Probabilistic Preferences

by Sven Nilsen, 2019

Due to hidden information states, it is not possible to predict preferences exactly as used in natural language. For example, a human might prefer A over B sometimes, but B over A other times.

The problem is, if preferences are probabilistic, how do we check for consistency? Is it possible to tell whether preferences are consistent even when there are hidden information states?

The answer to this is that there exists various degrees of consistency.

Assume that the preferences are given by a matrix `M`, like this:

	A	**B**	**C**
A	$P(A < A)$	$P(A < B)$	$P(A < C)$
B	$P(B < A)$	$P(B < B)$	$P(B < C)$
C	$P(C < A)$	$P(C < B)$	$P(C < C)$

The following law is required for a weak form of consistency:

$$P(A < B) = 1 - P(B < A)$$

At the diagonal, since objects are compared to themselves, the probability is `0.5`:

$$P(A < A) = 1 - P(A < A)$$
$$p = 1 - p$$
$$2p = 1$$
$$p = 0.5$$

If there are no hidden information states and preferences are consistent, then the matrix `M` can be rearranged into a upper or lower triangular matrix with ones in one part and zeroes in the other part (the diagonal is filled with `0.5`). A sorted list of unique objects in ascending order is upper triangular, while a sorted list of unique objects in descending order is lower triangular.

Preferences are probabilistic when there exists a number that is not on the diagonal which value is neither `0` or `1`. What the matrix `M` means is interpreted using the following iterative algorithm which loops infinitely and produces a distribution over permutations of some list `x` of indices:

```
i := floor(random() * len(x))
j := floor(random() * len(x))
if random() < M[x[i]][x[j]] { if i > j { swap(mut x, i, j) } }
else { if i < j { swap(mut x, i, j) } }
```

This distribution often converges for all `x`, making it possible to study stronger forms of consistency.