On the Generic Nonconvergence
of Bayesian Actions and Beliefs

Mark Feldman

BEBR

On the Generic Nonconvergence of Bayesian Actions and Beliefs

Mark Feldman, Assistant Professor
Department of Economics

ABSTRACT

Many authors have recently modeled learning by Bayesian decision-makers and by appeal to the Martingale Convergence Theorem proven that posterior beliefs converge a.s. from the point of view of the decision-maker. This paper emphasizes that this is not equivalent to a.s. convergence from the point of view of an observer who knows the true parameter value. For a variant of the two-armed bandit problem it is demontrated that for a residual family of parameter values and priors with objective probability one posterior beliefs of the Bayesian decision-maker never converge. *JEL* Classification Numbers: 022, 026, 211.

I. INTRODUCTION

Recently, there has been a revival of interest in studying the asymptotic dynamics of Bayesian learning and control in economic environments. One set of authors, including Easley and Kiefer [9, 10], Kiefer and Nyarko [16], McLennan [18], and Nyarko [19], have analyzed single agent decision problems for which there is a tradeoff between current period reward and the value of the information generated by the current period action. In another strand of the literature, Blume and Easley [5], Bray and Kreps [6], and Feldman [11, 12] attempt to characterize the tail of the sequence of beliefs and outcomes for economies with many passively learning agents. Specifically, these latter papers focus on whether Bayesian learning by agents with a correct specification of the underlying structure but uncertainty regarding the parameter values is a sufficient condition to assure convergence to a stationary rational expectations equilibrium.

These articles, as well as important earlier contributions of Cyert and DeGroot [6], Rothschild [23], and Townsend [26], have the following common framework. From the vantage point of the economic actors, the set of possible complete descriptions of the relevant time-invariant economic data can be represented as a separable metric space Θ with Borel σ-field B. The actors in the model, uncertain as to the "true" $\theta^0 \in \Theta$, have prior beliefs $\mu(0)$ on (Θ, B).[1] A result common to this literature is a "theorem" that the sequence $\{\mu(t)\}$ of posterior beliefs almost surely converges weakly to $\mu(\infty)$, the posterior belief conditioned on the limit sub-σ-field.

In all of the more recently authored papers the Martingale Convergence Theorem is invoked to establish a.s. convergence of the sequence of Bayesian beliefs, where the a.s. is with respect to the prior probability measures of the Bayesian agents. An essential point of this paper is that this is not necessarily equivalent to stating that a passive observer who is informed of the "true" parameter θ^0 would attach probability one to the event that the sequence of beliefs $\{\mu(t)\}$ of the Bayesian agents would converge to a limit belief $\mu(\infty)$.

To elaborate on this distinction, suppose that in period $t = 0, 1, \ldots$, agents observe outcomes in a separable metric space Y. Given the behavioral rules of the agents, each parameter value $\theta \in \Theta$ induces a probability measure Q_θ on the product space Y^∞. The prior $\mu(0)$ induces a measure $P_{\mu(0)}$ on Y^∞ defined by $P_{\mu(0)}(A) = \int Q_\theta(A) P_{\mu(0)}(d\theta)$. The application of the Martingale Convergence Theorem yields a.s. convergence of $\{\mu(t)\}$, where the a.s. statement is with respect to the probability measure $P_{\mu(0)}$. But this does not imply that for any particular θ that $\mu(t) \Rightarrow \mu^\infty$ with Q_θ probability one, even if θ is in the support of $\mu(0)$.

One might hope to establish a result that for a "large" class of priors, posteriors converge for a "large" class of parameter values. When Θ can be embedded in finite-dimensional Euclidean space, one has recourse to Lebesgue measure m (restricted to Θ) as a natural notion of size. Then if m $<<$ $\mu(0)$ the exceptional θ set, the set of θ such that $Q_\theta(\{\mu(t) \Rightarrow \mu^\infty\}) < 1$, has m measure zero. But in many naturally occurring

---

[1] Some authors allow for heterogeneous beliefs across agents.

settings Θ is not finite dimensional, and since there is no infinite dimensional analogue of Lebesgue measure, a measure-theoretic criterion is unavailable.

In lieu of a reference measure to evaluate size, the customary procedure is to resort to the topological notion of category. Residual subsets are deemed to be large, and subsets of first category (which are complements of residual subsets) are regarded as small.

The formal analysis in this paper focuses on the two-armed bandit problem with a countable number of potential outcomes associated with each arm. Building upon results of Freedman [13] we prove that there is a residual set of pairs (θ, μ) of parameter values and prior beliefs such that with Q_θ probability one posterior beliefs never converge. Furthermore, there exist reward structures with one arm more favorable than the other for an informed decision-maker; yet for a residual set of pairs (θ, μ) both arms will be played infinitely often with Q_θ probability one.

II. NOTATION AND MATHEMATICAL PRELIMINARIES

II.1. *Notational Conventions*

The set of real numbers is denoted by R. The set of natural numbers is denoted by $N = \{1, 2, \ldots\}$. A generic element of N is denoted by either i, n or t. The set of non-negative integers is denoted by $N_0 = \{0, 1, 2, \}$. If Z is an arbitrary set and $A \subset Z$, the indicator function for the set A is denoted by I_A. If W is a topological space, then C(W) is the Banach space of bounded, continuous real-valued functions defined on W with the sup norm. P(W) is the space of probability measures on W endowed with the topology of weak convergence. P(P(W)) is denoted by $P^2(W)$. Unless otherwise specified all σ-fields are the Borel σ-field. If (W,*W*) is a measurable space and $w \in W$ the Dirac measure $\delta_w \in P(W)$ is defined by $\delta_w(A) = 1$ if $w \in A$ and $\delta_w(A) = 0$ if $w \notin A$, for $A \in$ *W*.

II.2. *Category Theory*

I shall briefly review the concepts relating to Baire category. Standard references include Kelley [17, pp. 200-203], Oxtoby [20] and Royden [24, Section 7.8]. Let W be a metric space. A set $E \subset W$ is *nowhere dense* if $\bar{E}$ has empty interior. A set E is of *first category* or *meager* if it is the union of a countable collection of nowhere dense sets. If a set is not of first category then it is of *second category*. The complement of a set of first category is a *residual set*.

According to the theorem of Baire [24, Theorem 7.27], if W is a complete metric space then the intersection of a countable family of open dense subsets of W is itself a dense subset of W.

III. THE MODEL

III.1. *Preliminary Description*

There are a countable number of time periods, with periods indexed by $t = 0, 1, 2, \ldots$. At the beginning of period t the Bayesian decision-maker selects an *action* (or bandit-arm) $x(t) \in X = \{x_1, x_2\}$.

After choosing action x_k, an *outcome* $y(t) \in Y_k = \{y_{k1}, y_{k2}, \ldots\}$ is observed, and a period reward $R_k(y(t))$ is received. The range of R_k is bounded with $c_k = \inf\{R_k(y_{ki}): y_{ki} \in Y_k\}$.

The action $x(t)$ and outcome $y(t)$ are the realizations of the random variables $X(t)$ and $Y(t)$. The probability law specifying the distribution of $Y(t)$ conditional upon $X(t) = x_k$ is $\theta_k^0 \in P(Y_k)$, where Y_k is endowed with the discrete topology. The objective probabilities θ_1^0 and θ_2^0 are respectively elements of the *parameter subspaces* $\Theta_1 = P(Y_1)$ and $\Theta_2 = P(Y_2)$, and the vector $\theta^0 = (\theta_1^0, \theta_2^0)$ is an element of the *parameter space* $\Theta = \Theta_1 \times \Theta_2$. θ^0 is initially unknown to the decision-maker, and so is viewed from her perspective as the realization of a random element with distribution $\mu(0) \in P(\Theta)$. (When it is clear from the context that μ represents a prior belief, we sometimes write μ instead of $\mu(0)$.)

The objective of the decision-maker is to maximize, with discount factor $\alpha \in (0, 1)$, the expectation of the present value of the reward stream. Denoting for notational convenience the indicator function $I_{x_k}: X \to \{0, 1\}$ by I_k, for a given sequence $\{x(t)\}$ of actions and a given sequence $\{y(t)\}$ of outcomes, the total reward to the decision-maker is $\sum_{t=0}^{\infty} \alpha^t \cdot \{ \sum_{k=1}^{2} I_k(x(t)) \cdot R_k(y(t)) \}$. So the optimization problem to choose a policy $\{X(t)\}$ (a precise definition of a policy is provided in section III.3) that maximizes

$E[\sum_{t=0}^{\infty} \alpha^t \cdot \{ \sum_{k=1}^{2} I_k(X(t)) \cdot R_k(Y_k(t)) \}]$ where the expectation is taken with respect to the probability measure induced by her prior belief $\mu(0)$.

The policy or "rule" that the decision-maker abides by in choosing the sequence $\{x(t)\}$ depends upon α and $\mu(0)$. While α remains fixed throughout the paper, our goal is to study the sensitivity of the tail behavior to variation in the prior beliefs. To emphasize this dependence of actions upon the prior belief $\mu(0)$, we will often write $X(t, \mu(0))$ or $X(t, \mu)$ and $Y(t, \mu(0))$ or $Y(t, \mu)$ for respectively $X(t)$ and $Y(t)$.

III.2. *Specification of the Probability Spaces*

Endowed with the Prohorov metric, Θ is a complete and separable metric space. The Borel subsets of Θ_1, Θ_2, and Θ are denoted respectively by B_1, B_2, and B.

The set of possible outcomes from repeated play of arm k is Y_k^{∞}. Endowing Y_k^{∞} with the product topology, the product σ-field is denoted by Y_k^{∞}. To facilitate the analysis we elect, without loss of generality, to work in representation space (for further details see [8, Section 1.6]). Accordingly, we define the measurable spaces (Ω_1, F_1), (Ω_2, F_2) and (Ω, F) by $(\Omega_k, F_k) = (Y_k^{\infty}, Y_k^{\infty})$ and $(\Omega, F) = (\Omega_1 \times \Omega_2, F_1 \times F_2)$. A generic element of Ω_k is denoted as $\omega_k = (\omega_{k1}, \omega_{k2}, \ldots)$. Associated with each $\theta = (\theta_1, \in \Theta$, is an "objective" product probability measure $Q_\theta = Q_{\theta_1} \times Q_{\theta_2}$ on (Ω, F) defined by $Q_{\theta_k} = \theta_k \times \theta_k \times \ldots$. The interpretation is that: (i) ω_{kn} is the element $y_{ki} \in Y_k$ that is observed on the n'th

occasion that action x_k is chosen, and (ii) that conditional upon parameter θ_k, the outcomes $\{\omega_{k1}, \omega_{k2}, \dots\}$ from choosing x_k are i.i.d. with distribution θ_k.[2]

The product outcome space is $(\Omega, F) = (\Omega_1 \times \Omega_2, F_1 \times F_2)$. For $\theta = (\theta_1, \theta_2) \in \Theta$ the "objective" product probability is $Q_\theta = Q_{\theta_1} \times Q_{\theta_2}$. The set $\{(\Omega, F, Q_\theta): \theta \in \Theta\}$ is the *family of objective probability spaces*.

To avoid extraneous complications we make a restrictive independence assumption that the prior belief $\mu \in P(\Theta)$ can be decomposed as the product probability $\mu_1 \times \mu_2$ of the marginal distributions.[3] The induced prior probability P_μ on $(\Omega \times \Theta, F \times B)$ is defined for measurable rectangles $A \times B$, by $P_\mu(A \times B) = \int_B Q_\theta(A)\, \mu(0)(d\theta)$ for $A \in F$ and $B \in B$.[4] By the Product Measure Theorem [1, Theorem 2.6.2] there is a unique extension of P_μ to $(\Omega \times \Theta, F \times B)$. The *family of subjective probability spaces* is $\{(\Omega \times \Theta, F \times B, P_\mu): \mu \in P(\Theta)\}$. Slightly abusing notation, the marginal distribution of P_μ on (Ω, F) and (Θ, B) is also denoted by P_μ. So for example for $A \in F$, $P_\mu(A) = P_\mu(A \times \Theta)$.

III.3. *Histories and Policies*

A precise statement of the optimization problem is provided in this section. We start by specifying what constitutes an admissible plan. Included in the definition of an admissible plan are the *count* functions $C_k(t): \Omega \to N_0, k = 1, 2$. The realization $C_k(t)(\omega)$ represents the number of times arm k has been chosen through time t.

An *admissible plan* is defined as sequences of random variables $\{X(t)\}$, $\{Y(t)\}$, $\{C_1(t)\}$, and $\{C_2(t)\}$, and a sequence of sub-σ-fields H_t such that:

i) $H_0 = \{\emptyset, \Omega\}$

ii) $X(0)$ is H_0 measurable,

iii) If $X(0)(\omega) = x_k$, then $Y(0)(\omega) = \omega_{k1}$,

iv) for $t \geq 1$, $H_t = H_{t-1} \vee \sigma(X(t-1), Y(t-1))$,

v) $X(t)$ is H_t measurable,

vi) $C_k(0) = I_k(X(0))$, and $C_k(t) = C_k(t-1) + I_k(X(t))$, and

vii) If $X(t)(\omega) = x_k$ and $C_k(\omega) = n$, then $Y(t)(\omega) = \omega_{kn}$.

A *policy* is a sequence $\{X(t)\}$ for which there exists a (unique) corresponding admissible plan $[\{X(t), \{Y(t)\}, \{C_1(t)\}, \{C_2(t)\}, \{H_t\}]$. An *optimal policy for the prior* μ maximizes

---

[2]Technically, it is the projections $\{\text{Proj}(k, n)\}$, defined by $\text{Proj}(k, n)(\omega) = \omega_{kn}$ which are independent.

[3]Note the contrast with the assumptions of Rothschild [23]. Rothschild allows for the arms to be dependent but requires the outcome space be a two point set.

[4]For the integral to be well-defined it is necessary that the map $\theta \to Q_\theta(A)$ be measurable. This is a consequence of [21, Lemma 6.1].

$$\int_{\Omega} \sum_{t=0}^{\infty} \alpha^t \cdot \{ \sum_{k=1}^{2} I_k(X(t)(\omega)) \cdot R_k(Y(t)(\omega)) \} P_{\mu}(d\omega) \text{ over the set of policies.}$$

III.4. *Posterior Beliefs and The Optimal Bayesian Policy*

When possible, beliefs are updated according to Bayes rule. When an event of prior probability zero occurs, the updating is arbitrary. Since none of the results of this paper depend upon the choice of version of conditional probability, without loss of generality we shall assume that the decision-maker ignores any outcome to which she assigned a prior probability of zero. More precisely, the posterior map $\Gamma_k: P(\Theta_k) \times Y_k \to P(\Theta_k)$, $k = 1, 2$ is defined by:

For all $A \in B_k$,

$$\Gamma_k(\mu_k, y_{ki})(A) = \frac{\int_A \theta_k(y_{ki}) \mu_k(d\theta_k)}{\int_{\Theta_k} \theta_k(y_{ki}) \mu_k(d\theta_k)} \text{ whenever } \int_{\Theta_k} \theta_k(y_{ki}) \mu_k(d\theta_k) > 0, \text{ and}$$

$$\Gamma_k(\mu_k, y_{ki})(A) = \mu_k(A) \text{ whenever } \int_{\Theta_k} \theta_k(y_{ki}) \mu_k(d\theta_k) = 0.$$

For prior $\mu = \mu_1 \times \mu_2$, the sequences $\{v_1(n, \mu_1)\}_{n=0}^{\infty}$ and $\{v_2(n, \mu_2)\}_{n=0}^{\infty}$ of updating functions, $v_k(n, \mu_k): \Omega \to P(\Theta_k)$ are recursively defined by:

$$v_k(0, \mu_k)(\omega) = \mu_k,$$

$$v_k(n, \mu_k)(\omega) = \Gamma_k(v_k(n-1, \mu_k)(\omega), \omega_{kn}) \text{ for } n \geq 1.$$

The interpretation is that $v_k(n, \mu_k)(\omega)$ is the posterior measure on (Θ_k, B_k) after arm k has been chosen n times with prior measure μ_k.

Associated with a policy $\{X(t)\}$ and the prior belief $\mu = \mu_1 \times \mu_2$ is a sequence of posterior maps $\{\mu(t)\}_{t=1}^{\infty}$. The map $\mu(t): \Omega \to P(\Theta)$ is defined by $\mu(t)(\omega) = \mu_1(t)(\omega) \times \mu_2(t)(\omega)$ where $\mu_k(t)(\omega) = v_k(C_k(\omega), \mu_k)(\omega)$.

A policy $\{X(t)\}$ is *stationary* if there exists a measurable function $\pi: P(\Theta) \to X$ such that $X(t)(\omega) = \pi(\mu(t)(\omega))$ for all $\omega \in \Omega$. Identifying a function $\pi: P(\Theta) \to X$ with the stationary policy it induces, no confusion should result from referring to π as a stationary policy.

Invoking now well-known results (for details see [9], [15] or [21]) the Bayesian optimal control problem can be reformulated as a dynamic programming problem with state space $P(\Theta)$ and value function $V: P(\Theta) \to R$. Furthermore, there exists an optimal stationary policy [5, Theorem 7(b)].

Drawing on results of Gittins and Jones [14], with refinements by Berry and Fristedt [2], Ross [22] and Whittle [27], there is an optimal policy with a sharp characterization. Gittins and Jones proved the existence of functions $M_1: P(\Theta_1) \to R$ and $M_2: P(\Theta_2) \to R$ with the property that it is optimal to choose arm 1 in period t iff $M_1(\mu_1(t)) \geq M_2(\mu_2(t))$. The function M_k is often referred to as the *Gittins Index* for arm k.

To define M_1 consider a one-armed bandit problem where in the initial stage the decision maker has the option of playing arm 1 or stopping and collecting a terminal reward of m. In subsequent stages, if arm 1

has been played in all previous stages then the options remain selecting arm 1 or stopping and receiving a final payment m. The value of the optimal policy for beliefs $\mu_1 \in P(\Theta_1)$ is denoted as $V_1(\mu_1, m)$. The Gittins index is defined by $M_1(\mu_1) = \inf\{m: V_1(\mu_1, m) = m\}$. M_2 is defined analogously.

Throughout the remainder of the paper we assume that the Bayesian controller follows an optimal stationary policy $\pi^*: P(\Theta) \to X$ defined by:

$\pi^*(\mu_1 \times \mu_2) = x_1$ if $M_1(\mu_1) \geq M_2(\mu_2)$, and $\pi^*(\mu_1, \mu_2) = x_2$ if $M_2(\mu_2) > M_1(\mu_1)$.

For a more detailed exposition the reader is advised to consult Ross [22], Whittle [27], and especially the monograph of Berry and Fristedt [2].

IV. GENERIC LIMIT THEOREMS

IV.1. *Some Continuity Results*

Preparatory to proving the genericity theorems, some preliminary technical results are developed in this subsection. The principal result is the establishment of the continuity of the functions M_1 and M_2.

The first step is to develop characterizations of V_1 and V_2 by making use of the fact that the value functions V_k are solutions to the optimality equation. The expected one-period reward for choosing action x_k with beliefs μ_k is denoted by the function $U_k: P(\Theta_k) \to$ R defined by $U_k(\mu_k) =$

$\int [\sum_{i=1}^{\infty} R_k(y_{ki})\theta_k(y_{ki})] \mu_k(d\theta_k)$. Informally denoting expectation taken with respect to current beliefs μ_k by E_{μ_k} and denoting future beliefs with a prime symbol, the optimality equation can be written as:

$V_k(\mu_k, m) = \max \{m, U_k(\mu_k) + \alpha \cdot E_{\mu_k}[V_k(\mu_k', m)]\}$.

To make the above precise, it necessary to provide formal definitions of E_{u_k} and the mapping from current beliefs to probability distributions over future beliefs. So for k = 1, 2, define $\Psi_k: P(\Theta_k) \to P^2(\Theta_k)$ by $\Psi_k(\mu_k)(A) = \int_{\Theta_k} [\sum_{i=1}^{\infty} I_A(\Gamma_k(\mu_k, y_{ki})) \cdot \theta_k(y_{ki})] \mu_k(d\theta_k)$. Conditional upon period t beliefs $\mu_k(t)$ and arm k being selected, $\Psi_k(\mu_k(t))(A)$ can be interpreted as the probability from the point of view of the decision-maker that $\mu_k(t+1) \in A$. It will also be useful to have an expression for the conditional probability of realization y_{ki} given that arm k is chosen with beliefs μ_k. So we define the functions $\lambda_k: Y_k \times P(\Theta_k)$, k = 1, 2, by $\lambda_k(y_{ki}, \mu_k) = \int_{\theta_k} \theta_k(y_{ki}) \mu_k(d\theta_k)$. We can now prove the following results.

LEMMA 4.1. *The maps* $\mu_k \to \Gamma_k(\mu_k, y_{ki})$, k = 1, 2 *are continuous for all* i $\in$ N.

Proof. The proof which follows from routine arguments is for completeness included in the Appendix. ∎

PROPOSITION 4.2. Ψ_1 *and* Ψ_2 *are continuous.*

Proof. The proof is in the Appendix. ∎

The next step is to establish the continuity of V_1 and V_2. The boundedness of the functions R_1 and R_2 is necessary for this result. Berry and Fristedt [2, p.42] provide a counterexample to continuity when the reward function is unbounded.

LEMMA 4.3. V_1 *and* V_2 *are continuous functions.*

Proof. It suffices to prove that V_1 is continuous. Let $K \subset R$ be compact and define $E_K = C(P(\Theta_1) \times K)$. Define $T_K: E_K \to E_K$ by $T_K f(\mu_1, m) = \text{Max } \{m, U_1(\mu_1) + \alpha \cdot \int f(\mu_1', m) \, \Psi_1(\mu_1)(d\mu_1')\}$. Note that U_1 is bounded since R_1 is bounded. Invoking standard arguments, it is easily confirmed that T_K is a contraction mapping with fixed-point $\phi_K: P(\Theta_1) \times K$. Since ϕ_K is the solution to the optimality equation, it is also the restriction of $V_1: P(\Theta_1) \times R \to R$ to $P(\Theta_1) \times K$. Since ϕ_K is continuous, V_1 is continuous on every compact set and by extension V_1 is continuous. ∎

The expected loss from stopping when arm k is the only arm, is given by the function $L_k: P(\Theta_k) \times R \to R$ defined by $L_k(\mu_k, m) = V_k(\mu_k, m) - m$. The following Lemma collects some well-known results.

LEMMA 4.4. For $k = 1, 2$:

(i) the function $V_k(\mu_k, \cdot)$ is increasing and convex, and

(ii) the function $L_k(\mu_k, \cdot)$ is decreasing.

Proof. (i) See [20, Lemma VII.3.1]. (ii) Replacing sums with integrals, the proof provided in [20, Lemma VII.2.1] remains valid. ∎

Now consider the game for which the decision-maker must initially play arm k, and thereafter chooses between continuing to select arm k or permanently retiring with payoff of m. The value of pursuing the optimal strategy for this game is given by the function $W_k: P(\Theta_k) \times R \to R$ defined by $W_k(\mu_k, m) = U_k(\mu_k) + \alpha \cdot \int V_k(\mu_k', m) \, \Psi_k(\mu_k)(d\mu_k')$. With the following Lemma we prove that the set of terminal rewards m, for which the decision-maker is indifferent between continuing play (and possibly stopping at some future date with reward m received at that future date) and stopping immediately is a singleton.

LEMMA 4.5. (i) W_1 *and* W_2 *are continuous.* (ii) If $M_k(\mu_k) < m$, then $W(\mu_k, m) < m$.

Proof. (i) The proof follows from Proposition 4.2 and Theorem 5.5 of [3].

(ii) Suppose $M_k(\bar{\mu}_k) = \bar{m} < m$. Since $\bar{m}$ is a terminal payoff for which the controller is indifferent between continuing playing arm k and receiving payoff $\bar{m}$, $W_k(\bar{\mu}_k, \bar{m}) = \bar{m} = V_k(\bar{\mu}_k, \bar{m})$. According to Lemma 4.4 for all $\mu_k \in P(\Theta_k)$, $V_k(\mu_k, m) - V_k(\mu_k, \bar{m}) \leq (m - \bar{m})$, implying that $\int [V_k(\mu_k', m) - V_k(\mu_k', \bar{m})] \, \Psi_k(\bar{\mu}_k)(d\mu_k') \leq (m - \bar{m})$. Finally, note that

$$W_k(\bar{\mu}_k, m) = U_k(\bar{\mu}_k) + \alpha \cdot \int V_k(\mu_k', m) \Psi_k(\bar{\mu}_k)(d\mu_k')$$

$$= U_k(\bar{\mu}_k) +$$

$$\alpha \cdot \left\{ \int V_k(\mu_k', \bar{m}) \Psi_k(\bar{\mu}_k)(d\mu_k') + [\int V_k(\mu_k', m) \, \Psi_k(\bar{\mu}_k)(d\mu_k') - \int V_k(\mu_k', \bar{m}) \, \Psi_k(\bar{\mu}_k)(d\mu_k')] \right\}$$

$$= \bar{m} + \alpha \cdot [\int V_k(\mu_k', m) \, \Psi_k(\bar{\mu}_k)(d\mu_k') - \int V_k(\mu_k', \bar{m}) \Psi_k(\bar{\mu}_k)(d\mu_k')]$$

$$\leq \bar{m} + \alpha \cdot (m - \bar{m})$$

$$< m. \quad \blacksquare$$

The proof of continuity of M_1 and M_2 is now completed by verifying that these functions are both upper semicontinuous (u.s.c.) and lower semicontinuous (l.s.c.).

PROPOSITION 4.6. M_1 *and* M_2 *are continuous.*

Proof. First we establish that M_k is l.s.c. by verifying that for all $c \in R$ the set $\{\mu_k \in P(\Theta_k): M_k(\mu_k) > c\}$ is open. Suppose that $M_k(\bar{\mu}_k) > c$. So by definition of M_k, $V_k(\bar{\mu}_k, c) - c = L_k(\bar{\mu}_k, c) > 0$. From the continuity of L_k (Lemma 4.3), $\exists$ an open neighborhood N_k of $\bar{\mu}_k$ such that for all $\mu_k \in N_k$, $L_k(\mu_k, c) > 0$. Therefore, for all $\mu_k \in N_k$, $V_k(\mu_k, c) - c > 0$ and $M_k(\mu_k) > c$.

Upper semicontinuity is confirmed by demonstrating that the set $\{\mu_k \in P(\Theta_k): M_k(\mu_k) < c\}$ is open for all $c \in R$. Suppose that $M_k(\bar{\mu}_k) = \bar{m} < c$. By Lemma 4.5, $W_k(\bar{\mu}_k, c) < c$ and $\exists$ an open neighborhood J_k of $\bar{\mu}_k$ such that for $\mu_k \in J_k$, $W_k(\mu_k, c) < c$. But this implies that $M_k(\mu_k) < c$. $\quad \blacksquare$

IV.2.*Generic Outcomes when* $c_1 \neq c_2$

In this subsection we analyze the limit outcomes in the case when $c_1 \neq c_2$ (Recall that $c_k = \inf\{R_k(y_{ki}): i \in N\}$). Without loss of generality we assume that $c_1 < c_2$. To understand the next result, select $i^* \in N$ such that $R_1(y_{1i^*}) < c_2$. So if the decision-maker attaches sufficiently high probability to y_{1i^*} occurring if arm 1 is selected, arm 2 will be selected regardless of $\mu_2(t)$. Furthermore, the Proposition of Freedman [13] stated below implies that for a residual set of $(\theta_1, \mu_1(0))$ pairs, the probability of playing arm 1 infinitely often with $\int_{\Theta_1} \theta_1(y_{1i^*}) \, \mu_1(t)(d\theta_1)$ bounded away from one, is zero.

PROPOSITION 4.7 [Freedman]. *For* $k = 1, 2$, *the set* Σ_k *of pairs* $(\theta_k, \mu_k(0)) \subset \Theta_k \times P(\Theta_k)$ *with* $\lim \sup_{t \to \infty} \int_{\Omega_k} \mu_k(t)(\omega_k)(U) \, Q_{\theta_k}(d\omega_k) = 1$ *simultaneously for all nonempty open subsets of* Θ_k *is residual.*

So for a residual set $\Sigma_1 \subset \Theta_1 \times P(\Theta_1)$, if arm 1 is played sufficiently often the decision-maker will eventually (Q_θ a.s.) be sufficiently pessimistic regarding the realization θ_1 that arm 1 will never be tested again.

PROPOSITION 4.8. *Suppose $c_1 < c_2$. Then there exists a residual subset $\Sigma_1 \subset \Theta_1 \times P(\Theta_1)$ such that for $(\theta_1, \mu_1(0)) \in \Sigma_1$ and all $(\theta_2, \mu_2(0)) \in \Theta_2 \times P(\Theta_2)$, Q_θ a.s. $X(t, \mu_1(0) \times \mu_2(0))(\omega) = x_k$ only finitely often.*

Proof. For $i \in N$, define $\delta_i^2 \in P(\Theta_1)$ by $\delta_i^2(B) = 1$ if $\delta_i \in B$, else $\delta_i^2(B) = 0$. ($\mu_1 = \delta_i^2$ corresponds to the belief that outcome i occurs with probability one whenever action x_1 is chosen.) Choose $i^* \in N$ s.t. $R_1(y_{1i^*}) < c_2$. Observing that $M_1(\delta_i^2) = \dfrac{R_1(y_{1i^*})}{(1 - \alpha)}$ and recalling that M_1 is continuous (Proposition 4.6), there exists a neighborhood U of δ_i^2 s.t. for $\mu_1 \in U$, $M_1(\mu_1) < \dfrac{c_2}{1 - \alpha}$. Note that if $\mu_1(t)(\omega) \in U$, then for all $t' \geq t$, $X_t(\omega) = x_2$. The ω set for which $v_1(n, \mu_1)(\omega)$ is never in U is denoted as $D_1(\mu_1) = \{\omega: v_1(n, \mu_1)(\omega) \notin U, n = 1, 2, ...\}$. Define $\Sigma_1 = \{(\theta, \mu_1(0)): Q_\theta(D_1(\mu_1)) = 0\}$. By Proposition 4.7, Σ_1 is residual.

For $(\theta, \mu_1(0)) \in \Sigma_1$, define $n^*(\omega) = \inf\{n: v_1(n, \mu_1(0))(\omega) \in U\}$. n^* is Q_θ a.s. finite and so Q_θ a.s., for all t the count function $C_1(t)(\omega) \leq n^*(\omega) < \infty$. ∎

IV.3. *Generic Outcomes when $c_1 = c_2$*

To understand the next result suppose that $c_1 = c_2$ and $\dfrac{c_1}{1 - \alpha} < M_1(\mu_1(0)) < M_2(\mu_2(0))$. Then generically, Q_θ a.s. there will exist some time t, say τ_1, such that $M_2(\mu_2(\tau_1)) < M_1(\mu_1(0))$. Then reversing the roles of the arms, eventually at some time $\tau_2 > \tau_1$, $M_1(\mu_1(\tau_2)) < M_2(\mu_2(\tau_1))$, etc.. And so each arm will be chosen infinitely often. Observe that there exist consistent estimators of (θ_1, θ_2) and so from the point of view of a classical statistician, there will eventually be overwhelming evidence as to which arm offers superior prospects.

PROPOSITION 4.9. *Suppose that $c_1 = c_2$ and that the range of R_k, $k = 1, 2$, is not a singleton. Then there is a residual subset $\Re \subset \Theta \times P(\Theta)$ such that for $(\theta_1, \theta_2, \mu_1, \mu_2) \in \Re$, with Q_θ probability one $X(t, \mu(0)) = x_1$ infinitely often and $X(t, \mu(0)) = x_2$ infinitely often.*

Proof. For $\mu \in P(\Theta)$ define the set $\Xi_1(\mu) = \{\omega: X(t, \mu(0)) = x_2$ finitely often$\}$. The strategy is to first prove that there exists a residual subset $\Re_1 \subset \Theta \times P(\Theta)$ such that for $(\theta_1, \theta_2, \mu_1, \mu_2) \in \Re_1$, $Q_\theta(\Xi_1(\mu)) = 0$. Defining Ξ_2 analogously, by an identical argument there exists a residual subset $\Re_2 \subset \Theta \times P(\Theta)$ such that for $(\theta_1, \theta_2, \mu_1, \mu_2) \in \Re_2$, $Q_\theta(\Xi_2(\mu)) = 0$. Since the countable intersection of residual sets is a residual set, $\Re$ defined by $\Re = \Re_1 \cap \Re_2$ is residual.

We now verify the existence of such a set $\Re_1$. For $k = 1, 2$, define $\Theta_k^+ = \{\theta_k \in \Theta_k: \forall\ i \in N, \theta_k(y_{ki}) > 0\}$. Θ_k^+ is residual [13, Remark 1]. The sets $\Lambda_k \subset P(\Theta_k)$ defined by $\Lambda_k = \{\mu_k \in P(\Theta_k): \mu_k(\Theta_k^+) > 0\}$ are also residual [13, Remark 2], and so $\Lambda = \Lambda_1 \times \Lambda_2$ is residual [20, Theorem 15.3].

For $n \in N$, $\exists\ i(n)$ such that $R_1(y_{1i(n)}) < c_1 + n^{-1}$ and hence $M_1(\delta_{i(n)}^2) < \dfrac{c_1 + n^{-1}}{1 - \alpha}$. By the continuity of M_1, there exists a neighborhood N_n of $\delta_{i(n)}^2)$ such that for $\mu_1 \in U_n$, $M(\mu_1) < \dfrac{c_1 + n^{-1}}{1 - \alpha}$. Define N to be a family of such neighborhoods. So if $\mu_2(0) \in \Lambda_2$, then for all $\omega \in \Omega$ and $\forall\ n' \in N$, $\exists\ N_n \in N$ s.t. $\sup_{\mu_1 \in N_n} [M_1(\mu_1)] < M_2(v_2(n', \omega))$.

Define $\vartheta(\mu_1) = \{\omega \in \Omega: \exists\ n(\omega)\ \text{s.t. for all}\ n > n(\omega),\ v_1(j, \mu_1)(\omega) \in U_n\ \text{for only finitely many}\ j\}$. Define $\Re_1 = \{(\mu, \theta): \mu \in \Lambda, Q_\theta(\Xi(\mu_1, \mu_2)) = 0$ and observe that: (i) for $\theta = (\theta_1, \theta_2)$, $Q_\theta(\vartheta(\mu_1))$ does not depend upon θ_2, and (ii) for $\mu = (\mu_1, \mu_2) \in \Lambda$, $\Xi(\mu_1, \mu_2) \subset \vartheta(\mu_1)$. A direct consequence is that $\Re_1 \supset \{(\mu, \theta): \mu \in \Lambda, Q_\theta(\vartheta(\mu_1)) = 0\}$.

The proof is completed by confirming that $\{(\mu, \theta): \mu \in \Lambda, Q_\theta(\vartheta(\mu_1)) = 0\}$ is residual and hence $\Re_1$ is residual. Define $\aleph_1 = \{(\mu_1, \theta_1) \in P(\Theta_1) \times \Theta_1: Q_\theta(\vartheta(\mu_1)) = 0\}$. According to Proposition 4.7, $\aleph_1$ is residual. By application of the category analog of Fubini's theorem [20, Theorem 15.3], $\Lambda_1 \times \Theta_1$ is residual, and so $\aleph_1 \cap (\Lambda_1 \times \Theta_1)$ is residual. Again invoking [20, Theorem 15.3], $\{(\mu_1, \mu_2, \theta_1, \theta_2): (\mu_1, \theta_1) \in \aleph_1 \cap (\Lambda_1 \times \Theta_1), \mu_2 \in \Lambda_2\}$ is residual. And by definition, $(\mu_1, \mu_2, \theta_1, \theta_2): (\mu_1, \theta_1) \in \aleph_1 \cap (\Lambda_1 \times \Theta_1), \mu_2 \in \Lambda_2\} = \{(\mu, \theta): \mu \in \Lambda, Q_\theta(\vartheta(\mu_1)) = 0\}$. ■

V. CONCLUDING REMARKS

It is natural to inquire as to the generality of Propositions 4.7 and 4.8. Is there some peculiarity associated with having a countable outcome space that underlies these nonintuitive propositions? In cases when the parameter space is a smooth, finite dimensional set (as in [10], [16], and [19]) one could hope to modify the theorems of Schwartz [24] and prove a.s. convergence of beliefs for all $\theta \in \Theta$. (This work though, remains to be undertaken.) But the results of Freedman [13] and of this paper can be extended to other infinite dimensional settings. For instance, suppose for $k = 1, 2$ the outcome space $Y_k = [0, 1]$. Then if Θ_k is the set of all density functions with the $L_1[0, 1]$ topology, with minimal modification all proofs remain valid. Furthermore, I conjecture that the restriction that the decision-maker choose form a finite set of actions is inessential. So while a complete answer is not yet available, it appears that the only crucial assumption is that the parameter space Θ is infinite dimensional.

To assess the significance of the results contained in this paper with regard to economic modelling with Bayesian learning, requires addressing the issue of the "size" of the residual set R. The customary practice is to equate residual with "large" (even though in R^n there are residual sets of Lebesgue measure zero). If this equivalence is accepted and if prior beliefs are arbitrarily chosen, then one is lead to conclude as did Freedman [13], "for essentially any pair of Bayesians, each thinks the other is crazy."

An obvious objection to this stark prediction is that there is little basis for automatically classifying residual sets as "large". On the other hand, there are no grounds for asserting that the "bad" pairs of parameters and priors are in any sense small. A more cautious interpretation, relying only on the density of $\Re$ and of the Dirichlet priors, would stress the sensitivity of asymptotic beliefs to the prior. In particular, it

is illegitimate to restrict attention to a computationally convenient class of priors (even when dense) and presume that any resulting limit theorems extend beyond the initial family of distributions.

A second point of contention with the remark of Freedman, is that it may be appropriate to impose a priori restrictions on the family of admissible prior beliefs. Freedman himself now appears to support this stance. Diaconis-Freedman [7] advocate that Bayesian statisticians adopt what they label the "what if" method. This consists of "... after specifying a prior distribution generate imaginary data sequences, compute the posterior, and consider whether the posterior would be an adequate representation of the updated prior." Adapting this reccommendation to the context of economic modelling, it would be natural to require that agents have prior beliefs with full support and that the sequence of agent posterior beliefs converges almost surely with respect to the measure Q_θ for all $\theta \in \Theta$.

While this author is sympathetic to imposing such restrictions, these auxiliary conditions are ad hoc; Axioms of Bayesian decision theory provide no foundation for such constraints on prior beliefs. Furthermore, while the Dirichlet priors constitute an example of such priors within the specific framework of this paper, the existence of such priors for more general learning models (such as in [8]) has not yet been verified. Further research on this issue is needed.

Appendix

Proof of Lemma 4.1. It suffices to prove that $\Gamma_1(\cdot, y_{1i})$ is continuous. To simplify the notation for this proof, representative elements of Y_1, Θ_1, and $P(\Theta_1)$ are denoted respectively as y_i, θ, and μ. Let $F \subset \Theta_1$ be a closed set and suppose $\mu^n \Rightarrow \tilde{\mu} \in P(\Theta_1)$. Applying a standard characterization of weak convergence [B, Theorem 2.1], it suffices to show that $\overline{\lim}_n \Gamma_1(\mu^n, y_i)(F) \leq \Gamma_1(\mu, y_i)(F)$. Or equivalently that the map $\mu \rightarrow \Gamma_1(\mu, y_i)(F)$ is upper semicontinuous u.s.c..

Recall that $\Gamma_1(\mu, y_i)(F) = \dfrac{\int_F \theta(y_i)\mu(d\theta)}{\int_{\Theta_1} \theta(y_i)\mu(d\theta)}$. And note that since the map $\theta \rightarrow \theta(y_i)$ is bounded and continuous the map $\mu \rightarrow \int_{\Theta_1} \theta(y_i) \mu(d\theta)$ is continuous. So the proof will be completed by confirming that the map $\mu \rightarrow \int_F \theta(y_i) \mu(d\theta)$ u.s.c..

The map $\theta \rightarrow I_F(\theta) \cdot \theta(y_i)$ is u.s.c., and by [3, Exercise 7, p. 17] the map $\mu \rightarrow \int_{\Theta_1} I_F(\theta) \cdot \theta(y_i) \mu(d\theta)$ is u.s.c.. But $\int_F \theta(y_i) \mu(d\theta) = \int_{\Theta_1} I_F(\theta) \cdot \theta(y_i) \mu(d\theta)$, and so the map $\mu \rightarrow \int_F \theta(y_i) \mu(d\theta)$ u.s.c.. ∎

Proof of Proposition 4.2. It suffices to prove that Ψ_1 is continuous. As in the proof of Lemma 4.1, for notational simplicity representative elements of Y_1, Θ_1, and $P(\Theta_1)$ are denoted respectively as y_i, θ, and μ.

Let f: $P(\Theta_1) \to R$ be a bounded continuous function and $\{\mu^n\}$ a sequence with $\mu^n \Rightarrow \bar{\mu}$. To establish continuity of Ψ_1, we must verify that $\int_{P(\Theta)} f(\mu')\,\Psi_1(\mu^n)(d\mu') \to \int_{P(\Theta)} f(\mu')\,\Psi_1(\bar{\mu})(d\mu')$, or equivalently that

$$\int_\Theta [\sum_{i=1}^\infty f(\Gamma_1(\mu, y_i) \cdot \theta(y_i))]\, \mu^n(d\theta) \to \int_\Theta [\sum_{i=1}^\infty f(\Gamma_1(\mu, y_i) \cdot \theta(y_i))]\, \bar{\mu}(d\theta).$$

Define the functions g: $\Theta_1 \to R$ and g_n: $\Theta_1 \to R$ by $g(\theta) = \sum_{i=1}^\infty f(\Gamma(\bar{\mu}, y_i)) \cdot \theta(y_i)$ and $g_n(\theta) = \sum_{i=1}^\infty f(\Gamma(\mu^n, y_i)) \cdot \theta(y_i)$. Since the maps $y_i \to f(\Gamma(\bar{\mu}, y_i))$ and $y_i \to f(\Gamma(\mu^n, y_i))$ are bounded and continuous, g and g_n are continuous. Furthermore, as is now demonstrated, $g_n(\theta) \to g(\theta)$, $\bar{\mu}$ a.s.. Define $\lambda_\mu \in P(Y_1)$ by $\lambda_\mu(A) = \sum_{i=1}^\infty I_A(y_i) \cdot \lambda(y_i, \mu)$ for $A \subset Y_1$. The support of $\bar{\mu}$ is contained in $\Pi \subset \Theta_1$, defined by $\Pi = \{\theta \in \Theta_1 : \theta << \lambda_{\bar{\mu}}\}$. For $\theta \in \Pi$, $\theta(y_i) > 0$ implies that $\Gamma(\cdot, y_i)$ is continuous at $\bar{\mu}$ [Lemma 4.1], and hence $f(\Gamma(\cdot, y_i))$ is continuous at $\bar{\mu}$. Denoting the l_∞ norm of f by $\|f\|_\infty$, $\int_{Y_1} \|f\|_\infty\, \theta(dy_i) < \infty$. By a standard reult on the preservation of continuity under integration [4, Theorem 16.8], for $\theta \in \Pi$ the map

$\mu \to \sum_{i=1}^\infty f(\Gamma(\mu, y_i)) \cdot \theta(y_i)$ is continuous at $\bar{\mu}$ and so $g_n(\theta) \to g(\theta)$.

Finally, $\int_\Theta g_n(\theta)\, \mu_n(d\theta) \to \int_\Theta g(\theta)\, \bar{\mu}(d\theta)$ by [3, Theorem 5.5], or equivalently

$$\sum_{i=1}^\infty f(\Gamma_1(\mu, y_i) \cdot \theta(y_i))]\, \mu^n(d\theta) \to \int_\Theta [\sum_{i=1}^\infty f(\Gamma_1(\mu, y_i) \cdot \theta(y_i))]\, \bar{\mu}(d\theta). \blacksquare$$

References

1. Ash, R. (1972), Real Analysis and Probability, New York, Academic Press.

2. Berry, D. A., and B. Fristedt (1985), Bandit Problems: Sequential Allocation of Experiments, London: Chapman and Hall.

3. Billingsley, P. (1968), Convergence of Probability Measures, New York, Wiley.

4. Billingsley, P. (1979), Probability and Measure, New York, Wiley.

5. Blackwell, D. (1965), Discounted Dynamic Programming, *Annals of Mathematical Statistics* 36, 226-235.

6. Blume, L. and D. Easley (1984), "Rational Expectations Equilibrium: an Alternative Approach, " *Journal of Economic Theory* 34, 116-129.

5. Bray, M. and D. Kreps (1987), "Rational Learning and Rational Expectations," Arrow and the Ascent of Modern Economic Theory, edited by George Feiwel, New York, New York University Press.

6. Cyert, R. and M. DeGroot (1974), "Rational Expectations and Bayesian Analysis," *Journal of Political Economy* 82, 521-536.

7. Diaconis, P. and D. Freedman (1986), "On the Consistency of Bayes Estimates," Special Invited Paper, *Annals of Statistics* 14, 1-26.

8. Doob, J. (1953), Stochastic Processes, New York, Wiley.

9. Easley, D. and N. Kiefer (1988), "Controlling a Stochastic Process with Unknown Parameters," *Econometrica* 56, 1045-1064.

10. Easley, D. and N. Kiefer (1988), "Optimal Learning with Endogenous Data," forthcoming, *International Economic Review.*

11. Feldman, M. (1987a), "An Example of Convergence to Rational Expectations with Heterogeneous Beliefs," *International Economic Review* 28, 635-650.

12. Feldman, M. (1987b), "Bayesian Learning and Convergence to Rational Expectations," *Journal of Mathematical Economics* 16, 297-313.

13. Freedman, D. (1965), "On the Asymptotic Properties of Bayes Estimates in the Discrete Case II," *Annals of Mathematical Statistics* 36, 454-456.

14. Gittins, J. and D. Jones (1974), "A Dynamic Allocation Index for the Sequential Design of Experiments," in Progress in Statistics, edited by J. Gani, K. Sarkadi, and I. Vincze, 241-266, North-Holland, Amsterdam.

15. Hinderer, K. (1970), Foundations of Nonstationary Dynamic Programming with Discrete Time Parameter, Springer-Verlag, New York.

16. Kiefer, N. and Y. Nyarko (1988), "Control of a Linear Regression Process with Unknown Parameters," in Dynamic Econometric Modeling, edited by W. Barnett, E. Berndt, and H. White, Cambridge University Press, New York.

17. Kelley, J. (1955), General Topology, D. Van Nostrand, Princeton.

18. McLennan, A. (1987), "Incomplete Learning in a Repeated Statistical Decision Problem," working paper, University of Minnesota.

19. Nyarko, Y. (1988), "On the Convergence of Bayesian Posterior Processes in Linear Economic Models," working paper, Brown University.

20. Oxtoby, J. (1980), Measure and Category: A Survey of the Analogies between Topological and Measure Spaces, second edition, Springer-Verlag, New York.

21. Rieder, U. (1975), "Bayesian Dynamic Programming," *Advances in Applied Probability* 7, 330-348.

22. Ross, S. (1983), Introduction to Stochastic Dynamic Programming, New York, Academic Press.

23. Rothschild, M. (1974), "A Two-Armed Bandit Theory of Market Pricing," *Journal of EconomicTheory* 9, 185-202.

24. Royden, H. L. (1988), Real Analysis, third edition, Macmillan, New York.

25. Schwartz, L. (1965), "On Bayes Procedures," *Z. Wahr.* 4, 10-26.

26. Townsend, R. (1978), "Market Anticipations, Rational Expectations and Bayesian Analysis," *International Economic Review* 19, 481-494.

27. Whittle, P. (1982), Optimization Over Time: Dynamic Programming and Stochastic Control-I, John Wiley, New York.