Entropy

Let $\alpha=\{A_1,\dots,A_r\}$, $\xi=\{C_1,\dots,C_r\}$ and $\eta=\{D_1,\dots,D_r\}$ be finite partitions of $(X,\mathcal{A},\mu)$ and let T be a measure preserving transformation. That is $\mu(T^{-1}H)=\mu(H)$, $\forall H\in\mathcal{A}$.

A. Definitions:

- $\alpha\vee\xi:=\{A_i\cap C_j : 1\le i,j\le r\}$ is the join of α,ξ.
- $\phi(t):=-t\log t$, $I(\xi)(x):=-\sum\limits_{C\in\xi}\mathbb{1}_C(x)\cdot\log(\mu(C))$.

  $H(\xi):=\int I(\xi)(x)d\mu(x)=\sum\limits_{i=1}^{r}\phi(\mu(C_i))$.
- $H(\xi|\eta):=\sum\limits_{i=1}^{r}\mu(D_j)\cdot H(\xi_{D_j})=\sum\limits_{i,j}\mu(C_i\cap D_j)\log\frac{\mu(C_i\cap D_j)}{\mu(D_j)}$, where ξ_{D_j} is the partition generated by ξ on the set D_j and $H(\xi_{D_j}):=H_{\mu(\cdot|D_j)}(\xi_{D_j})$.
- $I(\widehat{\xi}|\mathcal{B})=-\sum\limits_{i=1}^{r}\mathbb{1}_{C_i}\log\mu(C_i|\mathcal{B})$, where $\mathcal{B}$ is a σ-algebra.
- $h(T,\eta):=\frac{1}{n}H\left(\eta\vee T^{-1}\eta\vee\cdots\vee T^{-(n-1)}\eta\right)$. Actually $H\left(\eta\vee T^{-1}\eta\vee\cdots\vee T^{-(n-1)}\eta\right)\downarrow h(T,\eta)$ also holds.
- $h(T):=\sup\{h(t,\eta):\eta \text{ is a finite partition }\}$.

B. Elementary properties of the entropy of partitions

(I): $H(\xi\vee\eta)=H(\xi|\eta)+H(\eta)$
(II): $H(\xi|\eta)=0$ if and only if η is a refinement of ξ (we denote it $\xi\le\eta$).
(III): $H(\xi\vee\eta)\ge H(\eta)$.
(IV): Let $\widehat{\xi}$ be a finite sub-σ-algebra of $\mathcal{A}$ with $\xi=\{C_1,\dots,C_r\}$. Further, let $\mathcal{F}$ be an arbitrary (not necessarily finite) sub-σ-algebra of $\mathcal{A}$.

$$H(\widehat{\xi}|\mathcal{F}) = \int-\sum_{i=1}^{r}\mathbb{E}(\mathbb{1}_{C_i}|\mathcal{F})\log\mathbb{E}(\mathbb{1}_{C_i}|\mathcal{F})d\mu = \int-\sum_{i=1}^{r}\mathbb{1}_{C_i}\log\mu(C_i|\mathcal{F})d\mu=\int I(\widehat{\xi}|\mathcal{F})d\mu. \tag{1}$$

(V): $H\left(\bigvee_{i=0}^{n-1}T^{-i}\xi\right)=H(\xi)+\sum\limits_{j=1}^{n-1}H\left(\xi|\bigvee_{i=1}^{j}T^{-i}\xi\right)$.

C. Some further elementary properties of the entropy

(a): $H(\alpha\vee\xi|\eta)=H(\alpha|\eta)+H(\xi|\alpha\vee\eta)$.
(b): If $\alpha\le\xi$ then $H(\alpha|\eta)\le H(\xi|\eta)$.
(c): If $\alpha\le\eta$ then $H(\xi|\alpha)\ge H(\xi|\eta)$.
(d): $H(\xi)\ge H(\xi|\eta)$.
(e): $H(\alpha\vee\xi|\eta)\le H(\alpha|\eta)+H(\xi|\eta)$.
(f): $H(\alpha\vee\xi)\le H(\alpha)+H(\xi)$.
(g): $H(\xi|\eta)=H(\xi)$ iff ξ and η are independent: $\mu(C_i\cap D_j)=\mu(C_i)\mu(D_j)$.
(h): $H(T^{-1}\xi|T^{-1}\eta)=H(\xi|\eta)$.
(i): $H(T^{-1}\xi)=H(\xi)$.

D. Elementary properties of the entropy of a transformation

(i): $h(T,\xi)\le H(\xi)$.
(ii): $h(T,\xi\vee\eta)\le h(T,\xi)+h(T,\eta)$.
(iii): If $\eta\le\xi$ then $h(T,\eta)\le h(T,\xi)$.
(iv): $h(T,\xi)\le h(T,\eta)+H(\xi|\eta)$.
(v): $h(T,T^{-1}\xi)=h(T,\xi)$.
(vi): $h(T,\xi)=h\left(T,\vee_{i=0}^{k-1}T^{-i}\xi\right)$ for all $k\ge1$
(vii): If T is invertible and $k\ge1$ then

$$h(T,\xi)=h\left(T,\bigvee_{i=-k}^{k}T^{i}\xi\right).$$

(viii): $h(T^k)=kh(T)$ for $k\ge0$
(ix): $h^{-1}(T)=h(T)$ if T is invertible.

E. Further important properties of the entropy of a transformation

0: Let ν_i be probability measures on $(X, \mathcal{A})$ and $\mathbf{p} = (p_1, \dots, p_\ell)$ be an arbitrary probability vector. That is $\sum_{i=1}^{\ell} p_i = 1$ and $p_i \geq 0$ for all i. We define $\nu := \sum_{i=1}^{\ell} p_i \nu_i$. Then $H_\nu(\xi) \geq \sum_{j=1}^{\ell} p_j \cdot H_{\nu_j}(\xi)$.

1: $h(T, \xi) = \lim_{n\to\infty} H\left(\widehat{\xi} | \bigvee_{i=1}^{n} T^{-i}\widehat{\xi}\right) = H\left(\widehat{\xi} | \bigvee_{i=1}^{\infty} T^{-i}\widehat{\xi}\right)$.

2: $h(T) = 0 \iff \forall \xi$ finite partition, $\widehat{\xi} \overset{\circ}{\subset} \bigvee_{i=1}^{\infty} T^{-i}\widehat{\xi}$

3: $h(T) = 0 \Longrightarrow \ \mathcal{A} = T^{-1}\mathcal{A} \Longrightarrow \ T$ is invertible mod0.

4: $H(\widehat{\xi}|\mathcal{B}) = 0 \iff \widehat{\xi} \overset{\circ}{\subset} \mathcal{B}$

5: For $i = 1, 2$ let $(X_i, \mathcal{A}_i, \mu_i)$ be probability spaces and let T_i be measure preserving on $(X_i, \mathcal{A}_i, \mu_i)$. Then $h(T_1 \times T_2) = h(T_1) + h(T_2)$.

6: The entropy of both of the one-sided and two sided Bernoulli Scheme $BS(p_1, \dots, p_r)$ is $-\sum_{i=1}^{r} p_i \log p_i$.

7: Theorem Ornstein: Two Bernoulli Schemes of the same entropy are isomorphic.

8: The entropy of both of the one-sided and two-sided Markov shifts $(\mathbf{p}, P)$ is $-\sum_{i,j=1} p_i p_{i,j} \log p_{i,j}$.

9: Theorem (Kolomogorov, Sinai) If $\mathcal{B}$ is a finite sub-algebra of $\mathcal{A}$ and if $\bigvee_{i=0}^{\infty} \mathcal{B} \overset{\circ}{=} \mathcal{A}$ then $h(T) = h(T, \mathcal{B})$.

10: Let T be **invertible** and we assume that $\exists\ \mathcal{B} \subset \mathcal{A}$ finite sub-algebra such that $\bigvee_{i=0}^{\infty} T^{-i}\mathcal{B} \overset{\circ}{=} \mathcal{A}$. Then $h(T) = 0$.

11: Theorem (Shannon-McMillian-Breiman) Assume that μ is ergodic and ξ is a finite or countable partition with $H(\xi) < \infty$. Let $\xi_k^n := \bigvee_{i=k}^{n} T^{-i}\xi$ and $\xi_k^n(x)$ be the element of ξ_k^n which contains x. Then

(2) $$I(\xi|\xi_1^n) \to I(\xi|\xi_1^\infty) \text{ a.e. and in } L^1, \text{ and } \frac{1}{n} I\left(\xi_0^{n-1}\right) \to \mathbb{E}\left[f|\mathcal{I}\right] \text{ a.e. and in } L^1,$$

where $f := I(\xi|\xi_1^n)$ and $\mathcal{I} := \left\{H \in \mathcal{A} : T^{-1}H \overset{\circ}{=} H\right\}$. The sequence of functions $h_n(x) := -\frac{1}{n}\log \mu(\xi_0^n(x))$ is convergent almost everywhere and in L^1. The limit is the constant function $h(T, \xi)$ if μ is ergodic.

12: Assume that for $i = 1, 2$ the map T_i is a measure preserving transformation of the Lebesgue space $(X_i, \mathcal{A}_i, \mu_i)$ and $\pi : X_1 \to X_2$ is a measure preserving map such that the following diagram commutes:

$$\begin{array}{ccc} (X_1, \mathcal{A}_1, \mu_1) & \xrightarrow{T_1} & (X_1, \mathcal{A}_1, \mu_1) \\ \pi\downarrow & & \downarrow\pi \\ (X_2, \mathcal{A}_2, \mu_2) & \xrightarrow[T_2]{} & (X_2, \mathcal{A}_2, \mu_2) \end{array}$$

Then we say that T_2 is a factor of T_1. In this case $h(T_1) \geq h(T_2)$.

13: Fix an $r \in \mathbb{N}$. Then

(3) $$\forall \varepsilon > 0 \exists \delta > 0 \text{ s.t. } \forall \xi := \{C_1, \dots, C_r\}, \eta := \{D_1, \dots, D_r\} \text{ if } \rho(\xi, \eta) < \delta \text{ then } d(\xi, \eta) < \varepsilon,$$

where

(4) $$\rho(\xi, \eta) := \sum_{i=1}^{r} \mu(C_i \Delta D_i), \quad d(\xi, \eta) := H(\xi|\eta) + H(\eta|\xi).$$

By (**D**. iv): $|h(T, \xi) - h(T, \eta)| \leq d(\xi, \eta)$. So,

(5) $$\forall \varepsilon > 0 \quad \exists \delta > 0, \text{ s.t. } \rho(\xi, \eta) < \delta \Longrightarrow |h(T, \xi) - h(T, \eta)| < \varepsilon.$$

14: Let A be a non-negative $N \times N$ irreducible matrix and $\lambda, \mathbf{u}, \mathbf{v}$ as in the Perron-Frobenius Theorem. (λ is the dominant eigenvalue and $\mathbf{u}^T \cdot A = \lambda \mathbf{u}^T$, $A \cdot \mathbf{v} = \lambda \cdot \mathbf{v}$ and $\sum_i u_i v_i = 1$.) We define:

(6) $$p_i := u_i v_i \quad p_{ij} := \frac{a_{ij} v_j}{\lambda v_i}.$$

Then $P = (p_{ij})$ is a stochastic matrix and for $\mathbf{p} := (p_1, \dots, p_N)$ we have $\mathbf{p}^T \cdot P = \mathbf{p}^T$. Let $\mu([i_1, \dots, i_n]) := p_{i_1} p_{i_1 i_2} \cdots p_{i_{n-2} i_{n-1}}$. That is μ is the stationary distribution corresponding to the stochastic matrix P. We say that μ is the Parry measure for the topological Markov chain determined by the matrix A. Then

(7) $$h_\mu(\sigma) = \log \lambda,$$

and μ is the only measure with maximal entropy (which is $\log \lambda$).