Approximate maximizers of intricacy functionals

J. Buzzi · L. Zambotti

Received: 14 September 2010 / Revised: 15 February 2011 / Published online: 2 March 2011
© Springer-Verlag 2011

Abstract   G. Edelman, O. Sporns, and G. Tononi introduced in theoretical biology the neural complexity of a family of random variables. A previous work showed that this functional is a special case of intricacy, i.e., an average of the mutual information of subsystems with specific mathematical properties. Moreover, its maximum value grows at a definite speed with the size of the system. In this work, we compute exactly this speed of growth by building “approximate maximizers” subject to an entropy condition. These approximate maximizers work simultaneously for all intricacies. We also establish some properties of arbitrary approximate maximizers, in particular the existence of a threshold in the size of subsystems of approximate maximizers: most smaller subsystems are almost equidistributed, most larger subsystems determine the full system. The main ideas are a random construction of almost maximizers with a high statistical symmetry and the consideration of entropy profiles, i.e., the average entropies of sub-systems of a given size. The latter gives rise to interesting questions of probability and information theory.

Keywords   Entropy · Complexity · Maximization · Discrete probability

Mathematics Subject Classification (2000)   94A17 · 92B20 · 60C05
1 Introduction

1.1 Neural complexity, a measure of complexity from theoretical biology

In [17], Edelman et al. introduced the so called neural complexity of a family of random variables. It has been considered from a theoretical and experimental point of view by a number of authors, see e.g. [1,4–7,11–19].

In order to define the neural complexity, we need to recall two classical definitions. If $X$ is a random variable taking values in a finite space $E$, then its entropy is defined by

$$H(X) := -\sum_{x \in E} P_X(x) \log(P_X(x)), \quad P_X(x) := \mathbb{P}(X = x).$$

Given two random variables defined over the same probability space, the mutual information between $X$ and $Y$ is

$$\text{MI}(X, Y) := H(X) + H(Y) - H(X, Y).$$

We refer to [3,5,9] for a review of the main properties of the entropy and the mutual information and for introductions to information theory and to the various roles of entropy in mathematical physics, respectively. For now, it suffices to recall that $\text{MI}(X, Y) \geq 0$ is equal to zero if and only if $X$ and $Y$ are independent, and therefore $\text{MI}(X, Y)$ is a measure of the dependence between $X$ and $Y$.

Edelman, Sporns and Tononi consider systems formed by a finite family $X = (X_i)_{i \in I}$ and define the following concept of complexity. For any $S \subset I$, they divide the system into two subsystems:

$$X_S := (X_i, i \in S), \quad X_{Sc} := (X_i, i \in Sc),$$

where $Sc := I \setminus S$. Then they compute the mutual information $\text{MI}(X_S, X_{Sc})$ and consider the average

$$\mathcal{I}(X) := \frac{1}{|I| + 1} \sum_{S \subset I} \frac{1}{|S| \binom{|I|}{|S|}} \text{MI}(X_S, X_{Sc}), \quad (1.1)$$

where $|I|$ denotes the cardinality of $I$ and $\text{MI}(X_{\emptyset}, X_I) = \text{MI}(X_I, X_{\emptyset}) := 0$.

In [2], we showed that the neural complexity belongs to a family of averages of mutual informations characterized by natural symmetry and additivity properties. More precisely, denote by $\mathcal{X}(d, N)$ the set of families $X = (X_i)_{1 \leq i \leq N}$ of random variables taking values in $\{0, 1, \ldots, d - 1\}$, where $d, N \in \mathbb{N}^* := \{1, 2, \ldots\}$ are positive integers. If $X \in \mathcal{X}(d, N)$ then we say that its law $\mu$ on $\Lambda_{d,N} := \{0, 1, \ldots, d - 1\}^N$ belongs to $\mathcal{M}(d, N)$. 

Springer
Definition 1.1  An intricacy is a function $I_c : \bigcup_{d,N} \mathcal{X}(d, N) \to \mathbb{R}_+$ given by:

$$I_c(X) := \sum_{S \subseteq I} c_S^I \text{MI}(X_S, X_{S^c}), \quad X \in \mathcal{X}(d, N),$$

where $I := \{1, \ldots, N\}$ and the coefficients $c := (c_S^I)_{S \subseteq I}$ are given by

$$c_S^I = \mathbb{E}(W_c^{|S|}(1 - W_c)^{|I\setminus S|}),$$

$W_c$ is a random variable in $[0, 1]$ such that $W_c$ and $1 - W_c$ have the same law $\lambda_c$ and we use the convention $0^0 = 1$. Moreover, we say that $I_c$ is null, if $c_S^I = 0$ whenever $S \not\in \{\emptyset, I\}$ (this occurs if and only if $\lambda_c([0,1]) = 0$).

The original Edelman-Sporns-Tononi neural complexity (1.1) is obtained by taking $W_c$ as the uniform r.v. over $[0, 1]$ so $c_S^I = (|I| + 1)^{-1} (|I| - 1)^{-1}$. Note the common abuse of language: $I_c(X)$, like the mutual information $\text{MI}(X_S, X_{S^c})$ and the entropy $H(X)$, is really a function of the law of $X$, not of its random realization. We sometimes more carefully write $I_c(\mu)$ for $\mu \in \bigcup_{d,N} \mathcal{M}(d, N)$.

Using a super-additivity argument, we showed in [2] that the maximum value of any intricacy over systems with a given size $N$ is asymptotically proportional to $N$. Observe that maximizing intricacy requires a compromise between randomness (without which there is no entropy) and strong correlations (without which the mutual information vanishes). We refer to [2] for a discussion of several properties and examples.

This paper analyzes this maximization problem in the limit $N \to \infty$. We compute exactly the speed of growth, i.e., the proportionality constant of the previous paragraph, by building “approximate maximizers”, that is, families of an increasing number of random variables taking value in a fixed set and achieving, in the limit, the maximum intricacy per variable. Moreover, we shall construct in this paper a sequence of approximate maximizers which works simultaneously for all intricacies.

Our construction is probabilistic in a fundamental way. We shall show that maximizers should approximately satisfy strong symmetries (see Proposition 1.5), that cannot be satisfied exactly (Lemma 3.1). However, we shall exhibit a random sequence of systems, which satisfy such symmetries in law, and approximately satisfy the same symmetries almost surely (Proposition 4.1).

1.2 Main results

We consider some intricacy $I_c$. Let $\lambda_c$ be the law of the random variable $W_c$ defining as above the coefficients $c_S^I$. Let, for $x \in [0, 1]$ and $N \in \mathbb{N}^*$,

$$i_N^c(x) := 2 \sum_{k=0}^N c_k^N \left( \binom{N}{k} \right) (k/N) \wedge x - x \quad \text{and} \quad i^c(x) := 2 \int_0^1 x \wedge t \lambda_c(dt) - x. $$

(1.2)
These two functions play a crucial role in this paper, and it will be seen below that $i_N^c$ converges uniformly to $i^c$ as $N \to \infty$. In the case of the neural complexity we have $i(x) = x(1 - x)$. We can now state our main results.

**Theorem 1.2** Let $I^c$ be an intricacy.

1. We have for all $\mu \in \mathcal{M}(d, N)$
   \[
   \frac{I^c(\mu)}{N \log d} \leq i_N^c(x), \quad \text{with} \quad x := \frac{H(\mu)}{N \log d}. \tag{1.3}
   \]

2. For any $x \in [0, 1]$, there exists a sequence $\mu^N \in \mathcal{M}(d, N)$ approaching the upper bound of point (1), i.e., satisfying:
   \[
   \lim_{N \to +\infty} \frac{H(\mu^N)}{N \log d} = x, \quad \lim_{N \to +\infty} \frac{I^c(\mu^N)}{N \log d} = i^c(x). \tag{1.4}
   \]

The case $x = 1/2$ plays a special role since it follows from Theorem 1.2 that, setting $I^c(d, N) := \max_{\mu \in \mathcal{M}(d, N)} I^c(\mu)$, we have
\[
\lim_{N \to \infty} \frac{I^c(d, N)}{N \log d} = i^c(1/2).
\]

While the upper bound (1.3) follows from direct computations, the existence of sequences $(\mu^N)_N$ approaching this bound is much more involved and is the main result of this paper. As shown in Proposition 1.5 below, such sequences must exhibit a non-trivial behavior, combining a large amount of local independence and of non-null correlation on a global level.

The main goal of this paper is the study of approximate maximizers, i.e., sequences $\mu^N \in \mathcal{M}(d, N)$ maximizing $\lim_{N \to \infty} I^c(\mu^N)/N \log d$ for a given intricacy $I^c$ and integer $d \geq 2$. It turns out to be more convenient and interesting to study approximate $x$-maximizers, i.e., sequences $\mu^N \in \mathcal{M}(d, N)$ satisfying (1.4). Their existence follows in our approach from a probabilistic construction: we shall prove that uniform distributions on appropriately chosen random sparse supports will have almost surely the desired properties: see Proposition 4.1 below. This construction is well-known in coding theory as the random code ensemble, see e.g. [3, 9].

In the course of the proof, we also obtain rather detailed information on the structure of approximate $x$-maximizers. A key notion is the following one.

**Definition 1.3** Given $X \in \mathcal{X}(d, N)$, its entropy profile is the function $h_X : [0, 1] \to [0, 1]$ such that $h_X(0) = 0,$
\[
h_X \left( \frac{k}{N} \right) = \frac{1}{\binom{N}{k}} \sum_{S \subset I, |S| = k} \frac{H(X_S)}{N \log d}, \quad k \in I := \{1, \ldots, N\}
\]
and $h_X$ is affine on each interval $[\frac{k-1}{N}, \frac{k}{N}]$, $k \in I$. 

\[ Springer \]
For $x \in [0, 1]$ and $N \geq 1$, the ideal entropy profile is

$$h_x^*(t) := t \wedge x = \min\{t, x\}. \quad (1.5)$$

**Proposition 1.4** Let $\mathcal{I}^c$ be an intricacy.

1. If $1/2$ is in the support of $\lambda_c$, then any approximate maximizer $(X^N)_N$ for $\mathcal{I}^c$ satisfies:

$$\lim_{N \to +\infty} \frac{H(X^N)}{N \log d} = \frac{1}{2}, \quad \lim_{N \to +\infty} \sup_{\text{supp}(\lambda_c)} |h_{X^N} - h_{1/2}^*| = 0. \quad (1.6)$$

2. Let $x \in [0, 1]$ and let $(X^N)_N$ be an approximate $x$-maximizer for $\mathcal{I}^c$. Then:

$$\lim_{N \to \infty} \sup_{\text{supp}(\lambda_c)} |h_{X^N} - h^*_x| = 0. \quad (1.7)$$

In particular, if $x \in \text{supp}(\lambda_c)$ then

$$\lim_{N \to \infty} \sup_{[0,1]} |h_{X^N} - h^*_x| = 0. \quad (1.8)$$

3. If $x \in \text{supp}(\lambda_c)$ then an approximate $x$-maximizer $(X^N)_N$ for $\mathcal{I}^c$ is an approximate $x$-maximizer for any other intricacy $\mathcal{I}^{c'}$.

4. Any approximate maximizer for an intricacy with $1/2 \in \text{supp}(\lambda_c)$, e.g., the neural complexity, is an approximate maximizer for any other intricacy.

The last point of the proposition asserts that, from the point of view of approximate maximizers, all intricacies with $1/2 \in \text{supp}(\lambda_c)$ have the same properties.

The extra assumptions about the support of $\lambda_c$ in (2), (3), (4) above, cannot be dropped as show easy computations in the case of $\lambda_c = \frac{1}{2}(\delta_p + \delta_{1-p})$ for $p \neq 1/2$.

We have the following consequence for approximate $x$-maximizers. We recall that $H(Y \mid Z) := H(Y, Z) - H(Z)$ denotes the conditional entropy.

**Proposition 1.5** Suppose that $x \in \text{supp}(\lambda_c)$ and let $(X^N)_N$ be an approximate $x$-maximizer for some $0 \leq x \leq 1$.

1. If $y \in ]0, x[$ then for all $\varepsilon > 0$

$$\lim_{N \to +\infty} \frac{1}{\binom{N}{\lfloor yN \rfloor}} \# \left\{ S \subset \{1, \ldots, N\} : |S| = \lfloor yN \rfloor, (1 - \varepsilon)|S| \log d < H(X^N_S) \leq |S| \log d \right\} = 1.$$

2. If $y \in [x, 1[$ then for all $\varepsilon > 0$

$$\lim_{N \to +\infty} \frac{1}{\binom{N}{\lfloor yN \rfloor}} \# \left\{ S \subset \{1, \ldots, N\} : |S| = \lfloor yN \rfloor, H(X^N \mid X^N_S) < \varepsilon x N \log d \right\} = 1.$$
This result can be loosely interpreted as follows: as \( N \to +\infty \),

1. if \( y \in ]0, x[ \) then for almost all subsets \( S \) with \( |S| = \lfloor yN \rfloor \), \( X_S \) is almost uniform over \( \{0, \ldots, d-1\}^S \);
2. if \( y \in [x, 1[ \) then for almost all subsets \( S \) with \( |S| = \lfloor yN \rfloor \), \( X \) is almost a function of \( X_S \).

This follows from the relation between entropy and conditional entropy on one side and independence versus dependence on the other side.

1.3 Further questions

We list a few questions that arise from this work.

1. The function \( i^c \) is symmetric with respect to \( x = 1/2 \) and independent of \( d \geq 2 \). However we do not know whether these simple symmetries reflect deeper properties and could be proved directly.
2. This work has focused on properties of systems with size tending to infinity. Notice that we know very little on the exact maximizers for fixed size. Exact maximizers are non-unique but we do not even know if there are only finitely many of them.
3. Our construction is global. However it would be interesting to build systems with maximum intricacy by means of a local approach, i.e., a “biologically reasonable” building process, using some type of local rules and/or evolution. That would amount to find a “reasonable” map \( T : \mathcal{M}(d, N) \to \mathcal{M}(d, N) \) such that the neural complexity of \( T^n(\mu) \) converges to the maximum as \( n \to \infty \) for “many” \( \mu \in \mathcal{M}(d, M) \).
4. Our Lemma 3.1 below seems to be relevant to recent results on \( k \)-independence, see [10], and leads to interesting questions in the theory of entropy and information, see Sect. 3 below.

2 Upper bounds on intricacies

In [2], it was proved that \( \mathcal{I}^c(X) < N \log d/2 \) if \( X \in \mathcal{X}(d, N) \). By comparison with the ideal entropy profile defined in (1.5), we prove asymptotically sharp upper bounds for systems with given size and entropy.

2.1 Definitions

Let \( \mathcal{I}^c \) be some intricacy. Recall from Definition 1.1 that its coefficients are defined by some random variable \( W_c \) with law \( \lambda_c \). Let \( (Y_i)_{i \geq 1} \) be a sequence of i.i.d. uniform random variables on \( [0, 1] \) and let

\[
D_N := \sum_{k=1}^{N} 1_{(Y_k \leq W_c)}, \quad \beta_N := \frac{D_N}{N}, \quad N \geq 1.
\]
Conditionally on $W_c$, $D_N$ is a binomial variable with parameters $(N, W_c)$. In particular, for all $g : \mathbb{N} \mapsto \mathbb{R}$, using the expression for the coefficients from Definition 1.1, for all bounded Borel $g : [0, 1] \mapsto \mathbb{R}$

$$
\mathbb{E}(g(\beta_N)) = \int \sum_{k=0}^{N} \binom{N}{k} x^k (1-x)^{N-k} g\left(\frac{k}{N}\right) \lambda_c(dx) = \sum_{k=0}^{N} c_N^k \binom{N}{k} g\left(\frac{k}{N}\right). \quad (2.2)
$$

Finally, the following norm is adapted to measure the distance between entropy profiles. For all bounded Borel $f : [0, 1] \mapsto \mathbb{R}$, let

$$
\|f\|_{c,N} := \sum_{k=0}^{N} c_N^k \binom{N}{k} |f(k/N)| = \mathbb{E}(|f(\beta_N)|). \quad (2.3)
$$

Notice that if $\mathcal{I}_c$ is non-null, $\lambda_c(0,1) > 0$ and therefore none of the coefficients $c_N^k = \mathbb{E}(W_c^k (1 - W_c)^{N-k})$ is zero.

2.2 Upper bounds and distance from the ideal profile

In this section we prove the following upper bounds

**Proposition 2.1** Let $\mathcal{I}_c$ be an intricacy.

1. $i^c : [0, 1] \mapsto [0, 1]$ is a concave function admitting the Lipschitz constant 1 and symmetric about 1/2: $i^c(1-x) = i^c(x)$. Moreover, $i^c(1/2) = \max_{x \in [0,1]} i^c(x)$ and $d^2 i^c / dx^2 = -2 \lambda_c$.

2. $|i^c(x) - i^c_N(x)| \leq 1/\sqrt{N}$.

3. All systems $X \in \mathcal{X}(d, N)$ with $\frac{H(X)}{N \log d} = x$ satisfy:

$$
\frac{i^c(X)}{N \log d} = i^c_N(x) - \|h_X - h^*_X\|_{c,N} \leq i^c_N(x). \quad (2.4)
$$

4. If $X^N \in \mathcal{X}(d, N)$ and $\lim_{N \to \infty} \frac{H(X^N)}{N \log d} = x$, then

$$
\limsup_{N \to \infty} \frac{i^c(X^N)}{N \log d} \leq i^c(x). \quad (2.5)
$$

Showing that $i^c(x)$ is indeed the value of the limit (2.5), rather than a mere upper bound, requires to prove the existence of sequences saturating the inequality. This is deferred to the next section. Before proving Proposition 2.1, we need some preliminary material which will also be useful later.
2.3 Intricacy as a function of the profile

Let us set

$$\Gamma := \{h : [0, 1] \mapsto [0, 1] : h(0) = 0, \ t \mapsto h(t) \text{ is non-decreasing and 1-Lipschitz}\}$$

and for any real number $x \in [0, 1]$

$$\Gamma_x := \{h \in \Gamma : h(1) = x\}.$$ 

These sets are endowed with the partial order: $h \leq g$ if and only if $h(t) \leq g(t)$ for all $t \in [0, 1]$. Each $\Gamma_x$ has a unique maximal element: the previously introduced ideal entropy profile, $h^*_x(t) = t \land x$.

**Lemma 2.2** For any $X \in \mathcal{X}(d, N)$, the entropy profile $h_X$, defined according to Definition 1.3, belongs to $\Gamma$.

**Proof** We recall that for a finite set $E$ and any $E$-valued random variable $Y$,

$$0 \leq H(Y) \leq \log |E|$$

and $H(Y) = \log |E|$ iff $Y$ is uniform over $E$. Moreover for any random vector $(X, Y)$ we have

$$\max\{H(X), H(Y)\} \leq H(X, Y) \leq H(X) + H(Y).$$

Let $X \in \mathcal{X}(d, N)$. Setting $I := \{1, \ldots, N\}$ and

$$H_k := \frac{1}{\binom{N}{k}} \sum_{S \subseteq I, |S| = k} H(X_S), \ k = 0, 1, \ldots, N$$

we must prove that

$$0 = H_0 \leq H_1 \leq \cdots \leq H_N = H(X), \ H_{k+1} - H_k \leq \log d, \ 0 \leq k < N.$$ 

The equalities $H_0 = 0$ and $H_N = H(X)$ are obvious. Let $0 \leq k < N$ and compute:

$$H_{k+1} = \frac{1}{\binom{N}{k+1}} \sum_{|S| = k+1} H(X_S) = \frac{1}{\binom{N}{k+1}} \sum_{|S| = k} \frac{1}{k+1} \sum_{i \in S'} H(X_{S \cup \{i\}})$$

$$\leq \frac{k!(N-k-1)!}{N!} \sum_{|S| = k} (N-k) (H(X_S) + \log d) = H_k + \log d$$

since $H(X_{S \cup \{i\}}) \leq H(X_S) + H(X_i)$ by the second inequality of (2.7). The same computation, since $H(X_{S \cup \{i\}}) \geq H(X_S)$ by the first inequality of (2.7), proves $H_k \leq H_{k+1}$. 

$\square$
Let for any $h \in \Gamma$

$$G^c_N(h) := 2 \sum_{k=0}^{N} c_k^N \binom{N}{k} h(k/N) - h(1) = 2 \mathbb{E} (h(\beta_N)) - h(1).$$

**Lemma 2.3** Fix $x \in [0, 1]$.

1. For all $X \in \mathcal{X}(d, N)$, $G^c_N(h_X) = \frac{\mathcal{I}(X)}{N \log d}$.
2. $G^c_N(h^*_x) = i^c_N(x)$ and $h^*_x$ is a maximizer of $G^c_N$ in $\Gamma_x$. It is the unique one if $\mathcal{I}^c$ is non-null.
3. For arbitrary $h \in \Gamma_x$, we have

$$\|h - h^*_x\|_{c,N} = |G^c_N(h) - G^c_N(h^*_x)|.$$

**Proof** Since $\mathbb{M}(X, Y) = H(X) + H(Y) - H(X, Y)$, $c^I_S = c^I_{Sc}$, and $\sum_S c^I_S = 1$,

$$\frac{\mathcal{I}(X)}{N \log d} = 2 \sum_{k=0}^{N} c_k^N \sum_{|S|=k} \frac{H(X_S)}{N \log d} - \frac{H(X)}{N \log d} = G^c_N(h_X).$$

Observe that $G^c_N : \Gamma_x \rightarrow \mathbb{R}$ is monotone, increasing if $\mathcal{I}^c$ is non-null, and that $h^*_x$ is the maximal element of $\Gamma_x$. Then (1), (2) and (3) follow by the definition of $G^c_N(h)$ and by (2.3).

**Proof of Proposition 2.1** The first point follows easily from the explicit definition of $i^c$. Notice that $i^c$ is the solution of the equation $d^2 i^c / dx^2 = -2\lambda_c$ with Dirichlet boundary condition at $\{0, 1\}$. In order to prove (2), we use the probabilistic representations (1.2) and (2.2). We obtain

$$|i^c_N(x) - i^c(x)| \leq \mathbb{E} \left( |h^*_N(\beta_N) - h^*_N(W_c)| \right) \leq \mathbb{E} (|\beta_N - W_c|) \leq \sqrt{\mathbb{E} (|\beta_N - W_c|^2)}.$$

Since $D_N = N\beta_N$ is, conditionally on $W_c$, a binomial variable with parameters $(N, W_c)$, we have that

$$\mathbb{E} \left( |\beta_N - W_c|^2 \right) = \mathbb{E} (\text{Var} (\beta_N \mid W_c)) = \mathbb{E} \left( \frac{W_c(1 - W_c)}{N} \right) \leq \frac{1}{4N}. \quad (2.8)$$

Formula (2.4) follows from Lemma 2.3, since for $\frac{H(X)}{N \log d} = x$

$$\frac{\mathcal{I}(X)}{N \log d} = G^c_N(h_X) = G^c_N(h^*_x) + G^c_N(h_X) - G^c_N(h^*_x) = i^c_N(x) - \|h_X - h^*_x\|_{c,N}.$$

Formula (2.5) follows easily.
3 No system with the ideal profile

We turn to the problem of maximizing $\mathcal{I}^c$ over $\mathcal{X}(d, N)$ at fixed $N$ for a prescribed value of the entropy $H(X)$. The results of Sect. 2 show that a system $X \in \mathcal{X}(d, N)$ such that $h_X(k/N) = h_N^*(k/N)$ ($k = 0, 1, \ldots, N$) with $x = \frac{H(X)}{N \log d}$ would be an exact maximizer. However, the next Lemma shows that such $X$ cannot exist except if $K$ or $N - K$ are bounded, independently of $N$. Thus, in general all we can hope is to find systems which approach the ideal profile. This will be done in Sect. 4.

We notice first that a system $X \in \mathcal{X}(d, N)$ such that $h_X(k/N) = h_N^*(k/N)$ ($k = 0, 1, \ldots, N$) with $x = \frac{H(X)}{N \log d}$ must satisfy $H(X_S)/\log d = |S| \wedge (xN)$ for all $S \subseteq I$. Indeed, for $|S| = k$ we have $H(X_S)/\log d \leq k \wedge (xN)$ and the average over all such $S$ is the ideal profile $h_X(k/N)$ which coincide with the upper bound: this is possible only if the inequality is an equality for all $S$.

Lemma 3.1 For each $d \geq 2$, there exists $H_\ast = H_\ast(d) < \infty$ with the following property. If $N \geq 1$ and $Y_1, \ldots, Y_N$ are random variables taking values in $\{0, \ldots, d - 1\}$ and defined on the same probability space such that, for some real number $H \in [0, N]$,

$$\frac{H(Y_{\sigma(1)}, \ldots, Y_{\sigma(k)})}{\log d} = k \wedge H, \ \forall \sigma \in S_N, \ \forall k = 1, \ldots, N,$$

then $H$ or $N - H \leq H_\ast$.

Proof Let $K := \lceil H \rceil + 1$. Without loss of generality, we assume that $K \geq 4$. Let us condition on the variables $(X_1, \ldots, X_K)$; in the following paragraphs we simply write “conditional” for “conditional on $(X_1, \ldots, X_K)$”. By assumption:

- $(X_1, X_2)$ belongs to $Z := \{0, \ldots, d - 1\}^2$;
- each $X_i$, $K < i \leq N$, is a function of $X_1, X_2$ as the conditional entropy of $(X_1, X_2, X_i)$ is not bigger than that of $(X_1, X_2)$. Moreover, the conditional entropy of $X_i$ is $\log d$. Hence, each such $X_i$ defines a partition $Z_i$ of $Z$ into $d$ subsets.
- For any pair $i \neq j$ in $\{1, 2, K + 1, \ldots, N\}$, $(X_i, X_j)$ has conditional entropy $(H - K + 2) \log d$, strictly greater than that of $X_i$ or $X_j$, both equal to $\log d$. In particular, $Z_i \neq Z_j$.

Thus, we have an injection from $\{1, 2, K + 1, \ldots, N\}$ into the set of partitions of $Z$ into $d$ subsets. This implies that $N - K + 2$ is bounded by a constant which depends only on $d$. \qed

We notice that this result can be interpreted in the context of $k$-independent families of variables, see [10]. In fact, it says that a finite family of $N$ discrete random variables can not have simultaneously the two following properties

1. any subset of $K$ variables is uniformly distributed over $\{0, \ldots, d - 1\}^K$;
2. any subset of $K + 1$ variables is a function of $K$ among them

unless $K$ or $N - K$ are bounded, uniformly in $N$.

Lemma 3.1 also leads to the following question in the theory of entropy and information.
Problem Describe the set of functions $h : \{0, \ldots, N\} \rightarrow \mathbb{R}$ obtained from picking $X \in \mathcal{X}(d, N)$ and setting $h(k)$ to be the average entropy of $X_S$ where $S$ ranges over the subsets of $\{1, \ldots, N\}$ with cardinality $k$.

Indeed, the basic properties of the entropy (2.6) and (2.7) (which together with their consequences are called Shannon inequalities) are satisfied by the ideal profile, which fails to the entropy profile of some probabilistic system. Thus, the set of entropy functions is not characterized by these conditions (see [8]).

4 Random construction of approximate maximizers

We are going to compute the maximum intricacy per site when the size goes to infinity, i.e., for $d \in \mathbb{N}^*$

$$I^c(d) := \lim_{N \to \infty} \sup \left\{ \frac{I^c(\mu^N_N)}{N \log d} : \mu^N_N \in \mathcal{M}(d, N) \right\}$$

Our analysis requires the consideration of the more precise entropy-intricacy function for $d \in \mathbb{N}^*$ and $0 \leq x \leq 1$:

$$I^c(d, x) := \sup \left\{ \limsup_{N \to \infty} \frac{I^c(X^N)}{N \log d} : X^N \in \mathcal{X}(d, N) \text{ s.t. } \lim_{N \to \infty} \frac{H(X^N)}{N \log d} = x \right\}.$$  (4.1)

Proposition 2.1 established that $I^c(d, x) \leq i^c(x)$, in the notation (1.2). Proposition 4.1 in this section shows that this inequality is in fact an equality.

In the rest of this section we construct approximate $x$-maximizers by choosing uniform distributions on random supports with the appropriate size: since $\frac{H(\mu^N_N)}{N \log d}$ must be close to $x$ and $\mu^N_N$ is uniform, then the size of the (random) support of $\mu^N_N$ must be close to $d^x$, recall (2.6).

It turns out that this simple construction, well known in information theory as the random code ensemble [9], yields the desired results.

4.1 Sparse random configurations

Let $N \geq 2$ and $0 \leq M \leq N$ be integers. We write $\Lambda_{d,n} := \{0, \ldots, d - 1\}^n, \forall n \geq 1$. We consider a family $(W_i)_{i \in \Lambda_{d,M}}$ of i.i.d. variables, each uniformly distributed on $\Lambda_{d,N}$, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We define a random probability measure on $\Lambda_{d,N}$

$$\mu^{N,M}(x) := d^{-M} \sum_{i \in \Lambda_{d,M}} 1_{(x=W_i)}, \quad x \in \Lambda_{d,N}. \quad (4.2)$$
In what follows we consider random variables $X^{N,M}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ such that
\[
\mathbb{P}\left(X^{N,M} = x \mid (W_i)_{i \in \Lambda_{d,M}}\right) = \mu^{N,M}(x), \quad x \in \Lambda_{d,N}.
\]
(4.3)

In other words,
\[
\text{conditionally on } (W_i)_{i \in \Lambda_{d,M}}, \quad X^{N,M} \text{ has law } \mu^{N,M}.
\]

We are going to prove the following

**Proposition 4.1** For integers $N \geq 1$, $0 \leq M \leq N$, let $X^{M,N}$ be the random systems defined above. Let $x \in [0, 1]$. For any intricacy $I^c$ we have, a.s. and in $L^1$
\[
\lim_{N \to +\infty} \frac{I^c(\mu^{N,\lfloor xN \rfloor})}{N \log d} = i^c(x)
\]
(4.4)
\[
\text{and}
\lim_{N \to +\infty} \frac{H(\mu^{N,\lfloor xN \rfloor})}{N \log d} = x.
\]
(4.5)

**Remark 4.2** We stress that in the following we write
\[
I^c(X^{N,M}) = I^c(\mu^{N,M}) \quad \text{and} \quad H(X^{N,M}) = H(\mu^{N,M})
\]
(4.6)
and that all these expressions are random variables which depend on $(W_i)_{i \in \Lambda_{d,M}}$. In other words, (4.6) indicates entropy and intricacy of the law of $X^{N,M}$ conditionally on $(W_i)_{i \in \Lambda_{d,M}}$. This abuse of notation seems necessary, to keep notation reasonably readable.

### 4.2 Average intricacy of sparse random configurations

We recall that $N \geq 2$, $M$ is an integer between 1 and $N$ and $S_N$ denotes the set of permutations of $\{1, \ldots, N\}$. By Lemma 2.3, $I^c(X^{N,M}) = 2 \sum_{k=0}^{N} c_k^N \binom{N}{k} h_X(k/N) - h_X(1)$, hence we get:
\[
\mathbb{E}\left(I^c(X^{N,M})\right) = \frac{1}{N!} \sum_{\sigma \in S_N} 2 \sum_{k=1}^{N} c_k^N \binom{N}{k} \mathbb{E}\left(H(X^{N,M}_{\sigma(1), \ldots, \sigma(k)})\right)
\]
\[
- \mathbb{E}\left(H(X^{N,M})\right).
\]
(4.7)

We are going to simplify this expression by exploiting the symmetries of our construction. Indeed, the random vector $X^{N,M} = (X_1^{N,M}, \ldots, X_N^{N,M}) \in \mathcal{X}(d, N)$ is
exchangeable, i.e. for all $\sigma \in S_N$ and any $\Phi : \Lambda_{d,N} \mapsto \mathbb{R}$

$$
\mathbb{E} \left( \Phi \left( X_{\sigma(1)}, \ldots, X_{\sigma(N)} \right) \right) = \mathbb{E} \left( \Phi \left( X_1, \ldots, X_N \right) \right).
$$

This follows easily from exchangeability of $(W_i)_{i \in \Lambda_{d,M}}$. Notice however that $\sum_\sigma \mu_{N,M}^\sigma$ has same law as $\mu_{N,M}$, but in general the two measures are not a.s. equal. In other words, $(X_1, \ldots, X_N)$ is exchangeable but not exchangeable conditionally on $(W_i)_{i \in \Lambda_{d,M}}$. In particular, for all $k \in \{1, \ldots, N\}$ and $\sigma \in S_N$

$$
\mathbb{E} \left( H \left( X_{\sigma(1)}, \ldots, X_{\sigma(k)} \right) \right) = \mathbb{E} \left( H \left( X_{\{1, \ldots, k\}} \right) \right), \quad (4.8)
$$

and we obtain by (4.7)

$$
\mathbb{E} \left( I^c \left( X_{N,M} \right) \right) = 2 \sum_{k=1}^N c_k^N \left( \begin{array}{c} N \\ k \end{array} \right) \mathbb{E} \left( H \left( X_{\{1, \ldots, k\}} \right) \right) - \mathbb{E} \left( H \left( X_{N,M} \right) \right). \quad (4.9)
$$

**Lemma 4.3** Let $y \in \Lambda_{d,k}$, $k \in \{1, \ldots, N\}$ and set

$$
\nu(y) := \sum_{z \in \Lambda_{d,N-k}} \mu_{N,M}(y, z). \quad (4.10)
$$

Then $d^M \cdot \nu(y)$ is a binomial variable with parameters $(d^M, d^{-k})$.

**Proof** Notice that, conditionally on $(W_i)_{i \in \Lambda_{d,M}}$, $X_{\{1, \ldots, k\}} = (X_1, \ldots, X_k) \in \Lambda_{d,k}$ has distribution

$$
\mathbb{P} \left( X_{\{1, \ldots, k\}} = y \mid (W_i)_{i \in \Lambda_{d,M}} \right) = \sum_{z \in \Lambda_{d,N-k}} \mu_{N,M}(y, z)
$$

$$
= d^{-M} \sum_{i \in \Lambda_{d,M}} \sum_{z \in \Lambda_{d,N-k}} 1_{((y,z) = W_i)},
$$

where $y \in \Lambda_{d,k}$ and $(y, z) \in \Lambda_{d,k} \times \Lambda_{d,N-k} = \Lambda_{d,N}$. For fixed $y \in \Lambda_{d,k}$, the family

$$
T_i := \sum_{z \in \Lambda_{d,N-k}} 1_{((y,z) = W_i)} , \quad i \in \Lambda_{d,M}
$$

is an i.i.d. family of Bernoulli variables with parameter $d^{-k}$. Indeed, if $\Pi_{N\mapsto k} : \Lambda_{d,N} \mapsto \Lambda_{d,k}$ is the natural projection, then the law of $\Pi_{N\mapsto k}(W_i)$ is uniform on $\Lambda_{d,k}$, so that

$$
\mathbb{P}(T_i = 1) = \mathbb{P}(\Pi_{N\mapsto k}(W_i) = y) = d^{-k}.
$$
Hence, for all \( y \in \Lambda_{d,k} \), \( d^M \cdot v(y) = \sum_{i \in \Lambda_{d,M}} T_i \) is the sum of \( d^M \) independent Bernoulli variables with parameter \( d^{-k} \), i.e. a binomial variable with parameters \((d^M, d^{-k})\). \( \square \)

Let us denote from now on by \( B_k \) a binomial variable with parameters \((d^M, d^{-k})\). Set \[ \varphi(x) := -\frac{x \log x}{\log d}, \quad \forall x > 0, \quad \varphi(0) := 0. \]

Notice that the function \( \psi(x) := -(1 + x) \log(1 + x) + x + \frac{x^2}{2} \) satisfies
\[
\psi(0) = \psi'(0) = 0, \quad \psi''(x) \geq 0, \quad \forall x \geq 0,
\]
so that \( \psi(x) \geq 0 \) for \( x \geq 0 \). Moreover, \( \varphi(1 + x) \geq 0 \) if \( x \in [-1, 0] \). Hence, for all \( x \geq -1 \),
\[
\varphi(1 + x) \geq -\mathbb{1}_{(x > 0)} \frac{x + x^2/2}{\log d}. \tag{4.11}
\]

Now, by (4.10)
\[
H \left( X_{1,\ldots,k}^{N,M} \right) = -\sum_{y \in \Lambda_{k,N}} v(y) \log v(y),
\]
then we obtain by Lemma 4.3 that
\[
h_k := \frac{1}{\log d} \mathbb{E} \left( H \left( X_{1,\ldots,k}^{N,M} \right) \right) = d^k \mathbb{E} \left( \varphi \left( B_k d^{-M} \right) \right). \tag{4.12}
\]

**Lemma 4.4** Away from \( k = M \), the entropy is nearly linear or constant:
\[
\begin{align*}
k - 2d^{k-M} &\leq h_k \leq k, \quad k = 1, \ldots, M, \\
M - d^{M-k} &\leq h_k \leq M, \quad k = M + 1, \ldots, N.
\end{align*}
\]

**Proof** The upper bounds are easy. Indeed, for \( k \leq M \) one uses (2.7), while for \( k > M \) we notice that the support of \( \mu_{N,M}^1 \) has cardinality at most \( d^M \), and apply (2.6) to conclude. We notice now that for any \( 0 \leq k \leq M \),
\[
\begin{align*}
h_k &= k + \mathbb{E} \left( \varphi \left( B_k d^{k-M} \right) \right) = M + d^{k-M} \mathbb{E} \left( \varphi(B_k) \right) \tag{4.13}
\end{align*}
\]
These identities follow from the formulae: \( \mathbb{E}(B_k) = d^{M-k} \), \( \varphi(d^{-j}) = j d^{-j} \) and \( \varphi(\alpha x) = \alpha \varphi(x) + x \varphi(\alpha) \) for \( \alpha > 0 \) applied to \( \varphi(B_k d^{k-M} \cdot d^{-k}) \) and \( \varphi(B_k \cdot d^{-M}) \).
Recall that $B_k$ is binomial with parameters $(d^M, d^{-k})$. Then $\mathbb{E}(B_k) = d^{M-k}$ and $\text{Var}(B_k) = d^M d^{-k} (1 - d^{-k})$. If we define $J_k := B_k \cdot d^{k-M} - 1$ then we obtain

$$\mathbb{E}(J_k^2) = d^{2(k-M)} \text{Var}(B_k) = d^{k-M} - d^{-M} \leq d^{k-M}.$$ 

Hence, using (4.11) we get

$$\mathbb{E}(\varphi(B_k \cdot d^{k-M})) = \mathbb{E}(\varphi(1 + J_k)) \geq -\frac{1}{\log d} \mathbb{E}\left(1_{(J_k > 0)} \left(J_k + \frac{J_k^2}{2}\right)\right) \geq -\mathbb{E}(|J_k| + J_k^2) \geq -\left(\sqrt{\mathbb{E}(J_k^2)} + \mathbb{E}(J_k^2)\right) \geq -2d^{k-M},$$

since $\mathbb{E}(|J_k|) \leq \sqrt{\mathbb{E}(J_k^2)}$ by Cauchy-Schwartz and $d^{k-M} \leq d^{k-M} \leq 1$. By (4.13) we obtain the desired lower bound for $k \leq M$.

Let us consider now the regime $k > M$. We have

$$h_k = d^k \mathbb{E}(\varphi(B_k d^{-M})) = M + d^{k-M} \mathbb{E}(\varphi(B_k)).$$

If $B_k \in \{0, 1\}$ then $\varphi(B_k) = 0$. Note that (4.11) implies that $\varphi(B_k) \geq -(B_k - 1)^2 - (B_k - 1)$ as the right hand side is zero whenever $B_k = 0, 1$ and less than $-(B_k - 1)^2/2 - (B_k - 1)$ otherwise. Thus,

$$d^{k-M} \mathbb{E}(\varphi(B_k)) \geq -d^{k-M} \mathbb{E}\left((B_k - 1)^2 + (B_k - 1)\right) = -d^{k-M} \mathbb{E}\left(B_k^2 - B_k\right)$$

$$= -d^{k-M} \left(d^{M-k} + d^{2(M-k)} - d^{M-2k} - d^{M-k}\right) \geq -d^{M-k}.$$

By (4.13) we obtain the lower bound for $k > M$. $\square$

### 4.3 Estimation of the expected intricacy

**Lemma 4.5** Let $x \in ]0, 1[, M := \lfloor xN \rfloor$ and $\alpha := d^{-1/2} < 1$. For all $N \geq 2$, using the notation (2.1),

$$-2d^{-(N-M)} \leq \mathbb{E}(2(D_N \wedge M) - M) - \frac{\mathbb{E}(I^c(X^{N,M}))}{\log d} \leq \mathbb{E}\left(4\alpha^{D_N-M}\right).$$

**Proof** By (4.9), (4.12) and (2.2),

$$\frac{\mathbb{E}(I^c(X^{N,M}))}{\log d} = 2 \sum_{k=1}^{N} c_k^N \binom{N}{k} h_k - h_N = 2 \mathbb{E}(h_{D_N}) - h_N.$$
So,
\[ E \left( I^c (X^{N,M}) \right) \frac{\log d}{-E (2 (D_N \wedge M) - M)} = 2E (h_{D_N} - D_N \wedge M) + M - h_N. \]

We conclude by Lemma 4.4. \(\square\)

**Lemma 4.6** Let \( x \in ]0, 1[ \) and \( M := \lfloor xN \rfloor \). Then for \( \alpha := d^{-1/2} < 1 \) we have
\[ \sum_{N \geq 1} \frac{1}{N} E \left( \alpha^{\lfloor D_N - M \rfloor} \right) < +\infty. \]

**Proof** First of all, we show that we can reduce to the case \( \lambda_c([0, 1]) = 0 \). Let us set \( \lambda^0 := \frac{1}{2} (\delta_0 + \delta_1) \). Then \( \lambda_c \) can be written as \( \theta \lambda^0 + (1 - \theta) \lambda_\pi \) where \( \theta = \lambda_c([0, 1]) \in [0, 1] \) and \( \lambda_\pi \) is a probability measure on \( ]0, 1[ \) associated with a system of coefficients \( \bar{c} \). By (2.2) and the expression for the coefficients in Def. 1.1, we have that
\[ P(\lfloor D_N = k \rfloor) = \binom{N}{k} c^N_k = \frac{\theta}{2} P(k \in \{0, N\}) + (1 - \theta) \binom{N}{k} \bar{c}^N_k, \text{ } k = 0, \ldots, N. \]

Therefore
\[ E \left( \alpha^{\lfloor D_N - M \rfloor} \right) = \frac{\theta}{2} (\alpha^{\lfloor M \rfloor} + \alpha^{\lfloor N - M \rfloor}) + (1 - \theta) \sum_{k=0}^{N} \binom{N}{k} \bar{c}^N_k \alpha^{\lfloor k - M \rfloor}. \]

It is clear that
\[ \sum_{N \geq 1} \frac{1}{N} \left( \alpha^{\lfloor M \rfloor} + \alpha^{\lfloor N - M \rfloor} \right) < +\infty. \]

Thus we can assume that \( \lambda_c([0, 1]) = 0 \), i.e. \( P(W_c \notin \{0, 1\}) = 1 \). We set now \( L := \lfloor \frac{N}{2} \rfloor \) and we claim that for all \( 0 \leq k \leq N \) we have \( P(D_N = k) \leq P(D_N = L) \). Indeed, \( \binom{N}{k} \leq \binom{N}{L} \), and since \( W_c \) and \( 1 - W_c \) are equal in law, we see that
\[ 2c^N_k = 2E \left( W^k_c (1 - W_c)^{N-k} \right) = E \left( W^k_c (1 - W_c)^{N-k} + W_c^{N-k} (1 - W_c)^k \right) \]

which, as \( W_c \neq 0, 1 \) a.s., attains its maximum for \( k = L \). Then \( N - L \geq L \geq \frac{N}{2} - 1 \) and we obtain for all \( 0 \leq k \leq N \)
\[ \binom{N}{k} c^N_k \leq \binom{N}{L} c^N_L = \frac{N!}{L!(N-L)!} E \left( W^L_c (1 - W)^{N-L} \right) \leq \frac{N!}{L!(N-L)!} 4^{N-L} \leq \frac{N!}{L!(N-L)!} 2^{-N+2}. \]
By Stirling’s formula \( n! = \sqrt{2\pi n} (n/e)^n (1 + \mathcal{O}(1/n)) \), there is a constant \( C \geq 0 \) such that
\[
2^{-N} \frac{N!}{L! (N-L)!} \leq C N^{-\frac{1}{2}}, \quad N \to +\infty, \quad L = \left\lfloor \frac{N}{2} \right\rfloor.
\]

Then, we obtain for some constants \( C_1, C_2 \geq 0 \)
\[
\mathbb{E} \left( \alpha \mid D_{N-M} \right) = \sum_{k=0}^{N} \binom{N}{k} c_k N^{k} \alpha^{k-M} \leq C_1 N^{-\frac{1}{2}} \sum_{k=0}^{N} \alpha^{k-M} \leq C_1 N^{-\frac{1}{2}} 2 \sum_{k=0}^{+\infty} \alpha^{k} = C_2 N^{-\frac{1}{2}},
\]
and the proof is finished. \( \square \)

**Proof of Proposition 4.1** Let \( x \in ]0, 1[ \) and \( M := \lfloor xN \rfloor \geq 1 \) \((N \text{ is large})\). By Lemma 4.4 for \( k = N \)
\[
\mathbb{E} \left( \left| \frac{M}{N} - \frac{H(X^N, M)}{N \log d} \right| \right) = \mathbb{E} \left( \frac{M}{N} - \frac{H(X^N, M)}{N \log d} \right) = \mathbb{E} \left( \frac{M - h_N}{N} \right) \leq \frac{d^{M-N}}{N} \leq d^{-N(1-x)}.
\]

Thus,
\[
\sum_{N \geq 1} \mathbb{E} \left( \left| \frac{M}{N} - \frac{H(X^N, M)}{N \log d} \right| \right) < +\infty,
\]
therefore a.s.
\[
\sum_{N \geq 1} \left| \frac{M}{N} - \frac{H(X^N, M)}{N \log d} \right| < +\infty,
\]
and in particular a.s.
\[
\lim_{N \to +\infty} \left| \frac{M}{N} - \frac{H(X^N, M)}{N \log d} \right| = 0.
\]
Setting \( x_N := \frac{H(X^N, M)}{N \log d} \), we have obtained
\[
\lim_{N \to +\infty} x_N = \lim_{N \to +\infty} \left( \frac{\lfloor xN \rfloor}{N} \right) = x
\]
a.s. and in $L^1$, namely we have proven (4.5). Now, by Proposition 2.1, this gives

$$|i_N^c(x_N) - i^c(x)| \leq N^{-1/2} + |x_N - x| \to 0$$

again a.s. and in $L^1$. On the other hand, by (1.2) and by Lemmas 4.5, 4.6

$$\sum_{N \geq 1} \mathbb{E} \left( \left| i_N^c(x_N) - \frac{I_c^c(X^N, M)}{N \log d} \right| \right) \leq C \sum_{N \geq 1} N^{-3/2} < \infty.$$ 

Arguing as above, it follows that $\frac{I_c^c(X^N, M)}{N \log d} \to i^c(x)$ a.s. and in $L^1$. This proves (4.4) and concludes the proof of Proposition 4.1.  

□

5 Proof of the main results

We now collect our results to prove Theorem 1.2 and Propositions 1.4 and 1.5. We consider some non-null intricacy $I_c^c$. Let $\lambda_c$ be the associated probability measure on $[0, 1]$ as in Def. 1.1. Recall from Eq. (4.1) the entropy-intricacy function $I_c^c(d, x)$ and the functions $i^c(x)$ and $i_N^c(x)$ from eq. (1.2).

The asymptotic behavior of the intricacy, i.e., Theorem 1.2, immediately follows from Propositions 2.1 and 4.1. The convergence of entropy profiles requires a little additional work.

Proof of Proposition 1.4 Let $(X_N^N)_{N \geq 1}$ be an approximate maximizer and set, for simplicity of notation:

$$x_N := \frac{H(X_N^N)}{N \log d}, \quad I_N := \frac{I_c^c(X_N^N)}{N \log d}.$$ 

If $1/2$ is in the support of $\lambda_c$, then by the first point of Proposition 2.1, it is the unique point where $i^c(x)$ achieves its maximum. Then, Theorem 1.2 implies that no $x \neq 1/2$ can be an accumulation point of $x_N$, $N \geq 1$. Thus an approximate maximizer is an approximate $1/2$-maximizer, proving the first half of (1.6). The second half will follow from the claim (2) of the proposition to which we now turn.

By definition of approximate $x$-maximizers, $x_N \to x$ and $I_N \to i^c(x)$. Using point (2) of Proposition 2.1, it follows that $|I_N - i_N^c(x)| \to 0$ and by (2.4) we have $\|h_{X_N} - h^*_x\|_{c,N} \to 0$. Notice now that for any $K$-Lipschitz function $f : [0, 1] \to \mathbb{R}$, by (2.8)

$$|\mathbb{E}(f(\beta_N)) - \mathbb{E}(f(W_c))| \leq \frac{K}{\sqrt{N}}.$$ 

As entropy profiles are 1-Lipschitz, we obtain

$$\int |h_{X_N} - h^*_x| d\lambda_c \to 0.$$ 

As all functions $(h_{X_N} - h^*_x)_N$ are 2-Lipschitz, (1.7) follows by a routine argument.
Assuming now \( x \in \text{supp}(\lambda_c) \), \( \lim_{N \to \infty} h_{X^N}(x) = h^*_x(x) = x \). On the one hand, as \( h_{X^N}(0) = 0 \) and \( h_{X^N} \) is 1-Lipschitz, it follows that the convergence \( \lim_{N \to \infty} h_{X^N}(t) = h^*_x(t) = t \) occurs for all \( t \in [0, x] \). On the other hand, all \( h_{X^N} \) being non-decreasing, \( h_{X^N}(x) = x \leq h_{X^N}(t) \leq h_{X^N}(1) \to x \). Hence the previous convergence occurs for all \( x \in [0, 1] \), proving (1.8).

Let us prove point (3). By (1.8) the profiles \( h_{X^N} \) converge to \( h^*_x \) uniformly on \([0, 1]\). Let \( T' \) be any other intricacy. By uniform convergence we have \( \|h_{X^N} - h^*_x\|_{c', N} \leq \sup_{[0, 1]} |h_{X^N} - h^*_x| \to 0 \). By (2.4) and Proposition 2.1-(2)

\[
\frac{T'(X)}{N \log d} = i_{N}^{c'}(x) - \|h_{X} - h^*_x\|_{c', N} \to i^{c'}(x), \quad N \to +\infty.
\]

By Theorem 1.2, \( i^{c'}(x) = T'(d, x) \) and therefore \((X^N)_N\) is an approximate x-maximizer for \( T' \).

Point (4) easily follows. \( 1/2 \in \text{supp}(\lambda_c) \), so any approximate maximizer for this intricacy, is an approximate \( 1/2 \)-maximizer for it and therefore for any intricacy by Proposition 2.1. By point (2) of Theorem 1.2, it is an approximate maximizer for this intricacy.

Finally, we analyze the consequences for approximate x-maximizers.

**Proof of Proposition 1.5** Let \( y \in ]0, 1[ \). By (1.8), \( h_{X^N}(y) \to h^*_x(y) = x \land y \) as \( N \to +\infty \). By Definition 1.3 of \( h_{X^N} \), we obtain, setting \( k_N := \lfloor yN \rfloor \),

\[
\frac{1}{N} \sum_{|S|=k_N} h^*_x(y) - \frac{H(X_S)}{N \log d} = h^*_x(y) - \frac{1}{N} \sum_{|S|=k_N} \frac{H(X_S)}{N \log d} \to h^*_x(y) - h_{X^N}(y) \to 0,
\]

since all terms in the sum are non-negative by Lemma 2.2. Let \( Z_N \), defined on \((\Omega, \mathcal{F}, \mathbb{P})\), be a random subset of \([1, \ldots, N]\) defined by

\[
\mathbb{P}(Z_N = S) = \frac{1}{N}, \quad \text{if } |S| = k_N.
\]

Then the above formula can be rewritten as follows

\[
\lim_{N \to +\infty} \mathbb{E} \left( \left| h^*_x(y) - \frac{H(X_{Z_N})}{N \log d} \right| \right) = 0.
\]

Since \( L^1 \) convergence implies convergence in probability, we obtain

\[
\lim_{N \to +\infty} \mathbb{P} \left( \left| h^*_x(y) - \frac{H(X_{Z_N})}{N \log d} \right| > \varepsilon h^*_x(y) \right) = 0, \quad \forall \varepsilon > 0.
\]

This readily implies the Proposition, by recalling that \( H(X^N | X^N_S) = H(X^N) - H(X^N_S) \) and that \( \frac{H(X^N)}{N \log d} \to x \) by assumption. \( \square \)
Acknowledgments  We thank the referee for a number of useful suggestions.

References

1. Barnett, L., Buckley, C.L., Bullock, S.: Neural complexity and structural connectivity. Phys. Rev. E 79, 051914 (2009)
2. Buzzi, J., Zambotti, L.: Mean mutual information and symmetry breaking for finite random fields. Ann. Inst. H. Poincaré (B), Probab. Stat. (2011, to appear)
3. Cover, T., Thomas, J.: Elements of Information Theory. Wiley, New York (2006)
4. Edelman, G., Gally, J.: Degeneracy and complexity in biological systems. Proc. Natl. Acad. Sci. USA 98, 13763–13768 (2001)
5. Greven, A., Keller, G., Warnecke, G.: Entropy. Princeton University Press, USA (2003)
6. Holthausen, K., Breidbach, O.: Analytical description of the evolution of neural networks: learning rules and complexity. Biol. Cybern. 81, 169–176 (1999)
7. Krichmar, J., Nitz, D., Gally, J., Edelman, G.: Characterizing functional hippocampal pathways in a brain-based device as it solves a spatial memory task. Proc. Natl. Acad. Sci. USA 102, 2111–2116 (2005)
8. Makarychev, K., Makarychev, Y., Romashchenko, A., Vereshchagin, N.: A new class of non-Shannon-type inequalities for entropies, Communications in Information and Systems, vol. 2, No. 2, pp. 147–166. International Press, USA (2002)
9. Mézard, M., Montanari, A.: Information, Physics, and Computation. Oxford University Press, USA (2009)
10. Peled, R., Yadin, A., Yehudayoff, A.: The maximal probability that k-wise independent bits are all 1. Random Structures and Algorithms (to appear). http://arxiv.org/abs/0801.0059
11. Seth, A., Izhikevich, E., Reeke, G., Edelman, G.: Theories and measures of consciousness: an extended framework. Proc. Natl. Acad. Sci. USA 103, 10799–10804 (2006)
12. Seth, A.K.: Models of consciousness. Scholarpedia 2(1), 1328 (2007)
13. Shanahan, M.P.: Dynamical complexity in small-world networks of spiking neurons. Phys. Rev. E 78, 041924 (2008)
14. Sporns, O., Tononi, G., Edelman, G.: Connectivity and complexity: the relationship between neuro-anatomy and brain dynamics. Neural Netw. 13(8–9), 909–922 (2000)
15. Sporns, O.: Networks analysis, complexity, and brain function. Complexity 8, 56–60 (2002)
16. Sporns, O.: Complexity. Scholarpedia 2(10), 1623 (2007)
17. Tononi, G., Sporns, O., Edelman, G.: A measure for brain complexity: relating functional segregation and integration in the nervous system. Proc. Natl. Acad. Sci. USA 91, 5033–5037 (1994)
18. Tononi, G., Sporns, O., Edelman, G.: A complexity measure for selective matching of signals by the brain. Proc. Natl. Acad. Sci. USA 93, 3422–3427 (1996)
19. Tononi, G., Sporns, O., Edelman, G.: Measures of degeneracy and redundancy in biological networks. Proc. Natl. Acad. Sci. USA 96, 3257–3262 (1999)