CASPEN Exit Report

Peter Taylor hosted by Benjamin Wandelt at the the Flatiron Institute

August 14, 2018

I visited the Flatiron Institute from the 28th of July through August 10th, after applying to visit in early July. The intention of my visit was to start developing the tools to perform a likelihood-free cosmic shear analysis, as it is known that the normal Gaussian likelihood assumption leads to bias. This problem will only get worse as the statistical power of experiments improve.

Producing a sufficiently large number of accurate forward models of the data, is the primary hurdle in likelihood-free inference. Before my visit we decided to forward model the data using log-normal simulations which offer a good compromise between speed and accuracy. At the beginning of my visit I spoke with Ben Wandelt, Justin Alsing and Stephen Feeney who are all members of the Flatiron Institute. We discussed our plans for producing the forward simulations, integrating these with the likelihood-free inference code developed at Flatiron called DELFI, and the ultimate goal of the project.

At this meeting Justin identified an issue with my original plan. I had intended to simply Poisson sample the shear field, to produce a catalog, but in order to accurately predict our error bars, we need to sample the shear field by mimicking the distribution of observed galaxies on the sky. We discussed possible solutions and we have now overcome this complication.

Over the next two weeks I implemented and refined my model with help from Ben. By the end of my visit, my simulations accurately reproduced the theoretical expectation for the two-point correlation function (see Figure 1).

My simulations are at an intermediate level of sophistication producing shear catalogs assuming a biased galaxy tracer model with consistent shear and density fields including a mask and shape noise. I still need to include redshift tomographic binning, photometric redshift error and intrinsic alignments to mimic a real survey. We also need to compress the catalogs to maps so that our method is computationally feasible for the full Euclid problem.

We have also begun to discuss integrating my simulations into the DELFI pipeline. Justin has provided me a few demo notebooks, and in principle it shouldn't be difficult to integrate my simulations with DELFI. Although the log-normal runs are fast, it still takes approximately a minute for a single point in cosmological parameter space on s single core for a Euclid-like simulation. The DELFI algorithm is very parallelizable so I have asked Stephen about MPI extensions. He has coded up an MPI implementation, so computational constraints should not be a problem going forward.

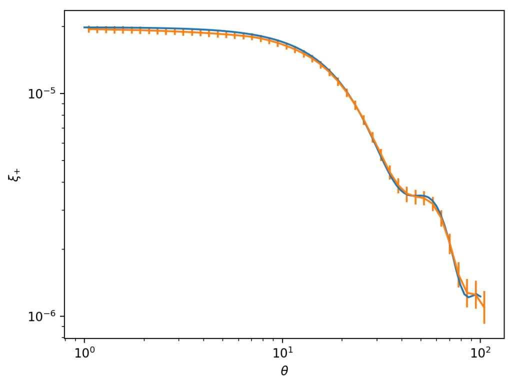

Figure 1: **Orange:** Shear correlation function computed from the galaxy catalogs in 100 realizations of my lognormal simulations. **Blue:** Theoretical expectation.

We now have the clearly defined goal of performing a likelihood-free analysis treating a single mock simulation as the data to: 1) demonstrate the feasibility of the likelihood-free approach 2) estimate the number of simulations needed for a real analysis and 3) quantify the extent of the bias in the standard Gaussian-likelihood approach. As an added bonus, since we are using a biased galaxy tracer model to populate the field with galaxies, it will only be a small extension to extend out analysis to a cosmic shear cross galaxy clustering analysis.

This visit was a fantastic opportunity which really helped us kick start this project. The progress I made over the two weeks would not have been possible without the trip. This project should result in a publication in the next 6 months.