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 $0 \cdot \text{undefined} = 0$.

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-σ-algebras $\mathcal{A}_1, \mathcal{A}_2$ of $\mathcal{B}$ as above we say that $\mathcal{A}_1$ is *contained mod μ in* $\mathcal{A}_2$ (notation: $\mathcal{A}_1 \subset_\mu \mathcal{A}_2$) if for all $A_1 \in \mathcal{A}_1$ there is $A_2 \in \mathcal{A}_2$ such that $\mu(A_1 \triangle A_2) = 0$. We say that $\mathcal{A}_1$ is *equivalent mod μ to* $\mathcal{A}_2$ (notation: $\mathcal{A}_1 =_\mu \mathcal{A}_2$) 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 $\Delta_n = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : \sum x_i = 1, x_i \geqslant 0\}$, $\Delta = \bigcup_n \Delta_n$. 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.

- $H(\vec{x}) \geqslant 0$ for all $\vec{x} \in \Delta$ and
$$H(\vec{x}) = 0 \iff \exists i \text{ such that } x_i = 1.$$
- H is continuous and invariant under permutations of the coordinates.
- The maximum of H on Δ 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, ξ, η are finite partitions of X, and $H_\mu(\xi) = H((\mu(A_i))_{A_i \in \xi})$, $H_\mu(\xi|\eta) = \sum_{B_j \in \eta} \mu(B_j) H_{\mu|_{B_j}}(\xi)$, where $\mu|_B(A) = \frac{\mu(B \cap A)}{\mu(B)}$, 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 $\sum_i x_i = 1$ 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 ξ, η 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 ξ and η are independent if and only if $H_\mu(\xi|\eta) = H_\mu(\xi)$.

For a countable sub-σ-algebra $\mathcal{C} \subset \mathcal{B}$, we say that ξ *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 ξ and $\mathcal{C}$ are independent if and only if $H_\mu(\xi|\mathcal{C}) = H_\mu(\xi)$.

4. Let $\xi = (A_1, A_2, \ldots)$ be a countable infinite partition of a probability space $(X, \mathcal{B}, \mu)$ and let $\{0,1\}^*$ denote all finite length sequences of 0's and 1's. A map $\mathbf{S} : \mathbb{N} \to \{0,1\}^*$ is called a *prefix-free code* if it is injective and if for any i, if $\mathbf{S}(i) = x_1 \cdots x_\ell$ 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 $\mathbf{S}$. The length of $w \in \{0,1\}^*$ is denoted by $|w|$. Set $L(\mathbf{S}) = \sum_i \mu(A_i)|\mathbf{S}(i)|$. Prove:

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

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 σ-invariant measure μ on X such that $\operatorname{supp} \mu = X$ and $h_\mu(\sigma) = 0$.

6. Give an example of a probability space $(X, \mathcal{B}, \mu)$ and of $S : X \to X$, $T : X \to X$, two $\mathcal{B}$-measurable maps that preserve μ, for which $h_\mu(S \circ T, \xi) > h_\mu(S, \xi) + h_\mu(T, \xi)$. Find such an example which also satisfies $S \circ T = T \circ S$.

7. (Construction of Markov partitions) Let $\mathbb{T} = \mathbb{R}^2/\mathbb{Z}^2$ and let $\pi : \mathbb{R}^2 \to \mathbb{T}$ be the natural projection. A *parallelogram* in $\mathbb{T}$ is the image under π of a parallelogram in $\mathbb{R}^2$. A *hyperbolic automorphism* of $\mathbb{T}$ is a map $T_A : \mathbb{T} \to \mathbb{T}$ given by $T_A(\pi(\vec{x})) = \pi(A\vec{x})$ where A is a matrix with integer coefficients, determinant ± 1, and $|\mathrm{tr}(A)| > 2$. A parallelogram is called A-*adapted* if its sides are parallel to eigenvectors of A. Prove that for any hyperbolic automorphism T_A of $\mathbb{T}$ there is a partition ξ of $\mathbb{T}$ 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 $\left[\frac{1}{c}, c\right]$ and in the other direction have length in $\left[\frac{1}{c|\lambda|^n}, \frac{c}{|\lambda|^n}\right]$, where λ is the larger eigenvalue of A (in absolute value). Conclude that for the Haar measure m on $\mathbb{T}$, $h_m(T_A) = \log|\lambda|$.

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

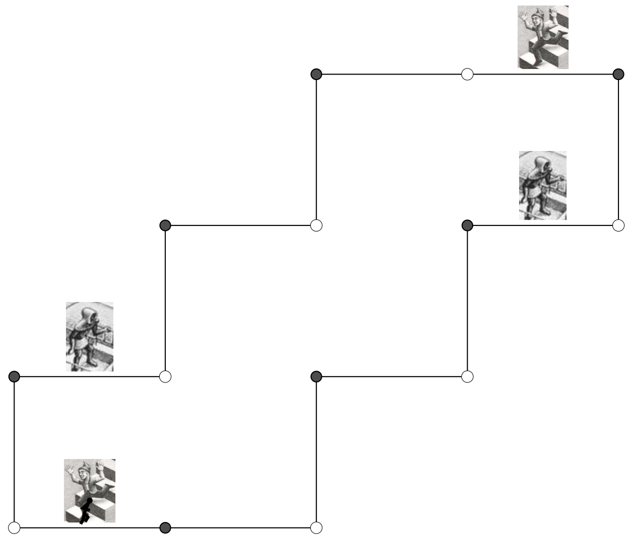

FIGURE 1. The Escher double staircase.

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 φ as a map on $\mathbb{R}^2$ and use this to define the derivative $d\varphi_x$ as a 2×2 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 $U \to M_0$, where $U \subset \mathbb{R}^2$ is open, and use this in order to define derivatives of maps $M_0 \to M_0$. We say that φ 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 P in $\mathbb{R}^2$ under a map $P \to M_0$ whose derivative is the identity. It is *φ-adapted* if its sides are parallel to eigenvectors of $D\varphi$.

(i) Exhibit affine automorphisms φ_1, φ_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 $|\mathrm{tr}(D\varphi)| > 2$.

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

(iii) Show that φ preserves the Lebesgue measure μ on M and compute $h_\mu(\varphi)$.

9. Find a probability space $(X, \mathcal{B}, \mu)$ and countably generated sub-σ-algebras $\mathcal{C}, \mathcal{A}, \mathcal{A}_1, \mathcal{A}_2, \ldots$ of $\mathcal{B}$ such that $\mathcal{A}_n \nearrow \mathcal{A}$ 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-σ-algebras of $\mathcal{B}$, $\{\mu_x^{\mathcal{A}} : x \in X'\}$ the conditional measures, where $X' \subset X$ is conull. Set $H_{\mu_x^{\mathcal{A}}}(\mathcal{C}) = \infty$ unless $\mathcal{C}$ is equivalent mod $\mu_x^{\mathcal{A}}$ with the σ-algebra generated by a finite entropy partition w.r.t. $\mu_x^{\mathcal{A}}$. Prove that

$$H_\mu(\mathcal{C}|\mathcal{A}) = \int_{X'} H_{\mu_x^{\mathcal{A}}}(\mathcal{C}) d\mu(x).$$

11. Let $(X, \mathcal{B}, \mu, T)$ be an invertible 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 $\mathrm{Prob}(X)^T$ denote the T-invariant Borel probability measures on X. Show that for every $f \in C(X)$ there is a decreasing sequence $\varepsilon_k \searrow 0$ such that for every $x \in X$, every $k \in \mathbb{N}$, and every $N > k$ we have

$$\inf_{\mu \in \mathrm{Prob}(X)^T} \int_X f d\mu - \varepsilon_k \leqslant \frac{1}{N} \sum_{n=0}^{N-1} f(T^k x) \leqslant \sup_{\mu \in \mathrm{Prob}(X)^T} \int_X f d\mu + \varepsilon_k.$$

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

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

this convergence is uniform in x, and μ is ergodic. (When $\mathrm{Prob}(X)^T$ 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 continuous 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 $A \in B$ 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 $\mathrm{Prob}(X)$ be the set of regular Borel probability measures on X. Define

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

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

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

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

- $(\mathrm{Prob}(X), D)$ 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 $\mu(\partial A) = 0$. Show that $\mu_n(A) \to \mu(A)$. 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, $\mathrm{Prob}(X)^T$ be as in question 12 and assume $h_{\mathrm{top}}(T) < \infty$. We say that $\mu \in \mathrm{Prob}(X)^T$ is *maximal for* T if $h_\mu(T) = h_{\mathrm{top}}(T)$.

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

17. Let A be an $k \times k$ integer matrix which is diagonalizable over $\mathbb{R}$, with eigenvectors $v_1, \ldots, v_k$. Define a toral automorphism by $X = \mathbb{R}^k/\mathbb{Z}^k$, $\pi : \mathbb{R}^k \to X$ the projection and $T(\pi(x)) = \pi(Ax)$. Let d' be the metric on $\mathbb{R}^k$ given by the sup norm with respect to the basis $v_1, \ldots, v_k$ and define a metric d on X by $d(x, y) = \min\{d'(x', y') : \pi(x') = x, \pi(y') = y\}$. Let d_n be the corresponding Bowen metric. What is the Bowen ball $\{x \in X : d_n(x, x_0) < r\}$ (for $x_0 \in X$ and r smaller than the diameter of X)? Compute $h_{\mathrm{top}}(T)$. Find a measure of maximal entropy. Is it unique?

18. Given a word w with symbols in $\{0,1\}$, define $\sigma(w)$ by replacing each digit x with the other digit $1-x$. Let $w_0 = 1$ and define a sequence of words w_n of length 2^n by letting w_{n+1} be the concatenation of w_n and $\sigma(w_n)$. Thus $w_1 = 10,\ w_2 = 1001,\ w_3 = 10010110, \ldots$ Note that for all $k < 2^n$, the k-th digit of w_n does not depend on n, and let w_∞ be the infinite word whose k-th digit is the k-th digit of each w_n with n large enough. Let $X = \{0,1\}^{\mathbb{N}}$, equipped with the Tychonov topology, let $T : X \to X$ be the one-sided shift, and let X' be the closure of the T-orbit of w_∞. Show that X' is T-invariant and let T' be the restriction of T to X'. Compute $h_{\mathrm{top}}(T')$.

19. Let (X,d) be a compact metric space and let $T : X \to X$ be an isometry, i.e. for all $x, y \in X$, $d(x,y) = d(Tx,Ty)$. Prove that $h_{\mathrm{top}}(T) = 0$.

20. Recall the notation of Host's theorem and its proof: let $X = \mathbb{R}/\mathbb{Z}$ be the torus, $p \geqslant 2$ an integer, $T_p : X \to X$ be the $\times p$ map $T_p(x) = px \mod \mathbb{Z}$, $D_n = \{0, 1/p^n, \ldots, 1 - 1/p^n\}$ the additive subgroup of order p^n, $S_x(y) = x + y \mod \mathbb{Z}$, $\mu \in \mathrm{Prob}(X)^{T_p}$, $\omega_n = \sum_{\alpha\in D_n} S_{\alpha*}\mu$, $\phi_n(x) = \frac{d\mu}{d\omega_n}(x)$. Assume that for μ-a.e. x we have $\phi_n(x)^{1/n} \to_{n\to\infty} \frac{1}{p}$. Prove that μ is Haar (Lebesgue) measure.

21. Let $(X, \mathcal{B}, \mu)$ be a Borel probability space and let $(t,x) \mapsto t.x$ be an action of $\mathbb{R}$ on X. We say that the action is *conservative w.r.t.* μ if for any $B \in \mathcal{B}$ with $\mu(B) > 0$ and any $t_0 > 0$ there is $t > t_0$ such that $\mu(t.B \cap B) > 0$. Suppose $X = \mathbb{T}$, $A \in M_2(\mathbb{Z})$ is a hyperbolic matrix, T_A is the corresponding hyperbolic automorphism as in Problem 7, T_A preserves μ, and $v^- \in \mathbb{R}^2$ is a contracting eigenvector of A. Define an action of $\mathbb{R}$ on X by $t.x = x + tv^-$ (addition mod $\mathbb{Z}^2$ on X). Show that $h_\mu(T_A) > 0$ if and only if the $\mathbb{R}$-action defined above is conservative w.r.t. μ.