EXCHANGEABILITY IN IP MODELS

GERT DE COOMAN AND ERIK QUAEGHEBEUR

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

More details about the material in these notes can be found in: [10, 9, 1] for general imprecise probability models, and [4, 2, 3, 5, 7] for exchangeability.

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 $\mathscr{X}$. A subject calls these variables *exchangeable* if he decides that the inferences and decisions he makes about these variables will not dependent on the order in which these variables are observed.

Our discussion will focus on the special case that $\mathscr{X}$ has only two elements, say $\mathscr{X} = \{H, T\}$, 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

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

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

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

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

Identifying the symmetry. 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 sequence of observations:

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

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

Date: 24 July 2014.

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

$$\begin{pmatrix}1 & 2 & 3\\ 1 & 2 & 3\end{pmatrix} \qquad HTH \to HTH$$
$$\begin{pmatrix}1 & 2 & 3\\ 3 & 1 & 2\end{pmatrix} \qquad HTH \to HHT$$
$$\begin{pmatrix}1 & 2 & 3\\ 2 & 3 & 1\end{pmatrix} \qquad HTH \to THH$$
$$\begin{pmatrix}1 & 2 & 3\\ 1 & 3 & 2\end{pmatrix} \qquad HTH \to HHT$$
$$\begin{pmatrix}1 & 2 & 3\\ 2 & 1 & 3\end{pmatrix} \qquad HTH \to THH$$
$$\begin{pmatrix}1 & 2 & 3\\ 3 & 2 & 1\end{pmatrix} \qquad HTH \to HTH$$

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 $\{H,T\}^3$:

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

and we will also denote this permutation by π. □

So the symmetry we are considering here, is the symmetry associated with the permutation group $\mathscr{P}_n$ on the space $\mathscr{X}^n$. 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. □

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

$$C\colon \mathscr{X}^n \to \mathscr{N}_n$$

with

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

and

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

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

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

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

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

Running example. The invariant atoms are

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

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

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

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

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 in total n balls, m_H of which are of type ‘heads’ and m_T of which of type ‘tails’.

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

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 $\mathscr{L}(\mathscr{X}^n)$ is exchangeable if and only if there is some coherent* count *lower prevision $\underline{Q}_n$ on $\mathscr{L}(\mathscr{N}_n)$ such that $\underline{P}_n = \underline{Q}_n \circ \mathrm{Hy}$.*

$$\begin{tikzcd}
\mathscr{L}(\mathscr{X}^n) \arrow[r, "\mathrm{inv}_{\mathscr{P}_n}"] \arrow[d, "\underline{P}_n"'] \arrow[dr, "\mathrm{Hy}"] & \mathscr{L}_{\mathscr{P}_n}(\mathscr{X}^n) \arrow[d, leftrightarrow] \\
\mathbb{R} & \mathscr{L}(\mathscr{N}_n) \arrow[l, "\underline{Q}_n"]
\end{tikzcd}$$

So inference—as represented by the coherent lower prevision $\underline{P}_n$—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 $m \in \mathscr{N}_n$, where the uncertainty about the composition is represented by the coherent lower prevision $\underline{Q}_n$, or equivalently, by a credal set $\mathscr{M}_n$ of mass functions q on $\mathscr{N}_n$.

Running example. We now ask ourselves what we can 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 $\underline{Q}_3 = \min$, and since

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

it follows from the representation theorem that

$$\underline{P}_3(A) = \min \mathrm{Hy}(I_A) = \min\Big\{0, \frac{2}{3}\Big\} = 0 \text{ and } \overline{P}_3(A) = \max \mathrm{Hy}(I_A) = \max\Big\{0, \frac{2}{3}\Big\} = \frac{2}{3}.$$

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. □

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 $\mathscr{N}_n$ of all count vectors for a total of n observations, such that $\underline{P}_n = \underline{Q}_n \circ \mathrm{Hy}$.

The problem here is that the representation space $\mathscr{N}_n$—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 of observing a sequence with count vector $m \in \mathscr{N}_n$ 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 $g \in \mathscr{L}(\mathscr{N}_n)$ on the counts m, then its *binomial expectation* is

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

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

$$\begin{aligned}
B_{(3,0)}(\theta) &= \binom{3}{3}\theta^3(1-\theta)^0 = \theta^3 \\
B_{(2,1)}(\theta) &= \binom{3}{2}\theta^2(1-\theta)^1 = 3\theta^2(1-\theta) = 3\theta^2 - 3\theta^3 \\
B_{(1,2)}(\theta) &= \binom{3}{1}\theta^1(1-\theta)^2 = 3\theta(1-2\theta+\theta^2) = 3\theta - 6\theta^2 + 3\theta^3 \\
B_{(0,3)}(\theta) &= \binom{3}{0}\theta^0(1-\theta)^3 = (1-\theta)^3 = 1 - 3\theta + 3\theta^2 - \theta^3,
\end{aligned}$$

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:

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

Also for general n, the Bernstein basis polynomials B_m, $m \in \mathscr{N}_n$ of degree n constitute a basis for the linear space $\mathscr{V}_n([0,1])$ 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 $\mathscr{N}_n$ into a polynomial $\mathrm{Mn}(g)$ on $[0,1]$, where

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

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.

$$\begin{tikzcd}
\mathscr{L}(\mathscr{X}^n) \arrow[r, "\mathrm{Hy}"] \arrow[d, "\underline{P}_n"'] & \mathscr{L}(\mathscr{N}_n) \arrow[dl, "\underline{Q}_n"'] \arrow[d, leftrightarrow, "\mathrm{Mn}"] \\
\mathbb{R} & \mathscr{V}_n([0,1]) \arrow[l, "\underline{R}_n"]
\end{tikzcd}$$

and therefore

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

so we can equally well 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 $\mathscr{V}([0,1])$ of all polynomials (of any degree).

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

$$\begin{tikzcd}
\mathscr{L}(\mathscr{X}^n) \arrow[r, "\mathrm{Hy}"] \arrow[d, "\underline{P}_n"'] & \mathscr{L}(\mathscr{N}_n) \arrow[dl, "\underline{Q}_n"'] \arrow[d, leftrightarrow, "\mathrm{Mn}"] & \\
\mathbb{R} & \mathscr{V}_n([0,1]) \arrow[l, "\underline{R}_n"'] \arrow[r, hook, "\mathrm{id}"] & \mathscr{V}([0,1]) \arrow[ll, bend left, "\underline{R}"]
\end{tikzcd}$$

Running example. Suppose we want to find out about the probability of the event A—'heads followed by tails in a run of three coin flips'—using the polynomial representation. Recall that, with $g = \mathrm{Hy}(I_A)$,

$$g(3,0) = 0 \text{ and } g(2,1) = \frac{2}{3} \text{ and } g(1,2) = \frac{2}{3} \text{ and } g(0,3) = 0$$

and therefore the count gamble g corresponds to the polynomial

$$\mathrm{Mn}(g|\theta) = \frac{2}{3}B_{(2,1)}(\theta) + \frac{2}{3}B_{(1,2)}(\theta) = \frac{2}{3}[3\theta^2(1-\theta) + 3\theta(1-\theta)^2] = 2\theta(1-\theta),$$

a polynomial of degree 2—this because the event A essentially involves only two observations.

Because *infinite* exchangeability is all we have assumed, we have a vacuous representing model, meaning that $\underline{R} = \inf$, and therefore

$$\underline{P}(A) = \inf_{\theta\in[0,1]} 2\theta(1-\theta) = 0 \text{ and } \overline{P}(A) = \sup_{\theta\in[0,1]} 2\theta(1-\theta) = \frac{1}{2},$$

and the upper probability is now different from the value $\frac{2}{3}$ obtained using the count representation. The reason for the difference is the following: with the count representation, we only assume the exchangeability of the first three coin flips, which allows for a representation with an urn of three balls and unknown composition; with the frequency representation, we assume the exchangeability of an infinity of coin flips, which requires a representation with an infinite number of balls and unknown composition. $\square$

REFERENCES

[1] Thomas Augustin, Frank P. A. Coolen, Gert de Cooman, and Matthias C. M. Troffaes, editors. *Introduction to Imprecise Probabilities*. John Wiley & Sons, 2014.

[2] Gert de Cooman and Erik Quaeghebeur. Exchangeability for sets of desirable gambles. In Thomas Augustin, Frank P. A. Coolen, Serafín Moral, and Matthias C. M. Troffaes, editors, *ISIPTA '09 – Proceedings of the Sixth International Symposium on Imprecise Probability: Theories and Applications*, pages 159–168, Durham, United Kingdom, 2009. SIPTA, SIPTA.

[3] Gert de Cooman and Erik Quaeghebeur. Exchangeability and sets of desirable gambles. *International Journal of Approximate Reasoning*, 53(3):363–395, 2012. Special issue in honour of Henry E. Kyburg, Jr.

[4] Gert de Cooman, Erik Quaeghebeur, and Enrique Miranda. Exchangeable lower previsions. *Bernoulli*, 15(3):721–735, 2009.

[5] Bruno de Finetti. La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, 7:1–68, 1937. English translation in [8].

[6] Bruno de Finetti. *Teoria delle Probabilità*. Einaudi, Turin, 1970.

[7] Bruno de Finetti. *Theory of Probability: A Critical Introductory Treatment*. John Wiley & Sons, Chichester, 1974–1975. English translation of [6], two volumes.

[8] Henry E. Kyburg Jr. and Henry E. Smokler, editors. *Studies in Subjective Probability*. Wiley, New York, 1964. Second edition (with new material) 1980.

[9] Matthias C. M. Troffaes and Gert de Cooman. *Lower Previsions*. Wiley, 2014.

[10] Peter Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.