

Contents

| | |
|---|---|
| 1. The Khinchin axioms for entropy | 2 |
| 1.1. Entropy axioms | 2 |
| 1.2. Properties of entropy | 2 |
| 2. A special case of Sidorenko's conjecture | 7 |

1. The Khinchin axioms for entropy

Note all random variables we deal with will be discrete, unless otherwise stated. We use $\log = \log_2$.

1.1. Entropy axioms

Definition 1.1 The **entropy** of a discrete random variable X is a quantity $H(X)$ that takes real values and satisfies the **Khinchin axioms**: [Normalisation](#), [Invariance](#), [Extendability](#), [Maximality](#), [Continuity](#) and [Additivity](#).

Axiom 1.2 (Normalisation) If X is uniform on $\{0, 1\}$ (i.e. $X \sim \text{Bern}(1/2)$), then $H(X) = 1$.

Axiom 1.3 (Invariance) If $Y = f(X)$ for some bijection f , then $H(Y) = H(X)$.

Axiom 1.4 (Extendability) If X takes values on a set A , B is disjoint from A , Y takes values in $A \sqcup B$, and for all $a \in A$, $\mathbb{P}(Y = a) = \mathbb{P}(X = a)$, then $H(Y) = H(X)$.

Axiom 1.5 (Maximality) If X takes values in a finite set A and Y is uniformly distributed in A , then $H(X) \leq H(Y)$.

Definition 1.6 The **total variance distance** between X and Y is

$$\sup_E |\mathbb{P}(X \in E) - \mathbb{P}(Y \in E)|.$$

Axiom 1.7 (Continuity) H depends continuously on X (with respect to total variation distance).

Definition 1.8 Let X and Y be random variables. The **conditional entropy** of X given Y is

$$H(X | Y) := \sum_y \mathbb{P}(Y = y) H(X | Y = y).$$

Axiom 1.9 (Additivity) $H(X, Y) := H((X, Y)) = H(Y) + H(X | Y)$.

1.2. Properties of entropy

Lemma 1.10 If X and Y are independent, then $H(X, Y) = H(X) + H(Y)$.

Proof (Hints). Straightforward. □

Proof. $H(X | Y) = \sum_y \mathbb{P}(Y = y) H(X | Y = y)$ Since X and Y are independent, the distribution of X is unaffected by knowing Y , so $H(X | Y = y)$ for all y , which gives the result. (Note we have implicitly used [Invariance](#) here). □

Corollary 1.11 If X_1, \dots, X_n are independent, then

$$H(X_1, \dots, X_n) = H(X_1) + \dots + H(X_n).$$

Proof (Hints). Straightforward. □

Proof. By Lemma [1.10](#) and induction. □

Lemma 1.12 (Chain Rule) Let X_1, \dots, X_n be RVs. Then

$$H(X_1, \dots, X_n) = H(X_1) + H(X_2 \mid X_1) + H(X_3 \mid X_1, X_2) + \dots + H(X_n \mid X_1, \dots, X_{n-1}).$$

Proof (Hints). Straightforward. \square

Proof. The case $n = 2$ is [Additivity](#). In general,

$$H(X_1, \dots, X_n) = H(X_1, \dots, X_{n-1}) + H(X_n \mid X_1, \dots, X_{n-1}),$$

so the result follows by induction. \square

Lemma 1.13 Let X and Y be RVs. If $Y = f(X)$, then $H(X, Y) = H(X)$. Also, $H(Z \mid X, Y) = H(Z \mid X)$.

Proof (Hints). Consider an appropriate bijection. \square

Proof. The map $g : x \mapsto (x, f(x))$ is a bijection, and $(X, Y) = g(X)$, so the first statement follows from [Invariance](#). Also,

$$\begin{aligned} H(Z \mid X, Y) &= H(Z, X, Y) - H(X, Y) \quad \text{by additivity} \\ &= H(Z, X) - H(X) \quad \text{by first part} \\ &= H(Z \mid X) \quad \text{by additivity} \end{aligned}$$

\square

Lemma 1.14 If X takes only one value, then $H(X) = 0$.

Proof (Hints). Use that X and X are independent. \square

Proof. X and X are independent (verify). So by Lemma [1.10](#), $H(X, X) = 2H(X)$. But by [Invariance](#), $H(X, X) = H(X)$. So $H(X) = 0$. \square

Proposition 1.15 If X is uniformly distributed on a set of size 2^n , then $H(X) = n$.

Proof (Hints). Straightforward. \square

Proof. Let X_1, \dots, X_n be independent RVs, uniformly distributed on $\{0, 1\}$. By Corollary [1.11](#) and [Normalisation](#), $H(X_1, \dots, X_n) = n$. So the result follows by [Invariance](#). \square

Proposition 1.16 If X is uniformly distributed on a set A of size n , then $H(X) = \log n$.

Proof (Hints). Straightforward. \square

Proof. Let $r \in \mathbb{N}$ and let X_1, \dots, X_r be independent copies of X . Then (X_1, \dots, X_r) is uniform on A^r , and $H(X_1, \dots, X_r) = rH(X)$. Now pick k such that $2^k \leq n^r \leq 2^{k+1}$. Then by Proposition [1.15](#), [Invariance](#) and [Maximality](#), $k \leq rH(X) \leq k+1$. So $\frac{k}{r} \leq \log n \leq \frac{k+1}{r}$ and $\frac{k}{r} \leq H(X) \leq \frac{k+1}{r}$ for all $r \in \mathbb{N}$. So $H(X) = \log n$, as claimed. \square

Theorem 1.17 (Khinchin) If H satisfies the Khinchin axioms and X takes values in a finite set A , then

$$H(X) = \sum_{a \in A} p_a \log(1/p_a) = \mathbb{E} \left[\log \frac{1}{P_X(X)} \right],$$

where $p_a = \mathbb{P}(X = a)$.

Proof (Hints).

- Explain why it is enough to prove for when the p_a are rational.
- Pick $n \in \mathbb{N}$ such that $p_a = \frac{m_a}{n}$, $m_a \in \mathbb{N}_0$. Let Z be uniform on $[n]$. Let $\{E_a : a \in A\}$ be a partition of $[n]$ into sets with $|E_a| = m_a$.

□

Proof. First we do the case where all $p_a \in \mathbb{Q}$. Pick $n \in \mathbb{N}$ such that $p_a = \frac{m_a}{n}$, $m_a \in \mathbb{N}_0$. Let Z be uniform on $[n]$. Let $\{E_a : a \in A\}$ be a partition of $[n]$ into sets with $|E_a| = m_a$. By **Invariance**, we may assume that $X = a \Leftrightarrow Z \in E_a$. Then

$$\begin{aligned}
 \log n = H(Z) &= H(Z, X) = H(X) + H(Z \mid X) \\
 &= H(X) + \sum_{a \in A} p_a H(Z \mid X = a) \\
 &= H(X) + \sum_{a \in A} p_a \log m_a \\
 &= H(X) + \sum_{a \in A} p_a (\log p_a + \log n) \\
 &= H(X) + \sum_{a \in A} p_a \log p_a + \log n.
 \end{aligned}$$

Hence $H(X) = -\sum_{a \in A} p_a \log p_a$.

The general result follows by **Continuity**.

□

Corollary 1.18 Let X and Y be random variables. Then $0 \leq H(X)$ and $0 \leq H(X \mid Y)$.

Proof (Hints). Trivial.

□

Proof. Immediate consequence of **Khinchin**.

□

Corollary 1.19 If $Y = f(X)$, then $H(Y) \leq H(X)$.

Proof (Hints). Straightforward.

□

Proof. $H(X) = H(X, Y) = H(Y) + H(X \mid Y)$. But $H(X \mid Y) \geq 0$.

□

Proposition 1.20 (Subadditivity) Let X and Y be RVs. Then $H(X, Y) \leq H(X) + H(Y)$.

Proof (Hints).

- Let $p_{ab} = \mathbb{P}(X = a, Y = b)$. Explain why it is enough to show for the case when the p_{ab} are rational.
- Pick n such that $p_{ab} = m_{ab}/n$ with each $m_{ab} \in \mathbb{N}_0$. Partition $[n]$ into sets E_{ab} of size m_{ab} . Let Z be uniform on $[n]$.
- Show that if X (or Y) is uniform, then $H(X \mid Y) \leq H(X)$ and $H(X, Y) \leq H(X) + H(Y)$.
- Let $E_b = \cup_a E_{ab}$ for each b . So $Y = b$ iff $Z \in E_b$. Now define an RV W as follows: if $Y = b$, then W is uniformly distributed in E_b . Use conditional independence to conclude the result.

□

Proof. Note that for any two RVs X, Y ,

$$\begin{aligned} H(X, Y) &\leq H(X) + H(Y) \\ \Leftrightarrow H(X | Y) &\leq H(X) \\ \Leftrightarrow H(Y | X) &\leq H(Y) \end{aligned}$$

by [Additivity](#). Next, observe that $H(X | Y) \leq H(X)$ if X is uniform on a finite set, since $H(X | Y) = \sum_y \mathbb{P}(Y = y) H(X | Y = y) \leq \sum_y \mathbb{P}(Y = y) H(X) = H(X)$ by [Maximality](#). By the above equivalence, we also have $H(X | Y) \leq H(X)$ if Y is uniform on a finite set. Now let $p_{ab} = \mathbb{P}(X = a, Y = b)$, and assume that all p_{ab} are rational. Pick n such that $p_{ab} = m_{ab}/n$ with each $m_{ab} \in \mathbb{N}_0$. Partition $[n]$ into sets E_{ab} of size m_{ab} . Let Z be uniform on $[n]$. WLOG (by [Invariance](#)), $(X, Y) = (a, b)$ iff $Z \in E_{ab}$.

Let $E_b = \cup_a E_{ab}$ for each b . So $Y = b$ iff $Z \in E_b$. Now define an RV W as follows: if $Y = b$, then $W \in E_b$, but then W is uniformly distributed in E_b and independent of X (and Z). So W and X are conditionally independent given Y , and W is uniform on $[n]$. Then $H(X | Y) = H(X | Y, W) = H(X | W)$ by conditional independence and by Lemma [1.13](#) (since W determines Y). Since W is uniform, $H(X | W) \leq H(X)$.

The general result follows by [Continuity](#). □

Corollary 1.21 $H(X) \geq 0$ for any X .

Proof (Hints). (Without using the formula) straightforward. □

Proof. (Without using the formula). By subadditivity, $H(X | X) \leq H(X)$. But $H(X | X) = 0$. □

Corollary 1.22 Let X_1, \dots, X_n be RVs. Then

$$H(X_1, \dots, X_n) \leq H(X_1) + \dots + H(X_n).$$

Proof (Hints). Trivial. □

Proof. Trivial by induction. □

Proposition 1.23 (Submodularity) Let X, Y, Z be RVs. Then

$$H(X | Y, Z) \leq H(X | Z).$$

Proof (Hints). Use that $H(X | Y, Z = z) \leq H(X | Z = z)$. □

Proof.

$$1. H(X | Y, Z) = \sum_z \mathbb{P}(Z = z) H(X | Y, Z = z) \leq \sum_z \mathbb{P}(Z = z) H(X | Z = z) = H(X | Z).$$

□

Remark 1.24 [Submodularity](#) can be expressed in several equivalent ways. Expanding using [Additivity](#) gives

$$H(X, Y, Z) - H(Y, Z) \leq H(X, Z) - H(Z)$$

and

$$H(X, Y, Z) \leq H(X, Z) + H(Y, Z) - H(Z)$$

and

$$H(X, Y, Z) + H(Z) \leq H(X, Z) + H(Y, Z).$$

Lemma 1.25 Let X, Y, Z be RVs with $Z = f(Y)$. Then $H(X | Y) \leq H(X | Z)$.

Proof (Hints). Straightforward. □

Proof. We have

$$\begin{aligned} H(X | Y) &= H(X, Y) - H(Y) = H(X, Y, Z) - H(Y, Z) \\ &\leq H(X, Z) - H(Z) = H(X | Z) \end{aligned}$$

by Submodularity. □

Lemma 1.26 Let X, Y, Z be RVs with $Z = f(X) = g(Y)$. Then

$$H(X, Y) + H(Z) \leq H(X) + H(Y).$$

Proof (Hints). Straightforward. □

Proof. By Submodularity, we have $H(X, Y, Z) + H(Z) \leq H(X, Z) + H(Y, Z)$, which implies the result, since Z depends on X and Y . □

Lemma 1.27 Let X be an RV taking values in a finite set A and let Y be uniform on A . If $H(X) = H(Y)$, then X is uniform.

Proof (Hints). Use Jensen's inequality. □

Proof. Let $p_a = \mathbb{P}(X = a)$. Then

$$H(X) = \sum_{a \in A} p_a \log(1/p_a) = |A| \cdot \mathbb{E}_{a \in A} p_a \log\left(\frac{1}{p_a}\right).$$

The function $x \mapsto x \log(1/x)$ is concave on $[0, 1]$. So by Jensen's inequality,

$$H(X) \leq |A| \cdot (\mathbb{E}_{a \in A} p_a) \cdot \log\left(\frac{1}{\mathbb{E}_{a \in A} p_a}\right) = \log|A| = H(Y),$$

with equality iff $a \mapsto p_a$ is constant, i.e. X is uniform. □

Corollary 1.28 If $H(X, Y) = H(X) + H(Y)$, then X and Y are independent.

Proof (Hints). Go through the proof of Subadditivity and check when equality holds. □

Proof. We go through the proof of subadditivity and check when equality holds. Suppose that X is uniform on A . Then

$$H(X | Y) = \sum_y \mathbb{P}(Y = y) H(X | Y = y) \leq H(X),$$

with equality iff $H(X | Y = y)$ is uniform on A for all y (by Lemma 1.27), which implies that X and Y are independent.

At the last stage of the proof, we said $H(X | Y) = H(X | Y, W) = H(X | W) \leq H(X)$, where W was uniform. So equality holds only if X and W are independent, which implies (since Y depends on W), that X and Y are independent. \square

Definition 1.29 Let X and Y be RVs. The **mutual information**

$$\begin{aligned} I(X : Y) &:= H(X) + H(Y) - H(X, Y) \\ &= H(X) - H(X | Y) \\ &= H(Y) - H(Y | X). \end{aligned}$$

Remark 1.30 Subadditivity is equivalent to the statement that $I(X : Y) \geq 0$, and Corollary 1.28 implies that $I(X : Y) = 0$ iff X and Y are independent.

Note that $H(X, Y) = H(X) + H(Y) - I(X : Y)$ (note the similarity to the inclusion-exclusion formula for two sets).

Definition 1.31 Let X, Y, Z be RVs. The **conditional mutual information** of X and Y given Z is

$$\begin{aligned} I(X : Y | Z) &:= \sum_z \mathbb{P}(Z = z) I(X | Z = z : Y | Z = z) \\ &= \sum_z \mathbb{P}(Z = z) (H(X | Z = z) + H(Y | Z = z) - H(X, Y | Z = z)) \\ &= H(X | Z) + H(Y | Z) - H(X, Y | Z) \\ &= H(X, Z) + H(Y, Z) - H(X, Y, Z) - H(Z). \end{aligned}$$

Submodularity is equivalent to the statement that $I(X : Y | Z) \geq 0$.

2. A special case of Sidorenko's conjecture

Definition 2.1 Let G be a bipartite graph with (finite) vertex sets X and Y and density α (defined to be $\frac{|E(G)|}{|X||Y|}$). Let H be another (think of it as small) bipartite graph with vertex sets U and V and m edges. Now let $\varphi : U \rightarrow X$ and $\psi : V \rightarrow Y$. We say that (φ, ψ) is a **homomorphism** if $\varphi(x)\varphi(y) \in E(G)$ for every edge $xy \in E(H)$.

Conjecture 2.2 (Sidorenko's Conjecture) For every G, H , for random $\varphi : U \rightarrow X, \psi : V \rightarrow Y$,

$$\mathbb{P}((\varphi, \psi) \text{ is a homomorphism}) \geq \alpha^m.$$

Remark 2.3 Sidorenko's Conjecture is not hard to prove when H is the complete bipartite graph $K_{r,s}$ (the case $K_{2,2}$ can be proved using Cauchy-Schwarz: exercise).

Theorem 2.4 Sidorenko's Conjecture is true if H is a path of length 3.

Proof. We want to show that if G is a bipartite graph of density α with vertex sets X, Y of size m and n , and we choose $x_1, x_2 \in X$, $y_1, y_2 \in Y$ independently at random, then $\mathbb{P}(x_1y_1, y_1x_2, x_2y_2 \in E(G)) \geq \alpha^3$.

It would be enough to let P be a path of length 3 chosen uniformly at random and show that $H(P) \geq \log(\alpha^3 m^2 n^2)$ (by Proposition [1.16](#)). Instead, we shall define a different RV taking values in the set of all paths of length 3 (including degenerate paths). To do this, let (X_1, Y_1) be a random edge of G (with $X_1 \in X$, $Y_1 \in Y$). Now let X_2 be a random neighbour of Y_1 and Y_2 be a random neighbour of X_2 . It will be enough to prove that

$$H(X_1, Y_1, X_2, Y_2) \geq \log(\alpha^3 m^2 n^2).$$

□