The economist places a prior over the space of agents' priors and observes the manner in which his own prior evolves with the system. We found it reasonable to assume that the economist's prior should have full support over the space of agent-priors. In summary, the economist specifies the underlying structure of the model, the decision and updating rules for his agents (the latter being Bayesian in the optimal learning framework), his own prior over agent-priors, and then studies the evolution of his own beliefs. We then

[2]Stochastic uncertainty is not necessary for the absence of learning in the presence of optimal behavior. In a perfectly deterministic situation, where a single distinguished action would permit complete identification immediately, Aghion, Bolton, and Jullien [1] provide a simple example where the identifying action is never taken, and the decision maker remains ignorant forever.

[3]To see this in an overly simplified framework, let the parameter to be learned be $\theta \in \Theta = \{\theta_1, \theta_2\}$, then let the agent's prior be defined by f_t^1, his prior that $\theta = \theta_1$. The Bayesian updating rule from his standpoint will be $f_{t+1}^1 = \frac{q(y_{t+1}|.,\theta_1)f_t^1}{q(y_{t+1}|.,\theta_1)f_t^1 + q(y_{t+1}|.,\theta_2)f_t^2}$ When we compute the expectation of f_{t+1}^1 with respect to the agent's prior distribution, it is trivial to see that it reduces to f_t^1, hence the martingale property holds. This is clearly of little interest to the modelling economist since he knows what the true θ is. The expectation of f_{t+1}^1 with respect to the economist's knowledge of the true θ is clearly not equal to f_t^1 except in very pathological cases.

[4]It is worth mentioning at this point that the numerical investigations of convergence of beliefs in Kiefer [25] are in the spirit of the question that we pose here and for which we develop the theory to follow. He generates states using the true parameter value which he (the economist) knows and the agent is trying to learn as a Bayesian and looks for convergence of the agent's beliefs in that experiment. It is clear that this is the question regarding objective convergence which we define and study below and not the subjective convergence of Easley and Kiefer [12, 14]. Moreover, a close look at Figure 11 in [25] clearly shows how neighboring priors can have very different behavior over time, a fact that we shall underline in section 6 of this paper.

examine the following questions:

i. Does the sequence of economist beliefs converge? An affirmative answer will force the economist to consider all agent-beliefs in the support of his/her limit belief (when the model has been running for a very long time) when studying the behavior of the modelled economy.

ii. If convergence of the economist's prior obtains, does the economist limit belief place point mass at some agent-belief? An affirmative answer will offer a partial justification of representative agent type models.

iii. If the answer to [ii.] is also affirmative, does the almost sure (with respect to the economist's limit belief) limit agent-belief itself place point mass at the true parameter value? An affirmative answer will be a full learning result that can partially justify equilibrium arguments such as the rational expectations hypothesis.

iv. If the answers to [ii.] and [iii.] are negative, then we would like to know what portion of the mass (or the economist-beliefs) is concentrated on which agent-beliefs under the particular parametrization that the modelling economist chooses. This will in general require some numerical simulations of the modelled economy.

In section 2, we set up the formal framework for analyzing these questions. This framework is a generalization of the standard repeated Bayesian learning framework [12, 14, 21]. Most economic environments are inherently *sequential* rather than *repeated* in nature. Stated differently, it is often the case that the actions today affect the environment for decision making tomorrow. By focussing on a model in which beliefs (via the Bayes updating map) form the only link between periods, the framework is unduly (and as we show in section 3, unnecessarily) restrictive. Our first results are obtained in section 3, where we show that beliefs - although no longer following a Markov process - do form, from the agent's point of view, a martingale. Intuitively, this clearly has to remain the case since the agents cannot expect their beliefs to change in a systematic way. The existence of limit beliefs now follows immediately from the martingale convergence theorem. In obviously suggestive terminology, we name this the subjective convergence of beliefs. We also show that under weak identifiability conditions of the true parameter, convergence to the true parameter takes place almost surely. This is in contrast to the non-learning results in the literature on the repeated framework where it may be optimal for the agent to repeat the same action forever which in that framework will insure non-learning.

In section 4, we take a first stab at answering the questions posed earlier. We show that when the set of parameter values is finite (or countable), agent-beliefs always converge [a.s.] from the point of view of the objective observer with knowledge of the true parameters. Surprisingly, this follows as a trivial corollary of the results of section

3 based on the martingale convergence theorem. If the parameter space is uncountable, the true parameter value may be such that agent-beliefs do not converge almost surely, but (in a sense to be defined in section 4) this is a zero probability event. Having obtained convergence of agent-beliefs, we first examine a weak version of question [ii.] above. In the spirit of that question, we ask whether we can argue that beyond a certain point, almost all the agents drawn at random from the economist-prior would be sufficietnly close together in their beliefs. The answer - not surprisingly - is no. At any point in time, there is always a positive mass of agents whose beliefs are far from the rest.

In section 5, we explicitely define the operator that determines the evolution of the economist-beliefs and review some mathematical preliminaries regarding conditions for the convergence of its iterates. It is clear that the limit of the iterates of that operator is the subject of questions [i.-iv.] above. It is also clear that with no further restrictions, the answers to questions [i.-iii.] can be of either kind depending on the circumstances. Very trivial examples that satisfy all the conditions in the paper can be constructed to show that 'anything can happen'. In section 6, we pick up where we left off in section 5 and consider a particular problem. We subject Kiefer's [25] monopolist model with his same parametrization to numerical analysis in the spirit of our framework. Our numerical results strongly confirm our original ideas. Convergence of the economist- and agent-beliefs occur rapidly. In that model, limit agent-beliefs split between full learning and convergence to the 'confounding' belief beyond which learning does not occur. The fact that neighboring priors can have very different evolutions over time (as seen in Figure 11 of [25]) becomes a lot more emphasized in our Figures 1-3. We then examine the limit economist-beliefs under different priors suggestive of tests of ergodicity à la Domowitz and El-Gamal [10].

2 The Repeated and the Sequential Statistical Decision Framework

In this section, we introduce the sequential statistical decision framework that was informally discussed in the introduction. To enable distinction from the repeated statistical decision problem, we first provide a description of the latter. Since the repeated framework has been extensively studied in the economics literature (see, e.g. McLennan [32], or Easley and Kiefer [12, 14]), our exposition is relatively informal and brief. For greater detail, the reader is referred to Easley and Kiefer [14]. The section is organized as follows: Subsection 2.1 introduces some basic notation that is maintained throughout the paper. Subsection 2.2 provides an outline of the repeated framework based on the Easley and Kiefer [14] exposition and summarizes the main results of their paper. In subsection 2.3, we introduce the sequential framework of "learning while doing". It is shown, extending the work of Easley and Kiefer [14] that once again

resorting to the techniques of stationary dynamic programming is possible. Standard results from the literature are invoked to prove the existence of optimal plans.

A word on nomenclature: strictly speaking, what we identify as a repeated decision framework in this paper is actually a sequential framework in the sense that uncertainty about the true parameter value of the stochastic law determining the state variable causes today's outcome to influence tomorrow's belief about the value of that parameter. Indeed, this is what enables Easley and Kiefer to cast their problem in a dynamic programming framework. However, in the absence of this uncertainty, there is no connection between periods - hence, the term 'repeated'. In what we call the 'sequential' framework, we allow the possibility that the action taken today affects tomorrow's environment regardless of whether the true parameter value is known or not. Uncertainty about the true parameter value in this case only provides an additional link between periods. We decide to use the term sequential for the latter model and not the former since the second case involves a natural sequential structure where the natural (technological, etc.) environment determining the stochastic law of motion for the state variable depends on the current state and action, whereas in the former, the dependence is artificially introduced by the agent's attempt to learn the true parameter.

2.1 Notation and Definitions

Throughout this paper, the following notation is maintained:

N 1 $(Y, \mathcal{F}(Y), \nu)$ *is a measure space where* Y *is a compact subset of a complete, separable metric space,* $\mathcal{F}(Y)$ *is the smallest* σ*-algebra generated by the Borel subsets of* Y*, and* ν *is a measure on* $(Y, \mathcal{F}(Y))$*. The support of* ν *is taken to be all of* Y*. Y will be referred to as the state space.*

N 2 $(\Theta, \mathcal{F}(\Theta))$ *is a measurable space.* Θ *is assumed to be a compact subset of a complete separable metric space, and* $\mathcal{F}(\Theta)$ *is the smallest* σ*-algebra generated by the Borel subsets of* Θ*.* Θ *will be referred to as the parameter space. There is a distinguished element* $\theta^* \in \Theta$ *which is the true parameter.*

N 3 $P(\Theta)$ *is the space of all probability measures on* $(\Theta, \mathcal{F}(\Theta))$*. A standard result (see, e.g. Parthasarathy [34, Theorem 6.4.1]) states that by the compactness of* Θ*,* $P(\Theta)$ *is compact under the (metrizable) topology of weak convergence. Elements of* $P(\Theta)$ *that place an atom of mass 1 at* $\theta \in \Theta$ *will be denoted by* δ_θ*. (This is a slight abuse of notation since that symbol will usually be reserved for the Dirac delta function at the value* θ *but we use it since it has been used in the economic literature*

on Bayesian learning). We shall denote the support of an element $\lambda \in P(\Theta)$, i.e. the subset of Θ on which λ places positive mass, by supp(λ).

N 4 *$(X, \mathcal{F}(X))$ is a measurable space. X is a compact subset of a complete separable metric space, and as usual, $\mathcal{F}(X)$ is the smallest σ-algebra generated by the Borel subsets of X. $(X, \mathcal{F}(X))$ will be referred to as the action space.*

N 5 *There is a bounded measurable function $r: X \times Y \times Y \to \Re$ referred to as the (instantaneous) reward function.*

N 6 *$\beta \in (0, 1)$ denotes a discount factor.*

2.2 The Repeated Statistical Decision Framework

Consider an agent who in each period picks an action $x \in X$, observes a state $y' \in Y$, and receives a reward $r(x, y')$ (there is a slight abuse of notation here; in the general sequential framework, this will be $r(x, y, y')$, in this special case, we mean by $r(x, y')$ that the reward only depends on the first and last arguments). The agent chooses the value of x to maximize expected discounted sum of rewards over an infinite horizon; i.e.

$$\max_{\{x_t\}} E_0 \sum_{t=0}^{\infty} \beta^t r(x_t, y_{t+1})$$

subject to a natural constraint defined by the stochastic law governing the distribution of the state y'. The conditional distribution of state y' given the chosen action x defining that law denoted by $f(y'|x, \theta^*)$ where $\theta^* \in \Theta$ is the true parameter defined in the previous subsection. If θ^* is known to the agent, then simple (one period) expected value maximization of $r(x, y')$ is optimal in the infinite horizon and entails solving a repeated problem, hence the title of this subsection. This is no longer true if θ^* is unknown to the agent, since the current action, and the observed state therein jointly reveal information about the value of θ^* driving the distribution of y' given x.

To elaborate, suppose the agent starts with a prior $\lambda_0 \in P(\Theta)$. Assume that $supp(\lambda_0) = \Theta$, or more generally that $\theta^* \in supp(\lambda_0)$. The agent's prior represents his belief about θ^*, i.e. $Pr\{\theta^* \in A\} = \lambda_0(A)$ for $A \in \mathcal{F}(\Theta)$. Assume further that the conditional density of y given $(x, \theta) \in (X \times \Theta)$ is given by $f(.|x, \theta)$, and that the agent perfectly knows the value of f for each pair (x, θ); and hence his only ignorance is the value θ takes and the realization of y' resulting form the draw from f. When the agent chooses an action $x \in X$, and observes a state $y' \in Y$, two things happen. First, a reward of $r(x, y')$ is received. Second, the agent uses the observed pair (x, y') to update his beliefs

about the value of θ^*. The posterior belief $\lambda_1 \in P(\Theta)$ will be achieved by applying Bayes' rule:

$$\lambda_1(A) = \frac{\int_A f(y'|x,\theta).\lambda_0(d\theta)}{\int_\Theta f(y'|x,\theta).\lambda_0(d\theta)}$$

for $A \in \mathcal{F}(\Theta)$. It is this ability of actions to generate information that converts the repeated problem into a sequential one, with state space $P(\Theta)$. This process now repeats itself with priors λ_1, choosing an action, observing a state, and updating to λ_2, and so on.

Easley and Kiefer [14] have shown that the agent's maximization problem may now be represented as a dynamic programming problem with state space $P(\Theta)$ and action space X. Given $\lambda \in P(\Theta)$ and $x \in X$, the density f now induces a distribution over $P(\Theta)$ for the next period's prior (which is the posterior that the agent obtains by applying Bayes' rule). Under some regularity conditions, an appeal to Maitra [29] shows the existence of a (stationary) optimal policy $g: P(\Theta) \to X$. Their main results may be summarized as follows:

i. Since the distribution of tomorrow's state depends only on today's state and action, and the latter under the optimal policy is itself a function only of today's state, the agent's beliefs follow a Markov process.

ii. The agent cannot expect his beliefs to change in any way he can predict; consequently, the state follows a Martingale; i.e. for all t, we have $E[\lambda_{t+1}(A)|\lambda_t] = \lambda_t(A)$ for all $A \in \mathcal{F}(\Theta)$.

iii. Construction of the underlying probability space, and an appeal to the martingale convergence theorem now reveal the existence of a $P(\Theta)$-valued random variable λ_∞ such that $\lambda_t \Rightarrow \lambda_\infty$ a.s.$[P_{\lambda_0}]$, where P_{λ_0} is the measure induced on sample paths by the initial belief λ_0.

iv. λ_∞ may not coincide with δ_{θ^*}, i.e. full learning need not occur. Notice that here, as below, we are abusing notation by letting λ_∞ denote both a random variable and its realization; but since we do not anticipate any confusion, and since this abuse of notation is also prevalent in the preceding literature, we shall continue using it. Full learning obtains, however, if the discount factor β is large enough (i.e. $\lambda_\infty = \delta_{\theta^*} a.s.[P_{\lambda_0}]$). In this context, results on incomplete learning in McLennan [32], and Feldman and McLennan [21] are also relevant.

v. Since $g(\lambda_t) \in X$ for all t, and X is compact, $g(\lambda_t)$ contains a convergent subsequence. If $g(\lambda_t) \to x_\infty$, then x_∞ maximizes one period expected payoff given beliefs λ_∞. Further, limit beliefs are invariant for limit actions; i.e. if the agent starts with prior λ_∞ and takes the action x_∞, then the updated beliefs will also be

λ_∞ a.s.$[P_{\lambda_\infty}]$. Indeed this is the source of possible incomplete learning that Easley and Kiefer refer to the most; i.e. if the action sequence converges to some x_∞ too rapidly, then the beliefs may get trapped at some $\lambda_\infty \neq \delta_{\theta^*}$ which is invariant to that x_∞.

All these convergence results relate to convergence under the prior belief of the decision maker on the set of sample paths. In other words, the almost sure convergence of λ_t to some limit belief $\lambda_\infty \in P(\Theta)$ guaranteed by the martingale convergence theorem states that the agents expect that they will converge almost surely with respect to the probability measure induced on sample paths by their initial belief. As argued in the introduction, this is not the same as the almost sure convergence with respect to the probability measure that a neutral observer who knows the value of θ^*, and the initial priors λ_0, will use. We shall return to this point in section 4 below.

2.3 The Sequential Statistical Decision Framework

The flavor of the sequential framework is perhaps best caught by a simple example. Consider the stochastic one-sector model of optimal growth. In each period, an agent observes a stock $y \geq 0$, decides on the fraction $x \in [0,1]$ to be consumed, and receives a reward of $r(x,y) = u(xy)$ where u is a utility function from consumption. The remaining stock $(1-x)y$ forms the investment for the period, and is converted into the next period's stock y' through a production function f which depends on parameter(s) θ^*, and the realization of a random variable ξ as $y' = f((1-x)y, \xi; \theta^*)$. The agent's objective is to maximize the expected discounted sum of utilities from consumption over the infinite horizon. When the parameter(s) θ^* of f, and the distribution of ξ are known, this fits into the standard stochastic dynamic programming framework. When θ^* is not known, however, an action x at a stock y has three effects. First, it yields a utility $u(x,y)$; second, it determines the distribution of the next period's stock; and finally, it yields information about the value of θ^* which will be used by the agent when determining next period's level of consumption.

It is the general formulation of problems of this sort that is carried out in this section. In each period, takes an action $x \in X$ after observing a state $y \in Y$. The state then evolves to a new state y' according to the probabilistic transition density $q(.|x,y,\theta^*)$, and the agent receives a reward of $r(x,y,y')$. The process then repeats with y' replacing y. The value of θ^* is, however, unknown to the agent, who starts with a prior λ over Θ. As before, we assume that $\theta^* \in supp(\lambda)$. If y' is the state observed in the next period, then the agent revises his prior λ to a posterior λ' using the Bayesian updating

formula

$$\lambda'(A) = \frac{\int_A q(y'|x,y,\theta).\lambda(d\theta)}{\int_\Theta q(y'|x,y,\theta).\lambda(d\theta)} \tag{2.1}$$

where $A \in \mathcal{F}(\Theta)$. Now, based on his knowledge of y' and his new prior λ', the decision maker takes an action $x' \in X$, and the process repeats.

To convert the agent's expected utility maximization problem into a dynamic programming problem à la Easley and Kiefer [14], some additional notation and assumptions are introduced. First, we denote the map in (2.1) by $\mathcal{B}: Y \times X \times Y \times P(\Theta) \to P(\Theta)$ so that if $(y_{t+1}, x_t, y_t, \lambda_t)$ represent respectively the period $t+1$ value of the state y, the period t action, the period t state y, and the period t belief, then we have

$$\lambda_{t+1}(A) = \mathcal{B}(y_{t+1}, x_t, y_t, \lambda_t)(A) = \frac{\int_A q(y_{t+1}|x_t, y_t, \theta).\lambda_t(d\theta)}{\int_\Theta q(y_{t+1}|x_t, y_t, \theta).\lambda_t(d\theta)} \tag{2.2}$$

for $A \in \mathcal{F}(\Theta)$.

Next, for the rest of this paper, we shall be considering the augmented state space $S = Y \times P(\Theta)$, and endow S with the product metric topology. Since Y and $P(\Theta)$ are compact, so is S. A generic element of S is denoted by $s = (y, \lambda)$. Now, given a state $s \in S$, and an action $x \in X$, the conditional transition density $q(.|x,y,\theta)$ induces a conditional distribution over S that we denote by $Q(.|s,x)$. This is defined by

$$Q(A|s,x) = \int_\Theta \int_Y I_A(y', \mathcal{B}(y', x, s)) q(y'|x,y,\theta) \nu(dy') \lambda(d\theta) \tag{2.3}$$

for $A \in \mathcal{F}(S)$ where $s = (y, \lambda)$, and I_A is the indicator random variable which equals 1 if $s \in A$ and 0 otherwise, and as usual, $\mathcal{F}(S)$ is the smallest σ-algebra containing the Borel subsets of S. For $s \in S$, we also define $r(x,s) = \int_\Theta \int_Y r(x,y,y') q(y'|x,y,\theta) \nu(dy') \lambda(d\theta)$. For the remainder of this paper, we shall be using r to refer to the expected reward function which now maps $X \times S$ into $\Re$.

Finally, we impose some regularity conditions on the various components:

A 1 *$r: X \times S \to \Re$ is jointly continuous on $X \times Y$.*

A 2 *For all $(x,y,\theta) \in X \times Y \times \Theta$, $supp(q(.|x,y,\theta)) = Y$.*

A 3 *$q: Y \times X \times Y \times \Theta \to \Re_+$ is jointly continuous in all its arguments.*

Our first result is the continuity of the Bayes map. This result is also stated and proved in Easley and Kiefer [14]. We adopt a slightly different method of proof by appealing directly to the standard integration to the limit result in Billingsley [3]:

Lemma 2.1 *$\mathcal{B}: Y \times X \times Y \times P(\Theta) \to P(\Theta)$ is a continuous map.* ○

Proof: Let $(y'_n, x_n, y_n, \lambda_n) \to (y', x, y, \lambda)$ in the appropriate product topology. Define $k^{(n)}(\theta) = q(y'_n|x_n, y_n, \theta)$, and $k(\theta) = q(y'|x, y, \theta)$. By **A.3**, $k^{(n)}(\theta_n) \to k(\theta)$ for all sequences $\theta_n \to \theta$. Since $\lambda_n \Rightarrow \lambda$ in the weak topology, Theorem 5.5 in Billingsley [3] implies that

$$\lim_{n\uparrow\infty} \int_\Theta q(y'_n|x_n, y_n, \theta)\lambda_n(d\theta) = \lim_{n\uparrow\infty} \int_\Theta k^{(n)}(\theta)\lambda_n(d\theta)$$

$$= \int_\Theta k(\theta)\lambda(d\theta) = \int_\Theta q(y'|x, y, \theta)\lambda(d\theta)$$

Obviously, this also holds when the integral is evaluated over any set $A \in \mathcal{F}(\Theta)$ with $\lambda(A) > 0$. From the form of the Bayes updating rule (2.1), taking the limits in (2.1) establishes the result. ■

Lemma 2.1 enables us to establish the following result regarding the weak continuity of the transition on the augmented state space.

Lemma 2.2 *Q is weakly continuous. That is, if $(x_n, s_n) \in X \times S$ converges to (x, s) in the product topology, then $Q(.|x_n, s_n) \Rightarrow T(.|x, s)$.* ○

Proof: Let $h: S \to \Re$ be continuous (and hence bounded by the compactness of S). Then we need to show that

$$\int_S h(\hat{s})Q(d\hat{s}|x_n, s_n) \to \int_S h(\hat{s})Q(d\hat{s}|x, s)$$

the L.H.S. may be written as

$$\int_\Theta \int_Y h(\hat{y}, \mathcal{B}(\hat{y}, x_n, y_n, \lambda_n))q(\hat{y}|x_n, y_n, \theta)\nu(d\hat{y})\lambda_n(d\theta)$$

and using the continuity of h, Lemma 2.1, **A.3**, and the weak convergence of λ_n to λ, another appeal to Theorem 5.5 in Billingsley [3] completes the proof. ■

Now, the quintuple (S, X, Q, r, β) defines a standard stochastic dynamic programming problem with compact state space S, compact action space X, weakly continuous transition Q, continuous reward function r, and a discount factor $\beta \in [0, 1]$. The main result of Maitra [29] directly applies and we have the main theorem of this section.

Theorem 2.1 *i. There exists a continuous function* $V: S \to \Re$ *such that for all* $s \in S$

$$V(s) = \max_{x \in X} \{r(x,s) + \beta \int_S V(s') Q(ds'|x,s)\} \tag{2.4}$$

ii. The set of maximizers in equation (2.4) *defines an upper hemicontinuous correspondence* $G: S \to X$. *Further,* G *admits a measurable selection* $g: S \to X$. ○

Proof: See Maitra [29].

In words, Theorem 2.3 says that the agent can do no better at maximizing expected sum of discounted rewards than by taking the action $g(s)$ whenever the state is given by $s \in S$. We refer to g in the rest of this paper as a stationary optimal policy function. Since g is an optimal policy function if and only if g is a measurable selection from G, in general, many stationary optimal policy functions may exist. However, all of them will give the same expected reward over the infinite horizon; namely that specified by V. Therefore, we assume from now on that the agent has selected some function g that we shall henceforth hold fixed.

3 The Martingale Property of Beliefs: Subjective Convergence

When the decision maker picks an optimal stationary policy $g: S \to A$, a Markov process arises on S, with the conditional distribution of s_{t+1} given s_t given by

$$Pr\{s_{t+1} \in A | s_t\} = \int_A Q(d\hat{s}|g(s_t), s_t)$$

The measurability of Q in s_t follows from our assumptions and the measurability of g. However, unlike the case of the repeated statistical decision problem, beliefs alone (which are only one component of the augmented state space) need not follow a Markov process, and in general, they will depend on the entire past of beliefs. On the other hand, since it is still true that the decision maker cannot expect his beliefs to change in any systematic way, beliefs will still follow a martingale with respect to the appropriate sigma-filtration. The purpose of this section is to demonstrate this, then use the martingale convergence theorem to show that beliefs converge to a limit belief, and to study the properties of the limit belief. The treatment in this section is simply attempting to demonstrate that the results of Easley and Kiefer [12, 14] still hold in this more general framework where the Markovian property for beliefs has been lost.

We start by introducing some new notation that will be useful in constructing the sigma-filtration with respect to which the sequence of beliefs will be shown to have the

Martingale property. Let $Z = X \times Y$, $Z_t = \times_{\tau=1}^{t} Z$, $Z_{-t} = \times_{\tau=t+1}^{\infty} Z$, and $Z_\infty = \times_{\tau=1}^{\infty} Z$. Define the sub sigma fields $\hat{Z}_t$ on Z_∞ by $\hat{Z}_t = \{A \subset Z_\infty : A = C \times Z_{-t}; C \in \mathcal{F}(Z_t)\}$; where as usual $\mathcal{F}(Z_t)$ is the smallest sigma field generated by the Borel subsets of Z_t. Also, let $\hat{Z}_\infty = \bigvee_{t=1}^{\infty} \hat{Z}_t$. Note that $\hat{Z}_t$ is the set of measurable events observable after t periods.[5]

For $\theta \in \Theta$ and $A \in \hat{Z}_t$, let $Pr_\theta(A)$ be the probability under the parameter θ of observing $(z_1, z_2, \ldots, z_t) \in C$, where $A = C \times Z_{-t}$. Pr_θ is calculable from knowledge of the underlying functions and the optimal policy g, and is measurable in θ since g is measurable. The measurable space of sample paths $((\Theta \times Z_\infty), \sigma(\mathcal{F}(\Theta) \times \hat{Z}_\infty))$ may now be endowed with the probability measure P_{λ_0} which is the extension of $Pr(D \times A) = \int_D Pr_\theta(A) \lambda_0(d\theta)$ for $D \in \mathcal{F}(\Theta)$, and $A \in \hat{Z}_\infty$.

The probability at time t that the agent places on the set $D \in \mathcal{F}(\Theta)$ can then simply be written as $\lambda_t(D) = E[I_{D \times Z_\infty} | Z_{t-1}^*]$ where $Z_{t-1}^* = \hat{Z}_{t-1} \times \{\emptyset, \Theta\}$. It is clear by construction that $Z_{t-1}^* \uparrow Z_\infty^*$ where as usual, $Z_\infty^* = \bigvee_{t=1}^{\infty} Z_t^*$, and it follows by Billingsley [4, example 35.5, p. 410] that $\lambda_t(D)$ is a martingale. As $I_{D \times Z_\infty}$ is an indicator random variable, $\lambda_t(D) \leq 1$ a.s. for all t. Therefore, by the martingale convergence theorem (Billingsley [4, p.416]), there exists a limit random variable denoted by λ_∞ such that $\lambda_t \Rightarrow \lambda_\infty$ a.s.$[P_{\lambda_0}]$.

This result is the same as that achieved in Easley and Kiefer [12, 14], and as in their case, it is true that in general, it need not be the case that $\lambda_\infty = \delta_{\theta^*}$ a.s.$[P_{\lambda_0}]$. As stated less rigorously in the introduction, the current state y influences next period's state and may hence force learning regardless of what the agent decides to do, (i.e. forces consistency results in the case of passive learning to come in action). In the repeated framework literature, full learning has been shown to hold under suitable conditions about how fast agents discount the future (for instance see Easley and Kiefer [12, 14]). All that we require here is the very mild identifiability condition **I.1** as well as the assumption **I.2** of the continuity of the optimal policy function g.

I.1 For all priors λ with support of more than one point, $\exists$ sets $A, B \in \mathcal{F}(\Theta)$ with $\lambda(A) > 0$ and $\lambda(B) > 0$ such that

$$q(y'|y, g(y, \lambda), \hat{\theta}) \neq q(y'|y, g(y, \lambda), \tilde{\theta})$$

for all $\hat{\theta} \in A$ and $\tilde{\theta} \in B$, and for some $(y, y') \in Y \times Y$.

I.2 The optimal policy selection g is continuous in λ.

Under that very mild condition we prove the full learning result:

[5] Note the slight error here in Easley and Kiefer [12, 14]. The appropriate σ-field after t-periods in their case should be the σ-field generated by the cylinder sets of the form $(C \times Z_{-t})$ for C a Borel subset of $(Z_{t-1} \times X)$, not as they claim, $(C \times Z_{-t})$ for C a Borel subset of Z_t.

Lemma 3.1 *Under conditions* **I.1** *and* **I.2**, *the limit belief* λ_∞ *achieved from the martingale convergence theorem is equal to* δ_{θ^*}. ○

Proof: Let the limit belief for simplicity of notation be λ. Choose a set A such that $\lambda(\delta A) = 0$, then

$$\lambda_{t+1}(A) = \frac{\int_A q(y_{t+1}|y_t, g(y_t, \lambda_t), \hat{\theta}).\lambda_t(d\hat{\theta})}{\int_\Theta q(y_{t+1}|y_t, g(y_t, \lambda_t), \theta).\lambda_t(d\theta)} \tag{3.5}$$

taking limits of both sides as $t \uparrow \infty$, and by the continuity of q and the continuity of g (assumption **I.2**), it is readily seen that the limit $\lambda_\infty = \lambda$ is invariant to the Bayesian updating rule $\mathcal{B}$. Now, by that invariance,

$$\lambda(A) = \frac{\int_A q(y'|y, g(y, \lambda), \hat{\theta}).\lambda(d\hat{\theta})}{\int_\Theta q(y'|g(y, \lambda), \theta).\lambda(d\theta)} \tag{3.6}$$

which can be rewritten as

$$\int_A 1.\lambda(d\hat{\theta}) = \int_A \frac{q(y'|y, g(y, \lambda, \hat{\theta})}{\int_\Theta q(y'|y, g(y, \lambda), \theta).\lambda(d\theta)}.\lambda(d\hat{\theta}) \tag{3.7}$$

and equality (3.2) has to hold for all sets $A \in \mathcal{F}(\Theta)$, and hence

$$\begin{aligned} q(y'|y, g(y, \lambda), \hat{\theta}) &= \int_\Theta q(y'|y, g(y, \lambda, \theta).\lambda(d\theta) \\ &= K(\lambda, y, y') \end{aligned} \tag{3.8}$$

almost surely $[\lambda]$; where $K(\lambda, y, y')$ is a constant that depends only on the shown arguments.

Now, by **A.2** and **A.3** (the assumptions of full support and continuity of the transition q), there is a positive probability of getting within a small neighborhood of $(y, y') \in Y \times Y$ of assumption **I.1**. Hence, by assumption **I.1**, the only possible almost sure $[P_{\lambda_0}]$ limit belief from the martingale convergence theorem used above has to have a one-atom support. But by the continuity of the Bayes updating map $\mathcal{B}$ and the full support assumption of the initial prior λ_0, convergence to a point mass at a "wrong" single point is a $[P_{\lambda_0}]$ measure zero event, and the warranted result is proved. ■

Assumption **I.2** was also used in the literature [14] and justified possibly on the basis of more fundamental assumptions including the strict concavity of the payoff function r. It is clear from the setup and proof of Lemma 3.1 that the same result can be obtained in the repeated framework of [12, 14] using the same two assumptions. It must be noted however, that assumption **I.1** in the repeated framework is extremely

restrictive since it can be violated by the agent choosing to play the same action forever, and hence not learn. In our sequential framework, however, condition **I.1** can only be violated if the optimal policy selection g is such that $g(y, \lambda)$ and y will move in such a perfectly offsetting way that the true parameter will be unidentifiable. We therefore conclude that the result of Lemma 3.1 is more useful in the sequential framework since its conditions are much milder in that framework.

4 Limit Beliefs from the Perspective of an Informed Objective Observer

In the previous section, we showed that in this generalized sequential statistical decision problem, with the augmented state space, we can still achieve the results of Easley and Kiefer [12, 14] of convergence of beliefs almost surely with respect to the probability measure on the space of sample paths generated by the decision maker's prior belief on Θ. As we described this result in the introduction, it is stating that the agents think that they will converge to a limit belief a.s. This is clearly not the interesting question to the economist who knows the true value of $\theta^\star$ (since he wrote the model), and can imagine a large class of possible prior beliefs λ_0 with which he may endow the agents in his model. This section introduces an alternative formalization of the modelling process. To study the economist's problem of assignment of priors to the agents, and study the evolution of measures in the model, we need to study the evolution of measures on the augmented state space. We leave that to the next section. In this section, we just introduce an objective observer who is equipped with the knowledge of the true parameter $\theta^\star$ and a prior on $P(\Theta)$, denoted by μ_0, where $\mu_0(A) = Pr\{\lambda_0 \in A\}$. The objective observer sits on the margin of $P(\Theta)$ in our augmented state space, and observes the evolution of μ_t.

Before we proceed with the study of what we may call 'objective convergence' (from the perspective of the informed observer with knowledge of the true $\theta^\star$) in general spaces, we note a simple though surprising result. In a finite or countable parameter space with the agents being assigned proper priors with full support (i.e. there is a positive probability assigned to each of the points in the parameter space), the subjective convergence result immediately yields an objective convergence result. It is clear, however, that such a convergence result is restricted to the case with an atomic parameter space since with continuous parameter spaces and atomless priors, the proof below will not work since any single $\theta \in \Theta$ (in particular $\theta^\star$) is of λ_0 measure zero, and $P_{\theta\star}$ (the probability under the true distribution determined by $\theta^\star$ of convergence of beliefs can be anything.

Theorem 4.1 *Let the parameter space be $\Theta = \{\theta_1, \theta_2, ...\}$, and $\theta^\star = \theta_k$ for some $k \in \mathcal{N}$. Then, under the true $\theta^\star$, the sequence of beliefs starting from any given λ_0*

converges a.s. $[P_{\theta^*}]$. ○

Proof Let $\tilde{Z} \in \hat{Z}_\infty$ be the set of sample paths for which beliefs λ_t starting from any given λ_0 converge. Then by the martingale convergence theorem result cited prior to Lemma 3.1 (subjective convergence of beliefs a.s. $[P_{\lambda_0}]$), and by the definition of P_{λ_0}, it follows that

$$P_{\lambda_0}(\tilde{Z}) = \int_\Theta Pr_\theta(\tilde{Z})\ \lambda_0(d\theta) = 1$$

which in our case with a countable parameter space (and writing λ_0 as a vector $\{p_1, p_2, ...\}$ of probabilities corresponding to $\{\theta_1, \theta_2, ...\}$) reduces to

$$\sum_{k \in \mathcal{N}} Pr_{\theta_k}(\tilde{Z}) p_k = 1 \tag{4.9}$$

but since we assume that λ_0 is a proper prior; i.e. $p_k > 0\ \forall k \in \mathcal{N}$ and $\sum_{k \in \mathcal{N}} p_k = 1$, it follows that the last equality above can only hold if $Pr_{\theta_k}(\tilde{Z}) = 1\ \forall k \in \mathcal{N}$. In particular, for the true θ^*, $Pr_{\theta^*}(\tilde{Z}) = 1$. ■

To study the evolution of the economist's sequence of measures in a general space, we will need to fully characterize the convergence of iterates of the relevant kernel, which is done in the next section. In this section, we ask a simpler question: Is it true that for some large T, and starting from a large (probability one) set of prior beliefs, all agents will be within some small neighborhood of some limit belief? In other words, given any particular value λ_∞, is it true that there exists an ϵ - neighborhood of λ_∞ and a time $T = T(\epsilon)$ such that $\forall t \geq T$, $\mu_t(N_\epsilon(\lambda_\infty)) = 1$. If such a result was achieved, it would justify the use of a single prior for all agents by means of an argument that the economy has been running long enough, and so almost all agents are arbitrarily close to some limit belief (an even more ambitious hope would be to also show that $\lambda_\infty = \delta_{\theta^*}$, and then the argument will state that 'the agents would have learnt the structure of the model'). Unfortunately, the result we achieve is a negative one. It turns out that given any fixed element (belief) in $P(\Theta)$, there will be positive mass outside a small neighborhood of that element for all time periods. With hindsight, this is not surprising under our assumptions of full support of μ_0, since starting with priors that are far enough from the desired neighborhood, the elements of some non-trivial neighborhood of that far away prior would not have converged by any given time period. Indeed, this is the idea and motivation of the proof provided below. Even though this result seems very natural once stated, we believe that the assumptions we provide are reasonable, and that this offers -to our knowledge- the first rigorous criticism of the loose approximation argument commonly used in economics to justify various versions of the rational expectations hypothesis.

The type of result we get in this section contrasts with the type of results sought and achieved in Feldman [20] based on inconsistency results of Bayesian inference in Freedman [22, 23] and Diaconis and Freedman [9] in more than one way. First, the results of Freedman and Freedman and Diaconis, and hence those of Feldman, are for the case with an infinite dimensional parameter space; whereas we are aiming at results that hold from the perspective of a modelling economist (or in this section, the objective observer who only looks at the margin along $P(\Theta)$) even in the standard finite dimensional parameter space. Second, the results in Freedman, and thus those in Feldman, show that convergence does not hold for a "large" (formally, residual) class of priors, but that class may very well be (as recognized by the authors) of measure zero. We are seeking a more measure theoretic result, since we want our economist to be no less intelligent than the agents he models; if they are Bayesian, so is he. Third, there is a discussion in Feldman [20] where he argues that an economist will endow his agents with priors that will converge, i.e. where learning is possible. However, such a requirement (ignoring the circular nature of the underlying argument) puts an immense burden on the shoulders of the modeler since in complicated models it may be prohibitively hard to show whether convergence and/or learning occurs. This point will be discussed further in the following sections. On the other hand, assuming that the economist solves the model for the class of priors that lead to learning and then restricts the class of priors with which he endows his agents to that class seems to us to be rather ad hoc. Such a procedure will be simply pushing the ad hoc assumption of rational expectations where the true distribution is known from period zero one step further. There is no epistemological gain from such a procedure, and we might just as well assume that $\lambda_0 = \delta_{\theta^*}$.

Now, we start with an economist who will endow agents with initial beliefs λ_0 drawn at random from a prior $\mu_0 \in P(P(\Theta))$, and inform the objective observer about the measure from which he drew those priors, as well as the true θ^*. Since in the previous section, we imposed the restriction that λ_0 has full support on Θ, we extend that characterization by assuming that the support of μ_0 is the set of λ's that have full support.

A 4 $supp(\mu_0) = \{\lambda \in P(\Theta): supp(\lambda) = \Theta\}$.

We also make the natural assumption that the economist is aware of the optimal policy function $g: Y \times P(\Theta) \to X$ that he lets his agents choose from the class G of section 2, and that he again informs his objective observer of that function. There is clearly no loss of generality in making this assumption since the economist is the one who assigns such a choice in the first place, and since he knows that such a choice g is stationary.[6]

[6]Another interesting issue, which we do not consider in this paper, arises here: Could different selections of g give different convergence behavior in the next section? and if so, how?.

We define for each time t the probability that $\lambda_t \in A$ by $\mu_t(A)$ where $A \in \mathcal{F}(P(\Theta))$. For example, for the first period,

$$\mu_1(A) = Pr\{(y', y) \in Y \times Y : \mathcal{B}(y', y, g(y, \lambda), \lambda) \in A\}$$

$$= \int_{P(\Theta)} \int_{\mathcal{B}^{-1}(A;y,\lambda)} \int_{Y} q(y'|y, g(y, \lambda), \theta^*)\nu(dy)\nu(dy')\mu_0(d\lambda)$$

We have implicitly assumed in the above transition that the initial value of y is drawn from a distribution on Y that is uniform w.r.t ν. Everything that follows holds equally well for a given fixed initial value y_0. And given the economist's prior at period t of where λ_t of a typical agent will be, he gets a posterior

$$\mu_{t+1}(A) = \int_{P(\Theta)} \int_{\mathcal{B}^{-1}(A;y,\lambda)} q(y'|y, g(y, \lambda), \theta^*)\nu(dy')\mu_t(d\lambda)$$

We ask the question posed above about the value of μ_t of a neighborhood of any given limit belief, and our result is:

Theorem 4.2 *Given any fixed value of* λ_∞, $\forall t$, $\exists \epsilon > 0$, *such that*

$$\mu_t(N_\epsilon(\lambda_\infty)) < 1$$

○

Proof: For a choice of 'small' numbers $\epsilon, \gamma > 0$, pick a $\theta \in \Theta$ such that $\|\lambda_\infty - \delta_\theta\| \geq \epsilon + 2\gamma$; where $\|.\|$ here refers to the metric of the weak topology. (Such a choice is always possible since we can choose a θ such that when we consider the distance between the integrals of a sequence of functions converging to δ_θ, then their integral with respect to the measure λ_∞ will be converging to zero and the integral with respect to the measure δ_θ will be converging to unity.) Denote the open γ neighborhood of δ_θ by $N_\gamma(\delta_\theta) \subset P(\Theta)$. We shall show that for all $t \geq 0$, $\mu_t(N_\gamma(\delta_\theta)) > 0$, thus proving the theorem.

Now, by assumption **A.4**, μ_0 has full support on the set of λ's that themselves have full support. Hence, by the argument in the previous paragraph, we can easily see that there is a choice of $\epsilon, \gamma > 0$ such that the two neighborhoods $N_\epsilon(\lambda_\infty)$ and $N_\gamma(\delta_\theta)$ are disjoint and $\mu_0(N_\gamma(\delta_\theta)) > 0$. By induction, it suffices to show that if $\mu_t(N_\gamma(\delta_\theta)) > 0$, then the same is true for μ_{t+1}. For ease of notation, let $A^* = N_\gamma(\delta_\theta)$. By definition,

$$\mu_{t+1}(A^*) = \int_{P(\Theta)} \int_{\mathcal{B}^{-1}(A^*;y_t,\lambda)} q(y'|y_t, g(y_t, \lambda), \theta^*)\nu(dy')\mu_t(d\lambda) \qquad (4.10)$$

By the assumption **A.2** that q has full support, $q(.|y_t, g(\lambda, y_t), \theta^*) > 0$. Since $\mathcal{B}$ is continuous on $Y \times X \times Y \times P(\Theta)$ (by Lemma 2.1), it follows that it is continuous in y' for a fixed augmented state (y_t, λ). Hence, $\mathcal{B}^{-1}(A^*; y_t, \lambda)$ is open in Y. It remains now to show that this set is non-empty. But that follows immediately since all Y maps δ_θ into itself under $\mathcal{B}$; hence, there exists a non-empty neighborhood in Y that maps into A^*. Hence, $\nu(\mathcal{B}^{-1}(A^*; y_t, \lambda)) > 0$, which establishes that all the sets over which the integral (4.1) is evaluated are of positive measure, and the integrand is positive over the entire state space, which establishes that $\mu_{t+1}(A^*) > 0$. This proves the result of the theorem. ■

As a special case, Theorem 4.2 tells us that there is a positive probability that almost sure full learning does not occur in finite time. This translates into the following obvious corollary.

Corollary 4.1

$$\mu_t(N_\epsilon(\delta_{\theta^*})) < 1$$

○

Proof: In the proof of Theorem 4.2 replace λ_∞ everywhere with δ_{θ^*}. ■

Notice that all we have shown in Theorem 4.2 is that the probability of almost sure convergence to any particular limit in finite time (and hence as a special case the probability of full learning in finite time) is less than 1. We did that by showing that some mass will be outside some neighborhood of that limit belief for all time. It may be the case however that the amount of mass outside such neighborhoods albeit positive for all t is converging to 0. In other words, it may still be the case that $\mu_t(N_\epsilon(\lambda_\infty)) \to 1$, and we write this as $\mu_t \Rightarrow \delta_{\lambda_\infty}$.

Even though this section shows that we cannot justify the assumption that the agents all have the same beliefs (usually the correct ones) on the basis that after enough time, all but a measure zero group are outside an arbitrarily small neighborhood of that belief, a weaker justification may be available. If we can show that $\mu_t(N_\epsilon(\lambda_\infty)) \to 1$, then the argument will be verbally phrased to say that beyond some point in time, an arbitrarily large mass of the agents (strictly less than 1) will be in an arbitrarily small neighborhood of the limit belief. This is a significantly weaker justification since there are two parameters being driven close to zero (in the proof of Theorem 4.2 those are ϵ and $\mu_t(N_\epsilon^c(\lambda_\infty))$). It is, however, one that may not be mathematically unfounded in all cases as the stronger version has been proven to be. To study the convergence of measures in the model of section 3, we need to use different tools, which we introduce in the next section.

5 Convergence of Measures on the Augmented State Space: Harris Ergodicity of the Transition Kernel

We now pick up the analysis where we left it in section 3. We know that the process λ_t does not by itself form a Markov process, and we need to study the convergence of measures in the Markov process on the augmented state space $Y \times P(\Theta)$. In section 3, we considered the transition kernel on that augmented state space from the view point of the individual agent with a prior λ given by:

$$Q(A|s) = \int_{\Theta} \int_{Y} I_A(y', \mathcal{B}(y', g(s), s)) q(y'|y, g(s), \theta) \nu(dy') \lambda(d\theta)$$

We shall continue to let the individuals solve their optimization problem as in section 3. The convergence of beliefs, however, will be studied from the view point of the economist with μ_0 as the prior on priors introduced in the previous section. The relevant stochastic kernel on the measurable augmented state space $(S, \mathcal{F}(S)) = (Y \times P(\Theta), \mathcal{F}(Y \times P(\Theta)))$, starting from the initial measure $\nu \otimes \mu$ is going to be different from Q. We think of measures on S at each time period of the form $\nu_t \otimes \mu_t$, and consider the Markov kernel

$$T(A|s) = \int_{P(\Theta)} \int_{Y} I_A(y', \mathcal{B}(y', g(s), s)) q(y'|y, g(s), \theta^*) \nu(dy') \mu(d\lambda)$$

We first notice the similarity between the transition from μ_t to μ_{t+1} and that of the cross measure $\nu_t \otimes \mu_t$. This is not surprising at all since $\nu_{t+1} = \int_{P(\Theta)} \int_Y q(y'|y_t, g(y_t, \lambda), \theta^*) \nu(dy') \mu_t(d\lambda)$ which is purely determined by y_t and μ_t. The kernel T satisfies the properties of a stochastic Markov kernel; i.e. $\forall A \in \mathcal{F}(S), T(A|.)$ is measurable, $\forall s = (\lambda, y) \in S, T(.|s)$ is a measure on $(S, \mathcal{F}(S))$, and of course, $\forall s \in S, T(S|s) = 1$. We think of iterates $T^{(n)}$ of T according to the iterative formulation $T^{(0)} = I$, and

$$T^{(n)}(A|s) = TT^{(n-1)}(A|s) = \int_S T^{(n-1)}(d\hat{s}|s) T(A|\hat{s})$$

The one period transition defined on any probability measure $\phi \in P(\Theta)$ by $\phi T(A) = \int_S \phi(ds) T(A|s)$ We say that the set $A \in \mathcal{F}(S)$ is attainable or reachable from s (written as $s \rightarrow A$) if for some $n \geq 1$, $T^{(n)}(A|s) > 0$. For any probability measure $\phi \in P(S)$, we say that the kernel T is ϕ-irreducible if for all sets A with $\phi(A) > 0$, and for all $s \in S$, $s \rightarrow A$ (and in this case, we call ϕ an irreducibility measure of T). If T is ϕ-irreducible with respect to some probability measure ϕ, then we say that T is irreducible.[7] We

[7]Perhaps the notion of irreducibility is more familiar (or standard) in the finite case where T is a probability transition matrix, and it is associated with the existence of a single ergodic class. The Markov chain is then said to be ergodic, and that is equivalent to the convergence from any initial probability vector to the unique stationary distribution. A similar result in more general spaces is what we present in this section.

define a *maximal irreducibility measure* as an irreducibility measure with respect to which all other irreducibility measures are absolutely continuous. The following result about the existence of maximal irreducibility measures can be found in Nummelin [33]

Proposition 5.1 *Suppose that T is ϕ-irreducible for some $\phi \in P(S)$, then*

i. There exists a maximal irreducibility measure for T.

ii. An irreducibility measure ψ is maximal if and only if $\psi T \ll \psi$. ○

Proof: See Nummelin [33, Proposition 2.4, p.13].

Now, we are ready for defining characteristics of T that will guarantee convergence of iterates of $T^{(n)}$. For a fuller discussion and derivation of the result of interest, the reader is again referred to Nummelin [33]. Intuitively, the conditions for convergence of iterates of $T^{(n)}$ is very similar to the case with a unidimensional discrete state space where T is a probability transition matrix, and results on convergence to a unique stationary probability vector are standard. With a more general state space, the Hopf decomposition theorem tells us that the resulting Markov chain $\{S_n\}$ can result in dividing the space S into a dissipative part S_d (which is a countable union of set A_i such that $Pr\{S_n \in A_i \quad i.o. | S_0 = s\} = 0; \forall s \in S$), and a conservative part S_c such that there exists a probability measure ϕ whereby for all sets A with $\phi(A) > 0$, and for $[\phi]$-almost all $s \in A$, $Pr\{S_n \in A \quad i.o. | S_0 = s\} = 1$. The conservative part of S is hence what we are after, since as in the discrete case, when all 'states' get visited infinitely often with probability one, we get the standard result that the proportion of time the process is in each of those states converges to the probability of that state under the unique stationary distribution. It is a more general result of that nature that we need. Conditions to get such a result are a little more complicated than the standard conditions on probability transition matrices for the unidimensional discrete case. We now define the conditions for such convergence, the definitions and major cited result here are again due to Nummelin [33].

Definition 5.1 *The kernel T is aperiodic if there does not exist a sequence $(T_0, T_1, \ldots, T_{m-1})$; $m > 1$ of nonempty set forming a period for the resulting Markov process S_n; i.e. s.t. for all $s \in T_i$, $T(T_j^c | s) = 0$ for $j = i + 1 \pmod m$.*

Definition 5.2 *An irreducible Markov chain $\{S_n\}$ (i.e. one that is generated by an irreducible kernel T) is said to be Harris recurrent if for all $s \in S$, $Pr\{S_n \in A \quad i.o. | S_0 = s\}$ for all sets A s.t. $\psi(A) > 0$, where again, ψ is a maximal irreducibility measure for T.*

Definition 5.3 *An irreducible Harris recurrent Markov chain generated by a kernel T is positive Harris recurrent T has a stationary probability measure π; i.e. $\pi T = \pi$.*

Definition 5.4 *A Markov chain $\{S_n\}$ generated by a kernel T is said to be Harris ergodic if it is aperiodic (def. 5.2), and positive Harris recurrent (defs. 5.3 & 5.4).*

The concepts specified in the above four definitions are what we need to get a result of convergence of iterates of $T^{(n)}$ much like the type of results we get for probability transition matrices that are irreducible. The following two results are the equivalent of ergodicity results in that simplified case. In the following two convergence results for iterations of the kernel, we use $\|.\|$ to be the norm of total variation.

Theorem 5.1 *If $\{S_n\}$ is an aperiodic Harris recurrent Markov chain, the for all $\phi, \psi \in P(S)$,*

$$\lim_{n\uparrow\infty} \|\phi T^{(n)} - \psi T^{(n)}\| = 0$$

○

Proof: See Nummelin [33, corollary 6.7, p.113].

In words, if we are not guaranteed that an invariant measure exists, but the process is aperiodic and Harris recurrent as described above, even though iterates of $T^{(n)}$ need not converge, starting with any two measures, their trajectories in $P(S)$ with the law of motion T will get to look more and more identical as the number of iterations increases. The next theorem tells us that with the existence of an invariant measure π, i.e. with Harris ergodicity, the iterates of T converge to that unique invariant measure, and vice versa.[8]

Theorem 5.2 *A Markov chain S_n generated by a kernel T is Harris ergodic if and only if there exists a probability measure $\pi \in P(S)$ on $(S, \mathcal{F}(S))$ such that for all $\phi \in P(S)$,*

$$\lim_{n\uparrow\infty} \|\phi T^{(n)} - \pi\| = 0$$

in other words, this can be written as: for all $s \in S$

$$\lim_{n\uparrow\infty} \|T^{(n)}(.|s) - \pi\| = 0$$

○

[8] We use the conditions for Harris ergodicity instead of the more familiar ergodicity for obvious reasons. Ergodic theorems of the Birkhoff and von Neumann type only guarantee the Cesàro convergence of measures to the unique invariant measure. For the learning result, however, we need strong convergence of the measures on our augmented state space. That, as stated in the next theorem, is equivalent to the stronger notion of Harris ergodicity. Of course all we really need is weak convergence, but that corresponds to mixing conditions that are much harder to characterize, and so we content ourselves with this perhaps over-restrictive condition.

Proof: See Nummelin [33, proposition 6.3, p.114].

Now, comparing the convergence of iterates of T in this section with the results of the previous section (which was mainly concerned with convergence of iterates of Q along the dimension of $P(\Theta)$). The main result of section 4 follows immediately when we notice that every set of positive initial measure in $Y \times P(\Theta)$ is reachable from some other set of positive measure, which yields the same result that the complement of a small neighborhood of any particular measure λ_∞ in $P(\Theta)$ will have positive measure at all time periods. The remaining question was whether or not μ_t converged to any measure, and if so, whether or not that measure was δ_{λ_∞}. We argued that to look at convergence of μ_t, we need to look at convergence of iterates of T, (for a discussion of the type of convergence we consider see previous footnote) and Theorem 5.2 above states that such convergence takes place if and only if T is Harris ergodic; and if the maximal irreducibility measure, and hence the limit measure, have full support, the rational expectations hypothesis will be rejected at all levels. Instead of proceeding to offer sufficient conditions to insure that such a result does or does not achieve, and then examining how stringent those conditions are, we content ourselves in this section with demonstrating -by means of trivial examples- that under the conditions imposed so far, anything can happen. In other words, learning, non-learning, or partial learning may all occur.

It is obvious that full learning may occur in the trivial case where $q(y'|y,x,\theta) = \tilde{q}(y|\theta)$; i.e. we are getting i.i.d. draws from the density q which depends only on the parameter θ. To make it painfully obvious, let $Y = (y^1, y^2)$ where $q(y^1|\theta) = \theta$. Then this is obviously a model of passive learning, and starting from any initial measure λ with full support, the Bayesian estimate of θ is equivalent asymptotically to $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} I_{y_i=y^1}$ which is consistent by the SLLN. It is also obvious in other instances that convergence does not guarantee uniformity of posteriors for different agents in the sense that $\mu_\infty \neq \delta_{\lambda_\infty}$,or if such uniformity exists, that $\lambda_\infty \neq \delta_{\theta^*}$. Absence of uniformity is obvious in the special case of $q(y'|y,x,\theta) = 1$ for all $(x,y,\theta) \in X \times Y \times \Theta$. In that case all priors stay the same, and $\mu_0 = \mu_t = \mu_\infty$ for all t, which by the full support of μ_0 rules out the uniformity of posteriors for all agents. For a more sophisticated result, let us make the assumption that q exhibits enough variation; i.e.

S.1 For some $\epsilon > 0$, and all values of θ, y, λ, and $x = g(y,\lambda)$, there exists a set $\hat{A} \subset Y$ with $\nu(\hat{A}) > 0$ s.t. $\forall y' \in \hat{A}$, $q(y'|y, g(y,\lambda),\theta) < 1 - \epsilon$, and a set $\hat{A} \subset Y$ with $\nu(\hat{A}) > 0$ s.t. $\forall y' \in \hat{A}$, $q(y'|y, g(y,\lambda),\theta) > 1 + \epsilon$.

This assumption is slightly stronger than what is needed to rule out pathological cases like the one above where q is uniform, or ones where the actions $x = g(y,\lambda)$ make q uniform so that no more learning can take place. We think that this assumption is reasonable, and one may think that under such assumption, we would achieve complete learning, or perhaps agreement on posteriors among agents. It is interesting, however,

that if we restrict attention to initial conditions in the measure-one subspace $\tilde{S} = Y \times D(\Theta)$, (where $D(\Theta)$ is the space of atomless measures on Θ with full support) to note that we get the opposite result; namely, that if T is Harris ergodic, the limit measure on S, which we call $\nu_\infty \otimes \mu_\infty$ will have full support.

Example 5.1 *Under the above conditions and assumption* **S.1**, *if T^n converges to a measure $\pi = \nu_\infty \otimes \mu_\infty$, then that limit measure has full support on $\tilde{S}$.* ○

Proof: Since T^n converges to $\pi = \nu_\infty \otimes \mu_\infty$, it follows that T is Harris ergodic by Theorem 5.7 above. Hence, there exists a maximal irreducibility measure ψ whereby for all sets $A \in Y \times D(\Theta)$ with $\psi(A) > 0$, $Pr\{S_n \in A \ i.o.\} = 1$. Now, for ψ to be a maximal irreducibility measure, all other irreducibility measures have to be absolutely continuous with respect to it. Hence, it remains to show that some irreducibility measure has full support on $\tilde{S}$. But by assumption S.1, there is a positive probability that we get a sequence of y's so that any set A with $\nu \otimes \mu$ (where $\nu \otimes \mu$ is some measure with full support) will ultimately be visited from any initial point $s \in \tilde{S}$. Hence, there exists an irreducibility measure with full support. Hence, the limit measure π must have full support. ■

In this section, we have shown that under reasonable conditions, even if measures μ_t converged, they may not converge to any particular δ_{λ_∞}, not to mention $\delta_{\delta_{\theta^*}}$. This leaves us with the conclusion that we should abandon trying to find "reasonable" sufficient conditions that insure that some form of the rational expectations hypothesis still holds, and consider any particular model (without the rational expectations assumption) on its own merit; studying how information is collected and processed optimally by agents, and how this influences asymptotic or equilibrium behavior in the economy.

6 Numerical Investigations of Harris Ergodicity in Kiefer's Monopolist Problem

As we stated earlier, the numerical investigations in this section are intended mainly as a suggestion of how to proceed where the theory abandons us. They are also very useful for outlining aspects of the model that one would not be able to see using standard methods. For the purposes of comparison, we follow exactly the model in Kiefer [25] of a monopolist maximizing his expected discounted profits while trying to learn which of two demand curves he faces. The demand curve gives a price which is distributed gaussian with mean $\alpha + \beta * q$ (where q is the quantity that the monopolist decides to produce) and variance unity. The parameters of the demand curve are allowed take one of two values $a_1 = 50$ and $b_1 = -5$, or $a_2 = 38.89$ and $b_2 = -3$. For the full description of the mathematical and numerical issues involved in solving this problem, we refer the reader to Kiefer [25]. We have reproduced Kiefer's results for a discount

factor of 0.25, and used them to further study the behavior of beliefs. As we argued in the introduction, Kiefer's numerical analysis (where he lets the monopolist choose the optimal quantity to produce given his prior, and then draws the price from the distribution that he knows is true (with $\alpha = a_1$ and $\beta = b_1$) and lets the agent update in a Bayesian fashion) is more in line with our study of objective convergence of beliefs.

Figure 1 shows the dynamics of beliefs for 50 agents whose priors (the probability of $\alpha = a_1$ and $\beta = b_1$, i.e. a number between 0 and 1) are drawn at random uniformly between 0 and 1. Figure 2 shows that after 20 iterations, almost all of the agents have converged to one of the two optimal invariant beliefs (1, the correct belief, or 1/2, the other optimal invariant belief - again see Kiefer [25] for details). We note how neighboring priors can result in very different time series behavior. This aspect could also be recognized in Figure 11 of [25] but it comes strongly into focus when we look at ensembles of agent-priors drawn from different economist-priors. Figure 3 shows that with 100 agents drawn at random, and after 20 iterations, very few of them have not yet converged to one of the optimal invariant beliefs, but seem to be on their way to convergence. The question we are interested in here is the value of the limit probability of a monopolist whose initial belief (λ_0 in the notation of this paper) is drawn at random from the economist's prior on priors (μ_0 in the notation of this paper taken to be the uniform prior over [0,1] in our numerical investigations) converging to one or the other of the invariant measures (or perhaps not converging to anything or cycling, etc. in more general cases). We therefore wish to start looking at those pictures in a different way, in terms of measures on the interval [0,1] (which happens to be a full representation of $P(\Theta)$) and their evolution over time.

This is done in Figure 4, where we started with an agent's prior on priors μ_0 being uniform over the interval [0,1], and generated 100 agents' priors λ_0 from that distribution. We then let the process run for 20 time periods as done in the previous 3 figures. Figure 4 shows clearly how all the mass ultimately gets accumulated at the two optimal invariant beliefs 1/2 and 1. One thing that we could not observe without this experiment is that the mass is almost equally allocated to those two points. In other words, an economist with no information about the initial prior that his monopolist will have (and hence by the principle of insufficient reason or maximum entropy chooses the uniform prior over priors) should be after a reasonably short period of time confident that there is a probability half of his economist having learnt the true demand curve, and a probability half of his never getting to learn it because of getting absorbed in the other optimal invariant belief. It is clear that the process is not Harris ergodic, however, since starting from different absolutely continuous (w.r.t. Lebesgue) priors on priors will yield different masses at the two optimal invariant beliefs. For instance, we demonstrate this in Figure 5 by starting from a prior on priors density of the form $f(x) = 2x$. Since that prior put more mass near the true belief of unity,

this reflected in the limiting behavior putting significantly more weight on that belief in the end. Applying a Kolmogorov-Smirnov test to the difference between the two limits obviously rejects at all reasonable significance levels, and the eyeball test here is more than sufficient. In more sophisticated models, one proceeds in the same manner, perhaps paying more attention to numerical variance reduction techniques to speed up convergence.

Figure 1
10 periods of beliefs of 50 agents drawn uniformly

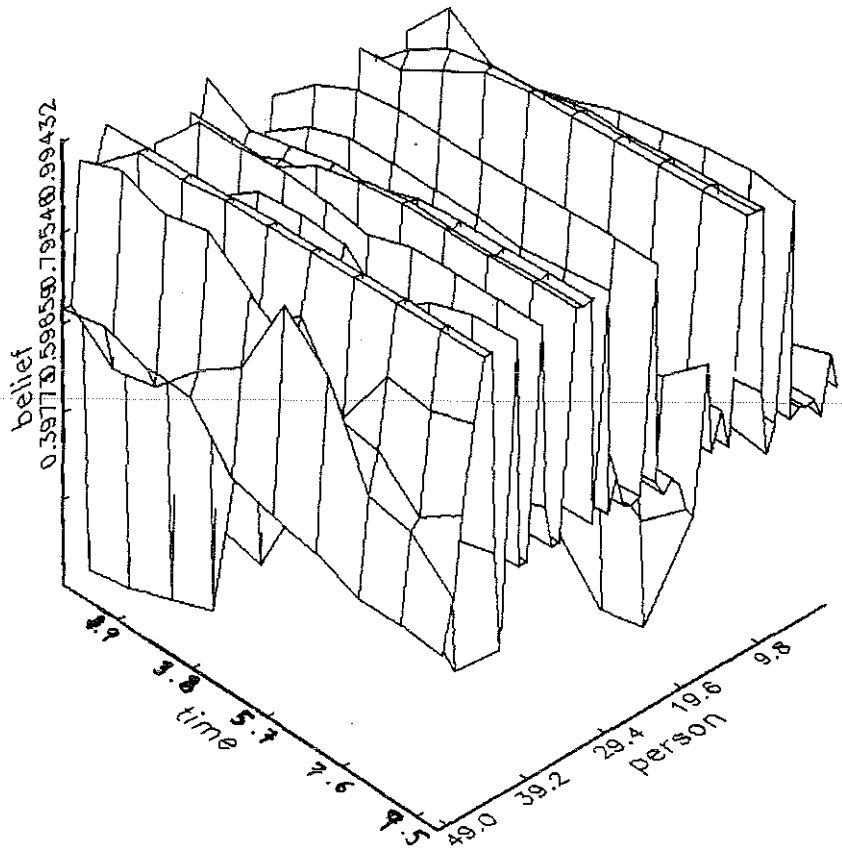


Figure 2
20 periods of beliefs of 50 agents drawn uniformly

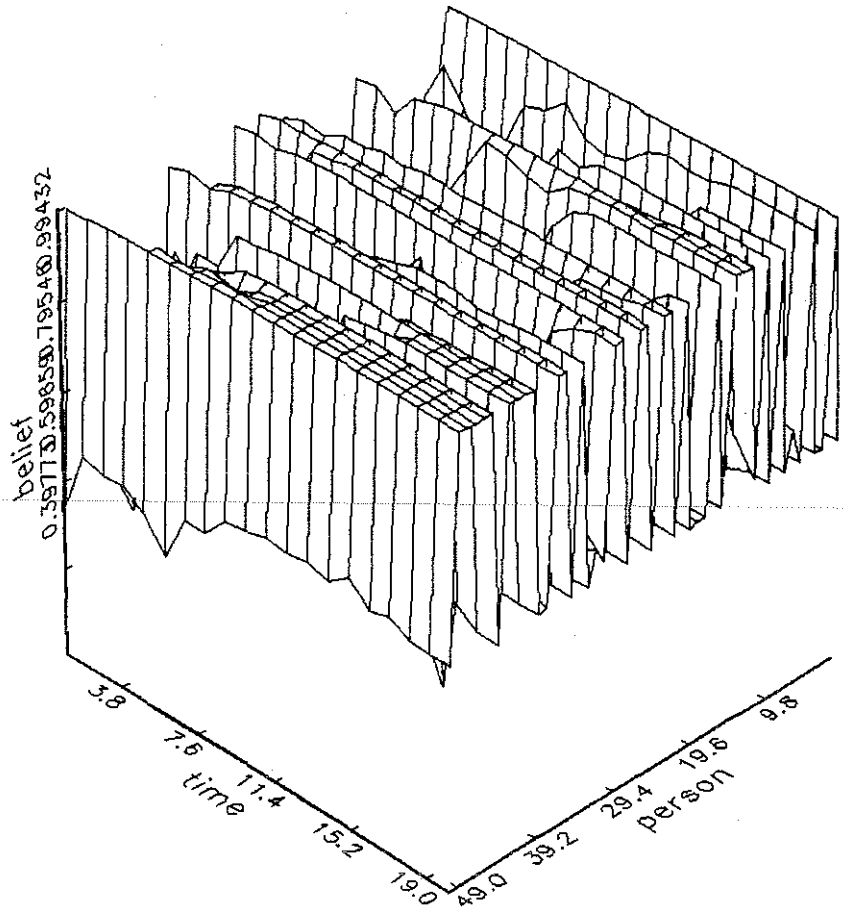


Figure 3
20 periods of beliefs of 100 agents drawn uniformly

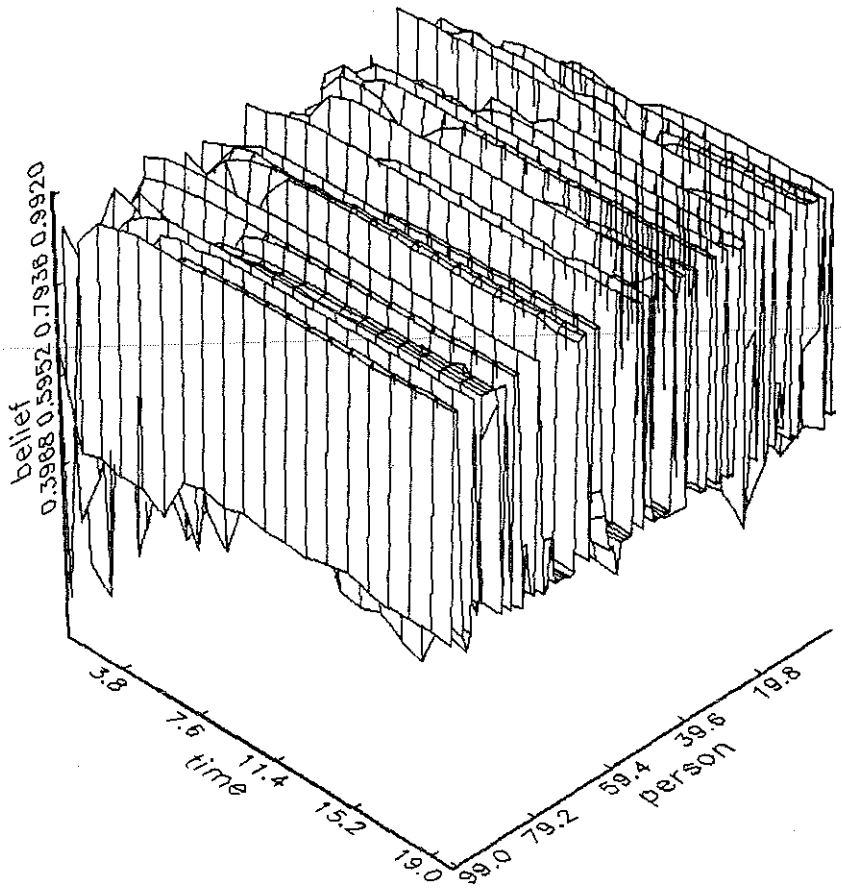


Figure 4
Evolution of beliefs from uniform prior on priors

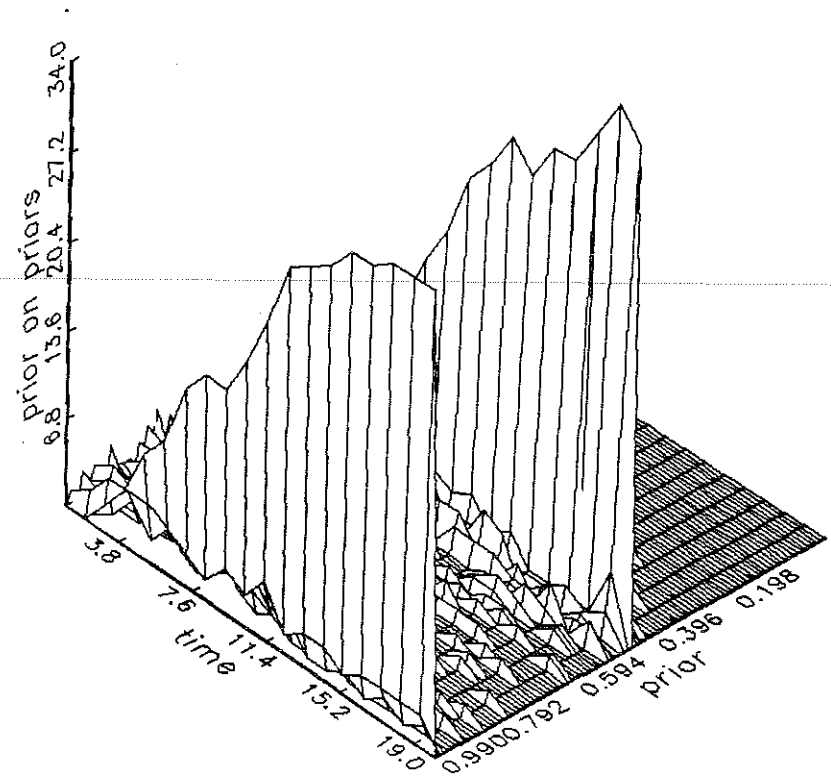
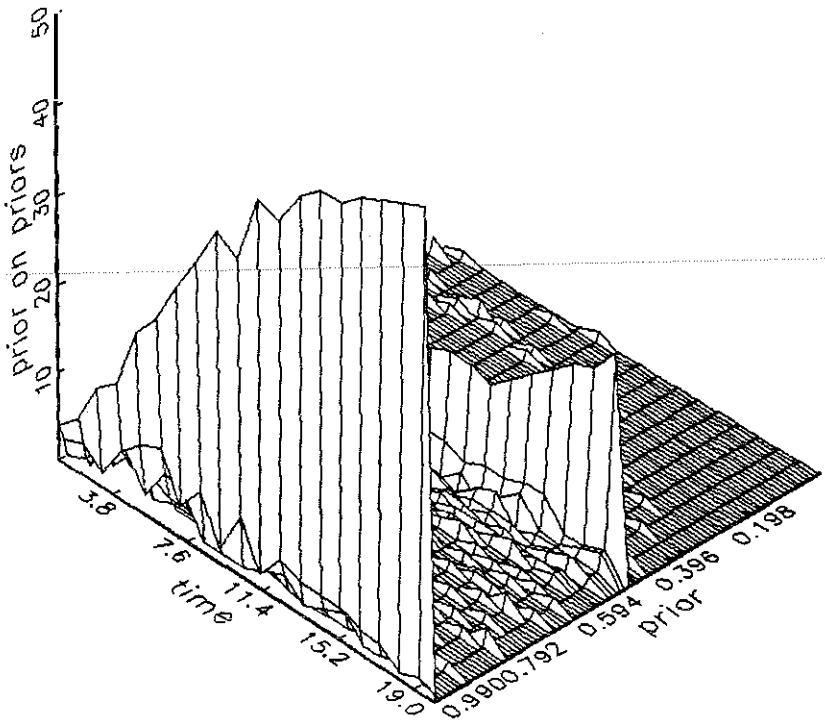


Figure 5
Evolution of beliefs from prior on priors density = 2x



References

[1] Aghion, P., P. Bolton, and B. Jullien, "Dynamic Price and Quantity Setting by a Monopolist Facing Unknown Demand", mimeo, Harvard University, 1986.

[2] Berry, D., and B. Fristedt, *Bandit Problems: Sequential Allocation of Experiments*, Chapman and Hall, London, 1985.

[3] Billingsley, P. *Convergence of Probability Measures*, John Wiley, N.Y., 1968.

[4] Billingsley, P. *Probability and Measure*, John Wiley, N.Y., 1978.

[5] Blume, L., and D. Easley, "Rational Expectations Equilibrium: An Alternative Approach", *Journal of Economic Theory*, vol.34, 1984, pp. 116-129.

[6] Blume, L., M. Bray, and D. Easley, "Introduction to the Stability of Rational Expectations Equilibrium ", *Journal of Economic Theory*, vol. 26, 1982, pp. 313-317.

[7] Brock, W., A. Marcet, J. Rust, T. Sargent " Informationally Decentralized Learning Algorithms for Finite Player, Finite Action Games of Incomplete Informatino", mimeo, 1988.

[8] Cornfeld, I., S. Fomin, and Ya. Sinai, *Ergodic Theory*, Springer-Verlag, N.Y., 1982.

[9] Diaconis, P. and D. Freedman, "On the Consistency of Bayes Estimates", (special invited paper) *Annals of Statistics*, vol. 14 1986, pp. 1-26.

[10] Domowitz, I., and M. El-Gamal, "Testing for Ergodicity in Stationary Markov Models", manuscript, University of Rochester, 1989.

[11] Dynkin, E. and A. Yushkevich, *Controlled Markov Processes*, Springer-Verlag, N.Y., 1979.

[12] Easley, D. and N. Kiefer, "Controlling a Stochastic Process with Unknown Parameters", working paper #372, Dept. of Econ., Cornell University, 1986.

[13] Easley, D. and N. Kiefer, "(p, ϵ) and ϵ- Optimality in Controlled Processes with Learning", CAE working paper, Cornell University, 1987.

[14] Easley, D. and N. Kiefer, "Optimal Learning with Endogenous Data", CAE working paper # 88-08, Cornell University, 1988.

[15] El-Gamal, M. "Real Expectations: A Harmony Theoretic Approach to Decision Making Under Uncertainty", RCER working paper # 176, University of Rochester, 1989.

[16] Feldman, M. "An Example of Convergence to Rational Expectations with Heterogeneous Beliefs", mimeo, UC Santa Barbara, 1986a.

[17] Feldman, M. "Bayesian Learning and Convergence to Rational Expectations", mimeo, UC Santa Barbara, 1986b.

[18] Feldman, M. "On the Generic Nonconvergence of Bayesian Actions and Beliefs", BEBR working paper #89-1528, University of Illinois- Urbana, Champaign, 1989.

[19] Feldman, M., and A. McLennan "Learning in a Repeated Statistical Decision Problem with Normal Disturbances", mimeo, University of Minnesota, 1989.

[20] Freedman, D. "On the Asymptotic Behavior of Bayes Estimates in the Discrete Case", *Annals of Mathematical Statistics*, vol. 34, 1963, pp. 1386-1403.

[21] Freedman, D. "On the Asymptotic Behavior of Bayes Estimates in the Discrete Case II", *Annals of Mathematical Statistics*, 1965, pp. 454-456.

[22] Jordan, J. "The Strong Consistency of the Least Squares Control rule and Parameter Estimates", manuscript, 1985.

[23] Kiefer, N. "A Value Function Arising in The Economics of Information", *Journal of Economic Dynamics and Control*, vol. 13, 1989.

[24] Kiefer, N. and Y. Nyarko, "Optimal Control of an Unknown Linear Process with Learning", working paper # 370, Dept. of Econ., Cornell University, 1986.

[25] Kihlstrom, R., L. Mirman, and A. Postlewaite, "Experimental Consumption and the 'Rothschild Effect' ", in *Bayesian Models in Economic Theory*, M. Boyer, and R. Kihlstrom (eds.), Elsevier Pubs., Amsterdam, 1984.

[26] Lai, T. and H. Robbins,"Iterated Least Squares in Multiperiod Control",*Advances in Applied Mathematics*, vol. 3, 1982.

[27] Maitra, "Dynamic Programming on Compact Metric Spaces", *Sankhya*, Series A, vol. 30, 1968, 211-221.

[28] Marcet, A., and T. Sargent, "Convergence of Least Squares Learning in Environments with Hidden State Variables and Private Information", mimeo, Carnegie-Mellon University, 1988.

[29] McLennan, A. "Price Dispersion and Incomplete Learning in the Long Run", *Journal of Economic Dynamics and Control*, vol. 7, 1984, pp. 331-347.

[30] McLennan, A. "Incomplete Learning in a Repeated Statistical Decision Problem", mimeo, University of Minnesota, 1987.

[31] Nummelin, E. *General Irreducible Markov Chains and Non-negative Operators*, Cambridge Univ. Press, Cambridge, 1984.

[32] Parthasarathy, K. *Probability Measures on Metric Spaces*, Academic Press, London, 1967.

[33] Pollard, D. *Convergence of Stochastic Processes*, Springer-Verlag, N.Y., 1984.

[34] Rothschild, M. "A Two-armed Bandit Theory of Market Pricing", *Journal of Economic Theory*, vol. 9, 1974, pp. 185-202.

[35] Shiryayev, A. *Probability*, Springer-Verlag, N.Y., 1984.

[36] Walters, P. *An Introduction to Ergodic Theory*, Springer-Verlag, N.Y., 1982.