Entropy, partitions, groups and association schemes, 1

Peter Cameron

CIRCA/Algebra seminar

February 2014

The purpose of this talk is to define the entropy function of a family of random variables on a finite probability space, and to prove the theorem that, up to a multiple, any such function can be approximated by one associated with a family of subgroups of a finite group.
Terence Chan, at the workshop on “Information Flows and Information Bottlenecks”, at QMUL in late 2012, stated the theorem without proof in his talk, and sketched the proof on the common room table; here are the details.

Finite probability spaces

A finite probability space consists of a finite set Ω with a function $\mathbb{P}$ from the subsets of Ω to the real numbers satisfying (simplified versions of) Kolmogorov's axioms:

- $\mathbb{P}(A) \geq 0$ for any $A \subseteq \Omega$;
- $\mathbb{P}(\Omega) = 1$;
- If $A \cap B = \emptyset$ then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

Of course it can be specified by defining $\mathbb{P}$ on elements of Ω (these values being non-negative and summing to 1) and then letting $\mathbb{P}(A)$ be the sum of the values of $\mathbb{P}$ over the elements of A.

A probability space Ω is uniform if $\mathbb{P}(a)$ is constant for $a \in \Omega$ (the constant, of course, being $1/|\Omega|$).

While the origin of probability theory is usually credited to Pascal, it was Fermat who proposed this model.

Random variables

Voltaire said of the Holy Roman Empire that it is "neither holy, nor Roman, nor an empire".
In a similar way, a random variable is neither random nor a variable; it is a function on the set Ω. In what follows, we are not interested in the codomain of the function.
A random variable f defines a partition $F = \{A_1, \ldots, A_k\}$ of Ω called the kernel of f: it is the partition of Ω into inverse images of points in the range of f.

Entropy

Let $F = \{A_1, \ldots, A_k\}$ be a partition of Ω. Let $p_i = \mathbb{P}(A_i)$, and define

$$h(p_1, p_2, \ldots, p_k) = -\sum_{i=1}^{k} p_i \log p_i.$$

The base of the logarithms is not too important; it is usually taken to be either e or 2, or in coding theory to be the size of the alphabet. Then the entropy of the partition F (or of the random variable f) is defined to be $H(F) = h(p_1, p_2, \ldots, p_k)$.
Entropy is a measure of our uncertainty about the value of f. In particular, if f takes k different values, then $H(f) \leq \log k$, with equality if and only if $p_1 = p_2 = \cdots = p_k = 1/k$. (We call a random variable satisfying the last condition uniform.)

Joint random variables

Now let $f_1, f_2, \ldots, f_r$ be random variables on Ω; suppose that the codomain of f_i is S_i. For $I \subseteq \{1, \ldots, r\}$, we define the joint random variable f_I to be the function from Ω to $\prod_{i \in I} S_i$ defined coordinatewise, so that the ith coordinate of $f_I(a)$ is $f_i(a)$ for $i \in I$.

Partitions of a set are ordered by refinement; they form a lattice, in which the meet of two partitions is their coarsest common refinement. The partition F corresponding to the random variable f_I is the meet of the partitions corresponding to all f_i for $i \in I$:

$$F_I = \bigwedge_{i \in I} F_i.$$

So, given a family of random variables, the partitions of the probability space corresponding to the joint random variables form the meet-semilattice generated by the partitions for the original variables.

Entropy of joint random variables

The entropy of the joint random variables defines a point in the positive orthant of $\mathbb{R}^{2^r}$, with one coordinate for each subset of $\{1, 2, \ldots, r\}$. Since the entropy of $f_\emptyset$ is 0, we can omit the first coordinate, and regard the point as lying in $\mathbb{R}^{2^r-1}$. Let Γ_r be the set of all such points (arising from all r-tuples of random variables on finite probability spaces).
We will also consider a "projective" version $P\Gamma_r$, where we represent a point by the line through the origin (a point in projective space of dimension $2^r - 2$). (This makes sense because the base of the logarithms is arbitrary.)
A big problem of information theory is:

Problem

Describe the set Γ_r.

Group random variables

Now let G be a finite group; we regard it as a uniform probability space. (This corresponds to the instruction "choose an element of G uniformly at random".) If K is any subgroup of G, then K defines a random variable F_K which maps each element $g \in G$ to the right coset Kg of K which contains it. This is a uniform random variable, so its entropy is $\log |G : K|$, where $|G : K|$ is the index of K in G (the number of cosets). We denote it by f_K, and the corresponding partition of G (into right cosets of K) by F_K.
We call f_k a group random variable.

Group families

If $K_1, \ldots, K_r$ are subgroups of G, then the joint random variable of f_{K_i} for $i \in I \subseteq \{1, \ldots, r\}$ is just f_{K_I}, where K_I is the subgroup

$$K_I = \bigcap_{i \in I} K_i,$$

and the corresponding partition is F_{K_I}, the coset partition of K_i. We will say that a family of random variables is a group family if it is defined in this way by a family of subgroups of a finite group. We will, by abuse of notation, often write F_i and F_I in place of F_{K_i} and F_{K_I} in this situation.

The main theorem

Theorem

The projective point corresponding to any point of Γ_r can be approximated arbitrarily closely by a group family of random variables.

Corollary

Any linear inequality satisfied by the entropy functions of all group families of random variables is satisfied by all entropy functions.

In what follows, I prove this theorem. We begin with a triviality:

Lemma

Any point of Γ_r can be approximated arbitrarily closely by a point arising from a family of random variables all of whose probabilities are rational.

Multinomial coefficients

Let $m_1, \ldots, m_k$ be positive integers summing to n. The multinomial coefficient $\binom{n}{m_1, m_2, \ldots, m_k}$ is defined to be $n!/(m_1!m_2! \cdots m_k!)$, the number of choices of subsets $A_1, A_2, \ldots, A_k$, with $|A_i| = m_i$ for $i = 1, \ldots, k$, which form a partition of $\{1, \ldots, n\}$.

Lemma

Let $p_1, p_2, \ldots, p_k$ be positive rational numbers summing to 1. Then

$$\log \binom{n}{p_1 n, p_2 n, \ldots, p_k n} \sim n h(p_1, p_2, \ldots, p_k)$$

as $n \to \infty$ through values such that all $p_i n$ are integers.

Warning: The symbol $\sim$ is used here in the sense of asymptotics, that is, $f \sim g$ means that $f/g \to 1$; not in the probability sense (same distribution).

Proof of the lemma

The proof is straightforward from a weak form of Stirling's formula:

$$\log n! = n \log n - n + O(\log n).$$

Then

$$\begin{aligned}
\log \binom{n}{p_1 n, \ldots, p_k n} &= n \log n - n - \sum p_i n \log p_i n + \\
&\qquad + \sum p_i n + O(\log n) \\
&= n \log n (1 - \sum p_i) - n \sum p_i \log p_i + O(\log n) \\
&= n h(p_1, \ldots, p_k) + O(\log n).
\end{aligned}$$

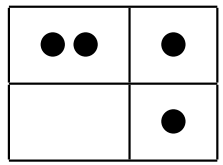
Construction

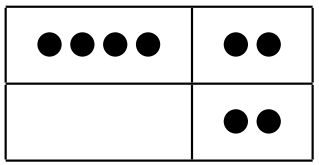
Now suppose we are given a family $\mathcal{F}$ of r random variables on the probability space Ω. By Lemma 3, we may assume that the probabilities associated with the random variables and their joint variables are all rational.
Next we show that we may assume that Ω is a uniform probability space. First we factor out the partition defined by the joint random variable corresponding to all the variables in $\mathcal{F}$. Now choose n such that, for every probability p which occurs, np is an integer. Finally, if a point of Ω has probability p, replace it by np points, each with probability $1/n$.

An example

This example shows what is going on. The first table is the joint distribution of two random variables, while the second and third show two uniform spaces which give the same probabilities and hence the same entropies.

$\frac{1}{2}$	$\frac{1}{4}$
0	$\frac{1}{4}$





Replacing each point in the middle picture by s points, for a fixed positive integer s, would give another space with the right properties.

Young subgroups of the symmetric group

Let $G = S_n$ be the symmetric group on the set $\{1, \ldots, n\}$. For any partition P of $\{1, \ldots, n\}$, the corresponding Young subgroup S_P consists of all permutations which fix the parts of P setwise.

Note that we have

$$S_P \cap S_Q = S_{P \wedge Q},$$

where $P \wedge Q$ is the meet of P and Q in the partition lattice. So the joint random variables corresponding to a subfamily of the subgroups gives the same partition (and so the same entropy) as the random variable corresponding to the intersection of these subgroups.

Moreover, G acts transitively on ordered partitions with given shape, and the stabiliser of a partition is the corresponding Young subgroup; so, if P has parts of size $m_1, \ldots, m_k$, then

$$|S_n : S_P| = \binom{n}{m_1, m_2, \ldots, m_k}.$$

Proof of the theorem

Given a family $\mathcal{F}$ of random variables defined on a uniform probability space of size n (where we take $\Omega = \{1, \ldots, n\}$), we consider the group family where $G = S_n$ and the subgroups H_i are the Young subgroups associated with the partitions.
The entropies in the group family are logarithms of multinomial coefficients; by Lemma 4, these (when divided by n) approximate the entropies of the given family of random variables.
Moreover, as we observed earlier, we can make n as large as we like by repeating each point a fixed number of times. So the theorem is proved.

Example

To return to our example: suppose that we repeat each point s times, and number the points from left to right and top to bottom. Then 4 is replaced by $4s$, and we have for the group family

$$\begin{aligned} H(F_1) &= \log \binom{4s}{3s, s} \sim 4sh(\tfrac{3}{4}, \tfrac{1}{4}), \\ H(F_2) &= \log \binom{4s}{2s, 2s} \sim 4sh(\tfrac{1}{2}, \tfrac{1}{2}), \\ H(F_{12}) &= \log \binom{4s}{2s, s, s} \sim 4sh(\tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{4}). \end{aligned}$$

Some problems

Problem

- *What families of groups other than symmetric groups give rise to entropy functions which can projectively approximate any entropy function?*
- *Given a family of groups, what are the restrictions on the entropy functions arising?*

For example, Terence Chan pointed out that, for abelian groups, a linear inequality holds for the entropy functions which is not valid in general (the Ingleton inequality).

Are these results useful to group theorists?

Remark

We can reverse the arguments to use entropy inequalities to give multiplicative inequalities for the indices of subgroups and their intersections.

For example, the basic Shannon inequality

$$H(f_1, f_2) \leq H(f_1) + H(f_2)$$

gives, for subgroups K_1, K_2 of a group G, the well-known result

$$|G : K_1 \cap K_2| \leq |G : K_1| \cdot |G : K_2|,$$

or more briefly

$$|G| \cdot |K_1 \cap K_2| \geq |K_1| \cdot |K_2|.$$

For example

More interesting are the recently discovered non-Shannon entropy inequalities, whose consequences for groups have not yet been investigated.
The first non-Shannon entropy inequality was found by Zhang and Yeung in 1998. It asserts that

$$\begin{aligned} & H(f_1,f_2)+H(f_1,f_3)+3((H(f_2,f_3)+H(f_2,f_4)+H(f_3,f_4)) \\ \geq\ & 2H(f_2)+2H(f_3)+H(f_4)+H(f_1,f_4)+H(f_1,f_2,f_3)+4H(f_2,f_3,f_4) \end{aligned}$$

or

$$\begin{aligned} & |K_1\cap K_2|\cdot|K_1\cap K_3|\cdot(|K_2\cap K_3|\cdot|K_2\cap K_4|\cdot|K_3\cap K_4|)^3 \\ \leq\ & (|K_2|\cdot|K_3|)^2\cdot|K_4|\cdot|K_1\cap K_4|\cdot|K_1\cap K_2\cap K_3|\cdot|K_2\cap K_3\cap K_4|^4. \end{aligned}$$

A question

By Chan's Theorem, the group-theoretic inequality on the last slide is equivalent to the entropy ineequality above it.
And yet the paper of Zhang and Yeung was a very important breakthrough. Had some group theorist found the other inequality, would it even have been regarded as worth publishing?
A prize to anyone who can find a group-theoretic application of it, or any of the hundreds of other non-Shannon inequalities that have been found in the context of network coding.