Shannon Entropy, Counting, and Shearer's Inequalities

Igal Sason, Technion - Israel Institute of Technology

December 16, 2024

Combinatorics Seminar
Einstein Institute of Mathematics
Hebrew University of Jerusalem

Shannon Entropy

Definition 1.1 (Shannon Entropy)

Let X be a discrete random variable, and let P_X denote its probability mass function (PMF) defined on a set $\mathcal{X}$. Then, the Shannon entropy of X is given by

$$\mathrm{H}(X) = -\sum_{x \in \mathcal{X}} \mathsf{P}_X(x) \log \mathsf{P}_X(x). \tag{1.1}$$

Throughout, logarithms are on base 2.

Definition 1.2 (Conditional Entropy)

Let X, Y be discrete random variables, and let $\mathsf{P}_{X,Y}$ denote its joint PMF defined on a set $\mathcal{X} \times \mathcal{Y}$. Then, the conditional entropy of X given Y is defined as

$$\mathrm{H}(X|Y) = \mathbb{E}_{y \sim \mathsf{P}_Y}\left[\mathrm{H}(X|Y=y)\right] \tag{1.2}$$

$$= -\sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \mathsf{P}_{X,Y}(x,y) \log \mathsf{P}_{X|Y}(x|y). \tag{1.3}$$

Some Useful Properties of the Shannon Entropy

- Maximality under the uniform distribution: If $|\mathcal{X}| < \infty$, then

$$0 \leq \mathrm{H}(X) \leq \log |\mathcal{X}|. \tag{1.4}$$

If X is uniform on its range (getting each value with probability $\frac{1}{|\mathcal{X}|}$), then the upper bound in (1.4) is attained, i.e., $\mathrm{H}(X) = \log |\mathcal{X}|$.

- Subadditivity:

$$\mathrm{H}(X_1, \ldots, X_n) \leq \sum_{j=1}^{n} \mathrm{H}(X_j), \tag{1.5}$$

with equality in (1.5) $\iff$ $X_1, \ldots, X_n$ are statistically independent.

- Chain rule:

$$\mathrm{H}(X_1, \ldots, X_n) = \sum_{j=1}^{n} \mathrm{H}(X_j | X_1, \ldots, X_{j-1}). \tag{1.6}$$

- Concavity: entropy is a concave functional.

Some Useful Properties of the Shannon Entropy (cont.)

- Massey's inequality: Let X be an integer-valued random variable with finite variance $\sigma_X^2 < \infty$. Then,

$$\mathrm{H}(X) \leq \tfrac{1}{2} \log\bigl(2\pi\mathrm{e}\,(\sigma_X^2 + \tfrac{1}{12})\bigr). \tag{1.7}$$

Binary Entropy Function

Definition 1.3

The binary entropy function is the function $\mathsf{H_b}\colon [0,1] \to [0,1]$ given by

$$\mathsf{H_b}(p) = -p \log p - (1-p) \log(1-p), \quad p \in [0,1], \tag{1.8}$$

with the convention that $0 \log 0 = 0$. Equivalently, $\mathsf{H_b}(p)$ is the entropy of a Bernoulli random variable with probabilities p and $1-p$.

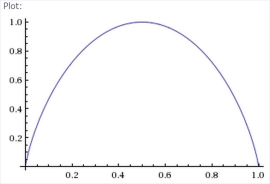

Figure 1: A plot of $\mathsf{H_b}(p)$ for $p \in [0,1]$.

Coin-Weighing Problem

The Coin-Weighing Problem (Erdós & Rényi, 1963)

- We are given n coins, which look quite alike but some are counterfeit.
- Weights of the authentic & counterfeit coins are known, and different.
- A scale enables to weigh any number of coins together.
- Each weighing $\rightarrow$ no. of counterfeit coins within the weighed coins.

The Coin-Weighing Problem (Erdós & Rényi, 1963)

- We are given n coins, which look quite alike but some are counterfeit.
- Weights of the authentic & counterfeit coins are known, and different.
- A scale enables to weigh any number of coins together.
- Each weighing $\rightarrow$ no. of counterfeit coins within the weighed coins.

Question

How many weighings are needed such that, for any constellation of the counterfeit coins among the n coins, one can decide with absolute certainty which of the coins are counterfeit ?

Remark: the sequence of weighings needs to be announced in advance, and a current weighing should not depend on earlier weighings.

The Coin-Weighing Problem

- Label the coins by the elements of the set $[n] \triangleq \{1, \ldots, n\}$.
- Denote the minimal number of weighings by $\ell(n)$.
- Let $\mathcal{S}_1, \ldots, \mathcal{S}_\ell \subseteq [n]$. Suppose that the coins whose labels are the elements of $\mathcal{S}_i$ are weighed together at the i-th weighing for $i \in [\ell]$.

The Coin-Weighing Problem

- Label the coins by the elements of the set $[n] \triangleq \{1, \ldots, n\}$.
- Denote the minimal number of weighings by $\ell(n)$.
- Let $\mathcal{S}_1, \ldots, \mathcal{S}_\ell \subseteq [n]$. Suppose that the coins whose labels are the elements of $\mathcal{S}_i$ are weighed together at the i-th weighing for $i \in [\ell]$.

Definition: Distinguishing Family

- Let Ω be a finite set.
- A collection $\{\mathcal{S}_1, \ldots, \mathcal{S}_\ell\}$ of subsets of Ω is called a distinguishing family of Ω if every subset $\mathcal{T} \subseteq \Omega$ is uniquely determined by the **cardinalities** of the intersections $\mathcal{S}_i \cap \mathcal{T}$ with $i \in [\ell]$.

The Coin-Weighing Problem

- Label the coins by the elements of the set $[n] \triangleq \{1, \ldots, n\}$.
- Denote the minimal number of weighings by $\ell(n)$.
- Let $\mathcal{S}_1, \ldots, \mathcal{S}_\ell \subseteq [n]$. Suppose that the coins whose labels are the elements of $\mathcal{S}_i$ are weighed together at the i-th weighing for $i \in [\ell]$.

Definition: Distinguishing Family

- Let Ω be a finite set.
- A collection $\{\mathcal{S}_1, \ldots, \mathcal{S}_\ell\}$ of subsets of Ω is called a distinguishing family of Ω if every subset $\mathcal{T} \subseteq \Omega$ is uniquely determined by the **cardinalities** of the intersections $\mathcal{S}_i \cap \mathcal{T}$ with $i \in [\ell]$.

$\{\mathcal{S}_1, \ldots, \mathcal{S}_\ell\}$ is a distinguishing family of subsets of a finite set Ω

$$\Updownarrow$$

for every distinct $\mathcal{A}, \mathcal{B} \subseteq \Omega$, $\exists\, i \in [\ell]$ such that $|\mathcal{A} \cap \mathcal{S}_i| \neq |\mathcal{B} \cap \mathcal{S}_i|$.

The Coin-Weighing Problem

Proposition

A necessary and sufficient condition for detecting the counterfeit coins, for any possible constellation among the n coins, is that the collection $\{\mathcal{S}_1, \dots, \mathcal{S}_\ell\}$ is a distinguishing family of $[n]$.

The Coin-Weighing Problem

Proposition

A necessary and sufficient condition for detecting the counterfeit coins, for any possible constellation among the n coins, is that the collection $\{\mathcal{S}_1, \ldots, \mathcal{S}_\ell\}$ is a distinguishing family of $[n]$.

Example

Label 4 coins by the elements $\{1, 2, 3, 4\} := [4]$, and let

$$\mathcal{S}_1 = \{1, 2, 3\}, \quad \mathcal{S}_2 = \{1, 3, 4\}, \quad \mathcal{S}_3 = \{1, 2, 4\}.$$

- Let f_1, f_2 and f_3 be, respectively, the number of counterfeit coins among those in $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$.
- Denote by $'-'$ an authentic coin, and by $'+'$ a counterfeit coin.

The Coin-Weighing Problem

Proposition

A necessary and sufficient condition for detecting the counterfeit coins, for any possible constellation among the n coins, is that the collection $\{\mathcal{S}_1, \ldots, \mathcal{S}_\ell\}$ is a distinguishing family of $[n]$.

Example

Label 4 coins by the elements $\{1, 2, 3, 4\} := [4]$, and let

$$\mathcal{S}_1 = \{1, 2, 3\}, \quad \mathcal{S}_2 = \{1, 3, 4\}, \quad \mathcal{S}_3 = \{1, 2, 4\}.$$

- Let f_1, f_2 and f_3 be, respectively, the number of counterfeit coins among those in $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$.
- Denote by $'-'$ an authentic coin, and by $'+'$ a counterfeit coin.
- The table on next slide shows that $\{\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3\}$ is a distinguishing family of $[4]$. This is the minimal number of weighings, $\ell(4) = 3$.

The Coin-Weighing Problem

f_1	f_2	f_3	1	2	3	4
0	0	0	$-$	$-$	$-$	$-$
1	1	1	$+$	$-$	$-$	$-$
1	0	1	$-$	$+$	$-$	$-$
1	1	0	$-$	$-$	$+$	$-$
0	1	1	$-$	$-$	$-$	$+$
2	1	2	$+$	$+$	$-$	$-$
2	2	1	$+$	$-$	$+$	$-$
1	2	2	$+$	$-$	$-$	$+$
2	1	1	$-$	$+$	$+$	$-$
1	1	2	$-$	$+$	$-$	$+$
1	2	1	$-$	$-$	$+$	$+$
3	2	2	$+$	$+$	$+$	$-$
2	2	3	$+$	$+$	$-$	$+$
2	3	2	$+$	$-$	$+$	$+$
2	2	2	$-$	$+$	$+$	$+$
3	3	3	$+$	$+$	$+$	$+$

The Coin-Weighing Problem

IT Lower Bound (Erdós & Rényi, '63 & Improvement: Pippenger, '77)

$$\ell(n) \geq \frac{2n}{\log_2 n}\left(1 + O\left(\frac{1}{\log n}\right)\right).$$

The Coin-Weighing Problem

IT Lower Bound (Erdós & Rényi, '63 & Improvement: Pippenger, '77)

$$\ell(n) \geq \frac{2n}{\log_2 n}\left(1 + O\left(\frac{1}{\log n}\right)\right).$$

Combinatorial Upper Bound (Lindenström '65, Cantor & Mills '66)

$$\ell(n) \leq \frac{2n}{\log_2 n}\left(1 + O\left(\frac{\log \log n}{\log n}\right)\right).$$

The Coin-Weighing Problem

IT Lower Bound (Erdós & Rényi, '63 & Improvement: Pippenger, '77)

$$\ell(n) \geq \frac{2n}{\log_2 n}\left(1 + O\left(\frac{1}{\log n}\right)\right).$$

Combinatorial Upper Bound (Lindenström '65, Cantor & Mills '66)

$$\ell(n) \leq \frac{2n}{\log_2 n}\left(1 + O\left(\frac{\log \log n}{\log n}\right)\right).$$

An instance of the power of the Shannon entropy in combinatorics !

Proof of IT Lower Bound

- Enumerate all subsets of $[n]$ by indices in $[2^n]$.

Proof of IT Lower Bound

- Enumerate all subsets of $[n]$ by indices in $[2^n]$.
- Let $\mathcal{A} \subseteq [n]$ be selected uniformly at random, and let $X \in [2^n]$ be the index that is assigned to the random subset $\mathcal{A}$.

Proof of IT Lower Bound

- Enumerate all subsets of $[n]$ by indices in $[2^n]$.
- Let $\mathcal{A} \subseteq [n]$ be selected uniformly at random, and let $X \in [2^n]$ be the index that is assigned to the random subset $\mathcal{A}$.
- $\mathcal{A} \leftrightarrow X \quad \Longrightarrow \quad \mathrm{H}(X) = \log_2(2^n) = n$ bits.

Proof of IT Lower Bound

- Enumerate all subsets of $[n]$ by indices in $[2^n]$.
- Let $\mathcal{A} \subseteq [n]$ be selected uniformly at random, and let $X \in [2^n]$ be the index that is assigned to the random subset $\mathcal{A}$.
- $\mathcal{A} \leftrightarrow X \quad \Longrightarrow \quad \mathrm{H}(X) = \log_2(2^n) = n$ bits.

$$\{\mathcal{S}_i\}_{i=1}^{\ell(n)} \text{ is a distinguishing family of } [n]$$

$$\Updownarrow$$

$$X \leftrightarrow \left(|\mathcal{A} \cap \mathcal{S}_1|, \ldots, |\mathcal{A} \cap \mathcal{S}_{\ell(n)}|\right).$$

Proof of IT Lower Bound

- Enumerate all subsets of $[n]$ by indices in $[2^n]$.
- Let $\mathcal{A} \subseteq [n]$ be selected uniformly at random, and let $X \in [2^n]$ be the index that is assigned to the random subset $\mathcal{A}$.
- $\mathcal{A} \leftrightarrow X \quad \Longrightarrow \quad \mathrm{H}(X) = \log_2(2^n) = n$ bits.

$$\{\mathcal{S}_i\}_{i=1}^{\ell(n)} \text{ is a distinguishing family of } [n]$$

$$\Updownarrow$$

$$X \leftrightarrow (|\mathcal{A} \cap \mathcal{S}_1|, \ldots, |\mathcal{A} \cap \mathcal{S}_{\ell(n)}|).$$

$$\begin{aligned} \mathrm{H}(X) &= \mathrm{H}(|\mathcal{A} \cap \mathcal{S}_1|, \ldots, |\mathcal{A} \cap \mathcal{S}_{\ell(n)}|) \\ &\leq \sum_{i=1}^{\ell(n)} \mathrm{H}(|\mathcal{A} \cap \mathcal{S}_i|). \end{aligned}$$

Proof of IT Lower Bound

- The subset $\mathcal{A}$ is selected uniformly at random from $[n]$.

$$\Updownarrow$$

$|\mathcal{A} \cap \mathcal{S}_i| \sim \text{Bin}(|\mathcal{S}_i|, \frac{1}{2})$ is binomially distributed for $i \in [\ell(n)]$.

Proof of IT Lower Bound

- The subset $\mathcal{A}$ is selected uniformly at random from $[n]$.

$$\Updownarrow$$

$|\mathcal{A} \cap \mathcal{S}_i| \sim \mathrm{Bin}(|\mathcal{S}_i|, \frac{1}{2})$ is binomially distributed for $i \in [\ell(n)]$.

- Let $Y_i \sim \mathrm{Bin}(|\mathcal{S}_i|, \frac{1}{2})$ for all $i \in [\ell(n)]$. Then,

$$\mathrm{H}(|\mathcal{A} \cap \mathcal{S}_i|) = \mathrm{H}(Y_i).$$

Proof of IT Lower Bound

- The subset $\mathcal{A}$ is selected uniformly at random from $[n]$.

$$\Updownarrow$$

$|\mathcal{A} \cap \mathcal{S}_i| \sim \mathrm{Bin}(|\mathcal{S}_i|, \frac{1}{2})$ is binomially distributed for $i \in [\ell(n)]$.

- Let $Y_i \sim \mathrm{Bin}(|\mathcal{S}_i|, \frac{1}{2})$ for all $i \in [\ell(n)]$. Then,

$$\mathrm{H}(|\mathcal{A} \cap \mathcal{S}_i|) = \mathrm{H}(Y_i).$$

- By Massey's inequality (1.7) for the entropy of a discrete random variable with finite variance, for all $i \in [\ell(n)]$,

$$\begin{aligned} \mathrm{H}(Y_i) &\leq \tfrac{1}{2} \log_2 \bigl(2\pi \mathrm{e} \, (\sigma^2_{Y_i} + \tfrac{1}{12})\bigr) \\ &= \tfrac{1}{2} \log_2 \bigl(2\pi \mathrm{e} \, (\tfrac{1}{4}|\mathcal{S}_i| + \tfrac{1}{12})\bigr) \quad (\sigma^2_{Y_i} = \tfrac{1}{4}|\mathcal{S}_i|) \\ &\leq \tfrac{1}{2} \log_2 \bigl(2\pi \mathrm{e} \, (\tfrac{n}{4} + \tfrac{1}{12})\bigr) \qquad (|\mathcal{S}_i| \leq n). \end{aligned}$$

Proof of IT Lower Bound

To conclude,

$$\begin{aligned} n &= \mathrm{H}(X) \\ &\leq \sum_{i=1}^{\ell(n)} \mathrm{H}(|\mathcal{A} \cap \mathcal{S}_i|) \\ &\leq \ell(n)\ \mathrm{H}(Y_n) \\ &\leq \tfrac{1}{2}\ell(n) \log_2\Bigl(\tfrac{1}{2}\pi \mathrm{e}\,(n+\tfrac{1}{3})\Bigr), \end{aligned}$$

from which the information-theoretic lower bound on $\ell(n)$ follows.

Proof of IT Lower Bound

To conclude,

$$\begin{aligned} n &= \mathrm{H}(X) \\ &\leq \sum_{i=1}^{\ell(n)} \mathrm{H}(|\mathcal{A} \cap \mathcal{S}_i|) \\ &\leq \ell(n)\ \mathrm{H}(Y_n) \\ &\leq \tfrac{1}{2}\ell(n) \log_2\Bigl(\tfrac{1}{2}\pi \mathrm{e}\,(n+\tfrac{1}{3})\Bigr), \end{aligned}$$

from which the information-theoretic lower bound on $\ell(n)$ follows.

Information-Theoretic Lower Bound (Explicit for Finite n)

For all $n \in \mathbb{N}$,

$$\ell(n) \geq \left\lceil \frac{2n}{\log_2\Bigl(\tfrac{1}{2}\pi \mathrm{e}\,(n+\tfrac{1}{3})\Bigr)} \right\rceil.$$

Combinatorial Upper bound (Lindenström '65)

Let $n = k2^{k-1}$ for $k \in \mathbb{N}$. Then, there exists a distinguishing family of $2^k - 1$ subsets of $[n]$.

Combinatorial Upper bound (Lindenström '65)

Let $n = k2^{k-1}$ for $k \in \mathbb{N}$. Then, there exists a distinguishing family of $2^k - 1$ subsets of $[n]$.

For fixed $n \in \mathbb{N}$, let $k_0 \in \mathbb{N}$ be the smallest integer satisfying $n \leq k_0 2^{k_0 - 1}$. Then, $\ell(n) \leq 2^{k_0} - 1$. Calculating the smallest such $k_0 = k_0(n) \in \mathbb{N}$ gives

$$k_0 = \left\lceil \frac{W_0(2n \ln 2)}{\ln 2} \right\rceil,$$

where $W_0 \colon [-\frac{1}{\mathrm{e}}, \infty) \to [-1, \infty)$ is the principal branch of the Lambert W function (and $x = W_0(u)$ is the solution of the equation $x\mathrm{e}^x = u$ for all $u > 0$, which is unique and positive).

Combinatorial Upper bound (Lindenström '65)

Let $n = k2^{k-1}$ for $k \in \mathbb{N}$. Then, there exists a distinguishing family of $2^k - 1$ subsets of $[n]$.

For fixed $n \in \mathbb{N}$, let $k_0 \in \mathbb{N}$ be the smallest integer satisfying $n \leq k_0 2^{k_0-1}$. Then, $\ell(n) \leq 2^{k_0} - 1$. Calculating the smallest such $k_0 = k_0(n) \in \mathbb{N}$ gives

$$k_0 = \left\lceil \frac{W_0(2n \ln 2)}{\ln 2} \right\rceil,$$

where $W_0 \colon [-\frac{1}{\mathrm{e}}, \infty) \to [-1, \infty)$ is the principal branch of the Lambert W function (and $x = W_0(u)$ is the solution of the equation $x\mathrm{e}^x = u$ for all $u > 0$, which is unique and positive).

Combinatorial Upper Bound (Explicit for Finite n)

For all $n \in \mathbb{N}$,

$$\ell(n) \leq \exp\left(\ln 2 \left\lceil \frac{W_0(2n \ln 2)}{\ln 2} \right\rceil \right) - 1.$$

Bounds on $W_0(x)$

For all $x \geq \mathrm{e}$,

$$\frac{x}{\ln x} \cdot \exp\left(\frac{1}{2}\frac{\ln \ln x}{\ln x}\right) \leq \exp\bigl(W_0(x)\bigr) \leq \frac{x}{\ln x} \cdot \exp\left(\frac{\mathrm{e}}{\mathrm{e}-1}\,\frac{\ln \ln x}{\ln x}\right),$$

which yields the asymptotic upper bound

$$\ell(n) \leq \frac{2n}{\log_2 n}\left(1 + O\Bigl(\frac{\log \log n}{\log n}\Bigr)\right).$$

Shearer's Lemma

Shearer's Lemma at a High-Level

- At a high level, Shearer's Lemma can be regarded as a combinatorial counterpart to the Loomis-Whitney inequality in geometry.

Shearer's Lemma at a High-Level

- At a high level, Shearer's Lemma can be regarded as a combinatorial counterpart to the Loomis-Whitney inequality in geometry.
- In a specialized form of Shearer's Lemma, we consider sets within a finite universe, discussing cardinalities rather than volumes and areas.

Shearer's Lemma at a High-Level

- At a high level, Shearer's Lemma can be regarded as a combinatorial counterpart to the Loomis-Whitney inequality in geometry.
- In a specialized form of Shearer's Lemma, we consider sets within a finite universe, discussing cardinalities rather than volumes and areas.
- Origin of Shearer's lemma:
  - Shearer's Lemma was initially developed as an information-theoretic tool to upper bound the size of any family of triangle-intersecting graphs of a given order (1986).
  - It marked the first significant progress toward resolving a conjecture proposed by Simonovits and Sós (1976).
  - That conjecture was proven, in a rather involved manner, using a combinatorial approach by Ellis, Filmus, and Friedgut (2012).

Shearer's Lemma at a High-Level

- At a high level, Shearer's Lemma can be regarded as a combinatorial counterpart to the Loomis-Whitney inequality in geometry.
- In a specialized form of Shearer's Lemma, we consider sets within a finite universe, discussing cardinalities rather than volumes and areas.
- Origin of Shearer's lemma:
  - Shearer's Lemma was initially developed as an information-theoretic tool to upper bound the size of any family of triangle-intersecting graphs of a given order (1986).
  - It marked the first significant progress toward resolving a conjecture proposed by Simonovits and Sós (1976).
  - That conjecture was proven, in a rather involved manner, using a combinatorial approach by Ellis, Filmus, and Friedgut (2012).
- Shearer inequalities have found extensive applications across various fields, including finite geometry, graph theory, Boolean functions analysis, and large-deviations analysis.

Shearer's Lemma

Shearer's lemma extends the subadditivity property of Shannon entropy.

Proposition 3.1 (Shearer's Lemma)

Let

- $n, m, k \in \mathbb{N}$,
- $X_1, \ldots, X_n$ be **discrete** random variables,
- $[n] \triangleq \{1, \ldots, n\}$,
- $\mathcal{S}_1, \ldots, \mathcal{S}_m \subseteq [n]$ be subsets such that each element $i \in [n]$ belongs to **at least** $k \geq 1$ of these subsets.
- $X^n \triangleq (X_1, \ldots, X_n)$, and $X_{\mathcal{S}_j} \triangleq (X_i)_{i \in \mathcal{S}_j}$ for all $j \in [m]$.

Then,

$$k \, \mathrm{H}(X^n) \leq \sum_{j=1}^{m} \mathrm{H}(X_{\mathcal{S}_j}). \tag{3.1}$$

Proof of Shearer's Lemma (Proposition 3.1)

- By assumption, $d(i) \geq k$ for all $i \in [n]$, where
$$d(i) \triangleq \Big|\Big\{j \in [m] : \, i \in \mathcal{S}_j\Big\}\Big|. \tag{3.2}$$
- Let $\mathcal{S} = \{i_1, \ldots, i_\ell\}$, $1 \leq i_1 < \ldots < i_\ell \leq n \implies |\mathcal{S}| = \ell$, $\mathcal{S} \subseteq [n]$.
- Let $X_{\mathcal{S}} \triangleq (X_{i_1}, \ldots, X_{i_\ell})$.
- By the chain rule and the fact that conditioning reduces entropy,
$$\begin{aligned} \mathrm{H}(X_{\mathcal{S}}) &= \mathrm{H}(X_{i_1}) + \mathrm{H}(X_{i_2}|X_{i_1}) + \ldots + \mathrm{H}(X_{i_\ell}|X_{i_1}, \ldots, X_{i_{\ell-1}}) \\ &\geq \sum_{i \in \mathcal{S}} \mathrm{H}(X_i|X_1, \ldots, X_{i-1}) \\ &= \sum_{i=1}^{n} \Big\{\mathbb{1}\{i \in \mathcal{S}\} \, \mathrm{H}(X_i|X_1, \ldots, X_{i-1})\Big\}. \end{aligned} \tag{3.3}$$

Proof of Shearer's Lemma (Cont.)

$$\begin{aligned}
\sum_{j=1}^{m} \mathrm{H}(X_{\mathcal{S}_j}) &\geq \sum_{j=1}^{m} \sum_{i=1}^{n} \Big\{ \mathbb{1}\{i \in \mathcal{S}_j\} \, \mathrm{H}(X_i | X_1, \ldots, X_{i-1}) \Big\} \\
&= \sum_{i=1}^{n} \left\{ \sum_{j=1}^{m} \mathbb{1}\{i \in \mathcal{S}_j\} \, \mathrm{H}(X_i | X_1, \ldots, X_{i-1}) \right\} \\
&= \sum_{i=1}^{n} \Big\{ d(i) \, \mathrm{H}(X_i | X_1, \ldots, X_{i-1}) \Big\} \\
&\geq k \sum_{i=1}^{n} \mathrm{H}(X_i | X_1, \ldots, X_{i-1}) \qquad (3.4) \\
&= k \, \mathrm{H}(X^n),
\end{aligned}$$

where inequality (3.4) holds due to the nonnegativity of the conditional entropies of discrete random variables, and under the assumption that $d(i) \geq k$ for all $i \in [n]$.

Special case: Subadditivity of the Shannon entropy

Let $n = m$ with $n \in \mathbb{N}$, and $\mathcal{S}_i = \{i\}$ (singletons) for all $i \in [n]$
$\Rightarrow$ every element $i \in [n]$ belongs to a single set among $\mathcal{S}_1, \ldots, \mathcal{S}_n$ (i.e., $k = 1$). By Shearer's Lemma, it follows that

$$\mathrm{H}(X^n) \leq \sum_{j=1}^{n} \mathrm{H}(X_j),$$

which is the subadditivity property of the Shannon entropy for discrete random variables.

Special case: Subadditivity of the Shannon entropy

Let $n = m$ with $n \in \mathbb{N}$, and $\mathcal{S}_i = \{i\}$ (singletons) for all $i \in [n]$
$\Rightarrow$ every element $i \in [n]$ belongs to a single set among $\mathcal{S}_1, \ldots, \mathcal{S}_n$ (i.e., $k = 1$). By Shearer's Lemma, it follows that

$$\mathrm{H}(X^n) \leq \sum_{j=1}^{n} \mathrm{H}(X_j),$$

which is the subadditivity property of the Shannon entropy for discrete random variables.

If every element $i \in [n]$ belongs to **exactly** k of the subsets $\mathcal{S}_j$ $(j \in [m])$, then Shearer's lemma also applies to continuous random variables $X_1, \ldots, X_n$, with entropy replaced by the differential entropy. Hence, Shearer's lemma yields the subadditivity property of the Shannon entropy for discrete and continuous random variables.

Special case: Han's Inequality

For all $\ell \in [n]$, let $\mathcal{S}_\ell = [n] \setminus \{\ell\}$. By Shearer's Lemma (Proposition 3.1) applied to these n subsets of $[n]$, since every element $i \in [n]$ is contained in exactly $k = n - 1$ of these subsets,

$$(n-1)\,\mathrm{H}(X^n) \leq \sum_{\ell=1}^{n} \mathrm{H}(X_1, \ldots, X_{\ell-1}, X_{\ell+1}, \ldots, X_n) \leq n\,\mathrm{H}(X^n). \quad (3.5)$$

An equivalent form of (3.5) is given by

$$0 \leq \sum_{\ell=1}^{n} \Big\{ \mathrm{H}(X^n) - \mathrm{H}(X_1, \ldots, X_{\ell-1}, X_{\ell+1}, \ldots, X_n) \Big\} \leq \mathrm{H}(X^n). \quad (3.6)$$

The equivalent forms in (3.5) and (3.6) are known as Han's inequality.

Proposition 3.2 (Shearer's Lemma: Probabilistic Version)

Let X^n be a discrete n-dimensional random vector, and let $\mathcal{S} \subseteq [n]$ be a random subset of $[n]$, independent of X^n, with an arbitrary probability mass function $\mathsf{P}_{\mathcal{S}}$. If there exists $\theta > 0$ such that

$$\Pr[i \in \mathcal{S}] \geq \theta, \quad \forall\, i \in [n], \tag{3.7}$$

then,

$$\mathbb{E}_{\mathcal{S}}\big[\mathrm{H}(X_{\mathcal{S}})\big] \geq \theta\, \mathrm{H}(X^n). \tag{3.8}$$

Proof of Proposition 3.2

By inequality (3.3), for any set $\mathcal{S} \subseteq [n]$,

$$\mathrm{H}(X_{\mathcal{S}}) \geq \sum_{i=1}^{n} \Big\{ \mathbb{1}\{i \in \mathcal{S}\} \, \mathrm{H}(X_i | X_1, \ldots, X_{i-1}) \Big\}.$$

Proof of Proposition 3.2 (cont.)

$$
\begin{aligned}
\Longrightarrow \; \mathbb{E}_{\mathcal{S}}\big[\mathrm{H}(X_{\mathcal{S}})\big] &= \sum_{\mathcal{S} \subseteq [n]} \mathsf{P}_{\mathcal{S}}(\mathcal{S}) \, \mathrm{H}(X_{\mathcal{S}}) \\
&\geq \sum_{\mathcal{S} \subseteq [n]} \left\{ \mathsf{P}_{\mathcal{S}}(\mathcal{S}) \sum_{i=1}^{n} \Big\{ \mathbb{1}\{i \in \mathcal{S}\} \, \mathrm{H}(X_i | X_1, \ldots, X_{i-1}) \Big\} \right\} \\
&= \sum_{i=1}^{n} \left\{ \sum_{\mathcal{S} \subseteq [n]} \Big\{ \mathsf{P}_{\mathcal{S}}(\mathcal{S}) \, \mathbb{1}\{i \in \mathcal{S}\} \Big\} \, \mathrm{H}(X_i | X_1, \ldots, X_{i-1}) \right\} \\
&= \sum_{i=1}^{n} \Pr[i \in \mathcal{S}] \, \mathrm{H}(X_i | X_1, \ldots, X_{i-1}) \\
&\geq \theta \sum_{i=1}^{n} \mathrm{H}(X_i | X_1, \ldots, X_{i-1}) \qquad (3.9) \\
&= \theta \, \mathrm{H}(X^n).
\end{aligned}
$$

Proposition 3.3 (Combinatorial Shearer's Lemma)

- Let $\mathscr{F}$ be a finite multiset of subsets of $[n]$ (possibly with repeats), where each element $i \in [n]$ is included in at least $k \geq 1$ sets of $\mathscr{F}$.
- Let $\mathscr{M}$ be a set of subsets of $[n]$.
- For every set $\mathcal{S} \in \mathscr{F}$, let the trace of $\mathscr{M}$ on $\mathcal{S}$, denoted $\text{trace}_{\mathcal{S}}(\mathscr{M})$, be the set of all possible intersections of elements of $\mathscr{M}$ with $\mathcal{S}$, i.e.,

$$\text{trace}_{\mathcal{S}}(\mathscr{M}) \triangleq \{\mathcal{A} \cap \mathcal{S} : \mathcal{A} \in \mathscr{M}\}, \quad \forall \mathcal{S} \in \mathscr{F}. \tag{3.10}$$

Then,

$$|\mathscr{M}| \leq \prod_{\mathcal{S} \in \mathscr{F}} \left|\text{trace}_{\mathcal{S}}(\mathscr{M})\right|^{\frac{1}{k}}. \tag{3.11}$$

Proof of Proposition 3.3

- Let $\mathcal{X} \subseteq [n]$ be a set that is selected uniformly at random from $\mathscr{M}$.
- Represent $\mathcal{X}$ by the random vector $X^n = (X_1, \ldots, X_n)$, where X_i (for all $i \in [n]$) denotes the indicator function of the event $\{i \in \mathcal{X}\}$.
- For every $\mathcal{S} \in \mathscr{F}$, let $X_{\mathcal{S}} = (X_i)_{i \in \mathcal{S}}$. Then,

$$\mathrm{H}(X_{\mathcal{S}}) \le \log \left| \mathrm{trace}_{\mathcal{S}}(\mathscr{M}) \right|. \tag{3.12}$$

- Applying Shearer's lemma (Proposition 3.1) gives

$$k \, \mathrm{H}(X^n) \le \sum_{\mathcal{S} \in \mathscr{F}} \log \left| \mathrm{trace}_{\mathcal{S}}(\mathscr{M}) \right|. \tag{3.13}$$

- $\mathrm{H}(X^n) = \log |\mathscr{M}|$ since X^n is in one-to-one correspondence with $\mathcal{X}$, which is a set selected uniformly at random from $\mathscr{M}$. Hence,

$$\log |\mathscr{M}| \le \frac{1}{k} \sum_{\mathcal{S} \in \mathscr{F}} \log \left| \mathrm{trace}_{\mathcal{S}}(\mathscr{M}) \right|, \tag{3.14}$$

  and exponentiation of both sides of (3.14) gives (3.11).

Shearer's Lemma in Finite Geometry

A Geometric Application of Shearer's Lemma

Example 4.1

Let $\mathcal{P} \subseteq \mathbb{R}^3$ be a set of points that has at most r distinct projections on each of the XY, XZ and YZ planes. How large can this set be ?

Example 4.1

Let $\mathcal{P} \subseteq \mathbb{R}^3$ be a set of points that has at most r distinct projections on each of the XY, XZ and YZ planes. How large can this set be ?

As we shall see in the next slide,

$$|\mathcal{P}| \leq r^{\frac{3}{2}}.$$

Furthermore, that bound on the cardinality of the set $\mathcal{P}$ is achieved by a grid of $\sqrt{r} \times \sqrt{r} \times \sqrt{r}$ points, provided that r is a square of an integer.

Example 4.1 (cont.)

- By Shearer's lemma,

$$2\,\mathrm{H}(X,Y,Z) \le \mathrm{H}(X,Y) + \mathrm{H}(X,Z) + \mathrm{H}(Y,Z). \tag{4.1}$$

- Let $(X,Y,Z) \in \mathcal{P}$ be selected uniformly at random in $\mathcal{P}$. Then,

$$\mathrm{H}(X,Y,Z) = \log|\mathcal{P}|. \tag{4.2}$$

- By assumption, the set $\mathcal{P}$ has at most r distinct projections on each of the XY, XZ, and YZ planes. Hence,

$$\mathrm{H}(X,Y) \le \log r, \quad \mathrm{H}(X,Z) \le \log r, \quad \mathrm{H}(Y,Z) \le \log r. \tag{4.3}$$

- Combining (4.1)–(4.3) gives

$$2\log|\mathcal{P}| \le 3\log r, \tag{4.4}$$

and then exponentiating both sides of (4.4) gives $|\mathcal{P}| \le r^{\frac{3}{2}}$.

Generalization of Example 4.1

- Let $\mathcal{P} \subseteq \mathbb{R}^n$ be a finite set with $|\mathcal{P}| = M$.
- Let $k \in [n-1]$.
- Let $\mathcal{S}_1, \ldots, \mathcal{S}_\ell$ be all the k-element subsets of $[n]$, where $\ell = \binom{n}{k}$. Then, every element $i \in [n]$ belongs to exactly $\binom{n-1}{k-1}$ of these subsets.
- By applying Shearer's lemma, it follows that

$$\binom{n-1}{k-1} \mathrm{H}(X^n) \leq \sum_{j=1}^{\ell} \mathrm{H}(X_{\mathcal{S}_j}). \tag{4.5}$$

- Let $X^n \in \mathcal{P}$ be a point that is selected uniformly at random in $\mathcal{P}$. Then,

$$\mathrm{H}(X^n) = \log M. \tag{4.6}$$

- Let M_j be the number of distinct projections of $\mathcal{P}$ on the k-dimensional subspace of $\mathbb{R}^n$ whose coordinates are the elements of the set $\mathcal{S}_j$. Then,

$$\mathrm{H}(X_{\mathcal{S}_j}) \leq \log M_j, \quad j \in [\ell]. \tag{4.7}$$

Generalization of Example 4.1 (cont.)

- Combining (4.5)–(4.7) gives

$$\binom{n-1}{k-1} \log M \leq \sum_{j=1}^{\ell} \log M_j. \tag{4.8}$$

- Let

$$R \triangleq \frac{\log M}{n}, \qquad R_j \triangleq \frac{\log M_j}{k}, \quad \forall\, j \in [\ell]. \tag{4.9}$$

- Combining (4.8), (4.9), and the identity $\frac{n}{k}\binom{n-1}{k-1} = \binom{n}{k} = \ell$, gives

$$R \leq \frac{1}{\ell} \sum_{j=1}^{\ell} R_j, \tag{4.10}$$

  and if $\sqrt[k]{M_j} \in \mathbb{N}$, for all $j \in [\ell]$, then (4.10) holds with equality for $\mathcal{P}$ that is an n-dimensional grid of points.

  Setting $n = 3$, $k = 2$, and $M_j = r$ for $j \in \{1, 2, 3\}$, gives Example 4.1.

Extremal Combinatorics: Intersecting Families of Graphs

Definition 5.1 (Triangle-Intersecting Families of Graphs)

Let $\mathcal{G}$ be a family of graphs on the vertex set $[n]$, with the property that for every $\mathsf{G}_1, \mathsf{G}_2 \in \mathcal{G}$, the intersection $\mathsf{G}_1 \cap \mathsf{G}_2$ contains a triangle (i.e, there are three vertices $i, j, k \in [n]$ such that each of $\{i, j\}$, $\{i, k\}$, $\{j, k\}$ is in the edge sets of both G_1 and G_2). The family $\mathcal{G}$ is referred to as a triangle-interesting family of graphs on n vertices.

Definition 5.1 (Triangle-Intersecting Families of Graphs)

Let $\mathcal{G}$ be a family of graphs on the vertex set $[n]$, with the property that for every $\mathsf{G}_1, \mathsf{G}_2 \in \mathcal{G}$, the intersection $\mathsf{G}_1 \cap \mathsf{G}_2$ contains a triangle (i.e, there are three vertices $i, j, k \in [n]$ such that each of $\{i, j\}$, $\{i, k\}$, $\{j, k\}$ is in the edge sets of both G_1 and G_2). The family $\mathcal{G}$ is referred to as a triangle-interesting family of graphs on n vertices.

Question (Simonovits and Sós, 1976)

How large can $\mathcal{G}$ (a family of triangle-intersecting graphs) be ?

Lower Bound on Largest Size

$|\mathcal{G}|$ can be as large as $2^{\binom{n}{2}-3}$.

Proof.

Consider the family $\mathcal{G}$ of all graphs on n vertices that include a particular triangle. □

Lower Bound on Largest Size

$|\mathcal{G}|$ can be as large as $2^{\binom{n}{2}-3}$.

Proof.

Consider the family $\mathcal{G}$ of all graphs on n vertices that include a particular triangle. □

Upper Bound on Largest Size

$|\mathcal{G}|$ cannot exceed $2^{\binom{n}{2}-1}$.

Proof.

A family of distinct subsets of a set of size m, where any two of these subsets have a non-empty intersection, can have a cardinality of at most 2^{m-1} ($\mathcal{A}$ and $\overline{\mathcal{A}}$ cannot be members of this family). The edge sets of the graphs in $\mathcal{G}$ satisfy this property, with $m = \binom{n}{2}$. □

Proposition 5.1 (Ellis, Filmus and Friedgut (2012))

The size of a family $\mathcal{G}$ *of triangle-intersecting graphs on* n *vertices satisfies* $|\mathcal{G}| \leq 2^{\binom{n}{2}-3}$.

This result was proved by using discrete Fourier analysis to obtain the sharp bound $|\mathcal{G}| \leq 2^{\binom{n}{2}-3}$, as conjectured by Simonovits and Sós.

Proposition 5.1 (Ellis, Filmus and Friedgut (2012))

The size of a family $\mathcal{G}$ of triangle-intersecting graphs on n vertices satisfies $|\mathcal{G}| \leq 2^{\binom{n}{2}-3}$.

This result was proved by using discrete Fourier analysis to obtain the sharp bound $|\mathcal{G}| \leq 2^{\binom{n}{2}-3}$, as conjectured by Simonovits and Sós.

- The first significant progress towards proving the Simonovits–Sós conjecture came from an information-theoretic approach by Chung, Graham, Frankl, and Shearer in 1986.
- Using the combinatorial Shearer lemma (Proposition 3.3), they derived a simple and elegant upper bound on the size of $\mathcal{G}$.
- Their bound was given as $2^{\binom{n}{2}-2}$, falling short of the Simonovits–Sós conjecture by a factor of 2.

Triangle-Intersecting Families of Graphs (cont.)

Proposition 5.2 (Chung, Graham, Frankl, and Shearer, 1986)

Let $\mathcal{G}$ be a family of K_3-intersecting graphs on a common vertex set $[n]$. Then, $|\mathcal{G}| \leq 2^{\binom{n}{2}-2}$.

Proof of Proposition 5.2

- Identify $\mathsf{G} \in \mathcal{G}$ with its edge set $\mathsf{E}(\mathsf{G})$, and let $\mathscr{M} = \big\{\mathsf{E}(\mathsf{G}) : \mathsf{G} \in \mathcal{G}\big\}$.
- Let $\mathcal{U} = \mathsf{E}(\mathsf{K}_n)$. For every $\mathsf{G} \in \mathcal{G}$, we have $\mathsf{E}(\mathsf{G}) \subseteq \mathcal{U}$, and $|\mathcal{U}| = \binom{n}{2}$.
- For every unordered equipartition $\mathcal{A} \cup \mathcal{B} = [n]$, which satisfies $\big||\mathcal{A}| - |\mathcal{B}|\big| \leq 1$, let $\mathcal{U}(\mathcal{A}, \mathcal{B})$ be the subset of $\mathcal{U}$ consisting of all those edges that lie entirely inside $\mathcal{A}$ or entirely inside $\mathcal{B}$.
- We apply Proposition 3.3 with $\mathscr{F} = \{\mathcal{U}(\mathcal{A}, \mathcal{B})\}$ with $\mathcal{A}, \mathcal{B}$ as above.
- Let $m = |\mathcal{U}(\mathcal{A}, \mathcal{B})|$, which is independent of the equipartition since

$$m = \begin{cases} 2\binom{n/2}{2} & \text{if } n \text{ is even,} \\ \binom{\lfloor n/2 \rfloor}{2} + \binom{\lceil n/2 \rceil}{2} & \text{if } n \text{ is odd.} \end{cases} \implies m \leq \frac{1}{2}\binom{n}{2}. \tag{5.1}$$

- By a simple double-counting argument, if k is the number of elements of $\mathscr{F}$ in which each element of $\mathcal{U}$ occurs, then

$$m\,|\mathscr{F}| = \binom{n}{2} k. \tag{5.2}$$

Proof of Proposition 5.2 (cont.)

- Let $\mathcal{S} \in \mathscr{F}$.
- Observe that $\text{trace}_{\mathcal{S}}(\mathscr{M})$ forms an intersecting family of subsets of $\mathcal{S}$; indeed, for any $\mathsf{G}, \mathsf{G}' \in \mathcal{G}$, $\mathsf{G} \cap \mathsf{G}'$ has a triangle $T = \mathsf{K}_3$, and since the complement of $\mathcal{S}$ (in $\mathcal{U}$) is triangle-free (viewed as a graph on $[n]$), at least one of the edges of T belongs to $\mathcal{S}$. So, since $|\mathcal{S}| = m$, we have

$$|\text{trace}_{\mathcal{S}}(\mathscr{M})| \leq 2^{m-1}.$$

- By Proposition 3.3 (and 1-to-1 correspondence between $\mathcal{G}$ and $\mathscr{M}$),

$$\begin{aligned} |\mathcal{G}| &= |\mathscr{M}| \\ &\leq \left(2^{m-1}\right)^{\frac{|\mathscr{F}|}{k}} && (5.3) \\ &= 2^{\binom{n}{2}\left(1-\frac{1}{m}\right)} && (5.4) \\ &\leq 2^{\binom{n}{2}-2}, && (5.5) \end{aligned}$$

where (5.4) relies on (5.2), and (5.5) holds due to (5.1).

Intersecting Families of Graphs (cont.)

Definition 5.2 (H-intersecting Families of Graphs)

Let $\mathcal{G}$ be a family of graphs on a common vertex set. Then, it is said that $\mathcal{G}$ is H-intersecting if for every two graphs $\mathsf{G}_1, \mathsf{G}_2 \in \mathcal{G}$, the graph $\mathsf{G}_1 \cap \mathsf{G}_2$ contains H as a subgraph.

Intersecting Families of Graphs (cont.)

Definition 5.2 (H-intersecting Families of Graphs)

Let $\mathcal{G}$ be a family of graphs on a common vertex set. Then, it is said that $\mathcal{G}$ is H-intersecting if for every two graphs $\mathsf{G}_1, \mathsf{G}_2 \in \mathcal{G}$, the graph $\mathsf{G}_1 \cap \mathsf{G}_2$ contains H as a subgraph.

Example 5.3

Let $\mathsf{H} = \mathsf{K}_t$ with $t \geq 2$. Then,

- $t = 2$ means that $\mathcal{G}$ is intersecting,
- $t = 3$ means that $\mathcal{G}$ is triangle-intersecting.

Intersecting Families of Graphs (cont.)

Definition 5.2 (H-intersecting Families of Graphs)

Let $\mathcal{G}$ be a family of graphs on a common vertex set. Then, it is said that $\mathcal{G}$ is H-intersecting if for every two graphs $\mathsf{G}_1, \mathsf{G}_2 \in \mathcal{G}$, the graph $\mathsf{G}_1 \cap \mathsf{G}_2$ contains H as a subgraph.

Example 5.3

Let $\mathsf{H} = \mathsf{K}_t$ with $t \geq 2$. Then,

- $t = 2$ means that $\mathcal{G}$ is intersecting,
- $t = 3$ means that $\mathcal{G}$ is triangle-intersecting.

Problem in Extremal Combinatorics

Given H and n, determine the maximum size of an H-intersecting family of graphs on n labeled vertices.

Intersecting Families of Graphs (cont.)

Generalized Conjecture (Ellis, Filmus, and Friedgut, 2012)

Every K_t-intersecting family of graphs on a common vertex set $[n]$ has size at most $2^{\binom{n}{2}-\binom{t}{2}}$, with equality for the family of all graphs containing a fixed clique on t vertices.

Intersecting Families of Graphs (cont.)

Generalized Conjecture (Ellis, Filmus, and Friedgut, 2012)

Every K_t-intersecting family of graphs on a common vertex set $[n]$ has size at most $2^{\binom{n}{2}-\binom{t}{2}}$, with equality for the family of all graphs containing a fixed clique on t vertices.

- For $t = 2$, it is trivial (since K_2 is an edge).
- For $t = 3$, it was proved by Ellis, Filmus & Friedgut ('12).
- For $t = 4$, it was recently proved by Berger and Zhao (2023).
- For $t \geq 5$, this problem is left open.

Shearer's Lemma and Cliques in Graphs

All graphs here are assumed to be finite, simple, and undirected.

Application of Proposition 3.2 to Graph Theory

Proposition 6.1

Let G be a simple graph on n vertices, and let m_ℓ be the number of cliques of order $\ell \in \mathbb{N}$ in G. Then, for all $s, t \in \mathbb{N}$ with $2 \leq s < t \leq n$,

$$(t!\, m_t)^s \leq (s!\, m_s)^t. \qquad (6.1)$$

Proof of Proposition 6.1

- Label the vertices in G by the elements of $[n]$, and let $2 \leq s < t \leq n$.
- Select a clique of order t in G uniformly at random, and then select the order of the vertices within that copy uniformly at random. This results in a random vector $(X_1, \ldots, X_t)$, reflecting the chosen order of the vertices.
- Let m_t be the number of cliques of order t in G. Then,

$$\mathrm{H}(X_1, \ldots, X_t) = \log(t! \, m_t), \tag{6.2}$$

  since the order of the vertices of a clique of order t in G can be selected in $t!$ equiprobable ways according to their order of selection.
- Let $\mathcal{S}$ be a uniformly selected subset of size s from $[t]$. Then,

$$\Pr[i \in \mathcal{S}] = \frac{s}{t}, \quad \forall \, i \in [t]. \tag{6.3}$$

- By Proposition 3.2, it follows from (6.2) and (6.3) that

$$\mathbb{E}_{\mathcal{S}}\big[\mathrm{H}(X_{\mathcal{S}})\big] \geq \frac{s \, \log(t! \, m_t)}{t}. \tag{6.4}$$

Proof of Proposition 6.1 (cont.)

- $\Longrightarrow \exists\, \mathcal{S}' \subset [t]$ with $|\mathcal{S}'| = s$, satisfying

$$\mathrm{H}(X_{\mathcal{S}'}) \geq \frac{s \, \log(t!\, m_t)}{t}. \tag{6.5}$$

- The random subvector $X_{\mathcal{S}'}$ is supported on a clique of order s in G (an induced subgraph of a clique is also a clique), so

$$\mathrm{H}(X_{\mathcal{S}'}) \leq \log(s!\, m_s), \tag{6.6}$$

since there are m_s cliques of order s in G, and the order of the vertices in a clique of order s can be selected in $s!$ ways.

- Combining (6.5) and (6.6) yields

$$\log(s!\, m_s) \geq \frac{s \, \log(t!\, m_t)}{t}, \tag{6.7}$$

which by exponentiating both sides of (6.7) gives (6.1).

Example 6.1

Let G be a simple graph on n vertices with $e(\mathsf{G})$ edges and $t(\mathsf{G})$ triangles. Substituting $s = 2$ and $t = 3$ into (6.1), with $m_2 = e(\mathsf{G})$ and $m_3 = t(\mathsf{G})$, gives

$$(6\,t(\mathsf{G}))^2 \le (2\,e(\mathsf{G}))^3, \tag{6.8}$$

which can be also derived by using spectral graph theory. Let $\mathbf{A}$ be the adjacency matrix of G, with spectrum $\{\lambda_j\}_{j=1}^n$, and $\underline{\lambda} = (\lambda_1, \ldots, \lambda_n)$. Then,

$$\sum_{j=1}^n \lambda_j^2 = \mathrm{Tr}(\mathbf{A}^2) = 2\,e(\mathsf{G}), \qquad \sum_{j=1}^n \lambda_j^3 = \mathrm{Tr}(\mathbf{A}^3) = 6\,t(\mathsf{G}), \tag{6.9}$$

$$(6\,t(\mathsf{G}))^2 = \left(\sum_{j=1}^n \lambda_j^3\right)^2 \le \|\underline{\lambda}\|_3^6 \le \|\underline{\lambda}\|_2^6 = \left(\sum_{j=1}^n \lambda_j^2\right)^3 = (2\,e(\mathsf{G}))^3, \tag{6.10}$$

where the second inequality in (6.10) holds since the norm $\|\cdot\|_p$ is monotonically decreasing in $p \ge 1$.

A Generalization of Shearer's Lemma

A Generalized Version of Shearer's Lemma

We next provide a generalized version of Shearer's Lemma. To that end, let Ω be a finite and non-empty set, and let $f\colon 2^{\Omega} \to \mathbb{R}$ be a real-valued set function (i.e., f is defined for all subsets of Ω).

Definition 7.1 (Sub/Supermodular function)

The set function $f\colon 2^{\Omega} \to \mathbb{R}$ is submodular if

$$f(\mathcal{T}) + f(\mathcal{S}) \geq f(\mathcal{T} \cup \mathcal{S}) + f(\mathcal{T} \cap \mathcal{S}), \qquad \forall\, \mathcal{S}, \mathcal{T} \subseteq \Omega \tag{7.1}$$

Likewise, f is supermodular if $-f$ is submodular.

Equivalent Condition for Submodularity

An identical characterization of submodularity is the diminishing return property, which is stated as follows.

Proposition 7.1

A set function $f\colon 2^{\Omega} \to \mathbb{R}$ is submodular if and only if whenever

$$\mathcal{S} \subset \mathcal{T} \subset \Omega, \;\; \omega \in \mathcal{T}^{\mathsf{c}} \implies f(\mathcal{S} \cup \{\omega\}) - f(\mathcal{S}) \geq f(\mathcal{T} \cup \{\omega\}) - f(\mathcal{T}). \quad (7.2)$$

Equivalent Condition for Submodularity

An identical characterization of submodularity is the diminishing return property, which is stated as follows.

Proposition 7.1

A set function $f \colon 2^{\Omega} \to \mathbb{R}$ is submodular if and only if whenever

$$\mathcal{S} \subset \mathcal{T} \subset \Omega, \;\; \omega \in \mathcal{T}^{\mathrm{c}} \implies f(\mathcal{S} \cup \{\omega\}) - f(\mathcal{S}) \geq f(\mathcal{T} \cup \{\omega\}) - f(\mathcal{T}). \tag{7.2}$$

The equivalent condition for the submodularity of f in (7.2) means that the larger is the set, the smaller is the increase in f when a new element is added.

Definition 7.2 (Monotonic set function)

The set function $f\colon 2^{\Omega} \to \mathbb{R}$ is *monotonically increasing* if

$$\mathcal{S} \subseteq \mathcal{T} \subseteq \Omega \implies f(\mathcal{S}) \leq f(\mathcal{T}). \qquad (7.3)$$

Likewise, f is *monotonically decreasing* if $-f$ is monotonically increasing.

Definition 7.2 (Monotonic set function)

The set function $f\colon 2^{\Omega} \to \mathbb{R}$ is *monotonically increasing* if

$$\mathcal{S} \subseteq \mathcal{T} \subseteq \Omega \implies f(\mathcal{S}) \leq f(\mathcal{T}). \tag{7.3}$$

Likewise, f is *monotonically decreasing* if $-f$ is monotonically increasing.

Definition 7.3 (Polymatroid, ground set and rank function)

Let $f\colon 2^{\Omega} \to \mathbb{R}$ be submodular and monotonically increasing set function with $f(\varnothing) = 0$. The pair (Ω, f) is called a polymatroid, Ω is called a ground set, and f is called a rank function.

Proposition 7.2 (Two Information-Theoretic Set Functions)

Let Ω be a finite and non-empty set, and let $\{X_\omega\}_{\omega \in \Omega}$ be a collection of discrete random variables. Then, the following holds:

1. The set function $f \colon 2^\Omega \to \mathbb{R}$, given by

$$f(\mathcal{T}) \triangleq \mathrm{H}(X_{\mathcal{T}}), \quad \mathcal{T} \subseteq \Omega, \tag{7.4}$$

   is a rank function.

2. The set function $f \colon 2^\Omega \to \mathbb{R}$, given by

$$f(\mathcal{T}) \triangleq \mathrm{H}(X_{\mathcal{T}} | X_{\mathcal{T}^c}), \quad \mathcal{T} \subseteq \Omega, \tag{7.5}$$

   is supermodular, monotonically increasing, and $f(\varnothing) = 0$.

There are more sub/supermodular information-theoretic set functions.

Proof.

We prove Item 1, in regard to the entropy as a set function $f\colon 2^{\Omega} \to \mathbb{R}$, given in (7.4). It is clear that $f(\varnothing) = 0$, and also f is monotonically increasing. The submodularity of f is next verified. Let $\mathcal{S} \subset \mathcal{T} \subset \Omega$ and $\omega \in \mathcal{T}^{\mathsf{c}} \triangleq \Omega \setminus \mathcal{T}$. Then,

$$\begin{aligned} f(\mathcal{T} \cup \{\omega\}) - f(\mathcal{T}) &= \mathrm{H}(X_{\mathcal{T} \cup \{\omega\}}) - \mathrm{H}(X_{\mathcal{T}}) \\ &= \mathrm{H}(X_{\omega} | X_{\mathcal{T}}) \\ &= \mathrm{H}(X_{\omega} | X_{\mathcal{S}}, X_{\mathcal{T} \setminus \mathcal{S}}) \\ &\leq \mathrm{H}(X_{\omega} | X_{\mathcal{S}}) \\ &= \mathrm{H}(X_{\mathcal{S} \cup \{\omega\}}) - \mathrm{H}(X_{\mathcal{S}}) \\ &= f(\mathcal{S} \cup \{\omega\}) - f(\mathcal{S}), \end{aligned} \tag{7.6}$$

which asserts the submodularity of $f \implies f$ is a rank function. □

Proposition 7.3 (I.S., 2022)

Let Ω be a finite set with $|\Omega| = n$. Let $f \colon 2^{\Omega} \to \mathbb{R}$ with $f(\varnothing) = 0$, and $g \colon \mathbb{R} \to \mathbb{R}$. Let the sequence $\{t_k^{(n)}\}_{k=1}^{n}$ be given by

$$t_k^{(n)} \triangleq \frac{1}{\binom{n}{k}} \sum_{\mathcal{T} \subseteq \Omega \colon |\mathcal{T}| = k} g\left(\frac{f(\mathcal{T})}{k}\right), \qquad k \in [n]. \tag{7.7}$$

- If f is submodular, and g is monotonically increasing and convex, then the sequence $\{t_k^{(n)}\}_{k=1}^{n}$ is monotonically decreasing, i.e.,

$$t_1^{(n)} \geq t_2^{(n)} \geq \ldots \geq t_n^{(n)} = g\left(\frac{f(\Omega)}{n}\right). \tag{7.8}$$

In particular,

$$\sum_{\mathcal{T} \subseteq \Omega \colon |\mathcal{T}| = k} g\left(\frac{f(\mathcal{T})}{k}\right) \geq \binom{n}{k} \, g\left(\frac{f(\Omega)}{n}\right), \qquad k \in [n]. \tag{7.9}$$

Proposition 7.3 (cont.)

- If f is submodular, and g is monotonically decreasing and concave, then the sequence $\{t_k^{(n)}\}_{k=1}^n$ is monotonically increasing.
- If f is supermodular, and g is monotonically increasing and concave, then the sequence $\{t_k^{(n)}\}_{k=1}^n$ is monotonically increasing.
- If f is supermodular, and g is monotonically decreasing and convex, then the sequence $\{t_k^{(n)}\}_{k=1}^n$ is monotonically decreasing.

I. Sason, "Information inequalities via submodularity, and a problem in extremal graph theory," *Entropy*, vol. 24, paper 597, pp. 1–31, April 2022. https://doi.org/10.3390/e24050597

Corollary 7.4

Let Ω be a finite set with $|\Omega| = n$, $f\colon 2^{\Omega} \to \mathbb{R}$, and $g\colon \mathbb{R} \to \mathbb{R}$ be convex and monotonically increasing. If

- f is a rank function,
- $g(0) > 0$ or there is $\ell \in \mathbb{N}$ such that $g(0) = \ldots = g^{(\ell-1)}(0) = 0$ with $g^{(\ell)}(0) > 0$,
- $\{k_n\}_{n=1}^{\infty}$ is a sequence such that $k_n \in [n]$, $\forall n \in \mathbb{N}$, with $k_n \underset{n\to\infty}{\longrightarrow} \infty$,

then

$$\lim_{n\to\infty} \left\{ \frac{1}{n} \log\left(\sum_{\mathcal{T}\subseteq\Omega\colon |\mathcal{T}|=k_n} g\left(\frac{f(\mathcal{T})}{k_n}\right)\right) - \mathsf{H_b}\left(\frac{k_n}{n}\right)\right\} = 0. \tag{7.10}$$

Furthermore, if $\lim\limits_{n\to\infty} \frac{k_n}{n} = \beta \in [0,1]$, then

$$\lim_{n\to\infty} \frac{1}{n} \log\left(\sum_{\mathcal{T}\subseteq\Omega\colon |\mathcal{T}|=k_n} g\left(\frac{f(\mathcal{T})}{k_n}\right)\right) = \mathsf{H_b}(\beta). \tag{7.11}$$

Corollary 7.5

Let Ω be a finite set with $|\Omega| = n$, and $f \colon 2^\Omega \to \mathbb{R}$ be submodular and nonnegative with $f(\varnothing) = 0$. Then,

- For $\alpha \geq 1$ and $k \in [n-1]$

$$\sum_{\mathcal{T} \subseteq \Omega \colon |\mathcal{T}|=k} \bigl(f^\alpha(\Omega) - f^\alpha(\mathcal{T})\bigr) \leq c_\alpha(n,k)\, f^\alpha(\Omega), \tag{7.12}$$

  with

$$c_\alpha(n,k) \triangleq \left(1 - \frac{k^\alpha}{n^\alpha}\right) \binom{n}{k}. \tag{7.13}$$

  For $\alpha = 1$, (7.12) holds with $c_1(n,k) = \binom{n-1}{k}$ regardless of the nonnegativity of f.

- If f is a rank function, then for $\alpha \geq 1$ and $k \in [n]$

$$\left(\frac{k}{n}\right)^{\alpha-1} \binom{n-1}{k-1} f^\alpha(\Omega) \leq \sum_{\mathcal{T} \subseteq \Omega \colon |\mathcal{T}|=k} f^\alpha(\mathcal{T}) \leq \binom{n}{k} f^\alpha(\Omega). \tag{7.14}$$

Specialization of Corollary 7.5 to a generalized Han's inequality

- Substituting $\alpha = 1$ and the entropy-set function of (7.4) into (7.12) gives that, for all $k \in [n-1]$,

$$\sum_{1 \le i_1 < \ldots < i_k \le n} \Big\{ \mathrm{H}(X^n) - \mathrm{H}(X_{i_1}, \ldots, X_{i_k}) \Big\} \le \binom{n-1}{k} \, \mathrm{H}(X^n), \quad (7.15)$$

  which is Fujishige's inequality (1978).

- Consequently, setting $k = n-1$ in (7.15) gives

$$\sum_{i=1}^{n} \Big\{ \mathrm{H}(X^n) - \mathrm{H}(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) \Big\} \le \mathrm{H}(X^n), \quad (7.16)$$

  which specialized to Han's inequality.

I. Sason, "Information inequalities via submodularity, and a problem in extremal graph theory," *Entropy*, vol. 24, paper 597, pp. 1–31, April 2022. https://doi.org/10.3390/e24050597

Proposition 7.4 (Generalized Version of Shearer's Lemma)

Let Ω be a finite set, let $\{\mathcal{S}_j\}_{j=1}^M$ be a finite collection of subsets of Ω (with $M \in \mathbb{N}$), and let $f \colon 2^\Omega \to \mathbb{R}$ be a set function.

1. If f is non-negative and submodular, and every element in Ω is included in at least $d \geq 1$ of the subsets $\{\mathcal{S}_j\}_{j=1}^M$, then
$$\sum_{j=1}^{M} f(\mathcal{S}_j) \geq d\, f(\Omega). \tag{7.17}$$
2. If f is a rank function, $\mathcal{A} \subset \Omega$, and every element in $\mathcal{A}$ is included in at least $d \geq 1$ of the subsets $\{\mathcal{S}_j\}_{j=1}^M$, then
$$\sum_{j=1}^{M} f(\mathcal{S}_j) \geq d\, f(\mathcal{A}). \tag{7.18}$$

Proposition 7.4 $\Longrightarrow$ Sherarer's Lemma in Proposition 3.1

Item 1 of Proposition 7.4 yields Sherarer's Lemma in Proposition 3.1 since the set function given in (7.4) is submodular, and it is also nonnegative for discrete random variables (in light of Item 1 of Proposition 7.2).

Proposition 7.4 $\Longrightarrow$ Sherarer's Lemma in Proposition 3.1

Item 1 of Proposition 7.4 yields Sherarer's Lemma in Proposition 3.1 since the set function given in (7.4) is submodular, and it is also nonnegative for discrete random variables (in light of Item 1 of Proposition 7.2).

Other Generalizations

- E. Friedgut, "Hypergraphs, entropy and inequalities," *The American Mathematical Monthly*, vol. 111, no. 9, pp. 749–760, November 2004.
- D. Gavinsky, S. Lovett, M. Saks, S. Srinivasan, "A tail bound for read-k families of functions," *Random Structures and Algorithms*, vol. 47, no. 1, pp. 1–10, August 2015.
- M. Madiman and P. Tetali, "Information inequalities for joint distributions, interpretations and applications," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2699–2713, June 2010.

Summary

Summary

- Entropy, counting, and coins weighing.
- Shearer's inequalities.
- Applications:
  - Finite Geometry.
  - Graph theory.
    - ★ cliques, and triangle-intersecting families of graphs,

Summary

- Entropy, counting, and coins weighing.
- Shearer's inequalities.
- Applications:
  - Finite Geometry.
  - Graph theory.
    - ⋆ cliques, and triangle-intersecting families of graphs,
  - Not covered in this talk:
    - ⋆ Probabilistic results in graph theory.
    - ⋆ Version of Shearer's lemma for the relative entropy.
    - ⋆ Read-k Boolean functions and Chernoff-like bounds for their sums.
    - ⋆ Counting independent sets in graphs.
    - ⋆ Counting graph homomorphisms.

Summary

- Entropy, counting, and coins weighing.
- Shearer's inequalities.
- Applications:
  - Finite Geometry.
  - Graph theory.
    - ⋆ cliques, and triangle-intersecting families of graphs,
  - Not covered in this talk:
    - ⋆ Probabilistic results in graph theory.
    - ⋆ Version of Shearer's lemma for the relative entropy.
    - ⋆ Read-k Boolean functions and Chernoff-like bounds for their sums.
    - ⋆ Counting independent sets in graphs.
    - ⋆ Counting graph homomorphisms.
- Generalizations of Shearer's and Han's inequalities:
  - Some Generalizations (I.S., 2022).
  - Not covered in this talk:
    - ⋆ Shearer's lemma on hypergraphs.
    - ⋆ Information-theoretic generalizations and counterparts.

My Related Papers on Shearer's Lemma and Its Extensions

1. I. S., "A generalized information-theoretic approach for bounding the number of independent sets in bipartite graphs," *Entropy*, vol. 23, no. 3, paper 270, pp. 1–14, 2021. https://doi.org/10.3390/e23030270
2. I. S., "Information inequalities via submodularity, and a problem in extremal graph theory," *Entropy*, vol. 24, no. 5, paper 597, pp. 1–31, 2022. https://doi.org/10.3390/e24050597