



Jamie Morton <jamietmorton@gmail.com>

Re: Coda discussion

Jamie Morton <jamietmorton@gmail.com>

Fri, Aug 18, 2017 at 9:43 AM

I'll add in my two cents in the discussion.

I think that ALR and ILR both have their practical use cases. We have seen that ALR performs quite well with sample classification in conventional multinomial regression models, and in terms of the graphical models that Michael presented at CoDaWork, ALR does seem to provide nice intuitive explanation (I don't completely grasp the graph partitioning idea that Alex presented, but would definitely be curious to see the details fleshed out).

However, I think there are definitely some questions / scenarios that ALR cannot cope with. Namely, trying to tweak out differentially abundant features between the samples. This is something that I try to highlight in my [paper](#). Identifying exactly which species is changing is a really hard problem -- looking at proportions alone, it is impossible to determine if a species is growing / declining / or even changing between the samples.

There have been a number of approaches using ALR-like techniques to try infer this with additional assumptions ([Sparcc](#) assumes a sparse correlation, and [ANCOM](#) assumes few species are changing). However, the crux of the problem lies in the fact that these approaches in the end are ill-define -- you are trying to solve an over-parameterized system of equations (this is acknowledged in the SparCC paper and briefly discussed in my [blogpost](#)). And this is not tractable without additional assumptions.

The ILR approach is quite different. Rather than focusing on which exact species are changing, we can relax the problem and instead focus on which partitions are changing. While this is a slightly different question, this is actually a tractable problem.

That being said, there are a ton of really cool things that you can do with the ILR transform. Because the ILR is really just an extended ALR with multiple parts in the numerator and the denominator of the log-ratio, you can represent this transform as a classifier. You can try identify and classify the parts that drive the separation between the groups of interest (this is basically what Javier's talk on optimal balances at CoDaWork was about). And because it forms an orthonormal basis, you can directly map the ILR transform to standard dimensionality reduction techniques (such as PCA, Orthogonal PLS, ...). And what makes this really cool is that at the end of the day, you can basically convert a dimensionality reduction problem in a sorting problem using the ILR transform. Given a clustering of your parts, you can essentially sort / group the parts followed by the ILR transform to obtain simple latent variables to summarize high-dimensional trends. This is what I was aiming for with the hierarchical clustering approach so that we can explicitly test for microbial partitioning due to pH.

In terms of future work -- I think we are just getting started realizing the implications of these techniques. And I'm really excited about the new method advances that have yet to come in the near future.

Best,
Jamie

[Quoted text hidden]