Reply to "**A Tale of Two Logratios: Rebuttal to Martin**" by M. Greenacre

J.A. Martín-Fernández 7 March 2018

	wine	beer	spirits
[1]	1.2	1.8	0.6
[2]	1.8	1.4	0.3
[3]	2.8	0.3	0.6
[4]	2.6	0.5	0.5
[5]	2.5	0.7	0.7
[6]	3.0	0.1	0.3
[7]	1.1	1.9	0.3
[8]	1.9	1.2	0.3
[9]	1.4	1.5	0.4
[10]	1.7	1.3	0.2
[11]	1.3	1.7	0.3
[12]	2.4	0.8	0.3
[13]	2.1	1.1	0.5
[14]	1.0	2.1	0.2
[15]	2.0	1.1	0.6
[16]	1.6	1.5	0.4
[17]	2.7	0.4	0.1
[18]	2.3	0.9	0.4
[19]	1.5	1.7	0.6
[20]	2.2	1.0	0.5
[21]	2.9	0.1	0.4

Once again, thanks to Michael Greeancre for continuing the discussion about his example with the CoDa "consumption of (wine, beer, spirits)" *(documents: 11.A Tale of Two Logratios and corresponding replies*).

However, I feel sad and worried because, after 7.5 pages of explanations (*see my document 13.Reply to A tale_MG-Martin.pdf*), I failed to show that there is no problem with ILRs when they involve geometric means.

Unfortunately, the document with the rebuttal (*14.Reply to Martin-MGreenacre; 1.5 pages*) doesn't include any new important material or explanation. Only one exhibit where one can see that the absolute and relative values are approximately proportional, because the total of the samples is approximately constant. However, as I showed in my reply, this fact is irrelevant when one compares the absolute and the relative information expressed in two vectors. The important issue for our discussion is that the absolute and relative information are different, and that they are both affected by the amalgamation of parts.

The rebuttal mostly includes a repetition of the arguments of the initial document (*11.A Tale of Two Logratios*) structured in six points and a final request. I will answer here these arguments with the following "point-by-point" list. I will try to be simple and not produce 7.5 pages… ;-)

- **Point 1:** I already answered this point (*see above*) about the differences between absolute and relative information and the effect of the amalgamation. It is a well-known discussion, but it is worth to be aware of this issue, particularly in those data sets where the total is approximately constant and one can doubt that the effect is present.
- **Point 2:** I used Euclidean distance with raw data because it is the most popular option for people that are not using CoDa. I really appreciate that Michael suggests a "log-transformation". Following this idea, when one is dealing with CoDa, the log-contrasts (linear combinations of log-transformed variables) are a "natural" option and, importantly, ILRs are a particular case of log-contrast.
- **Point 3:** There is no problem with the ILRs when they involve geometric means. They are a linear combination of log-transformed variables. ALRs and CLRs are linear combinations of log-parts as well. In addition, is there any problem with PCs in standard multivariate analysis? They are a linear combination of variables. I used parts and correlations just to answer the pairwise scatterplot included in the initial document (*11.A Tale of Two Logratios*), where a relation between parts is used to justify the correlation between ILRs.
- **Point 4:** Yes, I constructed two ILR-bases. One (SBP1) using Principal Balances, and the second (SBP2), "wine/beer" and "spirits/(wine&beer)", to analyse the goal in the example. I don't believe that log-linear combinations of variables (ILRs) are "mathematical niceties overruling the objective". Are PCs in standard multivariate analysis non-useful "mathematical niceties" too? I don't think so. I believe that they are part of the exploratory analysis of the problem.
- **Point 5:** I really appreciate that the explanation of the different scenarios for the change in a logratio is considered the "*heart of the matter*". I really appreciate this because this explanation also applies to the amalgamated logratio, not only to the ILRs. The same issue that Michael highlights "*We don't really know what the ILR by itself is actually measuring*" perfectly applies to the proposed amalgamated logratio. Then, where is the truly advantage of the amalgamated logratio? In my opinion, there is not advantage at all. The amalgamation distorts the distances, the angles … but statistical analysis is based on distances (variance) and angles (correlation). An amalgamated logratio is not a log-linear combination of parts:

log(spirits/(wine+beer))= log(spirits)-log(wine+beer), and it is impossible to relate this with the initial variables.

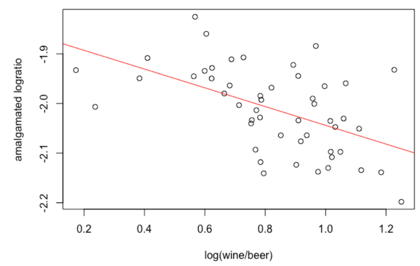
- **Final point:** this final point makes me think of a student asking: "why do we have to do PCA (multivariate analysis) if we can do simple pairwise scatterplots which are more intuitive?"… I have shown that amalgamated logratios have the same (even more!) difficulties of interpretation than ILRs. A final idea: which is the advantage to include the amalgamation of parts in the denominator with respect to the geometric mean?

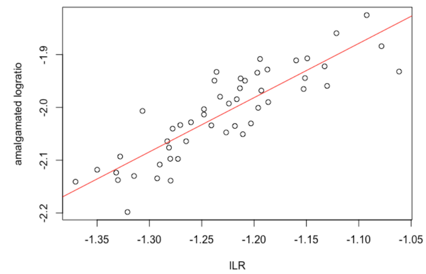
  log(spirits/(wine+beer)) = log(1/2) + log(spirits/((wine+beer)/2)),

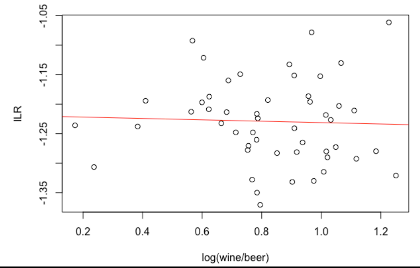
  the amalgamation is proportional to the arithmetical mean … is the arithmetical mean an appropriate operation with relative data? I don't think so. Well, when I visit a bank office asking for an "average rate" they really don't calculate the arithmetical mean … they calculate the geometrical mean… ;-)
- **The request:** I didn't prepare a different example because I argued that the Michael's example is perfect to show the dangers of the amalgamation. After my 7.5 pages of analysis I am convinced that the example is perfect and I will use it in my future lectures. However, because Michael asked again for a different example I include here a new example and the R code to reproduce it.

  This example shows the same situation as the example in the initial document (*11.A Tale of Two Logratios*), but now it is the amalgamation that has created an effect that does not exist. The first plot suggests that the amalgamated logratio (log(spirits/(wine+beer)) is related to log(wine/beer) (r = -0.5260762; p.value= 8.718905e-05). That is, following the Michael conclusion in the initial document: "*as the ratio of wine to beer increases, so the ratio of spirits to the amalgamation wine+beer, of lower alcohol drinks diminishes*". The second plot shows the relation between ILR and the amalgamated logratio (r= 0.8645423; p.value= 4.440892e-16). The third plot suggests that log(beer/wine) is not related with ILR (r= -0.03908718; p.value= 0.7875446). This result is as expected, because the simulation is based on two independent normal ILR: "wine/beer" and "spirits/(wine&beer)".

```
#####################################################################
### SIMULATION EXAMPLE
# means ilr1, ilr2 from M.Greenacre example
set.seed(1)
rilr1=rnorm(50,0.5653601,0.2)
rilr2=rnorm(50,-1.010777,0.06)
############# ilr data set
rilr=cbind(rilr1,rilr2)
cov(rilr)
cor(rilr) # independent ILRs
########### contrast matrix "wine/beer" and "spirits/(wine&beer)"
FI=rbind(c(sqrt(0.5),-sqrt(0.5),0),c(-sqrt(1/6),-sqrt(1/6),sqrt(2/3)))
########## clr coordinates
rclr=rilr%*%FI
######### raw data
rX=exp(rclr)/apply(exp(rclr),1,sum)
colnames(rX)<-c("wine",    "beer"  ,  "spirits")
head(rX,12)
######### spurious correlation produced by the amalgamation
plot(log(rX[,1]/rX[,2]),log(rX[,3]/(rX[,1]+rX[,2])),xlab="log(wine/beer)",
    ylab="amalgamated logratio")
abline(lm(log(rX[,3]/(rX[,1]+rX[,2]))~log(rX[,1]/rX[,2])), col="red")
cor(log(rX[,1]/rX[,2]),log(rX[,3]/(rX[,1]+rX[,2])))# -0.5260762
cor.test(log(rX[,1]/rX[,2]),log(rX[,3]/(rX[,1]+rX[,2])))$p.value # 8.718905e-05
#########  correlation between ilr and amalgamated logratio
plot(log(rX[,3]/sqrt(rX[,1]*rX[,2])),log(rX[,3]/(rX[,1]+rX[,2])),xlab="ILR",
    ylab="amalgamated logratio")
abline(lm(log(rX[,3]/(rX[,1]+rX[,2]))~log(rX[,3]/sqrt(rX[,1]*rX[,2]))), col="red")
cor(log(rX[,3]/sqrt(rX[,1]*rX[,2])),log(rX[,3]/(rX[,1]+rX[,2])))# 0.8645423
cor.test(log(rX[,3]/sqrt(rX[,1]*rX[,2])),log(rX[,3]/(rX[,1]+rX[,2])))$p.value # 4.440e-16
######### correlation between ilr-coordinates
plot(log(rX[,1]/rX[,2]),log(rX[,3]/sqrt(rX[,1]*rX[,2])),xlab="log(wine/beer)",
    ylab="ILR")
abline(lm(log(rX[,3]/sqrt(rX[,1]*rX[,2]))~log(rX[,1]/rX[,2])), col="red")
cor(log(rX[,1]/rX[,2]),log(rX[,3]/sqrt(rX[,1]*rX[,2])))# -0.03908718
cor.test(log(rX[,1]/rX[,2]),log(rX[,3]/sqrt(rX[,1]*rX[,2])))$p.value # 0.7875446
#####################################################################
```







I hope that this time I will convince Michael that the amalgamated logratio has the same or even more troubles than the ILR. That is, no advantage when using them. On the other hand, using ILR we have log-linear combinations of variables, as usual in multivariate analysis. I enjoyed a lot working with these data and I find they will be very useful for students in our CoDa-courses. Once more, thanks a lot!