Geometry of the entropy region - I

Laszlo Csirmaz

Central European University, Budapest

IHP, Paris, February 16, 2016

Outline

Entropy

Let A be a **random variable** taking k values with probability

$$p_1, p_2, \ldots, p_k, \qquad (p_1 + p_2 + \cdots + p_k = 1).$$

The **entropy** of A is

$$\boldsymbol{H}(A) \stackrel{\text{def}}{=} \sum_{i=1}^{k} -p_i \log_2(p_i).$$

Entropy

Let A be a **random variable** taking k values with probability

$$p_1, p_2, \ldots, p_k, \qquad (p_1 + p_2 + \cdots + p_k = 1).$$

The **entropy** of A is

$$\boldsymbol{H}(A) \stackrel{\text{def}}{=} \sum_{i=1}^{k} -p_i \log_2(p_i).$$

The outcome of A can be described by $\boldsymbol{H}(A)$ bits; this is the **information content** of the event A.

Coin-flipping is 1 bit: $-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1.$

Conditional entropy, mutual information

Let A, B and C be random variables. The **conditional entropy** of A given B is the *average entropy of the conditional distributions* $A\,|\,b$

$$\boldsymbol{H}(A\,|\,B) \stackrel{\text{def}}{=} \sum_{b\in B} p_b \cdot \boldsymbol{H}(A\,|\,b) = \boldsymbol{H}(AB) - \boldsymbol{H}(B) \ge 0.$$

The **mutual information** of A and B is

$$\begin{aligned}\boldsymbol{I}(A,B) &\stackrel{\text{def}}{=} \boldsymbol{H}(A) - \boldsymbol{H}(A\,|\,B) = \boldsymbol{H}(B) - \boldsymbol{H}(B\,|\,A) \\ &= \boldsymbol{H}(A) + \boldsymbol{H}(B) - \boldsymbol{H}(AB) \ge 0.\end{aligned}$$

The **conditional mutual information** of A and B given C is

$$\begin{aligned}\boldsymbol{I}(A,B\,|\,C) &\stackrel{\text{def}}{=} \sum_{c\in C} p_c \cdot \boldsymbol{I}(A\,|\,c, B\,|\,c) \\ &= \boldsymbol{H}(AC) + \boldsymbol{H}(BC) - \boldsymbol{H}(C) - \boldsymbol{H}(ABC) \ge 0.\end{aligned}$$

Outline

Before '98

Let A and B be *collection* of random variables.

Shannon inequalities

1. $\boldsymbol{H}(A) \geq 0$, $\boldsymbol{H}(\emptyset) = 0$
   – positive,
2. $\boldsymbol{H}(B) \geq \boldsymbol{H}(A)$ whenever $B \supseteq A$
   – monotone,
3. $\boldsymbol{H}(A) + \boldsymbol{H}(B) \geq \boldsymbol{H}(A \cup B) + \boldsymbol{H}(A \cap B)$
   – subadditive.

Subadditivity is equivalent to $\boldsymbol{I}(A, B|C) \geq 0$.

Before '98

Let A and B be *collection* of random variables.

Shannon inequalities

1. $\boldsymbol{H}(A) \geq 0$, $\boldsymbol{H}(\emptyset) = 0$
   – positive,
2. $\boldsymbol{H}(B) \geq \boldsymbol{H}(A)$ whenever $B \supseteq A$
   – monotone,
3. $\boldsymbol{H}(A) + \boldsymbol{H}(B) \geq \boldsymbol{H}(A \cup B) + \boldsymbol{H}(A \cap B)$
   – subadditive.

Subadditivity is equivalent to $\boldsymbol{I}(A, B|C) \geq 0$.

Are there more?

Outline

The case of one variable

Question

What values can take the entropy of a **single** variable?

The case of one variable

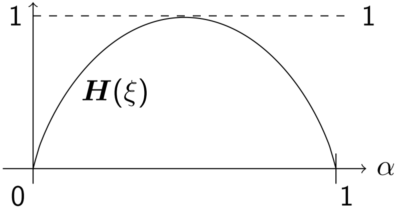
Question

What values can take the entropy of a **single** variable?

Answer

Any non-negative real value.

If $\mathrm{Prob}(\xi = 0) = \alpha$, $\mathrm{Prob}(\xi = 1) = 1 - \alpha$, then



moreover $\boldsymbol{H}(\xi\eta) = \boldsymbol{H}(\xi) + \boldsymbol{H}(\eta)$ when ξ and η are independent.

The case of two variables

Question

How does look like the **three** entropies of **two** variables?

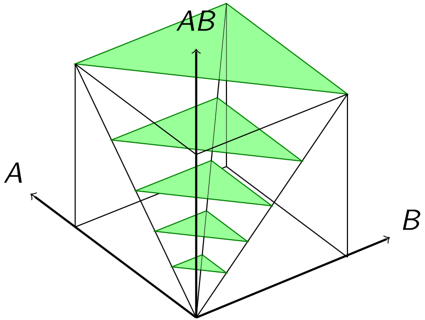
The case of two variables

Question

How does look like the **three** entropies of **two** variables?

Answer

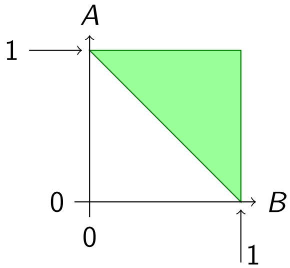
Anything is possible allowed by the Shannon-inequalities.



$$0 \le A \le AB,$$
$$0 \le B \le AB,$$
$$A + B \ge AB.$$

The case of two variables

Normalize this way: $AB = 1$



Let ξ, η, ζ be independent variables such that

$$\boldsymbol{H}(\xi) = AB - B \geq 0,$$
$$\boldsymbol{H}(\eta) = AB - A \geq 0,$$
$$\boldsymbol{H}(\zeta) = A + B - AB \geq 0.$$

Then

$$\boldsymbol{H}(\xi\zeta) = A, \quad \boldsymbol{H}(\eta\zeta) = B,$$

and

$$\boldsymbol{H}(\xi\zeta, \eta\zeta) = AB.$$

The case of three variables

Question

How does look like the **seven** entropies of **three** variables?

The case of three variables

Question

How does look like the **seven** entropies of **three** variables?

Answer

$\mathcal{C}$ is the 7-dimensional cone bounded by the Shannon inequalities.
The answer is $\mathcal{C}$ with some boundary points missing.

Research & PhD – problem

Describe the **boundary** of the cone $\mathcal{C}$.

The case of three variables

Question

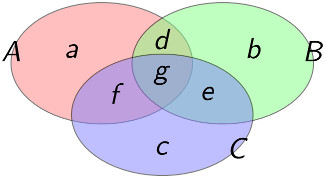
How does look like the **seven** entropies of **three** variables?

Answer

$\mathcal{C}$ is the 7-dimensional cone bounded by the Shannon inequalities.
The answer is $\mathcal{C}$ with some boundary points missing.

Research & PhD – problem

Describe the **boundary** of the cone $\mathcal{C}$.



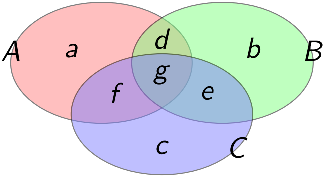
$a, b, c, d, e, f \geq 0$, $\quad g$ **can be** < 0.
$d+g, e+g, f+g \geq 0$.

E.g., $a = ABC - BC = (A \mid BC)$,

$$d = AC + BC - C - ABC = \\ = (A, B \mid C).$$

The case of three variables

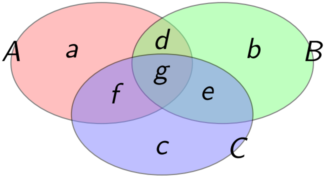
Normalize this way: $ABC=1$.



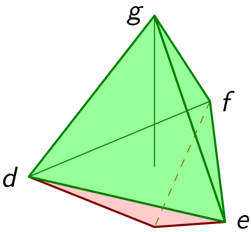
1. Take away the private info from A, B, C, i.e., set $a=b=c=0$.
2. Introduce the **barycentric** coordinates (d, e, f, g) as $d + e + f + g = 1$.
3. Visualize the possibilities

The case of three variables

Normalize this way: $ABC=1$.



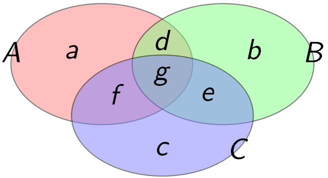
1. Take away the private info from A, B, C, i.e., set $a=b=c=0$.
2. Introduce the **barycentric** coordinates (d,e,f,g) as $d+e+f+g=1$.
3. Visualize the possibilities



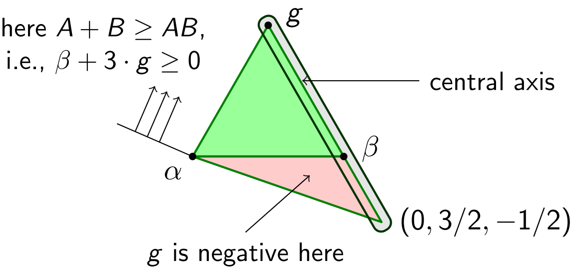
Every achievable point is a convex linear combination of these **five** extreme distributions – plus **three** accounting for the taken away private info.

The case of three variables

Alternate visualization:



1. Normalize as before: $ABC = 1$.
2. Look at the **symmetric core** $\alpha = a + b + c$, $\beta = d + e + f$.
3. Use the barycentric coordinates (α, β, g) as $\alpha + \beta + g = 1$.



The case of four variables

Question

How does look like the **fifteen** entropies of **four** variables?

The case of four variables

Question

How does look like the **fifteen** entropies of **four** variables?

Answer

No one knows exactly.

Some partial results:

- its closure is a convex cone, and only boundary points are missing – Zhang and Yeung (1997); Matus (2007)
- It is a **proper** subset of the cone determined by the Shannon inequalities – Zhang and Yeung (1998)
- It has a polyhedral inner core – the **Ingleton base**, which is surrounded by six isomorphic protrusions – Matus and Studeny (1999)
- The closure is **not** polyhedral, thus no finite set of inequalities determines it – Matus (2007)

The case of five and more variables

Question

How does look like the $2^N - 1$ entropies of $N \geq 5$ random variables?

The case of five and more variables

Question

How does look like the $2^N - 1$ entropies of $N \geq 5$ random variables?

Answer

Much less is known than even for 4 variables.

Lower bound on the share size in perfect secret sharing schemes is an optimization problem on the entropies of N random variables.
Conjecture: the best lower bound is **exponential** in N
Best known lower bound is **sublinear**, it's $N/\log N$, but exponential lower bound for **linear schemes** (*A. Gál*)

Research problem

Improve the log factor in the above estimate.

Outline

How to define a distribution?

Simplest method: list the values and the probabilities.

ξ_1	ξ_2	$\dots$	ξ_n	Prob
u_1	v_1	$\dots$	z_1	p_1
u_2	v_1	$\dots$	z_1	p_2
u_1	v_2	$\dots$	z_1	p_3
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
u_j	v_k	$\dots$	z_ℓ	p_s

The probabilities sum to 1:
$\sum p_i = 1$.

How to define a distribution?

Simplest method: list the values and the probabilities.

ξ_1	ξ_2	$\dots$	ξ_n	Prob
u_1	v_1	$\dots$	z_1	p_1
u_2	v_1	$\dots$	z_1	p_2
u_1	v_2	$\dots$	z_1	p_3
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
u_j	v_k	$\dots$	z_ℓ	p_s

An example: the *ringing bells*

How to define a distribution?

Simplest method: list the values and the probabilities.

ξ_1	ξ_2	$\dots$	ξ_n	Prob
u_1	v_1	$\dots$	z_1	p_1
u_2	v_1	$\dots$	z_1	p_2
u_1	v_2	$\dots$	z_1	p_3
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
u_j	v_k	$\dots$	z_ℓ	p_s

The probabilities sum to 1: $\sum p_i = 1$.

An example: the *ringing bells*

We have two ropes, c and d. When **any of them** is pulled, a rings, when **both** are pulled, b rings. Pull each rope independently with $1/2$ probability.

a	b	c	d	Prob
0	0	0	0	1/4
1	0	0	1	1/4
1	0	1	0	1/4
1	1	1	1	1/4

Entropies of the marginal

To get the marginal for a subset of variables: take their columns, merge identical rows, and sum the probabilities.

Original

a	b	c	d	Prob
0	0	0	0	1/4
1	0	0	1	1/4
1	0	1	0	1/4
1	1	1	1	1/4

Marginal on bc

b	c	Prob
0	0	1/2
0	1	1/4
1	1	1/4

Marginal on ab

a	b	Prob
0	0	1/4
1	0	1/2
1	1	1/4

The entropy is $\boldsymbol{H} = \sum_i -p_i \log_2(p_i)$. Since $-(1/4)\log_2(1/4) = 1/2, \quad -(1/2)\log_2(1/2) = 1/2$, thus we have

$$\boldsymbol{H}(abcd) = 2, \qquad \boldsymbol{H}(bc) = 3/2, \qquad \boldsymbol{H}(ab) = 3/2.$$

Outline

Common information

Definition

For random variables ξ and η their **common information** is another random variable c such that

1. both ξ and η determines c, that is

$$\boldsymbol{H}(c\xi) = \boldsymbol{H}(\xi), \quad \text{and} \quad \boldsymbol{H}(c\eta) = \boldsymbol{H}(\eta);$$

2. any information present both in ξ and η can be extracted from c alone: $\boldsymbol{I}(\xi, \eta) = \boldsymbol{H}(c)$, or, equivalently, $\boldsymbol{I}(\xi, \eta \mid c) = 0$.

Typically, common information does not need to exist. If it does, it has important consequences on the entropy structure.

Consequence of common information

Theorem

Suppose a, b, c, d are random variables; a and b have common information. Then

$$\boldsymbol{I}(a,b) \le \boldsymbol{I}(a,b\,|\,c) + \boldsymbol{I}(a,b\,|\,d) + \boldsymbol{I}(c,d).$$

Proof.

For five random variables a, b, c, d, e this is a Shannon inequality*:

$$\boldsymbol{H}(e) \le 2\boldsymbol{H}(e\,|\,a) + 2\boldsymbol{H}(e\,|\,b) + \boldsymbol{I}(a,b\,|\,c) + \boldsymbol{I}(a,b\,|\,d) + \boldsymbol{I}(c,d).$$

If e is the common information for a and b, then $\boldsymbol{H}(e) = \boldsymbol{I}(a,b)$, $\boldsymbol{H}(e\,|\,a) = \boldsymbol{H}(e\,|\,b) = 0$, and we are done. □

*`http://xitip.epfl.ch` or `https://github.com/lcsirmaz/minitip`

Consequence of common information

Theorem

Suppose a, b, c, d are random variables; a and b have common information. Then

$$\boxed{\boldsymbol{I}(a,b) \le \boldsymbol{I}(a,b\,|\,c) + \boldsymbol{I}(a,b\,|\,d) + \boldsymbol{I}(c,d).}$$

Proof.

For five random variables a, b, c, d, e this is a Shannon inequality*:

$$\boldsymbol{H}(e) \le 2\boldsymbol{H}(e\,|\,a) + 2\boldsymbol{H}(e\,|\,b) + \boldsymbol{I}(a,b\,|\,c) + \boldsymbol{I}(a,b\,|\,d) + \boldsymbol{I}(c,d).$$

If e is the common information for a and b, then $\boldsymbol{H}(e) = \boldsymbol{I}(a,b)$, $\boldsymbol{H}(e\,|\,a) = \boldsymbol{H}(e\,|\,b) = 0$, and we are done. □

This is the **Ingleton inequality**.

*`http://xitip.epfl.ch` or `https://github.com/lcsirmaz/minitip`

The Ingleton score

For the “bells“ distribution

$$\begin{aligned} \boldsymbol{H}(a) = \boldsymbol{H}(b) &= 0.8112\ldots \\ \boldsymbol{H}(c) = \boldsymbol{H}(d) &= 1 \\ \boldsymbol{H}(ab) &= 1.5 \\ \boldsymbol{I}(a,b) &= 0.1225\ldots \\ \boldsymbol{I}(a,b\,|\,c) = \boldsymbol{I}(a,b\,|\,d) &= 0 \end{aligned}$$

a	b	c	d	Prob
0	0	0	0	1/4
1	0	0	1	1/4
1	0	1	0	1/4
1	1	1	1	1/4

The **Ingleton score** for this distribution is

$$\frac{-\boldsymbol{I}(a,b) + \boldsymbol{I}(a,b|c) + \boldsymbol{I}(a,b|d) + \boldsymbol{I}(c,d)}{\boldsymbol{H}(abcd)} = -0.0612\ldots$$

Research problem

Give better lower and upper bounds on the Ingleton score.

Presently they are -0.15789 and -0.09243.