How to play with a biased coin?

Olivier Gossner [a,*] and Nicolas Vieille [b]

[a] *THEMA, UMR 7536, Université Paris X–Nanterre and CORE, Université Catholique de Louvain, Louvain, Belgium*

[b] *Laboratoire d'Économétrie de l'École Polytechnique et Département Finance et Économie, HEC, 1, rue de la Libération, 78351 Jouy en Josas, France*

Received 7 June 1999

Abstract

We characterize the max min of repeated zero-sum games in which player one plays in pure strategies conditional on the private observation of a fixed sequence of random variables. Meanwhile we introduce a definition of a strategic distance between probability measures, and relate it to the standard Kullback distance.
© 2002 Elsevier Science (USA). All rights reserved.

1. Introduction

In the game of *matching pennies* where the payoff of player one is given by the matrix

	Heads	Tails
Heads	1	0
Tails	0	1

the unique optimal strategy of each player is to toss a fair coin, and to play according to the outcome: *Heads* if the coin turns *Heads*, and *Tails* otherwise. Consider a repetition of matching pennies in which player one (the maximizer) can only condition his actions on the privately observed outcome of a biased coin,

* Corresponding author.
E-mail addresses: olivier.gossner@u-paris10.fr (O. Gossner), vieille@hec.fr (N. Vieille).

whose parameter p is common knowledge. Prior to each round of play, he gets to observe a toss which is independent of previous tosses and has to choose a (pure) action. We study the optimal strategies of player one and the corresponding value, namely the max min of that repeated game.

An obvious way to play is for player one to choose in any round an action according to the toss in *that* round. This strategy secures $\min(p, 1 - p)$ in the long run. We show that player one can actually do much better: the max min of the infinitely repeated game exists and is equal to half the entropy $H(\mathbf{p}) = -p \ln_2 p - (1 - p) \ln_2 (1 - p)$ of the distribution $\mathbf{p} = (p, 1 - p)$.

The game of matching pennies has specific features which make the proof a bit *ad hoc* (for instance, the unique optimal strategy in the one-shot game is the mixed action with maximal entropy). For the general case of a zero-sum game G, define $U(h)$ as the maximum that player one can guarantee in G when restricted to mixed actions of entropy at most h. Let cav U denote the smallest concave function that majorizes U. Assume that the sequence of private observations is i.i.d. (we shall argue that this assumption can be weakened to a great extent). We prove that the max min of the corresponding infinitely repeated game is cav $U(h)$ where h is the entropy of any private observation.

Tools borrowed from information theory, such as entropy or relative entropy, already found several applications into the literature of repeated games with bounded complexity. In a model of repeated games with stationary bounded memory, Lehrer (1994) makes use of mutual entropy as a measure of the information that a player's memory may contain about other player's strategies. Neyman and Okada (1999) introduce strategic entropy as a measure of the randomness contained in a player's strategy, and deduce bounds on the values of repeated games played by finite automata. Lehrer and Smorodinsky (2000) provide sufficient conditions for convergence of learning processes that are phrased in terms of the relative entropy between a utility maximizer's belief on the sequence of nature's moves and its true probability distribution. Our result is close in flavor to those of Neyman and Okada (2000). They proved that if player one can choose *any* strategy of per stage entropy h, then the maximum that player one can guarantee is cav $U(h)$. We thus obtain a similar result without assuming that player one can choose the stream of random variables on which his strategies can be conditioned. We essentially show that a player can approximate (through a deterministic coding) *any* strategy of per stage entropy h from *any* stream of random variables of per stage entropy h. We introduce the notion of *strategic distance* between two random variables as the appropriate measure for the quality of this approximation, and relate this notion with the classical Kullback distance. Shannon (1948) studied the question of coding information into the smallest sequence of bits (0s and 1s) as possible. The optimization problem we study is somehow dual since player one needs to use his random information during as much time as possible.

The paper is organized as follows. Section 2 contains the model and the statement of the results. Section 3 is a short reminder of basic properties of conditional entropy. In Section 4, we construct ε-optimal strategies of player one, in the game of matching pennies. We introduce the strategic distance and relate it to the Kullback distance in Section 5. Using this, we construct ε-optimal strategies for player one in the general case in Section 6. Section 7 deals with replies of player two in the general case. The model of bounded rationality we study may seem somehow unnatural and isolated from classical problems of game theory. In fact, we connect it with problems arising in repeated games with imperfect monitoring in Section 8.

2. Model and results

2.1. The repeated game—$\max \min$

A two player zero-sum game G given by finite sets of actions A and B for player one and player two and a payoff function $g : A \times B \to \mathbb{R}$ is fixed. We study repetitions of G in which player two's strategy choice is unrestricted whereas player one privately observes the successive realizations of a sequence of i.i.d. random variables, but has to play a pure strategy.

Our aim is to study how the $\max \min$ of the infinitely repeated game depends on the characteristics of this private source of uncertainty. Thus, we may and do consider only pure strategies for player two.

The repeated game G_∞ proceeds as follows. We let $\mathbf{X} = (X_n)_{n \geqslant 1}$ be a sequence of i.i.d., random variables, each with law $\mathbf{p}$, with values in a finite set C. In each stage $n \geqslant 1$, player one gets to observe X_n; players one and two choose actions; these are publicly announced, and the game proceeds to stage $n + 1$. The description of the game is common knowledge.

Given the monitoring assumptions, the behavior of player one in stage n is described by a function $\sigma_n : (C \times A \times B)^{n-1} \to A$. A strategy of player one is a sequence $\sigma = (\sigma_n)_{n \geqslant 1}$ of such maps. Similarly, a strategy of player two is a sequence $\tau = (\tau_n)_{n \geqslant 1}$, with $\tau_n : (A \times B)^{n-1} \to B$.

Any profile (σ, τ) (together with $\mathbf{X}$) induces a probability distribution $\mathbf{P}_{\sigma, \tau}$ on the set of plays $\mathbf{H} = (A \times B)^{\mathbb{N}}$ endowed with the product of discrete σ-algebras. Expectation with respect to $\mathbf{P}_{\sigma, \tau}$ is denoted by $\mathbf{E}_{\sigma, \tau}$.

We denote by $\mathbf{a}_n$, $\mathbf{b}_n$ the (random) actions of the players in stage n, and set $\mathbf{g}_n = g(\mathbf{a}_n, \mathbf{b}_n)$. Notice that, once σ and τ are fixed, $\mathbf{a}_n$ and $\mathbf{b}_n$ are simply functions of the random sequence $\mathbf{X}$ (more precisely, $\mathbf{a}_n$ depends on the first n components of $\mathbf{X}$, and $\mathbf{b}_n$ on the first $n - 1$ components, through player one's past play). Thus, $\mathbf{P}_{\sigma, \tau}(\mathbf{a}_n = a) = \mathbf{P}(x, \mathbf{a}_n(x) = a)$.

The expected average payoff γ_n for player one up to stage n is

$$\gamma_n(\sigma, \tau) = \mathbf{E}_{\sigma, \tau}\left[\frac{1}{n}\sum_{m=1}^{n}\mathbf{g}_m\right].$$

We recall the standard definition of the max min.

Definition 1. Let $v \in \mathbb{R}$.

- Player 1 can secure v if, for every $\varepsilon > 0$, there exists a strategy σ_ε, and a stage $N \geqslant 1$ such that:
  $$\forall \tau, \forall n \geqslant N, \quad \gamma_n(\sigma_\varepsilon, \tau) \geqslant v - \varepsilon.$$
- Player two can defend v if, given any σ and $\varepsilon > 0$, there exists a strategy τ_ε and a stage $N \geqslant 1$ such that
  $$\forall n \geqslant N, \quad \gamma_n(\sigma, \tau_\varepsilon) \leqslant v + \varepsilon.$$
- v is the max min of the game if player one can secure v, and player two can defend v.

2.2. Main result

In order to state our main result, we need to recall the definition of entropy. If $\mathbf{q}$ is a distribution over a finite set Ω, the entropy of $\mathbf{q}$ is

$$H(\mathbf{q}) = -\sum_{\omega \in \Omega}\mathbf{q}(\omega)\log\big(\mathbf{q}(\omega)\big),$$

where the logarithm is taken to be in base 2, and $0\log 0 = 0$ by convention. For $h \geqslant 0$, we let $U(h)$ be the max min of the (one-shot) game in which player one is restricted to mixed actions of entropy at most h

$$U(h) = \max_{H(x)\leqslant h}\min_{b}\mathbf{E}_x g(\cdot, b)$$

Finally, recall that the concavification $\operatorname{cav} u$ of a real mapping u defined on a convex set is the least concave function which majorizes u.

Theorem 2. *The* max min *of* G_∞ *is* $\operatorname{cav} U(H(\mathbf{p}))$.

The proof of the theorem is organized as follows: We need to prove both that player one can secure v and that player two can defend v. The most difficult part of the proof is the construction of ε-optimal strategies of player one. We construct such strategies for the game of matching pennies in Section 3. Next, we introduce some tools in Section 4, that we use in Section 5 to deal with general zero-sum games. That player two can defend v is a consequence of a previous analysis by Neyman and Okada (2000), we provide a proof for the sake of completeness in Section 6.
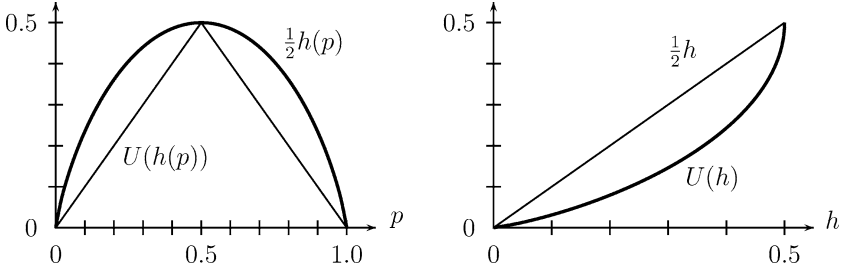
Fig. 1.

2.3. Example: matching pennies

In the game of matching pennies, $A = \{T, B\}$, $B = \{L, R\}$ and the payoff to player one is given by the array

	L	R
T	1	0
B	0	1

Assume $C = \{0, 1\}$ and $\mathbf{p}(1) = p$ (the sequence $\mathbf{X}$ is $\{0, 1\}$-valued). We set $h(p) = H(\mathbf{p})$. It is well known that h is a continuous, increasing, concave, bijective function from $[0, 1/2]$ to $[0, 1]$. Moreover, $h(p) = h(1 - p)$, for each p. Denote by $h^{-1} : [0, 1] \to [0, 1/2]$ the inverse map: h being concave, h^{-1} is convex.

Here, $U(h(p)) = \min\{p, 1 - p\}$, which is the payoff player one guarantees by playing either $pT + (1 - p)B$ or $(1 - p)T + pB$. Thus, $U(x) = h^{-1}(x)$. Given the properties of h^{-1}, cav $U(x) = x/2$. Theorem 2 shows that the max min of G_∞ exists and is equal to $(1/2)H(\mathbf{p})$.

The two graphs (see Fig. 1) above show visually that cav $U(h(p)) = (1/2)h(p)$. Clearly, $U(h(p))$ is concave in p whereas it is a convex function of $h(p)$.

3. Conditional entropy

The entropy of a random variable is usually seen as a good measure of the uncertainty contained in it, or of the amount of information gained by observing it. By definition, the entropy $H(Y)$ of a random variable Y with finite range is the entropy of its distribution.

Let (Y_1, Y_2) be a random vector with values in a finite set $\Omega_1 \times \Omega_2$, with distribution $\mathbf{q}$. For $\omega_1 \in \Omega_1$, denote by $\mathbf{q}(\cdot|\omega_1)$ the conditional distribution of Y_2, given $Y_1 = \omega_1$ and by $h(Y_2|\omega_1)$ the entropy of $\mathbf{q}(\cdot|\omega_1)$

$$h(Y_2|\omega_1) = - \sum_{\omega_2 \in \Omega_2} \mathbf{q}(\omega_2|\omega_1) \log \mathbf{q}(\omega_2|\omega_1).$$

The *conditional entropy* of Y_2 given Y_1 is defined to be

$$H(Y_2|Y_1) = E_{Y_1}\big[h(Y_2|Y_1)\big] = \sum_{\omega_1} \mathbf{q}(\omega_1)h(Y_2|\omega_1).$$

It is the average uncertainty on the value of Y_2, given that one observes Y_1.

An easy computation shows that

$$H(Y_1, Y_2) = H(Y_1) + H(Y_2|Y_1), \tag{1}$$

where $H(Y_1, Y_2)$ is the entropy of the variable (Y_1, Y_2). If $(Y_1, \ldots, Y_n)$ is a random vector with finite range $\Omega_1 \times \cdots \times \Omega_n$, the relation (1) yields inductively

$$H(Y_1, \ldots, Y_n) = H(Y_1) + \sum_{i=1}^{n} H(Y_i|Y_1, \ldots, Y_{i-1}). \tag{2}$$

Notice that, if $Y_1, \ldots, Y_n$ are i.i.d. variables, $H(Y_1, \ldots, Y_n) = nH(Y_1)$. Finally, remark that for any Y with finite range Ω, and any function $f : \Omega \to \Theta$, $H(f(Y)) \leqslant H(Y)$.

4. To guarantee: matching pennies

We construct an ε-optimal strategy for player one in the game of matching pennies, assuming $C = \{0, 1\}$. We shall show that what is optimal for player one is not to use his private endowment of uncertainty at a *constant* rate, but rather to *accumulate* it for a while, and use it to play in the *most* unpredictable way.

Fix $\varepsilon > 0$. We design a strategy σ of player one that depends on a parameter $\eta > 0$. We shall prove that, provided η is small enough, σ guarantees $\frac{1}{2}H(\mathbf{p})$ up to ε: there exists N, such that, for every τ and $n \geqslant N$,

$$\gamma_n(\sigma, \tau) \geqslant \frac{1}{2}H(\mathbf{p}) - \varepsilon, \tag{3}$$

where σ is defined by blocks of length l (l function of η). Under σ, player one's play is independent of player two's past play. Moreover, on *each* block, it is independent of the *play* in the past blocks; it depends only on the signals received by player one in the previous block (it plays repeatedly T in the first block).

Denote by $\mathbf{P} = \bigotimes_{\mathbb{N}} \mathbf{p}$ the distribution over $\{0, 1\}^{\mathbb{N}}$ induced by a sequence (x_n) of i.i.d. variables, distributed according to $\mathbf{p}$. For simplicity, we write H for $H(\mathbf{p})$.

Choose $\eta > 0$ such that $H + \eta$ is a rational number. Define a *typical sequence* of length n to be a sequence $x \in \{0, 1\}^n$, such that

$$\frac{1}{2^{n(H+\eta)}} \leqslant \mathbf{P}(x) \leqslant \frac{1}{2^{n(H-\eta)}}.$$

Let $C(n, \eta)$ be the set of typical sequences of length n. All typical sequences have roughly the same probability. The next lemma is known as the asymptotic

equipartition property in information theory. It asserts that the probability of the set of typical sequences goes to 1 as n goes to infinity. We refer to Cover and Thomas (1991) for proofs and comments.

Lemma 3. $\forall \eta > 0$, $\lim_{n \to \infty} \mathbf{P}(C(n, \eta)) = 1$.

Note. We shall later prove a stronger version of this result (Lemma 9) where we provide an estimate of η relative to n. This stronger result will be useful for the general zero-sum case, but Lemma 3 is sufficient for our proof in the case of matching pennies.

We choose the length l of the blocks so that $\mathbf{P}(C(l, \eta)) \geqslant 1 - \eta$ and $l(H + \eta) \in \mathbb{N}$. We define σ on a single block $b \geqslant 2$, and consider that, at the beginning of that block, player one has a private signal $\mathbf{x}$, drawn from $\{0, 1\}^l$ according to $\mathbf{P}$.

Since $\mathbf{P}(x) \geqslant 1/2^{l(H+\eta)}$, for $x \in C(l, \eta)$, the cardinality of $C(l; \eta)$ is at most $2^{l(H+\eta)}$. Therefore, there exists a one-to-one map $i : C(l, \eta) \to \{T, B\}^{l(H+\eta)}$.

Define σ as:

- if $\mathbf{x} \in C(l, \eta)$, play the sequence $i(\mathbf{x})$ of actions in the first $l(H + \eta)$ stages of the block, then repeatedly T;
- if $\mathbf{x} \notin C(l, \eta)$, play repeatedly T.

Notice that in the last $l - l(H + \eta)$ stages, player one plays repeatedly T; a best reply of player two will obviously be to play R in those stages, yielding a payoff of 0 to player one. We argue below that in the first $l(H + \eta)$ stages, player one's behavior is essentially unpredictable: in most of the stages, the distribution of player one's action, conditional on his past actions, is close to $(1/2, 1/2)$ (with high probability). We prove this formally.

For simplicity, we (re)label the stages of block b from 1 to l. We denote by $\mathbf{a} = (\mathbf{a}_1, \ldots, \mathbf{a}_{l(H+\eta)})$ the sequence of actions played by player one in the first part of the block. We also set $\mathbf{T} = (T, \ldots, T)$ (sequence of length $l - l(H + \eta)$). We first prove that the entropy of $\mathbf{a}$ is almost maximal.

Lemma 4. *One has* $H(\mathbf{a}) \geqslant l(H - 3\eta)$.

Proof.

$$
\begin{aligned}
H(\mathbf{a}) &= - \sum_{a \in \{T, B\}^{l(H+\eta)}} \mathbf{P}(\mathbf{a} = a) \log\big(\mathbf{P}(\mathbf{a} = a)\big) \\
&= - \sum_{a \neq \mathbf{T}} \mathbf{P}(\mathbf{a} = a) \log\big(\mathbf{P}(\mathbf{a} = a)\big) - \mathbf{P}(\mathbf{a} = \mathbf{T}) \log\big(\mathbf{P}(\mathbf{a} = \mathbf{T})\big) \\
&\geqslant - \sum_{\substack{x \in C(l; \eta) \\ i(x) \neq \mathbf{T}}} \mathbf{P}(x) \log\big(\mathbf{P}(x)\big)
\end{aligned}
$$

$$\geqslant l(H - \eta)\left(\mathbf{P}\big(C(l; \eta)\big) - \frac{1}{2^{l(H-\eta)}}\right)$$

$$\geqslant l(H - \eta)(1 - 2\eta)$$

$$\geqslant l(H - 3\eta).$$

The second inequality uses the fact that $-\log \mathbf{P}(x) \geqslant l(H - \eta)$ holds for every $x \in C(l, \eta)$, and the fact that *at most* one sequence $x \in C(l, \eta)$ is mapped to $\mathbf{T}$, the probability of this sequence being at most $1/2^{l(H-\eta)}$. The last two inequalities are valid, provided η is small enough. $\quad\square$

Lemma 5. *For every $\varepsilon > 0$, there exists $\eta_0 > 0$ such that for every $\eta < \eta_0$ and every τ*

$$\mathbf{E}_{\sigma,\tau}\left[\frac{1}{l}\sum_{n=1}^{l}\mathbf{g}_n\right] \geqslant \frac{1}{2}H - \varepsilon.$$

Thus, the expected average payoff on any block (except the first one) is at least $H/2 - \varepsilon$, provided η is small enough. This clearly implies (3), and concludes the proof that player one can guarantee $(1/2)H(\mathbf{p})$.

Proof. We give an estimate of

$$\alpha = \min_{\tau}\mathbf{E}_{\sigma,\tau}\left[\frac{1}{l}\sum_{n=1}^{l}\mathbf{g}_n\right].$$

Denote by $\mathbf{p}_n = (p_n, 1 - p_n)$ the conditional distribution of $\mathbf{a}_n$, conditional on $\mathbf{a}_1, \ldots, \mathbf{a}_{n-1}$: it is the belief held by player two in stage n over player one's next action, given player two's information (and knowing player one's strategy). Since player one's behavior is independent of player two's play, it is a best reply of player two to σ to play L whenever $p_n \leqslant \frac{1}{2}$, and R otherwise. This yields a payoff of $\min(p_n, 1 - p_n)$ in stage n.

Therefore,

$$\alpha = \mathbf{E}_{\sigma}\left[\frac{1}{l}\sum_{1}^{l}\min(p_n, 1 - p_n)\right].$$

For $n > l(H + \eta)$, $p_n = 1$, P_{σ}-a.s.; player one plays T in these stage, irrespective of x. Thus,

$$\alpha = (H + \eta)\mathbf{E}_{\sigma}\left[\frac{1}{l(H + \eta)}\sum_{1}^{l(H+\eta)}\min(p_n, 1 - p_n)\right].$$

Notice that $\min(p, 1 - p) = h^{-1}(H(\mathbf{p}))$, where h^{-1} has been defined in Section 2.3. Thus,

$$
\begin{aligned}
\alpha &= (H + \eta)\mathbf{E}_\sigma \left[\frac{1}{l(H + \eta)} \sum_1^{l(H+\eta)} h^{-1}\big(H(\mathbf{p}_n)\big) \right] \\
&\geqslant (H + \eta)h^{-1}\left(\mathbf{E}_\sigma \left[\frac{1}{l(H + \eta)} \sum_1^{l(H+\eta)} H(\mathbf{p}_n) \right] \right) \\
&= (H + \eta)h^{-1}\left(\frac{1}{l(H + \eta)} \sum_1^{l(H+\eta)} H(\mathbf{a}_n | \mathbf{a}_1, \ldots, \mathbf{a}_{n-1}) \right) \\
&= (H + \eta)h^{-1}\left(\frac{1}{l(H + \eta)} H(\mathbf{a}) \right).
\end{aligned}
$$

The first inequality is obtained by applying Jensen's inequality to the convex function h^{-1}. The second one uses the fact that the conditional entropy $H(\mathbf{a}_n | \mathbf{a}_1, \ldots, \mathbf{a}_{n-1})$ is defined as $\mathbf{E}_\sigma(H(\mathbf{p}_n))$. The last one is a consequence of the property (2) of conditional entropies.

Using Lemma 4, one deduces that $\alpha \geqslant (H + \eta)h^{-1}(1 - 4\eta/H)$. Since $h^{-1}(1) = 1/2$, one gets $\alpha \geqslant H/2 - \varepsilon$, provided η has been chosen small enough. $\square$

5. Tools for the general case

In order to prove our main result in the general case, we first develop here a number of tools. The general idea of our construction is that, by properly encoding his private information, player one tries to mimic some strategy (the optimal strategy in the case of matching pennies). It is therefore crucial to have a measure of the quality of the approximation. We shall define such a measure, the strategic distance between two probabilities over sequences of moves.

5.1. The extra difficulties

One feature of matching pennies that made its analysis simple is that the unique optimal strategy of player one is the distribution with maximal entropy, the uniform distribution $(1/2, 1/2)$. Assume that player one selects his moves over a block of length n according to the distribution $\mathbf{P} \in \Delta(\{T, B\}^n)$ which has an entropy close to the maximal entropy n. In the case of matching pennies, we could deduce the two facts:

- First, it must be the case that at almost every stage, the entropy of player one's move conditional to his past moves on this block must be almost maximal.

- Second, if the entropy conditional to the past moves is close to the maximal one, then the corresponding probability conditional to player two's information must be close to $(1/2, 1/2)$.

We could then conclude that player two could not do much better on average than $1/2$ over that block. In that sense **P** is a good approximation of the optimal strategy of player one in the n-stage matching pennies.

We need to elaborate on this. Let G be any zero-sum game, with action sets A and B, and payoff function g, which is being played n times. Assume that player 1 wishes to mimic the distribution $\mathbf{Q} \in \Delta(A^n)$, and selects its moves according to $\mathbf{P} \in \Delta(A^n)$ (both **P** and **Q** may be viewed as mixed strategies, which put positive probabilities only on pure strategies which are independent of player two's play).

We define a measure of how good the approximation of **Q** by **P** is, suited to game-theoretic purposes. It is clear that this measure cannot be defined as the difference between the entropies of **P** and **Q**. First, **P** and **Q** can have the same per stage entropy but have different entropies for a given stage and a given past history. Second, even in a one-stage game, two strategies can have the same entropy and guarantee different payoffs. Indeed, in the game G defined by

2	0
0	1

the two strategies $(1/3, 2/3)$ and $(2/3, 1/3)$ have the same entropy. The former is optimal and guarantees $2/3$, while the latter guarantees only $1/3$.

5.2. Strategic distance

Recall that $\mathbf{P}, \mathbf{Q} \in \Delta(A^n)$. Let $k \in \{1, \dots, n\}$. At stage k, given player one's past play h_k, the difference between what **P** and **Q** secure can be estimated by

$$\left| \min_b g\big(\mathbf{Q}(\cdot|h_k), b\big) - \min_b g\big(\mathbf{P}(\cdot|h_k), b\big) \right| \leqslant M \left\| \mathbf{P}(\cdot|h_k) - \mathbf{Q}(\cdot|h_k) \right\|_1, \qquad (4)$$

where $M = 2 \max_{A \times B} |g|$, and $\| \cdot \|_1$ is the L^1-norm on $\mathbb{R}^A$.

Thus, as for stage k, a good measure of the approximation is given by

$$\mathbf{E}_{\mathbf{P}}\big[\left\| \mathbf{P}(\cdot|\mathcal{H}_k) - \mathbf{Q}(\cdot|\mathcal{H}_k) \right\|_1 \big],$$

which we write $\| \mathbf{P}(\cdot|\mathcal{H}_k) - \mathbf{Q}(\cdot|\mathcal{H}_k) \|_{1,\mathbf{P}}$ ($\mathcal{H}_k$ being the algebra generated by player one's play in the first $k-1$ stages).

This motivates the following definition.

Definition 6. Let A be a finite set, $n \in \mathbb{N}$, and $\mathbf{P}, \mathbf{Q} \in \Delta(A^n)$. The strategic distance from **P** to **Q** is

$$d_S(\mathbf{P}\|\mathbf{Q}) = \frac{1}{n} \sum_{k=1}^n \left\| \mathbf{P}(\cdot|\mathcal{H}_k) - \mathbf{Q}(\cdot|\mathcal{H}_k) \right\|_{1,\mathbf{P}}.$$

Remark. This definition depends on G only through the action set of player one, A.

Remark. Beware that the strategic distance is not a distance, since it is not symmetric. The same abuse of language is common when speaking of the Kullback distance (see below).

Notation. For $n \in \mathbb{N}$, and $\mathbf{P} \in \Delta(A^n)$, we set

$$\underline{v}(\mathbf{P}) = \min_{b \in B^n} \frac{1}{n} \mathbf{E_P}\left(\sum_{k=1}^{n} g(a_k, b_k) \right).$$

Notice that $\underline{v}(\mathbf{P})$ is the amount secured by $\mathbf{P}$ in the n-stages version of G. For $n = 1$ and $p \in \Delta(A)$, the above notation particularizes to $\underline{v}(\mathbf{p}) = \min_b \mathbf{E_p} g(a, b)$. The lemma below is a straightforward consequence of the inequality (4).

Lemma 7. *Let $M = 2 \max_{A \times B} |g|$. For every $\mathbf{P}, \mathbf{Q} \in \Delta(A^n)$, with $Q = q^{\otimes n}$ for some $q \in \Delta(A)$ one has*

$$\left| \underline{v}(\mathbf{P}) - \underline{v}(\mathbf{Q}) \right| \leqslant M d_S(\mathbf{P} \| \mathbf{Q}).$$

5.3. Comparison of the strategic and Kullback distances

Let S be a finite set. Denote by Ω the set of $(\mathbf{p}, \mathbf{q}) \in \Delta(S) \times \Delta(S)$ such that $\mathbf{p} \ll \mathbf{q}$: $\mathbf{q}(s) = 0 \Rightarrow \mathbf{p}(s) = 0$. The Kullback distance between $\mathbf{p}$ and $\mathbf{q}$ is defined as

$$d(\mathbf{p} \| \mathbf{q}) = \mathbf{E_p} \log \frac{\mathbf{p}(s)}{\mathbf{q}(s)} = \sum_{s \in S} \mathbf{p}(s) \log \frac{\mathbf{p}(s)}{\mathbf{q}(s)},$$

with $0 \log \frac{0}{0} = 0$ by convention.

We shall construct strategies of player one which induce probabilities over action sequences which are close for Kullback distance to some fixed probability. In order to state that the payoffs guaranteed are also close, we shall rely on the following proposition, the proof of which is postponed to Appendix A.

Proposition 8. *Let $(n_k)_k$ be a sequence of positive integers, and consider $\mathbf{P}_k, \mathbf{Q}_k \in \Delta(A^{n_k})$ such that*

$$\lim_{k \to \infty} \frac{1}{n_k} d(\mathbf{P}_k \| \mathbf{Q}_k) = 0;$$

then

$$\lim_{k \to \infty} d_S(\mathbf{P}_k \| \mathbf{Q}_k) = 0.$$

5.4. A stronger version of the asymptotic equipartition property

The asymptotic equipartition property (Lemma 3) does not provide the rate of convergence of $\mathbf{P}(C(\eta, n))$. We shall use the following lemma.

Lemma 9. *For every $\varepsilon > 0$, there exists $K > 0$ such that*

$$\forall n > 0, \quad \mathbf{P}\left(C\left(\frac{K}{\sqrt{n}}, n \right) \right) \geqslant 1 - \varepsilon. \tag{5}$$

Proof. Let us write $\mathbf{x} = (\mathbf{x}_t)_{1 \leqslant t \leqslant n}$. Note that $\mathbf{x} \in C(\eta, n)$ if and only if

$$\left| \log \mathbf{P}(\mathbf{x}_1, \ldots, \mathbf{x}_n) + nh \right| \leqslant n\eta.$$

Observe that $\log \mathbf{P}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \sum_1^n \log p(\mathbf{x}_t)$ has mean $-nh$, by definition of h and variance nV^2 where V^2 is the variance of $\log p(\mathbf{x}_1)$. Tchebytchev's inequality yields directly

$$\mathbf{P}\big(\mathbf{x} \notin C(\eta, n) \big) \leqslant \frac{V^2}{n\eta^2}.$$

Hence the result with $K = V/\sqrt{\varepsilon}$. ☐

6. To guarantee: the general case

In this section we prove that player one can guarantee $\operatorname{cav} U(H(\mathbf{p}))$, whatever be the underlying one-shot game. For simplicity, we set $h = H(\mathbf{p})$. We start by recalling two general well-known properties of the cav operator.

Remark 10. For any mapping u, an equivalent definition of $\operatorname{cav} u$ is

$$\operatorname{cav} u(x) = \sup_{\substack{1 \geqslant \lambda \geqslant 0 \\ x = \lambda x_1 + (1-\lambda)x_2}} \lambda u(x_1) + (1 - \lambda)u(x_2), \tag{6}$$

where x, x_1, and x_2 are restricted to be in the domain of u.

Remark 11. If u is continuous, the supremum can be taken over any relatively dense subset of $\{(\lambda, x_1, x_2): 1 \geqslant \lambda \geqslant 0, \ x = \lambda x_1 + (1-\lambda)x_2\}$.

Since U is clearly continuous, we need only prove that player one can guarantee

$$\max_{\{1 \geqslant \lambda \geqslant 0: \ h = \lambda h_L + (1-\lambda)h_R\}} \lambda U(h_L) + (1 - \lambda)U(h_R), \tag{7}$$

where we impose that λ, h_L/h and h_R/h be rational numbers. That we shall prove as Proposition 12.

Proposition 12. *Let h_L, h_R, and λ be given. Assume that h_L/h, h_R/h, and λ are rational numbers, and $h = \lambda h_L + (1 - \lambda)h_R$. Let $\mathbf{p}_L$ and $\mathbf{p}_R$ be mixed actions in the one-shot game, such that $H(\mathbf{p}_L) = h_L$, $H(\mathbf{p}_R) = h_R$.*

Then player one can guarantee $\lambda \underline{v}(\mathbf{p}_L) + (1 - \lambda)\underline{v}(\mathbf{p}_R)$.

Proof. Let us briefly sketch the idea of the construction.

As for matching pennies, an ε-optimal strategy σ is defined by blocks of length l, and it uses on a given block only the private signals received during the previous block. The idea is to divide this stream of signals into two subsequences. The first one is encoded into a sequence of actions of length λl, which approximately mimics a sequence of i.i.d. variables, distributed according to $\mathbf{p}_L$. The second part is encoded into a sequence of actions of length $(1 - \lambda)l$, which mimics $\mathbf{p}_R$. As for matching pennies, the encodings are obtained by mapping typical sequences of signals to typical sequences of actions. This is feasible, provided the entropy of the subsequence of signals is roughly equal to the entropy of the distribution which player one tries to mimic.

We organize the proof as follows. For the moment, we let the length l of the blocks be any integer such that $\lambda l h_L/h$ is an integer, and we define σ (it is also convenient to assume that $\sqrt{l}$ is rational). σ depends on two small parameters η_L and η_R. We then prove that, provided l is large enough and η_l, η_R small enough, the average payoff guaranteed by σ on any single block exceeds $\lambda \underline{v}(\mathbf{p}_L) + (1 - \lambda)\underline{v}(\mathbf{p}_R) - \varepsilon$.

By stationarity, we deal only with the second block. We denote by $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_l)$ the random stream of signals available to player one at the beginning of the block. We set $l_1 = \lambda l h_L/h$, and denote by $\mathbf{x}_L$ ($\mathbf{x}_R$) the first l_1 (last $l - l_1$) components of $\mathbf{x}$. By construction, $H(\mathbf{x}) = lh$, $H(\mathbf{x}_L) = \lambda l h_L$, and $H(\mathbf{x}_R) = (1 - \lambda)l h_R$.

σ uses $\mathbf{x}_L$ to play in the first λl stages of the block, and $\mathbf{x}_R$ to play in the last $(1 - \lambda)l$ stages of the block. Since the two encodings of signals into actions are similar, we specify only the first part.

Step 1 (Coding sequences of signals into sequences of actions). Let $\eta_L > 0$ be small. It is convenient to assume that both $\alpha = \lambda l(h_L - \eta_L)(h + \eta_L)$ and $\beta = \lambda l$ are integers. This is consistent provided $(h_L - \eta_L)/(h + \eta_L)$ is rational; arbitrary small η_L can be chosen with this property. In the sequel, we simply write η for η_L.

Under σ, player one keeps the first α components of $\mathbf{x}_L$ (notice that $\alpha \leqslant l_1$), and encodes it into a sequence of actions of length β.

We address now the question of how to play $\mathbf{p}_L$ for β stages using a sequence of signals $\mathbf{x}_\alpha$ of length α.

We denote by $C(\eta, \alpha)$ the set of typical sequences of signals

$$C(\eta, \alpha) = \left\{ x \in C^\alpha : \frac{1}{2^{\alpha(h+\eta)}} \leqslant \mathbf{P}(x) \leqslant \frac{1}{2^{\alpha(h-\eta)}} \right\}.$$

Let $\mathbf{Q} = (\mathbf{p}_L)^{\otimes\beta}$ be the distribution on A^β of a sequence of β i.i.d. actions, each drawn according to $\mathbf{p}_L$. We denote by $A(\eta, \beta)$ the corresponding set of typical sequences

$$A(\eta, \beta) = \left\{ y \in A^\beta \colon \frac{1}{2^{\beta(h_L+\eta)}} \leqslant \mathbf{Q}(y) \leqslant \frac{1}{2^{\beta(h_L-\eta)}} \right\}.$$

Using Lemma 9, we now choose $K_0 > 0$ such that both $\mathbf{P}(C(\eta, \alpha)) > 1 - \varepsilon$ and $\mathbf{Q}(A(\eta, \beta)) > 1 - \varepsilon$ where $\eta = K_0/\sqrt{l}$.

Note that $C(\eta, \alpha)$ contains at most $2^{\alpha(h+\eta)}$ elements, and $A(\eta, \beta)$ contains at least $\mathbf{P}(A(\eta, \beta)) \times 2^{\beta(h_L-\eta)}$ elements. Hence, if $\mathbf{P}(A(\eta, \beta)) \geqslant 1/2$, $C(\eta, \alpha)$ contains at most twice as many points as $A(\eta, \beta)$.

Therefore, there exists $y_0 \in A^\beta$, and a map $i : C^\alpha \to A^\beta$ such that:

1. $i(C(\eta, \alpha)) \subseteq A(\eta, \beta)$, and does not contain y_0;
2. $i(\mathbf{x}_\alpha) = y_0$ for $\mathbf{x}_\alpha \notin C(\eta, \alpha)$;
3. $i^{-1}(y)$ contains at most two elements for $y \neq y_0$.

The strategy σ plays the sequence $i(x)$ of actions during the first β stages of the block. Denote by $\widehat{\mathbf{P}}$ the law of $i(\mathbf{x}_\alpha)$. We now prove that, provided l is large enough, $\underline{v}(\widehat{\mathbf{P}}) \geqslant \underline{v}(\mathbf{Q}) - K\varepsilon$, where K depends only on the payoff function g.

We denote by $\widehat{\mathbf{P}}_0$ the law of $i(\mathbf{x}_\alpha)$ conditional on $\mathbf{x}_\alpha$ being typical (i.e., $\mathbf{x}_\alpha \in C(\eta, \alpha)$).

Step 2 (A lower bound on the value guaranteed). We argue that if player two were informed whether $\mathbf{x}_\alpha \in C(\eta, \alpha)$ or not before playing in the block, then the best response of player two would lead to a lower payoff to player one. We rely here on the classical argument that more information corresponds to a broader strategy set, hence to a greater payoff to player two against the *same* strategy of player one.

Since $\mathbf{P}(C(\eta, \alpha)) > 1 - \varepsilon$, this shows that

$$\underline{v}(\widehat{\mathbf{P}}) \geqslant (1 - \varepsilon)\underline{v}(\widehat{\mathbf{P}}_0) - \varepsilon\|G\|,$$

where $\|G\| = \max_{i,a} |g^i(a)|$.

Step 3 (Comparison of $\widehat{\mathbf{P}}_0$ and $\mathbf{Q}$). It remains to prove that $\underline{v}(\widehat{\mathbf{P}}_0)$ is close to $\underline{v}(\mathbf{Q})$. This will be done by estimating the strategic distance between $\widehat{\mathbf{P}}_0$ and $\mathbf{Q}$ and using Lemma 7. We first estimate the standard Kullback distance between $\widehat{\mathbf{P}}_0$ and $\mathbf{Q}$.

Lemma 13. *There exists a constant K_1 such that, for every l*

$$d(\widehat{\mathbf{P}}_0\|\mathbf{Q}) \leqslant K_1 \sqrt{l}.$$

Proof.

$$
\begin{aligned}
d(\widehat{\mathbf{P}}_0 \| \mathbf{Q}) &= \mathbf{E}_{\widehat{\mathbf{P}}_0} \log \frac{\widehat{\mathbf{P}}_0(y)}{\mathbf{Q}(y)} \\
&= \mathbf{E}_{\widehat{\mathbf{P}}_0} \log \frac{\widehat{\mathbf{P}}(y)}{\mathbf{Q}(y)} - \log \mathbf{P}\big(C(\eta, \alpha)\big) \\
&\leqslant \big[-\alpha(h - \eta) + 1 \big] + \beta(h_L + \eta) - \log(1 - \varepsilon) \\
&\leqslant (\alpha + \beta)\eta + \beta h_L - \alpha h + 2 \\
&\leqslant \left(\frac{h_L - \eta}{h + \eta} + 1 \right) \lambda l \eta + \left(h_L - \frac{h(h_L - \eta)}{h + \eta} \right) \lambda l + 2
\end{aligned}
$$

the second equality uses the fact that no typical sequence is encoded into y_0, the first inequality uses the fact that $\widehat{\mathbf{P}}(y)$ is at most *twice* the maximal probability of a typical sequence of signals, and $\mathbf{Q}(y)$ is at least the minimal probability of a typical sequence of actions. The result follows, since $\eta = K_0 / l$.   □

We now complete the proof of Proposition 12. From Lemma 13, $(1/l) d(\widehat{\mathbf{P}}_0 \| \mathbf{Q})$ goes to 0 as l goes to infinity. Applying Proposition 8 implies that $d_S(\widehat{\mathbf{P}}_0 \| \mathbf{Q})$ also goes to zero as l goes to infinity.   □

7. Replies

We show in this section that player two can defend $\operatorname{cav} U(H(\mathbf{p}))$. This result is not new: it can be deduced from results by Neyman and Okada (2000). Since the proof is short, we provide it, for completeness. Let σ be a strategy of player one, known to player two. We define a strategy τ of player two as follows: in stage n, player two computes the distribution $\mathbf{p}_n$ of $\mathbf{a}_n$, conditional on past play $(\mathbf{a}_1, \mathbf{b}_1, \ldots, \mathbf{a}_{n-1}, \mathbf{b}_{n-1})$, and plays a pure best reply to this distribution.

This strategy τ defends $\operatorname{cav} U(H(\mathbf{p}))$ against σ.

Proposition 14. *For every n, $\gamma_n(\sigma, \tau) \leqslant \operatorname{cav} U(H(\mathbf{p}))$.*

Proof. For every $1 \leqslant m \leqslant n$, and play $(\mathbf{a}_1, \mathbf{b}_1, \ldots, \mathbf{a}_{m-1}, \mathbf{b}_{m-1})$,

$$
\mathbf{E}_{\sigma, \tau} \big[\mathbf{g}_m | \mathbf{a}_1, \ldots, \mathbf{b}_{m-1} \big] \leqslant U\big(H(\mathbf{p}_m) \big) \leqslant \operatorname{cav} U\big(H(\mathbf{p}_m) \big),
$$

since, conditional on past play, $\mathbf{a}_m$ has entropy $H(\mathbf{p}_m)$. By taking expectations and applying Jensen's inequality to $\operatorname{cav} U$, one gets

$$
\begin{aligned}
\mathbf{E}_{\sigma, \tau} \mathbf{g}_m &\leqslant \operatorname{cav} U\big(\mathbf{E}_{\sigma, \tau}\big(H(\mathbf{p}_m) \big) \big) \\
&= \operatorname{cav} U\big(H(\mathbf{a}_m | \mathbf{a}_1, \ldots, \mathbf{b}_{m-1}) \big) \\
&= \operatorname{cav} U\big(H(\mathbf{a}_m, \mathbf{b}_m | \mathbf{a}_1, \ldots, \mathbf{b}_{m-1}) \big),
\end{aligned}
$$

where the first equality follows by definition of the conditional entropy, and the second from the fact that $\mathbf{b}_m$ is a function of $\mathbf{a}_1, \ldots, \mathbf{b}_{m-1}$.

Now, by property of conditional entropy (and using Jensen's inequality again)

$$\frac{1}{n}\sum_{m=1}^{n}\mathbf{E}_{\sigma,\tau}\mathbf{g}_m \leqslant \operatorname{cav} U\left(\frac{1}{n}H(\mathbf{a}_1,\ldots,\mathbf{b}_n)\right).$$

Finally, notice that $(\mathbf{a}_1,\ldots,\mathbf{b}_n)$ is a function of $(X_1,\ldots,X_n)$, hence

$$H(\mathbf{a}_1,\ldots,\mathbf{b}_n) \leqslant H(X_1,\ldots,X_n) = nH(\mathbf{p}).$$

Since U is nondecreasing, so is $\operatorname{cav} U$, hence the result.

Remark. Of course, τ needs not be a best reply to σ in the long run.

Remark. Consider the example of matching pennies introduced above, and denote $\mathbf{p}_n = (p_n, 1 - p_n)$. If σ is to be an ε-optimal strategy of player one, it must be that the inequality $\gamma_n(\sigma, \tau) \leqslant (1/2)H(\mathbf{p})$ holds tightly. Therefore, it must be the case that $\min(p_n, 1 - p_n) \leqslant (1/2)H(\mathbf{p}_n)$ holds tightly, "most of the time." Equality holds only if $p_n = 0, 1$ or $p_n = 1/2$. Therefore, σ must be such that, "most of the time," player two either anticipates (almost) perfectly player one's next move ($p_n = 0$ or 1) or is (almost) completely ignorant of it (belief close to $(1/2, 1/2)$). These are the basic characteristics of the ε-optimal strategy for player one designed in Section 4. $\quad\square$

8. Extensions and concluding remarks

8.1. Markov chains

Remark that no use was made of the independence of the family $\{X_n\}$ except in the proof of Lemma 9. The notion of per stage entropy extends to the case of an irreducible and aperiodic Markov chain with finite state space. Assuming (X_n) is such a process with transition probabilities $(\pi_{i,j})$ and stationary probability measure (p_i), the entropy of (X_n) takes the value

$$h = -\sum_{i,j}\pi_{i,j}\log p_j.$$

Note that this definition is in accordance with the previous one when (X_n) is a i.i.d. sequence of random variables.

By the central limit theorem for functions of mixing processes (see, e.g., Theorem 21.1 in Billingsley, 1968), the sequence $(1/\sqrt{n})(\log \mathbf{P}(X_1,\ldots,X_n) + nh)$ either converges in distribution to a Gaussian variable, or converges in probability to zero (see p. 187 in Billingsley, 1968). In both cases, there exists K and large enough such that, for every $n \geqslant N$,

$$\mathbf{P}\left\{(x_1,\ldots,x_n) : -K \leqslant \frac{\ln \mathbf{P}(x_1,\ldots,x_n) + nh}{\sqrt{n}} \leqslant K\right\} \geqslant 1 - \varepsilon.$$

Therefore, the conclusion of Lemma 9 still holds for n large enough. Hence, Theorem 2 still holds.

An interesting extension (left open here) would certainly be to study the case in which player one can also control by his actions the transition probabilities of the Markov chain. This would lead us to study the trade-off between payoff acquisition and entropy acquisition.

8.2. Repeated games of imperfect monitoring

In a repeated game with imperfect monitoring, players can use the signaling structure of the game to generate correlation. These possibilities of internal correlation, first noticed by Lehrer (1991), result in a broader set of Nash equilibria of the repeated game. Let us informally discuss the question of characterizing the min max level y^i for player i in n-player repeated game with imperfect monitoring. Obviously, because players $-i$ (others that i) can play repeatedly according to any profile of mixed strategies, y^i is at most equal to the min max level v^i of player i in the one shot game (in mixed strategies). On the other hand, player i can always defend his minmax level in correlated strategies w^i. Hence $w^i \leqslant y^i \leqslant v^i$. Since players $-i$ may have possibilities to correlate their actions using the signaling structure of the game, we may have $y^i < v^i$. It is the case that $y^i = w^i$ when players $-i$ can be seen as a unique player that may choose any mixed strategy (over the tuples of actions of players $-i$) in the one shot game. This however is not true in general, and players $-i$ seen as a unique player may be limited in its strategy space.

Thus, in order to characterize y^i in the general case, one has to study the optimal trade-off between correlation acquisition and efficient punishment of player i. We hope that our work may serve as a first step towards such a characterization.

Acknowledgment

Appendix A. Proof of Proposition 8

A.1. Preliminaries

In the sequel, we shall often use functions $h : \mathbb{R}^+ \to \mathbb{R}^+$ that are non-decreasing, continuous, concave, and such that $h(0) = 0$. For simplicity we shall call them *nice* functions.

We first establish an existence result of nice functions.

Lemma 15. *Let K be a topological space and $f, g : K \to \mathbb{R}^+$. Assume there exists $\beta : \mathbb{R}^+ \to \mathbb{R}^+$ concave such that $g \leqslant \beta \circ f$, that $\{x : f(x) = 0\} \neq \emptyset$ and $\{x : f(x) = 0\} \subseteq \{x : g(x) = 0\}$. Then there exists a nice function α such that $g \leqslant \alpha \circ f$.*

Proof. The correspondence ψ from $\mathbb{R}^+$ to K defined by $\psi(y) = \{x \in K : f(x) \leqslant y\}$ is non-decreasing and non-empty valued, and g is bounded by $\beta \circ f$ on $\psi(y)$. Hence, the function α_0 given by $\alpha_0(y) = \sup_{\psi(y)} g$ is non-decreasing and takes its values on $\mathbb{R}^+$, and $\alpha_0 \leqslant \beta$. Hence the concave mapping $\alpha = \text{cav}\,\alpha_0$ is a well-defined $\mathbb{R}$-valued function and satisfies $g \leqslant \alpha \circ f$. Concavity of α on $\mathbb{R}^+$ implies continuity on $\mathbb{R}^{+*}$, and since 0 is an extreme point of $\mathbb{R}^+$, $\alpha(0) = \alpha_0(0) = 0$.

Recall that $\alpha = \inf\{\gamma : \gamma \text{ linear and } \gamma \geqslant \alpha_0\}$. Since α_0 is non-decreasing, all linear functions $\gamma \geqslant \alpha_0$ are non-decreasing, and so is α.

It remains to prove the continuity at 0. α being concave is dominated by an affine mapping: $\alpha(x) \leqslant ax + b$. Fix $1 > \varepsilon > 0$, and consider $\eta > 0$ such that $\max_{[0, \eta]} \alpha_0 \leqslant \varepsilon$. Take $x \in [0, \eta\varepsilon]$, $x = (1 - p)x_1 + px_2$ with $x_2 \geqslant x_1 \geqslant 0$. If $x_2 \leqslant \eta$ then clearly $(1 - p)\alpha_0(x_1) + p\alpha_0(x_2) \leqslant \varepsilon$. If $x_2 \geqslant \eta$ then $p \leqslant \varepsilon$ so that $p\alpha_0(x_2) \leqslant pax_2 + pb \leqslant ax + \varepsilon b$. Hence in both cases $(1 - p)\alpha_0(x_1) + p\alpha_0(x_2) \leqslant \varepsilon(a\eta + b + 1)$. Therefore in both cases $0 \leqslant \alpha(x) \leqslant \varepsilon(a\eta + b + 1)$ for $x \in [0, \eta\varepsilon]$, which implies continuity at 0. $\square$

A.2. Kullback and absolute Kullback distance

We recall some well-known elementary properties of the Kullback distance d. Recall that Ω is the set of $(\mathbf{p}, \mathbf{q})$ such that $\mathbf{p}$ is absolutely continuous w.r.t. $\mathbf{q}$.

Lemma 16. *The mapping d is separately convex in each variable on Ω, $d \geqslant 0$ on Ω, and $d(\mathbf{p} \| \mathbf{q}) = 0$ if and only if $\mathbf{p} = \mathbf{q}$.*

We define on Ω the *absolute* Kullback distance as

$$|d|(\mathbf{p} \| \mathbf{q}) = \mathbf{E_p} \left| \log \frac{\mathbf{p}(s)}{\mathbf{q}(s)} \right|.$$

The Kullback distance is a standard tool in statistics and information theory. However, we are not aware of any use of the absolute Kullback distance.

Obviously, $d \leqslant |d|$, and $\{d = 0\} = \{|d| = 0\}$.

Lemma 17. *There exists a nice function α_2, such that $|d| \leqslant \alpha_2 \circ d$ on Ω.*

For future use, it is crucial to make here the following observation. In this lemma, Ω (hence the underlying set S) is given. The nice function may thus depend on S. However, it will be clear from the proof below, that the nice function we exhibit is independent of S.

Proof. Given Lemma 15, we need only to prove that $|d|$ is majorized by a concave function of d. We prove that $|d| \leqslant d + 2$.

Let $(\mathbf{p}, \mathbf{q}) \in \Omega$. Set $S_1 = \{s, \mathbf{p}(s) > \mathbf{q}(s)\}$ and $S_2 = \{s, \mathbf{p}(s) \leqslant \mathbf{q}(s)\}$.

Set $f(\mathbf{p}, \mathbf{q}) = \sum_{s \in S_1} \mathbf{p}(s) \log \mathbf{p}(s)/\mathbf{q}(s)$ and $g(\mathbf{p}, \mathbf{q}) = \sum_{s \in S_2} \mathbf{p}(s) \log \mathbf{q}(s)/\mathbf{p}(s)$, so that $d(\mathbf{p} \| \mathbf{q}) = f(\mathbf{p}, \mathbf{q}) - g(\mathbf{p}, \mathbf{q})$ and $|d|(\mathbf{p} \| \mathbf{q}) = f(\mathbf{p}, \mathbf{q}) + g(\mathbf{p}, \mathbf{q})$. Thus, the result follows from the claim below.

Claim. $g(\mathbf{p}, \mathbf{q}) \leqslant 1$ *for every* $(\mathbf{p}, \mathbf{q})$.

For $\mathbf{p}$ fixed, we shall prove that $\max_{\mathbf{q}} g(\mathbf{p}, \mathbf{q}) \leqslant 1$. Let

$$\mathcal{Q} = \big\{ \mathbf{q} : (\mathbf{p}, \mathbf{q}) \in \Omega \big\} \quad \text{and} \quad S_2 = \big\{ s : \mathbf{q}(s) \geqslant \mathbf{p}(s) \big\}.$$

Write

$$\mathcal{Q} = \bigcup_{y \geqslant 1} \mathcal{Q}_y,$$

with $\mathcal{Q}_y = \mathcal{Q} \cap \{\mathbf{q}: \mathbf{q}(S_2) = y\mathbf{p}(S_2)\}$. We prove that $\max_{\mathbf{q} \in \mathcal{Q}_y} g(\mathbf{p}, \mathbf{q}) \leqslant 1$.

The map $\mathbf{q} \mapsto g(\mathbf{p}, \mathbf{q}) = \sum_{S_2} \mathbf{p}(s) \log \mathbf{q}(s)/\mathbf{p}(s)$ is concave on $\mathcal{Q}_y$. It is therefore immediate to check that it is maximized at any $\mathbf{q}^*$ which satisfies $\mathbf{q}^*(s) = y\mathbf{p}(s)$, for every $s \in S_2$. Notice that

$$g(\mathbf{p}, \mathbf{q}^*) = \sum_{S_2} \mathbf{p}(s) \log y = \mathbf{p}(S_2) \log y = \mathbf{q}(S_2)\frac{\log y}{y} \leqslant \frac{\log y}{y}.$$

The claim follows.   □

A.3. Strategic distance and Kullback distance

The purpose of this section is to prove the next result, which clearly implies Proposition 8.

Proposition 18. *There exists a nice function α such that*

$$d_S(\mathbf{P}\|\mathbf{Q}) \leqslant \alpha\left(\frac{1}{n}d(\mathbf{P}\|\mathbf{Q})\right)$$

for every $n \in \mathbb{N}$, $\mathbf{P}, \mathbf{Q} \in \Delta(A^n)$ such that $\mathbf{P} \ll \mathbf{Q}$.

Proof. Let first $k \in \{0, \dots, n-1\}$. At stage $k+1$, one has

$$\mathbf{E}_{\mathbf{P}}\big[\big\|\mathbf{P}(\cdot|\mathcal{H}_{k+1}) - \mathbf{Q}(\cdot|\mathcal{H}_{k+1})\big\|_1\big] = \sum_{a^k \in A^k} \mathbf{P}(a^k) \sum_{a \in A} |\mathbf{P}(a|a^k) - \mathbf{Q}(a|a^k)|.$$

Let $g(\mathbf{p}, \mathbf{q}) = \sum_{a \in A} |\mathbf{p}(a) - \mathbf{q}(a)|$ and $f(\mathbf{p}, \mathbf{q}) = \sum_{a \in A} \mathbf{p}(a)|\mathbf{p}(a) - \mathbf{q}(a)|$. Both f and g are continuous and nonnegative over the compact set $\Delta(A) \times \Delta(A)$. Moreover,

$$f(\mathbf{p}, \mathbf{q}) = 0 \quad \Rightarrow \quad \big(\forall a, \ \mathbf{p}(a) = 0 \text{ or } \mathbf{p}(a) = \mathbf{q}(a)\big).$$

Since $\sum_A \mathbf{p}(a) = \sum_A \mathbf{q}(a)$, this implies $\mathbf{p} = \mathbf{q}$, hence $g(\mathbf{p}, \mathbf{q}) = 0$.

Thus, $\{f = 0\} \subseteq \{g = 0\}$ and clearly $\{f = 0\} \neq \emptyset$. Furthermore, with β the constant function 2, $g \leqslant \beta \circ f$. Lemma 15 yields a nice function α_1 such that $g \leqslant \alpha_1 \circ f$. Therefore,

$$\sum_{a^k \in A^k} \mathbf{P}(a^k)\left(\sum_{a \in A}|\mathbf{P}(a|a^k) - \mathbf{Q}(a|a^k)|\right)$$

$$\leqslant \sum_{a^k \in A^k} \mathbf{P}(a^k)\alpha_1\left(\sum_{a \in A}\mathbf{P}(a|a^k)|\mathbf{P}(a|a^k) - \mathbf{Q}(a|a^k)|\right). \tag{8}$$

Hence

$$d_S(\mathbf{P}\|\mathbf{Q}) \leqslant \frac{1}{n}\sum_{k=0}^{n-1}\sum_{a^k \in A^k} \mathbf{P}(a^k)\alpha_1\left(\sum_{a \in A}\mathbf{P}(a|a^k)|\mathbf{P}(a|a^k) - \mathbf{Q}(a|a^k)|\right)$$

$$\leqslant \alpha_1\left(\frac{1}{n}\sum_{k=0}^{n-1}\sum_{a^k \in A^k} \mathbf{P}(a^k)\sum_{a \in A}\mathbf{P}(a|a^k)|\mathbf{P}(a|a^k) - \mathbf{Q}(a|a^k)|\right),$$

where the second step uses Jensen's inequality.

Since $\mathbf{P} \ll \mathbf{Q}$, whenever $\mathbf{P}(a^k) > 0$ one has

$$\mathbf{P}(a|a^k) > 0 \quad \Rightarrow \quad \mathbf{Q}(a|a^k) > 0.$$

Using the fact that $|x - y| \leqslant |\log x - \log y|$ for every $x, y \in]0, 1[$, one deduces

$$d_S(\mathbf{P}\|\mathbf{Q}) \leqslant \alpha_1\left(\frac{1}{n}\sum_{k=0}^{n-1}\sum_{a^k \in A^k} \mathbf{P}(a^k)|d|\big(\mathbf{Q}(\cdot|a^k)\|\mathbf{P}(\cdot|a^k)\big)\right).$$

By using Lemma 17, then twice Jensen's inequality

$$d_S(\mathbf{P}\|\mathbf{Q}) \leqslant \alpha_1\left(\frac{1}{n}\sum_{k=0}^{n-1}\sum_{a^k \in A^k} \mathbf{P}(a^k)\alpha_2\big(d\big(\mathbf{P}(\cdot|a^k)\|\mathbf{Q}(\cdot|a^k)\big)\big)\right)$$

$$\leqslant \alpha_1 \circ \alpha_2\left(\frac{1}{n}\sum_{k=0}^{n-1}\sum_{a^k \in A^k} \mathbf{P}(a^k)d\big(\mathbf{P}(\cdot|a^k)\|\mathbf{Q}(\cdot|a^k)\big)\right).$$

We now check that the argument of $\alpha_1 \circ \alpha_2$ is simply $(1/n)d(\mathbf{P}\|\mathbf{Q})$.

$$\frac{1}{n}\sum_{k=0}^{n-1}\sum_{a^k \in A^k} \mathbf{P}(a^k)d\big(\mathbf{P}(\cdot|a^k)\|\mathbf{Q}(\cdot|a^k)\big)$$

$$= \frac{1}{n}\sum_{k=0}^{n-1}\sum_{a^k \in A^k} \mathbf{P}(a^k)\sum_{a \in A}\mathbf{P}(a|a^k)\log\frac{\mathbf{P}(a|a^k)}{\mathbf{Q}(a|a^k)}$$

$$= \frac{1}{n}\sum_{k=0}^{n-1}\sum_{a^k \in A^k} \mathbf{P}(a^k)\sum_{a \in A}\mathbf{P}(a|a^k)\left\{\log\frac{\mathbf{P}(a^k, a)}{\mathbf{Q}(a^k, a)} - \log\frac{\mathbf{P}(a^k)}{\mathbf{Q}(a^k)}\right\}$$

$$= \frac{1}{n}\sum_{k=0}^{n-1}\left\{\sum_{a^k \in A^k}\sum_{a \in A}\mathbf{P}(a^k)\mathbf{P}(a|a^k)\log\frac{\mathbf{P}(a^k, a)}{\mathbf{Q}(a^k, a)} - \sum_{a^k \in A^k}\mathbf{P}(a^k)\log\frac{\mathbf{P}(a^k)}{\mathbf{Q}(a^k)}\right\}$$

$$= \frac{1}{n}\sum_{k=0}^{n-1}\big\{d(\mathbf{P}_{k+1}\|\mathbf{Q}_{k+1}) - d(\mathbf{P}_k\|\mathbf{Q}_k)\big\}$$

$$= d(\mathbf{P}\|\mathbf{Q}),$$

where $\mathbf{P}_k$ and $\mathbf{Q}_k$ are the marginals of $\mathbf{P}$ and $\mathbf{Q}$ over A^k. The result follows, since $\alpha_1 \circ \alpha_2$ is nice. $\quad\square$

References

Billingsley, P., 1968. Convergence of Probability Measures. In: Wiley Series in Probability and Mathematical Statistics. Wiley, New York.

Cover, T.M., Thomas, J.A., 1991. Elements of Information Theory. In: Wiley Series in Telecomunications. Wiley, New York.

Lehrer, E., 1991. Internal correlation in repeated games. Int. J. Game Theory 19, 431–456.

Lehrer, E., 1994. Finitely many players with bounded recall in infinitely repeated games. Games Econ. Behav. 7, 390–405.

Lehrer, E., Smorodinsky, R., 2000. Relative entropy in sequential decision problems. J. Math. Econ. 33, 425–440.

Neyman, A., Okada, D., 1999. Strategic entropy and complexity in repeated games. Games Econ. Behav. 29, 191–223.

Neyman, A., Okada, D., 2000. Repeated games with bounded entropy. Games Econ. Behav. 30, 228–247.

Shannon, C., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27, 379–423. 623–656.