Ergodic Theory - Week 11

Course Instructor: Florian K. Richter
Teaching assistant: Konstantinos Tsinas

1 Entropy

P1. Prove the following properties of entropy of partitions:

(a) If ξ is a finite partition with r atoms, then $H(\xi) \le \log r$, with equality if and only if $\mu(A) = 1/r$ for any $A \in \xi$.

> Let $\xi = \{A_1, A_2, ..., A_r\}$. Consider the strictly convex function $\phi(x) = x \log\, x$. By Jensen inequality we have that
>
> $$\frac{-\log r}{r} = \phi\left(\frac{1}{r}\sum_{i=1}^{r}\mu(A_i)\right) \le \frac{1}{r}\sum_{i=1}^{r}\phi(\mu(A_i)) = \frac{-H(\xi)}{r}.$$
>
> It follows that $H(\xi) \le \log r$, where the equality holds if and only it holds for the Jensen equality, which happens if and only if all the summands are equal (by strict convexity).

(b) $H(\xi \vee \eta) = H(\eta) + H(\xi \mid \eta)$.

> Let $\xi = \{A_1, A_2, ...\}$ and $\eta = \{B_1, B_2, ...\}$. We have
>
> $$\begin{aligned} H(\xi \vee \eta) &= -\sum_{i,j}\mu(A_i \cap B_j)\log\mu(A_i \cap B_j) \\ &= -\sum_{i,j}\mu(A_i \cap B_j)\log\frac{\mu(A_i \cap B_j)}{\mu(B_j)} - \sum_{i,j}\mu(A_i \cap B_j)\log\mu(B_j) \\ &= -\sum_{j}\mu(B_j)\sum_{i}\mu(A_i \mid B_j)\log\mu(A_i \mid B_j) - \sum_{j}\mu(B_j)\log\mu(B_j) \\ &= H(\xi \mid \eta) + H(\eta). \end{aligned}$$

(c) $H(\xi) \ge H(\xi \mid \eta)$.

> Let $\xi = \{A_1, A_2, ...\}$ and $\eta = \{B_1, B_2, ...\}$. Consider the strictly convex function $\phi(x) =$

$x\log\ x$ and by Jensen's equiality we have

$$\begin{aligned}
H(\xi \mid \eta) &= -\sum_j \mu(B_j) \sum_i \mu(A_i \mid B_j) \log\ \mu(A_i \mid B_j) \\
&= -\sum_i \sum_j \mu(B_j)\phi(\mu(A_i \mid B_j)) \\
&\leq -\sum_i \phi\Big(\sum_j \mu(B_j)\mu(A_i \mid B_j)\Big) \\
&= -\sum_i \phi\Big(\sum_j \mu(A_i \cap B_j)\Big) \\
&= -\sum_i \phi(\mu(A_i)) \\
&= -\sum_i \mu(A_i) \log\ \mu(A_i) \\
&= H(\xi).
\end{aligned}$$

(d) Two partitions ξ and η are independent if and only if $H(\xi \mid \eta) = H(\xi)$.

By c), $H(\xi \mid \eta) = H(\xi)$ if and only if for all i,

$$\mu(A_i \mid B_1) = \mu(A_i \mid B_2) = ... \leq \mu(A_i).$$

Assume that there exists some i_0 such that the inequality above is strict. Then for any fixed j, we have

$$\sum_i \mu(A_i \mid B_j) < \sum_i \mu(A_i) = 1.$$

On the other hand

$$\sum_i \mu(A_i \mid B_j) = \frac{1}{\mu(B_j)} \sum_i \mu(A_i \cap B_j) = \frac{1}{\mu(B_j)} \mu(B_j) = 1,$$

yielding a contradiction. Thus $H(\xi \mid \eta) = H(\xi)$ if and only if for all i, j $\mu(A_i \mid B_j) = \mu(A_i)$, which is equivalent to that ξ and η are independent.

P2. Prove that $d(\xi, \eta) = H(\xi \mid \eta) + H(\eta \mid \xi)$ defines a metric in the space of finite partitions (up to sets of measure 0).

Solution: Firstly, notice that given that $H(\xi \mid \eta), H(\eta \mid \xi) \geq 0$ (because the entropy is positive), we have that $d(\xi, \eta) \geq 0$ for every finite partitions ξ and η. Secondly, we have $d(\xi, \eta) = 0$ if and only if $H(\xi \mid \eta) = H(\eta \mid \xi) = 0$. It follows that $\eta = \xi$. Indeed, the two equalities imply that $H(\xi \vee \eta) = H(\xi) = H(\eta)$. Writinh $\xi = \{A_1, A_2, \ldots\}$ and $\eta = \{B_1, B_2, \ldots\}$, the first equality implies that

$$\sum_i \mu(A_i) \log \mu(A_i) = \sum_i \sum_j \mu(A_i \cap B_j) \log \mu(A_i \cap B_j).$$

Observe that

$$\sum_j \mu(A_i \cap B_j) \log \mu(A_i \cap B_j) \leq \sum_j \mu(A_i \cap B_j) \log \mu(A_i) = \mu(A_i) \log \mu(A_i)$$

and the equality holds if and only if there exists a single j_i such that $\mu(A_i \cap B_{j_i}) = \mu(A_i)$ and $\mu(A_i \cap B_j) = 0$ for all other j. Similarly, the other equality implies that for each j, there exists an i_j, such that $\mu(A_{i_j} \cap B_j) = \mu(B_j)$ and $\mu(A_i \cap B_j) = 0$ for all other i. Thus, for each A_i there exists a unique B_j for which $\mu(A_i \cap B_j) = \mu(A_i) = \mu(B_j)$, thus the partitions ξ, η are the same (up to sets of measure 0). Thirdly, the symmetry is direct from the definition.

Finally, we need to verify the triangle inequality. Let ξ, η, γ finite partitions. Notice that it is enough to show that

$$H(\xi \mid \gamma) \leq H(\xi \mid \eta) + H(\eta \mid \gamma).$$

Indeed, using properties (iii) and (iv) from Theorem 107 from the lecture notes, we have that

$$\begin{aligned}
H(\xi \mid \gamma) &\leq H(\xi \vee \eta \mid \gamma) \\
&= H(\xi \vee \eta \vee \gamma) - H(\gamma) \\
&= H(\xi \mid \eta \vee \gamma) + H(\eta \vee \gamma) - H(\gamma) \\
&= H(\xi \mid \eta \vee \gamma) + H(\eta|\gamma) \\
&\leq H(\xi \mid \eta) + H(\eta|\gamma)
\end{aligned}$$

P3. Let $r \in \mathbb{N}$. Prove that for every $\epsilon > 0$ there is $\delta > 0$ (depending on r, ε) such that $\xi = \{A_1, A_2, \dots, A_r\}$ and $\eta = \{B_1, B_2, \dots, B_r\}$ are two finite partitions with $\sum_{i=1}^r \mu(A_i \Delta B_i) < \delta$, then $d(\xi, \eta) < \epsilon$, where d is the metric defined in Problem 2.

Solution: It is enough to show that $H(\xi \mid \eta) < \epsilon$. For this, consider the partition

$$\gamma = \{A_i \cap B_j\}_{i \neq j} \cup \{\bigcup_{i=1}^r A_i \cap B_i\}.$$

Notice that $\xi \vee \eta = \gamma \vee \eta$, and so,

$$H(\xi \mid \eta) = H(\xi \vee \eta) - H(\eta) = H(\eta \vee \gamma) - H(\eta) = H(\gamma \mid \eta) \leq H(\gamma).$$

Thus, the goal is to show that this partition has small entropy. We see that this partition contains a set of large measure and all the other sets have relatively small measure. This will give that the entropy of this partition is small, which will eventually imply the desired result.

Observe that $A_i \cap B_j \subseteq \bigcup_{k=1}^r A_k \Delta B_k$ for every $i \neq j$. Thus

$$\mu(A_i \cap B_j) \leq \sum_{k=1}^r \mu(A_k \Delta B_k) < \delta.$$

On the other hand

$$\mu(\bigcup_{k=1}^r A_k \cap B_k) = \mu(\bigcup_{k=1}^r A_k \cup B_k) + \mu(\bigcup_{k=1}^r A_k \Delta B_k) > 1 - \delta.$$

Now we calculate the entropy of γ.

$$\begin{aligned}H(\gamma) &= -\sum_{i\neq j}\mu(A_j\cap B_j)\log(\mu(A_i\cap B_j)) - \mu(\bigcup_{i=1}^{r}A_i\cap B_i)\log(\mu(\bigcup_{i=1}^{r}A_i\cap B_i))\\ &= -\sum_{i\neq j}\phi(\mu(A_i\cap B_j)) - \phi(\mu(\bigcup_{i=1}^{r}A_i\cap B_i)),\end{aligned}$$

where $\phi(x) = x\log x$ is strictly decreasing in $(0, \frac{1}{e})$ and strictly increasing in $(\frac{1}{e}, 1)$, so we pick $\delta \in (0, \frac{1}{e})$. Then,

$$H(\gamma) < -r(r-1)\phi(\delta) - \phi(1-\delta) = -r(r-1)\delta\log(1-\delta),$$

which is less than ϵ for $\delta > 0$ small enough.