What is Krippendorff's Alpha?

David Honour

June 28, 2016

Krippendorff's alpha is non-parametric a reliability measure applicable to coding systems that yield one of a finite discreet set of valid responses.

It is often used to investigate the inter rater reliability of scoring systems. In such an investigation you first need to get a number of samples of the thing you want to score. Then you need a number of observers to look at each of these samples so you can compare their agreement. The results of observers for a given sample is called a unit.

In my research I found articles extolling the virtues of Krippendorff's alpha[1], and those explaining how to compute it[2]. However I found something of a deficit of explanations of what it actually is.

What is it?

Krippendorff's alpha looks at the pairs of responses given, and can be seen to be comparing the disagreement of the pairs observed within each unit to the disagreement expected by chance based on the responses given to all units.

Define R as the set of responses that an observation may yield. The responses in a given unit form a multiset u. Units with a single response cannot be compared for reliability, as there is nothing to compare to, and should be discarded. All the responses form a multiset whose items are these unit multisets, we define this multiset of all responses as U.

A way of quantifying the disagreement between ratings is required, and in order to support different classes of data (e.g. ordinal, nominal or interval) we require a function describing this which we denote $\delta(c, k)$. This represents the distance between responses in some sense:

$$\delta(c, c) = 0 \tag{1}$$

$$\delta(c, k) \geq 0 \tag{2}$$

$$\delta(c, k) = \delta(k, c) \tag{3}$$

Now that the problem is expressed in a mathematical form we can define the average disagreement within a multiset:

$$D(m) = \sum_{c \in R} \sum_{k \in R} \delta(c, k) \frac{W(m, c, k)}{P(|m|, 2)} \tag{4}$$

where $|m|$ is the cardinality (number of elements) of m, P is the number of permutations and the number of ways to make a pair containing c and k is:

$$W(m,c,k) = \begin{cases} c \neq k & \nu(m,c)\nu(m,k) \\ c = k & \nu(m,c)(\nu(m,c)-1) \end{cases} \tag{5}$$

where $\nu(m,c)$ is the multiplicity (number of occurrences) of c in m.

The average disagreement can be applied to each unit in turn and averaged across all units (weighted by the number of responses) to yield a quantity known as the observed disagreement:

$$D_o = \sum_{u \in U} \frac{|u|}{|V|} D(u) \tag{6}$$

where V is a combined multiset containing the responses of all units:

$$V = \sum_{u \in U} u \tag{7}$$

the sum is performed with $\uplus$ (the multiset sum).

The average disagreement of all possible response pairs is known as the expected disagreement and is given by:

$$D_e = D(V) \tag{8}$$

Krippendorff's alpha is defined as:

$$\alpha = 1 - \frac{D_o}{D_e} \tag{9}$$

An example

3 observers looked at 3 pictures of dresses, and asked if it was blue. Not all observers were asked about all pictures. In this example the set of possible responses $R = \{\mathrm{y}, \mathrm{n}\}$. We can choose a function for δ that tells us that they're different:

$$\delta(c,k) = \begin{cases} c = k & 0 \\ c \neq k & 1 \end{cases} \tag{10}$$

Collect the results of the observations to yield a multiset of unit multisets:

$$\{\{\mathrm{y}, \mathrm{n}, \mathrm{n}\}, \{\mathrm{y}, \mathrm{n}\}, \{\mathrm{n}\}\} \tag{11}$$

Removing any units with a single response (as they yield no pairs and thus do not contribute to the comparison):

$$U = \{\{\mathrm{y}, \mathrm{n}, \mathrm{n}\}, \{\mathrm{y}, \mathrm{n}\}\} \tag{12}$$

From this we can determine:

$$V = \{\mathrm{y}, \mathrm{n}, \mathrm{n}, \mathrm{y}, \mathrm{n}\} \tag{13}$$

Now that we have these we can calculate the expected disagreement:

$$D_e = D(\{\mathrm{y}, \mathrm{n}, \mathrm{n}, \mathrm{y}, \mathrm{n}\}) \tag{14}$$

the summations run over all possible pairs of values of R, which in this case are (y, y), (y, n), (n, y) and (n, n).

Consider the term for the pair (y, y):

$$\delta(\mathrm{y}, \mathrm{y})\frac{\mathrm{W}(\mathrm{V}, \mathrm{y}, \mathrm{y})}{\mathrm{P}(|\mathrm{V}|, 2)}) = 0 \tag{15}$$

since $\delta(\mathrm{y}, \mathrm{y}) = 0$. It can in fact be seen from equation 1 that the contributions for pairs where $c = k$ are always 0 for any valid choice of δ. Thus we know that the contribution from (n, n) is also 0.

Consider the term corresponding to the pair (y, n):

$$\begin{aligned}\delta(\mathrm{y}, \mathrm{n})\frac{\mathrm{W}(\mathrm{V}, \mathrm{y}, \mathrm{n})}{\mathrm{P}(|\mathrm{V}|, 2)}) &= 1\frac{\nu(V, \mathrm{y})\nu(\mathrm{V}, \mathrm{n})}{P(5, 2)} \\ &= 1\frac{2 \times 3}{20} \\ &= \frac{3}{10}\end{aligned} \tag{16}$$

note that we did not use the order of the pair and thus the contribution due to (n, y) is also $\frac{3}{10}$.

Using these results we determine that in this example:

$$D_e = 0 + \frac{3}{10} + \frac{3}{10} + 0 = \frac{3}{5} = 0.6 \tag{17}$$

Calculating the observed disagreement:

$$D_o = \left(\frac{3}{5}D(\{\mathrm{y}, \mathrm{n}, \mathrm{n}\}) + \frac{2}{5}\mathrm{D}(\{\mathrm{y}, \mathrm{n}\})\right) \tag{18}$$

$$= \left(\frac{3}{5}\left(2\frac{1 \times 2}{6}\right) + \frac{2}{5}\left(2\frac{1 \times 1}{2}\right)\right) \tag{19}$$

$$= \frac{4}{5} = 0.8 \tag{20}$$

Using these results we can calculate:

$$\alpha = 1 - \frac{0.8}{0.6} = -\frac{1}{3} \tag{21}$$

this tells us that the agreement is worse than would be expected by chance.

Equivalence to Krippendorff's form

The forms given by Krippendorff[2] do not match those I have presented here, however they can be seen to be equivalent.

Let's begin with:

$$D(m) = \sum_{c \in R} \sum_{k \in R} \delta(c,k) \frac{W(m,c,k)}{P(|m|,2)} \tag{22}$$

Taking the formula for the number of permutations of length r that may be made from a number of items n:

$$P(n,r) = \frac{n!}{(n-r)!} \tag{23}$$

setting $r = 2$ and simplifying yields:

$$P(n,2) = \frac{n!}{(n-2)!} \tag{24}$$

$$= \frac{n(n-1)(n-2)!}{(n-2)!} \tag{25}$$

$$= n(n-1) \tag{26}$$

If we note that the terms where $c = k$ are multiplied by 0 and thus do not contribute (equation 1) we can also substitute the $c \neq k$ branch of W:

$$D(m) = \sum_{c \in R} \sum_{k \in R} \delta(c,k) \frac{\nu(m,c)\nu(m,k)}{|m|(|m|-1)} \tag{27}$$

Thus the form of D_e has been recovered:

$$D_e = \frac{1}{|V|(|V|-1)} \sum_{c \in R} \sum_{k \in R} \nu(V,c)\nu(V,k)\delta(c,k) \tag{28}$$

note that my notation and Krippendorff's differ such that:

$$n = |V| \tag{29}$$

$$n_c = \nu(V,c) \tag{30}$$

$$n_k = \nu(V,k) \tag{31}$$

$$_{\text{metric}}\delta^2_{ck} = \delta(c,k) \tag{32}$$

To recover the form of D_o, begin by substituting, cancelling and reordering the sums:

$$\begin{aligned} D_o &= \sum_{u \in U} \frac{|u|}{|V|} D(u) \\ &= \sum_{u \in U} \frac{|u|}{|V|} \sum_{c \in R} \sum_{k \in R} \delta(c,k) \frac{\nu(u,c)\nu(u,k)}{|u|(|u|-1)} \\ &= \frac{1}{|V|} \sum_{c \in R} \sum_{k \in R} \delta(c,k) \sum_{u \in U} \frac{\nu(u,c)\nu(u,k)}{|u|-1} \end{aligned} \tag{33}$$

then noting that the quantity referred to as O_{ck} in Krippendorff's paper is given by:

$$O_{ck} = \sum_{u \in U} \frac{\nu(u,c)\nu(u,k)}{|u| - 1} \tag{34}$$

we are able to recover the original form:

$$D_o = \frac{1}{|V|} \sum_{c \in R} \sum_{k \in R} O_{ck} \tag{35}$$

Should you wish to look further at this and Krippendorff's paper there is one further piece of notational difference worth knowing:

$$m_u = |u| \tag{36}$$

References

[1] Andrew F Hayes and Klaus Krippendorff. Answering the call for a standard reliability measure for coding data. *Communication methods and measures*, 1(1):77–89, 2007.

[2] Klaus Krippendorff. Computing krippendorff's alpha reliability. *Departmental papers (ASC)*, page 43, 2007.