The Recent Amazing Result on Union-Closed Sets Conjecture

Thotsaporn Thanatipanonda
MUIC Math Seminar

February 22, 2023

Level 0. Statement of the Conjecture

The union closed conjecture is a well-known conjecture in combinatorics.

Definition 1 (Union closed set system)**.** A set system $\mathcal{F}$ is union closed if for all $A, B \in \mathcal{F}$ we have $A \cup B \in \mathcal{F}$.

Frankl conjectured that if a family of sets is union-closed, it must have at least one element (or number) that appears in at least half the sets of $\mathcal{F}$ (50% density bound).

Example 1. *Consider a union-closed family*

$$\mathcal{F} = \{\{2\}, \{1,3\}, \{2,4\}, \{1,2,3\}, \{1,2,3,4\}\}.$$

We see that there is an element that appears in at least half the sets of $\mathcal{F}$.

The 50% density bound was a natural threshold for two reasons.

- First, there are readily available examples of union-closed families in which all elements appear in exactly 50% of the sets. Like all the different sets you can make from the numbers 1 to 10, for instance. There are 1,024 such sets, which form a union-closed family, and each of the 10 elements appears in 512 of them.
- And second, at the time Frankl made the conjecture no one had ever produced an example of a union-closed family in which the conjecture didn't hold.

So 50% seemed like the right prediction.

The result we are presenting by Justin Gilmer (2022) is the first known constant lower bound

"There exists an $i \in [n](= \{1, 2, \ldots, n\})$ which is contained in 1% of the sets in $\mathcal{F}$".

The proof makes a clever use of the properties of entropy function and improves upon the $\Omega\left(\frac{1}{\log_2 |\mathcal{F}|}\right)$ bounds of Knill and Wojick.

0.1 Entropy Functions

Information theory developed in the first half of the 20th century, most famously with Claude Shannon's 1948 paper, "A Mathematical Theory of Communication." The paper provided a precise way of calculating the amount of information needed to send a message, based on the amount of uncertainty around what exactly the message would say. This link *between information and uncertainty* was Shannon's remarkable, fundamental insight.

Definition 2. Let X be a random variable taking values in some range B. Let p_b denote the probability that the value X is b. The *binary entropy* of X, denoted by $H(X)$ is just the expected information gain of X:

$$H(X) := \sum_{b \in B} p_b \log_2 (1/p_b).$$

Example 2. *Let X be Bernoulli random variables with probability p. Then*

$$H(p) := p\log_2(1/p) + (1-p)\log_2(1/(1-p)).$$

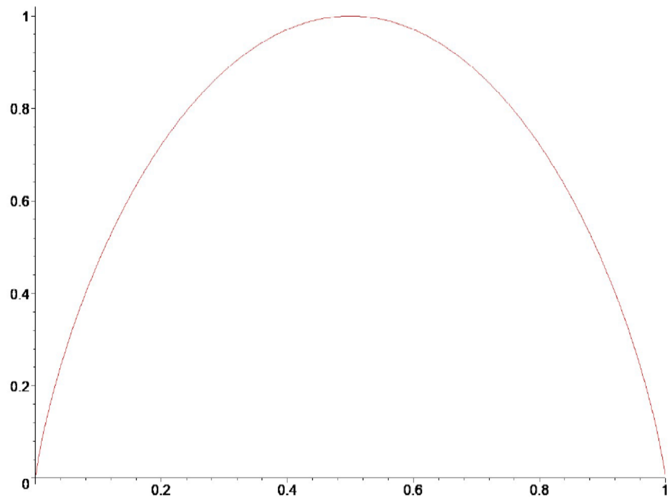

Figure 1: Graph of $H(p)$

The maximum of $H(p)$ is 1 at $p = 1/2$, while the minimum of $H(p)$ is 0 at $p = 0, 1$.

Example 3. *Given a distribution over subsets of* $[2]$ *as*

$$p_\emptyset = 0.2,\ p_{\{1\}} = 0.1,\ p_{\{2\}} = 0.5,\ p_{\{1,2\}} = 0.2.$$

Let A *be a sample from this distribution, then*

$$\begin{aligned} H(A) &= 0.2\log_2(1/0.2) + 0.1\log_2(1/0.1) + 0.5\log_2(1/0.5) + 0.2\log_2(1/0.2) \\ &= 1.760964047. \end{aligned}$$

On the other hand, if

$$p_\emptyset = p_{\{1\}} = p_{\{2\}} = p_{\{1,2\}} = 0.25,$$

then

$$H(A) = 4 \cdot 0.25\log_2(1/0.25) = \log_2 4 = 2.$$

Level 1. Outline of Gilmer's Proof

1.1 Main Theorem

Theorem 1. *Let A and B denote independent samples from a distribution over subsets of $[n]$. Assume that for all $i \in [n]$, $Pr[i \in A] \leq 0.01$. Then $H(A \cup B) \geq 1.26H(A)$.*

Theorem 1 will lead to a contradiction as we will show below. Then we conclude that there must be an element $i \in [n]$, such that $Pr[i \in A] > 0.01$.

- When $H(A) > 0$, Theorem 1 implies that $H(A \cup B) > H(A)$.
- However if we sample A, B independently and uniformly at random from a union-closed family $\mathcal{F}$, then $H(A \cup B) \leq H(A)$.

  This follows because $A \cup B$ is a distribution over $\mathcal{F}$ and the entropy of a distribution over $\mathcal{F}$ is maximized when it is the uniform distribution.

Example 4. *Consider $n = 4$, and a union-closed family $\mathcal{F} = \{\{2\}, \{1,3\}, \{2,4\}, \{1,2,3\}, \{1,2,3,4\}\}$. Assign the probability to each element of $\mathcal{F}$ uniformly. Then*

$$H(A) = \frac{1}{5}\log_2 5 + \cdots + \frac{1}{5}\log_2 5 = \log_2 5.$$

Meanwhile, the probability distribution of $A \cup B$ goes as $p_{\{2\}} = 1/25, p_{\{1,3\}} = 1/25, p_{\{2,4\}} = 3/25, p_{\{1,2,3\}} = 7/25, p_{\{1,2,3,4\}} = 13/25$. Hence

$$H(A \cup B) = \frac{1}{25}\log_2 25 + \cdots + \frac{13}{25}\log_2 (25/13) = 1.743372658\ldots < H(A).$$

1.2 First property of entropy functions

Property (0): Let $\mathcal{F} \subseteq \mathcal{B}$,

$$\sum_{b \in \mathcal{F}} p_b \log_2 1/p_b \leq \log_2 |\mathcal{F}|.$$

Moreover if $p_b = \dfrac{1}{|\mathcal{F}|}$ for all $b \in \mathcal{F}$ then $H(X) = \log_2 |\mathcal{F}|$.

Proof. We apply Jensen's Inequality, basing on the fact that $\log_2 x$ is concave.

$$\sum_{b \in \mathcal{F}} p_b \log_2 1/p_b \leq \log_2 \left(\sum_{b \in \mathcal{F}} p_b \cdot 1/p_b \right) = \log_2 |\mathcal{F}|.$$

$\square$

1.3 Result!

We then obtain as an immediate corollary from Theorem 1.

Theorem 2. *Let* $\mathcal{F} \subseteq 2^{[n]}$ *be a union-closed family,* $\mathcal{F} \neq \{\emptyset\}$*. Then there exists* $i \in [n]$ *that is contained in at least 1% of the sets in* $\mathcal{F}$*.*

Level 2. Proof of Theorem 1

2.1 More Examples

We could consider when A is a n-binary bits string. We use A as a short hand to represent the element in $\mathcal{F}$. For example $(0,0) \to \emptyset, (1,0) \to \{1\}, (0,1) \to \{2\}$ and $(1,1) \to \{1,2\}$.

Example 5. *Let $A = (A_1, A_2)$ be a random subset of $[2]$ such that each A_i are iid Bernoulli random variables with probability p. Then*

$$\begin{aligned}
H(A) &= p^2 \log_2 \frac{1}{p^2} + 2p(1-p) \log_2 \frac{1}{p(1-p)} + (1-p)^2 \log_2 \frac{1}{(1-p)^2} \\
&= 2p^2 \log_2 \frac{1}{p} + 2p(1-p) \left[\log_2 \frac{1}{p} + \log_2 \frac{1}{1-p}\right] + 2(1-p)^2 \log_2 \frac{1}{(1-p)} \\
&= 2p \log_2 \frac{1}{p} + 2(1-p) \log_2 \frac{1}{(1-p)} = 2H(p).
\end{aligned}$$

In general if $A = (A_1, A_2, \ldots, A_n)$ is a random subset of $[n]$ such that each A_i are iid Bernoulli random variables with probability p. Then $H(A) = nH(p)$.

We remark that we could not apply the linearity of expectation $H(A) = H(A_1) + H(A_2) + \cdots + H(A_n)$ for the entropy function, why?

The direct calculation of $H(A)$ is given below:

$$
\begin{aligned}
H(A) &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \left[\log_2 \frac{1}{p^k} + \log_2 \frac{1}{(1-p)^{n-k}}\right] \\
&= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \log_2 \frac{1}{p} + \sum_{k=0}^{n} (n-k) \binom{n}{k} p^k (1-p)^{n-k} \log_2 \frac{1}{1-p} \\
&= \left(n \log_2 \frac{1}{p}\right) \sum_{k=0}^{n} \binom{n-1}{k-1} p^k (1-p)^{n-k} + \left(n \log_2 \frac{1}{1-p}\right) \sum_{k=0}^{n} \binom{n-1}{k} p^k (1-p)^{n-k} \\
&= np \log_2 \frac{1}{p} + n(1-p) \log_2 \frac{1}{1-p} = nH(p).
\end{aligned}
$$

The followings are the motivational examples to the proof of union-closed sets conjecture.

Example 6. *Let A and B be two independent samples from the same Bernoulli distribution, i.e. $A = [n]$ with probability p and $A = \emptyset$ with probability $1 - p$. Then $H(A) = H(p)$ and $H(A \cup B) = H(2p - p^2)$.*

In terms of of binary digits, we use $A \cup B$ to denote $\max(A, B)$. Also

$$Pr(A \cup B = [n]) = Pr(A = [n] \text{ or } B = [n]) = p^2 + 2p(1-p) = p(p+2-2p) = p(2-p).$$

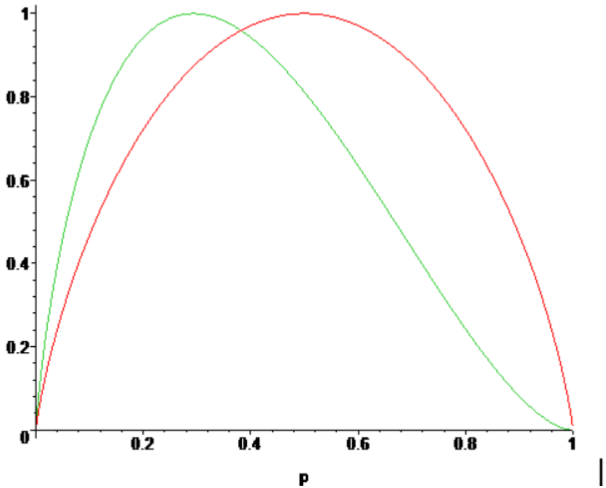

Figure 2: Graph of $H(p)$ and $H(2p - p^2)$

Note that $H(p)$ and $H(2p-p^2)$ intersect at $p = \dfrac{3-\sqrt{5}}{2} \approx 0.38196601125010\ldots$. That is $H(A \cup B) > H(A)$ if $p < 0.38196601125010$.

Example 7. *Let* $A = (A_1, A_2, \ldots, A_n)$ *be a random subset of* $[n]$ *such that each* A_i *are iid Bernoulli random variables with probability* p*. Then* $H(A) = nH(p)$ *and* $H(A \cup B) = nH(2p - p^2)$*.*

Remark 1: In the last example, we notice that

$$H(A \cup B) > H(A) \text{ if } p = Pr[A_i = 1] < \frac{3 - \sqrt{5}}{2} = 0.38196601125010.$$

Hence it might be possible that 1% density bound could be improved to 38.1966%, but no more.

2.2 More Definitions and Properties of entropy Functions

Here is a good point to mention about some useful definitions and properties of the entropy functions H that involves in the proof.

Definition 3. If E is some event, the *conditional entropy* of X given E is naturally defined by

$$H(X|E) = \sum_{b \in \mathcal{F}} Pr(X = b \,|\, E) \log_2 (1/Pr(X = b \,|\, E)).$$

Definition 4. If Y is another random variable taking values in some range A, the conditional entropy of X given Y is defined by

$$H(X|Y) = \sum_{a \in A} H(X \,|\, Y = a) \cdot Pr(Y = a).$$

We could think of $H(X \mid Y)$ as a weighted average of $H(X \mid Y = a)$ over all the possible values of Y. The following properties of $H(X \mid Y)$ are given without proof. The reader can prove it (by applying definitions) as an exercise.

1. $H(X \mid X) = 0$.
2. $H(X \mid Y) = 0$ if and only if $X = f(Y)$ for some function f.
3. $H(X \mid Y) = H(X)$ if and only if X and Y are independent.

The list of the main properties that we are going to use is given below:

1. $H(X, Y) = H(Y) + H(X \mid Y)$
2. $H(X_1, X_2, \ldots, X_n) \leq H(X_1) + H(X_2) + \cdots + H(X_n)$ with equality only holding when $X_1, X_2, \ldots, X_n$ are mutually independent.
3. **Chain Rule for Entropy:** For a sequence of random variables $X_1, \ldots, X_n$, denote $X_{<i} = (X_1, \ldots, X_{i-1})$. Then

$$\begin{aligned} H(X_1, \ldots, X_n) &= \sum_i H(X_i | X_{<i}) \\ &= H(X_1) + H(X_2 \mid X_1) + H(X_3 \mid X_1, X_2) + \ldots. \end{aligned}$$

4. For random variables X and Y and a function $f(Y)$,

$$H(X|Y) \leq H(X|f(Y)).$$

Note that we have seen property (2) holds with equality in Examples 5 and 7 earlier.

Proof. 1. Recall that

$$H(X,Y) = \sum_{a\in A}\sum_{b\in B} Pr(X=b, Y=a)\cdot \log_2 \frac{1}{Pr(X=b,Y=a)}.$$

and

$$\begin{aligned}
H(X\,|\,Y) &= \sum_{a\in A}\sum_{b\in B} Pr(X=b\,|\,Y=a)\cdot \log_2 \frac{1}{Pr(X=b\,|\,Y=a)}\cdot Pr(Y=a) \\
&= \sum_{a\in A}\sum_{b\in B} Pr(X=b, Y=a)\cdot \log_2 \frac{P(Y=a)}{Pr(X=b,Y=a)}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
H(X,Y) - H(X\,|\,Y) &= \sum_{a\in A}\sum_{b\in B} Pr(X=b, Y=a)\cdot \log_2 \frac{1}{P(Y=a)} \\
&= \sum_{a\in A} Pr(Y=a)\cdot \log_2 \frac{1}{P(Y=a)} = H(Y).
\end{aligned}$$

2. We show only $H(X,Y) \leq H(X) + H(Y)$. The rest will follow from induction.

$$\begin{aligned}
H(X) + H(Y) &= \sum_{b \in B} Pr(X=b) \cdot \log_2 \frac{1}{P(X=b)} + \sum_{a \in A} Pr(Y=a) \cdot \log_2 \frac{1}{P(Y=a)} \\
&= \sum_{a \in A} \sum_{b \in B} Pr(X=b, Y=a) \cdot \log_2 \frac{1}{P(X=b)} \\
&+ \sum_{a \in A} \sum_{b \in B} Pr(X=b, Y=a) \cdot \log_2 \frac{1}{P(Y=a)} \\
&= \sum_{a \in A} \sum_{b \in B} Pr(X=b, Y=a) \cdot \log_2 \frac{1}{P(X=b)P(Y=a)}.
\end{aligned}$$

Hence, if X and Y are independent then

$$H(X)+H(Y) = \sum_{a \in A} \sum_{b \in B} Pr(X=b, Y=a) \cdot \log_2 \frac{1}{P(X=b, Y=a)} = H(X,Y).$$

Otherwise, as

$$\sum_{a \in A} \sum_{b \in B} P(X=b)P(Y=a) = \sum_{a \in A} \sum_{b \in B} P(X=b, Y=a),$$

by Gibbs' inequality(below), we conclude that $H(X) + H(Y) \geq H(X, Y)$.

Gibbs' inequality:
Let $s_1, s_2, \ldots, s_n$ and $t_1, t_2, \ldots, t_n$ be positive real numbers such that $\sum_i s_i \leq \sum_i t_i$. Then

$$\sum_{i=1}^{n} t_i \log_2 t_i \geq \sum_{i=1}^{n} t_i \log_2 s_i.$$

3. By property (1),

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

   Then apply property (1) recursively to get the claim.

4. The sequence $X \to Y \to f(Y)$ forms a Markov chain. Thus by the data processing inequality:

$$\begin{aligned} I(X : f(Y)) &\leq I(X : Y) \\ H(X) - H(X \mid f(Y)) &\leq H(X) - H(X \mid Y) \\ H(X \mid Y) &\leq H(X \mid f(Y)). \end{aligned}$$

□

2.3. Lemma and Proof of Theorem 1

Theorem 1 Let A and B denote independent samples from a distribution over subsets of $[n]$. Assume that for all $i \in [n]$, $Pr[i \in A] \leq 0.01$. Then $H(A \cup B) \geq 1.26H(A)$.

Proof of Theorem 1: The proof strategy relies on revealing the bits of $A \cup B$ and A one at a time. For each step, we apply property (4) of entropy function:

$$H((A \cup B)_i \,|\, (A \cup B)_{<i}) \geq H((A \cup B)_i \,|\, A_{<i}, B_{<i}).$$

Then by the Main Lemma (below), we have

$$H((A \cup B)_i \,|\, A_{<i}, B_{<i}) \geq 1.26\, H(A_i \,|\, A_{<i}).$$

We then conclude that, for each $i = 1, \ldots, n$,

$$H((A \cup B)_i \,|\, (A \cup B)_{<i}) \geq 1.26 H(A_i \,|\, A_{<i}).$$

Sum over i and apply the chain rule, property (3), to conclude that $H(A \cup B) \geq 1.26H(A)$.

We will prove the main lemma in Level 3.

Lemma 3 (Main Lemma)**.** *Let C denote a random variable over a finite set S. For each $c \in S$, let p_c be a real number in $[0,1]$. Let X be a Bernoulli random variable sampled according to the following process: first sample $c \in C$, then sample X with $Pr[X = 1 \,|\, C = c] = p_c$. Assume further that $E[X] \leq 0.01$. Let C' be an iid copy of C, and sample X' conditioned on C' according to the same process (so $Pr[X' = 1 \,|\, C' = c] = p_c$, and X' is independent of X and C). Then*

$$H(X \cup X' \,|\, C, C') \geq 1.26 H(X \,|\, C).$$

Here, X, X' correspond to the random bits A_i, B_i respectively, and C, C' correspond to the histories $A_{<i}, B_{<i}$.

Level 3. Proof of the Main Lemma

We can forget all of the structure contained in the random variables $A_{<i}$ and $B_{<i}$. Lemma 3 only assumes that they are iid over some finite set S.

We note that Lemma 3 can be restated a bit more succinctly that assuming $\{p_c\}_{c \in S} \in [0, 1]$ is a finite sequence of real numbers satisfying $\mathbb{E}_c[p_c] \le 0.01$, then

$$\mathbb{E}_{c,c'}[H(p_c + p_{c'} - p_c p_{c'})] \ge 1.26\, \mathbb{E}_c[H(p_c)].$$

We note that this notation makes sense i.e. (Recall that X is a Bernoulli r.v.)

$$E[X] = P(X = 1) = \sum_c P(X = 1|C = c) \cdot P(C = c) = \sum_c p_c \cdot P(C = c) = E_c[p_c].$$

Similarly,

$$H(X\,|C) = \sum_c H(X|C = c) \cdot P(C = c) = \sum_c H(p_c) \cdot P(C = c) = E_c[H(p_c)].$$

3.1 Motivating Example

Example 8. *Sample $A \subseteq [n]$ in the following manner. First sample A_1 from a Bernoulli distribution with probability p. Then, conditioned on the event that $A_1 = 1$, sample each A_i from iid Bernoulli distributions with probability $q = 0.99$. Otherwise, if $A_1 = 0$ then each $A_i = 0$. To calculate $H(A)$, we apply the chain rule to get $H(A) = H(A_1, A_{>1}) = H(A_1) + H(A_{>1} \mid A_1)$. The conditional entropy can be computed as*

$$H(A_{>1} \mid A_1) = Pr[A_1 = 0] \cdot 0 + Pr[A_1 = 1] \cdot H(q)(n-1).$$

Thus $H(A) = H(p) + pH(q)(n-1)$. Via a similar calculation we get $H(A \cup B) = H(2p - p^2) + 2p(1-p)H(q)(n-1) + p^2 H(2q - q^2)(n-1)$.

Note from Example 8:

In this example, for n large and p small, $H(A \cup B)$ is dominated by the term $2p(1-p)H(q)(n-1)$. This corresponds to the event that exactly one of A_1, B_1 is equal to 1. It follows that $\dfrac{H(A \cup B)}{H(A)} \approx 2(1-p)$. Note in this case, the entropy $H(A \cup B \mid A_1 = B_1 = 1)$ is small relative to $H(A \mid A_1 = 1)$, i.e. $H(A \cup B \mid A_1 = B_1 = 1) = H(q(2-q))(n-1)$ and $H(A \mid A_1 = 1) = H(q)(n-1)$, where $H(q(2-q)) = 0.001473033527$ and $H(q) = 0.08079313575$.

3.2 Proof Sketch

We let $\mathcal{C}_0 = \{c \mid p_c \leq 0.1\}$ and let $\mathcal{C}_1 = \mathcal{C}_0^c$.

Using the assumption that $E[X] \leq 0.01$. we apply Markov's inequality to get that

$$Pr[c \in \mathcal{C}_1] = Pr[p_c > 0.1] \leq \frac{E_c[p_c]}{0.1} \leq 0.1.$$

This implies that $Pr[c \in \mathcal{C}_0] \geq 0.9$.

Details

We want to show that

$$H(X \cup X' \mid C, C') \geq 1.26 H(X \mid C),$$

under the assumptions that $E[X]$ and $E[X'] \leq 0.01$.

We write $H(X \cup X' \,|\, C, C')$ and $H(X \,|\, C)$ as a sum of disjoint events:

$$H(X \,|\, C) = H(X \,|\, C \in \mathcal{C}_0) \cdot Pr(C \in \mathcal{C}_0) + H(X \,|\, C \in \mathcal{C}_1) \cdot Pr(C \in \mathcal{C}_1).$$

and

$$\begin{aligned} H(X \cup X' \,|\, C, C') = {} & H(X \cup X' \,|\, C, C' \in \mathcal{C}_0) \cdot Pr(C, C' \in \mathcal{C}_0) \\ & + 2H(X \cup X' \,|\, C \in \mathcal{C}_0, C' \in \mathcal{C}_1) \cdot Pr(C \in \mathcal{C}_0, C' \in \mathcal{C}_1) \\ & + H(X \cup X' \,|\, C, C' \in \mathcal{C}_1) \cdot Pr(C, C' \in \mathcal{C}_1). \end{aligned}$$

We will show that

(1) $H(X \cup X' \,|\, C, C' \in \mathcal{C}_0) \cdot Pr(C, C' \in \mathcal{C}_0) \geq 1.26 H(X \,|\, C \in \mathcal{C}_0) \cdot Pr(C \in \mathcal{C}_0)$

(2) $2H(X \cup X' \,|\, C \in \mathcal{C}_0, C' \in \mathcal{C}_1) \geq 1.62 H(X \,|\, C \in \mathcal{C}_1) \cdot Pr(C \in \mathcal{C}_1)$

(3) $H(X \cup X' \,|\, C, C' \in \mathcal{C}_1) \cdot Pr(C, C' \in \mathcal{C}_1) \geq 0.$

Adding up these three events, we get the claimed statement of Lemma 3. Event (3) is trivial. Event (1) and (2) are based on the Lemmas below.

Given that both $C, C' \in \mathcal{C}_0$, the entropy $H(X \cup X')$ will be a constant factor larger than $\dfrac{H(X) + H(X')}{2}$.

Lemma 4. *Assume $p, p' \leq 0.1$. Then* $H(p + p' - pp') \geq 1.4 \left(\dfrac{H(p) + H(p')}{2} \right)$.

The plot from Maple is kind of verify this statement numerically.

In the event that exactly one of $c, c' \in \mathcal{C}_0$ we can show that $H(X \cup X') \geq 0.9\, H(X)$.

Lemma 5. *For any $p, p' \in [0, 1]$,* $H(p + p' - pp') \geq (1 - p)H(p')$.

Proof. We apply the Jensen's inequality from the concavity of H.

$$H(p \cdot 1 + (1 - p)p') \geq pH(1) + (1 - p)H(p') = (1 - p)H(p').$$

□

References

[1] Thomas M. Cover, Joy A. Thomas *Elements of Information Theory,* A John Wiley & Sons, 2nd edition.

[2] Justin Gilmer, *A constant lower bound for the union-closed sets conjecture*, arXiv:2211.09055v2

[3] Stasys Jukna, *Extremal Combinatorics,* Springer, 1st edition.