EMPIRICAL DISTRIBUTIONS OF BELIEFS UNDER IMPERFECT OBSERVATION

OLIVIER GOSSNER AND TRISTAN TOMALA

Abstract. Let $(\mathbf{x}_n)_n$ be a process with values in a finite set X and law P, and let $\mathbf{y}_n = f(\mathbf{x}_n)$ be a function of the process. At stage n, $\mathbf{p}_n = P(\mathbf{x}_n \mid \mathbf{x}_1, \ldots, \mathbf{x}_{n-1})$, element of $\Pi = \Delta(X)$, is the belief of a perfect observer of the process on its realization at stage n. A statistician observing $\mathbf{y}_1, \ldots, \mathbf{y}_n$'s holds a belief $\mathbf{e}_n = P(\mathbf{p}_n \mid \mathbf{x}_1, \ldots, \mathbf{x}_n) \in \Delta(\Pi)$ on the possible predictions of the perfect observer. Given X and f, we characterize set of limiting expected empirical distributions of the process $(\mathbf{e}_n)$ when P ranges over all possible laws of $(\mathbf{x}_n)_n$.

Date: January 2004.

1. Introduction

We study the gap in predictions made by agents that observe different signals about some process $(\mathbf{x}_n)_n$ with values in a finite set X and law P. Assume that a perfect observer observes $(\mathbf{x}_n)_n$, and a statistician observes a function $\mathbf{y}_n = f(\mathbf{x}_n)$. At stage n, $\mathbf{p}_n = P(\mathbf{x}_n \mid \mathbf{x}_1, \ldots, \mathbf{x}_{n-1})$, element of $\Pi = \Delta(X)$, is the best prediction that a perfect observer of the process can make on its next realization. To a sequence of signals $\mathbf{y}_1, \ldots, \mathbf{y}_n$ corresponds a belief $\mathbf{e}_n = P(\mathbf{p}_n \mid \mathbf{y}_1, \ldots, \mathbf{y}_{n-1})$ that the statistician holds on the possible predictions of the perfect observer. The information gap about the future realization of the process at stage n between the perfect observer and the statistician is seen in the fact that the perfect observer knows $\mathbf{p}_n$, whereas the statistician knows only the law $\mathbf{e}_n$ of $\mathbf{p}_n$ conditional to $\mathbf{y}_1, \ldots, \mathbf{y}_{n-1}$.

We study the possible limiting expected empirical distributions of the process $(\mathbf{e}_n)$ when P ranges over all possible laws of $(\mathbf{x}_n)_n$.

Call experiment elements of $E = \Delta(\Pi)$ and experiment distribution elements of $\Delta(E)$. We say that an experiment distribution δ is achievable if there is a law P of the process for which δ is the limiting expected empirical distributions of $(\mathbf{e}_n)$. Represent an experiment e by a random variable $\mathbf{p}$ with finite support and values in Π. Let $\mathbf{x}$ be a random variable with values in X such that, conditional on the realization p of $\mathbf{p}$, $\mathbf{x}$ has law p. Let then $\mathbf{y} = f(\mathbf{x})$. We define the entropy variation associated to e as:

$$\Delta H(e) = H(\mathbf{p}, \mathbf{x}|\mathbf{y}) - H(\mathbf{p}) = H(\mathbf{x}|\mathbf{p}) - H(\mathbf{y})$$

This operator measures the evolution of the uncertainty for the statistician on the predictions of the perfect observer.

Our main result is that an experiment distribution δ is achievable if and only if $\mathbf{E}_\delta(\Delta H) \geq 0$.

This result has applications to statistical problems and to game theoretic ones.

Assume that at each stage, both the perfect observer and the statistician take a decision, and the payoff to each decision maker is a function of his decision and the realization of the process. Then, given that both agents maximize expected utilities, their expected payoffs at stage n write as a function of $\mathbf{e}_n$. Consequently, their long-run expected payoffs is a function of the long-run expected empirical distribution of the process $(\mathbf{e}_n)$. One application of our result (under progress) is a characterization of the bounds on the value of information in repeated decision problems.

Information asymmetries in repeated interactions is also a recurrent phenomenon in game theory, and arise in particular when agents observe private signals, or have limited information processing abilities.

In a repeated game with private signals, each players observes at each stage of the game a signal that depends on the action profile of all the players. While public equilibria of these games (see e.g. Abreu, Pierce and Stachetti [APS90] and Fudenberg, Levine and Maskin [FLM94]), or equilibria in which a communication mechanism serves to resolve information asymmetries (see e.g. Compte [Com02] and Renault and Tomala [RT00]) are well characterized, endogenous correlation and endogenous communication gives rise to difficult questions that have only been tackled for particular classes of signalling structures (see Lehrer [Leh90] [Leh91], Renault and Tomala [RT98], and Gossner and Vieille [GV01]).

When agents have different information processing abilities, some players may be able to predict more accurately the future process of actions than others. These phenomena have been studied in the frameworks of finite automata (see Ben Porath [BP93], Neyman [Ney97] [Ney98], Gossner and Hernández [GH03], Bavly and Neyman [BN03], Lacôte and Thurin [LT03]), bounded recall (see Lehrer [Leh88] [Leh94], Piccione and Rubinstein [PR03], Bavly and Neyman [BN03], Lacôte and Thurin [LT03]), and time-constrained Turing machines (see Gossner [Gos98] [Gos00]).

Our result has already found applications to the characterization of the minmax values in classes of repeated games with imperfect monitoring (see Gossner and Tomala [GT03], Gossner, Laraki and Tomala [GLT03], and Goldberg [Gol03]).

Next section presents the model and main results, while the remain of the paper is devoted to the proof of our theorem.

2. Definitions and main results

2.1. Notations.

For a finite set S, $|S|$ denotes its cardinality.

For S compact, $\Delta(S)$ denotes the set of regular probability measures on S endowed with the weak-$*$ topology (thus $\Delta(S)$ is compact).

If $(\mathbf{x}, \mathbf{y})$ is a pair of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ such that $\mathbf{x}$ is finite, $P(\mathbf{x}|y)$ denotes the conditional distribution of $\mathbf{x}$ given $\{\mathbf{y} = y\}$ and $P(\mathbf{x}|\mathbf{y})$ is the random variable with value $P(\mathbf{x}|y)$ if $\mathbf{y} = y$.

ϵ_x denotes the Dirac measure on x, i.e. the probability measure with support $\{x\}$.

If $\mathbf{x}$ is a random variable with values in a compact subset of a topological vector space V, $\mathbf{E}(\mathbf{x})$ denotes the barycenter of $\mathbf{x}$ and is the element of V such that for each continuous linear form φ, $\mathbf{E}(\varphi(\mathbf{x})) = \varphi(\mathbf{E}(\mathbf{x}))$.

If p and q are probability measures on two probability spaces, $p \otimes q$ denotes the direct product of p and q, i.e. $(p \otimes q)(A \times B) = p(A) \times q(B)$.

2.2. Definitions.

2.2.1. *Processes and Distributions.*

Let $(\mathbf{x}_n)_n$ be a process with values in a finite set X such that $|X| \geq 2$ and let P be its law. A statistician gets to observe the value of $\mathbf{y}_n = f(\mathbf{x}_n)$ at each stage n, where $f \colon X \to Y$ is a fixed mapping. Before stage n, the history of the process is $x_1, \ldots, x_{n-1}$ and the the history available to the statistician is $y_1, \ldots, y_{n-1}$. The law of $\mathbf{x}_n$ given the history of the process is:

$$\mathbf{p}_n(x_1, \ldots, x_{n-1}) = P(\mathbf{x}_n | x_1, \ldots, x_{n-1})$$

This defines a $(\mathbf{x}_1, \ldots, \mathbf{x}_{n-1})$-measurable random variable $\mathbf{p}_n$ with values in $\Pi = \Delta(X)$. The statistician holds a belief on the value of $\mathbf{p}_n$. For each history $y_1, \ldots, y_{n-1}$, we let:

$$\mathbf{e}_n(y_1, \ldots, y_{n-1}) = P(\mathbf{p}_n | y_1, \ldots, y_{n-1})$$

This defines a $(\mathbf{y}_1, \ldots, \mathbf{y}_{n-1})$-measurable random variable $\mathbf{e}_n$ with values in $E = \Delta(\Pi)$. Following Blackwell [Bla51] [Bla53], we call experiments the elements of E.

The empirical distribution of experiments up to stage n is:

$$\mathbf{d}_n(y_1, \ldots, y_{n-1}) = \frac{1}{n} \sum\nolimits_{m \leq n} \epsilon_{\mathbf{e}_n(y_1, \ldots, y_{n-1})}$$

The $(\mathbf{y}_1, \ldots, \mathbf{y}_{n-1})$-measurable random variable $\mathbf{d}_n$ has values in $D = \Delta(E)$. We call D the set of *experiment distributions.*

Definition 1. *We say that the law P of the process* n-achieves *the experiment distribution δ if $\mathbf{E}_P(\mathbf{d}_n) = \delta$, and that δ is* n-achievable *if there exists P that n-achieves δ. D_n denotes the set of n-achievable experiment distributions.*

We say that the law P of the process achieves *the experiment distribution δ if $\lim_{n \to +\infty} \mathbf{E}_P(\mathbf{d}_n) = \delta$, and that δ is* achievable *if there exists P that achieves δ. D_∞ denotes the set of achievable experiment distributions.*

Proposition 2. (1) *For $n, m \geq 1$, $\frac{n}{n+m} D_n + \frac{m}{n+m} D_m \subset D_{m+n}$.*
(2) *$D_n \subset D_\infty$.*
(3) *D_∞ is the closure of $\cup_n D_n$.*
(4) *D_∞ is convex and closed.*

Proof. To prove (1) and (2), let P_n and P'_m be the laws of processes $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and $\mathbf{x}'_1, \ldots, \mathbf{x}'_m$ such that P_n n-achieves $\delta_n \in D_n$ and P'_m m-achieves $\delta'_m \in D_m$. Then any process of law $P_n \otimes P'_m$ $(n+m)$-achieves $\frac{n}{n+m}\delta_n + \frac{m}{n+m}\delta'_m \in D_{m+n}$, and any process of law $P_n \otimes P_n \otimes P_n \otimes \ldots$ achieves

$\delta_n \in D_\infty$. Point (3) is a direct consequence of the definitions and of (2). Point (4) now follows from (1) and (3). □

Example 1

Assume f is constant, let $(\mathbf{x}_n)_n$ be the the process on $\{0,1\}$ such that $(\mathbf{x}_{2n-1})_{n\leq 1}$ are i.i.d. uniformly distributed and $\mathbf{x}_{2n} = \mathbf{x}_{2n-1}$. At odd stages $e_{2n-1} = \epsilon_{(\frac{1}{2},\frac{1}{2})}$ a.s. and at the even stages $e_2 = \frac{1}{2}\epsilon_{(1,0)} + \frac{1}{2}\epsilon_{(0,1)}$ a.s. Hence the law of $(\mathbf{x}_n)_n$ achieves the experiment distribution $\frac{1}{2}\epsilon_{e_1} + \frac{1}{2}\epsilon_{e_2}$.

Example 2

Assume again f constant, a parameter p is drawn uniformly in $[0,1]$, and $(\mathbf{x}_n)_n$ is a family of i.i.d. Bernoulli random variables with parameter p. In this case, $p_n \to p$ a.s., and therefore e_n weak-$*$ converges to the uniform distribution on $[0,1]$. The experiment distribution achieved by the law of this process is thus the Dirac measure on the uniform distribution on $[0,1]$.

2.2.2. *Measures of uncertainty.* Let $\mathbf{x}$ be a finite random variable with values in X and with law P. Throughout the paper, log denotes the logarithm with base 2. By definition, the entropy of $\mathbf{x}$ is:

$$H(\mathbf{x}) = -\mathbf{E}\log P(\mathbf{x}) = -\sum\nolimits_x P(x)\log P(x)$$

where $0\log 0 = 0$ by convention. Note that $H(\mathbf{x})$ is non-negative and depends only on the law P of $\mathbf{x}$. The entropy of a random variable $\mathbf{x}$ is thus the the entropy $H(P)$ of its distribution P, with $H(P) = -\sum P(x)\log P(x)$.

Let $(\mathbf{x},\mathbf{y})$ be a couple of random variables with joint law P such that $\mathbf{x}$ is finite. The conditional entropy of $\mathbf{x}$ given $\{\mathbf{y} = y\}$ is the entropy of the conditional distribution $P(\mathbf{x}|y)$:

$$H(\mathbf{x} \mid y) = -\mathbf{E}[\log P(\mathbf{x} \mid y)]$$

The conditional entropy of $\mathbf{x}$ given $\mathbf{y}$ is the expected value of the previous:

$$H(\mathbf{x} \mid \mathbf{y}) = \int H(\mathbf{x} \mid y)dP(y)$$

If $\mathbf{y}$ is also finite, one has the following relation of additivity of entropies:

$$H(\mathbf{x}, \mathbf{y}) = H(\mathbf{y}) + H(\mathbf{x} \mid \mathbf{y})$$

Given an experiment e, let $\mathbf{p}$ be a random variable in Π with distribution e, $\mathbf{x}$ be a random variable in X such that the distribution of $\mathbf{x}$ conditional on $\{\mathbf{p} = p\}$ is p and $\mathbf{y} = f(\mathbf{x})$.

Definition 3. *The* **entropy variation** *associated to e is:*

$$\Delta H(e) = H(\mathbf{x}|\mathbf{p}) - H(\mathbf{y})$$

Remark 4. Assume that e has finite support. From the additivity formula:

$$H(\mathbf{p}, \mathbf{x}) = H(\mathbf{p}) + H(\mathbf{x}|\mathbf{p}) = H(\mathbf{y}) + H(\mathbf{p}, \mathbf{x}|\mathbf{y})$$

Therefore: $\Delta H(e) = H(\mathbf{p}, \mathbf{x}|\mathbf{y}) - H(\mathbf{p})$.

The interpretation is the following. The operator ΔH measures the evolution of the uncertainty of the statistician at a given stage. Assume e is the experiment representing the information gap between the perfect observer and the statistician that at stage n. The evolution of information can be seen with the following procedure:

- Draw $\mathbf{p}$ according to e;
- If $\mathbf{p} = p$, draw $\mathbf{x}$ according to p;
- Announce $\mathbf{y} = f(\mathbf{x})$ to the statistician.

The uncertainty for the statistician at the beginning of the procedure is $H(\mathbf{p})$. At the end of the procedure, the statistician knows the value of $\mathbf{y}$ and $\mathbf{p}, \mathbf{x}$ are unknown to him, the new uncertainty is thus $H(\mathbf{p}, \mathbf{x}|\mathbf{y})$. $\Delta H(e)$ is therefore the variation of entropy across this procedure. Note that it also writes as the difference between the entropy added to $\mathbf{p}$ by the procedure: $H(\mathbf{x}|\mathbf{p})$, and the entropy of the information gained by the statistician: $H(\mathbf{y})$.

Lemma 5. *The operator $\Delta H : E \to \mathbb{R}$ is continuous.*

Proof. $H(\mathbf{x}|\mathbf{p}) = \int H(\mathbf{x}|p)de(p)$ is linear-continuous in e, since H is a continuous on Π. The mapping that associates to e, the probability distribution of $\mathbf{y}$ is also linear-continuous. □

2.3. **Main results.** We characterize achievable distributions.

Theorem 6. *An experiment distribution δ is achievable if and only if $\mathbf{E}_\delta(\Delta H) \geq 0$.*

We also prove a stronger version of the previous theorem in which the transitions of the process are restricted to belong to an arbitrary subset of Π.

Definition 7. *The distribution $\delta \in D$ has support in $C \subset \Pi$ if for each e in the support of δ, the support of e is a subset of C.*

Definition 8. *Given $C \subset \Pi$ a process $(\mathbf{x}_n)_n$ with law P is a C-process if for each n, $P(\mathbf{x}_n|\mathbf{x}_1, \ldots, \mathbf{x}_{n-1}) \in C$, P-almost surely.*

Theorem 9. *Let C be a closed subset of Π. If δ has support in C and $\mathbf{E}_\delta(\Delta H) \geq 0$, then δ is achievable by the law of a C-process.*

Remark 10. If C is closed, the set of experiment distributions that are achievable by laws of C-processes is convex and closed. The proof is identical as for D_∞ so we omit it.

2.4. **Trivial observation.** We say that the observation is trivial when f is constant.

Lemma 11. *If the observation is trivial, any δ is achievable.*

This fact can easily be deduced from theorem 6. Since f is constant, $H(\mathbf{y}) = 0$ and thus $\Delta H(e) \geq 0$ for each $e \in E$. However, a simple construction provides a direct proof in this case.

Proof. By closedness and convexity, it is enough to prove that any $\delta = \epsilon_e$ with e of finite support is achievable. Let thus $e = \sum_k \lambda_k \epsilon_{p_k}$. Again by closedness, assume that the λ_k's are rational with common denominator 2^n for some n. Let $x \neq x'$ be two distinct points in X and $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be i.i.d. with law $\frac{1}{2}\epsilon_x + \frac{1}{2}\epsilon_{x'}$, so that $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is uniform on a set with 2^n elements. Map $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ to some random variable $\mathbf{k}$ such that $P(\mathbf{k} = k) = \lambda_k$. Construct then the law P of the process such that conditional on $\mathbf{k} = k$, $\mathbf{x}_{t+n}$ has law p_k for each t. P achieves δ. □

2.5. **Perfect observation.** We say that information is perfect when f is one-to-one. Let E_d denote the set of Dirac experiments, i.e. measures on Π whose support are a singleton. This set is a weak-$*$-closed subset of E.

Lemma 12. *If information is perfect, δ is achievable if and only if* $\operatorname{supp}\delta \subset E_d$.

We derive this result from thm. 6.

Proof. If $e \in E_d$, the random variable $\mathbf{p}$ associated to e is constant a.s., therefore $H(\mathbf{x}|\mathbf{p}) = H(\mathbf{x}) = H(\mathbf{y})$ since observation is perfect. Thus $\Delta H(e) = 0$, and $\mathbf{E}_\delta(\Delta H) = 0$ if $\operatorname{supp}\delta \subset E_d$. Conversely, assume $\mathbf{E}_\delta(\Delta H) = 0$. Since the observation is perfect, $H(\mathbf{y}) = H(\mathbf{x}) \geq H(\mathbf{x}|\mathbf{p})$ and thus $\Delta H(e) \leq 0$ for all e. So, $\Delta H(e) = 0$ δ-almost surely, i.e. $H(\mathbf{x}|\mathbf{p}) = H(\mathbf{x})$ for each e in a set of δ-probability one. For each such e, $\mathbf{x}$ and $\mathbf{p}$ are independent, i.e. the law of $\mathbf{x}$ given $\mathbf{p} = p$ does not depend on p. Hence e is a Dirac measure. □

2.6. **Example of a non-achievable experiment distribution.**

Example 3

Let $X = \{i, j, k\}$ and $f(i) = f(j) \neq f(k)$. Consider distributions of the type $\delta = \epsilon_e$.

If $e = \epsilon_{\frac{1}{2}\epsilon_j + \frac{1}{2}\epsilon_k}$, δ is achievable. Indeed, such δ is induced by an i.i.d. process with stage law $\frac{1}{2}\epsilon_j + \frac{1}{2}\epsilon_k$.

On the other hand, if $e = \frac{1}{2}\epsilon_{\epsilon_j} + \frac{1}{2}\epsilon_{\epsilon_k}$, under e the law of $\mathbf{x}$ conditional on $\mathbf{p}$ is a Dirac measure and thus $H(\mathbf{x}|\mathbf{p}) = 0$ whereas the law of $\mathbf{y}$ is the one of a fair coin and $H(\mathbf{y}) = 1$. Thus, $\mathbf{E}_\delta(\Delta H) = \Delta H(e) < 0$ and from thm. 6 δ is not achievable.

The intuition is as follows: if δ were achievable by P, only j and k would appear with positive density P-a.s. Since $f(j) \neq f(k)$, the statistician can reconstruct the history of the process given his signals, and therefore correctly guess $P(\mathbf{x}_n|x_1, \ldots, x_{n-1})$. This contradicts $e = \frac{1}{2}\epsilon_{\epsilon_j} + \frac{1}{2}\epsilon_{\epsilon_k}$ which means that at almost each stage, the statistician in uncertain about $P(\mathbf{x}_n|x_1, \ldots, x_{n-1})$ and attributes probability $\frac{1}{2}$ to ϵ_j and probability $\frac{1}{2}$ to ϵ_k.

3. Reduction of the problem

The core of our proof is to establish the next proposition.

Proposition 13. *Let $\delta = \lambda\epsilon_e + (1-\lambda)\epsilon_{e'}$ where λ is rational, e, e' have finite support and $\lambda\Delta H(e) + (1-\lambda)\Delta H(e') > 0$. Let $C = \operatorname{supp} e \cup \operatorname{supp} e'$. Then δ is achievable by the law of a C-process.*

Sections 4, 5, 6 and 7 are devoted to the proof of this proposition. We now prove theorems 6 and 9 from proposition 13.

3.1. The condition $\mathbf{E}_\delta\Delta H \geq 0$ is necessary.

We prove that any achievable δ must verify $\mathbf{E}_\delta\Delta H \geq 0$. Let δ be achieved by P. Recall that $\mathbf{e}_n$ is a $(\mathbf{y}_1, \ldots, \mathbf{y}_{n-1})$-measurable random variable in E. $\Delta H(\mathbf{e}_n)$ is thus a $(\mathbf{y}_1, \ldots, \mathbf{y}_{n-1})$-measurable real-valued random variable and:

$$\mathbf{E}_P\Delta H(\mathbf{e}_m) = H(\mathbf{p}_m, \mathbf{x}_m|\mathbf{y}_1, \ldots, \mathbf{y}_m) - H(\mathbf{p}_m|\mathbf{y}_1, \ldots, \mathbf{y}_{m-1}).$$

From the definitions:

$$\mathbf{E}_\delta(\Delta H) = \lim_n \frac{1}{n} \sum\nolimits_{m \leq n} \mathbf{E}_P\Delta H(\mathbf{e}_n)$$

We derive:

$$\begin{aligned}
\mathbf{E}_P \Delta H(\mathbf{e}_m) &= H(\mathbf{x}_1,\dots,\mathbf{x}_m|\mathbf{y}_1,\dots,\mathbf{y}_m) \\
&\quad - H(\mathbf{x}_1,\dots,\mathbf{x}_{m-1}|\mathbf{y}_1,\dots,\mathbf{y}_m,\mathbf{p}_m,\mathbf{x}_m) \\
&\quad - (H(\mathbf{x}_1,\dots,\mathbf{x}_{m-1}|\mathbf{y}_1,\dots,\mathbf{y}_{m-1}) \\
&\quad - H(\mathbf{x}_1,\dots,\mathbf{x}_{m-1}|\mathbf{y}_1,\dots,\mathbf{y}_{m-1},\mathbf{p}_m)) \\
&= H(\mathbf{x}_1,\dots,\mathbf{x}_m|\mathbf{y}_1,\dots,\mathbf{y}_m) - H(\mathbf{x}_1,\dots,\mathbf{x}_{m-1}|\mathbf{y}_1,\dots,\mathbf{y}_{m-1}) \\
&\quad + H(\mathbf{x}_1,\dots,\mathbf{x}_{m-1}|\mathbf{y}_1,\dots,\mathbf{y}_{m-1},\mathbf{p}_m) \\
&\quad - H(\mathbf{x}_1,\dots,\mathbf{x}_{m-1}|\mathbf{y}_1,\dots,\mathbf{y}_m,\mathbf{p}_m,\mathbf{x}_m) \\
&= H(\mathbf{x}_1,\dots,\mathbf{x}_m|\mathbf{y}_1,\dots,\mathbf{y}_m) - H(\mathbf{x}_1,\dots,\mathbf{x}_{m-1}|\mathbf{y}_1,\dots,\mathbf{y}_{m-1})
\end{aligned}$$

The first equality uses the additivity of entropies and the fact that $\mathbf{p}_m$ is $(\mathbf{x}_1,\dots,\mathbf{x}_{m-1})$-measurable and the second is a reordering of the first. The third equality uses again that $\mathbf{p}_m$ is $(\mathbf{x}_1,\dots,\mathbf{x}_{m-1})$-measurable and thus, conditionally on $\mathbf{p}_m$, $\mathbf{x}_m$ is independent of $(\mathbf{x}_1,\dots,\mathbf{x}_{m-1})$. It follows that:

$$\sum\nolimits_{m\le n} \mathbf{E}_P \Delta H(\mathbf{e}_m) = H(\mathbf{x}_1,\dots,\mathbf{x}_n|\mathbf{y}_1,\dots,\mathbf{y}_n) \ge 0$$

which gives the result.

3.2. ***C*-perfect observation.** In order to prove that $\mathbf{E}_\delta \Delta H \ge 0$ is a sufficient condition for δ to be achievable, we first need to study the case of perfect observation in details.

Definition 14. *Let C be a closed subset of* Π. *The mapping f is C-*perfect *if for each p in C, f is one-to-one on* $\operatorname{supp} p$.

We let $E_{C,d} = \{\epsilon_p, p \in C\}$ be the set of Dirac experiments with support in C. $E_{C,d}$ is a weak-$*$ closed subset of E and $\{\delta \in D, \operatorname{supp}\delta \subset E_{C,d}\}$ is a weak-$*$ closed and convex subset of D.

Lemma 15. *If f is C-perfect then:*

(1) *The experiment distribution δ is achievable the law of a C-process if and only if* $\operatorname{supp}\delta \subset E_{C,d}$

(2) *For each δ such that* $\operatorname{supp}\delta \subset E_{C,d}$, $\mathbf{E}_\delta(\Delta H) = 0$.

Proof. Point (1) Let $(\mathbf{x}_n)$ be a C- process P, δ achieved by P and p_1 be the law of $\mathbf{x}_1$. Since f is one-to-one on $\operatorname{supp} p_1$, the experiment $\mathbf{e}_2(y_1)$ is the Dirac measure on $p_2 = P(\mathbf{x}_2|x_1)$. By induction, assume that the experiment $\mathbf{e}_n(y_1, \ldots, y_{n-1})$ is the Dirac measure on $p_n = P(\mathbf{x}_n|x_1, \ldots, x_{n-1})$. Since f is one-to-one on $\operatorname{supp} p_n$, y_n reveals the value of x_n and $\mathbf{e}_{n+1}(y_1, \ldots, y_n)$ is the Dirac measure on $P(\mathbf{x}_n|x_1, \ldots, x_n)$. We get that under P, at each stage the experiment belong to $E_{C,d}$ P-a.s. and thus $\operatorname{supp}\delta \subset E_{C,d}$.

Conversely let δ be such that $\operatorname{supp}\delta \subset E_{C,d}$. Since the set of achievable distribution is closed, it is sufficient to prove that for any $p_1, \ldots, p_k$ in C, $n_1, \ldots, n_k$ integers, $n = \sum_j n_j$, $\delta = \sum_j \frac{n_j}{n}\epsilon_{e_j}$ is feasible where $e_j = \epsilon_{p_j}$. But then, $P_n = p_1^{\otimes n_1} p_2^{\otimes n_2} \cdots p_k^{\otimes n_k}$ n-achieves δ.

Point (2) If $e \in E_{C,d}$, the random variable $\mathbf{p}$ associated to e is constant a.s., therefore $H(\mathbf{x}|\mathbf{p}) = H(\mathbf{x}) = H(\mathbf{y})$ since f is C-perfect. Thus $\Delta H(e) = 0$, and $\mathbf{E}_\delta(\Delta H) = 0$ if $\operatorname{supp}\delta \subset E_{C,d}$.

□

3.3. **The condition $\mathbf{E}_\delta \Delta H \geq 0$ is sufficient.** According to proposition 13, any $\delta = \lambda\epsilon_e + (1-\lambda)\epsilon_{e'}$ with λ rational, e, e' of finite support and such that $\lambda\Delta H(e) + (1-\lambda)\Delta H(e') > 0$ is achievable by the law of a C-process with $C = \operatorname{supp} e \cup \operatorname{supp} e'$. We apply this result to prove theorem 9. Theorem 6 then follows using $C = \Pi$.

Proof of thm. 9 from prop. 13. Let $C \subset \Pi$ be closed, $E_C \subset E$ be the distributions with support in C, and $D_C \subset D$ be the distributions with support in E_C. Take $\delta \in D_C$ such that $\mathbf{E}_\delta(\Delta H) \geq 0$.

Assume first that $\mathbf{E}_\delta(\Delta H) = 0$ and that there exists a weak-$*$ neighborhood V of δ in D_C such that for any $\mu \in V$, $\mathbf{E}_\mu(\Delta H) \leq 0$. For $p \in C$, let

$\nu = \epsilon_{\epsilon_p}$. There exists $0 < t < 1$ such that $(1-t)\delta + t\nu \in V$ and therefore $\mathbf{E}_\nu(\Delta H) \leq 0$. Taking $\mathbf{x}$ of law p and $\mathbf{y} = f(\mathbf{x})$, $\mathbf{E}_\nu(\Delta H) = \Delta H(e) = H(\mathbf{x}) - H(\mathbf{y}) \leq 0$. Since $H(\mathbf{x}) \geq H(f(\mathbf{x}))$, we obtain $H(\mathbf{x}) = H(f(\mathbf{x}))$ for each $\mathbf{x}$ of law $p \in C$. This implies that f is C-perfect and the theorem holds by lemma 15.

Otherwise there is a sequence δ_n in D_C weak-$*$ converging to δ such that $\mathbf{E}_{\delta_n}(\Delta H) > 0$. Since the set of achievable distributions is closed, we assume $\mathbf{E}_\delta(\Delta H) > 0$ from now on. The set of distribution with finite support being dense in D_C (see e.g. [Par67] thm. 6.3 p. 44), again by closedness we assume:

$$\delta = \sum_j \lambda_j \epsilon_{e_j}$$

with $e_j \in E_C$ for each j. Let S be the finite set of distributions $\left\{\epsilon_{e_j}; j\right\}$. We claim that δ can be written as a convex combination of distributions δ_k such that:

- For each k, $\mathbf{E}_\delta(\Delta H) = \mathbf{E}_{\delta_k}(\Delta H)$.
- For each k, δ_k is the convex combination of two points in S.

This follows from the following lemma of convex analysis:

Lemma 16. *Let S be a finite set in a vector space and f be a real-valued affine mapping on $\operatorname{co} S$ the convex hull of S. For each $x \in \operatorname{co} S$, there exists an integer K, non-negative numbers $\lambda_1, \dots, \lambda_K$ summing to one, coefficients $t_1, \dots, t_K$ in $[0,1]$, and points (x_k, x'_k) in S such that:*

- $x = \sum_k \lambda_k (t_k x_k + (1-t_k) x'_k)$.
- *For each k, $t_k f(x_k) + (1-t_k) f(x'_k) = f(x)$.*

Proof. Let $a = f(x)$. The set $S_a = \{y \in \operatorname{co} S, f(y) = a\}$ is the intersection of a polytope with a hyperplane. It is thus convex and compact so by Krein-Milman's theorem (see e.g. [Roc70]) it is the convex hull of its extreme points. An extreme point y of S_a – i.e. a face of dimension 0 of S_a – must lie on a face of $\operatorname{co} S$ of dimension at most 1 and therefore is the convex combination of two points of S. □

We apply lemma 16 to $S = \left\{\epsilon_{e_j}; j\right\}$ and to the affine mapping $\delta \mapsto \mathbf{E}_\delta(\Delta H)$. Since the set of achievable distributions is convex it is enough to prove that for each k, δ_k is achievable. The problem is thus reduced to $\delta = \lambda\epsilon_e + (1-\lambda)\epsilon_{e'}$ such that $\lambda\Delta H(e) + (1-\lambda)\Delta H(e') > 0$. We approximate λ by a rational number and since C is closed, we may assume that the supports of e and e' are finite subsets of C. Proposition 13 now applies. □

4. Presentation of the proof of proposition 13

We want to prove that any $\delta = \lambda\epsilon_e + (1-\lambda)\epsilon_{e'}$, with $\lambda\Delta H(e) + (1-\lambda)\Delta H(e') > 0$ and λ rational is achievable. We construct a process P such that the induced experiment is close to e and to e' in proportions λ and $1-\lambda$ of the time. How can we design a process such that between periods T and $T+n$, the experiments are close to e?

A first idea is to define $\mathbf{x}_{T+1}$ up to $\mathbf{x}_{T+n}$ as follows: draw independently of the past $\mathbf{p}_T$ up to $\mathbf{p}_{T+n}$ i.i.d. according to e, and then draw independently $\mathbf{x}_t$ according to $\mathbf{p}_t$ for $T+1 \leq t \leq T+n$. This simple construction is not adequate since the induced experiment in stages $T+1 \leq t \leq T+n$ is the unit mass on $\mathbf{E}_e(p)$, and is different from e as soon as e is not a Dirac measure. We need to construct the process in such a way that a perfect observer knows which is the distribution p_t of $\mathbf{x}_t$ conditional to the past, but the statistician only knows it may be p with probability $e(p)$.

We thus amend the previous construction in order to take into account the information gap between a perfect observer of the process and the statistician before stage T. When the realized sequence of signals to the statistician up to stage T is $\tilde{y}_T = (y_1, \ldots, y_T)$, this information gap can be measured by the conditional probability $\mu(\tilde{y}_T) = P(\mathbf{x}_1, \ldots, \mathbf{x}_T | \tilde{y}_T)$.

Assume that the distribution $\mu(\tilde{y}_T)$ is close to that of n i.i.d. random variables and has entropy nh with $h > H(e)$. We explicit in this case a mapping

φ from X^T to Π^n such that the image distribution of $\mu(\tilde{y}_T)$ by φ is close to $e^{\otimes n}$.

We construct then the process at stages $T+1$ up to $T+n$ as follows:

Let $(\mathbf{p}_{T+1}, \ldots, \mathbf{p}_{T+n})$ be the image of $(\mathbf{x}_1, \ldots, \mathbf{x}_T)$ by φ, draw $\mathbf{x}_t$ for $T \leq t \leq T+n$ according to the realization of $\mathbf{p}_t$ and independently of the rest.

The realized sequence of experiments $e_{T+1}, \ldots, e_{T+n}$ is then close to e repeated n times, since the statistician does not know the realized value of p_t, whereas the perfect observer does.

Our construction of the process P mostly relies on the above idea. In order to formalize it, we need to define the notions of closeness that are useful for our purposes (closeness between $\mu(\tilde{y}_T)$ and the uniform distribution, closeness between the realized sequence of experiments and e). Once we have defined the conditions on $\mu(\tilde{y}_T)$ that allow us to construct the process for stages $T \leq t \leq T+n$ with $n = \lambda N$, we need to check that, with large enough probability, the construction can be applied once more for the block $T+n+1 \leq t \leq T+n+m$ with $m = (1-\lambda)N$. To do this, we prove that with high enough probability, $\mu(\tilde{y}_T)$ is close that of m i.i.d. random variables and has total entropy $n(h + \Delta H(e))$.

Section 5 presents the construction of the process for one block of stages, and establishes the necessary bounds on closeness of probabilities. In section 6, we iterate the construction, and show the full construction of the process P, including after sequences $\mu(\tilde{y}_T)$ for which the construction of section 5 fails. We terminate the proof by proving the weak-$*$ convergence of the sequence of experiments to $\lambda e + (1-\lambda)e'$ in section 7. In this last part, we first control the Kullback distance between the law of the process of experiments under P and an ideal law $Q = e^{\otimes n} \otimes e'^{\otimes m} \otimes e^{\otimes n} \otimes e'^{\otimes m} \otimes \ldots$, and finally relate the Kullback distance to weak-$*$ convergence.

5. The one block construction

5.1. **Kullback and absolute Kullback distance.** For two probability measures with finite support P and Q, we write $P \ll Q$ when Q is absolutely continuous with respect to P i.e. $(Q(x) = 0 \Rightarrow P(x) = 0)$.

Definition 17. *Let K be a finite set and P, Q in $\Delta(K)$ such that $P \ll Q$, the Kullback distance between P and Q is,*

$$d(P||Q) = \mathbf{E}_P\left[\log \frac{P(\cdot)}{Q(\cdot)}\right] = \sum\nolimits_k P(k) \log \frac{P(k)}{Q(k)}$$

We recall the absolute Kullback distance and its comparison with the Kullback distance from [GV02] for later use.

Definition 18. *Let K be a finite set and P, Q in $\Delta(K)$ such that $P \ll Q$, the absolute Kullback distance between P and Q is,*

$$|d|\,(P||Q) = \mathbf{E}_P\left|\log \frac{P(\cdot)}{Q(\cdot)}\right|$$

Lemma 19. *For every P, Q in $\Delta(K)$ such that $P \ll Q$,*

$$d(P||Q) \leq |d|\,(P||Q) \leq d(P||Q) + 2$$

See the proof of lemma 17, p. 223 in [GV02].

5.2. **Equipartition properties.** We say than a probability P with finite support verifies an Equipartition Property (**EP** for short) when all points in the support of P have close probabilities.

Definition 20. *Let $P \in \Delta(K)$, $n \in \mathbb{N}$, $h \in \mathbb{R}_+$,$\eta > 0$. P verifies an* **EP**(n, h, η)*, when*

$$P\{k \in K, |-\frac{1}{n}\log P(k) - h| \leq \eta\} = 1$$

We say than a probability P with finite support verifies an Asymptotic Equipartition Property (**AEP** for short) when all points in a set of large

P-measure have close probabilities.

Definition 21. *Let $P \in \Delta(K)$, $n \in \mathbb{N}$, $h \in \mathbb{R}_+$, $\eta, \xi > 0$. P verifies an* $\mathbf{AEP}(n, h, \eta, \xi)$*, when*

$$P\{k \in K, |-\frac{1}{n}\log P(k) - h| \leq \eta\} \geq 1 - \xi$$

Remark 22. Assume that P verifies an $\mathbf{AEP}(n, h, \eta, \xi)$ and let m be an integer then, P verifies an $\mathbf{AEP}(m, \frac{n}{m}h, \frac{n}{m}\eta, \xi)$.

5.3. **Types.** Given a finite set K and in integer n, we denote $\tilde{k} = (k_1, \ldots, k_n) \in K^n$ a finite sequence in K. The type of $\tilde{k}$ is the empirical distribution $\rho_{\tilde{k}}$ induced by $\tilde{k}$ that is, $\rho_{\tilde{k}} \in \Delta(K)$ and $\forall k$, $\rho_{\tilde{k}}(k) = \frac{1}{n}|\{i = 1, \ldots, n, k_i = k\}|$. The type set $T_n(\rho)$ of $\rho \in \Delta(K)$ is the subset of K^n of sequences of type ρ. Finally, the set of types is $\mathbb{T}_n(K) = \{\rho \in \Delta(K), T_n(\rho) \neq \emptyset\}$. The following estimates the size of $T_n(\rho)$ for $\rho \in \mathbb{T}_n(K)$ (see e.g. Cover and Thomas [CT91] Theorem 12.1.3 page 282):

$$\frac{2^{nH(\rho)}}{(n+1)^{|\mathcal{K}|}} \leq |T_n(\rho)| \leq 2^{nH(\rho)} \tag{1}$$

5.4. **Distributions induced by experiments and by codifications.** Let $e \in \Delta(\Pi)$ be an experiment with finite support and n be an integer.

Notation 23. *Let $\rho(e)$ be the probability on $\Pi \times X$ induced by the following procedure: First draw $\mathbf{p}$ according to e, then draw $\mathbf{x}$ according to the realization of $\mathbf{p}$. Let $Q(n, e) = \rho(e)^{\otimes n}$.*

We need to approximate $Q(n, e)$ in a construction where $(\mathbf{p}_1, \ldots, \mathbf{p}_n)$ is measurable with respect to some random variable $\mathbf{l}$ of law $P_{\mathcal{L}}$ in an arbitrary set $\mathcal{L}$.

Notation 24. *Let $(\mathcal{L}, P_{\mathcal{L}})$ be a finite probability space and $\varphi \colon \mathcal{L} \to \Pi^n$. We denote by $P = P(n, \mathcal{L}, P_{\mathcal{L}}, \varphi)$ the probability on $\mathcal{L} \times (\Pi \times X)^n$ induced by the following procedure. Draw $\mathbf{l}$ according to $P_{\mathcal{L}}$, set $(\mathbf{p}_1, \ldots, \mathbf{p}_n) = \varphi(\mathbf{l})$ and then draw $\mathbf{x}_t$ according to the realization of $\mathbf{p}_t$.*

We let $\tilde{P} = \tilde{P}(n, \mathcal{L}, P_{\mathcal{L}}, \varphi)$ *be the marginal of* $P(n, \mathcal{L}, P_{\mathcal{L}}, \varphi)$ *on* $(\Pi \times X)^n$*.*

Another point we need to take care of is that such a construction can be iterated, by relating properties of the "input" probability measure $P_{\mathcal{L}}$ with those of the "output" probability measure $P(\mathbf{l}, \mathbf{p}_1, \dots, \mathbf{p}_n, \mathbf{x}_1, \dots, \mathbf{x}_n | \mathbf{y}_1, \dots, \mathbf{y}_n)$.

Propositions 25 and 26 exhibit conditions on $P_{\mathcal{L}}$ such that there exists φ for which $\tilde{P}(n, \mathcal{L}, P_{\mathcal{L}}, \varphi)$ is close to $Q(n, e)$ and, with large probability under $P = P(n, \mathcal{L}, P_{\mathcal{L}}, \varphi)$, $P(\mathbf{l}, \mathbf{p}_1, \dots, \mathbf{p}_n, \mathbf{x}_1, \dots, \mathbf{x}_n | \mathbf{y}_1, \dots, \mathbf{y}_n)$ verifies an adequate **AEP**.

In proposition 25, the condition on $P_{\mathcal{L}}$ is an **EP** property, thus a stronger input property than the output property which is stated as an **AEP**. We strengthen this result by assuming that $P_{\mathcal{L}}$ verifies an **AEP** property in proposition 26.

5.5. **EP to AEP codification result.** We now state and prove our coding proposition when the input probability measure $P_{\mathcal{L}}$ verifies an **EP**.

Proposition 25. *For each experiment* e*, there exists a constant* $U(e)$ *such that for every integer* n *with* $e \in \mathbb{T}_n(\Pi)$ *and for every finite probability space* $(\mathcal{L}, P_{\mathcal{L}})$ *that verifies an* $\mathbf{EP}(n, h, \eta)$ *with* $n(h - H(e) - \eta) \geq 1$*, there exists a mapping* $\varphi \colon \mathcal{L} \to \Pi^n$ *such that, letting* $P = P(n, \mathcal{L}, P_{\mathcal{L}}, \varphi)$ *and* $\tilde{P} = \tilde{P}(n, \mathcal{L}, P_{\mathcal{L}}, \varphi)$*:*

(1) $d(\tilde{P} || Q(n, e)) \leq 2n\eta + |\mathrm{supp}\, e| \log(n+1) + 1$
(2) *For every* $\varepsilon > 0$*, there exists a subset* $\mathcal{Y}_\varepsilon$ *of* Y^n *such that:*
  (a) $P(\mathcal{Y}_\varepsilon) \geq 1 - \varepsilon$
  (b) *For* $\tilde{y} \in \mathcal{Y}_\varepsilon$*,* $P(\cdot | \tilde{y})$ *verifies an* $\mathbf{AEP}(n, h', \eta', \varepsilon)$

with $h' = h + \Delta H(e)$ *and* $\eta' = \frac{U(e)}{\varepsilon^2}(\eta + \frac{\log(n+1)}{n})$*.*

Proof of prop. 25.

Set $\rho = \rho(e)$ and $\tilde{Q} = Q(n, e)$.

Construction of φ: Since $P_{\mathcal{L}}$ verifies an $\mathbf{EP}(n,h,\eta)$,

$$2^{n(h-\eta)} \leq |\operatorname{supp} P_{\mathcal{L}}| \leq 2^{n(h+\eta)}$$

From the previous and equation (1), there exists $\varphi\colon \mathcal{L} \to T_n(e)$ such that for every $\tilde{p} \in (\operatorname{supp} e)^n$,

$$2^{n(h-\eta-H(e))} - 1 \leq |\varphi^{-1}(\tilde{p})| \leq (n+1)^{|\operatorname{supp} e|} 2^{n(h+\eta-H(e))} + 1 \tag{2}$$

Bound on $d(\tilde{P}||\tilde{Q})$: $\tilde{P}$ and $\tilde{Q}$ are probabilities over $(\Pi \times X)^n$ which are deduced from their marginals on Π^n by the same transition probabilities. It follows from the definition of the Kullback distance that the distance from $\tilde{P}$ to $\tilde{Q}$ equals the distance of their marginals on Π^n:

$$d(\tilde{P}||\tilde{Q}) = \sum\nolimits_{\tilde{p} \in T_n(e)} \tilde{P}(\tilde{p}) \log \frac{\tilde{P}(\tilde{p})}{\tilde{Q}(\tilde{p})}$$

Using equation (2) and the **EP** for $P_{\mathcal{L}}$, we obtain that for $\tilde{p} \in T_n(e)$:

$$\tilde{P}(\tilde{p}) \leq (n+1)^{|\operatorname{supp} e|} 2^{n(2\eta - H(e))} + 2^{-n(h-\eta)}$$

On the other hand, since for all $\tilde{p} \in T_n(e)$, $\tilde{Q}(\tilde{p}) = 2^{-nH(e)}$:

$$\frac{\tilde{P}(\tilde{p})}{\tilde{Q}(\tilde{p})} \leq (n+1)^{|\operatorname{supp} e|} 2^{2n\eta} + 2^{-n(h-\eta-H(e))}$$

Part (1) of the proposition now follows since $H(e) \leq h - \eta$.

Estimation of $|d|(\tilde{P}(\cdot|\tilde{y})||\tilde{Q}(\cdot|\tilde{y}))$: For $\tilde{y} \in Y^n$ s.t. $\tilde{P}(\tilde{y}) > 0$, we let $\tilde{P}_{\tilde{y}}$ and $\tilde{Q}_{\tilde{y}}$ in $\Delta((\Pi \times X)^n)$ denote $\tilde{P}(\cdot|\tilde{y})$ and $\tilde{Q}(\cdot|\tilde{y})$ respectively. Direct computation yields:

$$\sum\nolimits_{\tilde{y}} \tilde{P}(\tilde{y}) d(\tilde{P}_{\tilde{y}}||\tilde{Q}_{\tilde{y}}) = d(\tilde{P}||\tilde{Q})$$

Hence for $\alpha_1 > 0$:

$$P\left\{\tilde{y}, d(\tilde{P}_{\tilde{y}}||\tilde{Q}_{\tilde{y}}) \geq \alpha_1\right\} \leq \frac{2n\eta + |\operatorname{supp} e| \log(n+1) + 1}{\alpha_1}$$

and from lemma 19,

$$P\left\{\tilde{y}, |d|(\tilde{P}_{\tilde{y}}||\tilde{Q}_{\tilde{y}}) \leq \alpha_1 + 2\right\} \geq 1 - \frac{2n\eta + |\operatorname{supp} e|\log(n+1) + 1}{\alpha_1} \tag{3}$$

The statistics of $(\tilde{k}, \tilde{s})$ under $\tilde{P}$: We write that the type $\rho_{\tilde{p},\tilde{x}} \in \Delta(\Pi \times X)$ of $(\tilde{p}, \tilde{x}) \in (\Pi \times X)^n$ is close to ρ, with large P-probability. First, note that since φ takes its values in $T_n(e)$, the marginal of $\rho_{\tilde{p},\tilde{x}}$ on Π is e with P-probability one. For $(p, x) \in \Pi \times X$, the distribution under P of $n\rho_{\tilde{p},\tilde{x}}(p, x)$ is the one of a sum of $ne(p)$ independent Bernoulli variables with parameter $p(x)$. For $\alpha_2 > 0$ the Bienaymé-Chebyshev inequality gives:

$$P(|\rho_{\tilde{p},\tilde{x}}(p, x) - \rho(p, x)| \geq \alpha_2) \leq \frac{\rho(p, x)}{n\alpha_2^2}$$

Hence,

$$P\left(\|\rho_{\tilde{p},\tilde{x}} - \rho\|_\infty \leq \alpha_2\right) \geq 1 - \frac{1}{n\alpha_2^2} \tag{4}$$

The set of $\tilde{y} \in Y^n$ s.t. $\tilde{Q}_{\tilde{y}}$ verifies an AEP has large P-probability: For $(\tilde{p}, \tilde{x}, \tilde{y}) = (p_i, x_i, y_i)_i \in (\Pi \times X \times Y)^n$ s.t. $\forall i,\ f(x_i) = y_i$, we compute:

$$\begin{aligned}
-\frac{1}{n}\log Q_{\tilde{y}}(\tilde{p}, \tilde{x}) &= -\frac{1}{n}(\sum\nolimits_i \log\rho(p_i, x_i) - \log\rho(y_i)) \\
&= -\sum\nolimits_{(p,x)\in(\operatorname{supp} e)\times X} \rho_{\tilde{p},\tilde{x}}(p, x)\log\rho(p, x) \\
&\quad + \sum\nolimits_{y\in Y} \rho_{\tilde{p},\tilde{x}}(y)\log\rho_{\tilde{p},\tilde{x}}(y) \\
&= -\sum\nolimits_{(p,x)} \rho(p, x)\log\rho(p, x) + \sum\nolimits_y \rho(y)\log\rho(y) \\
&\quad + \sum\nolimits_{(p,x)}(\rho(p, x) - \rho_{\tilde{p},\tilde{x}}(p, x))\log\rho(p, x) \\
&\quad - \sum\nolimits_y(\rho(y) - \rho_{\tilde{p},\tilde{x}}(y))\log\rho(y)
\end{aligned}$$

Since

$$-\sum\nolimits_{(p,x)} \rho(p,x) \log \rho(p,x) = H(\rho)$$

and, denoting $f(\rho)$ the image of ρ on Y:

$$\sum\nolimits_{y} \rho(y) \log \rho(y) = -H(f(\rho))$$

letting $M_0 = -2|(\operatorname{supp} e) \times X| \log(\min_{p,x} \rho(p,x))$, this implies:

$$|-\frac{1}{n} \log \tilde{Q}_{\tilde{y}}(\tilde{p},\tilde{x}) - H(\rho) + H(f(\rho))| \leq M_0 \|\rho - \rho_{\tilde{p},\tilde{x}}\|_\infty \tag{5}$$

Define:

$$\begin{aligned} A_{\alpha_2} &= \{(\tilde{p},\tilde{x},\tilde{y}), |-\frac{1}{n} \log \tilde{Q}_{\tilde{y}}(\tilde{p},\tilde{x}) - H(\rho) + H(f(\rho))| \leq M_0 \alpha_2\} \\ A_{\alpha_2,\tilde{y}} &= A_{\alpha_2} \cap ((\operatorname{supp} e) \times X \times \{\tilde{y}\}),\ \tilde{y} \in Y^n \end{aligned}$$

Equations (4) and (5) yield:

$$\begin{aligned} \sum\nolimits_{\tilde{y}} P(\tilde{y}) \tilde{P}_{\tilde{y}}(A_{\alpha_2,\tilde{y}}) &= P(A_{\alpha_2}) \\ &\geq 1 - \frac{1}{n\alpha_2^2} \end{aligned}$$

Thus, for $\beta > 0$,

$$P\left\{\tilde{y}, \tilde{P}_{\tilde{y}}(A_{\alpha_2,\tilde{y}}) \leq 1-\beta\right\} \leq 1 - \frac{1}{n\alpha_2^2\beta} \tag{6}$$

Definition of $\mathcal{Y}_\varepsilon$ and verification of (2a): Set

$$\begin{cases} \alpha_1 &= \frac{4n\eta + 2|\operatorname{supp} e| \log(n+1) + 2}{\varepsilon} \\ \alpha_2 &= \frac{1}{4n\varepsilon^2} \\ \beta &= \frac{\varepsilon}{2} \end{cases}$$

and let:

$$\begin{cases} \mathcal{Y}^1_\varepsilon &= \left\{\tilde{y}, |d|(\tilde{P}_{\tilde{y}}||\tilde{Q}_{\tilde{y}}) \leq \alpha_1 + 2\right\} \\ \mathcal{Y}^2_\varepsilon &= \left\{\tilde{y}, \tilde{P}_{\tilde{y}}(A_{\alpha_2,\tilde{y}}) \leq 1 - \beta\right\} \\ \mathcal{Y}_\varepsilon &= \mathcal{Y}^1_\varepsilon \cap \mathcal{Y}^2_\varepsilon \end{cases}$$

Equations (3) and (6) imply

$$P(\mathcal{Y}_\varepsilon) \geq 1 - \varepsilon$$

Verification of (2b): We first prove that $\tilde{P}_{\tilde{y}}$ verifies an **AEP** for $\tilde{y} \in \mathcal{Y}_\varepsilon$. For such $\tilde{y}$, the definition of $\mathcal{Y}^1_\varepsilon$ and equation (3) imply:

$$\tilde{P}_{\tilde{y}}\left\{|\log \tilde{P}_{\tilde{y}}(\cdot) - \log \tilde{Q}_{\tilde{y}}(\cdot)| \leq \frac{2(\alpha_1+2)}{\varepsilon}\right\} \geq 1 - \frac{\varepsilon}{2}$$

From the definition of $\mathcal{Y}^2_\varepsilon$:

$$\tilde{P}_{\tilde{y}}\left\{|-\frac{1}{n}\log \tilde{Q}_{\tilde{y}}(\cdot) - H(\rho) + H(f(\rho))| \leq M_0\alpha_2\right\} \geq 1 - \frac{\varepsilon}{2}$$

The two above inequalities yield:

$$\tilde{P}_{\tilde{y}}\left\{|-\frac{1}{n}\log \tilde{P}_{\tilde{y}}(\cdot) - H(\rho) + H(f(\rho))| \leq \frac{2(\alpha_1+2)}{n\varepsilon} + M_0\alpha_2\right\} \geq 1 - \varepsilon \tag{7}$$

Remark now that $P(l, \tilde{p}, \tilde{x}|y) = P_y(\tilde{p}, \tilde{x})P(l|\tilde{p})$. If $\varphi(l) \neq \tilde{p}$, $P(l|\tilde{p}) = 0$ and otherwise, equation (2) and the **EP** for $P_{\mathcal{L}}$ imply:

$$\frac{2^{n(h-\eta)}}{2^{n(h+\eta)}((n+1)^{|\mathrm{supp}\, e|}2^{n(h-\eta-H(e))}+1)} \leq \frac{P(l)}{P(\tilde{p})} \leq \frac{2^{n(h+\eta)}}{2^{n(h-\eta)}(2^{n(h-\eta-H(e))}-1)}$$

From this we deduce, using $n(h - \eta - H(e)) \geq 1$:

$$|\log P(l) - \log P(\tilde{p}) - n(H(e) - h)| \leq 3n\eta + \log(n+1) + 1 \tag{8}$$

Let $P_{\tilde{y}}$ denote $P(\cdot|\tilde{y})$ over $\mathcal{L} \times (\Pi \times X)^n$. Using that $\Delta H(e) = H(\rho) - H(e) -$

$H(f(\rho))$, and setting $U(e) = \max(11, 4 + M_0)$, equations (7) and (8) yield:

$$P_{\tilde{y}}\left\{|-\frac{1}{n}\log P_{\tilde{y}}(\cdot) - (h + \Delta H(e))| \leq \frac{U(e)}{\varepsilon^2}(\eta + \frac{\log(n+1)}{n})\right\} \geq 1 - \frac{\varepsilon}{2}$$

which is the desired **AEP**. □

5.6. **AEP to AEP codification result.** Building on proposition 25, we now can state and prove the version of our coding result in which the input is an **AEP**.

Proposition 26. *For each experiment e, there exists a constant $U(e)$ such that for every integer n with $e \in \mathbb{T}_n(\Pi)$ and for every finite probability space $(\mathcal{L}, P_{\mathcal{L}})$ that verifies an* $\mathbf{AEP}(n, h, \eta, \xi)$ *with $n(h - H(e) - \eta) \geq 2$ and $\xi < \frac{1}{16}$, there exists a mapping $\varphi\colon \mathcal{L} \to \Pi^n$ such that, letting $P = P(n, \mathcal{L}, P_{\mathcal{L}}, \varphi)$ and $\tilde{P} = \tilde{P}(n, \mathcal{L}, P_{\mathcal{L}}, \varphi)$:*

1. $d(\tilde{P}||Q(n,e)) \leq 2n(\eta + \xi \log|\mathrm{supp}\, e|) + |\mathrm{supp}\, e| \log(n+1) + 1$
2. *For every $\varepsilon > 0$, there exists a subset $\mathcal{Y}_\varepsilon$ of Y^n such that:*
   (a) $P(\mathcal{Y}_\varepsilon) \geq 1 - \varepsilon - 2\sqrt{\xi}$
   (b) *For $\tilde{y} \in \mathcal{Y}_\varepsilon$, $P(\cdot|\tilde{y})$ verifies an* $\mathbf{AEP}(n, h', \eta', \xi')$

with $h' = h + \Delta H(e)$, $\eta' = \frac{U(e)}{\varepsilon^2}(\eta + \frac{\log(n+1)}{n}) + 4\frac{\sqrt{\xi}}{n}$ and $\xi' = \varepsilon + 3\sqrt{\xi}$.

We first establish the following lemma.

Lemma 27. *K is a finite set. Suppose that $P \in \Delta(K)$ verifies an* $\mathbf{AEP}(n, h, \eta, \xi)$*. Let the typical set of P be:*

$$C = \{k \in K, |-\frac{1}{n}\log P(k) - h| \leq \eta\}$$

Let $P_C \in \Delta(K)$ be the conditional probability given C: $P_C(k) = P(k|C)$. Then, P_C verifies an $\mathbf{EP}(n, h, \eta')$ *with $\eta' = \eta + 2\frac{\xi}{n}$ for $0 < \xi < \frac{1}{2}$.*

Proof. Follows immediately, since for $0 < \xi \leq \frac{1}{2}$, $-\log(1 - \xi) \leq 2\xi$. □

Proof of prop. 26. Set again $\rho = \rho(e)$ and $\tilde{Q} = Q(n, e)$. Let C be the typical set of $P_{\mathcal{L}}$. From lemma 27, $P'_{\mathcal{L}} = P_{\mathcal{L}}(\cdot|C)$ verifies an $\mathbf{EP}(n, h, \eta + 2\frac{\xi}{n})$. Since

$n(h - H(e) - \eta) \geq 2$, $n(h - H(e) - \eta - 2\frac{\xi}{n}) \geq 1$. Applying prop. 25 to e yields: a constant $U(e)$, a mapping $\varphi\colon C \to \Pi^n$, an induced probability P' on $\mathcal{L} \times (\Pi \times X)^n$, and subsets $(\mathcal{Y}'_\varepsilon)_\varepsilon$ of Y^n.

Choose $\bar{p} \in \arg\max e(p)$ and extend φ to $\mathcal{L}$ by setting it to $(\bar{p}, \ldots, \bar{p})$ outside C. With $P'' = P_{\mathcal{L}}(\cdot|C) \otimes (\epsilon_{\bar{p}} \otimes \bar{p})^{\otimes n}$, the probability induced by $P_{\mathcal{L}}$ and φ on $\mathcal{L} \times (\Pi \times X)^n$ is then $P = P_{\mathcal{L}}(C)P' + (1 - P_{\mathcal{L}}(C))P''$. Set $\tilde{P}$ as the marginal of P on $(\Pi \times X)^n$. To verify point (1), write:

$$\begin{aligned} d(\tilde{P}\|\tilde{Q}) &\leq P_{\mathcal{L}}(C)d(\tilde{P}'\|\tilde{Q}) + (1 - P_{\mathcal{L}}(C))nd(\epsilon_{\bar{p}} \otimes \bar{p}\|\rho) \\ &\leq d(\tilde{P}'\|\tilde{Q}) + \xi nd(\epsilon_{\bar{p}}\|e) \\ &\leq d(\tilde{P}'\|\tilde{Q}) + \xi n \log(|\operatorname{supp} e|) \end{aligned}$$

With $\mathcal{Y} = \{\tilde{y},\, P'(\tilde{y}) > \sqrt{\xi}P''(\tilde{y})\}$, $P'(\mathcal{Y}) \geq 1 - \sqrt{\xi}$ and then $P(\mathcal{Y}) \geq 1 - \xi - \sqrt{\xi}$. Let now $\mathcal{Y}_\varepsilon = \mathcal{Y}'_\varepsilon \cap \mathcal{Y}$. Point (2a) of the proposition is straightforward. To prove (2b), for $\tilde{y} \in \mathcal{Y}_\varepsilon$, let $C(\tilde{y})$ be the $(n, h', \frac{U(e)}{\varepsilon^2}(\eta + \frac{\log(n+1)}{n}))$ typical set of $P(\cdot|\tilde{y})$, and $A(\tilde{y}) = \{(l, \tilde{x}),\, P'(l, \tilde{x} \mid \tilde{y}) > \sqrt{\xi}P''(l, \tilde{x} \mid \tilde{y})\}$. Then $P'(A(\tilde{y})) \geq \sqrt{\xi}P''(A(\tilde{y}))$ and:

$$\begin{aligned} P(C(\tilde{y}) \cap A(\tilde{y}) \mid \tilde{y}) &= \frac{(P_{\mathcal{L}}(C)P' + (1 - P_{\mathcal{L}}(C))P'')(C(\tilde{y}) \cap A(\tilde{y}))}{(P_{\mathcal{L}}(C)P' + (1 - P_{\mathcal{L}}(C))P'')(\tilde{y})} \\ &\geq \frac{(1 - \xi)P'(C(\tilde{y}) \cap A(\tilde{y}))}{(\sqrt{\xi} + 1)P'(\tilde{y})} \\ &\geq (1 - 2\sqrt{\xi})P'(C(\tilde{y}) \cap A(\tilde{y}) \mid \tilde{y}) \\ &\geq 1 - 3\sqrt{\xi} - \varepsilon \end{aligned}$$

For $\tilde{y} \in \mathcal{Y}_\varepsilon$ and $(l, \tilde{x}) \in C(\tilde{y}) \cap A(\tilde{y})$, a similar computation shows that

$$|\log P(l, \tilde{x} \mid \tilde{y}) - \log P'(l, \tilde{x} \mid \tilde{y})| \leq -\log(1 - 2\sqrt{\xi}) \leq 4\sqrt{\xi}$$

Hence the result from (2b) of prop. 26. □

6. Construction of the process

Taking up the proof of proposition 13, let λ rational, e, e' having finite support be such that $\lambda\Delta H(e)+(1-\lambda)\Delta H(e') > 0$ and $C = \operatorname{supp} e \cup \operatorname{supp} e'$. We wish to construct a law P of a C-process that achieves $\delta = \lambda\epsilon_e + (1-\lambda)\epsilon_{e'}$. Again by closedness of the set of achievable distributions, we assume w.l.o.g. $e \in \mathbb{T}_{n_0}(C), e' \in \mathbb{T}_{n_0}(C)$ for some common n_0, $0 < \lambda < 1$ and $\lambda = \frac{M}{M+N}$ with M, N multiples of n_0.

Since $\lambda\Delta H(e) + (1-\lambda)\Delta H(e') > 0$, we assume w.l.o.g. $\Delta H(e) > 0$. Remark that for each $p \in \operatorname{supp} e$, $\Delta H(\epsilon_p) = H(\mathbf{x}|p) - H(\mathbf{y}|p)$, thus:

$$\mathbf{E}_e(\Delta H(\epsilon_p)) = H(\mathbf{x}|\mathbf{p}) - H(\mathbf{y}|\mathbf{p}) \geq H(\mathbf{x}|\mathbf{p}) - H(\mathbf{y}) = \Delta H(e) > 0.$$

Therefore, there exists $p_0 \in \operatorname{supp} e$ such that $\Delta H(\epsilon_{p_0}) > 0$ and we assume w.l.o.g. $\operatorname{supp} e' \ni p_0$. Hence $\max\{d(\epsilon_{p_0}\|e), d(\epsilon_{p_0}\|e')\}$ is well defined and finite.

We construct the process by blocks. For a block lasting from stage $T+1$ up to stage $T+M$ (resp. $T+N$), we construct $(\mathbf{x}_1, \ldots, \mathbf{x}_T)$-measurable random variables $\mathbf{p}_{T+1}, \ldots, \mathbf{p}_{T+M}$ such that their distribution conditional to $\mathbf{y}_1, \ldots, \mathbf{y}_T$ is close to that of M (resp. N) i.i.d. random variables of law e (resp. e'). We then take $\mathbf{x}_{T+1} \ldots, \mathbf{x}_{T+M}$ of law $\mathbf{p}_{T+1}, \ldots, \mathbf{p}_{T+M}$, and independent of the past of the process conditional to $\mathbf{p}_{T+1}, \ldots, \mathbf{p}_{T+M}$.

We define the process $(\mathbf{x}_t)_t$ and its law P over $\bar{N} = N_0 + L(M+N)$ stages, where (M, N) are multiples of (m, n), inductively over blocks of stages.

Definition of the blocks The first block labelled 0 is an initialization phase that lasts from stage 1 to N_0. For $1 \leq k \leq L$, the $2k$-th [resp. $2k+1$-th] block consists of stages $N_0+(k-1)(M+N)+1$ to $N_0+(k-1)(M+N)+M$ [resp. $N_0+(k-1)(M+N)+M+1$ to $N_0+k(M+N)$].

Initialization block During the initialization phase, $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N_0}$ are i.i.d. with law p_0, inducing a law P_0 of the process during this block.

First block Let S_0 be the set of $\tilde{y}_0 \in Y^{N_0}$ such that $P_0(\cdot|\tilde{y}_0)$ verifies an $\mathbf{AEP}(M, h_0, \eta_0, \xi_0)$. After histories in S_0 and for suitable values of the parameters h_0, η_0, ξ_0, proposition 26 allows to define random variables $\mathbf{p}_{N_0+1}, \ldots, \mathbf{p}_{N_0+M}$ such that their distribution conditional to $\mathbf{y}_1, \ldots, \mathbf{y}_{N_0}$ is close to that of M i.i.d. random variables of law e. We then take $\mathbf{x}_{N_0+1}, \ldots, \mathbf{x}_{N_0+M}$ of law $\mathbf{p}_{N_0+1}, \ldots, \mathbf{p}_{N_0+M}$, and independent of the past of the process conditional to $\mathbf{p}_{N_0+1}, \ldots, \mathbf{p}_{N_0+M}$. We let $\mathbf{x}_t$ be i.i.d. with law p_0 after histories not in S_0. This defines the law of the process up to the first block.

Second block Let $\tilde{y}_1$ be an history of signals to the statistician during the initialization block and the first block. Proposition 26 ensures that, given $\tilde{y}_0 \in S_0$, $P_1(\cdot|\tilde{y}_1)$ verifies an $\mathbf{AEP}(N, h_1, \eta_1, \xi_1)$ with probability no less that $1-\varepsilon$, where $h_1 = \frac{\lambda}{1-\lambda}(h_0 + \Delta H(e))$, $\eta_1 = \frac{\lambda}{1-\lambda}(\frac{U(e)}{\varepsilon^2}(\eta_0 + \frac{\log(M+1)}{M}) + 4\frac{\sqrt{\xi_0}}{M})$, $\xi_1 = \varepsilon + 3\sqrt{\xi_0}$. Thus, the set S_1 of $\tilde{y}_1$ such that $P_1(\cdot|\tilde{y}_1)$ verifies an $\mathbf{AEP}(N, h_1, \eta_1, \xi_1)$ has probability no less than $1 - 2\varepsilon$.

Inductive construction We define inductively the laws P_k for the process up to block k, and parameters h_k, η_k, ξ_k. We set $N_k = M$ if k odd and $N_k = N$ if k even. Let S_k be the set of histories $\tilde{y}_k$ for the statistician up to block k such that $P_k(\cdot|\tilde{y}_k)$ verifies an $\mathbf{AEP}(N_k, h_k, \eta_k, \xi_k)$. After $\tilde{y}_k \in S_k$, define the process during block k in order to approximate e i.i.d. if k is odd, and e' i.i.d. if k is even. After $\tilde{y}_k \notin S_k$, let the process during block k be i.i.d. with law p_0 conditional to the past. Proposition 26 ensures that, conditional on $\tilde{y}_k \in S_k$, $P_{k+1}(\cdot|\tilde{y}_{k+1})$ verifies an $\mathbf{AEP}(N_{k+1}, h_{k+1}, \eta_{k+1}, \xi_{k+1})$ with probability no less that $1-\varepsilon$, where h_{k+1}, η_{k+1} and ξ_{k+1} are given by the recursive relations:

$$\begin{cases} h_{k+1} &= \frac{\lambda}{1-\lambda}(h_k + \Delta H(e)) \\ \eta_{k+1} &= \frac{\lambda}{1-\lambda}(\frac{U(e)}{\varepsilon^2}(\eta_k + \frac{\log(M+1)}{M}) + 4\frac{\sqrt{\xi_k}}{M}) \\ \xi_{k+1} &= \varepsilon + 3\sqrt{\xi_k} \end{cases}$$

if k is even, and

$$\begin{cases} h_{k+1} &= \frac{1-\lambda}{\lambda}(h_k + \Delta H(e')) \\ \eta_{k+1} &= \frac{1-\lambda}{\lambda}(\frac{U(e')}{\varepsilon^2}(\eta_k + \frac{\log(N+1)}{N}) + 4\frac{\sqrt{\xi_k}}{N}) \\ \xi_{k+1} &= \varepsilon + 3\sqrt{\xi_k} \end{cases}$$

if k is odd.

The definition of the process for the $2L+1$ blocks is complete, provided for each k odd, $M(h_k - H(e) - \eta_k) \geq 2$, and for each k even, $N(h_k - H(e') - \eta_k) \geq 2$. We seek now conditions on $(\varepsilon, \eta_0, \xi_0, N_0, M, N, L)$ such that these inequalities are fulfilled. We first establish bounds on the sequences (ξ_k, η_k, h_k) and introduce some notations:

$$a(\varepsilon) = \frac{1}{\varepsilon^2}\max(\frac{\lambda}{1-\lambda}U(e); \frac{1-\lambda}{\lambda}U(e')) \tag{9}$$

$$c(\varepsilon, M, N) = \max(\frac{\lambda}{1-\lambda}\frac{U(e)}{\varepsilon^2}\frac{\log(M+1)}{M} + \frac{20}{M}; \frac{1-\lambda}{\lambda}\frac{U(e')}{\varepsilon^2}\frac{\log(N+1)}{N} + \frac{20}{N}) \tag{10}$$

Lemma 28. *For $k = 1, \dots, 2L$:*

(1) $\xi_k \leq \xi_{\max} = 11((\varepsilon)^{2^{-2L}} + (\xi_0)^{2^{-2L}})$.
(2) $\eta_k \leq \eta_{\max} = (a(\varepsilon))^{2L}(\eta_0 - \frac{c(\varepsilon,M,N)}{1-a(\varepsilon)}) + \frac{c(\varepsilon,M,N)}{1-a(\varepsilon)}$
(3) $h_k \geq h_0$ *for k even and* $h_k \geq h_1$ *for k odd.*

Proof. (1). Let θ be the unique positive number such that $\theta = 1 + 3\sqrt{\theta}$, one can check easily that $\theta < 11$ (numerically, $\theta \cong 10.91$). Using that for $x, y > 0$, $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$ and for $0 < x < 1$, $x < \sqrt{x}$, one verifies by induction that for $k = 1, \dots, 2L$:

$$\xi_k \leq \theta\varepsilon^{2^{-k}} + 3^{\sum_{j=0}^{k-1} 2^{-j}}(\xi_0)^{2^{-k}}$$

and the result follows.

(2). From the definition of the sequence (η_k), for each k:

$$\eta_{k+1} \leq a(\varepsilon)\eta_k + c(\varepsilon, M, N)$$

Using that $4\sqrt{\xi_{\max}} < 20$, the expression of $\eta_{\max}$ follows.

(3). For k even, $h_{k+2} = h_k + \frac{1}{\lambda}(\lambda \Delta H(e) + (1-\lambda)\Delta H(e')) > h_k$, similarly for k odd and the proof is completed by induction. □

The starting entropy h_0 comes from the initialization block.

Lemma 29. *For all $h_0, \eta_0, \xi_0, \varepsilon$, there exists $\bar{N}_0(h_0, \eta_0, \xi_0, \varepsilon)$ such that for any (N_0, M) that satisfy the conditions:*

$$N_0 \geq \bar{N}_0(h_0, \eta_0, \xi_0, \varepsilon) \tag{11}$$

$$\left| \frac{N_0}{M} \Delta H(\epsilon_{p_0}) - h_0 \right| \leq \frac{\eta_0}{3} \tag{12}$$

$$P(\{\tilde{y}_0, P_0(\cdot|\tilde{y}_0) \text{ verifies an } \mathbf{AEP}(M, h_0, \eta_0, \xi_0)\}) \geq 1 - \varepsilon$$

Proof. Since $\mathbf{x}_1, \ldots, \mathbf{x}_{N_0}$ are i.i.d. with law p_0, the conditional distributions $P_0(\mathbf{x}_i | f(\mathbf{x}_i))$ are also i.i.d. and for each $i = 1, \ldots, N_0$, $H(\mathbf{x}_i | f(\mathbf{x}_i)) = \Delta H(\epsilon_{p_0})$. Let $\bar{h} = \Delta H(\epsilon_{p_0}) > 0$, $\bar{\eta} = \frac{\bar{h}}{h_0}\frac{\eta_0}{3}$ and for each N_0:

$$C_{N_0} = \left\{ x_1, \ldots, x_{N_0}, \left| -\frac{1}{N_0} \log P(x_1, \ldots, x_{N_0} | f(x_1), \ldots, f(x_{N_0})) - \bar{h} \right| \leq \bar{\eta} \right\}$$

By the law of large numbers there is n_0 such that for $N_0 \geq n_0$, $P(C_{N_0}) \geq 1 - \xi_0^2$. For each sequence of signals $\tilde{y}_0 = (f(x_1), \ldots, f(x_{N_0}))$, define:

$$C_{N_0}(\tilde{y}_0) = \left\{ x_1, \ldots, x_{N_0}, \left| -\frac{1}{N_0} \log P(x_1, \ldots, x_{N_0} | \tilde{y}_0) - h_0 \right| \leq \bar{\eta} \right\}$$

and set:

$$S_0 = \{\tilde{y}_0, P_0(C_{N_0}(\tilde{y}_0)|\tilde{y}_0) \geq 1 - \xi_0\}$$

Then $P(C_{N_0}) = \sum_{\tilde{y}_0} P_0(\tilde{y}_0) P_0(C_{N_0}(\tilde{y}_0)|\tilde{y}_0) \leq P(S_0) + (1-\xi_0)(1-P(S_0))$ and therefore $P(S_0) \geq 1 - \xi_0$ which means:

$$P(\{\tilde{y}_0, P_0(\cdot|\tilde{y}_0) \text{ verifies an } \mathbf{AEP}(N_0, \bar{h}, \bar{\eta}, \xi_0)\}) \geq 1 - \varepsilon$$

Thus for each $\tilde{y}_0 \in S_0$, $P_0(\cdot|\tilde{y}_0)$ verifies an $\mathbf{AEP}(M, \frac{N_0}{M}\bar{h}, \frac{N_0}{M}\bar{\eta}, \xi_0)$. Choose then (M, N_0) such that condition (12) is fullfilled and from the choice of $\bar{\eta}$, $P_0(\cdot|\tilde{y}_0)$ verifies an $\mathbf{AEP}(M, h_0, \eta_0, \xi_0)$.

□

We give now sufficient conditions for the construction of the process to be valid.

Lemma 30. *If the following two conditions are fullfilled:*

$$M(h_0 - H(e) - \eta_{\max}) \geq 2 \tag{13}$$

$$N(h_1 - H(e') - \eta_{\max}) \geq 2 \tag{14}$$

then for $k = 0, \ldots, 2L$,

$$\begin{cases} M(h_k - H(e) - \eta_k) \geq 2 & \text{for } k \text{ odd} \\ N(h_k - H(e') - \eta_k) \geq 2 & \text{for } k \text{ even} \end{cases}$$

Proof. Follows from lemma 28. □

Summing up we get,

Lemma 31. *Under conditions (11), (12), (13), and (14), the process is well-defined.*

7. Bound on Kullback distance

Let P be the law of the process process $(\mathbf{x}_t)$ defined above. We estimate on each block the distance between the sequence of experiments induced by P with $b^{\otimes M}$ [resp $e'^{\otimes N}$]. Then, we show that these distances can be made small by an adequate choice of the parameters. Finally, we prove the weak-$*$ convergence of the distribution of experiments under P to $\lambda e + (1-\lambda)e'$.

Lemma 32. *There exists a constant $U(e,e')$ such that, if (11), (12), (13), and (14) are fulfilled, then for all k odd,*

$$\mathbf{E}d(P(\mathbf{p}_{t_k+1},\dots,\mathbf{p}_{t_{k+1}}|\tilde{y}_{k-1}))\|e^{\otimes M}) \le M(2\eta_{\max}+U(e,e')(\xi_{\max}+2\eta_{\max}+L\varepsilon))$$

and for all k even,

$$\mathbf{E}d(P(\mathbf{p}_{t_k+1},\dots,\mathbf{p}_{t_{k+1}}|\tilde{y}_{k-1}))\|e'^{\otimes N}) \le N(2\eta_{\max}+U(e,e')(\xi_{\max}+2\eta_{\max}+L\varepsilon))$$

where for each k, t_k denotes the last stage of the $(k-1)$-th block.

Proof. Assume k even. For $\tilde{y}_{k-1} \in S_{k-1}$, proposition 26 shows that:

$$\begin{aligned} & d(P(\mathbf{p}_{t_k+1},\dots,\mathbf{p}_{t_{k+1}}|\tilde{y}_{k-1})\|e^{\otimes M}) \\ \le\ & 2M(\eta_{\max}+\xi_{\max}\log(|\mathrm{supp}\, e|)) + |\mathrm{supp}\, e|\log(M+1)+1 \end{aligned}$$

For $\tilde{y}_{k-1} \notin S_{k-1}$,

$$d(P(\mathbf{p}_{t_k+1},\dots,\mathbf{p}_{t_{k+1}}|\tilde{y}_{k-1})\|e^{\otimes M}) = Md(\epsilon_{p_0}\|e)$$

The result follows, using $P(\cap_1^{2L} S_k) \ge 1 - 2L\xi_{\max}$ and with $U(e,e') = \max(2d(\epsilon_{p_0}\|e), 2d(\epsilon_{p_0}\|e')\}, |\mathrm{supp}\, e|+1, |\mathrm{supp}\, e'|+1)$. □

Lemma 33. *For any L and any $\gamma > 0$, there exists $(\varepsilon, \varepsilon_0, \eta_0)$, and $(\bar{M}, \bar{N})$ such that for all $(M,N) > (\bar{M},\bar{N})$, conditions (13) and (14) are fulfilled and for all N_0 such that (11) and (12) hold, for all k odd,*

$$\mathbf{E}d(P(\mathbf{p}_{t_k+1},\dots,\mathbf{p}_{t_{k+1}}|\tilde{y}_{k-1}))\|e^{\otimes M}) \le M\gamma$$

and for all k even,

$$\mathbf{E}d(P(\mathbf{p}_{t_k+1},\dots,\mathbf{p}_{t_{k+1}}|\tilde{y}_{k-1}))\|e'^{\otimes N}) \le N\gamma$$

Proof. We show how to choose the parameters to achieve the above result.

(1) Choose ε and ξ_0 such that ξ_{max} and $L\varepsilon$ are small.
(2) Choose η_0 and (M_η, N_η) such that $\eta_{\max}$ is small for all $(M,N) \geq (M_\eta, N_\eta)$.
(3) Choose $N_0 \geq N_0(h_0, \eta_0, \xi_0, \varepsilon)$
(4) Choose $(\bar{M}, \bar{N}) \geq (M_\eta, N_\eta)$ such that (13) and (14) are satisfied for $(M,N) \geq (\bar{M}, \bar{N})$.
(5) Choose $(M,N) \geq (\bar{M}, \bar{N})$ such that (12) holds.

Applying lemma 32 then yields the result. □

Lemma 34. *For any $\gamma > 0$, there exists $(\varepsilon, \xi_0, \eta_0, M, N, N_0, L)$ that fulfill (11), (12), (13), and (14) and such that*

(1) *for k odd, $\mathbf{E}d(P(\mathbf{p}_{t_k+1}, \ldots, \mathbf{p}_{t_{k+1}} | \tilde{y}_{k-1})) \| e^{\otimes M}) \leq M\gamma$*
(2) *for k even, $\mathbf{E}d(P(\mathbf{p}_{t_k+1}, \ldots, \mathbf{p}_{t_{k+1}} | \tilde{y}_{k-1})) \| e'^{\otimes N}) \leq N\gamma$*
(3) $\frac{N_0}{\bar{N}} \leq \gamma$

Proof. It is enough to use the previous lemma where L is chosen a large constant times $\frac{1}{\gamma}$. Then, remark that for (M,N) large enough, (12) is fulfilled for N_0 of an order constant times M, hence of the order constant times $\frac{\bar{N}}{L}$. □

7.1. **Weak-$*$ convergence.** Lemma 34 provides a choice of parameters for each $\gamma > 0$, hence a family of processes, and a corresponding family $(\delta_\gamma)_\gamma$ of elements of D_∞.

Lemma 35. *δ_γ weak-$*$ converges to $\lambda e + (1-\lambda)e'$ as γ goes to 0.*

Proof. With $\delta' = \frac{N_0}{\bar{N}} \epsilon_{p_0} + (1 - \frac{N_0}{\bar{N}})(\lambda e + (1-\lambda)e')$, since $\frac{N_0}{\bar{N}} \leq \gamma$, δ' converges weakly to $\lambda e + (1-\lambda)e'$ as γ goes to 0. Let $g \colon E \to \mathbb{R}$ continuous, we prove that $|\mathbf{E}_{\delta'} g - \mathbf{E}_{\delta_\gamma} g|$ converges to 0 as γ goes to 0.

$$
\begin{aligned}
|\mathbf{E}_{\delta'} g - \mathbf{E}_{\delta_\gamma} g| &\leq \frac{1}{\bar{N} - N_0} \sum\nolimits_{k \text{ odd}} \sum\nolimits_{t=t_k+1}^{t_{k+1}} \mathbf{E}|g(\mathbf{b}_t) - g(e)| \\
&+ \frac{1}{\bar{N} - N_0} \sum\nolimits_{k \text{ even}} \sum\nolimits_{t=t_k+1}^{t_{k+1}} \mathbf{E}|g(\mathbf{b}_t) - g(e')|
\end{aligned}
$$

By uniform continuity of g, for every $\bar{\varepsilon} > 0$, there exists $\bar{\alpha} > 0$ such that:

$$\|e_1 - e_2\|_1 \leq \bar{\alpha} \Longrightarrow |g(e_1) - g(e_2)| \leq \bar{\varepsilon}$$

We let $e_k = e$ for k odd and $e_k = e'$ for k even and $\|g\| = \max_{e''} |g(e'')|$. For t in the k-th block:

$$\begin{aligned} \mathbf{E}|g(\mathbf{e}_t) - g(e_k)| &\leq \bar{\varepsilon} + \frac{2\,\|g\|}{\bar{\alpha}} \mathbf{E}\|\mathbf{e}_t - e_k\|_1 \\ &\leq \bar{\varepsilon} + \frac{2\,\|g\|}{\bar{\alpha}} \sqrt{2 \ln 2 \cdot \mathbf{E} d(\mathbf{e}_t \| e_k)} \end{aligned}$$

since $\|p - q\|_1 \leq \sqrt{2 \ln 2 \cdot d(p\|q)}$ ([CT91], lemma 12.6.1 p.300) and from Jensen's inequality. Applying Jensen's inequality again:

$$\frac{1}{N_k} \sum\nolimits_{t=t_k+1}^{t_{k+1}} \mathbf{E}|g(\mathbf{e}_t) - g(e_k)| \leq \bar{\varepsilon} + \frac{2\,\|g\|}{\bar{\alpha}} \sqrt{\frac{2 \ln 2}{N_k} \sum\nolimits_{t=t_k+1}^{t_{k+1}} \mathbf{E} d(\mathbf{e}_t \| e_k)}$$

Now,

$$\begin{aligned} \sum\nolimits_{t=t_k+1}^{t_{k+1}} \mathbf{E} d(\mathbf{e}_t \| e_k) &= \sum\nolimits_{t=t_k+1}^{t_{k+1}} \mathbf{E} d(P(\mathbf{p}_t | \tilde{y}_{k-1}, y_{t_k+1}, \ldots, y_{t-1} \| e_k) \\ &\leq \sum\nolimits_{t=t_k+1}^{t_{k+1}} \mathbf{E} d(P(\mathbf{p}_t | \tilde{y}_{k-1}, \mathbf{p}_{t_k+1}, \ldots, \mathbf{p}_{t-1}) \| e_k) \\ &= \mathbf{E}_{\tilde{y}_{k-1}} d(P(\mathbf{p}_{t_k+1}, \ldots, \mathbf{p}_{t_{k+1}} | \tilde{y}_{k-1}) \| e_k^{\otimes N_k}) \\ &\leq N_k \gamma \end{aligned}$$

where the first inequality comes from the convexity of the Kullback distance. Reporting in the previous and averaging over blocks yields:

$$|\mathbf{E}_{\delta'} g - \mathbf{E}_{\delta_\gamma} g| \leq \bar{\varepsilon} + \frac{2\,\|g\|}{\bar{\alpha}} \sqrt{2 \ln 2 \cdot \gamma}$$

Thus, $|\mathbf{E}_{\delta'} g - \mathbf{E}_{\delta_\gamma} g|$ goes to 0 as γ goes to 0. □

References

[APS90] D. Abreu, D. Pearce, and E. Stacchetti. Toward a theory of discounted repeated games with imperfect monitoring. *Econometrica*, 58:1041–1063, 1990.

[Bla51] D. Blackwell. Comparison of experiments. In *Proceedings of the Second Berkeley*

Symposium on Mathematical Statistics and Probability, pages 93–102. University of California Press, 1951.

[Bla53] D. Blackwell. Equivalent comparison of experiments. *Annals of Mathematical Statistics*, 24:265–272, 1953.

[BN03] G. Bavly and A. Neyman. Online concealed correlation by boundedly rational players. Discussion Paper Series 336, Center for Rationality and Interactive Decision Theory, Hebrew University, Jerusalem, 2003.

[BP93] E. Ben Porath. Repeated games with finite automata. *Journal of Economic Theory*, 59:17–32, 1993.

[Com02] O. Compte. Communication in repeated games with imperfect private monitoring. *Econometrica*, 66:597–626, 2002.

[CT91] T. M. Cover and J. A. Thomas. *Elements of information theory.* Wiley Series in Telecomunications. Wiley, 1991.

[FLM94] D. Fudenberg, D. K. Levine, and E. Maskin. The folk theorem with imperfect public information. *Econometrica*, 62:997–1039, 1994.

[GH03] O. Gossner and P. Hernández. On the complexity of coordination. *Mathematics of Operations Research*, 28:127–141, 2003.

[GLT03] O. Gossner, R. Laraki, and T. Tomala. On the optimal use of coordination. mimeo, 2003.

[Gol03] Y. Goldberg. On the minmax of repeated games with imperfect monitoring: a computational example. mimeo, 2003.

[Gos98] O. Gossner. Repeated games played by cryptographically sophisticated players. DP 9835, CORE, 1998.

[Gos00] O. Gossner. Sharing a long secret in a few public words. DP 2000-15, THEMA, 2000.

[GT03] O. Gossner and T. Tomala. Entropy and codification in repeated games with signals. Cahiers du CEREMADE 309, Université Paris Dauphine, Paris, 2003.

[GV01] O. Gossner and N. Vieille. Repeated communication through the ‘and’ mechanism. *International Journal of Game Theory*, 30:41–61, 2001.

[GV02] O. Gossner and N. Vieille. How to play with a biased coin? *Games and Economic Behavior*, 41:206–226, 2002.

[Leh88] E. Lehrer. Repeated games with stationary bounded recall strategies. *Journal of Economic Theory*, 46:130–144, 1988.

[Leh90] E. Lehrer. Nash equilibria of n player repeated games with semi-standard information. *International Journal of Game Theory*, 19:191–217, 1990.

[Leh91] E. Lehrer. Internal correlation in repeated games. *International Journal of Game Theory*, 19:431–456, 1991.

[Leh94] E. Lehrer. Finitely many players with bounded recall in infinitely repeated games. *Games and Economic Behavior*, 7:390–405, 1994.

[LT03] G. Lacôte and G. Thurin. How to efficiently defeat strategies of bounded complexity. mimeo, 2003.

[Ney97] A. Neyman. Cooperation, repetition, and automata. In S. Hart and A. Mas-Colell, editors, *Cooperation: Game-Theoretic Approaches*, volume 155 of *NATO ASI Series F*, pages 233–255. Springer-Verlag, 1997.

[Ney98] A. Neyman. Finitely repeated games with finite automata. *Mathematics of Operations Research*, 23:513–552, 1998.

[Par67] K. R. Parthasaraty. *Probability Measures on Metric Spaces*. Academic Press, New York, 1967.

[PR03] M. Piccione and A. Rubinstein. Modeling the economic interaction of agents with diverse abilities to recognize equilibrium patterns. *Journal of European Economic Association*, 1:212–223, 2003.

[Roc70] R. T. Rockafellar. *Convex analysis*. Princeton University Press, 1970.

[RT98] J. Renault and T. Tomala. Repeated proximity games. *International Journal of Game Theory*, 27:539–559, 1998.

[RT00] J. Renault and T. Tomala. Communication equilibrium payoffs of repeated games with imperfect monitoring. Cahiers du CEREMADE 0034, Université Paris Dauphine, Paris, 2000.

CERAS, URA CNRS 2036

E-mail address: Olivier.Gossner@mail.enpc.fr

CEREMADE, UMR CNRS 7534 Université Paris 9 – Dauphine

E-mail address: tomala@ceremade.dauphine.fr