A Tale of Two Logratios: Rebuttal to Martin

This data set I provided in `11-A Tale of Two Logratios.pdf` and the researcher's objective was very simple. The researcher wanted to investigate the relationship between high alcohol consumption (spirits, let's call it "high") and low alcohol consumption (wine & beer, called "low"). I showed that there was actually no relationship at all, and I know there isn't because I constructed the data set! Any semblance of a relationship between "high" and "low" is either due to (i) the small random error I added to perturb the data to make it more realistic or (ii) the methodology used. In this case the use of an ILR indicates, as I showed, that there is something going on in the contrast between "high" and "low" - the ILR transformation has created an "effect" that doesn't exist. Hence, I find this a counter-example to the use of ILRs. In spite of this, Martin has managed to reply to a simple problem of one page with 7.5 pages of distances, correlations and other ILRs and I can see no direct contradiction of my assertion, only an affirmation (Point 5 below)!

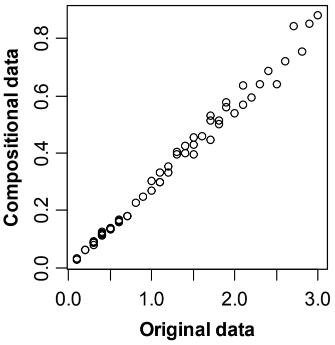


Exhibit 1: Plot of compositional data against original raw amounts, using my "Tale of Two Logratios" data set.

Point 1: Martin makes much about the difference between the raw and the relative values. Although I agree in general with the statement that "the absolute quantity matters" (see Greenacre 2017), I can assure everyone that there is almost no difference between raw and relative values in this example, which was specially constructed with this property to avoid the distinction Martin is making – so that cancels out almost 2 of the pages of Martin's reply. **Exhibit 1** shows the evidence, the plot of the relative (compositional) values vs. the raw absolute values. So there can be hardly any substance in Martin's saying (his page 3): "Note that the comparison using RELATIVE information is different from the ABSOLUTE case". Whatever "discovery" he is making must be very minor indeed and not worth highlighting.

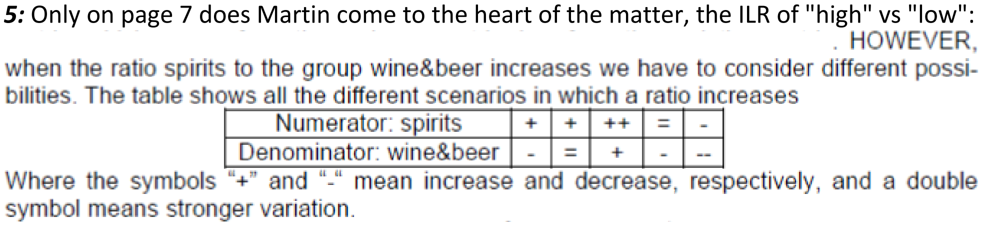
Point 2: I don't know why Martin uses Euclidean distances between the raw data, where the ranges are so different in the three columns. (range of wine is 4 times that of spirits!) This is not good practice. At least, data should be log-transformed, since these are clearly not interval-scale variables.

Point 3: My whole argument is about **logratios**, throughout all this debate, and the problem with ILRs when they involve geometric means. But Martin consistently gives examples of correlations involving log-transformed parts or log-transformed sums, e.g. " log(wine/beer) is not related to **log(spirits)** (r=0, p-value=0.98), and it is not related to **log(wine+beer)** (r=0.13; p-value=0.58), that is, THE RATIO wine/beer IS NOT RELATED TO THE TOTAL **wine+beer**." [Note: Wine+Beer is approximately constant (by construction, and quite realistically so as many people tend to alternate between beer and wine), so it can't be correlated with anything! So why is this obvious fact expressed in UPPER CASE, as if it is some major discovery?] All this correlation stuff involving parts is completely irrelevant, I am only interested in the behaviour of logratios. Also, I thought one was not allowed to use parts to compute correlations... Again, not good practice.

Point 4: Martin constructs two alternative "principal ILR balances", wine & spirits relative to beer and wine relative to spirits. But this is not what the researcher was interested in, her aim was to study the contrast between "high" alcohol and "low" alcohol drinks. Are mathematical niceties once again overruling the substantive objective of the researcher? As John Aitchison said (quoted from contribution #2 in this debate, by Vera and Juanjo, expressing exactly the point I have been trying to make):

> "*J. Aitchison (2003) stated that the ilr treatment, being mathematically sound, was generally not necessary or even useless. In certain cases, he proposed the use of amalgamations and the associated log-contrasts as a more intuitive and practical way of dealing with those problems.*"

Point 5: Only on page 7 does Martin come to the heart of the matter, the ILR of "high" vs "low":



. HOWEVER, when the ratio spirits to the group wine&beer increases we have to consider different possibilities. The table shows all the different scenarios in which a ratio increases

Numerator: spirits	+	+	++	=	-
Denominator: wine&beer	-	=	+	-	--

Where the symbols "+" and "-" mean increase and decrease, respectively, and a double symbol means stronger variation.

Note the following: "we have to consider different possibilities". **That's it!** We don't really know what the ILR by itself is actually measuring, it could be the five different scenarios given by Martin. This is exactly the reason that I asked Alecos at CODAWORK 2017 why the single ILR (in Antonella's paper, which he showed in his talk) was used instead of the amalgamation logratio, as conventionally used in Gibbs diagrams – we don't really know what's going on in the geometric mean used in that denominator unless we do a secondary analysis of all the logratios within that geometric mean (in our simple 3-part example, this is just the wine/beer logratio, but it becomes a very complex story if there are several parts involved – how many possibilities are there then?). And ditto for Jamie's single ILR in his paper, the same problem. This debate started with that question, and ends with it right here! Martin has hit the nail on the head!

Final point: I steadfastly maintain that my simple "A Tale of Two Logratios" example shows how misleading it can be to compute an ILR balance and how it can be misinterpreted, while an amalgamation balance is easy to understand since you know exactly what is in the numerator and what is in the denominator, being just an extension of a simple logratio, as well as being more intuitive, as Aitchison has expressed. My point is: when a researcher wants to combine parts for substantive reasons, the amalgamation of those parts should be computed and then considered in logratios with the other parts, or other amalgamations. The researcher's aim is paramount, not the pretty theoretical properties of an ILR balance: orthonormality, exact distance preservation, computing "principal balances" according to some theory rather than computing the ratios that are relevant to the study's objective, etc...

A request: I have asked for a similar counter-example of an amalgamation balance in favour of an ILR balance. There must be one, but it's not my job to find it, so can someone please make up one? Commenting on my counter-example Martin says "...one can easily create a similar example, where the nice plot is obtained for the geometric mean." So it sounds like it's easy to counter my amalgamation logratio argument. As far as I am concerned, this debate has run its course, it's taken a lot of my time, done only for the love of the subject and a desire to simplify compositional data analysts' lives (I am on leave and get no monthly salary these days...). I would now only be interested if anyone can produce as convincing a counter-example as mine in defence of an ILR balance in a practical context. And, please, keep it simple!

Reference:

Greenacre M (2017). "Size" and "shape" in the measurement of multivariate proximity. *Methods in Ecology and Evolution* 8: 1415-1424.

(If anyone wants a PDF, I can send it. There is also a video about this article, posted on the journal's website, or you find it easily on my YouTube channel **youtube.com/CARMEnetwork**)