Partial Diversity

by Sven Nilsen, 2019

In this paper I define a binary operation for diversity of partial discrete probability distributions.

A partial diversity of two discrete probability distributions[1] is the following binary operator:

> partial_diversity(a : partial_distribution, b : partial_distribution) =
> $\quad$ 1 – sqrt(1/2 ∑ i [0, n) { $(a_i - b_i)^2$ })
>
> partial_distribution(x : [prob; n]) = ∑ i [0, n) { x_i } <= 1

A partial probability distribution is when the sum is less or equal to one.

The name "diversity" comes from the motivation of talking about solutions of optimizations problems that have equal success rate, but where one solution might be preferred because of greater diversity in behavior. The diversity of a solution can be controlled by a second partial probability distribution.

When the partial diversity is `1`, the two partial probability distributions are identical.

When the partial diversity is `0`, the two partial probability distributions have both concentrated probability mass `1` in some unique state, but the two states are distinct.

For example:

> a := [1, 0, 0]
> b := [0, 1, 0]
>
> partial_diversity(a, b) == 0

More generally:

> a := [1, 0, 0, 0] $\qquad$ `a` has concentrated probability mass `1`
> b := [p, 1-p, 0, 0] $\qquad$ `b` has a binary probability mass (e.g. a biased coin flip)
>
> partial_diversity(a, b) == p

With other words, partial diversity has semantics that can be similar to probability.

The dual of partial diversity is partial orthogonality:

> partial_orthogonality <=> 1 – partial_diversity

When the partial orthogonality is `0`, the two partial distributions are identical.
When the partial orthogonality is `1`, the two partial distributions are concentrated and orthogonal.

The type of partial diversity is the following:

partial_diversity : partial_distribution^2 → prob

More precisely, the existential path[2] is logically equivalent to the unit interval of real numbers:

∃partial_diversity <=> prob

This works because a number between `0` and `1`, multiplied with itself, is equal or smaller:

$x^2 <= x$

x : prob

If this number is split into two parts, then the sum of the squared parts is smaller than the square:

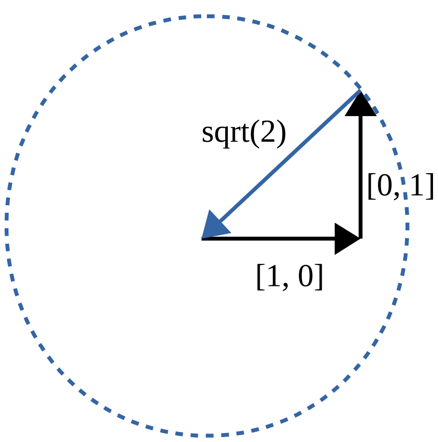
$x_0 + x_1 = x \quad => \quad x_0^2 + x_1^2 <= x^2$

x : prob
x_0 : prob
x_1 : prob

The difference between two numbers between `0` and `1` is itself between `0` and `1`.
The largest difference possible between two such numbers is `1`, so their largest square is `1`.
This means that one number must be `1` and the other number must be `0`.

The sum of squares of difference of two partial distributions can never be larger than `2`.
By dividing the sum of squares of differences by `2`, the value is normalized.

The square root of the sum of square differences is used to change the semantics into a vector:



This vector is normalized by its maximum length, which is `sqrt(2)`.
Since `sqrt(…) / sqrt(2) = sqrt(… / 2)`, the `2` is sum of squares of differences is divided by `2`.

The length of the vector is partial orthogonality.
One minus this length of the vector gives the partial diversity.

References:

[1] “Probability mass function”
Wikipedia
https://en.wikipedia.org/wiki/Probability_mass_function

[2] “Existential Paths”
Sven Nilsen, 2017
https://github.com/advancedresearch/path_semantics/blob/master/papers-wip/existential-paths.pdf