Entropy, Partitions and Association Schemes

R. A. Bailey
University of St Andrews /
Queen Mary, University of London (emerita)
rab@mcs.st-and.ac.uk, r.a.bailey@qmul.ac.uk

(Work inspired by conversations with Peter J. Cameron, Terence Chan, Max Gadouleau, Sune Jakobsen, Søren Riis)

Network Coding, Partitions and Security
Durham University, 20 November 2013

Finite probability space

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

- if $\Delta \subseteq \Omega$ then $\mathbb{P}(\Delta) \geq 0$;
- $\mathbb{P}(\Omega) = 1$;
- if $\Delta \cap \Gamma = \emptyset$ then $\mathbb{P}(\Delta \cup \Gamma) = \mathbb{P}(\Delta) + \mathbb{P}(\Gamma)$.

It suffices to specify $\mathbb{P}(\omega) = \mathbb{P}(\{\omega\})$ for all ω in Ω, with $\mathbb{P}(\omega) \geq 0$ and $\sum_{\omega \in \Omega} \mathbb{P}(\omega) = 1$. Then

$$\mathbb{P}(\Delta) = \sum_{\omega \in \Delta} \mathbb{P}(\omega).$$

Uniform finite probability space

- If necesssary, approximate each point probability $\mathbb{P}(\omega)$ by a rational number.
- Choose an integer N so that $N\,\mathbb{P}(\omega)$ is an integer for all ω in Ω.
- Replace each element ω by $N\,\mathbb{P}(\omega)$ elements, all with probablity $1/N$.
- Throw away any element ω for which $\mathbb{P}(\omega) = 0$.
- Rename Ω and all of its subsets, so that they consist of these new elements. In particular, $|\Omega| = N$.

If Δ is a renamed subset, it still has the same probability as before, but now we can write

$$\mathbb{P}(\Delta) = \frac{|\Delta|}{N}.$$

Now $(\Omega, \mathbb{P})$ is a uniform probability space. Assume $N > 1$.

Partitions and random variables

A partition of Ω is a set of mutually disjoint non-empty subsets of Ω whose union is Ω. The subsets are called parts.

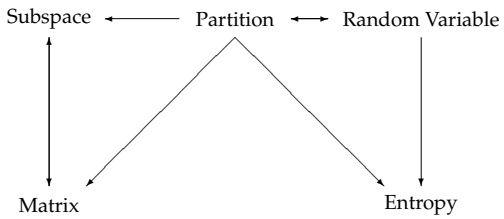
A random variable on Ω is any function on Ω.
Random variables are typically denoted $X, Y, \ldots$.

$$\mathbb{P}(X = x) \text{ means } \mathbb{P}\left(\{\omega \in \Omega : X(\omega) = x\}\right) = \mathbb{P}\left(X^{-1}(x)\right).$$

The random variable X defines a partition of Ω into the inverse images $X^{-1}(x)$ of the points x in the range of X.

The entropy (coming up) of the random variable X is defined entirely in terms of this partition.

Preview: things from partitions



Entropy of a random variable

Let X be a random variable.
Let the parts of the partition defined by X be $\Delta_1, \ldots, \Delta_n$.
The entropy $H(X)$ of X is defined by

$$\begin{aligned} H(X) &= -\sum_{i=1}^{n} \mathbb{P}(\Delta_i) \log(\mathbb{P}(\Delta_i)) \\ &= -\sum_{i=1}^{n} \frac{|\Delta_i|}{N} \log\left(\frac{|\Delta_i|}{N}\right). \end{aligned}$$

The base of the logarithms is not really important.

Entropy of a partition

Let F be a partition of Ω.
For $\omega \in \Omega$, let $X_F(\omega)$ be the part of F containing ω.
Then X_F is a random variable; if Δ is a part of F then

$$\mathbb{P}(X_F = \Delta) = \mathbb{P}(\Delta) = \frac{|\Delta|}{N}.$$

If the parts of F are $\Delta_1, \ldots, \Delta_n$ then

$$\begin{aligned} H(F) = H(X_F) &= -\sum_{i=1}^{n} \frac{|\Delta_i|}{N} \log\left(\frac{|\Delta_i|}{N}\right) \\ &= \frac{1}{N}\sum_{i=1}^{n} |\Delta_i| \left(\log(N) - \log(|\Delta_i|)\right) \\ &= \log N - \frac{1}{N}\sum_{i=1}^{n} |\Delta_i| \log(|\Delta_i|). \end{aligned}$$

Entropy of a uniform partition

The partition F is uniform if all of its parts have the same size.

If that size is k, then $N = nk$ and

$$\begin{aligned} H(F) &= \log N - \frac{1}{N}\sum_{i=1}^{n} |\Delta_i| \log(|\Delta_i|) \\ &= \log N - \frac{1}{N} nk \log k \\ &= \log N - \log k = \log n. \end{aligned}$$

Uncertainty

In general, if F has n parts then $H(F) \leq \log n$,
with equality if and only if F is uniform.

Entropy is a measure of uncertainty.
If there are more parts, we have less chance of guessing the right one.
If the parts have unequal sizes, we are better off if we bet on the larger parts.

Subspace defined by a partition

Consider the real vector space $\mathbb{R}^\Omega$ of all functions from Ω to $\mathbb{R}$. These are real random variables, but it is convenient to think of them as column vectors.

Partition F defines the subspace V_F of $\mathbb{R}^\Omega$ consisting of functions which are constant on each part of F.

Thus $\dim(V_F) = n$ if F has n parts.

Matrix defined by a partition

By an "$\Omega \times \Omega$ matrix" I mean a matrix that is not only of size $N \times N$ but also has its rows and columns labelled by the elements of Ω.
A partition F of Ω defines the $\Omega \times \Omega$ matrix P_F whose (α, β) entry is given by

$$P_F(\alpha, \beta) = \begin{cases} \frac{1}{k} & \text{if } X_F(\alpha) = X_F(\beta) \text{ and } |X_F(\alpha)| = k \\ 0 & \text{if } X_F(\alpha) \neq X_F(\beta). \end{cases}$$

If $y \in \mathbb{R}^\Omega$ then $(P_F y)(\alpha)$ is the average of the values $y(\beta)$ for β in $X_F(\alpha)$. Thus P_F is sometimes called the *F-averaging matrix*.
In fact, it is the matrix of orthogonal projection onto V_F under the standard inner product on $\mathbb{R}^\Omega$.

Two special cases

Equality partition E whose parts are singletons	Universal partition U which has a single part
X_E takes N values, with equal probability	X_U is constant
$H(E) = \log N$	$H(U) = 0$
$V_E = \mathbb{R}^\Omega$	V_U is the one-dimensional subspace of constant functions
$P_E = I$ (identity matrix)	$P_U = N^{-1}J$ where J is the all-1 matrix

Two partitions

Let F and G be two partitions of Ω.
If G has more parts than F,
it does not necessarily follow that $H(G) > H(H)$.

G | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 8 |

F | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |

$$H(F) = \log 8 = 3\log 2$$

$$H(G) = \log(16) - \frac{8}{16}\log 8 = 4\log 2 - \frac{1}{2}(3\log 2) = \frac{5}{2}\log 2 < H(F).$$

Partial order on partitions

Let F and G be partitions of Ω.
Then F is finer than G (written $F \preceq G$)
if every F-part is contained in a single G-part.

Consider spitting one part of G into two parts, of sizes x and y.

$$x\log x + y\log y < x\log(x+y) + y\log(x+y) = (x+y)\log(x+y).$$

Lemma
If $F \prec G$ then $H(F) > H(G)$.

Conditional probability

Let Θ be a non-empty subset of Ω and let X be a random variable on Ω.
The conditional random variable $X \mid \Theta$ is defined as follows.
The probability space is just Θ.
If x is in the range of X then

$$\begin{aligned}\mathbb{P}((X \mid \Theta) = x) &= \mathbb{P}(X = x \mid \Theta)\\ &= \frac{\mathbb{P}((X=x)\cap\Theta)}{\mathbb{P}(\Theta)}\\ &= \frac{|\{\omega \in \Theta : X(\omega) = x\}|}{|\Theta|}.\end{aligned}$$

Independence

Let X and Y be random variables on Ω.
Then X and Y are independent of each other if

$$\mathbb{P}(X = x \text{ and } Y = y) = \mathbb{P}(X = x)\,\mathbb{P}(Y = y)$$

for all x in the range of X and all y in the range of Y.

Equivalently, X and Y are independent if

$$\mathbb{P}(X = x \mid Y = y) = \mathbb{P}(X = x)$$

for all x and y.

Two partitions; two random variables

If F and G are partitions of Ω then
their infimum $F \wedge G$ is the partition each of whose parts is the non-empty intersection of an F-part with a G-part.

$$F \wedge G \preceq F \qquad \text{and} \qquad F \wedge G \preceq G.$$

If X and Y are random variables on Ω then they define a joint random variable (X, Y) by

$$(X, Y)(\omega) = (X(\omega), Y(\omega))$$

for ω in Ω.

If X defines the partition F and Y defines the partition G,
then the partition defined by (X, Y) is $F \wedge G$.

The entropy is $H((X, Y)) = H(X, Y) = H(F \wedge G)$.

Easiest case

Lemma
Let F and G be partitions of Ω.
The following statements are equivalent.

(i) $F \preceq G$.
(ii) $F \wedge G = F$.
(iii) *For every part Δ of F, the conditional random variable $X_G \mid (X_F = \Delta)$ takes a single value with non-zero probability.*
(iv) $H(X_F, X_G) = H(X_F)$.
(v) $V_G \leq V_F$.
(vi) $P_F P_G = P_G P_F = P_G$.

Mutual information

Given random variables X and Y on Ω,
their mutual information is

$$I(X,Y) = H(X) + H(Y) - H(X,Y).$$

If the corresponding partitions are F and G, this becomes

$$I(F,G) = H(F) + H(G) - H(F \wedge G).$$

We shall show that $I(X,Y) \geq 0$.

Information in two random variables

Let Γ_2 consist of all vectors of the form

$$(H(X), H(Y), H(X,Y))$$

as X and Y vary over all pairs of random variables over all uniform finite probability spaces.

If X and Y have partitions F and G, then $F \wedge G \preceq G$ so

$$H(X,Y) = H(F \wedge G) \geq H(G) = H(Y) \geq 0,$$

$$H(X,Y) \geq H(X) \geq 0,$$

$$H(X) + H(Y) - H(X,Y) = I(X,Y) \geq 0.$$

Do these inequalities define Γ_2?

Entropy of conditional random variables

Let Γ be a part of the partition defined by the random variable Y.
Then

$$H(X \mid \Gamma) = \log(|\Gamma|) - \frac{1}{|\Gamma|} \sum_{\Delta} |\Gamma \cap \Delta| \log(|\Gamma \cap \Delta|),$$

where the sum is over the parts Δ of the partition defined by X.

So we define

$$\begin{aligned} H(X \mid Y) &= \frac{1}{N} \sum_{\Gamma} |\Gamma| H(X \mid \Gamma) \\ &= \frac{1}{N} \sum_{\Gamma} |\Gamma| \log(|\Gamma|) - \frac{1}{N} \sum_{\Gamma} \sum_{\Delta} |\Gamma \cap \Delta| \log(|\Gamma \cap \Delta|) \\ &= H(X,Y) - H(Y). \end{aligned}$$

Two conditional entropies

$$H(X \mid Y) = H(X,Y) - H(Y).$$

Suppose that random variables Y_1 and Y_2 define partitions G_1 and G_2.
If $G_1 \preceq G_2$ then $Y_2 \mid Y_1$ is constant.
Then we can show that

$$H(X \mid Y_2) \geq H(X \mid Y_1).$$

Two partitions; two random variables (again)

If F and G are partitions of Ω then their supremum $F \vee G$ is the finest partition of Ω which is coarser than both F and G.

It is immediate that $F \wedge G \preceq F \preceq F \vee G$, $F \wedge G \preceq G \preceq F \vee G$, and that $V_F \cap V_G = V_{F \vee G}$.

If X and Y are random variables on Ω with corresponding partitions F and G then their common random variable is $X_{F \vee G}$.

Information

Let X and Y be random variables on Ω with corresponding partitions F and G.

Then $G \preceq F \vee G$, and so $H(F \mid F \vee G) \geq H(F \mid G)$.

$$H(F, F \vee G) - H(F \vee G) \geq H(F,G) - H(G)$$

$$H(F) - H(F \vee G) \geq H(F \wedge G) - H(G)$$

$$H(F) + H(G) - H(F \wedge G) \geq H(F \vee G)$$

$$I(X,Y) \geq H(F \vee G) \geq 0$$

Proportional meeting

Partitions F and G of Ω meet proportionately if, for all parts Δ of F and Γ of G,

$$|\Delta \cap \Gamma| = \frac{|\Delta|\,|\Gamma|}{N}.$$

Example

F gives the rows; G gives the columns.

1	3	4
2	6	8

Entropy under proportional meeting

Lemma

Let F and G be partitions of Ω. The following statements are equivalent.

(i) *The partitions F and G meet proportionately.*

(ii) *The random variables X_F and X_G are independent of each other.*

(iii) $H(X_F, X_G) = H(X_F) + H(X_G)$.

(iv) $P_F P_G = P_G P_F = P_U$.

Proof of (i) implies (iii)

If F and G meet proportionately, then

$$\begin{aligned}
H(X_{F\wedge G}) &= \log N - \frac{1}{N}\sum_i\sum_j \frac{|\Delta_i|\,|\Gamma_j|}{N}\log\left(\frac{|\Delta_i|\,|\Gamma_j|}{N}\right)\\
&= \log N - \frac{1}{N}\sum_j \frac{|\Gamma_j|}{N}\sum_i |\Delta_i|\log(|\Delta_i|)\\
&\quad - \frac{1}{N}\sum_i \frac{|\Delta_i|}{N}\sum_j |\Gamma_j|\log(|\Gamma_j|)\\
&\quad + \frac{1}{N^2}\sum_i |\Delta_i|\sum_j |\Gamma_j|\log N\\
&= 2\log N - \frac{1}{N}\sum_i |\Delta_i|\log(|\Delta_i|) - \frac{1}{N}\sum_j |\Gamma_j|\log(|\Gamma_j|)\\
&= H(X_F) + H(X_G).
\end{aligned}$$

Orthogonal partitions

Partitions F and G of Ω are orthogonal to each other (written $F \perp G$) if they meet proportionately within each part of $F \vee G$.

Example

5	0	0	0
0	1	3	1
0	2	6	2

Some special cases: F is orthogonal to itself; if $F \preceq G$ then F is orthogonal to G; and if F meets G proportionately then F is orthogonal to G.

Generalizing the previous two lemmas

Lemma

Let F and G be partitions of Ω. If $F \perp G$ then the following hold.

(i) *X_F and X_G are conditionally independent given $X_{F\vee G}$. (This means that $X_F \mid (F\vee G)(\omega)$ and $X_G \mid (F\vee G)(\omega)$ are independent for all ω in Ω.)*

(ii) $H(X_F, X_G) = H(X_F) + H(X_G) - H(X_{F\vee G})$. *(This means that $I(F,G) = H(F\vee G)$.)*

(iii) $\left(V_F \cap V_{F\vee G}^{\perp}\right) \perp \left(V_G \cap V_{F\vee G}^{\perp}\right)$.

(iv) $P_F P_G = P_G P_F = P_{F\vee G}$.

Information in three random variables

Let Γ_3 consist of all vectors of the form

$$(H(X), H(Y), H(X,Y), H(Z), H(X,Z), H(Y,Z), H(X,Y,Z))$$

as X, Y and Z vary over all triples of random variables over all uniform finite probability spaces.

What can we say about Γ_3?

What can we say about Γ_r?

It has been shown that it suffices to consider Ω as a finite group, with each relevant partition being the partition into cosets of some subgroup of Ω.

Do suprema matter?

Why do information theorists care about infima but not about suprema?

Suprema in statistics

Some toy data:

10.3	9.6		
9.9	10.2		
		6.1	5.8
		5.9	6.0

Statistician 1 says "I can see that there are differences between rows. If I fit the average value in each row and then subtract that from the data, then what is left is random noise, so I can see that columns have no effect."

Statistician 2 says "I can see that there are differences between columns. If I fit the average value in each column and then subtract that from the data, then what is left is random noise, so I can see that rows have no effect."

Statistician 3 says "The differences can be simply explained as the difference between the parts of Rows $\vee$ Columns."

Wilkinson's sweeping algorithm

The input is a vector y in $\mathbb{R}^\Omega$ and a sequence $F_1, \ldots, F_n$ of partitions of Ω ordered in such a way that if $F_i \succ F_j$ then $i < j$.

```
begin;
for i = 1 to n do
        begin z_i := P_{F_i} y;
                output z_i;
                y := y − z_i;
        end;
        output y;
end;
```

Does the order matter?

When does order matter?

Most people know that the output depends on the chosen ordering if there are any i and j such that F_i is not orthogonal to F_j.

It is less well known that the output depends on the chosen ordering if there are any i and j such that $F_i \vee F_j$ is not included. Tue Tjur pointed out the importance of the supremum in 1984 (he called it the 'minimum'); nevertheless, all statistical software includes infima but only Heiko Großmann's algorithm (in press for CSDA by late 2013) includes suprema.

Back to partitions

Let $\mathcal{F}$ be a collection of distinct partitions of Ω.
The relation $\preceq$ is a partial order on $\mathcal{F}$.

The *zeta function* of this partial order is the $\mathcal{F} \times \mathcal{F}$ matrix Z with entry $\zeta(F, G)$ equal to 1 if $F \preceq G$ and to 0 otherwise.
If the partitions are ordered in such a way that F precedes G if $F \preceq G$ then Z is upper triangular with all diagonal elements equal to 1.

Hence Z has an inverse matrix M, all of whose entries are integers. This is called the *Möbius function* of the partial order; its entries are written $\mu(F, G)$.

Entropy of a lattice?

The collection $\mathcal{F}$ is called a lattice if it is closed under $\vee$ and $\wedge$. What can we say about the entropy vector of a lattice of partitions?

Such a lattice in which all pairs of partitions are mutually orthogonal gives rise to a particularly nice family of random variables. What can we say about their entropy vector?

If, in addition, we insist that $\mathcal{F}$ include E and U and that all partitions be uniform, then we obtain an association scheme: coming up.

Orthogonal block structures

An orthogonal block structure is a finite set Ω together with a family $\mathcal{F}$ distinct partitions of Ω satisfying the following conditions.

(i) $U \in \mathcal{F}$.
(ii) $E \in \mathcal{F}$.
(iii) If $F \in \mathcal{F}$ then F is uniform, with n_F parts of size k_F.
(iv) If F and G are in $\mathcal{F}$ then $F \vee G \in \mathcal{F}$.
(v) If F and G are in $\mathcal{F}$ then $F \wedge G \in \mathcal{F}$.
(vi) If F and G are in $\mathcal{F}$ then $F \perp G$.

Relation matrices

Suppose that $(\Omega, \mathcal{F})$ is an orthogonal block structure.
For F in $\mathcal{F}$, define the $\Omega \times \Omega$ relation matrix R_F by

$$R_F(\alpha, \beta) = \begin{cases} 1 & \text{if } X_F(\alpha) = X_F(\beta) \\ 0 & \text{otherwise.} \end{cases}$$

Then condition (iii) (uniformity) implies that $R_F = k_F P_F$.

Adjacency matrices

Conditions (i) $U \in \mathcal{F}$ and (v) (infima) imply that, given any α and β in Ω, there is a finest partition F in $\mathcal{F}$ for which $X_F(\alpha) = X_F(\beta)$.
Define the $\Omega \times \Omega$ adjacency matrix A_F by

$$A_F(\alpha, \beta) = \begin{cases} 1 & \text{if } F \text{ is this finest partition} \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$R_F = \sum_{G \preceq F} A_G = \sum_{G \in \mathcal{F}} \zeta(G, F) A_G,$$

which can be inverted to give

$$A_F = \sum_{G \in \mathcal{F}} \mu(G, F) R_G.$$

Product of two adjacency matrices

Let $\mathcal{A}$ be the $\mathbb{R}$-linear span of $\{A_F : F \in \mathcal{F}\}$

$$= \text{ the } \mathbb{R}\text{-linear span of } \{R_F : F \in \mathcal{F}\}$$
$$= \text{ the } \mathbb{R}\text{-linear span of } \{P_F : F \in \mathcal{F}\}.$$

Conditions (vi) (orthogonality, so $P_F P_G = P_{F \vee G}$) and (iv) (suprema) show that $\mathcal{A}$ is closed under multiplication, and so forms an algebra.

Therefore, if F and G are in $\mathcal{F}$, there are real numbers $p(F, G; H)$ such that

$$A_F A_G = \sum_{H \in \mathcal{F}} p(F, G; H) A_H.$$

All the entries in $A_F A_G$ are non-negative integers.
Given α and β in Ω, there is a unique H in $\mathcal{F}$ with $A_H(\alpha, \beta) = 1$, while $A_{H'}(\alpha, \beta) = 0$ if $H' \neq H$. It follows that the coefficients $p(F, G; H)$ must be non-negative integers.

Association scheme

We have shown that $\{A_F : F \in \mathcal{F},\ A_F \neq 0\}$ satisfies the following conditions:

(0) All the entries of each A_F are 0 or 1, and they are not all 0.
(1) $A_E = I$.
(2) Each A_F is symmetric.
(3) $\sum A_F = R_U = J$.
(4) Each product $A_F A_G$ is a unique integer-linear combination of the A_H.

Any collection of matrices satisfying (0)–(4) is called an association scheme.

Bose–Mesner algebra

The algebra $\mathcal{A}$ is called the Bose–Mesner algebra of the association scheme. It is commutative, and all matrices are symmetric, so it has a basis of primitive idempotents.

For a general association scheme, there is no easy way to derive the primitive idempotents from the adjacency matrices. However, for an orthogonal block structure, the set of primitive idempotents is $\{Q_F : F \in \mathcal{F},\ Q_F \neq 0\}$, where

$$P_F = \sum_{G \in \mathcal{F}} \zeta(F, G) Q_G \quad \text{and} \quad Q_F = \sum_{G \in \mathcal{F}} \mu(F, G) P_G.$$

Back to entropy

Orthogonal block structures give nice families of partitions that may not come from groups.

Example

Suppose that there are $r-2$ mutually orthogonal Latin squares of order n. Let Ω consist of the n^2 cells in a square array.
Let F_1 and F_2 be the partitions of Ω into rows and columns respectively.
For $i = 3, \ldots, r-2$, let F_i be the partition of Ω according to the letters of square $i-2$.
Then, if $1 \leq i < j < r$, we have $F_i \vee F_j = U$ and $F_i \wedge F_j = E$.
Hence $H(F_i) = \log n$ for $1 \leq i \leq r$,
while $H(\bigwedge_{i \in I} F_i) = 2 \log n$ whenever $I \subseteq \{1, \ldots, r\}$ and $|I| \geq 2$.

A question

Is there anything special about the entropy vectors of orthogonal block structures?