On the Interpretability of Log-Ratios and Amalgamation

Justin Silverman

8/12/2017

Contents

Introduction

There have already been some very nice points brought up. As a result I have been slow to jump into this discussion. However, as I both promised Michael at CoDaWork that I would write up some of my comments on this issue and a recent prompt by Juanjo finally got me to take the plunge into this discussion.

Please excuse the informality of my writing. I have tried to air on the side of intuition rather than formality. However, if this comes at the cost of interpretability for some, please let me know and I will try to clarify.

I see the issues as two-fold:

1. A discussion on the appropriateness of amalgamation
2. A discussion of the interpretability and use of the ILR

To frame my biases: I have been a proponent of the ILR transform. My recent paper on a Phylogenetic ILR transform [1], is testament to this. That said, I very strongly believe that no tool is always good and I see some real downsides with the ILR transform and multiplicative amalgamations.

A comment on the utility of projections

Projections are one of the most useful tools when exploring datasets with more than 2 variables. Given a composition $(x_1, \ldots, x_D)$ simply plotting $log\frac{x_1}{x_2}$ should be thought of as a type of projection from the full D-dimensional dataset. A crucial component of having useful projections is that there should be a "meaningful" way in which looking at a projection informs our understanding of what the full D-dimensional dataset looks like. What is meaningful? That's really a difficult and subjective question so I will delay this discussion until it is needed in a minute.

[1] https://elifesciences.org/articles/21887

Comments on Amalgamation

Amalgamation of parts are often extremely useful practice if the amalgamation has meaning. That is if a composition is represented as $(x_1, \dots, x_D)$. A log-ratio of the form $\log \frac{x_1+x_2}{x_3}$ may be a very useful tool so long as $x_1 + x_2$ is a quantity that answers a question (again this relies on subjective assessments of "meaningful"). Just as an example, think about a composition that represents the probability of 3 mutually exclusive events A, B, and C given by the composition (x_A, x_B, x_C). If we want to ask about the probability of seeing events A or B it is simply $A + B$ (this has meaning).

However, amalgamation represents a non-trivial operation with respect to a log-ratio geometry. Bringing it back to projections, how does a D-dimensional dataset get projected into a new space given by the mapping $(x_1+x_2, x_3, \dots, x_D)$? In particular, the mapping $\phi : (\log \frac{x_1}{x_D}, \dots, \log \frac{x_{D-1}}{x_D}) \to (\log \frac{x_1+x_2}{x_D}, \log \frac{x_3}{x_D}, \dots, \log \frac{x_{D-1}}{x_D})$ is a non-trivial operation with respect to the Aitchison metric and is really not that easy to intuit.

So, skipping forward, my conclusion is that amalgamations are super useful if we are interested in the amalgamated quantity. However, if I am simply trying to explore a D-dimensional compositional dataset without much knowledge regarding higher order structure or specific amalgamations of interest, I probably would not start out by looking at an arbitrary nested series of amalgamated log-ratios. That is to say I would not recommend the following transformation of a 3 part composition $y_1 = \log \frac{x_1+x_2}{x_3}$ and $y_2 = \log \frac{x_2}{x_3}$ (it is simply too difficult for me to piece these projections together into a coherent picture of the D-dimensional dataset).

Now, why would anyone replace amalgamation with geometric means? For one thing, geometric means allow for simpler linear projections between log-ratios spaces. I am not arguing that geometric means are easy to interpret in terms of the original parts, just that geometric means make it easier to understand the shape of a D-dimensional dataset in terms of projections in a log-ratio space. Keep this idea in mind in the next section where I discuss the interpretation/utility of the ILR.

Comments on the ILR

So first the bad (because I need to establish credibility in the face of my apparent bias). **I still don't know how to interpret ILR balances.** In fact, I will take a step further and say that I really have no conception of what a geometric mean of a vector with more than 3 parts is (I would argue that you would need to be able to think about equivalent volumes in 4(+)-dimensional space for that). Beyond this, I still have no real idea what the $\sqrt{\frac{rs}{r+s}}$ means in the balance formula (I know how to derive it - e.g., Gram-Schmidt Orthogonalization, I even know why I want it there - e.g., variance stabilization and normalization of basis elements, *but this is very different than understanding what it means on an intuitive level*). **So the key question is why do I still think that the ILR is an important tool.**

Meter Stick Analogy

(Please bear with me through a sightly abstract argument). The power of units is that we can transport them and interpret them in new situations. I have an intuitive understanding of what 1 meter is. I can use that understanding to interpret what it means for a building to be 100 meters tall or for a cow to be 1.5 meters long even though in one case the meter stick may be oriented perpendicular to the ground and in the later case, parallel to the ground. When your units change as a function of another parameter (as in relativistic physics) humans get very confused and loose their ability to intuitively understand the meaning and instead end up just staring at meaningless numbers. I am going to argue that the ILR permits such a meaninful unit whereas the ALR is much more like relativistic physics, where the meaning of the units changes with another factor.

Ultimately I think the power of the ILR transform is that it is essentially an ALR transform (so you can use it for multivariate-normal modeling without dealing with singular covariances) **that permits**

a standard unit of interpretation [2]**.** To explain this comment I am going to first demonstrate how the ILR permits a meaningful unit while the ALR does not, then I am going to discuss what the unit is (even though my current understanding of its meaning is somewhat incomplete).

Limitations of the ALR

The following is just an excerpt from a few blog posts I did recently[3].

First off, I am going to write the ILR and ALR in a form that shows that the two are extremely similar. I am going to denote the following as "the ALR base d" (where $d \in 1, \ldots, D$):

$$alr_d(\mathbf{x}) = \mathbf{y} = \left(\ln \frac{x_1}{x_d}, \ldots, \ln \frac{x_D}{x_d}\right)$$

this is just the normal ALR with the d-th component taken as reference. Somewhat more revealingly, we can write this as

$$alr(\mathbf{x}) = \mathbf{y} = \ln(\mathbf{x}) \cdot B = \ln(\mathbf{x}) \cdot \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \cdots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

where the row of all -1s occurs in the d-th row and the matrix B is of dimension $D \times (D-1)$.

The matrix that I am calling B defines a set of vectors that we project our log-transformed data onto. That is, the columns of B can be used to form a coordinate system for us to interpret our data in. But not all coordinate systems are created equal, in general humans like to think in terms of Cartesian coordinates formed by an orthogonal basis (this is what makes intuitive sense to humans). Why? Well lets take "Ortho" and "-normal" separately. **Orthogonality** allows us to change the value of one variable without "requiring" a change of the other variables. For example, in an X-Y coordinate system you can change X without changing Y. This is not to say that X and Y cannot be correlated but any correlation we see is not going to be a triviality imposed by our coordinate system. **Normality** (or having basis vectors of unit length) implies that our units have the same intuitive meaning whether your units are lined up along the X or Y axis. Together orthonormality of a basis leads to a coordinate system that permits meaningful units that are invariant in the way I was alluding to earlier.

Its actually a simple linear algebra trick to turn the matrix B into an orthogonal basis. We can use the Gram-Schmidt orthogonalization procedure (effectively equivalent to the QR decomposition). We are going to rename B after we "orthonormalize" it, to V. Now the ILR transform base V is given by

$$ilr_V(\mathbf{x}) = \ln(\mathbf{x}) \cdot V$$

with inverse transform given by

$$ilr_V^{-1}(\mathbf{y}) = \mathcal{C}[exp(\mathbf{y} \cdot V^t)]$$

where $\mathcal{C}(\mathbf{a}) = \left(\frac{a_1}{\sum_i a_i}, \ldots, \frac{a_D}{\sum_i a_i}\right)$ is the closure operation.

Just as a point of order, note that the Inverse ALR is acctually not the same as the inverse ILR. With the inverse ILR we can "invert" the transform by simply reversing the operations and taking the transpose of the

[2]What I have been calling a unit of evidence-information since this term was coined in Egozcue and Pawlowsky-Glahn *Evidence information in Bayesian updating*, Proceedings of 4th International Workshop on Compositional Data (2011)

[3]In particular this one and this one. http://www.statsathome.com

matrix V. Note however, that for the ALR we acctually need to use a slightly different matrix than B. We need to use B^t_{inv} where B_{inv} is simply the matrix B but with the row of -1s replaced with 0s.

In this form the ILR really looks just like the ALR. We have just done some linear algebra on the matrix D to turn it into the orthonormal basis defined in V.

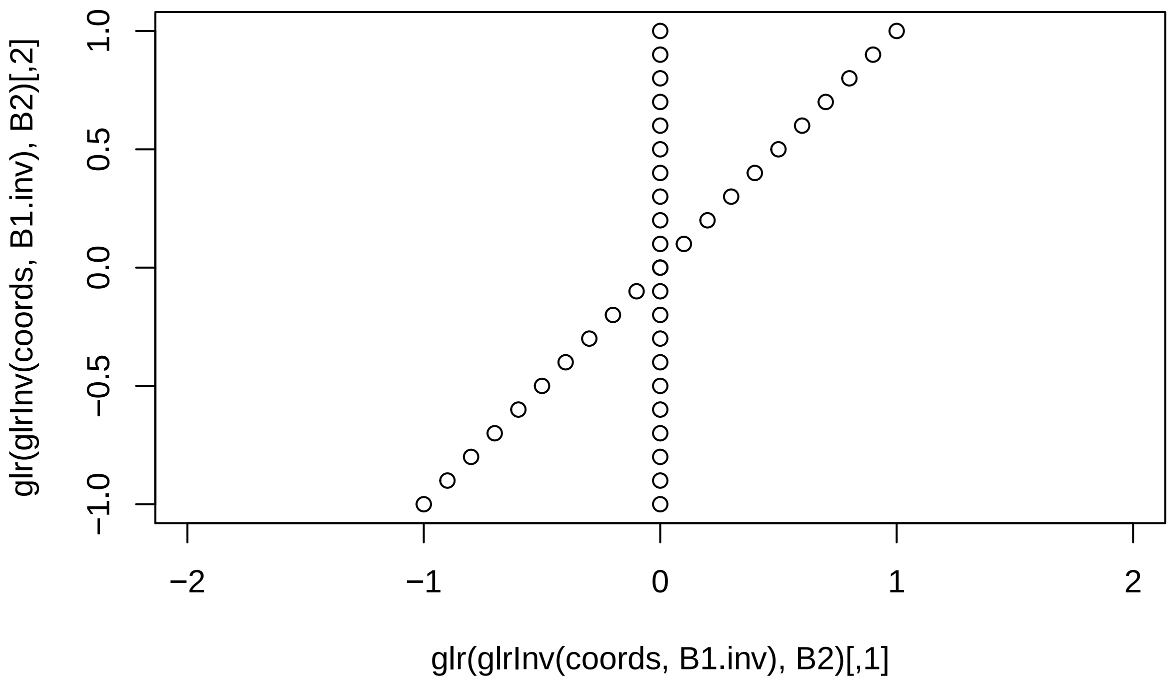
Now who cares? To answer this I am going to start by showing a shortcoming of the ALR. First I am going to define a few functions for the ILR, and ALR transforms.

```
# generic functions
miniclo <- function(x){
  if (is.vector(x)) x <- matrix(x, nrow = 1)
  (x/rowSums(x))
}

vec_to_mat <- function(x){
  if (is.vector(x)) return(matrix(x, nrow = 1))
  x
}

# BASE CODA
# generic log-ratio transform
glr <- function(x, V){
  log(x) %*% V
}

# generic log-ratio transform - Inverse
glrInv <- function(y, V){
  tmp <- exp(y %*% t(V))
  miniclo(tmp)
}

# ALR
create_alr_base <- function(D,d, inv=FALSE){
  if (d < 1 | d > D) stop("invalid d given D")
  B <- diag(D)
  if (!inv){
    B[d,] <- rep(-1, D)
  } else {
    B[d,] <- rep(0, D)
  }
  B[,-d]
}
```

Next I am going to create two ALR transforms and show how the coordinate system of the first gets represented in the second. We are going to see that the representation is "weird" in the sense that the coordinates do not map intuitively into each other.

```
B1 <- create_alr_base(3, 1)
B1.inv <- create_alr_base(3, 1, inv=T) # need to create an "inverse" version of this base
B2 <- create_alr_base(3, 3)

# Create Data for "Coordinates"
# Equally spaced pooints along coordinates
# of first ALR transform.
coords <- as.matrix(rbind(cbind(0, seq(-1, 1, by=0.1)),
```

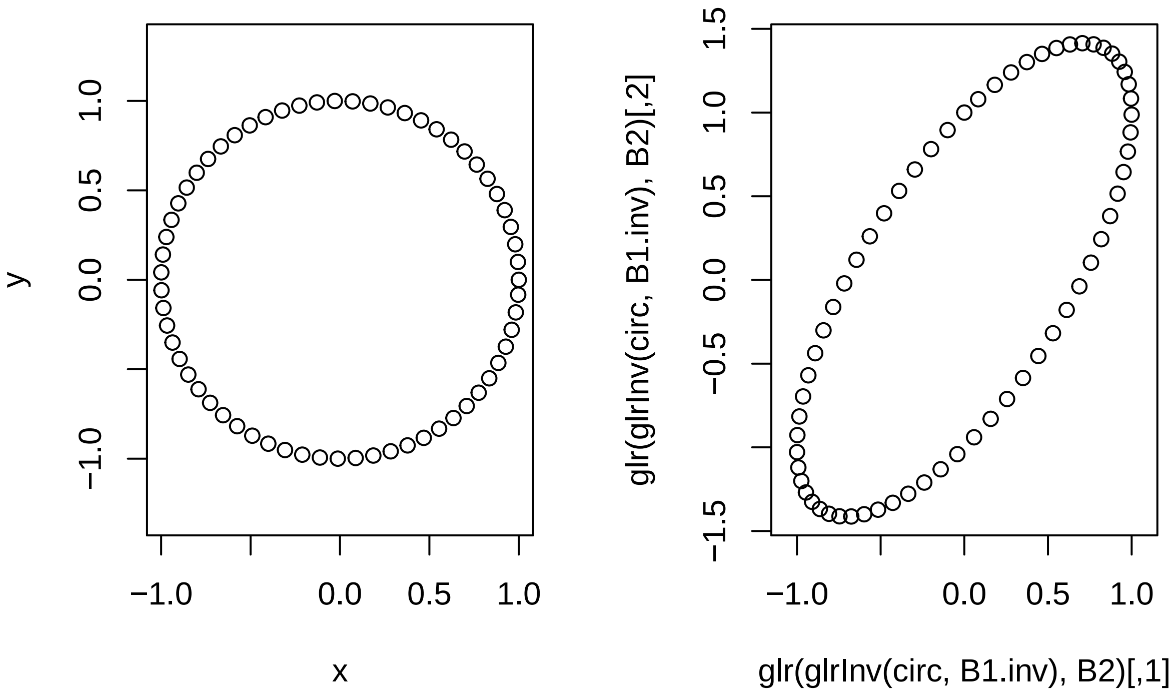
```
                          cbind(seq(-1, 1, by=0.1), 0)))

plot(glr(glrInv(coords, B1.inv), B2), asp=1)
```



The key point to notice here is that the coordinates of the first ALR transform intersect at an oblique angle (not orthogonal) in the second ALR transform's coordinates. As a result the euclidean distance measured between two points in one ALR basis is not the same as the Euclidean distance measured between two points in the other ALR basis. This highlights that any "units" we defined in the first ALR transform would not make sense in the second ALR transform. This is exactly what the CoDa Literature refers to when the ALR transform is referred to as "not preserving metric concepts".

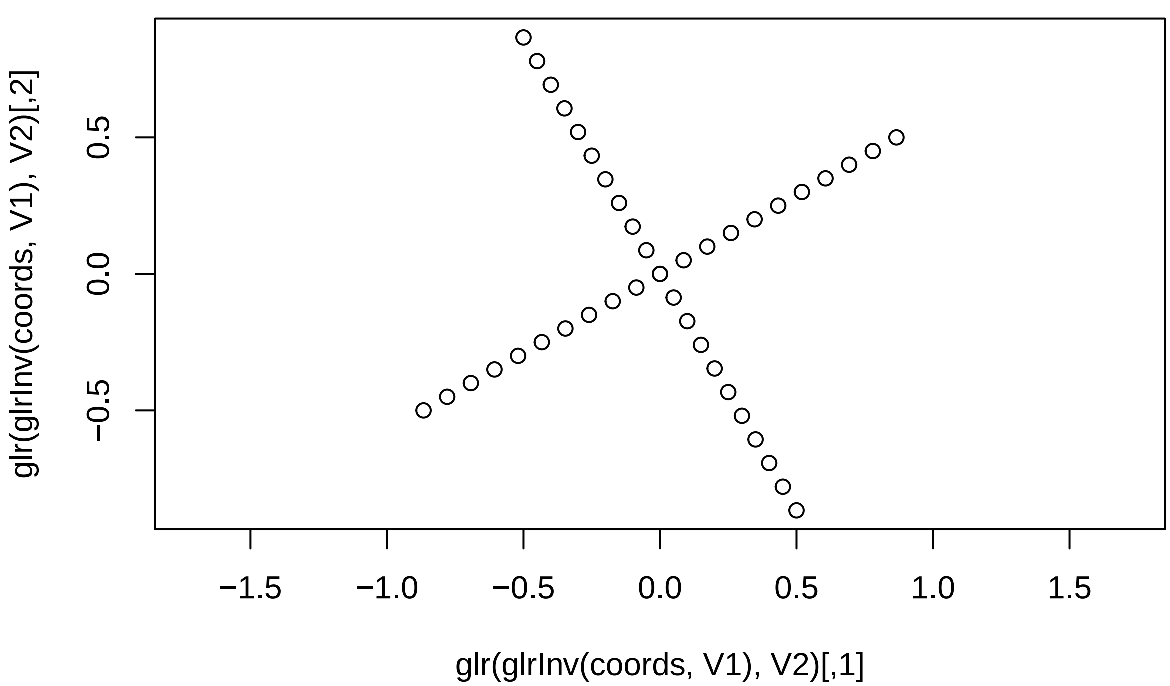
Just to show another example, look at how a circle defined in the first coordinate system warps in the second.

```
t <- seq(0, 2*pi, by = 0.1)
x <- cos(t)
y <- sin(t)
circ <- cbind(x, y)

par(mfrow=c(1,2))
plot(circ, asp=1) # Plot the circle in one ALR coordinate system
plot(glr(glrInv(circ, B1.inv),B2), asp=1) # Plot in a transormed ALR coordinate system.
```

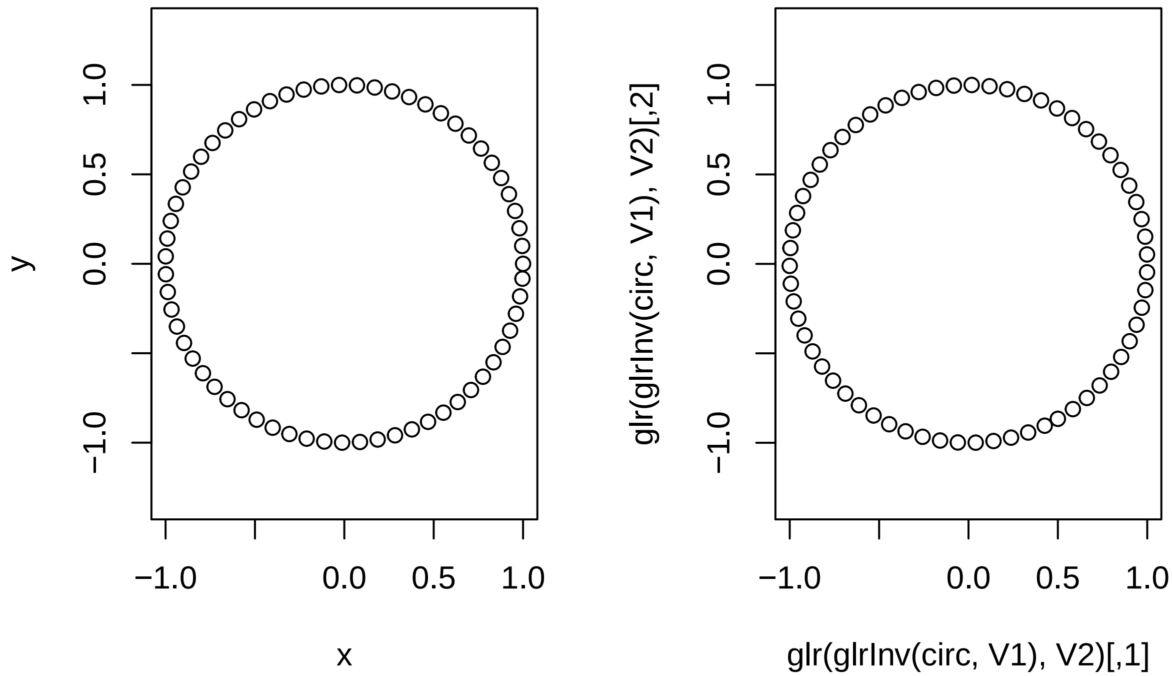


Now what about if we made an orthonormal basis V. In order to make two ILR transforms I am going to use the QR decomposition on each of the two ALR base matrices B that we just defined.

```
V1 <- qr.Q(qr(B1)) # The "Contrast Matrix" of the orthogonal basis
V2 <- qr.Q(qr(B2))

plot(glr(glrInv(coords, V1), V2), asp=1)
```




```
par(mfrow=c(1,2))
plot(circ, asp=1) # Plot the circle in one ALR coordinate system
plot(glr(glrInv(circ, V1),V2), asp=1) # Plot in a transormed ALR
```



Thus we see that our orthonormal basis (ILR coordinates) does not have the same problems as the ALR coordinates. Essentially with the ILR we can make use of the concept of intuitive units. All this is also inline with my earlier comments on the utility of projections. Different ILR coordinates represent a simple rotation (a special type of projection) that maintains the meaning of our units.

Evidence-Information / meaning of units

So far I have tried to give some results that show how an orthonormal basis provides a means of defining an intuitive unit. I have implicitly defined this "unit" to be a reference distance with respect to the Aitchison metric. However, I have yet to describe what that unit is or even more generally, what is the intuitive meaning of distance with respect to the Aitchison metric at all. Unfortunately I think this is a real outstanding problem in CoDa (or perhaps I just don't know the correct references).

Distance with respect to the Aitchison metric (or I will shorten to "Aitchison distance") is a really weird measurement that seems to have some similarities to units of Shannon-information, but also some crucial differences (for example it is scale-invariant and symmetric with respect to any CLR/ILR coordinate). Alternatively, it can be thought of as a measurement of "relative fold changes" between two compositions. Either way, I believe an intuitive description or understanding is somewhat missing. That said, I do believe such an intuition is possible and even if not, an intuitive understanding may not be required (Shannon information has seen widespread use and success despite not really being intuitive in the way a meter is).

What does a given balance mean?

As I already alluded to, the short answer is that I don't know yet. Now I probably have to explain this answer more. I know what a balance is on an intellectual level. For example, Alex Washburne gave a particularly nice description of it in terms of z-scores and variance stabilization. Alternatively, I have the "hand-wavy" description of it in terms of the "balance" (as in relative weight) of two groups of parts on either side of a binary partition. But in all the work I have done with the ILR I am still unable to simply look at a compositional dataset and say "I would eyeball that balance as having a value of being about 3" in the same way I can do that with the length of a cow (e.g., without a ruler I can estimate the length of a cow as being about 1.5 meters by eye).

Over the past few years of working with the ILR I have started to develop an intuitive understanding for what a large or small distance is (again, with respect to units or projections) and I think this is a very good sign regarding the utility of the ILR as a space where we can define intuitive units. Unfortunately, balances seem to be harder (in part because their scale depends with respect to the parts I may be interested in changes as a function of the number of parts on each side of the partition - think of the $\sqrt{\frac{rs}{r+s}}$ factor).

Now I know some researchers would say that my problem is that I am searching for an intuitive understanding of balance in terms of the parts rather than simply embracing the transformed nature of the data (i.e., just thinking of the data as being balances measured on a strange compositional scale). I like this mentality on some days (when I really just want a Real-space embedding of a dataset in which I am searching for patterns). On other days when I am trying to learn biology of microbes, this is really not helpful (remember, microbes can exist as a single unit and not as a composition; I could literally count all the microbes in a small agar plate under a microscope).

So what do I use the ILR for?

I find the ILR an indispensable tool for creating and manipulating models. By modeling in the ILR I get the benefits of the CLR and the ALR all rolled into one. I can literally treat the ILR transformed data just like it was in real-space with the classical multivariate normal central limit theorem (Lebesgue measure, Euclidean metric, etc…). Especially in a Bayesian setting it is incredibly useful to have my posterior samples in a form that I can do all kinds of calculations on without worrying about whether I need to transform so as to preserve the metric or transform so that my covariances are non-singular.

Beyond this, I find that looking at a high dimensional dataset in the ILR (even if you ignore the meaning of the coordinates) is incredibly useful. Because of the way projections behave with respect to the ILR it is incredibly easy for me to get an intuitive sense of what the data is doing ("its shape") by looking at the lower-dimensional projections in an ILR basis. To bring it back to units again, my ability to even gain such a "coordinate-independent" view of the data/results like this is arguably dependent on my ability to have a meaningful unit. *As a counter example, try envisioning a sphere hovering in the middle of a room, as you walk around the sphere (looking at it from different angles) the sphere changes shape so that it is no longer "spherical". In such a situation it is nearly impossible to gain much of an understanding of what that "sphere" actually is. I would argue this is precisely the situation I presented with the circles above.*

After I have gotten a feel for the data (or after I have done my modeling and have posterior samples or some other geometric result) I can also start looking at other lower dimensional projections. Often these lower dimensional projections are just simple pairwise log-ratios, sometimes I am interested in amalgamations of the parts, sometimes I am even interested in the raw proportions (blasphemous as this may be).

To Summarize my arguments in this section:

1. The ILR is a useful mathematical tool that makes model building easy.
2. The ILR is a useful tool for getting a sense of the data because of the way projections work with respect to an orthonormal basis.
3. I frequently use not only the ILR but many different representations of compositional data depending on the question I am trying to answer.

Finally, a requisite comment on CLR

I have been fairly quiet about the utility of the CLR. Between the ALR and the CLR it is possible to piece together much of the functionality of the ILR. That said there are few things I want to mention.

1. In my opinion the CLR provides a mediocre tool for understanding high-dimensional structure through projections. Unfortunately the CLR just has 1 too many dimensions and I find it is often no better than just looking at the log-transformed proportions as a result. While we may frequently look at a

2-dimensional image on a paper in our 3-dimensional world without problem, I find this sort of breaks down in higher dimensions and I have found I can run into problems with the CLR.

2. Although it probably doesn't have to be mentioned, the fact that modeling on CLR transformed data requires some type of singular covariance structure makes it extremely difficult to model with (appropriately) in practice. At least I have yet to learn how to do it well.

3. Related to the last point. The CLR really isn't much better than proportions, you still have a linear constraint that can cause problems in interpretation.

4. The CLR also has geometric means in its definition, but rather than just a geometric mean of a subset of the parts, its a geometric mean of all the parts! Arguably harder as it requires that you are capable of thinking of equivalent volumes of a D-dimensional hyper-rectangle, rather than a possibly lower-dimensional equivlanet volume problem.