

We will now look at a very specific example of symmetry, and see what the general results described above turn into.

We consider a *process*: a (finite or infinite) sequence of variables $X_1, X_2, X_3, \dots, X_n, \dots$ that assume values in some non-empty finite set \mathcal{X} . A subject calls these variables *exchangeable* if he decides that the inferences and decisions he makes about these variables will not depend on the order in which these variables are observed.

Our discussion will focus on the special case that \mathcal{X} has only two elements, say $\mathcal{X} = \{H, T\}$, but the extension to more general cases is straightforward. So envision we are flipping a coin repeatedly, and the subject assumes that the order in which the coin flips are observed is irrelevant for the inferences and decisions he will make about the coin flips.

How can we model this exchangeability assessment the subject makes?

1. EXCHANGEABILITY FOR FINITE SEQUENCES

Let us first look at a *finite* number n of coin flips.

The uncertain outcomes in this situation are now sequences of H and T of length n . The set of possible outcomes $x = (x_1, \dots, x_n)$ is

$$\mathcal{X}^n = \{(x_1, \dots, x_n) : x_k \in \{H, T\}\}$$

Running example. Let us look at the case $n = 3$. The possible outcomes x are:

$$HHT \quad HHT \quad HTH \quad HTT \quad THH \quad THT \quad TTH \quad TTT,$$

and all these outcomes make up the set of possible outcomes $\{H, T\}^3$. □

The fact that the subject assumes that the order of observations will not matter leads us to consider the following type of symmetry. Consider all permutations π of the index set $\{1, 2, \dots, n\}$, then we can use these permutations to turn a sequence of observations $x = (x_1, x_2, \dots, x_n)$ into a permuted sequences of observations:

$$\pi x := (x_{\pi(1)}, x_{\pi(2)}, \dots, x_{\pi(n)}).$$

This allows us to define a permutation group \mathcal{P}_n on \mathcal{X}^n .

Running example. The $3! = 6$ possible permutations of the index set $\{1, 2, 3\}$, and their actions on the sequence HTH are:

$$\begin{array}{ll} \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix} & HTH \rightarrow HTH \\ \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} & HTH \rightarrow HHT \\ \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} & HTH \rightarrow THH \\ \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} & HTH \rightarrow HHT \\ \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} & HTH \rightarrow THH \\ \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} & HTH \rightarrow HTH \end{array}$$

Similarly, with the permutation $\pi = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$ of the index set $\{1, 2, 3\}$ there corresponds the following permutation of the set of possible observations $\{H, T\}^3$:

$$\begin{aligned} HHT &\rightarrow \pi(HHT) = THH \\ HHT &\rightarrow \pi(HHT) = THH \\ HTH &\rightarrow \pi(HTH) = HTH \\ HTT &\rightarrow \pi(HTT) = TTH \\ THH &\rightarrow \pi(THH) = HHT \\ THT &\rightarrow \pi(THT) = THT \\ TTH &\rightarrow \pi(TTH) = HTT \\ TTT &\rightarrow \pi(TTT) = TTT, \end{aligned}$$

and we will also denote this permutation by π . \square

So the symmetry we are considering here, is the symmetry associated with the permutation group \mathcal{P}_n on the space \mathcal{X}^n . What we will now do, is apply the general results on symmetry representation to this particular case.

Running example. The invariant atoms are

$$\begin{aligned} [HHH] &= \{HHH\} & m_H = 3 \text{ and } m_T = 0 \\ [HHT] = [HTH] = [THH] &= \{HHT, HTH, THH\} & m_H = 2 \text{ and } m_T = 1 \\ [HTT] = [THT] = [TTH] &= \{HTT, THT, TTH\} & m_H = 1 \text{ and } m_T = 2 \\ [TTT] &= \{TTT\} & m_H = 0 \text{ and } m_T = 3 \end{aligned}$$

so there are four invariant atoms. \square

The invariant atoms are completely determined by their count vectors $m = (m_H, m_T)$. Indeed, consider the count map

$$C: \mathcal{X}^n \rightarrow \mathcal{N}_n$$

with

$$\begin{aligned} C(x)_H &= |\{k: x_k = H\}| = \text{number of heads in the sequence } x \\ C(x)_T &= |\{k: x_k = T\}| = \text{number of tails in the sequence } x \end{aligned}$$

and

$$\mathcal{N}_n := \{(m_H, m_T) \in \mathbb{N}_0^2: m_H + m_T = n\}.$$

All elements of an atom $[x]$ have the same count vector $m = C(x)$, and are completely determined by it:

$$[x] = \{y \in \mathcal{X}^n: C(y) = C(x)\}$$

so we also use the notation $[m]$. The number of elements in this atom is equal to the number of possible permutations of x , and therefore given by

$$|[m]| = \binom{n}{m_H \ m_T} = \binom{n}{m_H} = \frac{n!}{m_H!m_T!}.$$

Running example. The invariant atoms are

$$\begin{aligned}
 [(3, 0)] &= \{HHH\} && \text{has } \binom{3}{3} = 1 \text{ elements} \\
 [(2, 1)] &= \{HHT, HTH, THH\} && \text{has } \binom{3}{2} = 3 \text{ elements} \\
 [(1, 2)] &= \{HTT, THT, TTH\} && \text{has } \binom{3}{1} = 3 \text{ elements} \\
 [(0, 3)] &= \{TTT\} && \text{has } \binom{3}{0} = 1 \text{ elements}
 \end{aligned}$$

and the set of invariant atoms is in a one-to-one correspondence with the set of count vectors \mathcal{N}_3 for $n = 3$ observations, which has four elements, $(3, 0)$, $(2, 1)$, $(1, 2)$, and $(0, 3)$. \square

What are, in this case, the projection operator $\text{inv}_{\mathcal{P}_n}$ and the related uniform average expectation operator U ?

$$U(f|m) = \frac{1}{\binom{n}{m}} \sum_{y \in [m]} f(y) = \text{Hy}(f|m)$$

and this is the expectation operator associated with the *hypergeometric distribution*: independently taking balls, *without replacement*, from an urn whose composition is determined by the count vector m ; so there are m_H balls of type ‘heads’ and m_T balls of type ‘tails’.

We call a lower prevision \underline{P} on the sequences in \mathcal{X}^n *exchangeable* if it is strongly invariant with respect to the permutations in \mathcal{P}_n .

Our permutation invariance representation theorem now turns into Bruno de Finetti’s Representation Theorem for finite exchangeable sequences.

Theorem (De Finetti’s Finite Representation Theorem). *A coherent lower prevision \underline{P}_n on $\mathcal{L}(\mathcal{X}^n)$ is exchangeable if and only if there is some coherent count lower prevision \underline{Q}_n on $\mathcal{L}(\mathcal{N}_n)$ such that $\underline{P}_n = \underline{Q}_n \circ \text{Hy}$.*

$$\begin{array}{ccc}
 \mathcal{L}(\mathcal{X}^n) & \xrightarrow{\text{inv}_{\mathcal{P}_n}} & \mathcal{L}_{\mathcal{P}_n}(\mathcal{X}^n) \\
 \downarrow \underline{P}_n & \searrow \text{Hy} & \uparrow \\
 \mathbb{R} & \xleftarrow{\underline{Q}_n} & \mathcal{L}(\mathcal{N}_n)
 \end{array}$$

So inference—as represented by \underline{P} —about an finite exchangeable sequence X_1, X_2, \dots, X_n is equivalent to reasoning about drawing balls without replacement from an urn with an unknown composition $m \in \mathcal{N}_n$, where the uncertainty about the composition is represented by the lower prevision \underline{Q} , or equivalently, by a credal set \mathcal{M} of mass functions q on \mathcal{N}_n .

Running example. We now as ourselves, what can we say about the probability that in a sequence of three exchangeable coin flips, heads is followed by tails. So we are interested in the probability of the event

$$A = \{HTH, HTT, THT, HHT\}.$$

Since we know nothing else than that the sequence is exchangeable, the representing lower prevision is $\underline{Q} = \min$, and since

$$\begin{aligned}\text{Hy}(I_A|3,0) &= I_A(HHH) = 0 \\ \text{Hy}(I_A|2,1) &= \frac{1}{3} \left(I_A(HHT) + I_A(HTH) + I_A(THH) \right) = \frac{2}{3} \\ \text{Hy}(I_A|1,2) &= \frac{1}{3} \left(I_A(TTH) + I_A(THT) + I_A(HTT) \right) = \frac{2}{3} \\ \text{Hy}(I_A|0,3) &= I_A(TTT) = 0,\end{aligned}$$

it follows from the representation theorem that

$$\underline{P}(A) = \min \text{Hy}(I_A) = \min \left\{ 0, \frac{2}{3} \right\} = 0 \text{ and } \bar{P}(A) = \max \text{Hy}(I_A) = \min \left\{ 0, \frac{2}{3} \right\} = \frac{2}{3}.$$

This tells us that, even if we are completely ignorant about the composition of the—fictitious—urn, exchangeability alone makes sure that our inferences are not entirely vacuous. \square

2. EXCHANGEABILITY FOR INFINITE SEQUENCES

Let us now consider an *infinite* sequence $X_1, X_2, \dots, X_n \dots$ which our subject assumes to be exchangeable, which means that all its finite subsequences are assumed to be exchangeable.

So, in effect, this means that all the finite sequences

$$X_1, X_2, \dots, X_n \text{ for all } n \in \mathbb{N}$$

are assumed to be exchangeable.

So for any $n \in \mathbb{N}$, we can invoke the finitary representation theorem, which results in a representing lower prevision \underline{Q}_n on the set \mathcal{N}_n of all count vectors for a total of n observations, such that $\underline{P}_n = \underline{Q}_n \circ \text{Hy}$.

The problem here is that the representation space \mathcal{N}_n —the possible urn compositions—changes with the number of observations—the number of balls in the urn— n . If we want a single representation that is valid for all n , we will need to find a way to get rid of this n -dependency of the representation space. It turns out that there is a very elegant way to do this, by looking at *binomial distributions*.

Binomial expectations and Bernstein polynomials: Consider flipping the same coin independently n times, where the probability of heads on each coin flip is $\theta \in [0, 1]$.

Then the probability observing a sequence with count vector $m \in \mathcal{N}_n$ is

$$B_m(\theta) = \binom{n}{m_H} \theta^{m_H} (1 - \theta)^{m_T}.$$

Viewed as a function of θ , this B_m is a polynomial on $[0, 1]$ of degree n , called a *Bernstein basis polynomial* of degree n .

Similarly, if we consider a gamble $g \in \mathcal{L}(\mathcal{N}_n)$ on the counts m , then its *binomial expectation* is

$$\text{Mn}(g|\theta) = \sum_{m \in \mathcal{N}_n} g(m) B_m(\theta)$$

which, as a function of θ , is a polynomial of degree at most n on $[0, 1]$.

Running example. The probability of observing m_H heads and m_T tails in a run of $n = 3$ coin flips, is

$$B_{(3,0)}(\theta) = \binom{3}{3} \theta^3 (1 - \theta)^0 = \theta^3$$

$$B_{(2,1)}(\theta) = \binom{3}{2} \theta^2 (1 - \theta)^1 = 3\theta^2(1 - \theta) = 3\theta^2 - 3\theta^3$$

$$B_{(1,2)}(\theta) = \binom{3}{1} \theta^1 (1 - \theta)^2 = 3\theta(1 - 2\theta + \theta^2) = 3\theta - 6\theta^2 + 3\theta^3$$

$$B_{(0,3)}(\theta) = \binom{3}{0} \theta^0 (1 - \theta)^3 = (1 - \theta)^3 = 1 - 3\theta + 3\theta^2 - \theta^3,$$

which are all Bernstein basis polynomials of degree 3. Any polynomial of degree 3

$$p(\theta) = a + b\theta + c\theta^2 + d\theta^3$$

can be written uniquely as a linear combination of these basis polynomials:

$$p(\theta) = (a + b + c + d)B_{(3,0)}(\theta) + \left(a + \frac{2b}{3} + \frac{c}{3}\right)B_{(2,1)}(\theta) + \left(a + \frac{b}{3}\right)B_{(1,2)}(\theta) + aB_{(0,3)}(\theta).$$

Also for general n , the Bernstein basis polynomials B_m , $m \in \mathcal{N}_n$ of degree n constitute a basis for the linear space $\mathcal{V}_n([0, 1])$ of all polynomials of degree at most n .

So what we see, is that the multinomial expectation operator Mn turns any gamble g on the counts in \mathcal{N}_n into a polynomial $\text{Mn}(g)$ on $[0, 1]$, where

$$\text{Mn}(g|\theta) = \sum_{m \in \mathcal{N}_n} g(m)B_m(\theta).$$

What is more, because the Bernstein basis polynomials of degree n constitute a basis for all polynomials of degree at most n , the map Mn is *one-to-one*—a linear *isomorphism*, which preserves dimension.

$$\begin{array}{ccc} \mathcal{L}(\mathcal{X}^n) & \xrightarrow{\text{Hy}} & \mathcal{L}(\mathcal{N}_n) \\ \downarrow \underline{P}_n & \searrow \underline{Q}_n & \downarrow \text{Mn} \\ \mathbb{R} & \xleftarrow{\underline{R}_n} & \mathcal{V}_n([0, 1]) \end{array}$$

and therefore

$$\underline{Q}_n = \underline{R}_n \circ \text{Mn} \quad \text{and} \quad \underline{P}_n = \underline{Q}_n \circ \text{Hy} = \underline{R}_n \circ (\text{Mn} \circ \text{Hy})$$

so we can represent \underline{P}_n by a lower prevision \underline{R}_n on polynomials of degree up to n —the number of observations corresponds to the degree of the polynomials.

But we can easily make the representation independent of the number of observations n by looking at the set $\mathcal{V}([0, 1])$ of all polynomials (of any degree).

Theorem (De Finetti's Infinite Representation Theorem). *A coherent lower prevision \underline{P} on $\bigcup_{n \in \mathbb{N}} \mathcal{L}(\mathcal{X}^n)$ is exchangeable if and only if there is some coherent frequency lower prevision \underline{R} on $\mathcal{V}([0, 1])$ such that $\underline{P} = \underline{R} \circ \text{Hy} \circ \text{Mn}$.*

