

# Normalized Interplay Between PTMs

## Introduction:

Here we provide both a mathematical proof and theoretical justification for using the normalized interplay measure between two PTMs: PTM1 and PTM2. The normalization factor:  $-\log(p(PTM1PTM2))$  is explored in its application in PTM Crosstalk Scoring.

## Definitions:

Let:

- $p(PTM1)$  be the probability of occurrence of PTM1,
- $p(PTM2)$  be the probability of occurrence of PTM2,
- $p(PTM1PTM2)$  be the joint probability of both terms occurring together.

## Interplay:

The Interplay between the terms PTM1 and PTM2 is defined as:

$$\text{Interplay}(PTM1PTM2) = \ln \left( \frac{p(PTM1PTM2)}{p(PTM1)p(PTM2)} \right)$$

## Normalized Interplay:

Normalization involves the entropy of the joint event, computed as:

$$h(PTM1PTM2) = -\ln(p(PTM1PTM2))$$

The Normalized Interplay between PTM1 and PTM2 can then be expressed as:

$$\text{Normalized Interplay}(PTM1PTM2) = \frac{\text{Interplay}(PTM1PTM2)}{h(PTM1PTM2)}$$

## Mathematical Derivation:

### Properties of Logarithms and Probabilities

Since  $p(PTM1PTM2) \leq p(PTM1)$  and  $p(PTM2)$ , it follows:

$$\ln \left( \frac{p(PTM1PTM2)}{p(PTM1)p(PTM2)} \right) \leq 0$$

This inequality reflects the property that logarithms of values between 0 and 1 yield non-positive results.

### Bounding the Interplay:

The nature of logarithmic functions and the probabilities involved give:

$$\text{Interplay}(PTM1PTM2) \leq -\log(p(PTM1PTM2))$$

Furthermore, the theoretical minimum value of Interplay is achieved when PTM1 and PTM2 are present at abundance equal to that, that would be expected by random chance, thus reducing the Interplay to zero.

### Normalizing the Interplay:

Using the defined bounds and the entropy  $h(PTM1PTM2)$ :

$$-1 \leq \text{Normalized Interplay}(PTM1PTM2) \leq 1$$

This normalization ensures that the value of Normalized Interplay is confined within the interval  $[-1, 1]$ , scaling the strength of the association between PTM1 and PTM2 relative to the uncertainty of their joint occurrence.

### Justification for Normalization:

The choice of  $-\log(p(PTM1PTM2))$  as a normalization factor while a natural option, is grounded in information theory, where it represents the maximum possible surprisal or information content of observing both PTM1 and PTM2 together. This factor effectively scales the raw Interplay between two PTMs to a comparative measure across different contexts, providing a normalized value that reflects the proportion of the maximum possible surprisal that is realized in the observed joint occurrence of the PTM1 and PTM2. Thus, it allows for a consistent and interpretable comparison of the strength of association between different PTMs or terms, irrespective of their individual probability distributions.

## **Conclusion:**

This proof and exploration of theoretical limits demonstrate that Normalized Interplay is a theoretically grounded and effective measure for evaluating the degree of association between PTMs, scaled relative to the entropy of their joint probability.