Polyhedra in information theory

Tobias Boege

$$\begin{bmatrix}\text{Department of Mathematics}\\ \text{KTH Royal Institute of Technology}\end{bmatrix} \mapsto \begin{bmatrix}\text{Department of Mathematics and Statistics}\\ \text{UiT The Arctic University of Norway}\end{bmatrix}$$

Discrete Mathematics & Geometry seminar,
TU Berlin, 08 May 2024

Entropy

Let X be a random variable taking finitely many values $\{1, \ldots, d\}$ with positive probabilities. Its *Shannon entropy* is

$$H(X) := \sum_{i=1}^{d} p(X = i) \log {}^{1}/_{p(X=i)}.$$

- H is continuous on $\Delta(d)$ and analytic on the interior.

Entropy

Let X be a random variable taking finitely many values $\{1, \ldots, d\}$ with positive probabilities. Its *Shannon entropy* is

$$H(X) := \sum_{i=1}^{d} p(X = i) \log {}^{1}\!/_{p(X = i)}.$$

- H is continuous on $\Delta(d)$ and analytic on the interior.
- A random vector $X \in \Delta(d_i : i \in N)$ is a random variable in $\Delta(\prod_{i \in N} d_i)$, so the definition of H extends to vectors.
- For a random vector $X = (X_i : i \in N)$ we have 2^N marginals and we collect their entropies in an entropy profile $h_X : 2^N \to \mathbb{R}$.
  - For example (X, Y) has entropy profile $(0, H(X), H(Y), H(X, Y)) \in \mathbb{R}^4$.

Entropy as information

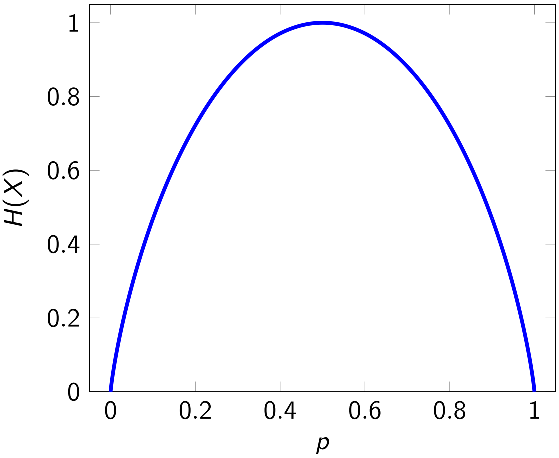


Figure: Entropy of a binary random variable X as a function of $p = p(X = \text{heads})$.

The entropy region and information inequalities

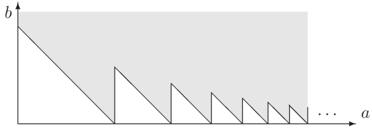
Let $\mathbf{H}_N^* \subseteq \mathbb{R}^{2^N}$ consist of all h_X where X is an N-variate discrete random vector. $\mathbf{H}_N^*$ is the image of $\bigcup_{d_1=1}^{\infty} \cdots \bigcup_{d_n=1}^{\infty} \Delta(d_1, \ldots, d_n)$ under the transcendental map $X \mapsto h_X$.

The entropy region and information inequalities

Let $\mathbf{H}_N^* \subseteq \mathbb{R}^{2^N}$ consist of all h_X where X is an N-variate discrete random vector. $\mathbf{H}_N^*$ is the image of $\bigcup_{d_1=1}^{\infty} \cdots \bigcup_{d_n=1}^{\infty} \Delta(d_1, \ldots, d_n)$ under the transcendental map $X \mapsto h_X$.

Problem

Find a description of the boundary of $\mathbf{H}_3^*$.

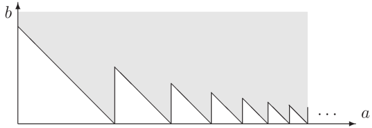


The entropy region and information inequalities

Let $\mathbf{H}_N^* \subseteq \mathbb{R}^{2^N}$ consist of all h_X where X is an N-variate discrete random vector. $\mathbf{H}_N^*$ is the image of $\bigcup_{d_1=1}^{\infty} \cdots \bigcup_{d_n=1}^{\infty} \Delta(d_1, \ldots, d_n)$ under the transcendental map $X \mapsto h_X$.

Problem

Find a description of the boundary of $\mathbf{H}_3^*$.



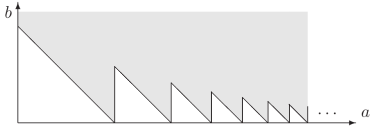
- Applications in cryptography, coding theory, engineering want to optimize linear functions over $\mathbf{H}_N^*$.

The entropy region and information inequalities

Let $\mathbf{H}_N^* \subseteq \mathbb{R}^{2^N}$ consist of all h_X where X is an N-variate discrete random vector. $\mathbf{H}_N^*$ is the image of $\bigcup_{d_1=1}^{\infty} \cdots \bigcup_{d_n=1}^{\infty} \Delta(d_1, \ldots, d_n)$ under the transcendental map $X \mapsto h_X$.

Problem

Find a description of the boundary of $\mathbf{H}_3^*$.



- Applications in cryptography, coding theory, engineering want to optimize linear functions over $\mathbf{H}_N^*$.

Theorem

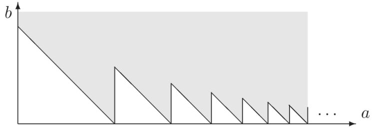
$\overline{\mathbf{H}_N^*}$ *is a convex cone of dimension* $2^N - 1$*. Furthermore* $\operatorname{relint}(\overline{\mathbf{H}_N^*}) \subseteq \mathbf{H}_N^*$.

The entropy region and information inequalities

Let $\mathbf{H}_N^* \subseteq \mathbb{R}^{2^N}$ consist of all h_X where X is an N-variate discrete random vector. $\mathbf{H}_N^*$ is the image of $\bigcup_{d_1=1}^{\infty} \cdots \bigcup_{d_n=1}^{\infty} \Delta(d_1, \ldots, d_n)$ under the transcendental map $X \mapsto h_X$.

Problem

Find a description of the boundary of $\mathbf{H}_3^*$.



- Applications in cryptography, coding theory, engineering want to optimize linear functions over $\mathbf{H}_N^*$.

Theorem

$\overline{\mathbf{H}_N^*}$ *is a convex cone of dimension* $2^N - 1$*. Furthermore* $\mathrm{relint}(\overline{\mathbf{H}_N^*}) \subseteq \mathbf{H}_N^*$.

- Elements of the dual cone (linear information inequalities) can give bounds for optimization problems.

Shannon inequalities

- A function $h : 2^N \to \mathbb{R}$ is a polymatroid if
  - $h(\emptyset) = 0$,
  - $h(I \mid K) := h(IK) - h(K) \geq 0$ for disjoint I and K,
  - $h(I : J \mid K) := h(IK) + h(JK) - h(IJK) - h(K) \geq 0$ for disjoint I, J, K.

Shannon inequalities

- A function $h : 2^N \to \mathbb{R}$ is a polymatroid if
  - $h(\emptyset) = 0$,
  - $h(I \mid K) := h(IK) - h(K) \geq 0$ for disjoint I and K,
  - $h(I : J \mid K) := h(IK) + h(JK) - h(IJK) - h(K) \geq 0$ for disjoint I, J, K.

- The set $\mathbf{P}_N$ of polymatroids is a polyhedral cone in $\mathbb{R}^{2^N}$ and $\mathbf{P}_N \supseteq \overline{\mathbf{H}_N^*} \to$ ITIP.
- The information inequalities in the dual cone of $\mathbf{P}_N$ are the Shannon inequalities.

Shannon inequalities

- A function $h : 2^N \to \mathbb{R}$ is a polymatroid if
  - $h(\emptyset) = 0$,
  - $h(I \mid K) := h(IK) - h(K) \geq 0$ for disjoint I and K,
  - $h(I : J \mid K) := h(IK) + h(JK) - h(IJK) - h(K) \geq 0$ for disjoint I, J, K.

- The set $\mathbf{P}_N$ of polymatroids is a polyhedral cone in $\mathbb{R}^{2^N}$ and $\mathbf{P}_N \supseteq \overline{\mathbf{H}_N^*} \to$ ITIP.
- The information inequalities in the dual cone of $\mathbf{P}_N$ are the Shannon inequalities.

Theorem ([Mat07])

$\overline{\mathbf{H}_N^*}$ *is not polyhedral for* $|N| \geq 4$.

- Conjecture: $\overline{\mathbf{H}_N^*}$ is not semialgebraic for $|N| \geq 4$.

Independence: geometry $\leftrightarrow$ information theory

Information-theoretical "special position" properties of discrete random variables can be formulated in terms of linear functionals on the entropy profile h_X:

Independence: geometry $\leftrightarrow$ information theory

Information-theoretical “special position” properties of discrete random variables can be formulated in terms of linear functionals on the entropy profile h_X:

Rank condition	Matroid concept	Information theory concept
$h(i) = 0$	loop	constant random variable
$h(N) = h(i) + h(N \setminus i)$	coloop	max. private information

Independence: geometry $\leftrightarrow$ information theory

Information-theoretical "special position" properties of discrete random variables can be formulated in terms of linear functionals on the entropy profile h_X:

Rank condition	Matroid concept	Information theory concept
$h(i) = 0$	loop	constant random variable
$h(N) = h(i) + h(N \setminus i)$	coloop	max. private information
$h(i \mid K) = 0$	closure operator	functional dependence
$h(K) = \sum_{k \in K} h(k)$	independent set	total independence
$h(i : j \mid K) = 0$	modular pair	conditional independence

Independence: geometry $\leftrightarrow$ information theory

Information-theoretical "special position" properties of discrete random variables can be formulated in terms of linear functionals on the entropy profile h_X:

Rank condition	Matroid concept	Information theory concept
$h(i) = 0$	loop	constant random variable
$h(N) = h(i) + h(N \setminus i)$	coloop	max. private information
$h(i \mid K) = 0$	closure operator	functional dependence
$h(K) = \sum_{k \in K} h(k)$	independent set	total independence
$h(i : j \mid K) = 0$	modular pair	conditional independence

All of these are **linear** on $\mathbf{H}^*$. Even though entropy is a transcendental function, many of these conditions are **polynomial** in the probabilities $\rightarrow$ algebraic statistics.

Beyond Shannon: Extension properties

All widely used polyhedral outer approximations to $\overline{\mathbf{H}_N^*}$ which improve upon $\mathbf{P}_N$ are derived from an extension property which is a theorem of the form:

- If $h \in \overline{\mathbf{H}_N^*}$, then there exists $\overline{h} \in \overline{\mathbf{H}_M^*}$ for some $M \supseteq N$ such that $\overline{h}|_N = h$ and some other linear conditions $\varphi(\overline{h}) \geq 0$ hold.

Beyond Shannon: Extension properties

All widely used polyhedral outer approximations to $\overline{\mathbf{H}^*_N}$ which improve upon $\mathbf{P}_N$ are derived from an extension property which is a theorem of the form:

- If $h \in \overline{\mathbf{H}^*_N}$, then there exists $\overline{h} \in \overline{\mathbf{H}^*_M}$ for some $M \supseteq N$ such that $\overline{h}|_N = h$ and some other linear conditions $\varphi(\overline{h}) \geq 0$ hold.
- The extension property is encapsulated in its cone $E^M_N = \{\, \overline{h} \in \boxed{\overline{\mathbf{H}^*_M}} : \varphi(\overline{h}) \geq 0 \,\}$.

Beyond Shannon: Extension properties

All widely used polyhedral outer approximations to $\overline{\mathbf{H}^*_N}$ which improve upon $\mathbf{P}_N$ are derived from an extension property which is a theorem of the form:

- If $h \in \overline{\mathbf{H}^*_N}$, then there exists $\overline{h} \in \overline{\mathbf{H}^*_M}$ for some $M \supseteq N$ such that $\overline{h}|_N = h$ and some other linear conditions $\varphi(\overline{h}) \geq 0$ hold.
- The extension property is encapsulated in its cone $E^M_N = \{\, \overline{h} \in \boxed{\overline{\mathbf{H}^*_M}} : \varphi(\overline{h}) \geq 0 \,\}$.

Extension principle: Let E^M_N be the cone of an extension property and $\pi^M_N : \mathbb{R}^{2^M} \to \mathbb{R}^{2^N}$ the canonical projection. Then $\overline{\mathbf{H}^*_N} \subseteq \pi^M_N(E^M_N)$.

Extension properties: Conditional product aka Copy lemma

- Consider $h \in \mathbf{P}_N$ and pick any $L \subseteq N$.
- An L-copy of N is a set M with $|N| = |M|$ and $N \cap M = L$ with a bijection $\sigma : N \to M$ fixing L pointwise.

Extension properties: Conditional product aka Copy lemma

- Consider $h \in \mathbf{P}_N$ and pick any $L \subseteq N$.
- An L-copy of N is a set M with $|N| = |M|$ and $N \cap M = L$ with a bijection $\sigma : N \to M$ fixing L pointwise. This induces an L-copy of h: $\sigma(h) \in \mathbf{P}_M$.

Extension properties: Conditional product aka Copy lemma

- Consider $h \in \mathbf{P}_N$ and pick any $L \subseteq N$.
- An L-copy of N is a set M with $|N| = |M|$ and $N \cap M = L$ with a bijection $\sigma : N \to M$ fixing L pointwise. This induces an L-copy of h: $\sigma(h) \in \mathbf{P}_M$.

The Copy lemma states:

Extension properties: Conditional product aka Copy lemma

- Consider $h \in \mathbf{P}_N$ and pick any $L \subseteq N$.
- An L-copy of N is a set M with $|N| = |M|$ and $N \cap M = L$ with a bijection $\sigma : N \to M$ fixing L pointwise. This induces an L-copy of h: $\sigma(h) \in \mathbf{P}_M$.

The Copy lemma states:

- Let $h \in \overline{\mathbf{H}_N^*}$ and $L \subseteq N$, fix an L-copy $\sigma : N \to M$ of N.

Extension properties: Conditional product aka Copy lemma

- Consider $h \in \mathbf{P}_N$ and pick any $L \subseteq N$.
- An L-copy of N is a set M with $|N| = |M|$ and $N \cap M = L$ with a bijection $\sigma : N \to M$ fixing L pointwise. This induces an L-copy of h: $\sigma(h) \in \mathbf{P}_M$.

The Copy lemma states:

- Let $h \in \overline{\mathbf{H}^*_N}$ and $L \subseteq N$, fix an L-copy $\sigma : N \to M$ of N.
- There exists $\overline{h} \in \boxed{\overline{\mathbf{H}^*_{NM}}}$ such that

$$\overline{h}|_N = h, \quad \overline{h}|_M = \sigma(h), \quad \overline{h}(N : M \mid L) = 0.$$

Extension properties: Conditional product aka Copy lemma

- Consider $h \in \mathbf{P}_N$ and pick any $L \subseteq N$.
- An L-copy of N is a set M with $|N| = |M|$ and $N \cap M = L$ with a bijection $\sigma : N \to M$ fixing L pointwise. This induces an L-copy of h: $\sigma(h) \in \mathbf{P}_M$.

The Copy lemma states:

- Let $h \in \overline{\mathbf{H}_N^*}$ and $L \subseteq N$, fix an L-copy $\sigma : N \to M$ of N.
- There exists $\overline{h} \in \boxed{\overline{\mathbf{H}_{NM}^*}}$ such that

$$\overline{h}|_N = h, \quad \overline{h}|_M = \sigma(h), \quad \overline{h}(N : M \mid L) = 0.$$

- Relaxation: only require $\overline{h} \in \boxed{\mathbf{P}_{NM}}$! This gives a tighter inner bound $\mathbf{P}_N \supseteq \cap_{L \subseteq N} \mathbf{S}_N^L \supseteq \overline{\mathbf{H}_N^*}$. Exploited numerous times: [DFZ11], [Boe23], ...

Extension properties: Ahlswede–Körner & Slepian–Wolf

The Ahlswede–Körner lemma states:

- Let $h \in \overline{\mathbf{H}^*_N}$ and $J, K \subseteq N$.
- There exists $\overline{h} \in \overline{\mathbf{H}^*_{Nz}}$ such that

$$\overline{h}|_N = h, \quad \overline{h}(z \mid K) = 0, \quad \overline{h}(I \mid z) = \overline{h}(I \mid J) \text{ for every } I \subseteq K.$$

Extension properties: Ahlswede–Körner & Slepian–Wolf

The Ahlswede–Körner lemma states:

- Let $h \in \overline{\mathbf{H}^*_N}$ and $J, K \subseteq N$.
- There exists $\overline{h} \in \overline{\mathbf{H}^*_{Nz}}$ such that

$$\overline{h}|_N = h, \quad \overline{h}(z \mid K) = 0, \quad \overline{h}(I \mid z) = \overline{h}(I \mid J) \text{ for every } I \subseteq K.$$

The Slepian–Wolf lemma states:

- Let $h \in \overline{\mathbf{H}^*_N}$ and $J, K \subseteq N$.
- There exists $\overline{h} \in \overline{\mathbf{H}^*_{Nz}}$ such that

$$\overline{h}|_N = h, \quad \overline{h}(z \mid K) = 0, \quad \overline{h}(z) = \overline{h}(K \mid J), \quad \overline{h}(K \mid Jz) = 0.$$

Outlook

- There exist many more extension properties for linear or algebraic representability of matroids (stronger properties than $\overline{\mathbf{H}^*}$).

Outlook

- There exist many more extension properties for linear or algebraic representability of matroids (stronger properties than $\overline{\mathbf{H}^*}$).
- Several infinite families of information inequalities are derived from only the Copy lemma. They have been tabulated but are not available FAIRly $\to$ GMM problem.

Rule [43] Given:

$$\begin{aligned} & aI(A;B) \\ \leq \quad & bI(A;B|C) + cI(A;C|B) + zI(B;C|A) \\ + \quad & eI(A;B|D) + fI(A;D|B) \\ + \quad & (b' + d' + z)I(B;D|A) + hI(C;D) \\ + \quad & iI(C;D|A) + zI(C;D|B) \end{aligned}$$

and

$$\begin{aligned} & a'I(A;B) \\ \leq \quad & b'I(A;B|C) + c'I(A;C|B) + d'I(B;C|A) \\ + \quad & e'I(A;B|D) + f'I(A;D|B) + g'I(B;D|A) \\ + \quad & h'I(C;D) + i'I(C;D|A) + j'I(C;D|B) \end{aligned}$$

Get:

$$\begin{aligned} & (a + a' + z)I(A;B) \\ \leq \quad & (a + b + c + f + b' + 2z)I(A;B|C) \\ + \quad & (-a + b + c + e + c' + z)I(A;C|B) \\ + \quad & (d' + z)I(B;C|A) + (e + e' + z)I(A;B|D) \\ + \quad & (f + f')I(A;D|B) \\ + \quad & (-a' + b' + e' + g' + i')I(B;D|A) \\ + \quad & (h + h' + z)I(C;D) + (i + i')I(C;D|A) \\ + \quad & (j')I(C;D|B) \end{aligned}$$

Using: RS is copy of CD over AB
Substitutions: $A\ C\ R$ S; $AD\ B\ R\ S$

Abbreviated Proof of (75): T: D-copy of A over BCRS.
L1: -a.c. +c.d. +r.cd.a +c.s.a +b.d.s +a.bs.d +2a.cr.bs +a.bs.cr +d.r.abcs +d.s.abcr
SL1: d.t.a +c.d.t +a.t.cd +c.r.t +a.t.cr +d.r.act +b.t.acdr +a.t.bs +c.s.at +b.t.acs +d.t.s +a.s.dt +b.d.ast +c.t.abds +a.r.bcst +r.ad.bcst +s.ad.bcrt +d.t.abcrs C2L1: 3t.ad.bcrs
S: C-copy of A over BDR.
L2: -2a.c. +2c.d. +a.b.cr +2a.c.br +c.ar.b +a.b.dr +4a.d.br +2a.br.d +2d.br.a +2r.cd.a +d.r.abc
SL2: c.s.b +a.b.cs +c.d.s +a.s.cd +d.s.abc +3a.s.br +3c.s.br +c.r.abs +d.r.s +a.s.dr +d.r.abs +d.br.as +c.r.ads +b.s.acdr +2c.s.abdr +2d.s.abcr
C2L2: 7s.ac.bdr
R: D-copy of C over AB.
S: c.r.a +3c.r.b +d.r.a +7d.r.b +c.d.r +2b.r.acd +r.ab.cd +9c.r.abd +3d.r.abc
C2: 16r.cd.ab

Outlook

- There exist many more extension properties for linear or algebraic representability of matroids (stronger properties than $\overline{\mathbf{H}^*}$).
- Several infinite families of information inequalities are derived from only the Copy lemma. They have been tabulated but are not available FAIRly $\rightarrow$ GMM problem.
- Want a framework to combine and iterate extension properties based on polyhedra and linear programming and certificates for the validity of information inequalities.

Outlook

- There exist many more extension properties for linear or algebraic representability of matroids (stronger properties than $\overline{\mathbf{H}^*}$).
- Several infinite families of information inequalities are derived from only the Copy lemma. They have been tabulated but are not available FAIRly $\to$ GMM problem.
- Want a framework to combine and iterate extension properties based on polyhedra and linear programming and certificates for the validity of information inequalities.

Thank you!

References

[BFP24] Michael Bamiloshin, Oriol Farràs, and Carles Padró. *A Note on Extension Properties and Representations of Matroids*. 2024. arXiv: 2306.15085 [math.CO].

[Boe23] Tobias Boege. "Selfadhesivity in Gaussian conditional independence structures". In: *Int. J. Approx. Reasoning* (2023). DOI: 10.1016/j.ijar.2023.109027.

[DFZ11] Randall Dougherty, Chris Freiling, and Kenneth Zeger. *Non-Shannon Information Inequalities in Four Random Variables*. 2011. arXiv: 1104.3602v1 [cs.IT].

[Mat06] František Matúš. "Piecewise linear conditional information inequality". In: *IEEE Trans. Inf. Theory* 52.1 (2006), pp. 236–238. DOI: 10.1109/TIT.2005.860438.

[Mat07] František Matúš. "Infinitely many information inequalities". In: *Proc. IEEE ISIT 2007*. 2007, pp. 41–44.

[Mat18] František Matúš. "Classes of matroids closed under minors and principal extensions". In: *Combinatorica* 38.4 (2018), pp. 935–954. DOI: 10.1007/s00493-017-3534-y.