Decision Problems in Information Theory

EE 667: Information Theory

Divyanshu Bhardwaj - 180253

Table of contents

Introduction

- Constraints on entropy are like ”laws of information theory”.
- This motivates the research of whether the following axioms complete the laws of information theory.
  1. $H(\phi) = 0$
  2. $H(X) \leq H(X \cup Y)$ (Monotonicity)
  3. $H(X) + H(Y) \geq H(X \cap Y) + H(X \cup Y)$ (Submodularity)
- In 1998, Zhang and Yeung answered negatively by finding a inequality that does not follow from the axioms.
- Now we know that there are multiple entropy inequalities that are not captured by the polymatroidal axioms.

- The linear inequalities are classified into Shannon and Non-Shannon inequality.
- While the Shannon inequalities are decidable, the Non-Shannon inequalities are undecidable.
- Now we know there are more important information laws that are non linear - conditional or use max, min.
- The paper initiates a study of algorithmic problem related to proving or disproving information inequalities.
- While provability of linear inequalities remain undecidable, an important question is whether more complex inequalities are any harder.

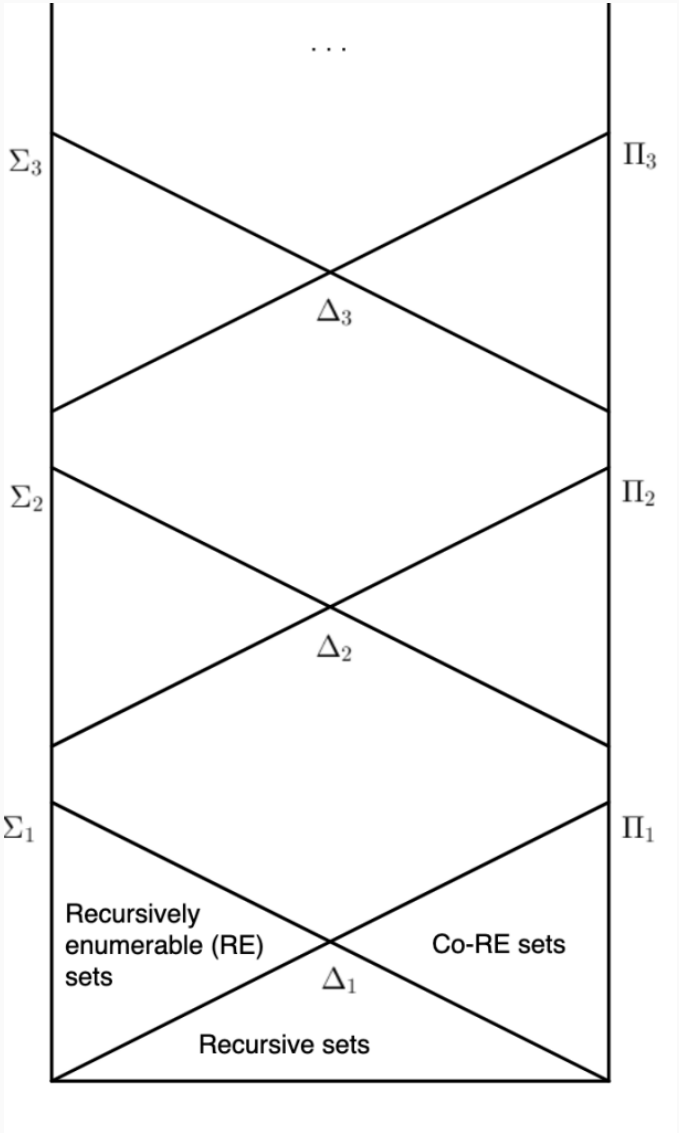
Preliminaries

- Fix a joint distribution over $n-$ random variables $V = \{X_1, \ldots, X_n\}$.
- For each $\alpha \subset [n]$, let X_α denote $(X_i : i \in \alpha)$.
- Define the set function $h : 2^{[n]} \to \mathbb{R}_+$ by $h(\alpha) = H(X_\alpha)$.
- h is entropic function and can be identified with a vector $\boldsymbol{h} = (h(X_\alpha) : \alpha \subset [n]) \in \mathcal{R}_+^{2^n}$ called entropic vector.
- The set of entropic vectors is denoted by $\Gamma_n^* \subset \mathbb{R}_+^{2^n}$.

Turing Machine: It is a computation device proposed by Alan Turing that captures the mathematical idea of solvability of a problem. The Church Turing thesis states that any problem that can be decided by algorithm is decidable by Turing Machine.

- Consider a set S of all the valid information inequalities. If there is an algorithm that decides if a given inequality belongs/does not belong to S, then S is called recursive.
- If there is an algorithm which can decide when any inequality is in S but may or may not decide if inequality is not in S, then S is called recursively enumerable.

- Arithmetic Heirarchy is a framework also known as Kleene-Mostowsky Heirarchy used to study and compare hardness of a problem.
- The class Σ_1^0 contains all recursively enumerable problems and Π_1^0 contains all co-recursively enumerable problems.
- The higher level of classes Σ_n^0 and Π_n^0 are defined using existential($\exists$) and universal($\forall$) quantifiers.
- The problems belonging to same class are considered equal in hardness and the problem in a higher class is considered harder to solve.


...
Σ_3
Π_3
Δ_3
Σ_2
Π_2
Δ_2
Σ_1
Π_1
Recursively enumerable (RE) sets
Co-RE sets
Δ_1
Recursive sets

The classifications Σ_n^0 and Π_n^0 are defined inductively for every natural number n using the following rules:

- If ϕ is logically equivalent to a formula of the form $\exists m_1 \exists m_2 \cdots \exists m_k \psi$, where ψ is Π_n^0, then ϕ is assigned the classification Σ_{n+1}^0.
- If ϕ is logically equivalent to a formula of the form $\forall m_1 \forall m_2 \cdots \forall m_k \psi$, where ψ is Σ_n^0, then ϕ is assigned the classification Π_{n+1}^0.

Problem Definition

- We break the various information inequalities into boolean combination of simpler information inequalities and check whether they are valid.

Definition: A boolean function is a function $F : \{0,1\}^m \to \{0,1\}$. We denote the input variables by $Z_1, \ldots, Z_m$ and $F(Z_1, \ldots, Z_m)$ is the output.
For example $F(Z_1, Z_2) = Z_1 \wedge Z_2$ is a boolean function.

- Consider 2 random variables X_1, X_2. Then

$$\boldsymbol{h} = (H(\phi), H(X_1), H(X_2), H(X_1, X_2))^T \in \mathbb{R}_+^4$$

- Any inequality $c_1 H(X_1) + c_2 H(X_2) + c_{12} H(X_{12}) \geq 0$ can be written as

$$\boldsymbol{c} \cdot \boldsymbol{h} \geq 0 \text{ where } c = [0, c_1, c_2, c_{12}] \in \mathbb{R}^4.$$

- This information inequality is said to be valid if it holds for all $\boldsymbol{h} \in \Gamma_2^*$. We can generalise this to any number of random variables.

Definition: To each Boolean function F with m inputs, and every m vectors $c_j \in \mathbb{R}^{2^n}, j \in [m]$, we associate following boolean information constraint

$$F(\boldsymbol{c}_1 \cdot \boldsymbol{h} \geq 0, \ldots, \boldsymbol{c}_m \cdot \boldsymbol{h} \geq 0)$$

The boolean constraint is said to be valid if it is valid for all $\boldsymbol{h} \in \Gamma_n^*$.

- Consider a 3 random variable system.

$$\boldsymbol{h} = (H(\phi), H(X), H(Y), H(Z), H(XY), H(YZ), H(XZ), H(XYZ))$$

- The following is an information inequality

$$h(XY) \leq \frac{2}{3}h(XYZ) \Rightarrow \max(h(YZ), h(XZ)) \geq \frac{2}{3}h(XYZ)$$

- The boolean constraint, with appropriate c_1, c_2, c_3, for this will be

$$F(z_1, z_2, z_3) := z_1 \Rightarrow z_2 \vee z_3$$

Main Results

Definition: Let F be a Boolean function. The entropic Boolean information constraint problem parameterized by F, denoted by $EBIC(F)$, is the following: given m integer vectors $\boldsymbol{c}_j \in \mathbb{Z}^{2^n}$. Check whether the constraint holds for all entropic vectors $\boldsymbol{h} \in \Gamma_n^*$.

- Let F be a boolean function. Then it can be written as conjunction of clauses i.e. $F = C_1 \wedge C_2 \wedge \ldots$. Checking validity of F is equivalent to checking validity of each C.
- A clause can be written as $C = (Z_1 \wedge Z_2 \ldots \wedge Z_k) \rightarrow (X_1 \vee X_2 \ldots \vee X_l)$.

- Max information inequality with it's equivalent representation:

$$\max\left(\boldsymbol{c}_1 \cdot \boldsymbol{h}, \boldsymbol{c}_2 \cdot \boldsymbol{h}, \ldots, \boldsymbol{c}_m \cdot \boldsymbol{h}\right) \geq 0$$

$$\left(\boldsymbol{c}_1 \cdot \boldsymbol{h} \geq 0\right) \vee \cdots \vee \left(\boldsymbol{c}_m \cdot \boldsymbol{h} \geq 0\right), \text{ where } \boldsymbol{c}_j \in \mathbb{R}^{2^n}$$

- Conditional information inequality will be represented as:

$$\left(\boldsymbol{c}_1 \cdot \boldsymbol{h} \geq 0 \wedge \cdots \wedge \boldsymbol{c}_k \cdot \boldsymbol{h} \geq 0\right) \Rightarrow \boldsymbol{c}_0 \cdot \boldsymbol{h} \geq 0$$

- Conditional independence is represented as:

$$I_h(C; D \mid A) = I_h(C; D \mid B) = I_h(A; B) = I_h(B; C \mid D) = 0 \Longrightarrow I_h(C; D) = 0$$

Theorem 1

7. Let F be a monotone boolean formula. Then $EBIC(F)$ is in Π_1^0 i.e. it is co-recursively enumerable.

Theorem 2

The CI implication problem is in Π_1^0.

Proof

- Fix $F = Z_1 \vee Z_2 \ldots \vee Z_m$ and $\boldsymbol{c_i} \in \mathbb{Z}^{2^n}$. We need to check:

$$\forall \boldsymbol{h} \in \Gamma^* \qquad \boldsymbol{c_1} \cdot \boldsymbol{h} \geq 0 \vee \ldots \vee \boldsymbol{c_m} \cdot \boldsymbol{h} \geq 0$$

Proposition

For every entropic vector $\boldsymbol{h} \in \Gamma_n^*$ and every $\epsilon > 0$, there exists a representable space Ω such that $||\boldsymbol{h} - \boldsymbol{h^\Omega}|| < \epsilon$.

Using this proposition, we claim that it is equivalent to

$$\forall \Omega \qquad \boldsymbol{c_1} \cdot \boldsymbol{h^\Omega} \geq 0 \vee \ldots \vee \boldsymbol{c_m} \cdot \boldsymbol{h^\Omega} \geq 0$$

Finally this is in Π_1^0 since formula after $\forall\Omega$ is decidable and is Σ_0^0 since h^Ω can be written as $\sum a_j log b_j$ where a_j and b_j are rationals.

We will utilise the result by Tarski. It states that the theory of Real numbers with $+, *$ is decidable. For the model $(\mathbb{R}, +, *)$: it is decidable that the formula $\Phi = \forall x \exists y \forall z (x^2 + 3y \geq z \wedge (y^3 + yz \leq xy^2)$ is true.

Consider a conditional inequality over a set of n joint random variables:

$$I_h(Y_1; Z_1|X_1) = 0 \wedge \ldots \wedge I_h(Y_k; Z_k|X_k) = 0 \implies I_h(Y; Z|X) = 0$$

We will see an algorithm that returns false if the inequality is false on any $\boldsymbol{h}$, proving the problem is in $\boldsymbol{\Pi}_0^1$.

- Iterate over all $N \geq 0$ and do following:
- Consider n joint random variables $X_1, \ldots, X_n$ where each has outcome in domain $[N]$.
- There are thus N^n outcomes and let probabilities of these outcomes $p_1, \ldots, p_{N^n}$ real variables respectively.
- Construct a formula $\triangle$ stating "there exists probabilities $p_1, \ldots, p_{N^n}$ for these outcomes whose entropy fails the conditional inequality". More precisely:
  - Convert each conditional independence statement in the antecedent $I_h(Y_i; Z_i | X_i) = 0$ into its equivalent statement on probabilities: $p(X_i Y_i Z_i)p(X_i) = p(X_i Y_i)p(X_i Z_i)$.
  - Replace each such statement with a conjunction of statements of the form
    $p(X_i = x, Y_i = y, Z_i = z)p(X_i = x) = p(X_i = x, Y_i = y)p(X_i = x, Z_i = z)$,
    , for all combinations of values x, y, z. If $X_i Y_i Z_i$ have in total k random variables then there are a total of N^k combinations of X, y, z, thus we create a conjunction of N^k equality statements.

- Each marginal probability is a sum of atomic probabilities, for example $P(X_i = x, Y_i = y) = p_{k_1} + p_{k_2} + \ldots$ where p_{k_i} are the probabilities of all outcomes that have $X_i = x$ and $Y_i = y$. Thus, the equality statement in the previous step becomes the following formula:

  $$(p_{i_1} + p_{i_2} + \ldots)(p_{j_1} + p_{j_2} + \ldots) = (p_{k_1} + p_{k_2} + \ldots)(p_{l_1} + p_{l_2} + \ldots)$$

  There is one such formula for every combination of values x, y, z; denote Φ_i the conjunction of all these formulas. Thus Φ_i asserts $I_h(Y_i; Z_i | X_i) = 0$.
- Let $\Phi = \Phi_1 \wedge \ldots \wedge \Phi_k$. Let Ψ be the similar formula for consequent; thus, Ψ asserts $I_h(Y; Z|X) = 0$.
- Finally, construct the formula $\triangle = \exists p_1, \ldots, \exists p_{N^n} (\Psi \wedge \neg \Phi)$.

- Check whether $\triangle$ is valid in $(\mathbb{R}, +, *)$. By Tarski's theorem, this step is decidable.
- If $\triangle$ is true, then return false, otherwise continue with $N + 1$.

- The implication problem for Conditional Independence statements has been extensively studied in the literature, but its complexity remains an open problem.
- The above theorem is the first upper bound on the complexity of the CI implication problem, placing it in Π^1_0.
- Hannula et al.[1] proved that, if all random variables are restricted to be binary random variables, then the CI implication problem is in EXPSPACE.

[1] Miika Hannula, Åsa Hirvonen, Juha Kontinen, Vadim Kulikov, and Jonni Virtema. Facets of distribution identities in probabilistic team semantics. CoRR, abs/1812.05873, 2018. arXiv:1812.05873

Mahmoud Abo Khamis, Phokion G. Kolaitis, Hung Q. Ngo, and Dan Suci

Decision Problems in Information Theory

47th International Colloquium on Automata, Languages, and Programming (ICALP 2020)

Thank You!