# Entropy for random group actions

Robert Burton[1], Karma Dajani[2] and Ronald Meester[2]

[1] Department of Mathematics
Oregon State University
Corvallis, OR 97331
U.S.A.
burton@math.orst.edu

[2] Department of Mathematics
University of Utrecht
P.O. Box 80.010
3508 TA Utrecht
The Netherlands
dajani@math.ruu.nl
meester@math.ruu.nl

**Abstract**

We consider the entropy of systems of random transformations, where the transformations are chosen from a set of generators of a $\mathbf{Z}^d$ action. We show that the classical definition gives unsatisfactory entropy results in the higher-dimensional case, i.e. when $d \geq 2$. We propose a definition of the entropy for random group actions which agrees with the classical definition in the one-dimensional case, and which gives satisfactory results in higher dimensions. This definition is based on the fibre entropy of a certain skew product. We identify the entropy by an explicit formula which makes it possible to compute the entropy in certain cases.

# 1 Random transformations and entropy

Let $(X, \mathcal{B}, \rho)$ be a probability space. In deterministic ergodic theory one is usually concerned with the study of a single ergodic measure-preserving transformation $S$ on $(X, \mathcal{B}, \rho)$. For example, the concept of entropy, as was defined by Kolomogorov and Sinai, plays an important role in the classification of ergodic systems (see e.g. Walters (1982)).

In this paper we are concerned with the entropy of random transformations. In Section 2, we consider the case where the dynamics or the random evolution of the system is generated by independent applications of transformations chosen at random according to some probability distribution. The concept and properties of such random systems have already been defined and studied (see Kifer (1986)). However, we show that the classical setup gives unsatisfactory entropy results when the set of transformations consists of generators of a higher-dimensional group. In this case, it is more natural to compare the random system with the deterministic group action. In such systems we do not think of picking transformations randomly one at a time, but rather according to a stationary and ergodic distribution. In Section 3, we develop a notion of random entropy which is based on this idea, and which gives satisfactory results in any dimension. In Section 4 we give an explicit formula for the calculation of the entropy of random group actions.

In the remaining part of this section we recall the classical definitions and results of random transformations, random entropy and random generators. All these are found in Kifer (1986) but we include them for the convenience of the reader.

Consider a probability space $(X, \mathcal{B}, \rho)$ and let $\mathcal{F}$ be a set of transformations acting on $X$. The set $\mathcal{F}$ is assumed to possess a measure structure such that the map from $\mathcal{F} \times X$ to $X$ defined by $(f, x) \to f(x)$ is measurable. Let $m$ be a probability measure on $\mathcal{F}$. Introduce a new probability space $(\Omega, \mu)$, where $\Omega = \mathcal{F}^{\mathbf{N}^+}$, $\mu = m^{\mathbf{N}^+}$, where $\mathbf{N}^+$ denotes the positive integers. The $\sigma$-algebra on $\Omega$ is the product $\sigma$-algebra. Thus, an element $\omega \in \Omega$ is a sequence of transformations $\omega = (f_1, f_2, \ldots)$. We denote the shift in $\Omega$ by $\sigma$: $(\sigma\omega)_i = f_{i+1}$, $i = 1, 2, \ldots$

All quantities defined below will depend on $m$, but we won't make this dependence explicit in the notation. Let $P$ be the operator acting on bounded functions of $X$ as follows:

$$P(g)(x) = \int_{\mathcal{F}} g(f(x)) dm(f).$$

The adjoint operator $P^*$ gives a new measure $P^*\rho$ on $X$ in the following way: For any measurable subset $G$ of $X$,

$$P^*\rho(G) = \int_X \int_{\mathcal{F}} \chi_G(f(x)) dm(f) d\rho(x),$$

where $\chi_G$ denotes the indicator function of $G$.

**Definition 1.1:** *The measure $\rho$ is said to be $P^*$-invariant if $P^*\rho = \rho$.*

**Definition 1.2:** *The measure $\rho$ is said to be $m$-invariant if $\rho(f^{-1}G) = \rho(G)$ for $m$ almost every $f$ and for every measurable $G \subset X$.*

We shall use the following notation:

(i) If $\xi$ is a finite partition of $X$, then $H(\xi) = H_\rho(\xi)$ denotes the entropy of $\xi$, i.e. $H_\rho(\xi) = -\sum_{A \in \xi} \rho(A) \log \rho(A)$.

(ii) For $\omega = (f_1, f_2, \ldots) \in \Omega$, we let ${}^i f(\omega) = f_i \circ f_{i-1} \circ \cdots \circ f_1$ where $\circ$ denotes composition. Note that $i = 0$ corresponds to the identity operator.

**Theorem 1.1:** *Suppose $\rho$ is $P^*$-invariant and let $\xi$ be a finite partition of $X$. Then*

$$h_\rho(m,\xi) := \lim_{n\to\infty} \frac{1}{n} \int_\Omega H_\rho(\vee_{i=0}^{n-1}({}^i f(\omega))^{-1}\xi) d\mu(\omega)$$

*exists. If $\rho$ is $m$-invariant, then*

$$h_\rho(m,\xi) = \lim_{n\to\infty} \frac{1}{n} H_\rho(\vee_{i=0}^{n-1}({}^i f(\omega))^{-1}\xi)$$

$\mu$ a.e.

**Definition 1.3:** *The random entropy of the system $(X,\rho;\mathcal{F},m)$ is defined by*

$$h_\rho(m) = \sup_\xi h_\rho(m,\xi)$$

*where the supremum is taken over all finite partitions of $X$.*

**Remark:** The deterministic entropy $h_\rho(S)$ (or $h_\rho(S,\xi)$) of a single measure preserving transformation $S$ can be viewed as a special case of the random entropy defined above when $\mu$ is concentrated on $S$.

**Definition 1.4:** *A finite partition $\xi$ of $X$ is said to be a random generator for $(X,\rho;\mathcal{F},m)$ if for $\mu$ a.e. $\omega \in \Omega$, $\vee_{i=0}^{\infty}({}^i f(\omega))^{-1}\xi$ generates the $\sigma$-algebra $\mathcal{B}$ on $X$, up to sets of $\rho$ measure zero.*

**Theorem 1.2:**

*(i) If $\xi$ is a random generator for $(X,\rho;\mathcal{F},m)$, then*

$$h_\rho(m) = h_\rho(m,\xi).$$

*(ii) If $\xi_1 \preceq \xi_2 \preceq \cdots$ is an increasing sequence of finite partitions generating the $\sigma$-algebra $\mathcal{B}$ on $X$ (i.e. $\vee_{i=1}^{\infty}\xi_i$ generates $\mathcal{B}$ up to sets of measure zero), then*

$$h_\rho(m) = \lim_{n\to\infty} h_\rho(m,\xi_n).$$

The random entropy defined above can be related to the fibre entropy of a skew product as follows. Consider

$$T : \Omega \times X \to \Omega \times X$$

defined by $T(\omega, x) = (\sigma\omega, f_1 x)$, where $\omega = (f_1, f_2, \ldots)$. One can now show that when $\rho$ is $P^*$-invariant,

$$h_{\mu\times\rho}(T) = h_\mu(\sigma) + h_\rho(m).$$

The first term in the right hand side is the entropy obtained by observing which transformations are chosen. The second term is the entropy obtained by the action of these random transformations. This point of view will be used later to generalise random entropy to higher dimensions.

## 2 Randomly chosen generators of group actions

First we consider the one-dimensional case. The underlying space $X$ is equipped with a measure $\rho$. Let $S$ denote an invertible $\rho$-invariant transformation. The space $\Omega := \mathcal{F}^{\mathbf{N}^+} = \{S^{-1}, S\}^{\mathbf{N}^+}$ is given the product measure $\mu$ which assigns probability $p$ to $S$ and probability $q = 1 - p$ to $S^{-1}$. Note that $\rho$ is $m$-invariant since $\rho$ is $S$-invariant. Thus for any finite partition $\xi$ of $X$, we have that $\mu$ a.e.

$$h_\rho(m,\xi) = \lim_{n\to\infty} \frac{1}{n} H_\rho(\vee_{i=0}^{n-1} ({}^i f(\omega))^{-1}\xi).$$

**Lemma 2.1:** *For any finite partition $\xi$ of $X$, we have*

$$h_\rho(m,\xi) = |p-q|h_\rho(S,\xi).$$

**Proof:** For $\omega \in \Omega$, define

$$X_i(\omega) = \begin{cases} 1 & \text{if } \omega_i = S, \\ -1 & \text{if } \omega_i = S^{-1}. \end{cases}$$

Let $U_0(\omega) = 0$ and for $n \geq 1$, let $U_n(\omega) = \sum_{i=1}^{n} X_i(\omega)$. For $n \geq 0$, let $K_n(\omega)$ be the set of distinct values of $U_0, U_1, \ldots, U_n$, and let $R_n(\omega) = |K_n(\omega)|$ be the cardinality of $K_n(\omega)$, that is, $R_n(\omega)$ is the range of $U_0(\omega), U_1(\omega) \ldots, U_n(\omega)$. Note that $K_n(\omega)$ is a subset of $\{-n, -(n-1), \ldots, (n-1), n\}$ consisting of $R_n(\omega)$ consecutive integers. It is well known (see Spitzer (1976)) that $\lim_{n\to\infty} \frac{R_n(\omega)}{n} = |p-q|$ $\mu$ a.e., say for $\omega \in A$ where $\mu(A) = 1$. Moreover, since $\rho$ is $m$-invariant, for any finite partition $\xi$ of $\Omega$ and any $\omega \in A$, we have

$$\begin{aligned} h_\rho(m,\xi) &= \lim_{n\to\infty} \frac{1}{n} H_\rho(\vee_{i=0}^{n-1}({}^i(f(\omega))^{-1}\xi)) \\ &= \lim_{n\to\infty} \frac{R_n(\omega)}{n} \frac{1}{R_n(\omega)} H_\rho(\vee_{i=0}^{R_n(\omega)-1} S^{-i}\xi) \\ &= |p-q|h_\rho(S,\xi). \end{aligned}$$

□

**Theorem 2.1:** *Suppose that $h_\rho(S) < \infty$. Then the random entropy of $(X, \rho; \mathcal{F}, m)$ is given by*

$$h_\rho(m) = |p-q|h_\rho(S).$$

**Proof:** Since the deterministic entropy $h_\rho(S) < \infty$ one can find, using Krieger's theorem (Krieger (1970)), a finite partition $\xi$ such that $h_\rho(S) = \lim_{n\to\infty} h_\rho(S,\xi_n)$ where $\xi_n = \vee_{i=-n}^{n} S^{-i}\xi$. Thus from Lemma 2.1 and Theorem 1.2 (ii) we see that

$$\begin{aligned} h_\rho(m) &= \lim_{n\to\infty} h_\rho(m,\xi_n) \\ &= \lim_{n\to\infty} |p-q| h_\rho(S,\xi_n) \\ &= |p-q| h_\rho(S). \end{aligned}$$

□

Next we show that Krieger's theorem is no longer true in the current setting. This is more than anything due to the fact that we consider only one-sided random choices of transformation. In the next section we shall work with the two-sided version.

**Theorem 2.2:** *Let $(X,\rho)$ be a probability space and suppose that $S : X \to X$ is invertible, $\rho$ invariant and satisfies $h_\rho(S) > 0$. Let $\mathcal{F} := \{S, S^{-1}\}$ with measure $m$ which assigns probability $p$ to $S$ and $q = 1-p$ to $S^{-1}$. If $p \neq q$, then $(X,\rho;\mathcal{F},m)$ has no random generator.*

**Proof:** If $\xi$ is a random generator, then for $\mu$ a.e. $\omega$, $\vee_{i=0}^{\infty}({}^i f(\omega))^{-1}\xi$ generates $\mathcal{B}$. Assume with no loss of generality that $p > q$. Then the set

$$A = \{\omega \in \Omega : \min_n U_n(\omega) = 0\}$$

has positive $\mu$ measure, with $U_n(\omega)$ as defined in the proof of Lemma 2.1. Let $\omega \in A$ be such that $\vee_{i=0}^{\infty}({}^i f(\omega))^{-1}\xi$ generates $\mathcal{B}$, and set $M = \max\{n : U_n(\omega) = 0\}$. Then for any $n \geq M$, we have

$$\vee_{i=0}^{n}({}^{i}f(\omega))^{-1}\xi \preceq \vee_{j=0}^{n}S^{-j}\xi.$$

Since the sequence $\vee_{i=0}^{n}({}^{i}f(\omega))^{-1})\xi$ generates $\mathcal{B}$, it follows that $\{\vee_{j=0}^{n}S^{-j}\xi\}$ also generates $\mathcal{B}$. This shows that $\xi$ is a one-sided generator for $S$ and so $S$ must have zero entropy (see Walters (1982), Corollary 4.18.1), which is a contradiction since $h_{\rho}(S) > 0$. □

The situation in the higher-dimensional case is quite different. To explain this we specialise to the case where $X = \{0,1\}^{\mathbf{Z}^2}$, $\rho$ is product measure and $\mathcal{F} = \{S^{-1}, S, T^{-1}, T\}$, where $S$ and $T$ denote the left- and downwards shifts respectively. We show that the random entropy is either 0 or $+\infty$ depending on whether $m$ is symmetric or not. (We say that $m$ is *symmetric* if $m(S) = m(S^{-1})$ and $m(T) = m(T^{-1})$, otherwise $m$ is called *nonsymmetric*.) This is the unsatisfactory fact referred to in the first section.

**Theorem 2.3:** *If $m$ is symmetric, then $h_{\rho}(m) = 0$.*

**Proof:** For any $\omega \in \Omega = \mathcal{F}^{\mathbf{N}^{+}}$, define

$$X_i(\omega) = \begin{cases} (1,0) & \text{if } \omega_i = S, \\ (-1,0) & \text{if } \omega_i = S^{-1}, \\ (0,1) & \text{if } \omega_i = T, \\ (0,-1) & \text{if } \omega_i = T^{-1}. \end{cases}$$

Let $U_0(\omega) = (0,0)$ and for $n \geq 1$, let $U_n(\omega) = \sum_{i=0}^{n} X_i(\omega)$. For $n \geq 0$, let $K_n(\omega)$ be the set of distinct values of $U_0(\omega), U_1(\omega), \ldots, U_n(\omega)$ and let $R_n(\omega) = |K_n(\omega)|$. For $i = 0, 1$, set $P_i = \{x \in X : x_{0,0} = i\}$ and let $\xi = \{P_0, P_1\}$. Since $m$ is symmetric, the random walk $\{U_n : n \geq 0\}$ is recurrent, so that $\lim_{n\to\infty} \frac{R_n(\omega)}{n} = 0$ and $\cup_{n=0}^{\infty} K_n(\omega) = \mathbf{Z}^2$ $\mu$ a.e., say on a set $A$ of $\mu$ measure one (see Spitzer (1976)). The latter implies that $\xi$ is a random generator for $(X, \rho; \mathcal{F}, m)$. Moreover, since both $S$ and $T$ are measure preserving with respect to $\rho$, it follows that $\rho$ is $m$-invariant. Thus, if $\omega \in A$ we have

$$\begin{aligned} h_\rho(m) &= h_\rho(m, \xi) = \lim_{n\to\infty} \frac{1}{n} H_\rho(\vee_{i=0}^{n-1} ({}^i f(\omega))^{-1} \xi) \\ &= \lim_{n\to\infty} \frac{R_n(\omega)}{n} \frac{1}{R_n(\omega)} H_\rho(\vee_{(i,j)\in K_n(\omega)} S^{-i} T^{-j} \xi) \\ &= \lim_{n\to\infty} \frac{R_n(\omega)}{n} h_\rho(S, T) = 0, \end{aligned}$$

where $h_\rho(S, T)$ denotes the entropy of the deterministic $\mathbf{Z}^2$ action generated by $S$ and $T$. (Note that $h_\rho(S, T) < \infty$ because of the product structure of $\rho$.) □

**Theorem 2.4:** *If $m$ is nonsymmetric, then $h_\rho(m) = +\infty$.*

Before proving this, we remark that the result is not surprising: If $m$ is nonsymmetric, the two-dimensional random walk has essentially a one-dimensional range.

**Proof of Theorem 2.4:** Consider the partition $\xi$ as defined in the proof of Theorem 2.3. For $k \geq 1$, let

$$\xi_k = \vee_{i=-k}^{k} \vee_{j=-k}^{k} S^{-i} T^{-j} \xi.$$

Thus, $\xi_k$ is the partition that specifies coordinates $x_{i,j}$ for $-k \leq i, j \leq k$; note that $\xi_k$ is a generating sequence for $\mathcal{B}^2$. For $\omega \in \Omega$, define $X_i(\omega)$, $U_i(\omega)$, $K_n(\omega)$ and $R_n(\omega)$ as in the proof of Theorem 2.3. Let $R_n^k(\omega)$ be the number of new boxes specified by $\vee_{i=0}^{n-1} ({}^i f(\omega))^{-1} \xi_k$. Since $m$ is nonsymmetric we can assume without loss of generality that the random walk $U_n$ has a drift to the right. For $n$ large enough, we have $R_n^k(\omega) \geq \epsilon n$. Thus, for $\omega \in A$ and $n$ large enough

$$\begin{aligned}\frac{1}{n} H_\rho(\vee_{i=0}^{n-1} ({}^i f(\omega))^{-1} \xi_k) &\geq \frac{1}{n} R_n^k(\omega) H_\rho(\xi_k) \\ &\geq \frac{1}{n} \epsilon n \, H_\rho(\xi_k).\end{aligned}$$

From this it follows that for all $k$,

$$h_\rho(m, \xi_k) \geq \epsilon \, H_\rho(\xi_k),$$

which tends to infinity when $k \to \infty$. Therefore,

$$h_\rho(m) = \lim_{k \to \infty} h_\rho(m, \xi_k) = +\infty.$$

□

## 3 Entropy formalism for random group actions

Let us reconsider the one-dimensional case once more, doing things two-sided from now on (this is just for convenience and not important). We let $\Omega = \mathcal{F}^{\mathbf{Z}}$ where $\mathcal{F} = \{S, S^{-1}\}$, and identify an element of $\Omega$ with the sequence of powers of $S$. Each combination of $\omega = (\ldots, \omega_{-1}, \omega_0, \omega_1, \ldots)$ and an integer $n \in \mathbf{Z}$ gives rise to a transformation $S^{f(n,\omega)}$, where

$$f(n, \omega) = \begin{cases} \sum_{i=0}^{n-1} \omega_i, & n > 0, \\ 0, & n = 0, \\ -\sum_{i=1}^{n} \omega_{-i}, & n < 0, \end{cases}$$

which is clearly a cocycle for the $\mathbf{Z}$-action given by $S$. This idea will now be generalised to higher dimensions.

Let $A$ be a finite set containing at least two elements, and consider a $\mathbf{Z}^2$ action $\psi$ on $X = A^{\mathbf{Z}^2}$ generated by two commuting and invertible $\rho$ measure-preserving transformations $S$ and $T$. It is perhaps more natural to define a notion of randomness in such a way that the resulting system can be compared with the deterministic group action $\psi$ on $X$. We shall restrict ourselves to the two-dimensional case, but the reader should note that generalisation to higher dimensions causes no difficulty.

Let $\mathbf{Z}^2$ be the graph whose vertices are the integer points $(k, l)$, and which has edges between vertices that are distance one apart. We denote the edge set of this graph by $\mathbf{E}^2$. We want each vertex $(k, l)$ to be associated with a transformation of the form $S^m T^n$, $(m, n) \in \mathbf{Z}^2$. To do this succesfully, we have to make a few definitions.

A *path* $\pi$ is a finite sequence of edges $\pi = (e_1, e_2, \ldots, e_k)$ in such a way that the endpoint of $e_i$ is the starting point of $e_{i+1}$ for all appropriate $i$, where begin- and endpoint are defined in the obvious way. When $\pi$ travels through an edge $e$ in the upwards or right direction, we say that $e$ is a positive edge for $\pi$; when $\pi$ travels through $e$ downwards or to the left, we say that $e$ is a negative edge for $\pi$. An edge may be traversed more than once by $\pi$ and in such a case it could be both positive and negative for $\pi$. We define $\pi^+$ to be the set of positive edges for $\pi$ and $\pi^-$ the set of negative edges, noting that a given edge $e$ can appear more than once in either $\pi^+$, $\pi^-$ or both.

The immediate analogue of the one-dimensional case would be to somehow label each edge by either $(1,0), (-1,0)$, $(0,1)$ or $(0,-1)$ and think of the labels as corresponding to $S$, $S^{-1}$, $T$ and $T^{-1}$ respectively. It turns out however, that a richer theory can be developed when we allow the labels to be any element of $\mathbf{Z}^2$, and think of the label $(m, n)$ as corresponding to $S^m T^n$. We denote the label of the edge $e$ by $\ell(e)$.

We shall not allow all possible configurations of labels. The restriction we impose is that for any two vertices $x$ and $y$ and any path $\pi$ from $x$ to $y$, the vector

$$\sum_{e \in \pi^+} \ell(e) - \sum_{e \in \pi^-} \ell(e)$$

is independent of the choice of $\pi$, and only depends on $x$ and $y$. The reason for this restriction will become apparent soon. (Note that in the one-dimensional case this would be no restriction at all.) At first sight, it is not clear that many labellings are possible under this restriction. A little thought however, reveals that it is sufficient (and necessary) to require the following. Denote the unit vectors by $e_1$ and $e_2$. Take a vertex $x$ and write $f_1$ for the edge between $x$ and $x + e_1$, $f_2$ for the edge between $x + e_1$ and $x + e_1 + e_2$, $f_3$ for the edge between $x$ and $x + e_2$ and $f_4$ for the edge between $x + e_2$ and $x + e_2 + e_1$. For a labelling to be allowed we now need to require that

$$\ell(f_1) + \ell(f_2) = \ell(f_3) + \ell(f_4).$$

for all vertices $x$.

A closed subset of the full shift space $\{\mathbf{Z}^2\}^{\mathbf{E}^2}$ (in the usual topology in which two configurations are close whenever they agree on a large part of the space) which is invariant under translations is in general called a *subshift*. We shall denote the subshift defined above by $\Omega$, to indicate that this set plays a similar role as $\Omega$ in the previous sections. For $\omega \in \Omega$, the label of the edge $e$ in $\omega$ is sometimes denoted by $\omega_e$. The subshift $\Omega$ will play an important role in our formalism, so we shall first convince ourselves that $\Omega$

contains many elements, i.e. there are many allowed configurations of the edge labels.

**Proposition 3.1:** *The subshift* $\Omega$ *has uncountably many elements.*

**Proof:** Consider a two by two square, the eight outer edges of which are labelled as follows, starting in the lower left corner and going clockwise: $(1,0),(0,1),(1,0),\ (0,1),(0,1),(1,0),(0,1),(1,0)$. Now tile the plane with these squares, noting that the labelling is such that this is possible. Now for the four remaining edges in the interior of each square, there are two possibilities, namely either $(0,1)$ for the left and lower, and $(1,0)$ for the right and upper edges, or vice versa. This implies that $\Omega$ contains uncountably many elements. □

**Remark:** Using the same argument as in the proof of Proposition 3.1, one actually shows that $\Omega$ has positive topological entropy.

We continue with the definition of the cocycle $f : \mathbf{Z}^2 \times \Omega \to \mathbf{Z}^2$ by

$$f((m,n),\omega) = (f_1((m,n),\omega), f_2((m,n),\omega)) = \sum_{e\in\pi^+} \ell(e) - \sum_{e\in\pi^-} \ell(e),$$

where $\pi$ is any path from $(0,0)$ to $(m,n)$. We have seen that for any configuration in $\Omega$, this is independent of the choice of $\pi$. Given $\omega$, the point $(m,n)$ should be thought of as associated with the map $S^{f_1}T^{f_2}$.

Next we introduce probability and random entropy. It turns out to be convenient to redefine $\Omega$ as follows

$$\Omega = \{\omega = ((\omega_z^1, \omega_z^2)_{z\in\mathbf{Z}^2}) : \forall\, i, z, \omega_z^i \in \mathbf{Z}^2,\ \text{and}\ \omega_z^1 + \omega_{z+e_1}^2 = \omega_z^2 + \omega_{z+e_2}^1\},$$

where $e_i$ denote the unit vectors in $\mathbf{Z}^2$. In this way, labels are associated to vertices rather than to edges. Let $\phi : \mathbf{Z}^2 \times \Omega \to \Omega$ be the group action given by the coordinate shift and let $\mu$ be a $\phi$-invariant ergodic probability measure on $\Omega$. Furthermore, we let $\psi : \mathbf{Z}^2 \times X \to X$ be the coordinate shift and $\rho$ be a $\psi$-invariant ergodic probability measure on $X$. The cocycle $f$ induces a $(\mu \times \rho)$-invariant $\mathbf{Z}^2$-action $\Phi : \mathbf{Z}^2 \times \Omega \times X \to \Omega \times X$ as follows:

$$\Phi_z(\omega, x) = (\phi_z(\omega), \psi_{f(z,\omega)}(x)).$$

Using this set up, there are now several (equivalent) ways to define random entropy and we choose one of them:

**Definition 3.1:** *Let $\mu$ be a stationary and ergodic probability measure on $\Omega$ which satisfies*

$$\int_\Omega \|\omega_0\|_\infty d\mu(\omega) < \infty.$$

*The random entropy $E_\rho(\mu)$ is defined by*

$$E_\rho(\mu) = h_{\mu\times\rho}(\Phi) - h_\mu(\phi).$$

This definition is of course inspired by the one-dimensional skew product of Section 1. Here is a very intuitive interpretation. If $H \subset \mathbf{Z}^2$ is a finite subset, let $P_H$ be the partition on $\Omega$ specifying the coordinates of $\omega \in \Omega$ indexed by elements of $H$. Similarly, we let $Q_H$ denote the partition on $X$ specifying the coordinates of $x \in X$ that are indexed by elements in $H$. We also write $B_n = \{0, 1, \ldots, n-1\}^2$.

Let, for $M \geq 0$,

$$L^M(n)(\omega) = \{u \in \mathbf{Z}^2 : u = u' + u'', u' \in \{-M, \ldots, M\}^2, u'' \in f(B_n, \omega)\},$$

and let $S^M(n)$ denote the cardinality of $L^M(n)$. We also define the partitions

$$A_M = (P_{\{0\}} \times X) \vee (\Omega \times Q_{\{-M,\ldots,M\}^2}).$$

With a slight abuse of notation we write $P_H$ instead of $P_H \times X$ and $Q_S$ instead of $\Omega \times Q_S$. It will be clear from the context which space is considered. Then, using the entropy addition formula from Ward and Zhang (1992) in the second equality below, we have

$$\begin{aligned} h_{\mu\times\rho}(\Phi) &= \lim_{M\to\infty}\lim_{n\to\infty}\frac{1}{n^2}H(\vee_{g\in B_n}\Phi_g^{-1}(A_M)) \\ &= \lim_{M\to\infty}\lim_{n\to\infty}\left\{\frac{1}{n^2}H(P_{B_n}) + \frac{1}{n^2}H(Q_{L^M(n)}|P_{B_n})\right\} \\ &= h_\mu(\phi) + \lim_{M\to\infty}\lim_{n\to\infty}\frac{1}{n^2}H(Q_{L^M(n)}|P_{B_n}). \end{aligned}$$

It follows that the random entropy $E_\rho(\mu)$ satisfies

$$E_\rho(\mu) = \lim_{M\to\infty}\lim_{n\to\infty}\frac{1}{n^2}H(Q_{L^M(n)}|P_{B_n}).$$

In words, to compute the random entropy, one looks in the box $B_n$, moves the square $\{-M,\ldots,M\}^2$ around according to the transformations in the box, computes the entropy of the corresponding partition, and finally divides by $n^2$ and takes the limit for $n\to\infty$. The answer will be independent of $\omega$, and the limit for $M\to\infty$ then corresponds to taking the supremum over all partitions in the classical definition of entropy. We shall see in the next section that this leads to an immediate generalisation of Theorem 2.1

# 4 Identification of the entropy

In this section we derive a formula for $E_\rho(\mu)$ which can be used to actually compute $E_\rho(\mu)$ in certain cases. We work with the same setup as in Section 3, and assume in particular that $\mu$ satisfies the integrability condition in Definition 3.1.

Consider the subshift $\Omega$. In a given realisation, each point $x = (k, l) \in \mathbf{Z}^2$ corresponds to a map $S^{f_1}T^{f_2}$. Consider a stationary and ergodic (under the whole group action) measure $\mu$ on $\Omega$. We define horizontal and vertical limits as follows (surpressing in the notation the fact that $f = (f_1, f_2)$ also depends on $\omega$):

$$h_S(k) = \lim_{n\to\infty} \frac{f_1(n,k) - f_1(0,k)}{n}, \quad h_T(k) = \lim_{n\to\infty} \frac{f_2(n,k) - f_2(0,k)}{n},$$

$$v_S(k) = \lim_{n\to\infty} \frac{f_1(k,n) - f_1(k,0)}{n}, \quad v_T(k) = \lim_{n\to\infty} \frac{f_2(k,n) - f_2(k,0)}{n}.$$

All these limits exist $\mu$ a.e. by stationarity. We first claim that $h_S(k)$ is independent of $k$ and similarly for the other quantities. To see this, we write $X_n$ for $f_1(n,k) - f_1(0,k)$ and $Y_n$ for $f_1(n,k+1) - f_1(0,k+1)$. We have that $E|X_n - Y_n| \leq K$ for some uniform $K > 0$. (This follows from the integrability condition on $\mu$.) Hence,

$$E\left(\left|\frac{X_n}{n} - \frac{Y_n}{n}\right|\right) \to 0$$

for $n \to \infty$ and it follows from Markov's inequality that $|\frac{X_n}{n} - \frac{Y_n}{n}|$ converges to 0 in probability and hence the a.e. limit (which we know exists) has to be 0 as well. This proves the claim. It follows that $h_S(k)$ is invariant under both horizontal and vertical translations and hence it is $\mu$ a.e. constant. Similar statements are valid for the other quantities. Therefore it makes sense to define $h_S = h_S(k)$, $h_T = h_T(k)$, $v_S = v_S(k)$ and $v_T = v_T(k)$.

Consider the parallellogram $\mathcal{P}$ with vertices $(0,0)$, $(h_S, h_T)$, $(v_S, v_T)$ and $(h_S + v_S, h_T + v_T)$. Most of the work is contained in the following theorem.

**Theorem 4.1:** *For any ergodic measure $\mu$ on $\Omega$ which satisfies*

$$\int_\Omega \|\omega_0\|_\infty d\mu(\omega) < \infty,$$

*we have that almost surely,*

$$\lim_{n\to\infty} \frac{S^M(n)}{n^2} = \lambda(\mathcal{P})$$
$$= |\det((h_S, h_T), (v_S, v_T))|.$$

*where $\lambda$ denotes two-dimensional Lebesgue measure.*

**Proof:** The fact that $S^M(n)/n^2$ converges a.e. is an immediate consequence of the multiparameter subadditive ergodic theorem (Krengel (1985), Theorem 6.2.9). Indeed, $S^M(n)$ represents a cardinality and is easily seen to be subadditive. The fact that the limit is $\mu$ a.e. constant follows from the fact that the limit is obviously invariant under translations, together with the ergodicity of $\mu$.

So what we have to do is identify, for all $M$, this limit $\lim_{n\to\infty} S^M(n)/n^2$. Interestingly, we shall identify the limit in *probability* of $S^M(n)/n^2$. This is then of course also the a.e. limit.

We shall first deal with a special case; we take $M = 0$, and assume that $\mu$ concentrates on the set $\{(1,0),(0,1),(-1,0),(0,-1)\}^{\mathbf{E}^2}$. At the end of the proof we shall indicate which changes are necessary to deal with the general case.

We start by fixing $\epsilon > 0$ and choosing $N = N(\epsilon)$ so large that

$$\mu\left(\left|\frac{f_1(n,0)}{n} - h_S\right| < \epsilon, \forall\, n > N\right) > 1 - \epsilon,$$

and similarly for the other three quantities. This choice of course implies that for all $n > N$ we have

$$\mu\left(\left|\frac{f_1(n,0)}{n} - h_S\right| > \epsilon\right) < \epsilon, \quad \mu\left(\left|\frac{f_2(n,0)}{n} - h_T\right| > \epsilon\right) < \epsilon,$$

$$\mu\left(\left|\frac{f_1(0,n)}{n} - v_S\right| > \epsilon\right) < \epsilon, \quad \mu\left(\left|\frac{f_2(0,n)}{n} - v_T\right| > \epsilon\right) < \epsilon.$$

We call a point $(k, l)$ *good* if we have

$$\left|\frac{f_1(k,0)}{k} - h_S\right| < \epsilon, \quad \left|\frac{f_2(k,0)}{k} - h_T\right| < \epsilon,$$

$$\left|\frac{f_1(k,l) - f_1(k,0)}{l} - v_S\right| < \epsilon, \quad \left|\frac{f_2(k,l) - f_2(k,0)}{l} - v_T\right| < \epsilon.$$

By the choice of $N$, we see that each point in the region $V := \{(k,l) : k \geq N, l \geq N\}$ has probability at least $1 - 4\epsilon$ to be good. Denote the set $V \cap [0, n-1]^2$ by $V_n$. We claim that uniformly in $n$, with probability at least $1 - \sqrt{4\epsilon}$, at least a fraction $1 - \sqrt{4\epsilon}$ of the points in $V_n$ are good. To see this, it is easiest to look at bad points (points which are not good). Given any collection $X_1, \ldots, X_r$ of $0, 1$-valued random variables, with $P(X_i = 1) \leq 4\epsilon$ for all $i$, we have

$$\begin{aligned} \mu\left(\frac{1}{r}\sum_{i=1}^{r} X_i \geq \sqrt{4\epsilon}\right) &\leq \frac{E(\frac{1}{r}\sum_{i=1}^{r} X_i)}{\sqrt{4\epsilon}} \\ &\leq \frac{4\epsilon}{\sqrt{4\epsilon}} = \sqrt{4\epsilon}, \end{aligned}$$

proving the claim.

Now that we know that with high probability most points in $V$ are good, let us look at the image of good points under the power map $f$. For any good point $(k, l) \in V_n$ we have

$$k(h_S - \epsilon) \leq f_1(k,0) \leq k(h_S + \epsilon),$$

$$l(v_S - \epsilon) < f_1(k,l) - f_1(k,0) \leq l(v_S + \epsilon),$$

and similarly for the $f_2$-images. Thus together this gives

$$f(k,l) \in (kh_S + lv_S, kh_T + lv_T) + [-2(n-1)\epsilon, 2(n-1)\epsilon]^2.$$

Hence each good point in the region $V_n$ is mapped under $f$ into the perturbed 'parallellogram' $(n-1)\mathcal{P} + [-2(n-1)\epsilon, 2(n-1)\epsilon]$, and we denote this last set by $\mathcal{P}_n(\epsilon)$.

We shall now give upper and lower bounds for the limit in probability of $S^0(n)/n^2$, using the notion of good and bad points. We start with the upper bound. Let $Z_n$ denote the number of bad points in $V_n$, and denote by $|\mathcal{P}_n(\epsilon)|$ the cardinality of $\mathcal{P}_n(\epsilon)$. We then have for $n$ large enough,

$$\begin{aligned} \frac{S^0(n)}{n^2} &\leq \frac{2nN}{n^2} + \frac{|\mathcal{P}_n(\epsilon)|}{n^2} + \frac{Z_n}{(n-N)^2} \cdot \frac{(n-N)^2}{n^2} \\ &\leq \frac{2N}{n} + (1+\delta(\epsilon,\mathcal{P}))^2\lambda(\mathcal{P}) + \frac{Z_n}{(n-N)^2} \cdot \frac{(n-N)^2}{n^2}, \end{aligned}$$

where $\delta(\epsilon, \mathcal{P}) \to 0$ when $\epsilon \to 0$. Note that $\frac{Z_n}{(n-N)^2}$ is the fraction of bad points in the region $V_n$, so that it follows by the claim above that $\mu(\frac{Z_n}{(n-N)^2} \leq \sqrt{4\epsilon}) \geq 1 - \sqrt{4\epsilon}$. Therefore, the constant to which $S^0(n)/n^2$ converges a.e. (and hence in probability) has to be at most $(1+\delta(\epsilon,\mathcal{P}))^2\lambda(\mathcal{P})$. Since $\epsilon > 0$ was arbitrary, we conclude that $\lim_{n\to\infty} S^0(n)/n^2 \leq \lambda(\mathcal{P})$.

The lower bound is a little more complicated. Note that the proof of the upper bound shows that most points in $V_n$ are mapped (under the power map) into $\mathcal{P}_n(\epsilon)$. For the lower bound, we need to show that most points inside $(n-1)\mathcal{P}$ are actually image points. To do this, we shall again only use the fact that $S^0(n)/n^2$ converges in probability to a constant. Fix again $\epsilon > 0$ and $N = N(\epsilon)$ as above. Consider the parallellogram obtained from $(n-1)\mathcal{P}$ by multiplying $(n-1)\mathcal{P}$ with center $((n-1)h_S + (n-1)v_S, (n-1)h_T + (n-1)v_T)$ with a factor $\frac{n-1-N}{n-1}$, and call this parallellogram $\tilde{\mathcal{P}}_n$. If all points on the boundary of $V_n$ (all vertices in $V_n$ which are adjacent to at least one vertex outside $V_n$) are good, we have that the union of the images of the points on this boundary contains a circuit in the set $\tilde{\mathcal{P}}_n(\epsilon) :=$

$\{x \in \mathbf{R}^2 : d(x, \partial(\tilde{\mathcal{P}}_n)) \leq 2(n-1)\epsilon\}$, where $\partial$ denotes boundary and $d$ the $L^\infty$ distance. We claim that, as a deterministic fact, if all points on the boundary of $V_n$ are good, then all vertices in the set $\tilde{\mathcal{P}}_n \backslash \tilde{\mathcal{P}}_n(\epsilon)$ are images of points of $V_n$. Assume for a moment that this claim is correct. It follows from the choice of $N$ that the probability that all vertices on the boundary of $V_n$ are good is at least $1 - 8\epsilon$. So if the claim is correct we have that

$$\mu\left(\frac{S^0(n)}{n^2} \geq \frac{|\tilde{\mathcal{P}}_n \backslash \tilde{\mathcal{P}}_n(\epsilon)|}{n^2}\right) \geq 1 - 8\epsilon,$$

from which it follows that

$$\mu\left(\frac{S^0(n)}{n^2} \geq \frac{n^2(1 - \delta'(\epsilon, \mathcal{P}))}{n^2}\lambda(\mathcal{P})\right) \geq 1 - 8\epsilon,$$

where $\delta'(\epsilon, \mathcal{P}) \to 0$ when $\epsilon \to 0$. Since $\epsilon > 0$ is arbitrary it follows that $\lim_{n\to\infty} S^0(n)/n^2 \geq \lambda(\mathcal{P})$.

Next we prove the claim, i.e. we show that if all points on the boundary of $V_n$ are good, then all vertices in the set $\mathcal{Q}_n(\epsilon) := \tilde{\mathcal{P}}_n \backslash \tilde{\mathcal{P}}_n(\epsilon)$ are image points of $V_n$.

Consider the continuous curve which connects $f(N, N), f(N, N+1), \ldots, f(N, n-1)$ by straight line segments and call this curve $\gamma_N$. Similarly, define $\gamma_{N+1}$ as the curve connecting $f(N+1, N), f(N+1, N+1), \ldots, f(N+1, n-1)$, and continue to define curves until $\gamma_{n-1}$, which is the curve connecting $f(n-1, N), f(n-1, N+1), \ldots, f(n-1, n-1)$. All the curves $\gamma_N, \ldots, \gamma_{n-1}$ have $n - N$ vertices, where we do count multiplicities. These vertices are denoted by $\gamma_i(1), \ldots, \gamma_i(n-N)$, for all $i = N, \ldots, n-1$. The assumption on $\mu$ implies that $|\gamma_i(k) - \gamma_i(k+1)| = 1$ for all relevant $i$ and $k$. Each side of the parallellogram $\mathcal{Q}_n(\epsilon)$ can be extended to a doubly infinite line which divides the plane into two half planes, one of which does not contain $\mathcal{Q}_n(\epsilon)$. The latter half planes are denoted $H_1, \ldots, H_4$, where the numbering is chosen such that $\gamma_N \subset H_1$, $\gamma_{n-1} \subset H_3$, $\gamma_i(1) \in H_2$ for all $i$ and $\gamma_i(n-N) \in H_4$ for all $i$. This can be done by the goodness of the points on the boundary of $V_n$. Next we are going to describe a process evolving in time as follows: at time $t = 0$, we place a particle at each of the vertices of $\gamma_N$, and label the particle at $\gamma_N(i)$ by the number $i$. Note that if a vertex appears more than once in $\gamma_N$, then there is more than one particle at this vertex. Between $t = 0$ and $t = 1$, particle 1 moves with unit speed from $\gamma_N(1)$ to $\gamma_{N+1}(1)$, therefore ending up at this last vertex at time 1. All other particles do not move in this time interval. (In fact, there will always be exactly one particle moving at any non-integer time.) Between time 1 and 2, particle 2 moves with unit speed from $\gamma_N(2)$ to $\gamma_{N+1}(2)$. We continue this, so that finally between time $n - N - 1$ and $n - N$, particle $n - N$ moves from $\gamma_N(n - N)$ to $\gamma_{N+1}(n - N)$. Note that at this point, all particles have moved from $\gamma_N$ to $\gamma_{N+1}$. Next, between time $n - N$ and $n - N + 1$, particle 1 moves from $\gamma_{N+1}(1)$ to $\gamma_{N+2}(2)$. After that, particle 2 moves from $\gamma_{N+1}(2)$ to $\gamma_{N+2}(2)$ and so on. So at time $2(n - N)$ particle $i$ is at the vertex $\gamma_{N+2}(i)$. We continue in the obvious way, until the final configuration is reached at time $(n - N)^2$ when particle $i$ is at $\gamma_n(i)$ for all $i = 1, \ldots, n - N$.

At all times $t$, the curve $\rho_t$ is defined by connecting particles $1, \ldots, n-N$ (in that order) with straight line segments. In particular, we have that $\rho_{k(n-N)}$ is just $\gamma_{N+k}$. Consider horizontal lines $l_k := \{(x, k) : x \in \mathbf{R}\}$ for integers $k$, and let $I(k)$ denote the intersection of $l_k$ and $\mathcal{Q}_n(\epsilon)$. Assume for the sake of concreteness that $\mathcal{Q}_n(\epsilon)$ is oriented in such a way that $H_1 \cap l_k$ is at the left of $H_3 \cap l_k$. If $I(k) \cap \rho_t \neq \emptyset$, we denote the leftmost point of this intersection by $r(k, t)$. If the intersection is empty, $r(k, t))$ is defined to be the leftmost point of $I(k)$.

The main observation is the following: as a function of $t$, $r(k,t)$ can only make 'jumps' from one vertex to a neighbouring vertex. Apart from these possible jumps, $r(k,t)$ is a continuous function of $t$. This observation is easy to verify by checking all possibilities. It follows from this observation, together with the fact that $r(k,0) \in H_1$ and $r(k,(n-N)^2) \in H_3$, that all vertices on the line segment $I(k)$ must belong to some $\rho_t$ for integer $t$, which means that all these vertices must belong to some $\gamma_i$. This proves the claim.

It remains to indicate the changes needed to deal with the general case. The assumption on $M$ is easily dispensed with: it follows from the proof just given that as long as $\mu$ is concentrated on $\{(1,0),(0,1),(-1,0),(0,-1)\}^{\mathbf{E}^2}$, we have $\lim_{n\to\infty} S^M(n)/n^2 = \lambda(\mathcal{P})$ for all $M \geq 0$, since an increase in $M$ is only a boundary effect which disappears when $n$ gets large.

To treat the general case, note that the proof just given consists of two parts. First we show that most points are mapped into $\mathcal{P}_n(\epsilon)$, and then we show that most points in $\mathcal{P}_n(\epsilon)$ are actually image points. When there is no restriction on $\mu$ (apart from the usual integrability condition), the first part of the proof goes through with only minor changes. The argument in the second part breaks down though since the function $r(k,t)$ can make big jumps now (as a function of $t$) and it is not clear that all vertices are image points. But we can save the proof by considering larger values of $M$: For $\epsilon > 0$, we take $M$ so large that $\int_{||\omega_0||_\infty \geq M} ||\omega_0||_\infty d\mu < \epsilon$. This choice implies that during the 'filling up' of the parallellogram with translates of $\{-M,\ldots,M\}^2$, the expected fraction of the parallellogram *not* covered goes down to zero with $\epsilon$. Also, the region outside the parallellogram covered by these translates is again a boundary effect which becomes negligible for large

$n$. These observations conclude the proof. $\square$

Finally we state and prove the immediate generalisation of Theorem 2.1.

**Theorem 4.2:** *Under the conditions of Theorem 4.1 we have*

$$\begin{aligned} E_\rho(\mu) &= \lambda(\mathcal{P})h_\rho(\psi) \\ &= |\det((h_S, h_T), (v_S, v_T))| h_\rho(\psi). \end{aligned}$$

**Proof:** According to the remarks following Definition 3.1, we need to identify

$$\lim_{M\to\infty} \lim_{n\to\infty} \frac{1}{n^2} H(Q_{L^M(n)}|P_{B_n}).$$

Let $R_n$ be the set of integer vectors in the parallellogram $n\mathcal{P}$. From the proof of Theorem 4.1 we have that for all $\epsilon > 0$,

$$\limsup_{n\to\infty} \frac{|L^M(n)(\omega) \bigtriangleup R_n|}{n^2} < \epsilon$$

a.e. and since the sequence is uniformly bounded, convergence is also in $L^1$. Also notice that if $C$ is an atom of $P_{B_n}$, then $L^M(n)(\omega)$ is the same for all $\omega \in C$ and we denote this set by $L^M(n)(C)$. Now for $M$ sufficiently large, we write

$$\begin{aligned} &\frac{1}{n^2}\left| H(Q_{L^M(n)}|P_{B_n}) - H(Q_{R_n})\right| \\ &= \frac{1}{n^2}\left|\sum_{C\in P_{B_n}} H(Q_{L^M(n)}|C)\mu(C) - H(Q_{R_n})\right| \\ &\le \frac{1}{n^2}\sum_{C\in P_{B_n}} \left|H(Q_{L^M(n)}|C) - H(Q_{R_n})\right|\mu(C) \\ &\le \log|A| \sum_{C\in P_{B_n}} \frac{|L^M(n)(C)\bigtriangleup R_n|}{n^2}\mu(C) \\ &= \log|A| \int_\Omega \frac{|L^M(n)(\omega)\bigtriangleup R_n|}{n^2} d\mu \le \epsilon \log|A| \end{aligned}$$

for $n$ large enough. Finally we note that is is easy to see that

$$\lim_{n\to\infty} \frac{H(Q_{R_n})}{n^2} = h_\rho(\psi)|\det((\alpha_1,\alpha_2),(\beta_1,\beta_2))|,$$

and this completes the proof. □

**Examples:** All our three examples below are in two dimensions and have $|A| = 2$. For $\rho$, we take product measure with equal marginals, so $h_\rho(\psi) = \log 2$. The edges will only be labelled by $(1,0)$ or $(0,1)$. The first label is identified with the transformation $S$, the second label with the transformation $T$. We start with two extremal cases.

**(A)** The entropy of a deterministic group action is a special case of Theorem 4.2.; just take $\mu$ the measure that concentrates on the realistion of labels in which all horizontal edges are labelled $S$ and all vertical edges are labelled $T$. We then have $h_S = v_T = 1$ and $h_T = v_S = 0$. This means that $E_\rho(\mu) = \log 2$.

**(B)** The example in the proof of Proposition 3.1 leads to the following measure $\mu$. First place the 'skeleton' of squares in a stationary way. (This can be done by choosing the position of the origin uniformly over the square and then tile the plane with adjacent squares.) Within each square, choose one of the two possibilities with probabilities $p$ and $1-p$ respectively, say independently of each other. It is easy to see that $h_S, h_T, v_S$ and $v_T$ are all equal to $\frac{1}{2}$ (independent of $p$). Hence $E_\rho(\mu) = 0$ in this case.

**(C)** We construct a measure $\mu$ on labels of edges in the first quadrant only, but this is not important, since we can for instance use Kolomogorov's consistency theorem to extend this to a measure on labels in the whole plane.

Choose $0 \le p \le 1$ and let $q = 1-p$. Label all edges of the $x$-axis $S$ with probability $q$ and $T$ with probability $p$, independently of each other.

For the $y$-axis we do the same with interchanged probabilities. Now denote the square $[n, n+1] \times [0, 1]$ by $W_n$, and denote the lower and upper edge of $W_n$ by $e_n$ and $f_n$ respectively. The labelling procedure is at follows: First label the remaining edges of $W_1$; if there are two possibilities for doing this we choose one of them with equal probabilities. At this point, the lower and left edge of $W_2$ are labelled, and we next complete the labelling of $W_2$, noting again that if there are two ways to do this, we choose one of them with equal probabilities. This procedure is continued and gives all labels in the strip $[0, \infty) \times [0, 1]$. Then we move one unit upwards, and complete in a similar fashion the labels in the strip $[0, \infty) \times [1, 2]$. (Of course, if you want to carry out this labelling, you never actually finish any strip. Instead, you start at some moment with the second strip which can be labelled as far as the current labelling of the first strip allows, etc.) We claim that this procedure yields a stationary and ergodic labelling, and in the next paragraph we indicate how to prove this. It is easy to see that $h_S = v_T = q$ and $h_T = v_S = p$ in this case, giving $E_\rho(\mu) = |p - q| \log 2$.

We end this example by a sketch of the proof that the labelling is stationary and ergodic. If we can show that the labelling of the edges $f_n$ has the same distribution as the labelling of the edges $e_n$, then we have shown that the labelling in the quadrant $[0, \infty) \times [1, \infty)$ has the same distribution as the labelling in $[0, \infty) \times [0, \infty)$ and we can use a similar argument for vertical lines plus induction to finish the argument. Therefore we only need to show that the labelling of the edges $f_n$ is i.i.d. with the correct marginals. To do this properly, consider the labels of the edges of $W_n$. There are six possible labellings of the edges of $W_n$. Four of these are such that $e_n$ and $f_n$ have the same label. The exceptional labellings are (starting at the lower left vertex and moving clockwise) $STTS$ and $TSST$. Denote the labelling of the edges of $W_n$ by $L_n$. Then it is not hard to see that $L_n$ is a Markov chain on the state space $\{STTS, TSST, TSTS, STST, TTTT, SSSS\}$. Take the transition matrix $P$ of $L_n$, interchange the rows and the columns corresponding to $STTS$ and $TSST$ to obtain $P'$, and consider the *backward* Markov chain corresponding to $P'$, denoted by $M_n$. An easy calculation then shows that $L_n$ and $M_n$ have the same transition matrix and that they are both in stationarity. But $M_n$ is by construction just the labelling of the strip $[0, \infty) \times [0, 1]$ rotated over 180 degrees. This implies that the labelling of the edges $f_n$ has the same distribution as the labelling of the edges $e_n$, which is what we wanted to prove. As for ergodicity of $\mu$: any invariant event $A$ can be approximated by an event $A'$ which depends only on the edges in a finite box. By moving this box to the right step by step we again obtain a Markov chain, which is ergodic (in fact, even mixing). This then implies that $A'$ is trivial, and hence so is $A$. □

We end with a corollary. The statement of the corollary could give rise to confusion: the entropy in the statement refers to the usual measure-theoretical entropy of $\mu$ and not to $E_\rho(\mu)$. A subset $K$ of $\mathbf{Z}^2$ is called *symmetric* if it is invariant under reflection in the origin.

**Corollary 4.1:**

*(i) Let $K \subset \mathbf{Z}^2$ be a symmetric finite set and let $\Omega_K$ be the subset of $\Omega$ of all configurations with labels in $K$. If $\Omega_K$ has a unique measure $\mu_K$ of maximal entropy, then $E_\rho(\mu_K) = 0$.*

*(ii) If $\mu$ is invariant under rotations, then $E_\rho(\mu) = 0$.*

**Proof:** The introduction of $\Omega_K$ in (i) is needed to make the statement not a priori empty, since $\Omega$ has no measure of maximal entropy. To prove (i), suppose we flip all labels from $S^mT^n$ to $S^{-m}T^{-n}$. We then obtain another measure $\mu'_K$, say, which is also concentrated on $\Omega_K$. It is clear that $h_S(\mu_K) = -h_S(\mu'_K)$ and similarly for the other quantities. But clearly, $\mu'_K$ has the same entropy as $\mu_K$, whence it follows from the assumption that $h_S(\mu_K) = h_S(\mu'_K)$ and similarly for the other quantities. it follows that $h_S(\mu_K) = h_T(\mu_K) = v_S(\mu_K) = v_T(\mu_K) = 0$, whence $\lambda(\mathcal{P}) = 0$.

For (ii), just observe that the assumption on $\mu$ implies that $h_S = v_S$ and $h_T = v_T$. $\square$

**Remark:** The phenomenon in (i) is exactly what happens in the one-dimensional case as well. The unique measure of maximal entropy on $\{0,1\}^{\mathbf{Z}}$ is the product measure with marginals 1/2. In this case, we saw already in Theorem 2.1 that the random entropy is equal to zero.

## References


[1] Conze, J.P. (1972) Entropie d'un groupe abélien de transformations, *Z. Wahrsch. Verw. Geb.* **25**, 18-30.

[2] Kifer, Y. (1986) *Ergodic Theory of Random Transformations*, Birkhäuser, Boston.

[3] Krengel, U. (1985) *Ergodic Theorems*, Walter de Gruyter, Berlin-New York.

[4] Krieger, W. (1970) On entropy and generators of measure-preserving transformations, *Trans. Amer. Math. Soc.* **149**, 453-464.

[5] Schmidt, K. (1995) *Dynamical Systems of Algebraic Origin*, Birkhäuser.

[6] Spitzer, F. (1976) *Principles of Random Walk*, 2nd edition, Springer-Verlag, New York.

[7] Walters, P. (1982) *An Introduction to Ergodic Theory*, Springer-Verlag, New York.

[8] Ward, T. and Zhang, Q. (1992) The Abramov-Rokhlin entropy addition formula for amenable group actions, *Monatshefte für Mathematik* **114**, 317-329.