Exercise sheet – Entropy and Ergodic Theory.
Tel Aviv University, Fall 2017

Conventions. In this exercise sheet some expressions are undefined (e.g. involve division by 0 or log 0) and they should be interpreted following a convention 0 · undefined = 0.

Definitions. A p.p.s. \((X, \mathcal{B}, \mu, T)\) is ergodic if \(A \in \mathcal{B}, T^{-1}(A) = A \implies \mu(A) \in \{0, 1\}\).

For sub-\(\sigma\)-algebras \(\mathcal{A}_1, \mathcal{A}_2\) of \(\mathcal{B}\) as above we say that \(\mathcal{A}_1\) is contained mod \(\mu\) in \(\mathcal{A}_2\) (notation: \(\mathcal{A}_1 \subset_\mu \mathcal{A}_2\)) if for all \(\mathcal{A}_1 \in \mathcal{A}_1\) there is \(\mathcal{A}_2 \in \mathcal{A}_2\) such that \(\mu(A_1 \Delta A_2) = 0\). We say that \(\mathcal{A}_1\) is equivalent mod \(\mu\) to \(\mathcal{A}_2\) (notation: \(\mathcal{A}_1 \equiv_\mu \mathcal{A}_2\)) if \(\mathcal{A}_1 \subset_\mu \mathcal{A}_2\) and \(\mathcal{A}_2 \subset_\mu \mathcal{A}_1\).

1. Let \(n \in \mathbb{N}\) and \(\Delta_n = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : \sum x_i = 1, x_i \geq 0\}, \Delta = \bigcup_n \Delta_n\). Prove that the function \(H : \Delta \to \mathbb{R}\) defined by \(H(\vec{x}) = -\sum_i x_i \log x_i\) for \(\vec{x} = (x_i) \in \Delta_n\), is the only function with the following properties.

   - \(H(\vec{x}) \geq 0\) for all \(\vec{x} \in \Delta\) and \(H(\vec{x}) = 0 \iff \exists i\) such that \(x_i = 1\).
   - \(H\) is continuous and invariant under permutations of the coordinates.
   - The maximum of \(H\) on \(\Delta\) is \(\log n\) and it is achieved only on the point \(x_1 = \cdots = x_n = 1/n\).
   - For all \(n \in \mathbb{N}\) and all \(\vec{x} \in \Delta_n\), \(H(x_1, \ldots, x_n) = H(x_1, \ldots, x_n, 0)\).
   - If \((X, \mathcal{B}, \mu)\) is a probability space, \(\xi, \eta\) are finite partitions of \(X\), and \(H_\mu(\xi) = H((\mu(A_i))_{A_i \in \xi})\), \(H_\mu(\xi|\eta) = \sum B_j \mu(B_j) H_\mu(\xi|\eta)\), where \(\mu|\mathcal{B}(A) = \frac{\mu(B \cap A)}{\mu(B)}\), then we have \(H_\mu(\xi \vee \eta) = H_\mu(\eta) + H_\mu(\xi|\eta)\).

2. Find a sequence \(x_1, x_2, \ldots\) of positive numbers such that \(\sum x_i = 1\) and \(-\sum_i x_i \log x_i = \infty\). Construct an ergodic p.p.s. \((X, \mathcal{B}, \mu, T)\) with \(h_\mu(T) = \infty\).

3. Two countable partitions \(\xi, \eta\) of a probability space \((X, \mathcal{B}, \mu)\) are called independent if

   \[A \in \xi, B \in \eta \implies \mu(A \cap B) = \mu(A)\mu(B)\].

Suppose \(H_\mu(\xi)\) and \(H_\mu(\eta)\) are finite. Show that \(\xi\) and \(\eta\) are independent if and only if \(H_\mu(\xi|\eta) = H_\mu(\xi)\).

For a countable sub-\(\sigma\)-algebra \(\mathcal{C} \subset \mathcal{B}\), we say that \(\xi\) and \(\mathcal{C}\) are independent if

\[A \in \sigma(\xi), C \in \mathcal{C} \implies \mu(A \cap C) = \mu(A)\mu(C)\].
Suppose $H_\mu(\xi) < \infty$ and show that $\xi$ and $C$ are independent if and only if $H_\mu(\xi|C) = H_\mu(\xi)$.

4. Let $\xi = (A_1, A_2, \ldots)$ be a countable infinite partition of a probability space $(X, B, \mu)$ and let $\{0, 1\}^*$ denote all finite length sequences of 0’s and 1’s. A map $S : \mathbb{N} \to \{0, 1\}^*$ is called a prefix-free code if it is injective and if for any $i$, if $S(i) = x_1 \cdots x_\ell$ where $\ell \in \mathbb{N}$ and $x_i \in \{0, 1\}$, for any $k < \ell$ the prefix $x_1 \cdots x_k$ is not in the image of $S$. The length of $w \in \{0, 1\}^*$ is denoted by $|w|$. Set $L(S) = \sum_i \mu(A_i)|S(i)|$. Prove:

- For any prefix-free code $S$, $L(S) \geq \frac{1}{\log 2} H_\mu(\xi)$.
- There exists a prefix-free code $S$ with $L(S) \leq \frac{1}{\log 2}(H_\mu(\xi) + 1)$.

These results are due to Shannon and form part of information theory.

5. Let $X = \{0, 1\}^\mathbb{Z}$ and let $\sigma : X \to X$ be the left shift. Prove that there is an ergodic $\sigma$-invariant measure $\mu$ on $X$ such that $\mu(A) = 0$.

6. Let $(X, B, \mu)$ be a probability space and let $S : X \to X$, $T : X \to X$ be two $B$-measurable maps that preserve $\mu$.

- Show that if $S \circ T = T \circ S$ then for any partition $\xi$, $h_\mu(S \circ T, \xi) \leq h_\mu(S, \xi) + h_\mu(T, \xi)$.
- Give an example of $X, B, \mu, S, T$ as above such that $S \circ T \neq T \circ S$ and $h_\mu(S \circ T, \xi) > h_\mu(S, \xi) + h_\mu(T, \xi)$.

7. (Construction of Markov partitions) Let $T = \mathbb{R}^2/\mathbb{Z}^2$ and let $\pi : \mathbb{R}^2 \to T$ be the natural projection. A parallelogram in $T$ is the image under $\pi$ of a parallelogram in $\mathbb{R}^2$. A hyperbolic automorphism of $T$ is a map $T_A : T \to T$ given by $T_A(\pi(x)) = \pi(Ax)$ where $A$ is a matrix with integer coefficients, determinant $\pm 1$, and $|\text{tr}(A)| > 2$. A parallelogram is called $A$-adapted if its sides are parallel to eigenvectors of $A$. Prove that for any hyperbolic automorphism $T_A$ of $T$ there is a partition $\xi$ of $T$ into finitely many $A$-adapted parallelograms, and there is $c > 0$ such that for each $n \in \mathbb{N}$, each element of $\xi \vee \cdots \vee T_A^{(n-1)}(\xi)$ is an $A$-adapted parallelogram whose sides in one direction have length in $[\frac{1}{c}, c]$ and in the other direction have length in $[\frac{1}{c^2}, c^2]$, where $\lambda$ is the larger eigenvalue of $A$. Conclude that for the Haar measure $m$ on $T$, $h_m(T_A) = \log \lambda$.

8. (Entropy of a pseudo-Anosov map via Markov partitions.) Let $M'$ be the compact subset of $\mathbb{R}^2$ bounded by the polygon with 14 sides shown in Figure 1, and let $M$ be the topological space obtained by identifying pairs of parallel edges in $\partial M'$ by translations, where each side marked with a stairclimber is identified with the side with identical marking, and unmarked sides are identified with the sides on
the opposite side of the polygon ($M$ is an example of a translation surface of genus 3 and is known as the Escher staircase). The points marked $\bullet$ and $\circ$ on $M$ are called singular points. Let $M_0 \subset M$ denote the nonsingular points. Let $\varphi : M \to M$ be a homeomorphism. Using the inclusion $M' \subset \mathbb{R}^2$, if both $x$ and $\varphi(x)$ are in the interior of $M'$, in a neighborhood of $x$ we can think of $\varphi$ as a map on $\mathbb{R}^2$ and use this to define the derivative $d\varphi_x$ as a $2 \times 2$ real matrix. This extends to the case when $x, \varphi(x) \in M_0$ (but may be in $\partial M'$). Similarly we can define derivatives of maps $\mathbb{R}^2 \to M_0$. We say that $\varphi$ is an affine automorphism if $\varphi(M_0) = M_0$, $d\varphi_x$ exists for all $x \in M_0$ and is an invertible matrix $D\varphi$ independent of $x$. A parallelogram in $M_0$ is the image in $M_0$ of a parallelogram in $\mathbb{R}^2$ under a map $\mathbb{R}^2 \to M_0$ whose derivative is the identity. It is $\varphi$-adapted if its sides are parallel to eigenvectors of $D\varphi$.

(i) Exhibit affine automorphisms $\varphi_1, \varphi_2$ of $M$ with $D\varphi_1 = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$, $D\varphi_1 = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$, and conclude that $\varphi = \varphi_1 \circ \varphi_2$ is an affine automorphism where $D\varphi$ satisfies $|\text{tr}(D\varphi)| > 2$.

(ii) Construct a partition $\xi$ of $M_0$ into finitely many $\varphi$-adapted parallelograms, and show there is $c > 0$ such that for each $n \in \mathbb{N}$, each element of $\xi \vee \cdots \vee \varphi^{-n}(\xi)$ is an $A$-adapted parallelogram whose sides in one direction have length in $[\frac{1}{c}, c]$ and in the other direction have length in $[\frac{1}{c\lambda}, \frac{c}{\lambda}]$, where $\lambda$ is the larger eigenvalue of $D\varphi$.
(iii) Show that \( \varphi \) preserves the Lebesgue measure \( \mu \) on \( M \) and compute \( h_\mu(\varphi) \).

9. Find a probability space \( (X, \mathcal{B}, \mu) \) and countably generated sub-\( \sigma \)-algebras \( \mathcal{C}, \mathcal{A}, \mathcal{A}_1, \mathcal{A}_2, \ldots \) of \( \mathcal{B} \) such that \( \mathcal{A}_n \not\subset \mathcal{A} \) but the sequence \( H_\mu(\mathcal{C}|\mathcal{A}_n) \) does not converge to \( H_\mu(\mathcal{C}|\mathcal{A}) \).

10. Let \( (X, \mathcal{B}, \mu) \) be a Borel probability space, \( \mathcal{C}, \mathcal{A} \) countably generated sub-\( \sigma \)-algebras of \( \mathcal{B}, \{\mu^A_x : x \in X'\} \) the conditional measures, where \( X' \subset X \) is conull. Set \( H_\mu(\mathcal{C}) = \infty \) unless \( \mathcal{C} \) is equivalent mod \( \mu^A_x \) with the \( \sigma \)-algebra generated by a finite entropy partition w.r.t. \( \mu^A_x \). Prove that

\[
H_\mu(\mathcal{C}|\mathcal{A}) = \int_{X'} H_\mu(\mathcal{C}) d\mu(x).
\]

11. Let \( (X, \mathcal{B}, \mu, T) \) be an ergodic p.p.s. on a Borel probability space, such that \( \mu(\{x\}) = 0 \) for every \( x \in X \). Let \( \mathcal{E} = \{A \in \mathcal{B} : T^{-1}(A) = A\} \). Prove that \( \mathcal{E} \) is not countably generated.

12. Let \( X \) be a compact metric space, let \( T : X \to X \) be a homeomorphism, and let \( \text{Prob}(X)^T \) denote the \( T \)-invariant Borel probability measures on \( X \). Show that for every \( f \in C(X) \) there is a decreasing sequence \( \varepsilon_k \downarrow 0 \) such that for every \( x \in X \) and every \( N > k \) we have

\[
\inf_{\mu \in \text{Prob}(X)^T} \int_X f d\mu - \varepsilon_k \leq \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) \leq \sup_{\mu \in \text{Prob}(X)^T} \int_X f d\mu + \varepsilon_k.
\]

Conclude that if \( \text{Prob}(X)^T \) contains only one measure \( \mu \), then for every \( x \in X \) and every \( f \in C(X) \) we have

\[
\sum_{n=0}^{N-1} f(T^n x) \to_{N \to \infty} \int_X f d\mu,
\]

this convergence is uniform in \( x \), and \( \mu \) is ergodic. (When \( \text{Prob}(X)^T \) is a singleton we say the system is uniquely ergodic.)

13. Let \( (X, \mathcal{B}, \mu, T) \) be a p.p.s. on a Borel probability space, let \( \mathcal{E} = \{A \in \mathcal{B} : T^{-1}(A) = A\} \) and \( \mathcal{E}' = \{A \in \mathcal{B} : \mu(A \triangle T^{-1}(A)) = 0\} \). Show that \( \mathcal{E} = \mu, \mathcal{E}' \).

Now let \( (X, \mathcal{B}, \mu) \) be a standard Borel space, let \( G \) be a locally compact second countable group, acting on \( X \) by measure preserving transformations. This means that there is a map \( G \times X \to X \), denoted by \( (g, x) \mapsto gx \), satisfying the group action laws \( ex = x, g_1(g_2 x) = (g_1 g_2)x \) for all \( g_1, g_2 \in G \) and \( x \in X \), and preserving the measure, i.e. for any \( A \in \mathcal{B} \) and \( g \in G, \mu(gA) = \mu(A) \). Define \( \mathcal{E} = \{A \in \mathcal{B} : \forall g \in G, g(A) = A\} \) and \( \mathcal{E}' = \{A \in \mathcal{B} : \forall g \in G, \mu(A \triangle g(A)) = 0\} \).
Prove or disprove: $\mathcal{E} = \mu \mathcal{E}'$.

14. Let $(X, d)$ be a compact metric space and let $\text{Prob}(X)$ be the set of regular Borel probability measures on $X$. Define

$$\text{Lip} = \{ f : X \to \mathbb{R} : \forall x_1, x_2 \in X, |f(x_1) - f(x_2)| \leq d(x_1, x_2) \}$$

and for $\mu, \nu \in \text{Prob}(X)$,

$$D(\mu, \nu) = \sup_{f \in \text{Lip}} \left| \int f \, d\mu - \int f \, d\nu \right| .$$

Prove that $D$ is a metric on $\text{Prob}(X)$, and prove or disprove:

- $(\text{Prob}(X), D)$ is a complete metric space.
- If $D(\mu_n, \mu) \to n \to \infty 0$ then $\mu_n \to \mu$ in the weak-* topology.
- If $\mu_n \to \mu$ in the weak-* topology then $D(\mu_n, \mu) \to n \to \infty 0$.

15. Suppose $X$ is a compact metric space and $\mu, \mu_1, \mu_2, \ldots$ are regular Borel probability measures on $X$ such that $\mu_n \to \mu$ in the weak-* topology. Suppose that $A \subset X$ has $\mu(\partial A) = 0$. Provide examples of open and closed sets $A$, and sequences of measures, for which the conclusion fails when we do not assume $\mu(\partial A) = 0$. Does the property ‘for any Borel set $A$ with $\mu(\partial A) = 0$, $\mu_n(A) \to \mu(A)$’ imply that $\mu_n \to \mu$ in the weak-* topology?

16. Let $X, T, \text{Prob}(X)^T$ be as in question 12. We say that $\mu \in \text{Prob}(X)^T$ is maximal for $T$ if $h_\mu(T) = h_{\text{top}}(T)$.

- Show that if $T : X \to X$ is expansive then there is a maximal measure for $T$.
- Show that if $\mu \in \text{Prob}(X)^T$ is the unique maximal measure for $T$, then the p.p.s. $(X, \mathcal{B}, \mu, T)$ is ergodic ($\mathcal{B}$ is the Borel $\sigma$-algebra).