

Krippendorff's Alpha

NLP-Taskforce

28.05.2025

1. Coincidence Matrix

| | sample_1 | sample_2 | sample_3 | sample_4 | sample_5 | sample_6 |
|-------------|----------|----------|----------|----------|----------|----------|
| Annotator 1 | 1 | 1 | 4 | 3 | 5 | * |
| Annotator 2 | 2 | 2 | 1 | 3 | 5 | 3 |



remove unpaired labels, create coincidence matrix:

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | | 2 | | 1 | |
| 2 | 2 | | | | |
| 3 | | | 1 | | |
| 4 | 1 | | | | |
| 5 | | | | | 1 |

2. Difference Functions

- *nominal*: agreement = 0, disagreement = 1
- *interval*: $(v_1 - v_2)^2$
- *ordinal*: $\left(\sum_{g=v_1}^{g=v_2} n_g - \left(\frac{nv_1 + nv_2}{2} \right) \right)^2$

example: weighing of coincidence matrix according to nominal and interval function

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 0 | 1 | 1 | 1 | 1 |
| 2 | 1 | 0 | 1 | 1 | 1 |
| 3 | 1 | 1 | 0 | 1 | 1 |
| 4 | 1 | 1 | 1 | 0 | 1 |
| 5 | 1 | 1 | 1 | 1 | 0 |

| | 1 | 2 | 3 | 4 | 5 |
|---|----|---|---|---|----|
| 1 | 0 | 1 | 4 | 9 | 16 |
| 2 | 1 | 0 | 1 | 4 | 9 |
| 3 | 4 | 1 | 0 | 1 | 4 |
| 4 | 9 | 4 | 1 | 0 | 1 |
| 5 | 16 | 9 | 4 | 1 | 0 |

3. Complete Formula

$$\alpha = 1 - \frac{D_o}{D_e}$$

Coincidence Matrix (see slide 1) with margins

| | 1 | 2 | 3 | 4 | 5 | Σ fr |
|-------------|---|---|---|---|---|-------------|
| 1 | | | 2 | | 1 | 3 |
| 2 | 2 | | | | | 2 |
| 3 | | | 1 | | | 2 |
| 4 | 1 | | | | | 1 |
| 5 | | | | | 1 | 2 |
| Σ fr | 3 | 2 | 2 | 1 | 2 | n=10 |

Weight for each value pair

| | 1 | 2 | 3 | 4 | 5 |
|---|----|---|---|---|----|
| 1 | 0 | 1 | 4 | 9 | 16 |
| 2 | 1 | 0 | 1 | 4 | 9 |
| 3 | 4 | 1 | 0 | 1 | 4 |
| 4 | 9 | 4 | 1 | 0 | 1 |
| 5 | 16 | 9 | 4 | 1 | 0 |

- D_o : sum of all observed disagreements in one triangle (weighed by difference function), example: $2*1 + 1*9 = 11$
- D_e : sum of all weighed expected disagreements (normalized):

$$\left(\frac{1}{n-1}\right) \sum_{v1=1, v2=1}^v n_{v1} n_{v2} \partial$$

example:

do = 11

$$de = \frac{1}{9} ((3 * 2 * 1) + (3 * 2 * 4) + (3 * 1 * 9) + (3 * 2 * 16) + (2 * 2 * 1) + (2 * 1 * 4) + (2 * 2 * 9) + (2 * 1 * 1) + (2 * 2 * 4) + (1 * 2 * 1)) =$$

$$\frac{1}{9} (6 + 24 + 27 + 96 + 4 + 8 + 36 + 2 + 16 + 2) = \frac{221}{9} = 24.56$$

$$\alpha = 1 - \frac{11}{24.56} = 1 - 0.45 = 0.55$$

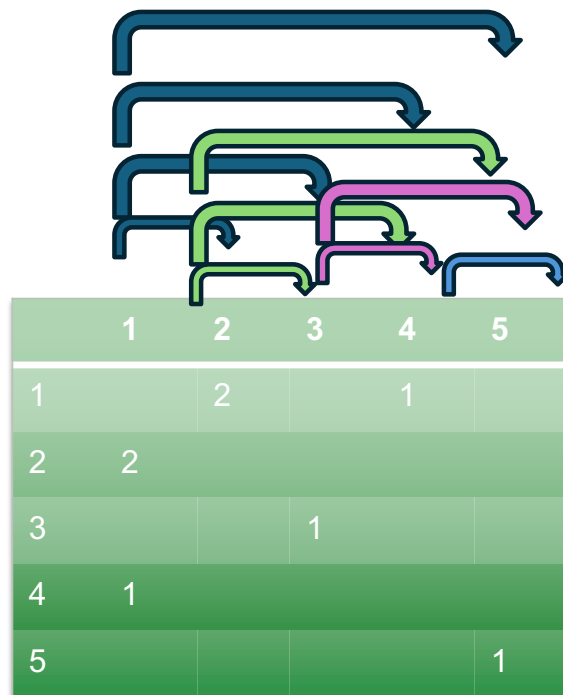
4. Focus: Expected Disagreement

| | 1 | 2 | 3 | 4 | 5 | Σ fr |
|------|---|---|---|---|---|------|
| 1 | | | 2 | 1 | | 3 |
| 2 | 2 | | | | | 2 |
| 3 | | | 1 | | | 2 |
| 4 | 1 | | | | | 1 |
| 5 | | | | | 1 | 2 |
| Σ fr | 3 | 2 | 2 | 1 | 2 | n=10 |

$$D_e = \left(\frac{1}{n-1} \right) \sum_{v1=1, v2=1}^v n_{v1} n_{v2} \partial$$

- n: number of annotations
- nv1, nv2: total number of label 1, total number of label 2
- ∂ : difference function

4. Focus: Expected Disagreement



$$De = \left(\frac{1}{n-1} \right) \sum_{v1=1, v2=1}^V n_{v1} n_{v2} \partial$$

1. Iterate through all possible label pairs:
1,2 / 1, 3 / 1,4 / ... / 4,5
2. For e.g. 1,2: calculate number of ways the pair 1,2 can be made $3*2$
3. Weigh this product by the difference function (e.g. $(1 - 2)^2 = 1^2$): $3*2*1$
4. Sum up all of these products for all label pairs
5. Normalize: divide sum by number of annotation samples minus one



Takeaway: more categories – more products/higher weights – higher expected disagreement;
intuition: less labels, higher chance of selecting the same label by chance

5. Overview

$$\alpha = 1 - \frac{D_o}{D_e}$$

- 1: perfect agreement
- 0: agreement no better than chance
- -1: systematic disagreement