COLLOQUIUM MATHEMATICUM

VOL. 126 2012 NO. 2

ENTROPY OF A DOUBLY STOCHASTIC MARKOV OPERATOR AND OF ITS SHIFT ON THE SPACE OF TRAJECTORIES

BY

PAULINA FREJ (Wrocław)

Abstract. We define the space of trajectories of a doubly stochastic operator on $L^1(X,\mu)$ as a shift space $(X^{\mathbb{N}},\nu,\sigma)$, where ν is a probability measure defined as in the Ionescu–Tulcea theorem and σ is the shift transformation. We study connections between the entropy of a doubly stochastic operator and the entropy of the shift on the space of trajectories of this operator.

1. Doubly stochastic operators, transition probabilities and entropy. Let $(X,\mathscr{B},\mu)$ be a probability space. By a *doubly stochastic operator* we mean a linear operator $T\colon L^1(\mu)\to L^1(\mu)$ which satisfies:

(i) Tf is positive for every positive $f\in L^1(\mu)$,
(ii) $T\mathbb{1}=\mathbb{1}$ (where $\mathbb{1}(x)=1$ for all $x\in X$),
(iii) $\int Tf\,d\mu=\int f\,d\mu$ for every $f\in L^1(\mu)$.

The class of all doubly stochastic operators contains Koopman operators of measure preserving transformations on X and, even more generally, Markov operators defined by measure preserving transition probabilities by the formula

$$Tf(x)=\int f(y)\,P(x,dy) \tag{1.1}$$

(we recall that a *transition probability* on $(X,\mathscr{B})$ is a function $P\colon X\times\mathscr{B}\to[0,1]$ such that $P(x,\cdot)$ is a probability measure for each $x\in X$ and $P(\cdot,B)$ is a measurable function for each $B\in\mathscr{B}$; a transition probability *preserves* μ if $\mu(A)=\int P(x,A)\,d\mu(x)$ for all $A\in\mathscr{B}$). In the latter case, the Ionescu–Tulcea theorem implies that there is a probability measure ν on the product space $X^{\mathbb{N}}$, $\mathbb{N}=\{0,1,2,\ldots\}$, whose values on cylinder sets are given by

$$\nu(A_0\times A_1\times\cdots\times A_n\times X^{\mathbb{N}})=\int_{A_0}\int_{A_1}\ldots\int_{A_n}P(x_{n-1},dx_n)\ldots P(x_0,dx_1)\,\mu(dx_0)$$

2010 *Mathematics Subject Classification*: Primary 28D20; Secondary 47A35.
Key words and phrases: doubly stochastic operator, entropy, shift space, Ionescu–Tulcea theorem.

DOI: 10.4064/cm126-2-5

for all $A_0, \ldots, A_n \in \mathscr{B}$ (see e.g. [N]). Note that in terms of the operator T the above formula takes the following form:

$$\nu(A_0 \times A_1 \times \cdots \times A_n \times X^{\mathbb{N}}) = \int \mathbb{1}_{A_0} \cdot T(\mathbb{1}_{A_1} \cdot T(\mathbb{1}_{A_2} \cdot \cdots \cdot T\mathbb{1}_{A_n}) \ldots)\, d\mu.$$

Let $\sigma\colon X^{\mathbb{N}} \to X^{\mathbb{N}}$ be the shift transformation, i.e. $(\sigma x)_n = x_{n+1}$ for $x = (x_n)_{n\in\mathbb{N}} \in X^{\mathbb{N}}$. It is easy to verify that since a transition probability P preserves μ, the measure ν is σ-invariant. The probability space $(X^{\mathbb{N}}, \nu)$ will be called the *space of trajectories* of the doubly stochastic Markov operator T defined by (1.1).

We continue the study of entropy of doubly stochastic operators started in [DF]. We take a closer look at the definition of entropy introduced in [FF], relating it to the entropy of the shift on the space of trajectories. In the literature one can find various generalizations of the notion of entropy: see e.g. [AF], [CNT], [GLW], [M], [MR] and [V]. By the results of [DF], all definitions of entropy of a doubly stochastic operator T which follow standard construction steps and satisfy some natural conditions concerning monotonicity, subadditivity, continuity and compatibility with the classical Shannon entropy, lead to the same quantity $h_\mu(T)$. Such are entropies defined in [AF], [DF], [GLW], [M]. However, these entropies differ in the choice of the basic entropy H_μ of a collection of functions, independent of the dynamics induced by an operator.

In [FF] it was shown that the formula which we use in the current paper (see Def. 1.1(3)) is also a version of operator entropy in the sense of [DF]. Our main Theorem 2.5 and its corollaries are valid only for this specific definition of H_μ, but it may serve as a tool in proving theorems on the general entropy $h_\mu(T)$. We remark that though we assume that the operator T is induced by a transition probability, the results apply directly to all doubly stochastic operators on standard Borel spaces, as it is well known that on such spaces each doubly stochastic operator is of the form (1.1) (see [N]). In the general case, one can represent a doubly stochastic operator T as a Markov operator on the space of all continuous functions on a compact Hausdorff space $\widetilde{X}$ (see [D] or [H] for details; a similar technique was also used in the proof of Lemma 3.2 in [DF]). It is easy to verify that such operators are always induced by transition probabilities. Moreover, if we transport a measure μ from X to $\widetilde{X}$, this representation will preserve the entropy of T.

Throughout the article, $(X, \mathscr{B}, \mu)$ is a probability space, $\|\cdot\|$ denotes the L^1 norm on the space of integrable functions on $(X, \mathscr{B}, \mu)$, λ is the Lebesgue measure on the unit interval and $\eta\colon [0,1] \to \mathbb{R}$ is the function defined by $\eta(x) = -x \log x$ for $x \in (0,1]$ and $\eta(0) = 0$ (log means logarithm to base 2). We replace partitions, used in the classical definition of the entropy of a measure preserving transformation, by finite sequences of functions from X

to $[0,1]$, which we call *families*. For a measurable function $f\colon X\to[0,1]$ let $A_f=\{(x,t)\in X\times[0,1]\colon t\le f(x)\}$ and denote by $\mathscr{A}_f$ the partition of $X\times[0,1]$ consisting of A_f and its complement. Recall that the *refinement* (or *join*) of partitions α and β is the partition $\alpha\vee\beta$ consisting of all sets $A\cap B$, where $A\in\alpha$, $B\in\beta$. For a finite collection of partitions $\alpha_1,\dots,\alpha_n$ we write $\bigvee_{i=1}^n\alpha_i$ or $\bigvee_{i\in I}\alpha_i$, $I=\{1,\dots,n\}$, for their refinement which is the collection of all sets $A_1\cap\dots\cap A_n$, where $A_i\in\alpha_i$. For a family $\mathscr{F}$ of measurable functions we define $\mathscr{A}_{\mathscr{F}}=\bigvee_{f\in\mathscr{F}}\mathscr{A}_f$. Denote by A^t the *t-section* of $A\subset X\times[0,1]$ at t, i.e. $A^t=\{x\in X:(x,t)\in A\}$, and by $\mathscr{A}^t_{\mathscr{F}}$ the partition of X consisting of all t-sections A^t, where $A\in\mathscr{A}_{\mathscr{F}}$. Let $H_\mu(\alpha)$ be the Shannon entropy of a partition α.

Definition 1.1. For a family $\mathscr{F}$ of measurable functions on X we define

(1) the *entropy* of $\mathscr{F}$ at level $t\in[0,1]$ by
$$H_\mu(\mathscr{F},t)=H_\mu(\mathscr{A}^t_{\mathscr{F}}),$$

(2) the *upper* and *lower entropy* of a doubly stochastic operator T with respect to $\mathscr{F}$ at level $t\in[0,1]$,
$$\overline{h}_\mu(T,\mathscr{F},t)=\limsup_{n\to\infty}\frac{1}{n}H_\mu\Big(\bigvee_{i=0}^{n-1}\mathscr{A}^t_{T^i\mathscr{F}}\Big),$$
$$\underline{h}_\mu(T,\mathscr{F},t)=\liminf_{n\to\infty}\frac{1}{n}H_\mu\Big(\bigvee_{i=0}^{n-1}\mathscr{A}^t_{T^i\mathscr{F}}\Big),$$

(3) the *static entropy* of $\mathscr{F}$ by
$$H_\mu(\mathscr{F})=\int_0^1 H_\mu(\mathscr{F},t)\,\lambda(dt),$$

(4) the *entropy* of a doubly stochastic operator T with respect to $\mathscr{F}$ by
$$h_\mu(T,\mathscr{F})=\lim_{n\to\infty}\frac{1}{n}H_\mu(\mathscr{F}^n),$$
where $\mathscr{F}^n$ is the concatenation of $\mathscr{F}$, $T\mathscr{F},\dots,T^{n-1}\mathscr{F}$. (The existence of the above limit was proved in [FF].)

Note that for each measurable set $A\subset X\times[0,1]$ the function $t\mapsto\mu(A^t)$ is measurable, hence $H_\mu(\mathscr{F},t)=-\sum_{A\in\mathscr{A}_{\mathscr{F}}}\mu(A^t)\log(\mu(A^t))$ is also measurable and the above integral is well defined.

2. Operator entropy vs. entropy on the space of trajectories. For a (measurable) function $g\colon X\to[0,1]$ and $t\in[0,1]$ let $\mathbb{1}_{\{g\ge t\}}\in L^1(\mu)$ be the characteristic function of the set $\{x\in X\colon g(x)\ge t\}$ and abbreviate $\mathbb{1}_X$ by $\mathbb{1}$.

Lemma 2.1. *Let $\mathscr{F}$ be a family of measurable functions on X with ranges in $[0,1]$. For every $\varepsilon > 0$ there exist $l \in \mathbb{N}$ and a set $\tau \subset [0,1]$ such that*

(1) $\|\mathbb{1}_{\{T^{l+n} f \geq t\}} - T^n \mathbb{1}_{\{T^l f \geq t\}}\| < \varepsilon$ *for every* $f \in \mathscr{F}$, $n \in \mathbb{N}$ *and* $t \in [0,1] \setminus \tau$,
(2) $\lambda(\tau) \leq \varepsilon$.

Proof. Fix $\varepsilon > 0$. Clearly, we may assume that $\varepsilon = 1/k$ for some positive integer k. Let r denote the cardinality of $\mathscr{F}$. For a function $g\colon X \to [0,1]$ and numbers a, b satisfying $0 \leq a < b \leq 1$ define $g_a^b := (g \vee a) \wedge b$, where $\vee$ and $\wedge$ denote pointwise maximum and minimum, respectively. By Lemma 2.3 from [DF] we can choose $l \in \mathbb{N}$ such that for each $g \in T^l \mathscr{F}$, $n \in \mathbb{N}$ and every pair $a < b$,

$$\|T^n(g_a^b) - (T^n g)_a^b\| < \frac{\varepsilon^3}{3^5 r}. \tag{2.1}$$

Take a partition ζ of $[0,1]$ into $3^3 k^2 r$ subintervals $I_0, I_1, \ldots$ of equal length $\varDelta = \varepsilon^2/(3^3 r)$. We assume that the intervals are enumerated so that $\inf I_0 = 0$ and $\sup I_{i-1} = \inf I_i$ for $i > 0$. Notice that for each $f \in \mathscr{F}$ there are at most $9k$ intervals I_i such that $\mu((T^l f)^{-1}(I_i)) \geq \varepsilon/9$. Thus, the number of intervals such that this inequality holds for at least one element of $\mathscr{F}$ is at most $9kr$.

For each $t \in [0,1]$ consider the interval $I(t) = I_{i-1} \cup I_i \cup I_{i+1}$ such that $t \in I_i$. Denote the endpoints of $I(t)$ by a_t and b_t ($a_t < b_t$). Let τ be the union of all I_is such that the measure of the preimage of at least one of I_{i-1}, I_i, I_{i+1} under some $T^l f$, $f \in \mathscr{F}$, is greater than or equal to $\varepsilon/9$. It is easy to see that $\lambda(\tau) \leq 3 \cdot \varDelta \cdot 9kr = \varepsilon$.

Fix $t \in [0,1] \setminus \tau$ and set

$$F_{T^l f}(x) = \frac{(T^l f)_t^{b_t}(x) - t}{b_t - t} \quad \text{and} \quad G_{T^l f}(x) = \frac{(T^l f)_{a_t}^t(x) - a_t}{t - a_t}.$$

Then

$$F_{T^l f} \leq \mathbb{1}_{\{T^l f \geq t\}} \leq G_{T^l f} \tag{2.2}$$

and

$$\|G_{T^l f} - F_{T^l f}\| \leq \mu((T^l f)^{-1}(I(t))) < \varepsilon/3.$$

Because the operator T is positive, from (2.2) we have

$$T^n F_{T^l f} \leq T^n \mathbb{1}_{\{T^l f \geq t\}} \leq T^n G_{T^l f}$$

for every $n \in \mathbb{N}$, and since T is an L^1-contraction it follows that

$$\|T^n \mathbb{1}_{\{T^l f \geq t\}} - T^n F_{T^l f}\| \leq \|T^n G_{T^l f} - T^n F_{T^l f}\| \leq \|G_{T^l f} - F_{T^l f}\| < \varepsilon/3.$$

Similarly, we can write

$$F_{T^{n+l}f} = \frac{(T^{n+l}f)_t^{b_t} - t}{b_t - t} \le \mathbb{1}_{\{T^{n+l}f \ge t\}} \le \frac{(T^{n+l}f)_{a_t}^t - a_t}{t - a_t} = G_{T^{n+l}f}.$$

We now prove that

$$\|F_{T^{n+l}f} - T^n F_{T^l f}\| < \varepsilon/9 \quad \text{and} \quad \|G_{T^{n+l}f} - T^n G_{T^l f}\| < \varepsilon/9.$$

Indeed, using (2.1) and the fact that $|b_t - t| \ge \Delta$ we obtain

$$\begin{aligned}\|F_{T^{n+l}f} - T^n F_{T^l f}\| &= \left\| \frac{(T^{n+l}f)_t^{b_t} - t}{b_t - t} - T^n\left(\frac{(T^l f)_t^{b_t} - t}{b_t - t}\right)\right\| \\ &= \frac{1}{|b_t - t|}\|(T^{n+l}f)_t^{b_t} - T^n(T^l f)_t^{b_t}\| < \frac{1}{\Delta}\frac{\varepsilon^3}{3^5 r} = \frac{\varepsilon}{9}.\end{aligned}$$

In an analogous way we can show that $\|G_{T^{n+l}f} - T^n G_{T^l f}\| < \varepsilon/9$. Now we obtain

$$\begin{aligned}\|F_{T^{n+l}f} - \mathbb{1}_{\{T^{n+l}f \ge t\}}\| &\le \|F_{T^{n+l}f} - G_{T^{n+l}f}\| \\ &\le \|F_{T^{n+l}f} - T^n F_{T^l f}\| + \|T^n F_{T^l f} - T^n G_{T^l f}\| \\ &\quad + \|T^n G_{T^l f} - G_{T^{n+l}f}\| < \frac{\varepsilon}{9} + \frac{\varepsilon}{3} + \frac{\varepsilon}{9} = \frac{5\varepsilon}{9}.\end{aligned}$$

Finally,

$$\begin{aligned}\|\mathbb{1}_{\{T^{n+l}f \ge t\}} - T^n \mathbb{1}_{\{T^l f \ge t\}}\| &\le \|\mathbb{1}_{\{T^{n+l}f \ge t\}} - F_{T^{n+l}f}\| + \|F_{T^{n+l}f} - T^n F_{T^l f}\| \\ &\quad + \|T^n F_{T^l f} - T^n \mathbb{1}_{\{T^l f \ge t\}}\| \\ &< \frac{5\varepsilon}{9} + \frac{\varepsilon}{9} + \frac{\varepsilon}{3} = \varepsilon. \ \blacksquare\end{aligned}$$

Below we show that the choice of τ in the previous lemma is in some sense uniform.

LEMMA 2.2. *Let $\mathscr{F}$ be a family of measurable functions on X with ranges in $[0,1]$. For every $\varepsilon > 0$ there exist $l_0 \in \mathbb{N}$ and a set $\tau \subset [0,1]$ such that*

(1) $\|\mathbb{1}_{\{T^{l+n}f \ge t\}} - T^n \mathbb{1}_{\{T^l f \ge t\}}\| < \varepsilon$ *for every* $f \in \mathscr{F}$, $l \ge l_0$, $n \in \mathbb{N}$ *and* $t \in [0,1] \setminus \tau$,
(2) $\lambda(\tau) \le \varepsilon$.

Proof. For $\frac{1}{2}\varepsilon$ we find τ and l_0 using the previous lemma. Then, for every $f \in \mathscr{F}$, $l \ge l_0$, $n \in \mathbb{N}$ and $t \in [0,1] \setminus \tau$,

$$\begin{aligned}&\|\mathbb{1}_{\{T^{l+n}f \ge t\}} - T^n \mathbb{1}_{\{T^l f \ge t\}}\| \\ &\le \|\mathbb{1}_{\{T^{l+n}f \ge t\}} - T^{l-l_0+n}\mathbb{1}_{\{T^{l_0}f \ge t\}}\| + \|T^{l-l_0+n}\mathbb{1}_{\{T^{l_0}f \ge t\}} - T^n \mathbb{1}_{\{T^l f \ge t\}}\| \\ &\le \|\mathbb{1}_{\{T^{l_0+(l-l_0+n)}f \ge t\}} - T^{l-l_0+n}\mathbb{1}_{\{T^{l_0}f \ge t\}}\| + \|T^{l-l_0}\mathbb{1}_{\{T^{l_0}f \ge t\}} - \mathbb{1}_{\{T^{l_0+(l-l_0)}f \ge t\}}\| \\ &< \varepsilon. \ \blacksquare\end{aligned}$$

The following definition was stated in [DF] in order to consider continuity of entropy.

DEFINITION 2.3. Let $r' \le r$. For two families of measurable functions, $\mathcal{F} = \{f_1, \dots, f_r\}$ and $\mathcal{G} = \{g_1, \dots, g_{r'}\}$, their L^1-*distance* is defined by

$$\operatorname{dist}(\mathcal{F}, \mathcal{G}) = \min_{\pi} \Big\{ \max_{1 \le i \le r} \int |f_i - g_{\pi(i)}| \, d\mu \Big\},$$

where the minimum is taken over all permutations π of $\{1, \dots r\}$ and where $\mathcal{G}$ is considered an r-element family by setting $g_i \equiv 0$ for $r' < i \le r$.

It is easy to prove that for any measurable functions f, f', g and g',

$$\|f \wedge g - f' \wedge g'\| \le \|f - f'\| + \|g - g'\|. \tag{2.3}$$

To simplify notation we denote the partition $\mathscr{A}^t_{T^l \mathscr{F}}$ of X by $\mathscr{A}^t_l$. Let $\mathbb{1}_{\mathscr{A}^t_l}$ denote the family of all functions $\mathbb{1}_A$, where $A \in \mathscr{A}^t_l$.

LEMMA 2.4. *Let $\mathscr{F}$ be a family of measurable functions on X with ranges in $[0,1]$. For every $\varepsilon > 0$ there exist $l_0 \in \mathbb{N}$ and a set $\tau \subset [0,1]$ such that*

(1) $\operatorname{dist}(T^n \mathbb{1}_{\mathscr{A}^t_l}, \mathbb{1}_{\mathscr{A}^t_{l+n}}) < \varepsilon$ *for every* $l \ge l_0$, $n \in \mathbb{N}$ *and* $t \in [0,1] \setminus \tau$,

(2) $\lambda(\tau) \le \varepsilon$.

Proof. Let $\mathscr{F} = \{f_1, \dots, f_r\}$. Fix $\varepsilon > 0$. We use the previous lemma to find $l_0 \in \mathbb{N}$ and $\tau \subset [0,1]$ for $\varepsilon/(r2^r)$ in place of ε. For every l each element of $\mathbb{1}_{\mathscr{A}^t_l}$ may be represented as the minimum of r functions $g_1^{l,t}, \dots, g_r^{l,t}$, where $g_i^{l,t}$ is either $\mathbb{1}_{\{T^l f_i \ge t\}}$ or $\mathbb{1} - \mathbb{1}_{\{T^l f_i \ge t\}}$. After r successive applications of (2.3) we obtain

$$\|g_1^{l+n,t} \wedge \dots \wedge g_r^{l+n,t} - T^n g_1^{l,t} \wedge \dots \wedge T^n g_r^{l,t}\| < \varepsilon/2^r$$

for all $l \ge l_0$, $n \in \mathbb{N}$ and $t \in [0,1] \setminus \tau$, provided that in the above formula the functions $g_i^{l+n,t}$ and $g_i^{l,t}$ are of the same type for $i = 1, \dots, r$, i.e.

$$\begin{cases} g_i^{l+n,t} = \mathbb{1}_{\{T^{l+n} f_i \ge t\}}, \\ g_i^{l,t} = \mathbb{1}_{\{T^l f_i \ge t\}}, \end{cases} \quad \text{or} \quad \begin{cases} g_i^{l+n,t} = \mathbb{1} - \mathbb{1}_{\{T^{l+n} f_i \ge t\}}, \\ g_i^{l,t} = \mathbb{1} - \mathbb{1}_{\{T^l f_i \ge t\}}. \end{cases} \tag{2.4}$$

Notice that $g_1^{l,t} \wedge \dots \wedge g_r^{l,t} \le g_i^{l,t}$ for every $i = 1, \dots, r$ and since T is positive we have $T^n(g_1^{l,t} \wedge \dots \wedge g_r^{l,t}) \le T^n g_i^{l,t}$ for each i. Hence

$$\begin{aligned} &T^n(g_1^{l,t} \wedge \dots \wedge g_r^{l,t}) \le T^n g_1^{l,t} \wedge \dots \wedge T^n g_r^{l,t} \\ &\quad \le g_1^{l+n,t} \wedge \dots \wedge g_r^{l+n,t} + |T^n g_1^{l,t} \wedge \dots \wedge T^n g_r^{l,t} - g_1^{l+n,t} \wedge \dots \wedge g_r^{l+n,t}|. \end{aligned}$$

In particular,

$$\int_A (T^n(g_1^{l,t} \wedge \dots \wedge g_r^{l,t}) - g_1^{l+n,t} \wedge \dots \wedge g_r^{l+n,t}) \, d\mu < \varepsilon/2^r \tag{2.5}$$

for any set $A \in \mathscr{B}$.

The sum of all functions of the form $g_1^{l+n,t} \wedge \cdots \wedge g_r^{l+n,t}$ and the sum of all $T^n(g_1^{l,t} \wedge \cdots \wedge g_r^{l,t})$ are equal (they are both $\mathbb{1}$). Suppose that for some pair of functions $f = g_1^{l+n,t} \wedge \cdots \wedge g_r^{l+n,t}$, $\widetilde{f} = T^n(g_1^{l,t} \wedge \cdots \wedge g_r^{l,t})$ we have

$$\int |f - \widetilde{f}|\, d\mu \geq \varepsilon.$$

Since

$$\int |f - \widetilde{f}|\, d\mu = \int\limits_{\{f \geq \widetilde{f}\}} (f - \widetilde{f})\, d\mu + \int\limits_{\{\widetilde{f} > f\}} (\widetilde{f} - f)\, d\mu,$$

and since by (2.5) we have $\int_{\{\widetilde{f} > f\}} (\widetilde{f} - f)\, d\mu < \varepsilon/2^r$, it follows that

$$\int\limits_{\{f \geq \widetilde{f}\}} (f - \widetilde{f})\, d\mu > \varepsilon(1 - 1/2^r).$$

But then summing over the remaining pairs (excluding $\widetilde{f} - f$) we obtain

$$\begin{aligned} 0 = &\int\limits_{\{f \geq \widetilde{f}\}} \sum (T^n(g_1^{l,t} \wedge \cdots \wedge g_r^{l,t}) - g_1^{l+n,t} \wedge \cdots \wedge g_r^{l+n,t})\, d\mu + \int\limits_{\{f \geq \widetilde{f}\}} (\widetilde{f} - f)\, d\mu \\ &< (2^r - 1) \cdot \varepsilon/2^r - \varepsilon(1 - 1/2^r) = 0, \end{aligned}$$

a contradiction.Thus

$$\int |g_1^{l+n,t} \wedge \cdots \wedge g_r^{l+n,t} - T^n(g_1^{l,t} \wedge \cdots \wedge g_r^{l,t})|\, d\mu < \varepsilon$$

for all pairs of functions satisfying (2.4). ∎

THEOREM 2.5. *Let σ be the shift transformation on the space of trajectories $(X^{\mathbb{N}}, \nu)$. For almost all $t \in [0,1]$,*

$$\underline{h}_\mu(T, \mathscr{F}, t) = \overline{h}_\mu(T, \mathscr{F}, t) = \lim_{l \to \infty} h_\nu(\sigma, \mathscr{A}_l^t \times X^{\mathbb{N}}).$$

Proof. Clearly,

$$\overline{h}_\mu(T, \mathscr{F}, t) = \overline{h}_\mu(T, T^l \mathscr{F}, t), \qquad \underline{h}_\mu(T, \mathscr{F}, t) = \underline{h}_\mu(T, T^l \mathscr{F}, t).$$

Fix $\varepsilon > 0$. For any $n, l \in \mathbb{N}$ and $t \in [0,1]$, from standard properties of entropy of a partition (see e.g. [P]) it follows that

$$\begin{aligned} &H_\nu\Big(\bigvee_{i=0}^{n-1} \sigma^{-i}(\mathscr{A}_l^t \times X^{\mathbb{N}})\Big) \\ &\quad \leq H_\nu\Big(\bigvee_{i=0}^{n-1} \mathscr{A}_{l+i}^t \times X^{\mathbb{N}}\Big) + H_\nu\Big(\bigvee_{i=0}^{n-1} \sigma^{-i}(\mathscr{A}_l^t \times X^{\mathbb{N}}) \,\Big|\, \bigvee_{i=0}^{n-1} \mathscr{A}_{l+i}^t \times X^{\mathbb{N}}\Big) \\ &\quad \leq H_\nu\Big(\bigvee_{i=0}^{n-1} \mathscr{A}_{l+i}^t \times X^{\mathbb{N}}\Big) + \sum_{i=0}^{n-1} H_\nu\Big(\underbrace{\sigma^{-i}(\mathscr{A}_l^t \times X^{\mathbb{N}})}_{\mathscr{L}_i} \,\Big|\, \underbrace{\mathscr{A}_{l+i}^t \times X^{\mathbb{N}}}_{\mathscr{P}_i}\Big). \end{aligned}$$

Notice that $\bigvee_{i=0}^{n-1} \mathscr{A}_{l+i}^t \times X^{\mathbb{N}} = \{A \times X^{\mathbb{N}} : A \in \bigvee_{i=0}^{n-1} \mathscr{A}_{l+i}^t\}$ and

$$\nu(A \times X^{\mathbb{N}}) = \int_A \mathbb{1}\, d\mu = \mu(A), \tag{2.6}$$

hence

$$H_\nu\Big(\bigvee_{i=0}^{n-1} \mathscr{A}_{l+i}^t \times X^{\mathbb{N}}\Big) = H_\mu\Big(\bigvee_{i=0}^{n-1} \mathscr{A}_{l+i}^t\Big).$$

We denote by μ_B and $\nu_{B\times X^{\mathbb{N}}}$ the conditional measures of μ and ν with respect to $B \subset X$ and $B \times X^{\mathbb{N}}$, respectively. By (2.6),

$$H_\nu(\mathscr{L}_i \,|\, \mathscr{P}_i) = \sum_{B\in\mathscr{A}_{l+i}^t} \mu(B) \cdot H_{\nu_{B\times X^{\mathbb{N}}}}(\mathscr{L}_i).$$

For any measurable $A, B \subset X$,

$$\nu(B \times X^{i-1} \times A \times X^{\mathbb{N}}) = \int \mathbb{1}_B \cdot T(\mathbb{1} \cdot \ldots \cdot T(\mathbb{1} \cdot T\mathbb{1}_A) \ldots)\, d\mu = \int_B T^i \mathbb{1}_A\, d\mu.$$

Hence

$$H_{\nu_{B\times X^{\mathbb{N}}}}(\mathscr{L}_i) = \sum_{A\in\mathscr{A}_l^t} \eta\Big(\frac{\nu(B \times X^{i-1} \times A \times X^{\mathbb{N}})}{\mu(B)}\Big) = \sum_{A\in\mathscr{A}_l^t} \eta\Big(\int_B T^i \mathbb{1}_A\, d\mu_B\Big).$$

We continue the calculation of the conditional entropy:

$$\begin{aligned} H_\nu(\mathscr{L}_i \,|\, \mathscr{P}_i) &= \sum_{B\in\mathscr{A}_{l+i}^t} \mu(B) \sum_{A\in\mathscr{A}_l^t} \eta\Big(\int T^i \mathbb{1}_A\, d\mu_B\Big) \\ &= -\sum_{B\in\mathscr{A}_{l+i}^t} \sum_{A\in\mathscr{A}_l^t} \Big(\int_B T^i \mathbb{1}_A\, d\mu \cdot \log \frac{\int_B T^i \mathbb{1}_A\, d\mu}{\mu(B)}\Big) \\ &= \sum_{B\in\mathscr{A}_{l+i}^t} \Big(\sum_{A\in\mathscr{A}_l^t} \eta\Big(\int_B T^i \mathbb{1}_A\, d\mu\Big) - \eta(\mu(B))\Big) \\ &= \sum_{B\in\mathscr{A}_{l+i}^t} \sum_{A\in\mathscr{A}_l^t} \eta\Big(\int_B T^i \mathbb{1}_A\, d\mu\Big) - H_\mu(\mathscr{A}_{l+i}^t). \end{aligned}$$

By Lemma 2.4, for any positive $\delta < \varepsilon$ there exists $l_0 \in \mathbb{N}$ and a subset $\tau \subset [0,1]$ with $\lambda(\tau) \le \delta$ such that $\operatorname{dist}(T^i \mathbb{1}_{\mathscr{A}_l^t}, \mathbb{1}_{\mathscr{A}_{l+i}^t}) < \delta$ for every $l > l_0$, $i \in \mathbb{N}$ and $t \in [0,1] \setminus \tau$. Then, for every pair of sets $A \in \mathscr{A}_l^t$, $B \in \mathscr{A}_{l+i}^t$ either $|\mu(B) - \int_B T^i \mathbb{1}_A\, d\mu| < \delta$ or $\int_B T^i \mathbb{1}_A\, d\mu < \delta$, so for a suitable choice of δ (and appropriate l_0 and τ) the entropies of the probability vectors $\{\mu(B) \colon B \in \mathscr{A}_{l+i}^t\}$ and $\{\int_B T^i \mathbb{1}_A\, d\mu \colon A \in \mathscr{A}_l^t, B \in \mathscr{A}_{l+i}^t\}$ differ by less than ε. Finally, we get

$$H_\nu\Big(\bigvee_{i=0}^{n-1} \sigma^{-i}(\mathscr{A}_l^t \times X^{\mathbb{N}})\Big) \le H_\mu\Big(\bigvee_{i=0}^{n-1} \mathscr{A}_{l+i}^t\Big) + n\varepsilon$$

for l large enough, $n \in \mathbb{N}$ and all t except for a set τ of measure ε. Dividing by n and taking the lower limit over n we get

$$h_\nu(\sigma, \mathscr{A}_l^t \times X^{\mathbb{N}}) \le \underline{h}_\mu(T, T^l \mathscr{F}, t) + \varepsilon = \underline{h}_\mu(T, \mathscr{F}, t) + \varepsilon$$

for l large enough and $t \notin \tau$. Notice that on the left hand side the lower limit is a limit, since we compute the entropy of a transformation. Letting l go to infinity, we obtain

$$\limsup_{l\to\infty} h_\nu(\sigma, \mathscr{A}_l^t \times X^{\mathbb{N}}) \le \underline{h}_\mu(T, \mathscr{F}, t) + \varepsilon$$

for $t \in [0,1] \setminus \tau$. Since ε was taken arbitrarily,

$$\limsup_{l\to\infty} h_\nu(\sigma, \mathscr{A}_l^t \times X^{\mathbb{N}}) \le \underline{h}_\mu(T, \mathscr{F}, t)$$

for almost all t.

On the other hand, we analogously estimate

$$H_\mu\Big(\bigvee_{i=0}^{n-1} \mathscr{A}_{l+i}^t\Big) \le H_\nu\Big(\bigvee_{i=0}^{n-1} \sigma^{-i}(\mathscr{A}_l^t \times X^{\mathbb{N}})\Big) + \sum_{i=0}^{n-1} H_\nu(\mathscr{A}_{l+i}^t \times X^{\mathbb{N}} \,|\, \sigma^{-i}(\mathscr{A}_l^t \times X^{\mathbb{N}})). \tag{2.7}$$

Now

$$\begin{aligned} &H_\nu(\mathscr{A}_{l+i}^t \times X^{\mathbb{N}} \,|\, \sigma^{-i}(\mathscr{A}_l^t \times X^{\mathbb{N}})) \\ &= \sum_{A\in\mathscr{A}_l^t} \bigg(\nu(X^{i-1} \times A \times X^{\mathbb{N}}) \sum_{B\in\mathscr{A}_{l+i}^t} \eta\bigg(\frac{\nu(B \times X^{i-2} \times A \times X^{\mathbb{N}})}{\nu(X^{i-1} \times A \times X^{\mathbb{N}})}\bigg)\bigg) \\ &= \sum_{A\in\mathscr{A}_l^t} \underbrace{\int T^i \mathbb{1}_A \, d\mu}_{\mu(A)} \cdot \sum_{B\in\mathscr{A}_{l+i}^t} \eta\bigg(\frac{\int_B T^i \mathbb{1}_A \, d\mu}{\int T^i \mathbb{1}_A \, d\mu}\bigg) \\ &= -\sum_{A\in\mathscr{A}_l^t} \sum_{B\in\mathscr{A}_{l+i}^t} \int_B T^i \mathbb{1}_A \, d\mu \Big(\log \int_B T^i \mathbb{1}_A \, d\mu - \log \mu(A)\Big) \\ &= \sum_{A\in\mathscr{A}_l^t} \sum_{B\in\mathscr{A}_{l+i}^t} \eta\Big(\int_B T^i \mathbb{1}_A \, d\mu\Big) - H_\mu(\mathscr{A}_l^t). \end{aligned} \tag{2.8}$$

Notice that

$$H_\mu(\mathscr{A}_l^t) = \sum_{A\in\mathscr{A}_l^t} \eta\Big(\int T^i \mathbb{1}_A \, d\mu\Big),$$

so, as before, there exist $\tau \subset [0,1]$ with $\lambda(\tau) \le \varepsilon$ and $l_0 \in \mathbb{N}$ such that the right hand side of (2.8) is less than or equal to ε if only $l \ge l_0$ and $t \in [0,1] \setminus \tau$. Dividing both sides of (2.7) by n and taking the upper limit over n we obtain

$$\overline{h}_\mu(T,\mathscr{F},t)=\overline{h}_\mu(T,T^l\mathscr{F},t)\le h_\nu(\sigma,\mathscr{A}_l^t\times X^{\mathbb{N}})+\varepsilon$$

for $t\in[0,1]\setminus\tau$ and so

$$\overline{h}_\mu(T,\mathscr{F},t)\le\liminf_{l\to\infty}h_\nu(\sigma,\mathscr{A}_l^t\times X^{\mathbb{N}})$$

for almost all t. Hence,

$$\begin{aligned}\overline{h}_\mu(T,\mathscr{F},t)&\le\liminf_{l\to\infty}h_\nu(\sigma,\mathscr{A}_l^t\times X^{\mathbb{N}})\\&\le\limsup_{l\to\infty}h_\nu(\sigma,\mathscr{A}_l^t\times X^{\mathbb{N}})\le\underline{h}_\mu(T,\mathscr{F},t)\qquad\lambda\text{-a.e.},\end{aligned}$$

and the assertion follows. ∎

It follows from the above theorem that for almost all t one can define the *entropy of T with respect to $\mathscr{F}$ at level t* by the formula

$$h_\mu(T,\mathscr{F},t)=\lim_{n\to\infty}\frac{1}{n}H_\mu(\mathscr{F}^n,t).\tag{2.9}$$

Corollary 2.6. *We have*

$$h_\mu(T,\mathscr{F})=\int_0^1 h_\mu(T,\mathscr{F},t)\,\lambda(dt)=\lim_{l\to\infty}\int_0^1 h_\nu(\sigma,\mathscr{A}_l^t\times X^{\mathbb{N}})\,\lambda(dt).$$

Proof. Let r be the cardinality of $\mathscr{F}$. Since the entropy of a partition is bounded by the logarithm of its cardinality and the partition $\mathscr{A}_{\mathscr{F}^n}$ consists of at most 2^{nr} elements, the function $n^{-1}H_\mu(\mathscr{F}^n,t)$ is bounded from above by a constant r. By Lebesgue's Dominated Convergence Theorem,

$$\begin{aligned}h_\mu(T,\mathscr{F})&=\lim_{n\to\infty}\frac{1}{n}\int_0^1 H_\mu(\mathscr{F}^n,t)\,\lambda(dt)=\int_0^1\lim_{n\to\infty}\frac{1}{n}H_\mu(\mathscr{F}^n,t)\,\lambda(dt)\\&=\int_0^1 h_\mu(T,\mathscr{F},t)\,\lambda(dt)=\int_0^1\lim_{l\to\infty}h_\nu(\sigma,\mathscr{A}_l^t\times X^{\mathbb{N}})\,\lambda(dt).\end{aligned}$$

As before, the cardinality of $\mathscr{A}_l$, hence of $\mathscr{A}_l^t$, is bounded by 2^r, so that $h_\nu(\sigma,\mathscr{A}_l^t\times X^{\mathbb{N}})\le r$. Using Lebesgue's theorem again, we can interchange the integration and taking the limit to obtain the assertion. ∎

Corollary 2.7. *If T is a doubly stochastic operator and σ the shift transformation on the space of its trajectories then*

$$h_\mu(T)\le h_\nu(\sigma).$$

Proof. Since for any partition $\mathscr{A}$ of X,

$$h_\nu(\sigma,\mathscr{A}\times X^{\mathbb{N}})\le h_\nu(\sigma),$$

the right hand side of the formula in Corollary 2.6 is less than or equal to the entropy of σ. Thus using Corollary 2.6 we obtain

$$h_\mu(T) = \sup_{\mathscr{F}} h_\mu(T, \mathscr{F}) \leq h_\nu(\sigma). \ \blacksquare$$

It is easy to see that equality holds in Corollary 2.7 for pointwise generated operators, because in this case the entropy of an operator is equal to the Kolmogorov–Sinai entropy of a generating transformation (see [DF]), which is isomorphic to the shift on the space of trajectories. On the other hand, equality fails to hold for nonpointwise operators even in the following simple example. Consider a space X consisting of two points and an operator T given by a 2×2 matrix with all entries $1/2$. Since each function is transformed by T to a constant, the entropy $h_\mu(T)$ is 0. But the space of all trajectories constitutes a Bernoulli shift with entropy 1 (recall that we use logarithm to base 2).

Acknowledgements. This research was supported by MENiSW grant N N201 394537 for 2009-2012.

REFERENCES

[AF] R. Alicki and M. Fannes, *Quantum Dynamical Systems*, Oxford Univ. Press, Oxford, 2001.

[CNT] A. Connes, H. Narnhofer and W. Thirring, *Dynamical entropy of C^*-algebras and von Neumann algebras*, Comm. Math. Phys. 112 (1987), 691–719.

[D] J. L. Doob, *A ratio operator limit theorem*, Z. Wahrsch. Verw. Gebiete 1 (1962/1963), 288–294.

[DF] T. Downarowicz and B. Frej, *Measure-theoretic and topological entropy of operators on function spaces*, Ergodic Theory Dynam. Systems 25 (2005), 455–481.

[FF] B. Frej and P. Frej, *An integral formula for entropy of doubly stochastic operators*, Fund. Math. 213 (2011), 271–289.

[GLW] E. Ghys, R. Langevin et P. G. Walczak, *Entropie mesurée et partitions de l'unité*, C. R. Acad. Sci. Paris Sér. I 303 (1986), 251–254.

[H] S. Horowitz, *Transition probabilities and contractions of L_∞*, Z. Wahrsch. Verw. Gebiete 24 (1972), 263–274.

[M] I. I. Makarov, *Dynamical entropy for Markov operators*, J. Dynam. Control Systems 6 (2000), 1–11.

[MR] P. Maličký and B. Riečan, *On the entropy of dynamical systems*, in: Proc. Conf. Ergodic Theory and Related Topics II (Georgenthal, 1986), Teubner-Texte Math. 94, Teubner, Leipzig, 1987, 135–138.

[N] J. Neveu, *Mathematical Foundations of the Calculus of Probability*, Holden-Day, 1965.

[P] K. Petersen, *Ergodic Theory*, Cambridge Univ. Press, 1983.

[V] D. Voiculescu, *Dynamical approximation entropies and topological entropy in operator algebras*, Comm. Math. Phys. 170 (1995), 249–281.

Paulina Frej
Institute of Mathematics and Computer Science
Wrocław University of Technology
Wybrzeże Wyspiańskiego 27
50-370 Wrocław, Poland
E-mail: Paulina.Frej@pwr.wroc.pl

Received 11 July 2011;
revised 27 February 2012 (5525)