ELSEVIER Journal of Mathematical Economics 33 (2000) 425–439

JOURNAL OF Mathematical ECONOMICS

www.elsevier.com/locate/jmateco

Relative entropy in sequential decision problems [1]

Ehud Lehrer [a,*], Rann Smorodinsky [b,2]

[a] *School of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Ramat-Aviv, Tel Aviv 69978, Israel*

[b] *Industrial Engineering, Technion, Haifa, 32000, Israel*

Received 22 June 1998; received in revised form 26 May 1999; accepted 17 June 1999

Abstract

Consider an agent who faces a sequential decision problem. At each stage the agent takes an action and observes a stochastic outcome (e.g., daily prices, weather conditions, opponents' actions in a repeated game, etc.). The agent's stage-utility depends on his action, the observed outcome and on previous outcomes. We assume the agent is Bayesian and is endowed with a subjective belief over the distribution of outcomes. The agent's initial belief is typically inaccurate. Therefore, his subjectively optimal strategy is initially suboptimal. As time passes information about the true dynamics is accumulated and, depending on the compatibility of the belief with respect to the truth, the agent may eventually learn to optimize. We introduce the notion of relative entropy, which is a natural adaptation of the entropy of a stochastic process to the subjective set-up. We present conditions, expressed in terms of relative entropy, that determine whether the agent will eventually learn to optimize. It is shown that low entropy yields asymptotic optimal behavior. In addition, we present a notion of pointwise merging and link it with relative entropy.

Keywords: Relative entropy; Sequential decision problems; Optimization

* Corresponding author. Tel.: +972-36408822; E-mail: lehrer@math.tau.ac.il

[1] This paper is partially based on a manuscript, by the same authors, entitled ''When is a Bayesian agent fortunate: relative entropy and learning''.

[2] E-mail: rann@ie.technion.ac.il.

0304-4068/00/$ - see front matter
PII: S0304-4068(99)00027-0

1. Introduction

Consider an agent holding a subjective belief over the distribution of an infinite stream of outcomes. At each stage the agent chooses an action and receives a payoff which depends on his action, the new outcome, and perhaps on all previous outcomes. The agent updates his belief in a Bayesian manner any time he observes a realized outcome. While the agent maximizes payoffs according to his belief, the realized outcomes are chosen according to the unknown real distribution. Typically, the agent's belief is inaccurate and his strategy is suboptimal. We say that the agent learns to optimize if his subjectively optimal strategy becomes optimal in the long run.

In many examples a subjective agent may never learn to optimize. In other examples agents always learn and there are cases where agents may or may not learn, depending on the specific realized sequence of outcomes. A fundamental problem in the Bayesian learning literature is to identify conditions under which learning occurs.

In an inspiring example, Blume and Easley (1992) consider a multi-stage economy, governed by an unknown stationary stochastic process. Each agent has a ''misspecified'' model (using their terminology). This means that the agents are undecided between two possible models regarding the stochastic process. Blume and Easley show that the logarithm of the likelihood ratio between a model's distribution and the true distribution, namely the relative entropy, is the critical parameter in determining whether learning occurs. In this paper we show how relative entropy is connected with learning to optimize in a more general context than implied by the example of Blume and Easley.

For every possible string of outcomes we consider the sequence of likelihood ratios between the agent's belief and the truth. We use this sequence to define the agent's relative entropy. We show that if the relative entropy is close to zero, then, in the long run, the payoff to a patient agent is almost equal to the payoff generated by the true optimal strategy.

The link we use to connect optimality and entropy is the notion of merging of opinions, originated in Blackwell and Dubins (1962). This notion captures the idea that a subjective belief becomes closer to the true distribution as sufficient data is observed. Kalai and Lehrer (1994) and later Lehrer and Smorodinsky (1996) introduce weaker versions of merging. These papers take a global approach and study the issue of merging on a set of full measure.

It turns out that the existing notions of merging are too strong for the optimization results discussed here. For our purposes we need to extend the merging concept in two ways. First, we introduce a pointwise notion of merging which allows us to study learning and optimization on specific sequences of outcomes. That is, such merging allows for learning to occur on some sequences of outcome while for other sequences merging and optimization fail. Second, since the focus of the paper is on ε-optimization, we resort to approximate merging (i.e.,

convergence of the subjective belief to the truth up to an ε, on most stages of the sequential decision process, while allowing for possibly significant mistakes on other stages). [3]

Merging is used as a link between entropy and optimality in the following manner: First, it is shown that whenever the relative entropy is close to zero approximate merging occurs. This, in turn, implies that the agent learns to optimize. A connection between entropy and optimization, in a different spirit, has been obtained by Sandroni (1996). He studies a model of asset pricing, similar to the well known ''Fruit Tree'' model of Lucas (1978). In Sandroni's model the agents are endowed with a subjective belief over the distribution of future prices and dividends. Sandroni ties down the agents' relative entropy to the long-run distribution of wealth. He shows that the agents with the least entropy are the only ones who remain in the market in the long-run.

As in other fields, the notion of relative entropy measures randomness and disorder. Clausis introduced entropy into thermodynamics in 1854. Later, Shannon (1948) made use of it in Information Theory to measure the capacity of a communication device. Kolmogorov (1958) carried it over to stochastic processes. Our definition is a natural extension of Kolmogorov's.

Entropy has been previously introduced into economic theory in a few studies. Lehrer (1988) used it to measure the richness of the information available to a boundedly rational player. Neyman and Okada (1996) used entropy as a measurement of the complexity of a strategy, as did Bollt and Jones (1996).

The paper is organized as follows. Section 2 provides motivating examples, Section 3 introduces the definitions of relative entropy and pointwise merging and establishes the linkage between relative entropy, merging and optimization. Proofs are given in Section 4 and Section 5 concludes.

2. Motivating examples

In the first four examples we focus on the following decision problem. Consider a decision maker (denoted DM) who takes, at every stage, one of three actions, a, b or c and receives a payoff. For simplicity let us assume that there are two states H and T and the payoffs are as follows: $u(a,H)=3$, $u(a,T)=0$, $u(b,H)=u(b,T)=2$, $u(c,H)=0$ and $u(c,T)=3$.

Nature tosses a fair coin at every round and chooses its state accordingly. In other words, nature evolves according to an i.i.d. process (typically, the state of

[3] The motivation for studying merging of opinions in the last decade has been the game-theoretic learning literature. Kalai and Lehrer (1993) and Nyarko (1994), used the notion of merging to show that rational learning entails convergence to equilibrium in a subjective set-up. Sandroni (1998b) studies merging and learning in finitely repeated games, and Lehrer and Smorodinsky (1997) broaden the class of examples under which rational learning leads to equilibrium.

nature is stochastically determined by the history of previous realized states). Obviously, the optimal strategy for an expected utility maximizer is to choose action b, resulting in a payoff of 2.

The DM, not knowing the true distribution, has some initial subjective belief regarding the future realizations. At each round he observes the realized state, updates his belief in a Bayesian manner and takes a decision which maximizes his (subjective) expected payoff.

In the first and third examples the relative entropy is not zero and the DM fails to optimize even in the long run. In contrast, the relative entropy in the second and the fourth examples is zero and indeed the DM learns to optimize.

Example 1. Assume that the DM believes that nature follows an i.i.d. process, but instead of the real probabilities he attributes to H the probability $1/4$ and to T the probability $3/4$. The DM will stay firm on his belief regardless of the observed history, and will always take the inferior action c. It turns out that the subjective belief, induced by $(1/4, 3/4)$, has a non-zero relative entropy with respect to the real one, induced by $(1/2, 1/2)$.

Example 2. As in the previous example, the DM believes that nature follows an i.i.d. process, but does not know the coin's parameter. Suppose that out of complete ignorance, the DM believes that the coin's parameter is selected according to the uniform distribution over the interval $[0,1]$. This induces a probability distribution over the set of all sequences consisting of H and T. After a while, the DM will observe histories where the frequency of H is nearly $1/2$. Therefore, the posterior over the set of parameters will be concentrated around $1/2$. As a result, the DM will eventually take the superior action b. This phenomenon, as we shall show, is explained by the zero relative entropy of the belief with respect to the truth.

Example 3. The third situation we describe is similar to the example by Blume and Easley (1992), referred to in Section 1. Consider an agent which, as before, believes that nature follows an i.i.d process (i.e., uses a coin). The agent believes that the parameter of the coin must be either $1/3$ or $3/4$, and puts some prior probability p on the former and $1-p$ on the latter. The analysis of the long run behavior is not straightforward. Does the agent's belief converge to the mid-point between the two coins, leading to optimal decisions? Does the agent swing back and forth between the two extreme possibilities, or rather, does the belief converge to one of the two parameters?

Blume and Easley show that the answer to this question can be given by comparing the *relative entropy* of both coins with respect to the true one. Since the relative entropy of the coin $1/3$ is closer to zero than the relative entropy of the coin $3/4$, the DM will eventually believe (with probability one) in the former possibility (i.e., $1/3$) inducing him to take the sub-optimal action a.

Example 4. Suppose that for any finite history of length t, with m heads, the agent's subjective probability is $\frac{2^{t+1}+3^{t-m}+3^m}{4^{t+1}}$. [4] We shall show in the sequel that the relative entropy of this belief with respect to the truth (the distribution induced by a fair coin) is zero. Our main theorem will imply that the agent will actually optimize in the long run.

In all the examples so far, we focused on standard sequential problems where the payoff function is deterministic. Our last example demonstrates the possibility of a stochastic payoff function, for which our results hold as well. In this example the payoff at each stage depends not only on the current action and outcome, but also on all previous outcomes.

Example 5. An agent facing a random walk on the integers has to decide at each stage on an action, $a \in A$. Assume that at each stage the random walk either goes one step right or left and that the agent's stage payoff depends only on the random walk's location. That is, the stage payoff function is a function, u: $A \times \mathbb{N} \to \mathbb{R}$. Here, the payoff function depends on the entire history of outcomes, yet the main theorem still applies.

3. Model and results

In this section we provide a simple model of decision making under uncertainty. We then introduce the notion of relative entropy of a subjective DM and eventually, in our main theorem, we point out the connection between entropy and optimality.

3.1. A model of sequential decision making

Let $t = 1,2,3,\ldots$ denote time and let Ω be the finite set of nature's possible outcomes at time t. Let $\Omega^{\mathbb{N}}$ be the set of all infinite strings of outcomes, and endow $\Omega^{\mathbb{N}}$ with the natural σ-algebra, $\mathcal{F}$, generated by all finite cylinders. That is, let $\mathcal{F}_t$ be the finite algebra generated by all finite strings in Ω^t (the Cartesian product of Ω with itself t times), and let $\mathcal{F} = \sigma(\cup_{t=1}^{\infty}\mathcal{F}_t)$. The event $P_t(\omega)$ is the smallest event in $\mathcal{F}_t$ that contains ω. Denote by μ the probability measure on $(\Omega^{\mathbb{N}}, \mathcal{F})$ according to which nature chooses $\omega \in \Omega^{\mathbb{N}}$.

At time t, prior to the revelation of nature's outcome $\omega_t \in \Omega_t$, an agent takes an action from a finite set A. The agent's stage-payoff at time t, u_t, depends on the agent's action, nature's outcome at this time and on all previous outcomes. In

[4] It is easy to show that this formula generates a well defined exchangeable distribution over $\{H, T\}^{\mathbb{N}}$.

other words, u_t: $A \times \Omega^{\mathbb{N}} \to \mathbb{R}$, such that $u_t(a, \cdot)$ is $\mathcal{F}_t$-measurable for any $a \in A$. Let $u = \{u_t\}_{t=1}^{\infty}$ denote the sequence of stage-payoff functions. The function u is called M-bounded if $|u_t(a, \omega)| < M$ for any $\omega \in \Omega$, $t \in \mathbb{N}$ and $a \in A$. An agent's pure strategy prescribes an action to be taken at each stage t, given the relevant information, i.e., all $t-1$ past actions and $t-1$ past outcomes. Formally, a pure strategy is a function $f: \cup_{t=1}^{\infty} (A \times \Omega)^{t-1} \to A$ (with $(A \times \Omega)^0$ interpreted as the set containing the null history). Let S denote the set of all pure strategies.

We restrict the discussion to agents who are expected utility maximizers. For any probability measure μ on $(\Omega^{\mathbb{N}}, \mathcal{F})$ we denote by $\mu_t(\cdot|\omega_1, \ldots, \omega_{t-1})$ the marginal on the tth coordinate of the conditional distribution $\mu(\cdot|\omega_1, \ldots, \omega_{t-1})$. We say that $f \in S$ is μ-optimal if $f(\omega_1, \ldots, \omega_{t-1}) \in \operatorname{argmax}_{a \in A} E_{\mu_t(\cdot|\omega_1, \ldots, \omega_{t-1})}(u_t(a, \omega))$, for all t and all $\omega \in \Omega$. In words, a strategy is μ-optimal if it maximizes the expected utility under μ. We use the notation f_μ to denote an arbitrary μ-optimal strategy.

3.2. Long-run payoffs

Given a sequence of stage payoffs, $\{u_t\}_{t=1}^{\infty}$, let $U_T = 1/T \sum_{t=1}^{T} u_t$ be the payoff of the finite horizon decision problem, of length T. For an infinite horizon decision problem let $U_r = (1-r) \sum_{t=1}^{\infty} r^t u_t$ be the discounted sum of payoffs.

For a fixed sequence of payoff function, u, we can think of U_T and U_r as functions from $S \times \Omega^{\mathbb{N}}$ into $\mathbb{R}$.

3.3. Subjective optimization

In this paper we depart from the traditional approach to optimization, where agents optimize with respect to the true distribution, μ. We consider an agent endowed with a subjective belief, $\tilde{\mu}$. Since the agent's belief typically does not coincide with the truth, an agent's strategy might be initially suboptimal with respect to μ. However, as the agent accumulates information when more observations become available, he may learn to optimize. As illustrated in Example 1 this is not necessarily the case.

Our goal in this paper is to identify conditions on the relation between the belief and the truth that ensure optimality in the long run. In other words we seek conditions under which the long run payoffs generated by f_μ will be equal those generated by $f_{\tilde{\mu}}$. We phrase these conditions solely in terms of the relative entropy of the belief with respect to the truth.

3.4. Entropy and optimality

Relative entropy is defined to be the following random variable which depends on likelihood ratios of the posteriors.

Definition 1. The *relative entropy* of $\tilde{\mu}$ with respect to μ at ω is [5]

$$h_{\mu}^{\tilde{\mu}}(\omega) = \liminf_{T\to\infty} \frac{1}{T} \log \frac{\tilde{\mu}(P_t(\omega))}{\mu(P_t(\omega))},$$

where $0/0 = 0$.

Remark 1. *A straightforward corollary of Lemma 1 in Lehrer and Smorodinsky (1996) is that $h_{\mu}^{\tilde{\mu}}(\omega)$ is μ-almost everywhere non-positive.*

Example 4 revisited. We calculate $h_{\mu}^{\tilde{\mu}}(\omega)$ of Example 4 on the event consisting of the sequences where the asymptotic relative frequency of Heads is $1/2$. Note that the μ-probability of this event is one.

$$h_{\mu}^{\tilde{\mu}}(\omega) = \liminf_{T\to\infty} \frac{1}{T} \log \frac{\left(\dfrac{2^{T+1} + 3^{T-\frac{T}{2}} + 3^{\frac{T}{2}}}{4^{T+1}} \right)}{2^T}$$

$$= \liminf_{T\to\infty} \frac{1}{T} \log \left(\frac{1}{2} + \frac{1}{2} \frac{\sqrt{3}^T}{2} \right).$$

Therefore,

$$h_{\mu}^{\tilde{\mu}}(\omega) \geq \liminf_{T\to\infty} \frac{1}{T} \log \frac{1}{2} = 0.$$

On the other hand,

$$h_{\mu}^{\tilde{\mu}}(\omega) \leq \liminf_{T\to\infty} \frac{1}{T} \log \left(\frac{1}{2} + \frac{1}{2} \frac{\sqrt{3}}{2} \right) = 0.$$

We conclude that $h_{\mu}^{\tilde{\mu}}(\omega) = 0$ μ-a.e. Our main result asserts that if the relative entropy of the DM's belief is close to zero, then his actual payoff is close to the payoff generated by the true optimal strategy. That is, $f_{\tilde{\mu}}$ generates a payoff close to the payoff generated by f_{μ}.

[5] (a) In the definition we use lim inf as the limit need not necessarily exist. For further discussion see Section 5.3. (b) Note that $h_{\mu}^{\tilde{\mu}}(\omega)$ is defined also in cases where $\mu(P_t(\omega)) = 0$ and $\tilde{\mu}(P_t(\omega)) > 0$, or when $\mu(P_t(\omega)) > 0$ and $\tilde{\mu}(P_t(\omega)) = 0$. In these cases the relative entropy may take the values infinity or minus infinity.

Our main result is the following.

Theorem 1. *For any pair of measures* μ *and* $\tilde{\mu}$ *on* $(\Omega^{\mathbb{N}}, \mathcal{F})$, *any* $M > 0$ *and* $\varepsilon > 0$ *there exists* $\delta > 0$ *and* $\overline{\Omega} \in \mathcal{F}$, $\mu(\overline{\Omega}) \geq 1 - \varepsilon$ *a stage* T_0 *and a discount factor* $1 > r_0$ *such that for any M-bounded payoff function u, any* $\omega \in \overline{\Omega}$, *if* $h_{\mu}^{\tilde{\mu}}(\omega) > -\delta$, *then: (a)* $|U_T(f_{\mu}, \omega) - U_T(f_{\tilde{\mu}}, \omega)| < \varepsilon$ *for any* $T > T_0$; *and (b)* $|U_r(f_{\mu}, \omega) - U_r(f_{\tilde{\mu}}, \omega)| < \varepsilon$ *for any* $r > r_0$.

3.5. Pointwise merging of opinions

In order to prove Theorem 1 we first introduce a variation of the notion of merging of opinions. Next, we link this notion to entropy (Proposition 1) and to optimality (Proposition 2).

Blackwell and Dubins (1962) introduce the notion of merging of opinions as a way to express long run accuracy of beliefs. This notion was weakened by Kalai and Lehrer (1993) and later by Lehrer and Smorodinsky (1996) who use it to study learning in games. [6] All these notions were global notions and referred to phenomena occurring on a set of measure one. Here we use a similar idea which is weaker than all its predecessors on the one hand, and is a point-wise notion on the other hand. For a subset $M \subset \mathbb{N}$ we use $\bar{d}(M)$ to denote the upper density of M. [7]

Definition 2. We say that a measure $\tilde{\mu}$, η-merges to μ on ω if

$$\bar{d}\{t| \text{ there exists } B \in \mathcal{F}_{t+1} | \tilde{\mu}(B|\mathcal{F}_t)(\omega) - \mu(B|\mathcal{F}_t)(\omega)| > \eta\} \leq \eta.$$

In other words, $\tilde{\mu}$ η-merges to μ if the accuracy of the one-step-ahead forecasts, according to $\tilde{\mu}$, is accurate up to η on a large proportion of stages.

The first step towards establishing our result is to show the connection between entropy and η-merging.

Proposition 1. *Fix* μ *and* $\tilde{\mu}$. *For every* $\eta > 0$ *exists* $\delta > 0$ *and a set* $\Omega_1 \in \mathcal{F}$ *with* μ*-measure* 1 *such that for any* $\omega \in \Omega_1$ $h_{\mu}^{\tilde{\mu}}(\omega) > -\delta$ *implies* $\tilde{\mu}$ η*-merges to* μ *on* ω.

The second step establishes the connection of η-merging to optimal strategies.

[6] Another reference connecting merging to learning in games is Sandroni (1998a).

[7] The *upper density* of M is $\bar{d}(M) = \limsup_{t \to \infty} |M \cap \{1, \ldots, t\}| / t$. The *lower density* of M is $\underline{d}(M) = \liminf_{t \to \infty} |M \cap \{1, \ldots, t\}| / t$.

Proposition 2. *For any* $M>0$ *and* $\varepsilon>0$ *there exists* $\eta=\eta(\varepsilon,M)>0$, *a set* $\Omega_2\in\mathscr{F}$, $\mu(\Omega_2)>1-\varepsilon$, *a time* T_0 *and a discount factor* $1>r_0>0$, *such that for any M-bounded payoff function u and any* $\omega\in\Omega_2$, *if* $\tilde{\mu}$ η-*merges to* μ *on* ω, *then:*
(a) $|U_T(f_\mu,\omega)-U_T(f_{\tilde{\mu}},\omega)|<\varepsilon$ *for any* $T>T_0$; *and (b)* $|U_r(f_\mu,\omega)-U_r(f_{\tilde{\mu}},\omega)|<\varepsilon$ *for any* $r>r_0$.

4. Proofs

In this section we provide proofs for Propositions 1 and 2. The main result follows as a corollary.

Proof of Proposition 1. This proof uses techniques developed in Lehrer and Smorodinsky (1996). Let $\lambda_t^k(\omega)=\max\{\tilde{\mu}_t(\omega|\mathscr{F}_{t-1}),\ 1/k\cdot\mu_t(\omega|\mathscr{F}_{t-1})\}$. Denote, $Y_t^k(\omega)=\log\ \lambda_k^t(\omega)/\mu_t(\omega|\mathscr{F}_{t-1})$. Note that $Y_t^k(\omega)$ is bounded by k. Consider the variable $Z_t^k(\omega)=Y_t^k(\omega)-E(Y_t^k|\mathscr{F}_{t-1})(\omega)$. Since the variable Y_t^k is bounded, so is the variable Z_t^k. In particular the variance of Z_t^k is finite for every t. Moreover, $E(Z_t^k|\mathscr{F}_{t-1})=0$ and for every $s<t$ $E(Z_t^kZ_s^k)=E(E(Z_t^kZ_s^k|\mathscr{F}_{t-1}))=E(E(Z_t^k|\mathscr{F}_{t-1})Z_s^k)=0$. Thus, Z_t^k and Z_s^k are uncorrelated and so strong law of large numbers for uncorrelated variables applies (see for instance, Feller, 1971, p. 243). Therefore, there exists an event $D_k\in\mathscr{F}$ with $\mu(D_k)=1$ such that for any ω in D_k

$$\frac{1}{t}\sum_{j=1}^{t}Y_j^k(\omega)-E\left(Y_j^k|\mathscr{F}_{j-1}\right)(\omega)\to 0.$$

Let $\Omega_1=\cap_{k=1}^{\infty}D_k$, obviously $\mu(\Omega_1)=1$. We claim that the event Ω_1 is the one referred to in the proposition. Suppose this is not true, then there exists $\eta>0$ and a sequence of $(\omega^i)_{i=1}^{\infty}\in\Omega_1$ such that $h(\omega^i)=h_\mu^{\tilde{\mu}}(\omega^i)\to_{i\to\infty}0$ and $\tilde{\mu}$ does not η-merge to μ on ω^i. In other words for any i there is a set of integers $N(i)\subset\mathbb{N}$ satisfying $\bar{d}(N(i))\geq\eta$ and for any $t\in N(i)$ there exists an event $B=B(t,i)$ in $\mathscr{F}_{t+1}$ satisfying

$$|\mu\left(B|\mathscr{F}_t\right)(\omega^i)-\tilde{\mu}\left(B|\mathscr{F}_t\right)(\omega^i)|>\eta.$$

Note that the complement of B satisfies the same inequality. Thus, we may decompose both B and its complement to the atoms of $\mathscr{F}_{t+1}$ in order to obtain,

$$\sum_{P_{t+1}}|\mu\left(P_{t+1}|\mathscr{F}_t\right)(\omega^i)-\tilde{\mu}\left(P_{t+1}|\mathscr{F}_t\right)(\omega^i)|>2\eta,$$

where the sum runs over all atoms P_{t+1} of $\mathscr{F}_{t+1}$. Since $\sum_{P_{t+1}}\mu(P_{t+1}|\mathscr{F}_t)(\omega^i)=\sum_{P_{t+1}}\tilde{\mu}(P_{t+1}|\mathscr{F}_t)(\omega^i)$ one can split $\sum_{P_{t+1}}|\mu(P_{t+1}|\mathscr{F}_t)(\omega^i)-\tilde{\mu}(P_{t+1}|\mathscr{F}_t)(\omega^i)|$ into two summations as follows. $\sum_{P_{t+1}}|\mu(P_{t+1}|\mathscr{F}_t)(\omega^i)-\tilde{\mu}(P_{t+1}|\mathscr{F}_t)(\omega^i)|=\sum'[\mu(P_{t+1}|\mathscr{F}_t)(\omega^i)-\tilde{\mu}(P_{t+1}|\mathscr{F}_t)(\omega^i)]+\sum''[\tilde{\mu}(P_{t+1}|\mathscr{F}_t)(\omega^i)-\mu(P_{t+1}|\mathscr{F}_t)(\omega^i)]$

where the first summation (Σ') runs over all P_{t+1} of $\mathscr{F}_{t+1}$ with $\mu(P_{t+1}|\mathscr{F}_t)(\omega^i) < \tilde{\mu}(P_{t+1}|\mathscr{F}_t)(\omega^i)$ and the other (Σ'') runs over all P_{t+1} of $\mathscr{F}_{t+1}$ with $\mu(P_{t+1}|\mathscr{F}_t)(\omega^i) > \tilde{\mu}(P_{t+1}|\mathscr{F}_t)(\omega^i)$. Moreover, these two summations are equal. As $\lambda_t^k(\omega^i) \le \tilde{\mu}(P_{t+1}|\mathscr{F}_t)(\omega^i)$, we obtain that $\sum_{P_{t+1}} |\mu(P_{t+1}|\mathscr{F}_t)(\omega^i) - \lambda_t^k(\omega^i)|$ is greater than or equal to the first summation (Σ'). Thus,

$$\sum_{P_{t+1}} |\mu(P_{t+1}|\mathscr{F}_t)(\omega^i) - \lambda_t^k(\omega^i)| > \eta \ \forall i \ \forall t \in N(i).$$

By part (ii) in Lemma 3 of Lehrer and Smorodinsky (1996), there exists $\delta > 0$ such that if $\sum_{P_{t+1}} \lambda_t^k(\omega^i)$ is sufficiently close to 1 (which is guaranteed when k is large enough), then

$$E(Y_{t+1}^k|\mathscr{F}_t)(\omega^i) < -\delta \ \forall t \in N(i).$$

Now fix $\alpha, \beta > 0$ such that

$$-\delta\frac{\eta}{2} + \left(1 - \frac{\eta}{2}\right)\alpha \le -\beta < 0.$$

Part (i) of Lemma 3 in Lehrer and Smorodinsky (1996) ensures that we can choose k large enough to satisfy

$$E(Y_{t+1}^k|\mathscr{F}_t)(\omega^i) \le \alpha \ \forall i \ \forall t \in \mathbb{N}.$$

For each i extract an infinite sequence $\bar{N}(i) \subset N(i)$ having the following two properties. (a) $t \in \bar{N}(i)$ implies $|\{\ell | \ell \le t, \ell \in N(i)\}|/t \ge (\eta/2)$. (This is possible since the upper density of $N(i)$ is at least η). (b) $1/t\sum_{j=1}^t Y_j^k(\omega^i) \le 1/t\sum_{j=1}^t E(Y_j^k|\mathscr{F}_{t-1})(\omega^i) + \beta/2$ (This is possible due to the law of large numbers).

For $t \in \bar{N}(i)$ the following then holds.

$$\begin{aligned}
\frac{1}{t}\log\frac{\tilde{\mu}(P_t(\omega^i))}{\mu(P_t(\omega^i))} &= \frac{1}{t}\sum_{j=1}^t \log\frac{\tilde{\mu}_j(\omega^i)}{\mu_j(\omega^i)} \le \frac{1}{t}\sum_{j=1}^t Y_j^k(\omega^i) \\
&\le \frac{1}{t}\sum_{j=1}^t E(Y_j^k|\mathscr{F}_{j-1})(\omega^i) + \frac{\beta}{2} \\
&\le \frac{1}{t}\sum_{\substack{j \le t \\ j \in N(i)}} -\delta + \frac{1}{t}\sum_{\substack{j \le t \\ j \notin N(i)}} \alpha + \frac{\beta}{2} \le -\frac{1}{t}\left(n\frac{\eta}{2}\cdot\delta\right) \\
&\quad + \frac{1}{t}t\left(1 - \frac{\eta}{2}\right)\alpha + \frac{\beta}{2} \\
&= -\delta\frac{\eta}{2} + \left(1 - \frac{\eta}{2}\right)\alpha + \frac{\beta}{2} \le -\frac{\beta}{2}.
\end{aligned}$$

Therefore, for any i and any $t \in \bar{N}(i)$

$$\frac{1}{t}\log\frac{\tilde{\mu}\left(P_t(\omega^i)\right)}{\mu\left(P_t(\omega^i)\right)} \leq -\frac{\beta}{2} < 0,$$

which in turn implies $h(\omega^i) \leq -\beta/2$ contradicting our hypothesis. ■

Before turning to the proof of Proposition 2, we state some auxiliary definitions and results.

Let $\Delta(\Omega)$ denote the space of all probability measures over Ω. For every $p \in \Delta(\Omega)$, $a \in A$ and u: $A \times \Omega \to \mathbb{R}$, let $E_p(u(a)) = \sum_{\omega \in \Omega} p(\omega)u(a, \omega)$ and let $E_p = \max_{a \in A} E_p(u(a))$

Lemma 1. *$E_p(u(a))$ and E_p are uniformly continuous, across all M-bounded payoff functions, w.r.t. the sup-norm topology.*

Proof. It is easily seen that for any fixed $a \in A$, $E_p(u(a))$ is uniformly continuous. Thus, E_p is uniformly continuous as a maximum of finitely many uniformly continuous functions. ■

The next claim we need is a trivial observation regarding sequences in $\mathbb{R}$.

Lemma 2. *Let $\{a_n\}$ and $\{b_n\}$ be two infinite bounded sequences in $\mathbb{R}$. For every $\varepsilon > 0$ there exists $\eta_1 > 0$ s.t. if $\mathbb{N}_1 \subset \mathbb{N}$ satisfies $\underline{d}(\mathbb{N}_1) > 1 - \eta_1$ and $\overline{\lim}_{n \in \mathbb{N}_1}|a_n - b_n| < \eta_1$*, then

$$\overline{\lim}\left|\frac{1}{N}\sum_{n=1}^{N} a_n - \frac{1}{N}\sum_{n=1}^{N} b_n\right| < \varepsilon.$$

The proof of Lemma 2 is straightforward and is therefore omitted. We are now ready to prove Proposition 2.

Proof of Proposition 2. (b) follows from (a) by standard arguments. To prove (a) note that the strong law of large numbers implies that the following is satisfied μ-a.e.

$$\lim_{T\to\infty}\left|\frac{1}{T}\sum_{t=1}^{T} u_t\left(f_\mu, \omega_t\right) - E_\mu\left(u_t\left(f_\mu\right)\right)\right| = 0$$

and

$$\lim_{T\to\infty}\left|\frac{1}{T}\sum_{t=1}^{T} u_t\left(f_{\tilde{\mu}}, \omega_t\right) - E_\mu\left(u_t\left(f_{\tilde{\mu}}\right)\right)\right| = 0.$$

So by Egoroff's Theorem one can choose T_1 large enough such that on a set $B \in \mathscr{F}$ of μ-measure $1 - \varepsilon/2$, if $\omega \in B$ and $T > T_1$, then

$$\left| \frac{1}{T} \sum_{t=1}^{T} u_t(f_\mu, \omega_t) - E_\mu\big(u_t(f_\mu)\big) \right| < \varepsilon/4 \tag{1}$$

and

$$\left| \frac{1}{T} \sum_{t=1}^{T} u_t(f_{\tilde{\mu}}, \omega_t) - E_\mu\big(u_t(f_{\tilde{\mu}})\big) \right| < \varepsilon/4. \tag{2}$$

By Lemma 1, for any $\eta_1 > 0$ one can choose η such that $\sup_{D \in \mathscr{F}_{t+1}} |\mu(D|\mathscr{F}_t) - \tilde{\mu}(D|\mathscr{F}_t)| < \eta$ implies

$$|E_\mu\big(u_t(f_\mu)\big) - E_{\tilde{\mu}}\big(u_t(f_{\tilde{\mu}})\big)| = |E_\mu - E_{\tilde{\mu}}| < \eta_1/2$$

and

$$|E_{\tilde{\mu}}\big(u_t(f_{\tilde{\mu}})\big) - E_\mu\big(u_t(f_{\tilde{\mu}})\big)| < \eta_1/2.$$

These inequalities yield

$$|E_\mu\big(u_t(f_\mu)\big) - E_\mu\big(u_t(f_{\tilde{\mu}})\big)| < \eta_1.$$

Without loss of generality we can choose $\eta \le \eta_1$. Therefore, if $\tilde{\mu}$ η-merges to μ on ω there exists a sequence of stages $\mathbb{N}_1 \subset \mathbb{N}$ s.t $\bar{d}(\mathbb{N}_1) > 1 - \eta \ge 1 - \eta_1$ and for all $t \in \mathbb{N}_1$

$$|E_\mu\big(u_t(f_\mu)\big) - E_\mu\big(u_t(f_{\tilde{\mu}})\big)| < \eta_1.$$

Now choose η_1 small enough, such that the sequences $\{E_\mu(u_t(f_\mu))\}_{t=1}^{\infty}$ and $\{E_\mu(u_t(f_{\tilde{\mu}}))\}_{t=1}^{\infty}$ satisfy the conditions of Lemma 2 for $\varepsilon/4$. This implies

$$\limsup_{t \to \infty} \left| \frac{1}{T} \sum_{t=1}^{T} E_\mu\big(u_t(f_\mu)\big) - E_\mu\big(u_t(f_{\tilde{\mu}})\big) \right| < \varepsilon/4.$$

Again, by Egoroff's theorem there exists $C \in \mathscr{F}$ with $\mu(C) > 1 - \varepsilon/2$ and T_2 such that for all $T > T_2$ and all $\omega \in C$ $\tilde{\mu}$ η-merges to μ on ω implies

$$\left| \frac{1}{T} \sum_{t=1}^{T} E_\mu\big(u_t(f_\mu)\big) - E_\mu\big(u_t(f_{\tilde{\mu}})\big) \right| < \varepsilon/2. \tag{3}$$

Let Ω_2 be the intersection of B and C. Obviously $\mu(\Omega_2) > 1 - \varepsilon$. Let $T_0 = \max(T_1, T_2)$. By inequalities (1), (2) and (3) we can deduce that for any $\omega \in \Omega_2$ if $\tilde{\mu}$ η-merges to μ on ω, then

$$\begin{aligned} |U_T(f_\mu, \omega) - U_T(f_{\tilde{\mu}}, \omega)| &= \left| \frac{1}{T} \sum_{t=1}^{T} \big(u_t(f_\mu, \omega_t) - u_t(f_{\tilde{\mu}}, \omega_t)\big) \right| \\ &\le \left| \frac{1}{T} \sum_{t=1}^{T} E_\mu\big(u_t(f_\mu)\big) - E_\mu\big(u_t(f_{\tilde{\mu}})\big) \right| + \varepsilon/4 + \varepsilon/4 \\ &\le \varepsilon/2 + \varepsilon/4 + \varepsilon/4 = \varepsilon. \end{aligned}$$

■

Proof of Main Theorem. Let $\overline{\Omega}$ be the intersection of the full measure set Ω_1 of Proposition 1 and the set Ω_2 of measure $1-\varepsilon$ of Proposition 2. Given $\varepsilon>0$ use Proposition 2 to find η such that η-merging will entail ε-optimality. Now by Proposition 1 we can find the minimal entropy, δ, which will ensure η-merging. ■

5. Final remarks

5.1. An alternative definition

Our main result connects relative entropy with optimization in a sequential decision problem. The relative entropy we introduced is defined in terms of the likelihood ratio of the belief and the truth. An alternative approach is to define the entropy in terms of the expectation of the logarithm of the likelihood.

Formally, let $e_{\mu}^{\tilde{\mu}}(\omega)=\lim\ \inf_{T\to\infty}\ \frac{1}{T}E_{\mu}\left(\log\frac{\tilde{\mu}(P_t(\omega))}{\mu(P_t(\omega))}\right)$. Under standard boundedness conditions the law of large numbers ensures that $e_{\mu}^{\tilde{\mu}}(\omega)$ is μ-almost surely equal to $h_{\mu}^{\tilde{\mu}}(\omega)$, which implies that both definitions are essentially equivalent. Note that Blume and Easley (1992) and Sandroni (1996) use the latter definition.

5.2. Day to day optimization

Another interesting issue is with regards to the day to day loss of utility resulting from subjective optimization, when compared to objective optimization. It turns out that our results imply that on almost all days the agent loses no more than some ε, providing that the entropy is close to zero. The formal statement is given, without a proof, in the following theorem.

Proposition 3. *For any $\varepsilon>0$ there exists $\delta>0$ such that with μ-probability* 1, *if $h_{\mu}^{\tilde{\mu}}(\omega)>-\delta$, then the lower density of the set of stages, D, is at least $1-\varepsilon$, where $D=\{t;\ |E_{\mu(\cdot|\omega_1,\ldots,\omega_{t-1})}(u_t(f_{\mu}))-E_{\mu(\cdot|\omega_1,\ldots,\omega_{t-1})}(u_t(f_{\tilde{\mu}}))|<\varepsilon\}$.*

5.3. Existence of the limit

In the definition of relative entropy we used the limit inferior as the limit does not exist in general. As the following example shows, even when both measures, μ and $\tilde{\mu}$, are stationary, the limit need not necessarily exist.

Example 6. Let $\Omega=\{0,\ 1\}^{\mathbb{N}}$ and μ be the measure generated by a fair coin. Construct $\tilde{\mu}$ as follows. $\tilde{\mu}=\sum_{k=1}^{\infty}p_k\,\mu_k$, where μ_k is generated by flipping a fair coin t_k times and then repeating the same stream of t_k outcomes deterministically

time after time. The sequence $\{t_k\}$ is chosen so that $t_{k+1} - t_k \to \infty$ fast. Finally, the probability vector $\{p_k\}_{k=1}^{\infty}$ should satisfy $p_k / \sum_{j>k} p_j \to \infty$ fast. Note that for any k, μ_k is stationary and, therefore, $\tilde{\mu}$ is. When the observation at date $t = t_k + 1$ is different from that of date 1 (this will happen infinitely often, μ-a.e.) one obtains $1/t \log \tilde{\mu}(P_t)/\mu(P_t) \approx 1/t \log 0/(1/2)^t = -\infty$ (The $\approx$ sign comes from $p_k / \sum_{j>k} p_j \to \infty$). On the other hand, when $t = t_k$, $1/t \log \tilde{\mu}(P_t)/\mu(P_t) \approx 0$ for large enough k.

However, when μ is stationary and $\tilde{\mu}$ is Markovian of any order, the limit exists (see Lehrer and Smorodinsky, 1995).

Acknowledgements

The author gratefully acknowledges BSF grants 96/00043 and 97/113, NSF grant SBR-9730385, and Technion MANLAM and research promotion grants.

References

Blackwell, D., Dubins, L., 1962. Merging of opinions with increasing information. Annals of Mathematical Statistics 38, 882–886.

Blume, L., Easley, D., 1992. Rational expectations and rational learning. Notes from the Economic Theory Workshop in Honor of Roy Radner, Cornell University.

Bollt, M.E., Jones, M.A., 1996. Developing Symbolic Dynamics to Measure the Complexity of Repeated Game Strategies by Topological Entropy. Math Center discussion paper No. 1165, Northwestern University.

Feller, W., 1971. An Introduction to Probability Theory and Its Applications, Vol. II. Wiley, New York.

Kalai, E., Lehrer, E., 1993. Rational learning lead to Nash equilibrium. Econometrica 61, 1019–1045.

Kalai, E., Lehrer, E., 1994. Weak and strong merging of opinions. Journal of Mathematical Economics 23, 73–86.

Kolmogorov, A.N., 1958. New metric invariants of transitive dynamical systems and automorphisms of lebesgue spaces. Doklady Akademii Nauk SSSR 119, 861–864.

Lehrer, E., 1988. Repeated games with stationary bounded recall strategies. Journal of Economic Theory 46, 130–144.

Lehrer, E., Smorodinsky, R., 1995. When is a Bayesian agent fortunate: relative entropy and learning, manuscript.

Lehrer, E., Smorodinsky, R., 1996. Compatible measures and merging. Mathematics of Operations Research 21 (3), 697–706.

Lehrer, E., Smorodinsky, R., 1997. Repeated large games with incomplete information. Games and Economic Behavior 18, 116–134.

Lucas, R., 1978. Asset pricing in an exchange economy. Econometrica 46, 1429–1445.

Neyman, A., Okada, D., 1996. Strategic entropy and complexity in repeated games. Discussion Paper #104, Center For Rationality and Interactive Decision Theory.

Nyarko, Y., 1994. Bayesian learning leads to correlated equilibrium in normal form games. Economic Theory 4, 821–842.

Sandroni, A., 1996. Do markets favor agents able to make accurate predictions? Mimeo, Northwestern University.

Sandroni, A., 1998a. Necessary and sufficient conditions for convergence to Nash equilibrium: the almost absolute continuity hypothesis. Games and Economic Behavior 22, 121–147.

Sandroni, A., 1998b. Does rational learning lead to Nash equilibrium in finitely repeated games?. Journal of Economic Theory 78, 195–218.

Shannon, C., 1948. The mathematical theory of communication. Bell System Technical Journal 27, 379–423.