ETH Zürich, HS 2022 — Coordinator
Prof. Vincent Tassion — Daniel Contreras

RANDOM WALKS ON TRANSITIVE GRAPHS (D-MATH)
EXERCISE SHEET 9

Definition. Let $\mathcal{X}$ be a finite set. Let $\mathcal{B}$ the set of binary numbers. A *binary code* is a mapping $C : \mathcal{X} \to \mathcal{B}$. For $x \in \mathcal{X}$, let $C(x)$ denote the codeword corresponding to x and let $l(x)$ the length of $C(x)$. We say that C is a *binary prefix code* if $\forall x, y \in \mathcal{X}$, $C(x)$ is not a prefix of $C(y)$ and $C(y)$ is not a prefix of $C(x)$.

For example, $C(\text{red}) = 0$, $C(\text{green}) = 10$, $C(\text{blue}) = 11$ is a binary prefix code for $\mathcal{X} = \{\text{red}, \text{green}, \text{blue}\}$ with codewords $\{0, 10, 11\}$. On the other hand $C(\text{red}) = 0$, $C(\text{green}) = 01$, $C(\text{blue}) = 10$ is a binary code but not a binary prefix code, since 0 is a prefix of 01.

($\star$) Exercise 1. (Kraft inequality) Let $l_1, l_2, \ldots, l_m$ the lengths of the codewords of a binary prefix code $C : \mathcal{X} \to \mathcal{B}$, as defined above. We want to show that

$$\sum_i 2^{-l_i} \leq 1. \tag{1}$$

(i) Consider a binary tree T of height $l_{\max} := \max\{l_1, \ldots, l_m\}$ where each vertex of the tree represents one of the codewords in the standard manner (root $\emptyset$, with children $\{0, 1\}$, each of those having children $\{00, 01\}$, and $\{10, 11\}$ respectively, and so on). Let U be a uniformly distributed random variable on the leaves of T. Let Z given by

$$\mathbb{P}[Z = C(x)] = \mathbb{P}[U \text{ contains a descendent of } C(x)], \text{ for all } x \in \mathcal{X}.$$

Show that Z is a well defined random variable on $C(\mathcal{X})$ and calculate its law. Use this to show (1).

(ii) Show that the converse is also true, that is, given a set of codeword lengths that satisfy (1), there exists a binary prefix code with these word lengths.

($\star$) Exercise 2. Let X be a random variable with values in a finite set $\mathcal{X}$. Let $C : \mathcal{X} \to \mathcal{B}$ be a binary prefix code.

(i) Show that

$$\mathbb{E}[l(X)] \geq H_2(X) := -\sum_{x \in \mathcal{X}} \mathbb{P}[X = x] \log_2(\mathbb{P}[X = x]).$$

with equality if and only if $\mathbb{P}[X = x] = 2^{-l(x)}$, for all $x \in \mathcal{X}$.

(ii) Show that $\mathbb{E}[l(X)] < H_2(X) + 1$.

Definition. Let X be a random variable with values in a finite set $\mathcal{X}$. The *Shannon entropy* of X is defined by

$$H(X) := -\sum_{x\in\mathcal{X}} \mathbb{P}[X = x] \log(\mathbb{P}[X = x]).$$

If Y is a random variable with values in a finite set $\mathcal{Y}$, we define the joint entropy of (X, Y) by

$$H(X, Y) := -\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \mathbb{P}[X = x, Y = y] \log(\mathbb{P}[X = x, Y = y]).$$

Remark. We use the convention $0 \cdot \log(0) = 0$.

($\star$) **Exercise 3.** Let X and Y as in the definition before. Show the following properties:

(i) $H(X) \geq 0$.

(ii) $H(X) \leq \log(|\mathcal{X}|)$ with equality if and only if $X \sim \text{Uniform}(\mathcal{X})$. *Hint*: Use the fact that log is a concave function.

(iii) $H(X, Y) \leq H(X) + H(Y)$ with equality if and only if X, Y are independent.

(iv) $H(X) \leq H(X, Y)$ with equality if and only if Y is $\sigma(X)$-measurable.

($\star$) **Exercise 4.** Let X and Y random variables in a common probability space $(\Omega, \mathcal{A}, \mathbb{P})$, taking values in finite sets $\mathcal{X}$ and $\mathcal{Y}$, respectively. Let $\mathcal{F} \subset \mathcal{A}$ be a sub-σ-algebra. We define the *conditional entropy* of X given $\mathcal{F}$, by

$$H(X \mid \mathcal{F}) := -\mathbb{E}\left[\sum_{x\in\mathcal{X}} \mathbb{P}[X = x \mid \mathcal{F}] \log(\mathbb{P}[X = x \mid \mathcal{F}])\right].$$

(i) Show that $H(X \mid \mathcal{F}) \leq H(X)$ with equality if and only if X is independent of $\mathcal{F}$.

(ii) We define $H(X \mid Y) := H(X \mid \sigma(Y))$. Show that

$$H(X \mid Y) = -\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \mathbb{P}[X = x, Y = y] \log(\mathbb{P}[X = x \mid Y = y]).$$

(iii) Show that $H(X \mid Y) = H(X, Y) - H(Y)$.