Available online at www.sciencedirect.com
ScienceDirect
Journal of Mathematical Economics 44 (2008) 24–32
JOURNAL OF Mathematical ECONOMICS
www.elsevier.com/locate/jmateco

Entropy bounds on Bayesian learning

Olivier Gossner [a], Tristan Tomala [b,*]

[a] *PSE, Paris, France and KSM-MEDS, Northwestern University, Evanston, USA*
[b] *CEREMADE, Université Paris Dauphine, Paris, France*

Received 24 October 2006; accepted 6 April 2007
Available online 24 April 2007

Abstract

An observer of a process $(\mathbf{x}_t)$ believes the process is governed by Q whereas the true law is P. We bound the expected average distance between $P(x_t|x_1, \ldots, x_{t-1})$ and $Q(x_t|x_1, \ldots, x_{t-1})$ for $t = 1, \ldots, n$ by a function of the relative entropy between the marginals of P and Q on the n first realizations. We apply this bound to the cost of learning in sequential decision problems and to the merging of Q to P.

AMS Classification: 62C10; 62B10; 91A26

Keywords: Bayesian learning; Repeated decision problem; Value of information; Entropy

1. Introduction

A bayesian agent observes the successive realizations of a process of law P, and believes the process is governed by Q. Following Blackwell and Dubins (1962), Q merges to P when the observer's updated law on the future of the process (given by Q) to the true one (given by P).

Different merging notions are defined depending on the type of convergence required, and merging theory studies conditions on Q and P under which Q merges to P under these different definitions (see e.g. Kalai and Lehrer, 1994; Lehrer and Smorodinsky, 1996). Merging theory has led to several applications such as calibrated forecasting (Kalai et al., 1999), repeated games with incomplete information (Sorin, 1999), and the convergence of plays to Nash equilibria in repeated games (Kalai and Lehrer, 1993).

When Q merges to P, the agent's predictions about the process become eventually accurate, but may be far from the truth during an arbitrarily long period of time. The present paper focuses on the average error in prediction during the first stages. Let e_n represent the (variational) distance between the agent's prediction and the true law of the stage n's realization of the process, and $(\bar{e}_n)_n$ denote the Cesaro means of $(e_n)_n$. Relying on Pinsker's inequality, we bound the expected average error in prediction up to stage n, $E_n = \mathbf{E}_P \bar{e}_n$, by a function of the relative entropy between the law P_n of the process and the agent's belief Q_n up to stage n. The advantage of the relative entropy expression is that it allows explicit computations in several cases.

We present applications to merging theory and to the cost of learning in repeated decision problems.

* Corresponding author.
E-mail address: tomala@ceremade.dauphine.fr (T. Tomala).

0304-4068/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jmateco.2007.04.006

A natural notion of merging is to require that the agent's expected average prediction errors vanish as time goes by. In this case we say that Q almost weakly merges on average (AWMA) to P. In Section 4 we relate AWMA to almost weak merging as introduced by Lehrer and Smorodinsky (1996). We show that AWMA holds whenever the relative entropy between P_n and Q_n is negligible with respect to n, i.e. $\lim_n d(P_n\|Q_n)/n = 0$ (Theorem 11) and derive rates of convergence for merging. It is worth noting that $\lim_n d(P_n\|Q_n)/n = 0$ does not imply absolute continuity of P with respect to Q, the only general condition in the literature for which a rate of convergence for merging is known (see Sandroni and Smorodinsky, 1999). We also derive conditions on a realization of the process for merging of Q to P to occur along this realization.

A decision maker in a n-stage decision problem facing a process of law P and whose belief on the process is Q is lead to use sub-optimal decisions rules, and suffers a consequential loss in terms of payoffs. In Section 5 we show that this loss can be bounded by expressions in E_n, thus in $d(P_n\|Q_n)$.

2. Preliminaries

Let X be a finite set and $\Omega = X^\infty$ be the set of sequences in X. An agent observes a random process $(\mathbf{x}_1, \ldots, \mathbf{x}_n, \ldots)$ with values in X whose behavior is governed by a probability measure P on Ω, endowed with the product σ-field. The agent believes that the process is governed by the probability measure Q.

Given a sequence $\omega = (x_1, \ldots, x_n, \ldots)$, $\omega_n = (x_1, \ldots, x_n)$ denotes the first n components of ω and we identify it with the cylinder generated by ω_n, i.e. the set of all sequences that coincide with ω up to stage n. We let $\mathcal{F}_n$ be the σ-algebra spanned by the cylinders at stage n and $\mathcal{F}$ the product σ-algebra on ω, i.e. spanned by all cylinders. We shall denote by $P(\cdot|\omega_n)$ the conditional distribution of $\mathbf{x}_{n+1}$ given ω_n under P (defined arbitrarily when $P(\omega_n) = 0$) and similarly for Q. By convention, $P(\cdot|\omega_0)$ is the distribution of $\mathbf{x}_1$.

The variational distance between two probability measures p and q over X is:

$$\|p - q\| = \sup_{A\subset X} |p(A) - q(A)| = \frac{1}{2}\sum_x |p(x) - q(x)|$$

Definition 1. The *variational distance* between P and Q at stage n at ω is:

$$e_n(P, Q)(\omega) = \|P(\cdot|\omega_{n-1}) - Q(\cdot|\omega_{n-1})\|$$

The *average variational distance* between P and Q at stage n at ω is:

$$\bar{e}_n(P, Q)(\omega) = \frac{1}{n}\sum_{m=1}^{n} e_m(P, Q)(\omega)$$

Recall that the relative entropy between p and q is

$$d(p\|q) = \sum_x p(x)\ln\frac{p(x)}{q(x)}$$

where $p(x)\ln(p(x)/q(x)) = 0$ whenever $p(x) = 0$, $(p(x) > 0, q(x) = 0 \Rightarrow p(x)\ln(p(x)/q(x)) = +\infty)$. This quantity is non-negative, equals zero if and only if $p = q$ and is finite if and only if $(q(x) = 0 \Rightarrow p(x) = 0)$. Pinsker's inequality bounds the variational distance by a function of the relative entropy as follows (see e.g. Cover and Thomas, 1991; Lemma 12.6.1, p. 300):

$$\|p - q\| \le \sqrt{\frac{d(p\|q)}{2}}$$

3. Relative entropy and average variational distance

Definition 2. The *local relative entropy* between P and Q at stage n at ω is:

$$d_n(P, Q)(\omega) = \sum_{m=1}^{n} d(P(\cdot|\omega_{m-1})\|Q(\cdot|\omega_{m-1}))$$

One has:

Proposition 3. *For each n and* ω:

$$\bar{e}_n(P,Q)(\omega) \le \sqrt{\frac{1}{2n}\, d_n(P,Q)(\omega)}$$

Proof. This follows directly from Pinsker's inequality and from the concavity of the square root function, by using Jensen's inequality. □

We denote by $E_n(P,Q)$ the expected average variational distance:

$$E_n(P,Q) := \mathbf{E}_P\, \bar{e}_n(P,Q)$$

We let P_n (resp. Q_n) be the marginal of P on the n first coordinates, i.e. P_n is the trace of P on $\mathcal{F}_n$. The expected average variational distance is bounded by the relative entropy as follows:

Proposition 4.

$$E_n(P,Q) \le \sqrt{\frac{1}{2n} d(P_n \| Q_n)}$$

Proof. From Proposition 3 and Jensen's inequality:

$$E_n(P,Q) \le \sqrt{\frac{1}{2n} \mathbf{E}_P\, d_n(P,Q)(\omega)}$$

Now, either by direct computation or by applying the chain rule for relative entropies (e.g. Cover and Thomas, 1991; Theorem 2.5.3, p. 23):

$$\mathbf{E}_P\, d_n(P,Q)(\omega) = d(P_n \| Q_n) \quad \square$$

4. Applications to merging theory

Merging theory studies whether the beliefs of the agent given by Q, updated after successive realizations of the process, converge to the true future distribution, given by P.

The next definitions are standard in merging theory (see Blackwell and Dubins, 1962; Kalai and Lehrer, 1993, 1994; Lehrer and Smorodinsky, 1996, 2000).

- *Q weakly merges to P* if $e_n(P,Q)(\omega)$ goes to zero P-a.s. as n goes to infinity.
- *Q almost weakly merges to P at ω* if $e_n(P,Q)(\omega)$ goes to zero on a full set of integers. That is, for every $\varepsilon > 0$, there is a set $N(\omega,\varepsilon)$ such that $\lim_n (1/n)|N(\omega,\varepsilon) \cap \{1,\ldots,n\}| = 1$ and $e_n(P,Q)(\omega) < \varepsilon$ for each $n \in N(\omega,\varepsilon)$.
- *Q almost weakly merges to P* if Q almost weakly merges to P at P-almost every ω.

The following shows that almost weak merging can be formulated through the average variational distance.

Proposition 5. *Q almost weakly merges to P at ω if and only if $\bar{e}_n(P,Q)(\omega)$ goes to zero as n goes to infinity.*

Proof. Let (a_n) be a bounded sequence of non-negative numbers. We say that (a_n) goes to zero with density one if for every $\varepsilon > 0$, the set M_ε of n's such that $a_n \le \varepsilon$ has density one: $\lim_n (1/n)|M_\varepsilon \cap \{1,\ldots,n\}| = 1$.

The proposition is a consequence of the following claim:

Claim 6. *The sequence (a_n) goes to zero with density one if and only if $(1/n)\sum_{m=1}^{n} a_m$ goes to zero as n goes to infinity.*

Proof of the claim. The Cesaro mean is:

$$\frac{1}{n}\sum_{m=1}^{n} a_m = \frac{1}{n}\sum_{m\,\in\, M_\varepsilon\cap\{1,\ldots,n\}} a_m + \frac{1}{n}\sum_{m\notin M_\varepsilon\cap\{1,\ldots,n\}} a_m$$

Letting $A = \sup_n a_n$, one has:

$$\varepsilon\left(1 - \frac{|M_\varepsilon \cap \{1,\ldots,n\}|}{n}\right) \leq \frac{1}{n}\sum_{m=1}^{n} a_m \leq \varepsilon + \left(1 - \frac{|M_\varepsilon \cap \{1,\ldots,n\}|}{n}\right) A$$

From the left-hand side, if $(1/n)\sum_{m=1}^{n} a_m$ goes to zero, for each $\varepsilon > 0$, $\lim_n (1/n)|M_\varepsilon \cap \{1,\ldots,n\}| = 1$, and from the right-hand side, if (a_n) goes to zero with density one, $(1/n)\sum_{m=1}^{n} a_m$ is less than 2ε for n large enough. □

We define a notion of merging in terms of expected average variational distance.

Definition 7. *Q* almost weakly merges on average (*AWMA*) to *P* if

$$\lim_n E_n(P, Q) = 0$$

AWMA amounts to the convergence of $\bar{e}_n(P, Q)(\omega)$ to 0 in L^1-norm or in *P*-probability and is weaker than *P*-almost sure convergence. AWMA is however not much weaker than almost weak merging, since the following proposition shows that if $E_n(P, Q)$ does not go to 0 too slowly, then *Q* almost weakly merges to *P*.

Proposition 8. *If $E_n(P, Q) \leq C/n^\alpha$ for $C > 0$ and $\alpha > 0$, then $\bar{e}_n(P, Q)(\omega) \to 0$, P-a.s.*

This is a direct consequence of the following lemma.

Lemma 9. *Let $(\mathbf{x}_n)$ be a sequence of random variables with range in $[0, 1]$ and let $\bar{\mathbf{x}}_n = (1/n)\sum_{m=1}^{n} \mathbf{x}_m$ be the arithmetic average. If $\mathbf{E}\bar{\mathbf{x}}_n \leq C/n^\alpha$ for $C > 0$ and $\alpha > 0$, then $\bar{\mathbf{x}}_n$ converges to 0 a.s.*

Proof. Let *p* be an integer. We first prove that $\bar{\mathbf{x}}_{n^p}$ converges to 0 a.s. when $p\alpha > 1$. It is enough to prove that for every $\varepsilon > 0$, $\sum_n P(\bar{\mathbf{x}}_{n^p} > \varepsilon) < +\infty$. By the Markov inequality,

$$P(\bar{\mathbf{x}}_{n^p} > \varepsilon) \leq \frac{\mathbf{E}(\bar{\mathbf{x}}_{n^p})}{\varepsilon} \leq \frac{C}{n^{p\alpha}\varepsilon}$$

Now for each integer *N*, there exists a unique *n* s.t. $n^p \leq N < (n+1)^p$. Then,

$$\bar{\mathbf{x}}_N = \frac{n^p}{N}\bar{\mathbf{x}}_{n^p} + \frac{N - n^p}{N} y$$

with $y \in [0, 1]$. Thus, $\bar{\mathbf{x}}_N \leq \bar{\mathbf{x}}_{n^p} + \left(1 + \frac{1}{n}\right)^p - 1$. □

Example 10 *(AWMA does not imply AWM).* Let $X = \{0, 1\}$ and construct *P* as follows. Take a family $(\mathbf{y}_k)_{k\geq 0}$ of independent random variables in *X* such that $P(\boldsymbol{y}_k = 0) = (1/k + 1)$, and set $\mathbf{x}_{2^k} = \mathbf{y}_k$. If $\boldsymbol{y}_k = 0$ then $\boldsymbol{x}_t = 0$ for $2^k < t < 2^{k+1}$. If $\boldsymbol{y}_k = 1$ then $(\mathbf{x}_t)_{2^k<t<2^{k+1}}$ are i.i.d. (1/2, 1/2) and independent of $\mathbf{x}_1, \ldots, \mathbf{x}_{2^k-1}$.

The belief *Q* is the distribution of an i.i.d. sequence of random variables (1/2, 1/2), so $q_t := Q(\boldsymbol{x}_t = 0|\omega_{t-1}) = 1/2$ for every *t* and ω_{t-1}.

We now compute $p_t := P(\boldsymbol{x}_t = 0|\omega_{t-1})$ and $e_t = 2|p_t - q_t|$. For $t = 2^k$, $p_t = (1/k + 1)$ and $e_t = 1 - (2/k + 1)$. For $2^k < t < 2^{k+1}$, $p_t = 0$ and $e_t = 1$ if $y_k = 0$, $p_t = 1/2$ and $e_t = 0$ if $\boldsymbol{y}_k = 1$.

On $\mathbb{N} - \cup_k\{2^k\}$, $\mathbf{E}_P e_t \to 0$ as *t* goes to $+\infty$, and $\mathbf{E}_P e_t \leq 1$ on $\cup_k\{2^k\}$. Therefore $\mathbf{E}_P \bar{e}_t \to 0$, and AWMA holds.

On the other hand, by Borel-Cantelli's lemma, $\boldsymbol{y}_k = 0$ infinitely often with *P*-probability one. Whenever $\boldsymbol{y}_k = 0$, $\bar{e}_{2^{k+1}-1} \geq (2^{k+1} - 2^k - 1/2^k) = 1/2 - 1/2^k$. Hence, on a set of *P*-probability one, $\bar{e}_t$ does not converge to 0.

Theorem 11.

(1) *If* $(1/n)d_n(P, Q)(\omega) \to 0$, *then* Q *almost weakly merges to* P *at* ω.
(2) *If* $(1/n)d(P_n \| Q_n) \to 0$, *then* Q *almost weakly merges on average to* P *and the speed of AWMA is* $\sqrt{(1/n)d(P_n \| Q_n)}$. *In particular, if* $d(P_n \| Q_n)$ *is bounded, AWMA occurs at the speed* $1/\sqrt{n}$.
(3) *If* $d(P_n \| Q_n) \leq Cn^{\beta}$ *for* $C > 0$ *and* $\beta < 1$, *then* $\bar{e}_n(P, Q)(\omega) \to 0$, P*-a.s., i.e.* Q *almost weakly merges to* P.

Proof. Follows from Propositions 3, 4, and 8. □

Remark 12. The only condition in the literature under which a rate of convergence of merging is known is absolute continuity, and convergence holds at a rate $1/\sqrt{n}$ in this case (see Sandroni and Smorodinsky, 1999). Note that condition (2) does not imply nor is implied by absolute continuity. Indeed, although $\sup_n d(P_n \| Q_n) < \infty$ implies absolute continuity, when $\sup_n d(P_n \| Q_n) = \infty$, absolute continuity may hold or fail for any rate of growth of $d(P_n \| Q_n)$.

Lehrer and Smorodinsky (1996) provide a sufficient condition for almost weak merging that generalizes absolute continuity. They prove that if $\lim(1/n)\ln(P(\omega_n)/Q(\omega_n)) = 0$ P-a.s. then Q almost weakly merges to P. Both absolute continuity and Lehrer and Smorodinsky's condition are global on the set of paths. Property (1) gives a condition on each ω for which almost weak merging at ω holds.

Example 13 *(Grain of truth).* An common assumption to models of reputation is *grain of truth* (see Sorin, 1999) : P and Q verify the grain of truth assumption if there exists $0 < \lambda \leq 1$ and a probability measure $\tilde{P}$ such that $Q = \lambda P + (1-\lambda)\tilde{P}$. In this case, for each ω, $(P(\omega_n)/Q(\omega_n)) \leq 1/\lambda$ so that $d(P_n \| Q_n) \leq -\ln\lambda$, and

$$E_n(P, Q) \leq \frac{\sqrt{-\ln\lambda}}{\sqrt{2n}}$$

Hence, under the grain of truth assumption, we obtain an explicit bound on $E_n(P, Q)$. Note that the speed of convergence is $1/\sqrt{n}$ and that the constant $\sqrt{-\ln\lambda/2}$ depends on λ only.

Example 14 *(Uniform prior on the parameter of a coin).* A coin is tossed infinitely often. Let X be $\{\text{Heads, Tails}\}$. The true distribution P is the one of an i.i.d. sequence of tosses with parameter $p \in [0, 1]$. The agent believes that the parameter of the coin is drawn from the uniform distribution and that the tosses are i.i.d. with the selected parameter. Here, the true distribution is not absolutely continuous with respect to the belief: under P, the empirical frequency of Heads converges to p almost surely, and this event has probability zero under Q. Yet, we can compute $d(P_n \| Q_n)$ and evaluate the speed of AWMA.

Denoting by h the number of Heads in ω_n,

$$P(\omega_n) = p^h(1-p)^{n-h}$$

and

$$Q(\omega_n) = \int_0^1 t^h(1-t)^{n-h}\mathrm{d}t = \frac{1}{(n+1)\binom{n}{h}}.$$

Then,

$$\begin{aligned} d(P_n \| Q_n) &= \sum_{\omega_n} P(\omega_n)\ln\frac{P(\omega_n)}{Q(\omega_n)} = \sum_{h=0}^{n}\binom{n}{h}p^h(1-p)^{n-h}\ln\left(p^h(1-p)^{n-h}(n+1)\binom{n}{h}\right) \\ &= \ln(n+1) + \sum_{h=0}^{n}\binom{n}{h}p^h(1-p)^{n-h}\ln\left(\binom{n}{h}p^h(1-p)^{n-h}\right) \\ &= \ln(n+1) - H(\mathcal{B}(p,n)) \leq \ln(n+1) \end{aligned}$$

where

$$H(\mathcal{B}(p,n)) = -\sum_{h=0}^{n} \binom{n}{h} p^h (1-p)^{n-h} \ln\left(\binom{n}{h} p^h (1-p)^{n-h}\right)$$

is the entropy of the binomial distribution. Thus, $d(P_n \| Q_n)$ is of order of magnitude $\log n$ and AWMA occurs at a speed no slower than $\sqrt{\ln n/n}$:

$$E_n(P,Q) \leq \sqrt{\frac{\ln(n+1)}{2n}}$$

Example 15 *(Parametric estimation).* The $\sqrt{\ln n/n}$ type of bound on $E_n(P,Q)$ of the previous example also holds in a general set-up. Consider a parameterized family of distributions $\{p_\theta, \theta \in \Theta\}$ on a measurable space, with $\Theta \subset \mathbb{R}^d$. The true law P of the process is i.i.d. with stage law p_{θ_0}, and the agent's prior belief on θ has density $w(\theta)$ w.r.t. the Lebesgue measure. Clarke and Barron (1990) present sufficient conditions under which

$$d(P_n \| Q_n) = \frac{d}{2}\ln\frac{n}{2\pi e} + \frac{1}{2}\ln\det I(\theta_0) + \ln\frac{1}{w(\theta_0)} + o(1)$$

where $I(\theta_0)$ is the Fisher information matrix. A bound on E_n follows using Proposition 4.

5. Bound on the cost of learning

A decision problem is given by a compact space of actions A and a continuous payoff function $\pi : A \times X \to \mathbb{R}$. We let $\|\pi\| = \max_{a,x}|\pi(a,x)|$. The agent chooses an action $a_n \in A$ at each stage n knowing $x_1, \ldots, x_{n-1}$ and receives a payoff $\pi(a_n, x_n)$ at stage n if x_n occurs. A strategy is a mapping $f : \cup_{n\geq 0} X^n \to A$, with $X^0 = \{\emptyset\}$ by convention. A P-optimal strategy is a strategy f_P such that for each stage t and history $\omega_{t-1} = (x_1, \ldots, x_{t-1})$, the action $\mathbf{f}_{P,t} := f_P(\omega_{t-1})$ maximizes $\sum_x P(x|\omega_{t-1})\pi(x,a)$ over $a \in A$.

5.1. Cost of learning, merging and relative entropy

Assuming the probability distribution governing states of nature is P and the decision maker maximizes according to a probability distribution Q, we define the cost of learning suffered by the decision maker in the n-stage decision problem as the difference between the payoff yielded by the optimal strategy f_P and the payoff yielded by the strategy f_Q actually played. Since there may exist several optimal strategies, we consider the worst case and define:

$$c_n(P,Q)(\omega) = \max_{f_Q} \sum_{t=1}^{n} \frac{1}{n} \mathbf{E}_P[\pi(\mathbf{f}_{P,t}, x_t) - \pi(\mathbf{f}_{Q,t}, x_t)|\omega_{t-1}]$$

$$C_n(P,Q) = \max_{f_Q} \sum_{t=1}^{n} \frac{1}{n} \mathbf{E}_P[\pi(\mathbf{f}_{P,t}, x_t) - \pi(\mathbf{f}_{Q,t}, x_t)]$$

where the maximum is taken over all Q-optimal strategies f_Q. Notice that the expressions of c_n and C_n do not depend on the choice of a particular P-optimal strategy.

The following result provides rates of convergence for the cost of learning.

Theorem 16.

(1) $0 \leq c_n(P,Q)(\omega) \leq 4\|\pi\|\bar{e}_n(P,Q)(\omega) \leq 2\sqrt{2}\|\pi\|\sqrt{d_n(P,Q)(\omega)/n}$ *for all n and ω.*
(2) $0 \leq C_n(P,Q) \leq 4\|\pi\|E_n(P,Q) \leq 2\sqrt{2}\|\pi\|\sqrt{d(P_n\|Q_n)/n}$ *for all n.*

Proof.

(1) Take a P-optimal strategy f_P and a Q-optimal strategy f_Q. For each ω_{t-1}, $\mathbf{E}_P[\pi(\mathbf{f}_{P,t}, x_t) - \pi(\mathbf{f}_{Q,t}, x_t)|\omega_{t-1}]$ is non-negative. Furthermore,

$$\begin{aligned}\mathbf{E}_P[\pi(\mathbf{f}_{P,t}, x_t) - \pi(\mathbf{f}_{Q,t}, x_t)|\omega_{t-1}] &= \mathbf{E}_P[\pi(\mathbf{f}_{P,t}, x_t)|\omega_{t-1}] - \mathbf{E}_Q[\pi(\mathbf{f}_{P,t}, x_t)|\omega_{t-1}]\\ &\quad + \mathbf{E}_Q[\pi(\mathbf{f}_{P,t}, x_t)|\omega_{t-1}] - \mathbf{E}_Q[\pi(\mathbf{f}_{Q,t}, x_t)|\omega_{t-1}]\\ &\quad + \mathbf{E}_Q[\pi(\mathbf{f}_{Q,t}, x_t)|\omega_{t-1}] - \mathbf{E}_P[\pi(\mathbf{f}_{Q,t}, x_t)|\omega_{t-1}]\end{aligned}$$

The second difference is non-positive since f_Q is Q-optimal. The first and third differences are both bounded by

$$\|\pi\|\sum_x |P(x|\omega_{t-1}) - Q(x|\omega_{t-1})| = 2\|\pi\| e_t(P, Q)(\omega)$$

Thus,

$$\mathbf{E}_P[\pi(\mathbf{f}_{P,t}, x_t) - \pi(\mathbf{f}_{Q,t}, x_t)|\omega_{t-1}] \le 4\|\pi\| e_t(P, Q)(\omega)$$

Averaging over time yields the desired inequality since the bound does not depend on the choice of the optimal strategies.

(2) This follows directly from the previous point by taking expectation and by noticing that $C_n(P, Q) = \mathbf{E}_P\, c_n(P, Q)$. Indeed, in the maximization problem defining $C_n(P, Q)$, the optimal f_Q should be such that $f_Q(\omega_{t-1})$ maximizes $\mathbf{E}_P[\pi(\mathbf{f}_{P,t}, x_t) - \pi(\mathbf{f}_{Q,t}, x_t)|\omega_{t-1}]$, and thus be also optimal for $c_n(P, Q)(\omega)$. □

Remark 17. Lehrer and Smorodinsky (2000) relate the limit log-likelihood ratio $\lim_n -(1/n)\ln(P(\omega_n)/Q(\omega_n))$ with the asymptotic cost of learning at ω. Theorem 16 provides a bound on the n-stage cost of learning for each n.

Remark 18. Theorem 16 provides a bound on expected payoffs and on conditional expected payoffs. We derive a result on the stream of realized payoffs as follows. For each pair of optimal strategies (f_P, f_Q) define,

$$c'_n(f_P, f_Q)(\omega) = \sum_{t=1}^{n} \frac{1}{n}(\pi(\mathbf{f}_{P,t}, x_t) - \pi(\mathbf{f}_{Q,t}, x_t))$$

$$c''_n(f_P, f_Q)(\omega) = \sum_{t=1}^{n} \frac{1}{n}\mathbf{E}_P[\pi(\mathbf{f}_{P,t}, x_t) - \pi(\mathbf{f}_{Q,t}, x_t)|\omega_{t-1}]$$

The difference $X_n := c'_n(f_P, f_Q) - c''_n(f_P, f_Q)$ is an average of uncorrelated random variables and since payoffs are bounded, from Bienaymé–Chebichev inequality, there exists a constant K depending on the payoff function only such that for each $\varepsilon > 0$, $P(X_n > \varepsilon) \le K/n\varepsilon^2$. Since $c''_n(f_P, f_Q)(\omega) \le c_n(P, Q)(\omega)$ we deduce from Theorem 16:

Claim 19. *There exists a constant K such that for every P-optimal strategy f_P, Q-optimal strategy f_Q and $\varepsilon > 0$,*

$$P\left(c'_n(f_P, f_Q)(\omega) > 2\sqrt{2}\|\pi\|\sqrt{\frac{d_n(P, Q)(\omega)}{n}} + \varepsilon\right) \le \frac{K}{n\varepsilon^2}$$

5.2. Fast convergence in regular decision problems

We get a faster rate of convergence under regularity conditions on the decision problem.

Theorem 20. *Assume* $v : p \mapsto \max_a \mathbf{E}_p \pi(a, \cdot)$ *is twice differentiable, and that* $\|v''\| = \max_p \|v''(p)\|$ *is finite. Then*:

(1) $c_n(P, Q)(\omega) \leq (\|v''\|/4)(d_n(P, Q)(\omega)/n)$ *for all* n *and* ω.
(2) $C_n(P, Q) \leq (\|v''\|/4)(d(P_n\|Q_n)/n)$ *for all* n.

Proof. Fix a P-optimal strategy f_P, a Q-optimal strategy f_Q, a history ω_{t-1} and set $p = P(\cdot|\omega_{t-1})$, $q = Q(\cdot|\omega_{t-1})$, $a = f_P(\omega_{t-1})$ and $b = f_Q(\omega_{t-1})$. Then,

$$\mathbf{E}_P[\pi(\mathbf{f}_{P,t}, x_t) - \pi(\mathbf{f}_{Q,t}, x_t)|\omega_{t-1}] = v(p) - \mathbf{E}_p \pi(b, \cdot) = v(p) - v(q) - (\mathbf{E}_p \pi(b, \cdot) - \mathbf{E}_q \pi(b, \cdot))$$

The mapping $p \mapsto \mathbf{E}_p \pi(a, \cdot)$ is linear, so its derivative with respect to p does not depend on p and we denote it π_a. From the envelope theorem, $v'(p) = \pi_a$ and $v'(q) = \pi_b$. Thus,

$$v(p) - \mathbf{E}_p \pi(b, \cdot) = v(p) - v(q) - (p - q)v'(q)$$

Since v is twice differentiable with second derivative bounded by $\|v''\|$,

$$v(p) - \mathbf{E}_p \pi(b, \cdot) \leq \tfrac{1}{2}\|v''\|\|p - q\|^2$$

From Pinsker's inequality, $\|p - q\|^2 \leq (1/2)d(p\|q)$. Thus,

$$\mathbf{E}_P[\pi(\mathbf{f}_{P,t}, x_t) - \pi(\mathbf{f}_{Q,t}, x_t)|\omega_{t-1}] \leq \tfrac{1}{2}\|v''\|(e_t(P, Q)(\omega))^2 \leq \tfrac{1}{4}\|v''\|d(p\|q)$$

The proof is concluded as for Theorem 16. □

Example 21. Consider a quadratic model where $A = [0, 1]$, $X = \{0, 1\}$ and $\pi(a, x) = -(x - a)^2$. Then,

$$v(p) = \max_a\{-pa^2 - (1 - p)(1 - a)^2\} = -p(1 - p)$$

From Theorem 20, $c_n(P, Q)(\omega) \leq d_n P, Q\omega/(2n)$ and $C_n P, Q) \leq dP_n\|Q_n/(2n)$.

Example 22. If the differentiability condition fails, the per-stage cost of learning might not be proportional to the square of the variational distance but to the variational distance itself, thus leading to a slower convergence rate.

Consider a "matching pennies" problem: $A = X = \{0, 1\}$ and the decision maker has to predict nature's move, $\pi(a, x) = \mathbf{1}_{\{a=x\}}$. Assume that the belief at some stage is $q = 1/2$ and that $p = 1/2 - \varepsilon$ (p and q are identified with the probability they put on 0). Let $b = 0$ be the action corresponding to a belief $> 1/2$. Then

$$v(p) - \mathbf{E}_p \pi(b, \cdot) = (1 - p) - p = 2\varepsilon = 2(q - p)$$

In this example, q is at a kink of the map v, therefore at a point where the "marginal value of information" is maximal.

5.3. The discounted case

Now we extend Theorems 16 and 20 to discounted problems. We define the cost of learning suffered by the decision maker in the δ-discounted decision problem ($0 < \delta < 1$) as:

$$C_\delta(P, Q) = \max_{f_Q} \sum_{t=1}^{\infty} (1 - \delta)\delta^{t-1} \mathbf{E}_P[\pi(\mathbf{f}_{P,t}, x_t) - \pi(\mathbf{f}_{Q,t}, x_t)]$$

where f_P is any P-optimal strategy and the maximum is taken over all Q-optimal strategies f_Q. Note that $C_\delta(P, Q)$ is always non negative.

Proposition 23. *If* $d(P, Q) = \sup_n d_n(P\|Q) < \infty$, *then*:

(1) $C_\delta(P, Q) \leq 2\sqrt{2}\|\pi\|\sqrt{d(P\|Q)}\sqrt{(1-\delta)}$.
(2) *If* $v : p \mapsto \max_a \mathbf{E}_p\pi(a, \cdot)$ *is twice differentiable and* $\|v''\| = \max_p \|v''(p)\| < \infty$, *then* $C_\delta(P, Q) \leq (\|v''\|/4)d(P\|Q)(1-\delta)$.

In particular, sufficiently patient agents suffer arbitrarily small costs of learning. More precisely, the cost is less than ε if $\delta \geq 1 - (\varepsilon^2/(8\|\pi\|^2 d(P\|Q)))$.

Proof.

(1) The discounted average of a sequence is a convex combination of the finite stage arithmetic averages: $C_\delta(P, Q) = \sum_m (1-\delta)^2\delta^{m-1} m C_m(P, Q)$. Then using Theorem 16,

$$C_\delta(P, Q) \leq 2\sqrt{2}\|\pi\|\sqrt{d(P\|Q)}(1-\delta)\sum_m (1-\delta)\delta^{m-1}\sqrt{m}$$

Jensen's inequality and the concavity of the square root function imply $\sum_m (1-\delta)\delta^{m-1}\sqrt{m} \leq 1/\sqrt{1-\delta}$ and the result follows.
(2) Follows from the same lines, using Theorem 20. □

References

Blackwell, D., Dubins, L., 1962. Merging of opinions with increasing information. The Annals of Mathematical Statistics 33, 882–886.
Clarke, B., Barron, A., 1990. Information-theoretic asymptotics of Bayes method. IEEE Transactions on Information Theory 36, 453–471.
Cover, T.M., Thomas, J.A., 1991. Elements of information theory. In: Wiley Series in Telecomunications. Wiley.
Kalai, E., Lehrer, E., 1993. Rational learning leads to Nash equilibrium. Econometrica 61, 1019–1945.
Kalai, E., Lehrer, E., 1994. Weak and strong merging of opinions. Journal of Mathematical Economics 23, 73–86.
Kalai, E., Lehrer, E., Smorodinski, R., 1999. Calibrated forecasting and merging. Games and Economic Behavior 29,151–169.
Lehrer, E., Smorodinsky, R., 1996. Compatible measures and merging. Mathematics of Operations Research 21, 306–697.
Lehrer, E., Smorodinsky, R., 2000. Relative entropy in sequential decision problems. Journal of Mathematical Economics 33, 425–440.
Sandroni, A., Smorodinsky, R., 1999. The speed of rational learning. International Journal of Game Theory 28, 199–210.
Sorin, S., 1999. Merging, reputation, and repeated games with incomplete information. Games and Economic Behavior 29, 274–308.